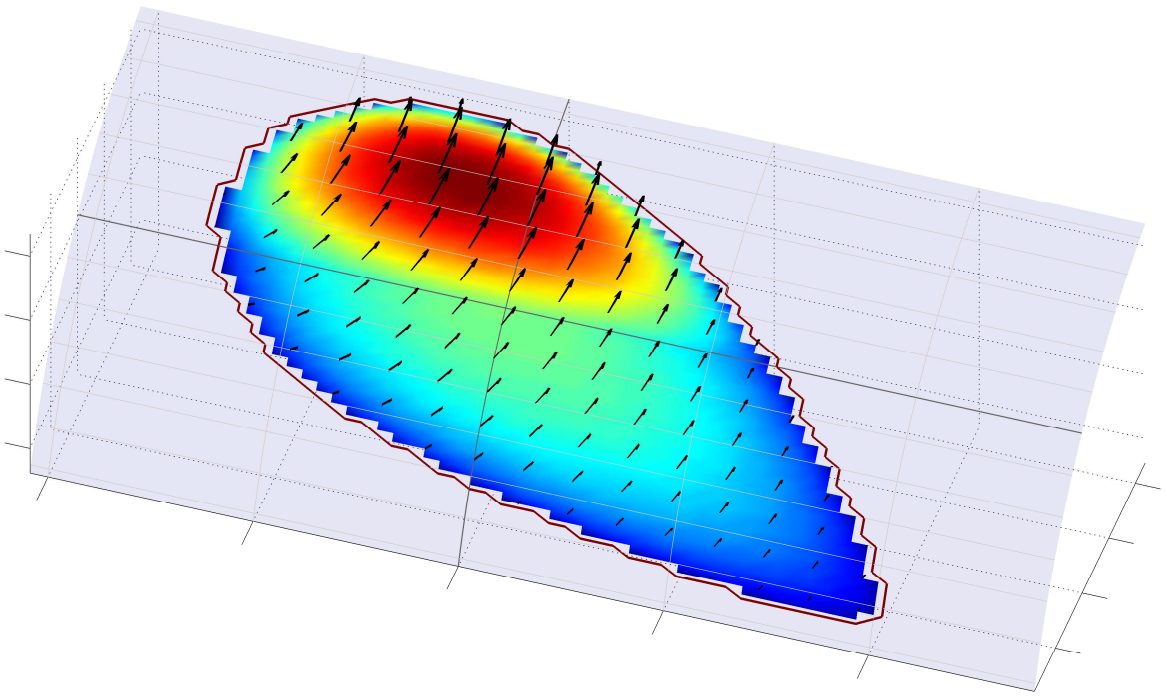
WHEEL–RAIL INTERACTION AND CONFORMAL CONTACT ANALYSIS

Ph.D. THESIS
presented to obtain the degree of
Doctor

Julio Blanco Lorenzo



THESIS ADVISORS

Dr. Javier Santamaría Manrique

Dr. Ernesto García Vadillo

September 2023, Bilbao

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

INGENIARITZA MEKANIKOA SAILA

DEPARTAMENTO DE INGENIERÍA MECÁNICA

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea
INGENIARITZA MEKANIKOA SAILA
DEPARTAMENTO DE INGENIERÍA MECÁNICA

Ph.D. Thesis

WHEEL–RAIL INTERACTION AND CONFORMAL CONTACT ANALYSIS

Presented by:

JULIO BLANCO LORENZO

in the

DEPARTMENT OF MECHANICAL ENGINEERING
BILBAO SCHOOL OF ENGINEERING
UNIVERSITY OF THE BASQUE COUNTRY (UPV/EHU)

to obtain the degree of:

DOCTOR

Thesis Advisors:

Dr. JAVIER SANTAMARÍA MANRIQUE

Dr. ERNESTO GARCÍA VADILLO

Bilbao, September 2023

AGRADECIMIENTOS

Quiero dar las gracias en primer lugar a los directores de esta tesis, Javier y Ernesto, por animarme a embarcarme en esta aventura, por proponerme el tema principal de esta investigación, por su guía, por su consejo y por su apoyo. Muchas gracias por esta oportunidad, por vuestra atención y por la confianza depositada en mí.

Agradezco la financiación recibida de Horizonte 2020 a través de los proyectos IN2TRACK-730841, IN2TRACK2-826255 e IN2TRACK3-101012456 dentro de la iniciativa Shift2Rail, así como al Ministerio de Ciencia e Innovación (MCI), al Ministerio de Economía y Competitividad (MINECO), a la Agencia Estatal de Investigación (AEI) y al Fondo Europeo de Desarrollo Regional (FEDER) por su financiación a través de los proyectos de investigación T 83/2006, TRA 2010-18386, TRA2014-59599-R y PID2019-109483RB-I00, y al Gobierno Vasco por las ayudas para grupos de investigación IT453-10, IT691-13, IT919-16, IT1764-22 y KK-2023/00029. Agradezco también la financiación recibida de la UPV/EHU a través de la Unidad de Formación e Investigación UFI-11/29.

Agradezco también a mis compañeros del grupo de investigación de Ferrocarriles del Departamento de Ingeniería Mecánica de la Escuela de Ingeniería de Bilbao, Nekane, Olatz, Rakel, Jesús y Josu, por las charlas, las meriendas y los buenos momentos compartidos, y por su ayuda.

I am very grateful to the people who became interested in my work, and with whom I have collaborated, developing part of the work presented in this thesis. Thank you very much to Dr.ir. Edwin A.H. Vollebregt from Vtech CMCC at Rotterdam, The Netherlands, with whom I delved into the application of the exact contact theory to conformal contact, and with whom I held fruitful discussions on conformal contact analysis. Many thanks to Prof. Paul A. Meehan and co-workers from the School of Mechanical and Mining Engineering of the University of Queensland in Brisbane, Australia, who proposed me to apply the models for conformal contact analysis developed in this work in the rolling bearing application. These collaborations have contributed to improving and expanding this work.

Y mi más sentido agradecimiento es para mi familia, y en especial para mis padres, Aurelio y Julia, que siempre han estado ahí. No solo me disteis la vida y un hogar, me habéis dado una gran educación y todas las oportunidades a vuestro alcance. Esto no habría sido posible sin vosotros.

ABSTRACT

The development of this thesis is focused on the study of wheel–rail conformal contact in linear elastostatics. To this end, comprehensive numerical models have been developed. These include a version of Kalker's exact contact theory extended for conformal contact, and detailed Finite Element (FE) models.

One of the major difficulties of the extension of the exact contact theory for conformal contact is the departure from the half-space assumption. To address this, the compliance of surfaces of elastic solids has been studied. For non-planar surfaces, the possibility of characterization by means of numerical analysis with FE models has been explored, and analytical approximations have been set out. New analytical results have been developed for the half-space as well.

With the developed contact models, a detailed investigation of wheel–rail 3D frictional conformal contact has been carried out. This covers both the normal and tangential parts of static (compression, shift) and rolling contact problems. In contrast, most of the past published literature on conformal contact has dealt with either 2D contact, or with frictionless contact. As a result of the investigation carried out, particular contact mechanics features resulting from conformity are described, differences between conformal and non-conformal contacts are discussed, and the adequacy of the hypothesis of non-conformity is assessed in situations with different degrees of conformity.

The developed detailed contact models have been further demonstrated in the rolling bearing application, contrasting with the much simpler contact mechanics models commonly used in this application. This has needed specific adaptations in the developed contact models. With these models, a study has been carried out of the roller–raceway rolling contact in a spherical roller bearing, only feasible by means of 3D partial slip contact models as the ones used in this work.

An additional part of the work reported in this thesis deals with the study of vertical vehicle–track dynamics, in which the wheel–rail contact is of great importance, by means of integrated multibody–FE models including a representation of the track dynamics, developed with commercial multibody and FE software packages. Models with different degree of complexity have

been constructed, linear as well as non-linear, and their adequacy has been assessed in different situations. With these models, a comparative study has been conducted of the dynamic performance of a ballasted track and three types of slab tracks.

RESUMEN

Este resumen está dirigido a los lectores hispanohablantes, siendo más extenso que el *Abstract* (resumen en inglés).

El contacto rueda-carril es uno de los elementos más diferenciadores de los vehículos ferroviarios, y el que determina en mayor medida su comportamiento dinámico. La elevada rigidez del contacto de acero sobre acero confiere al sistema rueda-carril una reducida resistencia a la rodadura y una alta capacidad portante. Por otra parte, esta elevada rigidez conduce a áreas de contacto pequeñas, del orden del centímetro, con niveles tensionales elevados. Esto implica una elevada exigencia en las condiciones de servicio de la interfaz rueda-carril, lo cual lleva a su vez a la necesidad de profundizar en su estudio y comprensión para mejorar sus capacidades y prestaciones. El análisis computacional es una parte importante de este proceso, máxime considerando la práctica imposibilidad de realizar mediciones reales detalladas de mecánica de contacto en condiciones realistas de operación de la interfaz rueda-carril.

La simulación precisa del contacto rueda-carril entraña una dificultad elevada, tanto desde el punto de vista de la determinación de los parámetros relevantes que influyen en el comportamiento físico de la interfaz, como de la resolución matemática del problema, que se presenta como fuertemente no lineal. Con el fin de simplificar la solución del problema de contacto rueda-carril, se adoptan hipótesis simplificativas. Una de las hipótesis habitualmente adoptadas es la de no conformidad en el contacto, que implica planitud del área de contacto y la asimilación del comportamiento local de los sólidos en contacto al del semiespacio infinito.

Si bien la hipótesis de no conformidad es a menudo adecuada, se presentan situaciones no poco habituales en condiciones realistas de operación en las que se dan condiciones de contacto conforme, comprometiendo la precisión de los resultados obtenidos mediante un análisis no conforme. Tal es el caso del contacto de la raíz de la pestaña de la rueda con la esquina de la cabeza del carril durante la inscripción en curvas cerradas, o el contacto entre perfiles de rueda y carril desgastados. Además, en estas situaciones frecuentemente se dan condiciones de contacto severas, con elevados niveles de tensión y deslizamiento, y se requieren soluciones de contacto precisas para evaluar

adecuadamente el daño resultante en la interfaz rueda-carril. Esto excede el alcance de los modelos de contacto rueda-carril utilizados habitualmente en simulaciones dinámicas, que priman la eficiencia computacional. Por otra parte, también se dan situaciones de contacto conforme en otras aplicaciones de gran interés tecnológico, como en rodamientos, en componentes de máquinas rotativas, y en uniones con bulones. Es por lo tanto de interés disponer de modelos que permitan hacer análisis de contacto conforme.

El desarrollo de esta tesis se centra en el estudio del contacto rueda-carril conforme en el marco de la elasticidad lineal y con un comportamiento cuasiestático del material en las cercanías del contacto. Para tal fin, se han desarrollado modelos numéricos detallados, que incluyen una versión de la teoría exacta de contacto de Kalker extendida para contacto conforme, y modelos de Elementos Finitos (EF) locales.

Una de las principales dificultades de la extensión de la teoría exacta a contacto conforme radica en que los sólidos en contacto dejan de verse como semiespacios infinitos. Esto representa un inconveniente importante, ya que las soluciones del semiespacio infinito sometido a cargas en su superficie están disponibles analíticamente, al contrario de lo que ocurre con sólidos con superficies no planas generales como los que se pueden tener en contacto conforme. Para tratar esta dificultad, se ha estudiado la flexibilidad de superficies de sólidos elásticos. Para superficies no planas, se ha explorado la posibilidad de caracterización mediante análisis numérico con modelos de EF, y se han planteado aproximaciones analíticas. Se han obtenido también nuevos resultados analíticos para el semiespacio infinito.

Con los modelos de contacto desarrollados, se ha llevado a cabo una investigación detallada del contacto rueda-carril conforme 3D y con fricción. Esto abarca las partes normal y tangencial de problemas de contacto estáticos (compresión, desplazamiento tangencial) y de rodadura. Por el contrario, la mayor parte de la literatura anterior sobre contacto conforme ha abordado contactos 2D, o contactos sin fricción. Como resultado de esta investigación, se describen características de mecánica de contacto particulares que resultan de la conformidad, se tratan las diferencias entre contactos conformes y no conformes, y se evalúa la adecuación de análisis no conformes en casos con distintos grados de conformidad.

Se ha mostrado la aplicabilidad de los modelos de contacto desarrollados también para rodamientos, contrastando con los modelos mucho más sencillos comúnmente empleados en esta aplicación. Esto ha requerido de adaptaciones específicas en los modelos de contacto desarrollados. Con estos modelos, se ha realizado un estudio del contacto en rodadura rodamiento-pista en un rodamiento esférico, solo factible con modelos 3D de deslizamiento parcial como los usados en este trabajo.

Las condiciones de contacto entre rodamientos y pistas determinan en gran medida los límites operacionales, prestaciones y vida de los rodamientos. La mecánica de contacto es por lo tanto una disciplina esencial en su diseño y análisis. Por su parte, las cuestiones relacionadas con la fricción en el contacto son importantes en diferentes aspectos del diseño y operación de los rodamientos, incluso con contactos adecuadamente lubricados, en los que las tensiones tangenciales transmitidas en el contacto son mucho menores que las presiones normales. Por ejemplo, la disipación friccional que se da en los contactos rodamiento-pista puede ser una fuente importante de calentamiento, y llevar a limitar la velocidad de operación para evitar un calentamiento excesivo del rodamiento. También es esencial para determinar el par de fricción y la evolución del desgaste de las superficies de rodadura del rodamiento.

Una parte adicional del trabajo reportado en esta tesis ha versado sobre el estudio de la interacción dinámica vertical vehículo-vía, en la cual el contacto rueda-carril es de gran importancia. La interacción dinámica vehículo-vía juega un papel fundamental en muchos problemas relacionados con la vía y con el material rodante, como el fallo prematuro por fatiga de elementos de estos sistemas, la generación de ruido y la transmisión de vibraciones al terreno. En esta tesis se presentan modelos integrados multicuerpo-EF en el dominio del tiempo, que incluyen una representación de la dinámica de la vía, desarrollados con paquetes de software comerciales para análisis de sistemas multicuerpo y de EF. Se han construido modelos de diferente grado de complejidad, tanto lineales como no lineales, y se ha evaluado su adecuación en distintas situaciones. Con estos modelos, se ha llevado a cabo un estudio comparativo del comportamiento dinámico de una vía de balasto y de tres tipos de vías en placa ante la circulación a alta velocidad de un vehículo de pasajeros.

La tesis se estructura como sigue:

- El Capítulo 1, Introducción, presenta los antecedentes y la motivación del trabajo, introduciendo el sistema rueda-carril, y particularmente el contacto rueda-carril, como posiblemente el elemento más distintivo e influyente en la interacción vehículo-vía, y entrañando también una complejidad considerable. Se habla también del contacto conforme, de la interacción vehículo-vía, y del análisis de contacto friccional en rodamientos. Siendo el desarrollo de modelos de mecánica de contacto rueda-carril avanzados el tema principal de esta tesis, se dedica un apartado de este capítulo a la formulación del problema de contacto. Estrechamente relacionado con esto, se trata asimismo el comportamiento mecánico del sólido elástico al nivel local del contacto. En los últimos apartados del capítulo, se listan los objetivos de la tesis, y se expone su estructura.
- En el Capítulo 2, Estado del arte, se proporciona una revisión del estado del arte de los temas principales estudiados en esta tesis, enfocada en el desarrollo de modelos numéricos detallados con la capacidad para tratar el contacto conforme. Esto incluye modelos de contacto rueda-carril, funciones de influencia del sólido elástico, análisis numérico del contacto en rodadura, contacto conforme, y análisis de mecánica de contacto en rodamientos.
- En el Capítulo 3, Interacción dinámica vehículo-vía en diferentes vías, se desarrollan modelos dinámicos para el estudio de la interacción vertical vehículo-vía con programas comerciales de análisis de sistemas multicuerpo y de EF. Se trata de modelos en el dominio del tiempo, con la capacidad para considerar características detalladas y no lineales tanto del vehículo como de la vía, como por ejemplo, traviesas con defecto de apoyo, y la interacción entre distintos ejes. Los modelos desarrollados se usan para llevar a cabo un estudio comparativo del comportamiento dinámico de distintas vías.
- El Capítulo 4, Análisis numérico del contacto conforme rueda-carril, está dedicado al desarrollo de modelos numéricos de mecánica de contacto para el análisis de contacto conforme. Estos incluyen una versión de la teoría exacta del contacto que se ha llamado CECT

(*Conformal Exact Contact Theory*; Teoría Exacta del Contacto Conforme en inglés) y modelos de EF para análisis de mecánica de contacto detallados. Se dan descripciones detalladas de las extensiones para contacto conforme implementadas en CECT, y de los ajustes necesarios en los modelos de EF para unos resultados fiables. Este desarrollo está dirigido inicialmente al caso rueda-carril, pero en el Capítulo 7 se muestra la aplicación al caso de los rodamientos.

- El Capítulo 5, Coeficientes de influencia del sólido elástico, trata del cálculo de los coeficientes de influencia para sólidos con superficies no planas, siendo esta una de las mayores dificultades asociadas a la aplicación de la teoría exacta del contacto al análisis del contacto conforme en lugar de concentrado. Los coeficientes de influencia caracterizan el comportamiento mecánico de los sólidos al nivel local del contacto, relacionando las tensiones transmitidas en el contacto con la respuesta elástica del sólido. El capítulo incluye nuevos resultados analíticos para el semiespacio infinito.
- El Capítulo 6, Características del contacto conforme rueda-carril, presenta un estudio detallado del contacto conforme rueda-carril, abarcando las partes normal y tangencial del problema de contacto. El estudio se basa en análisis numéricos realizados con los modelos y coeficientes de influencia desarrollados en los Capítulos 4 y 5. Se evalúan las características que resultan de la conformidad y las diferencias con respecto al contacto no conforme. Se realiza también un análisis geométrico de la distancia indeformada entre dos superficies no planas de sólidos de revolución.
- El Capítulo 7, Análisis de contacto en rodamientos, muestra la aplicación de los modelos numéricos de contacto desarrollados al análisis del contacto en rodadura en rodamientos. Esto contrasta con los modelos de mecánica de contacto mucho más simples comúnmente usados en esta aplicación. Con estos modelos, se lleva a cabo un estudio del contacto en rodadura rodamiento-pista en un rodamiento esférico.
- Por último, en el Capítulo 8, Conclusión, se compilan los resultados de la investigación llevada a cabo, se esbozan perspectivas de líneas

futuras de investigación, y se listan las publicaciones que han resultado de este trabajo.

ACRONYMS

1D, 2D, 3D: one dimensional, two dimensional, three dimensional.

ALE: arbitrary Lagrangian-Eulerian.

AP: analysis point.

BEM: Boundary Element Method.

CECT: Conformal Exact Contact Theory.

CEM: Contact Element Method.

CWD: current working directory.

DFT: Discrete Fourier Transform.

DOF: degree of freedom.

EMB, RMB: elastic multibody, rigid multibody.

EST: elastic slip tolerance.

FE, FEM: Finite Element, Finite Element Method, Finite Element Model.

FFT: Fast Fourier Transform.

H+C: ‘Hertz + Carter’.

IC, IF: influence coefficient, influence function.

MBS: multibody system.

N-R: Newton-Raphson.

PCS: potential contact surface.

RCF: rolling contact fatigue.

SAM: semi-analytical method.

SDEC: simple double-elliptical contact.

SRB : spherical roller bearing.

CONTENTS

Chapter 1

Introduction

Chapter summary

This chapter introduces the wheel–rail system, and particularly wheel–rail contact, as possibly the most distinctive and influential element in the railway vehicle–track interaction, and a one involving considerable complexity as well. Attention is drawn to issues related to frictional contact mechanics, playing an important role in wheel–rail interaction, and involving higher characterization and modelling complexity than frictionless contact. Conformal contact is also addressed, as a phenomenon occurring in the wheel–rail application and falling beyond the scope of the most usual contact models. Problems relevant to vehicle–track interaction are reviewed, and the importance of computer simulation as a way for their understanding and control is emphasized. The importance of frictional contact analysis in the rolling bearing application is also discussed. Being the development of advanced wheel–rail contact mechanics models the main topic of this thesis, a section of the chapter is devoted to the mathematical statement of the contact problem. Closely related to this, the mechanical behaviour of the contacting solid at the local contact level is also treated. In the last sections of the chapter, the objectives of this thesis are listed, and the organization of the thesis is set out.

1.1 Background and motivation

1.1.1 The wheel–rail system. Wheel–rail contact

The railway is a land transportation system in which the vehicle is guided on rails. The wheel–rail system (Figure 1.1) connects the vehicle with the track, providing support, guidance and traction to the railway vehicle. The steel wheel on steel rail system is predominant over others such as rubber-tired wheels or the Maglev [Wikipedia Maglev www]. The guiding by the rail and the circulation in trains of vehicles are the two distinctive features of the railway vehicle.

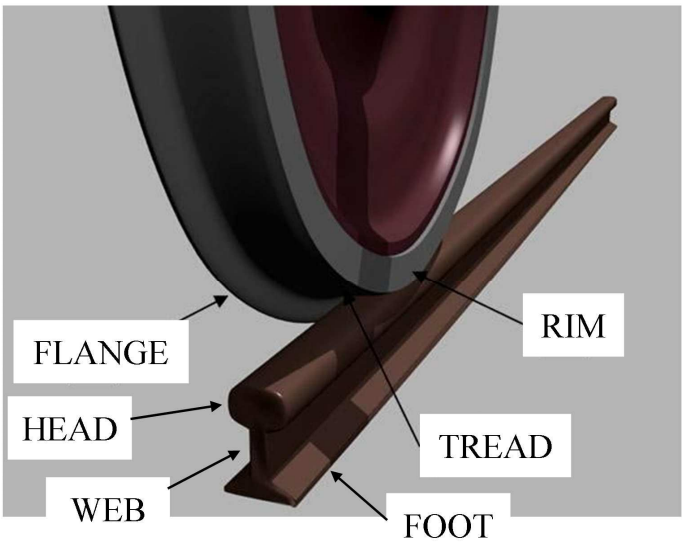


Figure 1.1. The wheel–rail system. Adapted from https://commons.wikimedia.org/wiki/File:Roue_ferro.jpg. List of authors in https://fr.wikipedia.org/w/index.php?title=Boudin%20(ferroviaire)&action=history. CC BY SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0/deed.en).

The high steel-on-steel wheel–rail contact stiffness provides the system a low rolling resistance and a high load bearing capacity, contributing to the high energy efficiency of railway transport. On the other hand, the high contact stiffness entails small contact patches and high contact stress levels. This implies severe operating conditions of the wheel–rail interface. The increasing vehicle speeds, axle loads and traffic densities, together with the demand for

reduced costs and environmental impact, and increased reliability and comfort, lead to more stringent quality and performance levels demanded to the wheel–rail system. All this brings the need for investigation and development, together with observation and learning from operating experience, to be able to match up to these expectations. Computer simulation is part of this process.

The wheel–rail interaction is the feature that determines to a greatest extent the dynamic behaviour of the railway vehicle. Its modelling is complex, due to the geometric non-linearities, as well as to the constitutive non-linearities of the contact itself. Contact problems are particularly highly non-linear and require significant computer resources to solve, as stated in the ANSYS Contact Technology Guide [ANSYS doc 2010]. Thus, the precise solution of the wheel–rail contact involves a high difficulty. This is so not only from the standpoint of the mathematical solution of the problem, but also from the determination of the relevant parameters that influence the physical behaviour of the interface. At the same time, the practical impossibility to carry out detailed contact mechanics field measurements in realistic operating conditions of the wheel–rail interface reinforce the importance of wheel–rail contact modelling.

To simplify the solution of the wheel–rail contact problem, simplifying assumptions are adopted. The wheel–rail contact problem usually conforms to the following hypotheses:

- The contacting solids are homogeneous and isotropic.
- The solids are massive, i.e., not hollow, at least around the contact.
- The first-order theory of elasticity holds.
- The contact is quasistatic. This implies that the effect of the inertia forces of the material of the solids near the contact is small in comparison with that of the stresses transmitted in the contact. That is, elastostatic contact problems are considered, in which the material of the solids near the contact has a quasistatic response under the action of the contact stresses, which are variable in time from the (Lagrangian) point of view of a material particle that traverses the contact even in steady-state contacts.

- The distance between the contacting surfaces is small compared to the contact patch dimensions. This follows from the assumption of linear elasticity, with small strains.
- The contact is concentrated. In this type of contact, the contact patch or region $\boldsymbol{C}$ has much smaller dimensions than the curvature radii of the surfaces of the solids around the contact, and much smaller as well than the dimensions of the solids. With these conditions, the contact area is approximately flat. This also implies that the local mechanical behaviour of the contacting solids at the contact level can be assimilated to that of the half-space.

Under these assumptions, the structural (global) problem of each solid is decoupled from the local contact problem. In other words, the general stress state of the solid does not produce appreciable changes in the local geometry of the contacting surfaces, and the overall geometry and supporting conditions of the solids do not influence their local mechanical behaviour around the contact.

In the following subsections, some concepts of wheel–rail rolling contact are introduced, that have a relevant role in the wheel–rail interaction. The focus is put in issues related to contact mechanics, and issues related to the wheel–rail geometry are not covered. Geometry related issues include the contact geometrical search (cf. §4.2.3.1), and the characterization of the wheel–rail and wheelset–track contact pairs by means of parameters (e.g. the equivalent conicity) that describe important properties of the railway wheelset as the basic constitutive unit of the railway vehicle, such as its guidance ability, comprising the self-centering and curve inscription abilities.

In the absence of plastic action, it is considered here that the term ‘rolling contact’ implies frictional interaction. The concepts reviewed next are related to this interaction.

1.1.1.1 Creepages

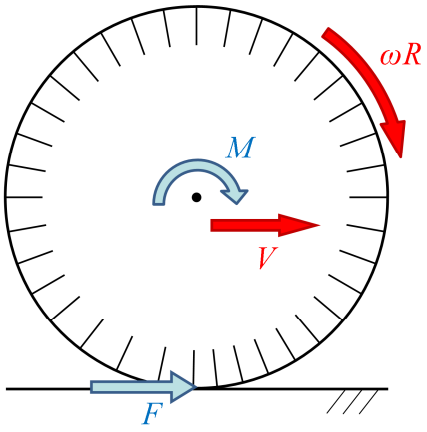
The creepages are defined as the relative rigid velocities between two bodies in rolling contact (known also as rigid slip velocities, cf. §1.2) divided by a reference speed, which is normally taken as the forward speed. The creep phenomenon is among the most essential concepts in frictional contact

mechanics, and particularly in rolling contact mechanics. Its first description is attributed to Reynolds [Reynolds 1875], who set out the coexistence of adhesion and slip areas in the contact patch of two bodies in rolling contact, as determined by the interrelationship between the friction forces and tangential strains.

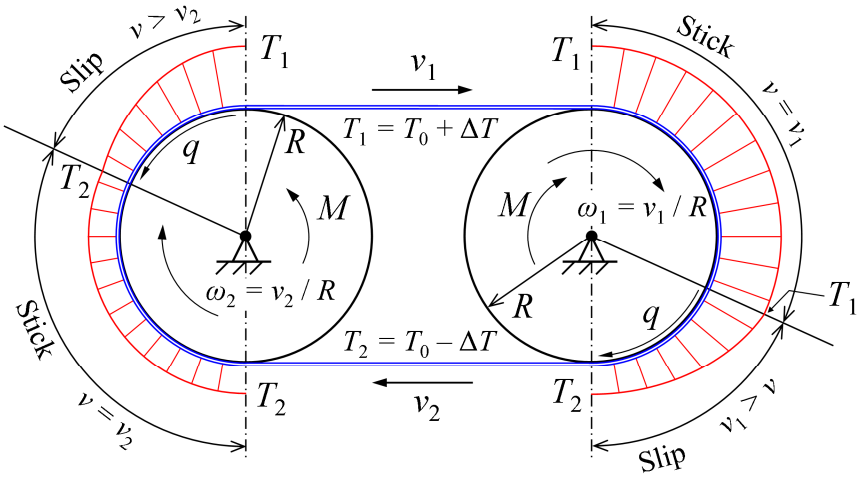
The creep phenomenon is illustrated here with the two examples shown in Figure 1.2. The first example, shown in Figure 1.2a, is a rolling wheel, in which a tractive moment is applied, and which is rolling without gross slip. As shown in the figure, the forward velocity of the wheel V is lower than its rigid solid velocity calculated as the wheel rotation velocity ω times the wheel radius R. The difference between V and ωR is generated due to the deformation of the contacting bodies in the rolling direction.

The radial lines drawn around the wheel in Figure 1.2a give an indication of the deformation of the wheel material in its circumferential direction, with closer lines indicating compressive deformation, and more separated lines tensile deformation. In the shown example, the ground may be considered to be rigid for simplicity.

The distance advanced by the wheel is determined by the deformed dimension in rolling direction of the portion of its perimeter in the adhesion part of the contact. In the case of the rolling wheel under traction of the example, the strain of the material about to enter in contact is compressive, and thus must remain in the adhesion ("locked") part of the contact. So, the distance advanced in one revolution by the wheel of the example subject to traction is lower than its undeformed perimeter. The effect is similar to reducing the rolling radius of the wheel. In the case of a braking wheel, the deformations change sign, and the distance advanced by the wheel in one revolution is larger than its undeformed perimeter. As this example shows, rigid velocity differences are compensated or absorbed by deformations at the contact level, without gross slip. In this situation, adhesion and slip areas coexist in the contact patch, as happens in the next example, and as shown in §2.1.2 for solids in rolling contact. In this situation, the contact is said to be in partial slip.



(a) Rolling wheel.



(b) Flexible transmission belt. Adapted from Fig. 8.1 of [Johnson 1987].

Figure 1.2. Examples of motion with creep.

Figure 1.2b shows a 1D example of motion with creep, corresponding to a flexible transmission belt. This example is shown in §8.1 of [Johnson 1987]. Here it is explained in a slightly different way. The belt connects two pulleys with the same radius R, the driving pulley being the one on the right, and the driven pulley the one on the left. The figure shows the moment M applied on each pulley, the frictional forces q transmitted from each pulley to the belt, the rotation velocities ω_1 and ω_2 of the driving and driven pulleys, respectively, and

the belt tensions. The tension on the tight side is T_1 and the tension on the slack side T_2. The belt tensions around the contact with each pulley are depicted by the red lines perpendicular to the contact surfaces. The figure also depicts the stick and slip arcs in each pulley. It is assumed that M is sufficiently small to not generate complete slip in the pulleys. Assuming the same coefficient of friction in the contact of the belt with both pulleys, the extension of the stick and slip arcs are the same in both pulleys.

It is to be demonstrated that the disposition of the stick and slip arcs in the pulleys is necessarily as shown in Figure 1.2b, i.e., with just one stick part in the side where the belt runs onto the pulley (which may be called the leading edge side of each belt–pulley contact) and just one slip part in the side where the belt runs out of the pulley (which may be called the trailing edge side). For this purpose, the relationship of the belt tension T is considered with the friction forces q transmitted between the belt and each pulley on the one hand, and with the belt linear velocity v on the other, as explained next:

- Regarding the belt as a massless funicular solid, the forces acting on a differential portion of belt in frictional contact with a pulley are shown in the free-body diagram of Figure 1.3. The belt tensions on its left and right sections are T and $T + \mathrm{d}T$, respectively. These are almost collinear in a differential portion of belt. The normal and tangential force transmitted by the pulley to the belt differential are $\mathrm{d}N$ and $\mathrm{d}q$, respectively. The equilibrium of forces in the longitudinal belt direction yields $\mathrm{d}T = -\mathrm{d}q$, according to positive directions shown for each force in the figure. I.e., the belt tension decreases in the direction of motion if the frictional forces push the belt forward, increases if the frictional forces drag the belt backwards, and is not changed if no frictional force is transmitted to it.
- Defining the belt extensibility λ as ε / T, being ε the longitudinal strain of the belt, the belt linear density ρ_l may be written as $\rho_l = \rho_{l,o} / (1 + \lambda T)$, being $\rho_{l,o}$ the linear density of the unstretched belt. Considering the mass balance of the belt in a control volume around one of the pulleys, Eq. (1.1) may be written:

$$\frac{\rho_{l,o}}{1+\lambda T_1} v_1 = \frac{\rho_{l,o}}{1+\lambda T_2} v_2 \rightarrow \frac{v_1}{v_2} = \frac{1+\lambda T_1}{1+\lambda T_2} \tag{1.1}$$

In this equation, v_1 and v_2 are the velocities of the belt in the tight and slack sides, respectively. So, it is seen that the belt velocity is higher in the tight side, to comply with mass equilibrium. A similar conclusion may be reached by direct application of Eq. (1.7).

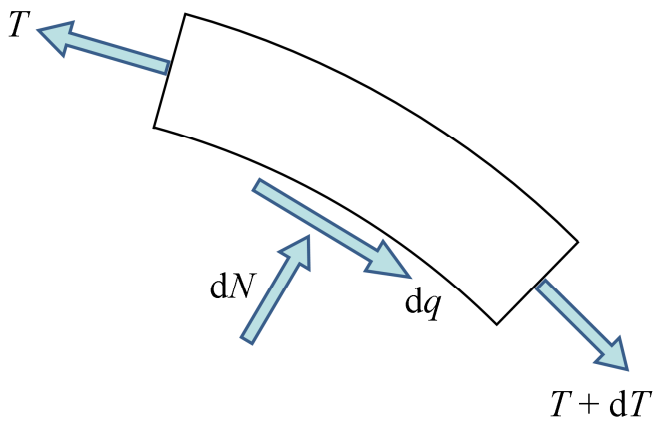


Figure 1.3. Equilibrium of a differential portion of belt in frictional contact with a pulley.

The evolution of the belt tension around the driving pulley is considered next, in the direction of motion, from the tight to the slack side. The rigid velocity of the driving pulley is v_1. The belt velocity when running into the driving pulley may be either larger, smaller, or equal to v_1. The first two possibilities are discarded by the following considerations:

- If the belt velocity were larger than v_1 at the leading edge, the belt would have a forward slip with respect to the pulley, and would receive a dragging frictional force from the pulley, as the frictional force must be opposing the relative slip (cf. §1.2). As a result of this frictional force, the belt tension would increase in the direction of motion, and consequently also the belt velocity. Therefore, there would be slip in the whole belt–pulley contact. Additionally, the frictional force transmitted by the belt to the pulley would be in the direction of motion in the

whole contact, so the moment M applied to the pulley to equilibrate it would have to be dragging, instead of driving. So, this is not possible for the driving pulley in steady state.

- If the belt velocity were smaller than v_1 at the leading edge, the belt would have a backwards slip with respect to the pulley, and would receive a forward frictional force from the pulley. This would make both the belt tension and velocity decrease in the direction of motion. Therefore, in this case again there would be slip in the whole belt–pulley contact. In this case, the frictional force transmitted by the belt to the pulley would be opposing the movement, and the moment M on the pulley would be driving. This configuration is possible only after M exceeds a certain threshold. But, as it has been initially assumed that M is sufficiently small to not generate complete slip in the pulleys, this must be discarded as well.

Therefore, it is concluded that the belt and pulley velocities must coincide at the leading edge, and therefore the adhesion arc must be located next to the leading edge. From the previous discussion, it may be seen as well that once there is a difference between the belt and pulley velocities, the belt velocity will keep diverging from the pulley velocity up to the trailing edge. In other words, once slip starts, slip will be maintained in the same direction up to the trailing edge.

Applying similar considerations to the driven pulley, it is concluded that the arrangement of the stick and slip arcs is similar to the driving pulley, as shown in Figure 1.2b. This is another illustrative example of how the difference of velocities of two rigid bodies may be compensated with elastic deformation, without gross slip. The arrangement of stick and slip contact areas in rolling contact of elastically similar solids is the same as the one of this 1D example. On the other hand, the tangential stress in the adhesion part is zero in this example, unlike in the rolling contact of solids.

Returning to the rolling wheel case, three creepages are defined, namely the longitudinal ξ, lateral η, and spin φ creepages. The first two describe translational relative movements in the longitudinal and lateral directions, and the spin rotational movement normal to the contact plane. In the wheel–rail case, the creepages may be approximated according to Eqs. (1.2)–(1.4).

$$\xi = \frac{w_{x,o}}{V} \approx \frac{V - \omega\, R_{\text{roll}} \pm \dot{\psi}\, l}{V} \tag{1.2}$$

$$\eta = \frac{w_{y,o}}{V} \approx \frac{V\psi + \dot{Y}}{V\cos(\delta)} \tag{1.3}$$

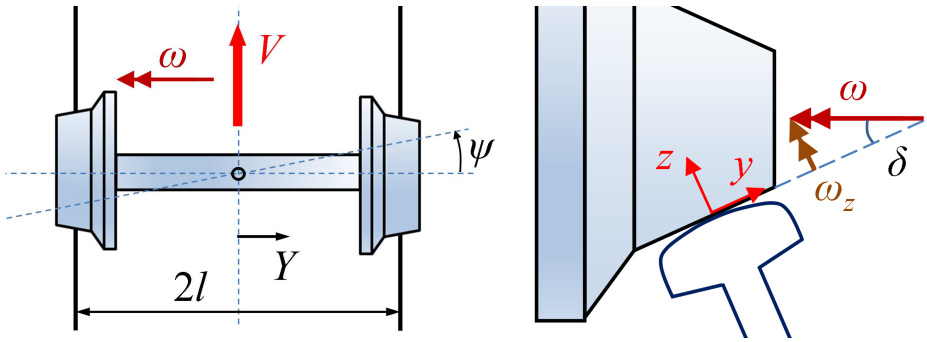
$$\varphi = \frac{\omega_z}{V} \approx \pm\frac{\sin(\delta)}{R_{\text{roll}}} \tag{1.4}$$

In Eqs. (1.2)–(1.4), some magnitudes and coordinates that are used in §1.2 are introduced. They are defined briefly below, and some of them are shown in Figure 1.4, depicting the rolling motion of a railway wheelset along a straight track. The ± signs in Eqs. (1.2) and (1.4) refer to the right and left wheels, respectively, as seen in Figure 1.4a.

- $w_{x,o}$ and $w_{y,o}$: components in the longitudinal and lateral directions of the contact plane of the rigid velocity difference of the wheel with respect to the rail, in the central contact point. Deformations of the bodies are not included in this rigid velocity.
- V: forward velocity of the wheelset.
- ω: magnitude of the rotation velocity of the wheel. ω_z is the component of ω perpendicular to the contact plane.
- R_{roll}: rolling radius of the wheel. It is the distance from the revolution axis of the wheel to the contact point (see Figure 2.4).
- l: half the wheel base. It is the distance between the contact point of each wheel of the wheelset with its respective rail.
- δ: contact angle. It is the angle between the rotation axis of the wheel and the contact plane.
- Y: lateral position of the wheelset in the track.
- ψ: yaw angle of the wheelset. It is a small angle.

In situations of partial slip, ξ, η, and φc (being c a characteristic dimension of the contact patch) are of the same order as the longitudinal strains of the

contacting surfaces. The surface strains are in turn much smaller than unity within linear elasticity.



(a) Plan view. (b) Front view. Detail of a wheel.

Figure 1.4. Imperfect rolling of a railway wheelset.

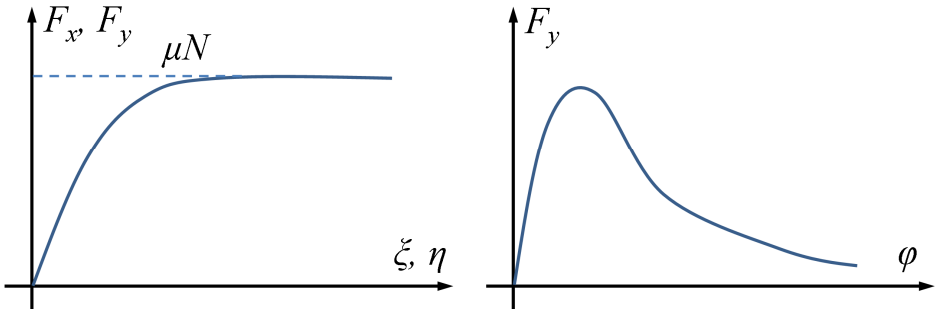
1.1.1.2 Tangential contact forces

When rolling is not perfect (i.e., in the presence of non-zero rigid velocity differences in the contact), tangential stresses are transmitted by friction in the wheel–rail contact, leading to resultant tangential forces. Thus, these forces are a direct consequence of creepages.

While the tangential wheel–rail contact forces are smaller in magnitude than the normal ones, they influence to a large extent the dynamic behaviour of the railway vehicle, its guidance and stability properties. They are dominant terms in the equations of motion of the railway wheelset, and are the primary responsible for wheelset steering [Wickens 2006]. As stated in [Wickens 2006], without consideration of these forces, very little can be deduced about the dynamical behaviour of railway vehicles apart from Redtenbacher's formula for the inscription of a coned wheelset on a curve [Redtenbacher 1855] and Klingel's formula for the kinematic oscillation of the wheelset [Klingel 1883]. The tangential contact stresses also play an important role in the degradation phenomena in the wheel–rail interface.

An objective of solving the wheel–rail rolling contact problem is to get the relationships between the creepages and the tangential contact forces. These relationships commonly resemble the creepage–creep force curves shown in Figure 1.5. Figure 1.5a shows the resultant longitudinal force F_x as a function of

the longitudinal creepage, in the absence of other creepages. The lateral creepage–creep force curve would be similar. Figure 1.5b shows the resultant lateral force F_y as a function of the spin creepage.



(a) Longitudinal or lateral creepage–creep force curve.

(b) Spin creepage–lateral force curve.

Figure 1.5. Creepage–creep force curves.

As may be seen both in Figure 1.5 a and b, the creepage–creep force curves may be approximated as linear for small creepages. For larger creepages, the curves are no longer linear, and are saturated. In the curve of Figure 1.5a, Coulomb's law is considered with a constant coefficient of friction. In this case, the force maintains a constant value once it becomes saturated above a certain creepage value, which corresponds to the Coulomb limit value, i.e. the coefficient of friction μ times the contact normal load N. The curve of Figure 1.5b in turn represents an effect equivalent to that of camber thrust in tire–road contact. For high spin values, the relevance of the elastic deformations vanishes, and the problem may be treated as a contact between rigid bodies. With a symmetrical contact patch and in the absence of viscoelastic or dynamic effects, the resultant tangential forces tend to zero with high spin values.

The tangential contact stresses and resultant forces depend on the creepages but not on the absolute velocities, as long as the coefficient of friction does not depend on the slip velocity. They also depend on the normal load and friction conditions.

The non-linear creepage–creep force curve relationships, together with the contact geometric non-linearities, are the two sources of non-linearity in the

wheel–rail contact. These non-linearities are usually one of the main difficulties in rail vehicle dynamic simulations.

1.1.2 Conformal contact

The most usual wheel–rail rolling contact theories rely on simplifying assumptions, some of which are listed in §1.1.1. One of these assumptions is that the contact is concentrated. As previously introduced, this implies that the contact patch is planar, and that the contacting bodies can be regarded as half-spaces at the local contact level. The latter brings an important advantage with regards to the mathematical solution of the contact problem, as the mechanical behaviour of the elastic half-space is known and there are closed-form solutions available (cf. §2.2.1), which can be applied in many different contact problems.

In many wheel–rail contact situations, the assumption of concentrated contact is reasonable and enables a simplification of the contact problem without introducing appreciable error. However, in some cases significantly non-planar wheel–rail contact patches may be encountered in realistic operating conditions. One of these cases is the contact between the wheel flange root and the rail gauge corner during inscription in tight curves, where the dimensions of the contact patch can be of similar magnitude as the lateral radii of curvature of wheel and rail, giving rise to conformal contact. Conformal contact could also take place with worn wheel–rail profiles. In these cases, the contact patch may become significantly curved in the lateral direction, and the adequacy of a concentrated or non-conformal contact analysis is compromised. Moreover, in these situations usually severe contact conditions are encountered, with high tractions and slip velocities, and accurate determination of these are necessary to properly assess the wear and fatigue damage in the wheel–rail interface.

Conformal contact is found in applications with technological relevance other than the wheel–rail case, as in rolling bearings, in piston and cylinder assemblies in reciprocating machines, in pinned joints, in ball–socket joints and in hip joints. It is therefore of interest to have methods for the solution of conformal contact problems.

Conformal contact entails other difficulties, apart from non-flatness in the contact and compromising the half-space assumption. Conforming surfaces with small separations imply that a precise geometric analysis becomes necessary. It

implies as well that small displacements and deformations may alter considerably the local contact geometry. Deformations that happen at the global or structural level in the solids could alter the local contact geometry, bringing about the possibility of coupling between the structural and contact problems, in contrast to concentrated contact problems.

This happens, for instance, in the contact problem of a slender flat plate on a plane or with a similar elastic plate, bending under the action of a load applied in its central part, illustrated in [Dundurs 1970] and [Ciavarella 2006]; cf. Figure 1.6. In this case, the contact extends initially in the whole surface of the plate, and as the load is increased and the plate bends, contact is lost in the borders of the plate. This type of conformal contact, in which the contact area decreases as the load increases, is called a receding contact. Another example of conformal contact in which the structural deformations influence the contact solution is shown in [Sundaram 2010a]. This is a 2D cylindrical contact, in which the contacting cavity deforms under the action of remote (structural) stresses, so that the contact stress distribution is altered.

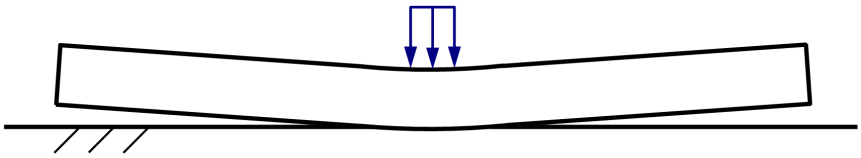


Figure 1.6. A receding conformal contact problem.

The phenomena of the above mentioned examples will not be present in the contact problems studied in this thesis. Here, advancing contacts between massive solids will be studied. So, the sections of the contacting bodies are sufficiently consistent to ensure that the geometry of the surfaces at the contact level is not significantly altered by the global deformations of the solids. As a consequence, in these cases the contact (local) and structural (global) problems may be treated as uncoupled. This can be done even if the contact occupies a region that may be appreciable with respect to the size of the contacting solids, and there is not a clear separation between the local and global stress fields.

Lastly, a distinction between conformal and non-planar is pertinent. Non-planar contact problems are necessarily conformal, in the framework of linear

elasticity. On the other hand, conformal contacts may be either planar or not. An example of planar conformal contact is the receding contact case of the slender beam given above. There may also be advancing, planar and conformal contacts, as the case with Hertzian conforming geometry and low load shown in §4.3.1.

1.1.3 Vehicle–track dynamic interaction

Mathematical modelling of the railway vehicle–track interaction has been stimulated by a wide range of problems. Stability, curve inscription performance (including derailment risk, curving forces, and wear index assessments) and passenger comfort analysis are some of the issues related to the railway vehicle. These analyses are carried out at low frequencies, at which the track and unsprung masses are not dynamically excited. Therefore, usually the track elasticity is neglected or modelled in a very simplified way in this type of analyses, while the vehicle is represented with multibody models, including details of both the unsprung and sprung masses and of the connections between the different bodies.

In the case of ride comfort analysis, the dynamics of the sprung masses of the vehicle are of primary interest, and these are uncoupled from the track dynamics by means of the primary and secondary suspensions. In the case of curve inscription analysis, the dynamics of the vehicle's axles normally occur at frequencies too low for the track to be excited dynamically, and thus, they are determined by the vehicle's dynamic properties, but not by the track's. However, such analyses are only valid for very low frequencies, normally below around 20 Hz, where the track behaves as a relatively stiff spring [Knothe 1993], [Popp 1999], [Nielsen 2003].

When phenomena at higher frequency ranges need to be studied, consideration of track dynamics becomes necessary. This phenomena includes wheel–rail interaction at short–wave irregularities such as welds ([Steenbergen 2008], [Correa 2015], [Correa 2018a], [Correa 2018b]) and irregular wear or corrugation ([Sato 2002], [Grassie 2009], [Baeza 2011], [Robles 2023b]), which can lead to increased wheel–rail forces. These in turn lead to increased deterioration of the rolling stock and track components —e.g., fatigue, loss of track geometry, and deterioration of the wheel and rail rolling surfaces in the

form of rolling contact fatigue (RCF) ([Ekberg 2002], [Ekberg 2005])—, and to environmental problems as increased rolling noise ([Thompson 2000], [Vincent 2000], [Thompson 2009]) and ground-borne vibration ([Jones 2000], [Lombaert 2006], [Xia 2010], [Galvín 2010b], [Alves Costa 2012]).

There are commercial software packages that enable the development of multibody models fit for the development of railway vehicle dynamic simulations, e.g. for the homologation of new railway vehicles (see e.g. [EN 14363 2017]), such as SIMPACK and VI-Rail, that are used both in the academy and in the industry. In contrast, integrated vehicle–track dynamic interaction models developed with commercial software packages are not so common. Traditionally, when there has been a need for models capable of considering the dynamic interaction between vehicle and track in a representative way, they have been developed exclusively by the interested research teams, tailored for each specific study. Many examples of this can be found in the literature, e.g., [Wu 2004], [Xie 2008], [Otero 2009], [Jin 2008], [Galvín 2010a], [Correa 2012], [Di Gialleonardo 2012], [Robles 2022], [Robles 2023a].

Comprehensive surveys of vehicle–track interaction modelling can be found in [Nielsen 2003] and [Popp 2005]. Detailed discussions on the vehicle and track modelling are provided in [Popp 1999] and [Knothe 1993], and on the utilization of FEM analysis for track models in [Dietz 2002]. [Bezin 2009] also provides a useful overview of the problem and summarizes the options available within commercial railway vehicle multibody (MBS) codes. [Bezin 2009], [Bezin 2010] and [Gonzalez 2008] show some interesting application examples in which multibody models including elastic bodies are used. ADAMS Rail (currently VI-Rail, [VI-grade www]) is used as the analysis software in [Bezin 2009] and [Bezin 2010], and SIMPACK ([SIMPACK Wheel Rail doc 2007]) in [Gonzalez 2008].

1.1.4 Frictional contact analysis in rolling bearings

The contact conditions between rolling elements and raceways determine to a large extent the operational limits, performance and life of rolling bearings. Contact mechanics is thus an essential discipline in their design and analysis. In well-lubricated contacts, the tangential stresses are much lower than the normal

pressures in the contact. As a result, the precise solution of the normal contact problem has traditionally received more attention than the tangential contact problem in rolling bearings.

However, even in well-lubricated contacts, the issues related to the tangential contact problem are relevant in several aspects of rolling bearing design and operation. For instance, the frictional dissipation occurring in the rolling element–raceway contacts may be an important heating source, and call for limiting the operating speed to avoid excessive temperature increase in the bearing. It is also essential for determining the rolling torque (as stated in §14 of [Harris 2001], of the many components that constitute the frictional resistance to motion in a ball–raceway contact, sliding is the most significant), the transmission efficiency and the wear evolution in the running surfaces of the bearing. This wear may alter the profiles of the contacting surfaces, possibly leading to increased peak normal pressures and reduced RCF life, as was shown in [Oloffson 2000]. As indicated in [Harris 2001], even though the shear stresses acting on the rolling elements and raceway surfaces in contact are small compared to the normal pressures, they cannot be neglected for the determination of the bearing endurance with regard to fatigue, and in many cases, they are the most significant factor in determining the endurance of a rolling bearing in a given application. By way of example, in [Slack 2010] it was shown how relatively small levels of friction can have a noticeable effect on the propagation pattern of RCF spalls. More recently, it was demonstrated that wear in grease-lubricated spherical roller bearings cause contamination of the grease, which in turn accelerates wear and degradation of the bearing to premature replacement [Pozzebon 2020], [Lin 2021]. The frictional contact analysis is an essential part of this process.

1.2 Statement of the contact problem

An elastostatic and concentrated contact between two massive, isotropic and homogeneous solids is considered. With these conditions, the contact area is approximately flat, as previously introduced.

The different relevant magnitudes of the problem are expressed in a local contact Cartesian coordinate system, with $\{x, y, z\}$ axes. The x and y axes are contained in the contact plane, and the z axis is normal to it. The x axis is

designated as longitudinal, and the y axis as the lateral axis. The x axis coincides with the rolling direction in rolling contact problems, and the forward or rolling velocity V is oriented along the positive x axis. The $\{x, y, z\}$ system follows the contact, so it becomes an Eulerian coordinate system in rolling contact problems. In partial slip rolling contact problems within linear elasticity (with small deformations), the material particles of both contacting bodies flow through the $\{x, y, z\}$ system at an approximate velocity of V in the negative x direction (cf. Figure 1.9b).

The contact problem is set out on the basis of two configurations of the contacting solids:

- Reference or undeformed configuration. It includes only the rigid body displacements of the solids. It may include structural deflections as well, which are produced at a global level in the solids, with much smaller spatial gradients than those produced at the local contact level.
- Final or deformed configuration. It is obtained adding the local contact deformations to the reference configuration.

Figure 1.7 depicts a section of the contacting solids, identified with numbers 1 and 2, through a perpendicular plane to the contact, showing both configurations. The reference configuration is shown in dashed lines, and the final configuration in thicker solid lines. The same figure shows an additional configuration in dotted lines, which may be termed as the initial one. The difference between the initial and reference configurations is a rigid body displacement and rotation.

To better appreciate the difference between the different configurations, the inclinations of the surfaces of the solids are exaggerated in the figure. Marked in the figure are different distances in the z axis from the contact plane to the surfaces of each contacting solid k, $k \in \{1, 2\}$, which will be referred to later. h_k and e_k are the distances from the contact plane to the surfaces in the reference and final configurations, respectively, u_{zk} the displacements due to elastic deformation, and d_k the normal approach of each solid (difference between the initial and reference configurations). The figure shows also the real contact area $\boldsymbol{C}$, and the apparent contact area $\boldsymbol{C_a}$. $\boldsymbol{C_a}$ is obtained as the intersection of both solids overlapped in the reference configuration.

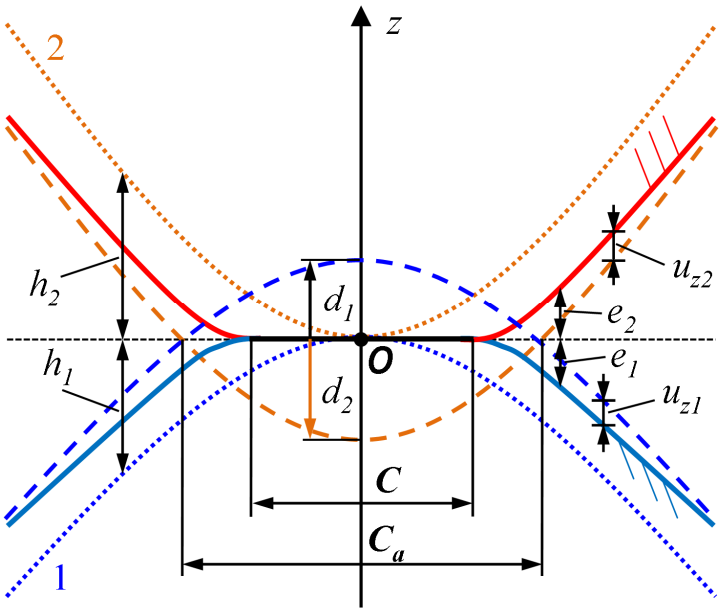


Figure 1.7. Configurations of the contacting solids, and relevant distances in the normal problem.

The contact problem is set out as follows:

- Given:
    - The geometry of the surface of each solid k, $k \in \{1, 2\}$, around the contact, described by the distance of the undeformed surface to the contact plane, $h_k = h_k(x, y)$.

      Under the adopted starting hypotheses, the relative displacements between the solids around the contact are much lower than the dimensions of the contact zone. Consequently, for each surface point in the surface of one of the solids, it is known beforehand the homonym point in the surface of the other solid which may enter in contact with it. This is the one with the same (x, y) coordinates in the local contact coordinate system. It is only necessary to consider the combined magnitudes of both solids (that is, the ones of a solid with respect to the other) in the formulation of the contact problem, not being necessary to consider the individual ones of each

solid. This applies to the distances, displacements and strains that enter in the contact problem, as well as to their derivatives. For this reason, in the remainder of this section the subindex k of each solid will be dropped, and combined magnitudes will be considered unless otherwise stated.

- The relative movement between the solids.

  It is defined considering the reference configuration of each solid (that is, considering the solids as rigid), and in general will be given by a displacement and a rotation. In rolling problems, velocities have to be considered as well. Considering a linear elastic and quasiestatic material behaviour, the velocities are of interest only in the tangential plane of the contact. The relative velocity field in this plane $\mathbf{w} = (w_x, w_y)(x, y)$, called also the rigid slip velocity, may be defined by only three normalized parameters, namely the ξ, η, and φ creepages. In planar contact, $\mathbf{w}$ is expressed as a function of the creepages according to Eq. (1.5). The superposition of the ξ, η and φ creepages is shown schematically in Figure 1.8.

$$\mathbf{w} / V = (w_x, w_y) / V = (\xi - \varphi\, y, \eta + \varphi\, x) \tag{1.5}$$

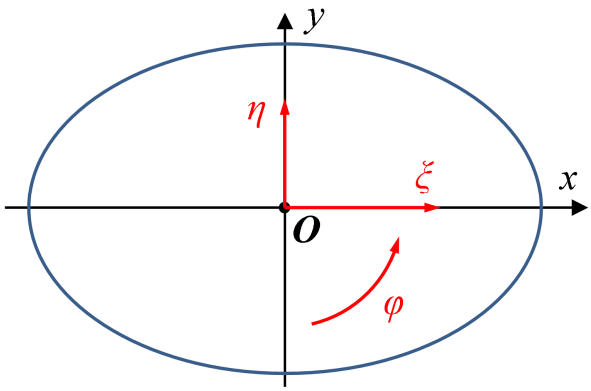


Figure 1.8. Superposition of longitudinal, lateral and spin creepages in the contact patch.

Alternatively, instead of part or all of the data of the relative movement, there may be the resultant loads transmitted in the contact in the specification of the problem.

- The coefficient of friction μ in the contact.

  It may be given through more or less complex formulations. A single, constant value may be considered in the whole contact, a static value and another dynamic one, or there may be values which are variable in space or as a function of some magnitudes that are part of the solution of the problem as the slip velocities.

- Find:
  - The contact area $\boldsymbol{C}$, and its division into the adhesion $\boldsymbol{A}$ and slip $\boldsymbol{D}$ zones.
  - The distributions of the stresses transmitted in the contact, $\mathbf{p} = (p_i)(x, y)$, with $i \in \{x, y, z\}$.

    Once these are obtained, other magnitudes of interest may be obtained as well, as the displacements due to elastic deformation $\mathbf{u} = (u_i)(x, y)$, with $i \in \{x, y, z\}$, and the slip $\mathbf{s} = (s_i)(x, y)$ or its velocities $\mathbf{v} = (v_i)(x, y)$, with $i \in \{x, y\}$.
- So that:
  - $p_z > 0$, $e = 0$ if $(x, y) \in \boldsymbol{C}$; and $p_z = 0$, $e > 0$ if $(x, y) \notin \boldsymbol{C}$.

    The constraints referred to the normal pressures p_z reflect that these may only be positive or zero, considering compressive normal pressures as positive. Contacts with adhesion are therefore excluded. The constraints referred to the normal distance e between the deformed surfaces express that there may not be interpenetration or overlap between the surfaces of the solids. Therefore, the constraints related to p_z and e are complementary, having that $e \geq 0$; $p_z \geq 0$; and $e \times p_z = 0$ for $\forall$ (x, y) in the boundary $\boldsymbol{\Gamma}$ of the solids (inside as well as outside the contact).

The constraints expressed here correspond to a "hard" relationship between p_z and e, in which no overlap is allowed for any p_z value. Other normal pressure–overlap constitutive relationships may be defined, to include the influence of roughness at lower scales than that of the modelled surfaces, or to smooth the numerical singularity of the "hard" relationship in the $(e, p_z) = (0, 0)$ point (cf. §5.1.2 of [Wriggers 2006] and §4.2.1.3).

- $\|\mathbf{p_t}\| \equiv \|(p_x, p_y)\| = g$, $\mathbf{v} \equiv (v_x, v_y) = -\lambda\, \mathbf{p_t}$ ($\lambda > 0$) if $(x, y) \in \boldsymbol{D}$; and $\|\mathbf{p_t}\| < g$, $\mathbf{v} = \mathbf{0}$ if $(x, y) \in \boldsymbol{A}$.

  The slip velocities **v** may be replaced by the slips **s** in these constraints. Here it has been chosen to use **v**, with a view to rolling contact problems.

  g is the adhesion limit for each contact point. It is defined as $\mu \times p_z$ according to Coulomb's law for dry friction. According to Coulomb's law as well, the slip takes place in the direction to the tangential stress $\mathbf{p_t}$ in each point. This could be generalized to an anisotropic friction law, considering different coefficients of friction in each direction. This could be relevant for textured surfaces for instance, with directional roughness patterns.

  As the constraints related to p_z and e, the constraints related to $\mathbf{p_t}$ and **v** are complementary, having that $\|\mathbf{v}\| \times (g - \|\mathbf{p_t}\|) = 0\ \forall\ (x, y) \in \boldsymbol{C}$, and $\boldsymbol{A} \cap \boldsymbol{D} = \varnothing$. Here also other constitutive relationships between $\mathbf{p_t}$ and **v** may be defined, different to the "hard" one expressed by the preceding constraints.

  These constraints introduce higher difficulty in the contact problem that the ones of p_z and e, because they are non-linear, and because in this case the number of unknowns is doubled as there are two-component vectors instead of scalars.

To complete the formulation of the problem, it is also necessary to know the constitutive relationships between the elastic displacements **u** and the stresses **p** in the surface of the solids. These may be expressed generally according to Eq.

(1.6). In this equation, ι_{ij} are the influence functions (IFs) of the solids, which relate the displacements in i direction u_i with the stresses in j direction p_j; $i, j \in \{x, y, z\}$. The integral in the equation is extended in the region $\boldsymbol{C}$ of the surface $\boldsymbol{\Gamma}$ of the solid which enters in contact, where the contact stresses p_j in general are non-zero. In concentrated contacts, the half-space IFs are applicable. These concepts are further elaborated in §1.3.

$$u_i(\mathbf{x}) = \iint_{C \in \Gamma} \sum_j \iota_{ij}(\mathbf{x}, \mathbf{x}')\ p_j(\mathbf{x}')\ dS(\mathbf{x}');\quad i, j \in \{x, y, z\} \tag{1.6}$$

The contact problem is divided into the normal and tangential parts.

- The normal part consists on obtaining the contact area $\boldsymbol{C}$ and the distribution of p_z inside it, taking into account the constraints of p_z and e. The fundamental inputs are the normal undeformed distance function $h(x, y)$, and either the normal approach between the solids d or the resultant normal load transmitted in the contact.

  Inside $\boldsymbol{C}$, the constraint $e = 0$ is fulfilled. Additionally, $e = h - d + u_z$, as may be observed in Figure 1.7. Therefore, in $\boldsymbol{C}$: $u_z = d - h$.

- The tangential part consists on obtaining the adhesion $\boldsymbol{A}$ and slip $\boldsymbol{D}$ areas inside $\boldsymbol{C}$, and the distribution of $\mathbf{p_t}$ in them, taking into account the constraints of $\mathbf{p_t}$ and $\mathbf{v}$. The fundamental inputs are the rigid slip velocity field $\mathbf{w}$ in the contact tangential plane, and the field of the adhesion limit g.

  The relative movement between the contacting solids may be decomposed into the solid rigid movement (corresponding to the reference configuration) and the one associated to the displacements due to elastic deformation (difference between the final and reference configurations). Then, Eq. (1.7) may be written for the total velocity $\mathbf{v}$:

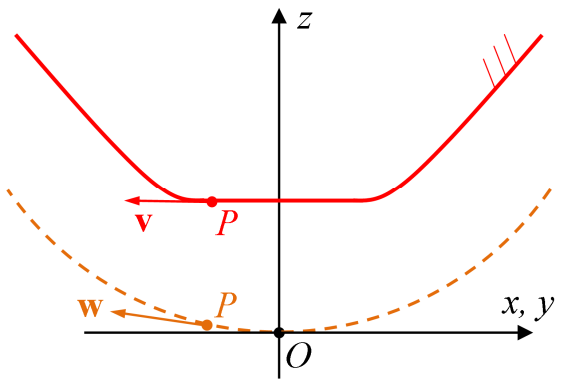
$$\mathbf{v} = \mathbf{w} + \frac{\mathrm{D}\mathbf{u}}{\mathrm{D}t} \approx \mathbf{w} - V\frac{\partial \mathbf{u}}{\partial x} + \frac{\partial \mathbf{u}}{\partial t} \tag{1.7}$$

  The derivative of the second term of the middle is written in capital D to denote that it is a Lagrangian derivative, i.e., associated to the

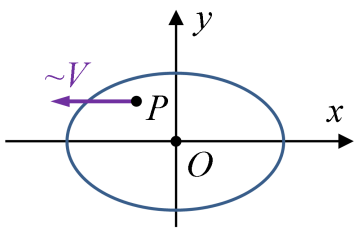
material particle. The rigid slip velocity **w** and total velocity **v** are represented for a material particle *P* in the reference and final configurations, respectively, in Figure 1.9a.

When applied to the adhesion part of the contact (with **v** = 0), Eq. (1.7) states that rigid velocity differences are compensated or absorbed by deformations at the contact level, as illustrated in the rolling wheel example of Figure 1.2a.

The approximation of the last term of Eq. (1.7) holds for partial slip contact problems with small deformations, in which case the velocity at which the material particles flow through the local contact system approaches V in the negative x direction, as previously stated and illustrated in Figure 1.9b. The convective term $-V\,\partial\mathbf{u}\,/\,\partial x$ appears in rolling contact problems, and is zero in static (compression or shift) contact problems. This term loses precision in rolling problems with gross slip. But in this case, the total velocity field is dominated by rigid slip as Kalker pointed out [Kalker 1990], [Kalker 2001], $\mathbf{v} \approx \mathbf{w}$, and it is not critical to approach correctly the D**u** / Dt term.



(a) Tangential velocities in the reference and final configurations of the solids.



(b) Flow of material particles through the Eulerian local contact coordinate system in rolling contact problems.

Figure 1.9. Tangential contact velocities.

A major difficulty for the contact problem solution is that the previously stated contact conditions are applicable to different regions (contact and exterior areas for the normal problem, and adhesion and slip areas for the tangential problem),

which are not known beforehand. So, in general, some iteration is necessary to find the correct division of areas and consequently the correct set of equations that define the problem.

In the most general case, the contact normal pressures cause tangential displacement differences u_x, u_y, in the contacting surfaces in addition to normal displacement differences u_z. And vice versa, the contact tangential stresses cause normal displacement differences in addition to tangential ones. If this is the case, the normal and tangential contact problems are coupled. However, this is not the case of many technical applications, with concentrated contacts and elastically similar materials, as the wheel–rail case. In this case, due to similarity, the normal pressures do not cause tangential displacement differences, and the tangential stresses do not cause normal displacement differences. Consequently, the normal and the tangential contact problems are uncoupled, and may be solved independently. This facilitates the solution of the frictional contact problem. The mentioned similarity principle is known as quasiidentity, and is illustrated in Figure 1.10 for the case of two contacting surfaces transmitting a tangential contact force F. Here it is shown how normal elastic displacements u_z caused in the surface of each solid cancel out with the corresponding u_z displacements caused in the other surface by the reciprocal tangential force.

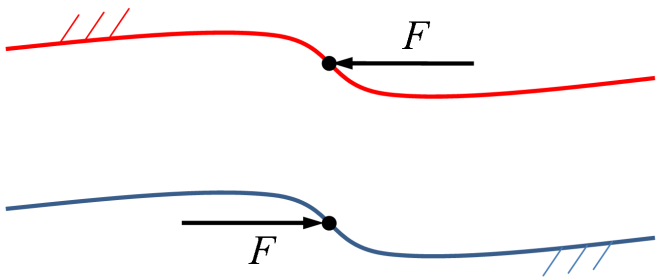


Figure 1.10. Illustration of quasiidentity for elastically similar solids subject to contact tangential stresses.

Due to the mentioned uncoupling between the normal and tangential contact problems, wheel–rail contact models are sometimes developed targeted at the solution of one of the parts of the contact problem, and may be designated as either normal or tangential contact models. Some of these models are reviewed

in §2.1.1 and 2.1.2, respectively. Normal contact models may be limited to the solution of the normal contact problem, while tangential contact models necessitate the previous solution of the normal contact problem to get the traction bound g.

1.3 Contact mechanical behaviour of the elastic solid

1.3.1 Influence functions and influence coefficients

As seen in §1.2, in the statement of the contact problem it is necessary to provide a relationship between the displacements due to elastic deformation in the contact surfaces and the stresses transmitted in the contact; cf. Eq. (1.6). ι_{ij} are the influence functions (IFs) which provide the constitutive relationship between the displacements and the stresses. Therefore, these functions characterize the mechanical behaviour of the elastic solid around the contact.

For contact problems, the displacement differences or relative displacements in the surface of the contacting bodies are of especial interest, as stated in §1.2. Designating the contacting bodies with numbers 1 and 2, these displacement differences u_i (with $i = x$, y or z) are defined according to Eq. (1.8), where $u^k_i(x, y)$ stands for the displacement in direction i of point (x, y) of the surface of body k (with $k = 1$ or 2). The positive direction of the z axis is defined here pointing into body number 2.

$$u_i(x, y) = u^2_i(x, y) - u^1_i(x, y) \tag{1.8}$$

Accordingly, both combined ι_{ij} and individual ι^k_{ij} IFs may be considered. The combined ι_{ij} IFs may be easily deduced from the ι^k_{ij} IFs, taking into account that the contact stresses act with opposite sign in each of the bodies (by Newton's 3rd law).

In general, the IFs ι^k_{ij} depend at least on the geometry of the solid and the mechanical properties of the material, and change as a function of point $\mathbf{x}$ where the load is applied and point $\mathbf{x}'$ where the displacements or other relevant magnitudes are observed. On the other hand, here it is considered that the material behaviour is linear elastic. In these conditions, the IFs do not depend on the stress levels nor on the applied loads in different points of the solid.

Moreover, the inertial effects in the vicinity of the contact are neglected. That is to say, it is considered that the stresses p^k_j in the surface of solid k are applied quasi-statically, in such a way that the IFs $\imath^k_{ij}$ are independent of the rate of load change and of its movement velocity on the surface of the solid. For this hypothesis to be valid, it is necessary on one hand that the applied loads are not shock-like, with very high change rates. §11.4 of [Johnson 1987] deals with the conditions under which this hypothesis may be considered valid. On the other hand, it is necessary that the movement velocities of the loads over the surface of the solid are much lower than the wave propagation velocities in the elastic solid, as is explained in [Johnson 1987] (§11.6) and in [Wang G 1989]. Taking into account that the wave propagation velocities in steel are on the order of thousands of meters per second, this premise is fulfilled for the case of wheel–rail contact at realistic travel speeds.

Under the above assumptions, Eq. (1.6) expresses a linear and instantaneous superposition. The influence coefficients (ICs) are defined from the IFs, upon the spatial discretization of the integral of Eq. (1.6) (cf. Eq. (2.100)). They relate the response of the elastic solid with distributed loads applied in finite elements of its surface, and may be obtained integrating the product of the corresponding IFs and the shape function of the distributed load in the loaded element. Making use of the superposition principle at the loaded element level, it may be written:

$$IC^k_{ij}(x,y,z) = \iint_{S_e} \imath^k_{ij}(X,Y,z)\, T(X,Y)\, dXdY \tag{1.9}$$

Being IC^k_{ij} the IC associated to the IF $\imath^k_{ij}$ (with $i, j = x, y$ or z) and $T(X, Y)$ the shape function which defines the distribution of the unitary load in the surface S_e of the loaded element, which centre is in the point $(x, y, 0)$. The observation point is defined here to be located at $(x, y, z) = (0, 0, z)$. The ICs of the displacement differences, IC_{ij}, are obtained in a similar way from the $\imath_{ij}$ IFs.

1.3.2 The elastic half-space

The solid domain delimited just by a plane in its free surface is termed as elastic half-space. For this particular case, the Green's functions or influence functions which provide the responses (displacements, strains or stresses) of the half-

space to point loads applied in the surface are available analytically. The IFs of the surface displacements of the half-space are reviewed in §2.2.1.

As previously stated, in concentrated contact problems between elastic solids in which the elastic displacements are sufficiently small for the linear theory of Elasticity to hold, the dimensions of the contact zone are much smaller than the characteristic dimensions of the contacting bodies and their curvature radii around the contact. In these conditions, the resulting elastic field is highly concentrated around the contact zone, so that the region of practical interest from the contact mechanics point of view is limited to this zone ([Johnson 1987] §2.1). As the contact zone is much smaller than the curvature radii of the surfaces and the contacting solids, it may be considered that this zone, as well as the adjacent parts of both solids, are nearly flat. From this it follows that it is valid to assume that the solids behave like half-spaces in the vicinity of the contact, regardless of their global geometry and supporting conditions. This hypothesis, which is used (not exclusively) in the Hertzian contact theory, is very convenient, as it allows to use directly the IFs of the elastic half-space, known and available in analytical form, in a great variety of contact problems.

1.4 Objectives

The following objectives are set out for this thesis:

- Develop comprehensive vehicle–track interaction models with commercial MBS and FE software packages. These models should provide a realistic representation of the track flexibility in a wide frequency range.
- Develop accurate wheel–rail contact mechanics models with the capacity to treat conformal contact.
- Carry out an analysis of the mechanical behaviour of non-planar elastic solids at the local contact level, assessing the differences with respect to the half-space. Related to this, the determination of ICs of solids with non-planar surfaces is to be addressed.
- Conduct an in-depth analysis of wheel–rail conformal contact, identifying the distinctive features of conformal with respect to non-

conformal contact. It is aimed as well to assess the errors incurred in different contact outputs when disregarding some of the effects of conformity in different conformal contact situations.

- Demonstrate the application of the developed contact mechanics models in the rolling bearing case.

1.5 Thesis layout

The thesis is organized as follows.

Chapter 2 provides a state-of-the-art review of the main topics studied in this thesis, focused on the development of comprehensive contact mechanics models with the ability to deal with conformal contact. This includes wheel–rail contact models, IFs of the elastic solid, numerical analysis of rolling contact, conformal contact, and contact mechanics analysis in rolling bearings.

In Chapter 3, dynamic models for the study of vertical vehicle–track interaction are developed with commercial MBS and FEM analysis packages. These are time-domain models, and have the capability to consider detailed and non-linear characteristics of both the vehicle and the track, e.g., hanging sleepers, and interaction between different wheelsets. The developed models are used to conduct a comparative study of the dynamic performance of different tracks.

Chapter 4 is devoted to the development of numerical contact mechanics models for conformal contact analysis. These include a version of Kalker's exact contact theory that has been named as CECT (Conformal Exact Contact Theory), and Finite Element (FE) models for detailed contact mechanics analyses. Detailed descriptions are given of the extensions for conformal contact implemented in CECT, and of the necessary tuning of the FE models for reliable results. The development is initially aimed at the wheel–rail case, but in Chapter 7 the application to the rolling bearing case is demonstrated.

Chapter 5 treats the calculation of ICs for non-planar solids, being this one of the major difficulties associated to the application of the exact contact theory to conformal contact analysis as opposed to concentrated contact. The chapter includes new analytical results for the half-space.

Chapter 6 presents a detailed study of wheel–rail conformal contact, covering both the normal and the tangential parts of the contact problem. The study is based on numerical analyses carried out with the models and ICs developed in Chapters 4 and 5. The characteristics brought about by conformity and the differences with respect to non-conformal contact are assessed. A geometrical analysis of the undeformed distance between two non-planar surfaces of bodies of revolution is also carried out.

Chapter 7 demonstrates the application of the developed numerical contact models for the analysis of rolling contact in rolling bearings. This contrasts with the much simpler contact mechanics models commonly used in this application. With these models, a study is carried out of the roller–raceway rolling contact in a spherical roller bearing.

Lastly, in Chapter 8 the outcomes from the investigation carried out in this thesis are compiled, perspectives for future works are outlined, and the publications that have resulted from this work are listed.

Chapter 2

State of the art

Chapter summary

This chapter provides a review of some contact mechanics analysis topics, mostly focused on the wheel–rail application. Firstly, the development of wheel–rail contact models is reviewed. Afterwards, an overview of known analytical solutions of influence functions for elastic solids is given. In addition, brief historical remarks are provided on the numerical analysis of rolling contact and on the analysis of conformal contact. Lastly, a review of contact mechanics analysis in rolling bearings is carried out.

2.1 Wheel–rail contact models

The rigorous solution of the contact problem involves considerable complexity, as discussed in §1.1.1. Wheel–rail contact models aim to simplify the contact problem solution, while retaining acceptable precision. The level of required precision and detail depends on the application and purpose of the study. A high level of detail and accurate contact stress and micro-slip distributions may be required for the study of contact degradation phenomena such as wear and rolling contact fatigue (RCF). On the other hand, in rail vehicle dynamic studies, usually interest for the output of the contact computation is limited to the resultant contact forces, getting reasonable representations of the creepage–creepage force relationships (cf. Figure 1.5). Additionally, each dynamic simulation involves the solution of many wheel–rail contact problems. This is so because the high contact stiffness calls for short time steps to preserve stability of the numerical integration of the dynamic equations, and multiple wheel–rail contact problems have to be solved in each step (the corresponding to each wheel–rail contact pair of the simulated vehicle). So, in this case emphasis is placed on saving computational cost.

Figure 2.1 shows the set out of the wheel–rail contact problem, having as inputs the wheel and rail geometries as well as their relative positions and velocities, and as final outputs the resultant contact forces. For its solution, the contact problem is usually split into its normal and tangential parts, as explained in §1.2. The normal problem is solved first, and subsequently the tangential problem. This is represented in the central part of Figure 2.1. The solution of the normal contact problem requires the previous solution of the contact search geometrical problem, to locate the wheel–rail contact patch. This review is focused on contact mechanics models, and algorithms for the geometric contact search are not covered. Some aspects of the geometric contact search are discussed in §4.2.3.1.

The tangential problem has as inputs part of the solution of the normal problem, in addition to the wheel–rail relative kinematics. When there is some coupling between the normal and tangential contact problems, it may be necessary to solve the normal problem again taking into account the previous solution of the tangential problem. This is represented by the arrow in parenthesis from the

tangential to the normal problem. It may then be necessary to solve iteratively both problems, until the correct solution to the coupled normal-tangential contact problem is found.

After the detailed solution of the normal and tangential contact problems, the resultant contact forces may be computed in a postprocessing step, by integration of the contact stresses. The resultant contact forces are represented in the right of Figure 2.1 by a normal force N, a tangential force T, and a moment M_z.

Some wheel–rail contact models aim at computing the resultant contact forces directly, without going through the detailed solution of the normal and tangential contact problems. This is represented by the dashed arrow from the left to the right hand side of the figure.

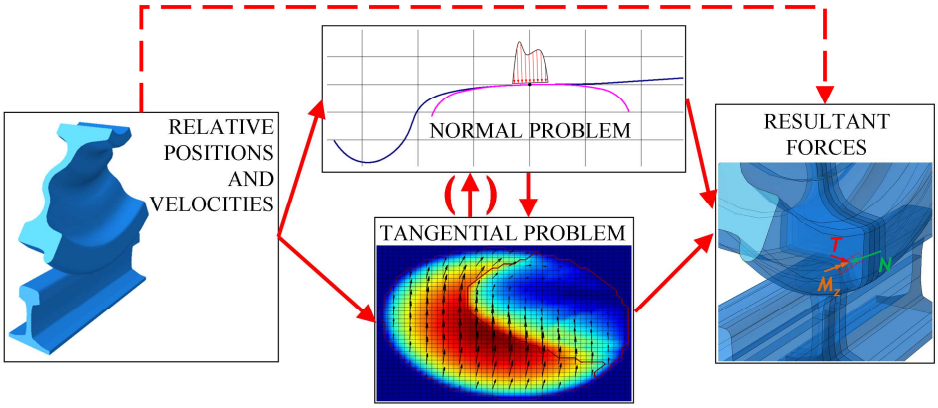


Figure 2.1. Set out of the wheel–rail contact problem.

In the following subsections, models for the solution of the wheel–rail normal and tangential contact problems are reviewed. Some of the models are aimed at providing just resultant contact forces, and others provide contact stress distributions as well. Another review of wheel–rail contact models is provided in [Meymand 2016]. Earlier reviews and historical briefs may be found in [Kalker 1979a], [Kalker 1991], [Piotrowski 2005], [Ayasse 2006] and [Knothe 2008].

The contact $\{x, y, z\}$ Cartesian coordinate system defined in §1.2 is used.

2.1.1 Normal contact

The Hertz theory and three popular non-Hertzian contact models used for the solution of the wheel–rail normal contact problem are reviewed here. The Hertz theory is the first known contact mechanics theory, and it is an exact and fast model. It is widely used for contact calculations in many applications, e.g. in the verification of wheel–rail contacts in cranes in [EN 13001-3-3 2014].

Non-Hertzian contact models arise for application in cases in which the Hertz theory cannot be accurately applied, i.e., with non-constant curvatures of the contact surfaces. The three non-Hertzian contact models reviewed in this section assume a semi-Hertzian representation of the undeformed normal distance h, with its principal directions aligned with the rolling (x) and lateral (y) directions of the contact coordinate system, described according to Eq. (2.1). In this equation, h_0 may be a general (not necessarily quadratic) function. R_x is the combined curvature radius in x direction, calculated as $1 / R_x = 1 / R_{x1} + 1 / R_{x2}$, with R_{x1} and R_{x2} being the longitudinal curvature radius of each contacting surface.

$$h(x,y) = h_0(y) + \frac{x^2}{2R_x} \tag{2.1}$$

All the normal contact models presented here are applicable for frictionless contact, or in cases where quasiidentity may be applied.

2.1.1.1 Hertz theory

Contact mechanics is considered to have started with the publication of Heinrich Hertz's paper "On the contact of elastic solids" [Hertz 1882]. In this work, the solution for the frictionless contact between two ellipsoids is given. Apart from the hypotheses for the contact problems set out in §1.1.1, it is considered that the contact surfaces are quadratic, i.e., with constant curvatures around the contact. The undeformed distance from the contact plane to the surface of each body, h_k, is described then according to Eq. (2.2). In this equation, x_k, y_k are the principal axes of the surface of each body, and A_k and B_k are half of the corresponding principal curvatures. I.e., $A_k = 1 / (2 R_{xk})$, and $B_k =$

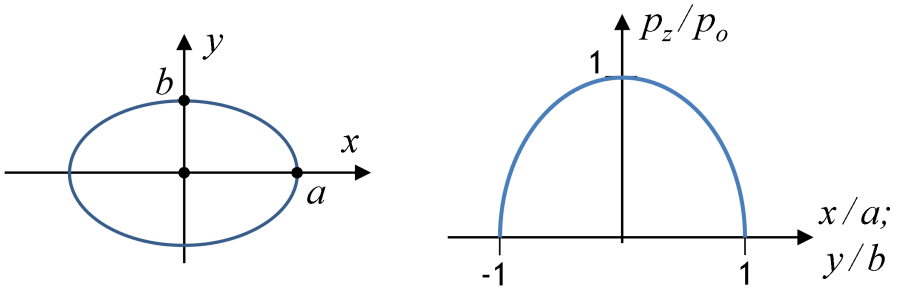
$1 / (2 R_{yk})$, being R_{xk} and R_{yk} the curvature radii of surface k ($k \in \{1, 2\}$) in x and y directions, respectively.

$$h_k(x_k, y_k) = A_k x_k^2 + B_k y_k^2; \quad k \in \{1, 2\} \tag{2.2}$$

It is assumed an elliptical contact area $\boldsymbol{C}$ with semi ellipsoidal contact pressure distribution p_z according to Eqs. (2.3) and (2.4); see Figure 2.2. The x and y axes in these equations correspond to the principal directions of the contact ellipse, which are in general different from the x_k, y_k axes of each surface. Note also that these x, y axes are in general different from the contact x, y axes defined in §1.2.

$$\boldsymbol{C} = \left\{ (x, y) \,|\, \left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 \le 1 \right\}; \quad k \in \{1, 2\} \tag{2.3}$$

$$p_z(x, y) = p_o \sqrt{1 - \left(\frac{x}{a}\right)^2 - \left(\frac{y}{b}\right)^2} \tag{2.4}$$



(a) Elliptical contact patch.

(b) Semi-ellipsoidal contact pressure distribution.

Figure 2.2. Hertzian normal contact solution.

The surface normal displacements are calculated applying Eq. (1.6), with the Boussinesq IF of the half-space (cf. Eqs. (2.107) and (2.112)) and the normal pressure distribution of Eq. (2.4). The resulting integral has a closed-form solution, as expressed in Eq. (2.5).

$$u_{zk}(x,y)=\iint_C \frac{(1-\nu_k)p_o}{2\pi G_k}\frac{\sqrt{1-\left(\frac{x}{a}\right)^2-\left(\frac{y}{b}\right)^2}}{\sqrt{(x-x')^2+(y-y')^2}}\,dx'\,dy'$$

$$=\frac{1-\nu_k^{\,2}}{\pi E_k}\left(K_L-K_M x^2-K_N y^2\right) \tag{2.5}$$

In this equation, ν_k, G_k and E_k are the coefficient of Poisson, the shear modulus, and the Young's modulus of body k, respectively. K_L, K_M, and K_N are given in Eqs. (2.6)–(2.8) in terms of elliptic integrals. In these equations, $\mathbf{K}(e)$ and $\mathbf{E}(e)$ are complete elliptic integrals of the first and second kind, respectively, and e is the eccentricity of the contact ellipse, given by Eq. (2.9). As can be seen in the equations, K_L, K_M and K_N depend only on the size and shape of the contact ellipse and on the maximum contact pressure p_o (which are unknowns in the normal contact problem).

$$K_L=\pi\, p_o b\,\mathbf{K}(e) \tag{2.6}$$

$$K_M=\frac{\pi\, p_o b}{e^2a^2}\left(\mathbf{K}(e)-\mathbf{E}(e)\right) \tag{2.7}$$

$$K_N=\frac{\pi\, p_o b}{e^2a^2}\left(\frac{a^2}{b^2}\mathbf{E}(e)-\mathbf{K}(e)\right) \tag{2.8}$$

$$e=\sqrt{1-b^2/a^2};\quad a\geq b \tag{2.9}$$

Eq. (2.10) gives the combined normal displacements u_z. This is obtained applying Eq. (2.5) to the two contact surfaces. In this equation, E^* is the combined effective Young's modulus of the two bodies, given in Eq. (2.11).

$$u_z(x,y)=\frac{1}{\pi E^*}\left(K_L-K_M x^2-K_N y^2\right) \tag{2.10}$$

$$\frac{1}{E^*} = \frac{1-\nu_1^{\,2}}{E_1} + \frac{1-\nu_2^{\,2}}{E_2} \tag{2.11}$$

On the other hand, applying the displacement compatibility equation for the normal contact problem (cf. §1.2), Eq. (2.12) is obtained. It is apparent in this equation that the principal axes of the normal undeformed distance function h are the same as the principal axes of the contact ellipse. The combined curvature parameters A and B are calculated according to Eqs. (2.20) and (2.21).

$$u_z(x, y) = d - h = d - Ax^2 - By^2 \tag{2.12}$$

Equating terms of Eqs. (2.10) and (2.12), the equalities of Eq. (2.13) are obtained, by which the normal contact problem is solved.

$$d = \frac{K_L}{\pi E^*}\,;\; A = \frac{K_M}{\pi E^*}\,;\; B = \frac{K_N}{\pi E^*} \tag{2.13}$$

In these equations, the usual inputs are A, B, and either the normal approach d or the normal load N, and the usual outputs are a, b, and either N or d. The difficulty in these equations is that the unknowns appear in the right-hand terms in a non-linear form, inside elliptic integrals, which cannot be analytically solved for. In practice, precalculated tables are used, that give non-dimensional coefficients m, n, and r as a function of the single non-dimensional geometry parameter $\cos(\theta)$, given in Eq. (2.14). $\cos(\theta)$ varies between 1 and 0, i.e., θ varies between 0º and 90º. The outputs a, b and d are proportional to m, n, and r, according to Eqs. (2.15)–(2.17).

$$\cos(\theta) = \frac{|B-A|}{B+A} \tag{2.14}$$

$$a = m \times \sqrt[3]{\frac{3N}{4(A+B)E^*}} = \frac{m}{\sqrt{r}}\sqrt{\frac{d}{A+B}} \tag{2.15}$$

$$b = n \times \sqrt[3]{\frac{3N}{4(A+B)E^*}} = \frac{n}{\sqrt{r}} \sqrt{\frac{d}{A+B}} \tag{2.16}$$

$$d = r \times \sqrt[3]{\left(\frac{3N}{4E^*}\right)^2 (A+B)} = \frac{r}{m^2}(A+B)\, a^2 \tag{2.17}$$

Precalculated tables for *m*, *n*, and *r* may be found e.g. in [Cooper 1968] (Table AI), [Roark's 2002] (Table 14.1), [Pascal 2007b] (Tables 1 and 2, including only *m* and *n* values), and [Thompson 2009] (Table 5.1). Equations for these parameters in terms of the complete elliptical integrals **K**(*e*) and **E**(*e*) and the contact ellipse eccentricity *e* may be found in [Thompson 2009] (Eqs. 5.41, 5.42, and 5.46 for *m*, *n*, and *r*, respectively).

The *a* and *b* semiaxes of the contact ellipse must be assigned to the *x* and *y* directions as a function of the geometry, considering that if $A > B$, then the ellipse semiaxis along *x* is smaller than the one along *y*. Extended tables may be constructed with θ going from 0 to 180º, using the property $m(A/B) = n(B/A)$ (and, obviously, $r(A/B) = r(B/A)$), with which it is not necessary to pay attention to this assignment. With these extended tables, *a* may be always assigned to the *x* axis, and *b* to the *y* axis.

p_o is related to *N* and the size of the contact ellipse according to Eq. (2.18), where p_m is the mean contact pressure. With this relation and Eqs. (2.15)–(2.17), other equations may be worked out relating p_o, *N*, E^*, *d*, and the size of the contact ellipse. Some useful Hertzian relations for 2D and 3D contact, as well as results for subsurface stresses, may be found in [Johnson 1987], [Roark's 2002] (§14), and [Pilkey 2005] (§9), for instance.

$$p_o = \frac{3}{2} p_m = \frac{3N}{2\pi\, ab} \tag{2.18}$$

From the Hertzian equations (2.15) and (2.16), it may be observed that the geometry of the contact ellipse depends primarily on the curvatures of the contact surfaces, and not so much on the load, which appears elevated to the power of (1/3). The surface curvatures may change broadly, being equal to the second derivative of the surface profiles. Another observation is that the shape

of the contact ellipse (the ratio between its principal axes a / b) does not depend on the load nor on the elastic properties. It only depends on the A / B ratio, i.e., on the ratio of the principal curvatures of the combined undeformed distance. For not too eccentric contact ellipses, a / b may be approximated with Eq. (2.19). The exponent of 2/3 is replaced by 0.63 in [Ayasse 2006] (Eq. 4.8).

$$\frac{a}{b} \approx \left(\frac{B}{A}\right)^{2/3} \tag{2.19}$$

The parameters A and B, which define the combined undeformed distance as $h = A\ x^2 + B\ y^2$, are calculated from the individual curvatures of the contact surfaces and their relative orientation with Eqs. (2.20) and (2.21). The subindexed C variables in these equations are the individual principal curvatures of each body (i.e., the inverse of the corresponding curvature radii). Concave curvatures are negative. γ is the angle between the principal planes of curvature of both contacting bodies, as shown in Figure 2.3.

$$A + B = \frac{C_{x1} + C_{y1} + C_{x2} + C_{y2}}{2} \tag{2.20}$$

$$|B - A| = \frac{1}{2}\sqrt{(C_{x1} - C_{y1})^2 + (C_{x2} - C_{y2})^2 + 2(C_{x1} - C_{y1})(C_{x2} - C_{y2})\cos 2\gamma} \tag{2.21}$$

The angle β defining the principal planes of the combined undeformed distance may be obtained with the geometrical analysis shown in Appendix 2 of [Johnson 1987]. Following that analysis, Eqs. (2.22) and (2.23) are obtained.

$$\left|\sin(2(\gamma - \beta))\right| = \left|\frac{(C_{x1} - C_{y1})\ \sin(2\gamma)}{2(B - A)}\right| \tag{2.22}$$

$$(C_{x1} - C_{y1})\ \sin(2\beta) = (C_{x2} - C_{y2})\ \sin(2(\gamma - \beta)) \tag{2.23}$$

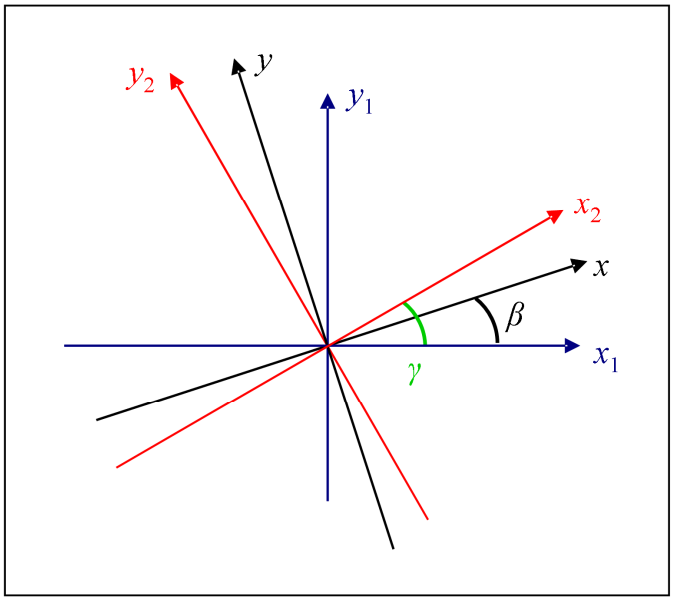


Figure 2.3. Orientation of the principal curvature planes of two contacting quadratic surfaces, and of their combined normal undeformed distance.

In the particular cases of a sphere of radius R on a plane, and of two crossed cylinders at 90º with the same radius R, $1 / R = A + B$. In the case of two crossed cylinders with the same radius crossed at an angle γ, $C_{x1} = C_{x2}$, and $C_{y1} = C_{y2} = 0$. In this case, it may be shown with Eqs. (2.20) and (2.21) that the angle θ defined in Eq. (2.14) is equal to γ.

In wheel–rail contact, see Figure 2.4, designating the rail as body 1 and the wheel as body 2, $R_{x1} = \infty$, $R_{x2} = R_{\text{roll}} / \cos(\delta)$, and $\gamma \approx 0$. R_{roll} and δ are the rolling radius and the contact angle, respectively. The approximation of Eq. (2.24) may be usually applied. In addition, as $\gamma \approx 0$, $\beta \approx 0$ may be assumed in some cases, implying that the principal directions of the undeformed distance function are assumed to be aligned with the principal directions of the wheel and rail surfaces, which are approximately aligned with the x and y directions of the contact coordinate system. These approximations are not precise in case of conformal contact in lateral direction, especially if the lateral curvature radii are much smaller than the longitudinal ones. The non-applicability of $\beta \approx 0$ in conformal contact is further treated in §6.1.2 and §6.2.1.2.

$$|A-B| \approx \frac{1}{2}\left(-C_{x1}+C_{y1}-C_{x2}+C_{y2}\right) \tag{2.24}$$

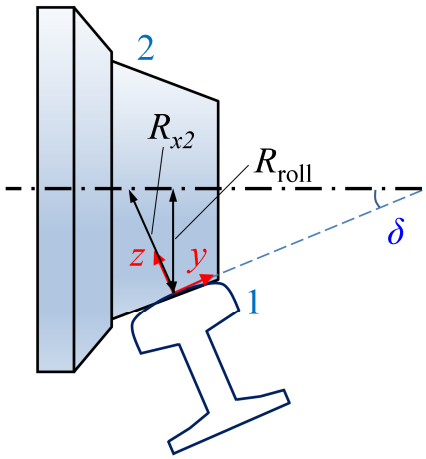


Figure 2.4. Rolling radius and contact angle in wheel–rail contact.

2.1.1.2 Kik & Piotrowski

The Kik & Piotrowski model ([Kik 1996], [Piotrowski 2008]) computes the overlap of the undeformed surfaces with some corrections as shown schematically in Figure 2.5, assuming a semi-Hertzian geometry as given in Eq. (2.1). The mentioned corrections are done so that in Hertzian cases the resulting contact patch approaches as closely as possible the corresponding Hertzian ellipse, as follows:

- Approach correction: a corrected, reduced approach, d_c, is used, instead of the theoretical one d. This correction is intended to adjust the size of the contact patch, considering that the apparent contact area $\boldsymbol{C_a}$ (obtained with the undeformed overlapped surfaces) is bigger than the real contact area $\boldsymbol{C}$ (cf. Figure 1.7). It is verified empirically that an adequate value for d_c is $d_c \approx 0.55d$, for not too eccentric contact ellipses. For a circular contact, it may be easily verified that the exact solution is obtained with $d_c = 0.50d$.
- Form correction: with a Hertzian undeformed distance function defined with its principal half curvatures A_0 and B_0 in x and y directions, the

width W to length L ratio of the obtained elliptical overlap with the uncorrected geometry complies with Eq. (2.25). On the other hand, the width to length ratio of the real contact patch is a different function of the A_0 / B_0 ratio, according to Hertzian theory (cf. §2.1.1.1), as expressed generically in Eq. (2.26). In this equation, n_0 and m_0 are the n and m Hertzian coefficients, respectively, defining each the size of each axis of the contact ellipse, corresponding to the Hertzian undeformed distance with A_0 and B_0 half curvatures.

$$\frac{A_0}{B_0} = \left(\frac{W}{L}\right)^2 \tag{2.25}$$

$$\frac{n_0}{m_0} = f\left(\frac{A_0}{B_0}\right) \tag{2.26}$$

The purpose of the form correction is to achieve an overlap with corrected width W_c and length L_c, with the same width to length ratio as the real contact patch. For a Hertzian case, this implies that W_c and L_c must comply with the relation given in Eq. (2.27).

$$\frac{L_c}{W_c} = \frac{m_0}{n_0} \tag{2.27}$$

The W_c and L_c values given in Eq. (2.28) comply with this relation, and in addition, $L_c\ W_c = L\ W$. Here, L and W are the dimensions of the overlap obtained with the uncorrected geometry and the corrected approach d_c. In a general non-Hertzian case, the n_0 / m_0 ratio necessary to calculate W_c and L_c is obtained from Hertzian theory as a function of $(W / L)^2$, i.e., replacing A_0 / B_0 with $(W / L)^2$ in Eq. (2.26), according to Eq. (2.25).

$$W_c = \sqrt{LWn_0/m_0}\ ;\ L_c = \sqrt{LWm_0/n_0} \tag{2.28}$$

To obtain the corrected overlap with W_c and L_c overall dimensions, a modified undeformed distance function is used, with corrected

longitudinal curvature radius $R_{x,c}$, and lateral profile $h_{0,c}(y')$, calculated according to Eqs. (2.29) and (2.30).

$$R_{x,c} = \frac{L_c^{\,2}}{8 d_c} \tag{2.29}$$

$$h_{0,c}(y') = h_0(y), \text{ with } y' = y - \frac{W_c - W}{W}(y_o - y) \tag{2.30}$$

Eq. (2.29) is easily verified as follows. The distance from the x = 0 position (i.e., the position with maximum overlap) to the leading and trailing limits of the contact patch, $x_l(y)$ and $x_t(y)$, is the same by symmetry along x, according to the assumed semi-Hertzian geometry. The longitudinal half curvature of the corrected geometry, A_c = 1 / (2 $R_{x,c}$), multiplied by either $x_l^{\,2}$ or $x_t^{\,2}$, must be equal to $d_c - h_{0,c}(y')$ in any lateral position. At the lateral position of maximum interpenetration, $|x_l|$ and $|x_t|$ are maximum and equal to L_c / 2, $h_{0,c}(y')$ = 0, and it is verified that $A_c\, x_l^{\,2} = d_c$.

In Eq. (2.30), the corrected lateral profile $h_{0,c}$ is calculated mapping the original profile h_0 on a modified lateral coordinate y', which is obtained by factoring the original lateral coordinate y with the W_c / W ratio. y_o in this equation is the lateral coordinate with maximum interpenetration.

$x_l(y)$ and $x_t(y)$ are calculated in each lateral position applying Eq. (2.1) together with the condition $d_c = h_c$, with the corrected parameters $R_{x,c}$, d_c, and $h_{0,c}(y')$. This results in Eq. (2.31).

$$x_l(y) = -x_t(y) = \sqrt{2R_{x,c} \times \max\left(0; d_c - h_{0,c}(y)\right)} \tag{2.31}$$

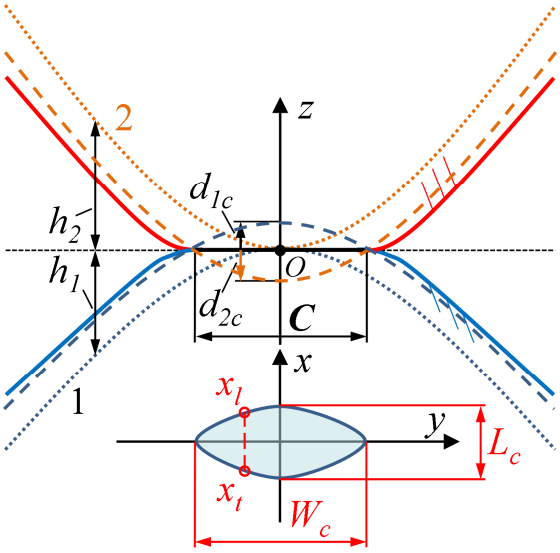


Figure 2.5. Computation of the contact patch as the interpenetration region of the undeformed surfaces with corrected approach, longitudinal and lateral dimensions in the Kik & Piotrowski normal contact model.

The normal pressure distribution in the contact patch is calculated assuming elliptical variation in longitudinal direction, and Hertzian proportionality with the longitudinal dimension of the contact patch in each lateral position. I.e., being p_o the maximum normal pressure in the contact patch (which takes place at the point of maximum overlap), and $x_l(y)$ the half longitudinal dimension of the contact patch at lateral position y, the maximum normal pressure along the strip located at lateral position y_i is given by Eq. (2.32). It follows that the pressure distribution throughout the contact patch is given by Eq. (2.33).

$$\max\left(p_z(x, y = y_i)\right) = p_o \frac{x_l(y_i)}{L_c/2} \tag{2.32}$$

$$p_z(x, y) = \frac{p_o}{L_c/2}\sqrt{x_l^{\,2}(y_i) - x^2} \tag{2.33}$$

Once the form of the pressure variation is known, its maximum p_o may be determined by imposing the displacement compatibility condition that the elastic deformation at the point of maximum overlap (where h = 0) must be

equal to the approach d. For this purpose, the product of $p_z(x, y)$ with the Boussinesq IF of the normal surface displacements is integrated in the computed contact patch (for a general non-Hertzian case, this integration is carried out numerically), leaving p_o as the only unknown in the resulting linear equation. After determining p_o, the normal load N is determined by integration of the normal pressures. In this way, the relation between N and d is obtained.

2.1.1.3 STRIPES

In a similar way as the Kik & Piotrowski method, the STRIPES method ([Ayasse 2005], [Quost 2006]) is based on the virtual interpenetration of the undeformed wheel and rail surfaces, with some corrections. In this case, the contact patch is discretized in longitudinal strips, and each strip i may have its particular values of longitudinal and lateral half-curvatures, A_i and B_i, and contact angle α_i. Thus, non-planar contact areas may be considered.

The B_i values are limited to a small positive value, so cases with concave lateral curvatures like the one shown in Figure 2.6 are out of the scope of this model.

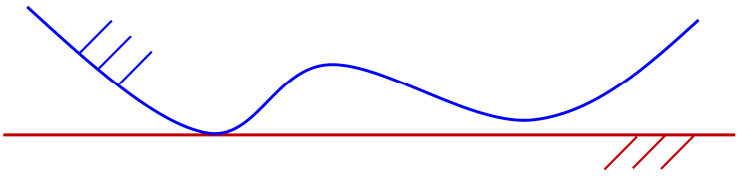


Figure 2.6. Contact surface with concave lateral curvature.

Figure 2.7 shows schematically the division in strips of the undeformed overlapped contact surfaces, as well as the overlaps and surface inclinations α_1, α_2, of each strip.

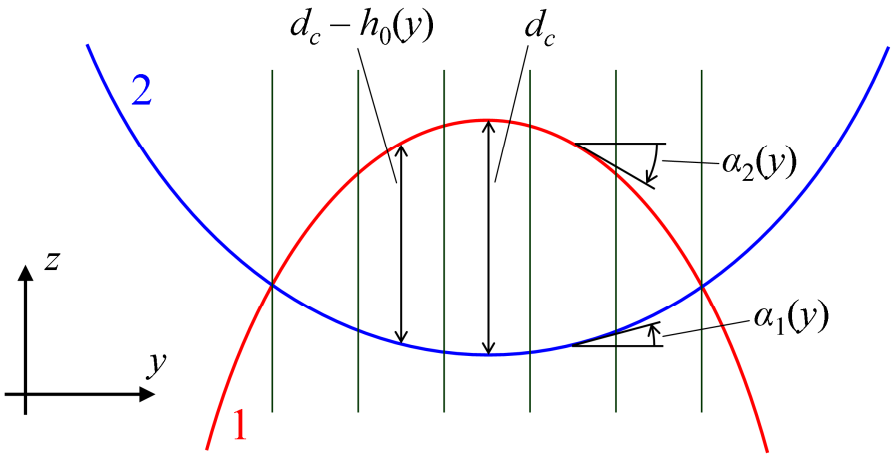


Figure 2.7. Overlap and surface inclination of each contact patch strip in STRIPES method. Adapted from [Quost 2006].

The corrected approach d_c is defined as $d_c = \varepsilon\, d$, being ε a factor depending on the geometry. The following two options for the approach and form corrections are proposed:

- Correction of only longitudinal half-curvatures A_i, and filtering of the lateral curvatures B_i. This is carried out in such a way that the contact patch obtained with a Hertzian geometry corresponds exactly to the Hertzian contact ellipse. ε is calculated according to Eq. (2.34). The subscript 0 in this equation denotes parameters corresponding to the strip with maximum overlap. n and r are non-dimensionless Hertzian coefficients that together with m define the contact ellipse and the approach in the Hertzian theory according to Eqs. (2.15)–(2.17).

$$\varepsilon = \frac{n_0^{\,2}}{r_0\left(1 + A_0/B_0\right)} \tag{2.34}$$

  The corrected longitudinal curvature in each strip, A_{ci}, is calculated according to Eq. (2.35). The subindex i denotes parameters calculated with the local geometry of each strip i.

$$A_{ci} = \frac{n_i^{\,2}}{m_i^{\,2}} B_i \tag{2.35}$$

The filtering on the lateral half-curvatures B_i is physically justified by the interaction between the different strips.

- Compensation of longitudinal and lateral half-curvatures, A_i and B_i. This is carried out imposing that the sum of the corrected longitudinal and lateral curvatures in each strip is equal to the sum of the uncorrected values, i.e., $A_{ci} + B_{ci} = A_i + B_i$. Eqs. (2.36)–(2.38) give the expressions for ε, A_{ci} and B_{ci}.

$$\varepsilon = \frac{n_0^{\,2}}{r_0 \left(1 + (n_0/m_0)^2 \right)} \tag{2.36}$$

$$A_{ci} = \frac{(A_i + B_i)\,(n_i/m_i)^2}{1 + (n_i/m_i)^2} \tag{2.37}$$

$$B_{ci} = \frac{A_i + B_i}{1 + (n_i/m_i)^2} \tag{2.38}$$

The semi-Hertzian normal pressure distribution is given by Eq. (2.39).

$$p_z(x,y) = p_{oi}(y) \sqrt{1 - \left(\frac{x}{x_l(y)} \right)^2} \tag{2.39}$$

p_{oi} in Eq. (2.39) is the maximum normal pressure in strip i. It is given in Eq. (2.40), which is based on the Hertzian relationship given in Eq. (2.41). In the latter equation, dN_i / dy is the integrated normal load in strip i per unit strip width.

$$p_{oi}(y) = \frac{2E^*\left(d_c - h_{0,c}(y)\right)}{\pi\, n_i\, r_i\, \varepsilon_i\, x_l(y)} \tag{2.40}$$

$$\frac{\mathrm{d}N_i}{\mathrm{d}y} = \frac{E^*\left(d_c - h_{0,c}(y)\right)}{n_i\, r_i\, \varepsilon_i} \tag{2.41}$$

2.1.1.4 ANALYN

The ANALYN method [Sichani 2014] is based on the virtual overlap between the wheel and rail semi-Hertzian surfaces. The correction of the undeformed normal distance function $h(x, y)$ for the calculation of the virtual overlap is carried out adding a term $\zeta(x, y)$ which varies in the same proportion as h. This correction is defined generalizing the surface deformations for the Hertzian case (with $h = A\,x^2 + B\,y^2$) given in Eq. (2.5) to a semi-Hertzian case, in which h is given by $h = A(y)\,x^2 + h_0(y)$. The correction term ζ and the corrected h function used for the calculation of the virtual overlap, h_{vo}, are given in Eqs. (2.42) and (2.43), respectively.

$$\zeta(x,y) = \alpha(y)\; A(y)\; x^2 + \beta(y)\; h_0(y) \tag{2.42}$$

$$h_{vo}(x,y) = h(x,y) + \zeta(x,y) = \left(1+\alpha(y)\right)\; A(y)\; x^2 + \left(1+\beta(y)\right)\; h_0(y) \tag{2.43}$$

The $\alpha(y)$ and $\beta(y)$ correction functions are given by Eqs. (2.44) and (2.45), respectively. These are calculated so that in a Hertzian case the obtained contact patch matches with the Hertzian contact ellipse. The contact patch half-length in each lateral position y is given by Eq. (2.46). The dependence of the different parameters with the lateral position y is indicated explicitly in the equations. The Hertzian coefficients m, n, and r are calculated accordingly with the local curvatures.

$$\alpha(y) = \frac{r(y)}{m^2(y)}\left(1+\frac{B(y)}{A(y)}\right) - 1 \tag{2.44}$$

$$\beta(y) = \frac{r(y)}{n^2(y)}\left(1+\frac{A(y)}{B(y)}\right) - 1 \tag{2.45}$$

$$x_l(y) = \sqrt{\frac{d - (1+\beta(y))\, h_0(y)}{(1+\alpha(y))\, A(y)}} \tag{2.46}$$

In this case, a correction for the approach d is not carried out. I.e., d in Eq. (2.46) corresponds directly to the approach, and coincides with the maximum overlap, as illustrated in Figure 2.8.

The correction term $\zeta(x, y)$ is designated as an approximation of the surface deformations in [Sichani 2014], though this term is not really a deformation: it is 0 at the point of maximum overlap, increases with h, and the displacement compatibility condition ($e = h - d + u_z$) inside the contact is not fulfilled with the deformations approached with this term.

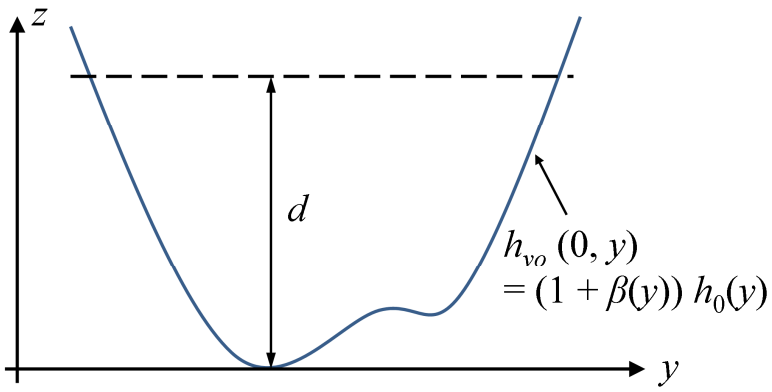


Figure 2.8. Corrected undeformed distance profile and overlap in ANALYN method.

As in the STRIPES method, the contact patch is discretized in longitudinal strips. The B_i half-curvatures are also limited to a minimum positive value. The semi-Hertzian pressure distribution is defined with Eq. (2.39), with p_{oi} being given by Eq. (2.47) in this case.

$$p_{oi}(y) = \frac{2E^*\left(d - (1+\beta(y))\, h_0(y)\right)}{\pi\, n_i\, r_i\, x_l(y)} \tag{2.47}$$

2.1.2 Tangential contact

The tangential contact problem is more difficult to solve than the normal problem, because the related constraints are non-linear and the number of involved unknowns is twice as much as in the normal problem, as stated in §1.2. Tangential contact models are therefore necessary in many wheel–rail interaction related studies.

In this section, the evolution of theories for wheel–rail tangential contact is reviewed, from the first analytical solution due to Carter [Carter 1926] and Fromm [Fromm 1926] to some of the most commonly used models nowadays for rail vehicle dynamic simulation.

Commonly adopted simplifying assumptions in wheel–rail tangential contact models are as follows:

- Steady-state rolling contact. This implies that the $\partial \mathbf{u} / \partial t$ term in the right-hand side of Eq. (1.7) vanishes. This is justified because the distribution of contact stresses reaches a quasi-stationary state within very short rolling distances, on the order of few times the longitudinal contact patch dimension, as shown by experience [Kalker 1979b]. On the other hand, transient contact models may be necessary in the study of high-frequency phenomena such as may be encountered with short-pitch corrugation, see e.g. [Baeza 2007], [Baeza 2011].
- Quasiidentity (cf. Figure 1.10). This is justified because the material of both wheel and rail is steel.
- The rigid slip velocity field is described with the creepages according to Eq. (1.5), corresponding to a planar contact area.
- In some contact models, Hertzian geometry is assumed. I.e., the contact patch is assumed to be an ellipse, aligned with the contact x and y directions. This assumption, though directly related to the normal contact problem, also has implications in the tangential problem solution.

The models revised in the following subsections adopt all these simplifying assumptions, unless otherwise stated. All these models, except for Kalker's linear theory, are partial slip models, which imply that they consider the

coexistence of adhesion and slip areas in the contact patch, in addition to taking into account the effect of the contact surface deformations.

2.1.2.1 Carter / Fromm's theory

Carter and Fromm's theory ([Carter 1926], [Fromm 1926]) is an exact theory for 2D Hertzian steady rolling contact with similar materials, in the presence of a longitudinal creepage and Coulomb friction. In the 2D case, the number of unknowns of the tangential problem is the same as in the normal problem, and the tangential contact conditions are linear. Therefore, the solution of the 2D tangential problem presents less complication than in the 3D case. Additionally, there is an analytical solution for the steady-state contact with similar materials, given by this theory.

The distribution of longitudinal stress in the contact p_x is found superposing two elliptical distributions p_{xi}, of the form given in Eq. (2.48), as represented in Figure 2.9. The field of tangential displacement differences u_{xi} caused by each of these distributions in a pair of quasiidentical half-spaces in plane strain has the analytical solution given in Eq. (2.49) within the domain where each distribution is applied, i.e., in $|x - x_{oi}| < a_i$. In these equations, τ_{oi} is a maximum value of tangential stress, x_{oi} the point where the stress distribution is centred, and a_i the half-width (in x direction) of the stress distribution. The subindex i refers to each stress distribution i, $i \in \{1, 2\}$. In Eq. (2.49), $const_i$ is a constant term for each stress distribution, and E^* is the combined effective Young's modulus of the two bodies (cf. Eq. (2.11)), which is equal to $E / [2(1 - \nu^2)]$ for elastically similar materials in contact.

$$p_{xi} = \tau_{oi}\sqrt{1-\frac{\left(x-x_{oi}\right)^2}{a_i^{\,2}}},\ \ \left|x-x_{oi}\right| < a_i \tag{2.48}$$

$$u_{xi} = const_i - \frac{\tau_{oi}}{a_i E^*}\left(x-x_{oi}\right)^2,\ \ \left|x-x_{oi}\right| < a_i \tag{2.49}$$

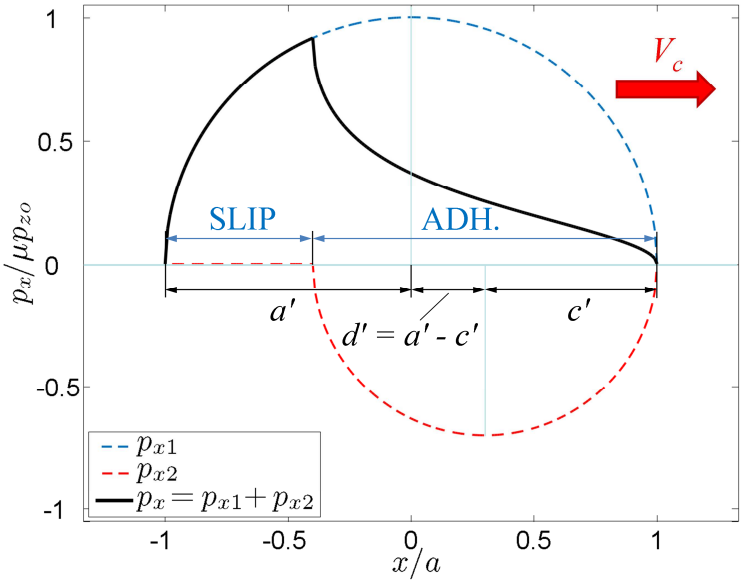


Figure 2.9. Tangential contact stresses in Carter's 2D tangential contact solution.

The V_c velocity in Figure 2.9 is the velocity at which the contact moves on the contacting surfaces. This coincides approximately with the forward velocity of the wheel V in the case of a wheel running on a straight rail.

According to the kinematic equation for the tangential contact (Eq. (1.7)) particularized to steady contact with only longitudinal creepage, in the adhesion region ($\mathbf{v} = 0$), the condition $\xi = \partial u_x / \partial x$ must be fulfilled. As ξ is constant in the whole contact, the quadratic terms with x of Eq. (2.49) must vanish in the adhesion region when superposing the effect of the two stress distributions. This is fulfilled when the ratio $|\tau_{oi}| / a_i$ is the same in both p_{xi} distributions. τ_{o1} is determined from the saturation condition in the slip region, i.e., $p_x = p_{x1} = g$, being g the traction bound, $\mu\, p_z$, and μ the coefficient of friction. Taking this into account, $\tau_{o1} = \mu\, p_{zo}$, being p_{zo} the maximum normal contact pressure. The position of the adhesion zone, adjacent to the leading edge, is the only possible one that complies with the condition that the slip is in the opposite direction to the tangential stress in the slip region. The remaining condition for the determination of the necessary parameters of the tangential stress distribution is given by the equality $\xi = \partial u_x / \partial x$.

Taking all this into account, it is verified that the contact conditions and constraints of the adhesion and slip regions are satisfied with the $p_x = p_{x1} + p_{x2}$ distribution with the following parameters: $\tau_{o1} = \mu\ p_{zo}$; $x_{o1} = 0$; $a_1 = a$; $\tau_{o2} = -c/a\,\mu\ p_{zo}$; $a_2 = c$; and $x_{o2} = d = a - c$. a and c are the half-widths of the contact patch and of the adhesion zone, respectively. d is the distance from the centre of the contact patch to the centre of the adhesion zone (not to be confused with the normal approach, also designated as d). With these values, the solution of p_x is given in Eq. (2.50). The primed dimensions a', c', and d', and the coordinate x' are obtained dividing the corresponding non-primed quantities by a.

$$-\operatorname{sgn}(\xi)\times\frac{p_x}{\mu\ p_{zo}}$$

$$=\begin{cases}\sqrt{1-x'^2}-c'\sqrt{1-\left(\dfrac{x'-(1-c')}{c'}\right)^2} & \text{for } 1-2c'<x'\le 1\\ \sqrt{1-x'^2} & \text{for } -1\le x'\le 1-2c'\end{cases} \tag{2.50}$$

The relationship between ξ and the dimensions of the adhesion and slip zones is given by Eq. (2.51). The sign of ξ is chosen so that the rigid slip velocity is in the opposing direction to p_x.

$$\frac{\partial u_x}{\partial x}=\xi=-\frac{2\mu\ p_{zo}d}{aE^*} \tag{2.51}$$

The integration of the p_x distribution of Eq. (2.50) in the contact yields the resultant longitudinal force in the contact F_x, according to Eq. (2.52), where N is the total normal load. The c / a proportion may be solved for in this equation, with the result given in Eq. (2.53).

$$F_x=\mu N\left(1-\frac{c^2}{a^2}\right) \tag{2.52}$$

$$\frac{c}{a} = \sqrt{1 - \frac{F_x}{\mu N}} \tag{2.53}$$

Taking into account the previous equations and Eq. (2.51), the creepage–creep force relationship is given in Eq. (2.54) and plotted in Figure 2.10 in normalized axes. In this equation, R is the combined longitudinal curvature radius. Figure 2.10 includes the asymptote of the curve for small creepages. The slope of this asymptote, which is independent of μ, is given in Eq. (2.55). The saturation (full slip) condition is reached with a creepage equal to twice the value predicted with this initial slope.

$$\xi = -K\left\{1 - \sqrt{1 - \frac{F_x}{\mu N}}\right\};\ K = \frac{2\mu\, p_{zo}}{E^*} = \frac{\mu\, a}{R} \tag{2.54}$$

$$\left.\frac{\partial F_x}{\partial \xi}\right|_{\xi \to 0} = \frac{2NR}{a} \tag{2.55}$$

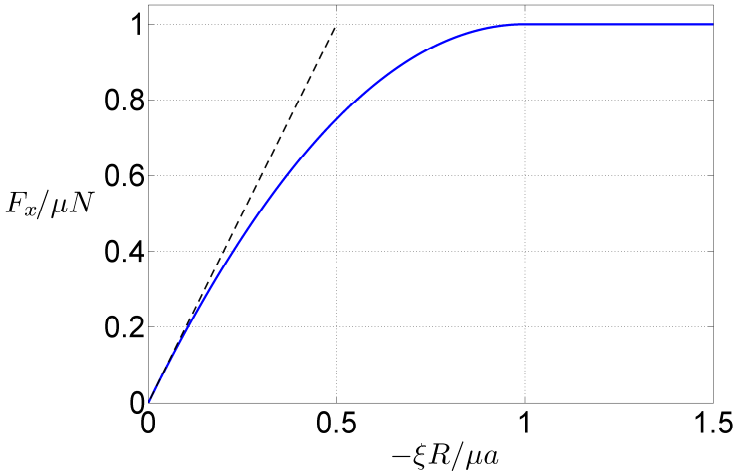


Figure 2.10. Creepage–creep force curve of Carter's rolling contact problem.

To obtain the slip velocities in the slip region, the elastic deformations caused by an ellipsoidal p_x distribution of the form given in Eq. (2.48) are needed. Eq. (2.49) gives the elastic displacements caused by such stress distribution inside

its application region. The displacements outside the application region of p_x are obtained making the convolution of the p_x stress distribution with the Flamant IF [Flamant 1892] of the horizontal surface displacements of the elastic half-plane under tangential loading. The displacement gradients are obtained using the derivative of the mentioned Flamant IF with respect to x. This convolution is expressed in Eq. (2.56). In this equation, an ellipsoidal p_x distribution with a maximum value of τ_o is assumed, applied in a region of half-width a centred at $x = 0$. The analytical solution of this equation is given in Eq. (2.57). Here, the change of variable $X' = X / a$ is used.

$$\frac{\partial u_x}{\partial x} = -\frac{2\tau_o}{\pi E^*}\int_{-a}^{a}\frac{\sqrt{1-\left(\frac{X}{a}\right)^2}}{x-X}dX \tag{2.56}$$

$$-\frac{2\tau_o}{\pi E^*}\int_{-1}^{1}\frac{\sqrt{1-X'^2}}{x-aX'}adX' = -\frac{2\tau_o}{\pi E^*}\int_{-1}^{1}\frac{\sqrt{1-X'^2}}{\frac{x}{a}-X'}dX'$$

$$= -\frac{2\tau_o}{\pi E^*}\left[-\sqrt{\left(\frac{x}{a}\right)^2-1}\,\operatorname{asin}\left(\frac{\frac{x}{a}X'-1}{\left|X'-\frac{x}{a}\right|}\right)+\frac{x}{a}\operatorname{asin}(X')-\sqrt{1-X'^2}\right]_{-1}^{1}$$

$$= -\frac{2\tau_o}{E^*}\left(\frac{x}{a}-\operatorname{sgn}\left(\frac{x}{a}\right)\sqrt{\left(\frac{x}{a}\right)^2-1}\right) \tag{2.57}$$

Now, the u_x displacement gradients caused by the p_{x1} and the p_{x2} stress distributions in the whole contact patch may be written, using Eq. (2.57) and the derivative with respect to x of Eq. (2.49), appropriately modified taking into account the centre, width and maximum value of each stress distribution. The gradients caused by p_{x1} are given in Eq. (2.58), and the ones caused by p_{x2} in Eq. (2.59). The subindices 1 and 2 in these equations are not to be interpreted as related to bodies 1 and 2 (what is being expressed are combined quantities of

both bodies), but as related to p_{x1} and p_{x2}, respectively. The first line of Eq. (2.59) corresponds to the adhesion zone, and the second one to the slip zone.

$$\frac{\partial u_{x1}}{\partial x} = -\frac{2\mu\ p_{zo}}{E^*}\frac{x}{a}, \quad \text{for } \left|\frac{x}{a}\right| \le 1 \tag{2.58}$$

$$\frac{\partial u_{x2}}{\partial x} = \begin{cases} \dfrac{2c\mu\ p_{zo}}{aE^*}\dfrac{(x-d)}{c}, & \text{for } \left|\dfrac{x-d}{c}\right| \le 1 \\ \dfrac{2c\mu\ p_{zo}}{aE^*}\left(\dfrac{x-d}{c} + \sqrt{\left(\dfrac{x-d}{c}\right)^2 - 1}\right), & \text{for } x < d-c \end{cases} \tag{2.59}$$

The displacement gradients given in Eqs. (2.58) and (2.59) are plotted in Figure 2.11, together with the total displacement gradient, for c / a = 0.70 (the same case for which the p_x distribution is shown in Figure 2.9). The K' value represented here is equal to K / a, being K defined in Eq. (2.54).

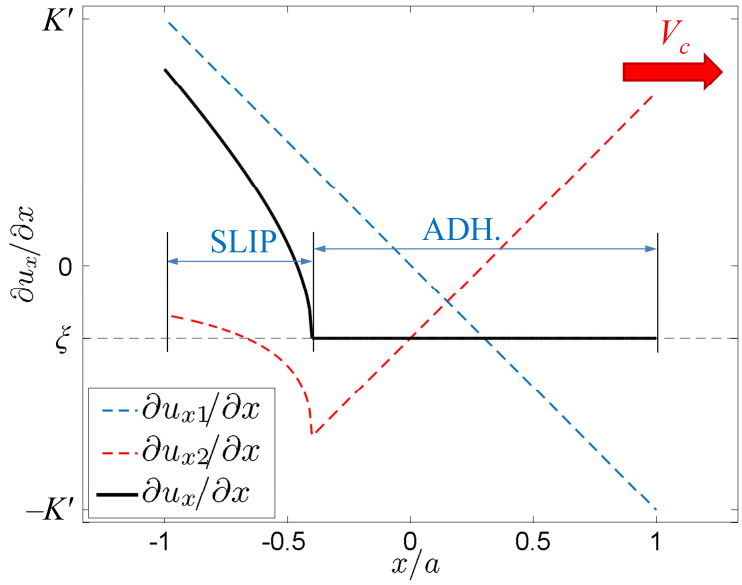


Figure 2.11. Longitudinal displacement gradients in Carter / Fromm's 2D tangential contact solution.

The slip velocity is obtained applying the kinematical condition of the tangential problem (v_x / V_c = ξ – ∂u_x / ∂x) in the slip zone, and replacing the ξ

value from Eq. (2.51) and the displacement gradients from Eqs. (2.58) and (2.59). The slip velocity, of course, is 0 in the adhesion region. The result for the slip region is given in Eq. (2.60). This expression is valid only if there is some adhesion. An alternative form of this result is given in Eq. (2.61), with the normalized creepage ξ^* being equal to ξ / K. ξ^* is equal to d' (with the appropriate sign) in partial slip.

$$\frac{v_x}{V_c} = -K'\sqrt{(x-d)^2 - c^2} \tag{2.60}$$

$$\mathrm{sgn}(\xi) \times \frac{v_x}{V_c}$$

$$= \begin{cases} 0 & \text{for } 1-2c' < x' \le 1 \\ \xi\left(1 + \frac{1}{\xi^*}\left(c' - 1 + \sqrt{(x' - (1-c'))^2 - c'^2}\right)\right) & \text{for } -1 \le x' \le 1-2c' \end{cases} \tag{2.61}$$

2.1.2.2 Johnson; Vermeulen & Johnson

Johnson presented the first 3D tangential contact model. It was aimed to circular contacts with pure creepage [Johnson 1958b] or with spin [Johnson 1958a]. This was extended later for elliptical contacts without spin by Vermeulen and Johnson in [Vermeulen 1964].

The model is based on an extrapolation of Carter / Fromm's 2D theory (cf. §2.1.2.1) to 3D, assuming an elliptical adhesion area with the same shape as the contact patch and tangent to the leading edge, as represented in Figure 2.12. In a similar way as in Carter / Fromm's theory, the tangential stress distribution is composed of two ellipsoidal stress distributions with contrary sign, one of them (the biggest in magnitude) acting in the whole contact patch, and the other one in the adhesion region. The assumed adhesion area is only approximate. The condition of opposing directions of the slip velocity and tangential stress vectors is not fulfilled in the shadowed region adjacent to the leading edge marked in Figure 2.12.

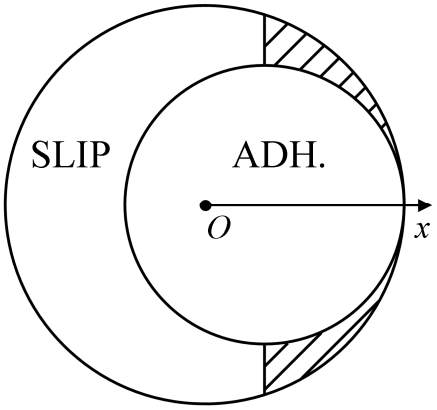


Figure 2.12. Assumed distribution of adhesion and slip areas in the contact patch in the Johnson and Vermeulen & Johnson tangential contact models.

The model has the advantage of leading to simple closed-form expressions for the creepage–creep force curve, as well as for the distribution of tangential stresses. The main results of the model are revised next. Normalized creepages ξ' and η' are defined according to Eq. (2.62). In this equation, G is the shear modulus, and C_{11} and C_{22} are non-dimensional creepage coefficients, for which expressions are provided in [Vermeulen 1964]. A resultant normalized creepage ζ' is defined as $\zeta' = \|(\xi', \eta')\|$. The modulus of the resultant tangential contact force, $\|\mathbf{F_t}\|$, is defined as a function of ζ', according to Eq. (2.63). $\mathbf{F_t}$ has the same direction as the normalized creepage vector (ξ', η').

$$(\xi', \eta') = -\frac{(C_{11}\xi, C_{22}\eta)\, abG}{3\mu N} \tag{2.62}$$

$$\frac{\mathbf{F_t}}{\mu N} = \begin{cases} \left[1-(1-\zeta')^3\right] & \text{for } w' < 1 \\ 1 & \text{for } w' \geq 1 \end{cases} \tag{2.63}$$

2.1.2.3 Strip theory

The strip theory for contact mechanics was set out in [Haines 1963] and in [Halling 1964]. It consists on replacing the 3D contact mechanics problem for a set of 2D contact problems. For this purpose, the contact patch is divided in

longitudinal strips, and Carter / Fromm's theory is applied in each strip. The basic assumption of this model is that there is not interaction between different strips of the contact patch. This works well for wide contact patches (in the direction transverse to rolling).

The theory was originally set out for pure longitudinal creepage, and in [Kalker 1967b] it was extended for lateral creepage and small spin. The former case is revised here. The same creepage is considered in the whole contact patch. The p_{zo} / a ratio between the maximum normal pressure and longitudinal dimension of each strip is also the same in the whole contact patch, according to Hertzian proportionality. Then, according to Eq. (2.51), the distance d from the centre of the contact patch to the centre of the adhesion zone must be the same in the whole contact patch. Hence, the trailing edge of the adhesion zone is a reflection of the leading edge about the straight line crossing the contact patch at $x = d$, as illustrated in Figure 2.13. The most important advance of this theory is to be the first 3D contact model that properly approximates the correct form of the adhesion zone.

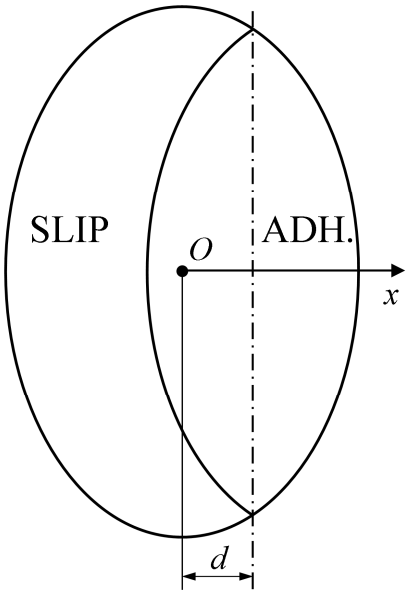


Figure 2.13. Assumed distribution of adhesion and slip areas in the contact patch in the strip theory.

2.1.2.4 Kalker's linear theory

Kalker's linear theory, developed in [Kalker 1967a], is the first exact theory for 3D tangential contact. It is applicable for elliptical contact patches with infinitesimally small creepages (which may be a combination of longitudinal, lateral and spin creepage) or with infinite coefficient of friction, in which the whole contact patch is in adhesion. The theory is useful in linear analyses, and to calibrate simplified non-linear models.

The development of the theory starts extracting the equations of the elastic displacement field in the contact patch, based on the kinematic equation of the tangential problem. The problem is reduced to its elasticity part, seeking for the stress distribution that leads to the previously formulated displacement field. The problem is solved combining stress distributions of polynomial form (truncating in a finite number of terms), which produce known displacement fields in the elliptical contact area (polynomial as well), and equating to 0 the stresses in a finite number of points on the leading edge of the contact patch.

As an outcome of the theory, linear relationships between the resultant contact forces and creepages are obtained. These relationships are expressed in matrix form in Eq. (2.64).

$$\begin{Bmatrix} F_x \\ F_y \\ M_z \end{Bmatrix} = -Gab \begin{bmatrix} C_{11} & 0 & 0 \\ 0 & C_{22} & \sqrt{ab}\, C_{23} \\ 0 & -\sqrt{ab}\, C_{23} & ab\, C_{33} \end{bmatrix} \begin{Bmatrix} \xi \\ \eta \\ \varphi \end{Bmatrix} \tag{2.64}$$

C_{ij} in Eq. (2.64) are dimensionless creepage coefficients, which depend on the eccentricity of the contact ellipse and on the coefficient of Poisson ν. They may be found tabulated e.g. in [Kalker 1967a], [Kalker 1990], [Garg 1984] and [Ayasse 2006] for ν values of 0, 0.25 and 0.5. They are also tabulated and plotted for ν = 0.30 in [Johnson 1987] and in [Thompson 2009], respectively, and a polynominal fit for the C_{11}, C_{22} and C_{23} coefficients for ν = 0.27 is proposed in [Ayasse 2006]. In some references (e.g. in [Kalker 1991]), the C_{33} coefficient is not tabulated, due to its reduced relevance in usual wheel–rail interaction problems.

As may be seen in Eq. (2.64), the longitudinal creepage only produces longitudinal resultant force, and the other creepages do not contribute to the longitudinal resultant force. This is due to the symmetries of the elliptical contact patch and of the rigid slip velocity field. In general, this is not the case for non-elliptical contact patches.

The following analogies with the tyre–road case may be made of the terms related to the C_{23} coefficient:

- The contribution of the spin to the lateral thrust is analogous to the camber thrust.
- The contribution of the lateral creepage to the spin moment is analogous to the self-aligning torque, which the driver may feel in the steering wheel. In the wheel–rail application it is sometimes neglected, being on the order of tens of N.m.

Eq. (2.64) may also be expressed in normalized form, normalizing both the creepages and the contact resultant forces. The normalized longitudinal and lateral creepages, ξ' and η', are given in Eq. (2.62), and the normalized spin, φ', is given in Eq. (2.65).

$$\varphi' = -\frac{\sqrt{(ab)^3}\,G\,C_{23}}{\mu N}\varphi \tag{2.65}$$

$$m_z = \frac{M_z}{\mu N\sqrt{ab}} \tag{2.66}$$

The normalized contact resultant forces, f_x and f_y, are calculated dividing the corresponding forces by μN. The normalized spin moment, m_z, is calculated according to Eq. (2.66). Eq. (2.67) is the normalized form of Eq. (2.64).

$$\begin{Bmatrix} f_x \\ f_y \\ m_z \end{Bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & -3C_{23}/C_{22} & C_{33}/C_{23} \end{bmatrix} \begin{Bmatrix} \xi' \\ \eta' \\ \varphi' \end{Bmatrix} \tag{2.67}$$

2.1.2.5 Shen-Hedrick-Elkins' heuristic method

The Shen-Hedrick-Elkins' heuristic method [Shen 1983] is a simple saturation law applied on Kalker's linear theory. It is summarized in Eqs. (2.68) and (2.69). In this equation f^l is the normalized contact resultant force obtained with Kalker's linear theory, i.e., $f^l = F^l / (\mu N)$, with $F^l = \|(F^l_x, F^l_y)\|$, being F^l_x and F^l_y the contact resultant longitudinal and lateral forces computed with Eq. (2.64).

$$\varepsilon = \begin{cases} f^l - \dfrac{f^{l2}}{3} + \dfrac{f^{l3}}{27} & \text{for } f^l \le 3 \\ 1 & \text{for } f^l > 3 \end{cases} \tag{2.68}$$

$$\left(F_x, F_y\right) = \varepsilon \frac{\mu N}{F^l}\left(F^l{}_x, F^l{}_y\right) \tag{2.69}$$

This model is similar to the Vermeulen & Johnson model, but with correct C_{ij} creepage coefficients, and taking into account the influence of the spin. The main advantage of the model is its simplicity. Its main drawback is its precision loss with high spin values (as may be encountered when there is contact in the wheel flange, which may happen during negotiation of tight curves), not being able to predict the decrease of the lateral force with high spin values.

2.1.2.6 Simplified theory of rolling contact. FASTSIM

The FASTSIM algorithm [Kalker 1982] of Kalker's simplified rolling contact theory is a widely used tangential contact model in the wheel–rail application. It is able to handle situations with arbitrary combinations of creepages with satisfactory precision, including situations with high spin, while having a low enough computational cost for practical application in rail vehicle dynamic simulations.

The simplification of this model consists in assuming that the elastic tangential displacement in each point in the contact patch depends only on the contact tangential stress at that point. Thus, the tangential behaviour of the contacting surfaces is considered equivalent to that of a bed of springs or 'wire brush' model, in which the individual bristles deform independently from one another. This is expressed in Eq. (2.70), and represented schematically in Figure 2.14. In

this equation, $\mathbf{u}_t$ and $\mathbf{p}_t$ are the tangential elastic displacement and stress vectors, respectively, and L is a tangential surface compliance.

$$\mathbf{u}_t(x, y) = L\,\mathbf{p}_t(x, y) \tag{2.70}$$

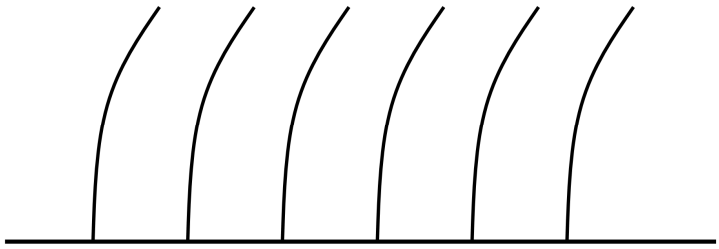


Figure 2.14. 'Wire brush' tangential model of the contacting surfaces used in the simplified theory of rolling contact. Adapted from [Kalker 2001].

The kinematic equation of the tangential problem (Eq. (1.7)) is discretized in time according to Eq. (2.71). In this equation, **u** and **u'** are the elastic displacements of a particle in the contacting surface at the current and previous time instants, t and t', respectively, being $t - t' = \Delta t$. The subindex t (of tangential) in **u** is omitted here for brevity.

$$\mathbf{v} = \mathbf{w} + \frac{\mathrm{D}\mathbf{u}}{\mathrm{D}t} \approx \mathbf{w} + \frac{\mathbf{u} - \mathbf{u}'}{\Delta t} \tag{2.71}$$

For a particle occupying position $P = (x, y)$ at t, $\mathbf{u} = \mathbf{u}(x, y, t)$. The (x, y) coordinates are given in the Eulerian contact coordinate system described in §1.2. Considering the flow of material through this system (cf. Figure 1.9b), and considering the surface elastic displacements much smaller than the position change of the particle in the Δt time interval, the position P' of the same particle at t', is approximately $P' \approx (x + V\,\Delta t, y)$. I.e., from t' to t, the particle displaces approximately a distance of $V\,\Delta t$ along the negative x axis. So, the displacement **u'** of the same particle may be written as $\mathbf{u}' \approx \mathbf{u}(x + V\,\Delta t, y, t - \Delta t)$. The time dependence vanishes in steady-contact, so $\mathbf{u} = \mathbf{u}(x, y)$, and $\mathbf{u}' \approx \mathbf{u}(x + V\,\Delta t, y)$.

The tangential surface compliance L of Eq. (2.70) is determined so that the results of the simplified theory for vanishing creepages coincide with those of

the linear theory (cf. §2.1.2.4), as outlined next. Eq. (1.7) for no slip ($\mathbf{v} = 0$) yields $\partial\mathbf{u} / \partial x = \mathbf{w} / V = (\xi - \varphi\, y,\ \eta + \varphi\, x)$ (cf. also Eq. (1.5)). This is integrated with respect to x, obtaining Eq. (2.72). In this equation $k(y)$ and $l(y)$ are integration constants. They are determined in Eq. (2.73), imposing the no traction condition ($\mathbf{p_t} = \mathbf{0}$) at the leading edge, located at $(x, y) = (x_l(y), y)$. Here, use is made of the tangential constitutive relation of the simplified theory (Eq. (2.70)), which implies that $\mathbf{u_t} = \mathbf{0}$ where $\mathbf{p_t} = \mathbf{0}$.

$$\begin{cases} u_x = \xi\, x - \varphi\, y\, x + k(y) \\ u_y = \eta\, x + \varphi\ x^2/2 + l(y) \end{cases} \tag{2.72}$$

$$\begin{cases} k(y) = -\xi\, x_l(y) + \varphi\, y\, x_l(y) \\ l(y) = -\eta\, x_l(y) - \varphi\ x_l(y)^2/2 \end{cases} \tag{2.73}$$

Thus, Eqs. (2.72) and (2.73) determine the tangential elastic displacement field $\mathbf{u_t}$. From this, the tangential stress field is determined using Eq. (2.70), as $\mathbf{p_t} = \mathbf{u_t} / L$. The contact resultant forces are determined integrating $\mathbf{p_t}$ in the contact ellipse (of longitudinal and lateral semiaxes a and b, respectively), as indicated in Eq. (2.74). The results are given in Table 2.1, together with the results of the linear theory.

$$\left(F_x, F_y\right) = \int_{-b}^{b} \int_{-a(y)}^{a(y)} \mathbf{p_t}\, dx\, dy \tag{2.74}$$

Table 2.1. Tangential contact resultant forces according to the simplified theory and to the linear theory.

	Simplified theory	**Linear theory**
F_x	$-\dfrac{8a^2b}{3L}\xi$	$-abGC_{11}\,\xi$
F_y	$-\dfrac{8a^2b}{3L}\eta - \dfrac{\pi\, a^3b}{4L}\varphi$	$-abGC_{22}\eta - (ab)^{1.5}GC_{23}\varphi$

L is determined equating the results of both theories given in Table 2.1. However, different L values result when considering the terms associated to the different creepages (ξ, η, or φ), as indicated in Eq. (2.75). A solution is to

calculate a weighted L value with the three possible L values given in this equation, as indicated in Eq. (2.76). Another possibility is to use different L values in x and y directions. This would still leave two L values to choose from for the y direction (the ones associated to the η and φ creepages). An additional drawback of using multiple L values is that the tangential stress and slip vectors are no longer parallel, as may be seen from the explanation that follows. Anyway, some versions of the simplified theory use 3 independent L values, each associated to a different creepage. In this way, better results may be obtained for small creepages. On the other hand, the saturation part of the curves (for large creepages) is not reproduced correctly with multiple L values, and this is achieved with a single L.

$$\begin{cases} \xi \to L_1 = 8a/(3GC_{11}) \\ \eta \to L_2 = 8a/(3GC_{22}) \\ \varphi \to L_3 = \pi\, a^2/\left(4G\sqrt{ab}C_{23}\right) \end{cases} \tag{2.75}$$

$$L = \frac{L_1|\xi| + L_2|\eta| + L_3|\varphi|\sqrt{ab}}{\sqrt{\xi^2 + \eta^2 + ab\varphi^2 + 10^{-20}}} \tag{2.76}$$

Using the described tangential constitutive model, the contact solution is obtained numerically, in a 2D grid of rectangular elements organized in parallel strips aligned with the rolling direction, as shown in Figure 2.15. The length of each strip of the mesh is equal to the length of the corresponding longitudinal strip of the contact ellipse, and each strip is divided into the same number of equal rectangular elements. Therefore, the length of the elements of the different strips is variable. In the same way, Δt is variable across the contact, changing in proportion to the length of the mesh elements.

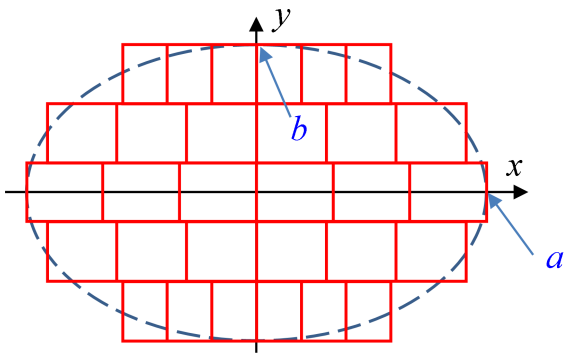


Figure 2.15. Mesh of the contact ellipse of the FASTSIM algorithm.

The tangential stresses are computed along each strip in a recursive, non-iterative way, moving from the leading to the trailing edge. Therefore, it is still much faster than exact contact models. This calculation is done applying Eq. (2.77), obtained inserting the relationship of the simplified model given in Eq. (2.70) on the kinematic equation (2.71). In Eq. (2.77), $\mathbf{p'}_t$ is the tangential stress at instant t'. $\mathbf{p'}_t$ is $\mathbf{0}$ at the leading edge.

$$\mathbf{v} \approx \mathbf{w} + L\frac{\mathbf{p}_t - \mathbf{p'}_t}{\Delta t} \rightarrow \mathbf{p}_t = \mathbf{p'}_t + \frac{\Delta t}{L}(\mathbf{v} - \mathbf{w}) \tag{2.77}$$

The algorithm starts assuming adhesion ($\mathbf{v} = \mathbf{0}$) in the first element of each strip, and calculating a tentative adhesion tangential stress, $(\mathbf{p}_t)_A$, according to Eq. (2.78). The adhesion condition ($\|\mathbf{p}_t\| \leq g$) is checked with $(\mathbf{p}_t)_A$. If $\|(\mathbf{p}_t)_A\| \leq g$, the adhesion assumption is correct, and $\mathbf{p}_t = (\mathbf{p}_t)_A$. Otherwise (if $\|(\mathbf{p}_t)_A\| > g$), the element is in slip, and its $\mathbf{p}_t$ is calculated scaling down $(\mathbf{p}_t)_A$ so that $\|\mathbf{p}_t\| = g$. I.e., $\mathbf{p}_t = (g / \|(\mathbf{p}_t)_A\|)\,(\mathbf{p}_t)_A$.

$$(\mathbf{p}_t)_A = \mathbf{p'}_t - \frac{\Delta t}{L}\mathbf{w} \tag{2.78}$$

For elements in slip, Eq. (2.79) demonstrates that the slip condition that the slip velocity vector is in the opposite direction to the tangential stress vector is verified exactly only if a single L value is used. For an element in slip, as $\|(\mathbf{p}_t)_A\| > g$, the $(1 - g / \|(\mathbf{p}_t)_A\|)$ factor in the right-hand side of this equation is positive.

$$\mathbf{v} \approx \mathbf{w} + L\frac{\mathbf{p}_t - \mathbf{p}'_t}{\Delta t} = \frac{L}{\Delta t}\left(-\left(\mathbf{p}_t\right)_A + \mathbf{p}_t\right) = -L\frac{\left(\mathbf{p}_t\right)_A}{\Delta t}\left(1 - \frac{g}{\left\|\left(\mathbf{p}_t\right)_A\right\|}\right) \tag{2.79}$$

The adhesion limit g used in this solution of the tangential problem is based on an assumed parabolic distribution of normal pressures, according to Eq. (2.80). The maximum normal pressure, $p_{z,o}$, is equal to $2N / (\pi ab)$ in this case. With this, there is an analogy between the normal and the tangential contact problems, assuming in both proportionality between the elastic displacements and contact stresses. Experience shows that considering the true Hertzian (ellipsoidal) normal pressure distribution to compute g leads to worse results of the simplified theory for tangential contact, especially with regards to predicting the correct limits between the adhesion and slip zones. The normal contact problem is otherwise solved with Hertzian theory. The direct application of the simplified theory for the normal contact problem does not work well.

$$p_z(x,y) = p_{z,o}\left(1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}\right) \tag{2.80}$$

FASTSIM achieves generally good precision in the contact force resultants, usually with errors below 15% with respect to the exact contact theory (§2.1.3), while being much faster. In this respect, FASTSIM has been considered to provide the best precision / cost compromise for rail vehicle dynamic simulations for many years.

Though commonly used for planar contacts, the FASTSIM algorithm may be readily extended to non-planar contacts assigning a different orientation to each strip, provided the assumption of half-space-like elastic behaviour and elastic similarity of the contacting bodies is retained. Also, owing to the contact patch division into strips, it may be readily extended to non-elliptical contact patches as well, the main difficulty lying in the calculation of proper L parameters. These are some of the extensions set out for FASTSIM:

- Non-elliptical contact patches: this is used for the tangential contact problem in [Kik 1996] and in [Ayasse 2005], with the normal contact

models explained in §2.1.1.2 and §2.1.1.3, respectively. Another extension for non-elliptical contact patches is also set out in [Alonso 2007].

- Variable coefficient of friction: the extension of FASTSIM for a friction coefficient variable with the slip velocity is introduced in [Giménez 2005].
- Transient rolling: a way to solve transient rolling contact problems with FASTSIM is proposed in [Guiral 2013].

2.1.2.7 Polach's method

Polach's method [Polach 1999] is based on integrating in the contact patch tangential stress distributions similar to those computed with FASTSIM, but with ellipsoidal adhesion limit g, as depicted in Figure 2.16 for a case without spin. The method is aimed at obtaining contact resultant forces, rather than stress distributions inside the contact patch.

The integration in the contact ellipse of a stress distribution like the one shown in Figure 2.16 yields the resultant force vector of Eq. (2.81). The first term inside the parenthesis in the right-hand side of this equation is associated to the adhesion area, and the second one to the slip area. In this equation, ζ_c is a corrected effective creepage, equal to $\|(\xi, \eta_c)\|$. η_c is a corrected lateral creepage; $\eta_c = \eta + \varphi\ a$ if $|\eta + \varphi\ a| > |\eta|$, $\eta_c = \eta$ otherwise. ε, which is equal to the longitudinal gradient of the resultant tangential stress in the adhesion region, is given in Eq. (2.82), with the weighted tangential contact rigidity C given in Eq. (2.83). ζ in Eq. (2.83) is equal to $\|(\xi, \eta)\|$. The rest of the symbols in these equations are as previously defined.

$$\left(F_{x,1}, F_{y,1}\right) = \frac{2N\mu}{\pi}\left(\frac{\varepsilon}{1+\varepsilon^2} + \operatorname{atan}\varepsilon\right) \times \frac{(\xi,\eta)}{|\zeta_c|} \tag{2.81}$$

$$\varepsilon = \frac{2C\pi\, a^2 b}{3N\mu}\zeta_c \tag{2.82}$$

$$C = \sqrt{\left(\frac{1}{L_1}\frac{\xi}{\zeta}\right)^2 + \left(\frac{1}{L_2}\frac{\eta}{\zeta}\right)^2} \tag{2.83}$$

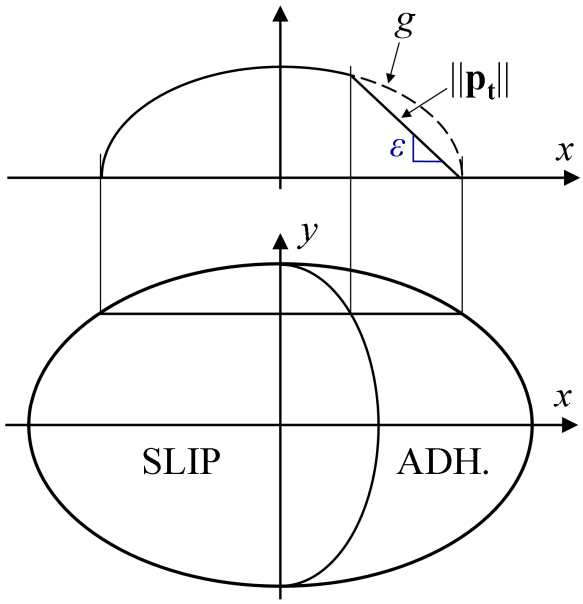


Figure 2.16. Adhesion and slip areas and tangential stress distribution in Polach's method for a rolling case with pure creepage. Adapted from [Polach 1999].

Eq. (2.81) gives the components of the resultant tangential force associated to the ξ and η creepages. On the other hand, the component of the lateral force due to the spin, $F_{y,\varphi}$, is calculated according to Eq. (2.84). This equation results from the analytical solution for vanishing longitudinal dimension of the contact patch ($a \rightarrow 0$), corrected empirically with the factor R given in Eq. (2.85). K_M is given in Eq. (2.86), and the factors ε_φ and δ of this equation are given in Eqs. (2.87) and (2.88), respectively, with $C_\varphi = 1 / L_3$ in Eq. (2.87).

$$F_{y,\varphi} = -\frac{9}{16} aN\mu K_M R \frac{\varphi}{\zeta_c} \tag{2.84}$$

$$R = \left[1 + 6.3\left(1 - \mathrm{e}^{-\frac{a}{b}}\right)\right] \tag{2.85}$$

$$K_M = \left|\varepsilon_\varphi\right|\left(\frac{\delta^3}{3} - \frac{\delta^2}{2} + \frac{1}{6}\right) - \frac{1}{3}\sqrt{\left(1-\delta^2\right)^3} \tag{2.86}$$

$$\varepsilon_\varphi = \frac{2\ C_\varphi\ \pi\ a^2 b}{3N\mu}\frac{\eta_c}{R} \tag{2.87}$$

$$\delta = \frac{\varepsilon_\varphi - 1}{\varepsilon_\varphi + 1} \tag{2.88}$$

The components of the resultant tangential force are calculated as $F_x = F_{x,1}$, and $F_y = F_{y,1} + F_{y,\varphi}$.

Polach's method provides the resultant tangential forces through analytical expressions, as Shen-Hedrick-Elkins' heuristic method, so it is very fast. A notable advance of Polach's method is that it is able to predict the decay of the lateral contact force at high spin values. On the other hand, the results for certain combinations of creepages are less precise than those of Shen-Hedrick-Elkins' method.

An extension of Polach's method for variable friction coefficient with the slip velocity was presented in [Polach 2005].

2.1.2.8 FASTRIP

Sichani's FASTRIP method ([Sichani 2016a], [Sichani 2016b]) is a combination of the strip theory to predict the non-linear growth of tangential stress along the x direction in the adhesion area, with a modified version of FASTSIM to determine the tangential stress direction in the slip area (the latter being necessary in cases with spin).

The contact patch is discretized in longitudinal strips, as in the strip theory and FASTSIM. The tangential stress distribution along each strip is computed as the sum of two elliptical stress distributions of opposing signs (similar to the

distribution shown in Figure 2.9), according to a generalized version of the strip theory for combined creepage and a not necessarily much lower than b. This generalization is done in such a way that the results coincide with those of Kalker's linear theory for small creepages. The components of tangential contact stress p_j, $j \in \{x, y\}$, are calculated according to Eq. (2.89). The subindex i in this equation refers to strip i. $p_{z,o,i}$ is the maximum normal pressure in each strip, and a_i the strip half-length. The factors $k_{j,i}$ and $k'_{j,i}$ are defined in Eq. (2.94) and (2.95), respectively.

$$p_j = \frac{\mu\, p_{z,o,i}}{a_i}\left(k_{j,i}\sqrt{{a_i}^2 - x^2} - {k'}_{j,i}\sqrt{(a_i - d_i)^2 - (x - d_i)^2}\right) \tag{2.89}$$

d_i, the longitudinal distance from the centre of the contact patch to the centre of the adhesion region of strip i, is calculated according to Eq. (2.90). η^* and φ^* are non-dimensional lateral and spin creepages, defined in Eqs. (2.91) and (2.92), respectively, together with the non-dimensional longitudinal creepage ξ^*. Ξ is a non-dimensional longitudinal slip, which is defined together with the non-dimensional lateral slip H in Eq. (2.93).

$$d_i = \frac{\sqrt{{\eta^*}^2 + \left(1 - {\varphi^*}^2\right)\Xi^{*2}} + \eta^*\varphi^*}{\left(1 - {\varphi^*}^2\right)(1 - \nu)}\, a_i \tag{2.90}$$

$$\left(\xi^*, \eta^*\right) = \frac{4G(1-\nu)}{2\pi^2 \mu\, p_{z,o,i}}\left(C_{11}\xi, C_{22}\eta\right) \tag{2.91}$$

$$\varphi^* = \frac{3}{\pi}\sqrt{\frac{b}{a}}\,\frac{G a_i C_{23}}{2\mu\, p_{z,o,i}}\,\varphi \tag{2.92}$$

$$(\Xi, \mathrm{H}) = \left(\xi^*, \eta^*\right) + \frac{\varphi^*}{a_i}(-y, d_i) \tag{2.93}$$

$$\left(k_{x,i}, k_{y,i}\right) = \frac{(\Xi, \mathrm{H})}{\|(\Xi, \mathrm{H})\|} \tag{2.94}$$

$$\left(k'_{x,i}, k'_{y,i}\right) = \left(k_{x,i}, k_{y,i}\right) - \left(0, \varphi^*\right) \tag{2.95}$$

In the slip area, a modified version of FASTSIM is used to determine the direction of the tangential stresses. Two features of this FASTSIM version are as follows:

- Different effective tangential compliance L values are used in each strip, depending on the saturation condition of the strip: for strips completely in slip, a weighted L is used, calculated according to Eq. (2.76). Otherwise, individual L_1, L_2 and L_3 values are used associated to each creepage (cf. Eq. (2.75)).
- A correction is performed in the adhesion limit g for contact ellipses with $a / b < 1$, which is given by Eq. 16 of [Sichani 2016a]. This affects only in the calculation of the direction of the resultant tangential stress in the slip area, but not its magnitude, which is calculated as the actual normal pressure times the coefficient of friction.

FASTRIP improves the precision of the contact resultant forces with respect to FASTSIM, and additionally provides realistic tangential stress distributions in a wide variety of creepage situations. As a drawback, the foreseen adhesion regions are always located in the leading part of the contact patch, while with high spin the adhesion region may be located in the central part, with the rest of the contact patch in slip. Nevertheless, in cases with high spin, the resultant contact forces are normally small.

The application of FASTRIP to non-elliptical contact areas, together with ANALYN (§2.1.1.4) for the normal part, is shown in [Sichani 2016b]. However, details are not given on the calculation of the flexibility parameters L_i and of the creepage coefficients C_{ij} for non-elliptical contact areas.

2.1.2.9 Books of tables

Pre-calculated or look-up tables are used as well to provide quick and precise tangential contact resultant forces in rail vehicle dynamic simulations. The tables store tangential contact resultant forces for different wheel–rail contact situations. The tables are first built offline (before the dynamic simulation), which may be done with more precise (and computationally costly) models than

those normally used in the dynamic simulations, such as Kalker's exact contact theory. Then, during the dynamic simulation, the tables are evaluated by some interpolation method. A drawback of this approach is that the tables are normally built storing just the resultant forces, so that it is not possible to retrieve more detailed information, like contact stress distributions.

The first known book of tables for wheel–rail tangential contact was created by Kalker in the 1970s with his program DUVOROL, predecessor of the CONTACT program, where Kalker implemented his exact contact theory. These tables were created for British Rail, and were not published. In [Kalker 1996], a book of tables for elliptical contact patches was presented. More recently, tables covering a broader range of contact patch shapes have been calculated in [Piotrowski 2017]. In this reference, so called "SDEC" (simple double-elliptical contact) regions are covered, which are made up of two half-ellipses, as shown in Figure 2.17.

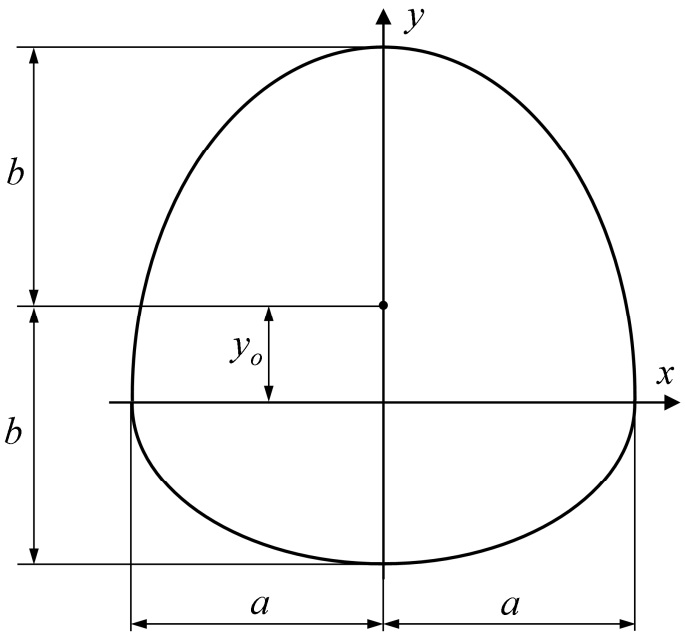


Figure 2.17. SDEC or falling drop region. Adapted from [Piotrowski 2017].

The tables are multi-dimensional, as they need to cover different creepage situations and contact patch shapes. The involved variables are suitably

normalized, to minimize the number of table dimensions. Each independent variable involves a table dimension. In the case of the tables presented in [Piotrowski 2017], the number of table dimensions is 6, with the following set of independent variables: ξ^*, η^*, χ, ψ, a/b, and v. ξ^*, η^*, and χ are normalized creepages defined in Eqs. (2.96) and (2.97), with the c and ρ dimensions being defined in Eqs. (2.98) and (2.99), respectively. ψ is a shape factor of the SDEC region, equal to y_o / b. A_c in Eq. (2.99) is the contact patch area, and m and n are the Hertzian coefficients introduced in §2.1.1.1 for the best fitting ellipse. The data stored in each entry of the table are normalized tangential resultant forces $f_i = F_i / \mu N$, $i \in \{x, y\}$.

$$\left(\xi^*, \eta^*\right) = \frac{\rho}{\mu\, c}\left(\xi, \eta\right) \tag{2.96}$$

$$\chi = \frac{\rho}{\mu}\varphi \tag{2.97}$$

$$c = \sqrt{ab} \tag{2.98}$$

$$\rho = \frac{4E}{3\left(1-\nu^2\right) N}\left(\frac{A_c}{\pi\, m\, n}\right)^{3/2} \tag{2.99}$$

For Hertzian contacts, and assuming a single value of the material coefficient of Poisson v is of interest, the number of table dimensions may be reduced to 4: ξ^*, η^*, χ, and a/b. In this case, the definition of ρ of Eq. (2.99) may be replaced by $\rho = 4 / (C_{x2} + C_{y1} + C_{y2})$ (assuming a straight rail, with $C_{x1} = 0$).

Given the number of table dimensions, the size of the table may easily become excessive. To avoid this, the number of points along each dimension has to be adequately adjusted, while maintaining an adequate resolution and range. A large table size translates into increased cost, not only for its offline calculation, but also for its later evaluation, in addition to increased memory requirements.

2.1.3 Exact contact theory

Kalker developed the first numerical method aimed at solving the wheel–rail rolling contact problem [Kalker 1979b], [Kalker 1990]. It is known as the exact contact theory, because it uses the exact surface stress–elastic displacement constitutive relationship expressed in Eq. (1.6). The method is generally applicable to elastostatic contact problems, not being limited by the usual assumptions stated at the beginning of §2.1.2. Thus, it is valid for non-Hertzian and non-steady state contacts, and for contacts with dissimilar materials.

The exact contact theory uses an Eulerian mesh of the contact surface, associated to the Eulerian coordinate system following the contact described in §1.2, in which the kinematic equations of the contact problem are set out. In this mesh, the exact integral relationships between surface stresses and elastic displacements are discretized, and the normal and tangential contact conditions are enforced.

The meshed region is called the potential contact surface (PCS). It must circumscribe the contact patch(es), while being as adjusted in size as possible so that a fine spatial resolution can be provided. An estimate of an adequate size for it may be commonly obtained by means of the intersection of the undeformed contact surfaces in the reference configuration. As depicted in Figure 1.7, the apparent contact area given by this intersection encompasses the true contact area.

The discretization of the PCS is commonly carried out with equal rectangular elements aligned with the x and y directions, as shown in Figure 2.18. Compared to discretizations with non-uniform meshes like the one used in the FASTSIM algorithm (cf. Figure 2.15), a uniform discretization, with equal elements, provides computational advantages in the calculation of the influence coefficients (ICs) and in their convolution with the contact stresses to obtain the elastic displacements, as explained in §4.1.1.4. Other element types are possible apart from the rectangular, maintaining regular mesh structure and associated advantages. For example in [Kalker 1972], triangular elements with linear pressure distributions are considered in a numerical solution to the frictionless contact problem.

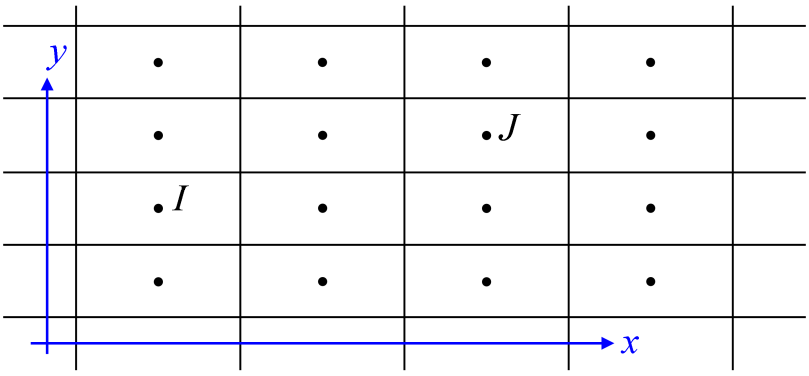


Figure 2.18. Discretization of the PCS with equal rectangular elements.

The space discretization of the exact surface stress–displacement relationship given in Eq. (1.6) leads to Eq. (2.100), used in the exact contact theory. Subindices *I* and *J* in this equation refer to elements of the mesh, and subindices *i* and *j* to directions of the contact coordinate system. IC_{IiJj} is the IC that gives the displacement in *i* direction of element *I* due to a unitary load in *j* direction on element *J*. It is obtained integrating the product of the corresponding IF with the shape function of the distributed load in the loaded element, as explained in §1.3.1 (Eq. (1.9)) and §5.1.

$$u_{Ii} \approx \sum_{j\in\{x,y,z\}} \sum_{J=1}^{N} IC_{IiJj}\, p_{Jj};\;\; I \in \{\, 1...NE \,\},\;\; i \in \{\, x,y,z \,\} \tag{2.100}$$

In general, $IC_{IiJj} = IC_{IiJj}(\mathbf{x_I}, \mathbf{x_J})$, i.e. IC_{IiJj} generally depends on the position of both *I* and *J*. In the case of the half-space, due to the homogeneity and similarity of the medium in the direction parallel to the surface and to the regularity of the mesh, IC_{IiJj} depends only on the relative position between *I* and *J*. So, it may be written $IC_{IiJj} = IC_{IiJj}(\mathbf{x_I} - \mathbf{x_J})$. For a 3D contact problem, this implies reducing the number of IC_{IiJj} ICs from $O(N^4)$ to $O(N^2)$, being *N* the number of elements in each direction of the mesh. Eq. (2.100) represents a matrix-vector multiplication between a matrix of ICs $\mathbf{[IC_{ij}]}$ and a vector of contact stresses $\mathbf{\{p_j\}}$.

The points chosen to represent each element of the mesh will be designated here as analysis points (APs). These are the points for which the problem unknowns (contact stresses) are sought, and on which the contact conditions are checked.

The APs of the mesh form a regular pattern, like the elements. In the case of the commonly used rectangular elements with uniform load distribution, the AP of each element is its geometric centre, as depicted in Figure 2.18 with black circles. Then, in this case, the number of APs is the same as the number of elements. The APs of the contact mesh may be assigned a 2D numbering, defining their position along the x and y directions of the mesh. A 1D numbering for both the elements and APs may be defined as well, as illustrated in Figure 2.19. According to this, the number e corresponding to an AP located at position index (ix, iy) in the mesh is $e = ix + n_x \times (iy - 1)$. n_x and n_y are the number of APs in x and y directions of the mesh, and ix and iy the position indices in each direction. The concept of AP is treated again in §4.1.1.1.

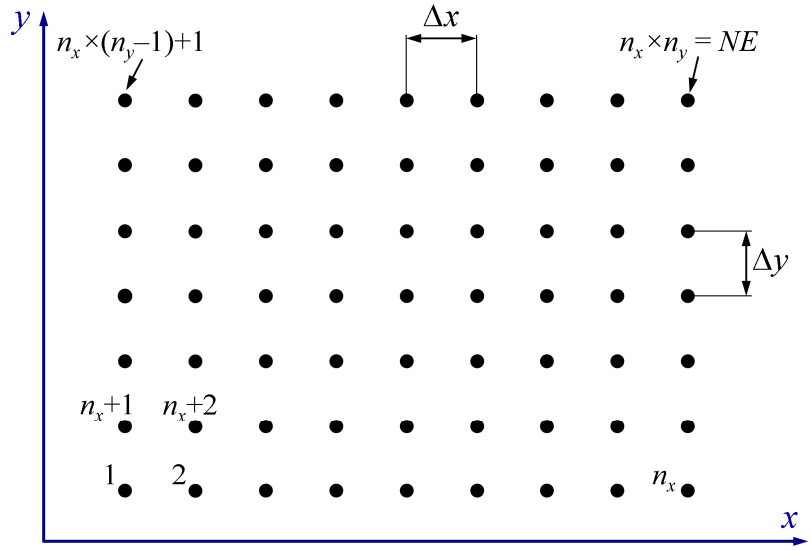


Figure 2.19. Numbering of APs of the mesh of the PCS.

The contact problem is divided into the normal and tangential parts for its solution. The algorithms for the solution of each part are called NORM and TANG, respectively.

The algorithm NORM is summarized as follows:

1. A set of APs inside the contact ***C*** is assumed. For each AP I in ***C***, the displacement compatibility condition $e = 0$ is formulated, according to

Eq. (2.101), and making use of Eq. (2.100). The number of APs assumed to be in $\boldsymbol{C}$ is designated as N_C.

$$d - h_I = u_{Iz} = \sum_{J \in \boldsymbol{C}} IC_{IzJz} p_{Jz} + \sum_{j \in \{x,y\}} \sum_{J \in \boldsymbol{C}} IC_{IzJj} p_{Jj}; \;\; I \in \boldsymbol{C} \qquad (2.101)$$

The first term of the right-hand side of Eq. (2.101) is the contribution of the normal pressures to u_{Iz}, and the second term the contribution of the tangential stresses. Normally, the first term is dominant. If the tangential stresses were zero, or their contribution to u_{Iz} zero or known, Eq. (2.101) would represent a linear system of N_C equations for the N_C p_{Iz} unknowns of the normal contact problem (it is known that $p_{Iz} = 0$ outside $\boldsymbol{C}$). To facilitate the foregoing discussion, it will be assumed for the moment that the second term of the right-hand side of Eq. (2.101) is zero. This happens in cases where quasiidentity can be applied, as in concentrated wheel–rail contact. Afterwards, the solution process for the more general case in which this is not fulfilled will be outlined.

2. The equation system (2.101) is solved, and the normal contact inequalities are checked: $p_{Iz} > 0$ in $\boldsymbol{C}$, and $e_I > 0$ outside $\boldsymbol{C}$. If some AP does not fulfil its corresponding inequality, it is changed from set, adding or removing it from $\boldsymbol{C}$ as applicable.

3. If the set of APs in $\boldsymbol{C}$ has changed in the previous step, the algorithm returns to step 1.

Steps 1–3 are repeated until the correct set of APs in $\boldsymbol{C}$ is found, with which the applicable normal contact inequalities are fulfilled in all APs of the mesh. Kalker provided a rigorous mathematical proof of the convergence of the NORM algorithm when a single AP is changed from set in each iteration. In practice, it is observed that it also works changing at the same time all APs that do not fulfil their corresponding inequality in each iteration, as described here. The fact that the matrix of the system, formed by the IC_{IzJz} ICs, is of dominant diagonal (in other words, due to the markedly local character of the influence of the stresses in the deformations of the solid), and positive in half-space contact problems, facilitates the convergence of the NORM algorithm.

Regarding the tangential contact problem, the particle time derivative of the elastic displacements that appears in the kinematic Eq. (1.7), D**u** / d*t*, is discretized in time considering the displacements **u** and **u'** of the particle at two nearby time instants *t* and *t'*, as in Eq. (2.71). Considering additionally the space discretization of the integral surface stress–displacement constitutive relationship expressed in (2.100), Eq. (2.102) is obtained. In this equation, the contributions of the normal pressures and tangential stresses to u_{Ii} are separated, as in Eq. (2.101). As done previously for the normal problem, it will be assumed that the contributions due to the normal pressures are known, to facilitate the explanation.

$$\frac{\mathrm{D}u_{Ii}}{\mathrm{D}t} \approx \frac{u_{Ii} - u'_{Ii}}{t - t'} = \frac{\left(\sum_{j \in \{x,y\}} \sum_{J \in \boldsymbol{C}} IC_{IiJj}\, p_{Jj} + \sum_{J \in \boldsymbol{C}} IC_{IiJz}\, p_{Jz} \right)}{\Delta t}$$

$$- \frac{\left(\sum_{j \in \{x,y\}} \sum_{J \in \boldsymbol{C}} IC'_{IiJj}\, p'_{Jj} + \sum_{J \in \boldsymbol{C}} IC'_{IiJz}\, p'_{Jz} \right)}{\Delta t}; \; I \in \boldsymbol{C}; \; i \in \{ x, y \} \tag{2.102}$$

The primed quantities in Eq. (2.102) are referred to the previous time instant *t'*. p'_{Jj} are the stresses acting on element *J* at *t'*, and IC'_{IiJj} are the ICs of element *J* over the position occupied at *t'* by the particle located at AP *I* at the current time instant *t*. The positions of the particles corresponding to a set of APs of a contact mesh at *t* and *t'* are depicted in Figure 2.20 with black circles and red squares, respectively; recall also the material flow through the Eulerian contact coordinate system illustrated in Figure 1.9b. From *t'* to *t*, the particles have a rigid body displacement $\Delta q = V_c\,(t - t') = V_c\,\Delta t$.

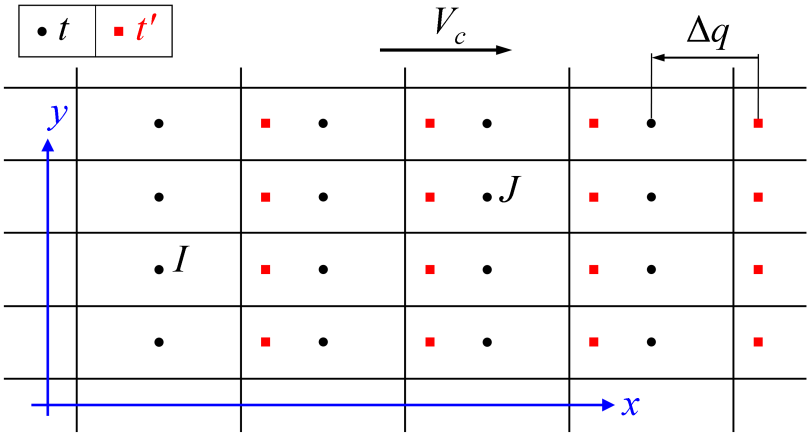


Figure 2.20. Rigid body displacement of material particles in the Eulerian contact coordinate system.

Taking into account the equality conditions of the tangential problem for $\mathbf{p_t}$ and $\mathbf{v}$, there are two equations for each of the N_C APs in contact. For the elements in adhesion, the condition $\mathbf{v} = \mathbf{0}$ is used, developed according to Eqs. (1.7) and (2.102). For the elements in slip, the used conditions are $\|\mathbf{p_t}\| = g$, and $p_x\, v_y - p_y\, v_x = 0$. The latter condition, expressing the parallelism between $\mathbf{p_t}$ and $\mathbf{v}$, is used in place of $\mathbf{v} = -\lambda\, \mathbf{p_t}$, avoiding adding new unknowns λ. So, there are $2{\times}N_C$ equations, and the $2{\times}N_C$ unknowns are the p_x and p_y components of the tangential stress of each AP. In transient problems, the stresses $\mathbf{p'_t}$ of the previous instant t' must be known, and in steady rolling problems, $\mathbf{p_t} = \mathbf{p'_t}$ in the Eulerian contact mesh.

The division of the APs into the adhesion $\boldsymbol{A}$ and slip $\boldsymbol{D}$ sets is not known beforehand. Therefore, as in the normal problem, it is necessary to iterate until, in addition of fulfilling the previous equations, the inequalities corresponding to each AP are fulfilled: for the APs in $\boldsymbol{A}$, $\|\mathbf{p_t}\| < g$; and for the APs in $\boldsymbol{D}$, $\mathbf{p_t} \cdot \mathbf{v} < 0$. The steps of the TANG algorithm are similar to those of the NORM algorithm: based on assumed sets $\boldsymbol{A}$ and $\boldsymbol{D}$, the corresponding system of equations is formed and solved, and subsequently the corresponding inequalities are checked in each AP, changing from set the non-compliant APs. This is repeated iteratively until convergence is achieved (i.e., until all APs fulfil their corresponding inequality or contact condition). Unlike in the normal problem,

in this case not all the equations are linear; the ones of the APs in slip are non-linear. Convergence is also more difficult than in the normal problem.

When the normal and tangential problems are coupled, a possibility is to solve them following the Panagiotopoulos process ([Panagiotopoulos 1975], [Antes 1992]). This consists on solving the normal and the tangential problems alternatively until the last two normal-tangential solutions are close enough to consider that convergence has been achieved. In each Panagiotopoulos iteration, the normal problem is solved first, with tangential tractions fixed in the values obtained in the tangential solution of the previous iteration. Then the tangential problem is solved, with normal pressures fixed in the values obtained in the previous normal solution. The process is looped until convergence is achieved, which is not guaranteed.

The Panagiotopoulos normal-tangential contact iteration is illustrated in Figure 2.21, including the main inputs for both the normal and tangential problems that depend partly on the solution of the other problem, and hence change during the solution process. As depicted in the figure, the influence of the tangential stresses for the normal problem solution is collected in a modified undeformed distance h^*, and the influence of the normal pressures for the tangential problem solution is collected in a modified rigid slip velocity $\mathbf{w}^*$. In transient contact problems, $\mathbf{w}^*$ also includes the influence of the tangential stresses of the previous instant $\mathbf{p'}_t$ (not depicted in the figure), which are known.

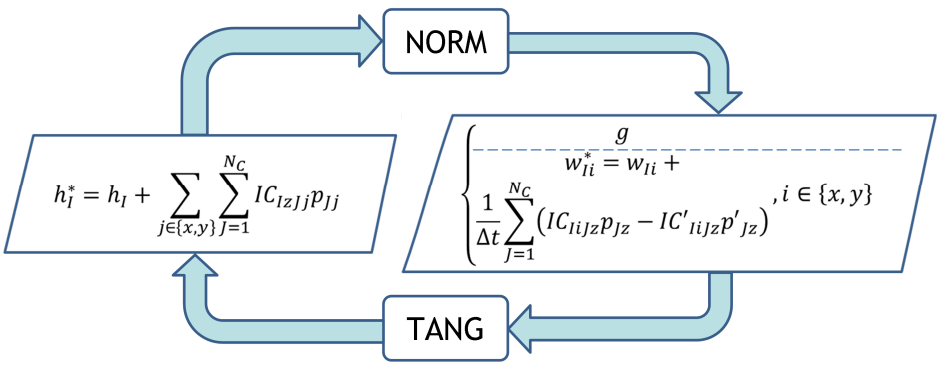


Figure 2.21. Panagiotopoulos process for the solution of the coupled normal-tangential contact problem.

The complete normal-tangential coupled contact problem may also be solved at the same time. This involves setting up systems of equations which include the contact equations for the normal and the tangential contact problems. A method doing this named KOMBI was described in [Kalker 1990]. The method currently used in the exact contact theory implemented in the CONTACT program is the Panagiotopoulos process.

More details on the solution algorithms of the exact contact theory are given in §4.1.4, as implemented in the version developed in this thesis for the solution of conformal contact problems. More efficient solvers than the NORM and TANG solvers described here have been developed by Vollebregt and co-workers; cf. [Vollebregt 2014b] and [Vollebregt 2014c] for the normal problem, and [Zhao J 2015] for the tangential problem. A key of their efficiency is performing the convolutions of the ICs with the contact stresses in the Fourier domain (cf. §4.1.1.4.2), and they avoid the explicit formation of the entire systems of equations described here.

In addition to providing accurate numerical solutions, the exact contact theory provides flexibility to model complex interfacial behaviours, that may not be readily accommodated in other simplified contact models. These include consideration of third body layer, and variable friction coefficient with slip velocity.

At the software level, the consideration of the third body layer is quite straightforward, involving just the addition of constant IC_{IiIi} flexibility terms in the local direct ICs of each element with itself (under the common application to the third body layer of the 'wire brush' model depicted in Figure 2.14). A limiting shear stress may also be easily implemented, limiting if necessary the traction bound g.

The consideration of a coefficient of friction dependent on the slip velocity is somewhat more involved. As the slip velocity is part of the solution, an additional, outer iteration becomes necessary for the solution of the tangential problem, updating the coefficient of friction in each position as successive, tentative slip velocities are computed. Implementations in the exact contact theory of this capability are reported in [Croft 2011] and in [CONTACT UG 2013]. In the latter reference, it is shown that decaying friction laws with pure dependence on the local slip velocity lead to unwanted effects as dependence of

the contact solution in the initial estimate. To avoid this, the possibility for a memory effect in the friction law is implemented, by which the coefficient of friction depends on its previous history (in a small distance) in addition to the local slip velocity. It is shown that the mentioned unwanted effects and high-frequency stick-slip oscillations are suppressed with this memory effect.

2.2 Influence functions of the elastic solid

In this section, the available analytical results of influence functions (IFs) of the elastic solid are introduced. To start with, the case of the elastic half-space is considered, as the reference case for which the theoretical foundations for its study have been developed to a greater extent, and of great practical relevance especially in concentrated contact problems. Afterwards, particular cases with non-planar surfaces are treated.

In accordance with the hypotheses for the target contact problems set out in §1.1.1, particularly massive (i.e., not hollow) solids are considered, in which their smallest characteristic dimensions is at least of the same order of magnitude than the contact patch dimensions. In these cases, the stresses that appear as a consequence of the contact are highly concentrated around the contact zone, and the stress state in the vicinity of the contact typically presents a relatively high hydrostatic component.

Additionally, attention is limited here to homogeneous and isotropic solids. In the literature studies have been carried out as well with solids made up of multiple layers of different elastic properties. In practice, sometimes cases are encountered of contact between solids that have applied coatings with different elastic properties from the substrate. In these cases, it is not valid to assume that the solids are homogeneous, if precise contact mechanics analyses are pursued. These cases are out of the scope of this thesis, and here only a small sample of the related literature is cited.

In [Burmister 1945] the mathematical approach of the problem of elastic half-spaces of 2 and 3 layers subject to normal and axisymmetric loads on the surface was developed. In [Chen 1971] the development was extended to non-axisymmetric loads. Johnson introduces the contact problem between layered solids in [Johnson 1987] (§5.8), presenting some results for some particular

cases of elastic layers on rigid substrates in plane strain, and a bibliography review is carried out of contact problems with layered solids. In [O'Sullivan 1988] the problem of a layered half-space under arbitrary normal and tangential surface loads was solved. The obtained formulation was applied to a contact problem with a rigid sphere in full slip, although the coupling between the normal and the tangential contact problems was not taken into account. In other works such as [Nogi 1997] and [Wang Z-J 2010], other application examples may be found including numerical contact calculations between an elastic half-space with a top layer and a rigid solid. Nyqvist ([Nyqvist 2012], [Nyqvist 2015]) studied the frictional non-conformal contact between multilayered solids with arbitrary number of layers. Analytical expressions have been obtained of the IFs of the layered half-space in the Fourier domain. In contrast to the case of the homogeneous half-space, it is not possible to express these IFs analytically in the space domain, and it is necessary to evaluate them numerically from the expressions in the Fourier domain, calculating the integrals which define the inverse transformation.

2.2.1 The elastic half-space

The influence functions (IFs) of the elastic homogeneous half-space related to point loads applied in the surface may be obtained analytically from the Boussinesq [Boussinesq 1885] and Cerruti [Cerruti 1882] potentials, as is cited in [Johnson 1987] (§3.1), and may be found in the literature about Elasticity and contact mechanics, see e.g. [Landau & Lifshitz 1986], [Johnson 1987], [Kalker 1990] or [Popov 2010].

The solutions for the half-space may also be used to solve quarter-space problems. Hetényi [Hetényi 1970] devised a way to superimpose the solutions of two half-spaces with perpendicular free surfaces to compute the elastic field in the quarter-space. Based on this concept, in [Zhang 2013] the procedure to obtain the IC matrices of the quarter-space was given, explicitly and with no iterations. Particularly the quarter-space with normal pressure applied on its free surface was considered in that reference, but the method may be readily applied with surface shear stresses as well.

Next, the IFs of the elastic displacements of the half-space produced by unit point or concentrated surface forces are given. These are denoted here with

letter *a*. Here a homogeneous half-space is considered, with a single material of linear elastic and isotropic behaviour.

The nomenclature followed to express these IFs is defined first. A Cartesian {*x*, *y*, *z*} coordinate system is considered, with its origin located in the free surface, the *x* and *y* axes parallel to this surface, and the *z* axis pointing into the half-space, completing the right-handed Cartesian system. The *x*, *y* and *z* axes are designated as longitudinal, lateral and normal, respectively. The IF $a^k_{ij}(x, y, z)$ is defined as the displacement produced in *i* direction at the point (0, 0, *z*) of the half-space designated with number *k*, by a unit point load applied in direction *j* at point (*x*, *y*, 0) of its surface (with *i*, *j* = *x*, *y*, or *z*).

The IFs a^k_{xx}, a^k_{xy}, a^k_{xz}, a^k_{zx} and a^k_{zz} are given in Eqs. (2.103)–(2.107).

$$a^k_{xx}(x,y,z) = \frac{1}{4\,\pi\,G_k}\left(\frac{1}{\rho} + \frac{1-2\nu_k}{\rho+z} + \frac{x^2}{\rho^3} - \frac{(1-2\nu_k)\,x^2}{\rho(\rho+z)^2}\right) \tag{2.103}$$

$$a^k_{xy}(x,y,z) = \frac{1}{4\,\pi\,G_k}\left(\frac{xy}{\rho^3} - \frac{(1-2\nu_k)\,xy}{\rho(\rho+z)^2}\right) \tag{2.104}$$

$$a^k_{xz}(x,y,z) = \frac{-1}{4\,\pi\,G_k}\left(\frac{xz}{\rho^3} - \frac{(1-2\nu_k)\,x}{\rho(\rho+z)}\right) \tag{2.105}$$

$$a^k_{zx}(x,y,z) = \frac{-1}{4\,\pi\,G_k}\left(\frac{xz}{\rho^3} + \frac{(1-2\nu_k)\,x}{\rho(\rho+z)}\right) \tag{2.106}$$

$$a^k_{zz}(x,y,z) = \frac{1}{4\,\pi\,G_k}\left(\frac{z^2}{\rho^3} + \frac{2(1-\nu_k)}{\rho}\right) \tag{2.107}$$

Being G_k and ν_k the shear modulus and the coefficient of Poisson of the half-space *k*, and ρ the distance between the point where the load is applied and the point where the displacements are observed:

$$\rho = \sqrt{x^2 + y^2 + z^2} \tag{2.108}$$

The rest of the IFs may be obtained from the ones given above, interchanging the x and y coordinates in the corresponding equations. For example, the IFs a^k_{yy} and a^k_{yz} may be obtained from a^k_{xx} and a^k_{xz} respectively, changing the x variable for the y variable in Eqs. (2.103) and (2.105). It is verified that $a^k_{xz} / a^k_{yz} = x / y$. That is to say, the projections in a plane parallel to the surface of the displacements caused by a point normal load are oriented radially with respect to the point where the load is applied, as it must be due to the axisymmetry of the problem with a point normal load.

Clearly, for a homogeneous half-space, the IFs depend only on the relative positions in the x and y axes of the point where the load is applied with respect to the point where the displacements are observed, as well as on the coordinate z of the latter. That is to say, the displacements produced in any point (x', y', z) of the solid by a load applied at the point $(x' + x, y' + y, 0)$ of its surface, are equal to the displacements produced at point $(0, 0, z)$ of the solid by the same load applied at the point $(x, y, 0)$. This property allows considerable computational savings in the computations related to the ICs in concentrated contact problems (in which additionally the elastic properties of the solids do not change in the direction parallel to the surface in the vicinity of the contact); cf. §4.1.1.4.

Next, the IFs a_{xx}, a_{xy}, a_{xz} and a_{zz} of the relative surface displacements between two contacting elastic half-spaces are given. These are deduced from the corresponding a^k_{ij} IFs in Eqs. (2.103)–(2.107), taking into account that in the surface $z = 0$.

$$a_{xx}(x,y) = \frac{1}{\pi G}\left(\frac{1-\nu}{\rho} + \nu\frac{x^2}{\rho^3}\right) \tag{2.109}$$

$$a_{xy}(x,y) = \frac{\nu}{\pi G}\frac{xy}{\rho^3} \tag{2.110}$$

$$a_{xz}(x,y) = \frac{K}{\pi G}\frac{x}{\rho^2} \tag{2.111}$$

$$a_{zz}(x,y)=\frac{1-\nu}{\pi G}\frac{1}{\rho} \tag{2.112}$$

In Eqs. (2.109)–(2.112), dealing with displacements in the surface (z = 0), the dependency of coordinate z is suppressed, and $\rho=\sqrt{x^2+y^2}$. The combined elastic constants G, ν and K used in these equations are defined as follows:

$$G=\frac{2}{\frac{1}{G_1}+\frac{1}{G_2}} \tag{2.113}$$

$$\nu=\frac{G}{2}\left(\frac{\nu_1}{G_1}+\frac{\nu_2}{G_2}\right) \tag{2.114}$$

$$K=\frac{G}{4}\left(\frac{1-2\nu_2}{G_2}-\frac{1-2\nu_1}{G_1}\right) \tag{2.115}$$

Again, the rest of the IFs may be obtained from the ones given above, changing x for y in the corresponding equations. For example, the IFs a_{yy} and a_{yz} may be obtained from a_{xx} and a_{xz} respectively, changing x for y in Eqs. (2.109) and (2.111). On the other hand, a_{zx} is deduced by reciprocity from a_{xz}:

$$a_{zx}(x, y) = a_{xz}(-x, -y) = -a_{xz}(x, y) \tag{2.116}$$

ICs derived from the above given IFs are given in §5.1, including results available in the literature, and new ones developed in this thesis.

The half-plane is the 2D equivalent of the half-space. Eqs. (2.117)–(2.121) give the IFs for the surface displacements of the elastic half-plane subject to a unit concentrated line[1] load on the surface (the remaining ones may be deduced by

[1] Here line load means a distributed load applied in an infinite longitude in the direction perpendicular to the plane where the 2D geometry is located, uniformly as it

reciprocity). In these equations, the superindex "*hp*" stands for half-plane. A Cartesian coordinate system similar to the one used before for the half-space is considered: the *x* axis is contained in the plane and parallel to the surface, the *z* axis is normal to the surface and pointing to the inside of the half-plane, and the *y* axis completes the right-handed system.

$$a_{zz}^{hp} = \frac{1-\nu}{\pi G} \ln \left| \frac{x_0}{x} \right| \tag{2.117}$$

$$a_{xx}^{hp} = a_{zz}^{hp} \tag{2.118}$$

$$a_{xz}^{hp} = \mathrm{sgn}(x) \frac{1-2\nu}{4G} \tag{2.119}$$

The x_0 variable in Eq. (2.117) is the necessary reference point in the *x* axis to define the origin of the normal displacements due to the normal line load, as in the case of the half-plane the absolute magnitude of these is undetermined.

The following relations apply for the displacements normal to the plane, as developed in [Johnson 1987] (§2.9):

$$a_{yy}^{hp} = \frac{a_{zz}^{hp}}{1-\nu} \tag{2.120}$$

$$a_{xy}^{hp} = a_{zy}^{hp} = 0 \tag{2.121}$$

The ICs associated to line loads distributed uniformly or linearly in a surface element parallel to the *x* axis may be easily obtained integrating the previous IFs in the *x* coordinate, and may be found in [Johnson 1987], §2 (Eqs. (2.30), (2.32) and (2.37))[2].

---

corresponds to a 2D problem. The adjective "concentrated" or "point" is used when the zone where the load is applied is punctual inside the plane.

[2] Note that the cited reference gives the expressions for the displacements in *x* corresponding to loads applied in the origin ($x = 0$), while here the displacements

2.2.2 Non-planar solids with available analytical results

In this section the cases of 2D cylindrical geometries and of the sphere are examined, for which analytical IFs are available. It is also worth mentioning the case of 3D infinitely long cylindrical geometries subject to axially periodic normal loading, studied by Chiu ([Chiu 1963], [Chiu 1964], [Chiu 1987]). In [Chiu 1987] the procedure for the computation of the ICs of the cylinder and the cylindrical cavity under the action of a uniform normal pressure distribution in a rectangular surface element is explained. In this case, there is no closed form solution available for the IFs, and the ICs are computed as a doubly infinite series of elemental solutions for bisinusoidal load in the cylindrical surface.

2.2.2.1 Cylindrical geometries under line loads

There are analytical solutions for the 2D geometries of the cylinder and the cylindrical cavity in the infinite space subject to line loads in the surface. These cases have great practical interest, as contacts between components with this geometry are frequently found in applications such as joints with pins and in diverse machine elements as pistons, shafts and bearings for instance.

In this section, the case of plane strain will be considered. The solutions for plane stress are similar to those for plane strain, as the equations of linear elasticity in statics for homogeneous and isotropic bodies apply in a similar way for both states, with the only difference of the elastic constants that appear in them in each case (cf. [Muskhelishvili 1977], §25 and 26). The solutions for plane stress may therefore be obtained from those for plane strain, modifying the elastic constants in the following way, as explained e.g. in [Woodward 1976] – app. I, [Ciavarella 2001a], or [Hou 2001]:

$$E' = \frac{E(1+2\nu)}{(1+\nu)^2} \tag{2.122}$$

produced in the origin by loads applied in x are considered. Therefore, it is necessary to change the sign of the expressions of A_{xz} and A_{zx} of the cited reference to obtain the expressions of the corresponding ICs as defined here.

$$\nu' = \frac{\nu}{1+\nu} \tag{2.123}$$

This applies generally, not only in the particular case of cylindrical geometries. In the previous equations, E is the Young's modulus and ν the coefficient of Poisson of the solid. E' and ν' are the modified elastic constants which may be used in the solutions for plane strain, replacing the original ones E and ν, to obtain the solutions for plane stress[3]. It is not necessary to modify the shear modulus G, as may be verified with the previous equations (2.122) and (2.123).

[Muskhelishvili 1977], §80a gives the solution for the problem of the cylinder under the action of a couple of opposed and collinear (not necessarily radial) surface forces. The resulting expressions of the surface displacements are particularized in [Woodward 1976] for the case of diametrically opposed forces. The applicable solutions for the case of a single force, normal or tangential, applied in the surface of the cylinder, equilibrated with a point force and moment in the centre or with a uniformly distributed load in the whole section of the cylinder, may be found in [Sundaram 2010b]. The complex potentials[4] for

[3] Commonly, in practical situations neither the condition of plane strain nor that of plane stress is strictly fulfilled, as for example in pinned joints, in which typically the pin diameter is of the same order of magnitude as the thickness where the contact is produced. In these cases, the real solution will presumably be somewhere between the solutions for plane strain and for plane stress. Moreover, it may be verified that the differences between the solutions for plane strain and plane stress turn out to be not very big, mostly with small values of the coefficient of Poisson. So it is reasonable to assume that in these cases the hypothesis of plane strain as well as that of plane stress will lead to a good approximation of the real solution (without considering 3D effects as e.g. misalignments between the axes of both contacting bodies).

[4] The complex potentials are the complex representation of the stress function for 2D linear elasticity problems. The stress function or Airy function applicable in plane linear elasticity problems in the absence of volumetric forces or with conservative volumetric forces, is an auxiliary scalar function, which second order derivatives represent the stress components, and that considerably simplifies the treatment of these problems. In 3D elasticity the displacement functions are more usual, being in general necessary at least three functions for a complete definition of the problem. More details may be

the case of the cylindrical cavity were derived in [Rothman 1950] for the first time, as cited in [Sundaram 2010a]. [Woodward 1976] gives the expression for the surface displacements in the cylindrical cavity due to a couple of opposed normal forces applied in two diametrically opposed points of the surface. On the other hand, in [Sundaram 2010a] the expressions for the displacements due to a single point force are given, normal or tangential, applied in the surface as well.

Next, the IFs of the surface displacements for the previously cited 2D problems with cylindrical geometries are exposed. Figure 2.22 illustrates the different problems considered for the cylinder, and Figure 2.23 the ones corresponding to the cylindrical cavity.

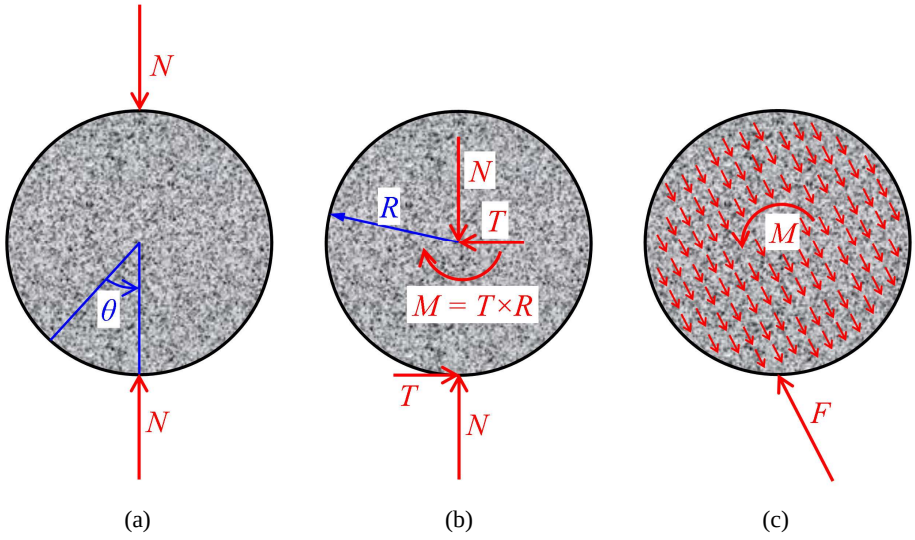
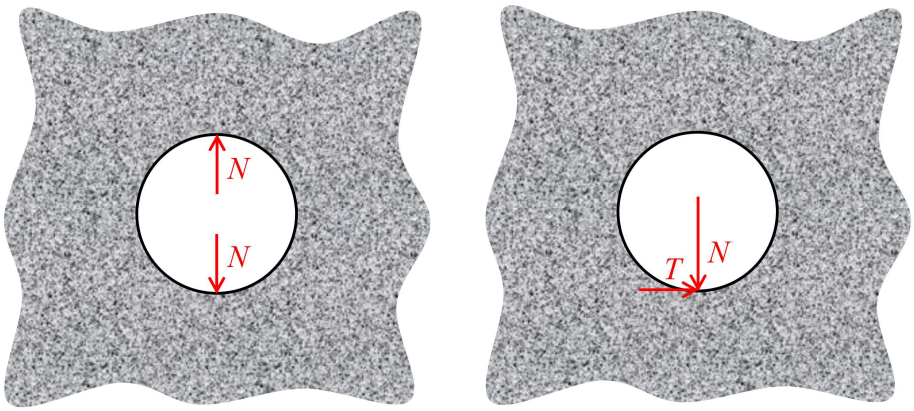


Figure 2.22. 2D problems for the cylinder subject to point line forces in the surface. (a) Two normal, diametrically opposed forces. (b) and (c): a single force (decomposed into its normal N and tangential T components) equilibrated with a moment; and with an

found about the principles and application of the stress functions and displacement functions in linear elasticity problems in the literature about Elasticity (e.g. [Love 1906], [Timoshenko 1951], [Muskhelishvili 1977] or [Barber 2002]), as well as examples of their use to solve practical problems. [Muskhelishvili 1977], §5, presents a detailed development of the complex representation of the general solution of the equations of plane linear elasticity.

opposed force (b) in the centre of the cylinder or (c) uniformly distributed in the whole section of the cylinder.



(a) Two normal, diametrically opposed forces. (b) A single force.

Figure 2.23. 2D problems for the cylindrical cavity subject to point line forces in the surface.

The IFs, given in Eqs. (2.124)–(2.133), are expressed in a curvilinear reference system, tangent in each point to the surface of the cylinder or of the cylindrical cavity, with radial or normal coordinate n and lateral or tangential s. On the other hand, the surface points are located according to a polar reference system with origin in the centre of the cylinder or of the cylindrical cavity, with angular coordinate θ. The angular coordinate is defined positive clockwise, as the tangential displacements, and the positive direction of the normal displacements is defined towards the inside of the solid in each case. These IFs are designated with the a letter, as those of the half-space, because they are exact analytical. The $a_{ij}(\theta)$ IF, with i, j = n or s, represents the displacement produced in i direction in the surface point with 0 angular coordinate, due to a unitary line concentrated force (with dimension of force per unit length in the direction perpendicular to the plane) applied in j direction in the surface point with θ angular coordinate, and to the corresponding reactions that equilibrate it.

$$a_{nn}^{cyl,a} = -K_1\left(\cos\theta \, \ln\left|\tan\left(\frac{\theta}{2}\right)\right| + 1\right) + K_2 \sin|\theta| \tag{2.124}$$

$$a_{nn}^{cyl,b} = \frac{-1}{4\pi G}\Bigg((\kappa-1)\sin\theta\, F\left(\tilde{\theta}\right) + \left(\frac{\kappa+1}{2}\right)\left(1+\cos\theta\, \ln\left(2-2\cos\theta\right)\right)+\cos\theta \Bigg) \tag{2.125}$$

$$a_{nn}^{cyl,c} = a_{nn}^{cyl,b} + \frac{1}{4\pi G}\frac{\cos\theta}{2} \tag{2.126}$$

$$a_{sn}^{cyl,a} = K_1 \sin\theta\, \ln\left|\tan\left(\frac{\theta}{2}\right)\right| + K_2 \cos\theta \,\mathrm{sgn}\left(\tilde{\theta}\right) \tag{2.127}$$

$$a_{sn}^{cyl,b} = \frac{-1}{4\pi G}\Bigg(\sin\theta - \left(\frac{\kappa+1}{2}\right)\sin\theta\, \ln\left(2-2\cos\theta\right) + (\kappa-1)\cos\theta\, F\left(\tilde{\theta}\right)\Bigg) \tag{2.128}$$

$$a_{sn}^{cyl,c} = a_{sn}^{cyl,b} + \frac{1}{4\pi G}\frac{\sin\theta}{2} \tag{2.129}$$

$$a_{nn}^{cav,a} = -K_1 \cos\theta\, \ln\left|\tan\left(\frac{\theta}{2}\right)\right| - K_2 \sin|\theta| \tag{2.130}$$

$$a_{nn}^{cav,b} = \frac{1}{4\pi G}\left((\kappa-1)\sin\theta\, F\left(\tilde{\theta}\right) - \left(\frac{\kappa+1}{2}\right)\cos\theta\, \ln\left(2-2\cos\theta\right)\right) \tag{2.131}$$

$$a_{sn}^{cav,a} = -K_1 \sin\theta\, \ln\left|\tan\left(\frac{\theta}{2}\right)\right| + K_2 \cos\theta \,\mathrm{sgn}\left(\tilde{\theta}\right) \tag{2.132}$$

$$a_{sn}^{cav,b} = \frac{-1}{4\pi G}\left(\left(\frac{\kappa+1}{2}\right)\sin\theta\, \ln\left(2-2\cos\theta\right) + (\kappa-1)\cos\theta\, F\left(\tilde{\theta}\right)\right) \tag{2.133}$$

The first superindex that identifies the IFs in the left side of the previous equations indicates the considered solid (cylinder or cylindrical cavity), and the second ($^{\cdots,l}$; with l = a, b or c) the load and supporting condition, according to the sub-figure with the same letter in the previous figures. The constants K_1 and K_2 used in these equations are defined in the following way:

$$K_1 = \frac{2(1-\nu)\,(1+\nu)}{\pi E} = \frac{(1-\nu)}{\pi G} = \frac{\kappa+1}{4\pi G} \tag{2.134}$$

$$K_2 = \frac{(1-2\nu)\,(1+\nu)}{2E} = \frac{\pi(1-2\nu)}{4\pi G} = \frac{\pi}{2}\frac{\kappa-1}{4\pi G} \tag{2.135}$$

With κ being Kolosov's constant, equal to $3 - 4\nu$ in plane strain, and G the shear modulus of the cylinder or the cavity. The function $F\left(\tilde{\theta}\right)$ is defined in Eq. (2.136), and the angle $\tilde{\theta}$ in Eq. (2.137).

$$F\left(\tilde{\theta}\right) = \frac{\tilde{\theta} - \pi \times \mathrm{sgn}\left(\tilde{\theta}\right)}{2} \tag{2.136}$$

$$\tilde{\theta} = \theta - 2\pi\left\lfloor \frac{\theta/\pi + 1}{2} \right\rfloor \tag{2.137}$$

The previous equations (2.124)–(2.133) differ slightly from the ones found in the literature (e.g. in [Woodward 1976], [Sundaram 2010 a and b], [Liu 2012] and [Liu 2013]), as explained next. In order to be able to use the previous expressions for any θ, the corresponding $\tilde{\theta}$ angle is defined inside the interval $(-\pi, \pi)$, according to Eq. (2.137), where $\lfloor\ \rfloor$ indicates rounding towards $-\infty$. Additionally, in the expressions for the cylinder and the cylindrical cavity subject to two normal, diametrically opposed forces, the absolute values of the tangents are taken, for the $a_{nn}^{\cdots,a}$ IFs the absolute values of the sines are taken, and for the $a_{sn}^{\cdots,a}$ IFs the cosines are multiplied by the sign of the $\tilde{\theta}$ angle. On the other hand, for the cylinder subject to two normal, diametrically opposed forces, with respect to the equations (4.36) or (B.10) and (4.37) or (B.11) of

[Woodward 1976], in the previous equations (2.124) and (2.127) the second term in the right side is corrected, omitting the division with π.

Regarding the a_{ss} IFs, it is verified that $a_{ss} = a_{nn}$ in each case, as it happens in the half-plane, except for a rigid body movement term for the cylinder, depending on the loading and supporting condition.

2.2.2.2 Sphere under opposing radial loads

The solution to the problem of the elastic sphere subject to a pair of normal and diametrically opposed point forces may be found in [Sternberg 1952], as cited in [Woodward 1976]. On the other hand, Titovich [Titovich 2012] presents a compact form of the IFs of the sphere loaded axisymmetrically, using quickly convergent series, and apart from the case of normal diametrically opposed point forces, the formulation for more general axisymmetric load distributions is developed. Next the IFs of the normal and tangential surface displacements of the sphere are reproduced, which may be found in [Woodward 1976] for the problem of the sphere loaded with a pair of normal and diametrically opposed point forces.

$$
\begin{aligned}
a_{nn}^{esf} = \frac{1}{4\pi GR}\Bigg\{ & \frac{m-2}{m+1} + \frac{m-1}{m}\left(\frac{1}{\sin(\theta/2)} + \frac{1}{\cos(\theta/2)} - 4\right) \\
& + \left(\frac{m-2}{m}\right)^2 \left(\ln\left(\cot\frac{\theta}{4}\cot\frac{\pi-\theta}{4}\right) - 2\right) \\
& - \frac{3m^2 - 20m + 16}{m^3}\left[\sin^2\frac{\theta}{2}\ln\left(\frac{1+\sin(\theta/2)}{\sin(\theta/2)}\right)\right. \\
& \left. + \cos^2\frac{\theta}{2}\ln\left(\frac{1+\cos(\theta/2)}{\cos(\theta/2)}\right) + \frac{1}{2} - \sin\frac{\theta}{2} - \cos\frac{\theta}{2}\right] \\
& + \sum_{k=1}^{\infty} A_{2k} P_{2k}(\cos\theta)\Bigg\}
\end{aligned}
\tag{2.138}
$$

$$a_{sn}^{esf} = \frac{1}{4\pi GR}\left\{\frac{m-2}{m}\left(\frac{\cos(\theta/2)-\sin(\theta/2)}{\sin\theta}\right)\right.$$

$$+\frac{m^2+8m-8}{m^2}\left[\left(1-\cos\left(\frac{\theta}{2}\right)-\sin\left(\frac{\theta}{2}\right)\right)\cot\theta\right.$$

$$\left.\left.+\frac{1}{2}\sin\theta\ln\left(\cot\left(\frac{\theta}{2}\right)\frac{1+\sin(\theta/2)}{1+\cos(\theta/2)}\right)\right]+\sum_{k=1}^{\infty}B_{2k}\frac{dP_{2k}(\cos\theta)}{d\theta}\right\} \qquad (2.139)$$

The IFs of Eqs. (2.138) and (2.139) are expressed in a curvilinear reference system, with direction *n* normal to the sphere in each point and pointing to the inside of the sphere, and the *s* direction tangential, meridional with respect to the "poles" defined by the points where the loads are applied, and with the same positive direction as the θ angle, which is the zenithal angle with respect to one of the points where the load is applied. The meaning of the $a_{ij}(\theta)$ IFs for the sphere is similar to that indicated in the previous section for the cylindrical geometries, being the θ angle bounded in this case in the (0, π) interval.

In the previous equations, *R* is the radius of the sphere, *G* its shear modulus, and *m* the inverse of the coefficient of Poisson *v*. $P_{2k}(\cos\theta)$ are the Legendre polynomials of degree $2k$ in the variable $\cos\theta$, and the coefficients A_{2k} and B_{2k} are defined in the following equations (2.140) and (2.141):

$$A_n = \frac{1}{m^4(n-1)(n+1)(n+2)\Delta'}\left\{\left(m^4+7m^3+30m^2-64m+32\right)n^2\right.$$

$$+\left(7m^4+22m^3-39m^2+44m-16\right)n$$

$$\left.+10m^4-11m^3+9m^2+20m-16\right\} \qquad (2.140)$$

$$B_n = \frac{1}{m^3(n-1)(n+1)(n+2)\Delta'}\left\{\left(7m^3+11m^2-28m+16\right)(n+2)^2 + \left(-18m^3-48m^2+132m-72\right)(n+2) + 9(m-1)\left(m^2+8m-8\right)\right\} \quad (2.141)$$

The variable Δ' that appears in the previous equations is defined as follows:

$$\Delta' = n(n-1)+(2n+1)\frac{m+1}{m} \quad (2.142)$$

The Legendre polynomials $P_n(x)$ of degree 0 and 1 are given by the functions $P_0(x) = 1$ and $P_1(x) = x$, and those of successive degrees may be calculated by their known recurrence law:

$$(n+1)P_{n+1}(x)-(2n+1)\,x\,P_n(x)+n\,P_{n-1}(x)=0 \quad (2.143)$$

On the other hand, the derivatives of the Legendre polynomials that appear in Eq. (2.139) may be evaluated differentiating the expression (2.143), which leads to Eq. (2.144):

$$(n+1)\frac{dP_{n+1}(\cos\theta)}{d\theta} = (2n+1)\times \left(-\sin\theta\times P_n(\cos\theta)+\cos\theta\times\frac{dP_n(\cos\theta)}{d\theta}\right)-n\frac{dP_{n-1}(\cos\theta)}{d\theta} \quad (2.144)$$

2.3 Numerical analysis of rolling contact

Generally, numerical methods lead to more precise contact solutions for problems that cannot be treated with analytical methods, at the expense of increased computational cost. Comprehensive surveys of numerical methods for general contact problems are provided in [Wriggers 1995], in the books

[Kikuchi 1988], [Laursen 2003], and [Wriggers 2006], and in the PhD Thesis of Yastrebov [Yastrebov 2011]. Here a brief review is provided focused on rolling contact problems.

The numerical computation of rolling contact problems was pioneered by Kalker with his well-known exact contact theory [Kalker 1979b], [Kalker 1990], introduced in §2.1.3. The method is aimed at solving contact problems in linear elastostatics, and was implemented first in the program DUVOROL and later in the program CONTACT. It is widely known in the railway community and used as a reference to validate simpler contact models. CONTACT has continued being developed by Vollebregt ([Vollebregt 2012], [CONTACT UG 2013], [Vollebregt 2020], [Vollebregt 2021]), and several authors have programmed their implementations as well, [Jin 2005], [Baeza 2011] and [Kaiser 2012] being some examples.

[Wang G 1993] and [Wang Z 2012] are other examples of methods similar to Kalker's exact contact theory for rolling contact problems. In [Wang G 1993], the 2D problem of steady-state rolling of two viscoelastic cylinders was analysed. In [Wang Z 2012], 3D steady-state rolling problems with surfaces including micro-irregularities were analysed. The method used here was limited to cases with zero spin.

Some authors classify the exact contact theory as a boundary element method (BEM). Others (e.g. [Wang Z 2012]) designate this type of method as a semi-analytical method (SAM), referring to the calculation of the ICs, which relate surface displacements and stresses. This type of method is indeed commonly used for half-space contact problems, and the half-space ICs are calculated analytically. However, work has been done to extend the exact contact theory to more general geometries with numerically calculated ICs (see e.g. [Li Z 2002], [Vollebregt 2014a]), so the designation of SAM, in this case, is now less appropriate. Here, the designation of 'contact element method' (CEM) will be used for methods like Kalker's exact contact theory, in which just the contact surface is discretized, to distinguish them from the BEM in which the whole surfaces of the contacting solids are discretized. The latter was done e.g. in [Abascal 2007], to solve the structural (global) and the contact (local) problems in a coupled way. The CEM designation encompasses methods in which the ICs

are either known a priori or are fed as input data (not being necessarily calculated analytically), extending the SAM designation.

Rolling contact problems have also been studied with BEM. In [González 1998] and [González 2000] a method was developed to analyse 2D steady rolling contact problems with BEM. This was subsequently extended to 3D, for a deformable wheel on a rigid base in [Abascal 2007], and allowing for the two contacting bodies to be deformable in [Rodríguez-Tembleque 2010]. In the latter work, each body could be meshed with either BEM or FEM. The methodology developed in these references allows for studying coupled structural and contact problems, though the development was aimed at linear elastostatic problems. In [González 1998], [González 2000] and in one of the procedures presented in [Abascal 2007], the equilibrium equations were condensed to the degrees of freedom in the PCS, which could be classified as a CEM wherein the ICs are calculated numerically with BEM.

CEM methods are best suited to contact problems not coupled with the structural problem, as is the case in linear elastostatic problems in which the remote loads are not influenced by the specific distribution of contact stresses. In cases in which this premise is not fulfilled due e.g. to large displacements or inelastic material behaviour, FE models are used to get numerical solutions to rolling contact problems. FEM has found much wider application in the analysis of contact problems than BEM, being more suited to general non-linear problems. FEM for contact mechanics analyses commonly involve much more degrees of freedom than other contact analysis methods and higher computational costs, as the volume of the contacting solids has to be meshed, and not just the contact surface. In turn, they offer more versatility to include in the analysis complex material behaviour, non-linear geometry, and dynamic effects, for example.

When the rolling is steady state, it is advantageous to set out the problem in a static mesh relative to the motion of the wheel axis of revolution, where the problem can be described in a purely spatial way, without explicit time dependence. In this setting, the displacement fields are decomposed into a part of rigid body motion, which is described in an Eulerian way, and a part of deformation, which is described in a Lagrangian way. This has the advantage of limiting the mesh refinement for detailed contact analysis to the region occupied

by the contact patch and of avoiding time stepping. This approach has been used mainly in the tire–road application. The tire undergoes large deformations and has a complex structure and inelastic material behaviour, while the road is commonly assumed to be rigid. This methodology was applied for 2D in [Zeid 1981], [Padovan 1984], and [Oden 1986] and extended to 3D in [Bass 1987], [Padovan 1987], [Oden 1988], [Faria 1989, 1992], [Nackenhorst 1993], and [Le Tallec 1994]. In this framework, the inertial effects are accounted for, in addition to non-linear geometry and material behaviour. [Nackenhorst 1993] noted the relationship between this type of analysis and the Arbitrary Lagrangian Eulerian (ALE) methods ([Donea 2004]), widely used in fluid–structure interaction problems, among others. [Kabe 2000], [Nackenhorst 2000], [Hu 2002], [Damme 2003], [Nackenhorst 2004], [Ziefle 2008], and [Rafei 2018] are further examples of its development and application in 3D rolling contact problems. In [Kabe 2000], comparisons between solutions obtained for steady rolling contact with this methodology and with Lagrangian simulations were made, showing the substantial computational advantage of the former.

In [Nackenhorst 2000] and [Damme 2003], the application of the ALE formulation for 3D wheel–rail rolling contact analyses was demonstrated, where the two contacting bodies have to be regarded as deformable. As stated in [Nackenhorst 2004], future work has to be spent on the development of reliable and efficient numerical algorithms. A difficulty of the ALE formulation for rolling contact problems, which is relevant also in the case of a single deformable body, is that the particle path-dependent variables are not computed inherently. This becomes necessary for the representation of the kinematics of the tangential contact problem and inelastic material behaviour. These issues were addressed in [Ziefle 2008], and the issue related to the tangential contact problem, viz the computation of the convective term of the contact slip velocities, in [Laursen 2006] as well. In later works using the ALE formulation ([Suwannachit 2013], [Behnke 2015], [Draganis 2015], [Wollny 2016], [Draganis 2017]), thermo-mechanical rolling contact problems have been studied.

For transient rolling contact problems, a total Lagrangian description seems preferable [Wriggers 2001]. [Telliskivi 2001], [Li Z 2008], [Chongyi 2010], [Zhao X 2011], [Bian 2013], [Pletz 2014b], [Vo 2014], [Vo 2015], [Zhao X

2014], [Zhao X 2015], [Yang 2016], [Toumi 2016], and [Gao 2022] are examples of Lagrangian 3D wheel–rail rolling contact analyses in which detailed frictional contact solutions were computed with FEM. This type of analysis involves a non-linear time-stepping solution, which is computationally costly.

2.4 Conformal contact

Early analyses of conformal contact are due to [Steuermann 1940] and [Persson 1964]. They studied the two-dimensional problem of conformal contact of bodies with cylindrical geometry. In [Goodman 1965] the contact between a sphere and a spherical seat was studied. Being in this case the geometry axisymmetric, the problem is also 2D. 2D conformal cylindrical contact is characteristic of pinned joints, which are present in numerous engineering applications. Following the work of Persson, in [Ciavarella 2001] the formulation of the 2D conformal cylindrical contact without friction was developed, obtaining the load–contact angle variation relationship in closed form, and extending the range of validity of the formulation to any value of Dundurs' first material parameter.

The development presented in [Ciavarella 2001a] was extended for the case of non-zero Dundurs' second material parameter in [Ciavarella 2001b], solving numerically the governing integral equations of the problem. Further, an analytical approximation was proposed for the load–contact angle variation relationship, based on the assumption, verified by the numerical calculations in the work, that the impact of Dundurs' second material parameter on the pressure distribution (for a given contact angle variation) can be neglected. The analytical load–contact angle variation relationship was also derived in [Hou 2001] for the case of identical materials in contact, and a fracture mechanics study was conducted considering a radial crack emanating from the surface of the circular hole of the plate. The main analytical results for 2D frictionless cylindrical contact of the above-cited references is summarized next.

A cylinder of radius R_1 is considered journalled in an infinite plate or cavity with a hole of radius R_2, as shown in Figure 2.24. The difference in radius $\Delta R = R_2 - R_1$ is positive for a clearance fit, and negative for an interference fit. The cylinder is designated as body 1, and the cavity as body 2. Being the Young's

modulus of body k E_k and its coefficient of Poisson ν_k, the elastic constants G_k, E^*_k and κ_k of the same body are calculated according to Eqs. (2.145)–(2.147). These are the shear modulus, the equivalent Young's modulus for plane strain, and Kolosov's constant for plane strain of body k, respectively.

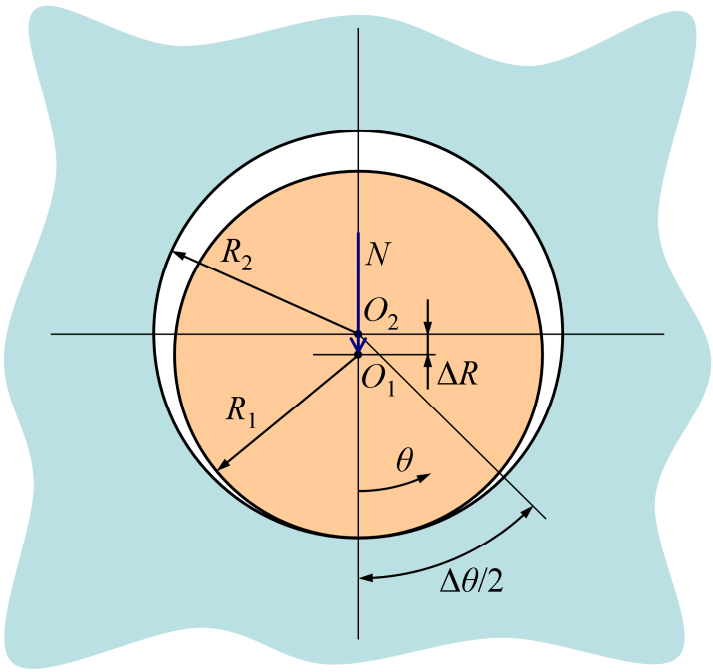


Figure 2.24. 2D cylindrical contact problem.

$$G_k = \frac{E_k}{2(1-\nu_k)} \tag{2.145}$$

$$E^*_k = \frac{E_k}{1-\nu_k^2} \tag{2.146}$$

$$\kappa_k = 3-4\nu_k \tag{2.147}$$

Dundurs' first and second material parameters α and β for the pair of materials in contact are defined according to Eqs. (2.148) and (2.149).

$$\alpha = \frac{\frac{\kappa_1 + 1}{G_1} - \frac{\kappa_2 + 1}{G_2}}{\frac{\kappa_1 + 1}{G_1} + \frac{\kappa_2 + 1}{G_2}} \tag{2.148}$$

$$\beta = \frac{\frac{\kappa_1 - 1}{G_1} - \frac{\kappa_2 - 1}{G_2}}{\frac{\kappa_1 + 1}{G_1} + \frac{\kappa_2 + 1}{G_2}} \tag{2.149}$$

When a radial load N is applied on the pin, a contact patch is formed spanning a total angle $\Delta\theta$. The resulting contact surface may be assumed to be cylindrical as well, with the radius R given in Eq. (2.150). A non-dimensional load parameter may be defined as $E^*_1 \, \Delta R / N$.

$$R = \frac{E^*_1 \, R_1 + E^*_2 \, R_2}{E^*_1 + E^*_2} \tag{2.150}$$

The contact angle spanned by the contact patch $\Delta\theta$ is found by iterative solution of Eq. (2.151). In this equation, $b = \tan(\Delta\theta/4)$.

$$\frac{E^*_1 \, \Delta R}{N} = \frac{(\alpha - 1)\left(\ln\left(1 + b^2\right) + 2b^4\right) + 2}{\pi(1 + \alpha)\left[1 + b^2\right] b^2} - \frac{4\beta}{\pi(1 + \alpha)} \tag{2.151}$$

As $E^*_1 \, \Delta R / N$ goes to 0 (i.e., as N goes to infinity and/or ΔR goes to 0), $\Delta\theta$ tends to a limiting value, being the same for clearance and interference contact. In the case of interference contact, the limiting value of the load, N_{lim}, at which the contact begins to recede at $\theta = \pi$ is given in Eq. (2.152).

$$N_{\text{lim}} = \frac{\pi \, (\alpha + 1)}{2 \, (\alpha - 1 - 2\beta)} \; E^*_1 \, \Delta R \tag{2.152}$$

The contact normal pressure p_n around the cylindrical contact is given by Eq. (2.153) for $N < N_{lim}$ in an interference contact, and by Eq. (2.154) for both a clearance contact and an interference contact with $N > N_{lim}$. In these equations,

$b' = \tan(\theta/4)$. Eqs. (2.151)–(2.154) may be used in plane stress with appropriate modification of the elastic constants (cf. §2.2.2.1). For the case of clearance, the results for cylindrical contact converge towards the Hertzian solution for $\theta \to 0$. As θ increases, the cylindrical contact tends to become more concentrated than predicted with Hertzian theory.

$$p_n R = \frac{-4 G_1 G_2}{2 G_1 + G_2(\kappa_1 - 1)} \min(0, \Delta R) + \frac{N}{\pi}\cos\theta \tag{2.153}$$

$$\frac{p_n R}{N} = \frac{2\sqrt{\max(0, b^2 - b'^2)}}{\pi\sqrt{1+b^2}(1+b'^2)} + \frac{1}{2\pi b^2(1+b^2)}$$

$$\times \ln\left(\frac{\sqrt{1+b^2} + \sqrt{\max(0, b^2 - b'^2)}}{\sqrt{1+b^2} - \sqrt{\max(0, b^2 - b'^2)}}\right) \tag{2.154}$$

Double conforming cylindrical contacts without friction, with an intermediate annular elastic body between the pin and the infinite elastic plate, were studied in [Liu 2012] and [Liu 2013]. The displacements of the pin and the plate were calculated making use of their corresponding Green's functions, and those of the ring were calculated by means of a Fourier series technique. In [Liu 2013] the model was extended to cope with interference in the contact, and the influence of geometric irregularities on the resulting contact pressure distributions was studied.

In [Sundaram 2010a] and [Sundaram 2010b] the formulation for the 2D conformal cylindrical contact problem with friction was developed, for the case of rigid and elastic pin, respectively. The influence of remote stresses in the plate was incorporated in the analysis. The analytical Green's function of the displacements of the pin and the hole subject to boundary point loading were used, and a numerical method was proposed to solve the governing singular integral equations of the problem, valid provided the contact is not split in more than one contact patch.

Other references of studies about 2D conformal contact may be found in [Fagan 2001] and [Ciavarella 2006]. In [Fagan 2001], an overview is provided of

analytical and numerical methods to study conformal contact problems, and a few references of 3D conformal contact analyses are provided as well.

The works about 3D conformal contact in the literature are comparatively few compared with 2D contact. [Woodward 1976], [Paul 1979] and [Paul 1981] are early examples of wheel–rail frictionless conformal contact studies. In [Marshek 1984], numerical solutions were obtained of 3D frictionless contact with cylindrical surfaces. In these works, a discretization of the non-planar contact surface was used, which can be classified as a CEM method. Considering only frictionless surfaces reduces the complexity of the contact problem significantly [Woodward 1976]. Thus, frictional contact has been less studied than frictionless contact.

In [Piotrowski 1999] an analysis procedure for conformal rolling contact problems was presented. The normal contact problem was solved with a simplified procedure assuming an elastic behaviour of the contacting solids similar to the elastic half-space. The tangential contact problem was solved with FASTSIM, accounting for the non-flat geometry in the calculation of the rigid slip velocities. A simplified method based on the strip theory was used in §8.5 of [Johnson 1987] to study conformal rolling contact problems as well. STRIPES (cf. §2.1.1.3) is another rolling contact algorithm, able to deal with non-planar contacts. A distinct approach is used by Pascal ([Pascal 2007a], [Pascal 2016], [Pascal 2019]), by means of a multi-Hertzian technique. More recently, other researchers have addressed the development of simplified wheel–rail conformal contact models for vehicle dynamics simulations ([Boccini 2016], [Marques 2018], [Marini 2022]).

A version of Kalker's exact contact theory extended for conformal contact was developed by Li [Li Z 2002] and was used in [Burgelman 2014] in a case study of conformal rolling contact. Vollebregt has also extended Kalker's exact contact theory implemented in the CONTACT programme for conformal contact ([Vollebregt 2014a], [Vollebregt 2018], [Vollebregt 2021]).

Both in [Li Z 2002] and in [Vollebregt 2014a], comparisons were presented between rolling contact results obtained in cases with different conformity levels with the exact contact theory extended for conformal contact on the one hand, and with simpler analysis methods on the other hand. In [Li Z 2002], a wheel running on a conforming groove was considered, and the results obtained

with the strip theory were used for comparison. In [Vollebregt 2014a], 800 wheel–rail contact cases with different combinations of longitudinal and lateral creepages were considered. In this work, the results obtained with the exact contact theory assuming planar Hertzian geometry were considered for comparison, and the contribution of different sources to the differences between the planar and the conformal results was studied.

Other examples of 3D conformal contact analysis are cited in §2.5, referred to the rolling bearing application.

2.5 Contact mechanics analysis in rolling bearings

Contact analysis methods of different complexity are used in rolling bearings depending on the application and purpose of the study. [Oh 1976] and [Hartnett 1979] are early examples of numerical solutions of the normal part of 3D contact problems in rolling bearings, based on a discretization of the contact surface. In these works, the PCS was taken as flat, and the contact surfaces were assumed to behave like half-spaces in the normal direction. The computational cost grew rapidly with the refinement of the mesh, as indicated in [Oh 1976].

FE models are also used to obtain numerical contact solutions in rolling bearings. Some examples may be found in [Ju 2000], [Bomidi 2014], [Li F 2018] and [Abdullah 2020]. In [Ju 2000], the contact pressure profiles in the roller–raceway contact obtained with rollers with different crowned profiles were studied with 3D FEM. In [Bomidi 2014], detailed simulations of the initiation and propagation of subsurface initiated rolling contact fatigue (RCF) cracks were carried out with 3D elastic-plastic FEM. Evolving contact pressure profiles between a flexible half-space with growing cracks simulating the raceway and a rigid roller were calculated, and it was shown how taking into account more realistic contact pressure profiles in the analysis instead of purely Hertzian prescribed ones could have a considerable impact in the computed RCF life, especially in heavily loaded contacts. The frictional contact between the crack faces was explicitly included in the simulations. On the other hand, a full slip situation was considered in the roller–raceway contact, prescribing the contact shear stresses as directly proportional to the normal pressures with an assumed coefficient of friction. In [Li F 2018], RCF simulations were carried out as well. In this case, a 2D explicit FE model was used to simulate the line

contact between a cylindrical roller and a raceway in plane strain, modelling both bodies as elastic. A friction coefficient was specified for the contact in the analysis, but a situation of pure rolling was considered, so nearly frictionless contact solutions were obtained. In [Abdullah 2020], experimental and numerical results of a 4-ball RCF test were reported. The numerical results were obtained with a 3D FE model with which the frictionless contact between an elastic-plastic ball and rigid plates rotating over the contact track of the ball was simulated.

Regarding the contact mechanics part of the simulation, the above-cited works dealt mainly with the solution of the normal part of the rolling element–raceway contact. Frictional contact problems involve a greater difficulty than frictionless contact problems, and simplified contact models are usually employed to facilitate their solution. A commonly used simplifying assumption is that of full slip, wherein the tangential stresses are determined as the normal pressures times a coefficient of friction. This is used for instance in [Houpert 1999], [Houpert 2002] and [Leblanc 2007], following [Jones 1959]. As explained in [Houpert 1999] and [Houpert 2002], the coefficient of friction is actually variable across the contact (see also [Kleckner 1982]). Still, a single value may be used for the full contact provided it is properly calculated according to the average contact pressure and sliding speeds in the contact. Besides the shearing of the lubricant film between the contact surfaces, the possible contribution of the solid direct contact between the surfaces should be considered. The full slip approach is also commonly used in rigid body dynamics simulations of other mechanical systems involving frictional contacts, to achieve fast computation of the contact resultant forces and moments ([Zhuravlev 1998], [Leine 2003], [Kireenkov 2008], [Karapetyan 2009], [Kudra 2013], [Zobova 2019]).

The underlying assumption in full slip tangential contact solutions is that the elastic displacement gradients of the contacting surfaces are small in relation to the rigid slip velocities and can be neglected. The surfaces are thus regarded as rigid in the tangential direction of the contact. However, the effects of the tangential elasticity of the bodies may be significant in certain situations, as shown in [Halling 1966–67] and §8.5 of [Johnson 1987]. In these works, the tangential contact problem was solved with the strip theory (cf. §2.1.2.3). The strip theory was also used in [Oloffson 2000] and [Meehan 2017] to obtain

tangential contact solutions to calculate wear profiles in spherical roller thrust bearings and spherical roller bearings, respectively. This being an analytical solution, its computational costs are much lower than those of numerical methods. On the other hand, it is limited to cases in which the direction of the rigid slip velocities is approximately constant in the rolling direction, as it was initially conceived for situations with pure creepage.

The FASTSIM algorithm [Kalker 1982] of Kalker's simplified rolling contact theory is another simplified partial slip model for the tangential contact problem (cf. §2.1.2.6), which overcomes the limitation of constant direction of the rigid slip velocities along the contact. This model was applied in [Legrand 1995] and [Heras 2018] to different types of ball bearings, seeking to improve the tangential contact solutions obtained with the usual full slip approach.

In more recent works ([Tonazzi 2017], [Heras 2017a], [Schwack 2018], [Schwack 2021], [Fallahnezhad 2019]), detailed frictional contact solutions in rolling bearings have been computed with FEM in Lagrangian, time-stepping simulations. In these works, it was relevant to obtain a detailed solution of the tangential contact problem in order to get representative friction torques, plastic or wear profiles in the computations. A relevant difficulty in rolling contact analyses of rolling bearings with respect to the wheel–rail or tire–road applications is that in many cases more than two contacting bodies have to be modelled.

In [Tonazzi 2017], a 3D elastic-plastic FE model of the ball–raceway contact region of an oscillating ball bearing was used to simulate oscillating rolling contact cases with different levels of friction, focusing the analysis in the contact of the ball with the inner raceway. It was shown that the tangential contact stresses altered the equivalent plastic strain distributions around the contact. [Heras 2017a] studied the friction torque in four-point contact slewing bearings, comparing the results obtained with a detailed 3D FE model against the full-slip solution. In some of the studied cases significant differences were observed in the tangential stress and slip distributions computed with the different models, as well as non-negligible differences in the resulting friction torques. In [Schwack 2018], the patterns of slip and frictional dissipation which may lead to a fretting wear damage termed as "false brinelling" in stationary bearings subject to vibration or to oscillatory motion were studied in an angular

contact ball bearing under oscillatory motion with different coefficients of friction and bearing geometries. The study was carried out both numerically, with 3D elastic FE models, and experimentally, observing good correspondence between the computed frictional work distributions and the fretting marks observed in the physical tests. The same modelling approach has been used in [Schwack 2021] to study the effect of the bearing oscillation amplitude on the accumulated frictional work densities in the ball–raceway contacts. In [Fallahnezhad 2019], 2D FE models were used to study false brinelling in a cylindrical roller bearing under different oscillatory shift and rolling contact cases that could take place during transportation. The patterns of stick and slip areas in the contact were shown to influence the wear profiles, causing W-shaped profiles in some of the studied cases.

Chapter 3

Vehicle–track dynamic interaction in different tracks

Chapter summary

In this chapter, models to study the vertical dynamic vehicle–track interaction in the time domain are developed. Two commercial analysis software packages are used for this purpose: a multibody system (MBS) analysis software, and a Finite Element Method (FEM) analysis software. As a result, a combined MBS-FEM representation of the vehicle–track model is integrated into the MBS software, which allows for the study of dynamic phenomena in a wide frequency range. Other simpler approaches for modelling the dynamic vehicle–track interaction are also considered, such as pure multibody or FE representations of the whole vehicle–track system. The quality of the results obtained with the different types of models used is analysed, and some conclusions are put forth regarding the possible validity of simpler train–track interaction model types under certain conditions, as well as the most suitable configuration of the most complex models. The developed models are used to analyse and compare the dynamic performance of a ballasted track and three types of slab track.

3.1 Description of the studied track types

Four track types have been considered in the studies presented here: a ballasted track, and three types of slab tracks: the RHEDA 2000 track, the STEDEF track, and a floating slab track. In the RHEDA 2000 and STEDEF tracks, the slab is directly supported on the soil, and in the floating slab track, the slab is isolated from the soil by means of an elastomeric mat.

Ballast has been used since the beginning of railways to serve as a transition element between the sleepers and the soil, providing compliance and vibration damping, as well as surfacing and draining capabilities to the track. The increased maintenance costs and reduced life cycle of the track associated with higher transportation speeds, axle loads, and traffic densities led to the appearance of the slab track in the 1960s.

Some of the advantages of the slab track with respect to the ballasted track are its higher geometric stability and reduced maintenance costs. The main disadvantage of the slab track is its higher installation cost. Apart from that, railway administrations and design engineers often prefer to use traditional ballasted track designs instead of newer slab track designs, because greater experience has been gained with the use of the former track type.

Within the different slab track types, floating slab tracks provide and effective way to reduce vibration transmission form railway traffic to the ground [Lombaert 2006], [Hussein 2006], [Kuo 2008]. By adding an elastic layer beneath the slab, the natural frequency of the system is reduced at the expense of increasing its cost.

The relevant parameters that have been used in the simulations for each of the studied track types are usual parameters of new high-performance tracks in Europe, and are listed in Table 3.1. These are typical values based on standard configurations of each type of track and may vary within a certain range. Consequently, the results and conclusions presented in this study must be understood as applicable specifically with the parameter values chosen here, and may not be extended straightforwardly to other possible configurations of each type of track.

Table 3.1. Parameters of the different track types considered in the simulations.

Parameter	Ballasted	RHEDA 2000	STEDEF	Floating slab
Railpad vertical stiffness [kN/mm]	100	24.6[e]	150	213.2
Railpad vertical (relative) damping [kN s/m] or [%]	15	10%[f]	10%	14.8
Sleeper mass [kg]	320	138	221.7	—
Rail support spacing [m]	0.60	0.65	0.60	0.60
Sleeper support vertical stiffness [kN/mm]	100	—	50	—
Sleeper support vertical relative damping [%]	20%	—	10%	—
Slab cross-section dimensions (width×height) [m]	—	3.2×0.24	2.9×0.49	2.5×0.55
Slab support stiffness per unit area [kN/m^3]	—	—	—	15000
Slab support damping per unit area [kN s/m^3]	—	—	—	30
Slab concrete Young's modulus [GPa]	—	34	29	30
Slab concrete shear modulus [GPa]	—	14.2	12.1	11
Slab concrete density [kg/m^3]	—	2500	2500	2500
Slab concrete relative viscous damping [%]	—	1%	1%	1%

[e] Combined stiffness of the upper railpad and lower base plate pad in series. The stiffness values for each of the two pads are 450 kN/mm and 26 kN/mm, respectively.

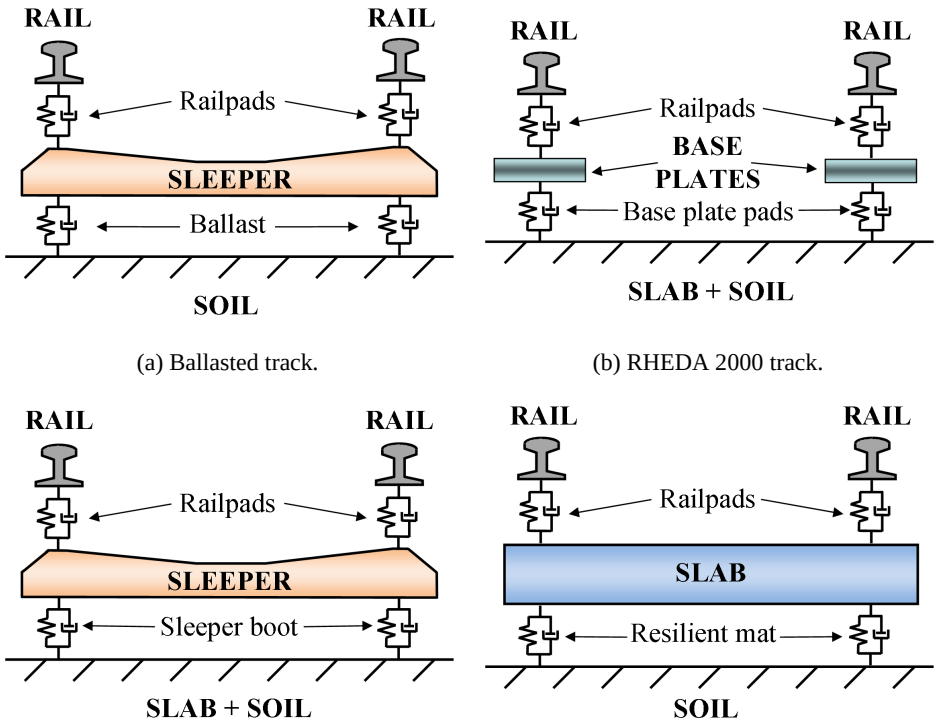
[f] For both the railpad and the base plate pad.

In all the cases, the modelled tracks are considered to rest on an infinitely stiff soil. The parameters chosen for the ballasted track correspond to a high-performance ballasted track adequate for high-speed traffic.

The sleeper support vertical stiffness and relative damping in Table 3.1 refer to the properties of the elastic level below the sleeper, i.e. the ballast for the ballasted track and the elastomeric material above which each sleeper rests in the STEDEF track.

As for the rail properties, a 60E1 rail section is considered in all cases. The mechanical properties of the rail material and the geometric properties of the rail section are the same as those given in [Wu 2004].

Figure 1.5 shows the dynamic models of the different track types studied. A brief description of the three studied slab track types is given in the following subsections.



(a) Ballasted track. (b) RHEDA 2000 track.

(c) STEDEF track. (d) Floating slab track.

Figure 3.1. Cross-sections of the dynamic models of the studied tracks.

3.1.1 RHEDA 2000 track

This is the type of slab track that has been most used in Europe. The rails are discretely supported above each sleeper, and the sleepers are directly embedded in the slab, forming a monolithic assembly. Therefore, most of the vertical flexibility and damping of the track is provided in the rail fastening system, which is usually of the type VOSSLOH IOARV 300 [VOSSLOH www]. This

fastening system comprises two elastic levels: one is given by the railpad, located between the rail and the base plate, and the other by the base plate pad, located between the base plate and the sleeper. Therefore, for modelling this track, two elastic layers are considered between the rails and the slab. The base plate has a mass of 4.4 kg.

The reinforced concrete slab of this track is continuous and is executed on the field. The slab rests on a layer of gravel and cement or lean concrete, which are materials of very high stiffness compared to the elements in other elastic levels of the track. Therefore, the slab can be considered to be directly supported on the soil with no elasticity beneath it.

3.1.2 STEDEF track

In the STEDEF slab track, there are two elastic levels above the slab. One, like in the other track types studied, is given in the rail fastening system. In this case, the rail is also discretely supported above each sleeper. Different options are available for the rail fastening system, the simplest being the NABLA type. This fastening system has a single elastic level.

The second elastic level above the slab is given between the sleepers and the slab, since these are not directly built into the slab like in the RHEDA 2000 track, but they rest above an elastomeric material.

As in the case of the RHEDA 2000 track, the reinforced concrete slab of the STEDEF track is continuous and executed on the field. In this case again, the slab rests on a layer of gravel and cement or lean concrete, and no elastic level is considered beneath it.

3.1.3 Floating slab track

The application of floating slab tracks in mainly focused to reducing vibration transmission from railway traffic to the ground [Lombaert 2006], [Hussein 2006], [Kuo 2008]. The slab is isolated from the ground by means of a resilient layer, which may consist of single support bearings, steel springs, or a continuous elastomeric mat. Therefore, unlike in the other two types of slab tracks previously described, in this type of track, an elastic layer is considered beneath the slab.

By maximizing the resilience of the elastic layer beneath the slab, the cut-on frequency of the system is reduced, and a more effective reduction in the transmission of vibrations to the ground can be achieved. The slab mat resilience, however, is limited by the maximum allowable static rail deflection [Lombaert 2006].

Above the slab, the track has a single elastic level given by the discrete support pads between the rails and the slab.

The parameters chosen for this track are taken from [Lombaert 2006]. These parameters correspond to a stiff pad for the rails and soft mat for the slab. This combination of parameters can be adequate to reduce wheel–rail force and track deflection, to ensure riding quality and train stability and to reduce force transmission to the subgrade due to environmental concerns [Kuo 2008].

3.2 Description of the vehicle–track interaction models

In this study, different types of models have been constructed for the study of vertical vehicle–track dynamic interaction in the time domain, that account for the elasticity of the track, combining the use of the commercial FEM software NASTRAN and the commercial multibody system (MBS) software SIMPACK. Three different approaches have been followed for this purpose, as explained in the next subsections. The use of a commercial MBS software makes it possible to include, in a reliable way, models of advanced non-linear wheel–rail contact as well as complex elements or joints in the vehicle model. On the other hand, the FEM allows taking into account the flexibility of continuous longitudinal track components (the rail and the slab).

3.2.1 Rigid multibody models

Rigid multibody (RMB) is the simplest approach to model a vehicle–track interaction model in a MBS environment, and is the standard option that is available in most MBS analysis programs, including SIMPACK, for modelling an elastic track.

The idea consists of employing discrete elements of mass, stiffness, and damping for modelling the different elastic levels of the track in a way that is

similar to the one for the vehicle. A group of lumped mass, spring, and damper elements are placed below every axle of the vehicle, each of those groups always remaining below the axle to which it is associated [Bezin 2009]. Therefore, each of those groups travels along the running path of the vehicle at the same speed as each of the vehicle's axles.

The physical parameters of mass, stiffness, and damping assigned to each lumped element beneath the vehicle's wheelsets are found in this study by adjusting the vertical receptance of the mass-spring-damper assembly beneath each wheelset to that of the track, the latter being calculated in a frequency analysis with a FE model of the track. The FE track models used for this purpose are the same as the ones used in the FE vehicle–track interaction models, described in §3.2.2.

The models of all the tracks considered have two elastic levels. Therefore, the discrete assemblies representing the track beneath each wheelset are modelled as a two-degree of freedom (DOF) system, as depicted in Figure 3.2 (in the figure, the horizontal displacement is the only DOF of each of the two masses of the system; in the RMB models the DOF of each mass of the track model is the vertical displacement). The mass designated as m_1 in the figure represents the rail level, and the mass designated as m_2 the sleeper in the ballasted and STEDEF tracks, the base plate pad in the RHEDA 2000 track, and the slab in the floating slab track.

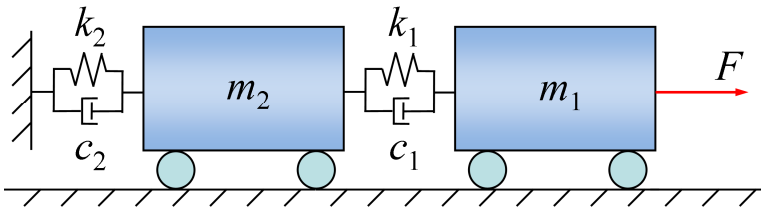


Figure 3.2. 2-DOF dynamic system representing the track beneath each wheelset in the RMB models (depicted with horizontal instead of vertical displacement as the DOF of each mass).

The dynamic matrix **[D]** of the 2-DOF system shown in Figure 3.2 is given in Eq. (3.1). In this equation, **[K]**, **[C]** and **[M]** are the stiffness, damping and mass matrices, ω the forcing frequency, and $i = \sqrt{-1}$.

$$[\mathbf{D}] = [\mathbf{K}] + \mathrm{i}\omega\,[\mathbf{C}] - \omega^2\,[\mathbf{M}]$$

$$= \begin{bmatrix} \left(k_1 - \omega^2 m_1\right) + \omega\, c_1\,\mathrm{i} & -k_1 - \omega\, c_1\,\mathrm{i} \\ -k_1 - \omega\, c_1\,\mathrm{i} & \left(k_1 + k_2 - \omega^2 m_2\right) + \omega\left(c_1 + c_2\right)\mathrm{i} \end{bmatrix} \tag{3.1}$$

The receptance $R(\omega)$ is calculated as the steady-state displacement amplitude of m_1 when a unitary sinusoidal force is applied on it, as depicted in Figure 3.2. Therefore, the force vector is $(1, 0)^T$. Applying Cramer's rule, $R(\omega)$ may be written analytically according to Eq. (3.2), where $|\mathbf{D}|$ stands for the determinant of **[D]**. The analytical expressions for the real and imaginary parts of $|\mathbf{D}|$ are given in Eqs. (3.3) and (3.4), respectively. For the undamped system, $\mathrm{Im}(|\mathbf{D}|) = 0$, and the natural frequencies may be found as the two roots of the biquadratic equation $\mathrm{Re}(|\mathbf{D}|) = 0$.

$$R(\omega) = \frac{\left(k_1 + k_2 - \omega^2 m_2\right) + \omega\left(c_1 + c_2\right)\mathrm{i}}{|\mathbf{D}|} \tag{3.2}$$

$$\mathrm{Re}\left(|\mathbf{D}|\right) = m_1 m_2 \omega^4 + \left(-k_1 m_2 - c_1 c_2 - m_1\left(k_1 + k_2\right)\right)\omega^2 + k_1 k_2 \tag{3.3}$$

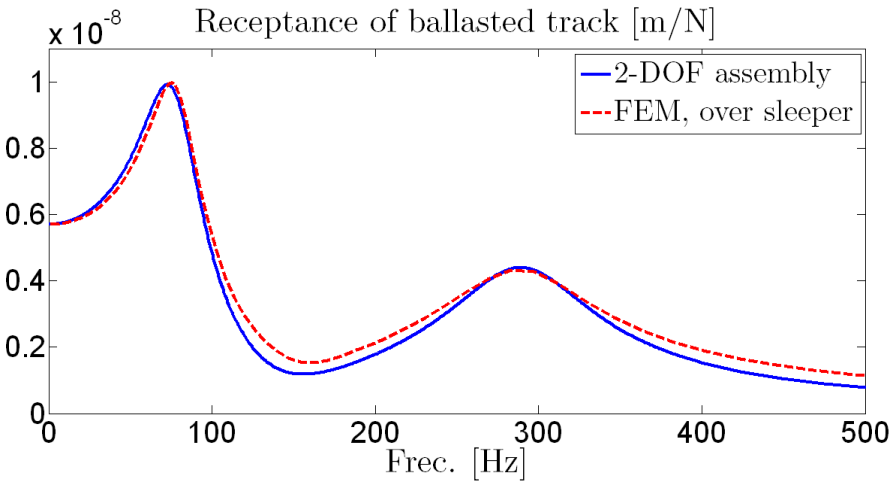
$$\mathrm{Im}\left(|\mathbf{D}|\right) = \omega\left(k_1 c_2 + k_2 c_1 - \left(c_1 m_2 + m_1\left(c_1 + c_2\right)\right)\omega^2\right) \tag{3.4}$$

For each track, the adjustment of parameters of the 2-DOF assemblies beneath each wheelset to approximate the corresponding reference track receptance calculated with a FE model of the track is carried out following these steps:

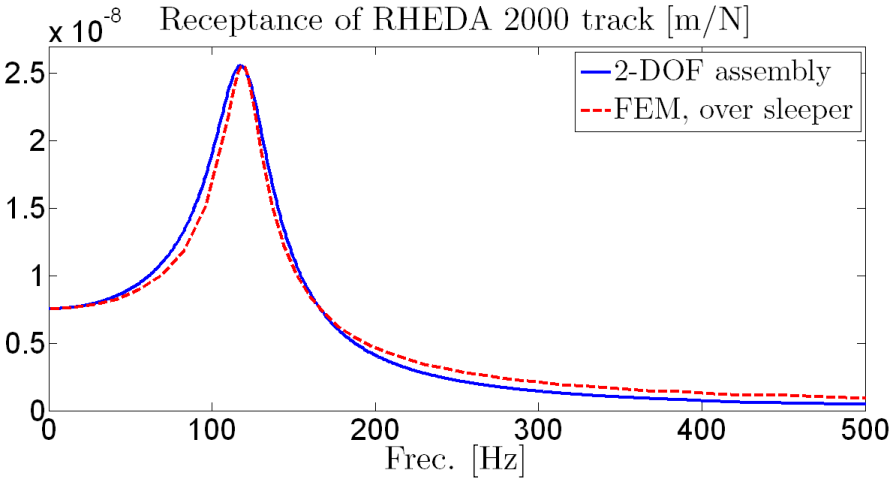
- The parameters of mass, stiffness and damping of the track corresponding to a single sleeper bay are calculated. These parameters are identified with the appendix "*,sb*" in their subindex, and may be calculated with the data given in Table 3.1. For example, $m_{1,sb}$ is calculated as the mass of the two rails in the sleeper spacing length, and $m_{2,sb}$ is the mass of a sleeper in the case of the ballasted track, the mass of two base plate pads in the case of the RHEDA 2000 track (one for each rail of the track), and the mass of the slab in the sleeper spacing length in the floating slab track.

- An equivalent number of sleeper bays is defined for each elastic level of the 2-DOF system, n_1 and n_2. These may be interpreted as a measure of the track lengths participating in the track motion in the frequency range in which the receptance adjustment is carried out. The parameters of mass of the 2-DOF system are calculated as $m_i = n_i \times m_{i,sb}$, and similarly for k_i and c_i, with $i \in \{1, 2\}$.
- n_1 and n_2 are adjusted iteratively, so that the track receptance of the equivalent 2-DOF system calculated analytically according to Eqs. (3.2)–(3.4) matches as well as possible the reference track receptance calculated with FEM. Specifically, it is aimed at adjusting the modulus of the track receptance, in the frequency range of roughly 0 to 500 Hz.
- The relative damping of one or both levels of the equivalent 2-DOF system are modified if necessary to improve the receptance adjustment. Here, the relative damping ξ_i is defined as $\xi_i = c_{i,sb}\ \omega_i / (2\ k_{i,sb})$, with $i \in \{1, 2\}$. ω_i are the natural frequencies of the undamped 2-DOF system (with the parameters corresponding to a single sleeper bay), ω_1 being the natural frequency at which m_1 and m_2 vibrate out of phase, and ω_2 the natural frequency at which m_1 and m_2 vibrate in phase. Another possibility could have been to define ξ_i as $\xi_i = c_i \Big/ \left(2\sqrt{k_{i,sb} m_{i,sb}}\right)$. For a system with more than one DOF, this not always renders similar results to the previous formula.

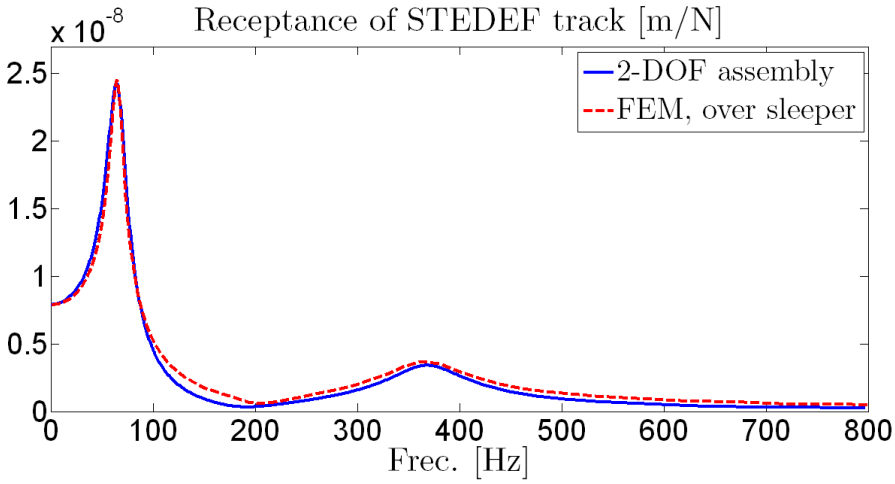
With the described adjustment, a quite good correspondence between the receptance of the track calculated with a FE model and the receptance of the 2-DOF assemblies beneath each wheelset can be achieved at low and mid frequencies up to a few hundred Hertz. Figure 1.2 shows the adjustment of receptances achieved for each track. The receptance of the track calculated with FEM shown in each graph corresponds to the position above a sleeper, not being large differences with the receptance calculated at mid-span in the frequency range shown.



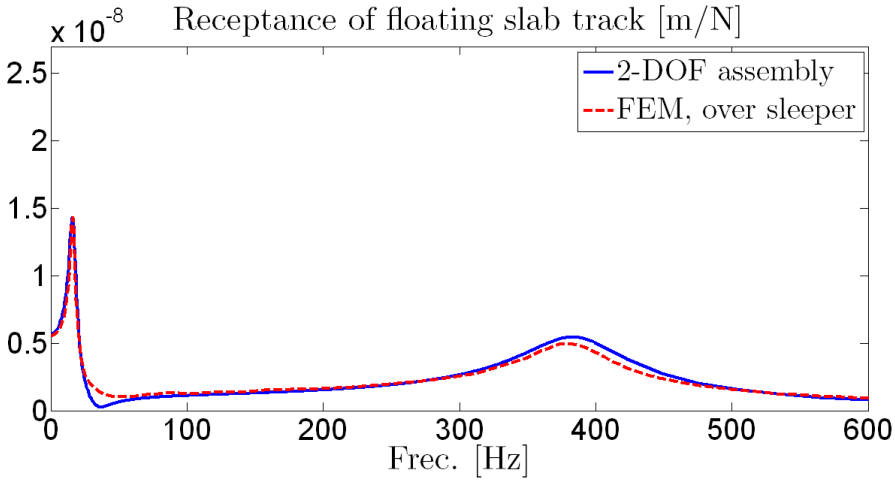
(a) Ballasted track.



(b) RHEDA 2000 track.



(c) STEDEF track.



(d) Floating slab track.

Figure 3.3. Comparison of the modulus of the receptance of each track calculated with a FE model and with an adjusted 2-DOF mass-spring-damper assembly.

The parameters corresponding to the shown receptances are listed in Table 3.2. n_1 and n_2 are around 2.7 in the ballasted, RHEDA 2000 and STEDEF tracks. In the floating slab track, $n_1 = 2$, and $n_2 = 9.92$. This result is explained as follows. The reason for the higher n_2 in this track is the high bending stiffness of the slab. The reason for the lower n_1 is the low degree of rail bending allowed, as a result of both the high bending stiffness of the slab and high railpad stiffness (see also the pad forces in Figure 3.17 and related discussion).

Table 3.2. Parameters of the RMB track models.

Parameter	Ballasted	RHEDA 2000	STEDEF	Floating slab
$m_{1,sb}$ [kg]	72.4	78.4	72.4	72.4
$k_{1,sb}$ [kN/mm]	200	900	300	426.4
ξ_1 (original | adjusted) [-]	0.139 | 0.139	0.10 | 0.10	0.10 | 0.10	0.086 | 0.108
$c_{1,sb}$ (adjusted) [kN s/m]	30.0	16.5	25.5	37.3
$m_{2,sb}$ [kg]	320	8.8	221.7	2062.5
$k_{2,sb}$ [kN/mm]	100	52.0	50.0	22.5
ξ_2 (original | adjusted) [-]	0.20 | 0.25	0.10 | 0.156	0.10 | 0.15	0.102 | 0.160
$c_{2,sb}$ (adjusted) [kN s/m]	100	21.4	36.6	70.1
ω_1 [rad/s]	1857.3	10913	2356.7	2469.1
ω_2 [rad/s]	500.2	754.7	410.2	102.7
n_1 [-]	2.50	2.69	2.70	2.00
n_2 [-]	2.70	2.69	3.00	9.92
m_1 [kg]	181	211	195	145
k_1 [kN/mm]	500	2420	810	853
c_1 [kN s/m]	75	44.4	68.7	74.6
m_2 [kg]	864	23.7	665	20500
k_2 [kN/mm]	270	140	150	223
c_2 [kN s/m]	270	57.8	110	695

The real behaviour of the longitudinally continuous elements of the track, namely the rails, and if it is the case, also the slab, cannot be properly represented with this type of model. The possible effect in the vehicle–track interaction of the rail bending between two consecutive sleepers in the case of discretely supported rails is also neglected. Local characteristics/imperfections in the track, such as voided sleepers, cannot be properly dealt with either.

The contact condition between wheel and rail is achieved with the standard wheel–rail pair available in the SIMPACK rail environment [SIMPACK Wheel Rail doc 2007]. A Hertzian contact model is used with the possibility of multiple contact patches in the same wheel–rail pair. In this study, the tangential forces and torque in each contact are calculated with the FASTSIM algorithm with a discretization of the contact patch of 11×11 elements.

A full three-dimensional (3D) model of the vehicle with 50 DOFs and a simplified model consisting of a single wheelset with a static force applied into it have been used with this type of train–track interaction model to obtain the results shown in §3.5and 3.6.

3.2.2 Finite element models

In this type of model, both train and track are completely modelled in a FE environment. The elements used for modelling the different components of the track are similar to those used in the elastic multibody (EMB) vehicle–track models that are described in §3.2.3: BEAM elements are used for the rails, lumped linear viscoelastic elements for each railpad, sleeper support, and flexible material beneath slab, lumped masses for each sleeper, and either 1D BEAM or solid HEXA elements for the slab.

The vehicle is modelled by means of lumped mass, spring and damper elements. A model of a quarter of the vehicle has been considered, that consists of three elastic levels (wheelset, bogie, and car body). The vertical displacement is the only DOF in each elastic level. As a variation to this single wheelset model, a model of two wheelsets has also been used, which consists simply of two of the assemblies described here for the single wheelset model, separated by the bogie wheelbase distance. With this second model, the possible effect of wave reflections between different wheelsets of the vehicle can be studied. These wave reflections can have a considerable effect on the wheel–rail contact forces, particularly at frequencies corresponding to vibration modes of the track with the rail vibrating with two, three, and upper multiple half-wavelengths between the two bogie's wheelsets [Nielsen 2003]. The leading axle of the vehicle is placed at a rail location either directly above a sleeper, or at mid-sleeper bay, in the different analyses that have been conducted.

The contact between vehicle and track is accomplished by means of a linear BEAM element, that provides stiffness only in the vertical direction. The normal contact law is linearized according to the Eq. (3.5). In this equation, k_H is the linearized contact stiffness, N and d the normal wheel–rail force and approach, respectively, N_0 the reference normal wheel–rail force, taken as the wheel–rail force in static conditions, and C_H the Hertzian contact constant, with which the normal force (N)–approach (d) relationship is expressed as $N = C_H d^{3/2}$. Taking into account the Hertzian relationship between N and d (cf. Eq. (2.17)), C_H is given by Eq. (3.6), with nomenclature as indicated in §2.1.1.1. Here C_H is taken as equal to 1.04E+11 N/m$^{3/2}$. As this study is focused on vertical vehicle–track interaction, and given the small usual conicity values of

wheel–rail contact with the wheelset centred in the track, it is assumed that the normal wheel–rail approach takes place in the vertical direction.

$$k_H = \left.\frac{\mathrm{d}N}{\mathrm{d}d}\right|_{N=N_0} = \frac{3}{2} C_H^{2/3} N_0^{1/3} \tag{3.5}$$

$$C_H = \frac{4E^*}{3\sqrt{r^3(A+B)}} \tag{3.6}$$

The FE models of vehicle and track are fully linear. Therefore, possible non-linear effects in the vehicle–track system, like wheel or sleeper lift-off, and non-linear characteristics of vehicle or track components, are not taken into account. However, it has been checked that with the level of excitation used in the simulations, wheel or sleeper lift-off does not occur. It would be possible to include non-linear characteristics in these FE vehicle–track models. This however would entail a considerable increase in computation costs, in these models that have a number of DOFs at least one order of magnitude higher than the other two types of multibody vehicle–track models that have been used.

This type of vehicle–track interaction model is a moving irregularity model, since the vehicle remains static with respect to the track, while a strip of irregularity passes between wheel and rail. The irregularity is entered by means of a temperature excitation in the BEAM element that represents the vertical contact stiffness between wheel and rail: the rail irregularity at each point in time is converted to a temperature value, so that the thermal expansion of the BEAM element matches the desired value of rail irregularity.

This type of model is simpler to build than the EMB vehicle–track models that are described in §3.2.3, and, since it is a linear model, its computational costs are considerably lower than in the EMB vehicle–track models, despite the higher number of DOF. Particularly, considerable savings in computation times are achieved, thanks to the substitution of the sophisticated contact algorithm inside the MBS models, that calculates both the contact position and the contact forces in a non-linear way, by a simple linear spring between two fixed nodes. In addition, the track can also be represented with a high level of detail.

However, with these FE models some capabilities are lost with respect to the EMB vehicle–track models described in §3.2.3. First, this model can only be used for the study of vertical dynamics, because wheel–rail contact pairs, necessary to properly account for the tangential forces at the wheel–rail interface, cannot be introduced (the contact condition is simplified by a linear spring, as previously indicated). Additionally, the possible effect of the parametric excitation resulting from the varying dynamic stiffness of the track at different points along each sleeper bay is dropped out of the analysis, since the vehicle remains stationary at one given position on the track. Finally, the same fact of the vehicle being stationary instead of moving with respect to the track, gives rise to another source of error; the rail vibration that reaches a moving wheel is not the same as the one that reaches a stationary wheel. This error is negligible if the wave propagation speed on the track is much higher than the train speed [Wu 2004], but it can be important if the two speeds are of the same order.

3.2.3 Elastic multibody models

These are the most comprehensive vehicle–track interaction models that have been built in this study. Like the RMB train–track models (§3.2.1), SIMPACK is the main modelling environment. The models of the vehicle are also the same. The track is modelled in a more realistic way, using continuous elastic bodies. In addition, the relative longitudinal motion of the vehicle with respect to the track is properly taken into account, because the track does not travel together with the vehicle as in the RMB models, and the vehicle is not stationary at one point on the track, as in the FEM models.

In its 8.9 version, SIMPACK provides two main modules to integrate elastic bodies into MBS models, which are named FLEXBODY and FLEXTRACK. FLEXTRACK is oriented to the study of vehicle–track interaction at very low frequencies, e.g. the dynamic interaction of a vehicle with a flexible bridge structure. It has been checked in this study that it is not adequate to reproduce dynamic phenomena at higher frequencies. Therefore, in this study, FLEXBODY has been used to construct the EMB vehicle–track models, and the procedure followed to build the models with this module is explained.

By means of FLEXBODY, the rails and slab of the different tracks are represented as elastic bodies inside the EMB vehicle–track models. Each elastic body is represented within the MBS model by means of a number of its natural modes, which are obtained in a previous real eigenvalue modal analysis of a FE model of the elastic component. In this case, NASTRAN has been used as the FE processor to perform the required modal analyses.

The output results of the modal analysis imported into the MBS model are the stiffness and mass matrices of the elastic track component, as well as its mode shapes. The dynamic behaviour of the elastic component is represented by means of modal superposition; i.e., each of the considered mode shapes of the elastic component is an additional DOF in the MBS model, and the equation of motion governing each of these DOFs is the uncoupled dynamic equation corresponding to that natural mode. For an elastic rail, the equation of motion for the *i*th mode can be written according to Eq. (3.7). In this equation, $z_i(t)$ is the modal displacement of the *i*th mode, ω_i its natural frequency, m_i its generalized mass, ξ_i its damping ratio, and $\Phi_i(x)$ its shape function; $f_{wr,j}(t)$ is the wheel–rail force at the *j*th wheel, $x_{w,j}$ the longitudinal location along the rail of the *j*th wheel, and *m* the number of wheelsets; $f_{p,k}$ is the force of the *k*th railpad, $x_{p,k}$ the location of that railpad along the rail, and *n* the total number of railpads in the modelled rail section.

$$\ddot{z}_i(t) + 2\xi_i\omega_i\dot{z}_i(t) + \omega_i^{\,2} z_i(t) = \frac{1}{m_i}\left[\sum_{j=1}^{m} f_{wr,j}(t)\,\Phi_i\left(x_{w,j}\right) - \sum_{k=1}^{n} f_{p,k}(t)\,\Phi_i\left(x_{p,k}\right)\right] \tag{3.7}$$

The movement of the elastic rail is described by superposition of the movements of the *q* modes describing its elasticity, as expressed in Eq. (3.8). In this equation, u_R describes the movement of one of the DOFs of the elastic rail as a function of the longitudinal coordinate *x* and time *t*, and Φ_i contains the displacements of mode shape *i* in that DOF. A similar equation can be written for each of the DOFs of the rail. In this study in which only vertical vibration is considered, and with 1D models of the rail, only vertical displacement and pitch rotation are taken as relevant DOFs of the rail.

$$u_R(x,t) = \sum_{i=1}^{q} z_i(t)\ \Phi_i(x) \tag{3.8}$$

In this way, a compact and efficient representation of the elastic component within the MBS model is achieved with relatively few DOFs —the considered normal modes of the elastic component. To increase the efficiency of the elastic body representation in the MBS model, it is convenient and often necessary to use some condensation techniques in the FE model. In this way, the size of the mass and stiffness matrices and the eigenvectors that will be imported into the MBS model can be reduced to a reasonable extent for efficient handling by the MBS integration algorithm. It has to be taken into account that the size of MBS models is much more limited than the size of FE models: while FE models usually reach sizes of hundreds of thousands of DOFs, the size of MBS models is typically limited to some few hundred DOFs.

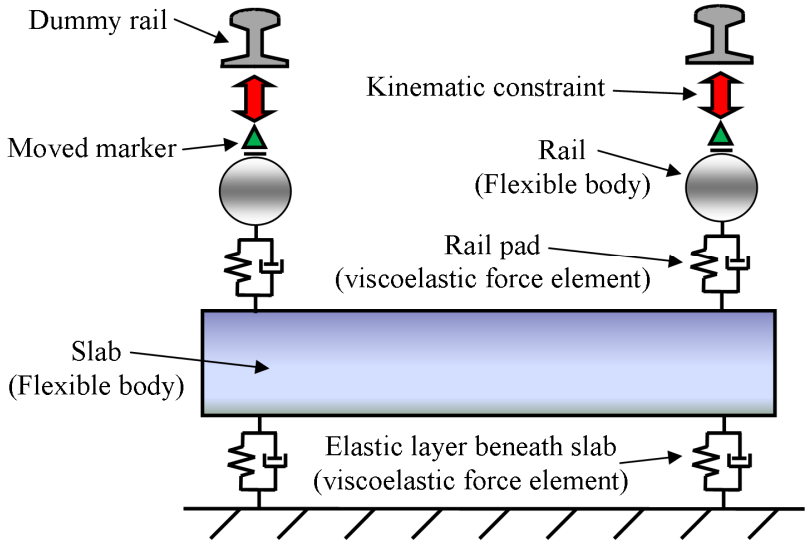
For the purposes of condensation, the generalized dynamic reduction available in NASTRAN [Herting 2004] has been used. It has been checked that this condensation method improves the results obtained with the static Guyan condensation [Guyan 1965] and that the error in the calculation of normal modes derived from the condensation step is small.

As will be discussed in §3.2.3.1, both the necessary number of modes and number of nodes of the condensed model of each elastic component have been studied to obtain reasonable accuracy in its representation inside the EMB train–track models.

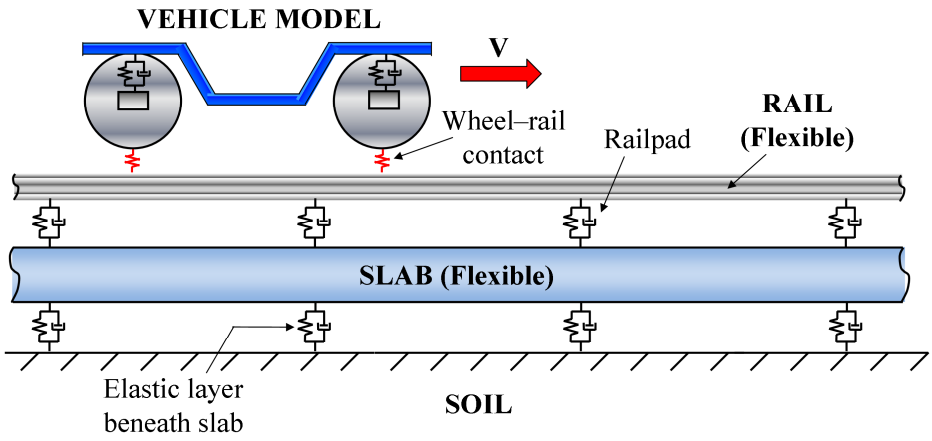
Concerning the coupling between vehicle and track, the standard wheel–rail pair element available in the SIMPACK Rail module is used here again, but this time, some additional auxiliary elements need to be defined. The reason is that the wheel–rail pair element can only be defined between rigid bodies with no relative longitudinal motion, and the rails are modelled in this case as elastic bodies that do not travel along the track with the wheels.

The additional elements defined are: one moved marker in the elastic rail bodies for each of the vehicle's wheelsets, one dummy rigid rail body for each wheel–rail pair, and one kinematic constraint for each dummy rigid rail body. Figure 3.4a shows a schematic cross-section of this type of track model, that describes

the different elements used and their arrangement. A side view of the train–track interaction model is depicted in Figure 3.4b, in which the described auxiliary elements are omitted and one bogie of the vehicle and the wheel–rail contact elements are added.



(a) Track cross-section.



(b) Side view.

Figure 3.4. Sketch of the EMB vehicle–track interaction model, containing both elastic bodies and lumped elements (one bogie represented only in the side view).

The moved markers move along each elastic rail, their longitudinal motion being defined with the longitudinal motion of the wheelset to which they are associated, and their vertical motion by linear interpolation of the vertical displacements and pitch rotations of the two nearest interface nodes defined in the elastic rail.

The dummy rigid rail bodies also move along the track together with each of the vehicle's wheels. Their only purpose is to contain the rail profile geometry, and to transfer the load from each wheel to the rails, serving as the rail body in each wheel–rail pair. They are given a negligible mass, to not alter the dynamic properties of the track.

The kinematic constraints are used to link the vertical displacement of each dummy rigid rail body to that of its corresponding moved marker. There is another option to couple the dummy rail bodies with the moved markers in the flexible rail, by means of stiff springs [Gonzalez 2008]. To not alter the dynamic properties of the model, a high value for the stiffness of this spring should be used. However, the high stiffness value of these springs, together with the low mass of the dummy rail bodies, introduces high-frequency oscillations in the simulation and compromises the stability of the numerical time-step integration process. These problems are avoided using the mentioned kinematic constraint, as is done in this study.

The damping for each elastic body of the track is introduced by means of modal damping for each of the considered natural modes of the elastic component. As each elastic body is modelled separately, lumped force elements representing the concentrated stiffness and damping provided in each of the track's discrete supports (i.e. railpads, sleeper supports, and ballast) can be defined between the different track bodies. Thus, the non-uniformly distributed and hence non-proportional stiffness and damping properties of a discretely supported track can be adequately modelled without the need for a complex eigenvalue modal analysis.

One limitation present in the EMB vehicle–track interaction models pertains to the length of flexible track simulated. While in the case of the previously described RMB vehicle–track models, the length of simulated elastic track is unlimited, here a finite length of elastic track has to be considered to limit excessive computational costs. The spurious effects from the boundaries can

affect the quality of the results in the middle of the track [Baeza 2006]. Those spurious effects can be eliminated by simulating a sufficient length of track. In the simulation results presented in this chapter, track lengths up to 200 sleeper bays have been considered.

In the following subsections, the modelling of each individual component of the track is explained in more detail. Since computational costs can become very high in this kind of model [Bezin 2009], it is necessary to exercise proper judgement in defining the degree of detail that is going to be considered in each component of the model to achieve a reasonable precision–cost relationship. For this purpose, different options, with varying degree of complexity, have been considered for representing the elastic bodies inside the EMB vehicle–track models.

3.2.3.1 Rails

The rails are modelled as continuous elastic bodies by means of 1D FE models, based on Rayleigh–Timoshenko beam elements. This model allows for a reasonably good representation of the vertical dynamic rail properties up to 2.5 kHz [Knothe 1993]. If lateral dynamics were to be considered, a more detailed representation of the cross-section of the rail could be necessary, since the level of the rail cross-section distortion in the lateral vibration modes is much higher than in the vertical vibration modes [Knothe 1993], [Nielsen 2003].

The FE models of the rails contain 12–24 beam elements per sleeper bay, but the model is reduced by condensation. The number of chosen nodes for the condensed model depends on the frequency of the upper natural vibration mode of the rail that needs to be considered; with increasing natural frequencies, more nodes per sleeper bay need to be taken in the condensed model to be able to calculate the higher modes with little error. In the condensed models, at least the nodes above each sleeper position need to be included (i.e. one node per sleeper bay), to serve as connection points for the railpads. If more nodes need to be incorporated in the condensed model, these nodes should be taken equidistant, because in this way, the error in the calculation of eigenmodes with the condensed model is minimized.

When calculating the rail's eigenmodes, no intermediate support is considered in the finite length of the rail modelled; only simple supports are defined as

boundary conditions in each of its two ends. Additionally, the motion out of the vertical plane and twist of the rails is constrained. To represent adequately in the EMB vehicle–track models the dynamic behaviour of the rail under the influence of the multiple supports existing along its length, on the basis of the eigenmodes calculated for the rail without intermediate supports, the wavelength of the upper eigenmode considered has to be sufficiently small. As a consequence, a large number of eigenmodes need normally to be imported into the MBS models; the more eigenmodes the longer the distance of the flexible track modelled.

The parametric excitation in tracks with discretely supported rails, resulting from the difference in the dynamic track stiffness between different positions along each sleeper bay, may have a significant effect in the wheel–rail contact forces [Wu 2004]. The difference in the dynamic track stiffness arises from the rail bending between supports. The wavelength of the upper eigenmode considered should be in the order of twice the sleeper distance at most, to be able to properly account for this rail bending. The resulting eigenfrequency of the upper vibration mode considered may be much higher than the upper frequency of interest in the analysis. As importing a large number of eigenmodes in the MBS models can entail a considerable rise in computational costs (it has to be remembered that each of the imported eigenmodes for the elastic bodies represents one more DOF in the MBS models), the upper frequency of the necessary eigenmodes for the rail has to be properly chosen to obtain reasonable precision in the frequency range of interest.

As an example of this, Figure 3.5 compares the receptances of the STEDEF track above sleeper calculated with a FE model of the track (the reference result in this case), and with EMB models, the latter ones including the rail as a flexible body with natural modes up to 500 Hz in one case, and up to 2500 Hz in the other. The receptances are compared up to a frequency f of 2500 Hz. As could be expected, the EMB model in which more rail modes are included approaches the FEM result better. The EMB model including rail modes only up to 500 Hz does not represent the pinned-pinned anti-resonance at all, which takes place at a frequency slightly above 1000 Hz.

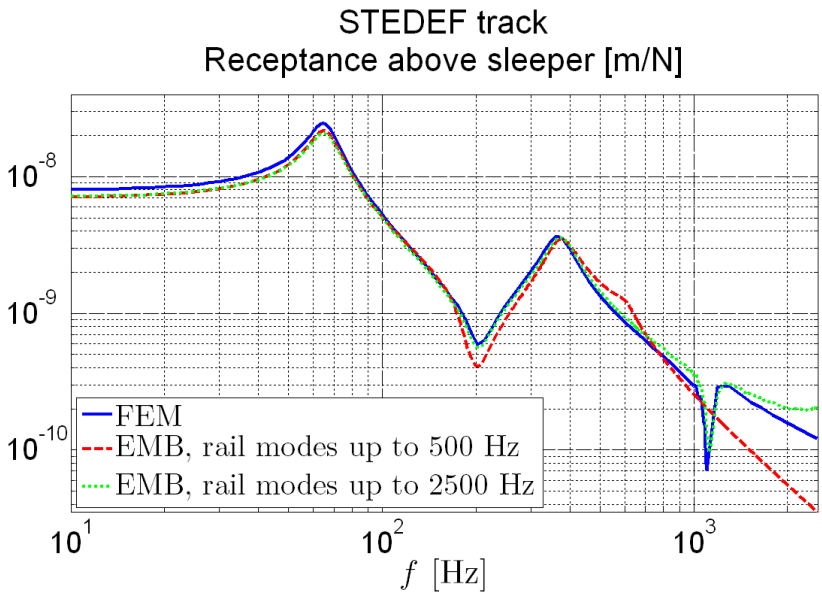


Figure 3.5. Comparison of the modulus of the receptance of the STEDEF track calculated with a FE model and with EMB models of the track with different number of rail modes.

3.2.3.2 Discrete supports and sleepers

The discrete supports and sleepers of the track are modelled as lumped elements. Their modelling is similar in the RMB, FE and EMB models, except that in the first case only equivalent mass-spring-damper assemblies are modelled beneath each wheelset, while in the other cases all the sleepers and discrete supports of the track are explicitly modelled (in the finite section of track modelled).

Each discrete railpad and sleeper support in the track models is modelled as a lumped viscoelastic element. The sleeper supports represent the stiffness and damping of the ballast beneath each sleeper in the case of the ballasted track, and the stiffness and damping of the elastomeric material between the slab and each sleeper in the case of the STEDEF track. The elastomeric mat beneath the slab in the floating slab track is also represented by means of discrete independent viscoelastic elements.

Even though a considerable degree of non-linear behaviour is typically observed in these elastomeric elements, with stiffness values changing with

preload level and frequency of excitation, the assumption of linear stiffness and damping properties is considered sufficient for the purposes of this study. Another possible source of non-linearity within the track may come from the possibility of sleeper lift-off. However, it has been checked that with the level of vertical excitation used in the simulations, this does not occur. Nevertheless, it has be noted that in this type of model, non-linear stiffness and damping properties depending on the load level can be assigned to each support, if desired.

Each sleeper is modelled as a lumped rigid mass, with one DOF (vertical displacement). In the case of the RHEDA 2000 track, the base plates in each of the rail supports are also modelled as lumped rigid masses.

3.2.3.3 Slab

The slab is modelled as a continuous elastic body, like the rails. Two types of meshing have been considered for this element: (a) a 3D meshing, using HEXA8 solid elements, and (b) a 1D meshing, similar to that used for the rails, using BEAM2 elements. With a 3D mesh, the possible 2D vibration of the slab can be accounted for, as it is done in [Zhai 2009]. In many other references as [Lombaert 2006], [Hussein 2006], and [Kuo 2008], a 1D representation of the slab similar to the one proposed here is used.

As in the case of the rails, no support condition is considered beneath or above the slab when calculating its eigenmodes. In the case of the 3D meshed slab models, a large number of nodes need to be considered to adequately calculate and represent the 3D vibration modes of the slab. On the contrary, in the case of the 1D meshed slab models, the number of needed nodes is much lower, as only 1D vibration modes are calculated. Additionally, the number of calculated eigenmodes for a given frequency range is lower than in the 3D meshed slab models, as only the vibration modes contained in the longitudinal vertical plane are considered in the analysis.

3.3 Description of the vehicle model

The railway vehicle considered in this chapter is a high-speed passenger car. The physical parameters for the vehicle used in the simulations are listed in Table 2.1. The suspension parameters have been taken from [Melis 2008].

Table 3.3. Parameters of the considered high-speed vehicle.

Parameter	Value	Units
Car body mass (unloaded condition)	53500	kg
Bogie mass	3500	kg
Wheelset mass	1800	kg
Wheel nominal rolling radius	0.46	m
Bogie spacing	17.325	m
Bogie wheelbase	2.5	m
Primary vertical stiffness	1.83E+06[a]	N/m
Primary vertical damping	16000[a]	N s/m
Secondary vertical stiffness	0.14E+06[a]	N/m
Secondary vertical damping	12000[a]	N s/m

[a] Value per wheelset.

The vehicle is modelled as a MBS with discrete elements of mass, stiffness, and damping. The full 3D model of the vehicle that has been used in most of the simulations consists basically of car body, and two identical bogies, each of them having two wheelsets. The wheels have a S1002 profile. All these bodies are represented with rigid masses and are interconnected with different spring and damper elements, representing the primary and secondary suspensions. The complete vehicle model has 50 DOFs.

Simpler models of the vehicle have also been used in some of the simulations, and the validity of the simplified models has been analysed in the different simulations. In some of the simulations, the vehicle model is reduced either to a single wheelset with a static force equivalent to the nominal axle load of the vehicle acting on it, or to three equivalent masses (one for the wheelset, another for the bogie, and another for the car body) with equivalent vertical stiffness and damping elements between them representing the primary and secondary suspensions.

3.4 Definition of track irregularities

Numerical simulations of the vehicle running on straight track with vertical rail irregularities have been carried out with the track and vehicle models described in previous sections. The vehicle speed is fixed at 250 km/h in all the simulations presented in this chapter.

Two types of rail irregularities have been considered as the vertical excitation input: a randomly distributed low frequency broadband rail irregularity on one hand, and a high-frequency irregularity on the other hand. The low frequency broadband irregularity used corresponds to low intensity vertical rail irregularities, according to [ERRI B 176/3 1993]. The high-frequency excitation is defined according to the power spectral density function given in Eq. (3.9), taken from [Zhai 2009]. The units of $G_V(f)$ in this equation are $mm^2/(1/m)$, and f is the spatial frequency in cycle/m. This type of high-frequency irregularity could represent a corrugated rail, which often provokes significant track vibrations and wheel–rail force fluctuations [Oyarzabal 2009]. Wavelengths between 40 and 200 mm are considered for this second type of excitation.

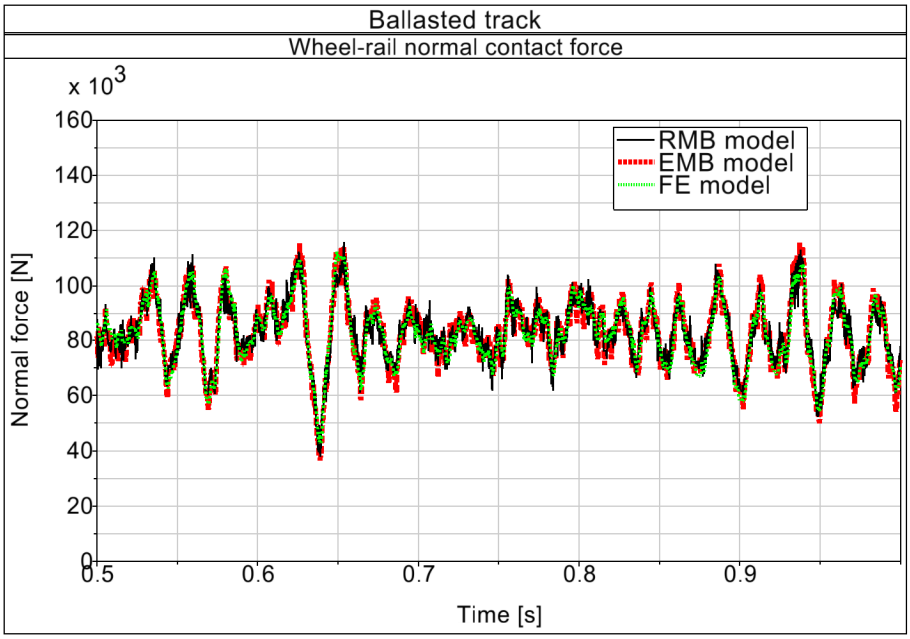
$$G_V(f) = 0.036 f^{-3.15} \tag{3.9}$$

3.5 Comparison of results with different vehicle–track models

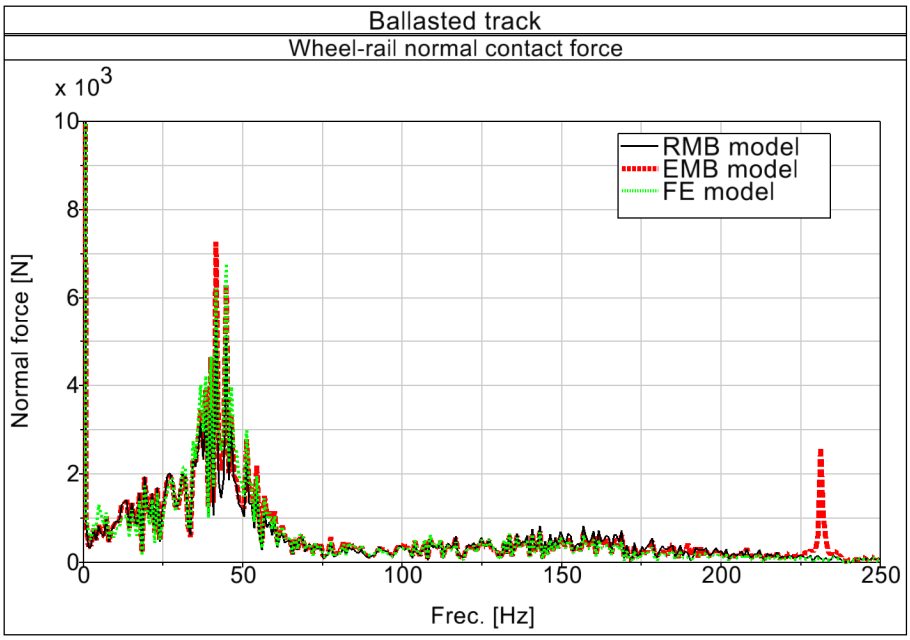
In this section, the results obtained with the different vehicle–track models are compared, and some conclusions are obtained in relation to the validity of the simpler vehicle–track models in some situations.

3.5.1 Comparison between the RMB, FE, and EMB models

Concerning the simulations with the low frequency broadband rail irregularity, comparable results are obtained with the simpler RMB and FEM vehicle–track models, and with the more complete EMB models, in the case of the ballasted, RHEDA 2000, and STEDEF tracks. Results obtained with the different vehicle–track models for the case of the ballasted track are shown in Figure 3.6 as an example.



(a) Time histories.



(b) Frequency spectra.

Figure 3.6. Wheel–rail contact forces obtained with different vehicle–track models in the ballasted track with the low frequency broadband track irregularity.

These results are justified because the necessary conditions for these simpler models to be valid are fulfilled. On one hand, the influence of the parametric excitation in these tracks is small compared to the influence of the rail roughness. On the other hand, the interaction between different wheelsets of the vehicle is small in this frequency range, because the only longitudinal coupling between them is provided by the rails, with a relatively low bending stiffness. Finally, the influence of the relative longitudinal motion of the vehicle with respect to the track is also negligible, because the wave propagating speeds at 25 Hz and above are much higher than the forward speed of the vehicle.

It is particularly remarkable to see the high similarity between the results obtained with the RMB models, and those obtained with the much more costly EMB models, despite the great difference of the track modelling in both cases. It has to be taken into account, however, that a parameter adjustment for the discrete track elements in the RMB models has been necessary. This is based on a previous frequency analysis of a FEM model of the track as indicated in §3.2.1, to approximate the results obtained with the EMB models. With a proper parameter adjustment, a very good correspondence of the receptance of the discrete assembly representing the track beneath each of the vehicle's wheelsets, to the receptance calculated with full FE models of the track, can be achieved. Consequently, the much simpler models of the track in the RMB models can well approximate the results obtained with the more detailed EMB models.

In the case of the floating slab track, however, different results are obtained with the different vehicle–track models. The reason is that in this case, the influence of the parametric excitation is much higher than in the other tracks, due to the fact that the sleeper passing frequency is close to the frequency of the second vibration mode of the wheelset–track system, which is easily excited at this frequency range. Also, in this case, the relative difference of the track receptance at different positions inside a sleeper bay is higher than in the other tracks, and this contributes to further increase the influence of the parametric excitation.

The wheel–rail contact forces calculated with the EMB vehicle–track models (including the full vehicle model) for the vehicle running on each of the tracks with no irregularities are plotted in Figure 3.7. Therefore, in this case, the only

dynamic effect comes from the parametric excitation. This effect is seen to be much higher in the case of the floating slab track than in the other tracks, as has been stated.

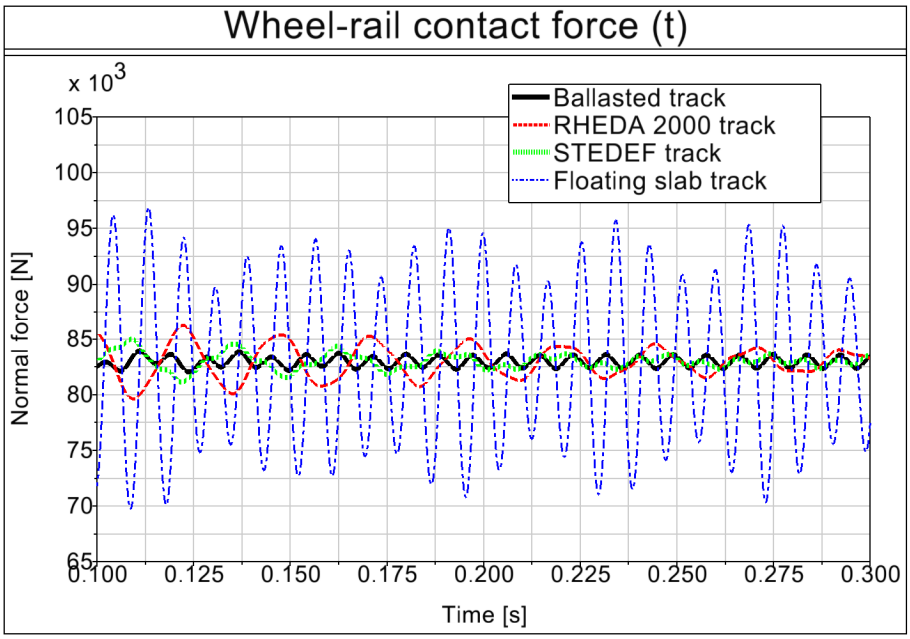
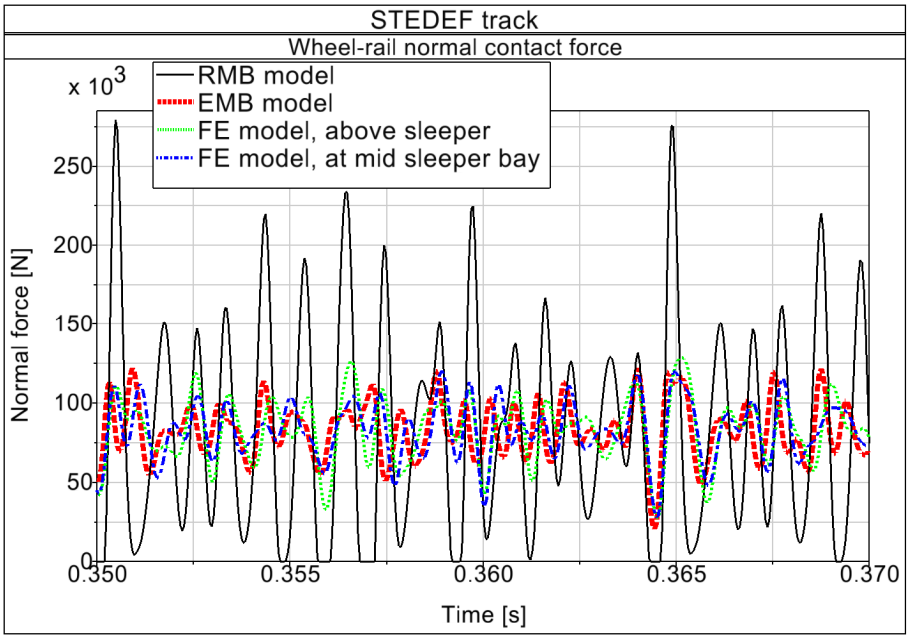
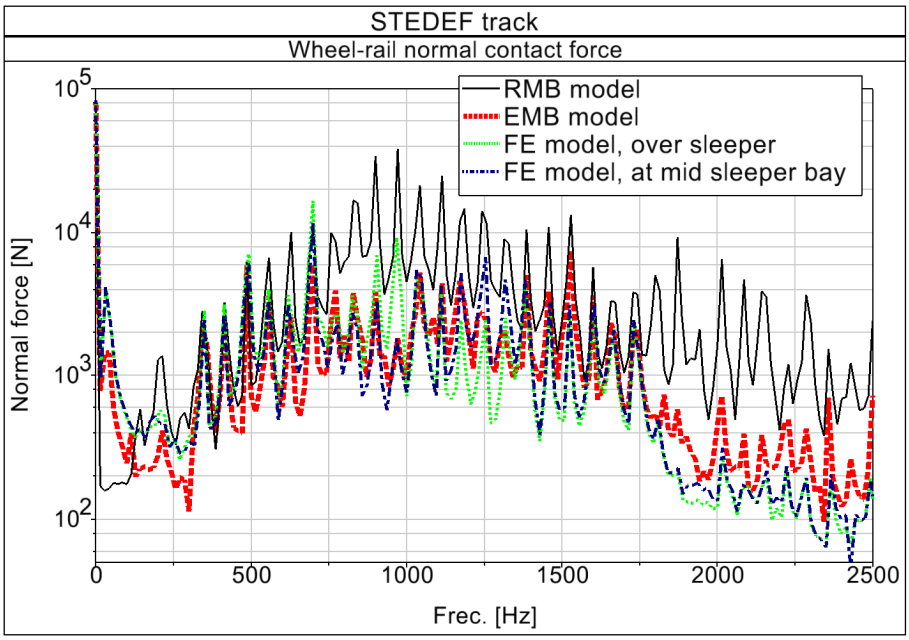


Figure 3.7. Time histories of the wheel–rail contact forces with no rail irregularities for the different tracks.

Concerning the simulations with the high frequency broadband rail irregularity, this time, different results are obtained with the different models, as can be seen in Figure 3.8, in which the wheel–rail forces obtained with the different vehicle–track models for the STEDEF track are depicted, both in time and frequency domain. The results for the other track types are qualitatively similar.



(a) Time histories.



(b) Frequency spectra.

Figure 3.8. Wheel–rail contact forces obtained with different vehicle–track models in the STEDEF track with the high frequency broadband track irregularity.

Referring to Figure 3.8, the most different results compared to those of the EMB models are obtained with the RMB models. In these models, as explained before, a reasonably good approximation of the track receptance can be obtained with the discrete assemblies representing the track beneath each wheelset, for frequencies up to a few hundred Hertz. At high frequencies, however, the receptances of these discrete assemblies do not approximate well the receptance calculated with the complete FE models of the tracks.

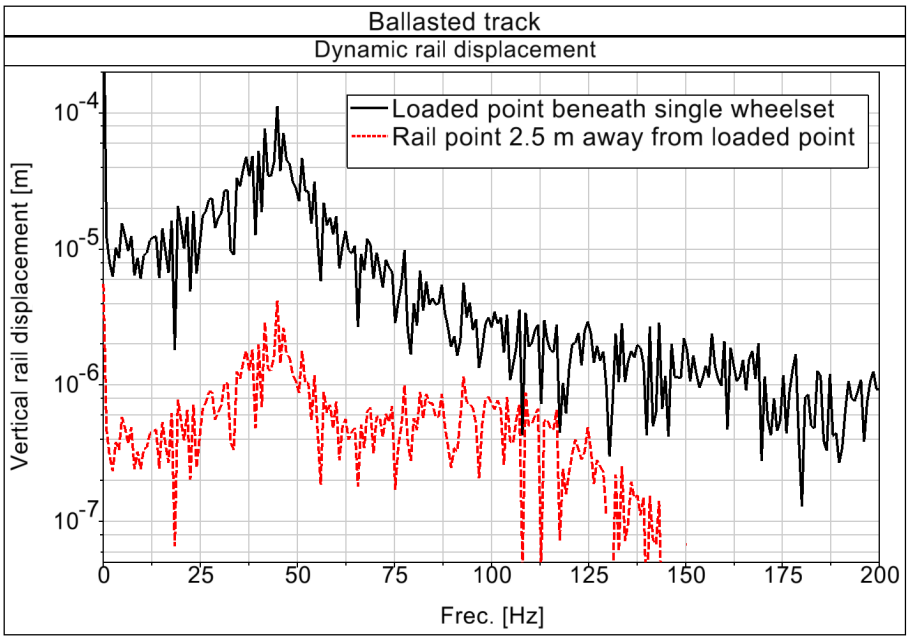
The results obtained with the FE vehicle–track models approximate better those obtained with the EMB models, but still some differences are noted between the results of both types of models. These differences are due to the differences in the track receptance at different points along each sleeper bay, which are greatest around the pinned–pinned resonance frequency. The FE models used here do not account for this variation in the track receptance, because the vehicle is static at one point in the track. For this same reason, in this case, different results are obtained also within the FE models, depending on the locations of the vehicle's wheelsets on the track; i.e., above sleeper or at mid-sleeper bay, as can be noted in Figure 3.8.

3.5.2 Comparison of models with single or multiple wheelsets

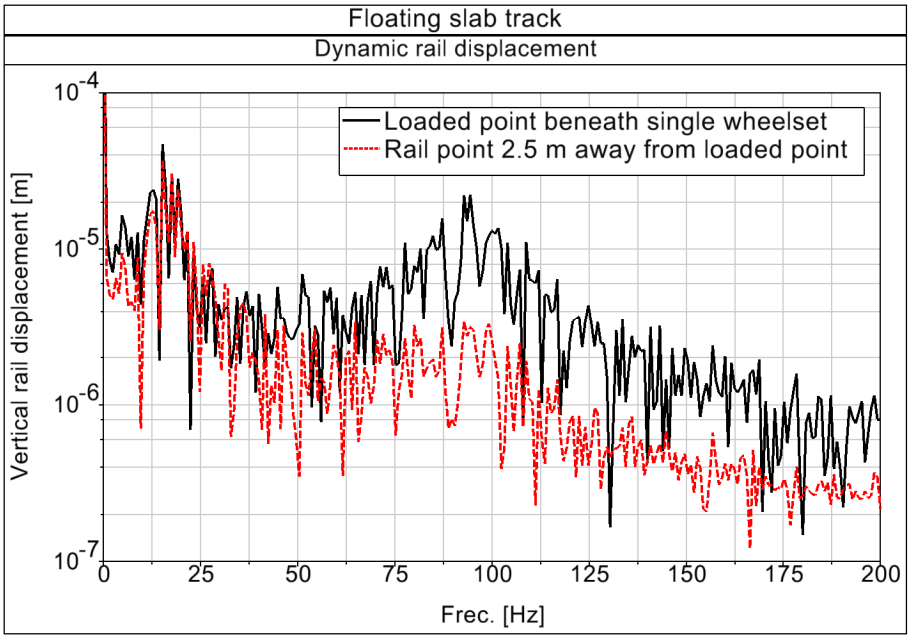
With the low frequency broadband rail irregularity, similar results are obtained in the simulations with a single or with multiple wheelsets for the ballasted, RHEDA 2000 and STEDEF tracks. This is because the interaction between different wheelsets is small at low frequencies.

On the other hand, in the case of the floating slab track, similar results are also obtained with a single wheelset or with multiple wheelsets. In this track, the floating slab couples the different wheelsets of the vehicle with much higher stiffness than the rails in the other tracks. However, this coupling only takes place at low frequencies. At higher frequencies, the vibration of the rail is decoupled from the vibration of the slab. Although the vibrations of the slab are transmitted along a great distance in the track due to its high stiffness, the vibration levels seen in the slab at frequencies above the natural frequency of the slab vibrating above the elastic mat beneath it are quite lower than those seen in the rails.

Figure 3.9 shows the frequency spectra of the dynamic displacement of the rails, obtained from simulations with a single vehicle wheelset. The displacements of the rail point beneath the vehicle's wheelset, and the displacements of the rail point where the other wheelset of the same bogie would have been, are represented for the case of the ballasted and floating slab tracks. The results for the other two types of track are qualitatively similar to those of the ballasted track. It can be seen that the displacements of the point where the other wheelset of the bogie would have been are much smaller than the displacements beneath the modelled wheelset in the case of the ballasted track at all frequencies. In the case of the floating slab track, this is also true at frequencies above about 75 Hz. Below this frequency, the displacements of both points are of similar magnitude. However, at these lower frequencies the wheel–rail contact forces are not being excited in this track, as will be seen in §3.6.1. From this, it is explained that the dynamic interaction between the different wheelsets of the vehicle is small for the studied track types with the low frequency broadband rail irregularity considered.



(a) Ballasted track.



(b) Floating slab track.

Figure 3.9. Frequency spectra of the dynamic rail displacements with the low frequency broadband track irregularity, obtained from simulations with the FE vehicle–track models and with a single vehicle wheelset.

In the case of the simulations with the high frequency broadband rail irregularity, some differences can be seen between the results obtained with one, or with multiple wheelsets, though small. As an example, Figure 3.10 compares the wheel–rail contact forces computed for the floating slab track with the EMB vehicle–track models, with a single and with multiple wheelsets, in the time domain. The differences seen in the frequency domain are much smaller. As seen in the figure, only moderate differences in the wheel–rail contact forces are noted in the time domain, but the amplitude and frequency content of the force oscillations is very similar with one or with multiple wheelsets. Therefore, in many applications regarding the study of tangent track with purely vertical dynamics, a single wheelset model can provide representative results of the vehicle–track interaction.

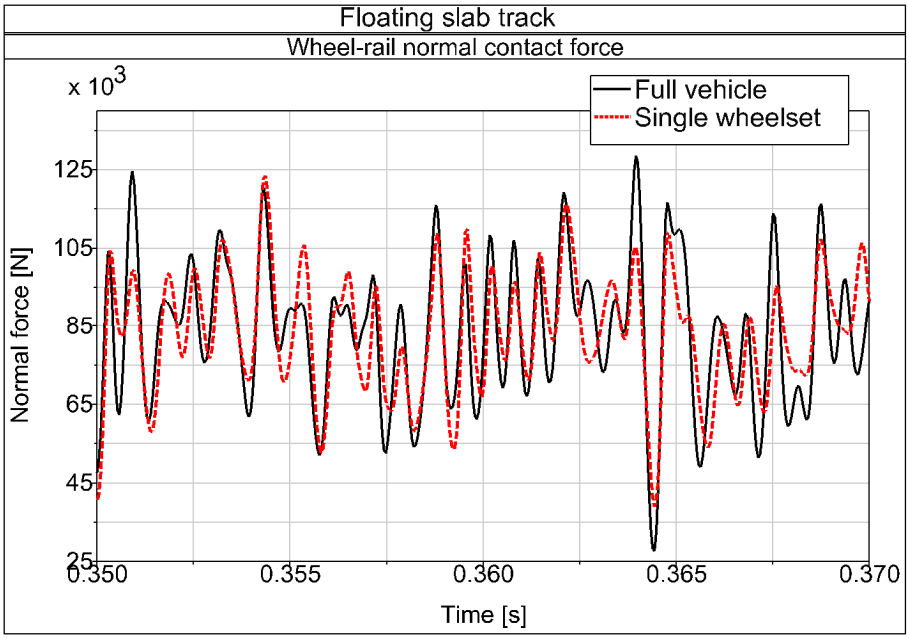


Figure 3.10. Time histories of the wheel–rail contact forces in the floating slab track with the high frequency broadband track irregularity, obtained from simulations with the EMB vehicle–track models with a single and with multiple wheelsets.

3.5.3 Comparison of 1D and 3D mesh for the slab in the floating slab track

Both 1D and 3D meshings for the slab have been tested in the FE and EMB vehicle–track models of the floating slab track.

Figure 3.11 shows the time histories of the vertical displacements of three points of the slab during a bogie passage in the case of the floating slab track with 3D slab meshing. The three points considered belong to the same vertical transversal plane of the slab, and are located at different lateral distances from the track centre. It can be seen that the vibration of the slab is fundamentally 1D in the vertical direction, since the displacements of different points across its width are very similar.

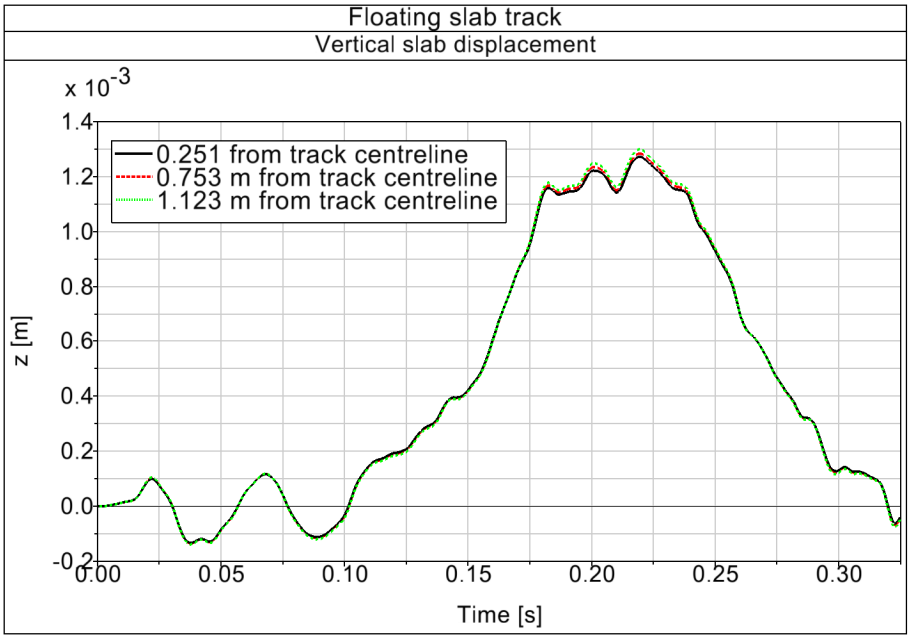


Figure 3.11. Vertical displacement of the slab in the floating slab track model with 3D meshing of the slab.

To further check the validity of the 1D meshing of the slab, FE models of the floating slab track have been constructed using both the 1D and the 3D meshings for the slab, and their receptances have been calculated in a frequency range from 0 to 2500 Hz. In these FE models, the errors associated to the mode truncation and condensation steps of the EMB models are avoided. The receptances calculated for this track with 1D and 3D meshings of the slab are compared in Figure 3.12, both above sleeper and at mid sleeper bay. As shown in the figure, the receptances calculated with the much simpler 1D meshing for the slab show very good agreement with those calculated with the 3D slab meshings in the whole frequency range. As an exception, in the 20–200 Hz frequency range an oscillation is noted in the receptances calculated with the model with solid slab meshing, which is not seen in the model with 1D slab meshing. This may be due to boundary effects of the former model, in which a shorter length of the track is modelled (71 sleeper bays are included in the model with 3D slab meshing, and 200 in the model with 1D slab meshing).

Therefore, it can be concluded that a 1D representation of the slab is suitable for purely vertical vehicle–track interaction studies in tangent tracks, and that the

much more costly 3D meshing does not bring any additional characteristic not already discerned by the 1D meshing. However, this should be assessed in different cases if the characteristics of the slab (e.g. its cross-sectional dimensions) change significantly from those considered here.

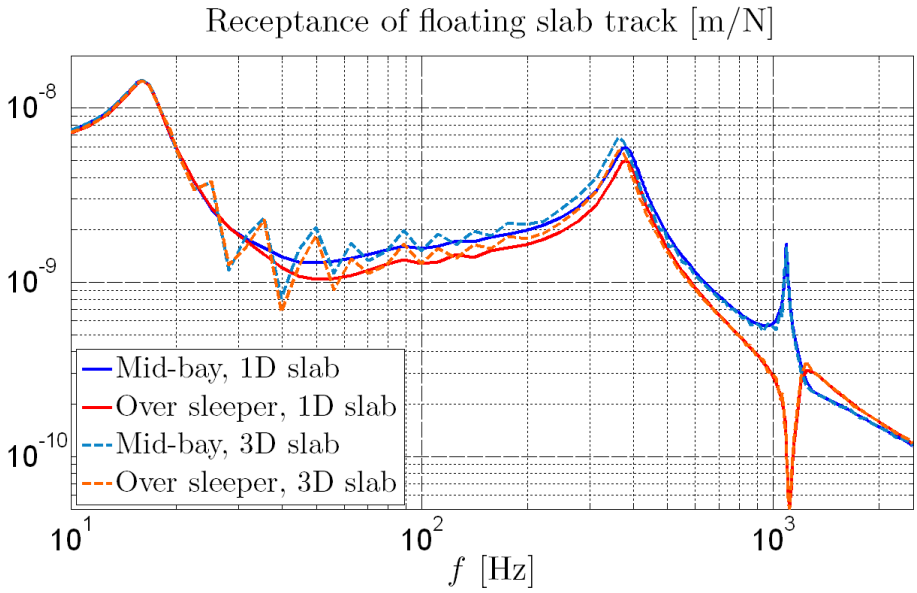


Figure 3.12. Receptances of the floating slab track calculated with FE models, with 1D and 3D meshings of the slab. The length of the models with 1D and 3D slab meshings are 200 and 71 sleeper bays, respectively.

3.5.4 Comparison of computational costs

In Figure 3.13, the computation times with different types of vehicle–track models are compared (note the logarithmic scale). The EMB models have the highest computational costs. The computational costs of the FE and RMB models are much lower, and of the same order, although it has to be said that their relative computational costs can vary substantially, depending, for example, on the number of DOFs of each model or on the type of wheel–rail contact model used. Note, for example, that the wheel–rail contact used in FE models is more simplistic than that used in the RMB models, and that including a complex wheel–rail contact model in the FE models could increase significantly the computation time. Within the EMB models, the computation

times can also vary substantially, depending on the degree of detail with which the elastic bodies are represented.

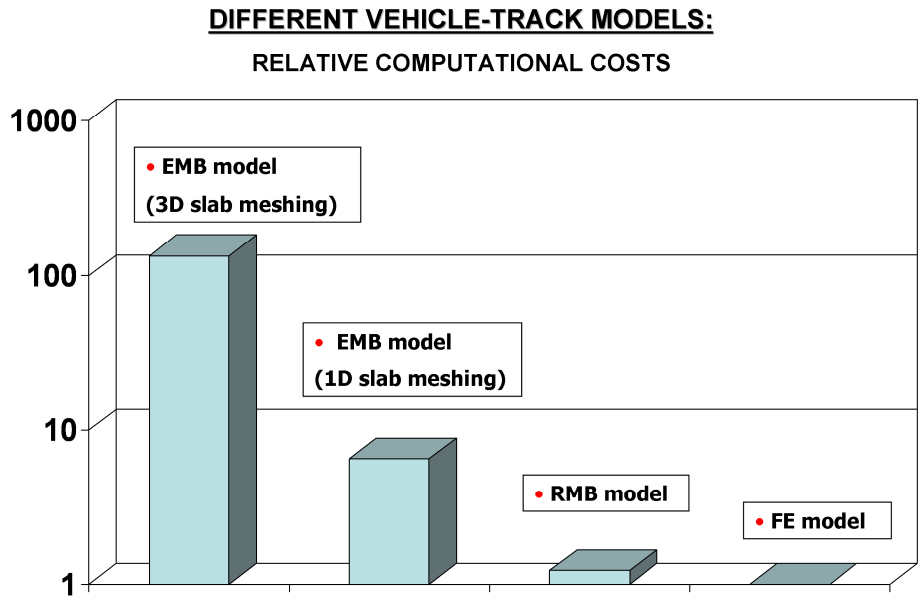


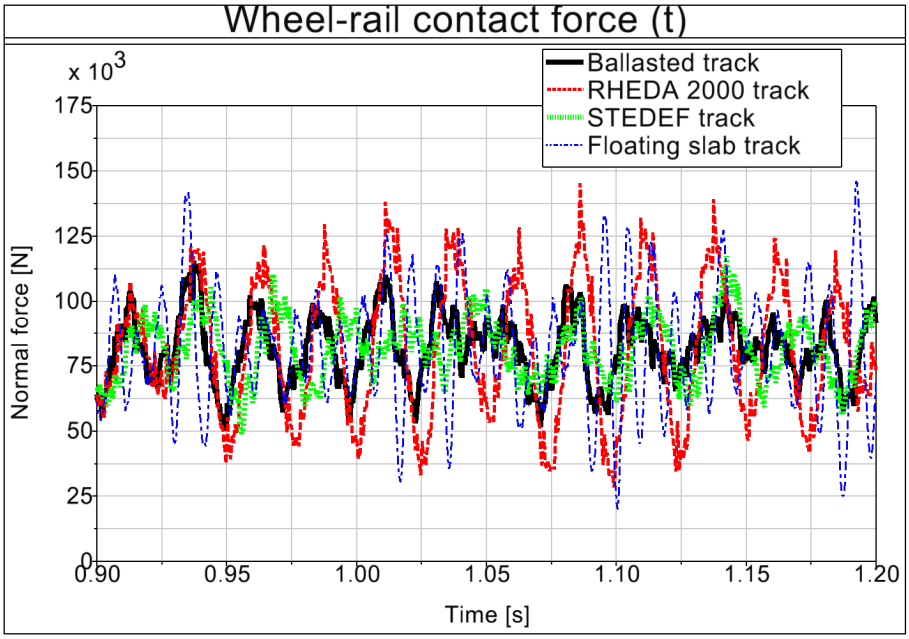
Figure 3.13. Comparison of computational costs with different vehicle–track models.

3.6 Comparison of vehicle–track interaction in different tracks

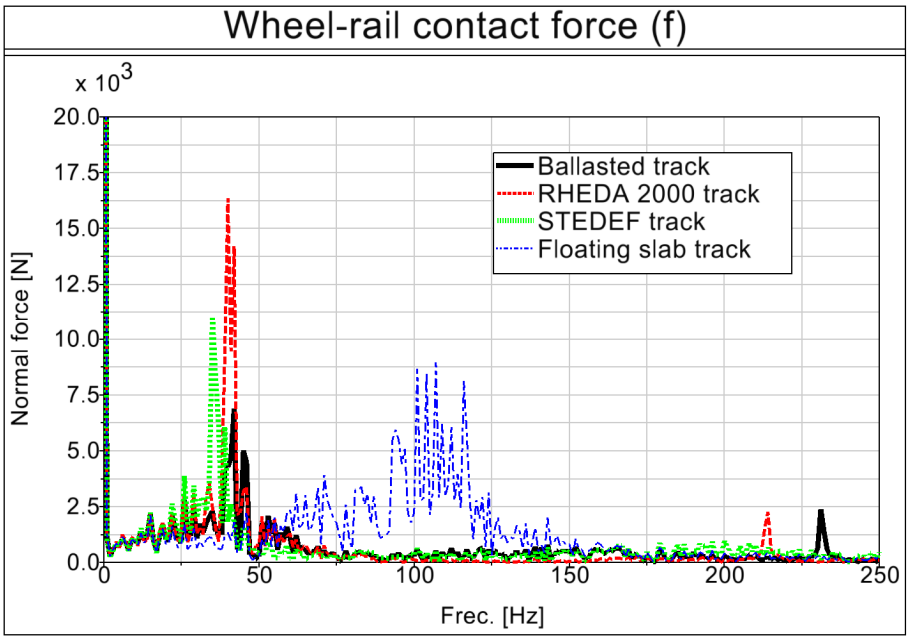
In this section, the results obtained for the different track types studied are presented, with two types of broadband vertical rail irregularities as described previously, at low and at high frequencies. All the vehicle–track interaction results presented in this section have been obtained with the EMB vehicle–track models and the complete model of the vehicle.

3.6.1 Wheel–rail contact forces

Figure 3.14 shows the time histories and corresponding frequency spectra of the wheel–rail contact forces for the different track types studied with the low frequency broadband vertical rail irregularity.



(a) Time histories.



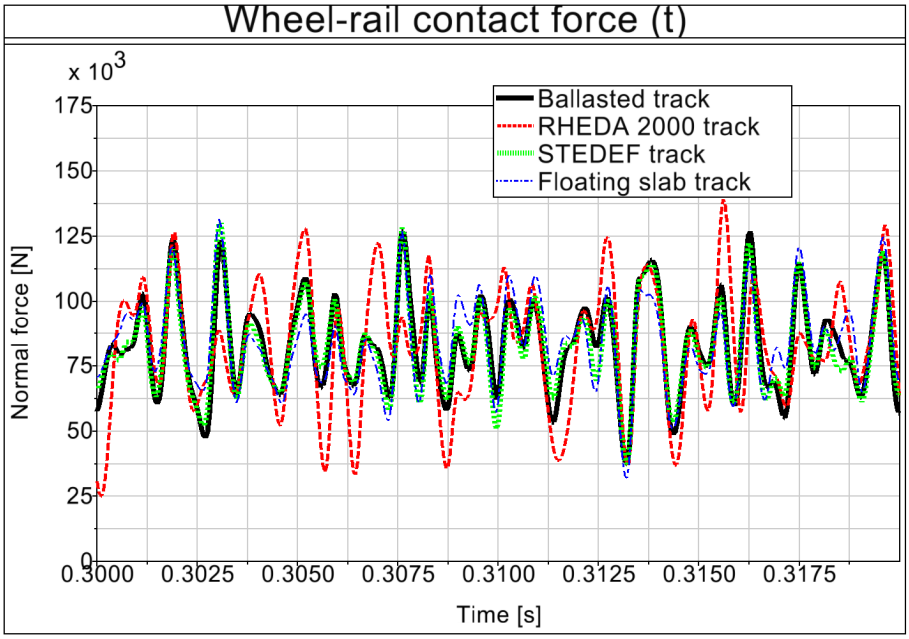
(b) Frequency spectra.

Figure 3.14. Wheel–rail contact forces with the low frequency broadband irregularity in the different tracks studied.

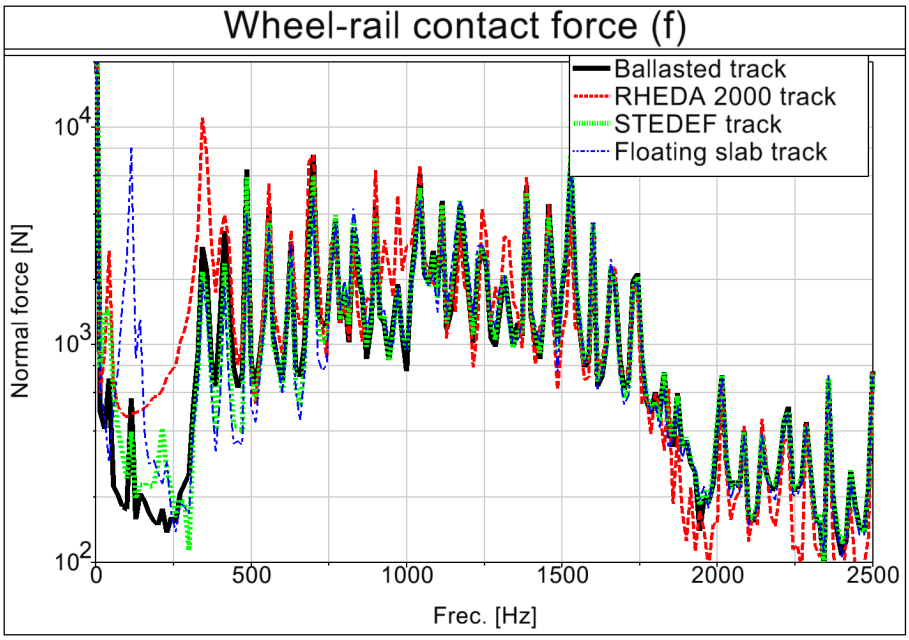
Fluctuations in the wheel–rail contact forces are larger in the RHEDA 2000 and floating slab tracks than in the ballasted and STEDEF tracks. In these latter two tracks, the first natural mode of the tracks, in which the sleepers vibrate in phase with the rails, is being excited. In the REDHA 2000 track as well, the first natural mode of the track is the one that is the most excited. In this mode, rails and base plates vibrate in phase above the lower elastic base plate pad. In the case of the floating slab track, the second mode of the track is being excited, in which the rails vibrate out of phase with the slab.

In all four tracks, the frequencies at which peaks occur in the frequency spectra of the wheel–rail contact forces match well with the frequencies of the corresponding natural vibration modes of the track calculated analytically from equivalent two-DOF systems, similar to the ones that are employed to model the track in the RMB models as described previously. To calculate these frequencies, the mass of the wheelset is added to the mass of the elastic level representing the rails, to properly account for the vibration of the whole vehicle–track system.

Figure 3.15 shows the time histories and corresponding frequency spectra of the wheel–rail contact forces for the case of the high frequency broadband vertical rail irregularity. In this case, the computed wheel–rail contact forces are very similar for the ballasted track, STEDEF track, and floating slab track. For these tracks, only significant differences are seen in the contact force spectra at frequencies related to the vibration modes of the different tracks and at the sleeper passing frequency. On the contrary, for the RHEDA 2000 track, differences in the wheel–rail forces are also noted at higher frequencies. To explain this, the receptances of the different tracks are compared in Figure 3.16.



(a) Time histories.



(b) Frequency spectra.

Figure 3.15. Wheel–rail contact forces with the high frequency broadband irregularity in the different tracks studied.

It is seen that for the ballasted, STEDEF, and floating slab tracks, for frequencies above about 500 Hz, once the different resonance frequencies of the tracks have been passed, the receptances are very similar. In these tracks, the receptances at high frequencies converge to that of the unsupported free rail (with the exception of the pinned–pinned resonance, characteristic of a discretely supported rail). In the case of the RHEDA 2000 track, the high frequency eigenmode related with the stiff railpad between the rail and the light base plate alters the track receptance at higher frequencies. Additionally, in this track, the pinned–pinned resonance happens at a slightly lower frequency than in the other tracks, due to the higher sleeper spacing (cf. Table 3.1).

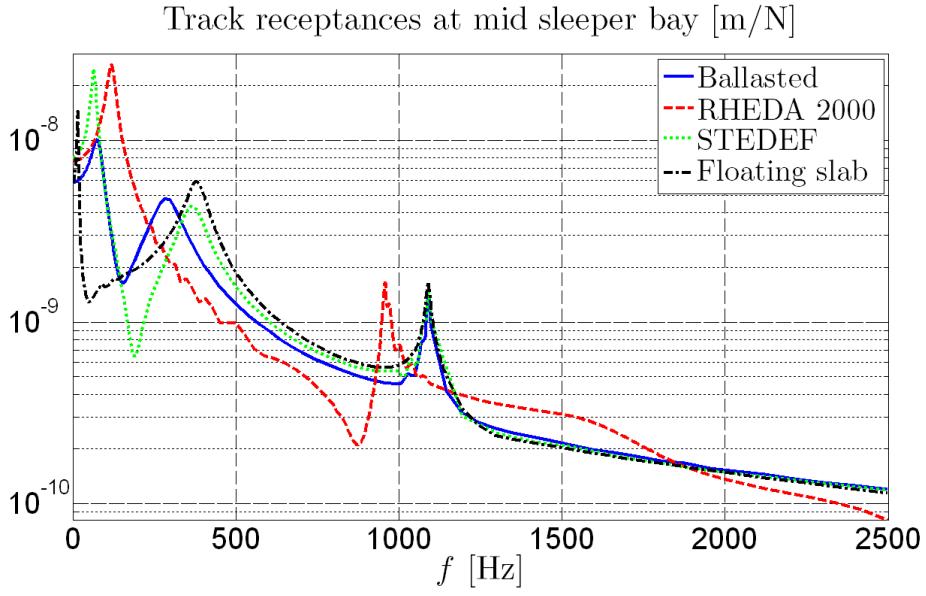


Figure 3.16. Receptances of the different tracks studied.

3.6.2 Pad forces

In this section, the pad forces are studied. Figure 3.17 shows the pad forces obtained for each type of track during a vehicle passage with low frequency broadband rail irregularity. For each type of track, the pad with the highest dynamic force is depicted.

The pad forces obtained in the floating slab track are much larger than those obtained in the other types of tracks studied. In this track, a much lower degree

of rail bending is allowed due to the high stiffness of the railpad and the slab. As a result, the number of pads sharing the vehicle load at any given time is reduced, and the peak value of the force seen by each pad is increased.

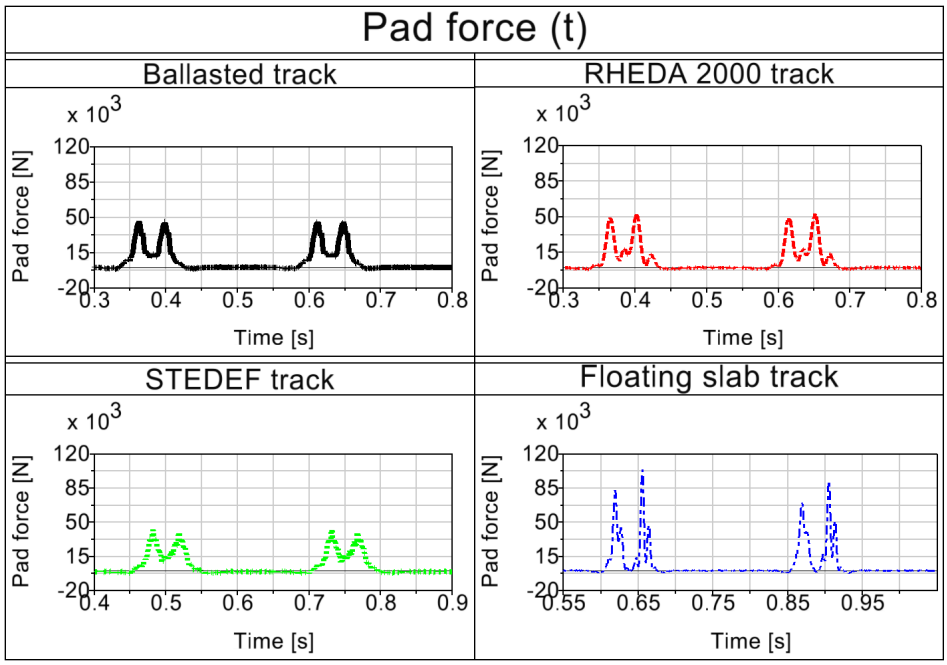


Figure 3.17. Time histories of the pad forces with the low frequency broadband irregularity in the different tracks studied.

In Figure 3.18, the pad forces obtained with the high frequency broadband rail irregularity are shown. Similar tendencies to the ones observed with the low frequency rail irregularity are seen with the high frequency rail irregularity when comparing the peak forces obtained in each type of track. The slightly larger pad force seen in the RHEDA 2000 track compared to the ballasted and STEDEF tracks is justified because, in this track, the number of railpads per sleeper length is lower than in the other tracks. In the case of this track, the force in the upper stiff railpad has been plotted. In Figure 3.19, the forces in the upper stiff railpad and in the lower flexible base plate pad are compared. Obviously, the high-frequency content in the lower base plate pad is much lower, as it is filtered by the base plate.

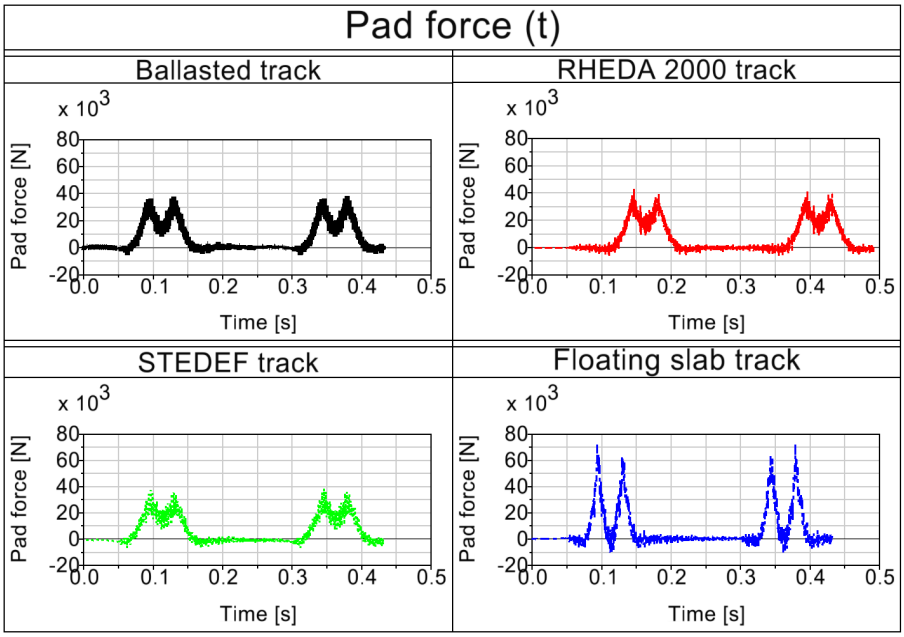


Figure 3.18. Time histories of the pad forces with the high frequency broadband irregularity in the different tracks studied.

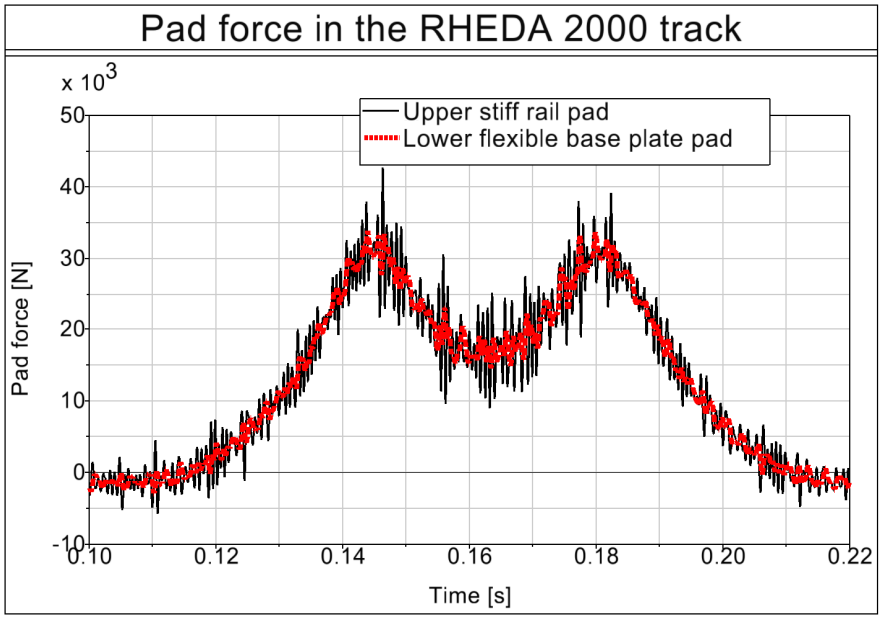


Figure 3.19. Comparison of forces in the upper stiff railpad and lower flexible base plate pad in the RHEDA 2000 track with the high frequency broadband irregularity.

Chapter 4

Numerical analysis of wheel–rail conformal contact

Chapter summary

In this chapter, comprehensive numerical models developed for the study of wheel–rail elastic conformal contact are described. These are an implementation of Kalker's exact contact theory with the necessary extensions for conformal contact, and FEM contact models. With these models it is aimed at producing detailed solutions of 3D frictional contact problems, both static and rolling. One of the main contributions of this thesis is the development of the version of the exact contact theory extended for conformal contact described here, which has been named CECT (Conformal Exact Contact Theory). The final part of the chapter includes several validation cases to demonstrate the numerical accuracy of the models.

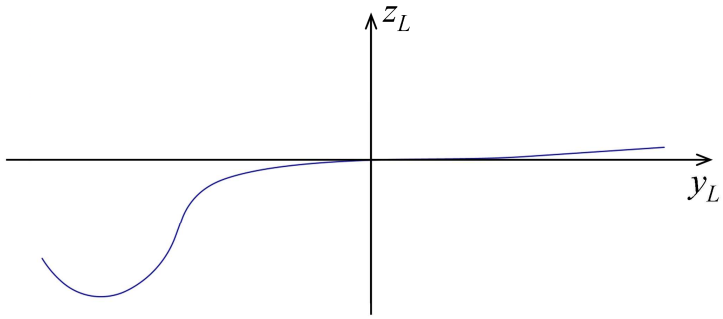
4.1 Conformal Exact Contact Theory — CECT

CECT is the implementation of the exact contact theory with extensions for non-planar conformal contact developed in this thesis. It has the capability to solve frictional contact problems, both static (compression, shift) and rolling, transient and steady state. It has been implemented in MATLAB [MATLAB www].

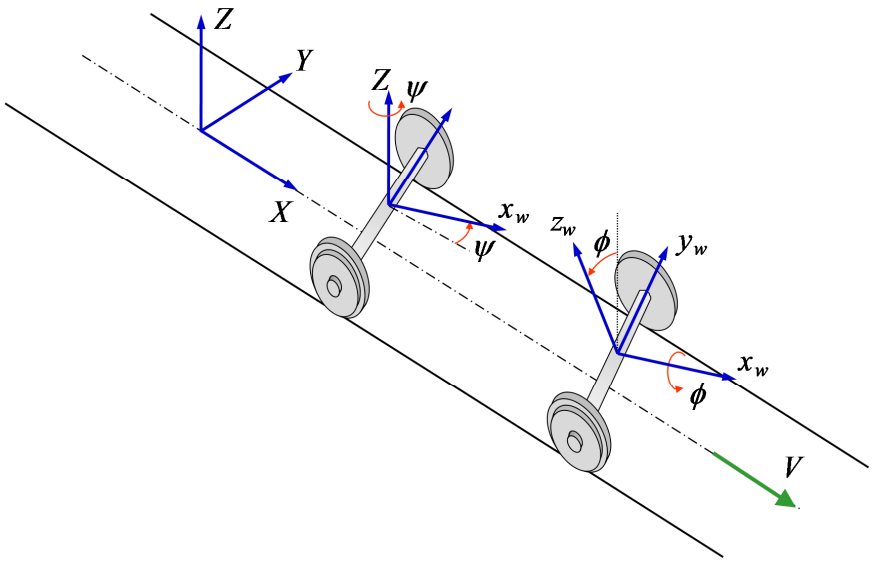
Some coordinate systems used in the specification of the contact problem are defined next, which may be used both in planar and in non-planar contact.

The wheel and rail profiles are described by a planar curve in local (y_L, z_L) Cartesian systems associated to each profile. Figure 4.1a shows an example profile of a railway wheel, with its associated local reference system. The origin of this system is typically chosen at the nominal contact point.

The positions of the rail and the wheelset are expressed in a global or track {X, Y, Z} Cartesian coordinate system. Its origin is in the middle of the two rails, at the same height as the origin of the local system of the rail profile. The X axis is aligned with the rolling direction, the Y axis points laterally to the left rail, and the Z axis completes the right-handed system, pointing vertically upwards. The X and Y axes are contained in the (usually horizontal) track plane. In addition, a {x_w, y_w, z_w} coordinate system is defined associated to the wheelset, with origin in its centre, and with the y_w axis aligned with the wheelset axis. Its orientation is obtained with a first yaw rotation ψ around the Z axis, followed by a roll rotation ϕ around the x_w axis. The global and wheelset reference systems and these rotations are depicted in Figure 4.1b.



(a) Example wheel profile and associated local (y_L, z_L) Cartesian system.



(b) Global and wheelset coordinate systems.

Figure 4.1. Coordinate systems used for the specification of the contact problem.

4.1.1 Extensions for non-planar contact

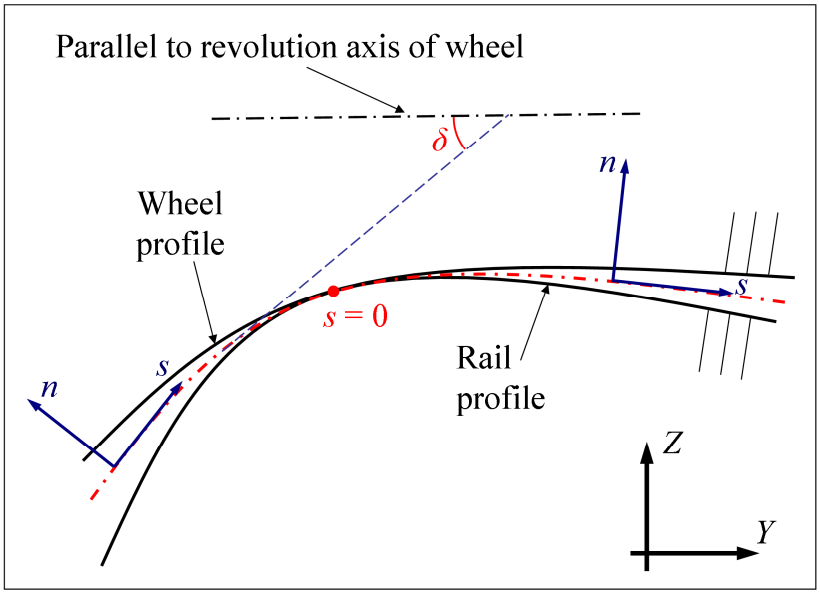
The basic hypotheses of the target contact problems are the same as those set out in §1.1.1, except for the hypothesis of concentrated contact, which is now removed. The major extensions needed to accommodate this are a more elaborate geometrical analysis to obtain the normal undeformed distance and

rigid slip velocities accurately, and the influence coefficients (ICs) for non-planar solids. These extensions are described in this section.

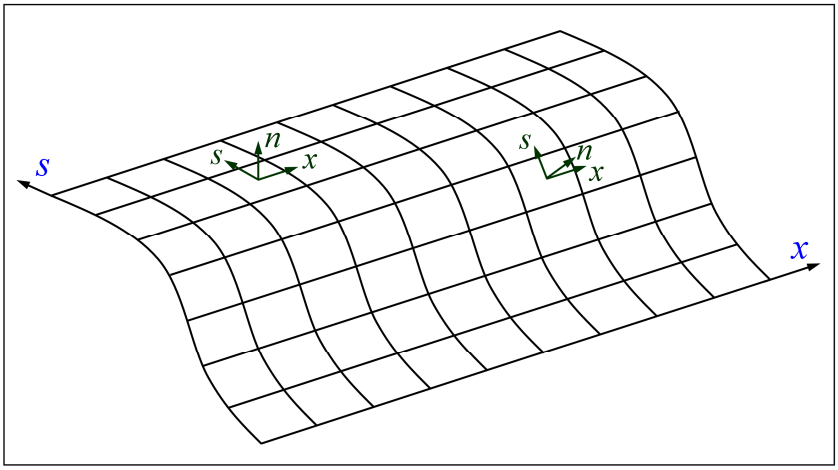
4.1.1.1 Contact surface geometry, discretization and coordinate system

A potential contact surface (PCS) is defined, which must encompass the contact patch as in planar contact. Figure 4.2 illustrates the geometry of the PCS, its discretization, and the local contact coordinate system in which the non-planar contact problem is formulated.

The tangent contact plane used in planar contact is now generalized to a curved surface. The contact is assumed to be conformal only in the plane perpendicular to the rolling direction. Accordingly, the contact surface is assumed to be prismatic, i.e. extruded, along the rolling direction, as depicted in Figure 4.2b. This is a valid assumption in the wheel–rail case, and in other relevant rolling contact applications with both bodies having curvature in the rolling direction, such as in rolling bearings, as shown in Chapter 7. The profile of the PCS is located between the wheel and rail profiles, as may be seen in Figure 4.2a. This figure shows the profile of the PCS in red chain line. This profile is calculated as an interpolation between the wheel and rail profiles, weighted with the stiffness of each of the contacting bodies. The quantity $E_k / (1 - \nu_k^2)$ is taken as the measure of the stiffness of each contacting body k for this weighting. In Figure 4.2a the contact angle δ is represented as well, a relevant geometric parameter in the contact problem, which is variable in curved contact surfaces.



(a) Section perpendicular to rolling direction.



(b) Perspective view including mesh (surface profile different to that shown in (a)).

Figure 4.2. Non-planar PCS and local $\{x, s, n\}$ curvilinear coordinate system.

A curvilinear system with orthogonal $\{x, s, n\}$ axes is defined associated to the PCS, as a generalization to the Cartesian $\{x, y, z\}$ system used in planar contact. The local vectorial quantities in the contact such as surface displacements and stresses, as well as the ICs, are expressed in this coordinate system, with x, s, n

subindices. The *s* and *n* axes are tangent and normal respectively at each point to the PCS. The *x* axis is aligned with the rolling direction, and is also tangent to the PCS. The *x* and *s* directions define the tangent plane to the contact surface in each point. The direction of the *s* and *n* axes is generally different in each lateral position of the PCS, and the *x* axis has a constant direction, as shown in Figure 4.2 a and b. Therefore, the orientation change of the *s* and *n* axes in two different lateral positions is produced by means of a rotation around the *x* axis. The origin of the *x* and *s* axes is arbitrarily defined, being most conveniently chosen at some point around the centre of the foreseen contact patch, such as the rigid contact point.

Homologous contact points, i.e. pairs of points of the contacting surfaces which may come into contact, are defined beforehand as points with the same (*x*, *s*) coordinates, as in planar contact. This is adequate in the framework of linear elasticity, with elastic displacements much lower than the contact patch dimensions, and with smooth contact surfaces. The same criteria was used in [Marshek 1984], taking as homologous points those that lie at equal distances along the profiles of their respective surfaces from the initial contact point. In [Woodward 1976] a more elaborate, iterative approach was set out, termed as "point-mating procedure", to find the correct pairs of contact points. In each iteration, the contact problem was solved with an assumed set of homologous points, and their precise locations were corrected for the next iteration according to their tangential displacements in the last obtained contact solution. In the first iteration, homologous points as defined here were considered, with the same (*x*, *s*) coordinates. The iteration could be repeated until meeting the desired tolerances on the tangential position differences of the homologous contact points.

The discretization of the PCS is carried out with a regular pattern of equal rectangular elements aligned with the *x* and *s* directions, cf. Figure 4.2b, analogously to that shown in Figure 2.18 for planar contact. The necessary parameters for its definition are the upper and lower bounds in the longitudinal and lateral direction, and the number of elements in each direction.

Based on the chosen discretization, the set of analysis points (APs) for the numerical formulation and solution of the contact problem is defined, as explained in §2.1.3. In this thesis two types of elements have been used, namely

elements with uniform and with bilinear stress shape functions. Expressions of the half-space ICs for both element types are given in §5.1. The APs are the centres of the elements in the case of the uniform elements, and the corners or nodes in the case of the bilinear elements. Therefore a number n of elements in the mesh in a given direction corresponds to n APs when using uniform elements, and to $n + 1$ APs when using bilinear elements. The number of APs in x and s directions are n_x and n_s, respectively[8]. The total number of APs is $n_x \times n_s = NE$. The concept of IC is interpreted now as related with APs rather than with elements: the IC_{IiJj} is the displacement along direction i caused at AP I due to the load distribution associated to a unit pressure along direction j at AP J (e.g. a "tent-shaped" distribution spanning 4 elements in the case of bilinear rectangular elements, c.f. §5.1.2). In foregoing discussions, the term "element" will be used in place of the term "analysis point" in some cases, when the distinction is not important for the subject discussed.

4.1.1.2 Normal undeformed distance

The normal undeformed distance h is the fundamental geometric input for the normal contact problem. Particularly in conformal contact, where the contacting surfaces are very close to one another, it is essential to carry out a precise geometric analysis to obtain a realistic representation of h. Small changes in the relative position or orientation of the contacting solids may lead to a substantial variation in the form of h.

Here the wheel is assumed to be a perfect body of revolution, and the rail straight in the rolling direction. Both contacting surfaces, generally non-planar, are considered smooth, without roughness. Roughness could be considered superposing it in a final step to the normal undeformed distances computed following the process described below.

The input data are as follows:

[8] In some cases, the number of APs in each direction will be designated as m_x and m_s, while n_x and n_s will designate the number of elements.

a) Lateral profiles of both contacting bodies. These are given by a list of value pairs representing the coordinates of the set of points chosen to represent them, which in general do not obey any analytical expression.

b) Widths of track and wheelset.

c) Nominal rolling radius of the wheel.

d) Position and orientation of the wheel with respect to the rail. This is generally given by lateral and vertical displacements u_y and u_z, a yaw angle ψ and a roll angle ϕ. A rail cant angle may be specified as well.

The procedure followed to calculate h between two non-planar conforming surfaces is based on calculating the intersections of the solids with planes perpendicular to x. Afterwards the normal undeformed distance is calculated between homologous points in the obtained intersection curves for each solid belonging to the same plane. It is not necessary to calculate the intersection curves for the rail, as they are the same as its profile. For the wheel, the starting point is the development for the calculation of the contact locus of a yawed wheel with the rail due to Wang [Wang K 1984]. A description of it may be found in Appendix D of [Li Z 2002]. The key in this development is that the distance between the intersection of the wheel axis from the normal to the possible contact point on the wheel surface, to the centre of the rolling circle of that possible contact point, is known. Additionally, the possible contact point is in the same longitudinal position as the mentioned intersection, and therefore the longitudinal distance from the possible contact point to the centre of its rolling circle is known.

The basis of this development is shown in Figure 4.3. The figure depicts a conical wheel with yaw angle ψ and roll angle ϕ. Given the orientation of the wheel, its conicity angle δ, and a rolling circle on its surface with radius R and centre at O, the problem consists on finding the possible contact point C with the rail on that rolling circle. The normal to the wheel surface on any point of the considered rolling circle intersects the wheel axis at point A. The distance from O to A, denoted as H in the figure, is equal to $R \tan(\delta)$. The rail is prismatic in X direction, and therefore the normal to any point on its surface is contained in the YZ plane. Since the normals of both contacting surfaces must be aligned at the contact point, the normal to the wheel surface at point C is also

contained in the *YZ* plane, and points *C* and *A* have the same *X* coordinate. Therefore the distance from *O* to *C* is also $R \tan(\delta)$. This, together with the known equation of the considered rolling circle, makes possible to determine the coordinates of point *C* in closed form.

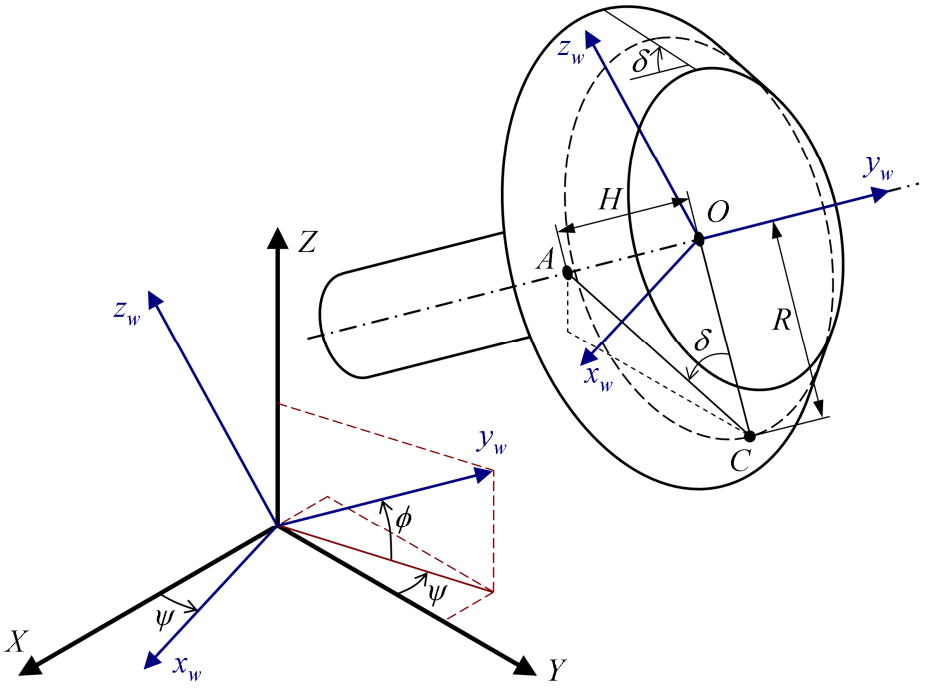


Figure 4.3. Wheel–rail contact locus.

The wheel depicted in Figure 4.3 is conical, but this development is readily applicable to a body of revolution of any shape, because finding the possible contact point on a given rolling circle of a general surface of revolution is equivalent to finding it on the cone tangent to it. Next, this development is adapted to find the intersection of the rolling circle on the wheel surface with a plane *i* perpendicular to *X*. For this purpose, the only necessary change is replacing the projection on *X* of the distance *H* with the distance Δx_i from *O* to the considered intersection plane.

The considered rolling circle is the intersection of a sphere with radius R centred at O (Eq. (4.1)) and a plane perpendicular to the wheelset axis y_w which passes through O (Eq. (4.2)).

$$(x - x_O)^2 + (y - y_O)^2 + (z - z_O)^2 = R^2 \tag{4.1}$$

$$l_x(x - x_O) + l_y(y - y_O) + l_z(z - z_O) = 0 \tag{4.2}$$

In these equations, x_O, y_O and z_O are the coordinates of O in the global $\{X, Y, Z\}$ reference system. l_x, l_y and l_z in Eq. (4.2) are the director cosines defining the orientation of the wheelset axis in the global reference system. They are given by Eq. (4.3).

$$(l_x,\ l_y,\ l_z) = (-\sin\psi\ \cos\phi,\ \cos\psi\ \cos\phi,\ \sin\phi) \tag{4.3}$$

Additionally, the distance from O to the intersection plane i is fixed:

$$x - x_O = \Delta x_i \equiv K \tag{4.4}$$

With the above equations, the coordinates $(x, y, z) = (x_{sc,i}, y_{sc,i}, z_{sc,i})$ of the intersection point of the rolling circle with the intersection plane may be calculated in closed form. From Eqs. (4.2) and (4.4):

$$(y - y_O) = \frac{-Kl_x - l_z(z - z_O)}{l_y} \tag{4.5}$$

Replacing Eqs. (4.4) and (4.5) in Eq. (4.1):

$$\frac{K^2 l_x^{\,2} + l_z^{\,2}(z - z_O)^2 + 2Kl_x l_z(z - z_O)}{l_y^{\,2}} + (z - z_O)^2 = R^2 - K^2 \tag{4.6}$$

$$\left(1 + \frac{l_z^{\,2}}{l_y^{\,2}}\right)(z - z_O)^2 + 2K\frac{l_x l_z}{l_y^{\,2}}(z - z_O) + \left[\left(1 + \frac{l_x^{\,2}}{l_y^{\,2}}\right)K^2 - R^2\right] = 0 \tag{4.7}$$

Eq. (4.7) is quadratic in ($z - z_O$). Its two possible solutions are given by Eq. (4.8). According to the geometry, the solution with minus sign in front of the square root is considered.

$$(z - z_O) = \frac{-2K\frac{l_x l_z}{l_y^2} \pm \sqrt{\frac{4K^2 l_x^2 l_z^2}{l_y^4} - 4\left(1+\frac{l_z^2}{l_y^2}\right)\left[\left(1+\frac{l_x^2}{l_y^2}\right)K^2 - R^2\right]}}{2\left(1+\frac{l_z^2}{l_y^2}\right)} \tag{4.8}$$

The term inside the square root of Eq. (4.8) is rewritten as follows:

$$\frac{4K^2 l_x^2 l_z^2}{l_y^4} - 4\left(K^2 + K^2\frac{l_x^2}{l_y^2} - R^2 + K^2\frac{l_z^2}{l_y^2} + \frac{K^2 l_x^2 l_z^2}{l_y^4} - R^2\frac{l_z^2}{l_y^2}\right) =$$

$$= \frac{-4}{l_y^2}\left(K^2 - R^2 l_y^2 - R^2 l_z^2\right) = \frac{-4}{l_y^2}\left(K^2 - R^2\left(1 - l_x^2\right)\right) \tag{4.9}$$

The equality $l_x^2 + l_y^2 + l_z^2 = 1$ has been used in Eq. (4.9). Eq. (4.8) is simplified into Eq. (4.10):

$$(z - z_O) = \frac{-2K\frac{l_x l_z}{l_y^2} - \frac{2}{l_y}\sqrt{R^2\left(1 - l_x^2\right) - K^2}}{2\left(1+\frac{l_z^2}{l_y^2}\right)}$$

$$= \frac{-K l_x l_z - l_y\sqrt{R^2\left(1 - l_x^2\right) - K^2}}{l_y^2 + l_z^2} \equiv R' \tag{4.10}$$

The quantity in the right hand side of Eq. (4.10) is now designated as R'. ($y - y_O$) is calculated replacing ($z - z_O$) by R' in Eq. (4.5). The coordinates (y_O, z_O) of O in the global reference system are given in Eqs. (4.11) and (4.12). In Eq.

(4.11), y_w is the axial coordinate of O in the wheelset $\{x_w, y_w, z_w\}$ reference system. The origin of the wheelset system is initially located at coordinates (0, 0, $R_{roll,n}$) in the global system, being $R_{roll,n}$ the nominal rolling radius of the wheel. Its final position is obtained with translations (u_y, u_z) in the global Y and Z directions.

$$y_O = u_y + l_y \times y_w \tag{4.11}$$

$$z_O = R_{roll,n} + u_z + l_z \times y_w \tag{4.12}$$

Finally, the (global) lateral and vertical coordinates of the intersection point with plane i of the rolling circle located at lateral position y_w of the wheelset are given in Eqs. (4.13) and (4.14). This intersection is shown schematically in Figure 4.4, with some of the relevant parameters in the calculation. Δx_i, R and R' are in general variable with y_w.

$$y_{sc,i}(y_w) = u_y + l_y \times y_w - \frac{l_x \times \Delta x_i(y_w) - l_z R'}{l_y} \tag{4.13}$$

$$z_{sc,i}(y_w) = R_{roll,n} + u_z + l_z \times y_w - R' \tag{4.14}$$

The equations in the original development for the wheel contact locus, reproduced in Eqs. (4.78)–(4.80), are singular for $\delta = 90^o$, due to the tan(δ) term in the distance H. This problem is not encountered in Eqs. (4.13) and (4.14), because the distance Δx_i is defined instead. So, the method explained here for the calculation of h may be used with unrestricted δ values, as e.g. in the cylindrical contact cases of §4.3.7. This method may also be used with large ψ and ϕ rotation angles.

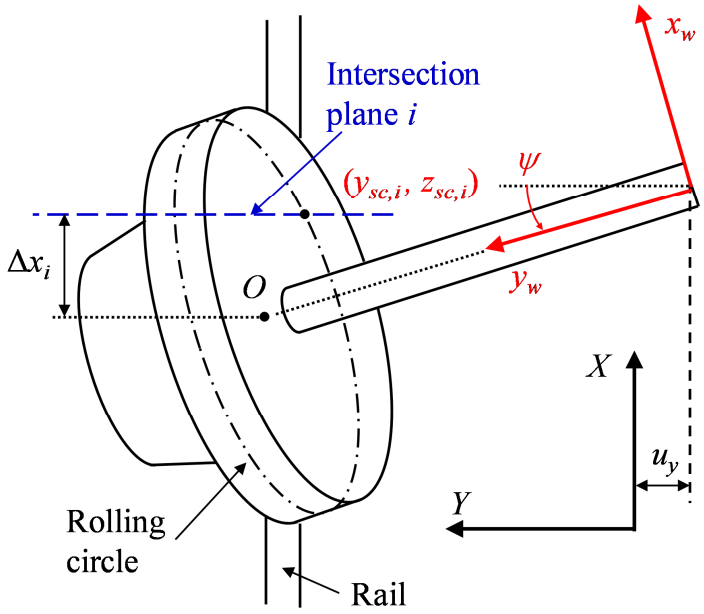


Figure 4.4. Intersection of a rolling circle of the wheel with a plane perpendicular to the longitudinal direction.

The main steps followed for the calculation of the normal undeformed distances between non-planar surfaces are listed below. These steps are repeated for each longitudinal position x_i in the discretization of the PCS.

1) Calculation of the signed longitudinal distance $\Delta x_i\ (y_w)$ from each longitudinal position i of the mesh ($i \in \{1, \ldots, n_x\}$), defining the location of each of the planes with which the intersection of the wheel surface is to be calculated, to the points in the revolution axis of the wheelset corresponding to each of the points representing the given profile of the wheel $z_w\ (y_w)$. These points in the revolution axis of the wheel are the centres of the rolling circles of the wheel containing the points in which its profile is discretized. If X_i is the X coordinate of the longitudinal position i of the mesh in the global coordinate system (and also in the wheelset coordinate system), $\Delta x_i\ (y_w) = X_i - l_x\ y_w$.

2) Calculation of the intersection curve *xsec_w* (x_i) of the wheel with the plane perpendicular to x located in the current longitudinal position x_i of

the mesh, with Eqs. (4.13) and (4.14). This curve has as many points ($y_{sc,i}$, $z_{sc,i}$) as the input wheel profile.

3) Parameterization of the calculated intersection curve *xsec_w* with the lateral curvilinear coordinate *s*, which represents its length. With this, the intersection curve is mapped in the *s* coordinate. $s = 0$ is assigned for the first point, and the *s* values for the rest of the points are calculated recursively, summing length increments of the curve between consecutive points. For a point *j*, $s_j = s_{j-1} + \Delta l_j$, being s_j and s_{j-1} the *s* coordinates of consecutive points *j* and *j*–1, respectively, and Δl_j the distance between these points. The intersection curve is taken as piecewise linear in this calculation.

4) Search for the origin of the *s* coordinate on the intersection curve *xsec_w*, which is defined to be on the point at the minimum distance from the chosen origin point of the local contact coordinate system. This is a search of zeros in a non-linear 1D curve, that is done iteratively. The *s* values obtained in the previous step are shifted so that $s = 0$ at the contact origin.

5) Interpolation of the intersection curve *xsec_w* (x_i) in the *s* coordinates of the APs of the mesh, obtaining the interpolated curve *xsec_w_i* (x_i).

   Depending on the geometry and the sampling of the profiles with respect to the discretization of the PCS, attention hast to be paid to this interpolation to avoid noise in the solution of the normal problem, as this is sensitive to small variations in the underformed distance. As an example of this, Figure 4.5 shows a conformal contact case in which using linear interpolations in this step of the computation of *h* leads to some noise in the computed normal pressure (p_n) distribution. This is the "circ. a." case described in §6.2 (cf. Table 6.1). In this case, the lateral curvature radii of both contacting surfaces are constant in the contact patch, and smooth p_n distributions are expected. Figure 4.5 shows the p_n profiles along *s* in a cross-section close to the centre of the contact patch, computed with CECT for this case using two interpolation types in this step. With spline interpolation, the noise in the computed p_n distribution is avoided, as may be seen in the figure. In this computation, the contact patch was discretized with approximately

36 elements in the lateral direction, and the wheel and rail profiles were defined with approx. 85 and 150 points respectively in the same zone.

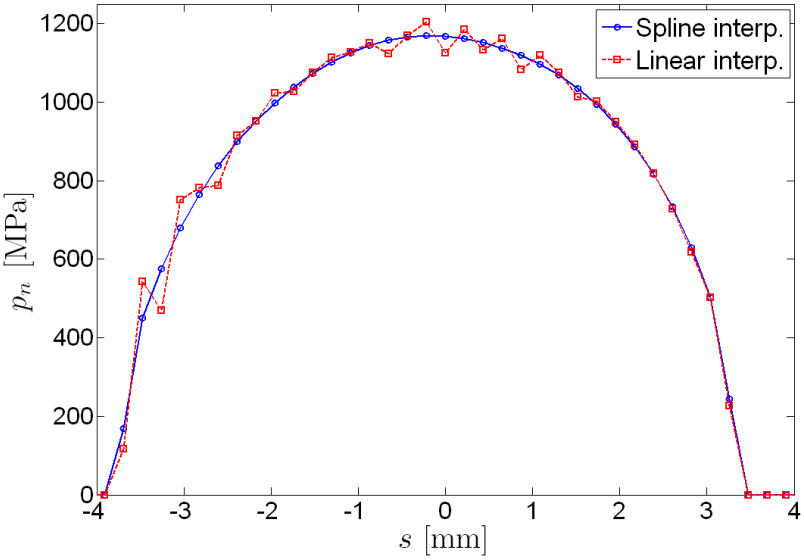


Figure 4.5. Effect of interpolation for the calculation of h on the resulting distribution of p_n in a conformal contact case.

Steps 3) to 5) above are performed in a similar way for the rail. For a straight rail, they are performed only once and not at each longitudinal position, since the rail cross-section *xsec_r* does not change, and the intersection curve is given directly by the rail profile (with a rigid body translation and rotation if so defined, cf. `incl_r` and `hw_tr` input parameters in §4.1.2.1.1). For a curved rail, Eqs. (4.13) and (4.14) for the intersection are particularized for the case of zero yaw angle as explained in §7.2.2.1.1. The profile of the PCS, *prf_pcs_i*, is calculated as an average of the wheel and rail interpolated cross-sections obtained at x_0, the longitudinal position where the origin of the local contact coordinate system is located, *xsec_w_i* (x_0) and *xsec_r_i* (x_0), weighted with the elastic constant $E_k / (1 - \nu_k^2)$ of each body k as indicated in §4.1.1.1.

6) h is calculated as the projection over the local n direction of the PCS of the vector between homologous points of the intersection curves of the wheel and the rail. Figure 4.6 shows this calculation schematically. For two points on the wheel and rail surfaces located at coordinates ($Y_{w,j}$, $Z_{w,j}$) and ($Y_{r,j}$, $Z_{r,j}$) in the YZ plane, considering a n direction defined by

its cosine directors $(-\sin(\alpha_j), \cos(\alpha_j))$ in the same plane, h is equal to $(Z_{w,j} - Z_{r,j}) \cos(\alpha_j) - (Y_{w,j} - Y_{r,j}) \sin(\alpha_j)$. This calculation is done for each lateral position of the discretization. This projection step serves to avoid jumps in h from non-zero positive to negative values in regions in which the undeformed surfaces change from separation ($h > 0$) to overlap ($h < 0$).

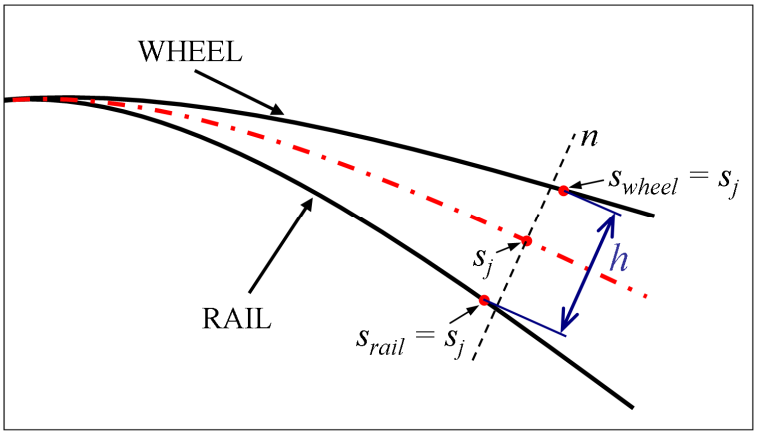


Figure 4.6. Calculation of the normal undeformed distance h between non-planar surfaces.

Two approaches are used to calculate the local n direction of the PCS in each lateral position of the mesh:

a) If n_s is higher than the number of points within the mesh range in s direction of both the wheel and rail reference cross-sections *xsec_w* (x_0) and *xsec_r* (x_0), the regularly discretized profile of the PCS calculated above *prf_pcs_i* is used directly. The n direction at lateral position i_s ($i_s \in \{1, \ldots, n_s\}$) is determined as the normal to the line joining the APs of the PCS at positions $i_s - 1$ and $i_s + 1$. As an exception, the n direction at $i_s = 1$ is determined with APs at positions 1 and 2, and the n direction at $i_s = n_s$ with APs at positions $n_s - 1$ and n_s. This calculation is shown schematically in Figure 4.7, where the circles mark positions of APs of the mesh in the YZ plane. This scheme is used both with uniform and with bilinear elements; thus, the lines joining adjacent positions of the mesh in the figure are element edges only for the case of bilinear elements,

in which the APs are located at the element corners. The *n* direction is defined with an angle α in the *YZ* plane in the $[-\pi, \pi]$ range, being 0 for *n* aligned with the positive *Z* axis, and increasing counter clock-wise as looked towards the negative *X* axis, as depicted in Figure 4.7.

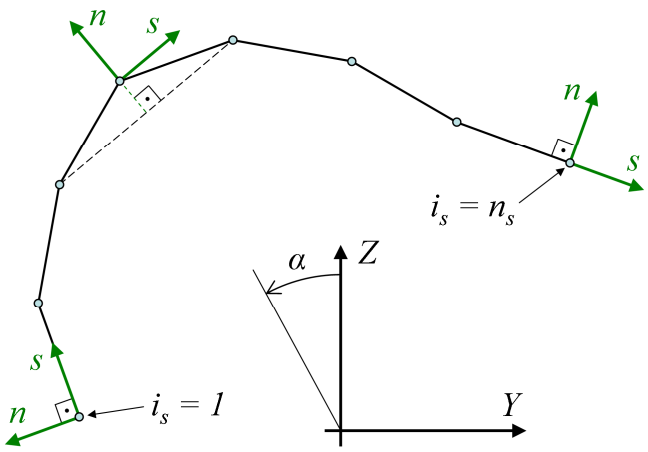
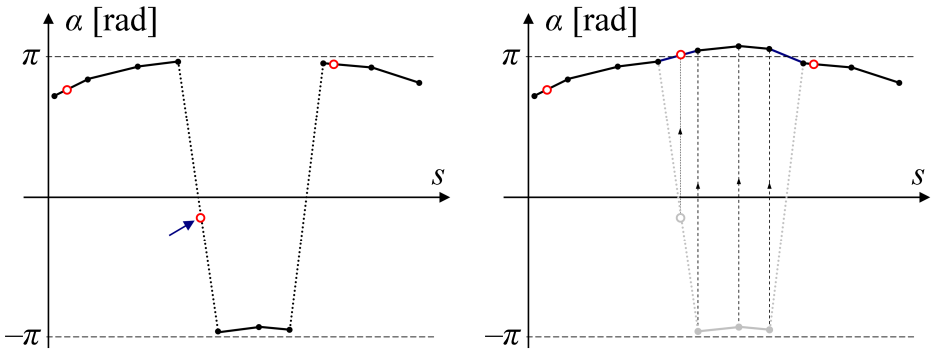


Figure 4.7. Calculation of the *n* direction at each lateral position of the mesh of the PCS.

b) If n_s is lower than the number of points within the mesh range in *s* direction of either cross-section *xsec_w* (x_0) or *xsec_r* (x_0), a more refined profile of the PCS is used, *prf_pcs_ref*. To calculate it, the *s* coordinates of the cross-section *xsec_w* (x_0) or *xsec_r* (x_0) with more points within the mesh range are used. The other cross-section is sampled at these *s* coordinates, and *prf_pcs_ref* is computed as a weighted average of the sampled cross-section and the other one, in a similar way as *prf_pcs_i*.

The α angles defining the orientation of *prf_pcs_ref* in the *YZ* plane are then calculated at each point of *prf_pcs_ref*, in a similar way as depicted in Figure 4.7 for *prf_pcs_i*. In this way, a curve of angles α_{ref} (s_{ref}) is obtained on *prf_pcs_ref*. The α angles on *prf_pcs_i* are obtained by sampling (i.e. interpolating) α_{ref} (s_{ref}) on the *s* positions of the mesh. This interpolation is problematic if the point of discontinuity of α (i.e., the point at which α changes from $-\pi$ to π,

which, according to the convention for α explained above, is where the n direction points in the negative Z axis) is within the range of the mesh. The problem is illustrated in Figure 4.8a, with linear interpolation. Here, the points of *prf_pcs_ref* are represented with filled circles, and the points of the mesh with red unfilled circles. If a point of the mesh falls between two discontinuous points of *prf_pcs_ref*, as the point marked with an arrow in the figure, the interpolated α may be very bad. This may lead to a large error in the computed undeformed distance at that s position of the mesh, and subsequently to a large peak in the normal pressure when solving the normal contact problem.



(a) Interpolation error at discontinuity point. (b) Redefinition of α_{ref} curve for continuity.

Figure 4.8. Treatment of a discontinuous α_{ref} curve.

To avoid this, the α_{ref} curve is made continuous by redefining it in an extended angular range, as in the example depicted Figure 4.8b. If a discontinuity is detected between two adjacent α_i and α_{i+1} values, α_{i+1} and subsequent α values up to the next discontinuity or to the end of the curve are modified, adding 2π if $\alpha_{i+1} < \alpha_i$, or subtracting 2π otherwise. In the example of Figure 4.8b, there are two discontinuities, one from π to $-\pi$, and the other from $-\pi$ to π. Translating the section of the curve between both discontinuities by $+2\pi$ in the vertical axis, both discontinuities are resolved at once. Taking this into account, it could be thought of considering only the odd-numbered discontinuities in the curve as the starting points of the sections of the curve to be translated. But this would not be

correct; the possibility of having more than one consecutive discontinuity in the same direction (from $-\pi$ to π or vice versa) has to be considered. This happens in a "heart-shaped" closed profile, for example.

The algorithm to make the α_{ref} curve continuous is as follows:

1. Detect the discontinuity points in the α_{ref} curve, as the ones in which the absolute difference between two adjacent angles is higher than π. The number of these discontinuities is denoted as n_{discn}.
2. Determine the sign *sgn* of each discontinuity. *sgn* = –1 for a discontinuity from $-\pi$ to π, and *sgn* = 1 for a discontinuity from π to $-\pi$.
3. Set counter i = 1. The following loop is carried out until i surpasses the number of discontinuities in the curve n_{discn}:
    - Determine the position index in the curve of the current discontinuity i, and its sign *sgn*.
    - Search for the next discontinuity in the curve with opposite sign to the current one i. If there is such discontinuity, define i_2 as its position index in the set of discontinuities. Otherwise, $i_2 = n_{discn} + 1$.
    - Set counter $ii = 0$. If, in the search of the previous step, it is determined that there is one or more discontinuities adjacent to the current one with the same sign, the following is carried out $i_2 - i - 1$ times:
        - Translate the section of the curve from the position of discontinuity $i + ii$ to the end by $sgn \times 2\pi$.
        - ii += 1.
    - Translate the section of the curve between the discontinuity with index $i + ii$ and the next, which is the one with index i_2 (or up to the end of the curve, if

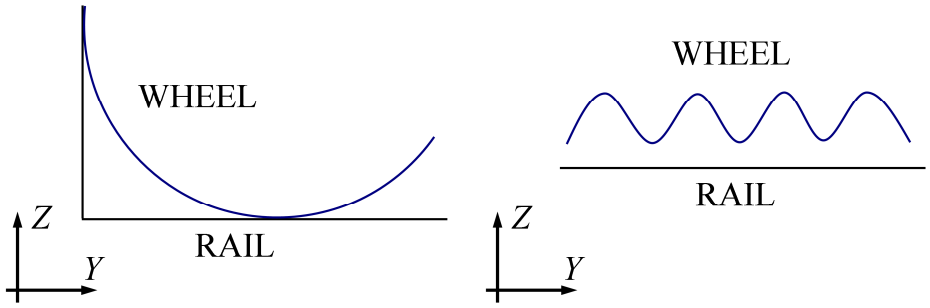
discontinuity *i* + *ii* is the last one of the curve), by $sgn \times 2\pi$.

- $i = i_2 + 1$.

After making α_{ref} continuous, the interpolation for α is carried out safely. In a final step, the interpolated α values are bounded again to the $[-\pi, \pi]$ range.

In 2D problems with $n_s = 1$, the *n* direction points towards the positive *Z* axis.

Homologous contact points are defined as points with the same (*x*, *s*) coordinates, as indicated before. By using the *s* coordinate for the location of homologous contact points in the *YZ* plane, multi-valued wheel and rail profiles (i.e. profiles with more than one point in a given lateral position, or profiles with vertical sections) may be handled without problem. But, in order for pairs of points with common *s* coordinates to really be homologous points, both contacting surfaces have to be close to one another in the whole PCS, and they have to be smooth, i.e., without roughness or waviness at a scale lower than the contact dimensions. Figure 4.9 illustrates two examples that do not fulfil these conditions. These cases cannot be correctly treated in CECT. It would be relatively straightforward to include roughness, superposing it in a final step to the "smooth" *h* previously calculated following the method described above.



(a) Profiles that separate and join again (adapted from [Vollebregt 2021]).

(b) Profiles with waviness.

Figure 4.9. Pairs of profiles not valid for the location of homologous contact points as those with common *s* coordinates.

The above referenced normal undeformed distance h corresponds to the initial configuration (cf. §1.2). The undeformed distance in the reference configuration, i.e. after application of the approach d between the contacting bodies, is also of interest. This will be designated as h'. Figure 4.10 illustrates the effect of the approach in h'. In a non-planar contact surface, this effect is variable with the surface orientation. Following the figure, h' is calculated from h and d according to Eq. (4.15). In this equation, $\Delta\alpha$ is the angular difference in the YZ plane between the direction in which the approach is applied and the local normal of the contact surface.

$$h' = h - d\cos(\Delta\alpha) \tag{4.15}$$

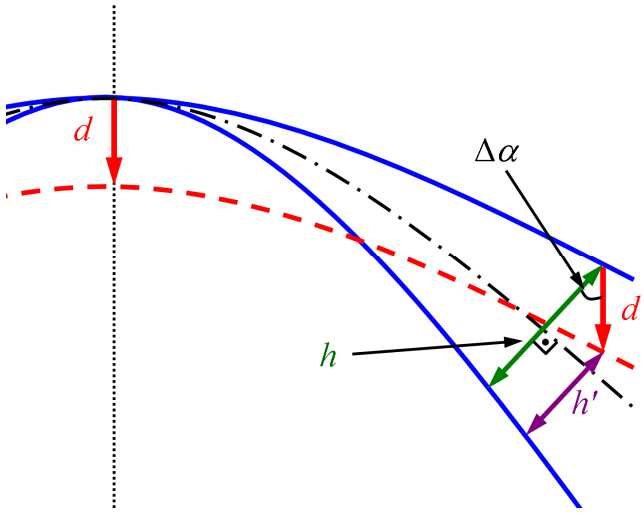


Figure 4.10. Effect of the approach in the normal undeformed distance.

4.1.1.3 Rigid slip velocities

The rigid slip velocities **w** or imposed rigid shifts are the fundamental kinematic input for the tangential contact problem. In planar contact surfaces the field of **w** may be described in terms of the longitudinal, lateral and spin creepages, ξ, η and φ, according to Eq. (1.5). In non-planar contact surfaces, the notion of creepages as applicable to the whole contact area loses its sense (though it is still possible to consider creepages referred to particular points in the contact patch), and it becomes necessary to compute the local rigid shifts or slip

velocities at each point, as explained in [Piotrowski 1999]. This is done computing the rigid body velocities of all the APs of the mesh of the PCS as belonging to each of the contacting solids, and projecting the resulting velocity difference vector of each AP on the PCS. The projection at a point *I* on a curved surface is equal to the projection on the plane tangent to that surface at *I*.

The necessary input data for this calculation are as follows:

- $\boldsymbol{\omega}$, the angular velocity vector of the wheel, with components ω_X, ω_Y, and ω_Z in the global {*X*, *Y*, *Z*} system.
- $\mathbf{v_P}$, the linear velocity vector of a point *P* of the wheel, as well as the position of that point. $\mathbf{v_P}$ has components $v_{P,X}$, $v_{P,Y}$ and $v_{P,Z}$ in the {*X*, *Y*, *Z*} system.
- The points of the mesh of the PCS on which the rigid slip velocities are to be computed. The vector from point *P* to an AP *I* in the PCS is $\mathbf{r_{I-P}}$.
- The angles of inclination α of the PCS in the *YZ* plane. The angle at AP *I* is α_I. These angles are measured from the *Y* axis to the tangent to the PCS directed in the positive *s* direction, counter clock-wise as viewed looking towards the negative *x* direction. They have the same value as the α angle defining the direction of the local normal depicted in Figure 4.7. As the PCS is prismatic, the α angles vary only in the *s* direction.

The above data does not include rail velocities, because the rail is considered to have no rigid body motion. This does not imply any loss of generality; if the rail is moving, the wheel motion relative to the rail is considered to define the necessary kinematic input, and the problem is solved as if the rail were at rest.

A new coordinate system {*u'*, *v'*, *w'*} is defined, with its *v'* axis parallel to $\boldsymbol{\omega}$. This system is obtained from the global system {*X*, *Y*, *Z*} with the following steps:

- Rotation of ψ_ω around the *Z* axis. After this rotation the orientation of the *u'* axis, which is obtained from the *X* axis, is attained.
- Rotation of ϕ_ω around the *u'* axis.
- The origin of the system {*u'*, *v'*, *w'*} is translated to point *P* on which the velocity $\mathbf{v_P}$ is given.

The angles ψ_ω and ϕ_ω, which define the orientation of $\boldsymbol{\omega}$, are analogous to the angles ψ and ϕ associated to the orientation of the revolution axis of the wheel, shown in Figure 4.1b. These angles, depicted in Figure 4.11, may be computed from the components of $\boldsymbol{\omega}$ according to Eqs. (4.16) and (4.17):

$$\psi_\omega = \operatorname{atan}\left(-\omega_X / \omega_Y\right) \tag{4.16}$$

$$\phi_\omega = \operatorname{sgn}\left(\omega_Z\right) \times \operatorname{acos}\left(\operatorname{sgn}\left(\omega_Y\right) \frac{\sqrt{{\omega_X}^2 + {\omega_Y}^2}}{\|\boldsymbol{\omega}\|} \right) \tag{4.17}$$

According to the defined rotations, the relationship between the unitary vectors along the axes of the systems $\{u', v', w'\}$ and $\{X, Y, Z\}$ is given by Eq. (4.18). **i**, **j**, **k** are the unitary vectors along the axes of the global system, and **i'**, **j'**, **k'** the unitary vectors along the axes of the system $\{u', v', w'\}$. $\mathbf{[R_a}(\gamma)\mathbf{]}$ is the rotation matrix of angle γ around axis **a** (see also §7.1.1), $\mathbf{[A]}^{\mathrm{T}}$ is the transpose of matrix **[A]**, and s_γ and c_γ are the sine and the cosine of γ.

$$\begin{Bmatrix} \mathbf{i'} \\ \mathbf{j'} \\ \mathbf{k'} \end{Bmatrix} = \left[\mathbf{R}_{\mathbf{u'}}\left(\phi_\omega\right)\right]^{\mathrm{T}} \left[\mathbf{R}_{\mathbf{Z}}\left(\psi_\omega\right)\right]^{\mathrm{T}} \begin{Bmatrix} \mathbf{i} \\ \mathbf{j} \\ \mathbf{k} \end{Bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{\phi_\omega} & s_{\phi_\omega} \\ 0 & -s_{\phi_\omega} & c_{\phi_\omega} \end{bmatrix} \begin{bmatrix} c_{\psi_\omega} & s_{\psi_\omega} & 0 \\ -s_{\psi_\omega} & c_{\psi_\omega} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{Bmatrix} \mathbf{i} \\ \mathbf{j} \\ \mathbf{k} \end{Bmatrix}$$

$$= \begin{bmatrix} c_{\psi_\omega} & s_{\psi_\omega} & 0 \\ -c_{\phi_\omega} s_{\psi_\omega} & c_{\phi_\omega} c_{\psi_\omega} & s_{\phi_\omega} \\ s_{\phi_\omega} s_{\psi_\omega} & -s_{\phi_\omega} c_{\psi_\omega} & c_{\phi_\omega} \end{bmatrix} \begin{Bmatrix} \mathbf{i} \\ \mathbf{j} \\ \mathbf{k} \end{Bmatrix} \tag{4.18}$$

Figure 4.11 illustrates the projection of a velocity vector $\mathbf{v_{I/P}}$ on an *xs* plane with inclination α_I with respect to the *XY* plane. $\mathbf{v_{I/P}}$ is the velocity of a point *I* relative to point *P*, considering both points as solidary to the wheel. Thus $\mathbf{v_{I/P}} = \boldsymbol{\omega} \times \mathbf{r_{I\text{-}P}}$, and the equalities of Eq. (4.19) hold.

$$\mathbf{v}_{\mathbf{I/P}} = \boldsymbol{\omega} \times \mathbf{r}_{\mathbf{I\text{-}P}} = \boldsymbol{\omega} \times \mathbf{r}_{\mathbf{I\text{-}P'}} = \|\boldsymbol{\omega}\| \cdot \|\mathbf{r}_{\mathbf{I\text{-}P'}}\| \cdot \left(-c_{\theta_I}\,\mathbf{i}' - s_{\theta_I}\,\mathbf{k}'\right) \tag{4.19}$$

$\mathbf{r}_{\mathbf{I\text{-}P'}}$ is the projection of $\mathbf{r}_{\mathbf{I\text{-}P}}$ on a plane perpendicular to $\boldsymbol{\omega}$, and P' is the projection of P on the plane perpendicular to $\boldsymbol{\omega}$ passing through I. θ_I is the azimuthal coordinate of point I in the $\{u', v', w'\}$ system, measured as shown in Figure 4.11, with origin in the negative w' axis, and positive clock-wise while looking towards the negative v' direction. The circumference centred at P', perpendicular to $\boldsymbol{\omega}$ and with radius $\mathbf{r}_{\mathbf{I\text{-}P'}}$ is represented in dashed line in the figure. $\mathbf{v}_{\mathbf{I/P}}$ is tangent to this circumference at I.

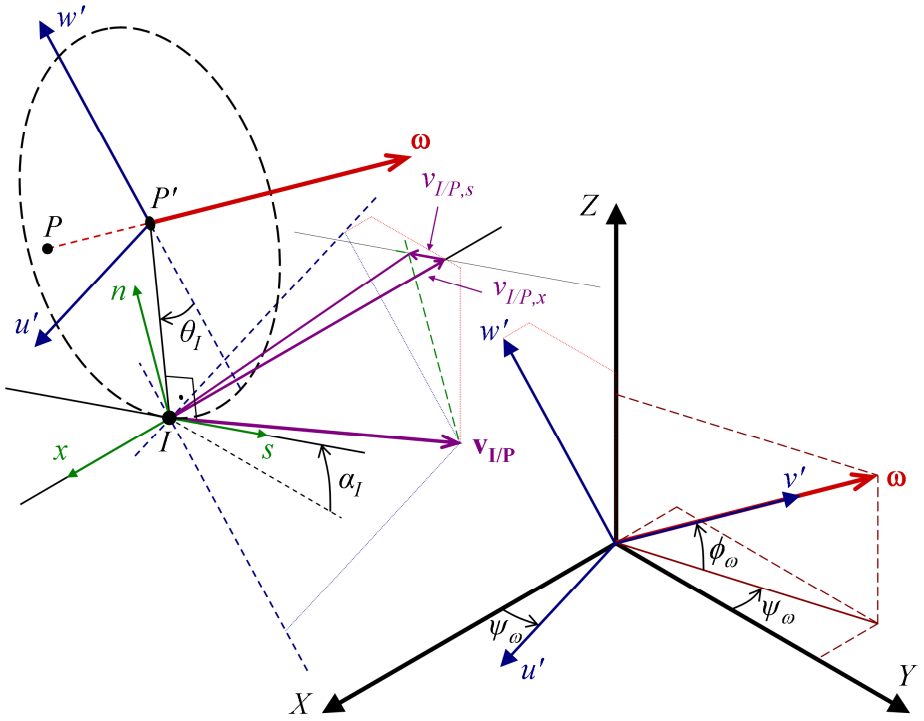


Figure 4.11. Projection of rigid slip velocity on tangent *xs* plane. Isometric view.

The transformation between the systems $\{x, s, n\}$ and $\{X, Y, Z\}$ is defined with the rotation matrix $\mathbf{[R_X(\alpha_I)]}$ according to Eq. (4.20). The unitary vectors along the axes of the system $\{x, s, n\}$ are $\mathbf{e_1}$, $\mathbf{e_2}$ and $\mathbf{e_3}$.

$$\begin{Bmatrix}\mathbf{e_1}\\ \mathbf{e_2}\\ \mathbf{e_3}\end{Bmatrix}=\left[\mathbf{R_X}(\alpha_I)\right]^{\mathrm{T}}\begin{Bmatrix}\mathbf{i}\\ \mathbf{j}\\ \mathbf{k}\end{Bmatrix}=\begin{bmatrix}1 & 0 & 0\\ 0 & c_{\alpha_I} & s_{\alpha_I}\\ 0 & -s_{\alpha_I} & c_{\alpha_I}\end{bmatrix}\begin{Bmatrix}\mathbf{i}\\ \mathbf{j}\\ \mathbf{k}\end{Bmatrix} \tag{4.20}$$

Combining Eqs. (4.18)–(4.20), $\mathbf{v_{I/P}}$ is written on the axes $\{x, s, n\}$. The total rigid velocity of point I of the wheel, $\mathbf{v_I}$, is the sum $\mathbf{v_P}$ and $\mathbf{v_{I/P}}$. As the rail is considered at rest, this is equal to the rigid slip velocity at the same point I, $\mathbf{w_I}$. The components of $\mathbf{w_I}$ along axes x and s, $w_{I,x}$ and $w_{I,s}$, are given in Eqs. (4.21) and (4.22). The terms in the right-hand side of these equations multiplied by $\|\boldsymbol{\omega}\|\cdot\|\mathbf{r_{I\text{-}P'}}\|$ are due to $\mathbf{v_{I/P}}$. These are represented in Figure 4.11 as components $v_{I/P,x}$ and $v_{I/P,s}$. The component of $\mathbf{w_I}$ in normal direction, $w_{I,n}$, is accommodated by normal deformations, and does not enter in the specification of linear elastostatic contact problems. For a continuous contact, $w_{I,n}$ is small, otherwise the contacting surfaces would be colliding or separating.

$$w_{I,x}=v_{P,X}+\|\boldsymbol{\omega}\|\cdot\|\mathbf{r_{I-P'}}\|\cdot\left(-c_{\theta_I}c_{\psi_\omega}-s_{\theta_I}s_{\psi_\omega}s_{\phi_\omega}\right) \tag{4.21}$$

$$\begin{aligned}w_{I,s}=&\,v_{P,Y}c_{\alpha_I}+v_{P,Z}s_{\alpha_I}+\|\boldsymbol{\omega}\|\cdot\|\mathbf{r_{I-P'}}\|\\ &\times\left(-c_{\alpha_I}c_{\theta_I}s_{\psi_\omega}+c_{\alpha_I}s_{\theta_I}c_{\psi_\omega}s_{\phi_\omega}-s_{\alpha_I}s_{\theta_I}c_{\phi_\omega}\right)\end{aligned} \tag{4.22}$$

Besides the rigid slip velocities due to the rigid body motion of the contacting solids, in transient contact problems the rigid shifts or slip velocities due to the change in the (normal) approach between the solids have to be considered as well. As illustrated in Figure 4.12, a change in the approach of Δd causes a rigid shift in s direction $W_{I,s}$ equal to $-\Delta d \times \sin(\Delta\alpha_{I\text{-}d})$, $\Delta\alpha_{I\text{-}d}$ being the angle between the direction of Δd and the normal of the PCS at I. This is a form of coupling between the normal and the tangential parts of the contact problem, which does not appear in planar contact. Related examples are shown in §6.2.1.5.

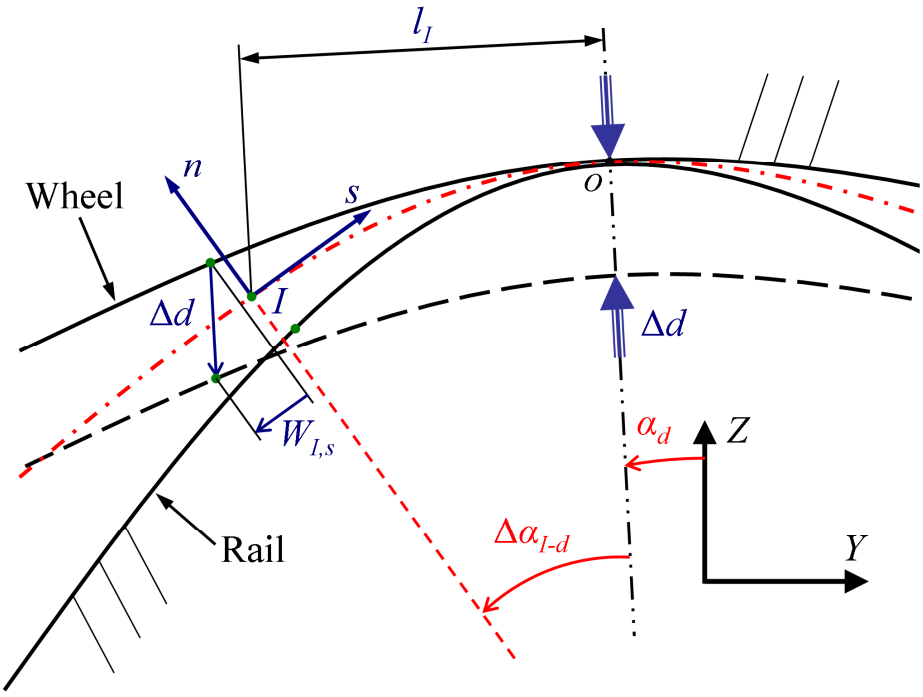


Figure 4.12. Contribution of the variation of the approach to the rigid tangential shift.

In transient rolling contact problems, this rigid shift $W_{I,s}$ is divided by the time increment $\Delta t = t - t'$ between the previous and the current time step and added to the velocity $w_{I,s}$ of Eq. (4.22). In static contact problems, the relative motion between the contacting surfaces is assumed to be composed of the approach d (which may be in any direction in the sn plane), a shift in x direction Sft_x, and a pivoting rotation Ω_s. The axis of the rotation Ω_s is assumed to be aligned with the direction of Δd, and its pole is assumed to be the origin o of the local contact x, s coordinates. The total rigid displacement differences or shifts in the tangential directions of the contact surface, $W_{I,x}$ and $W_{I,s}$, are computed according to Eqs. (4.23) and (4.24).

$$W_{I,x} = Sft_x - \Omega_s \cdot l_I \tag{4.23}$$

$$W_{I,s} = -\Delta d \cdot \sin(\Delta\alpha_{I-d}) + \Omega_s \cdot x_I \cdot \cos(\Delta\alpha_{I-d}) \tag{4.24}$$

l_I in Eq. (4.23) is the signed distance from the contact origin o to I, measured perpendicular to the direction of the approach as shown in Figure 4.12. It is calculated according to Eq. (4.25), where Y_I, Z_I, Y_o, Z_o are the coordinates of points I and o in the YZ plane, and α_d is the angle defining the direction of the approach measured from the vertical as shown in Figure 4.12. The quantities that are passed to the tangential contact solver are the increments of rigid displacement differences $\mathbf{W_I}$ between the previous and the current time step, as explained in §4.1.4.2.

$$l_I = (Y_I - Y_o)\cos(\alpha_d) + (Z_I - Z_o)\sin(\alpha_d) \tag{4.25}$$

4.1.1.3.1 Rolling radii for the calculation of rigid velocities

The rigid slip velocities are usually a small fraction of the rolling velocities in practical linear elastic rolling contact problems. The rigid velocities of the wheel are the superposition of its translation and rotation. For points on the contact surface and with no gross slip, the velocities due to the rotational motion have approximately the same magnitude as the translation velocity, and opposite direction. That is, the velocities due to wheel translation and rotation nearly cancel out each other. For given wheel translational and rotational velocities, small differences in the rolling radii may then lead to considerable variations in the resulting wheel rigid slip velocities. So it is necessary to define precisely the geometry for an accurate calculation of the rigid slip velocities. This is not specific of conformal contact; in fact, it is relevant mostly in planar rolling contact problems with low levels of rigid slip.

It has been stated that the rigid slip velocities are calculated at points on the PCS, and projected on the PCS. This is further elaborated here. Figure 4.13 shows the undeformed surface of a cylindrical wheel of radius R rolling on a plane. The wheel rolls along the positive X direction with translation velocity V and rotation velocity $\boldsymbol{\omega}$, with its rotation and revolution axis parallel to the Y direction. With no gross slip, $V \approx \omega R$. Considering a point P on the wheel axis of revolution, the relative velocity of a point I_W on the wheel surface with respect to P, $\mathbf{v_{IW/P}}$, is $\boldsymbol{\omega}\times\mathbf{r_{IW\text{-}P}}$. Its magnitude is ωR at any point on the wheel surface. On the other hand, its projection on the PCS changes with the cosine of the azimuthal coordinate of I_W, $\theta_{IW} = \theta_I + \varepsilon_I$. For non-conformal contact in

rolling direction, $x_{IW} - x_P << R$ for any point I_W inside the contact, θ_{IW} is small, and $\cos(\theta_{IW}) \approx \cos(\theta_I) \approx 1 - \theta_I^2 / 2 \approx 1 - ((x_{IW} - x_P) / R)^2 / 2 \approx 1$. The projected rigid slip velocity of I_W along x, $w_{IW,x}$, is equal to $V - \omega R \cos(\theta_{IW})$. For the particular case of $V = \omega R$ (i.e. with nominally no rigid slip), $w_{IW,x} = V (1 - \cos(\theta_{IW}))$. The difference $1 - \cos(\theta_{IW})$, though small, is not negligible with respect to creepages usually encountered in wheel–rail rolling contact problems. For example, for $x_{IW} - x_P = 10$ mm and $R = 460$ mm, $1 - \cos(\theta_{IW}) \approx 0.02\%$.

So if the rigid velocities of the wheel are calculated at points I_W of its undeformed surface and then projected on the PCS, there is a component of the rigid slip velocity changing in proportion to $\cos(\theta_{IW})$ along the contact. It is considered that this variation, which comes from the difference between the undeformed geometry of the wheel and the PCS on which the rigid velocities are projected, is not realistic and has to be corrected. This is done in CECT by computing the rigid velocities of the wheel at points I on the PCS instead of at points on the actual undeformed wheel surface. Referring to Figure 4.13, this means that the rigid slip velocities of the wheel are computed at points I instead of at points I_W. It is verified that $v_{I/P,x} = -\|\mathbf{v_{I/P}}\| \cos(\theta_I) = -\omega \|\mathbf{r_{I\text{-}P}}\| \cos(\theta_I) = -\omega (R / \cos(\theta_I)) \cos(\theta_I) = -\omega R$, constant along the PCS. The variation with $\cos(\theta_{IW})$ which appears if the calculation is carried out at point I_W is thus avoided. Also, when $V = \omega R$, $w_{I,x} = V + v_{I/P,x} = 0$, as expected.

In [Vollebregt 2021], the exposed unintended variation of the rigid velocity of the wheel along the contact is avoided in a different way, introducing a so-called “mid-plane correction”. This consists on considering a plane midway between the undeformed wheel and rail surfaces (and thus with variable inclination in the XZ plane along the contact), in place of the prismatic PCS, for projection of the rigid velocities.

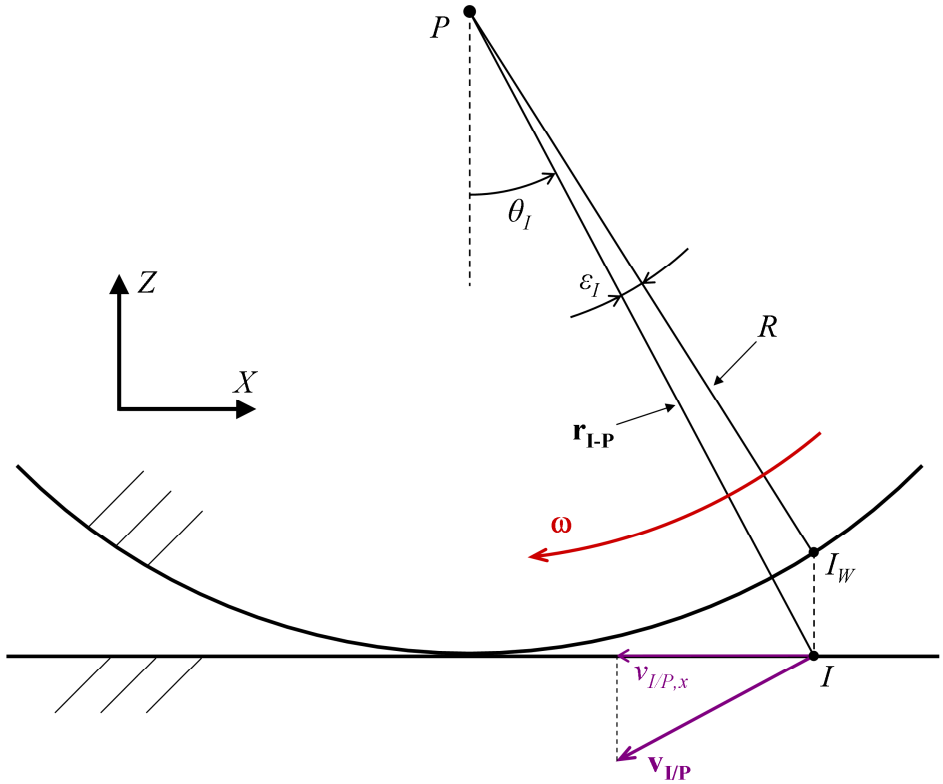


Figure 4.13. Calculation of wheel rigid velocities on the PCS as opposed to on the wheel surface.

Another question is which geometry should be considered for the calculation of the rigid velocities. Should it be the undeformed or the deformed geometry? Figure 4.14 illustrates the difference between the undeformed and the deformed rolling radii of a wheel rolling on a plane, $R^{(ud)}$ and $R^{(d)}$. The axis of revolution of the wheel is located at point P in the undeformed configuration, and moves to point P' in the deformed configuration. In elastic contact of steel wheel on steel rail, the difference between the undeformed and the deformed rolling radii of the wheel in the contact patch is of order 0.01%. This is not negligible with respect to the magnitude of the creepages in practical wheel–rail rolling contact problems.

Bearing this in mind, here it has been questioned whether the **w** term in Eq. (1.7) should be computed on the deformed rather than on the undeformed

configuration of the wheel. That is, considering a purely linear superposition of velocities between the rigid body motion of the solids and the local contact field, such as is done with the normal displacements in the normal part of the contact problem, might not be completely correct. Tests have been done with FE models to try to clarify this issue. These are described in §4.3.9.1. These tests have not proved that the deformed geometry should be used instead of the undeformed one. Currently the field of rigid velocities **w** is computed in CECT using the initial undeformed configuration of the contacting bodies.

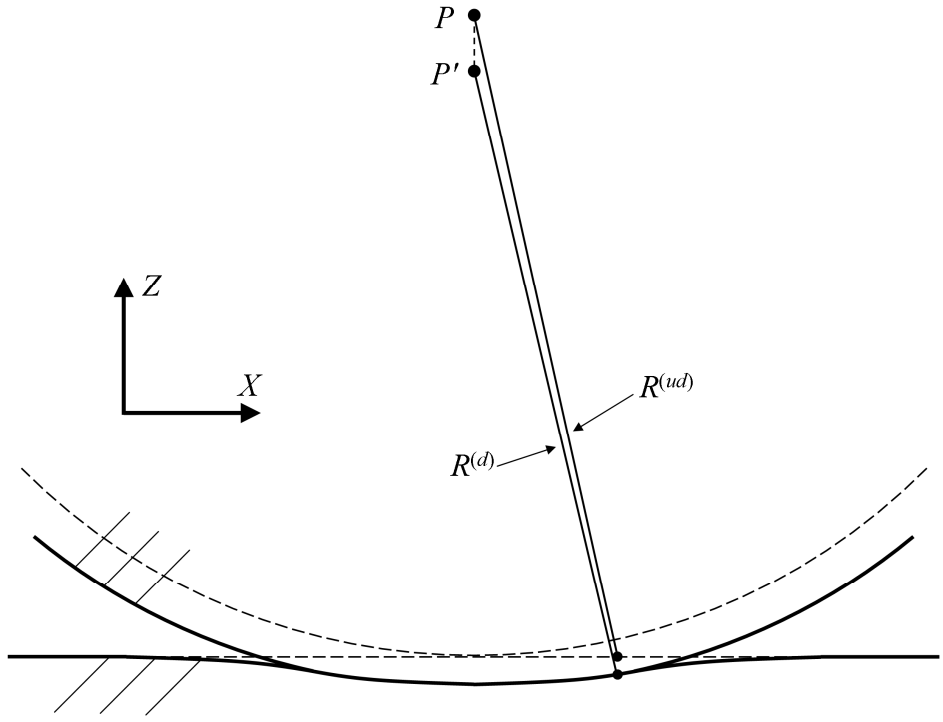


Figure 4.14. Rolling radii in the undeformed and deformed configurations.

4.1.1.4 Influence coefficients

The non-flatness of the contact surface has consequences in the mechanical behaviour of the contacting solids, which are no longer seen as half-spaces as in planar concentrated contact. This represents an important difficulty, as the ICs for solids with general non-planar surfaces are not available analytically, as there are those of the half-space.

The ICs for non-planar solids may be calculated numerically, with FE models for example. However, the process for this calculation is involved, and the validity of the numerically calculated ICs is limited to a particular geometry, a coefficient of Poisson and to a mesh of the PCS. In this thesis an analytical approximation of the ICs for non-planar solids is proposed, in order to avoid their numerical calculation. These topics are dealt with in Chapter 5. Here the structure of the IC matrices of non-planar solids is discussed and contrasted with the planar case (i.e. the half-space). Along this, implications in the scope of the required calculations to obtain the complete IC matrices, and in their convolution with the contact stresses to obtain the elastic displacements, are addressed.

As was seen in Chapter 2, the spatial discretization of the integral equation (1.6), which provides the constitutive relationship between elastic displacements and stresses in the contact surface, leads to a matrix equation in which the vector $\mathbf{\{u_i\}}$ of elastic displacements in all the mesh points is obtained as the product of the IC matrices $\mathbf{[IC_{ij}]}$ and the vectors $\mathbf{\{p_j\}}$ with the stresses in each principal direction in all the mesh points (Eq. (2.100)). In 3D problems, the subindices *i, j* of these vectors and matrices may have three possible values; one for each direction of the coordinate system defined for the problem. In planar contacts, $i, j = x, y,$ or z, and in non-planar contacts $i, j = x, s,$ or n. Therefore there will be three vectors $\mathbf{\{u_i\}}$ and $\mathbf{\{p_j\}}$, and 9 matrices $\mathbf{[IC_{ij}]}$.

Designating as *NE* the total number of APs in the mesh, the vectors $\mathbf{\{u_i\}}$ and $\mathbf{\{p_j\}}$ have *NE* elements, and $\mathbf{[IC_{ij}]}$ is a square matrix with dimension $NE \times NE$. Element (I, J) of this matrix is the displacement u_i obtained in the I^{th} AP of the mesh, due to a unit pressure in direction j applied in the J^{th} AP. Each complete $\mathbf{[IC_{ij}]}$ matrix has therefore NE^2 values, but they are not all independent, as explained next. In what follows, the $\mathbf{[IC_{ij}]}$ matrices of the current time instant t are considered. Afterwards additional considerations for the $[\mathbf{IC'_{ij}}]$ matrices of the previous time instant t' (cf. §2.1.3) are set out, which apply similarly in planar and in non-planar contact.

In the case of the half-space, the ICs between any two I, J APs depend on the relative position between those two APs in each direction of the surface, but not

on the absolute position of each of the points, as explained in §2.1.3. Consequently, only *NE* values of each complete $\mathbf{[IC_{ij}]}$ matrix are independent[9], and the rest are equal to one of these, with the same or with changed sign, due to similarity and (anti)symmetry. By way of example, the ICs between elements[10] I_0 and J_0 of Figure 4.15a are equal to the ICs between elements I_1 and J_1, as the relative position between each of these pairs of elements is the same. On the other hand, referring to the ICs for the current time instant *t*, the ICs between elements I_1 and J_2 have the same absolute value as the ICs between elements I_0 and J_0, and will have the same or the contrary sign depending on the particular IC being symmetric or antisymmetric in the longitudinal direction.

The mentioned *NE* necessary values for each IC matrix may be obtained by means of a single load case, observing the resulting displacements in the *NE* elements of the mesh when the element in one of the corners of the mesh is loaded. That is, the loaded element is one of the four elements highlighted in red in Figure 4.15a, so that between that element and the rest of the elements of the mesh there are all the possible combinations of relative positions in both mesh directions that may occur between any pair of elements of the mesh[11]. The necessary values for the nine $\mathbf{[IC_{ij}]}$ matrices may be obtained in a similar way, observing the displacements of the *NE* elements of the mesh in each of the 3 space directions under 3 different load cases, with the load applied in a different direction in each of them. Taking into account the reciprocity relations $IC_{IiJj} = IC_{JjIi}$, it is not necessary to store the 9×*NE* values thus obtained, but only 6×*NE*.

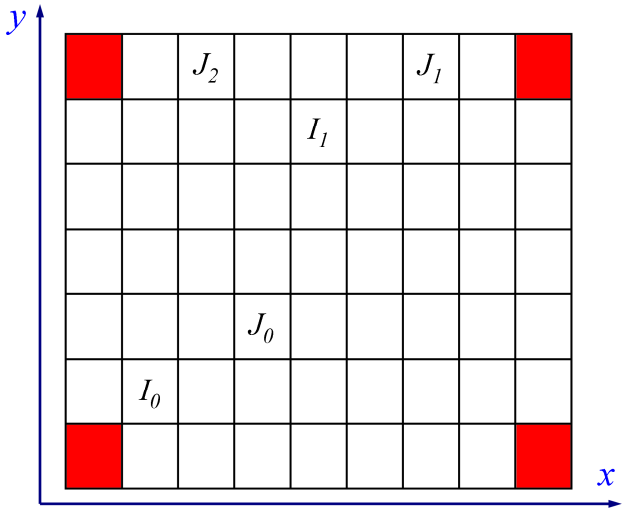
---

[9] Roughly, without discounting the zero values, which are n_y in $\mathbf{[IC_{xz}]}$, n_x in $\mathbf{[IC_{yz}]}$ and n_x+n_y-1 in $\mathbf{[IC_{xy}]}$; idem in their reciprocals.

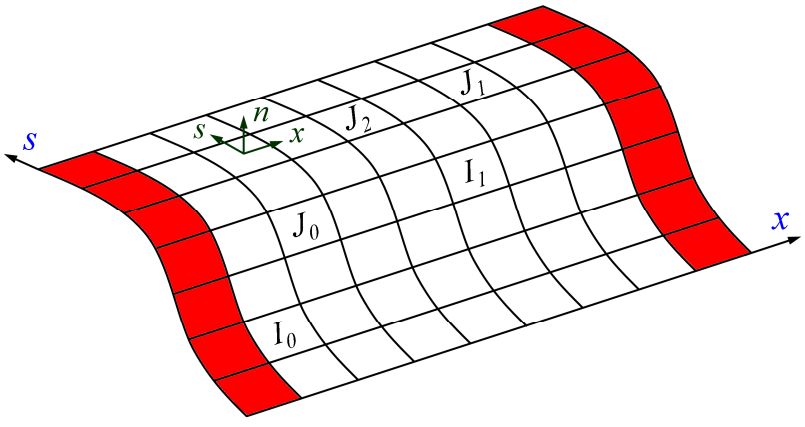
[10] "Element" is being used in place of "analysis point" here. In Figure 4.15 uniform elements are assumed, without loss of generality.

[11] If the calculation of the ICs is carried out analytically, in such a way that it costs the same evaluating the displacements in the *NE* elements of the mesh in a single load case, or evaluating the displacements in a single element in *NE* load cases, alternatively the displacements of one of the four corner elements may be evaluated loading each time a different element of the mesh in *NE* different load cases, to obtain the *NE* necessary ICs for a matrix.

This may be reduced to 4×*NE* in meshes with square elements (i.e. with $\Delta x = \Delta y$), storing for example only the IC_{xx}, IC_{zz}, IC_{xy} and IC_{xz} ICs.



(a) Plane contact surface.



(b) Curved contact surface.

Figure 4.15. Similarities of ICs in regular meshes, and possible load cases to compute the necessary values for each **[IC$_{ij}$]** matrix.

The *NE* independent values of each **[IC$_{ij}$]** matrix may be stored in a vector `v_IC`, with element *e* of the array being equal to $IC_{(e)i(1)j}$. *e* is the 1D index of the AP in the 2D grid of the PCS, as illustrated in Figure 2.19. This numbering

is applied similarly in curved contact surfaces. A possible generic code to construct the complete *NE*×*NE* **[IC$_{ij}$]** matrix from `v_IC` is shown in Figure 4.16. The **[IC$_{ij}$]** matrix is stored in the `ICij` variable in this code. The exponents *fx*, *fy* depend on the (anti)symmetry of the IC_{ij} in question along directions *x* and *y* with respect to the loaded AP, as given in Table 4.1. *fj* = 2 if the IC_{ij} is symmetric, and *fj* = 1 if the IC_{ij} is anti-symmetric along direction *j* (*j* = *x* or *y*). It has to be noted that the size of each of these matrices is of $O(n_j^4)$; *j* = *x* or *y*, assuming $n_x \approx n_y$. The construction of these complete **[IC$_{ij}$]** matrices may be avoided as explained in §4.1.1.4.2.

```
NE = nx * ny;
sz = [nx ny];
v_ei = 1:NE;
[sub_x_ei, sub_y_ei] = ind2sub(sz, v_ei');
ICij = zeros(NE, NE);
for ej=1:NE
   [sub_x_ej, sub_y_ej] = ind2sub(sz, ej);
   dif_sub_x = sub_x_ei - sub_x_ej;
   sgn_dx = sign(dif_sub_x);
   sgn_dx(abs(sgn_dx)<0.1) = 1;
   dif_sub_y = sub_y_ei - sub_y_ej;
   sgn_dy = sign(dif_sub_y);
   sgn_dy(abs(sgn_dy)<0.1) = 1;
   inds = sub2ind(sz, abs(dif_sub_x)+1, ...
                      abs(dif_sub_y)+1);
   ICij(:, ej) = v_IC(inds) .* sgn_dx.^fx .* ...
                 sgn_dy.^fy;
end % for ej=1:NE
```

Figure 4.16. Possible MATLAB code to construct a complete half-space [IC$_{ij}$] matrix.

Table 4.1. Factors *fx* and *fy* for each IC_{ij}.

IC_	***fx***	***fy***
xx, *yy*, *zz*	2	2
xy, *yx*	1	1
xz, *zx*	1	2
yz, *zy*	2	1

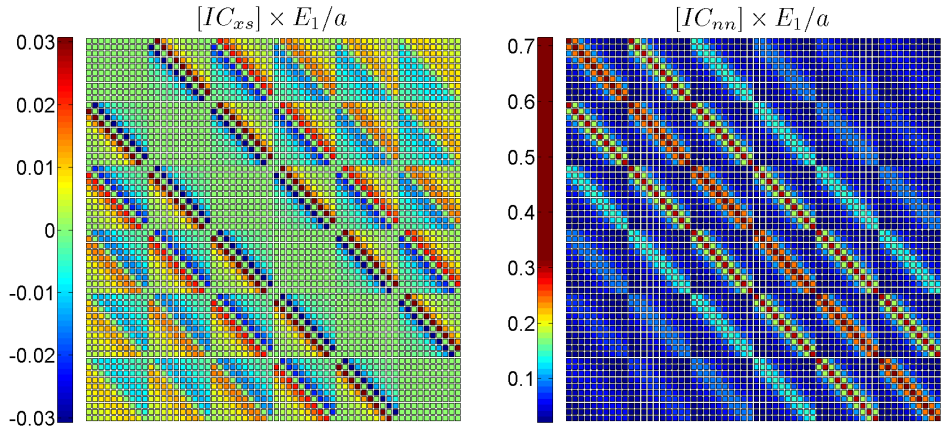
Due to the similarities in x and y directions of the mesh, the $\mathbf{[IC_{ij}]}$ matrix is Block Toeplitz[12] with Toeplitz Blocks (BTTB), with n_y independent Toeplitz blocks, each of which has n_x independent values, according to the numbering of APs adopted. Figure 4.17 shows the structure of this matrix. Each of the represented cells is a Toeplitz matrix of dimension $n_x \times n_x$, which is symmetric for the ICs which are symmetric in x, and anti-symmetric with zeros in its main diagonal for the ICs which are anti-symmetric in x. The whole cells in the main diagonal of the $\mathbf{[IC_{ij}]}$ matrix are also zero for the ICs which are anti-symmetric in y. Each cell is related to a pair of rows of APs in the mesh, and is numbered with two *iy_I*, *iy_J* indices, indicating the lateral positions of each of these rows in the mesh. The first one corresponds to the row of APs where the displacements are observed, and the second one to the row of loaded APs. The factor ky is 1 for ICs which are symmetric in y, and –1 for ICs which are anti-symmetric in y.

(1, 1)	$ky \times$ (2, 1)	$ky \times$ (3, 1)	…	$ky \times$ (n_y–1, 1)	$ky \times$ (n_y, 1)
(2, 1)	(1, 1)	$ky \times$ (2, 1)	…	$ky \times$ (n_y–2, 1)	$ky \times$ (n_y–1, 1)
(3, 1)	(2, 1)	(1, 1)	…	$ky \times$ (n_y–3, 1)	$ky \times$ (n_y–2, 1)
…	…	…	…	…	…
(n_y–1, 1)	(n_y–2, 1)	(n_y–3, 1)	…	(1, 1)	$ky \times$ (2, 1)
(n_y, 1)	(n_y–1, 1)	(n_y–2, 1)	…	(2, 1)	(1, 1)

Figure 4.17. BTTB structure of a half-space $[IC_{ij}]$ matrix.

[12] A $N \times N$ matrix **[T]** is Toeplitz when all the elements in each of its (upper-left to lower-right) diagonals are equal; i.e. when each element of the matrix $T(i, j) = h_{i-j}$, for all $i, j \in \{1, \ldots, N\}$, with a given set of $2N - 1$ scalars $\{h_{-N+1}, \ldots, h_0, \ldots, h_{N-1}\}$.

Figure 4.18 shows example half-space **[IC$_{xs}$]** and **[IC$_{nn}$]** matrices for a mesh with $n_x \times n_s$ = 10×6 APs, with bilinear elements, and with a ratio of lateral to longitudinal dimensions of the elements, b/a, of 0.70. These have been computed for a half-space with a coefficient of Poisson v of 0.30. The data in the figure are normalized with the Young's modulus E_1 of the half-space and with the longitudinal dimension of the loaded element a. More details about the dimension of the ICs are given in Chapter 5, and in Figure 5.2 and subsequent figures similar magnitudes are plotted.



(a) IC_{xs}. (b) IC_{nn}.

Figure 4.18. Example half-space **[IC$_{ij}$]** matrices of a mesh with 10×6 APs.

Figure 4.19 shows an alternative code to build each complete **[IC$_{ij}$]** matrix, which exploits the BTTB structure of the matrix. This code is found to be about 8-10 times faster than the code in Figure 4.16 on a 60×60 mesh, with the factor increasing for smaller meshes. Once built, these matrices are reused many times in the solution process of the contact problem, and the time required to build one of them (with the less efficient algorithm shown) is usually small in relation to the total time for the solution of a contact problem. So, the benefit of this efficiency gain in the whole contact solution process may be limited, depending among other things on the number of complete matrices to be built.

```
ICij = zeros(NE, NE);
if fy>1.5, ky=1; else ky=-1; end
for iy=1:ny
   col = v_IC(1+(iy-1)*nx:iy*nx);
   if fx>1.5
      toep = toeplitz(col, col);
   else
      row = -col; row(1) = col(1);
      toep = toeplitz(col, row);
   end
   for i=1:ny-iy+1
      c = 1+(i-1)*nx;
      r = c + (iy-1)*nx;
      ICij(r:r+nx-1, c:c+nx-1) = toep;
      if iy>1.1
         ICij(c:c+nx-1, r:r+nx-1) = ky * toep;
      end
   end
end % for iy=1:ny
```

Figure 4.19. MATLAB code to construct a complete half-space **[IC_{ij}]** matrix exploiting its BTTB structure.

In the case of conformal contact, the similarity related to the relative position of a pair of elements in the lateral direction is lost. Therefore, the ICs between elements I_0 and J_0 of Figure 4.15b are no longer equal to the ICs between elements I_1 and J_1, despite both pairs of elements have the same relative position in the mesh in its both principal directions (x and s). That is, the ICs between two elements will depend in general not only on their relative positions in the mesh, but also on their absolute positions in lateral direction. On the other hand, the ICs between elements I_1 and J_2 continue being the same in absolute value as the ICs between elements I_1 and J_1 (having the same or the contrary sign depending on whether the particular IC is symmetric or antisymmetric in longitudinal direction), as in the plane case, because in this case the lateral positions in the mesh of both pairs of elements does not change, and the similarity in the longitudinal direction is maintained[13]. In the case of contact

[13] In this thesis, contacts between bodies which are either prismatic or of revolution are studied (or at least which may be approached as such around the contact). Considering

surfaces with circular profile, it may be assumed as a first approximation that there is also similarity in the lateral direction as in the plane case, according to the surface orientation based approximation of ICs for non-planar solids exposed in §5.2.3.1. The most general case of no lateral similarity is assumed in what follows.

According to the previous discussion, the necessary data for each $\mathbf{[IC_{ij}]}$ matrix may no longer be obtained with a single load case as in the plane case, but in general n_s load cases will be necessary (n_s = 7 in Figure 4.15b). These load cases are defined loading each time an element located in each of the n_s lateral positions of the mesh, and in one of the extreme positions in longitudinal direction; see the elements marked in red in Figure 4.15b. To obtain the necessary data for the 9 IC matrices of each of the contacting bodies, $3\times n_s$ load cases are necessary, as indicated in [Li Z 2002] (one for each loading direction and lateral position of the mesh). As these calculations have to be carried out for each of the two contacting bodies, a total of $6\times n_s$ load cases are necessary to obtain the IC matrices of the displacement differences in the contact surface. The number of lateral positions n_s to be considered here can be reduced in cases where symmetries or similarities can be used in the lateral direction.

---

that the longitudinal direction is aligned with the axis of the prismatic body or with the circumferential direction of the body of revolution, there is similarity in the IFs in longitudinal direction for each of the bodies separately. But if there is a yaw angle between both contacting bodies, so that their respective longitudinal directions are no longer aligned, strictly speaking this similarity would be lost in the combined ICs of both solids (the ones related to the displacement differences of both in the contact surface), associated to the common mesh of the PCS. However, given the small magnitude of the yaw angle in practical wheel–rail contact problems (on the order of tens of mrad at the highest), it may be assumed with little error that there is similarity of the combined ICs in longitudinal direction regardless of this angle. Besides, in order for a conformal contact to be actually feasible between prismatic bodies and/or bodies of revolution, with a curved surface in lateral direction, in linear elasticity problems where the displacements are small, the yaw angle must necessarily be small. With big yaw angles, the possible contacts are concentrated, and in this case it applies again the similarity of the ICs in both directions of the (plane) contact surface.

The number of independent values in each **[IC$_{ij}$]** matrix is $NE \times n_s = n_x \times n_s^2$ in the case of the crossed IC_{ij}-s with $i \neq j$; $i, j \in \{x, s, n\}$. In the case of the 3 direct IC_{ii}-s this is reduced to $n_x \times n_s \times (n_s+1)/2$ on account of reciprocity. Additionally, only 3 out of the 6 sets of crossed IC_{ij}-s need to be calculated, being able to deduce the remaining ones by reciprocity as in the plane case. This makes a total of $3 \times n_x \times n_s \times (3n_s+1)/2$ independent values of ICs, in place of the $6 \times n_x \times n_s$ of the plane case.

The independent values of each **[IC$_{ij}$]** matrix may be stored in an array `a_IC`, with element (*e*, *is*) of the array being equal to $IC_{(e)i(e1\_is)j}$, where *e1_is* is the 1D index of the first AP in the *is*th lateral position of the mesh, i.e. $e1\_is = 1 + n_x \times (is - 1)$. The chosen loaded APs here are those with indices *e1_is*. All `a_IC` arrays are defined with the same size of $NE \times n_s$ for convenience, so $e = 1, \ldots, NE$; $is = 1, \ldots, n_s$, in all these arrays. In the current implementation of CECT, arrays `a_IC` may be either computed numerically (outside CECT), cf. §5.2.2, or approximated analytically within CECT, cf. §5.2.3. In the latter case, previously calculated vectors `v_IC` with ICs of the half-space are used, each with *NE* values.

The resulting **[IC$_{ij}$]** matrix is no longer BTTB, as a result of the lack of similarity in the lateral direction. Each of the $n_s \times n_s$ blocks of the matrices of the crossed IC_{ij}-s (each of them, with dimension $n_x \times n_x$, corresponding to a pair of lateral positions in the mesh with indexes *is_I*, *is_J*) is different in general. On the other hand, the matrices of the IC_{ii} direct ICs have a symmetric block structure due to reciprocity between different rows of the mesh, i.e. block (*is_I*, *is_J*) is the same as block (*is_J*, *is_I*), for all $is\_I, is\_J \in \{1, \ldots, n_s\}$. According to the surface orientation based approximation of ICs exposed in §5.2.3, in a general case it could be assumed as a first approximation for all the **[IC$_{ij}$]** matrices that all the (*is*, *is*) blocks in the main diagonal are equal. Regarding the structure of each $n_x \times n_x$ block, as there is similarity in the longitudinal direction as in the planar case, each of the blocks of all the **[IC$_{ij}$]** matrices is also a Toeplitz matrix, with n_x independent values and symmetric or anti-symmetric.

Figure 4.20 shows a code to build a **[IC$_{ij}$]** matrix for a curved contact surface from the $NE \times n_s$ `a_IC` array, taking into account its characteristics described above. The `sym` variable in this code is set to 1 for matrices of the direct IC_{ii}-s,

with symmetric block structure. Some rough comparative values of the times required to build these matrices are given next, based on a mesh with 60×60 APs. With smaller meshes, it is observed that the relative differences between the different times increase. In the case of the block-symmetric **[IC$_{ii}$]** matrices, the time taken by the algorithm in Figure 4.20 is approx. twice as much as the corresponding algorithm shown in Figure 4.19 for the planar case. In this case the algorithm in Figure 4.20 is around 3 times faster than the less efficient code (not shown here), similar to that shown Figure 4.16 for the planar case. In the case of the non block-symmetric **[IC$_{ij}$]** matrices (i.e. with `sym` = 0) this advantage is reduced to a factor of 2. The less efficient code is somewhat faster (around 10-20%) working with the `a_IC` array than the code in Figure 4.16 working with the `v_IC` vector of *NE* values.

```
 1 ICij = zeros(NE, NE);
 2 if fx>1.5, kx=1; else kx=-1; end
 3 nsc=ns;
 4 for isr=1:ns
 5    r = 1+(isr-1)*nx;
 6    if sym, nsc=isr; end
 7    for isc=1:nsc
 8       c = 1+(isc-1)*nx;
 9       col = a_IC(1+(isr-1)*nx:isr*nx, isc);
10       row = kx*col; row(1) = col(1);
11       toep = toeplitz(col, row);
12       ICij(r:r+nx-1, c:c+nx-1) = toep;
13       if sym && isr-0.1>isc
14          ICij(c:c+nx-1, r:r+nx-1) = toep;
15       end
16    end % for isc=1:nsc
17 end % for isr=1:ns
```

Figure 4.20. MATLAB code to construct a complete **[IC$_{ij}$]** matrix for a curved contact surface taking into account that it is composed of Toeplitz blocks.

4.1.1.4.1 IC matrices of the previous time instant

Up to now, the matrices **[IC$_{ij}$]** for the current time instant *t* have been considered. In rolling contact problems, **[IC'$_{ij}$]** matrices for the previous time instant *t'* have to be computed as well. In general, a higher number of independent values of ICs are necessary to construct these matrices than the

[IC$_{ij}$] matrices, because more longitudinal position differences ($x_I - x_J$) between pairs of elements *I*, *J* in the mesh have to be considered.

For the **[IC$_{ij}$]** matrices, the number of differences ($x_I - x_J$) to be considered is equal to the number of elements in the longitudinal direction n_x. A possible set of these is ($x_{ix} - x_1$) = ($ix - 1$)×Δx, with ix = 1, 2, …, n_x; x_{ix} being the longitudinal coordinate of the ix^{th} longitudinal position of the discretization, and Δx the longitudinal dimension of the elements of the mesh. The other $n_x - 1$ possible differences with reversed sign ($x_1 - x_{ix}$) are omitted here, because the ICs for them may be deduced from the ones for the differences ($x_{ix} - x_1$) by (anti)symmetry, as indicated before.

For the **[IC'$_{ij}$]** matrices, the discretization parameter Δq, cf. Figure 2.20 in §2.1.3, takes part in the longitudinal position differences. For any two elements *I*, *J*, being *J* the loaded element, the difference to be considered is ($x_I + \Delta q - x_J$). As a result, there are up to $2 \times n_x - 1$ different values of $|x_I + \Delta q - x_J| = |(ix\_I - ix\_J) \times \Delta x + \Delta q|$ (with ix_I, $ix\_J \in \{1, 2, \ldots, n_x\}$), in place of the n_x different values for the **[IC$_{ij}$]** matrices. Consequently, the number of independent values of ICs in these matrices is increased by a maximum factor of $(2 \times n_x - 1) / n_x$. For the most general case, all the necessary longitudinal position differences are covered with the sets ($x_I + \Delta q - x_1$) and ($x_I + \Delta q - x_{nx}$), i.e. with $ix\_I \in \{1, 2, \ldots, n_x\}$ and $ix\_J \in \{1, n_x\}$. If Δq is a multiple of $\Delta x/2$ special cases arise. If $\Delta q = \Delta x/2$, the number of different $|x_I + \Delta q - x_J|$ values in the mesh is reduced to n_x, and they are covered with the set ($x_I + \Delta q - x_1$), i.e. with $ix\_I \in \{1, 2, \ldots, n_x\}$ and $ix\_J = 1$, as for the **[IC$_{ij}$]** matrices. If $\Delta q = \Delta x$, the number of different $|x_I + \Delta q - x_J|$ values is $n_x + 1$. Also in this case a similar set of data as for the **[IC$_{ij}$]** matrices may suffice for the **[IC'$_{ij}$]** matrices, provided a complete row of elements is left out of the contact either at the trailing or at the leading edge of the mesh.

The **[IC'$_{ij}$]** matrices have the same block structure as with $\Delta q = 0$, in the planar as well as in the non-planar case. As a difference, now each block is a Toeplitz matrix with $2n_x - 1$ independent values in the most general case. With $\Delta q = \Delta x$ the diagonals are shifted upwards one position, so that e.g. for the direct IC_{ii}-s the dominant diagonal is the first one (above the main), instead of the main.

A complete **[IC'$_{ij}$]** matrix for a non-planar contact surface may be constructed with a code similar to that in Figure 4.20, changing the definition of the `col` and `row` variables to form each Toeplitz block as shown in the code of Figure 4.21. Thus, lines 9 and 10 of the code in Figure 4.20 are substituted with the code in Figure 4.21. The *if–else* block of the code in this figure covers the cases of $\Delta q \neq 0$ for the **[IC'$_{ij}$]** matrices (line 11), and that of $\Delta q = 0$ for the **[IC$_{ij}$]** matrices (line 13). The arrays `a_IC_a` and `a_IC_b` used in this code are similar to the array `a_IC` introduced before for the **[IC$_{ij}$]** matrices. Each of these arrays have $NE \times n_s$ elements, and contain the ICs associated to x_J positions in the first and last longitudinal positions of the mesh respectively. Element (*e*, *is*) of array `a_IC_a` is equal to $IC'_{(e)i(e1\_is)j}$, i.e. it gives the influence of the first AP of the *is*th lateral position of the mesh (the position indices of which are (*ix*, *is*) = (1, *is*)) over the position occupied by the *e*th AP at the previous time instant *t'*. Element (*e*, *is*) of array `a_IC_b` is equal to $IC'_{(e)i(enx\_is)j}$, i.e. it gives the influence of the last AP of the *is*th lateral position of the mesh (the position indices of which are (*ix*, *is*) = (n_x, *is*)) over the position occupied by the *e*th AP at the previous time instant *t'*.

```
 9 col = a_IC_a(1+(isr-1)*nx:isr*nx, isc);
10 if dq
11    row = a_IC_b(isr*nx:-1:1+(isr-1)*nx, isc);
12 else
13    row = kx*col;
14 end
15 row(1) = col(1);
```

Figure 4.21. MATLAB code to define the elements of each Toeplitz block of a complete **[IC'$_{ij}$]** matrix for a curved contact surface. The complete **[IC'$_{ij}$]** matrix is built up with such Toeplitz blocks according to the code of Figure 4.20.

If the ICs are approached analytically as explained in §5.2.3, arrays `a_IC_a` and `a_IC_b` may be formed using vectors `v_IC_a` and `v_IC_b` in the most general case. These are similar to the vector `v_IC` introduced before for the **[IC$_{ij}$]** matrices, each with the *NE* half-space ICs over the positions occupied at *t'* by all APs. `v_IC_a` has the ICs of the first AP of the mesh over all APs, and `v_IC_b` has the ICs of the last AP of the mesh over all APs.

For $\Delta q = i \times \Delta x$ with integer i, arrays `a_IC_a` and `a_IC_b` may be formed from array `a_IC` with the ICs for the current time instant t as shown in the code of Figure 4.22, reducing the number of longitudinal positions of the mesh in i. That is, if array `a_IC` contains the ICs for a mesh with n_x longitudinal positions, arrays `a_IC_a` and `a_IC_b` that may be formed from it contain the ICs for a mesh with $(n_x - i)$ longitudinal positions at most.

```
rq = dq/dx;
i_rq = round(rq);
ai_rq = abs(i_rq);
c_IC = reshape(a_IC, nx, ns, ns);
nx = nx - ai_rq;
NE = nx * ns;
if dq > 0
   a_IC_a = reshape(c_IC(1+ai_rq:end, :, :), NE, ns);
   c_IC_b = zeros(nx, ns, ns);
   c_IC_b(end-ai_rq:end, :, :) = ...
               c_IC(1:ai_rq+1, :, :);
   if ai_rq < nx-1.1
      c_IC_b(end-ai_rq-1:-1:1, :, :) = ...
        c_IC(2:nx-ai_rq, :, :) * kx;
   end
   a_IC_b = reshape(c_IC_b, NE, ns);
elseif dq < 0
   a_IC_b = kx * ...
         reshape(c_IC(end:-1:1+ai_rq, :, :), NE, ns);
   c_IC_a = zeros(nx, ns, ns);
   c_IC_a(1:1+ai_rq, :, :) = ...
      c_IC(1+ai_rq:-1:1, :, :) * kx;
   if ai_rq < nx-1.1
      c_IC_a(2+ai_rq:end, :, :) = ...
           c_IC(2:nx-ai_rq, :, :);
   end
   a_IC_a = reshape(c_IC_a, NE, ns);
end
```

Figure 4.22. MATLAB code to form `a_IC_a` and `a_IC_b` arrays for code in Figure 4.21 from array `a_IC` with ICs for current time instant, for the case that Δq is a multiple of Δx.

Both the cases of $\Delta q > 0$ and $\Delta q < 0$ are considered in the code of Figure 4.22. The constant `kx` used here is 1 for IC symmetric in x and –1 for ICs anti-

symmetric in *x*, as defined in line 2 of the code of Figure 4.20 from the value *fx* of the particular IC considered (cf. Table 4.1).

4.1.1.4.2 Convolution with contact stresses

Surface elastic displacements have to be computed many times during the solution of contact problems with the exact contact theory. This may be done directly in the space domain following Eq. (2.100), multiplying the complete (*NE*×*NE*) $\mathbf{[IC_{ij}]}$ matrices with the vectors of contact stresses $\mathbf{\{p_j\}}$. But making use of the similarities of ICs in a regular mesh, it may be done more efficiently in the Fourier domain, using fast Fourier transforms (FFTs), and without building the complete $\mathbf{[IC_{ij}]}$ matrices, as explained here. FFT algorithms are widely described in the literature, see e.g. [Brigham 1988], [Chu 2000] and §12 of [Press 2007]. Multilevel methods are also used to perform efficiently this type of matrix multiplications. These are not covered here. [Brandt 1990], [Polonsky 1999], [Polonsky 2000], [Venner 2000] and [Sainsot 2011] are some references on the subject.

The discrete convolution theorem states that given two discrete signals **{r}** and **{s}**, one of them periodic with period *N*, and the other of finite duration *N*, the discrete Fourier transform (DFT) of their cyclic convolution **{r}**⊗**{s}** is equal to the element-wise product of the DFT of each of them. This is expressed in Eq. (4.26). The signal **{r}** may be identified as the response or echo function, and the signal **{s}** as the excitation or input function. r_k is the effect produced on channel $J+k$ by the unit input of channel J, s_J.

$$\mathrm{DFT}\left(\{\mathbf{r}\}\otimes\{\mathbf{s}\}\right)=\mathrm{DFT}\left(\{\mathbf{r}\}\right)\cdot\mathrm{DFT}\left(\{\mathbf{s}\}\right) \tag{4.26}$$

The cyclic convolution of the two described discrete signals is defined in Eq. (4.27), where N_1 and N are integers. The I subindex in the equation refers to the point for which the output of the convolution (surface displacements in the case of contact analyses) is calculated. Due to the periodicity of **{s}**, $s_k = s_{k+iN}$ for any integer i. Convolutions are commutative, so that **{r}**⊗**{s}** = **{s}**⊗**{r}**. The cost of computing the convolution according to Eq. (4.27) in N points I (e.g. $I = 1, \ldots, N$) is $O(N^2)$. The cost of computing it in the Fourier domain using the discrete convolution theorem (Eq. (4.26)) is reduced to $O(N \log N)$, including the cost of

the transformation and inverse transformation operations, thanks to the efficiency of FFT algorithms.

$$y_I = \left(\{\mathbf{r}\} \otimes \{\mathbf{s}\}\right)_I = \sum_{k=-N_1}^{N-N_1-1} r_k \cdot s_{I-k} \tag{4.27}$$

Cyclic convolutions arise in the analysis of domains with circular geometry, e.g. bolted flanges and rolling bearings. But this is not the case of the linear convolutions appearing in concentrated contact problems as that expressed by Eq. (2.100). In contact problems, **{r}** is identified as the ICs, and **{s}** as the contact surface stresses. **{s}** is not periodic, but extends over a finite duration *NE*, according to the chosen discretization of the PCS with *NE* elements. Still, the discrete convolution theorem may be applied to perform the convolution expressed in Eq. (2.100) in the Fourier domain exactly, with a proper arrangement of **{r}** and **{s}**. This is explained e.g. in §13.1 of [Press 2007] for 1D signals, and Liu [Liu 2000] illustrates the application of this process to 2D integrals of ICs with contact surface stresses in concentrated contact problems.

This process is demonstrated now for a 1D (line) contact computed on a PCS meshed with *N* APs, which are numbered in order from 1 to *N*, located in an *x* axis. Note that here 1 is chosen as the starting index, in order to keep consistency with previously shown equations and code excerpts. In contrast, 0 is often chosen as the starting index in the related literature. The starting point are the vectors **{IC}** and **{p}** given in space domain, containing each one the necessary values of ICs and contact stresses in a given direction (subindices *x*, *s*, *n* indicating directions are omitted for clarity). p_k is the contact stress in AP *k*. IC_k is the influence of the load on AP in position *J* over the AP in position $J + k$. In a 1D mesh with *N* APs, *k* ranges from $-N + 1$ to $N - 1$ to complete the vector **{IC}**. Therefore **{IC}** has $2N - 1$ values, IC_{-N+1}, …, IC_0, …, IC_{N-1}, corresponding to the $2N - 1$ different relative distances (sign included) between APs in the mesh. Not always all of these ICs are independent: as discussed previously, in the most general case of ICs for the previous time instant *t'*, the $2N - 1$ ICs are all independent; on the other hand, in the case of ICs for the current time instant *t*, there are just *N* independent values, and the remaining $N - 1$ are symmetric/anti-symmetric, among other cases.

Firstly, vectors **{IC}** and **{p}** are arranged according to the following steps:

- Extend **{p}** from N to M elements, with $M \geq 2N - 1$, filling with zeros the new elements of **{p}** (zero padding). That is, set $p_k = 0$ for $k = N + 1, \ldots, M$ in the extended vector.

  In choosing the new dimension M, it is advisable to round up the minimum necessary $2N - 1$ to the nearest integer power of 2, for greater efficiency in subsequent FFT operations.

- Arrange **{IC}** in wraparound order, and extend it if necessary to M elements. This means values $IC_0, \ldots, IC_{N-1}$ are placed in the first N positions of the vector, and values $IC_{-N+1}, \ldots, IC_{-1}$ in positions $M - N + 2, \ldots, M$ (in the stated order). Any additional elements between the first N and the last $N - 1$ are left as zero.

  In the case of the ICs for the current time instant t, the (anti)symmetry relations $IC_{M-k+1} = k_x\, IC_{k+1}$ may be applied for $k = 1, \ldots, N - 1$. Here $M-k+1$ and $k+1$ subindices denote position indices in the vector, starting from 1. The constant k_x has the same meaning as `kx` in the codes of Figure 4.20 and Figure 4.21, i.e. it is 1 for ICs which are symmetric and –1 for ICs which are anti-symmetric in the considered axis.

The described arrangement of vectors **{IC}** and **{p}** is illustrated in Figure 4.23. In the figure, the particular cases of symmetric/antisymmetric ICs are represented, but this is not necessary to apply the discrete convolution theorem. It is neither necessary to have symmetry in the duration of the response function **{IC}**, with N elements at the beginning of the vector and $N - 1$ at the end, though this happens to be the case in the problem at hand.

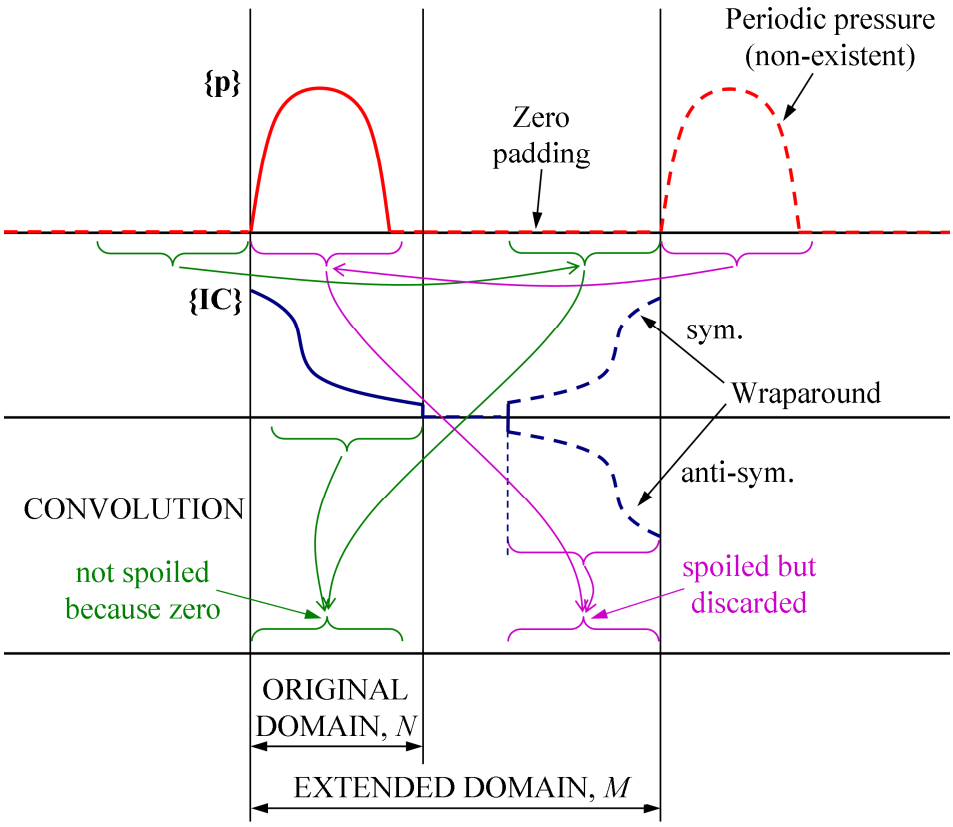


Figure 4.23. Extension and zero padding of vectors **{IC}** and **{p}** in a line contact analysis, so that the cyclic convolution in the extended domain is equivalent to the linear convolution in the original domain. Adapted from Figure 13.1.4 of [Press 2007].

The cyclic convolution of **{IC}** and **{p}** in the extended domain of length M may be written according to Eq. (4.28), where H is the Heaviside unit step function, $H(x) = 0$ for $x \le 0$ and $H(x) = 1$ for $x > 0$. It may be easily verified that the portion of this cyclic convolution located within the original domain of length N is equal to the desired linear convolution in the original domain, expressed by Eq. (4.29).

$$\left(\{\mathbf{IC}\}\otimes\{\mathbf{p}\}\right)_I = \sum_{k=1}^{M} IC_{I-k+(M+1)H(k-I)} \cdot p_k, \quad I \in \{1, 2, \ldots, M\} \tag{4.28}$$

$$\left(\{\mathbf{IC}\} * \{\mathbf{p}\}\right)_I = \sum_{k=1}^{N} IC_{I-k} \cdot p_k, \quad I \in \{1, 2, \ldots, N\} \tag{4.29}$$

Firstly, the summation in the right hand side of Eq. (4.28) may be restricted to $k = 1, \ldots, N$, since $p_k = 0$ for $k = N + 1, \ldots, M$ as a result of zero padding. Secondly, the element in the position $I - k + (M+1)\, H(k - I)$ of the extended vector **{IC}** in Eq. (4.28) is equal to IC_{I-k} for $I, k = 1, \ldots, N$; $I - k = -N + 1, \ldots, N - 1$, as a result of the wraparound arrangement of the extended vector **{IC}**. The subindexed *IC* element in the right hand side of Eq. (4.28) refers to the element of the extended vector **{IC}** located in the position indicated by the index. Conversely, the subindex of the *IC* element in the right hand side of Eq. (4.29) is not referred to a position in a vector of ICs, but rather identifies an IC itself, IC_{I-k} denoting the effect of the load on AP k over AP I. Note that the cyclic convolution of Eq. (4.28) is done with the same extended **{IC}** vector for all I elements, while the linear convolution of Eq. (4.29) is done for a different set of ICs for each I element. E.g., given $N = 10$, the set of ICs used for $I = 2$ is $\{IC_1, IC_0, \ldots, IC_{-8}\}$; and for $I = 6$ is $\{IC_5, IC_4, \ldots, IC_{-4}\}$.

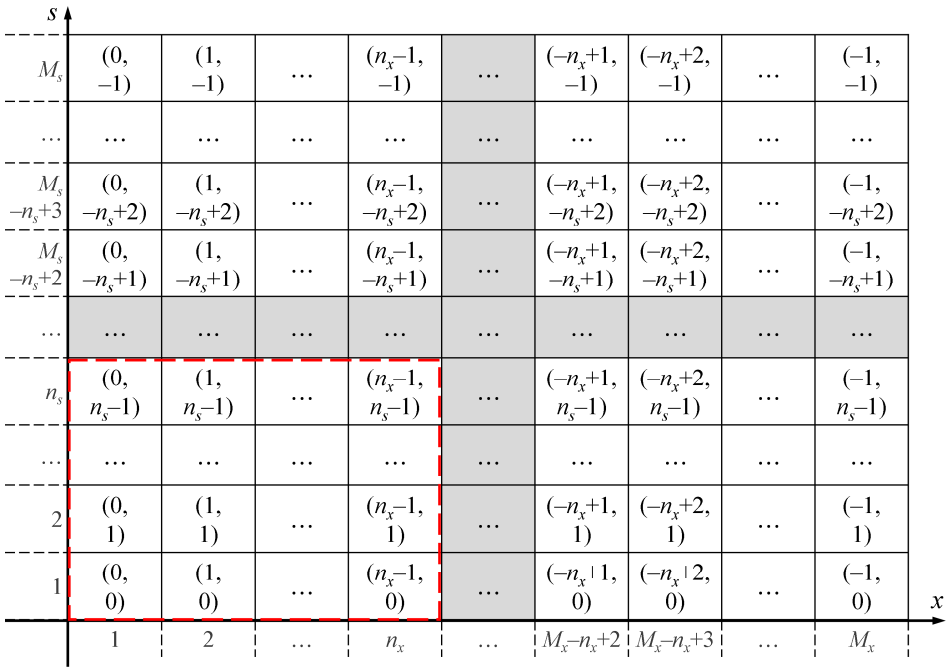
The differences between the intended linear convolution and the cyclic convolution are sketched in Figure 4.23. On the one hand, an incorrect influence is carried from the pressures applied in positions $M - N + 2, \ldots, M$ at the right side of the extended domain to positions $1,\ldots, N - 1$ of the original domain, as if those pressures were applied in the left side adjacent to the original domain. But as those pressures are zero due to zero padding, this incorrect influence is also zero. This is shown with the green marks in the figure. On the other hand, an incorrect influence is carried from the pressures applied in positions $1, \ldots, N - 1$ in the original domain to positions $M - N + 2, \ldots, M$ at the right side of the extended domain, as if those pressures were applied in the right side adjacent to the extended domain (marked as non-existent in the figure). This is shown with the purple marks in the figure. These are non-zero pressures which effectively carry an incorrect influence to the right side of the extended domain. The APs closer to the right end receive more incorrect influences: the AP at position M is influenced by $N - 1$ non-existent pressures, the AP at position $M - 1$ by $N - 2$, and so on. The incorrect influences do not reach the original domain, so this is solved by just disregarding the results of the convolution outside the original domain, which are not needed for the solution of the contact problem.

In conclusion, by extending **{IC}** and **{p}** to $M \geq 2N - 1$ elements, zero padding **{p}** and arranging **{IC}** in wraparound order, the contributions of the non-existent "wraparound" pressures from both ends of the extended domain are suppressed in the original (target) domain, thereby achieving the equivalence between the cyclic convolution and the linear convolution in the original domain.

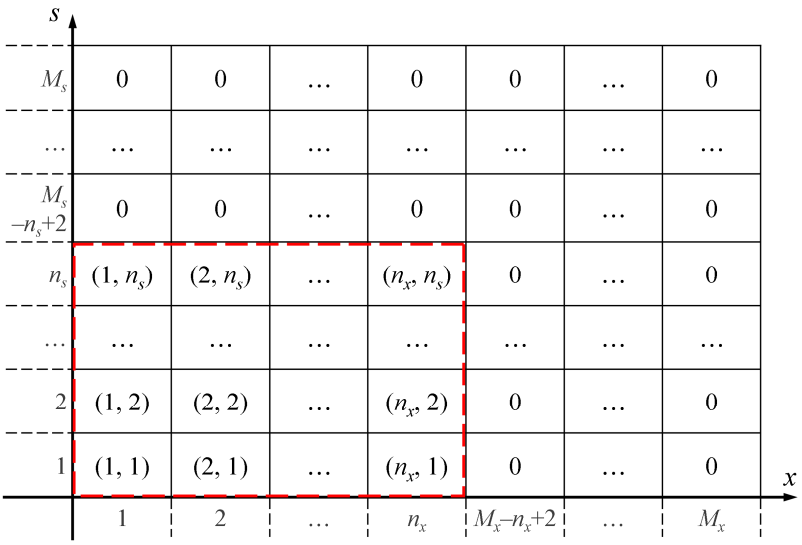
Once **{IC}** and **{p}** are properly arranged, the surface elastic displacements **{u}** in the original domain are calculated by efficient computation of the convolution **{IC}**⊗**{p}**, applying the discrete convolution theorem:

- The individual DFTs of the extended **{IC}** and **{p}** are calculated, DFT(**{IC}**) and DFT(**{p}**).
- The convolution is done in the Fourier domain by element-wise multiplication of DFT(**{IC}**) and DFT(**{p}**). The result is DFT(**{u}**).
- **{u}** is calculated by performing the inverse DFT of the vector obtained in the previous step. Only the elements of **{u}** within the original domain (numbered from 1 to N) are retained as the sought surface displacements.

The described process is analogous for multiple dimensions. In the case of 3D planar contact, the convolutions of the ICs with the contact stresses, as well as the FFT operations, are 2D. Figure 4.24 depicts the preparation of the arrays of ICs **[m_IC]** and contact stresses **[m_p]**, including the padding with zeros of **[m_p]** and the wraparound arrangement of **[m_IC]**. These arrays have a similar meaning to the vectors **{IC}** and **{p}** seen for the 1D case. The original **[m_IC]** and **[m_p]** arrays have $n_x \times n_s$ elements, coinciding with the dimension of the contact mesh. **[m_IC]** is a reshaped version of the vector `v_IC` used in the codes of Figure 4.16 and Figure 4.19. The original portion of each array is enclosed inside a dashed red rectangle in Figure 4.24. The figure shows only the subindices of the elements of the extended **[m_IC]** and **[m_p]** matrices. The subindices for **[m_IC]** in Figure 4.24a denote position differences in directions (x, s) of the mesh. A subindex (k_x, k_s) makes reference to the IC of an AP located at position (ix, is) over another located at position ($ix+k_x$, $is+k_s$). On the other hand, the subindices for **[m_p]** in Figure 4.24b denote the absolute position in the contact mesh (starting with index 1).



(a) **[m_IC]**.



(b) **[m_p]**.

Figure 4.24. Preparation of the arrays of ICs and contact stresses in 3D contact for the efficient computation of their 2D linear convolution in the Fourier domain.

[m_IC] and **[m_p]** are extended to dimension $M_x \times M_s$, with $M_x \geq 2n_x - 1$ and $M_s \geq 2n_s - 1$. Normally both M_x, $M_s << n_x \times n_s = NE$. So, apart from the computational efficiency, the convolution of ICs with contact stresses in the Fourier domain provides an advantage regarding memory usage, because the construction of the complete $NE \times NE$ $\mathbf{[IC_{ij}]}$ matrix is avoided, using **[m_IC]** in its place. The row and column shadowed in grey in Figure 4.24a represents the intermediate rows and/or columns which may be present if $M_s > 2n_s - 1$ and/or $M_x > 2n_x - 1$. The values of those possible additional rows and columns are left as zero in the extended **[m_IC]** and **[m_p]** matrices.

In the general non-planar case, the similarity of the ICs in the *s* direction of the mesh is lost, i.e. the ICs between two lateral positions s_I and s_J no longer depend only on the difference $s_I - s_J$. Therefore, the linear convolution of ICs with contact stresses may be performed only in longitudinal direction. The above described process for the planar case is adapted according to the code in Figure 4.25 for the non-planar case.

```
 1 Mx = 2^ceil(log2(2*nx-1));
 2 m_p = [m_p; zeros(Mx-nx, ns)];
 3 ft_m_p = fft(m_p);
 4 m_u = zeros(nx, ns);
 5 for i = 1:ns
 6    m_IC = zeros(Mx, ns);
 7    m_IC(1:nx, :) = a_IC_a(1+(i-1)*nx:i*nx, :);
 8    m_IC(Mx-nx+2:end, :) = ...
          a_IC_b(1+(i-1)*nx:i*nx-1, :);
 9    ft_m_IC = fft(m_IC);
10    ft_m_u = ft_m_IC .* ft_m_p;
11    ft_v_u = sum(ft_m_u, 2);
12    v_u = ifft(ft_v_u);
13    m_u(:, i) = v_u(1:nx);
14 end % for i = 1:ns
```

Figure 4.25. MATLAB code to perform the linear convolution of the contact stresses and ICs in the Fourier domain in non-planar contact.

The process is explained next, with reference to the code of Figure 4.25:

- Line 1: M_x is calculated as the minimum integer power of 2 greater than $2n_x - 1$.

- Line 2: **[m_p]** is expanded only in longitudinal direction, with zero padding in longitudinal positions $ix = n_x+1, \ldots, M_x$. The expanded **[m_p]** has $M_x \times n_s$ elements. Thus it has rows 1 to n_s in the s direction of the matrix **[m_p]** shown in Figure 4.24b.
- Line 3: The 1D DFT of **[m_p]** is calculated, column-wise. Thus, column j of the result `ft_m_p` contains the 1D DFT of the contact stress vector in longitudinal strip j of the contact mesh.
- Line 4: The matrix **[m_u]** of the sought displacements in the contact mesh is initialized.
- `for` loop for each strip `i` of the contact mesh (`i` = 1, …, n_s):
    - Lines 6–8: **[m_IC]** is defined, expanded only in longitudinal direction. This matrix, with $M_x \times n_s$ elements as **[m_p]**, contains the ICs of all the strips of the contact mesh over the current strip `i`. A given column j of the matrix contains the ICs of strip j over strip `i`.

      The ordering in the x subindex of the matrix is the same as that shown in Figure 4.24a for the planar case. So, the first (x) subindex of each element of the matrix is the same as in Figure 4.24a. Note that the matrices are represented in Figure 4.24 with the x axis horizontal, while the x direction runs through the first dimension, i.e. column-wise, in the matrices of the code of Figure 4.25.

      The second (s) subindex of each element of the matrix is `i` – j in each column j of the matrix. But this subindex may not be interpreted in the same general sense as explained for matrix **[m_IC]** in Figure 4.24a. In this case, as there are not similarities in the lateral direction, ICs with the same (`i` – j) s subindex are in general different for different `i` or j values.
    - Line 9: The 1D DFT of **[m_IC]** is calculated, column-wise. The result is stored in matrix `ft_m_IC`.

- Line 10: The DFT of the 1D circular convolution in the longitudinal direction of each column of **[m_IC]** with each column of **[m_p]** is computed as the element-wise multiplication of `ft_m_IC` and `ft_m_p`. Column *j* of the resulting matrix `ft_m_u` contains the 1D DFT of the contribution of the loads on strip *j* to the displacements of strip `i`.
- Line 11: The 1D DFT of the total displacements of strip `i` is computed as the sum of the contributions of all strips *j* of the contact mesh, summing the n_s columns of `ft_m_u` in a single vector `ft_v_u` of M_x elements.
- Line 12: `ft_v_u` is transformed to the space domain computing its inverse DFT. The result is stored in vector `v_u`, of M_x elements.
- Line 13: The first n_x elements of `v_u` are retrieved as the displacements of strip `i`.

Evidently, this process for the general non-planar case, calculating the influences between each pair of contact strips with 1D convolutions, is computationally more expensive than the process for the planar case, where a single 2D convolution is carried out for the whole contact mesh.

In the case of 2D non-planar contact problems with one AP in *x* direction, it is not possible to take advantage of convolutions in the Fourier domain to compute the surface elastic displacements. In this case, complete $\mathbf{[IC_{ij}]}$ matrices are formed (of dimension $NE \times NE = n_s \times n_s$), and the surface displacements are computed by multiplication of $\mathbf{[IC_{ij}]}$ with the contact stresses. In the current implementation of the program, this is done also with $n_x > 1$ for some of the ICs. Table 4.2 indicates which IC matrix is formed in each case, for the current and previous time instants, t and t'. Some details on the use of the complete $\mathbf{[IC_{ij}]}$ matrices in the normal and tangential solvers of this implementation are given in §4.1.4.

Table 4.2. IC matrices formed in the current implementation of CECT for $n_x > 1$.

IC_	***t***	***t'***
xx, *ss*, *xs*, *sx*	[IC_{ij}]	[m_IC']
nn	[IC_{ij}]	None
xn, *sn*	[m_IC]	[m_IC']
nx, *ns*	[m_IC]	None

4.1.1.4.3 Change of contact mesh during analysis

In a transient contact analysis with changing load, whether planar or not, it may be interesting to change the contact mesh in different steps during the analysis, adjusting it to the contact patch size at different stages of the analysis. Here it is described how this is implemented in CECT. Currently, the capability of mesh change during the analysis is implemented only for ICs calculated analytically in CECT.

Figure 4.26 shows contact meshes for two consecutive steps $k - 1$ and k in a contact analysis. The elements and nodes of each mesh are shown, the latter marked with circles. The mesh of step $k - 1$ is drawn in red, and the mesh of step k in black and grey.

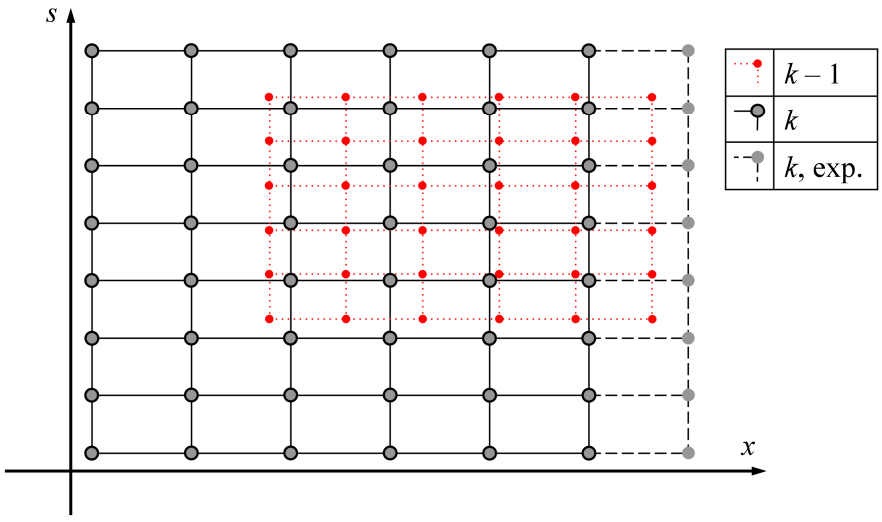


Figure 4.26. Change of contact mesh in transient contact analysis.

It is necessary to calculate the tangential elastic displacements on the APs of the mesh of step k, u'_τ, $\tau \in \{x, s\}$, due to the contact stresses acting in the previous

step $k-1$, which have been computed in a different mesh. If the mesh of step k is circumscribed in the mesh of step $k-1$, this may be done directly interpolating the (discrete) displacement field obtained in the mesh of step $k-1$ at the APs of the mesh of step k. But this is not possible if the mesh of step k is not circumscribed in the mesh of step $k-1$, as in the example shown in Figure 4.26. The procedure followed to cover this case is based on mapping the contact stresses obtained in step $k-1$ into the mesh of step k, as described next.

- The mesh of step k is extended as necessary to circumscribe the mesh of step $k-1$. In the example shown in the figure, this is accomplished by adding one column of elements in the right part of the mesh (i.e. in the positive x direction). The added elements and nodes are labelled as "k, exp.". The original (not extended) mesh of step k is labelled as "k".

  The numbers of new columns of elements to be added in the negative and positive x directions, $n_{add,x-}$ and $n_{add,x+}$, are computed according to Eqs. (4.30) and (4.31). X_{inf} and X_{sup} are the lower and upper boundaries of the mesh in the global X axis, and the terms inside parenthesis (k) and ($k-1$) indicate the step to which each value is referred. $\lceil\ \rceil$ indicates rounding towards $+\infty$. Similar equations are used for the expansion of the mesh in the s direction.

$$n_{add,x-} = \left\lceil \max\left(0;\ \ X_{\inf}(k) - X_{\inf}(k-1)\right) \Big/ \Delta x(k) \right\rceil \tag{4.30}$$

$$n_{add,x+} = \left\lceil \max\left(0;\ \ X_{\sup}(k-1) - X_{\sup}(k)\right) \Big/ \Delta x(k) \right\rceil \tag{4.31}$$

  The number of APs of the original mesh of step k is $n_x \times n_s = NE$, and the number of elements of the extended mesh is $n_{x,e} \times n_{s,e} = NE_e$.

  The indices of the APs of the original mesh of step k in the extended mesh are collected in a vector *ie_e*. The number of these elements is evidently *NE*.

- The contact stresses of step $k-1$ are interpolated in the APs of the extended mesh of step k, forming matrices $\mathbf{[p_{j,o,e}]}$, $j \in \{x, s, n\}$ (the _e suffix refers to the extended mesh, and the _o suffix to quantities of

step k – 1). Stepwise (‘*nearest*’) interpolation is done with uniform elements, and bilinear with bilinear elements. 0 values of stress in the [$\mathbf{p_{j,o.e}}$] matrices are assumed for the APs outside the domain of the mesh of step k – 1.

- According to the previously obtained $\mathbf{[p_{j,o,e}]}$ matrices, the APs of the extended mesh of step k that were in contact in step k – 1 are determined as those where the normal pressure $p_{n,o,e}$ is non-zero. The number of these elements is *NEC_o_e*, and their indices in the extended mesh of step k are collected in a vector *iec_o_e*. This is used in the construction of complete $\mathbf{[IC'_{ij}]}$ matrices in the case of contact problems with one AP in x direction.

- The arrays of ICs for the previous time instant t' (t' corresponding to step k – 1) are generated for the APs of the extended mesh of step k. Arrays `a_IC_a` and `a_IC_b`, of dimension $NE_e \times n_{s,e}$, are generated from vectors of half-space ICs `v_IC_a` and `v_IC_b`, of length NE_e, see §4.1.1.4.1.

  The sought tangential displacements u'_τ are calculated by convolutions of the ICs with matrices $\mathbf{[p_{j,o,e}]}$, all magnitudes being expressed in the extended mesh of step k. The necessary values corresponding to the *NE* APs of the original mesh of step k, with indices *ie_e*, are retained for use in the tangential problem solution.

  The relevant ICs for this calculation are IC'_{xx}, IC'_{xs}, IC'_{sx}, IC'_{ss}, IC'_{xn} and IC'_{sn}. Since the normal contact problem is instantaneous in linear elastostatics, previous normal displacements u'_n are not needed, and ICs IC'_{nj}, $j \in \{x, s, n\}$, are not computed.

  In the case of 2D contact problems with one AP in x direction, complete $\mathbf{[IC'_{ij}]}$ matrices are constructed instead of arrays `a_IC_a` and `a_IC_b`. The displacements u'_τ are calculated multiplying matrices $\mathbf{[IC'_{ij}]}$ with matrices $\mathbf{[p_{j,o,e}]}$. Each $\mathbf{[IC'_{ij}]}$ matrix, of dimension $NE \times NEC\_o\_e$, is formed from vectors `v_IC_a` and `v_IC_b`. The rows of the matrix correspond to the APs of the original mesh of step k, and the columns to the APs of the extended mesh of step k that were in contact in step k – 1.

With this procedure, the calculation of ICs between elements of different meshes is avoided. This would be expensive, since in general there are not similarities in the relative positions between elements belonging to different meshes (with different element sizes) neither in longitudinal nor in lateral direction. So it would be necessary to compute complete $NE \times NEC\_o$ $\mathbf{[IC'_{ij}]}$ matrices (NEC_o being the number of elements in contact in the mesh of step $k-1$), and calculate the displacements by direct multiplication of $\mathbf{[IC'_{ij}]}$ with the contact stresses of step $k-1$.

In addition to the calculation of u'_z, the following quantities initially computed in the mesh of step $k-1$ are mapped into the mesh of step k, for use in the solution of the tangential contact problem:

- The elements that were in slip in step $k-1$, to be fed as initialization values for the tangential contact solver. The array of the binary flag [slip/no slip] obtained in step $k-1$ is interpolated in the APs of the extended mesh of step k, by means of stepwise interpolation. The state of adhesion is initially assigned for the APs outside the domain of the mesh of step $k-1$.
- In the case of static or shift (not rolling) contact problems, the rigid shifts of step $k-1$, by means of spline interpolation.

4.1.2 Input and output files

In this section the structure, contents and format of the CECT input and output files is explained.

4.1.2.1 Input files

The input for CECT is provided in text files. The following format conventions apply to all input files:

- Any number of blank and comment lines may be introduced, anywhere in the file. Comments may also be added in the right part of any line. The beginning of comments is marked with '%'.
- Numbers may be introduced in any format commonly admitted. For example, 1 may be introduced as 1, 1., 1.0, 1E+0, … In the case of the

main input file, equivalent operations may be also introduced, e.g. 2/2 in place of 1.

- The different elements of a line may be separated with spaces (one or several consecutive) or commas.
- Any number of blank spaces may be introduced at the beginning and at the end of a line.
- String type data must be enclosed in single quotes ('…').

The different input files are explained in the following subsections. The units in which each input parameter must be provided are specified where applicable.

4.1.2.1.1 Main input file

Multiple analysis cases may be specified in a single input file. Each case refers to a time step in a transient contact analysis (shift or rolling), or to a steady rolling contact analysis. The main input file has a first section with common data for all analysis cases, followed by the data for each case. The input variables are organized in a defined order throughout the file, according to Table 4.3 and Table 4.4. The lines from which input data is read by CECT are numbered in the left side of these tables. Some of the items are mandatory (must be always supplied) and others are optional. This is indicated in these tables in the following way:

- In bold: mandatory items.
- Within brackets ([]): optional items.
- Otherwise: items that are mandatory for particular values of previous items.

Next each variable of the input file is explained.

Common data

Table 4.3. Structure of CECT main input file I: common data for all analysis cases.

	1	2	3	4	5	6	7	8
1	**Ncases**	**c_o_n**	[ics]					
2	**E_w**	**E_r**	**nu_w**	**nu_r**	[mat_dgt]			
	% If mat_dgt = 1:							
3	G_inf	r_inf_G	exp1_G	exp2_G	nu_inf	r_inf_nu	exp1_nu	exp2_nu
4	K_inf	r_inf_K	exp1_K	exp2_K	[dum]	[dum]	[dum]	[dum]
	% If mat_dgt = 2:							
3	E_w_inf	r_inf_E_w	exp1_E_w	exp2_E_w	nu_w_inf	r_inf_nu_w	exp1_nu_w	exp2_nu_w
4	E_r_inf	r_inf_E_r	exp1_E_r	exp2_E_r	nu_r_inf	r_inf_nu_r	exp1_nu_r	exp2_nu_r

- 1st line.

  `Ncases`: total number of cases to be calculated. The input file must contain data at least for this number of cases. If there are data for more cases, the data for the excess cases is not used.

  `c_o_n`: bit indicating if the mesh will use uniform or bilinear elements (see §5.1). The APs are the centres (c) of the elements in the case of the uniform elements, and the corners or nodes (n) in the case of the bilinear elements. Possible values:

  0: uniform elements.

  1: bilinear elements.

  `ics` (optional): digit or string defining the influence coefficients (ICs) for the generally non-planar contacting bodies. The same definition is used for all the ICs. Default: 0. Options:

  0: surface orientation-based analytical approximation, according to Eq. (5.72) in §5.2.3.1.

  1: half-space ICs. The curvilinear *s* coordinate is used in place of the lateral *y* coordinate of a plane contact surface.

2: similar to option 0, adding a factor of $|\pi - |\alpha|| / \pi$ (with the angle α defined in Figure 5.28) in the term A_{sn} or A_{ns} of the approximated ICs B_{ss}, B_{nn}, B_{sn} and B_{ns}, cf. §5.2.3.3.

3: $\hat{B}$ variant of surface orientation-based analytical approximation, according to Eq. (5.77) in §5.2.3.4.

4: similar to option 3, adding the $|\pi - |\alpha|| / \pi$ factor as in option 2.

i + 10: similar to i, with $i \in \{1, 2, 3, 4\}$, except for IC_{xn} and IC_{nx}, which are equated to the half-space A_{xn} and A_{nx}, respectively.

20, 21, 22: 2D cylindrical ICs, cf. §5.2.1.1. The different options specify different support and loading conditions for the cylinder. The cavity is single-loaded in all cases.

- 20: central restriction, and unique load on cylinder (cf. Figure 2.22b).
- 21: uniformly distributed restriction in whole section, and unique load on cylinder (cf. Fig. Figure 2.22c).
- 22: two diametrically opposed loads on cylinder (cf. Figure 2.22a).

The following limitations apply for `ics` = 20–22:

- `c_o_n` = 0
- `mat_dgt` = 0
- `nx` = 1
- `tdgt` = 0
- The wheel and rail profiles (`prf_w` and `prf_r`) must be defined with 3 or more points. The profiles should be circular or close to circular for correct results.

If any of these conditions is not fulfilled, CECT prints an error message and aborts execution.

Character string: root, without extension, of the names of the files containing the IC arrays (`a_IC`, cf. §4.1.1.4) calculated numerically beforehand. CECT will search for 9 txt files with this root and suffixes *_Bij*, with $i, j \in \{x, n, s\}$. The structure of these files is explained in §4.1.2.1.2.

- 2nd line. Elastic properties of materials.

  `E_w`, `E_r`: Young's modulus of wheel (w) and rail (r). [GPa]

  `nu_w`, `nu_r`: coefficient of Poisson of wheel (w) and rail (r). [-]

  `mat_dgt` (optional): digit specifying space variation model for effective elastic properties of contacting bodies. Cf. §5.2.3.5. Default: 0. Options:

  0: uniform elastic properties.

  1: spatially variable elastic properties, specified for the combined elastic properties of both bodies.

  2: spatially variable elastic properties, specified for the individual elastic properties of each body.

- 3rd and 4th lines (optional; must be provided if `mat_dgt` = 1 or 2). Parameters defining the space variations of the elastic properties of the contacting bodies. If `mat_dgt` = 1, the numbers in columns 5 to 8 of line 4 are not used. Designating generically the different elastic properties as `f`; `f` ∈ {`G`, `nu`, `K`, `E_r`, `nu_r`, `E_w`, `nu_w`}, and with reference to Eqs. (5.78) and (5.79):

  `f_inf`: effective elastic property `f` far from the loaded zone. Equal to f^{∞} in Eq. (5.78). Units: GPa for `G` and `E*`, dimensionless for `nu*` and `K`.

  `r_inf_f`: limit distance for the variation function of the elastic property `f`. Beyond this distance from the loaded zone, `f` does not change and is equal to `f_inf`. Equal to $r_{\infty,f}$ in Eq. (5.79). [mm]

  `exp1_f`, `exp2_f`: exponents for the variation function of `f`. Equal to m_f and n_f in Eq. (5.78), respectively. [-]

Data for each case

Table 4.4. Structure of CECT main input file II: particular data for each analysis case.

	1	2	3	4	5	6	7	8	9
0	**CASE**	**c**	**r**						
	% Geometry								
1	**gdgt**	prf_w	prf_r	mtd_intrp	incl_r	r_nom_w	[r_nom_r]	hw_ws	hw_tr
2	despl_y	despl_z	psi	phi					
	% Mesh								
3	**mdgt**	x_0	s_0	x_inf	x_sup	s_inf	s_sup	nx	ns
	% Normal								
4	**ndgt**	**d_o_N**	**dir_d**	**r_d_ini**					
	% Tangential								
5	**tdgt**	**dq**	**sftx_o_vel**	**omega**	**f**	[bit_l6]			
6	x_v	y_v	z_v	dir_vx	dir_vy	dir_vz	dir_omx	dir_omy	dir_omz

- 0th line. The beginning of the specification of a case is marked with the keyword `CASE`. This line is completed with the following two parameters:

  `c`: integer number of the current case. It must follow the progression 1, 2, 3, … for all the cases given in the input file. This number is used as part of the output .mat and .prf file names for the case, if they are created.

  `r`: bit specifying if .mat output file (cf. §4.1.2.2.1) with detailed results for the current case is to be printed (1) or not (0).

- 1st line. Geometry.

  `gdgt`: bit specifying if new geometry is specified for the current case (1) or not (0). If `gdgt` = 0, no more data must be provided in the 1st and 2nd lines, and the data of the last case in which `gdgt` was 1 is used. If `gdgt` = 1, the rest of the parameters of the 1st and 2nd lines must be specified, except for `r_nom_r` which may optionally be omitted. `gdgt` must be 1 in the 1st case.

Note: no transition is carried out between the data of different cases. If new data are introduced in a case, it is assumed that they apply to the whole PCS, and the data from the previous case are "forgotten".

`prf_w`, `prf_r`: character strings specifying the names of the text files defining the wheel (w) and rail (r) profiles, file extension included. Cf. §4.1.2.1.3.

Note: Attention: currently the program is not prepared to deal with changes in the input file for the rail profile `prf_r` during transient analyses, either shift (`tdgt` = 0) or rolling (`tdgt` = 1). The reason is that the possible changes in the absolute spatial position that may take place in the PCS, even when its origin in lateral direction `s_0` is not changed (note that `s_0` is measured over the input rail profile), would not be correctly taken into account when the input rail profile is changed.

`mtd_intrp`: string specifying the interpolation method for the extraction of points from the given wheel and rail profiles, for the calculation of the normal undeformed distances. Options (see documentation of MATLAB interp1 and interp1q functions): [], `'nearest'`, `'linear'`, `'spline'`, `'pchip'`, `'cubic'`, `'v5cubic'`. If an empty value is given ([]), the quick linear interpolation method of MATLAB is used (interp1q). The `'spline'` method is recommended for best results.

`incl_r`: tangent of the rail inclination (cant) angle. Positive towards the track centre. The rail profile defined in the `prf_r` file is rotated the angle defined with this parameter with respect to the (0, 0) point of the profile. The rail inclination angle is in the [–90º, 90º] range.

`r_nom_w`: nominal rolling radius of the wheel. It must be always positive; that is, the wheel is always convex in the rolling direction. $R_{roll}(y_L)$ = `r_nom_w` – $z_L(y_L)$; being $R_{roll}(y_L)$ the local rolling radius of the wheel at a lateral position y_L in its local coordinate system (cf.

Figure 4.1a), and $z_L(y_L)$ the vertical coordinate of the wheel profile in the same lateral position. [mm]

`r_nom_r` (optional): nominal rolling radius of the rail or raceway. Positive if convex and negative if concave in rolling direction. If omitted or 0, a straight rail is assumed. [mm]

Note: the revolution axis of the rail is parallel to the global *Y* axis. Vertically it is located at $Z = -$`r_nom_r`. Longitudinally it is located at $x = 0$ of the local contact coordinate system. `incl_r` does not have any effect over its orientation.

`hw_ws`: half-width of the wheelset. It is the lateral distance between the centre of the wheelset and the origin of the local wheel profile coordinate system. [mm]

`hw_tr`: half-width of the track. It is the lateral distance between the centreline of the track and the origin of the local rail profile coordinate system, which is positioned at (`hw_tr`, 0) in the global *YZ* plane. [mm]

- 2nd line. Geometry; definition of the wheel position and orientation. Before the displacements and rotations defined in this line, the origin of the local wheel profile coordinate system is located at (0, `hw_ws`, 0), and the centre of the wheelset at (0, 0, `r_nom_w`), in the global {*X*, *Y*, *Z*} coordinate system.

  `despl_y, despl_z`: displacements of the wheelset centre in the *YZ* plane. After these displacements, the centre of the wheelset is located at (0, `despl_y`, `r_nom_w` + `despl_z`) in the global coordinate system. [mm]

  `psi`: wheelset yaw angle. Cf. ψ in Figure 4.1b. [rad]

  `phi`: wheelset roll angle. Cf. ϕ in Figure 4.1b. [rad]

  Note: the parameters of this line are intended to define the position of the wheel in the undeformed, unloaded configuration. The wheel should be positioned so that it is just touching the rail in the undeformed configuration. For this purpose, the geometric contact search problem

is solved as outlined in §4.2.3.1. The final wheel position may change from the position defined in this line as a result of the specification of the normal and tangential problems. The most directly related parameters are `d_o_N`, `dir_d`, and `sftx_o_vel` (the latter mostly for `tdgt` = 0).

- 3rd line. Mesh; definition of the domain of the PCS and its discretization.

  `mdgt`: bit specifying if a new domain and discretization of the PCS is defined (1) or not (0). If `mdgt` = 0, no more data must be provided in this line, and the data of the last case in which `mdgt` was 1 is used. If `mdgt` = 1, the rest of the parameters of this line must be specified. `mdgt` must be 1 in the 1st case.

  `x_0`: position in global *X* axis of the origin of the local contact coordinate system. [mm]

  `s_0`: lateral curvilinear coordinate on the rail profile of the origin of the local contact coordinate system. It is measured from the first point defining the rail profile given in the `prf_r` file. The point defined with `x_0` and `s_0` should normally be at or around the position of the initial or geometric contact point. [mm]

  `x_inf`, `x_sup`: limits of the domain of the PCS (between element borders) in the local contact *x* axis. [mm]

  `s_inf`, `s_sup`: limits of the domain of the PCS (between element borders) in the local contact *s* axis. [mm]

  `nx`, `ns`: number of elements of the mesh of the PCS, in the longitudinal and lateral directions, respectively. The number of APs in each direction is equal to `nx` + `c_o_n` and `ns` + `c_o_n`.

- 4th line. Normal contact.

  `ndgt`: digit specifying the type of input `d_o_N` and output `N_o_d` (cf. §4.1.2.2.1). Options: 0, 1.

  `d_o_N`:

  If `ndgt` = 0: approach, *d*. [mm]

If `ndgt` = 1: normal load, *N*. [kN]

Note: the normal load is the integral of the normal pressures in the contact area, which do not have a common direction. Its direction is not known beforehand. If `ndgt` = 1, what is given as input through `d_o_N` is the projection of the normal load on the direction defined by `dir_d`, and not its magnitude. On the other hand, if `ndgt` = 0, the output `N_o_d` is the magnitude of the normal load.

`dir_d`: angle defining the direction of the approach. It is measured from the *Z* axis, counter clock-wise while looking towards the negative *X* direction (cf. the α_d angle in Figure 4.12). A value of 0 stands for the vertical descending direction. [rad]

`r_d_ini`: factor multiplying the approach, to obtain the initial guess of the set of elements in contact in the first iteration of the normal solution. More details are given in §4.1.4.1. Recommended value around 0.55. [-]

- 5th line. Tangential contact.

`tdgt`: digit specifying the type of tangential problem. Options:

0: shift, static.

1: transient rolling.

2: steady rolling.

`dq`: in rolling problems, longitudinal distance travelled between the previous and the current time instants. It should have a value around the longitudinal dimension of the elements of the contact mesh `dx`. [mm]

Note 1: in shift problems (`tdgt` = 0) this parameter is not used, and the program sets `dq` = 0.

Note 2: in steady rolling problems (`tdgt` = 2), CECT tries to use the SteadyGS tangential solver (cf. §4.1.4.2). For this purpose, `dq`

must be equal to `dx`. If it is not, CECT prints a warning message and changes the value of `dq` to `dx`.

Note 3: if the rail has curvature in the longitudinal direction (i.e. if `r_nom_r` $\neq 0$), `dq` will in general be variable in the *s* direction. In this case, the `dq` value specified as input is assigned to the $s = 0$ position. More details are given in §7.2.2.1.2.

`sftx_o_vel`: if `tdgt` = 0, imposed rigid shift on the wheel in *x* direction [mm]. If `tdgt` = 1 or 2, velocity of the wheelset [m/s].

Note 1: it is assumed that the rail has no rigid body motion.

Note 2 (applicable for `tdgt` = 0): this input represents the imposed displacement (shift) to the wheel in the longitudinal direction. In the *YZ* plane, the imposed displacement is determined by `d_o_N` and `dir_d`. The imposed displacements have to be introduced with respect to the initial undeformed position at the beginning of the first case of a consecutive series of cases with `tdgt` = 0.

Note 3 (applicable for `tdgt` = 1 or 2): this input is the velocity of a point of the wheel. If `bit_l6` = 0, it is assumed that its direction is along positive *x*, and that it corresponds to a point in the wheelset axis; i.e., it is the travelling velocity of the wheelset. If `bit_l6` = 1, this velocity corresponds to the point of the wheelset located in the position (`x_v`, `y_v`, `z_v`) in the local coordinate system of the wheelset $\{x_w, y_w, z_w\}$, and its direction is defined by the cosine directors (`dir_vx`, `dir_vy`, `dir_vz`).

`omega`:

If `tdgt` = 0: pivoting rotation imposed to the wheel with respect to the rail (total with respect to the initial undeformed position of the first case of a consecutive series of cases with `tdgt` = 0). It is equal to Ω_s in Eqs. (4.23) and (4.24). The direction of the rotation axis is defined with `dir_d`, the positive sense pointing into the wheel. The pivoting point is the origin of the local contact coordinate system, defined with `x_0` and `s_0`. [rad]

If `tdgt` = 1 or 2: rotation velocity of the wheelset. If `bit_l6` = 0, it is assumed to be aligned with the wheelset axis, i.e. with the y_w axis. If bit_l6 = 1, its direction is defined by the cosine directors (`dir_omx`, `dir_omy`, `dir_omz`). [rad/s]

`f`: coefficient of friction. A number or a string is expected. If a number is input, this is the uniform value of the coefficient of friction in the whole PCS. If a string is input, it is the name of a file (extension included) in which spatially variable coefficients of friction may be defined. The format of this file is specified in 4.1.2.1.4. [-]

`bit_l6` (optional): bit indicating if a 6th line is provided in the case. Default: 0.

- 6th line. Tangential contact; specification of general velocity and rotation velocity vectors. This line must be provided only if `bit_l6` = 1, and is taken into account only in rolling problems (`tdgt` = 1 or 2).

  `x_v`, `y_v`, `z_v`: coordinates in the wheelset local coordinate system {x_w, y_w, z_w} of the point for which the velocity (`sftx_o_vel`) is given. [mm]

  `dir_vx`, `dir_vy`, `dir_vz`: cosine directors, in the global coordinate system, of the velocity vector of the wheelset point located at (`dir_vx`, `dir_vy`, `dir_vz`) in the wheelset local coordinate system {x_w, y_w, z_w}. [-]

  `dir_omx`, `dir_omy`, `dir_omz`: cosine directors, in the global coordinate system, of the rotation velocity vector of the wheel. [-]

4.1.2.1.2 Files of externally calculated ICs

These files contain `a_IC` arrays of ICs calculated outside CECT. These ICs may be calculated numerically with FE models. The process of such calculation is detailed in §5.2.2. 9 of these files (one for each IC_{ij}, with $i, j \in \{x, n, s\}$) must be supplied when the `ics` parameter in the main input file is a character string. The structure of these files is as follows.

- 7 first lines (comment lines excluded): head lines with the data specified in Table 4.5 and Table 4.6. The first two lines are padded with zeros to the right summing up a total of `ns` numbers in each line (this padding is not shown in Table 4.5).

  Note 1: generally, these lines have to be equal in the 9 IC files. Otherwise, CECT will print an error message and abort execution. As an exception, differences between the Young's and shear moduli are permitted, if the coefficients of Poisson and the *K* constant are equal to the data provided in the main input file. If there are differences in the Young's and shear moduli, CECT will scale the imported ICs to correspond to the elastic constants given in the main input file.

  Note 2: the data of the mesh of the PCS given in these IC files (`dx`, `nx`, `x_sup` – `x_inf`, `s_inf`, `s_sup`, `ns`, `ds`, `c_o_n`) must match with the data of the mesh of the PCS given in the main input file (`c_o_n`, `x_inf`, `x_sup`, `s_inf`, `s_sup`, `nx`, `ns`). Otherwise, CECT will print an error message and abort execution.

Table 4.5. Structure of the first two head lines of the files containing the externally calculated IC arrays for CECT input.

	1	2	3	4	5	6	7	8	9
	% Elastic constants of the bodies:								
1	E_r	nu_r	G_r	E_w	nu_w	G_w	G	nu	K
	% Data of the objective mesh:								
2	x_sup – x_inf	nx	dx	s_inf	s_sup	ns	ds	c_o_n	cor

Table 4.6. Structure of the 3rd to 7th head lines of the files containing the externally calculated IC arrays for CECT input.

	1	2	...	ns
	% s [mm]			
3	s_1	s_2	...	s_ns
	% yr [mm]			
4	yr_1	yr_2	...	yr_ns
	% zr [mm]			
5	zr_1	zr_2	...	zr_ns
	% yw [mm]			
6	yw_1	yw_2	...	yw_ns
	% zw [mm]			
7	zw_1	zw_2	...	zw_ns

- 1st line: elastic constants of the bodies. `E`: Young's modulus [GPa]; `nu`: coefficient of Poisson [-]; `G`: shear modulus (redundant) [GPa]. Subindex `r` for the rail and `w` for the wheel. The last `G`, `nu` and `K` are the combined elastic constants of the two bodies.
- 2nd line: data of the discretization:
  - `x_sup – x_inf`: longitudinal dimension of the mesh (between leading and trailing element borders). [mm]
  - `nx`, `ns`: number of elements in *x* and *s* directions.
  - `dx`, `ds`: dimension of each element in *x* and *s* directions. [mm]
  - `s_inf`, `s_sup`: lower and upper limits of the mesh in *s* direction (between element borders). [mm]
  - `c_o_n`: bit to indicate if the elements are uniform (0) or bilinear (1).
  - `cor`: bit to indicate if the ICs of the file need to be corrected (0) or if they are ready for use in CECT as they are (1). In the former case, the correction for the ICs at the loaded and surrounding elements described

in Figure 5.24 (§5.2.2) is applied inside CECT to the imported ICs.

- 3rd line: local contact curvilinear *s* coordinate at each lateral position of the mesh. [mm]
- 4th to 7th lines: local contact *y* and *z* Cartesian coordinates of the rail and wheel cross-sections in the plane perpendicular to the *x* direction, at each lateral position of the mesh. [mm]

  Note: a profile is calculated as a weighted interpolation of the wheel and rail cross-sections defined in these lines, in a similar way as it is done in the calculation process of the normal undeformed distances with the wheel and rail geometry data provided in the main input file (cf. §4.1.1.2). CECT prints a warning message in the .log file if the difference between the coordinates of the profile calculated with this data and the profile of the PCS is larger than 1 μm.

- Rest: `a_IC` array, with dimension `((nx+c_o_n)*(ns+c_o_n), ns+c_o_n)`. Its structure is as indicated in §4.1.1.4. [mm/GPa]

4.1.2.1.3 Files of wheel and rail profiles

The files specified with the `prf_w` and `prf_r` strings in the main input file contain the definition of the wheel (w) and rail (r) profiles. These are specified as a set of points in the $\{y_L, z_L\}$ coordinate system of each profile. The structure of the files containing wheel and rail profiles is identical. It is illustrated in Table 4.7.

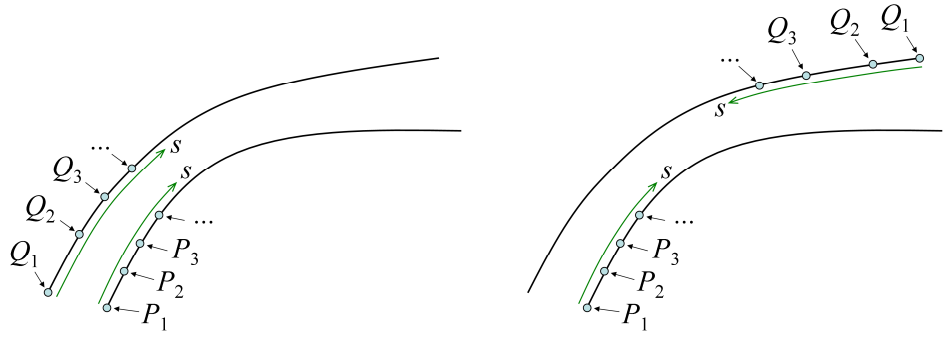
Table 4.7. Structure of input file with a wheel or rail profile.

	1	2
	% Profile description (optional)	
1	yL,1	zL,1
2	yL,2	zL,2
...	...	...
Np	yL,Np	zL,Np

Each file contains a set of data lines. Each data line contains two numbers, which are the $(y_{L,i}, z_{L,i})$ coordinates of a point *i* of the profile defined in the file.

Any number of `Np` points may be chosen to discretize the profile; the number of data lines in the file is equal to the number of these points. The coordinates must be given in mm. The following notes apply:

- The points in the file have to be given in consecutive order.
- Neither the y_L nor the z_L coordinates need be monotonically increasing or decreasing; there may be vertical sections for example.
- The direction of the local contact *s* coordinate is determined based on the order of the given points of the profiles. This order must be consistent in the wheel and rail profiles, taking into account the relative position of both profiles in the contact analysis, so that the direction of the *s* coordinate in both profiles matches. Figure 4.27 shows examples of definition of rail and wheel profiles defined with sets of points P_i and Q_i, with consistent and inconsistent order.



(a) Consistent order. (b) Inconsistent order.

Figure 4.27. Examples of pairs of rail and wheel profiles with consistent and inconsistent order.

- The input profiles have to be smooth, with no measurement noise and roughness, so that the *s* curvilinear coordinate over both profiles represents similar distances in the *sn* plane. In this way, it is possible to define homologous points as those with the same (*x*, *s*) coordinates, as explained in §4.1.1.1. Figure 4.9b illustrates a case in which this condition is not fulfilled.
- For non-zero `hw_ws` and `hw_tr` values, the profiles correspond to a right wheel/rail pair as viewed looking towards the negative *X* direction,

according to the positioning of the profiles in the global coordinate system explained under the definition of `hw_tr` and the 2nd line with the geometry data for each case in §4.1.2.1.1.

4.1.2.1.4 File with coefficients of friction

Spatially variable coefficients of friction may be defined by means of this file. The coefficients of friction defined here depend only on the position; they do not depend on other field variables like slip velocities for example.

The file contains a numerical matrix that represents the field of the coefficient of friction. The data are given in a rectangular mesh, with coordinates X_a, S_a:

- X_a: longitudinal coordinate, parallel to the global X axis. Its origin is the position of the wheelset centre in the first case of the input file. It is different from the global X coordinate in that its origin does not move along the track with the wheelset centre; it remains in the same point. In this way, variable coefficients of friction may be readily defined for a length of track in the file. This is useful for a set of transient rolling cases for example, in which the wheelset travels over a track spot with different friction conditions.
- S_a: lateral curvilinear coordinate on the rail profile, with origin in the first point of the rail profile given in the `prf_r` file of each case.

The matrix of coefficients of friction given in the file is rectangular, with `(NS+1)×(NX+1)` elements. Its structure is depicted in Table 4.8. `NX` and `NS` are the number of points discretizing the field of the coefficient of friction in X_a and S_a directions, in a grid of rectangular elements with sides aligned with these directions.

Table 4.8. Structure of input file with spatially variable coefficients of friction.

	1	2	3	...	NX+1
	% Description (optional)				
1	0	Xa_1	Xa_2	...	Xa_NX
2	Sa_1	f_1_1	f_2_1	...	f_NX_1
3	Sa_2	f_1_2	f_2_2	...	f_NX_2
...	...	...	...	...	...
NS+1	Sa_NS	f_1_NS	f_2_NS	...	f_NX_NS

The first row of the matrix contains the X_a coordinates of the mesh, starting in the second column; and the first column contains the S_a coordinates, starting in the second row. The coordinates have to be given in consecutive increasing order in each direction, and they may be not uniformly spaced. So, the elements of the mesh may have different sizes. The coordinates are given in mm, and the coefficient of friction is non-dimensional. Position (`i+1`, `j+1`) of the matrix contains the coefficient of friction at point with coordinates (`Xa_j`, `Sa_i`). The value in the position (1, 1) of the matrix is not used, but this position has to be filled with some numerical value. The field described by this matrix is mapped into the APs of the mesh of the PCS of each analysis case by means of bilinear interpolation in CECT. Zero coefficient of friction is assumed for the APs falling outside the domain of this matrix.

4.1.2.2 Output files

The output from CECT is printed to different text files. All the files begin with three or four similar header lines, which include the name of the main input file, the CECT version, and the date/time of creation of the output file.

The root name of all the output files is the same as that of the main input file. In addition, the names of the .mat and .prf files, which have data for a single case, have the number of the case included in %04d format (i.e. zero-padded to the left to four digits). For example, the name of the .mat file for the 12th case of a main input file with name `file_name.txt` would be `file_name.0012.mat`. If upon creation of a new file, CECT detects that a file with the same name as the one initially assigned to the new file already exists in the current working directory (CWD), the current system date is appended to the name of the new file in yyyy-mm-dd format. Continuing with the previous example, the name of the new .mat file would have the following structure: `file_name.yyyy-mm-dd.0012.mat`. If a file with this latter name already exists as well in the CWD, CECT overwrites it when creating the new file.

These are the sign conventions of the returned output quantities:

- The resultant forces and stresses are those acting on the wheel.

- Displacement, velocity and elastic deformation differences are defined as those of the wheel with respect to the rail.

In addition to the printed output, the following output variables are returned to the MATLAB workspace:

- `sol`: structure including solution arrays with data for each AP of the PCS, and some mesh and input data. The following scalars and vectors are included:
    - `c_o_n`: cf. §4.1.2.1.1.
    - `mx`, `ms`: number of APs of the PCS in *x* and *s* directions, respectively. `mi` = `ni` + `c_o_n`, with `i` ∈ {`x`, `s`}.
    - `x`, `s`: vectors with the coordinates of the APs of the mesh in *x* and *s* directions, with `mx` and `ms` elements, respectively. [mm]
    - `tdgt`: cf. §4.1.2.1.1.
    - `veloc`:
        - If `tdgt` = 0: `sftx_o_vel` (cf. §4.1.2.1.1). [mm]
        - If `tdgt` = 1 or 2: contact velocity over rail surface, $V_{c,r}$ (cf. §7.2.2.1.2). For a straight rail, equal to `sftx_o_vel` if `bit_l6` = 0 (cf. §4.1.2.1.1). [m/s]

          Note: presently, only positive contact velocities are admitted (i.e. contact advancing along positive *x* axis).

    The following arrays complete the structure. The dimensions of each array are `mx`×`ms`. Position (*ix*, *is*) of each array contains the data for the AP located in the ix^{th} longitudinal position and is^{th} lateral position of the mesh.

    - `eldiv`: integer specifying the state. Outside contact (0), inside contact and sticking (1), or inside contact and sliding (2). [-]
    - `h`: undeformed distance, before application of approach. [mm]
    - `mu`: coefficient of friction. [-]

  - `pn`, `px`, `ps`: contact stresses in *n*, *x* and *s* directions. [MPa]
  - `un`: normal elastic displacement differences due to normal pressures. [mm]
  - `ux`, `us`: tangential elastic displacement differences in *x* and *s* directions due to all contact stresses (normal and tangential). As an exception, if the ConvexGS solver (cf. §4.1.4.2) is used with `tdgt` = 2, what is output are the increments between the previous and the current time instant of these tangential elastic displacement differences. [mm]
  - `srel`:
    - If `tdgt` = 0:
      - If `sftx_o_vel` = 0: absolute slipped distance between previous and current time instants. [mm]
      - If `sftx_o_vel` ≠ 0: absolute slipped distance between previous and current time instants divided by `sftx_o_vel`. [-]
    - If `tdgt` = 1 or 2: non-dimensional relative slip velocity v_{rel}. $v_{rel} = \|\mathbf{v}\| / V_c$, being **v** the slip velocity and V_c the contact velocity. [-]
  - `wx`, `ws`:
    - If `tdgt` = 0: rigid shift increments between previous and current time instants in *x* and *s* directions. [mm]
    - If `tdgt` = 1 or 2: rigid slip velocities in *x* and *s* directions. [m/s]

- `prf_pcs`: matrix with the (*Y*, *Z*) coordinates of the profile of the PCS, with dimensions `ms`×2. The first column of the matrix contains the *Y* coordinates, and second column the *Z* coordinates. Row *is* of the matrix corresponds to the is^{th} lateral position of the mesh. [mm]

- `ang_prf_pcs`: angles of the profile of the PCS in the *YZ* plane. Column vector with `ms` elements. The angles are in the range (–π, π), and are measured counter clock-wise while looking towards the negative *X* direction, with origin in the positive *Y* axis. An angle of 0 corresponds to a horizontal tangent in which the wheel is above the rail, and an angle of –π or π to a horizontal tangent in which the wheel is below the rail. [rad]
- `N_o_d`: complementary to `d_o_N` input, cf. §4.1.2.1.1:
    - If `ndgt` = 0: normal load, *N*. It is the magnitude of the resultant of the normal pressures acting on the PCS. [kN]
    - If `ndgt` = 1: approach, *d*. [mm]
- `dir_N`: angle of *N* in the *YZ* plane. It is in the range (–π, π), and is measured positive clock-wise while looking towards the negative *X* direction, with origin in the positive *Z* axis (like the α angle shown in Figure 4.7). [rad]
- `Fx`: signed resultant force along the *x* axis due to the contact tangential stresses. [kN]
- `Flat`: modulus of the resultant force in the *YZ* plane due to the contact tangential stresses. [kN]
- `dir_Flat`: angle of `Flat` in the *YZ* plane. It is in the range (–π, π), and is measured positive counter clock-wise while looking towards the negative *X* direction, with origin in the positive *Y* axis. [rad]
- `Msp`: modulus of the resultant moment due to the contact tangential stresses about the origin of the local contact coordinate system. [N.m]
- `dir_Msp`: cosine directors of `Msp` in the global coordinate system. Vector with 3 components. [-]

The `sol`, `prf_pcs` and `ang_prf_pcs` variables are returned only for the last analyzed case of the main input file. The rest of the variables are returned as arrays, including the values for each of the analyzed cases.

4.1.2.2.1 .mat file

This is the main output file from CECT. Its structure is based on that of the .mat output file from CONTACT [CONTACT UG 2013]. It consists of three head lines that include aggregated results for the whole PCS and some of the input variables, and the main body which contains the solution arrays with AP data. Each .mat file contains data for a single case, and it is created only when input `r` (cf. line 0 in Table 4.4) = 1 for the case, and a converged solution has been achieved for the case.

The structure of the head lines is shown in Table 4.9 and Table 4.10.

Table 4.9. Structure of the head lines of the .mat output file. First 6 columns.

	1	2	3	4	5	6
1	mx	ms	xl	sl	dx	ds
2	c_o_n	gdgt	ndgt	mdgt	tdgt	G_w
3	veloc	-	N_o_d	dir_N	Fx	Flat

Table 4.10. Structure of the head lines of the .mat output file. Last 7 columns.

	7	8	9	10	11	12	13
1	chi	dq	idgt	mat_dgt	fmt	r_nom_r	-
2	G_r	nu_w	nu_r	d_o_N	dir_d	-	-
3	dir_Flat	Msp	dir_Msp_1	dir_Msp_2	dir_Msp_3	-	-

A commented line precedes each head line, with the identifiers of the data included in the line and their units. The content of each head line is described next. Most of the variables have been already introduced in §4.1.2.1.1 and beginning of §4.1.2.2. Here only not previously introduced variables are described. Zeros are printed in the positions without identifier in Table 4.9 and Table 4.10.

- 1st line. It includes the contact mesh description.
    - `xl`, `sl`: (x, s) coordinates of the 1st AP of the contact mesh. For `c_o_n` = 1, they are equal to inputs `x_inf` and `s_inf`, respectively (cf. §4.1.2.1.1). [mm]
    - `dx`, `ds`: size of each element of the mesh in x and s directions. [mm]
    - `chi`: currently not used. 0 is printed.

- `idgt`: `ics`, if `ics` is integer. Else, –1.
- `fmt`: integer specifying the format of the .mat file. For use by loader function.

- 2nd line. It includes some input digits and elastic constants. Cf. §4.1.2.1.1.
- 3rd line. It includes resultant forces and moments.

The AP variables listed in Table 4.11 are printed in the main body of the file. Only results for the *NEC* APs falling inside the contact (`eldiv` = 1 or 2) are printed. Each line contains the data corresponding to one AP. The variable `I` printed in the 1st column is the index of the AP (1D numbering, according to Figure 2.19). The rest of the data is as explained at the beginning of §4.1.2.2 under the solution arrays of the `sol` output variable (except that the arrays in `sol` include the data for all APs of the mesh, not only for the ones inside the contact). A commented line preceding the main body is printed as well, with the identifier and units of the data in each column.

Table 4.11. Content of the main body of the .mat output file.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1 ... NEC	I	eldiv	h	mu	pn	px	ps	un	ux	us	srel	wx	ws

4.1.2.2.2 .prf file

The .prf file contains the profile of the PCS. The printed data lines are arranged in a `ms`×3 matrix. The first two columns correspond to the `prf_pcs` variable, and the 3rd column to the `ang_prf_pcs` variable; both of them described at the beginning of §4.1.2.2.

CECT prints a new .prf file for a case when the following two conditions are met, and in addition a converged solution has been achieved for the case:

- Input bit `r` = 1 in the current case AND
- `gdgt` = 1 OR `mdgt` = 1 (i.e. either the geometry or the mesh have changed) in any case since the one after the previous .prf file was printed to the current one.

4.1.2.2.3 .log file

This file is created as soon as the main input file is read, and is printed along the execution of the cases included in the main input file. A single .log file is created for a main input file, including information of the execution of all cases read from the input file.

Information messages related to the main steps of the execution of each case are printed in the file, together with elapsed times. In addition, different warning and error messages are printed.

Information about the following steps is included:

- Loading of data from the main input file.
- Calculation of normal undeformed distances.
- Calculation of IC matrices (except for tangential contact).
- Normal problem solution. Information for each NORM iteration: number of elements 1) in contact, 2) changing from exterior to interior and 3) changing from interior to exterior.
- Calculation of IC matrix of current time instant for tangential contact.
- Tangential problem solution. Information for each TANG iteration: number of elements in slip, error norms (slip, adhesion limit, slip direction) and corresponding tolerances.
- Panagiotopoulos loop. Information for each iteration: error norms and corresponding tolerances.
- Completion of each case. Number of APs in contact, slipping and total.

In addition, the head lines of the .mat file (cf. Table 4.9 and Table 4.10) are printed at the end of each analysis case, regardless of the value of `r`.

These are some of the cases in which error messages are printed. CECT aborts execution when an error message is printed.

- Some necessary input file not found.
- Incorrect type/value of some input parameter.

- `ics` = 20 to 22 (cylindrical ICs) used with some non-allowed `c_o_n`, `mat_dgt`, `nx` or `tdgt` value, or with incorrectly defined wheel and rail profiles (cf. §4.1.2.1.1).
- Data of input files of ICs not matching data in main input file or in other input files of ICs (cf. §4.1.2.1.2).
- Incorrect definition of wheel or rail profile (with a single point; with two adjacent coincident points).
- Some AP out of the range of the wheel or rail profile in *s* direction.
- Negative contact velocity.

These are some of the cases in which warning messages are printed. CECT continues running when a warning message is printed, and a solution may be achieved, but the solution may not be valid and it may be necessary to revise some input data.

- `dq` not equal to `dx` (if `tdgt` = 2).
- Shear modulus in input file of ICs different from shear modulus in main input file.
- `bit_l6` = 1 with `tdgt` = 0. Data of 6th line of the case (cf. Table 4.4) is ignored.
- Wheel or rail profile far from circular. This is applicable for `ics` = 20 to 22 (cylindrical ICs), and is printed when the difference between the calculated maximum and minimum profile radii exceeds 3% of the calculated best-fit profile radius.
- Absolute value of minimum normal undeformed distance bigger than 1 μm.
- Distance between the profiles of the PCS calculated with the input wheel and rail profiles and with the data in the files of ICs bigger than 1 μm.
- Maximum number of NORM iterations reached.
- Not possible to use SteadyGS tangential solver because there is some element inside contact in the trailing edge of the mesh of the PCS.

- Within the tangential solver:
    - Maximum number of Newton-Raphson iterations reached in the solution of the non-linear equations for an AP in slip.
    - Maximum number of outer iterations reached.
- Maximum number of iterations reached in Panagiotopoulos loop. Restarting with slight increase in `d_o_N`.
- Panagiotopoulos process appears to be diverging. Restarting loop with slow method.
- Tangential problem not converged in current Panagiotopoulos iteration. Restarting with slight increase in `d_o_N`.
- Contact border reached some border of the domain of PCS. Consider if it is necessary to extend the domain. This is applicable if the number of APs in the concerned mesh direction is greater than 2.

4.1.3 Program structure

Table 4.12 outlines the program execution pipeline, with the main steps of the solution process of the contact problem. Related sections of this chapter are also shown for some of the steps.

Table 4.12. Main steps of the solution process of the contact problem in CECT.

No.	Step	Reference	Notes
1	Read input	§4.1.2.1.1	
2	Load `a_IC` arrays from file	§4.1.2.1.2	(1)
	For each case `c`:		
3	With cylindrical ICs, calculate transverse profile radii		(2)
4	Construct rectangular contact mesh	§4.1.1.1	(3)
5	Calculate h	§4.1.1.2	(4)
6	Load μ data	§4.1.2.1.4	(5)
7	Calculate IC_{ij}-s	§4.1.1.4, §5	(6)
8	Update **W'**	§4.1.1.3	(7)
9	If `mdgt(c)`=1: map results of case `c-1` to current mesh	§4.1.1.4.3	(8)
10	Calculate IC'_{ij}-s	§4.1.1.4.1	(9)
11	Calculate $\mathbf{u_t'}^*$	§4.1.4.2	(10)
12	Calculate V_c, Δt, **{Δq}**	§7.2.2.1.2	(11)
13	Panagiotopoulos loop	§4.1.4.3	
14	Organize output arrays and print output files	§4.1.2.2	
15	Update values of previous instant for next case		(12)

Notes to Table 4.12:

(1) Applicable if `ics` is a string.

(2) Applicable for `ics` = 20 to 22, in each case in which `gdgt(c)` > 0.9. The transverse profile radii are used to calculate the ICs. These are the main steps followed to calculate the transverse radius of a profile given by a set of (y_L, z_L) points:

   a. Calculate the intersection c_0 between the normals to the lines l_a, l_b formed by the first and mid point, and by the mid and last point of the profile, passing through the mid-point of each line. This is a first estimate of the centre of the circumference *cir* which best approaches the given profile. If both normals are parallel (which should not happen for a circular profile), the mid point of the line joining the mid points of lines l_a, l_b is calculated instead.

   b. Calculate the centre c of *cir*. This is done searching for c as the centre of a polar coordinate system, with respect to which the difference between the maximum and the minimum radial

coordinates R_i of the points i of the given profile is minimized. This is a two-variable unconstrained minimization problem, where the two variables are the coordinates of c. The previously calculated c_0 point is used as starting point.

c. The radius R of *cir* is calculated as the value with which the rms of the differences $(R - R_i)$ is minimized.

(3) If `mdgt(c)` > 0.9.

(4) If `mdgt(c)` > 0.9 OR `gdgt(c)` > 0.9. min($|h_l|$) should be near zero. If min($|h_l|$) > 1 μm, a warning message is printed.

(5) If `f` is a string.

(6) If `mdgt(c)` > 0.9 OR `gdgt(c)` > 0.9 OR `c` < 1.1. If `ics` is not a string, `v_IC` vectors are calculated for each IC. Matrices of IC_{xn}, IC_{sn}, IC_{nx}, IC_{ns} and IC_{nn} ICs are formed.

(7) The total rigid shift of the previous time instant, **W'**, is assigned the value of **W** of the previous case (or zero if `c`=1). This is relevant only if `tdgt(c)`=0. The quantities passed to the tangential solver are $\Delta\mathbf{W} = \mathbf{W} - \mathbf{W}'$ if `tdgt(c)`=0, and $\Delta\mathbf{W} = \mathbf{w}\,\Delta t$ otherwise.

(8) If `mdgt(c)` > 0.9 AND `c` > 1.1.

(9) Calculation of IC'_{ij}-s.

a. `If mdgt(c)<0.1 AND (NOT(dq(c)*tdgt(c) = dq(c_o)*tdgt(c_o)) OR mdgt(c_o)>0.9 OR tdgt(c_o)>1.9) AND tdgt(c)<1.9`: calculation of IC'_{xx}, IC'_{xs}, IC'_{sx}, IC'_{ss}, IC'_{xn} and IC'_{sn} ICs.

b. `If tdgt(c)>1.9 AND (mdgt(c)>0.9 OR gdgt(c)>0.9 OR NOT(dq(c)=dq(c_o)) OR tdgt(c_o)<1.9 OR c<1.1)`: calculation of IC'_{xn} and IC'_{sn} ICs.

Note: `c_o` = max(`c`−1, 1).

(10) Calculation of elastic displacements of previous instant $\mathbf{u_t}'^*$ in all the APs of the current mesh. If `tdgt(c)<1.9 AND c>1.1`. Otherwise, $\mathbf{u_t}'^* = \mathbf{0}$ in this step, and a different $\mathbf{u_t}'^*$, with the influence of only the current

normal pressures (cf. Table 4.13), is calculated inside the Panagiotopoulos loop.

(11) $\Delta t = 1$ if `tdgt(c)`<0.1, and $\Delta t =$ `dq` / $V_{c,o}$ otherwise. $V_{c,o}$ is the contact velocity V_c at s=0. The calculation of **{Δq}** is relevant for large rolling radii variations in the contact patch, cf. §7.2.2.1.2. `If |r_nom_r| > 0 AND tdgt(c)=1 AND c>1.1`: u'^{*}_{Ii} values are interpolated at points $x_I + \Delta q_I(s)$, from available values at points $x_I + \Delta q_o$, as explained in §7.2.2.1.2.

(12) Previous instant quantities for the next case (`c+1`) are updated with the values computed in the current case `c`. The updated previous instant values include the contact stresses, the ***C*** set, the slopes (α angles) of the PCS, and the longitudinal and lateral mesh positions.

4.1.4 Solution algorithms

In this section, some details are given about the contact solution algorithms currently implemented in CECT. Most of the material discussed here applies similarly to planar and to non-planar contact, but some of the equations have to be adapted for non-planar contact.

4.1.4.1 Normal contact

As described in §1.2, the solution of the normal problem consists in finding the set of APs in contact ***C*** and the distribution of normal pressures p_{In} in it, subject to the normal contact restrictions:

$$p_{In} > 0,\ e_I = 0 \text{ if } I \in \boldsymbol{C} \tag{4.32}$$

$$p_{In} = 0,\ e_I > 0 \text{ if } I \notin \boldsymbol{C} \tag{4.33}$$

The normal contact solver is roughly the same as the NORM solver described in [Kalker 2001]. Figure 4.28 shows the main steps of the algorithm.

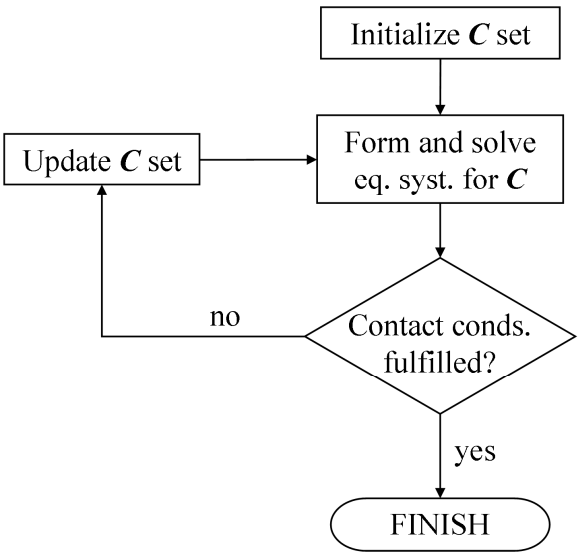


Figure 4.28. Flowchart of normal solver.

Each step of the algorithm is detailed next.

1. Initialization of the set of APs in contact.

   A good initialization serves to reduce the number of iterations and hence the time required to solve the normal problem. The following situations are distinguished:

   - `c > 1.1 AND ite_p < 1.1 AND gdgt(c) = 0 AND mdgt(c) = 0 AND ndgt(c) = ndgt(c_o) AND d_o_N(c) = d_o_N(c_o) AND dir_d(c) = dir_d(c_o)`

     This condition states that the first normal contact solution is being carried out for the current case `c` (i.e. the first Panagiotopoulos iteration `ite_p` is running, cf. §4.1.4.3), and the geometry, mesh and input data for the normal problem in the current case `c` are the same as in the previous case `c_o`.

     In this situation, the starting ***C*** set is defined as the final ***C*** set obtained in the previous case.

   - `ite_p > 1.1`

The starting ***C*** set is defined as the one obtained in the previous Panagiotopoulos iteration `ite_p-1`.

- Otherwise:

  The overlap between the undeformed contacting surfaces is calculated, considering a scaled approach d_{sc}. That is, an undeformed distance $h'_{I,sc}$ is calculated for each of the $I \in \{1, NE\}$ APs, in a virtual reference configuration with an approach d_{sc}, as:

$$h'_{I,sc} = h^*_I - (d_{sc} + h^*_{I,m} / \cos(\Delta\alpha_{I,m\text{-}d})) \cos(\Delta\alpha_{I\text{-}d}) \tag{4.34}$$

  h^*_I is the modified initial undeformed distance that incorporates the effect of the tangential stresses; cf. Figure 2.21. The tangential stresses are prescribed (fixed) during the solution of the normal problem. $\Delta\alpha_{I\text{-}d}$ is as defined in Figure 4.12, cf. also Eq. (4.15). The term $h^*_{I,m} / \cos(\Delta\alpha_{I,m\text{-}d})$ accompanying d_{sc} in Eq. (4.34) is meant to cancel out the minimum h^*_I, $h^*_{I,m}$, if it is non-zero. The AP with $h^*_I = h^*_{I,m}$ is located at coordinates $(x, s) = (x_{I,m}, s_{I,m})$, and $\Delta\alpha_{I,m\text{-}d}$ is the particular value of $\Delta\alpha_{I\text{-}d}$ for that AP.

  The starting ***C*** set is formed with the APs in which $h'_{I,sc}$ is negative, i.e. where there is overlap. This is a simplified approach of the virtual penetration methods for the normal contact problem discussed in §2.1.1.

  d_{sc} is calculated as follows:

  - If `ndgt` = 0:

    d_{sc} = `r_d_ini` × `d_o_N`

  - If `ndgt` = 1:

    The quadratic undeformed distance which best fits $h^*_I - h^*_{I,m} \cos(\Delta\alpha_{I\text{-}I,m})$ is calculated, being $\Delta\alpha_{I\text{-}I,m}$ the α angle between each AP I and the AP with $h^*_I = h^*_{I,m}$. For this purpose, the profiles of $h^*_I - h^*_{I,m} \cos(\Delta\alpha_{I\text{-}I,m})$ in x and s directions, and passing through (x_{Im}, s_{Im}), are extracted, and the best fitting parabola for each profile is found. The coefficients of these parabolas are designated as A_{Hz}

and B_{Hz}. A quadratic undeformed distance h^*_{Hz} is assumed with these coefficients, as:

$h^*_{Hz} = A_{Hz} (x - x_{I,m})^2 + B_{Hz} (s - s_{I,m})^2$

The Hertzian approach d_{Hz} corresponding to h^*_{Hz}, the input load `d_o_N`, and material elastic constants is calculated. Finally, the scaled approach d_{sc} is calculated as:

$d_{sc} =$ `r_d_ini` $\times d_{Hz}$

2. Set-up and solution of the system of equations for APs in ***C***.

The system of equations is based on the condition $e_I = 0$ for the *NEC* APs I in ***C*** ($I = c_1, c_2, \ldots, c_{NEC}$), cf. Eq. (4.32). This leads to the following equation for each AP in ***C***:

$$e_I = h^*_I - d\cos(\Delta\alpha_{I-d}) + u_{In} = 0 \tag{4.35}$$

u_{In} is calculated summing up the contributions of all elements in ***C***, as in planar contact:

$$u_{In} = \sum_{J \in \boldsymbol{C}} IC_{InJn}\, p_{Jn} \tag{4.36}$$

Eqs. (4.35) and (4.36) define *NEC* linear equations. The unknowns in these equations are the *NEC* p_{Jn} pressures for all APs in ***C***, and also the approach d for `ndgt` = 1. Thus the resulting system of equations, of the form **[A]{x}** = **{b}**, changes slightly for each `ndgt` value:

- For `ndgt` = 0:

  The matrix **[A]** of the system is the part of the complete $\mathbf{[IC_{nn}]}$ matrix corresponding to the APs in ***C***, with dimension *NEC*×*NEC*. From Eq. (4.35), the independent vector **{b}** is made with the quantities $d\cos(\Delta\alpha_{I\text{-}d}) - h^*_I$ for each AP in ***C***.

- For `ndgt` = 1:

  An additional equation is needed to complete the system of equations with the additional unknown d. This is given by the condition of

normal force equilibrium along the direction defined by `dir_d`. The integral of the projection of normal pressures along that direction must equal the given input normal force projection `d_o_N` = *N*. This leads to Eq. (4.37), where A_E is the element area (equal to *dx*×*ds*):

$$\sum_{J\in\boldsymbol{C}} p_{Jn}\cos\left(\Delta\alpha_{I-d}\right) = N / A_E \tag{4.37}$$

Eqs. (4.35)–(4.37) result in the following linear system of equations:

$$\begin{bmatrix} IC_{c_1c_1} & \dots & IC_{c_1c_{NEC}} & -\cos\left(\Delta\alpha_{c_1-d}\right) \\ \dots & \dots & \dots & \dots \\ IC_{c_{NEC}c_1} & \dots & IC_{c_{NEC}c_{NEC}} & -\cos\left(\Delta\alpha_{c_{NEC}-d}\right) \\ -\cos\left(\Delta\alpha_{c_1-d}\right) & \dots & -\cos\left(\Delta\alpha_{c_{NEC}-d}\right) & 0 \end{bmatrix}$$

$$\times\begin{Bmatrix} p_{c_1} \\ \dots \\ p_{c_{NEC}} \\ d \end{Bmatrix} = \begin{Bmatrix} -h^*_{c_1} \\ \dots \\ -h^*_{c_{NEC}} \\ -N / A_E \end{Bmatrix} \tag{4.38}$$

The subindex *n* is omitted in Eq. (4.38) for brevity. The matrix of this system is of (*NEC*+1) × (*NEC*+1) dimension. The sign of the last equation is reversed, so that the matrix is kept symmetric, as with `ndgt` = 0. The *NEC*×*NEC* first elements of the matrix are the same as the matrix for `ndgt` = 0.

In the current implementation in CECT, the entire system of equations is formed explicitly, and solved with the MATLAB mldivide operator (\), both for `ndgt` = 0 and 1.

3. Check of contact conditions.

   Both inequalities of Eqs. (4.32) and (4.33) are checked for the solution obtained in the preceding step, i.e. $p_n > 0$ for all APs in $\boldsymbol{C}$ (the contact pressures are compressive), and $e > 0$ for all APs outside $\boldsymbol{C}$ (there is no interpenetration of the deformed surfaces). As a difference with the

algorithm described in [Kalker 2001], both conditions are checked simultaneously, and the ***C*** set is updated accordingly in the same iteration. The APs inside ***C*** in which the $p_n > 0$ condition is not fulfilled are removed from ***C***, and the elements outside ***C*** in which the $e > 0$ condition is not fulfilled are included in ***C*** for the next iteration.

If there has been any change in the ***C*** set, the previous step 2 and this one are repeated with the new ***C*** set. Otherwise, the solution of the normal problem is finished. In the latter case, the modulus of the resultant normal force and its direction in the *YZ* plane `dir_N` are computed in a final step, through the vector sum of the contributions of all elements.

Profiling of the version of the NORM algorithm described here shows that most of the time is spent in the solution of the linear system of equations in step 2 above. With this solver, it is more efficient in terms of computation time to have the complete $\mathbf{[IC_{nn}]}$ matrix (of *NE*×*NE* dimension) from the beginning than to work with the `a_IC` array (of *NE*×*ns* dimension, cf. §4.1.1.4). In the latter case, considerably more effort is needed to form the matrix of the system of equations in each iteration.

In contrast with non-conformal contact, the $\mathbf{[IC_{nn}]}$ matrix is not necessarily positive. This is not favourable with regard to the solution of the normal problem. Nevertheless, the matrix is of dominant diagonal, and no convergence difficulties have been experienced in the solution of the normal part of the contact problems encountered during the elaboration of this thesis. The situation is more complicated for the tangential part. Details about how the IC_{nn} ICs change in non-planar surfaces are given in §5.2.3.6.1 and other parts of Chapter 5.

4.1.4.2 Tangential contact

As described in §1.2, the solution of the tangential problem consists in finding the division of APs in ***C*** into adhesion ***A*** and slip ***D***, and the distribution of tangential stresses $\mathbf{p_{It}} = (p_{Ix}, p_{Is})$ in ***C***, subject to the tangential contact restrictions:

$$\|\mathbf{p_{It}}\| \le g_I,\ \mathbf{S_{It}} = 0 \text{ if } I \in \boldsymbol{A} \tag{4.39}$$

$$\|\mathbf{p}_{\mathbf{It}}\| = g_I,\ \mathbf{S}_{\mathbf{It}} = -\lambda_I\, \mathbf{p}_{\mathbf{It}}\ (\lambda_I > 0) \text{ if } I \in \boldsymbol{D} \tag{4.40}$$

In these equations, the slip velocity **v** used in §1.2 is replaced with the slipped distance $\mathbf{S_t} = (S_x, S_s)$ in a time increment Δt; $\mathbf{S_t} = \mathbf{v}\,\Delta t$; with the aim of obtaining a common formulation valid for the different types of tangential problems studied (static, transient rolling and steady rolling). The equations are developed as follows.

For the APs in adhesion (***A***):

$$S_{Ii} = \Delta W_{Ii} + u_{Ii} - u'_{Ii} = 0, \quad i \in \{x, s\} \tag{4.41}$$

For the APs in slip (***D***):

$$g_I{}^2 - p_{Ix}{}^2 - p_{Is}{}^2 = 0 \tag{4.42}$$

$$p_{Ix}\, S_{Is} - p_{Is}\, S_{Ix} = 0 \tag{4.43}$$

ΔW_{Ii} in Eq. (4.41) is the increment of the rigid slip in the time increment $\Delta t = t - t'$ from the previous to the current time instant. In rolling contact problems, it is obtained multiplying the rigid slip velocity w_{Ii} by Δt. Eqs. (4.21) and (4.22) provide the expressions for w_{Ii}, together with the additional component $-\Delta d \times \sin(\Delta\alpha_{I\text{-}d})$ for transient problems given next to them. The rigid slip W_{Ii} for static problems is defined in Eqs. (4.23) and (4.24).

The variations of the tangential elastic displacement differences $u_{Ii} - u'_{Ii}$ in Eq. (4.41) are computed according to Eq. (4.44). The meaning of the terms IC^*_{IiJj} and u'^*_{Ii} of this equation is different for the different types of tangential problems, as specified in Table 4.13. In all cases, these terms are fixed and known at the beginning of the solution of the tangential problem.

$$u_{Ii} - u'_{Ii} = \sum_{j=x,s} \sum_{J \in \boldsymbol{C}} IC^*_{IiJj}\, p_{Jj} + \sum_{J \in \boldsymbol{C}} IC_{IiJn}\, p_{Jn} - u'^*_{Ii}, \quad i \in \{x, s\} \tag{4.44}$$

Table 4.13. Terms for the calculation of $u_{Ii} - u'_{Ii}$ in Eq. (4.44).

`tdgt`	IC^*_{IiJj}	u'^*_{Ii} $(i \in \{x, s\})$
0 or 1	IC_{IiJj}	$\sum_{j=x,s,n} \sum_{J \in \boldsymbol{C}} IC'_{IiJj}\ p'_{Jj}$
2	$IC_{IiJj} - IC'_{IiJj}$	$\sum_{J \in \boldsymbol{C}} IC'_{IiJn}\ p'_{Jn}$

p'_{Jj} are the contact stresses of the previous time instant; i.e. the p_{Jj} values computed in the previous case (interpolated if necessary to the current mesh; cf. §4.1.1.4.3). As indicated in Table 4.13, for `tdgt` = 2, the u'^*_{Ii} term only includes the influence of the normal pressures (and in this case, $p'_{Jn} = p_{Jn}$). This is because the tangential stresses of the previous time instant are initially unknown (form part of the solution) in this case. The IC^*_{IiJj} term for `tdgt` = 2 shown in the table does not apply for the SteadyGS solver treated in §4.1.4.2.3. In that case, the differences $u_{Ii} - u'_{Ii}$ are computed convolving ICs with differences of stresses, instead of differences of ICs with stresses, as explained in that section.

The normal pressures are prescribed (fixed) during the solution of the tangential problem. The traction bounds g_I in Eq. (4.42) are also fixed. These are defined as a prescribed coefficient of friction μ_I times the normal pressure p_{In}. Some friction models consider the dependence of the coefficient of friction on the slip velocity. Currently this is not provided for in CECT.

Eq. (4.43) is used to express the parallelism between the $\mathbf{p_{It}}$ and $\mathbf{S_{It}}$ vectors. It is used in place of the $\mathbf{S_{It}} = -\lambda_I\ \mathbf{p_{It}}$ condition in Eq. (4.40), to avoid the introduction of new unknowns λ_I and related constraints ($\lambda_I > 0$). The first equality of Eq. (4.41), together with Eq. (4.44), which apply anywhere in the contact, is used to replace the quantities S_{Ii} in Eq. (4.43). The resulting expression is given in Eq. (4.45). This is applicable only for the APs in slip, as Eq. (4.43).

$$p_{Ix}\left(\Delta W_{Is} + \sum_{j=x,s} \sum_{J \in \boldsymbol{C}} IC^*_{IsJj} p_{Jj} + \sum_{J \in \boldsymbol{C}} IC_{IsJn} p_{Jn} - u'^*_{Is}\right)$$

$$- p_{Is}\left(\Delta W_{Ix} + \sum_{j=x,s} \sum_{J \in \boldsymbol{C}} IC^*_{IxJj} p_{Jj} + \sum_{J \in \boldsymbol{C}} IC_{IxJn} p_{Jn} - u'^*_{Ix}\right) = 0 \qquad (4.45)$$

Eqs. (4.41)–(4.45) define the system of equations for the tangential part of the contact problem. They are 2×*NEC* equations with 2×*NEC* unknowns, which are the p_{Ix} and p_{Is} tangential stresses for each AP I in contact. Together with these, the inequality constraints of Eqs. (4.39) and (4.40) have to be considered, developed as follows:

$$p_{Ix}^2 + p_{Is}^2 \leq 1.001^2 \, g_I^2, \text{ if } I \in \boldsymbol{A} \tag{4.46}$$

$$p_{Ix} \, S_{Ix} + p_{Is} \, S_{Is} \leq \varepsilon_{ps}, \text{ if } I \in \boldsymbol{D} \tag{4.47}$$

In Eqs. (4.46) and (4.47) small tolerance values are introduced, instead of forcing the exact inequalities of Eqs. (4.39) and (4.40), to avoid possible chattering problems in the iterative search for the ***A*** and ***D*** sets (i.e. APs changing alternately from one set to the other in successive iterations). These are realized by the 1.001^2 factor introduced in the right-hand side of Eq. (4.46) (in place of 1.0), and the ε_{ps} value introduced in the right-hand side of Eq. (4.47) (in place of 0). ε_{ps} is assigned a fixed value equal to $0.001 \times \max(\|\mathbf{\Delta W_I}\|) \times \max(g_I)$. $\max(\|\mathbf{\Delta W_I}\|)$ and $\max(g_I)$ are the maximum $\|\mathbf{\Delta W_I}\|$ and g_I values of all APs in contact.

In the current implementation of CECT, Eqs. (4.41)–(4.47) are applied in the same set of APs where the unknowns $\mathbf{p_{It}}$ are computed. It would be possible to check the contact conditions in a different set of APs. For example, checking the adhesion condition (Eq. (4.41)) at points with longitudinal coordinates $x_I + \Delta q/2$ would be more precise than at x_I, because the increments of elastic displacement differences $u_{Ii} - u'_{Ii}$ would be calculated by central differences instead of by backward differences. I.e., $(u_{Ii} - u'_{Ii})(x_I + \Delta q/2, s_I, t) \approx u_{Ii}(x_I, s_I, t) - u_{Ii}(x_I + \Delta q, s_I, t')$ would be more precise than $(u_{Ii} - u'_{Ii})(x_I, s_I, t) \approx u_{Ii}(x_I, s_I, t) - u_{Ii}(x_I + \Delta q, s_I, t')$. However, this may be worse for the elements in slip. In the particular case of having ΔW_{Ii} much larger than $u_{Ii} - u'_{Ii}$, it would be clearly better to evaluate the parallelism condition between $\mathbf{p_{It}}$ and $\mathbf{S_{It}}$ (Eq. (4.43)) at the same positions x_I where $\mathbf{p_{It}}$ are calculated. In the case of evaluating this at points $x_I + \Delta q/2$ with an averaged value of $\mathbf{p_t}$, an incorrect trivial solution could appear. E.g. with $\Delta q = \Delta x$, the averaged $\mathbf{p_t}$ would be calculated as $\mathbf{p_t}(x_I + \Delta q/2) = (\mathbf{p_t}(x_I) + \mathbf{p_t}(x_{I+1})) / 2 = (\mathbf{p_{It}} + \mathbf{p_{I+1t}}) / 2$, where the subindex I+1 expresses the position in the mesh adjacent to I towards the positive x axis. In this case, a

solution with $\mathbf{p_{It}} = -\mathbf{p_{I+1t}}$ would lead to zero averaged $\mathbf{p_t}(x_I + \Delta q/2)$, with which the parallelism condition of Eq. (4.43) would be trivially fulfilled.

The equations for the tangential problem involve a greater difficulty than those of the normal problem, because the equations corresponding to the APs in slip are non-linear, and because the system size is doubled, as now the unknowns are vectors with (x, s) components, instead of scalars. Additionally, the matrix of ICs for the tangential part, including the IC_{xx}, IC_{ss}, IC_{xs} and IC_{sx} ICs, is less favourable for convergence in the iterative search for the correct division of $\boldsymbol{C}$ into the $\boldsymbol{A}$ and $\boldsymbol{D}$ sets than the matrix of IC_{nn} ICs for the normal problem. The IC_{xs} and IC_{sx} crossed ICs introduce more coupling and complexity in the relationship between the fields of contact surface stresses and elastic displacements. The ICs have changing signs, and a less local characteristic than the direct IC_{ii} ($i \in \{x, s, n\}$) ICs. On the other hand, they are considerably smaller in magnitude. The $IC_{IiJj} - IC'_{IiJj}$ IC differences arising in steady rolling problems involve further difficulties, as explained in [Vollebregt 1995]. This motivated the development of the SteadyGS solver [Vollebregt 2010], treated here in §4.1.4.2.3.

The non-linear set of equations for the tangential problem defined above are solved with a Newton-Raphson procedure. The Jacobian of these equations is needed for this purpose. Next the necessary expressions are given, derived analytically from the preceding equations. The equations for an AP I are designated as follows:

- $F_{I1} \equiv$ Eq. (4.41) with $i = x$
- $F_{I2} \equiv$ Eq. (4.41) with $i = s$
- $F_{I3} \equiv$ Eq. (4.42)
- $F_{I4} \equiv$ Eq. (4.43) (developed to Eq. (4.45))

$$\frac{\partial F_{I1}}{\partial p_{Jj}} = IC^{*}_{IxJj} \tag{4.48}$$

$$\frac{\partial F_{I2}}{\partial p_{Jj}} = IC^{*}_{IsJj} \tag{4.49}$$

$$\frac{\partial F_{I3}}{\partial p_{Jj}} = -2\,p_{Jj}\,, \quad \text{for } J = I;\ 0 \text{ otherwise} \tag{4.50}$$

$$\frac{\partial F_{I4}}{\partial p_{Jj}} = p_{Ix} IC^{*}_{IsJj} - p_{Is} IC^{*}_{IxJj}, \quad \text{for } J \neq I \tag{4.51}$$

$$\frac{\partial F_{I4}}{\partial p_{Ix}} = \Delta W_{Is} + \sum_{J \in \boldsymbol{C}} IC_{IsJn}\, p_{Jn} - u'^{*}_{Is} + IC^{*}_{IsIx}\, p_{Ix} + \sum_{j=x,s} \sum_{J \in \boldsymbol{C}} IC^{*}_{IsJj}\, p_{Jj} - IC^{*}_{IxIx}\, p_{Is} \tag{4.52}$$

$$\frac{\partial F_{I4}}{\partial p_{Is}} = -\Delta W_{Ix} - \sum_{J \in \boldsymbol{C}} IC_{IxJn}\, p_{Jn} + u'^{*}_{Ix} - IC^{*}_{IxIs}\, p_{Is} - \sum_{j=x,s} \sum_{J \in \boldsymbol{C}} IC^{*}_{IxJj}\, p_{Jj} + IC^{*}_{IsIs}\, p_{Ix} \tag{4.53}$$

The following subsections describe the different tangential solvers implemented in CECT. The first tangential solver implemented, described in §4.1.4.2.1, has been found to be quite ineffective in conformal contact problems, failing to converge not infrequently. The failure rate in conformal problems has been observed to be considerably higher than in non-conformal problems. It is believed that the more rapid variations in the direction of the slip velocities encountered in conformal contact, coming from the variable rolling radii and geometric spin, is a factor contributing to the increased convergence difficulties in the tangential part. As a result of these difficulties, this first solver was abandoned soon, in favour of the more robust algorithms ConvexGS and SteadyGS, described in §4.1.4.2.2 and §4.1.4.2.3.

The biggest difficulties have been experienced in the solution of frictional contact problems in rolling bearings (Chapter 7), where the changes in the slip velocities are more abrupt (particularly around the zero slip lines), as a result of the bigger relative variations of rolling radii. New difficulties were found in the tangential problem when frictional contact problems in rolling bearings where tackled with the initial versions of the ConvexGS and SteadyGS algorithms

implemented in CECT, and also in the outer iteration for the coupled normal-tangential problem (§4.1.4.3). The initial versions of these algorithms, which worked relatively well in the wheel–rail application, were found to be less effective for contact in rolling bearings. This promoted the improvement of the initial versions of the code, and currently most of these difficulties have been overcome, at least in the type of smooth contact problems treated in this thesis. Relevant features of the algorithms that provide for better robustness are described in the corresponding sections.

These are common aspects for all solvers:

- Initialization of ***A*** and ***D*** sets and $\mathbf{\{p_t\}}$.

  As for the normal solver, an initial estimate of the ***A*** and ***D*** sets is passed to the tangential solver, with the aim of speeding up the iterative search of the correct ***A*** and ***D*** sets. In this case, also initial values of the tangential stresses are provided. The tangential stresses are gathered in the vector $\mathbf{\{p_t\}}$, which has $2\times NEC$ elements (the tangential stress components p_{Ix} and p_{Is} for each AP in ***C***). The initial estimates are based on the solution of the previous calculated tangential problem.

  If no tangential problem has been previously calculated (i.e. if the case being run is the first one, the Panagiotopoulos iteration is the first one, and the Panagiotopoulos loop has not been restarted), $\mathbf{\{p_t\}}$ is inizialized as **{0}**, and all the APs are assumed to be initially in adhesion. It has to be noted that it is necessary to place in the ***A*** set all I APs with $\mathbf{p_{It}} = \mathbf{\{0\}}$, because otherwise the Jacobian would be singular due to Eq. (4.50).

  If a previous solution for the tangential problem is available:

  - $\mathbf{\{p_t\}}$ is initialized with the previous available value, scaled down element-wise if necessary to ensure that $\|\mathbf{p_{It}}\| \le g_I$ for all APs.
  - The ***A*** and ***D*** sets are initialized with the previous available solution, updated if necessary to account for possible changes in the traction bound:
    - The APs for which $\mathbf{p_{It}}$ has been scaled down so that $\|\mathbf{p_{It}}\| = g_I$, are placed in the initial ***D*** set.

- The APs for which $\|\mathbf{p}_{It}\| \leq 0.90\, g_I$, are placed in the initial $\boldsymbol{A}$ set.

- Maximum number of iterations.

  There are two iterative loops in the solvers: an inner loop for the solution of the non-linear systems of equations (Newton-Raphson), and an outer loop for the search of the division of APs into the $\boldsymbol{A}$ and $\boldsymbol{D}$ sets. The parameters *maxite_nr* and *maxite_t* are defined, limiting the maximum number of iterations in inner and outer loops, respectively. Their values are fixed at *maxite_nr* = 25 and *maxite_t* = 200 for the TANG, ConvexGS and SteadyGS solvers.

- Tolerances for the whole contact equality constraints.

  A $\{\mathbf{p}_t\}$ vector is the solution to a given tangential problem if the following conditions are fulfilled, in addition to the inequality constraints of Eqs. (4.46) and (4.47):

$$F_{I1}(\{\mathbf{p}_t\}) = 0,\ F_{I2}(\{\mathbf{p}_t\}) = 0, \quad \forall I \in \boldsymbol{A} \tag{4.54}$$

$$F_{I3}(\{\mathbf{p}_t\}) = 0,\ F_{I4}(\{\mathbf{p}_t\}) = 0, \quad \forall I \in \boldsymbol{D} \tag{4.55}$$

  In practice, numerically exact solutions are not achieved for the non-linear systems of equations that arise when there is some AP in slip. The preceding "hard" solution conditions (4.54)–(4.55) are replaced by the following "weak" ones:

$$\text{norm}(\{\mathbf{F}_{1,2}(\{\mathbf{p}_t\})\}_{\boldsymbol{A}}) < tol_s \tag{4.56}$$

$$\text{norm}(\{\mathbf{F}_3(\{\mathbf{p}_t\})\}_{\boldsymbol{D}}) < tol_g \tag{4.57}$$

$$\text{norm}(\{\mathbf{F}_4(\{\mathbf{p}_t\})\}_{\boldsymbol{D}} \,./\, \{\|\mathbf{p}_{It}\|\}_{\boldsymbol{D}}) < tol_{ds} \tag{4.58}$$

  The norm(·) function in Eqs. (4.56)–(4.58) is the Euclidean norm. The vector $\{\mathbf{F}_{1,2}\}_{\boldsymbol{A}}$ contains the values F_{I1} and F_{I2} of all the APs in $\boldsymbol{A}$, and the vectors $\{\mathbf{F}_3\}_{\boldsymbol{D}}$ and $\{\mathbf{F}_4\}_{\boldsymbol{D}}$ contain, respectively, the values F_{I3} and F_{I4} of all the APs in $\boldsymbol{D}$. The ./ operator in Eq. (4.58) stands for the element-wise division of two vectors. tol_s, tol_g and tol_{ds} are different tolerance limits, which are defined as follows:

$$tol_s = \varepsilon_s \cdot \max\left(\left\|\mathbf{\Delta W_I}\right\|\right) \cdot \sqrt{NEC} \tag{4.59}$$

$$tol_g = 2 \cdot \varepsilon_g \cdot \max\left(g_I\right)^2 \cdot \sqrt{NEC} \tag{4.60}$$

$$tol_{ds} = \varepsilon_{ds} \cdot \max\left(\left\|\mathbf{\Delta W_I}\right\|\right) \cdot \sqrt{NEC} \tag{4.61}$$

The factors ε_s, ε_g and ε_{ds} in Eqs. (4.59)–(4.61) currently have the following fixed values: ε_s = 5.0E–5; ε_g = 1.0E–4; ε_{ds} = 5.0E–4.

The error metrics and tolerance criteria of Eqs. (4.56)–(4.61), as defined in these equations, apply to the TANG solver, treated in §4.1.4.2.1. In this solver, each equation system has 2×*NEC* equations, corresponding to the *NEC* APs in ***C***. In the ConvexGS and SteadyGS solvers treated in §4.1.4.2.2 and §4.1.4.2.3, each equation system corresponds to a single AP. So, each equation system has 2 equations. In this case, the error metric of Eq. (4.56) and associated tolerance tol_s, corresponding to APs in ***A***, disappears, because the equations for the APs in ***A*** are linear and are solved in a numerically exact way; they do not require tolerances. And the error metrics of Eqs. (4.57) and (4.58), for the APs in ***D***, now are calculated on a single equation at a time. So the norm(·) operator in them amounts to the absolute value of each equation. The tolerance criteria in Eqs. (4.59)–(4.61) are applied similarly in the different solvers, replacing the $\sqrt{NEC}$ factor by 1 for the ConvexGS and SteadyGS solvers.

- Relaxation of Newton-Raphson (N-R) tolerances in the solution of non-linear equations in the first iterations for the search of the correct ***A*** and ***D*** sets.

  In the iterative search for the correct division of APs in contact into the ***A*** and ***D*** sets, it is likely that several iterations will be needed before finding the correct ***A*** and ***D*** sets. Changing the ***A*** and ***D*** sets implies building and solving a new system of equations. Taking this into account, variable tolerances are considered in the N-R solution procedure of the non-linear equations. The idea is to use more relaxed tolerances in the first iterations, with the aim of spending less effort in solving non-linear equations which are unlikely to be the correct ones. But this relaxation of

tolerances should not be too strong, because too degraded solutions of the equations could hinder the search for the ***A*** and ***D*** sets and even cause divergence in this search.

Here the relaxation of tolerances is implemented through a factor *tf* by which all tolerance limits (Eqs. (4.59)–(4.61)) are equally scaled, and which is varied in each iteration as a function of some metric related to the change of the solution between the last two iterations. The change metric used for this purpose in the different solvers is as follows:

- In the TANG solver is the ratio (*neid* + *neqd*) / *NEC*, *neid* being the number of APs changed from set ***A*** to ***D*** (the ones which do not comply with the inequality of Eq. (4.46)), and *neqd* the number of APs changed from ***D*** to ***A*** (the ones which do not comply with the inequality in Eq. (4.47)) in the last iteration.
- In the ConvexGS and SteadyGS solvers it is the ratio $\|\mathbf{\{p_t\}}^{(n)} - \mathbf{\{p_t\}}^{(n-1)}\| / \max(1.0E{-}12, \|\mathbf{\{p_t\}}^{(n-1)}\|)$, $\mathbf{\{p_t\}}^{(n)}$ and $\mathbf{\{p_t\}}^{(n-1)}$ being the tangential stress vectors computed for the whole ***C*** set in the current (n) and the previous (n–1) iteration.

tf varies with the change metric according to a linear function with lower and upper bounds, as shown in Figure 4.29. The parameters of this function, depicted in the figure, are fixed at the following values: *min_cm* = 0.01; *max_cm* = 0.2; *max_tf* = 10 for ConvexGS and SteadyGS solvers and *max_tf* = 1.0 for TANG solver (i.e. no tolerance scaling is currently applied in the TANG solver). In any case, a necessary condition to validate a solution for the tangential problem is that *tf* = 1.0 in the last iteration.

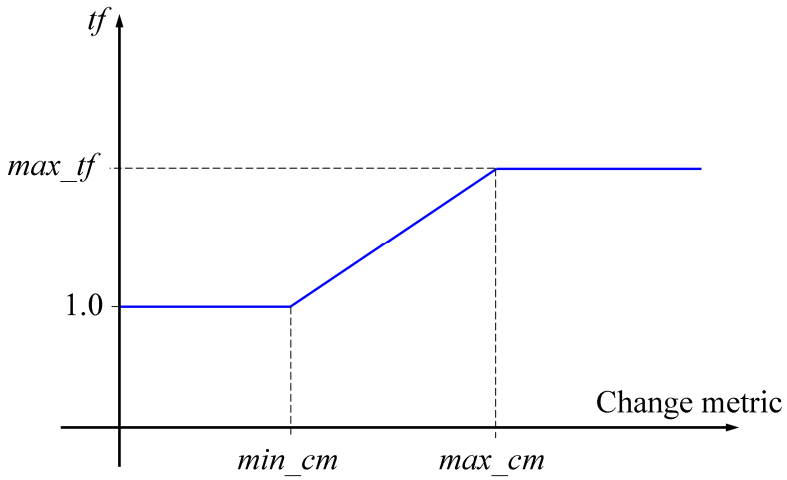


Figure 4.29. *tf* tolerance scaling factor for the Newton-Raphson solution of the non-linear equations of the tangential problem.

After a solution to the tangential problem is achieved, obtaining the tangential stress field $\{\mathbf{p_t}\}$, in a final step the tangential resultant forces and moment and elastic displacement differences are computed. The basic formulas for the resultant forces and moment are given next. $F_{lat,Y}$ and $F_{lat,Z}$ are the components in the Y and Z axes of the resultant `Flat` force, and $M_{sp,X}$, $M_{sp,Y}$ and $M_{sp,Z}$ the components in the global axes of the resultant moment `Msp`. `Fx`, `Flat` and `Msp` are defined in §4.1.2.2. x_I, Y_I, Z_I are the coordinates of each AP I in the contact x axis and global Y and Z axes, respectively; and Y_0, Z_0 are the Y and Z coordinates of the origin of the contact reference system.

$$F_x = A_E \sum_I p_{Ix} \tag{4.62}$$

$$F_{lat,Y} = A_E \sum_I p_{Is} \cos\alpha_I \tag{4.63}$$

$$F_{lat,Z} = A_E \sum_I p_{Is} \sin\alpha_I \tag{4.64}$$

$$M_{sp,X} = A_E \sum_I \left(p_{Is} \sin\alpha_I \left(Y_I - Y_0\right) - p_{Is} \cos\alpha_I \left(Z_I - Z_0\right) \right) \tag{4.65}$$

$$M_{sp,Y} = A_E \sum_I \left(p_x \left(Z_I - Z_0 \right) - p_{Is} \sin \alpha_I \; x_I \right) \tag{4.66}$$

$$M_{sp,Z} = A_E \sum_I \left(p_{Is} \cos \alpha_I \; x_I - p_x \left(Y_I - Y_0 \right) \right) \tag{4.67}$$

4.1.4.2.1 TANG solver

Figure 4.30 shows the basic steps of the TANG solver. This algorithm is similar to the TANG algorithm described in [Kalker 2001]. The basic steps are analogous to the one for the normal solver depicted in Figure 4.28. A major difference now is that the matrix equation to be solved for each division of APs in ***A*** and ***D*** sets is non-linear if there is any AP in slip, and has to be solved iteratively.

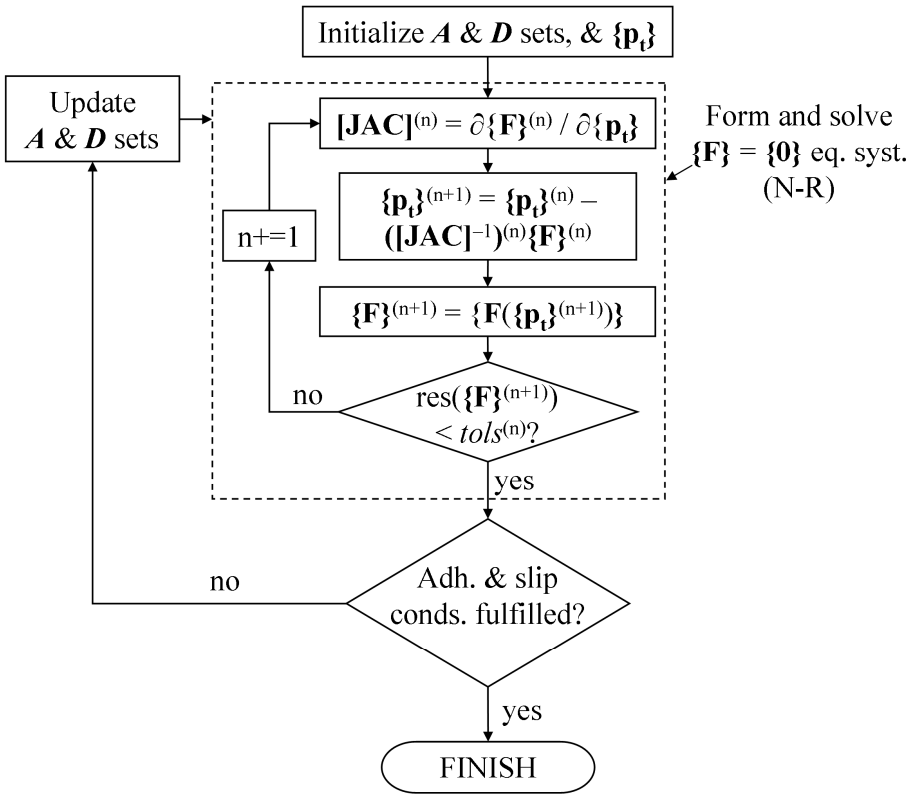


Figure 4.30. Flowchart of TANG solver.

In the outer loop of the algorithm, the adhesion and slip conditions (Eqs. (4.46) and (4.47), respectively) are checked at the same time for all the APs in each iteration, in a similar way as the inequalities for the APs in and outside ***C*** of Eqs. (4.32) and (4.33) are checked at the same time in the normal solver, as explained in §4.1.4.1. The APs which do not fulfil the inequality of their current set (***A*** or ***D***) are changed to the other set.

As mentioned previously, this solver has turned out to be ineffective in the application to conformal contact problems. On the contrary, it has been found to be more effective than the ConvexGS and SteadyGS solvers in 1D rolling contact problems with fine meshes (say `nx` > 500, and `ns` = 1), in which the tangential contact stresses have only longitudinal component and the tangential contact constraints are linear. The following features, not shown in Figure 4.30, have been implemented trying to improve its robustness:

- Scaling of the equations.

  Eqs. F_{I1} to F_{I4} compute different physical variables, which may have very different orders of magnitude. Eqs. F_{I1} and F_{I2} compute increments of slip, Eq. F_{I3} computes a tangential stress squared, and Eq. F_{I4} the scalar product of an increment of slip and a tangential stress. A scaling of the equations is optionally carried out, with the aim of avoiding large magnitude differences between the different equations and potential numerical conditioning problems of the resulting Jacobian matrices. This scaling is carried out dividing each equation by the following factors:

  - F_{I1} and F_{I2}: $\max(\|\mathbf{\Delta W_I}\|)$
  - F_{I3}: $\max(g_I)^2$
  - F_{I4}: $\max(g_I) \times \max(\|\mathbf{\Delta W_I}\|)$

  This scaling, if applied, is compensated when comparing the norms of the residuals of the equations with their tolerances (Eqs. (4.56)–(4.58)).

- Relaxation in the Newton-Raphson process for the solution of the non-linear systems of equations.

  The equation for the update of $\{\mathbf{p}_t\}$ in each N-R iteration shown in Figure 4.30 is modified as follows:

$$\mathbf{\{p_t\}}^{(n+1)} = \mathbf{\{p_t\}}^{(n)} - rf\,\mathbf{[JAC]}^{(n)}\mathbf{\{F\}}^{(n)} \tag{4.68}$$

The factor *rf* in this equation may be 1 (no relaxation), smaller (under-relaxation) or greater (over-relaxation). In [Vollebregt 2009] it is indicated that, in its application in the ConvexGS solver, it is better to perform the relaxation in a different way for the APs in slip, maintaining the modulus of the stress vector and applying the relaxation in the change of its direction instead. This has not been implemented in this solver.

- Restarts of the N-R loop for the solution of the non-linear system of equations.

  If the N-R loop (inner loop in Figure 4.30) fails to achieve a converged solution in the specified maximum number of iterations *maxite_nr*, two restarts are attempted, from different initial points, in the following order:

  - Restart assuming all APs in $\boldsymbol{A}$ and zero initial tangential stresses $\mathbf{\{p_t\}}^{(0)}$. This restart is only attempted if in the current iteration all APs are not in $\boldsymbol{A}$.
  - Restart assuming all APs in $\boldsymbol{D}$ and $\mathbf{p_{It}} = -g_I\,\mathbf{\Delta W_I} / \|\mathbf{\Delta W_I}\|$. This restart is only attempted if in the current iteration all APs are not in $\boldsymbol{D}$.

  The program does not stop if no converged solution to the non-linear system of equations has been achieved after consuming the maximum number of N-R iterations and all possible restarts. It continues with the check of the adhesion and slip conditions using the available (non-converged, potentially bad) tangential solution in a new outer iteration, in the same way as if a converged solution had been achieved in the inner loop.

These modifications have not served to solve the convergence problems of this solver.

4.1.4.2.2 ConvexGS solver

This algorithm was published in [Vollebregt 1995]. The fundamental difference between the ConvexGS and SteadyGS solvers on the one hand, and the TANG solver on the other, is in the inner loop of Figure 4.30 for the solution of the generally non-linear equations of the tangential problem. In the ConvexGS and SteadyGS solvers, individual equation systems for each AP are set-up and solved each time, instead of systems with 2×*NEC* simultaneous equations for the whole contact.

These are the basic steps of the algorithm:

1. Initialize ***A*** and ***D*** sets, and $\{\mathbf{p_t}\}$.

   Cf. §4.1.4.2.

2. For each AP: solve the individual equations of the tangential problem.

   Individual systems of 2 equations are set-up and solved sequentially for each AP. The unknowns are the tangential stresses p_{Ix} and p_{Is}. The 2×2 Jacobian of each system is formed with Eqs. (4.48) and (4.49) (setting $J = I$) for APs in ***A***, and with Eqs. (4.50), (4.52) and (4.53) for APs in ***D***. During the solution of the equations for each AP, the stresses of the rest of APs are held constant. This implies that the matrix-vector product of the tangential ICs and stresses $ICp_{Ii} = \sum_{j=x,s} \sum_{J \in \boldsymbol{C}} IC^{*}_{IiJj}\, p_{Jj}$ (cf. Eq. (4.44) and Table 4.13) may be split in a variable term due to the influence of the current AP *I*, and a constant term due to all other APs *J* ($J \neq I$), $ICpk_{Ii}$. $ICpk_{Ii}$ is computed at the beginning of the individual solution process for each AP, according to Eq. (4.69). To facilitate computations, also the constant component of the slip in the AP *I* due to the influence of all other APs, Sk_{Ii}, is computed. This is done following Eqs. (4.41) and (4.44), replacing ICp_{Ii} with $ICpk_{Ii}$ in Eq. (4.44).

$$ICpk_{Ii} = ICp_{Ii} - \sum_{j=x,s} IC_{IiIj}\, p_{Ij} \,, \quad i \in \{x, s\} \tag{4.69}$$

   In each outer iteration, all the APs are visited, in an order determined by the quantity $\|\mathbf{p_{It}}^{(n)} - \mathbf{p_{It}}^{(n-1)}\|$ (i.e. the modulus of the vector difference

between the tangential stress in the AP in the last two outer iterations). The APs with higher $\|\mathbf{p}_{It}^{(n)} - \mathbf{p}_{It}^{(n-1)}\|$ value are visited first. It has been observed that this ordering is relevant for the effectiveness of the algorithm.

Each individual system of equations is solved directly for APs in ***A***, and iteratively (with N-R) for APs in ***D***. In the latter case, the previously introduced *tf* factor is applied for the relaxation of tolerances of the non-linear system. After solving each individual system, the applicable inequality constraint (Eq. (4.46) for adhesion, and Eq. (4.47) for slip) is checked for the AP in question.

- If the condition is fulfilled, the solution is correct for the AP.
- Otherwise, the AP is changed to the other set, and the individual system of 2 equations corresponding to the new set is set-up and solved. For APs which are changed here from ***A*** to ***D***, the initial $\mathbf{p}_{It}$ for the individual N-R loop of the AP is defined as the one from the previous solution (which does not fulfil the adhesion condition), scaled down so that $\|\mathbf{p}_{It}\| = g_I$. The inequality constraint for this second solution is checked only for APs that were changed from ***A*** to ***D***. If an AP initially in ***D*** is changed to ***A***, it is maintained in ***A*** regardless of fulfilling the adhesion condition or not.
    - If the slip condition is fulfilled, this second solution (with the AP in ***D***) is retained for the AP.
    - Otherwise, the previously obtained first solution is recovered, and the AP is returned to ***A***.

For APs that are initially in ***D*** in a given outer iteration, the parallelism between the tangential stress and slip vectors of the AP is checked in each iteration of the (inner) N-R loop. If in any iteration after iteration number *maxite_nr* / 2 the angle between the two vectors is lower than 0.1 rad, the individual N-R loop of the AP is aborted, and the AP is changed to ***A***. It has been observed that frequently the N-R loop for an AP initially in ***D*** fails to converge when the correct state of the AP is ***A***.

The program does not stop if the individual N-R loop for an AP in ***D*** fails to converge (either if the AP was initially in ***D*** in the current outer iteration, or if it was changed from ***A*** to ***D***). Additionally, the last tried stress vector for the AP is retained, even if the convergence criteria were not fulfilled with it. On the other hand, a non-convergence flag, *converg_nr*, is reset to 0 for the current outer iteration. Resetting *converg_nr* to 0 implies that the overall solution for the tangential problem will not be validated in the current outer iteration. *converg_nr* is set to 1 at the beginning of each outer iteration.

After achieving a new solution for an AP, the variables **{ICp}**, ($\mathbf{p_{It}}^{(n)} - \mathbf{p_{It}}^{(n-1)}$), and $\|\mathbf{\{p_t\}}^{(n)} - \mathbf{\{p_t\}}^{(n-1)}\|$ are updated. **{ICp}** is the vector containing the $2\times NEC$ ICp_{Ii} values for the whole contact. This variable, as well as $\|\mathbf{\{p_t\}}^{(n)} - \mathbf{\{p_t\}}^{(n-1)}\|$, are updated taking into account that only the stress $\mathbf{p_{It}}$ for the current AP has changed, avoiding needless computations. This is done according to Eqs. (4.70) and (4.71), where $\Delta_I(\cdot)$ represents the change in the concerned quantity when $\mathbf{p_{It}}$ is updated in an AP *I*. The full computation of **{ICp}** at a time is carried out only once, at the beginning of the program.

$$\Delta_I\left(ICp_{Ii}\right) = \sum_{j=x,s} IC^*_{IiIj}\left(p_{Ij}^{\,(n)} - p_{Ij}^{\,(n-1)}\right),\ \forall I \in \boldsymbol{C};\ i \in \{x, s\} \tag{4.70}$$

$$\begin{aligned}\Delta_I(\|\mathbf{\{p_t\}}^{(n)} - \mathbf{\{p_t\}}^{(n-1)}\|^2) = {} & (p_{Ix}^{(n)} - p_{Ix}^{(n-1)})^2 + (p_{Is}^{(n)} - p_{Is}^{(n-1)})^2 \\ & - (p_{Ix}^{(n-1)} - p_{Ix}^{(n-2)})^2 - (p_{Is}^{(n-1)} - p_{Is}^{(n-2)})^2\end{aligned} \tag{4.71}$$

Afterwards, the loop continues with the next AP, until having visited all the APs in the current outer iteration.

3. Check error tolerances of the whole tangential solution.

   Cf. Eqs. (4.56)–(4.58). No *tf* scaling is applied in the tolerance limits of Eqs. (4.59)–(4.61) in this global check. If all the tolerances are not fulfilled or if *converg_nr* = 0, return to step 2 (outer iteration). Otherwise, the calculation is finished.

A more efficient tangential solver than ConvexGS has been developed, named TangCG [Zhao J 2015].

4.1.4.2.3 SteadyGS solver

This algorithm, which was published in [Vollebregt 2010], is used for steady rolling problems. It is differenced from ConvexGS in the computation of the variations of elastic displacement differences $u_{Ii} - u'_{Ii}$. ConvexGS uses the differences of ICs $IC_{IiJj} - IC'_{IiJj}$ for this purpose, cf. Eq. (4.44) and Table 4.13. As explained in [Vollebregt 1995], this brings some difficulties for the convergence of the algorithm, especially for small values of Δq. Setting $\Delta q = \Delta x$, and taking into account the similarity of the ICs in longitudinal direction (cf. Figure 4.15b and related discussion), it is possible to reorganize the first term in the right-hand side of Eq. (4.44) (ICp_{Ii}) according to Eq. (4.72). In this way, the use of the differences $IC_{IiJj} - IC'_{IiJj}$ is avoided, and only the ICs for the current time instant, IC_{IiJj}, are used. Δp_{Jj} in this equation is equal to the stress p_{Jj} ($j \in \{x, s\}$) at AP J, with position indices $(ix, is) = (ix_J, is_J)$ in the mesh of the PCS, minus $p_{J'j}$ at the AP J' adjacent to it towards the leading edge, with position indices $(ix_J + 1, is_J)$. For APs at the leading edge of the mesh ($ix_J = \mathtt{mx}$) or at a leading edge of $\boldsymbol{C}$, $\Delta p_{Jj} = p_{Jj} - 0 = p_{Jj}$.

$$ICp_{Ii} = \sum_{j=x,s}\sum_{J\in\boldsymbol{C}} \left(IC_{IiJj} - IC'_{IiJj}\right) p_{Jj} = \sum_{j=x,s}\sum_{J\in\boldsymbol{C}^*} IC_{IiJj}\Delta p_{Jj} \,,$$

$$\text{if } \Delta q = \Delta x;\ i \in \{x, s\} \tag{4.72}$$

The sought unknowns in the APs in $\boldsymbol{A}$ are now the differences Δp_{Ij}, instead of p_{Ij}. Accordingly, Eqs. (4.48) and (4.49) are replaced by Eqs. (4.73) and (4.74).

$$\frac{\partial F_{I1}}{\partial \Delta p_{Jj}} = IC_{IxJj} \tag{4.73}$$

$$\frac{\partial F_{I2}}{\partial \Delta p_{Jj}} = IC_{IsJj} \tag{4.74}$$

A necessary condition for the use of this algorithm is that all APs at the trailing edge of the mesh ($ix_J = 1$) are outside $\boldsymbol{C}$. This is because the ICs that provide the

influences of the APs at the trailing edge on the APs at the leading edge for the previous instant t' are not available within the set of IC_{IiJj} ICs for the current instant t. This limitation is manifested in Eq. (4.72) in that the summation over the J influencing APs is now extended to a new set $\boldsymbol{C^*}$, instead of the contact set $\boldsymbol{C}$. $\boldsymbol{C^*}$ is equal to the union of $\boldsymbol{C}$ with the set of adjacent APs in the trailing edge, $\boldsymbol{T}$. This is exemplified in Figure 4.31. There may be more than one AP in $\boldsymbol{T}$ in a given lateral position of the mesh, as illustrated in the figure. It would be feasible to use this algorithm also in the case of having some APs at the trailing edge of the mesh inside $\boldsymbol{C}$ and all APs at the leading edge outside $\boldsymbol{C}$.

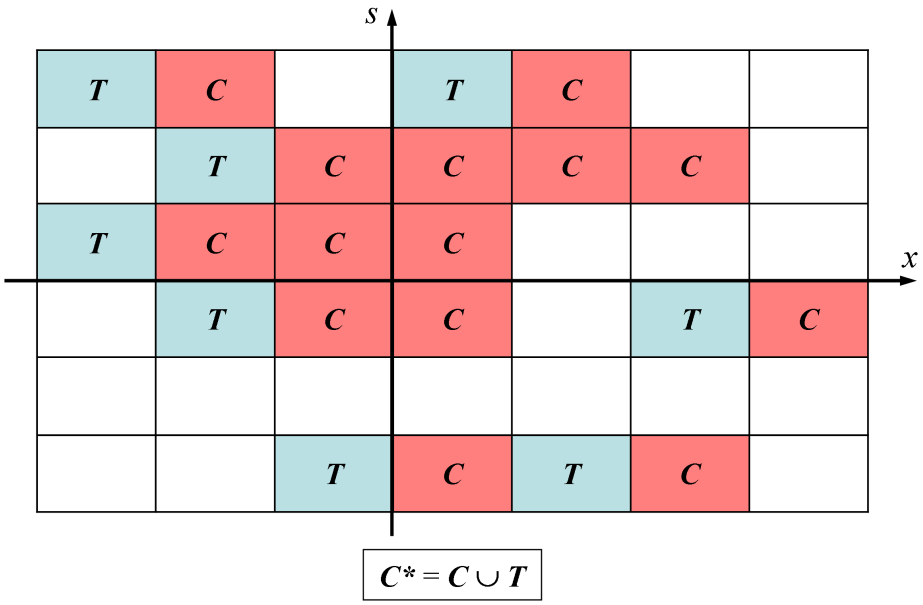


Figure 4.31. $\boldsymbol{C^*}$ set, equal to the $\boldsymbol{C}$ set plus the adjacent APs in the trailing edge $\boldsymbol{T}$.

Some features of SteadyGS which are not shared with ConvexGS are commented next:

- The APs are visited ordered by entire rows or strips (a strip is formed by the set of APs located in the same s position), from the trailing to the leading edge, as advised in [Vollebregt 2010]. The ordering of the strips in each outer iteration is determined by the change metric $\Delta p_t(is)$, defined in Eq. (4.75). $\boldsymbol{C}(is)$ in this equation is the subset of APs in $\boldsymbol{C}$ located in

the is^{th} lateral position of the mesh. The rows with higher $\Delta p_t(is)$ are visited first.

$$\Delta p_t(is) = \sum_{I \in \boldsymbol{C}(is)} \left\| \mathbf{p_{It}}^{(n)} - \mathbf{p_{It}}^{(n-1)} \right\| \tag{4.75}$$

- The unknown variables are now different for the APs in ***A*** on the one hand, and for those in ***D*** on the other hand. The unknowns for APs in ***A*** are the differences Δp_{Ij}, and the ones for APs in ***D*** are p_{Ij}. The values of each of these variables for all the APs of ***C*** are collected in vectors $\{\Delta \mathbf{p_t}\}$ and $\{\mathbf{p_t}\}$, respectively. The values of both vectors are updated in unison, so that the equality $\Delta p_{Ij} = p_{Ij} - p_{I'j}$ is fulfilled (I' is the AP adjacent to I towards the leading edge). Each time the individual problem for an AP I is solved, both vectors are updated in the following way:
  - APs in ***A***: new Δp_{Ij} ($j \in \{x, s\}$) values are obtained. With these, the positions corresponding to AP I of $\{\Delta \mathbf{p_t}\}$ are updated, and the following additional modifications are carried out:
    - The values in $\{\mathbf{p_t}\}$ of AP I are updated, as $p_{Ij} = p_{I'j} + \Delta p_{Ij}$. The adhesion condition is checked with these values.
    - The values in $\{\mathbf{p_t}\}$ of the APs J located between I and the nearest trailing edge in the same row are updated recursively, going from I towards the trailing edge, to try to maintain their current Δp_{Jj} values. After updating the p_{Jj} values for each J AP, the adhesion condition is checked for that AP. If it is not fulfilled, the p_{Jj} values are scaled down so that $\|\mathbf{p_{Jt}}\| = g_J$, and the values Δp_{Jj} are updated accordingly. Additionally, that AP J is changed to ***D***.
    - The values Δp_{ITj} in $\{\Delta \mathbf{p_t}\}$ of AP IT, the trailing edge nearest to I in the same row, are updated. If, after the previous update of $\{\mathbf{p_t}\}$, no AP has changed to ***D***, the change in Δp_{ITj} is equal to the change in Δp_{Ij} with contrary sign.

  It has to be noted that, due to the changes in $\{\Delta \mathbf{p_t}\}$ additional to the changes of the Δp_{Ij} values of the current AP I, the Jacobian defined by Eqs. (4.73) and (4.74) is not exact. If no APs changed to ***D*** (i.e., reach saturation) after the updates in $\{\mathbf{p_t}\}$, an exact Jacobian could be

obtained as $IC_{IiJj} - IC_{ITiJj}$, $i, j \in \{x, s\}$. But these differences of ICs are precisely what are wanted to be avoided in this solver. Additionally, it is not known beforehand if more values will change in $\mathbf{\{\Delta p_t\}}$ as a result of reaching saturation in some AP between *I* and *IT*.

- APs in ***D***: new p_{Ij} ($j \in \{x, s\}$) values are obtained. With these, the values in $\mathbf{\{\Delta p_t\}}$ of AP *I* and the one preceding it in the same row (adjacent to *I* towards the trailing edge) are updated.

Apart from these features, the algorithm is differenced from ConvexGS also in the way in which not small rolling radii variations across the contact are handled, as explained in §7.2.2.1.2.

4.1.4.3 Coupled normal-tangential contact

The coupled normal tangential contact problem is solved through the so called Panagiotopoulos process ([Panagiotopoulos 1975], [Antes 1992]), wherein the normal and tangential problems are solved alternately, as explained in §2.1.3 and shown schematically in Figure 2.21. The main steps of the process are shown in Figure 4.32. It has to be mentioned that in conformal contact, the normal and tangential contact problems are coupled even with elastically similar materials as in the wheel–rail case; i.e. elastic quasiidentity does not apply.

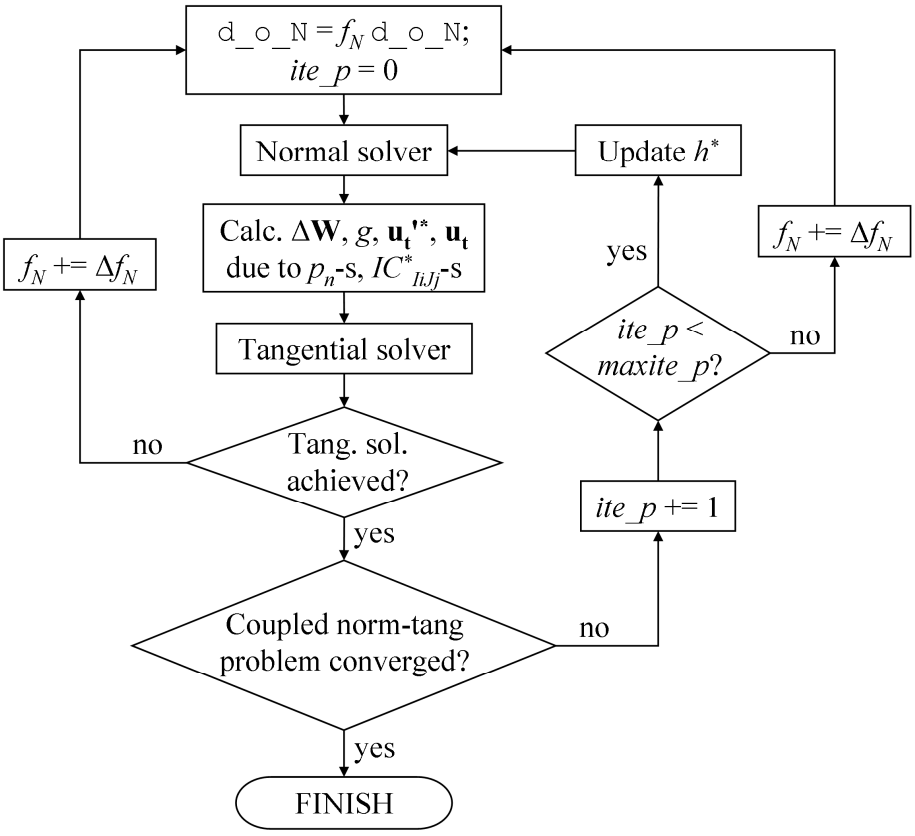


Figure 4.32. Flowchart of the Panagiotopoulos algorithm.

Some details of the algorithm as implemented in CECT are given next.

- Factor f_N for modification of `d_o_N`.

  A factor f_N is defined, which is initialized with a value of 1.0. In the event of any of the following convergence failure cases, f_N is increased by a step Δf_N, and the whole Panagiotopoulos process is restarted with the input `d_o_N` modified with f_N as shown in the first step of Figure 4.32.

  - A converged tangential solution has not been achieved by the tangential solver.
  - The maximum number of allowed Panagiotopoulos iterations (*maxite_p*) has been reached without achieving a converged solution

for the coupled normal-tangential contact problem. *maxite_p* has currently a fixed value of 15.

Δf_N is given a small value, and a maximum number of such restarts is allowed, so that the modified `d_o_N` is kept close to the original one. The maximum number of restarts is defined by a maximum value $f_{N,max}$ that f_N is allowed to reach. These parameters have the following fixed values. If `ndgt`=1, $\Delta f_N = 0.0005$, and $f_{N,max} = 0.002$. If `ndgt`=0, Hertzian proportionality is applied to obtain equivalent load variations. That is, $\Delta f_N = (1.0005)^{(1/3)} - 1$, and similarly for $f_{N,max}$.

If, after consuming all possible restarts, no converged solution has been achieved, CECT does not print any .mat nor .prf files for the current case, and continues with the next case in the input file. However, in transient problems (`tdgt`=0 or 1), in which the solution of a case depends on the previous one, the solutions obtained for cases after a case for which no converged solution has been achieved may not be correct.

- The normal solver uses modified normal undeformed distances h^* as input, which incorporate the effect of the tangential stresses, as indicated in §4.1.4.1 (cf. also Figure 2.21). h^* is updated in each Panagiotopoulos iteration, with the tangential solution obtained in the previous iteration.
- The following quantities which are inputs for the tangential solver are calculated for each AP in ***C***:
  - $\Delta\mathbf{W}$. Eqs. (4.23)–(4.25) are used for static problems (`tdgt`=0), and Eqs. (4.21)–(4.22) for rolling problems (`tdgt`=1 or 2). A $-\Delta d \times \sin(\Delta\alpha_{I\text{-}d}) / \Delta t$ term is added in Eq. (4.22) if `tdgt`=1, cf. Figure 4.12. The calculation of the rigid slips has been implemented inside the Panagiotopoulos loop so that deformed geometry could be used in the computation (cf. §4.1.1.3.1 and Figure 4.14). Currently the rigid slips are calculated with the undeformed geometry, and this calculation could be carried out outside the Panagiotopoulos loop.
  - Traction bound. $g_I = \mu_I p_{In}$.
  - Tangential elastic displacement differences of the previous time instant, $\mathbf{u_t}'^*$. Cf. Eq. (4.44) and Table 4.13. The calculation itself is

carried out here only for `tdgt`=2. If `tdgt`=0 or 1, the $\mathbf{u_t}'^*$ values are calculated before the Panagiotopoulos loop for the whole PCS, and in this step the values corresponding to the $\boldsymbol{C}$ set are selected. In the first case (`c=1`), if `tdgt`=0 or 1, $\mathbf{u_t}'^* = \mathbf{0}$.

- Tangential elastic displacement differences $\mathbf{u_t}$ due to current normal pressures; i.e. $\sum_{J\in\boldsymbol{C}} IC_{IiJn}\, p_{Jn}$ term in Eq. (4.44).

- $\mathbf{[IC^*_{ij}]}$ complete IC matrices for the tangential problem, with $i, j \in \{x, s\}$; cf. Eq. (4.44) and Table 4.13. These matrices are formed for different sets of APs for the TANG and ConvexGS solvers on the one hand, and for the SteadyGS solver on the other (cf. §4.1.4.2). Each of these matrices has *NEC*×*NEC* elements for the TANG and ConvexGS solvers, and *NEC*×*NECT* for the SteadyGS solver. *NECT* is the number of APs in the $\boldsymbol{C}^*$ set (cf. Figure 4.31). These matrices are calculated only under the following conditions (otherwise, previously calculated matrices are used, avoiding their redundant calculation):

  - `NOT(((mdgt(c)<0.1 AND gdgt(c)<0.1) OR ite_p > 1.1) AND NOT var_C)`, OR
  - `c < 1.1 AND ite_p < 1.1`, OR
  - `(tdgt(c)>1.9 OR tdgt(c-1)>1.9) AND ABS(tdgt(c)-tdgt(c-1))>0.1`

  The first condition states that in the first Panagiotopoulos iteration of a case (`ite_p`=1), new IC^*_{ij}-s are calculated if either the mesh, the geometry or the $\boldsymbol{C}$ set have changed from the previous case. For next Panagiotopoulos iterations, new IC^*_{ij} -s are calculated if the $\boldsymbol{C}$ set has changed from the previous iteration. The condition of change of the $\boldsymbol{C}$ set is checked with the `var_C` bit, which is set to 1 if the $\boldsymbol{C}$ set is different from the one with which the last IC^*_{ij}-s were calculated.

  If `tdgt`=2 and the SteadyGS solver is not used, $IC^*_{ij} = IC_{IiJj} - IC'_{IiJj}$. In this case, two further necessary conditions to avoid the calculation of the IC'_{IiJj}-s in the first Panagiotopoulos iteration of the case is that `dq(c) = dq(c_o) AND tdgt(c-1) = 2`.

- The tangential solver is run only if there is friction anywhere inside $\boldsymbol{C}$ in the current Panagiotopoulos iteration. Initial estimates of tangential stresses $\{\mathbf{p}_t\}$ and AP division into the $\boldsymbol{A}$ and $\boldsymbol{D}$ sets are passed to the tangential solver, based on the previous available tangential solution, as explained in §4.1.4.2. The previous available tangential solution is that of the previous case (`c_o`) for the first Panagiotopoulos iteration of a case `c`, and that of the previous Panagiotopoulos iteration with converged tangential solution otherwise.
- The following conditions are used to check the convergence of the coupled normal-tangential solution:

$$\|\{\mathbf{p}\}^{(n)} - \{\mathbf{p}\}^{(n-1)}\| \,/\, \max(1.0\text{E}{-12}, \|\{\mathbf{p}\}^{(n)}\|) < tol\_p\_norm \tag{4.76}$$

$$\max(|\{\mathbf{p}\}^{(n)} - \{\mathbf{p}\}^{(n-1)}|) \,/\, \max(1.0\text{E}{-12}, \max(|\{\mathbf{p}\}^{(n)}|)) < tol\_p\_max \tag{4.77}$$

  $\{\mathbf{p}\}$ is a $3\times NE$ element vector that collects the contact stresses (p_n, p_x and p_s) for the whole PCS. *tol_p_norm* and *tol_p_max* have the following fixed values: *tol_p_norm* = 1.0E–5; *tol_p_max* = 1.0E–4.

  These conditions are checked only if there is friction somewhere inside $\boldsymbol{C}$, or if the difference between the h and h^* values of the current iteration is non-zero. If Eqs. (4.76) and (4.77) are fulfilled, or if they are not checked, the solution to the coupled normal-tangential problem is considered to have converged, and the Panagiotopoulos loop is finished.

4.2 Finite element models

The FE models constructed for contact mechanics analyses of wheel–rail conformal contact are described in this section. These have served to contrast and validate the results obtained with CECT, and to study aspects of the subsurface stress field in conformal contact that could not be studied with simpler models. As the FE models are the most comprehensive contact models used in this thesis, the FEM results are considered generally the reference. However, from this description (especially from §4.2.2) it will become apparent

that the FE models, being more complex, generally involve greater simulation cost and difficulties. They have more numerical and tolerance settings and simulation setup possibilities, and they may require considerable tuning to achieve reliable results as well as robustness and affordable simulations; see [Toumi 2016] for example. Even after their tuning, some degree of inaccuracy in the FEM has to be allowed to get results at reasonable cost. This being so, in non-conformal contact problems with well-established numerical solutions, in which the exact contact theory is known to yield precise results, the CECT results have been taken as the reference for tuning and validation of the FEM.

4.2.1 Description of models

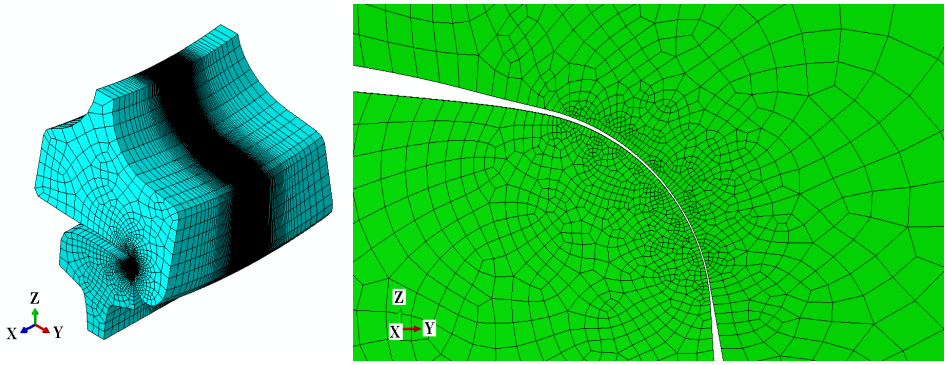
These FE models have been constructed with the commercial FE package Abaqus/Standard [Abaqus doc 2012]. Some FE models were also constructed with the commercial FE package ANSYS [ANSYS doc 2010], with which some initial non-conformal, static contact test cases were solved. The description given here refers mostly to the Abaqus models, but the analysis methodology used with the ANSYS models was similar.

A major difficulty of the contact FE models described here is the high computational cost. This results from the large size of the models and the numerous substeps in each simulation. Each rolling contact simulation with these models has taken typically on the order of a few tens of hours on a high performance computer, with an Intel Xeon E5-2687W @ 3.10 GHz eight-core processor and memory of 128 GB (8×16 GB), using 5 cores in parallel.

4.2.1.1 Geometry and mesh

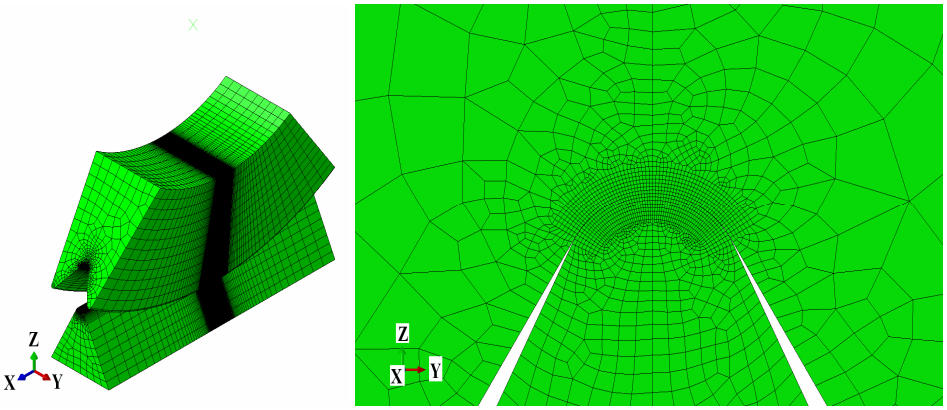
Only a portion of the wheel and the rail including the contact zone is meshed, with sufficient extension to ensure that the boundaries of the model are far enough not to affect the local stress field in the contact. In some wheel–rail contact related studies with finite element models, e.g. [Chongyi 2010], [Bian 2013], [Zhao X 2014], [Vo 2015] and [Yang 2016], the entire wheel and a larger section of the rail are modelled, and details of the vehicle and of the track structure are included. This is avoided in the models described here because they are intended to focus on the local contact problem, and it is not aimed at including the general structural behaviour of the contacting bodies in the study.

Two of the constructed FE models, designated as A and B, are shown in Figure 4.33 and Figure 4.34. Model A has a total of 322,162 elements and 347,415 nodes, and model B 435,552 elements and 466,698 nodes.

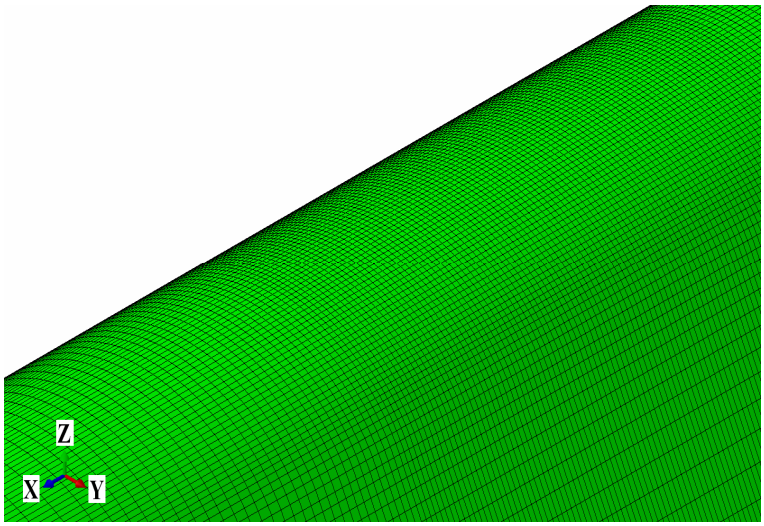


(a) General isometric view. (b) Cross-section. Zoomed view in the contact area.

Figure 4.33. FE model A for wheel–rail conformal contact analysis.



(a) General isometric view. (b) Cross-section. Zoomed view in the contact area.



(c) Detail of the mesh of the rail surface (with the wheel hidden).

Figure 4.34. FE model B for wheel–rail conformal contact analysis.

The meshes are Lagrangian. The modelled domains include the whole wheel tyre and rail head as may be seen in Figure 4.33, or comparatively larger volumes in the case of other geometries like the one in Figure 4.34. The geometry of model A is similar to a S1002 wheel on a 60E1 rail. The modelled dimensions in the rolling direction are enough for the distances between the contact patch and model boundaries to be above five times the longitudinal contact patch dimension. Thus, the wheel and rail portions included in the model are much larger than the contact patch.

The volumes of the modelled portion of the contacting bodies are meshed with linear hexahedron elements. In most of the models C3D8 elements have been used, and in some cases C3D8R elements have been used instead. The C3D8R elements work well when the mesh of the contact patch is sufficiently fine, but may exhibit hourglassing problems if very few nodes enter in contact or if there are large discontinuities in the contact forces between adjacent nodes; cf. §5.2.2.

The rail is aligned with the global X axis of the model, defining the rolling direction. The contact surfaces are considered smooth, without irregularities. The mesh of the rail is extruded along the rolling direction, and the mesh of the wheel is swept or rotated around its axis. Thus, the surfaces of the bodies are

discretized in rectangular element faces (nearly rectangular in the case of the wheel), aligned with the rolling direction.

The mesh is heavily focused in the contact zone, as may be seen in Figure 4.33 and Figure 4.34, and the element size is increased rapidly out of the contact vicinity to minimize the mesh size. The mesh refinement is similar in both contacting bodies. The element size in the most refined zone is defined to have a resolution of at least 8-10 elements along each semiaxis of the contact patch in each contacting surface, and typically around 20. The most refined zone is meshed with elements of uniform size, and in rolling contact simulations it is extended for a length of 2–3 times the foreseen contact patch length in rolling direction. The element size defined in the depth direction is of the same order as the most refined surface sizes, though the degree of refinement is decreased faster in this direction. Based on the study presented in [Zhao X 2011], it is considered that the defined mesh density should suffice to avoid significant discretization errors. This refinement is considered also adequate to allow the representation of the local contact field in sufficient detail for the purposes of contrasting the most relevant features of conformal and non-conformal contact mechanics at the macroscopic level, with smooth surfaces.

In most of the models, a few layers of structured mesh are provided in the depth direction in the contact surfaces of both bodies, with parallelepiped elements with nearly rectangular faces. These may be appreciated in Figure 4.34b for model B. They are not present in model A. It is advisable to have elements with regular shape at the contact surface to get smoother contact solutions. Numerical noise in the contact solutions is more prone to appear with more irregular elements, mostly in the tangential part of the contact problem. Elements with nearly right angles also tolerate better large aspect ratios, with less impact in their accuracy. In this way, the element dimensions may be defined according to the contact patch size in each surface direction, without having to keep an aspect ratio close to 1.

Appropriate mesh sizing is required to achieve the desired resolution in the contact vicinity, and at the same time contain the mesh size within affordable limits. The mesh size is determined primarily by the element size in the most refined zone, and the distance in which the refined zone is extended in the rolling direction to allow a sufficient rolled length (cf. §4.2.1.5 and §4.2.2.1).

The size of the modelled wheel and rail portions is of secondary importance to the mesh size. Variable element sizes are defined in the rolling direction, as may be appreciated in Figure 4.34c. The most refined zone is placed where the contact patch will be located in the final part of the simulation. This most refined zone is sized to accommodate just one or two contact patch widths, to limit the mesh size.

4.2.1.2 Material

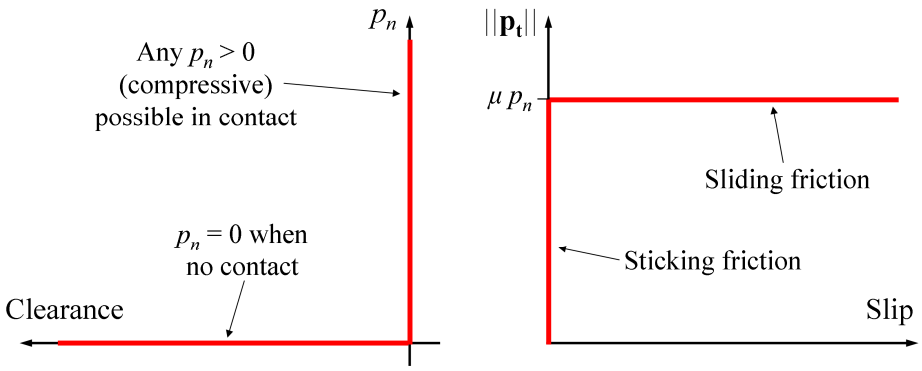
Steel is considered as the material for the wheel and the rail, with homogeneous properties in the modelled domains. The material behaviour is modelled as isotropic and linear elastic. In this way, the FEM and CECT simulations are comparable. The values assigned to the elastic constants are E (Young's modulus) = 210 GPa and ν (coefficient of Poisson) = 0.30, unless otherwise noted.

4.2.1.3 Contact definition

Wheel–rail contact is defined with a mechanical contact pair in Abaqus/Standard, using the `*Contact Pair` keyword. These contact pairs provide capability for contact between two deformable bodies and finite sliding.

Surface-to-surface contact discretization is used. This usually produces more accurate contact stress solutions than node-to-surface discretization, but with fine meshes like the ones in the models described here, the differences should be small. The convex body is designated as the slave, and the concave as the master. The rail is always the convex body in these models. Finite-sliding contact tracking is used, which is the most general formulation, with the default path-based tracking algorithm.

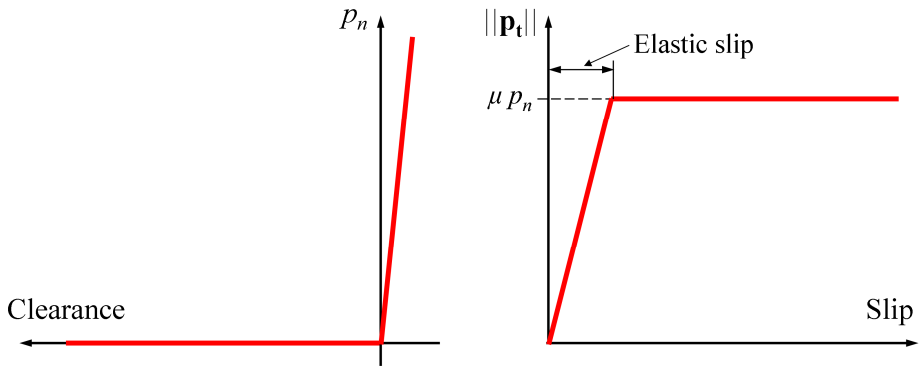
Regarding the contact interfacial constitutive behaviour, hard pressure-overclosure and tangential stress-slip relationships are considered in the normal and tangential contact, respectively. These are illustrated in Figure 4.35. In the tangential part, the basic, isotropic Coulomb friction model is used, with a prescribed constant coefficient of friction, equal in all directions. These constitutive behaviours are analogous to the ones implemented in CECT.



(a) Normal contact. (b) Tangential contact.

Figure 4.35. Hard contact interface constitutive relationships.

The contact constraints are enforced with the penalty method, with the default linear behaviour (i.e. with constant stiffness). In this way, regularisation of the hard interfacial behaviour laws (see e.g. Chapter 36 of the Abaqus Analysis User's Manual [Abaqus doc 2012]) is achieved, as depicted in Figure 4.36 — specifically, a piecewise polynomial regularisation as described in §5.2.3 of [Wriggers 2006].



(a) Normal contact. (b) Tangential contact.

Figure 4.36. Regularization of the hard contact interface constitutive relationships with the penalty method.

The penalty stiffness values have to be defined so that a good compromise between precision and ease of convergence is achieved. The contact relationships are fulfilled more closely with a high penalty stiffness, but at the

expense of increasing the computational cost, because the convergence is made more difficult and more iterations are needed in each substep to achieve a converged solution. The adjustment of the penalty stiffness values in these models is discussed in §4.2.2.3.

The frictional contact conditions produce unsymmetric terms in the stiffness matrix of the system. Unsymmetric matrix storage and solution scheme are used in the FEM simulations, as recommended in §36.1.5 of the Abaqus Analysis User's Manual [Abaqus doc 2012], to improve the convergence rate of these simulations.

4.2.1.4 Loads and boundary conditions

Figure 4.37 shows schematically the loads and boundary conditions applied in these FE models. Global bending of the bodies in the vertical plane is inhibited with these boundary conditions, so that the local contact field under mainly vertical loads is practically independent of the size of the modelled wheel and rail portions, considering that these are considerably larger than the contact patch dimensions as explained in §4.2.1.1.

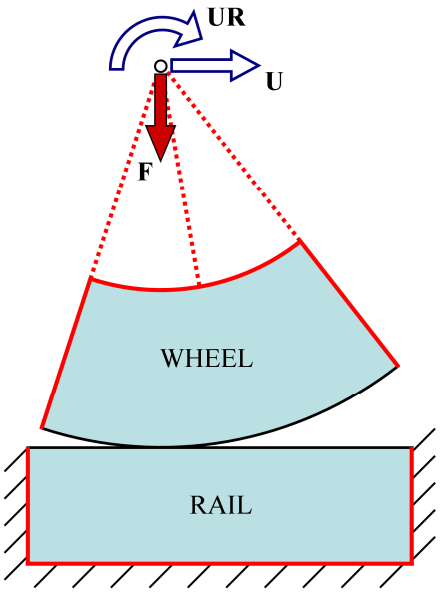


Figure 4.37. Loads and boundary conditions in the FE model of wheel–rail contact.

The lower surface of the rail is built-in, i.e. the nodes in this surface have all their degrees of freedom (DOF-s) constrained to zero. The vertical cut surfaces of the rail (one at each extreme in longitudinal direction of the modelled rail portion) are also built-in in some of the models, while in others they have only their longitudinal displacement constrained. Almost the same contact results are obtained with either conditions, due to the long distance between the contact patch and these cut surfaces.

A reference point is defined for the wheel, located in its axis of revolution. This point is rigidly tied to the cut surfaces of the wheel. The loads (**F**) and the rigid body motion of the wheel (displacements **U** and rotations **UR**) are applied through this point.

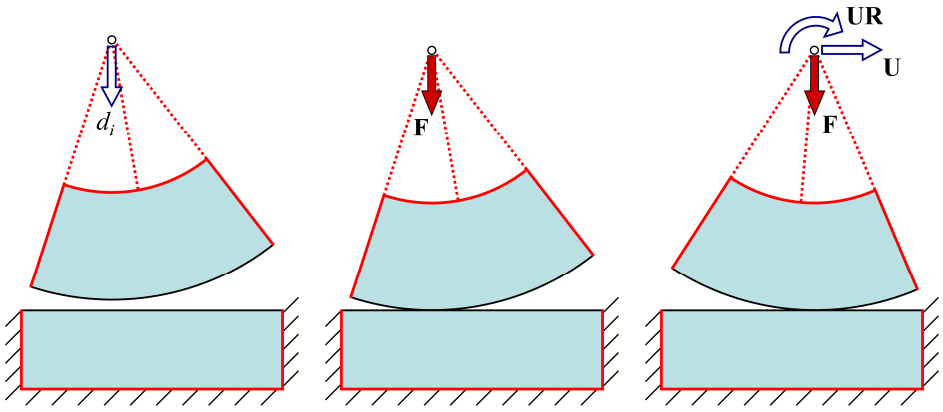
4.2.1.5 Simulation steps

Static contact, transient, or steady state rolling contact cases may be simulated with these FE models. This is accomplished with non-linear static simulations in all cases, with implicit solver, and following a step-by-step technique. Rolling contact is simulated applying prescribed displacements and rotations to the wheel in a plurality of small increments or substeps, representing the desired rigid body motion of the wheel on the rail. Non-linear geometry is used in the simulations, to accommodate finite wheel motions.

In the previously cited works [Chongyi 2010], [Bian 2013], [Zhao X 2014], [Vo 2015] and [Yang 2016], explicit dynamic transient models are used. With those models, inertial forces in the rolling contact process are taken into account, and the explicit integration is effective in dealing with the severe nonlinearities associated with the contact conditions. The use of static models in this work is justified because here it is not intended to study any dynamic phenomena, and the local behaviour of the contacting bodies can be assumed to be unaffected by inertial effects at conventional rolling speeds, as stated in §1.1.1. In this way, some drawbacks associated with transient dynamic models are avoided (e.g. the transient vibrations at the beginning of the simulation, and the need to model longer sections in the rolling direction and to simulate the rolling through a longer distance until these are faded out), and quite detailed results are achieved in the contact zone without incurring in too high computational costs. [Pletz

2014a] is another example in which a quasistatic 3D FE model is used in a detailed wheel–rail contact mechanics study.

The main steps making up a FEM contact simulation as carried out in this work are illustrated in Figure 4.38.



(a) Initial approach. (b) Load. (c) Rolling.

Figure 4.38. Main simulation steps of rolling contact with FEM.

Each step is explained next:

- Initial approach.

  Starting from a state with no contact between wheel and rail, an initial approach d_i is applied to the wheel towards the rail, so that an initial contact patch is formed. This displacement-controlled step is necessary because before achieving the initial contact closure the system is singular if the rigid body motion of the wheel is not constrained.

  The wheel is translated and oriented previously so that at the beginning of this step it is nearly touching the rail at the intended location and with the intended orientation. The necessary wheel position is calculated beforehand, solving the wheel–rail geometric contact search problem as outlined in §4.2.3.1. In some cases, the wheel orientation is applied after the loading step, when it is possible to reuse the same contact solution obtained after the loading step in different rolling contact runs.

A small value is given to d_i, so that the resulting wheel–rail contact load in this step is small and the majority of the contact load is built-up in the next step. The wheel DOF-s not involved in the application of the initial approach are constrained to zero.

The longitudinal position and pitch angle of the wheel in this step depend on the type of simulation carried out. These are defined so that the most refined zone of the meshes of both contacting surfaces match, and to have the contact patch positioned in the most refined zone of the mesh in the final part of the simulation, as indicated in §4.2.1.1. In static contact simulations, the contact patch is formed right in the centre of the most refined zone of the mesh in this step. In rolling contact simulations it is formed upstream of this zone, taking into account that it will translate (longitudinally over the rail and circumferentially over the wheel surface) during the rolling step of the simulation.

- Load.

In this step, the displacement DOF-s in the reference point of the wheel in which the approach d_i was applied in the previous step are released, and the wheel–rail normal load is applied in these DOF-s. The wheel remains stationary on the rail, with the rest of its DOF-s constrained to zero.

The way the normal wheel–rail load is increased in this step and the chosen incrementation (number of substeps) are relevant mostly in static frictional contact problems. In Hertzian contact, a constant rate of increase of the linear dimensions of the contact patch is achieved with a cubic load increase. This proportion changes in conformal contact. With circular cross-sections, as the total contact angle variation in the contact patch $\Delta\delta$ (cf. Figure 6.9) increases, the longitudinal dimension increases faster, and the lateral dimension slower. Bearing this in mind, in these models usually a quadratic load increase is applied in this step, as a compromise between avoiding large increases of the contact patch size in the first substeps, and large increases of the load in the last substeps. The number of substeps must be enough to adequately represent the rigid shift increments in each point of the contact patch from the moment of entering in contact, and the evolution of the elastic displacement differences in the contact surface. This step is usually divided in a

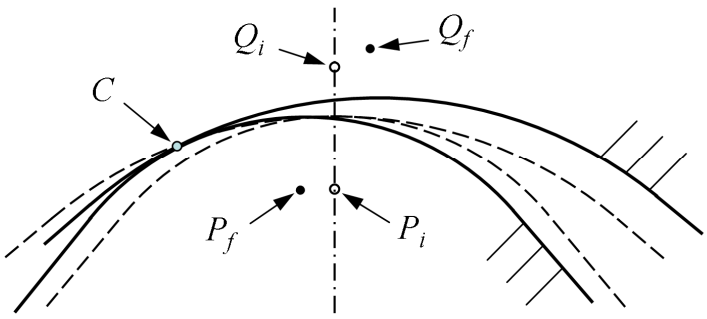
number between 20 and 40 substeps in static frictional contact problems. This number is normally chosen in proportion to the mesh resolution in the contact patch. These considerations apply equally to static frictional contact problems computed with CECT. In steady rolling contact problems they are much less relevant, because the tangential stress field developed in the contact patch in this step is washed away in the rolling step of the simulation.

Sometimes a second load step is defined, to add longitudinal or lateral preload in the contact. This helps to accelerate the convergence to the steady rolling state in rolling contact simulations. The value of the preload is defined as a fraction of the foreseen resultant contact tangential force in the final steady state, usually between 50% and 75%. To apply the preload, the concerned displacement DOF-s of the wheel reference point are released, and the preload is applied along the wanted direction. This second load step is divided in up to five substeps.

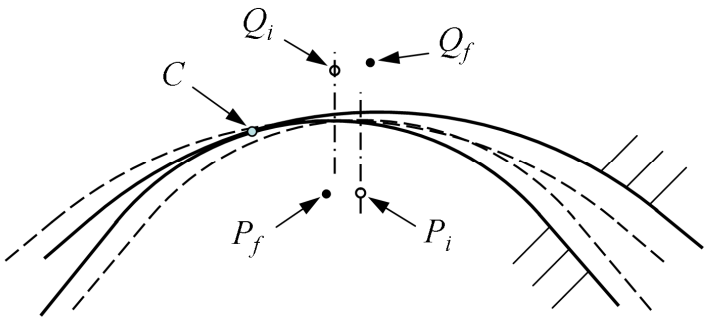
It has to be considered that in rolling contact problems with lateral forces, the final contact result may change appreciably depending on the magnitude of the lateral preload introduced in this step. As in conformal contact the contacting surfaces are close to one another, relatively small deflections or displacements of the bodies may produce significant changes in the position of the contact patch in the *s* direction. This in turn may change the spin creepages and resultant frictional contact lateral force F_{lat}, as a result of the different contact surface orientations.

Figure 4.39 illustrates how the lateral position of the contact patch may change with the lateral preload and global deflections of the contacting bodies. The figure shows an upper concave body on a lower convex body in rolling contact, with a F_{lat} force that pushes the upper body to the right and the lower body to the left. The undeformed configurations of the bodies are shown in dashed lines, and the deformed configurations, which are obtained adding the global deflections to the undeformed configurations, in solid lines. A point in each body is represented in both configurations, to better visualize the changes in the relative position of the surfaces. These are point P for the lower body and point Q for the upper body. Subindex i refers to the undeformed configuration, and

subindex *f* to the deformed configuration. The bodies considered in this example have a lateral symmetry plane, and points *P* and *Q* lie in this plane.



(a) No lateral preload applied.



(b) Lateral preload applied.

Figure 4.39. Change of lateral position of the contact patch with the lateral preload applied in the load step.

In Figure 4.39a both bodies are initially positioned with their lateral symmetry planes aligned, and no lateral preload is applied before the rolling step. In the rolling step, the upper body bends to the right and the lower body to the left, under the action of F_{lat}. As a result, the central contact point *C* shifts to the left. In Figure 4.39b a lateral preload is applied before the rolling step, in the same direction as the F_{lat} force developed in the rolling step, and lower in magnitude. Upon the introduction of this preload, the undeformed configuration of the upper body is translated to the left, as may be observed by the lateral translation

of Q_i with respect to P_i. Combining this translation with the lateral bending of the bodies, the shift of the contact point C is now lower than with no lateral preload.

- Rolling.

  In this step, simultaneous translations and rotations are applied to the reference point of the wheel in multiple small increments (substeps), to simulate the rolling motion of the wheel along the rail.

  For a wheel running on a straight rail in steady rolling, the following applies in relation to the wheel axis of revolution:

  - Its direction remains fixed.
  - It is aligned with the axis of rotation of the wheel.
  - It translates along the global X direction.

  The necessary boundary condition in the reference point of the wheel to define its prescribed motion may be readily defined either with the `type=velocity` option or without it (the default is `type=displacement`) in the `*Boundary` entry defining the motion of the reference point of the wheel in the Abaqus input deck, taking into account the first point above. The magnitude of the total rotation applied to the wheel in the rolling step, $\|\Delta\boldsymbol{\Omega}\|$ is calculated as $\|\Delta\boldsymbol{\Omega}\| = \Delta U_x / R_0 \times (1 - \xi_0)$, being ΔU_x the total translation of the wheel in the rolling step, R_0 the rolling radius of a given point in the wheel, and ξ_0 the longitudinal creepage prescribed for that point (negative if tractive). The components of $\Delta\boldsymbol{\Omega}$ in the global coordinate system are proportional to the cosine directors defining the orientation of the axis of revolution of the wheel in the same system, taking into account the second point above.

  The displacement DOF-s of the reference point of the wheel in the YZ plane may be constrained at the values attained at the end of the loading step if the normal wheel–rail load remains constant during rolling, or they may be left free, bound by the wheel–rail normal load and contact pair. The first option, restraining more the wheel motion, has the advantage of facilitating the convergence of the FEM solution. Its downside is that the normal wheel–rail load will have some uncontrolled change as the

tangential stress field evolves during the rolling step, due to normal–tangential contact coupling. This unintended change is avoided with the second option.

The rolling phase of the simulation may be split into several rolling steps, to traverse regions of the surface with different mesh densities in the longitudinal direction. The number of increments in each step is then defined according to the element lengths in the rolling direction. The criteria followed here to define the number of increments in each rolling step is to make the advanced distance by the wheel in each increment similar to the mean longitudinal dimension of the mesh elements in the surface section traversed during the step. This is analogous to the criteria for the definition of dq in the CECT analyses set out in §4.1.2.1.1. A sensitivity to this incrementation in a particular case is presented in §4.2.2.2.

For steady rolling, the applied load and creepages are held constant in the whole rolling phase of the simulation, and the wheel is rolled over the rail in a sufficiently long distance to allow the local contact field to stabilize. This is analogous to the method used by Kalker in DUVOROL to solve steady rolling problems with the exact contact theory [Kalker 1979b], wherein the steady rolling contact problem is solved as a sequence of transient rolling cases with constant loads and creepages, until the steady state sets in. Experience with the exact contact theory shows that the steady state may be achieved after a rolled distance of 1–2 contact lengths [Kalker 1979b]. A larger rolled distance may be necessary with the FE models used here, because of the global deflections that develop during the rolling phase. These influence the rigid body motion at the local contact level, delaying the convergence to the steady rolling state. The rolled distance to reach the steady rolling state is one of the aspects in the tuning of these FE models, as discussed in §4.2.2.

4.2.2 Tuning

These aspects of the tuning of the contact FE models are discussed in the following subsections:

- Total rolled distance.

- Size of each rolling substep.
- Penalty stiffness.

These subsections refer to the FE contact models developed in Abaqus, with which most of the investigation presented here has been carried out. Afterwards, the most important points are compiled in §4.2.2.4. In this last subsection, the conclusions reached in the tuning of the FE contact models constructed in ANSYS are also included.

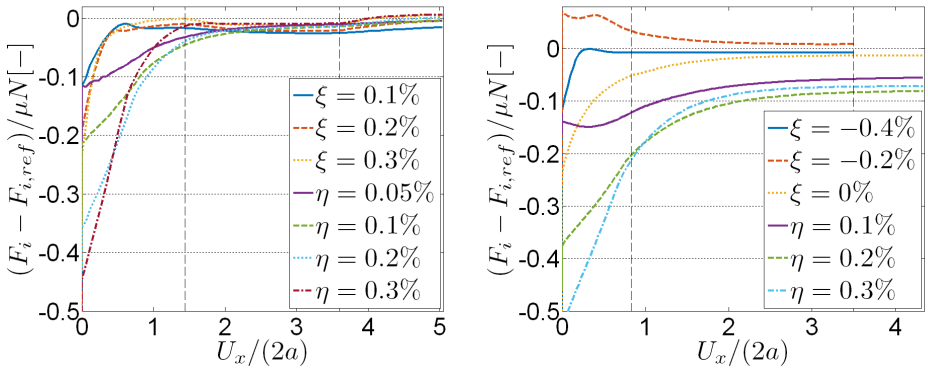
4.2.2.1 Total rolled distance

In a steady rolling contact simulation, the total rolled distance has to be long enough to ensure that the contact field has approached sufficiently the steady state. On the other hand, the total rolled distance has a significant influence on the cost of the contact FEM simulation. Firstly, because it implies a longer distance with a refined mesh, and hence a bigger mesh, and secondly, because more rolling substeps are needed. So, this is one of the most important aspects to adjust. Additionally, it has to be considered right from the preprocessing phase, when constructing the FE mesh.

Figure 4.40 shows the evolution of the contact resultant frictional forces with the rolled distance in different FEM steady rolling contact simulations. The trends of these resultant forces (CFSn, n = 1, 2, 3, history output variables in Abaqus contact FE models, cf. §4.2.1 of Abaqus Analysis User's Manual [Abaqus doc 2012]) are good indicators of the stabilization of the contact solution in these simulations. They are revised in each rolling contact FEM simulation to check the stationarity of the solution. Figure 4.40a corresponds to the non-conformal 3D rolling case studied in §4.3.9, with geometric and load parameters given in Table 4.28. Figure 4.40b corresponds to the conformal rolling case with geometry designated as "circ. s." studied in §6.2.4.1, with the same N and μ as the previous case, cf. Table 6.1 for other related data. Different steady rolling situations are simulated in each case, with pure longitudinal (ξ) or lateral (η) creepage as indicated in the legend of each figure. The creepages in Figure 4.40b are referred to the rigid contact point.

The abscissa axis of each graph is the rolled distance U_x divided by the longitudinal dimension of the contact patch $2a$. The ordinate axis is the

difference of the contact resultant frictional force during the FEM simulation F_i, with its corresponding reference value $F_{i,ref}$, divided by the coefficient of friction μ times the normal contact load N. The considered F_i is the longitudinal resultant frictional force F_x in the cases with pure longitudinal creepage, and the resultant frictional force in the YZ plane F_{lat} in the cases with pure lateral creepage. Their reference values have been calculated with CONTACT [CONTACT UG 2013] in the non-conformal case of Figure 4.40a, and with CECT in the conformal case of Figure 4.40b.



(a) Case of non-conformal rolling. (b) Case of conformal rolling.

Figure 4.40. Evolution of contact resultant forces in FEM rolling contact simulations.

The rolling phase in the cases shown in the figure is started with a preload of about 75% of the reference F_x and 50% of the reference F_{lat} in each case. Referring to Figure 4.40a, it is observed that the cases with longitudinal creepage converge to the steady state considerably faster than the cases with lateral creepage. This is attributed to the influence of the global lateral deflections of the bodies present in the cases with lateral creepage, which are bigger than the longitudinal ones in the cases with longitudinal creepage. A second aspect observed in these results is that stabilization is reached earlier in the cases with higher saturation level. This is mostly observed in the cases with lateral creepage. Similar comments are made to Figure 4.40b, except that the cases with longitudinal creepage and lower saturation level take more time to stabilize. Also, there are higher deviations between the F_{lat} values obtained from the FEM simulations and the reference values.

The necessary rolled distances in these simulations to get rolling contact solutions close to steady may be quoted in less than 2 times the contact patch length for the cases with longitudinal creepage, and between 2 and 3 times for the cases with lateral creepage.

The vertical dashed lines in the graphs of Figure 4.40 mark different rolling steps in the rolling phase of the simulations. Thus, there are up to 3 rolling steps. The mesh density and size of the substeps are different in each of the steps. The finest mesh density and substep size are those of the 2nd step. The size of each element of the mesh in rolling direction in the zone traversed in this step is such that the longitudinal axis of the contact patch accommodates about 36 elements in the two cases considered in the figure. The distance advanced by the wheel in each rolling substep of the 2nd rolling step equals the element length in the case of Figure 4.40a, and 0.96 times the element length in the case of Figure 4.40b. The substep size in the other steps is bigger than that of the 2nd step by a factor of 2.1 times in the case of Figure 4.40a, and 1.2 times in the case of Figure 4.40b. Performing rolling steps with coarser meshes and substep sizes prior to the most refined one in which the steady contact solutions are attained serves to minimize the rolled distance in the most refined step, which accounts for the greatest part of the cost of the simulation. The coarser mesh and biggest substep size in the 1st and 3rd rolling steps of the first case has an effect on the resultant frictional forces that may be appreciated in the right part of Figure 4.40a. It can be seen that the trends of the resultant forces change at the beginning of the 3rd step: after having reached stable values during the 2nd step, in the 3rd step they increase by about 2% of the μN value. The effect of the substep size on the contact results is discussed in more detail in §4.2.2.2.

4.2.2.2 Size of each rolling substep

The effect of the substep size in the rolling phase of the simulation on the computed FEM rolling contact solutions is studied here for a conformal contact case. The considered case has the “circ. a.” geometry defined in Table 6.1, with $\mu = 0.30$, $N = 80$ kN, zero yaw angle, and zero reference longitudinal and lateral creepages. Different simulations are run for this case on the same mesh and with the same rolled distance, with different number of substeps in the rolling phase.

Figure 4.41 shows the resultant F_{lat} force obtained at the end of the different simulations, after a rolled distance of 32 mm, which is about 1.9 times the contact patch length. It is plotted against the relative substep size, expressed as the distance advanced by the wheel in each substep, ΔU_x, divided by the length Δx of each element in the most refined zone of the mesh, where the rolling is simulated. Δx equals 0.372 mm, with which the longitudinal axis of the contact patch is discretized in about 45 elements.

F_{lat} decreases as the substep size is reduced, as may be seen in Figure 4.41. It is considered that the results obtained with higher ΔU_x / Δx above 1 are less precise, because the material elastic deformation velocities (the term D**u** / Dt in Eq. (1.7)) is less well discretized. The difference between the obtained F_{lat} for $\Delta U_x / \Delta x = 1$ and the extrapolated value for $\Delta U_x / \Delta x = 0$ is about 2.5% (or 1.5% of the μN value). It is not clear whether the extrapolated value for $\Delta U_x / \Delta x = 0$ is more precise than the value for $\Delta U_x / \Delta x = 1$. Tests with CECT with `tdgt`=2 in a 1D Carter / Fromm rolling problem suggest that the $\Delta U_x / \Delta x$ value to obtain the most precise solution could be around the range of 0.5–1. In this work $\Delta U_x / \Delta x$ values around 1 are used, as previously mentioned.

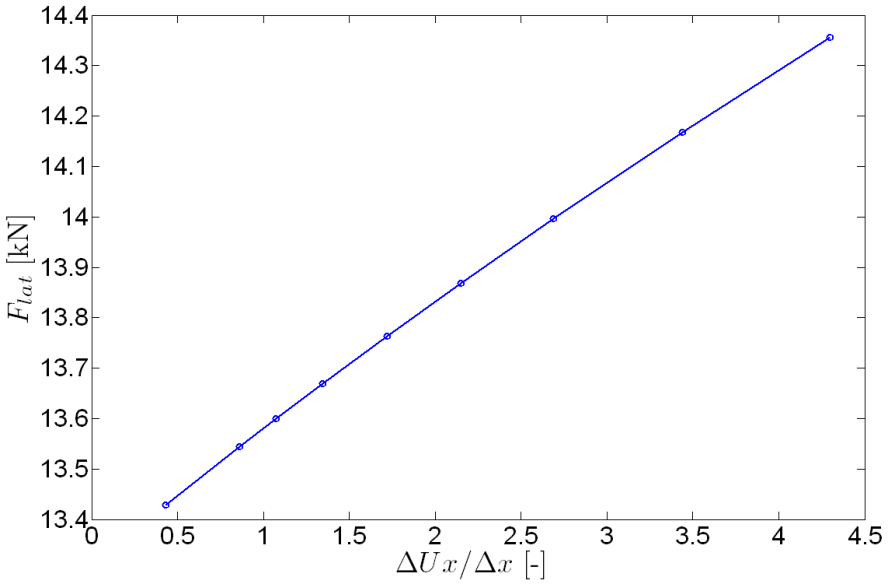
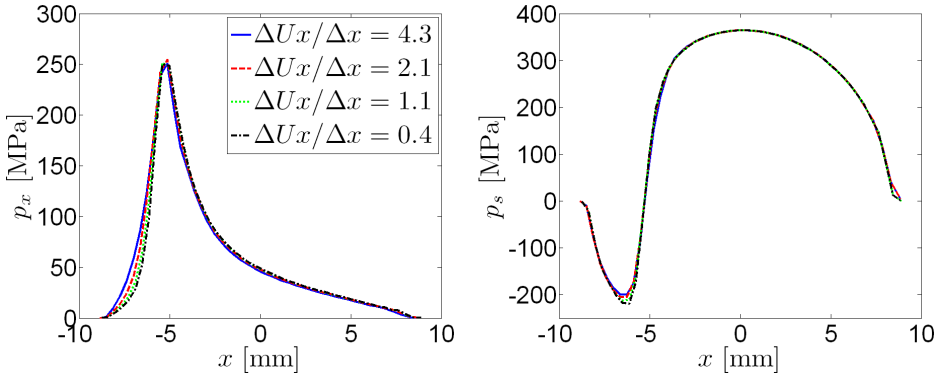


Figure 4.41. F_{lat} at the end of FEM rolling contact simulations with different substep sizes in rolling phase.

Figure 4.42 shows more detail of the frictional contact stress results of some of the FEM simulations considered here. The figure shows the profiles of the p_x and p_s components of tangential stress along the contact patch, close to its centreline, obtained after a rolled length of about 1.4–1.5 times the contact patch length. In this case, the saturation level is high and most of the contact patch is in slip, due to the high contact angle and associated spin. In consequence, the differences between the different solutions are rather limited. In cases with lower saturation, the substep size could have a greater impact on the results.



(a) p_x. (b) p_s.

Figure 4.42. Tangential contact stresses along the contact patch at s = –0.18 mm.

4.2.2.3 Penalty stiffness

In this section, the sensitivity of the FEM contact results to the penalty stiffness for the tangential contact is investigated. This penalty stiffness is defined in terms of an elastic slip tolerance (EST), which specifies the allowable elastic slip (cf. Figure 4.36b) as a fraction of a characteristic contact element dimension. Regarding the normal contact, the default penalty stiffness setting provided in the Abaqus FE models has been found to yield precise results, so no sensitivity has been carried out with it.

The case of conformal rolling studied in §6.2.4.2 is considered, with the "circ. l." geometry defined in Table 6.1, with reference creepages $\xi_0 = 0$, $\eta_0 = 0.001$ and $\varphi_0 = 0$. The curvatures in the central contact point are the same as in the case considered in §4.2.2.2, so the overall contact patch dimensions are similar

to that case. As a difference, in this case the contact patch is symmetric both longitudinally and laterally for zero yaw angle. The same total rolled distance and a similar mesh refinement as in the simulations reported in §4.2.2.2 are used here.

Figure 4.43 shows the normalized f_{lat} force, $f_{lat} = F_{lat} / \mu N$, obtained at the end of the FEM rolling contact simulations with different EST, namely EST = 0.0025, 0.0005 and 0.0001. The default value of the EST in the Abaqus contact FE models is 0.005. The EST values are plotted in the abscissa axis of the graph in logarithmic scale. As shown in the figure, F_{lat} decreases when increasing the EST. The longitudinal resultant F_x, smaller in magnitude in this case and not shown in the figure, follows a similar trend.

The differences between the solutions with EST of 0.0001 and 0.0005 are much smaller than those between the solutions with EST of 0.0005 and 0.0025. Still, the f_{lat} value obtained with the most stringent EST of 0.0001 falls far from the f_{lat} value computed with CECT for this case, which is 0.49. The differences between the CECT and FEM results in different conformal rolling contact cases with lateral creepage are examined further in §6.2.4. The drawback of increasing the penalty stiffness is increasing the number of iterations for convergence, as indicated in §4.2.1.3. This is illustrated in Table 4.14, which lists the number of iterations taken to complete the FEM rolling contact simulations with different EST values. The severe discontinuity iterations are related to the enforcement of the contact conditions, the equilibrium iterations to the fulfilment of the force equilibrium, and the total iterations are the sum of the former two. The highest increment is produced in the severe discontinuity iterations when reducing the EST from 0.0005 to 0.0001.

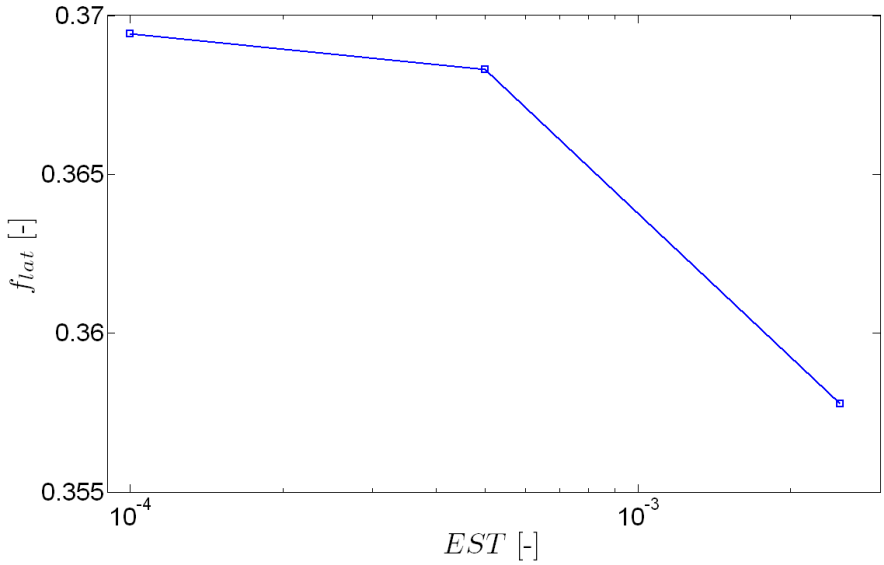
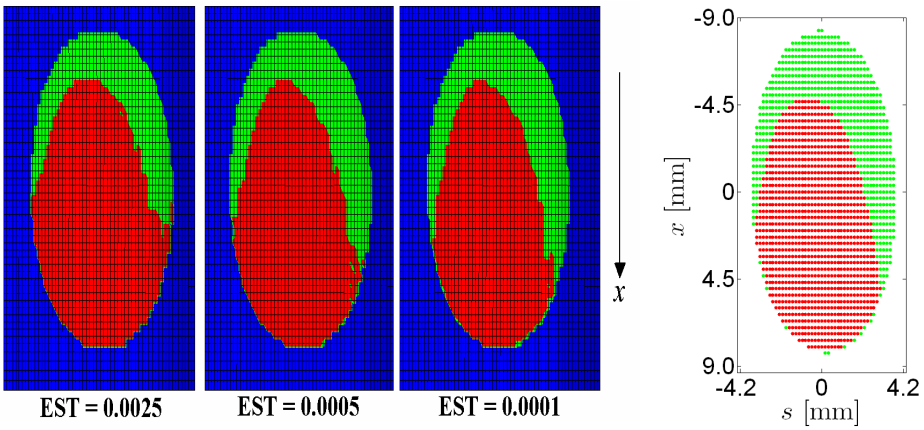


Figure 4.43. f_{lat} at the end of FEM rolling contact simulations with different EST.

Table 4.14. Iterations to complete FEM rolling contact simulations with different EST values.

EST	Severe discontinuity iterations	Equilibrium iterations	Total iterations
0.0025	306	114	420
0.0005	348	132	480
0.0001	485	138	623

Figure 4.44a shows the adhesion and slip areas in the contact patch obtained from the FEM simulations after a rolled distance of about 1.5 times the contact patch length, with the different EST values. The contact patches are shown as viewed from the top of the rail. The rolling (x) direction is also shown in the figure. Higher EST values result in more extended adhesion areas. The differences are noted mainly at the sides of the leading edge of the slip area. Figure 4.44b shows the adhesion and slip areas for the same steady rolling contact case computed with CECT. The slip area in the rear part of the contact patch obtained with CECT is larger than the slip area in the same zone of the FEM solutions. Otherwise, the adhesion and slip area distribution computed with CECT compares well with the FEM solutions obtained with the lower EST values.



(a) FEM solutions. (b) CECT solution.

Figure 4.44. Adhesion (in red) and slip (in green) areas obtained with different EST.

Considering these results, the EST has been set to 0.0005 in the FE contact models used in this thesis. The FEM contact results previously presented have been obtained with this setting as well.

4.2.2.4 Summary

The most relevant aspects which have been seen necessary to pay attention to for a proper tuning of the FE contact mechanics models described here are summarized in Table 4.15 for the Abaqus models, and in Table 4.16 for the ANSYS models. Suggested values for some numerical parameters are included, particularly when the values considered appropriate have been found to be considerably different from the program default ones. All these are to be regarded as orientative guidelines, and specific to the particular contact problems and objectives pursued here.

Table 4.15. Settings in Abaqus FE wheel–rail contact mechanics models.

Concept	Suggested setting / action	Notes
Contact constraint enforcement	Penalty (default).	
Initial contact closure	Define displacement-controlled step.	
Normal load incrementation	- No. increments ≈ no. elms in contact patch semi-axis. - Quadratic/cubic load ramp.	Relevant in static frictional contact analyses.
Rolled distance	$U_x / (2a) \approx 2$	- Lower distances may suffice with pure long. creep / higher saturation levels; longer distances may be necessary with lat. creep / lower saturation levels. - Tangential preload prior to rolling may help accelerate convergence to steady state. - Check evolution of frictional contact forces.
Substep size	$\Delta U_x / \Delta x \approx 1$	- Define rolling step(s) with coarser mesh / incrementation prior to the most refined one to minimize rolled distance in the latter.
Penalty stiffness	- Normal: default is OK. - Tangential: EST ≈ 0.0005.	Lower EST values may be necessary in 2D contact analyses (1 or 2 orders of magnitude).

Table 4.16. Settings in ANSYS FE wheel–rail contact mechanics models.

Concept	Settings
KEYOPTS	2: 0 5: 0 or 1 6: 0 10: 0 or 2
Real Constants [1]	FKN (3): 1–3 FTOLN (4): 0.03–0.1 ICONT (5): 0 FKT (12): 1–1.5 [2] SLTO (23): 3E–5 – 1 E–4 [3][4] TNOP (24): 0
Solver	NROPT,UNSYM

(1) Indicated values are factors applied on the default values.

(2) Even though seen as <auto> in the Friction menu of the Contact Wizard, the used value if left unchanged is 0, and not 1 which is its default value according to the Contact Technology Guide [ANSYS doc 2010]. It is important to ensure a correct setting for this parameter, because with a value of 0 the quality of the tangential solution may be seriously degraded.

(3) With the KEYOPT 10 of the CONTA174 elements set to 0, this setting is irrelevant.

(4) Convergence failed in a case in which this parameter was decreased to 1E–5.

The KEYOPTS and Real Constants listed in Table 4.16 apply to the CONTA174 contact (slave) elements. These are 3D 8-node elements which are used with TARGE170 “target” (master) elements to define surface-to-surface contact pairs. A brief description of the listed KEYOPTS and Real Constants is given below. More information may be found in the ANSYS Element Reference and Contact Technology Guide [ANSYS doc 2010].

- KEYOPTS:
  - 2: contact algorithm. Options:
    - 0: augmented Lagrangian.
    - 1: penalty.
    - 2: internal MPC.
    - 3: Lagrange multiplier on normal, penalty on tangent.
    - 4: pure Lagrange multiplier on normal and tangent.
  - 5: CNOF/ICONT automated adjustment. Options:
    - 0: no automated adjustment.
    - 1: close gap with auto CNOF.
    - 2: reduce penetration with auto CNOF.
    - 3: close gap / reduce penetration with auto CNOF.

        - 4: auto ICONT.
    - 6: contact stiffness variation. Options:
        - 0: use default range for stiffness updating.
        - 1: make a nominal refinement to the allowable stiffness range.
        - 2: make an aggressive refinement to the allowable stiffness range.
    - 10: contact stiffness update. Options:
        - 0: each load step if FKN is redefined during load step (pair based).
        - 1: each substep based on mean stress of underlying elements from the previous substep (pair based).
        - 2: each iteration based on current mean stress of underlying elements (pair based).
        - 3: each load step if FKN is redefined during load step (individual element based).
        - 4: each substep based on mean stress of underlying elements from the previous substep (individual element based).
        - 5: each iteration based on current mean stress of underlying elements (individual element based).

- Real constants:
    - FKN (3): normal penalty stiffness factor.
    - FTOLN (4): penetration tolerance factor.
    - ICONT (5): initial contact closure.
    - FKT (12): tangent penalty stiffness factor.
    - SLTO (23): allowable elastic slip.
    - TNOP (24): maximum allowable tensile contact pressure.

One of the differences between the Abaqus and the ANSYS models has been the adjustment of the penalty stiffness values. In the ANSYS models it was necessary to adjust the normal penalty stiffness settings (particularly the FTOLN Real Constant) to obtain precise results in the normal part.

Another issue experienced both with the Abaqus and with the ANSYS FEM contact models has been a relatively high level of noise in the tangential contact stress distributions. The following actions have mitigated the problem:

- Refine and increase mesh quality in contact area. Provide some layers of parallelepiped elements with nearly rectangular faces.
- Increase number of substeps.

4.2.3 Some pre- and post-processing calculations

4.2.3.1 Geometric contact search

It is necessary to solve the geometric wheel–rail contact search problem as a previous step to the solution of the contact mechanics problem, either if the contact mechanics problem is solved with FEM or with other contact models. In some cases, the wheel position relative to the rail is given as input, and the contact search is carried out to define the contact mesh (or the mesh of the PCS for CECT, for instance). In other cases, the intended location of the rigid contact point is the input, and some DOF of the wheel (normally its lateral position) has to be determined to produce the contact at that location. In these cases, the contact search problem is solved for a number of tentative wheel lateral positions, until the rigid contact point is found where intended.

Except for very simple wheel and rail geometries, the geometric contact search problem has to be carried out numerically. For a wheel with non-zero yaw angle, the contact search is 3D, but it may be reduced to 2D calculating the contact locus on the wheel surface as explained in [Wang K 1984]. This applies for the contact of a wheel on a straight rail. This procedure has been implemented as explained below. An extended procedure to calculate the wheel contact locus with a curved rail may be found in §4.2 of [Vollebregt 2022]. The 2D contact search problem consists on finding the minimum vertical distance between two curves; the rail profile, and the curve defined by the wheel contact locus. This is represented schematically in Figure 4.45.

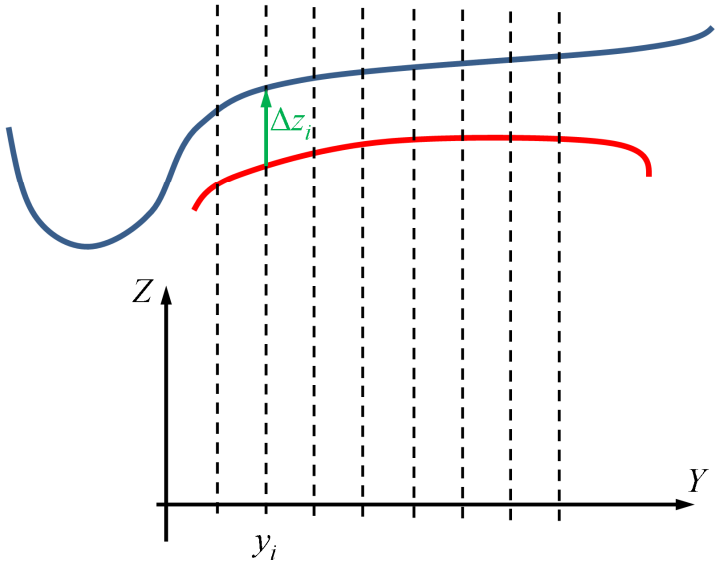


Figure 4.45. Geometric wheel–rail contact search.

The necessary data for the definition of the contact search problem are the parameters that define the geometry of the wheelset–track pair, and the necessary independent DOF-s of the wheel with respect to the rail. The geometry of the wheelset–track pair is defined by the wheel and rail profiles, the rail inclination, the nominal rolling radius of the wheel, and the track and wheelset widths (cf. the 1st line for each case in the main input file for CECT, §4.1.2.1.1). The relevant DOF-s of the wheel are given by the displacements u_y and u_z in the YZ plane, and the roll and yaw angles ϕ and ψ. Two calculation types are implemented:

- Wheel–rail pair. A single wheel–rail contact pair is taken into account (left or right). In this case, the independent (input) wheel DOF-s are u_y, ϕ, and ψ; and u_z is to be determined. The main steps of the calculation are as follows:
  - Calculate the wheel contact locus with Wang's method. The coordinates in the global system of the possible contact point C of the rolling circle located at the lateral position y_w in the wheelset local coordinate system are given by Eqs. (4.78)–(4.80).

$$x_C(y_w) = l_x \times y_w - l_x \times R \times \tan\delta \tag{4.78}$$

$$y_C(y_w) = u_y + l_y \times y_w + \frac{R}{l_y}\left[l_x^{\,2} \tan\delta - l_z\chi \right] \tag{4.79}$$

$$z_C(y_w) = R_{roll,n} + u_z + l_z \times y_w + R\chi \tag{4.80}$$

The angle δ in these equations is the contact angle of the wheel profile in its local coordinate system, positive as depicted in Figure 4.3 (the corresponding angle δ for the other wheel of the wheelset, not shown in the figure, is also positive). The quantity χ used in Eqs. (4.79) and (4.80) is given in Eq. (4.81), and the rest of the parameters are as explained for Eqs. (4.1)–(4.14) in §4.1.1.2. These equations are valid for the wheel located in the positive y_w side, as the one shown in Figure 4.3. For the other wheel, these equations are applied changing the sign of δ. In Eq. (4.78) it is assumed that the wheelset centre is located at $x = 0$ along the global X axis.

$$\chi = \frac{l_x^{\,2} l_z \tan\delta - l_y \sqrt{1 - l_x^{\,2}\left(1 + \tan^2\delta\right)}}{1 - l_x^{\,2}} \tag{4.81}$$

- Position the rail profile in the global coordinate system. The coordinates $\{y, z\}$ in the YZ plane of a point with coordinates $\{y_L, z_L\}$ in the rail profile local coordinate system are given by Eqs. (4.82) and (4.83). In these equations, the ± sign applies to the rails located in the positive and negative side of the Y axis, respectively. It is applied to the y_L coordinate to mirror the rail profile in the negative side of the Y axis, assuming that the profile given in the local $\{y_L, z_L\}$ coordinates corresponds to the positive side. ϕ_r is the rail inclination angle, and *hw_tr* the track half-width (cf. the input parameters `incl_r` and `hw_tr` in §4.1.2.1.1). Normally ϕ_r will have nominally equal and opposite values for both rails.

$$y = \pm y_L \cos(\phi_r) - z_L \sin(\phi_r) \pm hw\_tr \tag{4.82}$$

$$z = \pm y_L \sin(\phi_r) + z_L \cos(\phi_r) \tag{4.83}$$

- Map the wheel contact locus and rail profile to common y_i positions in the global Y axis, interpolating as necessary from the sets of $\{y, z\}$ coordinates given by Eqs. (4.79) and (4.80) for the wheel, and (4.82) and (4.83) for the rail. Other quantities like the slopes and curvatures associated to each point may likewise be interpolated. These may serve to estimate the extent of the contact patch.

- Calculate the vertical distance Δz_i between both profiles at each y_i position, as depicted in Figure 4.45. The position y_i at which the contact takes place is found as the one with the minimum Δz_i value. The contact condition is that the minimum $\Delta z_i = 0$, so the wheel vertical displacement u_z to comply contact in the wheel–rail pair is equal to its current value minus the minimum Δz_i.

  The contact location may be refined beyond the resolution provided by the chosen set of y_i lateral positions, using the condition that at the rigid contact point the slopes in the YZ plane of the wheel contact locus and the rail profile coincide. If the exact Y coordinate of the rigid contact point is y^*, in the neighbourhood of y^* the slope $\tan(\alpha) = dz / dy$ of the upper profile (the wheel contact locus) in the global coordinate system must be equal or lower than that of the lower profile for $y < y^*$, and equal or higher for $y > y^*$, for the contact point to be feasible. This enables to determine the pair of y_i points between which y^* may be found by interpolation. The calculation of the angles α of the rail profile is straightforward, as $\alpha_r = \pm\alpha_{L,r} + \phi_r$; being $\alpha_{L,r}$ the angle of the rail profile in its local $\{y_L, z_L\}$ system; i.e. $\alpha_{L,r} = \text{atan}(dz_L / dy_L)$, and α_r the angle in the global system. $\alpha_{L,r}$ is used with + sign for the rail in the positive side of the Y axis, and with – sign for the rail in the negative side, as y_L in Eqs. (4.82) and (4.83), so that the rail profile is mirrored as explained before.

  The slopes $\tan(\alpha_w)$ for the wheel may be approached by finite differences $\Delta z / \Delta y$, or else they may be calculated more precisely

based on the formulation of the normal **n** to the wheel surface, depicted in Figure 4.46. Here a conical wheel is depicted for ease of illustration, but this formulation is applicable to any surface of revolution, as the normal to a surface of revolution of any shape is equal to the normal to the tangent cone to it at the considered point. **n** is seen in true magnitude only in view *A* of the figure.

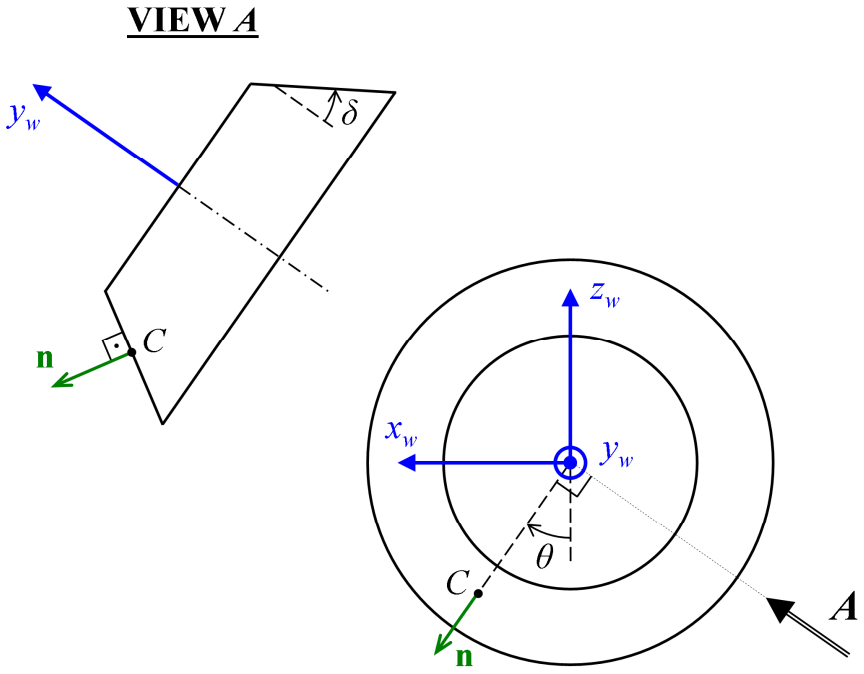


Figure 4.46. Normal to the wheel surface.

n is given in the wheelset $\{x_w, y_w, z_w\}$ coordinate system according to Eq. (4.84). Here $\mathbf{i_w}$, $\mathbf{j_w}$ and $\mathbf{k_w}$ are the unitary vectors along the axes x_w, y_w and z_w.

$$\mathbf{n} = c_\delta \, (s_\theta \, \mathbf{i_w} - c_\theta \, \mathbf{k_w}) + s_\delta \, \mathbf{j_w} \tag{4.84}$$

The transformation between the $\{x_w, y_w, z_w\}$ and $\{X, Y, Z\}$ coordinates is similar to that given in Eq. (4.18), with angles ϕ and ψ in place of ϕ_ω and ψ_w (cf. Figure 4.3). The components of **n** in the $\{X, Y, Z\}$ axes,

n_X, n_Y, and n_Z, are given in Eqs. (4.85)–(4.87). Eqs. (4.84)–(4.87) apply for the wheel located in the positive y_w axis. For the other wheel, the sign of δ is changed.

$$n_X = c_\delta\, s_\theta\, c_\psi - s_\delta\, s_\psi\, c_\phi - c_\theta\, c_\delta\, s_\phi\, s_\psi \tag{4.85}$$

$$n_Y = c_\delta\, s_\theta\, s_\psi + s_\delta\, c_\psi\, c_\phi + c_\theta\, c_\delta\, s_\phi\, c_\psi \tag{4.86}$$

$$n_Z = s_\delta\, s_\phi - c_\theta\, c_\delta\, c_\phi \tag{4.87}$$

At the possible contact point C, the normal to the wheel surface is aligned with the normal to the rail surface, which is contained in the YZ plane at any point for a straight rail. So, n_X must be 0. Eq. (4.85) with this condition could also be used to calculate the wheel contact locus. But this equation requires iterative solution, so it would be less efficient than using the closed form solution given directly by Eqs. (4.78)–(4.80).

Finally, the slopes $\tan(\alpha_w)$ are calculated as $\tan(\alpha_w) = -n_Y / n_Z$.

- Wheelset–track pair. Both wheel–rail contact pairs of the wheelset are taken into account. In this case, the independent (input) wheel DOF-s are u_y and ψ. u_z and ϕ are to be determined so that each wheel contacts with the corresponding rail. This is done solving both individual wheel–rail pairs as outlined above with fixed tentative $\phi^{(k)}$ values, which are iterated until the same minimum Δz_i value results in both wheel–rail pairs. Figure 4.47 shows the flowchart with the basic steps of the process.

  The + and – subindices refer to the wheel–rail pair in the + and – sides of the Y axis, and the superindices in parentheses to the iteration number. The first step in each iteration is solving both wheel–rail pairs (each of them individually), with a tentative ϕ. This yields the lateral position of the contact point (y^*) and the minimum vertical distance ($\min(\Delta z_i)$) in each wheel–rail contact pair. In this step, the wheelset is not moved. In the next step, the vertical displacement u_z of the wheelset is updated, so that contact is achieved in one of the wheel–rail pairs, without interpenetration in the other. If, after this movement, the new minimum

vertical distances are within a small tolerance value (e.g. 2.0E–3 μm) in both wheel–rail pairs, the solution of the wheelset–track pair has been accomplished. Otherwise, ϕ is updated, trying to equalize the minimum vertical distances of both wheel–rail pairs, and a new iteration is carried out.

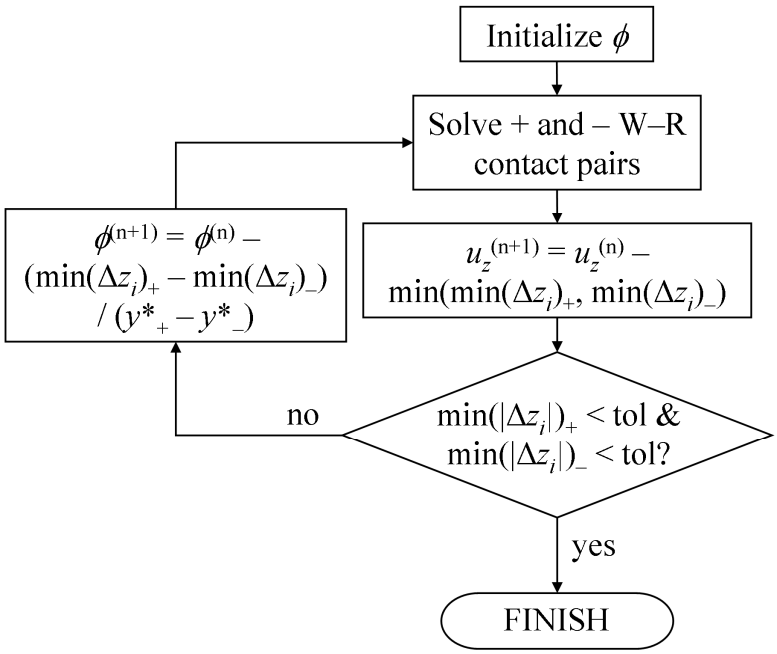


Figure 4.47. Flowchart of the geometrical contact problem solution for a wheelset–track pair.

Figure 4.48 shows an example graphical output of the solution of the geometric contact search in a wheel–rail contact pair. This includes the representation of the contacting rail profile and wheel contact locus in the *YZ* plane, the location of the rigid contact point, and the vertical distances between both profiles.

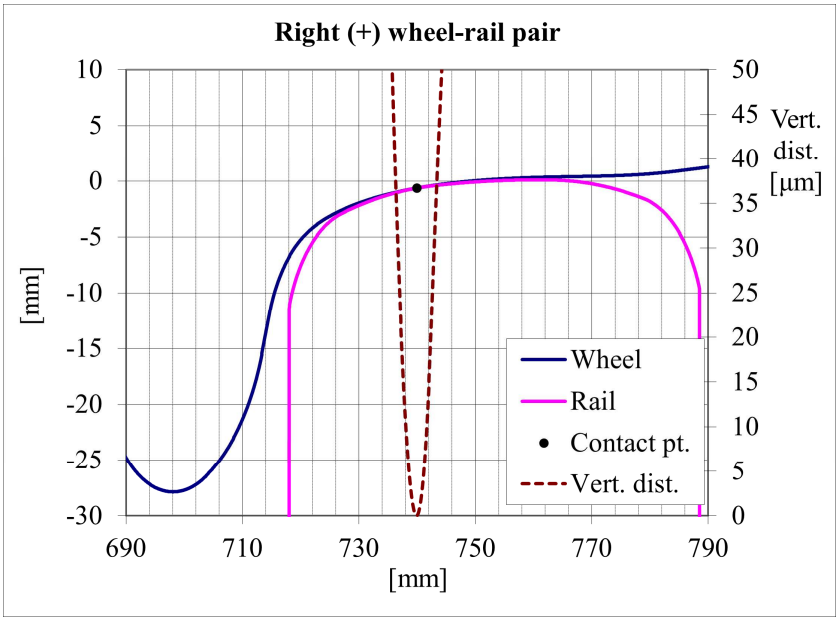


Figure 4.48. Graphical representation of the solution to the geometric contact search in a wheel–rail pair.

4.2.3.2 Adaptation of output to CECT format

Part of the post-processing of the FEM results consists of arranging the nodal contact surface results in a format similar to the CECT output, so that the results from the different sources can be compared side by side. Table 4.17 lists the contact surface nodal field output quantities from the Abaqus FE models that are used for this purpose, and their correspondence with the CECT output arrays (cf. §4.1.2.2.1). These are retrieved from the rail (prismatic) surface for convenience in these wheel–rail contact FEM models, which is defined to be the slave surface in the contact pair with the wheel. On the other hand, in the rolling element–raceway contact FEM model for rolling bearings described in §7.2.3, the results are retrieved from the rolling element surface, defined as the slave surface in those models.

Table 4.17. Correspondence of Abaqus contact surface nodal field outputs with CECT output arrays.

Abaqus[(1)]	**CECT**
CPRESS	`pn`
CSHEAR*n*[(2)]	`px`, `ps`
CSTATUS	`eldiv`
CSLIP*n*[(2)(3)]	`srel`

Notes to Table 4.17:

(1) See §4 of Abaqus Analysis User's Manual [Abaqus doc 2012].

(2) n = 1, 2. The correspondence of the 1 and 2 directions of the contact surface with the x and s directions is determined based on the orientation of the contact surface in the global coordinate system; cf. §35.3.1, 37.1.1, 1.2.2, and Figure 1.2.2-1 of the Abaqus Analysis User's Manual [Abaqus doc 2012].

(3) The calculation of the slip velocities based on the FEM output CSLIP*n* is explained below.

Having the sides of the rectangular element faces in the FEM contact surface aligned with the contact x and s directions facilitates the post-processing of the tangential results, as the main directions of the surface correspond directly to the x and s directions. Consequently, no transformation is necessary of the FEM tangential contact results, given by their two components in the contact surface main directions, except for a sign change in some cases. The arrangement of the contact surface element faces also facilitates the numbering of nodes and elements, which is done in a similar way as in the mesh used in the exact contact theory (cf. Figure 2.19), though in this case the mesh may not be uniform.

The derivation of some relevant tangential contact quantities from the FEM simulations is explained next.

- Slip velocities.

  The tangential slip velocities v_x, v_s are calculated using the contact nodal field outputs CSLIP*n*. CSLIP*n* are the accumulated slipped distances in each node. Therefore, the slip velocities are proportional to the increments of CSLIP*n*. These increments are calculated subtracting the

CSLIP*n* values of two consecutive simulation substeps at each node. The slip velocities are then calculated dividing these increments by the time increment between the two substeps. In a rolling contact simulation, this time increment is equal to ΔU_x / V_c. V_c is defined arbitrarily, externally to the FEM simulation (it is not used in the FEM simulation, which is static).

- Frictional work.

In some contact wear models, wear is assumed to be directly related to the frictional work density, see e.g. [Popov 2010]. Therefore, this magnitude is of primary interest in the study of wear in the contact interface. The frictional power density at each point (x, s) in the contact surface, P_{fric}, is equal to $\|\mathbf{p_t}\|\cdot\|\mathbf{v}\|$. This is defined in units of power per unit area. The frictional work in the rail after one wheel passage, W_{fric}, is calculated integrating P_{fric} along x according to Eq. (4.88). In this equation x_{ledg} and x_{tredg} are the x coordinates of the leading and trailing edge of the contact patch.

$$W_{fric}(s) = \frac{1}{V_{c,r}} \int_{x_{ledg}(s)}^{x_{tredg}(s)} P_{fric}(x,s)\, dx \tag{4.88}$$

The discretization of this equation leads to Eq. (4.89). The summation in this equation is extended over the contact surface elements e with nodes in a given s_i lateral position of the mesh. $x_1(e)$ and and $x_2(e)$ are the longitudinal coordinates of the leading and trailing nodes of an element e.

$$W_{fric}(s_i) = \frac{1}{V_{c,r}} \sum_{e(s_i)} \frac{P_{fric}(x_1(e),s_i) + P_{fric}(x_2(e),s_i)}{2} (x_1(e) - x_2(e)) \tag{4.89}$$

This computation is done similarly with the CECT results. In the case of CECT, ($x_1 - x_2$) = `dx` is constant. On the other hand, for `c_o_n` = 0, the averaging on P_{fric} expressed in Eq. (4.89) is avoided, using instead the value computed at the centre of each element, assumed uniform in the whole element.

Calculating P_{fric} as $\|\mathbf{p_t}\| \cdot v_{rel} \cdot V_{c,r}$ instead of $\|\mathbf{p_t}\| \cdot \|\mathbf{v}\|$, the dependence of W_{fric} on $V_{c,r}$ is seen to vanish in Eqs. (4.88) and (4.89).

- Rigid slip velocities.

  These are not derived from the FEM contact outputs, but they are calculated externally based on the defined rigid body motion for the wheel. The calculation is carried out as explained in §4.1.1.3, starting for convenience from a point P on the wheel axis of revolution (which is defined as the reference point of the wheel), of which its position along the simulation is output.

Some of the contact surface FEM results are written to a .mat text file, similar in format to the CECT .mat output file (cf. §4.1.2.2.1). The following arrays are written to this file: `I`, `eldiv`, `s`, `mu`, `pn`, `px`, `ps`, `X_def`, `Y_def`, `Z_def`, `srel`, `wx`, `ws`. `X_def`, `Y_def` and `Z_def` are the deformed coordinates of each node, in a coordinate system with origin in the contact reference point and axes aligned with the global {*X*, *Y*, *Z*} system. The contact reference point is prescribed externally, being normally the rigid contact point or some point near the centre of the contact patch. Its global coordinates are (X_{ref}, Y_{ref}, Z_{ref}). `s` is the lateral (*s*) coordinate, with origin as well in the contact reference point, and calculated with deformed coordinates as explained below. The rest of the arrays are as explained in §4.1.2.2.1 for CECT. `X_def`, `Y_def`, `Z_def` and `s` are output for all the nodes of the contact patch, because with deformed coordinates they are in general unique for each node. E.g., with `ms` lateral mesh positions, there are not only `ms` unique values of `s`. However, being the surface element faces rectangular, or nearly so, and the displacement gradients small, all the nodal coordinates will fall near one of the `mx` values in longitudinal direction, or `ms` values in lateral direction, defining a mesh of rectangular elements.

The necessary data to produce this results file are as follows:

- FEM nodal data:
  - The coordinates (at least the undeformed and optionally also the deformed ones) of each node in the contact patch.
  - The contact field outputs listed in Table 4.17.

- Specified rigid body kinematics of the wheel: forward velocity, rotation velocity, yaw and roll angles, and position of a point of the wheel axis of revolution.
- The size of each substep.
- The position of the contact reference point, and angle of the contact surface in the *YZ* plane at that point.

All these data is retrieved for the substep of interest, and in the case of CSLIP*n*, also for the previous substep, if slip velocities are to be calculated. The main steps for the production of the .mat file with the FEM outputs are listed next.

- Selection of the set of nodes in the rail (slave) surface covering the whole contact patch, and export of the tables containing the nodal coordinates and contact field output data. The data of these tables is then brought together to a single table.
- Extraction of the profile in the *YZ* plane of the contact surface encompassing the whole contact patch, similar to the PCS shown in Figure 4.2b. This profile is represented with `ms` points in the *YZ* plane. The extraction of these points is carried out in the following steps:
  - Location of the node with the maximum contact pressure in the contact patch. This node has coordinates (X_c, Y_c, Z_c) in the global coordinate system. With a smooth contact pressure distribution, the longitudinal and lateral cross-sections of the contact patch passing through this node are normally the longest ones in each direction, or close to the longest.
  - Retrieval of the coordinates in the *YZ* plane (both undeformed and deformed) of all the nodes with undeformed longitudinal coordinate within the range $X_c \pm \Delta x_{min}$, being Δx_{min} the smallest longitudinal dimension of the elements of the contact patch. That is, the profile of the contact surface passing through $X = X_c$ is extracted.
  - It must be taken into account that for contact patches that may not be symmetric in x nor in s directions, like the ones that may arise in conformal contact (e.g. like those shown in §6.2.1.1 and 6.2.1.2), the entire profile of the contact surface circumscribing the whole contact

patch may not be covered by the set of nodes contained in a single longitudinal position. An example is shown in Figure 4.49, marking a cross-section of the contact patch in red, and the entire range in the *s* direction spanned by the contact patch. So, all the nodes are checked, to make sure that all existing mesh positions in the *YZ* plane are covered. For each node, the following steps are carried out for this purpose:

- Check if the undeformed global *Y* coordinate of the current node, Y_i, already exists in the set of *Y* coordinates of the nodes extracted in the previous step around $X = X_c$.
  - If it exists (within a predefined small tolerance value), no further action is carried out on this node and the loop continues with the next node.
  - Otherwise, new *YZ* coordinates are added to the profile of the contact surface. For this purpose, first the set of nodes with (approximately) the same undeformed *Y* coordinate as the current node is retrieved. From this set, the node closest to $X = X_c$ is selected, and the *YZ* coordinates of that node are added to the set of *YZ* coordinates representing the profile of the contact surface.

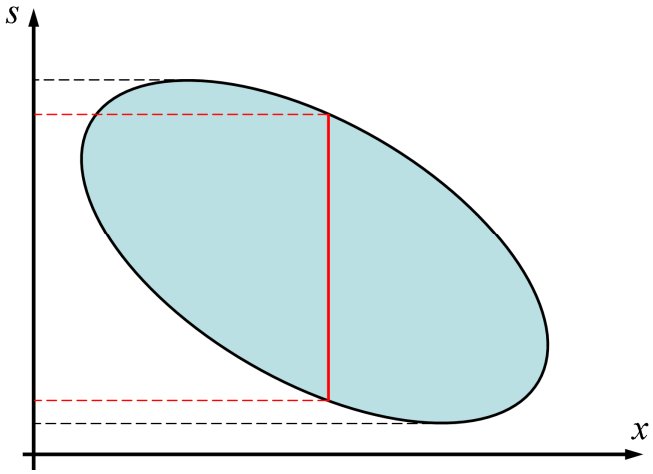


Figure 4.49. Example in which the entire range of the contact patch in the *s* direction cannot be covered by a single cross-section.

- Sorting of the set of *YZ* coordinates representing the profile of the contact surface by ascending *Y* coordinate.
- Computation of the *s* coordinates on this profile (both undeformed and deformed). The *s* coordinates are computed incrementally, considering a piecewise linear representation of the profile, according to Eq. (4.90). In this equation, the subindices *i* and *i*+1 refer to the node number, in ascending *Y* coordinate order. In a first step, the $s = 0$ coordinate is located in the first node of the profile. Secondly, the s_{ref} coordinate of the contact reference point is calculated, interpolating on the $s(Y)$ curve with its Y_{ref} lateral coordinate. Finally, the origin of the *s* coordinate is located on the contact reference point, subtracting s_{ref} to the *s* values calculated firstly.

$$s_{i+1} = s_i + \sqrt{\left(Y_{i+1} - Y_i\right)^2 + \left(Z_{i+1} - Z_i\right)^2} \tag{4.90}$$

In the case of the rolling bearing FEM contact models described in §7.2.3, the process for the extraction of the profile of the contact surface is similar to that described here, but it is done in a cylindrical system aligned with the axis of revolution of the rolling element, instead of the global $\{X, Y, Z\}$ system. In this process, nodal coordinates given in the $\{X, Y, Z\}$ system are transformed to the mentioned cylindrical system of the rolling element, using the formulation presented in §7.1.1.

- Calculation of the s_i coordinate for all the nodes, interpolating in the $s(Y)$ profile obtained in the previous step with the Y_i coordinate of each node. This is done both with deformed and undeformed coordinates. The s_i values computed with deformed coordinates are written to the .mat output file as the previously listed `s` values. The s_i values computed with undeformed coordinates are used to determine the *is* index of each node, i.e. its lateral position index in the rectangular mesh. This is equal to the node number on the profile computed in the previous step having the nearest *s* coordinate.
- Extraction of the set of *x* coordinates circumscribing the contact patch. A first set of *x* coordinates is extracted from the *is* lateral position (longitudinal strip) of the mesh having more nodes. As in the lateral

direction, all the existing *x* positions may not be covered by the set of nodes in a single *is* position. Therefore, the coordinates of all nodes are checked to form the entire set of *x* positions of the mesh, in a similar way as explained above for the *s* positions.

- Calculation of the 1D `I` index for each node. Firstly, its *ix* index is determined, in a similar way as for the *is* index explained above. Then `I` is computed as `I` = *ix* + (*is* – 1)`mx`, with `mx` being the number of *x* positions of the mesh.

4.3 Validation of models

This section shows results obtained with CECT and FEM for different contact test cases for which results are available in the literature, with the aim of validating the numerical contact models described in this chapter. Test cases with different characteristics have been considered for this purpose, as listed in Table 4.18. The following features are specified in the table for each case:

- Dim.: dimensionality of the problem, 2D or 3D. If 2D, the contact surface direction in which the contact is analyzed is indicated in parenthesis (the contact is considered infinite in the other direction, i.e., plane strain is considered).
- `tdgt`: cf. §4.1.2.1.1. A new integer value of –1 is added here, to distinguish problems with and without friction. Thus, in this table `tdgt` = –1 stands for frictionless contact problems, and `tdgt` = 0 for static problems with friction.
- Sim. mat.: elastically similar materials. Yes (Y) or No (N).
- Conformal: conformal contact. Yes (Y) or No (N). Non-conformal contacts are always planar, while conformal contacts may be planar or non-planar. A P in parenthesis is added in this column for conformal contacts that are planar.

Table 4.18. Validation cases for numerical contact models.

Test case	Dim.	`tdgt`	Sim. mat.	Conformal
Hertzian conformal	3D	–1	Y	Y (P)
Cattaneo shift	3D	0	Y	N
Oblique compression	3D	0	Y	N
Spence compression	3D	0	N	N
2D rolling	2D (*x*)	2	Y	N
Cattaneo to Carter	3D	1	Y	N
Cylindrical contact	2D (*s*)	–1	Y	Y
3D rolling without slip	3D	2	Y	N
Steady rolling	3D / 2D	2	Y	N

In the next subsections results for each test case are reported. Except where noted, steel is the considered material for both contacting bodies, with E (Young's modulus) = 210 GPa and ν (coefficient of Poisson) = 0.30.

4.3.1 Hertzian conformal contact

This is a Hertzian case, in which the contacting surfaces are conformal in lateral (s) direction, and in which the normal load N is small enough to have a nearly planar contact patch. A small yaw angle is applied to the wheel (the upper body), so that the principal curvature planes of both contacting bodies are not aligned. The parameters of the case are listed in Table 4.19. Convex curvature radii are positive, and concave ones negative.

Table 4.19. Input parameters of Hertzian conformal test case.

Parameter	Description	Value	Units
R_{sr}	Lateral radius of curvature of rail	10.0	mm
R_{xw}	Longitudinal radius of curvature of wheel	500	mm
R_{sw}	Lateral radius of curvature of wheel	–10.5	mm
N	Normal load	100	N
γ	Yaw angle	20	mrad
δ	Contact angle	0	mrad

As $\delta = 0$, the yaw rotation is perpendicular to the contact surface, and coincides with the angle between the principal planes of curvature of both contacting bodies. The principal curvatures of the resulting combined undeformed distance between the two surfaces, and the orientation of the planes containing them, may be computed with Eqs. (2.20)–(2.23).

Figure 4.50 shows the contour of the contact patch computed with CECT and the corresponding Hertzian ellipse, which closely matches it. The maximum normal pressure and the approach computed with CECT are 57.0356 MPa and 0.206689 μm, respectively. These values are also in close agreement with the Hertzian results, which are 57.0160 MPa and 0.206695 μm. It is verified as well that the normal pressure distribution computed with CECT is very close to ellipsoidal.

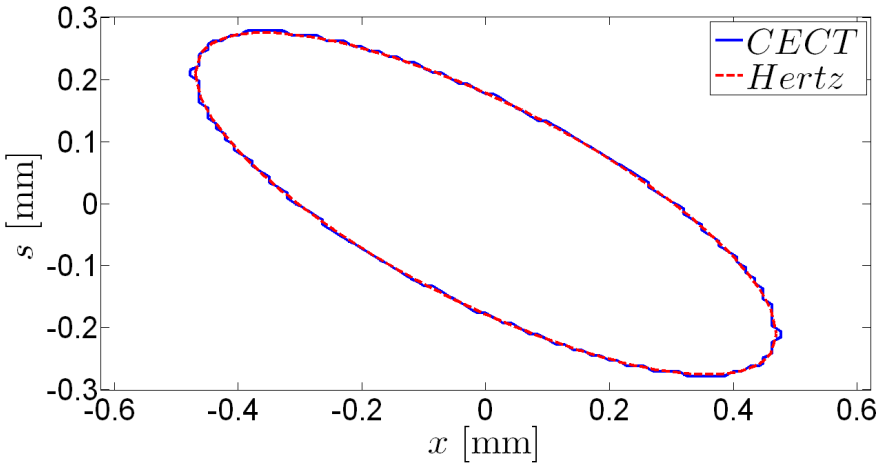


Figure 4.50. Contact patch of Hertzian conformal case computed with CECT and with Hertzian formulae.

A noteworthy point in this example is the large angle β of the contact ellipse with respect to the rotation γ applied to the wheel (cf. Figure 2.3). In this case, $\beta \approx -23.5\gamma$. This effect is discussed further in §6.2.1.2.

4.3.2 Cattaneo shift

This is a shift contact problem, in which two elastically similar bodies of Hertzian geometry are first brought into contact by a normal load N, and subsequently pulled tangentially. While the tangential force is below the limiting value μN, part of the contact area remains in adhesion, while the other part undergoes a small slip. The solution to the 2D problem was first presented by Cattaneo [Cattaneo 1938]. The tangential stress distribution is found as the superposition of two concentric and proportionally scaled elliptical distributions, as depicted in Figure 4.51. There is a central stick region, and slip

regions in the adjacent to the borders of the contact patch, which grow progressively towards the centre as the tangential load increases. The component p_{x1} shown in the figure is equal to the traction bound g. This tangential stress distribution resembles the one for 1D steady rolling (cf. §2.1.2.1), with the stick region being in the central part of the contact patch instead of adjacent to the leading edge.

The solution for the 3D case may be found similarly, superposing two ellipsoidal tangential stress distributions aligned with the shift direction, as explained in §7.2(d) of [Johnson 1987] for the case of a circular contact. The contact patch is divided in a central adhesion region and a peripheral slip region, similar to the 2D case. This solution produces uniform tangential elastic displacement differences in the adhesion region, with values $u_x = W_x$, $u_s = 0$, and thus complies exactly with the stick condition. Not only the tangential displacement differences, but also the tangential displacements u^k_x of each body $k = 1, 2$ are uniform, and $u^k_s = 0$, inside the stick region[14]. On the other hand, the slip condition is not fulfilled exactly, since the lateral displacement differences u_s are non-zero in the slip region, and thus the slip is not aligned with the tangential stress. Nevertheless, these lateral displacements, and the associated error, are small, and this analytical solution is a good approximation of the exact solution.

[14] This, which is evident for contact between elastically similar solids, as well as between an elastic solid and a rigid solid, is also fulfilled for the general case of two elastic solids with different elastic constants. It is evident for two elastically similar solids, because any non-uniform distribution of tangential elastic displacements in both solids, fulfilling the necessary (anti)symmetries, would result in a non-uniform displacement difference distribution (twice the value of the individual distributions). In the case of non-similar elastic solids, the tangential elastic displacements are taken as the differences from the base state of normal compression before the tangential shift, assumed to have been attained without friction. The uniformity in the tangential elastic displacements is fulfilled taking into account the contributions of the tangential stresses, and the difference of normal pressures from the base state of normal compression to the final state after applying the shift with friction.

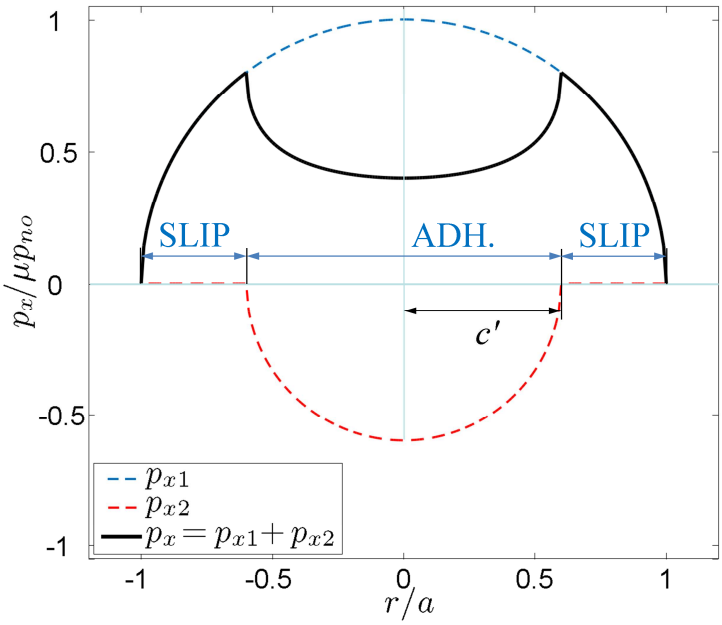


Figure 4.51. Tangential stress solution of the Cattaneo shift contact problem.

Eqs. (4.91)–(4.93) describe the analytical solution of the tangential stresses for a circular contact. In these equations r' is equal to the radial coordinate from the centre of the contact patch divided by the contact patch radius a, c' is the radius of the central adhesion region c divided by a, and p_{no} is the maximum Hertzian normal pressure, which is found in the contact centre. G_r and G_w are the shear moduli of each contacting body, and v_r and v_w their coefficients of Poisson. The maximum tangential stress is found at the limit between the adhesion and slip zones (cf. Figure 4.51), and is equal to $\mu\ p_{no}\sqrt{1-c'^2}$.

$$-\mathrm{sgn}(W_x)\times\frac{p_x}{\mu p_{no}}=\begin{cases}\sqrt{1-r'^2}-c'\sqrt{1-\left(\dfrac{r'}{c'}\right)^2} & \text{for } r'<c' \\ \sqrt{1-r'^2} & \text{for } c'\le r'\le 1\end{cases} \tag{4.91}$$

$$\frac{F_x}{\mu N} = 1 - \left(1 - \frac{16 a W_x}{3 \mu N \left(\frac{2-\nu_r}{G_r} + \frac{2-\nu_w}{G_w} \right)} \right)^{3/2} \leq 1 \tag{4.92}$$

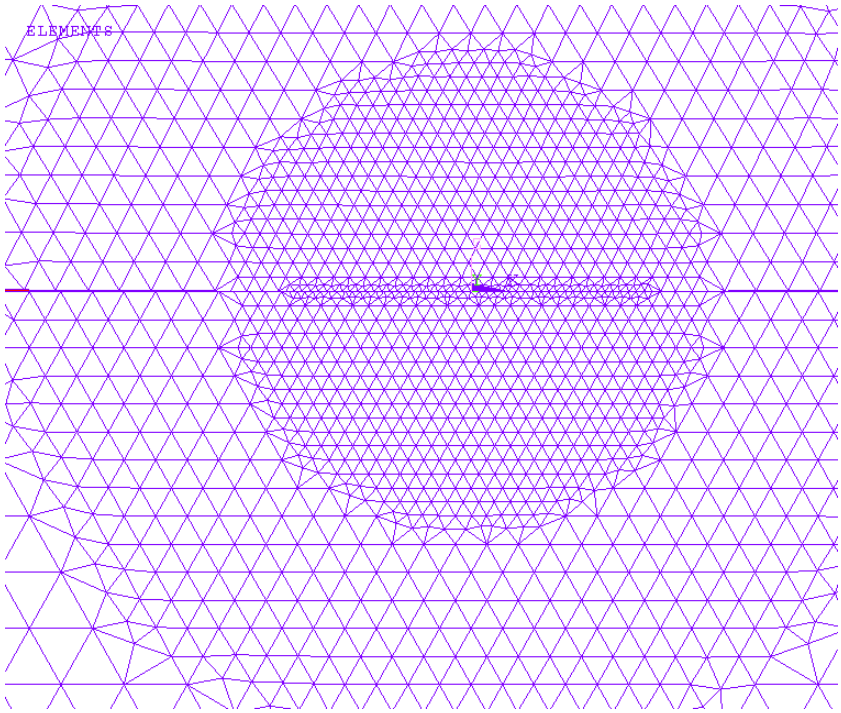
$$c' = \frac{c}{a} = \left(1 - \frac{F_x}{\mu N} \right)^{1/3} \tag{4.93}$$

The contact of a sphere on a plane is considered for this test case, with the input parameters given in Table 4.20.

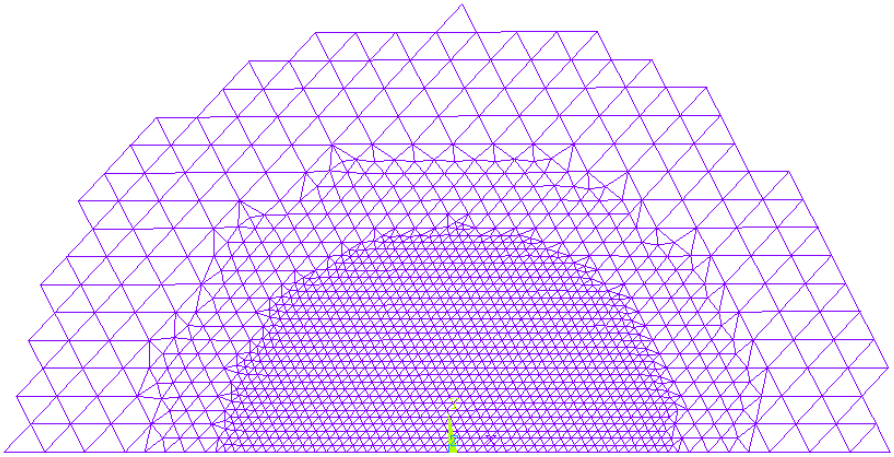
Table 4.20. Input parameters of circular contact shift test case.

Parameter	Description	Value	Units
R_x, R_s	Combined principal radii of curvature	500.0	mm
d	Approach	6.0	μm
W_x	Tangential shift	1.0; 2.0	μm
μ	Coefficient of friction	0.30	-

The case is computed with CECT and with FEM. The used FEM in this case is built in ANSYS. As a difference with the models described in §4.2.1, in this case the volume is meshed with quadratic tetrahedron elements (SOLID187). Only half of the geometry is meshed, making use of the longitudinal symmetry plane. Figure 4.52 shows part of the mesh of the FE model near the contact.



(a) Side view.



(b) Plan view.

Figure 4.52. Mesh around the contact of the FE model for the sphere-on-plane shift test case.

The computation with CECT is carried out in a single step or increment for the normal loading and another one for each of the two tangential shifts indicated in Table 4.20. In the FEM analysis, each step is divided in two substeps. The numerical results obtained with CECT and FEM are compared with the reference (Cattaneo) analytical results in Figure 4.53 and Table 4.21. Where two values are given in the table, these correspond to each of the W_x values listed in Table 4.20. The agreement of the CECT results with the reference is excellent, while the FEM results show higher differences, mostly in the tangential part for the case with higher W_x. In this case, the differences in the magnitude of the tangential stresses with respect to the reference results are around 4%.

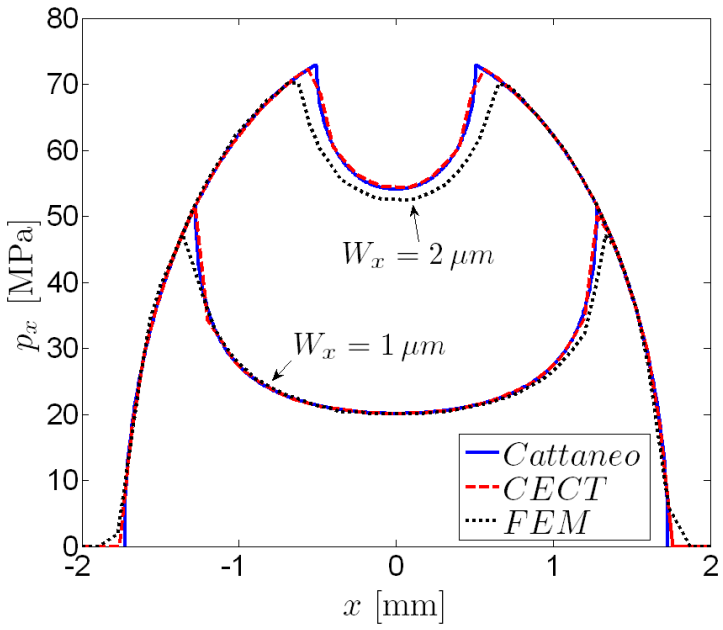


Figure 4.53. Comparison of p_x along the contact centreline computed for the circular contact shift test case.

Table 4.21. Results of circular contact shift test case.

Result	Units	Cattaneo	CECT	FEM
N	kN	1.5988	1.59932	1.6034
p_{no}	MPa	254.46	254.52	254.70
F_x	N	288.0 | 467.8	288.2 | 468.0	286.2 | 469.2
max(p_x)	MPa	51.35 | 72.94	51.47 | 73.03	47.28 | 70.15

4.3.3 Oblique compression

This case is differenced from the Cattaneo shift in that the normal and tangential loads change simultaneously. In the most general form, there is an initial normal load N_0, and after it, a force F is applied, inclined at an angle α with respect to the contact plane normal, as depicted in Figure 4.54.

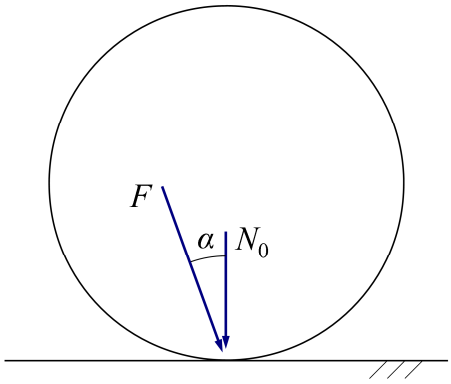


Figure 4.54. Oblique compression contact problem.

The solution to this problem may be found in §7.3 of [Johnson 1987]. The solution for the case with $N_0 = 0$ is derived in problem 2 of §8 of [Popov 2010] as well. The solution of this problem resembles that of the Cattaneo shift, the tangential stress field being composed of two concentric ellipsoidal distributions superposed. If tan $\alpha < \mu$, there is no slip in the contact area. The test case considered here fulfils this condition. The parameters of this test case are listed in Table 4.22. As in the Cattaneo shift test case of §4.3.2, the contact of a sphere on a plane is considered.

Table 4.22. Input parameters of circular contact oblique compression test case.

Parameter	Description	Value	Units
R_x, R_s	Combined principal radii of curvature	500.0	mm
N_0	Initial normal load	3.342	kN
F	Inclined load	23.705	kN
α	Angle of inclination of F	0.1639	rad
μ	Coefficient of friction	0.30	-

This test case has been solved with CECT in 1+40 steps, 1 for the application of N_0 and 40 for the application of F. The PCS has been discretized in a 50×50 element mesh, and it has been changed two times during the analysis, adapting it to the contact patch size. In step 1 it has been defined with a 6×6 mm size, in step 11 it has been increased to 8×8 mm, and in step 26 to 10×10 mm. This serves to test the mesh change functionality treated in §4.1.1.4.3.

The problem has been specified in two different ways in the CECT input, as listed in Table 4.23 (cf. §4.1.2.1.1 for the meaning of the different input parameters). These different input specifications are designated as 1 and 2. In input 1, the inclined load is specified to be in the *YZ* plane, and in input 2, in the *XZ* plane. So, the obtained solutions for each input are rotated 90º with respect to each other. In both cases, the loads are not specified directly, but through imposed displacements, with the parameters `d_o_N`, and either `dir_d` (in input 1) or `sftx_o_vel` (in input 2). It is necessary to determine the magnitude and direction of these imposed displacements. This is straightforward, because the ratio of the tangential to normal contact compliances, C_T / C_N, is independent of the load in this case with no slip, and equal to (2–ν) / 2(1– ν); cf. Eqs. (7.43) and (7.44) of [Johnson 1987]. So, the direction of the resultant imposed displacements (the increments with respect to the state after the application of N_0) remains constant. The angle α_d which defines their direction with respect to the vertical (similar to α for F in Figure 4.54) is given by Eq. (4.94). In this equation ΔW_h is the increment of rigid shift in the plane of contact, and Δd the increment of normal approach. ΔW_h is equal to the increment of `d_o_N` sin(`dir_d`) in input 1, and to the increment of `sftx_o_vel` in input 2. Δd is equal to `d_o_N` cos(`dir_d`) in both inputs.

$$\tan(\alpha_d) = \frac{\Delta W_h}{\Delta d} = \frac{C_T F}{C_N N} = \frac{2-\nu}{2(1-\nu)} \tan\alpha \tag{4.94}$$

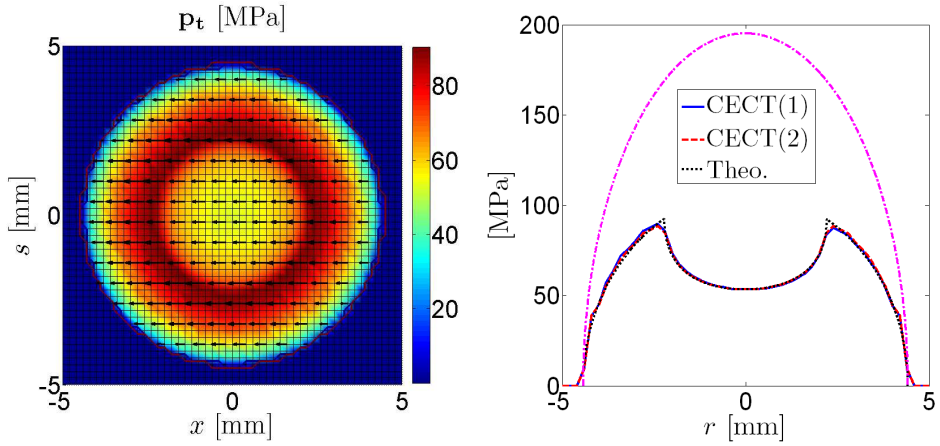
Table 4.23. CECT input specification for oblique compression test case.

Parameter	Value	
	Input 1	**Input 2**
`ndgt`	0	0
`f` (1)	0	0.3
`f` (2:)	0.3	0.3
`dir_d`	α_d	0
`d_o_N`	$\frac{C_N\left(N_0 + F\cos\alpha\right)}{\cos\alpha_d}$	$C_N\left(N_0 + F\cos\alpha\right)$
`sftx_o_vel` (0)	0	0
`sftx_o_vel` (2:)	0	$C_N F\cos\alpha\tan\alpha_d$

The numbers in parenthesis in the first column of Table 4.23 denote the steps of the analysis. The applied displacements have been increased following a quadratic function from the second to the last step, so that the rate of change of the contact patch radius is constant. This leads to a cubic increase of the load.

In input 1, `dir_d` is set to α_d from the first step. In this way, the direction of the total applied displacement `d_o_N` from the beginning of the analysis remains constant, and it is not necessary to calculate a different α_d for each step. To avoid tangential stresses in the first step, `f` is set to 0 in this step.

Figure 4.55 shows the tangential stress solutions obtained for this test case. Figure 4.55a shows the tangential stress field in the whole contact patch computed with CECT for input 2. Figure 4.55b shows the directional tangential stresses along perpendicular central sections of the contact patch computed with CECT for both inputs (p_s along s for input 1, and p_x along x for input 2), together with the theoretical result. The traction bound is shown as well, in magenta chain line. The agreement of the numerical results with the theoretical is very good, as can be seen in the figure. The CECT result for input 1 shows a slight asymmetry, which is due to the slight non-flatness of the profile of the PCS. In the CECT solutions a few APs appear in slip in the perimeter of the contact area (32 out of 1557 in contact for solution 2).



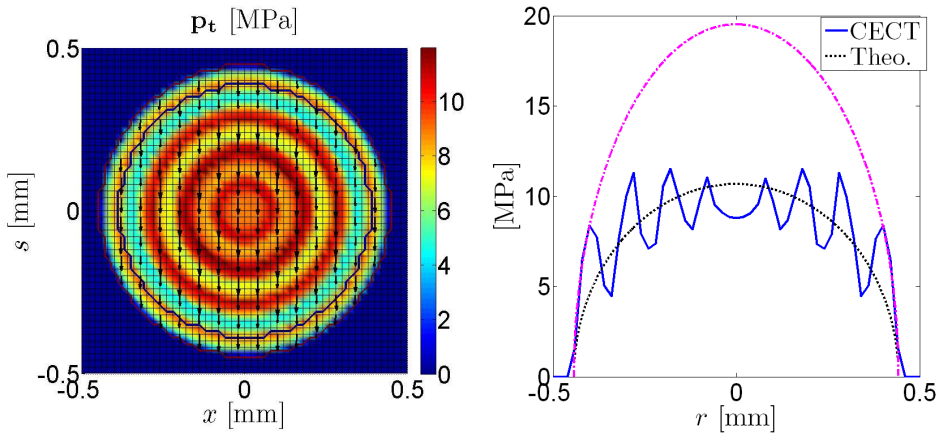
(a) CECT solution for input 2. $\mathbf{p_t}$ magnitude and direction.

(b) Tangential stresses along the contact centrelines.

Figure 4.55. Tangential stresses in the oblique compression test problem.

The numerical computation of this case with CECT is more troublesome than the previous cases, especially for the limit case with $N_0 = 0$. In the first compression steps, the tangential stresses exhibit a rippled pattern, where each ripple is associated to a compression step. This is a numerical artefact caused by the fact that the new APs that enter in contact accumulate excessive rigid slip, because the increment of rigid slip between the previous t' and current t instants is calculated for all APs in contact in instant t as if they had been in contact in the whole t'–t interval. So, for example, in the case of solving the oblique compression problem with $N_0 = 0$ in a single step, the obtained solution with the current implementation of CECT would be the same as for a Cattaneo shift problem. Figure 4.56 shows an example of the ripples arising in the first steps of an oblique compression case with similar parameters as the one calculated above, except for N_0, which is set to 0. In this case, F is also increased cubically, reaching a value of 27.2 N in the step shown. Figure 4.56a shows the tangential stress field computed with CECT in the whole contact patch and the contours of the adhesion and slip zones. Figure 4.56b shows the lateral tangential stresses p_s along s at $x = 0$, computed with CECT and analytically, together with the traction bound shown in magenta chain line. In addition to the ripples, in Figure 4.56a it is observed that there is a considerable number of APs

slipping in the whole perimeter in the CECT solution (377 out of 1557 APs in contact), while in the theoretical solution all the contact is in adhesion.



(a) $\mathbf{p_t}$ magnitude and direction. (b) p_s across the contact centreline.

Figure 4.56. CECT solution for oblique compression case with N_0 = 0, in step 4 of the analysis.

The error in the solution is progressively reduced as the contact grows and the influence of the new elements that enter in contact in the overall solution decreases (as long as the prescribed load incrementation does not lead to an accelerating growth of the contact patch), but a relatively high number of steps is necessary to achieve a sufficiently smooth solution. It could be interesting to investigate the possibility to reduce this numerical error, by adequately scaling down the inter-step increments of rigid slip and elastic displacement differences for the new APs that enter in contact in each step. This could be done on the basis of the precise instant (in the (t'–t) interval) in which each new AP enters in contact. An estimate of this instant could be achieved using the deformed contact surfaces of t'.

4.3.4 Spence compression

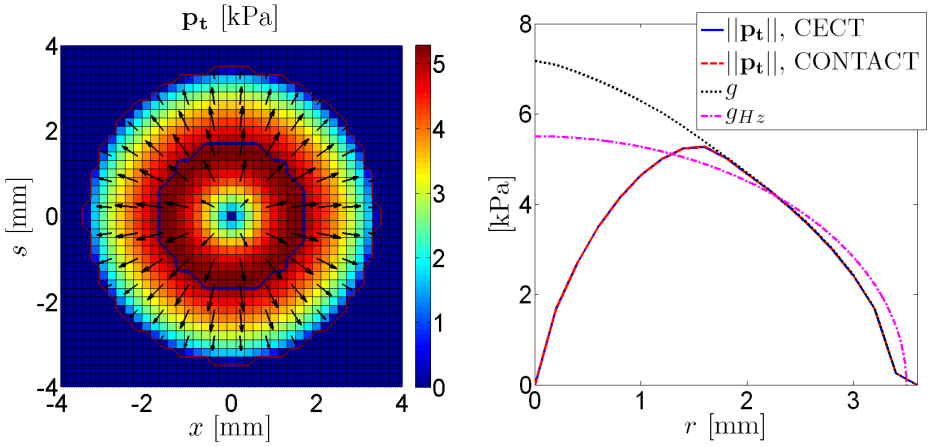
In this compression case there is no net tangential load, but tangential stresses appear due to the elastic dissimilarity of the contacting bodies. A solution to this problem was presented in [Spence 1975].

The parameters of the test case considered are listed in Table 4.24. This test case is the same as the Spence compression test case supplied with the CONTACT software [CONTACT UG 2013].

Table 4.24. Input parameters of circular contact compression test case with dissimilar materials.

Parameter	Description	Value	Units
E_r	Young's modulus of plane	200	GPa
E_w	Young's modulus of sphere	0.001	GPa
ν_r	Coefficient of Poisson of plane	0.0	-
ν_w	Coefficient of Poisson of sphere	0.0	-
R_x, R_s	Combined principal radii of curvature	121.5	mm
N	Normal load	0.4705	N
μ	Coefficient of friction	0.2986	-

This case has been solved with CECT and with CONTACT (free version v13.1; [CONTACT UG 2013]). The same mesh and incrementation has been used in both cases. The PCS is a square with 9 mm side discretized in a 45×45 mesh, and the load is increased cubically in 35 steps. As in the oblique compression case, it is necessary to use a high number of steps to obtain a correct solution.



(a) CECT solution. $\mathbf{p_t}$ magnitude and direction. (b) Stresses along contact centreline.

Figure 4.57. Stresses in the compression test problem with dissimilar materials.

The computed tangential stresses are shown in Figure 4.57, together with the contours of the adhesion and slip zones. Figure 4.57b compares the tangential stresses computed with CECT and with CONTACT, which are seen to agree almost exactly. The non-Hertzian traction bound, g, and the Hertzian one, g_{Hz},

are also included in this graph, showing that the normal–tangential coupling alters the normal pressure profile significantly.

4.3.5 2D steady rolling

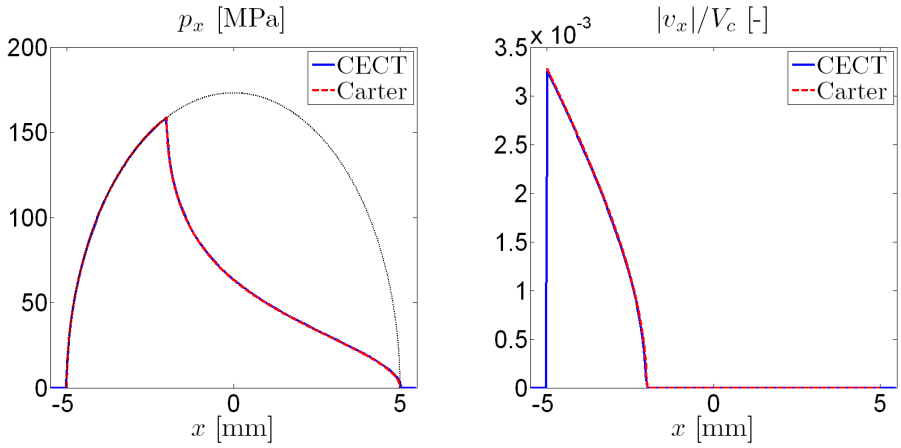
The steady rolling of a cylinder on a plane is considered, which results in a 2D Hertzian contact. This problem has analytical solution, that has been reviewed in §2.1.2.1. The parameters of the test case considered are listed in Table 4.25. With these parameters, 30% of the contact area is in slip according to the analytical solution, and the traction ratio $t_x = F_x / N$ is equal to 0.1530.

Table 4.25. Input parameters of 2D steady rolling test case.

Parameter	Description	Value	Units
R	Cylinder radius	500	mm
N	Normal load	4.53	kN/mm
ξ	Longitudinal creepage	–0.0009	-
μ	Coefficient of friction	0.30	-

The 2D problem is specified in the CECT input setting `c_o_n` = 0, `ns` = 1, and the lateral dimension of the mesh much larger than the longitudinal (in this case, it has been set about 180 times larger). The rigid contact point has to be located in the defined unique position of the mesh in *s* direction; the wheel and rail profiles are not relevant as long as this condition is fulfilled. `ndgt` = 1 is used for the normal problem, and `d_o_N` is defined as the desired normal load per unit length *N* multiplied by the lateral mesh dimension (`s_sup–s_inf`). It has to be taken into account that in 2D half-plane contact problems the absolute value of the approach is undefined. In the numerical approximation to the 2D problem carried out here, this is reflected in that for a fixed load per unit length *N*, the value of the approach changes with the dimension (`s_sup–s_inf`). So, it would not be appropriate to use `ndgt` = 0 in this case.

The tangential stresses and slip velocities computed with CECT are compared with the analytical Carter / Fromm solution (cf. §2.1.2.1) in Figure 4.58. The CECT solution has been obtained with approx. 200 elements in contact. The agreement of both solutions is very good. The traction ratio calculated with CECT is equal to 0.1532, also very close to the analytical result.



(a) Tangential stresses. (b) Slip velocities.

Figure 4.58. Solution of 2D steady rolling test case.

4.3.6 3D transient rolling: from Cattaneo to Carter

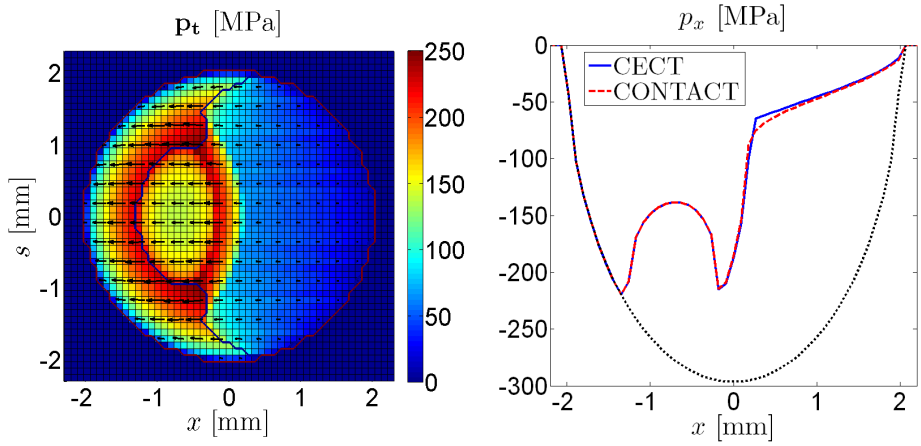
This is a 3D transient rolling case of a sphere on a plane. Starting from a state of static contact with a resultant longitudinal contact force $F_{x,0}$, the sphere starts rolling, with a pure longitudinal creepage ξ which is kept constant. ξ is calculated to produce a steady F_x equal to $F_{x,0}$. The parameters of the test case considered are listed in Table 4.26.

Table 4.26. Input parameters of Cattaneo to Carter test case.

Parameter	Description	Value	Units
R_x, R_s	Combined principal radii of curvature	200	mm
N	Normal load	6.25	kN
$F_{x,0}$	Initial tangential force	–1.193	kN
ξ	Longitudinal creepage	0.001	-
μ	Coefficient of friction	0.4013	-

This test case has been computed with CECT and with CONTACT, with the same discretization. The PCS is defined as a 4.59 mm side square discretized in 51×51 uniform elements (c_o_n = 1), and a step $\Delta q = \Delta x = 0.09$ mm is used. Figure 4.59 shows the tangential stresses computed after 20 rolling steps, which corresponds to a rolled distance of about 0.45 times the contact patch diameter in this case. Figure 4.59a shows the tangential stress field and contours of adhesion and slip zones computed with CECT. The solution computed with

CONTACT is similar. Figure 4.59b compares the p_x stresses along the contact centreline computed with CECT and with CONTACT, together with the traction bound in dotted line. It is verified that the results obtained with both programs are similar.



(a) CECT solution. $\mathbf{p_t}$ magnitude and direction.

(b) p_x along contact centreline.

Figure 4.59. Tangential stresses in 3D transient rolling test case after 20 rolling steps.

Figure 4.60 shows the evolution of the normalized resultant longitudinal force $f_x = F_x / \mu N$ in the transient rolling simulation. The simulation has been run for a total rolled distance U_x of approximately 2.3 times the contact patch diameter $2a$. The solution has almost reached a steady state at the end of the simulation. It is observed that there is a slight difference in the final f_x values obtained with CECT and with CONTACT, of about 0.2% with respect to the traction bound μN, which is possibly due to the slight non-flatness of the profile of the PCS.

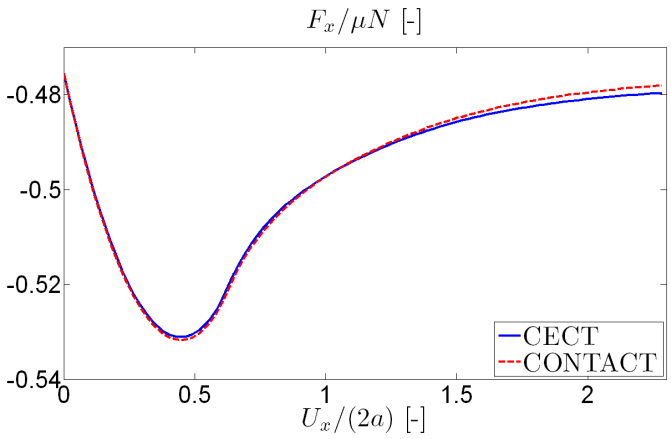


Figure 4.60. Evolution of f_x in transient rolling test case.

4.3.7 Cylindrical contact

This is a 2D cylindrical conformal contact case without friction. Details of the analytical solution of this problem are given in §2.4. This problem has been solved with CECT, considering a cavity radius of 100 mm and two different cylinder radii: 99.8 mm (clearance fit) and 100.1 mm (interference fit). Both cases have been solved for different load levels.

`dir_d` has been set to 0 in the first case, and to $\pi/2$ in the second. I.e., the load is applied towards the negative *Z* axis in the first case, and towards the positive *Y* axis in the second. In the first case, the profiles are positioned so that the undeformed distance is 0 at the initial contact point, as usual. In the second case, both profiles are positioned concentrically. A discretization with `ns` = 720 elements has been used in both cases. The mesh has been changed twice during each analysis case, spanning sectors of 60º, 120º and 180º of the circumference in the clearance fit case, and 360º, 280º and 200º in the interference fit case. The 2D mesh for the problem is defined with `c_o_n` = 0 and `nx` = 1. In this case exact 2D cylindrical ICs are used with `ics`=20 (cf. §4.1.2.1.1), so the longitudinal dimension of the contact mesh may be set to an arbitrary value. The normal problem is specified with `ndgt`=1, and the total load `d_o_N` has to be defined as the desired load per unit thickness *N* multiplied by the longitudinal mesh dimension.

The results obtained with CECT are compared with the analytical solution in Figure 4.61 and Figure 4.62. Figure 4.61 shows the contact patch size as a function of the non-dimensional load parameter $E^*_w\,|\Delta R|\,/\,N$, where E^*_w is the equivalent Young's modulus of the cylinder for plane strain (`E_w` / (1 – `nu_w`^{2})), ΔR the radial clearance between the cylinder and the cavity (negative for interference fit) and N the load per unit thickness. The results of both the clearance fit and of the interference fit cases are shown in the figure. The contact patch size is expressed as the total angle spanned in the cavity circumference, $\Delta\theta$ (cf. Figure 2.24)[15]. It is verified that $\Delta\theta$ tends asymptotically to the same limit value in both cases. This limit value depends only on the two Dundurs' material parameters α and β for the pair of contacting materials, as indicated in §2.4.

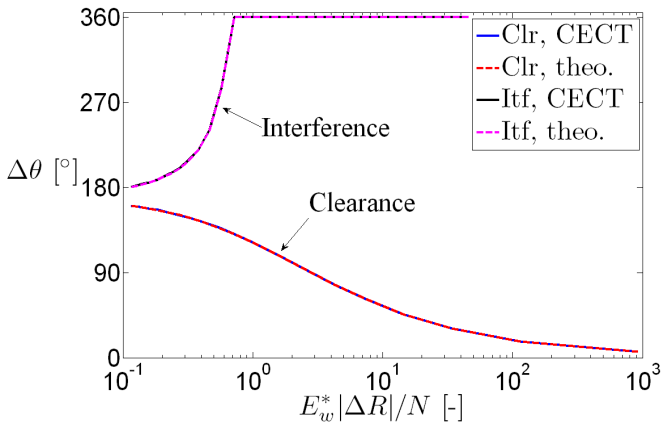
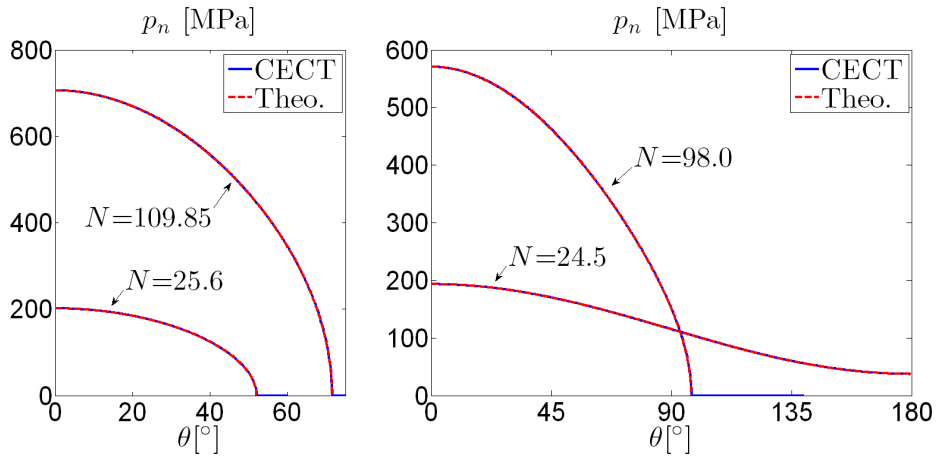


Figure 4.61. Contact patch size as a function of load in frictionless cylindrical contact.

In Figure 4.62, the normal pressure distributions in the contact patch are graphed for two different load levels in the clearance fit and interference fit cases for half the contact surface (the results being symmetric in the other half). The abscissa axis shows the angle θ with respect to the direction of application of the load. The load per unit thickness N corresponding to each case is annotated in the figure in kN/mm. It is verified that the CECT solutions match very well the theoretical solutions. The results obtained with `ics`=21 are

[15] $\Delta\theta$ is also called the total contact angle variation, designated as $\Delta\delta$ in this thesis.

similar to the ones obtained with `ics`=20. The results are different with `ics`=22.



(a) Clearance fit case. (b) Interference fit case.

Figure 4.62. p_n distributions for different loads in frictionless cylindrical contact.

In the interference fit case, the mesh is initially defined spanning the whole 360º of the circumference, as previously mentioned. In this case, it is intended to model a contact surface with closed profile. No special treatment is done for this case in the current implementation of CECT, and the extent of the mesh in *s* direction has to be defined with caution to get correct results if some of the borders of the mesh in *s* direction is in contact. This should be done so that the distance between the first and last positions of the mesh in *s* direction, Δs_{ns-1}, is equal to the distance between any other two adjacent *s* positions of the mesh, Δs. This is achieved approximately in this case by defining the `s_sup`, `s_inf` limits of the mesh as equal to $\pm\pi R_m$, being R_m the average radius of the cavity and cylinder. The profile defining the smallest circumference has to be extended as necessary to accommodate this *s* range. If Δs_{ns-1} is larger than Δs, the borders of the mesh will exhibit a pressure peak if there is contact in them; and if Δs_{ns-1} is smaller than Δs, they will exhibit a pressure valley. Noticeable pressure perturbations in the mesh borders are prone to be caused by small deviations between Δs_{ns-1} and Δs. For the CECT computations, the cylinder and cavity profiles are defined so that the mesh borders are located at the point which first loses contact when the load is increased sufficiently. In this way, this problem is

not present after contact is first lost. Small pressure perturbations are observed in the mesh borders of the CECT solutions for the cases in which there is contact in the full circumference, which are due to this problem, though this is not appreciated in Figure 4.62b.

4.3.8 3D steady rolling without slip

Here the 3D steady rolling with Hertzian geometry and no slip is considered. The no-slip condition is approached with a high coefficient of friction of 1.0 and low creepage levels in the computations with CECT. The reference results are provided by Kalker's linear theory of rolling [Kalker 1967].

The chosen surface curvature radii are $R_{sw} = R_{sr} = 300$ mm (transverse curvature radii of wheel and rail) and $R_{xw} = 323.3$ mm (longitudinal curvature radius of wheel). As the same convex transverse curvatures are defined for both bodies, the resulting profile of the contact surface is exactly planar, and no rigid slip is generated due to contact surface curvature. The normal load N is set to 30 kN. With these inputs, the dimensions of the Hertzian contact ellipse are $a = 4.48$ mm and $b = 2.69$ mm. The lateral b to longitudinal a semi-axis ratio of the contact ellipse is 0.60.

3 cases with different creepages are computed: pure longitudinal (ξ), pure lateral (η) and pure spin (φ) creepage. The applied values in each case are $\xi =$ –1.0E–05, $\eta =$ –1.0E–05, and $\varphi = \tan(0.003) / 323.3 = 9.27933\text{E–}06$ mm^{-1}. In the cases with pure longitudinal and lateral creepages, the mean contact angle δ_0 is zero. The longitudinal creepage is imposed in the CECT input by defining `sftx_o_vel` and `omega` so that `omega` = `sftx_o_vel` / $R_{roll} \times (1 - \xi)$, being R_{roll} the wheel rolling radius, equal to R_{xw} in this case. The lateral creepage is imposed by setting the wheel yaw angle `psi` = 1.0E–05 rad. The spin creepage is imposed by displacing the wheel with respect to the rail so that $\delta_0 = 3$ mrad. In all the cases, the wheel forward velocity `sftx_o_vel` is aligned with the x axis.

The numerical results computed with CECT are compared with the reference results in terms of the contact resultant forces and moment. The numerical and reference results are listed in Table 4.27, in the form of the creepage coefficients C_{ij}, computed according to Eq. (2.64). The pure longitudinal, lateral

and spin creepage situations considered are identified in the table as i = 1, 2, and 3, respectively. Two values are given in each case in the table, the first being the numerical value computed with CECT, and the second the reference tabulated value. The numerical results have been obtained on a PCS of approximately 10×6 cm, discretized in 60×50 elements with c_o_n = 1 for i = 1 and 2, and in 90×54 elements for i = 3. The reference C_{ij} values may be found tabulated e.g. in [Kalker 1967] and [Kalker 1990] for ν = 0, 0.25 and 0.50. The values for the ν value considered here of 0.30 have been taken from Appendix 5 of [Johnson 1987]. They may also be found graphed in Figure 5-11 of [Thompson 2009].

The best agreement between the CECT and reference values is found for C_{11} and C_{22}, and the highest discrepancy is found in C_{33}, with a difference above 6%. This case has been computed with CONTACT as well, obtaining similar results. The numerical value for C_{33} (i.e. for the resultant moment due to spin) is found to converge slowly towards the reference value with mesh refinement.

Table 4.27. Creepage coefficients of Kalker's linear theory computed with CECT for b/a = 0.60 and ν = 0.30 (computed | reference tabulated values).

	i=1 (ξ)	i=2 (η)	i=3 (φ)
C_{ii} [-]	5.01 | 4.99	4.67 | 4.65	0.92 | 0.86
C_{23} ($-C_{32}$) [-]		2.26 | 2.29	2.27 | 2.29

4.3.9 Steady rolling

In this section a case of 3D partial-slip steady rolling with Hertzian geometry is considered, with parameters listed in Table 4.28. The contact angle δ_0 is zero, so there is no spin creepage. Problems with pure longitudinal or lateral creepage are solved, changing the creepage in one direction while maintaining the creepage in the other direction at zero.

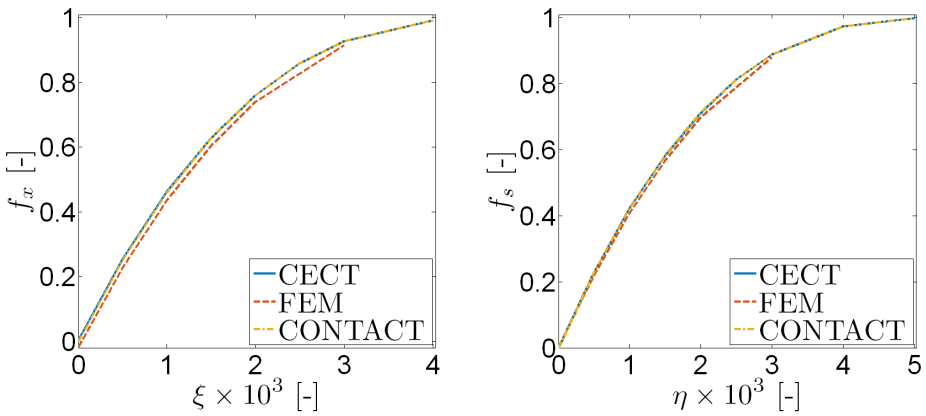
Table 4.28. Input parameters of 3D steady rolling test case.

Parameter	Description	Value	Units
R_{xw}	Longitudinal radius of curvature of wheel	500	mm
R_{sw}	Lateral radius of curvature of wheel	∞	mm
R_{sr}	Lateral radius of curvature of rail	300	mm
N	Normal load	80	kN
μ	Coefficient of friction	0.30	-
δ_0	Contact angle	0	º

CECT, FEM and CONTACT are used to solve equivalent problems. The PCS is meshed with 0.25×0.20 mm bilinear elements in the case of CECT, with 0.375×0.30 mm elements in the case of FEM, and with 0.246×0.20 mm uniform elements in the case of CONTACT (longitudinal × lateral dimension).

The longitudinal creepage is imposed as explained in §4.3.8 in the CECT input. The lateral creepage is imposed differently in CECT in this case, changing the direction of the wheel forward velocity instead of defining a wheel yaw angle. For this purpose, `bit_l6` is set to 1, and the direction of the wheel velocity vector is defined as (`dir_vx`, `dir_vy`, `dir_vz`) = (1, η, 0), with (`x_v`, `y_v`, `z_v`) = (0, 0, 0) and (`dir_omx`, `dir_omy`, `dir_omz`) = (0, 1, 0). On the other hand, in FEM the lateral creepage is imposed through the wheel yaw angle.

Figure 4.63 shows the computed longitudinal and lateral creepage–creep force curves, with f_x and f_s being the contact resultant forces in x and s directions normalized with μN. There is an excellent agreement between the curves obtained with CECT and with CONTACT. The curves obtained with FEM fall slightly below. The greatest differences are found in the longitudinal curve, with a maximum difference of about 0.027 μN between the FEM and the other two curves.

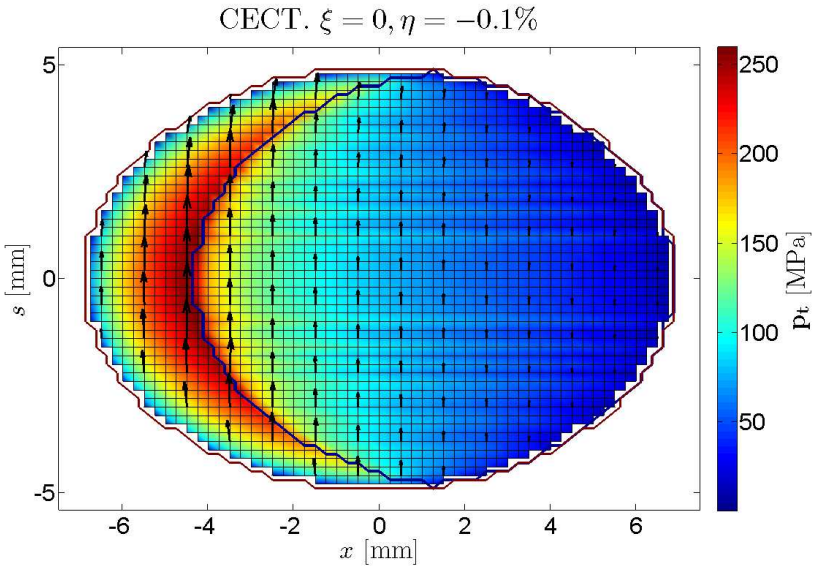


(a) Longitudinal. (b) Lateral.

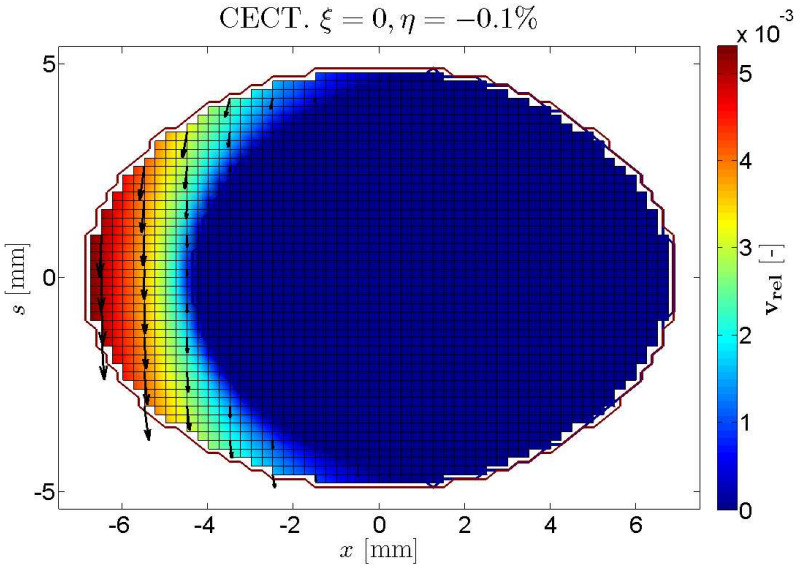
Figure 4.63. Creep force curves in 3D steady rolling test case.

The different solutions for $\xi = 0$ and $\eta = -0.1\%$ are compared in more detail hereafter. Figure 4.64 to Figure 4.66 show the contours of adhesion and slip zones in the contact patch, with the distributions of tangential stresses and relative slip velocities, computed with CECT, FEM and CONTACT, respectively. Some relevant quantities of the contact solutions are listed in Table 4.29.

The correspondence between the results of the different models is very good. The maximum differences are seen in the magnitudes related to the tangential part of the contact problem. The maximum difference in the results shown in the table occurs in the maximum W_{fric} value, and is around 10% between FEM and CONTACT. As can be seen in the table, the magnitudes related to the tangential part of the contact problem computed with FEM are slightly lower than those computed with CONTACT and CECT. This is attributed mostly to the use of the penalty formulation in the FE model to enforce the contact constraints, which introduces some flexibility in the contact interface that is not present in CECT and CONTACT; cf. §4.2.2.3. Another part of the differences are attributed to the coarser discretization used in the FEM analyses. The FEM solutions have been obtained with an elastic slip tolerance value of 0.0005, and with a rolling substep size ΔU_x / Δx (cf. §4.2.2.2) of 1.0. Both the contact resultant forces and the contact solution of Figure 4.65 and Table 4.29 for FEM are reported after a rolled distance of about three times the longitudinal contact patch dimension $2a$. After this rolled distance, the contact resultant forces are stabilized, cf. Figure 4.40a. Regarding the contact patch dimensions, it has to be noted that the accuracy of the results is limited by the element size. Nevertheless, the contact patch shapes obtained with the different analysis methods are nearly coincident when plotted one over the other.

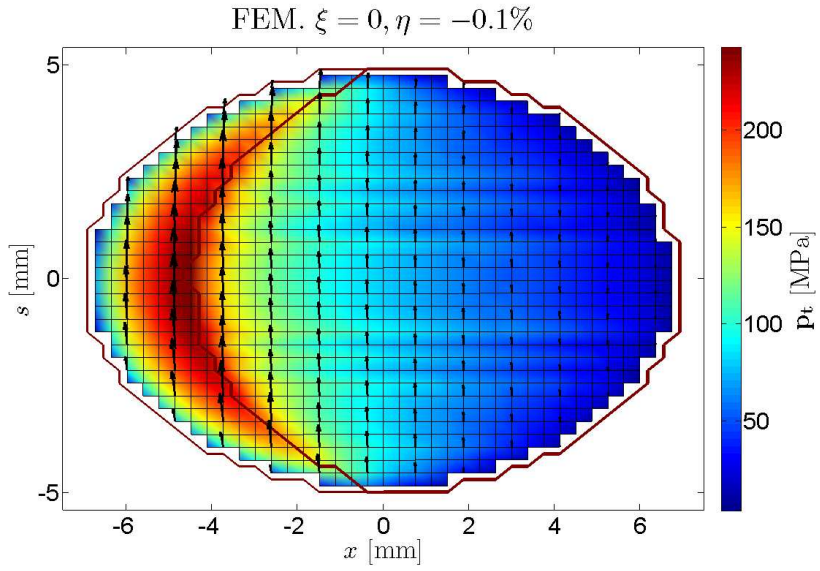


(a) $\mathbf{p_t}$.

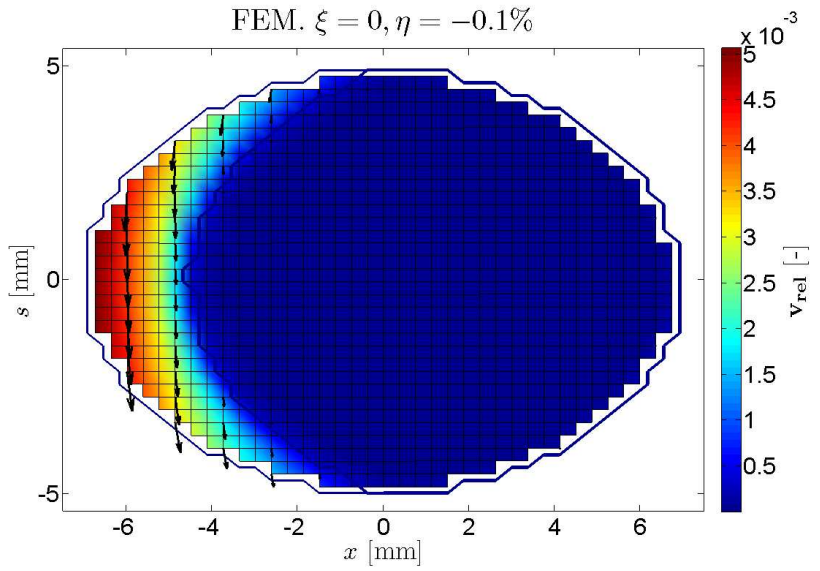


(b) $\mathbf{v_{rel}}$.

Figure 4.64. CECT solution for 3D steady rolling test case with ξ = 0 and η = –0.1%.

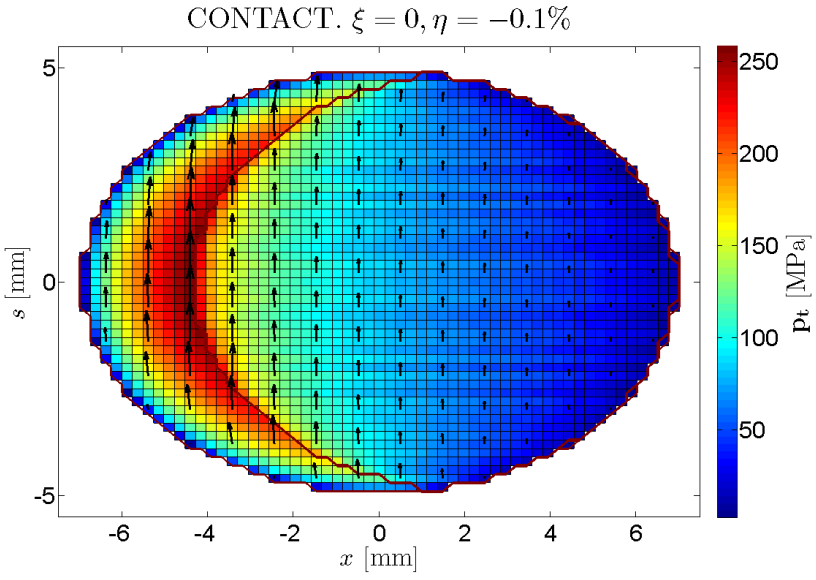


(a) $\mathbf{p_t}$.

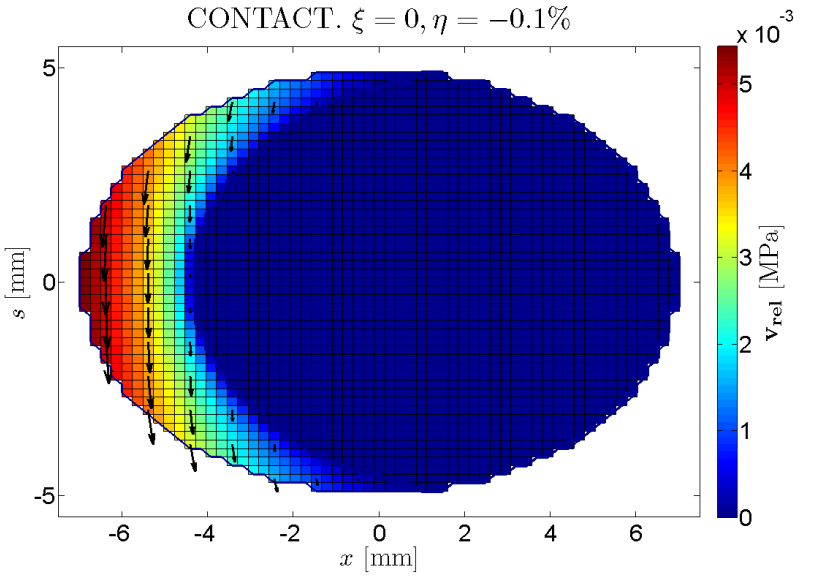


(b) $\mathbf{v_{rel}}$.

Figure 4.65. FEM solution for 3D steady rolling test case with ξ = 0 and η = –0.1%.



(a) $\mathbf{p_t}$.



(b) $\mathbf{v_{rel}}$.

Figure 4.66. CONTACT solution for 3D steady rolling test case with ξ = 0 and η = – 0.1%.

Table 4.29. Results of 3D steady rolling test case with $\xi = 0$ and $\eta = -0.1\%$.

Result	Hertz	CECT	FEM	CONTACT
a [mm]	6.923	6.875	6.932	7.011
b [mm]	4.927	4.900	4.945	4.900
d [μm]	88.39	88.38	88.74	88.39
max(p_n) [MPa]	1120	1120	1121	1120
max($\|\|\mathbf{p_t}\|\|$) [MPa]		259.6	243.0	258.3
max($\|\|\mathbf{v_{rel}}\|\|$) × 10^3 [-]		5.313	5.064	5.435
max(W_{fric}) [mJ/mm^2]		1.469	1.336	1.486

4.3.9.1 Free rolling

Additional free rolling tests have been carried out with FEM, to try to elucidate the question set out in §4.1.1.3.1 of whether the rigid slip velocities should be computed using the undeformed or the deformed geometry (cf. Figure 4.14). For this purpose, the 3D FE model built for the case specified in Table 4.28 has been used on the one hand, and on the other, a new 2D FE model has been built.

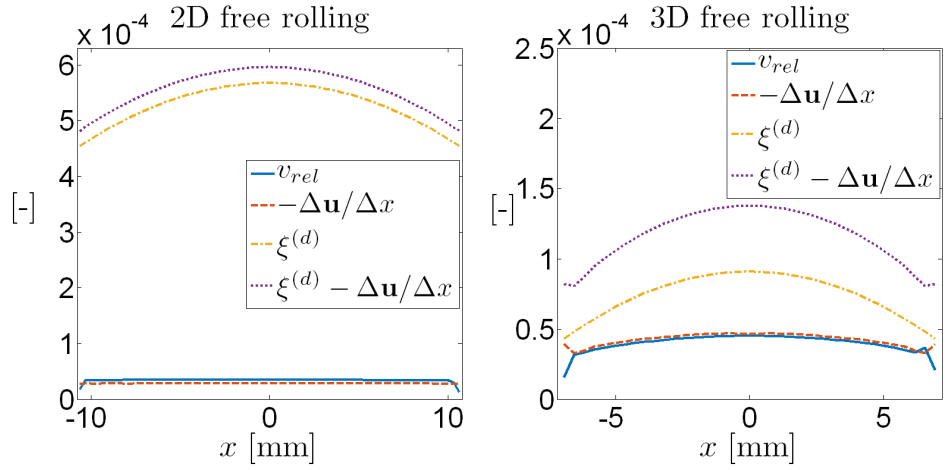
The 2D model represents a wheel with 500 mm radius rolling on a plane in plane strain. The load in the 3D case is 80 kN as specified in Table 4.28. In the 2D case, the load is 20 kN/mm, and the resulting contact patch half-length is 10.5 mm. The element size in the 2D model is 0.3 mm in the contact zone.

The slip velocities computed in these FEM analyses are to be compared with the rigid slip velocities calculated taking into account both the undeformed and the deformed geometry. The FEM analyses are carried out with non-linear geometry (cf. §4.2.1.5), so they gather the influence of deformations. The free rolling condition, with nominally null rigid slip velocities, has been chosen for these tests to focus on the relatively small slip velocity differences object of this study. Additionally, frictionless contact has been considered in these analyses, to facilitate the retrieval of the interesting slip velocities aside from the influence of the tangential tractions and numerical elastic slip tolerance present with friction.

Figure 4.67 shows the relative slip velocities v_{rel} calculated in the 2D and 3D free rolling test cases along x, together with the relative rigid slip velocities calculated with the deformed geometry, designated as $\xi^{(d)}$. The results of Figure 4.67b for the 3D case correspond to the contact longitudinal centreline (at s = 0). The lateral slip velocities are zero in the results shown for both the 2D and

3D cases. v_{rel} is calculated with the FEM contact outputs CSLIP*n*, as outlined in §4.2.3.2. $\xi^{(d)}$ is calculated projecting the rigid slip velocities calculated with the deformed rolling radii of the wheel on the deformed contact surface, taking into account the non-flatness of the deformed contact surface in the longitudinal direction, according to Eq. (4.95). α_x in this equation is the inclination of the deformed contact surface in longitudinal direction, as represented in Figure 4.68. The rest of the quantities of Eq. (4.95) are represented in this figure as well. The *P'* point is the wheel centre in the deformed configuration, as shown in Figure 4.14.

$$\xi^{(d)} = \frac{\left(V - \omega R^{(d)} \cos\theta\right)\cos\alpha_x - \omega R^{(d)} \sin\theta \sin\alpha_x}{V} \tag{4.95}$$



(a) 2D case. (b) 3D case.

Figure 4.67. Slip velocities calculated with FEM in free rolling test cases.

Figure 4.67 includes the relative velocities associated to the elastic deformations, that in steady rolling are equal to the $-\partial\mathbf{u} / \partial x$ term of Eq. (1.7). In these frictionless test cases, the elastic deformations are caused only by the contact normal pressures p_n. The associated elastic displacement differences, though small, are not zero in these concentrated contact cases with elastically similar materials, because the deformed contact surfaces are slightly non-planar. The longitudinal gradient $-\partial\mathbf{u} / \partial x$ is approximated with the differences between adjacent nodes in longitudinal direction as $-\Delta\mathbf{u} / \Delta x$. Both the horizontal and

vertical displacements are taken into account, and they are projected onto the deformed contact surface (with inclination α_x in longitudinal direction) similar to what is done in Eq. (4.95). I.e., Δu_x is calculated as $\Delta u_x = \Delta u_X \cos \alpha_x + \Delta u_Z \sin \alpha_x$, being Δu_X and Δu_Z the differences between adjacent nodes in longitudinal direction of elastic displacement differences in X and Z directions, respectively, and Δu_x the resulting difference in the longitudinal direction of the deformed contact surface.

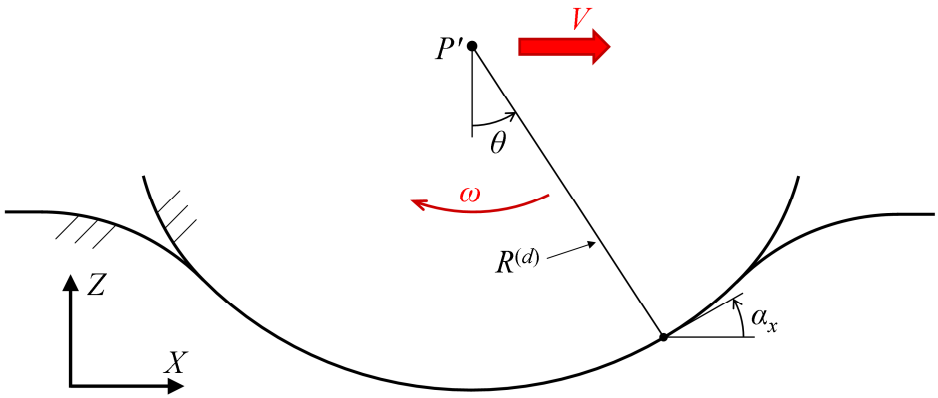


Figure 4.68. Elastic wheel rolling on an elastic plane, with the deformed geometry parameters for the calculation of rigid slip velocities.

To confirm that the rigid slip velocities should be calculated with the deformed geometry, v_{rel} should match with $\xi^{(d)} - \Delta\mathbf{u} / \Delta x$. Both in the 2D and 3D test cases, the best match of v_{rel} is with $-\Delta\mathbf{u} / \Delta x$. This seems to indicate that the rigid slip velocities should be calculated with the undeformed geometry. These being zero in the considered test cases leads to the equality $v_{rel} = -\Delta\mathbf{u} / \Delta x$. To obtain the match of v_{rel} with $-\Delta\mathbf{u} / \Delta x$ in the results of these FEM test cases, it is important to take into account both Δu_X and Δu_Z in the calculation of $\Delta\mathbf{u}$, and to carry out the projection on the deformed contact surface as explained before.

Chapter 5

Influence coefficients of the elastic solid

Chapter summary

In this chapter, the determination of the influence coefficients (ICs) of the elastic solid is tackled. To start with, the case of the elastic half-space is treated, of great practical relevance as a wide variety of problems in contact mechanics may be solved assuming that the mechanical behaviour of the contacting solids is similar to that of the half-space around the contact. Available analytical results are reviewed, and new ones are developed for rectangular surface elements with bilinear load distribution. On the other hand, more general cases of solids with non-planar surfaces are addressed, as the ones that may be found in conformal contact, with no analytical solution. For these cases, the numerical calculation is considered, and an approximation is proposed to avoid the numerical calculation.

5.1 The elastic half-space

In this section, the ICs of the elastic half-space are calculated following Eq. (1.9), and starting from the IFs given in §2.2.1. The same {*x*, *y*, *z*} Cartesian coordinate system defined in that section is considered. Now these axes are designated with uppercase letters in the formulae that follows, and the corresponding lowercase letters are used for particular coordinates along them. The following integral expressions are defined with the shape function $T(X, Y)$ in the surface S_e of the element in which the distributed load is applied, generalizing the development presented in [Kalker 2001] to arbitrary load distributions:

$$J_1 = \iint_{S_e} T(X,Y)\frac{1}{\rho}\,dXdY = J_2 + J_4 \tag{5.1}$$

$$J_2 = \iint_{S_e} T(X,Y)\frac{X^2}{\rho^3}\,dXdY \tag{5.2}$$

$$J_3 = \iint_{S_e} T(X,Y)\frac{XY}{\rho^3}\,dXdY \tag{5.3}$$

$$J_4 = \iint_{S_e} T(X,Y)\frac{Y^2}{\rho^3}\,dXdY \tag{5.4}$$

$$J_5 = \iint_{S_e} T(X,Y)\frac{X}{\rho^2}\,dXdY \tag{5.5}$$

$$J_6 = \iint_{S_e} T(X,Y)\frac{Y}{\rho^2}\,dXdY \tag{5.6}$$

In these equations, $\rho = \sqrt{X^2 + Y^2}$. In view of Eqs. (2.109)–(2.112) and (1.9), the combined ICs of the elastic half-space may be expressed as a function of these integral expressions J_1 to J_6 according to the following equations:

$$A_{xx} = \frac{1}{\pi G}\left((1-\nu)\ J_1 + \nu\ J_2\right) \tag{5.7}$$

$$A_{xy} = \frac{\nu}{\pi G} J_3 \tag{5.8}$$

$$A_{yy} = \frac{1}{\pi G}\left((1-\nu)\ J_1 + \nu\ J_4\right) \tag{5.9}$$

$$A_{xz} = \frac{K}{\pi G} J_5 \tag{5.10}$$

$$A_{yz} = \frac{K}{\pi G} J_6 \tag{5.11}$$

$$A_{zz} = \frac{(1-\nu)}{\pi G} J_1 \tag{5.12}$$

Here the exact analytical ICs of the half-space are denoted with upper case letter *A*. The remaining ICs may be deduced by reciprocity, as indicated next:

$$A_{yx} = A_{xy} \tag{5.13}$$

$$A_{zx} = -A_{xz} \tag{5.14}$$

$$A_{zy} = -A_{yz} \tag{5.15}$$

In the next sections the analytical expressions of the half-space ICs for two particular cases with different shape functions $T(X, Y)$ are developed, both of them with the distributed load applied in rectangular surface elements aligned with the *X* and *Y* axes of the adopted Cartesian reference system. Taking the loaded rectangular element aligned with the *X* and *Y* axes does not imply generality loss, as the expressions of the ICs in other coordinate systems may be obtained with a coordinate transformation from their expressions in this coordinate system, and combining appropriately as necessary the ICs associated to loads along the *X* and *Y* directions.

Expressions will be given for some of the ICs, being able to deduce the remaining ones with the following considerations:

- The A_{yy} and A_{yz} ICs may be obtained by means of similarity relationships from the A_{xx} and A_{xz} ICs respectively, taking into account the (anti)symmetry of the geometry and loads in longitudinal and lateral direction with respect to the central point of the loaded domain.
- The reciprocity relationships (5.13)–(5.15) may be applied for the A_{yx}, A_{zx} and A_{zy} ICs.

5.1.1 Uniform rectangular elements

A rectangular element centred in the point (x, y) of the surface is considered. The dimension of the half-side of the rectangle aligned with the X axis is a, and that of the half-side aligned with the Y axis is b. There is a unitary distributed load p_j applied in this element in direction j (with $j = x$, y or z), which is uniform inside the element and zero outside it; that is, $p_j = 1$ in all (X, Y) where $x–a \le X \le x+a$ and $y–b \le Y \le y+b$, and $p_j = 0$ outside the rectangle.

The analytical expressions for this type of element may be found in [Kalker 2001] for instance. The expressions for the A_{xx}, A_{xy}, A_{xz} and A_{zz} ICs are given next.

$$A_{xx}(x,y) = \frac{1}{\pi G}\left[\left[Y \ln(X+\rho) + (1-\nu)\ X \ln(Y+\rho)\right]\right] \tag{5.16}$$

$$A_{xy}(x,y) = \frac{-\nu}{\pi G}\left[\left[\rho\right]\right] \tag{5.17}$$

$$A_{xz}(x,y) = \frac{-K}{\pi G}\left[\left[Y \ln(\rho) + X\ \mathrm{atan}\left(\frac{Y}{X}\right)\right]\right] \tag{5.18}$$

$$A_{zz}(x,y) = \frac{1-\nu}{\pi G}\left[\left[Y \ln(X+\rho) + X \ln(Y+\rho)\right]\right] \tag{5.19}$$

The double brackets [[ ]] in the previous equations (5.16)–(5.19) denote the evaluation of the integral expression f inside them in the limit points ($-x-a$, $-x+a$) in X and ($-y-b$, $-y+b$) in Y (ρ is also evaluated in each of the limit points), i.e.:

$$[[\, f \,]] = f(-x+a,\, -y+b) - f(-x-a,\, -y+b) - f(-x+a,\, -y-b) + f(-x-a,\, -y-b) \tag{5.20}$$

Each term in Eqs. (5.16)–(5.19) tends to 0 at the singularity points. The A_{yy} and A_{yz} ICs may be deduced by similarity from A_{xx} and A_{xz}, according to the following equations:

$$A_{yy}(x, y, a, b) = A_{xx}(y, x, b, a) \tag{5.21}$$

$$A_{yz}(x, y, a, b) = A_{xz}(y, x, b, a) \tag{5.22}$$

Being $A_{ij}(x, y, a, b)$ the A_{ij} IC of the element centred in (x, y) with longitudinal half-side a and lateral b. That is to say, the ICs A_{yy} and A_{yz} may be obtained applying Eq. (5.20) to the expressions of the A_{xx} and A_{xz} ICs respectively, and interchanging x with y, and a with b.

5.1.2 Bilinear rectangular elements

In this section the analytical expressions of the ICs for rectangular elements with bilinearly varying load distribution are given. These expressions have been developed in this thesis, not having found them published in the literature before. It is interesting to have these analytical expressions and avoid the numerical integration of the ICs, which presents difficulties due to the singularity of the IFs at the origin. This is a common difficulty in the Boundary Element Method, due to which usually adaptive quadratures are necessary to obtain valid results in the numerical integration of the ICs.

In [Svec 1971] and [Kalker 1972] analytical expressions of the A_{zz} IC for triangular elements with non-uniform load distributions may be found. In [Svec 1971] polynomial pressure distributions are considered (which may be particularized to uniform and linear for example), and in [Kalker 1972] linear pressure distributions are considered. In [Li J 2001] the necessary expressions

for all the ICs of triangular elements with uniform, linear and bilinear load distributions are given.

In [Kaiser 2012] rectangular elements with bilinear load distribution are used, although the expressions of the ICs are not given. This type of element is also used in [Vollebregt 2009], where the calculation of the ICs is carried out numerically, subdividing each element in turn in a mesh of sub-elements, each of them with uniform load. The ICs are obtained as the sum of the contributions of each of the sub-elements, which may be expressed analytically with the formulae of the previous Section 5.1.1.

[Dydo 1995] provides the ICs for the displacements and stresses in any point of the elastic half-space due to uniform, linear and bilinear load distributions in a surface rectangle, in terms of harmonic functions which have to be differentiated to obtain the sought ICs. However, the differentiation of these harmonic functions is not an easy task.

As it is mentioned in [Zhao J 2016], it is difficult to obtain the expressions of the ICs for rectangular elements with bilinear load distribution. The procedure followed here to obtain these expressions has consisted on performing directly the analytical integration of Eqs. (5.1)–(5.6) in a rectangular element, inserting the bilinear load distribution function $T(X, Y)$ in them. For this purpose, use has been made of handbooks of tables of integrals, as those which may be found in [Jeffrey 2008] and [Spiegel 2009]. During this integration, functions with multiple branches appear, and care must be taken to choose the appropriate branch in the primitive function in each case. To assist in this task, the different integrals have been carried out both analytically and numerically, and the resulting primitive functions have been graphed together for different positions of the loaded domain in the x-y plane. Details of the integrals are provided in the Appendix. Here the final analytical expressions are provided, with which the ICs are completely defined. These expressions have been published in [Blanco-Lorenzo 2016]. More recently, expressions of the ICs of the subsurface stresses for this type of bilinear element have been published in [van der Wekken 2019].

Rectangular elements of the surface with longitudinal dimension a and lateral b are considered[16]. A unitary load in a node located in the coordinates (x, y) results in a bilinear load distribution in each of the four rectangular elements surrounding the node, which has unit value in that node and zero value in the rest of the nodes of the four elements. The situation is depicted in Figure 5.1.

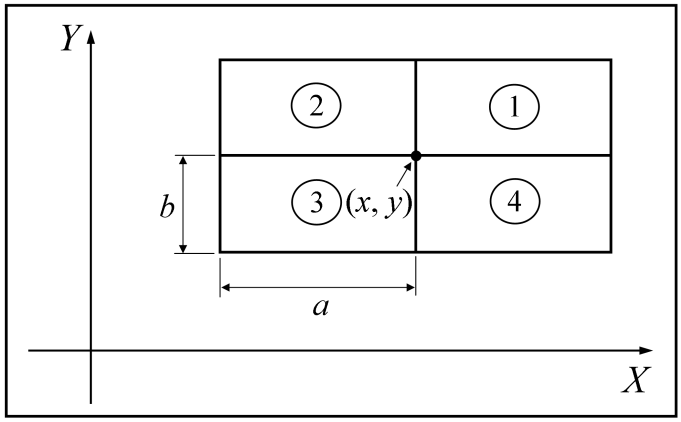


Figure 5.1. Definition of the coordinate system and element dimensions for the calculation of the ICs for rectangular elements with bilinearly varying load distributions.

For example, the bilinear load distribution $T(X, Y)$ in element number 1 of Figure 5.1 due to a unit nodal load on the node at (x, y) follows Eq. (5.23):

$$T(X,Y)=\left(1-\frac{X-x}{a}\right)\left(1-\frac{Y-y}{b}\right) \tag{5.23}$$

The contributions J_2^1, J_3^1 and J_5^1 of element 1 to the integral expressions J_2, J_3 and J_5 of Eqs. (5.1)–(5.6) associated to the unit nodal load on the node at (x, y) are given in the following equations. The integral expressions for J_4 and J_6 may

[16] Note that in this section the dimensions of the sides of the rectangles are a and b, while in the previous section 5.1.1 a and b designated the dimensions of the half-sides. However, in both cases the dimensions of the domain with non-zero load is the same, $2a \times 2b$.

be obtained from the ones corresponding to J_2 and J_5 respectively, interchanging x with y and a with b.

$$J_2^{\,1}(x,y,a,b)=\frac{1}{ab}\left[\left[(x+a)\left\{\left((y+b)Y-\frac{Y^2}{2}\right)\ln(X+\rho)\right.\right.\right.$$
$$\left.\left.\left.+\frac{1}{2}X\rho\right\}-(y+b)Y\rho+\frac{2}{3}\rho^3-\rho X^2\right]\right] \tag{5.24}$$

$$J_3^{\,1}(x,y,a,b)=\frac{1}{ab}\left[\left[(x+a)\left(\frac{Y}{2}-(y+b)\right)\rho\right.\right.$$
$$+\left(\frac{X^3}{3}-\frac{1}{2}(x+a)X^2\right)\ln(Y+\rho)+\left(\frac{1}{2}(y+b)-\frac{Y}{3}\right)X\rho$$
$$\left.\left.+\left(\frac{Y^3}{3}-\frac{1}{2}(y+b)Y^2\right)\ln(X+\rho)\right]\right] \tag{5.25}$$

$$J_5^{\,1}(x,y,a,b)=\frac{1}{ab}\left[\left[\left\{(x+a)\left((y+b)Y-\frac{\rho^2}{2}\right)+\frac{X^3}{3}\right\}\ln\rho\right.\right.$$
$$+(x+a)(y+b)X\ \operatorname{atan}\left(\frac{Y}{X}\right)$$
$$+\left(\frac{1}{2}(y+b)\rho^2-\frac{Y^3}{3}\right)\operatorname{atan}\left(\frac{X}{Y}\right)$$
$$\left.\left.-\frac{\pi}{4}(y+b)\operatorname{sgn}(XY)X^2\right]\right]+\frac{y-b}{6} \tag{5.26}$$

The sgn() function that appears in Eq. (5.26) gives the sign of the argument; i.e. 1 for argument equal to or greater than 0, and –1 for negative arguments.

The double brackets in Eqs. (5.24)–(5.26) have a different meaning from that in Eqs. (5.16)–(5.19): in this case they denote the evaluation of the integral expression f inside them at the limit points $(x, x+a)$ for X and $(y, y+b)$ for Y (ρ is also evaluated in each of the limit points, but calculated with the X and Y variables, instead of with x and y as it appears in Eq. (2.108)), i.e.:

$$[[f]] = f(x+a, y+b) - f(x, y+b) - f(x+a, y) + f(x, y) \tag{5.27}$$

Each term of Eqs. (5.24)–(5.26) tends to 0 at the singularity points.

The superindex 1 in the left term of Eqs. (5.24)–(5.26) denotes the contribution of element number 1. The contributions of the other elements surrounding the node are deduced easily by similarity considerations as a function of the respective contributions of the element number 1, according to the following equations:

$$J_2^{\,2}(x, y, a, b) = J_2^{\,1}(-x, y, a, b) \tag{5.28}$$

$$J_2^{\,3}(x, y, a, b) = J_2^{\,1}(-x, -y, a, b) \tag{5.29}$$

$$J_2^{\,4}(x, y, a, b) = J_2^{\,1}(x, -y, a, b) \tag{5.30}$$

$$J_3^{\,2}(x, y, a, b) = -J_3^{\,1}(-x, y, a, b) \tag{5.31}$$

$$J_3^{\,3}(x, y, a, b) = J_3^{\,1}(-x, -y, a, b) \tag{5.32}$$

$$J_3^{\,4}(x, y, a, b) = -J_3^{\,1}(x, -y, a, b) \tag{5.33}$$

$$J_5^{\,2}(x, y, a, b) = -J_5^{\,1}(-x, y, a, b) \tag{5.34}$$

$$J_5^{\,3}(x, y, a, b) = -J_5^{\,1}(-x, -y, a, b) \tag{5.35}$$

$$J_5^{\,4}(x, y, a, b) = J_5^{\,1}(x, -y, a, b) \tag{5.36}$$

The total integral expressions J_i due to the unit nodal load in the node located at (x, y) are obtained adding the contributions of the four elements surrounding the node, as it is indicated in the following equation:

$$J_i(x,y,a,b)=\sum_{k=1}^{4} J_i^{\,k}(x,y,a,b) \tag{5.37}$$

After evaluating the previous integral expressions (5.24)–(5.26) in the four elements surrounding the node according to Eqs. (5.27)–(5.37), simplifying and grouping terms, one arrives to the following expressions for each of the integrals:

$$\begin{aligned} J_2(x,y,a,b)=\frac{1}{ab}\Bigg(& (x+a)\Bigg(\frac{1}{2}\left(F_1(x,y+b,a)+F_1(x,-y+b,a)\right) \\ & -F_1(x,y,a)\Bigg)+(-x+a)\Bigg(-F_1(-x,y,a) \\ & +\frac{1}{2}\left(F_1(-x,y+b,a)+F_1(-x,-y+b,a)\right)\Bigg) \\ & -\frac{1}{2}\Big((x+a)^2\left(r_{11}-2r_{12}+r_{41}\right)-2x^2\left(r_{13}-2r_{14}+r_{33}\right) \\ & +(-x+a)^2\left(r_{21}-2r_{22}+r_{31}\right)\Big)-\Big((y+b)^2\left(r_{11}-2r_{13}+r_{21}\right) \\ & -2y^2\left(r_{12}-2r_{14}+r_{22}\right)+(-y+b)^2\left(r_{31}-2r_{33}+r_{41}\right)\Big) \\ & +\frac{2}{3}\left(r_{11}^{\,3}+r_{21}^{\,3}+r_{31}^{\,3}+r_{41}^{\,3}-2\left(r_{12}^{\,3}+r_{13}^{\,3}+r_{22}^{\,3}+r_{33}^{\,3}\right)+4r_{14}^{\,3}\right)\Bigg) \end{aligned} \tag{5.38}$$

$$J_3(x,y,a,b)=\frac{1}{3ab}\Big((x+a)\big(-(y+b)r_{11}+(-y+b)r_{41}+2y\,r_{12}\big)$$

$$+F_2(x,-y,b)+F_2(y,-x,a)$$

$$+(-x+a)\big((y+b)r_{21}-(-y+b)r_{31}-2y\,r_{22}\big)$$

$$+2x\big((y+b)r_{13}-2y\,r_{14}-(-y+b)r_{33}\big)$$

$$+\frac{1}{2}\big(F_2(-x+a,-y,b)+F_2(x+a,y,b)$$

$$+F_2(-y+b,-x,a)+F_2(y+b,x,a)\big)\Big) \tag{5.39}$$

$$J_5(x,y,a,b)=\frac{1}{ab}\Bigg(\frac{1}{2}\Big(F_3(-x+a,y,b)-F_3(x+a,y,b)+F_5(x,y+b,a)$$

$$+F_5(x,-y+b,a)\Big)+F_3(x,y,b)+F_4(x+a,y,b)$$

$$-F_4(-x+a,y,b)-2F_4(x,y,b)-F_5(x,y,a)$$

$$+\frac{\pi}{4}\Big\{\operatorname{sgn}(x+a)(x+a)^2\big(-|y+b|-|-y+b|+2|y|\big)$$

$$+2\operatorname{sgn}(x)x^2\big(|y+b|+|-y+b|-2|y|\big)$$

$$+\operatorname{sgn}(-x+a)(-x+a)^2\big(|y+b|+|-y+b|-2|y|\big)\Big\}\Bigg) \tag{5.40}$$

The following auxiliary functions and variables are defined for Eqs. (5.38)–(5.40):

$$F_1(X,Y,L) = Y^2 \ln\left(\frac{X+L+\sqrt{(X+L)^2+Y^2}}{X+\sqrt{X^2+Y^2}}\right) \tag{5.41}$$

$$F_2(X,Y,L) = X^3 \ln\left(\frac{\left(-Y+L+\sqrt{X^2+(-Y+L)^2}\right)\left(Y+\sqrt{X^2+Y^2}\right)}{\left(-Y+\sqrt{X^2+Y^2}\right)\left(Y+L+\sqrt{X^2+(Y+L)^2}\right)}\right) \tag{5.42}$$

$$\begin{aligned} F_3(X,Y,L) = X\Bigg(&\frac{1}{2}\left(\frac{X^2}{3}-(Y+L)^2\right)\ln\left(X^2+(Y+L)^2\right) \\ &+\frac{1}{2}\left(\frac{X^2}{3}-(-Y+L)^2\right)\ln\left(X^2+(-Y+L)^2\right) \\ &-\left(\frac{X^2}{3}-Y^2\right)\ln\left(X^2+Y^2\right)\Bigg) \end{aligned} \tag{5.43}$$

$$\begin{aligned} F_4(X,Y,L) = X^2\Bigg(&(Y+L)\,\operatorname{atan}\left(\frac{Y+L}{X}\right) \\ &+(-Y+L)\,\operatorname{atan}\left(\frac{-Y+L}{X}\right)-2Y\operatorname{atan}\left(\frac{Y}{X}\right)\Bigg) \end{aligned} \tag{5.44}$$

$$F_5(X,Y,L) = Y\left(\left(\frac{Y^2}{3}+(X+L)^2\right)\operatorname{atan}\left(\frac{X+L}{Y}\right)\right.$$

$$-\left(\frac{Y^2}{3}+(-X+L)^2\right)\operatorname{atan}\left(\frac{-X+L}{Y}\right)$$

$$\left.-2\left(\frac{Y^2}{3}+X^2\right)\operatorname{atan}\left(\frac{X}{Y}\right)\right) \tag{5.45}$$

$$r_{11} = \sqrt{(x+a)^2+(y+b)^2} \tag{5.46}$$

$$r_{12} = \sqrt{(x+a)^2+y^2} \tag{5.47}$$

$$r_{13} = \sqrt{x^2+(y+b)^2} \tag{5.48}$$

$$r_{14} = \sqrt{x^2+y^2} \tag{5.49}$$

$$r_{21} = \sqrt{(-x+a)^2+(y+b)^2} \tag{5.50}$$

$$r_{22} = \sqrt{(-x+a)^2+y^2} \tag{5.51}$$

$$r_{31} = \sqrt{(-x+a)^2+(-y+b)^2} \tag{5.52}$$

$$r_{33} = \sqrt{x^2+(-y+b)^2} \tag{5.53}$$

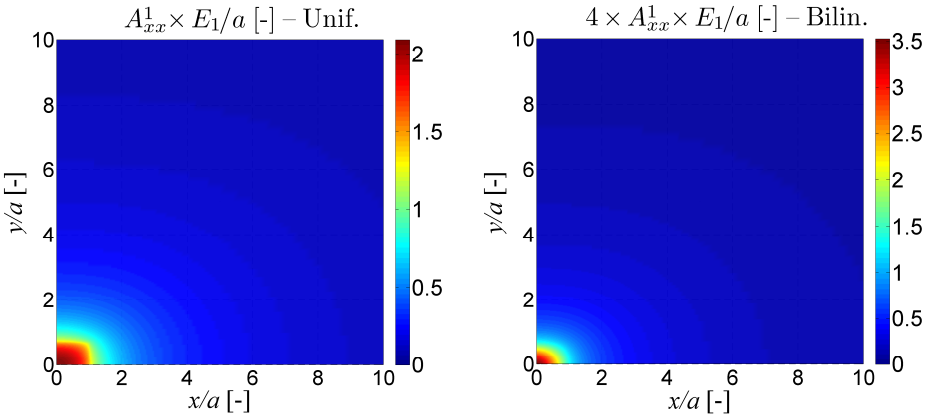
$$r_{41} = \sqrt{(x+a)^2+(-y+b)^2} \tag{5.54}$$

5.1.3 Comparisons between uniform and bilinear rectangular elements

In this section, some comparisons are made of the half-space ICs of rectangular surface elements with uniform and bilinear load distributions. On the other hand, the precision of some results of the normal and the tangential parts of the contact problem obtained with the exact contact theory with both types of elements are compared.

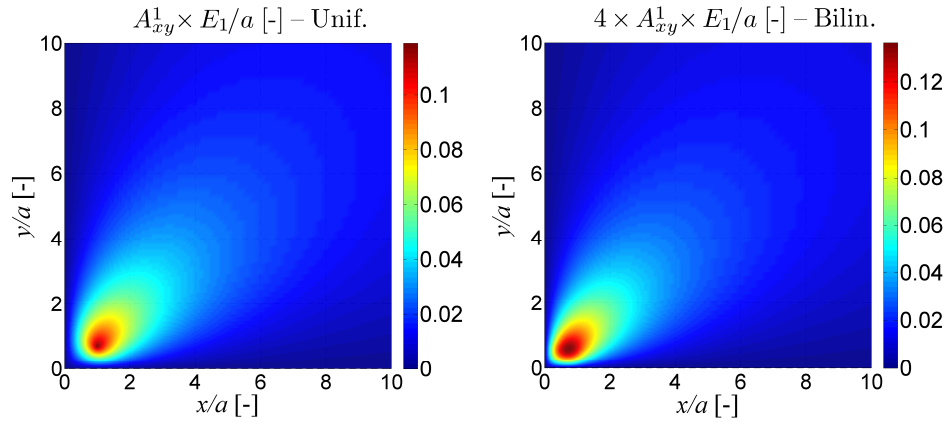
In order to perform a more direct comparison between the ICs with both types of load distribution, the unit load is multiplied by four in the case of the bilinear distribution, so that the total applied force is equal to the case of uniform load distribution. The ICs are presented in non-dimensional form, multiplied by the Young's modulus of the half-space E_1 and divided by the longitudinal dimension a of the loaded element. A value of the coefficient of Poisson ν_1 of 0.30 is considered, and a b/a ratio of the sides of the loaded element of 0.70.

In the following figures the A^1_{xx}, A^1_{xy}, A^1_{xz} and A^1_{zz} ICs are graphed as a function of the position of the centre of the loaded domain in the surface. In each figure one of the ICs is shown, with both types of load distribution, only in the first quadrant of the surface, indicating in each case the symmetry or antisymmetry relations in the rest of quadrants.



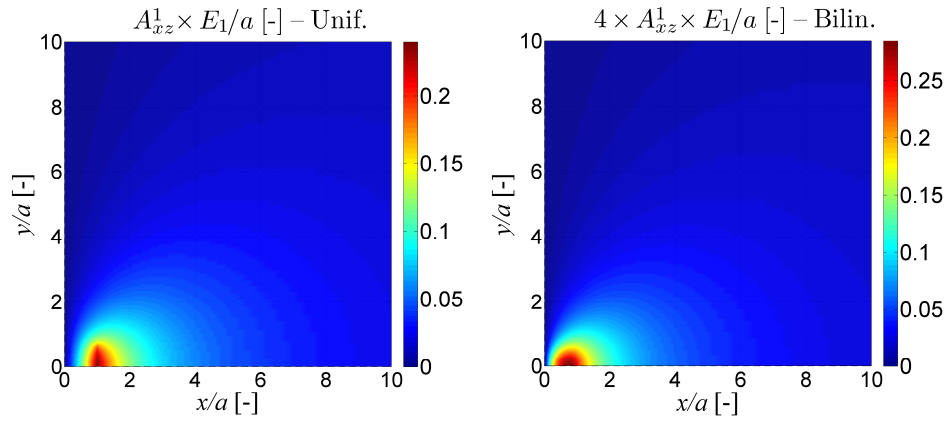
(a) Uniform load distribution. (b) Bilinear load distribution.

Figure 5.2. A^1_{xx} IC with ν_1 = 0.30, symmetric with respect to x = 0 and y = 0. Rectangular loaded element with b/a = 0.70.



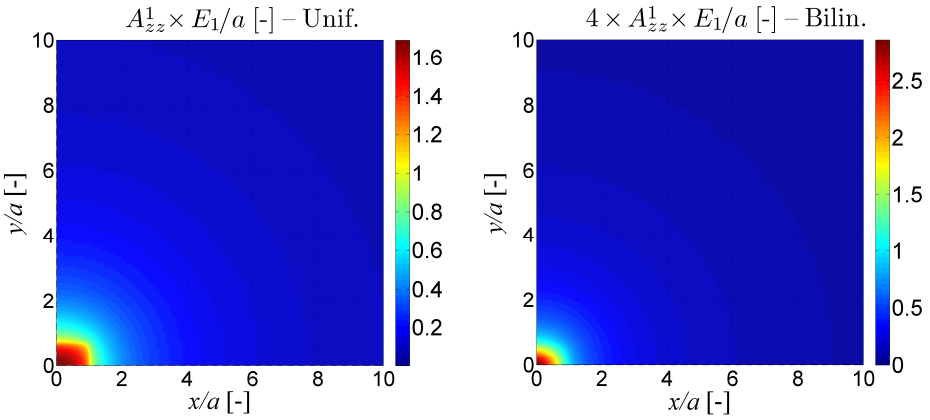
(a) Uniform. (b) Bilinear.

Figure 5.3. A^1_{xy} IC with ν_1 = 0.30, antisymmetric with respect to x = 0 and y = 0. Rectangular loaded element with b/a = 0.70.



(a) Uniform. (b) Bilinear.

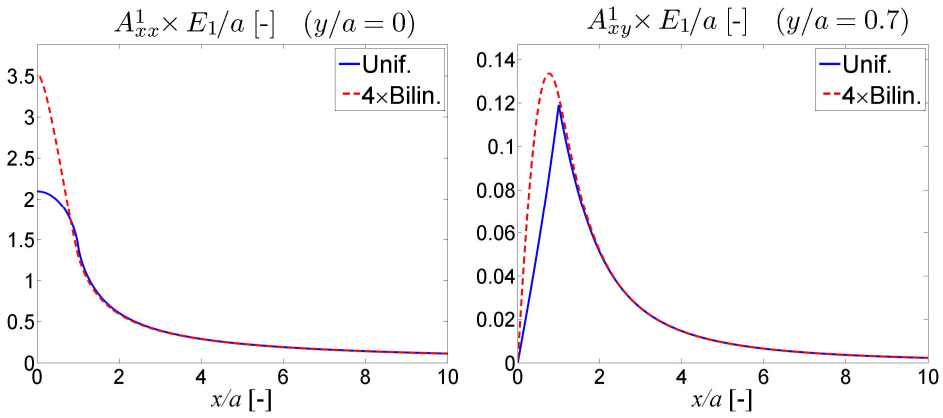
Figure 5.4. A^1_{xz} IC with ν_1 = 0.30, antisymmetric with respect to x = 0 and symmetric with respect to y = 0. Rectangular loaded element with b/a = 0.70.



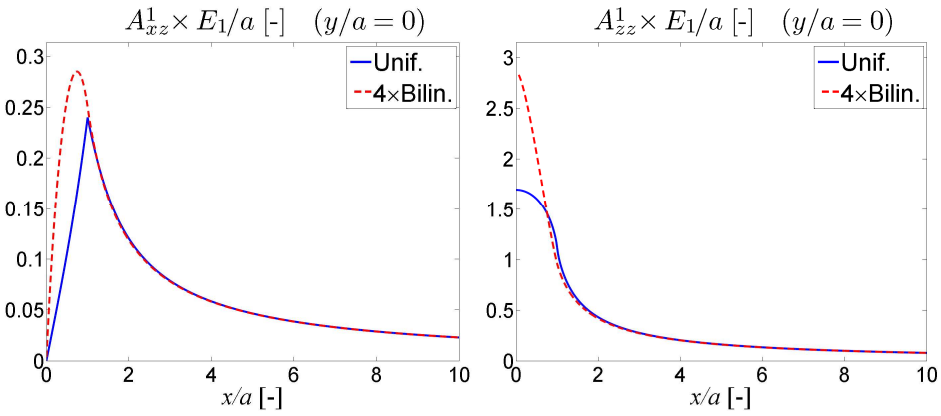
(a) Uniform. (b) Bilinear.

Figure 5.5. A^1_{zz} IC with ν_1 = 0.30, symmetric with respect to x = 0 and y = 0. Rectangular loaded element with b/a = 0.70.

In order to better appreciate the differences between the ICs with both load distribution types, in the Figure 5.6 the ICs shown in the previous figures are plotted together with both load distribution types in each case, as a function of the non-dimensional longitudinal coordinate x/a, in the given y/a lateral positions.



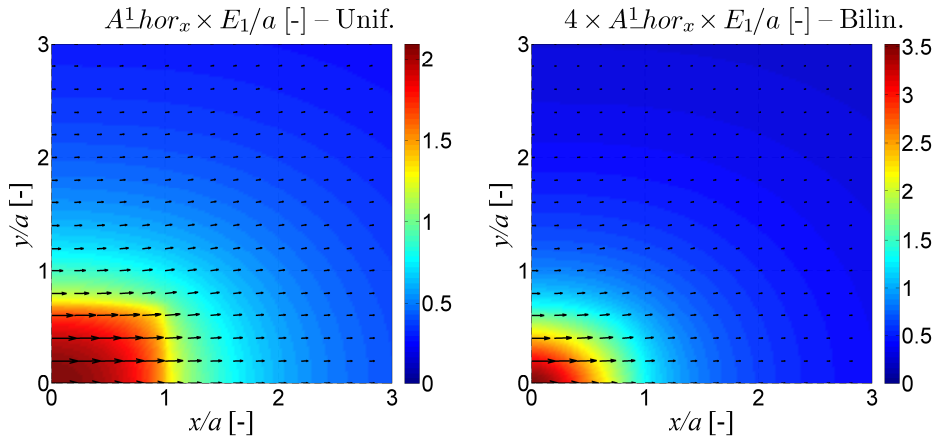
(a) A^1_{xx}. (b) A^1_{xy}.



(c) A^1_{xz}. (d) A^1_{zz}.

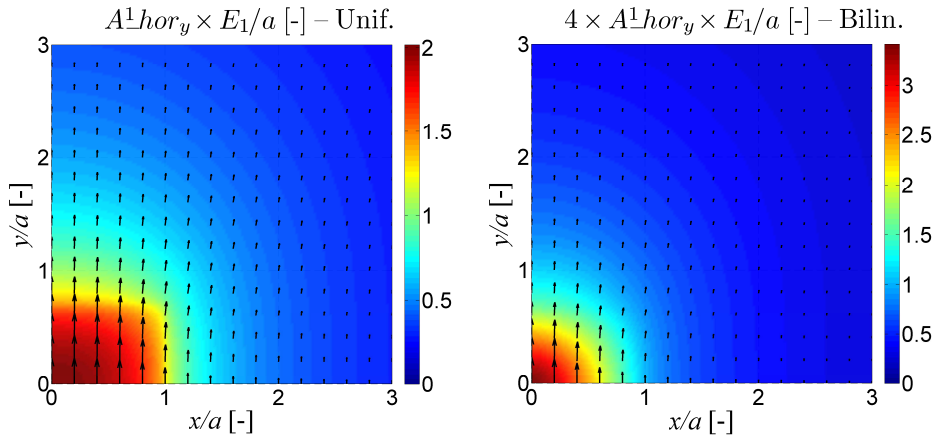
Figure 5.6. Comparison of the ICs with uniform (solid blue lines) and bilinear (dashed red lines) load distribution, as a function of the longitudinal position *x*/*a* of the centre of the loaded domain, for lateral positions *y*/*a* indicated in each case.

As may be seen in Figure 5.6, the variations of the ICs inside the loaded domain are more pronounced with the bilinear load distribution, in line with the variation of the applied load inside the element. On the other hand, as the loaded domain is moved away, the ICs with both load distributions tend to become equal, as may be expected by Saint-Venant's principle. Another point which may be appreciated in the previous figures is that the direct ICs (the ones which provide the displacements in a given direction due to loads applied in the same direction) are considerably larger than the crossed ICs (the ones which provide the displacements in a given direction due to loads applied in a different direction). This may also be seen in the following figures, which show the magnitude and direction of the horizontal displacements $A^1\_hor_i$ generated in the origin by elements centred in different (*x*, *y*) positions, loaded with unitary stresses acting in different *i* = *x*, *y* or *z* directions.



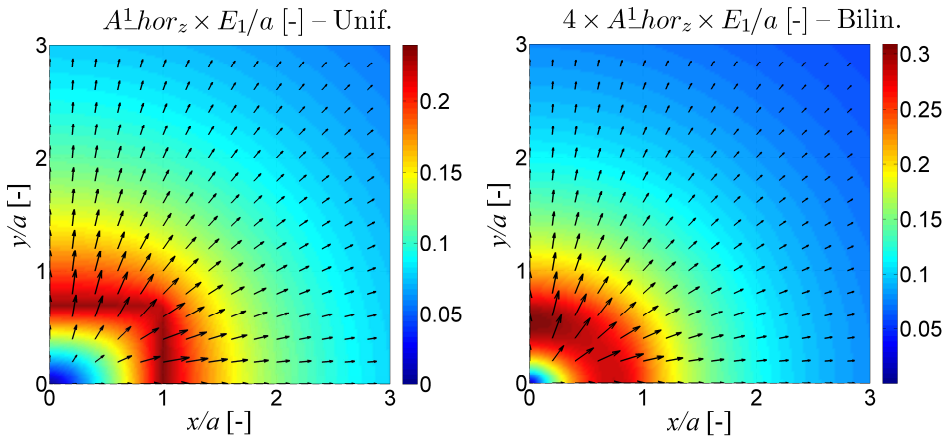
(a) Uniform. (b) Bilinear.

Figure 5.7. $A^1\_hor_x$ horizontal displacement generated at the origin of the half-space with $\nu_1 = 0.30$ by a distributed load acting in x direction, antisymmetric with respect to $x = 0$ and symmetric with respect to $y = 0$. Rectangular loaded element with $b/a = 0.70$.



(a) Uniform. (b) Bilinear.

Figure 5.8. $A^1\_hor_y$ horizontal displacement generated at the origin of the half-space with $\nu_1 = 0.30$ by a distributed load acting in y direction, symmetric with respect to $x = 0$ and antisymmetric with respect to $y = 0$. Rectangular loaded element with $b/a = 0.70$.



(a) Uniform. (b) Bilinear.

Figure 5.9. $A^1\_hor_z$ horizontal displacement generated at the origin of the half-space with $\nu_1 = 0.30$ by a distributed load acting in z direction, symmetric with respect to $x = 0$ and $y = 0$. Rectangular loaded element with $b/a = 0.70$.

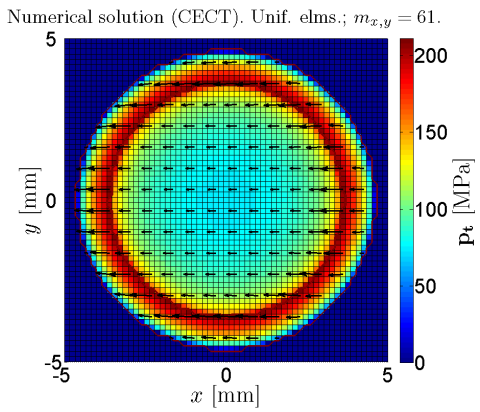
Next a comparison is presented of the performance of both types of elements in the solution of a frictional contact problem with CECT, the implementation of the exact contact theory developed in this thesis (cf. §4.1). The problem considered for the comparison is a Cattaneo shift. The input data of the considered case are as follows:

- Material of the contacting bodies: steel, with Young's modulus E of 210 MPa and coefficient of Poisson ν of 0.30.
- Combined curvature radii of the contacting surfaces, R_x and R_y: 300 mm in both principal directions. The contact patch is therefore circular.
- Normal load, N: 50 kN.
- Tangential shift, W_x: 0.01 mm.
- Coefficient of friction, μ: 0.30.

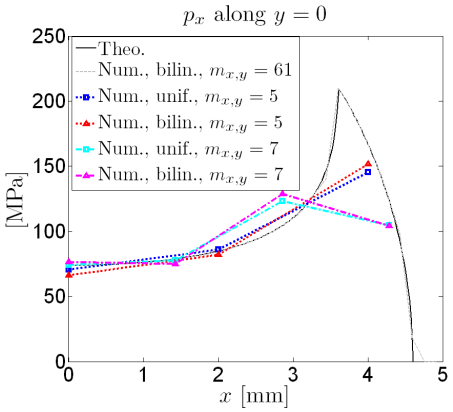
The theoretical radius of the contact patch in this case is 4.603 mm. To obtain the numerical solution of the problem with the exact contact theory, the PCS is defined as a square with 10 mm side, and is discretized in an equal number of elements in both directions. Results are computed for different mesh densities. With each mesh density, the meshes are adjusted to have the same number of

analysis points[17] (APs) and the same element size with both uniform and bilinear element types, and are located so that the central AP is located on the central position of the contact patch[18].

Figure 5.10 shows some theoretical and numerical results of this problem.



(a) Contact patch contours of adhesion and slip zones, and magnitude and direction of the tangential stresses $\mathbf{p_t}$. Numerical solution obtained with uniform elements and $m_{x,y}$ = 61 APs in each direction.



(b) Distribution of longitudinal tangential stresses p_x along the contact patch centreline. Theoretical ("Theo.") and numerical ("Num.") solutions with different element types and mesh densities as indicated in the legend.

Figure 5.10. Results of a Cattaneo shift problem.

Table 5.1 shows some representative results of the normal part of the contact problem, and Table 5.2 some of the tangential part. Results obtained with different mesh densities are presented, defined with the number of APs in each direction $m_{x,y}$, and with uniform and bilinear elements (indicated in the tables as "Unif." and "Bilin.", respectively). The results are presented in non-dimensional

[17] The APs are the centres of the elements in the case of the uniform elements, and the nodes in the case of the bilinear elements, as explained in §4.1.1.1.

[18] Therefore, with the uniform elements meshes with an odd number of elements in each direction are used here, and with the bilinear elements meshes with an even number of elements (one less in each direction than the number of the corresponding mesh with uniform elements).

form, divided by the corresponding analytically calculated theoretical result indicated in the captions of the tables.

Table 5.1. Ratios between numerical and theoretical results of the normal contact problem for the considered Cattaneo shift, with different mesh densities and element types. *pn_max*: maximum normal pressure; δ: approach. Theoretical results: *pn_max* = 1127 MPa; δ = 7.061E–02 mm.

	pn_max		δ	
$m_{x,y}$ [-]	Unif.	Bilin.	Unif.	Bilin.
5	1.0059	1.0042	0.9893	0.9860
7	1.0039	1.0117	0.9950	0.9924
9	1.0043	1.0067	0.9990	0.9969
11	1.0027	1.0053	0.9991	0.9985
15	1.0018	1.0024	0.9999	0.9988
21	1.0006	1.0009	0.9995	0.9989
31	1.0005	1.0008	0.9998	0.9997
41	1.0004	1.0005	1.0000	0.9998
51	1.0002	1.0003	0.9999	0.9999
61	1.0005	1.0006	1.0000	0.9999

Table 5.2. Ratios between numerical and theoretical results of the tangential contact problem for the considered Cattaneo shift, with different mesh densities and element types. F_x: resultant tangential force; *pt_centr*: tangential stress at the contact patch centre. *pt_max*: maximum tangential stress in the contact patch. Approximate theoretical results: F_x = 7.832 kN; *pt_centr* = 73.76 MPa; *pt_max* = 210.8 MPa.

	F_x		*pt_centr*		*pt_max*	
$m_{x,y}$ [-]	Unif.	Bilin.	Unif.	Bilin.	Unif.	Bilin.
5	1.0204	1.0006	0.9582	0.9005	0.6899	0.7419
7	0.9957	0.9911	1.0072	1.0401	0.7055	0.8034
9	0.9952	0.9962	1.0018	0.9887	0.8499	0.9651
11	1.0024	0.9996	0.9961	0.9961	0.9562	1.0080
15	0.9976	0.9986	1.0020	0.9996	0.9318	0.9943
21	1.0002	1.0003	0.9995	1.0011	0.9694	0.9882
31	1.0003	1.0003	1.0001	0.9949	0.9783	0.9997
41	1.0003	1.0002	1.0021	1.0000	0.9887	1.0000
51	1.0003	1.0003	1.0013	1.0014	0.9974	1.0002
61	1.0003	1.0004	1.0004	1.0045	0.9998	1.0019

The numerical results of both the normal and the tangential parts tend to approach their theoretical values as the mesh is refined, with both element types. As may be seen in Table 5.1, the results obtained in the normal part are good with both element types, with errors below 2% in the normal pressures

and approaches even with the coarsest meshes. The errors of the numerical results in the tangential part are somewhat higher, as may be observed in Table 5.2, especially those of maximum tangential stress *pt_max*. This is related with the highly localized character of the maximum tangential stress in this problem, which may be appreciated in Figure 5.10. The errors in *pt_max* are lower with bilinear elements. However, the tangential stresses obtained with bilinear elements tend to present a somewhat higher oscillation or noise than those obtained with uniform elements, as may be seen in Figure 5.10b. Regarding the values of the resultant longitudinal force F_x and the tangential stress at the centre *pt_centr*, no clear improvement is observed with the bilinear elements with respect to the uniform elements.

On the other hand, when applying the exact contact theory with each of these types of elements, in general more convergence difficulties have been observed in the iterative algorithms for the solution of the tangential part of contact problems with bilinear than with uniform elements. This may possibly be related to the fact that the direct ICs reach higher values in the centre of the loaded domain with respect to the rest of the mesh with uniform elements, as may be observed in Figure 5.11. In each curve of this figure, the ICs at the centre of the loaded domain (at $x/a = 0$), and at the adjacent AP of the mesh along the x axis (at $x/a = 1$) are marked. In this way, the terms in the diagonal of the matrices of ICs turn out to be more dominant with uniform elements, and this contributes to a better numerical conditioning of the matrices.

In conclusion, no significant precision improvement has been verified in the results obtained in the contact surface when using the exact contact theory with bilinear elements, with respect to the uniform elements. This is in line with the results reported in [Vollebregt 2009]. On the other hand, van der Wekken [van der Wekken 2019] shows that the bilinear elements have the merit of avoiding the numerical artefacts in the computed stresses near the surface present with the uniform elements close to the element boundaries.

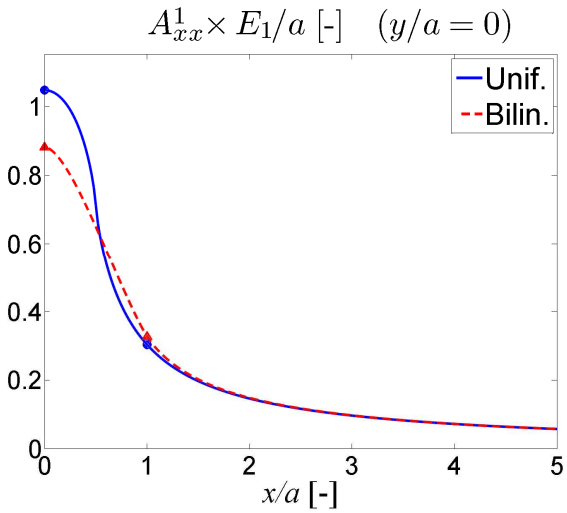


Figure 5.11. Comparison of A^1_{xx} ICs of the half-space with ν_1 = 0.30, with uniform and bilinear rectangular elements. b/a = 0.70. In this case a and b are the total longitudinal and lateral dimensions for both element types[19].

5.2 Solids with non-planar surfaces

In conformal contact problems, the dimensions of the contact zone may no longer be much smaller than the curvature radii of the contacting surfaces, and may reach a comparable size. In these conditions, the contact surfaces are no longer flat, and the hypothesis that the contacting bodies behave as half-spaces around the contact becomes questionable. In this section, the elastic behaviour of solids with non-planar surfaces is studied. As in the case of the half-space, particularly the IFs and ICs of the surface displacements of the solid due to surface loads are considered. Firstly, two particular cases with available

[19] In this figure, the same size is considered for both element types, with the aim of locating the APs in the same positions for both of them. In contrast, in the previous figures of this chapter where the different ICs are shown, it has been chosen to consider the same size of the loaded domain for both types of load distribution, being said domain composed by a single element in the case of the uniform load distribution, and by 4 elements (two in x direction by two in y direction) in the case of the bilinear load distribution.

analytical results are studied. Afterwards the numerical calculation of the ICs for general cases is undertaken. Lastly, an analytical approximation to avoid the generally costly numerical calculation is proposed.

The characteristics of the solids considered set out in §2.2 are summarized and specified for the non-planar case as follows:

- The solids are massive (not hollow) around the contact, and the contact is far from surface borders or constraint locations as compared to contact patch dimensions. In this way, the load transmission beneath the contact surfaces resembles that of the half-space.
- The solids are homogeneous, isotropic, and with linear elastic behaviour.
- The surfaces are approximately prismatic (i.e. extruded in one direction) around the contact.
- The surfaces are smooth, not having local curvature radii small in relation to the contact patch dimensions.

These assumptions are usually fulfilled in relevant rolling contact applications, like in the wheel–rail case or in rolling bearings.

5.2.1 Cases with available analytical results

The cases of 2D cylindrical geometries and of the sphere introduced in §2.2.2 are considered. Comparisons of the IFs and ICs of these geometries with those of the corresponding planar geometries are presented, identifying common features in their differences that will be a basis for the analytical approximation proposed in §5.2.3.1.

5.2.1.1 Cylindrical geometries under linear loads

In this section, the ICs derived from the IFs given in §2.2.2.1 are worked out. The superindexes used for the different ICs follow the same notation as in §2.2.2.1, identifying the considered solid (cylinder or cylindrical cavity) and the load and supporting condition, making reference to Figure 2.22 and Figure 2.23.

The $A_{ij}(\theta)$ ICs associated to a distributed line load (with units of pressure) uniformly in a surface element with angular dimension $\Delta\theta$ (not to be confused

with the total contact angle in 2D cylindrical contact, designated as well as $\Delta\theta$ in §2.4) and centred in θ are obtained evaluating the primitive functions $I_{ij}(\theta)$ of the $a_{ij}(\theta)$ IFs given in §2.2.2.1, at the points ($\theta + \Delta\theta/2$) and ($\theta - \Delta\theta/2$), as indicated in Eq. (5.55):

$$A_{ij} = R \times \left[I_{ij} \right]_{\theta-\Delta\theta/2}^{\theta+\Delta\theta/2} = R \times \left(I_{ij}\left(\theta + \Delta\theta/2\right) - I_{ij}\left(\theta - \Delta\theta/2\right) \right) \tag{5.55}$$

In this equation R is the radius of the cylinder or of the cylindrical cavity.

The following equations provide the primitive functions for the different IFs given in previously. In each case, appropriate additional terms are introduced in order for the resulting primitive functions to be continuous, so that they may directly be used in the evaluation of the integral expressions that define the ICs according to Eq. (5.55).

$$I_{nn}^{cyl,a} = -K_1 \sin\theta \, \ln\left|\tan\left(\frac{\theta}{2}\right)\right| - K_2\left(\cos\theta \, \mathrm{sgn}\left(\tilde{\theta}\right) - \mathrm{sgn}\left(\tilde{\theta}\right) - 4\left\lfloor \frac{\theta+\pi}{2\pi} \right\rfloor \right) \tag{5.56}$$

$$I_{nn}^{cyl,b} = \frac{-1}{4\pi G}\left(\left(\frac{\kappa+1}{2} \right) \sin\theta \, \ln\left(2 - 2\cos\theta\right) - \left(\kappa - 1\right)\left(\cos\theta \, F\left(\tilde{\theta}\right) + \pi \left\lfloor \frac{\theta}{2\pi} \right\rfloor \right) \right) \tag{5.57}$$

$$I_{nn}^{cyl,c} = I_{nn}^{cyl,b} + \frac{1}{4\pi G}\frac{\sin\theta}{2} \tag{5.58}$$

$$I_{sn}^{cyl,a} = -K_1\left(\cos\theta \, \ln\left|\tan\left(\frac{\theta}{2}\right)\right| - \ln\left|\sin\theta\right| \right) + K_2 \sin\theta \, \mathrm{sgn}\left(\tilde{\theta}\right) \tag{5.59}$$

$$I_{sn}^{cyl,b} = \frac{-1}{4\pi G}\left\{-\cos\theta \right.$$

$$-\left(\frac{\kappa+1}{2}\right)\left[\cos\theta\,(1-\ln(2-2\cos\theta))+\ln(1-\cos\theta)\right]$$

$$\left.+(\kappa-1)\left(\frac{\cos\theta}{2}+\sin\theta\,F(\tilde{\theta})\right)\right\} \tag{5.60}$$

$$I_{sn}^{cyl,c} = I_{sn}^{cyl,b} - \frac{1}{4\pi G}\frac{\cos\theta}{2} \tag{5.61}$$

$$I_{nn}^{cav,a} = -K_1\left(\sin\theta\ \ln\left|\tan\left(\frac{\theta}{2}\right)\right|-\theta\right)$$

$$+K_2\left(\cos\theta\,\mathrm{sgn}(\tilde{\theta})-\mathrm{sgn}(\tilde{\theta})-4\left\lfloor\frac{\theta+\pi}{2\pi}\right\rfloor\right) \tag{5.62}$$

$$I_{nn}^{cav,b} = \frac{-1}{4\pi G}\left\{\left(\frac{\kappa+1}{2}\right)\left(\sin\theta\ \ln(2-2\cos\theta)-\theta\right)\right.$$

$$\left.-(\kappa-1)\left(\sin\theta-\cos\theta\,F(\tilde{\theta})-\pi\left\lfloor\frac{\theta}{2\pi}\right\rfloor\right)-\sin\theta\right\} \tag{5.63}$$

$$I_{sn}^{cav,a} = K_1\left(\cos\theta\ \ln\left|\tan\left(\frac{\theta}{2}\right)\right|-\ln|\sin\theta|\right)+K_2\sin\theta\,\mathrm{sgn}(\tilde{\theta}) \tag{5.64}$$

$$I_{sn}^{cav,b} = \frac{-1}{4\pi G}\left\{\left(\frac{\kappa+1}{2}\right)\left[\cos\theta\,(1-\ln(2-2\cos\theta))+\ln(1-\cos\theta)\right]\right.$$

$$\left.+(\kappa-1)\left(\frac{\cos\theta}{2}+\sin\theta\,F(\tilde{\theta})\right)\right\} \tag{5.65}$$

The previous primitive functions include terms with singularity points when θ tends to 0. Finite limits are obtained in all cases, and these are given in Eqs. (5.66)–(5.69):

$$\lim_{\theta\to 0}\left\{\cos(\theta)\,\left[1-\ln(2-2\cos\theta)\right]+\ln(1-\cos\theta)\right\}=1-\ln(2) \tag{5.66}$$

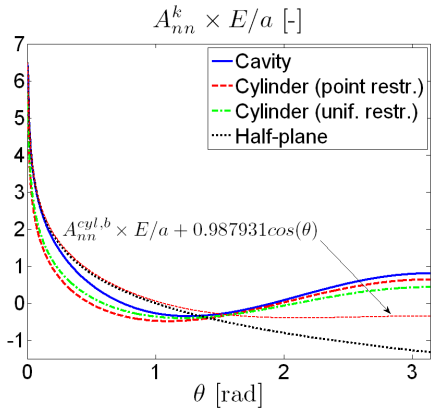
$$\lim_{\theta\to 0}\left\{\sin\theta\ \ln(2-2\cos\theta)\right\}=0 \tag{5.67}$$

$$\lim_{\theta\to 0}\left\{\cos\theta\ \ln\left|\tan\left(\frac{\theta}{2}\right)\right|-\ln\left|\sin\theta\right|\right\}=-\ln(2) \tag{5.68}$$

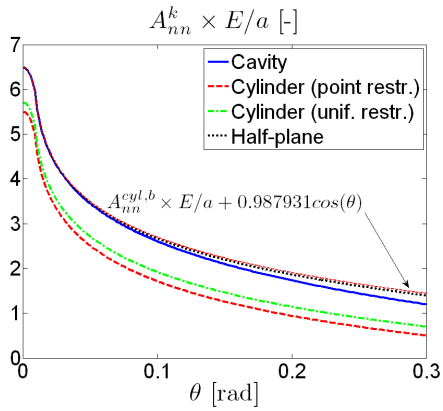
$$\lim_{\theta\to 0}\left\{\sin\theta\ \ln\left|\tan\left(\frac{\theta}{2}\right)\right|\right\}=0 \tag{5.69}$$

Next, some comparisons are shown of the ICs associated to uniformly distributed line loads. Figure 5.12 compares the A^{k}_{nn} ICs (with k equal to *cyl*, *cav* or *hp*) of the cylinder and cylindrical cavity with those of the half-plane, and in Figure 5.13 similar comparisons are made for the A^{k}_{sn} ICs. The ICs are presented in non-dimensional form, multiplied by the Young's modulus E and divided by the half-width $a = R\times \Delta\theta/2$ of the loaded band. The total angle of the loaded band $\Delta\theta$ is fixed at 20 mrad. A value of the coefficient of Poisson of 0.30 is considered. For the case of the half-plane, the angular distances and coordinates are converted to their linear counterparts multiplying them by the considered unit radius R.

In Figure 5.12 $x_0 = R$ is taken as the origin point for the normal displacements of the half-plane, so that the normal displacement in the central position of the loaded band ($\theta = 0$) of the half-plane coincides with that of the cavity. On the other hand, an additional curve is included for the case of the cylinder with a point restriction in its centre, adding the necessary rigid body displacement for the normal displacement in $\theta = 0$ to coincide with that of the cavity and of the half-plane. In this way, a more direct comparison between the different curves may be made. The added rigid body displacement in the additional curve for the cylinder with punctual centre restriction is indicated in the equation included in the figure.

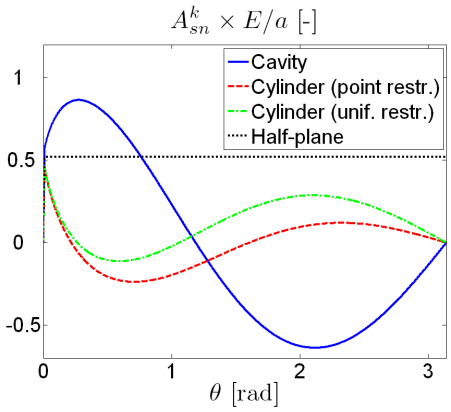


(a) Variation in the whole circumference (A^k_{nn} symmetrical in θ = 0, and in the cases of the cylinder and cavity also in $\theta = \pi$ rad).

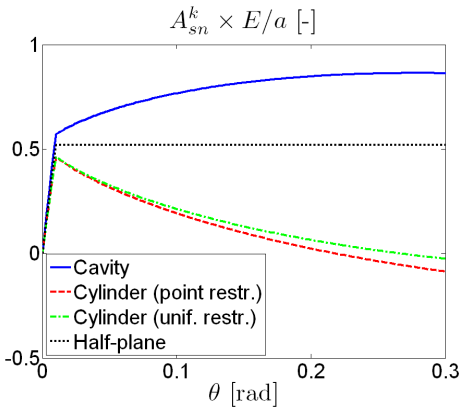


(b) Detail from θ = 0 to 0.3 rad.

Figure 5.12. A^k_{nn} ICs of cylindrical geometries and of the half-plane, with a single uniformly distributed line load in a total angular range $\Delta\theta$ of 20 mrad, and with ν = 0.30.



(a) Variation in the whole circumference (A^k_{sn} antisymmetric in θ = 0, and in the cases of the cylinder and cavity also in $\theta = \pi$ rad).



(b) Detail from θ = 0 to 0.3 rad.

Figure 5.13. A^k_{sn} ICs of cylindrical geometries and of the half-plane, with a single uniformly distributed line load in a total angular range $\Delta\theta$ of 20 mrad, and with ν = 0.30.

As may be seen in Figure 5.12b, for small angular distances between the point of load application and the point where the displacements are observed, the trends of the A^k_{nn} of cylindrical geometries are similar to the case of the half-

plane. The A_{nn} ICs of the cylinder exhibit a slightly lower decrease than those of the half-plane around the loaded zone, while the decrease for those of the cavity is somewhat higher. The latter are the ones that present bigger differences with respect to the A_{nn} ICs of the half-plane.

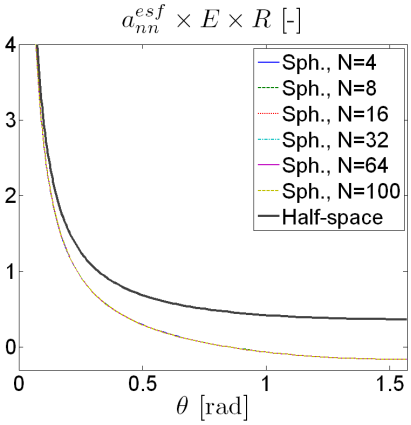
However in the case of the the A_{sn} ICs, the differences between the cylindrical geometries and the half-plane are much more pronounced, as may be seen in Figure 5.13. Already for θ values of 25 mrad, differences above 20% are reached between the A^k_{sn} ICs of the cylindrical geometries and those of the half-plane. Near the loaded zone, the A^k_{sn} of the cavity are bigger than those of the half-plane, and the ones of the cylinder are smaller. Additionally, considering a conformal contact problem between a cylinder and a cylindrical cavity, these differences are added up when calculating the A_{sn} ICs associated to the displacement differences between both bodies. In contrast, in the case of the A_{nn} ICs, the differences between the A^k_{nn} of the cylinder and the cavity with respect to the half-plane tend to cancel each other. On the other hand, comparing the values of Figure 5.12 with those of Figure 5.13, it is seen that the magnitude of the A^k_{sn} ICs is considerably smaller than that of the A^k_{nn} ICs, in line with what was seen in §5.1.3 for the half-space.

According to the above observations, it may be anticipated that with not too high conformity levels, the use of the A_{nn} ICs of the half-plane (in 2D) or of the half-space (in 3D) may be a good approximation for the study of frictionless conformal contact problems in which only the normal part of the contact is of interest (where only the A^k_{nn} ICs take part). On the contrary, in the tangential part of the contact problem, the use of the solutions corresponding to the half-plane or to the half-space for ICs such as the A_{sn} may lead to appreciable errors already with relatively low conformity levels, as is seen in §6.2.2.2.

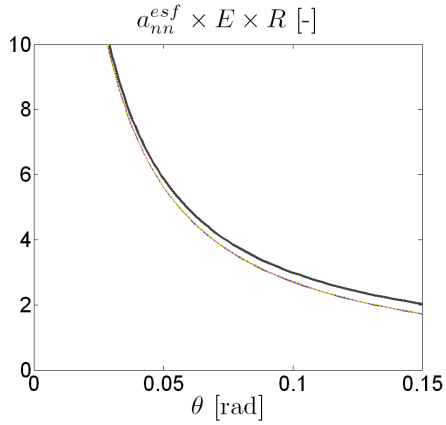
5.2.1.2 Sphere under opposing radial loads

Figure 5.14 represents the a^k_{nn} IFs, and Figure 5.15 the a^k_{sn} IFs, of the sphere and of the half-space. Equations for them have been given in §2.2.2.2. The IFs are represented in non-dimensional form, multiplied by the Young's modulus E and by the radius of the sphere R. For the case of the half-space, two unitary point loads are considered acting in normal direction in points of the surface separated by a distance of πR. In this case, the displacements produced in the

line joining those two points are represented, dividing by the unitary value of the considered radius R to convert from linear to angular distances in the abscissa of the graphs.

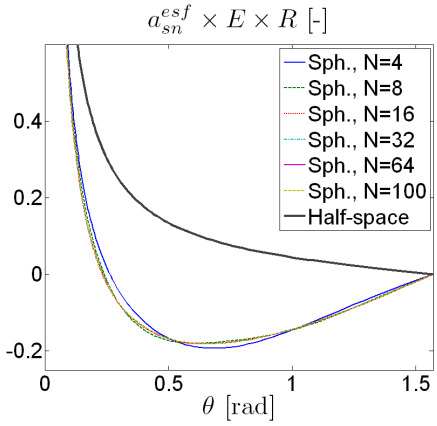


(a) Variation from θ = 0 to π/2 rad (a^{k}_{nn} symmetric with respect to θ = π/2 rad).

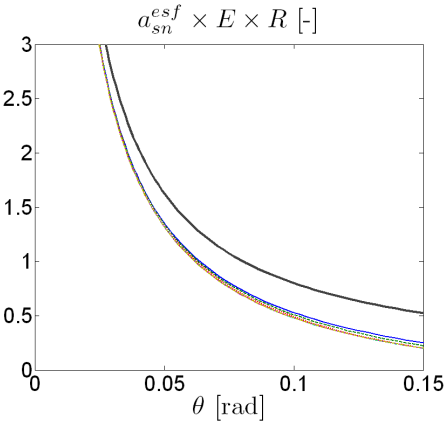


(b) Detail from θ = 0 to 0.15 rad.

Figure 5.14. a^{esf}_{nn} IF of the sphere, with two diametrically opposed normal point forces, together with the equivalent IF for the half-space. v = 0.30.



(a) Variation from θ = 0 to π/2 rad (a^{k}_{sn} antisymmetric with respect to θ = π/2 rad).



(b) Detail from θ = 0 to 0.15 rad.

Figure 5.15. a^{esf}_{sn} IF of the sphere, with two diametrically opposed normal point forces, together with the equivalent IF for the half-space. v = 0.30.

The previous figures show several curves of the sphere IFs, calculated with different values of the maximum degree N of the Legendre polynomials at

which the infinite summations that appear in the final part of Eqs. (2.138) and (2.139) are truncated, as indicated in the figure legends, up to a maximum N value of 100. It may be seen that the a^{esf}_{nn} IF converges quickly with N, having already with N=4 nearly undistinguishable differences with respect to N = 100, while for the a^{esf}_{sn} IF the convergence with N is somewhat slower (in Figure 5.15 the curves for N = 4 and N = 8 may be distinguished, but the curves for N = 16 and higher are superposed with that for N = 100).

As it happens with the IFs of cylindrical geometries, in the case of the sphere also the differences with respect to the half-space are much more pronounced in the a^{esf}_{sn} IF than in the a^{esf}_{nn}, and already from relatively reduced θ values.

5.2.2 Numerical calculation with FEM

In the case of general geometries for which no analytical solutions for the ICs are available, their calculation may be carried out by means of some numerical method, like the Finite Element Method. This section details the procedure followed in this thesis to calculate numerically with FEM the IC matrices necessary as input for the exact contact theory. The numerically calculated ICs will be designated here with the C letter.

5.2.2.1 Arrangement of FE models

Next, some practical questions are treated about the construction of the necessary FE models for the numerical calculation of the ICs. Li [Li Z 2002] presented an analysis of the influence of different aspects of the FE models in the obtained numerical values of the ICs, as the necessary dimensions of the meshed domain to avoid alteration of the results by the proximity of the limits of the mesh, the boundary conditions in the limit surfaces of the mesh, the order of the elements (linear or quadratic) and the size of the elements in the loaded zone. Some important aspects to be taken into account in the computation of the ICs were pointed out as well, as the need to remove the global displacements from the FE solution, and the need to correct the FE solution in the vicinity of the loaded element. The possibility of truncating the influence region at each point in the contact patch discretization to reduce the computational cost of the subsequent contact analysis with the exact contact theory, i.e., taking into account at each element only the influence from the nearest elements around it,

was also briefly discussed. The need for improvement of the FE models used for the calculation of the ICs was set out as a possible topic for further research, this need arising from computer hardware limitations at the time that work was carried out.

In [Vollebregt 2014a] and in [Zhao J 2016], the numerical calculation of ICs for non-planar solids with FE models was further investigated. The effect on the precision and computation cost of the ICs of factors as the size of the model with respect to the loaded zone, the type of solid element (hexahedron or wedge) and its order (linear or quadratic), the degree of local refinement around the loaded element, and the rate of element size increase out of the loaded zone was studied in detail. In [Vollebregt 2014a], the propagation of the errors in the calculated ICs to the final numerical results of contact problems was assessed, as well as the effect of truncating the influence region in the contact patch. Additionally, individual ICs for both convex and concave contacting bodies were shown, and the consequences of their combined differences for the contact problem were pointed out. With this, it was recognized that in conformal contact, there is some coupling between the normal and tangential parts of the contact problem even with elastically similar materials, in addition to the coupling at the global equilibrium level already shown in [Li Z 2002].

Here some comparisons will be presented between the C_{nn} ICs obtained numerically with FEM with different types of elements and forms of application of the unit load, for a prismatic body with a non-planar cross-section as the one shown in Figure 5.16. It is a convex section, symmetric with respect to the vertical symmetry axis shown in the figure, with a central circular part flanked by two straight parts. Table 5.3 lists some of the relevant data of the different FE models used in these calculations, including the values of the dimensions represented in the figure. The second to fourth columns of the table include the following data:

- 2: element type used in the mesh.
- 3: Number of elements NE_C in which the rectangular zone loaded with a uniform pressure is discretized in each principal direction of the surface (longitudinal×lateral), together with the number of nodes NN_C in which the load is distributed

- 4: Longitudinal (a) and lateral (b) half-sides of the rectangular surface of the element nearest to the loaded zone.

The FE models have been constructed with the commercial FE software Abaqus [Abaqus doc 2012]. In all the cases the total longitudinal dimension of the models is 600 mm (but in all the cases except for the last one listed in Table 5.3, use is made of symmetries in the longitudinal and lateral direction, and the longitudinal dimension of the meshed domain is 300 mm), big enough in comparison with the dimensions of the zone of interest where the variations of the ICs are studied, so as to avoid the results being altered by the boundaries of the model.

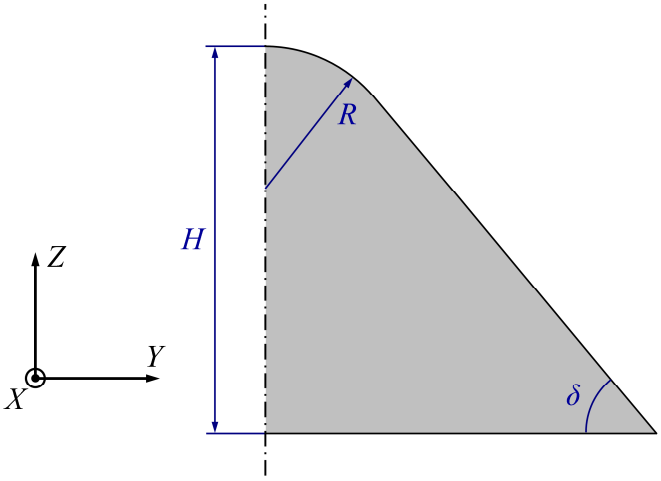


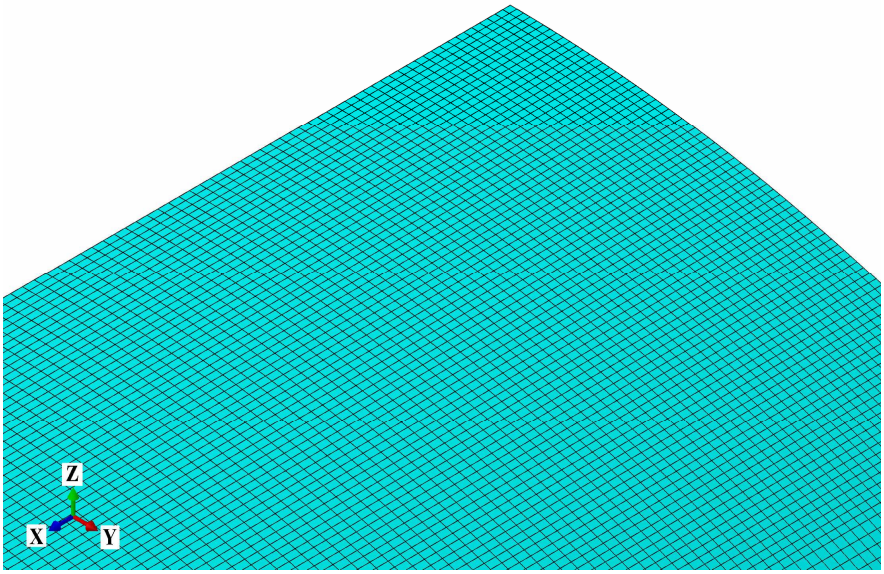
Figure 5.16. Cross-section of the considered prismatic body to compare the ICs obtained numerically with different FE models.

In the comparisons shown below, only cases in which the loaded zone is centred in the convex surface of the model are considered, symmetrically located with respect to the longitudinal and lateral symmetry planes of the model. As an exception, in test no. 6 the loaded zone is located with two of its sides contained in the symmetry planes, instead of centred with respect to these; but this detail is not relevant in the results shown. The load is applied on a single node in tests nos. 1 and 3, and by means of a uniform stress distributed on a rectangular zone of the surface (which is translated to point loads applied on the nodes of the free faces of the loaded elements), either on 4 elements (2 in each direction) in tests 2, 4 and 5, or on a single element in test 6.

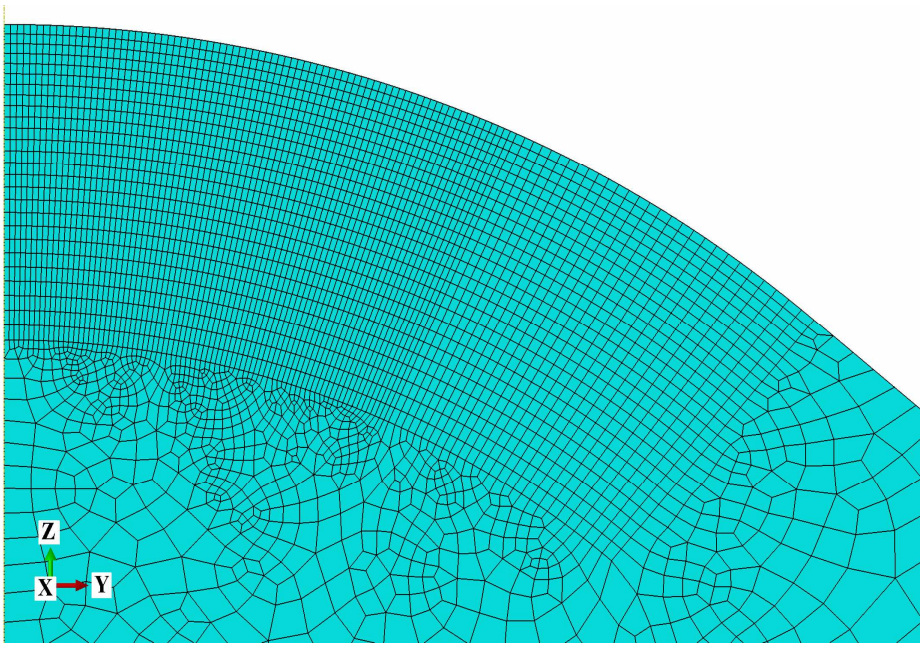
Table 5.3. Parameters of the different FE models used to calculate the ICs of the prismatic body of Figure 5.16.

Test no.	Elm. type	NE_C / NN_C [-]	$a \times b$ [mm]	R [mm]	δ [º]	H [mm]
1	C3D8R	- / 1	0.060×0.052	10	40	100
2	C3D8R	2×2 / 9	0.060×0.052		40	100
3	C3D8	- / 1	0.060×0.052		40	100
4	C3D8	2×2 / 9	0.060×0.052		40	100
5	C3D20	2×2 / 21	0.120×0.105		40	100
6	C3D8	1 / 4	0.167×0.158		65	150

In all the cases, the meshes have been built with solid hexahedral elements, being the surface divided in faces of rectangular elements, as the solid meshes are made by extrusion in the longitudinal direction of the mesh of the cross-section. Additionally, in the zones of interest of the surface where the displacements are observed, and up to a certain depth, structured meshes are used, with nearly parallelepiped hexahedra. The topology of these meshes is the same as that of the FE models constructed for detailed contact analyses described in §4.2.1. Figure 5.17 shows two views of the mesh near the loaded zone of one of the constructed FE models. The elements of the loaded zone and the ones adjacent to them are similar to the rest of the nearby mesh; there is not a local refinement in the loaded domain with respect to the adjacent meshing.



(a) Isometric view of the convex surface.



(b) Cross-section.

Figure 5.17. Details of one of the meshes used in the calculation of the ICs of the prismatic body of Figure 5.16.

From the variables related to the mesh listed in Table 5.3, the ones which have an appreciable influence on the quality of the calculated ICs are the type of element on the one hand, and the number of elements or nodes in which the loaded zone is discretized on the other. The dimensions of the elements of the mesh around the loaded zone, being in all the cases much lower than the characteristic dimensions of the solid (including the curvature radius of the circular zone in the central part of the cross-section), have nearly no influence in the precision of the results. Regarding the parameters related to the geometry of the solid listed in the table, once fixed the radius R, neither the angle δ nor the total section height H have an appreciable influence in the variations[20] of the C_{nn}

[20] The mentioned geometric features do influence in the total computed displacements. But the main ICs obtained in test no. 6 of Table 5.3, which geometric parameters are different from those of the rest of the tests listed in the table, differ from those obtained in the rest of tests in a global displacement term in each case, which has nearly no

in the zone of interest that is considered here, of similar dimension to the radius R.

As may be seen in Table 5.3, in these comparisons three solid hexahedra element types have been considered. The C3D8R is a linear element with reduced integration, the C3D8 is a linear element with full integration[21], and the C3D20 is of quadratic order and full integration.

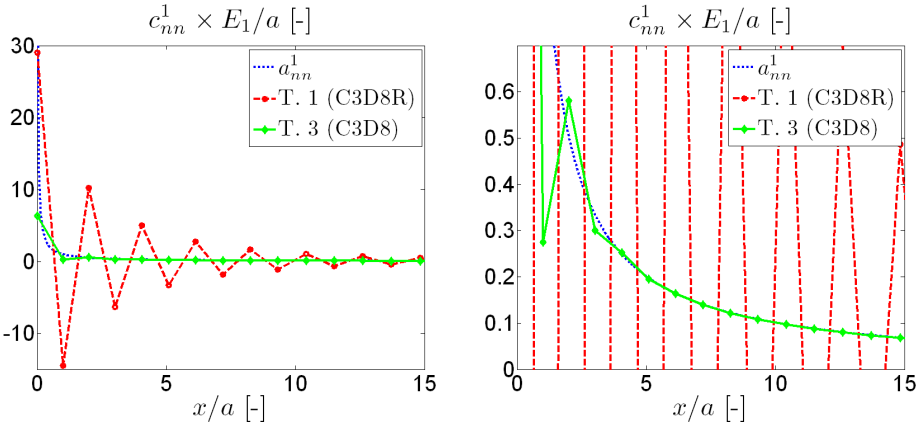
The next figures compare the C_{nn} obtained in the different tests listed in Table 5.3, along the vertical longitudinal plane of symmetry. The values obtained in the nodes of the mesh used in each test are marked with markers in each curve. Each figure includes as well the curves for the half-space ICs. For $x/R << 1$, the variations along the longitudinal direction of the C_{nn} of the considered prismatic solid with non-planar surface are similar to those of the half-space.

The results are presented in non-dimensional axes: the longitudinal coordinate, with origin in the centre of the loaded domain, is divided by the longitudinal half-side of the loaded domain a; and the elastic displacements are expressed as ICs (with units of displacement per unit pressure) multiplied by the Young's modulus E_1 and divided by a. This normalization is adequate near the loaded domain, and for values of $x/R << 1$, to keep the geometric similarity with respect to the loaded zone. Far from the loaded zone, and for x/R values of order unity or higher, it would be more adequate to keep the geometric similarity with respect to the cross-section or global geometry of the solid; while the

variation in the studied zone of interest, and which may therefore be considered in this context as a rigid body displacement. Considering as relevant in contact mechanics the variations or gradients of the surface displacements around the contact, rather than the total displacements, the C_{nn} obtained in test no. 6 are comparable to the ones obtained in the other tests (leaving aside the fact that the sides ratio a/b is slightly different from that of the rest of tests), compensating with the appropriate rigid body displacement term.

[21] Actually, these elements use selectively reduced integration. Reduced integration is used on the volumetric terms, replacing the real volume changes at the Gauss points of the element by the average volume change of the element (cf. §3.2.4-7 of the Abaqus Theory Manual [Abaqus doc 2012]).

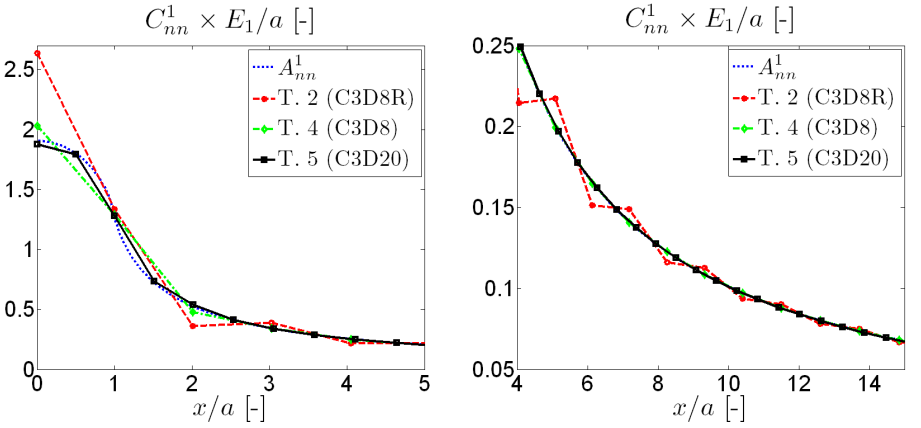
dimensions of the loaded zone are no longer relevant, neither the related geometric similarity, according to Saint-Venant's principle. In that case, the x coordinate could be made non-dimensional dividing it by the curvature radius of the cross-section R; and the elastic displacements, by division by the total load applied and multiplication by $E_1 R$.



(a) General view for x/a in the (0, 15) interval.

(b) Detail of results of test 3 in the same interval.

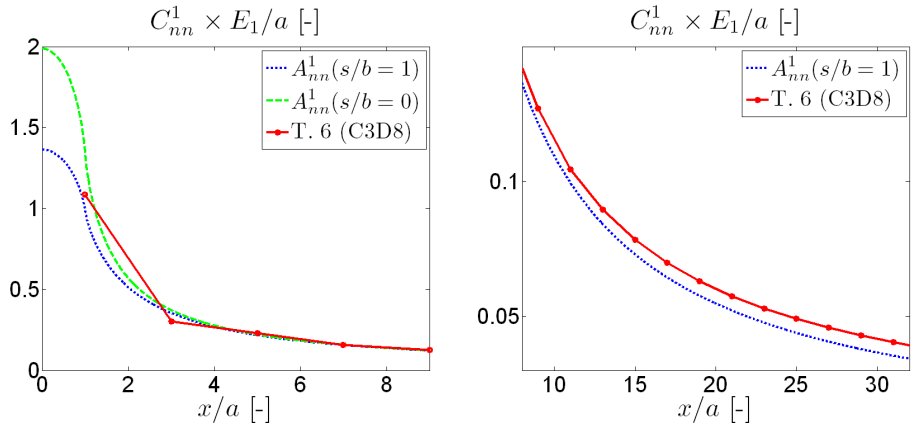
Figure 5.18. c_{nn} IFs of tests 1 and 3 of Table 5.3, with point load, together with the a_{nn}.



(a) Nearest part to loaded zone, for x/a in the (0, 5) interval.

(b) Detail for x/a in the (4, 15) interval (the A_{nn} curve is hidden behind the curves of tests 4 and 5).

Figure 5.19. C_{nn} ICs of tests 2, 4 and 5 of Table 5.3, with load distributed in four elements, together with the corresponding A_{nn}.



(a) Nearest part to loaded zone, for *x*/*a* in the (0, 9) interval.

(b) Detail for *x*/*a* in the (9, 32) interval.

Figure 5.20. C_{nn} ICs of test 6 of Table 5.3, with load distributed in a single element, together with the corresponding A_{nn} at lateral positions *s*/*b* = 1 (where the C_{nn} are calculated), and *s*/*b* = 0.

Figure 5.18 shows the c_{nn} IFs of tests 1 and 3, in which the load is applied in a single node. In this figure, the half-space a_{nn} curve represents the product $4 \times a^1_{nn} \times b \times E_1$. The results obtained with the C3D8R elements are very irregular, presenting big oscillations that propagate strongly in the mesh in a considerable distance from the loaded node, which render them almost useless. This is a consequence of a numerical instability phenomenon termed hourglassing, which is related to the zero energy deformation modes of the element. This is a pathology of reduced integration elements, mostly of those of linear order; and as mentioned in the Abaqus Theory Manual [Abaqus doc 2012], is much more likely to originate with concentrated than with distributed loads. The results obtained with the C3D8 elements exhibit an appreciable oscillation in the nodes adjacent to the loaded one, as may be appreciated in Figure 5.18b, but two or three nodes beyond it the trend of the a_{nn} is correctly reproduced.

Figure 5.19 shows the C_{nn} obtained in tests 2, 4 and 5, where the applied load is distributed in 4 elements. In this case, the results obtained with the C3D8R elements (test 2) are much more regular than in test 1, although they still present a certain level of numeric oscillation, which gradually diminishes with the distance to the loaded domain. In spite of the observed bad performance of the C3D8R for the calculation of the ICs (mostly in test 1), it has to be mentioned

that good results have been obtained with these elements in contact problems in which the contact patch encompasses a considerable number of elements and the contact stresses vary smoothly inside it. These results have been validated with available analytical solutions and with results obtained with C3D8 elements. With respect to the C3D8 elements, the results of test 4 in the zone nearest to the loaded domain also improve with respect to the results obtained with the same type of element in test 3, as can be seen comparing Figure 5.19a with Figure 5.18b. The best results are obtained with quadratic elements, as was to be expected: in test 5, there is a difference of just about 1.5% in the C_{nn} obtained in the centre of the loaded zone, with respect to the theoretical A_{nn} value.

The C_{nn} obtained in test 6 are graphed in Figure 5.20. The precision achieved in the nearest node to the centre of the loaded zone, at $x/a = s/b = 1$, is not much worse than that obtained in test 4 at the centre of the loaded zone, in spite that the load changes from being distributed in 9 nodes in test 4, to just 4 in test 6. The deviation with respect to the theoretical value for the half-space is of 8.8% in test 6 at $x/a = s/b = 1$, and of 6.5% in test 4 at $x/a = s/b = 0$. As in test 4, the precision improves quickly with the distance to the loaded zone. Additionally, the errors in the nodes nearest to the centre of the loaded zone have alternating signs (being the C_{nn} obtained in the nearest zone bigger than the corresponding A_{nn}, and smaller in the adjacent node), so that it may be thought that they will tend to compensate each other.

On the other hand, it has to be considered that the APs where the ICs are necessary in a surface mesh in the exact contact theory based on elements with uniform load distribution, are located at even values of x/a and s/b, whereas in test 6 the ICs are obtained at odd values of x/a and s/b. Therefore, it is necessary to deduce somehow the ICs at the even values of the non-dimensional coordinates, from the values originally obtained at the odd coordinates. This may be done by means of interpolations, which is equivalent to averaging in each element the results of its 4 nodes, as the positions where the results are sought are the centres of the surface faces of each element, and these faces are rectangular. But in the loaded element, the interpolation of its nodal values (equal the 4 of them, by symmetry), amounts to translating directly those values to the centre. In this way, the error obtained in the C_{nn} of the loaded element

increases up to 46% (error of the value at $x/a = 1$ of the curve of test 6 in Figure 5.20a, with respect to the A_{nn} at $x/a = s/b = 0$), being the interpolated value just above half the correct value. The ICs obtained with this interpolation are not valid for use in the solution of contact problems with the exact contact theory, but they are valid after applying a correction on them, as described later.

Far from the loaded zone, the curve obtained in test 6 follows the same trend as the curve of the half-space, but it is above it, as may be seen in Figure 5.20b. However, the difference between both curves is nearly constant in a wide range of distances from the loaded zone, so it does not have consequences in the results which may be obtained with these ICs in a contact problem (cf. footnote no. 20).

5.2.2.2 Computation of numerical IC matrices

Next, it is explained the procedure followed to obtain the IC matrices for solids with non-planar surfaces with FE models like those described in §5.2.2.1. These IC matrices will be used as input in the exact contact theory to solve conformal contact problems. Specifically the relevant parameters of the mesh of the FE model which has been used to produce a complete set of IC matrices are those of test 6 of Table 5.3. Therefore, models with C3D8 elements have been used, and the loaded zone is represented with a single element. These models use directly the same meshing as the one successfully used to obtain detailed contact solutions with FEM contact analyses.

As explained in §4.1.1.4, in general it will be necessary to load each one of the n_s lateral positions in which the PCS is discretized, to get the necessary data for the $\mathbf{[C_{ij}]}$ matrices of solids with non-planar cross-section and regular geometry in longitudinal direction.

In the case of prismatic solids, the 3D mesh is built extruding the cross-section mesh in longitudinal direction. In the case of solids of revolution, if the variation of the local radius of revolution in the zone of interest of the surface is much smaller than the mean radius of revolution in that zone, as is typically the case of the wheel in the railway application, the simplest way is to do something similar to the case of the prismatic solid, sweeping in this case the mesh around the axis of revolution. In both cases, the nodes of the surface are distributed in a rectangular pattern (or nearly rectangular in the case of the solid of revolution),

as may be seen in Figure 5.17a, being located at the corners of the elements. The dimensions of the sides of the surface rectangular faces of the elements will not be necessarily equal in the whole mesh, neither in longitudinal nor in lateral direction, and the FE meshes of each body in general will not be equal. It would not be necessary to have the surface nodes arranged in a rectangular pattern either, but this facilitates the interpolations that have to be done later.

After finishing the calculations with the FE model of the two solids that take part in the contact, the nodal displacements in the complete mesh of each solid for each computed load case will be available, in the three space directions. For the calculation of the IC matrices, only the surface displacements are of interest, in a region enclosing the potential contact surfaces of the contact problems foreseen to be solved with these matrices. This region will be rectangular, as the PCS. In the lateral direction it will be delimited by the set of lateral positions of the mesh (adjacent or not) that have been loaded in the computed load cases. It will not be possible to obtain complete IC matrices in other lateral positions. In the longitudinal direction, a region as broad as the mesh of the model allows may be taken, even though later the PCS and corresponding IC matrices are limited to a smaller zone. Taking into account the symmetry in the longitudinal direction, it will be necessary to retrieve the displacements in one of the sides of the loaded zone, being possible to deduce the ones of the other side by symmetry or antisymmetry relations. With the displacements of a region located at a distance between 0 and L_x in the longitudinal direction with respect to the centre of the loaded zone, it will be possible to build IC matrices for PCS-s up to a total longitudinal dimension of L_x (remind the set of possible loaded elements for a mesh of a laterally curved PCS marked in red in Figure 4.15b).

It must be mentioned that the displacements obtained from a FE analysis include the structural or global displacements of the meshed part of each solid. To produce valid IC matrices for a contact model to be incorporated in an analysis which takes into account the structural flexibility of the bodies out of the contact itself (for example, including its natural modes in a multibody model, as explained in Chapter 3), it would be necessary to subtract the structural displacements (already taken into account outside the contact model) from the total ones that are obtained in the FE models, to avoid accounting for them twice in the analysis, as noted in [Li Z 2002] and [Vollebregt 2014a].

However, for an independent local contact analysis, it is not necessary to make this separation, and the total displacements obtained with the FE models may be retained (as in contact mechanics the relevant quantities are the variations of the surface displacements around the contact, rather than the total displacements, as mentioned in the footnote no. 20).

Once selected for each solid the set of nodes of which the displacements will be taken, each node is indexed according to its position in the longitudinal and lateral directions of the set. For each row of nodes, according to its lateral position in the mesh, a $\{x, s, n\}$ local curvilinear coordinate system is defined, taking into account the coordinates in the transversal plane of the nodes of the adjacent elements. The displacements obtained normally in a common global system of the FE model will be transformed to this local curvilinear system.

From this point on, a number of search operations, transformations between coordinate systems, interpolations, scaling operations, superpositions and corrections are carried out, with the aim of translating the displacements available at the nodes of the FE meshes to the APs of the regular objective mesh of the PCS (the objective mesh from now on; composed of equal rectangular elements) and of obtaining finally the combined ICs of both solids in that mesh. As the FEM results are obtained with single first order loaded elements with uniform load, the elements of the objective mesh are also taken as uniform, with the APs located at the element centres.

The next steps are repeated for each load case to be computed in the objective mesh, for the two solids that participate in the contact:

1. The two loaded elements of the FE mesh which centres are nearest at each side in the lateral direction from the centre of the loaded element in the objective mesh are searched. Figure 5.21 illustrates this search, for a given loaded element of the objective mesh. In the figure also the longitudinal and lateral dimensions of the elements of the objective mesh are indicated, Δx and Δs.

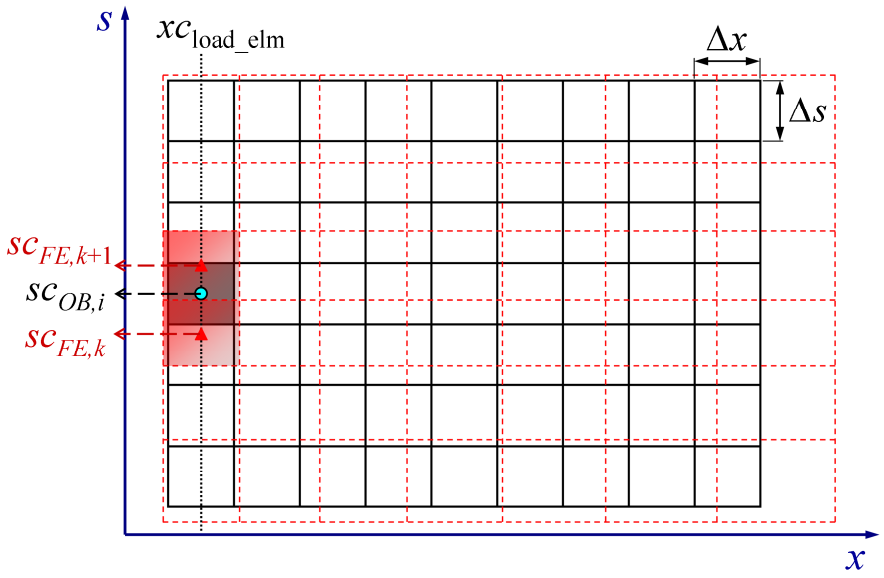


Figure 5.21. Possible FE mesh (in dashed red lines) and objective mesh (in solid black lines). A loaded element of the objective mesh is marked in grey, and the two nearest loaded elements of the FE mesh at each side in red with transparency. The longitudinal position of the centres of the loaded elements $xc_{\text{load\_elm}}$ is marked as well, coincident in both meshes, and the lateral positions of each of the marked elements.

For the two corresponding load cases of the FE model, the displacements are transformed to the local curvilinear coordinate system of each node of the FE mesh, which defines its principal directions $\{x, s, n\}$.

2. With the displacements of the previous step, calculated at the nodes of the FE mesh of each body, the necessary interpolations are carried out to obtain the displacements at the APs of the objective mesh, in two steps. Firstly, the displacements of the FE mesh are translated by means of two-dimensional interpolation to a grid of points with the same arrangement as the APs of the objective mesh, but laterally shifted according to the vector going from the centre of the loaded element of the objective mesh to the centre of the loaded element of the FE mesh. This in turn is carried out in two steps: firstly, the FE nodal displacements are translated to the centres of the elements of the FE mesh, averaging the four nodal displacements of each element.

Secondly, the displacements at the APs of the objective mesh are interpolated from the displacements at the centres of the elements of the FE mesh. Figure 5.22 illustrates this interpolation, which can be carried out easily thanks to the rectangular arrangement of the nodes of the FE mesh. In this case, it has been opted for bilinear interpolations. The displacements at the APs of the objective mesh could also be directly interpolated from the nodal FE displacements.

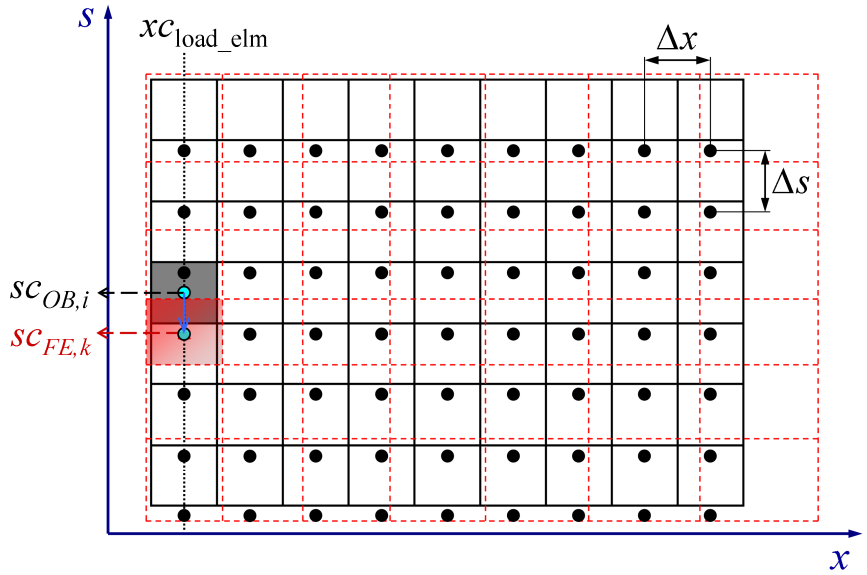


Figure 5.22. 2D interpolation to obtain the displacements in the pattern of APs of the objective mesh, from the displacements calculated at the nodes of the FE mesh.

The interpolation is carried out in relative coordinates with respect to the centre of the loaded element. On the other hand, the displacements are scaled with the area ratio between the elements of the objective mesh and the current loaded element of the FE mesh. In other words, the displacements are scaled in proportion to the applied forces in the element in each mesh, as the ICs are calculated per unit stress, and the applied force is then proportional to the area of the loaded element. In this way, the possibility of having elements with different sizes and proportions in each mesh is allowed for. However, to obtain good results, the loaded elements of both meshes should be of not too

different sizes and proportions, as obviously this simple scaling will not reproduce correctly the displacement field near the loaded element in the objective mesh, when this is of a different size from the loaded element with which the displacements were originally calculated. A more sophisticated scaling could be devised, taking into account the changes of the half-space ICs with each element for example. But this would increase the complexity of this step, as it would imply calculating a scaling factor for each AP of the objective mesh, and for each IC matrix. Additionally, it would be necessary to separate adequately the scaling of the global displacement component (which would scale in direct proportion to the total applied force), and that of the variable displacement at the contact zone scale.

The 2D interpolation described above is done for the displacements of the two load cases of the FE model identified in the previous step, each time applying a different lateral shift to the pattern of APs of the objective mesh as explained above.

Secondly, a linear interpolation is carried out in relation to the lateral position of the loaded element. This second interpolation, one-dimensional, is done between the two sets of interpolated displacements obtained above, corresponding to the two loaded elements of the FE mesh adjacent in the lateral direction to the current loaded element of the objective mesh, to obtain the displacements in the actual positions of the APs of the objective mesh, associated to said loaded element of the objective mesh. This interpolation is expressed in Eq. (5.70), designating with superindices FE,k and $FE,k+1$ the two sets of interpolated displacements obtained above, and with superindex OB,i the set of displacements resulting from this last one-dimensional interpolation. The different lateral coordinates s that appear in the equation are illustrated in Figure 5.21.

$$\mathbf{u}^{OB,i}\left(x_r, s_r\right) = \frac{\left(sc_{FE,k+1} - sc_{OB,i}\right)\mathbf{u}^{FE,k}\left(x_r, s_r\right)}{sc_{FE,k+1} - sc_{FE,k}}$$

$$+ \frac{\left(sc_{OB,i} - sc_{FE,k}\right)\mathbf{u}^{FE,k+1}\left(x_r, s_r\right)}{sc_{FE,k+1} - sc_{FE,k}} \tag{5.70}$$

This interpolation must be done as well in relative coordinates (x_r, s_r) with respect to the centre of the loaded element corresponding to each set of displacements as indicated in Eq. (5.70), and not in absolute coordinates. In Figure 5.23 it may be easily visualized the error incurred when doing the interpolations of the displacements from the FE mesh to the objective mesh in absolute coordinates, instead of in relative coordinates with respect to the centre of the loaded element. The figure shows the displacement fields $\mathbf{u}^{FE,k}$ and $\mathbf{u}^{FE,k+1}$ obtained when loading separately the elements of the FE mesh with centres located at $sc_{FE,k}$ and $sc_{FE,k+1}$, respectively. Typical trends of some of the direct ICs near the loaded element are represented, which reach their maximum in the centre of the loaded element and decrease rapidly in magnitude outside it. To simplify, it is assumed in this explanation that the two loaded elements of the FE mesh as well as the element of the objective mesh are of the same size and shape, that the same displacement curves are obtained loading those two elements of the FE mesh, and that the centre of the loaded element of the objective mesh is located midway between the two elements of the FE mesh. Considering for instance the displacement in the centre of the loaded element of the objective mesh itself, it may be seen that if it is calculated interpolating with the values of the FE mesh corresponding to its same absolute position, the obtained value will be considerably lower than the two obtained in the centres of the loaded elements of the FE mesh. However, the correct result is an intermediate value between these two.

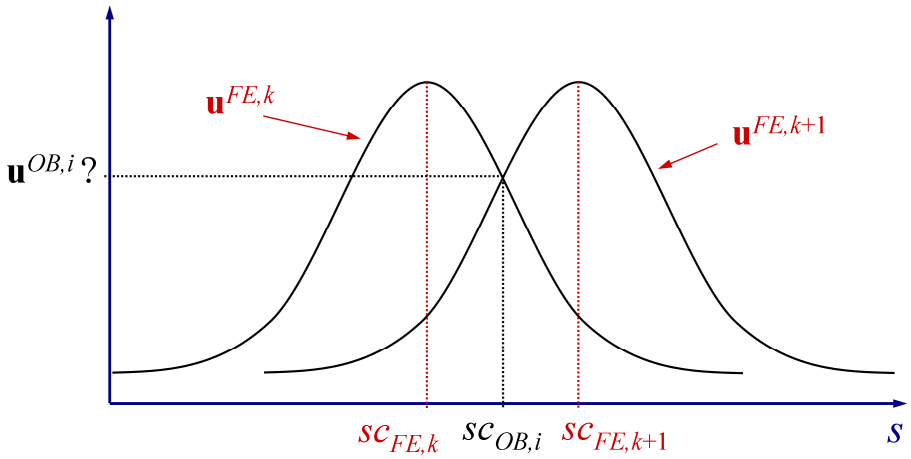


Figure 5.23. Incorrect interpolation of the two sets of displacements obtained for the two elements of the FE mesh adjacent to the loaded element of the objective mesh, in absolute instead of relative coordinates with respect to the centre of the loaded element.

3. The ICs are obtained superposing, with positive or negative sign depending on the case, the displacements obtained in the previous step for the two bodies in contact, divided by the applied pressure in each case. That is, the surface displacement differences between the two bodies are calculated with Eq. (1.8), taking into account that the stress acts in opposite direction in each body (Newton's third law) to define correctly the sign to be used in each case. The obtained results are tabulated as a function of the absolute coordinates of the objective mesh.

4. As previously discussed, when using a single first order element to represent the loaded zone, a very high error is made in estimating the displacement at the central point of the loaded zone by direct averaging of the available displacements at the nodes, located at the four corners of the loaded zone.

   In [Zhao J 2016] careful attention is paid to the precision of the calculated displacements in the loaded zone, and that zone is locally refined with the aim of achieving a good precision there. In this way, on the one hand there is the discretization of the PCS, formed by equal rectangular elements, each of which is equivalent to a loaded zone. On

the other hand, there is the discretization of the FE models used for the calculation of the ICs, in which what could be called a sub-meshing of the loaded zone is done, subdividing it in more elements in turn. After performing a sensitivity analysis with different FE meshes, some recommendations related to different meshing parameters are given, aimed at getting an error below 1% in the computed displacement at the centre of the loaded zone. Regarding the local refinement of the loaded zone, it is recommended to subdivide it in at least 4 elements in each direction, and to extend the region of this local refinement at least a 25% beyond the loaded zone. Additionally, it is recommended to use quadratic elements.

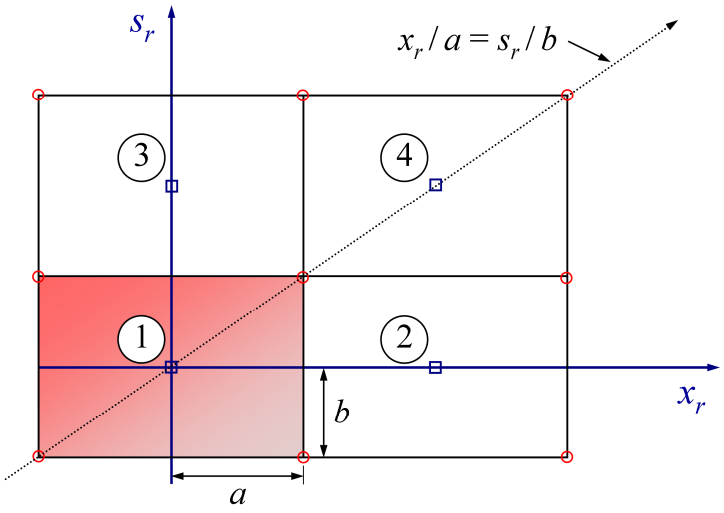
The approach of locally refining the loaded zone has as a drawback, apart from the increase in the computational cost, the need to build a different FE mesh for each lateral position to be loaded, refining a loaded zone each time (unless all the lateral positions to be loaded are refined simultaneously, assuming the additional computational cost). The procedure followed here, without this local refinement, avoids these drawbacks. For this purpose, instead of using a high mesh density in the loaded zone to explicitly reproduce in the FE solution the high displacement gradients existing in that zone, what is done is to extrapolate the displacement at the centre of the loaded zone from the available values at the nodes, using the known variation patterns of the half-space. This is justified because with an adequate refinement (at the PCS level, not at the local level of the loaded zone which represents one of its elements), the size of the loaded zone is considerably smaller than the characteristic dimensions of the cross-section of the solid (even in the study of conformal contact between solids with non-planar surfaces). In this way, it is possible to assume without much error a half-space-like behaviour within the loaded zone itself; i.e. it may be assumed that the variations of the displacements inside the loaded zone are not very different to those of the half-space.

With this idea, the displacements previously obtained by interpolating from the nodal values of the FE model, are corrected with the differences between the theoretical values of the half-space at the points

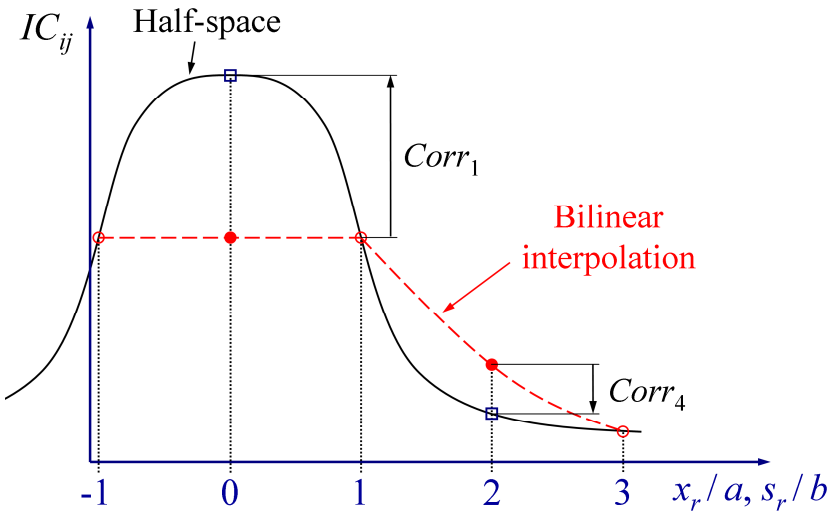
where the interpolations are done, and the interpolated values (with the same type of interpolation as previously used with the FEM nodal values, bilinear in this case) at the same points from the theoretical values of the half-space (not from the numerical FEM values) at the corners of the elements of the PCS. In this way, the errors due to the interpolations from the nodal positions of the FE mesh to the positions of the APs of the objective mesh are compensated.

Figure 5.24 illustrates this correction, which is done for the loaded element of the PCS, and for its adjacent elements as well. Therefore, four correction types are done for each IC: one for the loaded element, and other three for the three types of elements adjacent to the loaded element. The related elements are shown in Figure 5.24a. The corrections for the rest of elements adjacent to the loaded one that are not shown in the figure are obtained from those of the represented elements, by means of (anti)symmetry relations. By way of example, Figure 5.24b represents the corrections for elements 1 and 4, numbered with the same subindex as the element number. The theoretical half-space curve of one of the direct ICs along the diagonal $x_r/a = s_r/b$ is shown, together with the curve obtained by bilinear interpolation from the theoretical half-space values of the nodes.

The most important corrections are those of the direct ICs of the loaded element. The corrections for the crossed ICs of the loaded element are zero, because the theoretical half-space value at the centre of the loaded element, and the value interpolated bilinearly from its four nodal values, are zero.



(a) Identification of the loaded element (shaded in red) and their adjacent elements. The nodal positions are marked with red circles, and the APs with blue squares.



(b) Schematic representation of the corrections for elements 1 and 4. Marked with red circles the nodal values with which the interpolations are done, with red points the interpolated values at the APs of the objective mesh, and with blue squares the theoretical half-space values at the same points.

Figure 5.24. Corrections of the errors due to the bilinear interpolations from the nodal positions of the FE mesh to the APs of the objective mesh.

Here it is assumed that the loaded and adjacent elements of the FE mesh are similar to the elements of the objective mesh, so that the nodal positions of the FE mesh correspond approximately to the corners of the elements of the objective mesh. A more refined procedure could be set out for this correction, interpolating directly the displacements of the APs of the objective mesh from the FEM nodal displacements in the previous step 2, and defining the corrections from the FEM nodal positions to the APs of the objective mesh. In general, this would result in different corrections for each lateral position of the objective mesh. Moreover, in the most general case, there would be six different corrections to cover each loaded element of the objective mesh and its neighbouring elements, as the symmetries in *s* direction between the FE and objective meshes would no longer apply, and only the symmetries in *x* direction would remain.

The steps listed above may be repeated as many times as wished to produce sets of IC matrices for different meshes of the PCS, without the need to repeat FEM calculations and starting from the same nodal displacement result files of the FE analyses. In the next Section 5.2.3 some curves of ICs of solids with non-planar surfaces are shown, numerically calculated with FEM following the procedure described here, together with ICs estimated by analytical approximations detailed in that section.

Figure 5.26 demonstrates the importance of the correction described in step 4 of the above list to be able to use the ICs obtained with FE models as described here (without local refinement in the loaded zone) in the exact contact theory. The figure shows two lateral profiles of normal pressures in the central section of the contact patch of a simple conformal contact problem obtained with CECT, with two sets of numerically calculated IC matrices. The only difference between both sets of IC matrices is that in one of them this correction is included, and in the other it is not. The considered problem is a frictionless contact case between a prismatic solid (designated as rail) and a solid of revolution (designated as wheel), with transversal circular profiles in the contact zone, with a convex radius of 10.0 mm and a concave radius of 10.1 mm. Figure 5.25 shows schematically the geometry of the zone of interest of the cross-sections of the solids, and Table 5.4 lists the relevant geometric parameters and

elastic properties. Figure 5.25 shows as well the orientation of the *X*, *Y* and *Z* axes of the global Cartesian coordinate system.

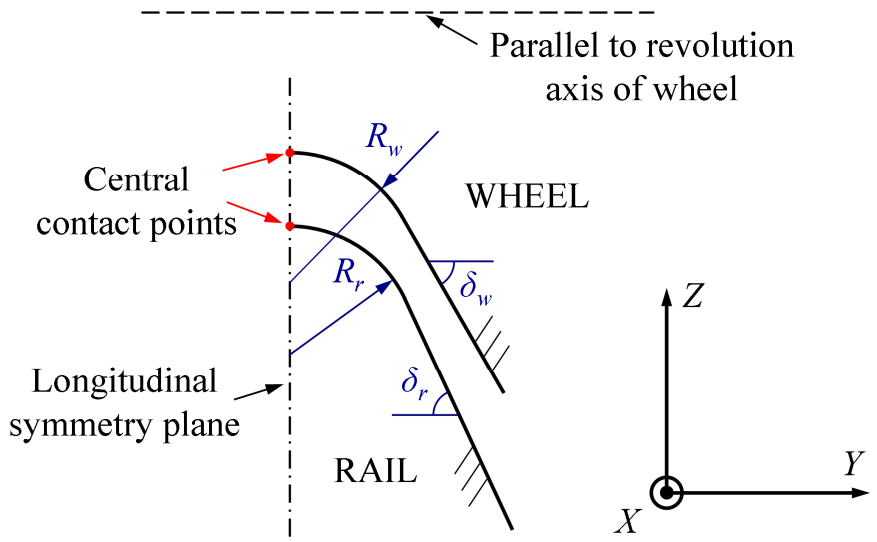


Figure 5.25. Cross-section geometry of the two solids for the frictionless conformal contact problem considered in Figure 5.26.

Table 5.4. Geometric parameters and elastic properties of the solids for the frictionless conformal contact problem considered in Figure 5.26.

Parameter	Units	Value	Description
R_w	mm	–10.1	Radius of circular part of cross-section of wheel.
R_r	mm	10.0	Radius of circular part of cross-section of rail.
R_{roll}	mm	500	Rolling radius of the wheel in the central contact point.
δ_w	º	60	Geometric parameter of the wheel cross-section.
δ_r	º	65	Geometric parameter of the rail cross-section.
δ_0	º	0	Contact angle at the central contact point.
E_w, E_r	GPa	210	Young's modulus of the material of the solids.
ν_w, ν_r	-	0.30	Coefficient of Poisson of the material of the solids.

The normal load between both solids is 80 kN. The mean contact angle is zero, as well as the yaw angle between both solids. In this way, the contact has a longitudinal symmetry plane, and a lateral one. Advantage has been taken of the lateral symmetry of the geometry in this case to cut to half the number of necessary load cases in the FE analyses to produce the IC matrices. For this purpose, the ICs associated to half of the lateral positions of the objective mesh

have been inferred from symmetry and antisymmetry relations from the ICs associated to the corresponding positions symmetrically located with respect to the longitudinal symmetry plane.

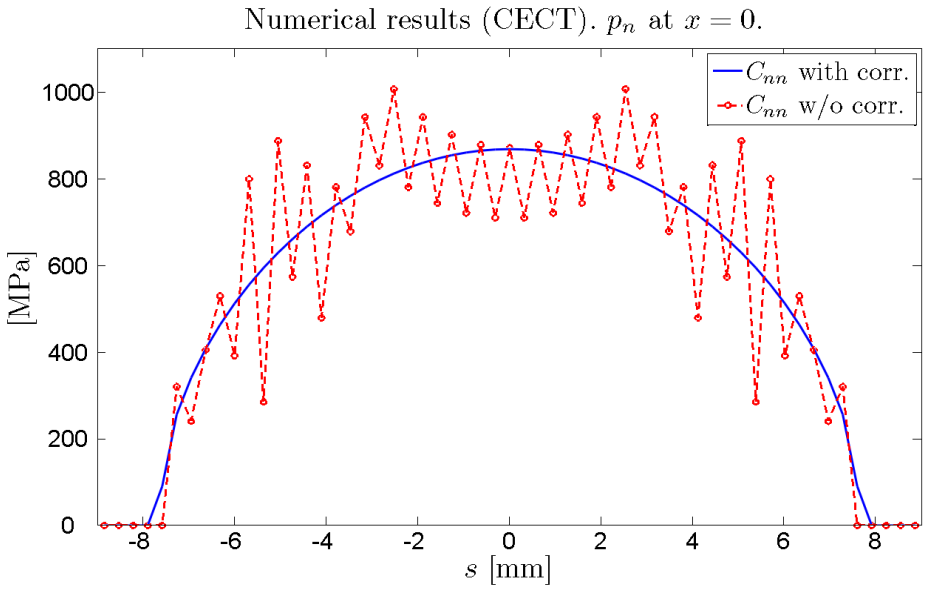


Figure 5.26. Lateral profiles of normal pressures p_n in the central cross-section of the contact patch in a frictionless conformal contact problem obtained with CECT, with two sets of C_{nn} matrices, with and without the previously described correction to compensate the errors resulting from the interpolations in the loaded element and its adjacent ones.

These results have been obtained with a mesh of a similar density to the ones used to produce the majority of the contact numerical solutions shown in this thesis, both with FE models and with the exact contact theory, with around 40 elements in each direction of the contact patch. As can be seen in the figure, the results obtained with the set of ICs without correction are not valid. The C_{nn} matrix without correction leads to a numerically ill conditioned problem (probably because the terms in the matrix diagonal become much less dominant), and a highly irregular normal pressure distribution is obtained, with large variations between adjacent elements. On the other hand, the result obtained with the set of ICs with correction, is good, and has been validated with FEM. It is easily understandable the importance of a good estimation of the ICs of the loaded element with respect to itself, particularly in the case of the direct ICs, as they are the greatest in magnitude. Consider that in the

simplified contact theory it is assumed that the elastic displacements in each element depend only on the load in the same element, ignoring the influence of other elements; in other words, the IC matrices in the simplified contact theory are diagonal, with the ICs of the loaded element with respect to itself (properly scaled) as the only non-zero terms.

Figure 5.27 gives an idea of the precision obtained in the numerical calculation of the ICs with the models and procedure described in this section. The figure represents the differences between the numerically calculated combined direct C_{ii} ICs of the two solids with geometries and elastic properties described in Figure 5.25 and Table 5.4, and the corresponding half-space A_{ii} ICs, in the centre of the loaded element and in the next three adjacent in longitudinal direction. The loaded element is located in the longitudinal symmetry plane of the cross-section of the solids, where the origin of the lateral s coordinate is positioned. The dimensions a and b of the elements of the objective mesh where the ICs have been calculated are the same as those of the loaded elements in the FE models used in the calculation, indicated in Table 5.3 (test 6). The total longitudinal dimension $2a$ of the loaded element is 0.333 mm, so the distance from the centre of the loaded element to the third adjacent element is 1 mm. As this distance is quite below the minimum curvature radius of the surfaces in the zone where the loaded element is located (10 mm), it is assumed that the behaviour of the direct ICs in the considered elements is approximately equivalent to that of the corresponding half-space ICs. Therefore, the half-space ICs provide a reference to assess the precision of the numerically obtained ICs.

For comparison with the A_{ii}, in the C_{ii} the necessary global displacement is subtracted. This global displacement is calculated in each case so that the resulting C_{ii} equates the A_{ii} in the fourth element adjacent to the loaded one in longitudinal direction. This distance is used as a compromise between big enough so that the numerical errors in the computed C_{ii} may be considered low, and yet small enough so that the variations of the C_{ii} and the A_{ii} may still be assumed equivalent. Table 5.5 lists these global displacements in non-dimensional form. These are expressed in the $\{X, Y, Z\}$ global Cartesian system. For the considered loaded element, the directions of the local s lateral and n normal axes coincide with the directions of the Y and Z axes, so it is understood that the load directions Y and Z reference the load cases with the stress applied

in *s* and in *n* directions in that element. As a result of the geometry and load symmetry conditions, in each case the global displacement is aligned with the applied load; that is, there is a single component of the global displacement in each load case, in the same direction as the applied load.

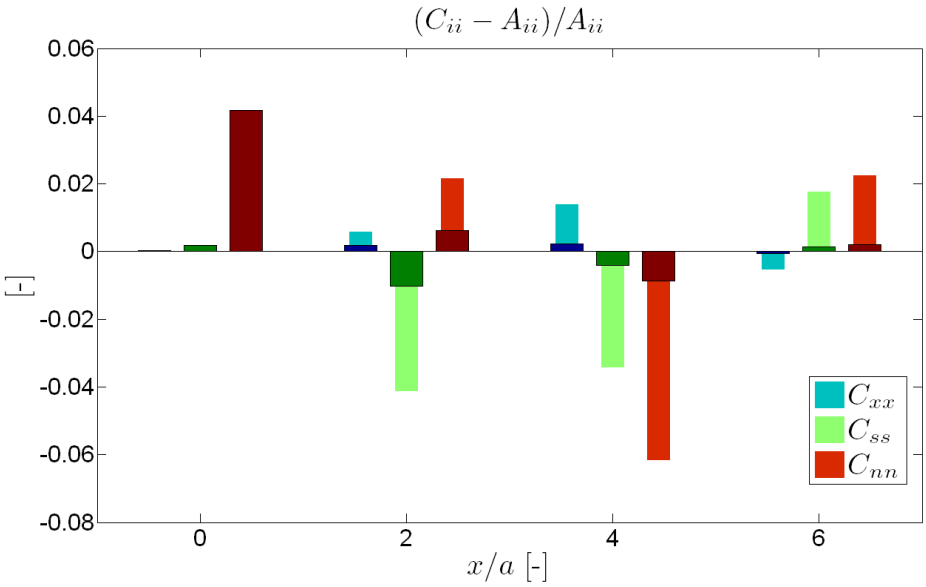


Figure 5.27. Differences between the C_{ii} direct ICs obtained with FEM for the geometry described in Figure 5.25 and Table 5.4, and the corresponding half-space ICs. The differences are given in normalized form, on the one hand with respect to the half-space value of each element (with the narrower and lighter bars), and on the other hand with respect to the half-space value of the loaded element (with the wider and darker bars).

Table 5.5. Non-dimensionalized global displacements, $u_{i,g}\times E/(p\times a)$, in the C_{ii} ICs calculated with the FE models of test 6 of Table 5.3 for the solids with the geometries and elastic properties described in Figure 5.25 and Table 5.4. *E* is the Young's modulus, *p* the uniform pressure and *a* the longitudinal half-side of the loaded element. *i* = *X*, *Y* or *Z*. The longitudinal and lateral half-sides *a* and *b* of the elements of the objective mesh are respectively 0.167 and 0.158 mm.

$u_{i,g}\times E/(p\times a)$	Direction of load and global displacement		
	Longitudinal (*X*)	Lateral (*Y*)	Vertical (*Z*)
RAIL	3.03E–02	1.88E–01	4.67E–03
WHEEL	–1.31E–02	–9.12E–04	–8.32E–03

The positive values in Table 5.5 indicate that the FE model of the solid in question is more flexible in the direction and under the considered load than the

half-space, and the negative values that it is more rigid. As could be expected, the convex solid is more flexible than the half-space in the three directions, especially in lateral direction; and the concave solid, more rigid. Therefore, the stiffness differences of each solid with respect to the half-space tend to cancel each other in the combined direct ICs of the two solids.

As may be seen in Figure 5.27, the errors of each C_{ii} change sign at least once in the considered elements, so it is reasonable to think that they will tend to compensate between them at least in part. The biggest error with respect to the theoretical value of the loaded element is of 4%, and is found in the C_{nn} of the loaded element itself. However, the relative error in the second adjacent element is bigger. On the other hand, and contrary to what could be expected, in the case of the C_{xx} and C_{ss} the errors are bigger in the first and second adjacent elements than in the loaded element; both the relative to the C_{ii} of each element, and the relative ones to the C_{ii} of the loaded element. In any case, except for the mentioned error in the C_{nn} of the loaded element, the rest of the errors relative to the C_{ii} of the loaded element are on the order of 1% or lower.

An alternative to calculate the ICs with the FE models used here, is based on applying the loads punctually, in a single node each time. In this way, the loaded elements become directly APs in the objective mesh, and the interpolations from the nodes to the centres of the elements are avoided, with the large associated errors that have been seen previously. The ICs thus obtained would in principle be valid to be used in contact problems with the exact contact theory without the need of any correction. Even if they had significant errors around the singular loaded node[22] (cf. the considerable oscillation in the nodes adjacent to the loaded one that may be seen in the curve of test 3 in Figure 5.18b), with an adequate mesh density in the PCS these errors should tend to compensate each other. After all, the IC matrices thus obtained are a condensation of the complete matrix of the FE model in the APs of the objective mesh. So it seems reasonable to think that if with some given FE

[22] The nodal values obtained applying the load “distributed” in a single element (translating to four nodal loads) also have non-negligible errors around the loaded zone, due to the coarseness of the discretization of the variation of the load in the surface, as has been seen in Figure 5.27.

models good contact solutions are obtained, the ICs obtained by loading the individual degrees of freedom of the same FE models should also be valid to obtain good results in contact analyses with the exact contact theory. Nevertheless, in this thesis the calculation of the ICs has not been performed in this way. The considerable irregularity in the curve of the ICs around the loaded node that may be seen in Figure 5.18b might give rise to some numerical problem or instability when using these ICs in the exact contact theory.

5.2.3 Analytical approximation

5.2.3.1 Background

While the ICs may be numerically calculated with FEM precisely for any geometry as has been explained in the previous section, this is an arduous task, and additionally the results obtained in each case are limited to a particular geometry of the contacting bodies, to a given coefficient of Poisson and to an analysis mesh with elements of specific dimensions or which may vary only in a narrow range. Therefore, it is of interest to develop analytical expressions that allow at least approximating the real ICs in a sufficiently precise way to be used satisfactorily in detailed contact analysis with the exact contact theory, and avoid its numerical calculation.

In this section, an approach is presented for the analytical approximation of the ICs for 3D non-planar solids. The aim is to get a simple way of approximating their ICs. The proposed approach is based in a simple concept of force decomposition that is represented schematically in Figure 5.28.

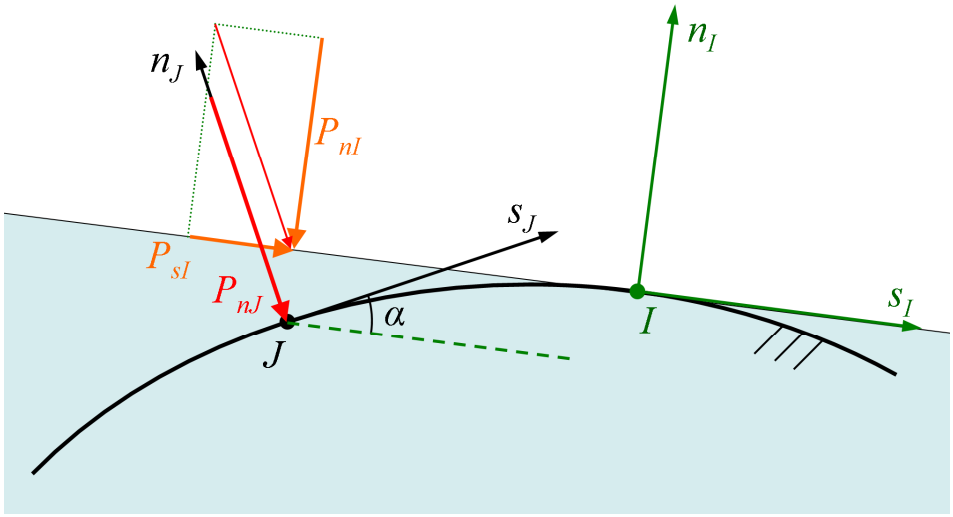


Figure 5.28. Sketch of the force decomposition concept for the analytical approximation of the ICs for 3D solids with curved surfaces.

The figure shows a force P_{nJ} acting in the local normal direction of a given surface point J. This force may be decomposed in the principal axes of another surface point I, the lateral s_I and the normal n_I, as indicated in Eq. (5.71):

$$\overrightarrow{P_{nJ}} = \overrightarrow{P_{nI}} + \overrightarrow{P_{sI}} = -\left|P_{nJ}\right| \times \cos\alpha \times \overrightarrow{n_I} + \left|P_{nJ}\right| \times \sin\alpha \times \overrightarrow{s_I} \tag{5.71}$$

Forces in other directions are decomposed in a similar way. Based on this force decomposition, the approximated ICs are set out as the superposition of the contributions of each force component in the principal axes of the point where the displacements are observed, considering the contribution of each component as equivalent to that of the half-space. In this way, the approximated ICs in a given point I due to the loads applied in a point J are defined in Eq. (5.72). The analytically approximated ICs are denoted here with letter B.

$$\begin{bmatrix} B_{xx} & B_{xs} & B_{xn} \\ B_{sx} & B_{ss} & B_{sn} \\ B_{nx} & B_{ns} & B_{nn} \end{bmatrix} \approx \begin{bmatrix} A_{xx} & A_{xs} & A_{xn} \\ A_{sx} & A_{ss} & A_{sn} \\ A_{nx} & A_{ns} & A_{nn} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}$$

$$= \begin{bmatrix} A_{xx} & A_{xs}c_\alpha + A_{xn}s_\alpha & A_{xn}c_\alpha - A_{xs}s_\alpha \\ A_{sx} & A_{ss}c_\alpha + A_{sn}s_\alpha & A_{sn}c_\alpha - A_{ss}s_\alpha \\ A_{nx} & A_{ns}c_\alpha + A_{nn}s_\alpha & A_{nn}c_\alpha - A_{ns}s_\alpha \end{bmatrix}$$

Or, in compact form: $[B_{ij}] \approx [A_{ij}] \cdot [\mathbf{R}_\mathbf{x}(\alpha)]$ (5.72)

In this approximation it is assumed that the ICs depend only on the relative position and orientation between the point where the load is applied and the point where the displacements are observed. The α angle, represented in Figure 5.28, is the difference in orientation of the principal *s* and *n* directions of points *I* and *J*. Considering prismatic surfaces aligned in the *x* direction, a single rotation α around the *x* axis suffices to define the difference in surface orientation between different points. c_α and s_α stand for $\cos(\alpha)$ and $\sin(\alpha)$ respectively, and $\mathbf{[R_x}(\alpha)\mathbf{]}$ is the rotation matrix of angle α around the *x* axis. The α angle is calculated as the slope of point *J* minus the slope of point *I*. The slope in each point is measured from the horizontal to the tangent of the cross-section directed in the positive *s* direction, counter clock-wise while looking towards the negative *x* axis. On the other hand, the sign of α changes depending of the direction of the *n* axis: if the *n* axis points into the solid the sign is maintained, and if it points out of the solid the sign changes. Eq. (5.72) may be considered for the individual ICs of each solid as well as for the combined ICs of the two contacting solids, and also for the corresponding IFs. For the combined ICs of the differences in the surface displacements of solid 2 with respect to solid 1 (cf. Eq. (1.8)), if the *n* axis points into solid 2, the α angle as defined here keeps its sign.

According to Eq. (5.72), the ICs which may change more in relation to the half-space ICs around the loaded point are the B_{sn} and B_{ns}, due to the elevated magnitude of the direct ICs (the A_{nn} for the B_{sn}, and the A_{ss} for the B_{ns}) in relation to the crossed ICs, and to the rapid variation of the sine function around

the zero values of the angle. This is in agreement with the characteristics of the ICs of the cylindrical geometries and of the sphere seen in §5.2.1.

5.2.3.2 Influence of curvature in longitudinal direction

In the approximation proposed in Eq. (5.72) it is assumed that the surfaces of the solids are curved only in lateral direction, but the presented reasoning could be extended for doubly curved surfaces, formulating the force decomposition in the three space directions, and considering the differences in orientation in lateral and longitudinal directions. Nevertheless, in practical rolling contact problems, and particularly in the wheel–rail application, the conformity in the longitudinal or rolling direction is much lower than what may be in lateral direction. This justifies the assumption of considering that the surfaces of the contacting bodies are curved only in lateral direction, and not in the longitudinal, with regard to the calculation of the ICs. The IC which may be affected to a greater extent by the curvature in longitudinal direction is expected to be the IC_{xn}, for the same reason for which the IC_{sn} and IC_{ns} are the most affected ICs by the curvature in lateral direction. According to the presented reasoning of force decomposition, applied now to differences in the surface orientation in longitudinal direction, the IC_{xn} of a cylindrical wheel (with no curvature in lateral direction) could be approached by Eq. (5.73), to take into account the effect of the curvature in longitudinal direction:

$$B_{xn} \approx A_{xn} \times \cos(\alpha_x) - A_{xx} \times \sin(\alpha_x) \tag{5.73}$$

In this equation the α_x angle represents the difference of the surface orientation in longitudinal direction between the point of load application J and the point where the displacements are observed I, analogously to the α angle in the lateral plane. For small values of the longitudinal coordinate x in relation to the revolution radius of the wheel R_x, as verified in the contact patch in practical rolling contact problems, $\alpha_x \approx \Delta x / R_x$, being Δx the difference of longitudinal position between points J and I[23].

[23] In this approximation the undeformed geometry of the wheel is being considered. If the deformed geometry is considered, or an intermediate configuration between the

To illustrate the influence of the longitudinal curvature in the IC_{xn}, in Figure 5.29 the individual C^k_{xn} are depicted for the prismatic solid and for the solid of revolution with the geometry and elastic properties defined in Figure 5.25 and Table 5.4. These have been numerically calculated with FEM models. Together with the numerically calculated values, estimated values are shown for each of the solids, for different values of their curvatures in longitudinal direction: on the one hand for the prismatic solid as if it had the same curvature in longitudinal direction as the solid of revolution, and on the other hand for the solid of revolution as if it were prismatic. The estimated values are obtained adding or subtracting the second term in the right hand side of Eq. (5.73) to the numerically calculated values, as the values of α_x are small in the zone of interest, so that $\cos(\alpha_x) \approx 1$ and it is assumed that the first term in the right hand side of the equation remains constant regardless of the curvature in longitudinal direction.

The displacements produced in the same lateral position as the applied load are considered, so that there is no difference in orientation in the lateral direction. Specifically the central position in the cross-section of each solid is considered, in its longitudinal symmetry plane, where the local normal directions n are perpendicular to the axis of revolution in the case of the solid of revolution. The applied load is distributed uniformly in a rectangular element with longitudinal a and lateral b half-sides of 0.1667 and 0.158 mm respectively. Outside the loaded zone, C_{xx} is approximately proportional to the inverse of the distance from the centre of the loaded zone, in this case $1 / \Delta x$. On the other hand, $\sin(\alpha_x) \approx \alpha_x \approx \Delta x / R_x$, and therefore the second term of Eq. (5.73) modifies the C_{xn} in an approximately constant quantity in x and proportional to the longitudinal curvature $1 / R_x$.

undeformed and the final deformed one, the α_x angles and their influence in the IC_{xn} according to Eq. (5.73) are lower than shown here.

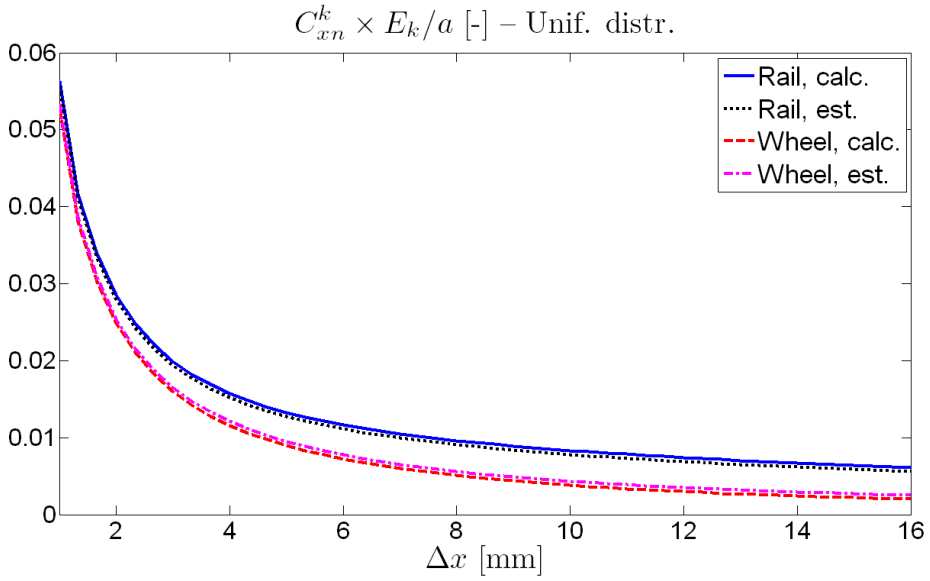


Figure 5.29. Influence of the curvature in longitudinal direction in the IC_{xn}. Variation in x, in the lateral position of $\Delta s = 0$.

Figure 5.29 shows the C^k_{xn} in non-dimensional form, for $\Delta x/a$ values greater than 6. Closer to the loaded zone the different curves tend to approach each other. The graph is extended up to $\Delta x = 16$ mm in the abscissa axis, a typical value of the size of the contact patch in the wheel–rail application. The maximum value of the non-dimensional quantity $A^k_{xn} \times E_k/a$ for the half-space in this case is 0.28, and is produced at the $(\Delta x/a, \Delta s/b) = (1, 0)$ coordinates. As may be seen in the figure, the differences in the calculated and estimated C^k_{xn} of each solid with different R_x values are relatively reduced. This confirms that the simplification of not taking into account the effect in the ICs of the curvature in longitudinal direction is adequate. On the other hand, the differences between the C^k_{xn} of each solid, related with the geometry of the cross-section of each solid (one being convex and the other concave) are quite bigger. These last differences have appreciable consequences in the tangential contact in cases of conformal rolling contact, as will be explained in Chapter 6. These differences are further addressed in §5.2.3.5, proposing a method for their better approximation.

5.2.3.3 Analogy with cylindrical ICs

Recalling the equations of the ICs of cylindrical geometries, a parallelism may be appreciated between that 2D case and the approximation for 3D solids with non-planar surfaces proposed here. As a matter of fact, the ICs of the cylinder and the cylindrical cavity under a single force in the surface are composed of a term multiplied by the cosine of the angular difference between the point where the load is applied and the point where the displacements are taken (that angular difference, designated as α here, was designated as θ in §2.2.2 and 5.2.1), and of another term multiplied by the sine of that angular difference, as in Eq. (5.72). Additional terms appear in the cylinder, different as a function of the support conditions, which turn out to be rigid body displacements. Therefore, they have no affection at the contact mechanics level, and may be disregarded in the present discussion (cf. footnote no. 20).

For example, in Eq. (2.131) for the a_{nn} IF of the cylindrical cavity under a single normal force, the first term in the right which is multiplied by $\sin(\theta)$ may be assimilated to the a_{sn}^{hp} of the corresponding plane geometry (that is, the half-plane), and the second term in the right multiplied by $\cos(\theta)$, to a_{nn}^{hp}, in line with the approximation for B_{nn} in Eq. (5.72). Taking the limit for $\theta \to 0$, and the adequate sign of θ in each case, it is verified that:

$$-\frac{\kappa-1}{4\pi G}\lim_{\theta\to 0}\left(F\left(\tilde{\theta}\right)\right)=\operatorname{sgn}(\theta)\frac{1-2\nu}{4G}=a_{xz}^{hp}\equiv a_{sn}^{hp} \tag{5.74}$$

$$-\frac{\kappa+1}{8\pi G}\lim_{\theta\to 0}\left(\ln\left(2-2\cos\theta\right)\right)=\frac{1-\nu}{2\pi G}\times 2\lim_{\theta\to 0}\left(\ln\left(\frac{1}{\theta}\right)\right)\propto a_{zz}^{hp}\equiv a_{nn}^{hp} \tag{5.75}$$

$F\left(\tilde{\theta}\right)$ and $\tilde{\theta}$ in Eq. (5.74) are as defined in Eqs. (2.136) and (2.137). In Eq. (5.75), the series expansion of the cosine function has been used truncated in the term $\theta^2/2$ (neglecting higher order infinitesimal terms for $\theta \to 0$), and that $\ln(\theta)$ varies in the same proportion as $\ln(1/\theta)$. After the second term, the proportionality symbol is used (with the aim of denoting similar variation or asymptotic behaviour, and not proportionality) instead of the equality, because as it is known the absolute magnitude of the a_{nn}^{hp} IF of the half-plane is

undetermined, being necessary to choose an arbitrary x_0 reference to define it (cf. Eq. (2.117)).

Besides the equalities of Eqs. (5.74) and (5.75), it is observed also that the term which in Eq. (2.131) for $a_{nn}^{cav,b}$ is multiplied by sin(θ), in Eq. (2.133) for $a_{sn}^{cav,b}$ is multiplied by cos(θ); and the term which is multiplied by cos(θ) in $a_{nn}^{cav,b}$, is multiplied by sin(θ) in $a_{sn}^{cav,b}$, with the adequate sign in each case. This is also in line with the idea of the force decomposition presented here, taking into account that the direct a_{ss} IF of the half-plane is equal to the a_{nn}. The cylinder has the same expressions, changing the sign of the θ angle.

Therefore, in the cylindrical geometries a force decomposition law similar to the one proposed here for 3D general geometries is fulfilled exactly. As a difference, in the expressions for the ICs of the cylindrical geometries, the terms assimilated to the a_{sn}^{hp} include a factor variable with θ, equal to $|\pi - |\theta|| / \pi$. It has been tried to include this factor in the A_{sn} half-space ICs for use in the B_{ij} ICs, so as to make them more similar to the exact ICs for cylindrical geometries, but with this it has not been found any consistent precision improvement of the B_{ij} ICs for 3D prismatic solids. Moreover, if the materials of the two contacting bodies are elastically similar, as in the wheel–rail application, the combined half-space A_{sn} are zero, and in this case the factor $|\pi - |\theta|| / \pi$ becomes irrelevant for the B_{ij} approximation of the combined ICs.

5.2.3.4 Moving the tangent point of the equivalent half-space

In the approximation of Eq. (5.72) it has been assumed that the behaviour of the non-planar solid is equivalent to a half-space tangent to its surface at *I*. A variant of this approximation is obtained by considering a tangent plane at the location *J* where the load is applied, as illustrated in Figure 5.30. This gives approximated ICs denoted by $\overline{B}_{ij}$ according to Eq. (5.76).

$$\left[\overline{B}_{ij}\right] \approx \left[\mathbf{R}_{\mathbf{x}}(\alpha)\right] \cdot \left[A_{ij}\right]$$

$$= \begin{bmatrix} A_{xx} & A_{xs} & A_{xn} \\ A_{sx}c_\alpha - A_{nx}s_\alpha & A_{ss}c_\alpha - A_{ns}s_\alpha & A_{sn}c_\alpha - A_{nn}s_\alpha \\ A_{sx}s_\alpha + A_{nx}c_\alpha & A_{ss}s_\alpha + A_{ns}c_\alpha & A_{sn}s_\alpha + A_{nn}c_\alpha \end{bmatrix} \tag{5.76}$$

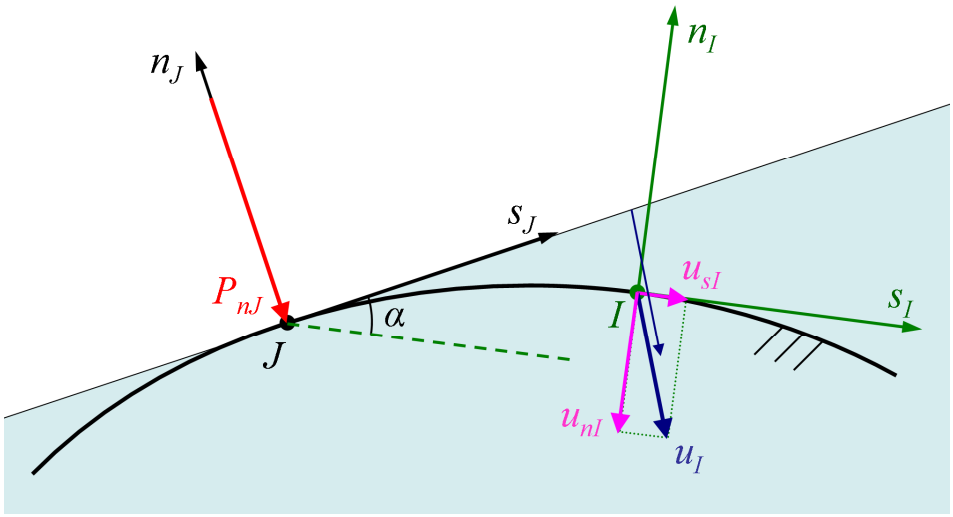


Figure 5.30. $\overline{B}_{ij}$ variant of surface orientation-based approximated ICs.

A third variant $\hat{B}_{ij}$ is set out imagining the tangent half-space halfway between points I and J, oriented at an angle $\beta = \alpha/2$ apart from the tangents at those points. This variant, suggested by Vollebregt [Blanco-Lorenzo 2021], could be thought of as a mix between the previous B_{ij} and $\overline{B}_{ij}$ variants. Defining $A_{s+n} = (A_{ss} + A_{nn}) / 2$, and $A_{s-n} = (A_{ss} - A_{nn}) / 2$, the resulting expressions for the $\hat{B}_{ij}$ ICs are given in Eq. (5.77).

$$\left[\hat{B}_{ij}\right] \approx \left[\mathbf{R}_{\mathbf{x}}(\beta)\right] \cdot \left[A_{ij}\right] \cdot \left[\mathbf{R}_{\mathbf{x}}(\beta)\right]$$

$$= \begin{bmatrix} A_{xx} & A_{xs}c_\beta + A_{xn}s_\beta & -A_{xs}s_\beta + A_{xn}c_\beta \\ A_{xs}c_\beta + A_{xn}s_\beta & A_{s+n}c_\alpha + A_{s-n} + A_{sn}s_\alpha & -A_{s+n}s_\alpha + A_{sn}c_\alpha \\ A_{xs}s_\beta - A_{xn}c_\beta & A_{s+n}s_\alpha - A_{sn}c_\alpha & A_{s+n}c_\alpha - A_{s-n} + A_{sn}s_\alpha \end{bmatrix}$$

(5.77)

This $\hat{B}_{ij}$ variant has the merit of preserving the symmetries $\hat{B}_{sx} = \hat{B}_{xs}$, $\hat{B}_{xn} = -\hat{B}_{nx}$ and $\hat{B}_{sn} = -\hat{B}_{ns}$; and hence verifying reciprocity.

In [Marshek 1984] the problem of the approximation of the B_{nn} IC was addressed in the context of 3D frictionless conformal contact analyses with cylindrical surfaces. Here a different approximation for the B_{nn} IC was proposed, according to the scheme of Figure 5.31, as $B_{nn} \approx A_{nn} \times \cos(\alpha/2)$.

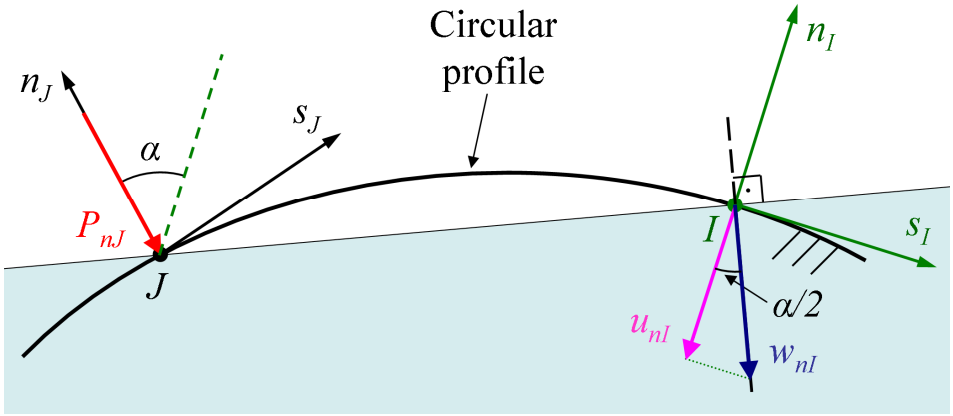


Figure 5.31. Scheme of the approximation for B_{nn} proposed in [Marshek 1984] (adapted from the cited reference).

In the remainder, the B_{ij} and $\hat{B}_{ij}$ variants (Eqs. (5.72) and (5.77)) will be considered. The notation B will be used generically also, designating any of the variants of ICs presented here as well as the extension of the following section. The majority of the CECT results reported in subsequent chapters of this thesis

have been obtained with the $\hat{B}_{ij}$ variant of approximated ICs described in this section.

5.2.3.5 Extension of approximated ICs for the geometric differential stiffness

The load/displacement decomposition effect, which is the basis for the calculation of the previously exposed *B* ICs based on the surface orientation, explains many of the differences between the ICs of non-planar solids and those of the half-space, but there are other differences which are not explained by this effect. The differences between the C^{k}_{xn} of a convex and a concave body seen in Figure 5.29 are an example of this. In §5.2.3.6 considerable differences between the combined B_{xn} and C_{xn} ICs of two conforming bodies with elastically similar materials are illustrated as well.

These discrepancies are attributed to the differences between the stiffnesses of the convex and concave conforming bodies, that come as a result of the different geometry of their cross-sections: the concave body, which has more material, tends to be stiffer than the convex one. This characteristic is not captured by the B_{ij} ICs treated in the previous sections. This section provides an extension for the B_{ij} ICs to overcome this limitation. For most of the ICs, the effect of the surface orientation proves to be dominant in the deviation of the ICs of non-planar solids from those of the half-space, and the B_{ij} ICs previously presented provide a good approximation. However, in the particular case of the C_{xn}, the effect of the geometric differential stiffness between the two conforming bodies plays a more important role. Consequently, that approximation is not satisfactory for the C_{xn}, and it becomes necessary to take into account the effect of the geometric differential stiffness to improve it.

The approach followed here to incorporate the differential stiffness effect is to define modified, effective elastic properties of the conforming bodies, that vary as a function of the distance from the loaded element in the surface. Close to the loaded element, at a distance much lower than the typical dimensions of the general features of the cross-sections of the bodies, the bodies look like half-spaces, their geometric differential stiffness tends to zero, and hence their effective elastic properties will be their real ones. As the distance increases, a bigger portion of the material and (unequal) cross-sections of the bodies

influences their behaviour, and more difference between their stiffnesses will be noted. Their approximated ICs are then calculated with modified elastic properties instead of the real ones of the bodies to approximate the variation of the geometric differential stiffness between the conforming bodies.

The modified, effective elastic properties of the bodies vary between their real values at zero distance from the loaded element, to some limit values at a given limit distance from the loaded element related to the geometry of the cross-sections of the bodies. This is represented here using the function defined in Eq. (5.78). This form of empirical equation has been chosen with the aim of providing a continuous transition in space for the values of the effective elastic properties, having the possibility of adjusting the rate of change at both limits of the variation interval.

$$f(r) = f^{0} + \left(f^{\infty} - f^{0}\right) \times \left[(1-\rho)\cdot \rho^{m_f} + \rho \cdot \left(1-(1-\rho)^{n_f}\right)\right] \tag{5.78}$$

$$r = \sqrt{x_r^{\,2} + s_r^{\,2}}\ ,\ \rho = \min\left(1, r/r_{\infty,f}\right). \tag{5.79}$$

The nomenclature used in Eq. (5.78) is as follows: f refers to an elastic property, such as the combined G, ν or K of both bodies, or an individual elastic constant of one of the bodies. For each property, f^0 is its effective value at zero distance from the centre of the loaded element, equal to its real value, and f^∞ its effective value at infinite distance from the loaded element. (x_r, s_r) is a position from the centre of the loaded element in x and s directions, r the actual surface distance from the centre of the loaded element as defined in Eq. (5.79), and $r_{\infty,f}$ a limit distance beyond which the elastic property f no longer changes and is equal to its effective value at infinity f^∞. m_f and n_f are two constant exponents for each elastic property f. The effective elastic properties obtained with Eq. (5.78) are used to calculate modified A_{ij}, designated as A'_{ij}. These A'_{ij} are then used in Eq. (5.77) for the $\hat{B}$ ICs presented in §5.2.3.4. This extended version of the $\hat{B}_{ij}$ is designated as $\hat{B}'$.

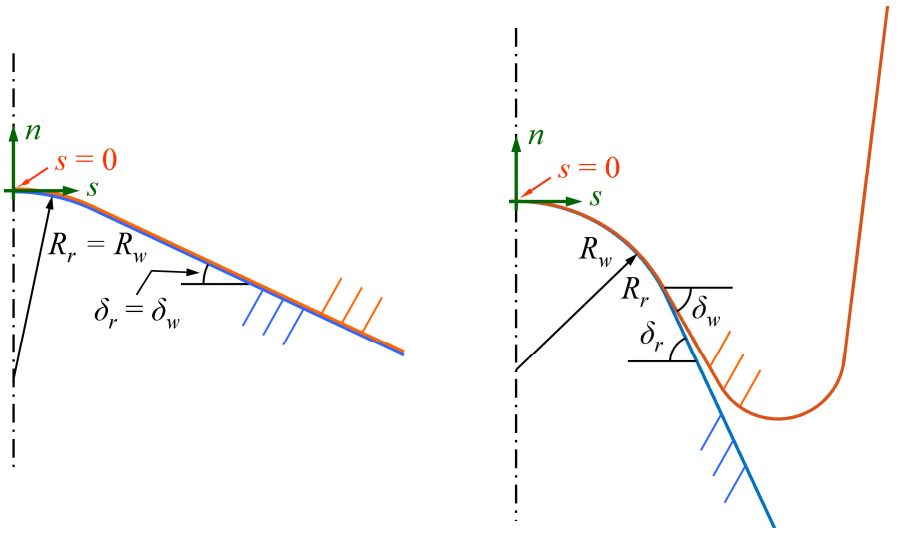
An empirical approach is followed to determine the parameters f^∞, $r_{\infty,f}$, m_f and n_f for each elastic property. The C ICs are calculated numerically with FEM for a single lateral position of the load in the cross-sections of the contacting bodies,

and these are used as a reference to adjust the parameters in Eq. (5.78). A drawback is that numerical calculation of some of the C_{ij} ICs is involved, but this is less costly than computing the full C_{ij} matrices. The lateral position used for the loaded element in the calculation of the C_{ij} ICs is chosen in the central zone of the expected contact patches.

With the $\hat{B}'_{ij}$ ICs, it is aimed specifically at improving the approximation of the C_{xn}, where the differential stiffness effects are more prominent. Taking this into account, here the $\hat{B}'_{ij}$ will be computed on the basis of adjusting just the elastic mismatch constant of both bodies K, seeking to improve the approximation for the C_{xn} while affecting the other B_{ij} ICs as little as possible. Thus, the $\hat{B}_{xn}$ and the $\hat{B}_{sn}$ ICs will be affected most by this extension, as well as their reciprocals, and to a lesser extent the $\hat{B}_{xs}$, $\hat{B}_{ss}$ and the $\hat{B}_{nn}$ ICs through the contributions of the A'_{xn} and A'_{sn}.

5.2.3.6 Comparison of approximated and numerically calculated ICs

Next, some comparisons are done between different variants of approximated B ICs proposed here, half-space A ICs, and C ICs numerically computed with FEM. Two geometries are considered for this purpose, with cross-sections depicted in Figure 5.32 and related parameters listed in Table 5.6. Two prismatic bodies are considered in geometry 1, and a concave wheel on a convex rail in geometry 2. Geometry 2 is the same as that described in Figure 5.25 and Table 5.4. In both geometries the lower convex body is designated as the rail, and the upper concave one as the wheel. The materials of the conforming bodies are elastically similar in both cases. As listed in the table, slightly different values of v have been considered for each of the geometries. This difference does not have significant effects in the comparisons shown here.



(a) Geometry 1. (b) Geometry 2.

Figure 5.32. Cross-sections of conforming bodies, symmetric about $s = 0$, considered for comparison of ICs.

Table 5.6. Parameters of conforming bodies considered for comparison of ICs.

Symbol	Description	Units	Geometry no.	
			1	2
δ_r	Surface inclination of rail (lower body)	º	25.5	65
δ_w	Surface inclination of wheel (upper body)	º	25.5	60
R_r	Radius of circular part of cross-section of rail	mm	10.0	10.0
R_w	Radius of circular part of cross-section of wheel	mm	10.0	10.1
R_{roll}	Nominal rolling radius of wheel (at $s = 0$)	mm	∞	500
E_w, E_r	Young's modulus	GPa	210	210
ν_w, ν_r	Coefficient of Poisson	–	0.28	0.30
a	Longitudinal half-side of loaded element	mm	0.200	0.167
b	Lateral half-side of loaded element	mm	0.200	0.158

Both geometries have a cross-section with circular central zone delimited by inclined straight sections, with different subtended angles in each case. Besides testing the performance of the *B* ICs in each of these cases, the comparison of their respective *C* ICs allows to appreciate the effect of their different geometries. An objective is to observe the effect of the different overall cross-sections of each geometry (with different inclinations of the side straight parts) on the ICs in the central circular zone, common in both geometries. This serves as a test for the consistency of the surface orientation based approach of the *B*

ICs, which takes into account specifically surface orientation differences in the region of interest.

Comparisons are made of the combined ICs of two bodies, as well as of the individual ones for each of the bodies of geometry 2. The geometries are identified with superindex 1 for the geometry of Figure 5.32a and 2 for the geometry of Figure 5.32b. For the individual ICs of each of the bodies of geometry 2, superindex "2,*r*" is used for the rail and "2,*w*" for the wheel. Specifically the ICs associated to the loaded element located in the lateral position $s = 0$ are compared, in the longitudinal symmetry plane of the cross-sections of both solids.

The *C* ICs have been calculated with FE models for both geometries shown in Figure 5.32. The *C* ICs for geometry 1 have been provided by Edwin A.H. Vollebregt. Some of them were shown in [Vollebregt 2014a]. The FE mesh has been refined locally around the loaded element for these ICs, to provide adequate resolution for the step load variation and capture the relatively high displacement gradients. This local mesh refinement has been omitted for the *C* ICs corresponding to geometry 2. Instead, extrapolation is used on the basis of the trends of the half-space for the displacements in the loaded element and its neighbouring elements, as explained in §5.2.2.2.

To adequately compare the numerically obtained ICs with the analytical ones, appropriate global displacements are subtracted from the numerical ICs in each case. On the one hand the global displacements listed in Table 5.7 are deducted. The values for geometry 2 are the same in each direction as the sum of the individual values listed in Table 5.5 for the rail and the wheel, after scaling according to the different normalizations of the quantities listed in each table. On the other hand, in the case of lateral tangential load, some global rotations $\theta_{x,g}$ around the longitudinal axis are deducted as well for geometry 2. These rotations are calculated so that the vertical displacements at points located in the same longitudinal position as the loaded element, symmetrically with respect to the longitudinal symmetry plane, and sufficiently far from the loaded element, are equated. For this purpose, specifically the displacements at the lateral extremes of the objective mesh in which the *C* ICs have been obtained in this case are considered, located at $s = \pm 8.8$ mm. The values of these global rotations, normalized with the Young's modulus E, the rail radius R_r, and the

uniform pressure p and dimensions a and b of the loaded element as $\theta_{x,g} \times ER_r^2/(4ab \times p)$, are –1.036 rad for the rail and 0.122 rad for the wheel. For geometry 1, the values are lower due to the lower δ_r and δ_w angles, and are neglected here. The resulting global displacements from Table 5.7 and the $\theta_{x,g}$ rotations are subtracted from the numerical ICs after transforming them to the local (x, s, n) coordinate system of each element. The global displacements associated to the load in direction j are subtracted from the C_{ij} ICs (with $i, j = x$, s or n).

Table 5.7. Non-dimensionalized global displacements, $u_{i,g} \times ER_r/(4ab \times p)$ (with $i = X$, Y or Z), in the C ICs calculated with FEM for the solids with the geometries and elastic properties described in Figure 5.32 and Table 5.6.

$u_{i,g} \times ER_r/(4ab \times p)$	Direction of load and global displacement		
Geometry no.	**Longitudinal (X)**	**Lateral (Y)**	**Vertical (Z)**
1	6.47E–02	5.04E–01	–8.57E–02
2	2.72E–01	2.96E+00	–5.77E–02

The trends of the ICs are graphed in the longitudinal direction x and lateral direction s, along trajectories with a separation of 1.6 mm from the centre of the loaded element. This offset is used to avoid the relatively high displacements gradients at the singular loaded element masking the variations of the ICs outside it. Besides, the trends of the ICs close to the loaded element approach to those of the half-space. Even though the IC_{ij}, and particularly the direct IC_{ii}, of the loaded element are considerably bigger than the ones shown in the following figures[24], the latter are still relevant in the behaviour of the solids at the local contact level, and observing their differences with respect to the half-space ICs some qualitative differences between conformal and non-conformal

[24] Recall as a reference the maximum values of the direct A^1_{xx} and A^1_{zz} of the half-space for a rectangular element with uniform load distribution and sides ratio $a/b = 0.7$ shown in Figure 5.6 a and d. The non-dimensionalised A^1_{xx} and A^1_{zz} for the uniform element in that figure reach values around 2 in the centre of the loaded element. As those are individual ICs, they should be multiplied by two to compare them with the combined ICs shown here. Additionally, as they are non-dimensionalised with one of the sides of the loaded element, they should be multiplied by the current a/b relation and divided by the a/b relation with which they were obtained, to make a uniform comparison with the ICs shown here.

contacts may be explained, as will be seen in Chapter 6. Consider for example that in a non-conformal circular contact, if the pressure was uniform in the whole contact patch[25], the contribution of all the concentric differential rings in which the contact patch may be divided, to the normal elastic displacement at the central point, is the same: being r the distance to the central point, the b_{nn} change in proportion to $1/r$, and the differential area of each ring changes in proportion to r; therefore the contribution of each ring, proportional to the product of its b_{nn} with its differential area, would be independent of r and constant in this case.

The ICs are given per unit load in the uniformly loaded rectangular element, and are normalized additionally by multiplication with E and R_r. The dimensions of the loaded element are given in Table 5.6 for each geometry. The ICs are given for a unit total load rather than for a unit pressure in each loaded element, to compare directly the ICs of both geometries.

The ICs are graphed in non-dimensional relative coordinates $(x, s)/R_r$ with respect to the centre of the loaded element, taking it as the origin. Thus, the IC_{xn} and IC_{sn} ICs reverse their sign with respect to curves shown in previous graphs, in which the position coordinates were in the opposite direction. Some of the ICs are omitted for clarity in some graphs, when they are equivalent or very similar to others shown. In the graphs along the lateral direction s of the combined ICs, the limit between the circular and straight sections in geometry 1 is marked with a vertical dotted line. This change in geometry has an appreciable effect on the trends of several of the ICs as may be appreciated in the different graphs.

In the following, the direct IC_{ii} and the IC_{xs} ICs are examined in §5.2.3.6.1, and the IC_{xn} and IC_{sn} ICs are discussed in §5.2.3.6.2. The deviations of the ICs of convex bodies on the one hand and of concave bodies on the other, with respect to the corresponding half-space ICs, are generally opposite to each other. These deviations counteract each other in the combined ICs of the two conforming

[25] By way of example, in a Hertzian circular contact the variation of the normal pressure with respect to the maximum value in the centre is below 15% inside half the radius of the contact patch.

bodies in the case of the IC_{ii} and IC_{xs} ICs, and are added together in the case of the IC_{xn} and IC_{sn} ICs.

As a summary, it is seen that the approximated ICs have different accuracies for each type of IC. Also, the accuracy of the approximated ICs depends among others on the shape of the non-planar surfaces; the affection of the sudden curvature change in the cross-section of geometry 1 of Figure 5.32 and of the solid of Figure 5.48 on the accuracy of the B_{sn} beyond the lateral position of this change is an example of this. Overall, it may be concluded that the surface orientation-based approximation gives a good first order estimation of the main deviations between the ICs of non-planar solids and those of the half-space. Additionally, with a simple reasoning it provides insight into the way in which these deviations occur. However, some limitations of the approximation are apparent, and additional work is needed to overcome them.

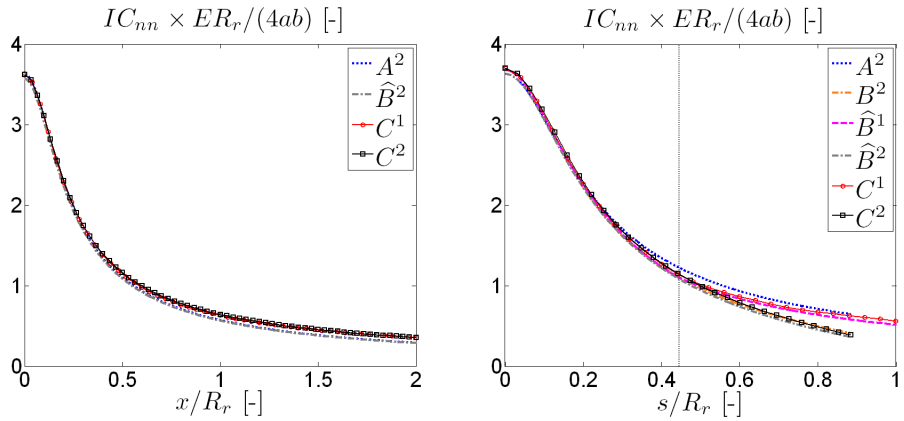
In Chapter 6, some effects of the differences between the half-space ICs and those of solids with curved surfaces will be seen, both in the normal and in the tangential part of the contact problem, and to which extent they are captured with the *B* approximated ICs.

5.2.3.6.1 Comparison of IC_{ii}-s and IC_{xs}

First the direct IC_{ii}-s will be revised. Their curves in *s* and *x* are symmetric with respect to the origin of each coordinate.

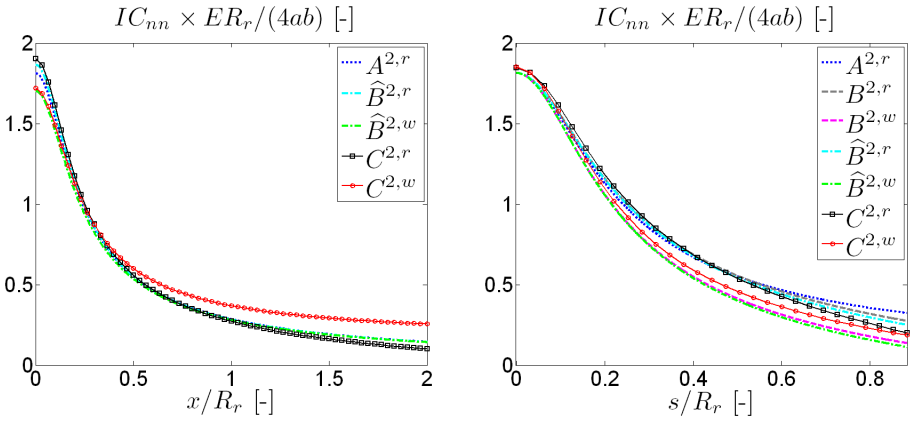
The greatest differences between the combined C_{nn} and the A_{nn} are produced in the lateral direction, as may be seen in Figure 5.33. Both the combined (Figure 5.33) and the individual (Figure 5.34) C_{nn} tend to become smaller than the corresponding A_{nn} as the difference in orientation α from the point of load application increases, particularly in the case of the combined C_{nn} and in the individual ones of the concave body. The B_{nn} capture well these deviations in the geometry considered, as shown in Figure 5.33b and Figure 5.34b. The deviations between the individual C_{nn} and A_{nn} of the convex body are smaller, as may be seen in Figure 5.34b. This is again well represented by either the B_{nn} or the $\hat{B}_{nn}$. Referring to Eq. (5.72), the second term in the equation for B_{nn} is positive in the case of the convex body (including its minus sign), and opposes the decrease of the first and principal term with respect to A_{nn}. The different

cross-sections of both geometries do not produce big differences in the variations of the combined IC_{nn} in the longitudinal direction, as seen in Figure 5.33a. The most notable differences in this direction occur in the individual IC_{nn} of the concave body as seen in Figure 5.34a, with the $C^{2,w}{}_{nn}$ having less pronounced variations than either the corresponding *A* or *B* ICs.



(a) Along *x* at *s* / R_r = 0.16. (b) Along *s* at *x* / R_r = 0.16.

Figure 5.33. Combined IC_{nn} for geometries 1 and 2 of Figure 5.32.

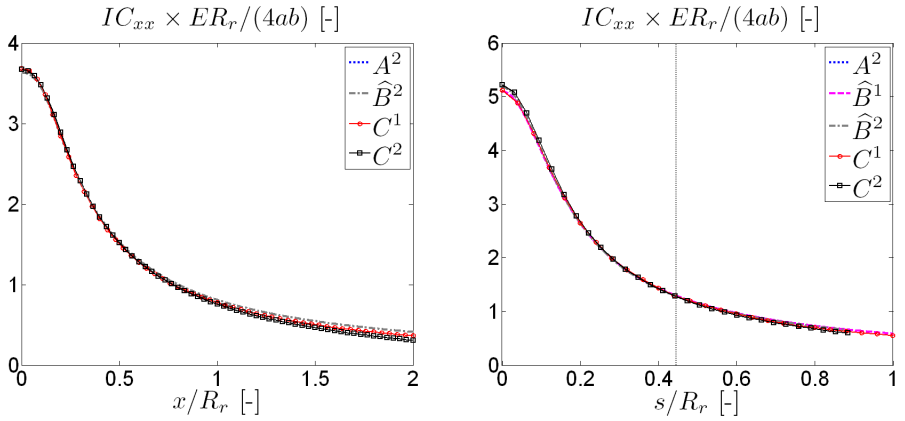


(a) Along *x* at *s* / R_r = 0.16. (b) Along *s* at *x* / R_r = 0.16.

Figure 5.34. Individual IC_{nn} for geometry 2 of Figure 5.32.

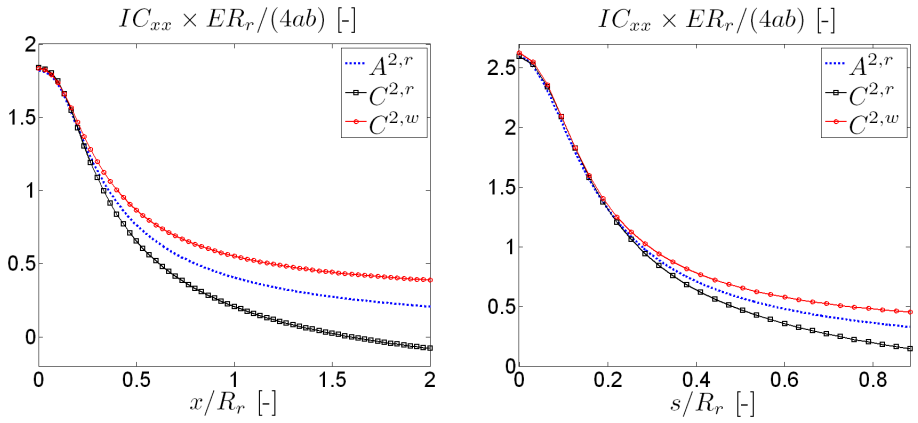
The deviations between the combined C_{xx} and A_{xx} are very limited in the longitudinal as well as in the lateral direction, cf. Figure 5.35, and the combined C_{xx} of both geoemtries are very similar. The surface orientation based

approximation, which in this case amounts to the equality of the B_{xx} and the A_{xx}, works well for the combined C_{xx} of both solids. On the contrary, there are considerable differences between the individual C_{xx} and A_{xx} of each body, both in longitudinal and in lateral directions, as can be seen in Figure 5.36. The C_{xx} of the convex body have higher gradients than the A_{xx}, and those of the concave body, smaller. These deviations are higher in the convex body. The deviations are of opposite sign in each of the bodies, and tend to oppose each other in the combined C_{xx}. In the case considered here with elastically similar materials, they turn out to nearly cancel out with each other, but it has to be taken into account that the extent to which this compensation occurs depends on the relative stiffness of both bodies. This observation is also applicable for the IC_{nn}, IC_{ss} and IC_{xs} ICs.



(a) Along x at s / R_r = 0.16. (b) Along s at x / R_r = 0.16.

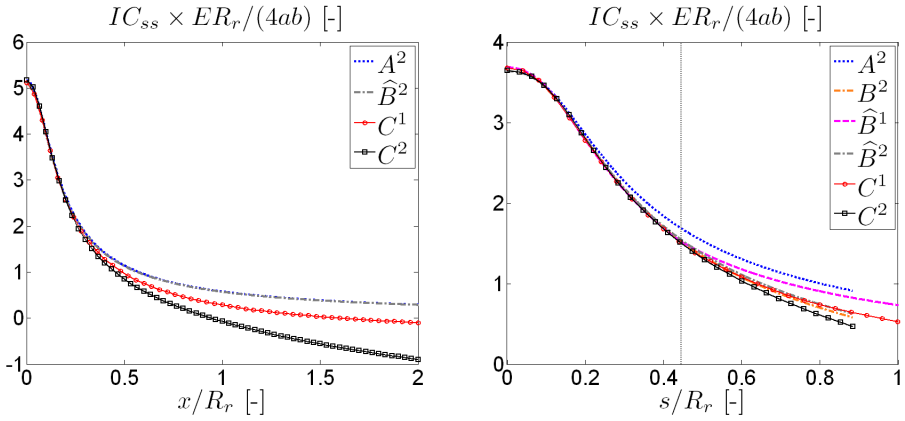
Figure 5.35. Combined IC_{xx} for geometries 1 and 2 of Figure 5.32.



(a) Along x at s / R_r = 0.16. (b) Along s at x / R_r = 0.16.

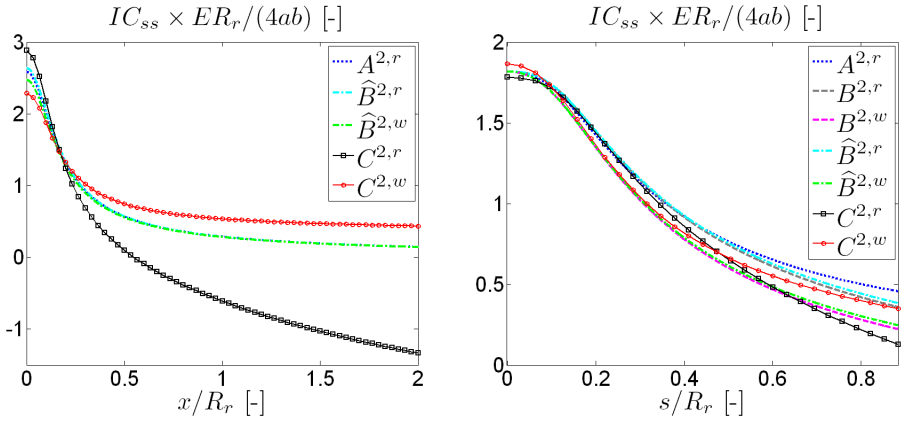
Figure 5.36. Individual IC_{xx} for geometry 2 of Figure 5.32.

The combined C_{ss}, Figure 5.37, have significant deviations with respect to the A_{ss} in both principal directions of the surface, tending to become smaller as the distance with respect to the loaded element increases. The trend in the lateral direction is relatively well reproduced with the B_{ss} (Figure 5.37b), but not in longitudinal direction (Figure 5.37a). The C_{ss} maintain a considerably higher gradient than the B_{ss} as the longitudinal distance from the loaded element increases, especially for geometry 2, with the highest difference between the convex and the concave cross-sections. This is attributed to the twist of the convex body under lateral load, which causes sections away from the loaded one to be more uncoupled from it than in the case of the half-space. The deviations between the individual C_{ss} and both the corresponding A and B ICs are considerably higher, especially for the convex body in the longitudinal direction, as can be seen in Figure 5.38a.



(a) Along x at s / R_r = 0.16. (b) Along s at x / R_r = 0.16.

Figure 5.37. Combined IC_{ss} for geometries 1 and 2 of Figure 5.32.

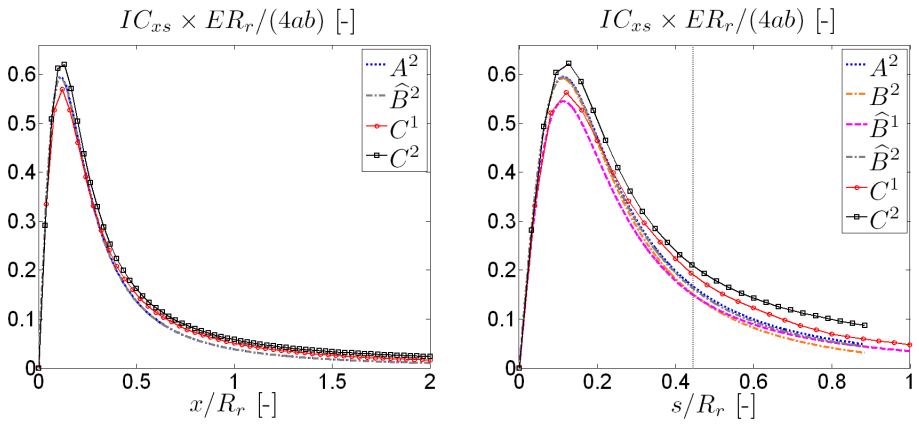


(a) Along x at s / R_r = 0.16. (b) Along s at x / R_r = 0.16.

Figure 5.38. Individual IC_{ss} for geometry 2 of Figure 5.32.

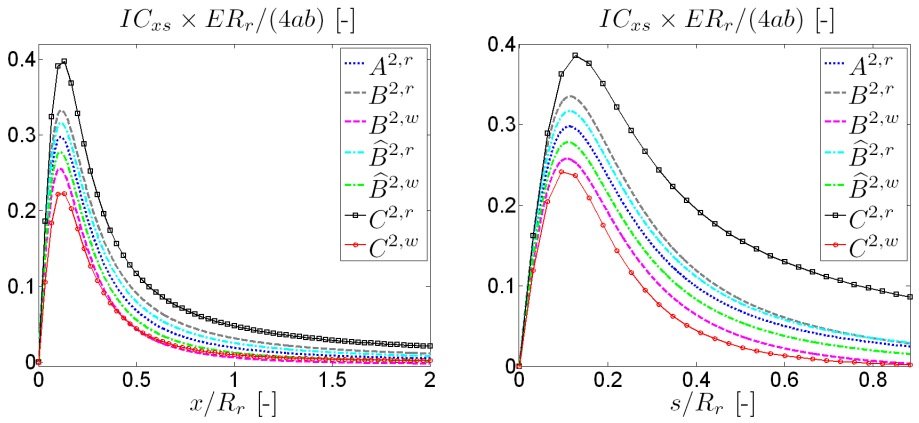
Next, the IC_{xs}-s are revised in Figure 5.39 and Figure 5.40. Their curves are antisymmetric both in s and x. The highest differences between the combined C_{xs} and both the A_{xs} and the B_{xs} happen in the lateral direction (Figure 5.39b). These are attributed to the difference in longitudinal stiffness of the two bodies, resulting from the cross-sections. Referring to the individual C_{xs} in Figure 5.40, it is seen that the individual C_{xs} of the convex body deviate more from the half-space A_{xs} than those of the concave body. This indicates that the material missing in the convex cross-section (with respect to the half-space), has a

greater effect than the additional material present in the concave section, like with the C_{ss}. The B_{xs} and $\hat{B}_{xs}$ capture only part of the differences between the A_{xs} and C_{xs}, with the B_{xs} performing better in this case than the $\hat{B}_{xs}$. The remaining differences in the individual IC_{xs} are higher in the lateral direction, like in the combined IC_{xs}. The performance of the B_{xs}-s is not fully satisfactory. The consequences of this are limited because the magnitude of the IC_{xs} is significantly smaller than that of the direct IC_{xx} and IC_{ss} ICs.



(a) Along x at s / R_r = 0.16. (b) Along s at x / R_r = 0.16.

Figure 5.39. Combined IC_{xs} for geometries 1 and 2 of Figure 5.32.



(a) Along x at s / R_r = 0.16. (b) Along s at x / R_r = 0.16.

Figure 5.40. Individual IC_{xs} for geometry 2 of Figure 5.32.

5.2.3.6.2 Comparison of IC_{xn} and IC_{sn}

Next, the IC_{xn} and IC_{sn} ICs are examined. Comparisons with the $\hat{B}'$ ICs introduced in §5.2.3.5 are included for the combined ICs. The curves of the IC_{xn} are symmetric in s and antisymmetric in x, and those of the C_{sn} are antisymmetric in s and symmetric in x. While these ICs have also a considerably lower magnitude than the direct IC_{ii}-s, their influence can be significant in the tangential contact problem as shown in Chapter 6. This is because the normal pressures usually reach much higher values than the tangential stresses. Additionally, for the case of the IC_{xn}, which are considerably lower in magnitude than the IC_{sn}, it has to be considered that they are antisymmetric with respect to the longitudinal/rolling direction. This increases the gradients of the related elastic displacements in that direction, relevant in the tangential part of rolling contact problems, upon performing the convolution with the normal pressures.

The combined IC_{xn} are plotted in Figure 5.41 for geometry 1 and in Figure 5.42 for geometry 2. The C_{xn} of both geometries are included in both figures to better appreciate the differences between them. Table 5.8 lists the parameters of the effective elastic property K used for the $\hat{B}'$ ICs shown in these figures.

Table 5.8. Parameters of adjusted elastic property K used for the combined $\hat{B}'$ ICs of the two geometries described in Figure 5.32 and Table 5.6.

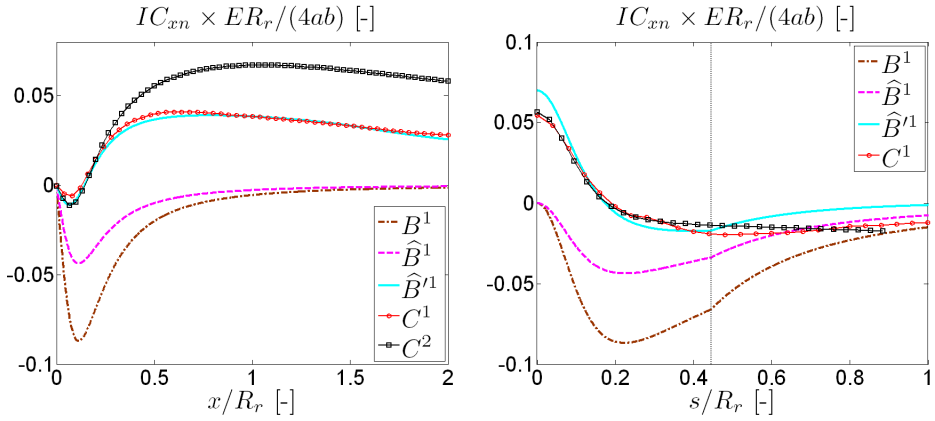
	Geometry 1	**Geometry 2**
K^{∞}	–0.065	–0.145
$r_{\infty,K} / R_r$	1.60	2.25
m_K	0.67	0.90
n_K	1.80	2.00

As expected, the differential stiffness effect is stronger in geometry 2, and consequently its C_{xn} reach higher values than the C_{xn} of geometry 1. This is seen in the plots along x (Figure 5.41a and Figure 5.42a). The B_{xn} are a poor approximation of the C_{xn}, though they reproduce to a certain extent some of their characteristics. In fact, the B_{xn} turn out to be even a worse approximation for the C_{xn} than the A_{xn}, due to the effect of the differential stiffness previously discussed. Upon examination of the individual IC_{xn} in Figure 5.43, it may be noted that the B_{xn} and the $\hat{B}_{xn}$ predict the highest displacements in the concave

body (cf. the second term of the approximated B_{xn} in Eqs. (5.72) and (5.77), with different sign in each solid), while the highest individual C_{xn} (in magnitude) are those of the convex body. The difference is higher in the longitudinal trend (Figure 5.43a) for the convex body.

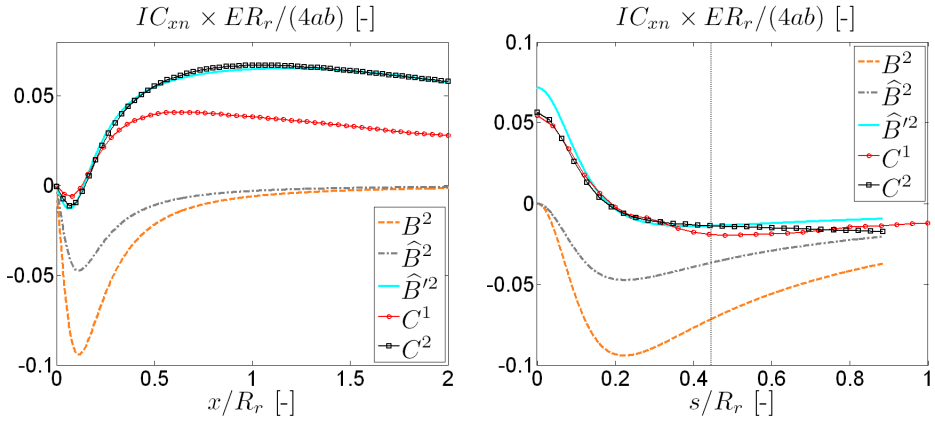
The adjusted $\hat{B}'_{xn}$ provide a much better approximation of the combined C_{xn}, especially along x, although the values are somewhat overestimated close to the loaded element. This happens mostly for geometry 1, as seen in the longitudinal trend in Figure 5.44 through $s = 0$ (i.e., at the same lateral position as the loaded element). On the other hand, it has been tried to adjust different individual effective elastic properties seeking to properly approximate the combined and the individual C_{xn} at the same time, but this has not been achieved with the $\hat{B}'$ ICs.

Part of the differences between the C_{xn} of both geometries may be attributed to the difference in the longitudinal curvature of the upper bodies, using a prismatic body in geometry 1 and a body of revolution in geometry 2. The curve labelled as "$C + A^{2,w}{}_{x+n} \sin(\alpha_x)$" in Figure 5.44 is an estimate of the combined C_{xn} of geometry 2 if the wheel were replaced by a prismatic body with the same cross-section. This has been computed applying the $\hat{B}$ approximation in longitudinal direction, in a similar way as discussed in §5.2.3.2. As may be seen in Figure 5.44, the estimated effect of the longitudinal curvature of the wheel on the C_{xn} computed for geometry 2 explains part of the differences between the C_{xn} computed for both geometries. However, this effect is small compared to the remaining differences between the C_{xn} of each geometry, attributed to the different magnitude of the geometric differential stiffness in each case. Also, the longitudinal curvature of the wheel explains only a small part of the differences of the C_{xn} of geometry 2 with the rest of IC_{xn}-s.



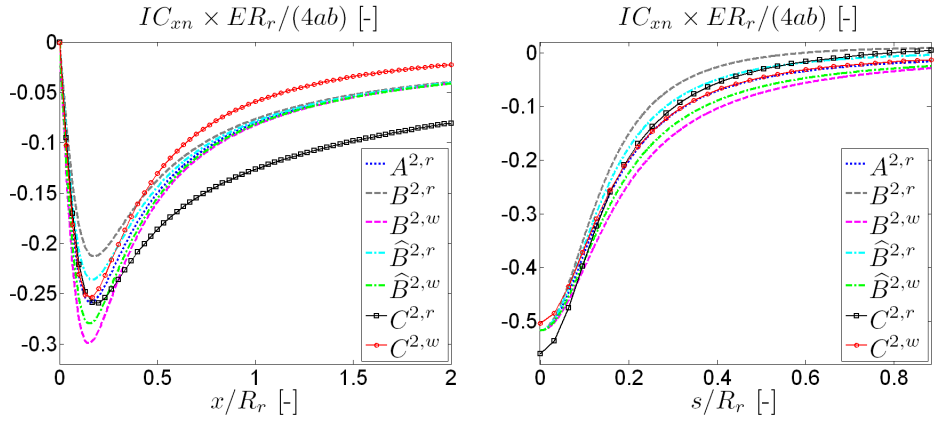
(a) Along x at s / R_r = 0.16. (b) Along s at x / R_r = 0.16.

Figure 5.41. Combined IC_{xn} for geometry 1 of Figure 5.32.



(a) Along x at s / R_r = 0.16. (b) Along s at x / R_r = 0.16.

Figure 5.42. Combined IC_{xn} for geometry 2 of Figure 5.32.



(a) Along *x* at *s* / R_r = 0.16. (b) Along *s* at *x* / R_r = 0.16.

Figure 5.43. Individual IC_{xn} for geometry 2 of Figure 5.32.

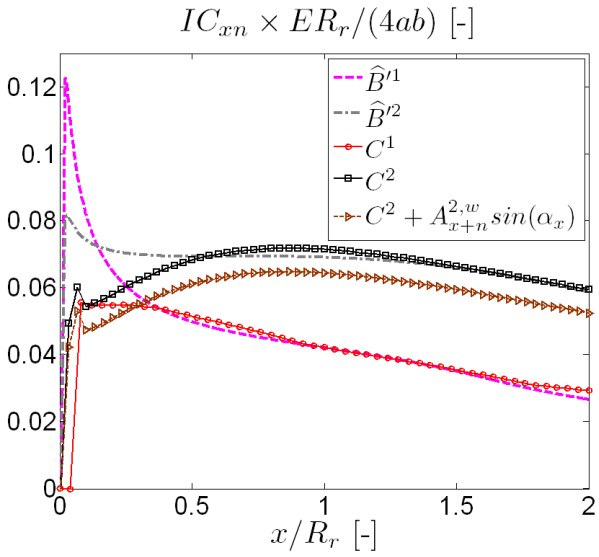


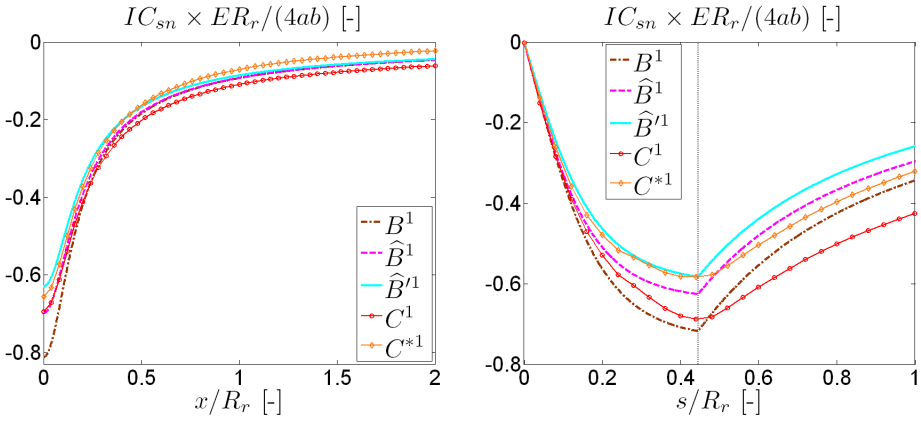
Figure 5.44. Combined IC_{xn} for geometries 1 and 2 of Figure 5.32, along *x* at *s* = 0.

The most notable differences between the A_{ij} and C_{ij} crossed ICs (with $i \neq j$) occur in the IC_{sn}. These are examined in Figure 5.45 to Figure 5.47. As it happens with the a_{sn} of cylindrical geometries, the individual C^k_{sn} of the convex body are considerably lower than those of the half-space in the vicinity of the loaded zone, and the C_{sn} of the concave body are higher in magnitude, cf. Figure 5.13. The elastic sphere shows this same behaviour, as seen in Figure 5.15. In

the combined C_{sn}, these differences, of opposite sign, are added; they do not counteract each other. As will be seen in Chapter 6, using directly the A_{sn} in place of the C_{sn} in the solution of conformal contact problems with the exact contact theory may give rise to important errors in the tangential part of the contact problem.

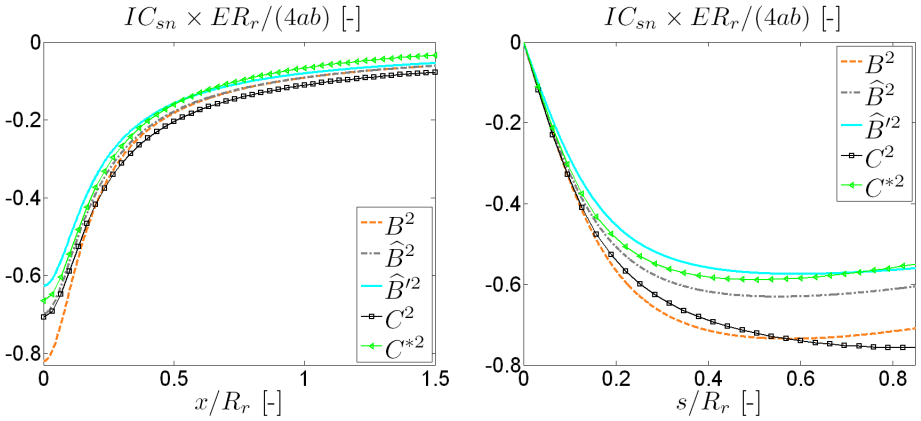
The C_{sn} are adequately approximated with both the B_{sn} and the $\hat{B}_{sn}$, for the combined (Figure 5.45 and Figure 5.46) as well as for the individual (Figure 5.47) IC_{sn} of each body, especially inside the circular zone of the cross-sections. The greatest discrepancies are seen in the lateral trends beyond the change from circular to straight profile in geometry 1, cf. Figure 5.45b. The B_{sn} approximate better the C_{sn} than the $\hat{B}_{sn}$ in both geometries considered, and the approximation of the $\hat{B}'_{sn}$ is worse than that of the $\hat{B}_{sn}$.

A considerable part of the differences between the C_{sn} on the one hand, and $\hat{B}_{sn}$ and $\hat{B}'_{sn}$ on the other, turns out to be equivalent to a rigid body vertical displacement. This does not affect the solution of contact problems except for the relative displacement of the bodies at a given load (cf. footnote no. 20). This is shown with the C^* curves in Figure 5.45 and Figure 5.46, that are obtained by adjusting the global vertical displacements discounted to the C_{sn} to match better with the $\hat{B}'_{sn}$. In this case, the non-dimensional global vertical displacements $u_{i,g} \times ER_r / (4ab \times p)$ for geometry 1 and 2 have been changed from –8.57E–02 and –5.77E–02 (cf. Table 5.7) to 1.57E–01 and 2.19E–01, respectively. The C^r_{sn} curves in Figure 5.47 may be adjusted as well to match better with the corresponding B^r_{sn} curves, though not shown in the figure. In this case, this is achieved by changing the non-dimensional global vertical displacement of the rail from the value shown in Table 5.5 (4.67E–03) to 1.32E–02. On the other hand, differences between different IC_{sn} give rise to considerable differences in some of the results in the tangential part of the contact problem as shown in §5.2.3.7.



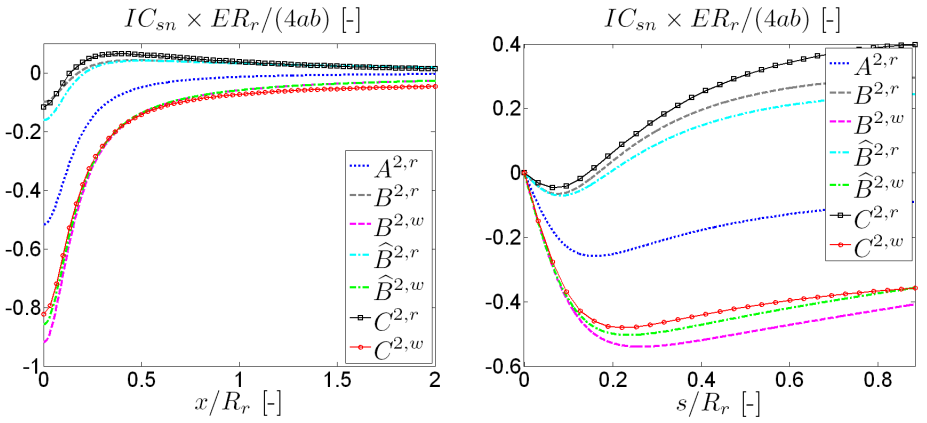
(a) Along *x* at *s* / R_r = 0.16. (b) Along *s* at *x* / R_r = 0.16.

Figure 5.45. Combined IC_{sn} for geometry 1 of Figure 5.32.



(a) Along *x* at *s* / R_r = 0.16. (b) Along *s* at *x* / R_r = 0.16.

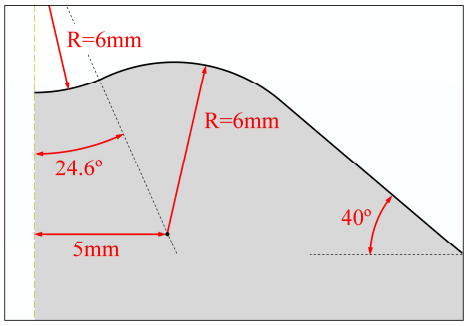
Figure 5.46. Combined IC_{sn} for geometry 2 of Figure 5.32.



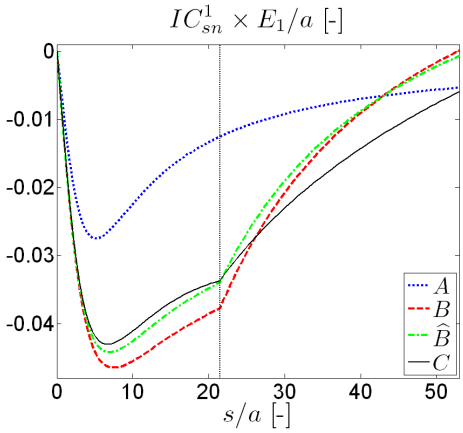
(a) Along x at s / R_r = 0.16. (b) Along s at x / R_r = 0.16.

Figure 5.47. Individual IC_{sn} for geometry 2 of Figure 5.32.

The next figures show how the approximated B_{sn} ICs work for other two individual prismatic solids with different geometries in their cross-section. In these cases, the loaded element is located also in the longitudinal symmetry plane of the cross-sections, but its dimensions are different, as indicated in the caption of each figure. In Figure 5.48 a solid with double curvature is considered, concave in the central part and convex in the adjacent lateral parts of the section. The curvature change between the different parts is produced in a single point, with no transition. The s coordinate of this point is marked in the graph of Figure 5.48b with a vertical dotted line. On the other hand, in Figure 5.49 a convex solid is considered with its central part of elliptical form instead of circular, with the long axis oriented vertically.

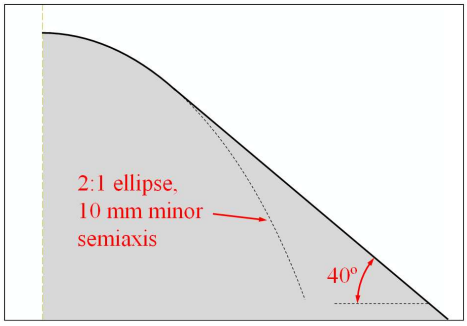


(a) Geometry of the cross-section (showing only part of half of the section, symmetric with respect to the yellow vertical axis in the figure).

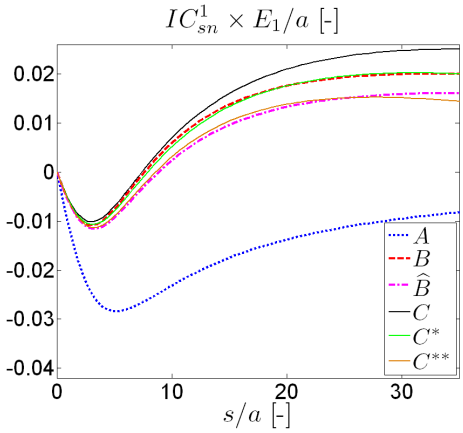


(b) Lateral variation of the ICs (antisymmetric with respect to s = 0) at x / a = 5.18.

Figure 5.48. Individual IC_{sn} of a non-planar prismatic solid with double curvature (concave and convex) in its cross-section. Loaded element located at s = 0, in the symmetry axis of the section. Dimensions of the loaded element: a = 0.12 mm, b = 0.103 mm.



(a) Geometry of the cross-section (showing only part of half of the section, symmetric with respect to the yellow vertical axis in the figure).



(b) Lateral variation of the ICs (antisymmetric with respect to s = 0) at x / a = 5.18. In the "C*" and "C**" curves non-dimensional global vertical displacements $u_{z,g}$×E_1/(p×a) of 7.875E–03 and 1.6625E–02 resp. are deducted.

Figure 5.49. Individual IC_{sn} of a non-planar prismatic solid with elliptical geometry in its cross-section. Loaded element located at s = 0, in the symmetry axis of the section. Dimensions of the loaded element: a = 0.12 mm, b = 0.106 mm.

As can be seen in Figure 5.48b, for the solid with double curvature the B_{sn} do not approach as well the C_{sn} as with the simpler geometry previously considered. Nevertheless, an important part of the deviations between the C_{sn} and the A_{sn} are corrected, and the sudden change in the lateral trend of the B_{sn} foreseen at the point of the curvature change in the section is also produced in the C_{sn}, although in a less pronounced way. The B_{sn} work much better with the next geometry, as can be seen in Figure 5.49b. Three C_{sn} curves are plotted in the same figure; one without deducting any global displacement, and other deducting the global vertical displacements indicated in the figure caption, with which a very good adjustment between the C_{sn} and each B_{sn} is achieved in the shown section. In the trends in longitudinal direction, not shown in these figures, a good adjustment is achieved as well between the B_{sn} and the C_{sn} of both geometries.

5.2.3.7 Performance of approximated ICs in the solution of conformal contact problems

Here numerical computations are reported for different wheel–rail conformal contact test cases, assessing the performance of the different variants of approximated ICs in the normal and the tangential parts of the contact problem. CECT is used for this purpose. Each test case is solved with different sets of ICs, to evaluate to which extent errors in the ICs affect contact related outputs. For the normal part of the contact problem, treated in §5.2.3.7.1, additional reference results are shown obtained with FEM contact analyses.

5.2.3.7.1 Normal contact

This section evaluates the errors made in the solution of the normal part of the contact problem when using approximated ICs in cases of conformal contact between bodies with different cross-sections.

Description of geometries used

Four test cases are considered, with a concave wheel placed on a convex rail aligned in longitudinal direction x, with zero yaw angle and mean contact angle. Therefore, the geometries are symmetric along the overall longitudinal and lateral directions at the central point of the contact at $(x, s) = (0, 0)$.

In the first two test cases the bodies have circular cross-sections, and in the last two test cases they have different elliptical cross-sections in the vertical plane perpendicular to x. The geometries are defined with the semi-axes b_{el} and d_{el} of their circular or elliptical cross-sections in the overall lateral (y) and vertical (z) directions, i.e.

$$\left(\frac{y-y_c}{b_{el}}\right)^2+\left(\frac{z-z_c}{d_{el}}\right)^2=1,$$

$$\text{or } y = y_c + b_{el}\cos(\theta),\ z = z_c + d_{el}\sin(\theta). \tag{5.80}$$

y_c and z_c in Eq. (5.80) are the lateral and vertical coordinates of the centre of the ellipse, and θ is the parameter of the curve. The surface inclination α is obtained from the latter parametrization. The s coordinate is defined implicitly using the arc length along the surface. Elliptical integrals are involved in its analytical calculation for the elliptical cross-sections.

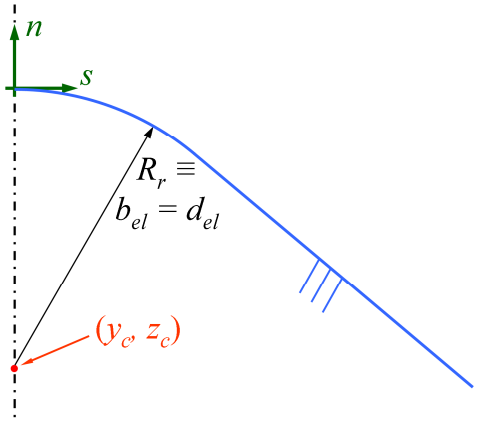
The relevant parameters for each case are listed in Table 5.9. In all test cases the wheel and rail material is steel, with E = 210 GPa and ν = 0.30. The test cases with circular cross-sections are labelled as “circ. s.” (“circ. short”) and “circ. l.” (“circ. long”), referring to the length to width ratio a/b of the resulting contact patches in each case. An objective of these test cases is to study the effect of the a/b ratio on the results. On the other hand, the test cases with elliptical cross-sections are labelled as “el. wide” and “el. tall”, referring to the ratio b_{el} / d_{el} of the cross-sections in each case.

Table 5.9. Parameters of test cases for normal contact. R_{roll} is the nominal rolling radius of the wheel.

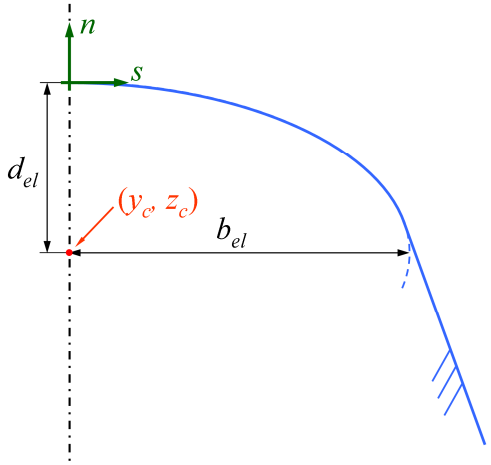
	b_{el} / d_{el} [mm]		R_{roll}
	Wheel	Rail	[mm]
Case “circ. s.”	10.1 / 10.1	10.0 / 10.0	500.0
Case “circ. l.”	10.5 / 10.5	10.0 / 10.0	653.7
Case “el. wide”	13.7 / 6.49	12.97 / 6.485	653.7
Case “el. tall”	6.985 / 14.7	6.485 / 12.97	653.7

The rail cross-sections for test cases “circ. l.”, “el. wide” and “el. tall” are depicted in Figure 5.50. Test case “circ. s.” uses the same geometry described in Figure 5.25 and Table 5.4. The rail used in test case “el. wide” is flatter than the circular geometry, such that the surface inclination changes slower across the s

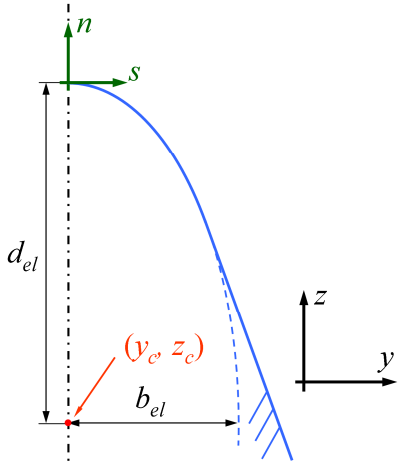
direction than for the circular rail. In this case, the rate of change $\partial\alpha/\partial s$ is lowest at the top of the rail and increases towards the sides. Test case “el. tall”, on the other hand, exhibits a faster change of the contact angle than the circular geometry at the top of the rail, and a reducing rate of change towards the sides.



(a) Test case “circ. l.”.



(b) Test case “el. wide”.



(c) Test case “el. tall”.

Figure 5.50. Rail cross-sections with circular or elliptical central zone.

The b_{el} / d_{el} ratio of the rail is 2/1 in test case “el. wide”, and 1/2 in test case “el. tall”. The b_{el} and d_{el} dimensions of the wheel have been adjusted in each test case with elliptical cross-sections so that the a/b ratios of the resulting contact

patches are around the same values in each case. In this way, the effects the a/b ratio are studied with the "circ.*" test cases on the one hand, and the effects of the form of the contact angle variation across s are studied with the "el.*" test cases on the other, to allow a more clear interpretation of the exposed results.

Results computed with FEM

For each test case, the frictionless contact has been calculated under different loads. The FE contact models have been built with the commercial FE package Abaqus/Standard [Abaqus doc 2012]. The geometries have been meshed with C3D8 linear hexahedron elements. The FE mesh structure used in these models is similar to that shown in Figure 5.17, for both the wheel and the rail. The surface dimensions of the elements in the most refined zone of each mesh are listed in Table 5.10.

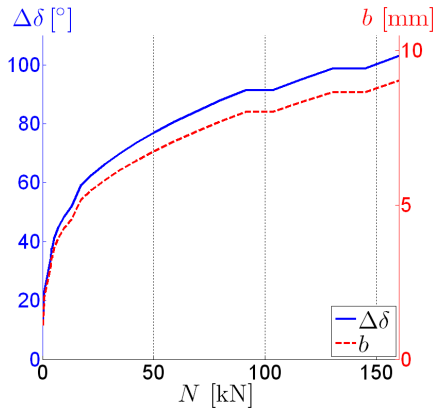
Table 5.10. Surface dimensions of elements (longitudinal × lateral) in the most refined zone of each FE mesh.

Test case	Element surface dimensions [mm]
"Circ. s."	0.33 × 0.31
"Circ. l."	0.36 × 0.29
"El. wide"	0.36 × 0.34
"El. tall"	0.30 × 0.11

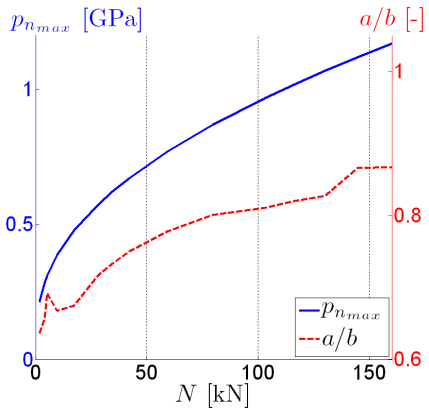
Figure 5.51 to Figure 5.54 show representative results obtained with FEM for the different test cases, such as the contact patch sizes and angle variation, as a function of the compressive normal load N between the contacting bodies. Contact patch shapes are shown in subfigures c for different total angle variations in each test case. The N values corresponding to these contact patches are indicated in the legend and caption of each figure. Comparing the N values in the "circ.*" cases it is seen that case "circ. l.", with less conformity than case "circ. s." (i.e. with greater difference between the convex and concave transversal radii of the contacting surfaces; cf. Table 5.9), requires much bigger loads to produce the same angle variation. Referring to the "el.*" cases, case "el. wide", using a flatter rail, requires bigger loads than case "el. tall".

It has to be mentioned that the loads used for test cases "circ. l." and "el. wide" are higher than what could be considered realistic in practical applications. Also, the resulting contact pressures are above usual levels to initiate yield. The

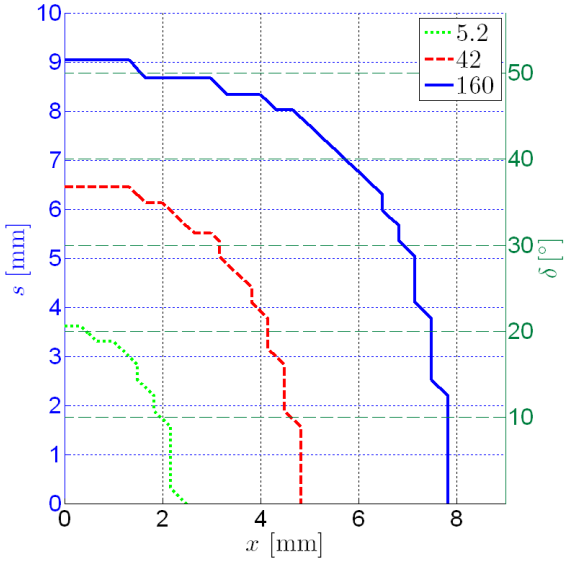
purpose of these test cases is to illustrate the influence of different factors on the performance of the approximated ICs in the solution of contact problems.



(a) Total contact angle variation $\Delta\delta$ and lateral dimension of contact patch b.

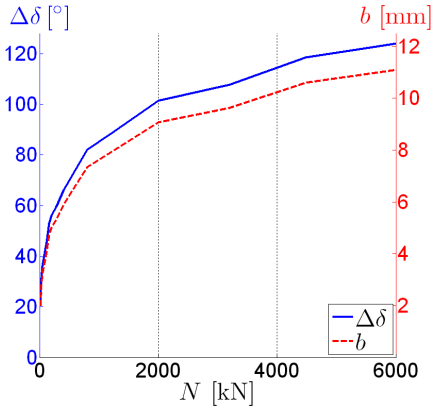


(b) Maximum normal pressure $p_{n,max}$ and a/b ratio.

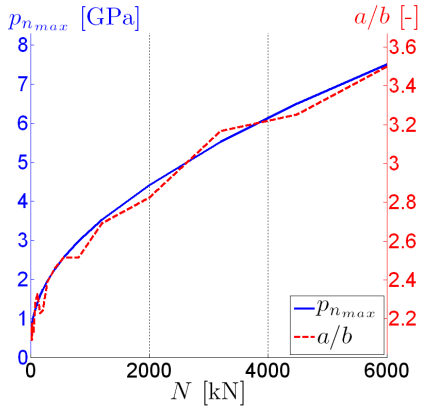


(c) Contact patch contours (symmetric about $x = 0$ and $s = 0$) for N = 5.2, 42 and 160 kN corresponding to $\Delta\delta \approx 40, 70, 100^{o}$.

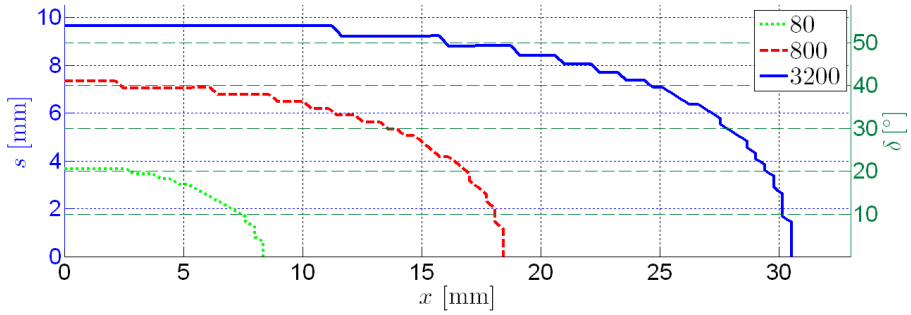
Figure 5.51. Results of frictionless compression test case “circ. s.” calculated with FEM.



(a) Total contact angle variation $\Delta\delta$ and lateral dimension of contact patch b.

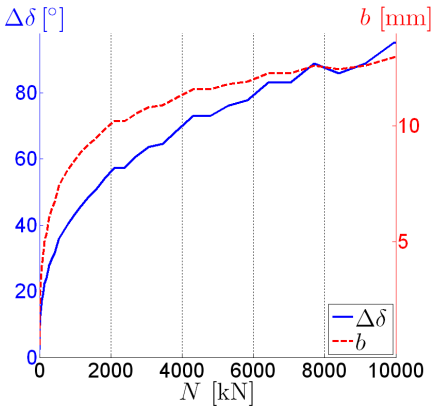


(b) Maximum normal pressure $p_{n,max}$ and a/b ratio.

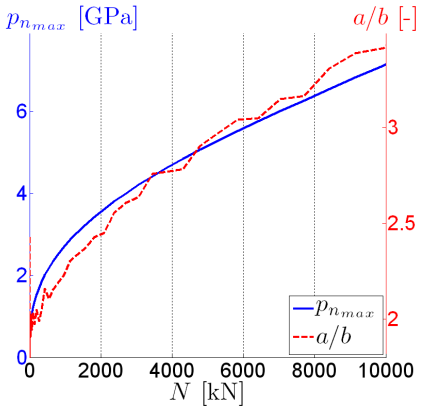


(c) Contact patch contours (symmetric about x = 0 and s = 0) for N = 80, 800 and 3200 kN corresponding to $\Delta\delta \approx$ 40, 80, 110º.

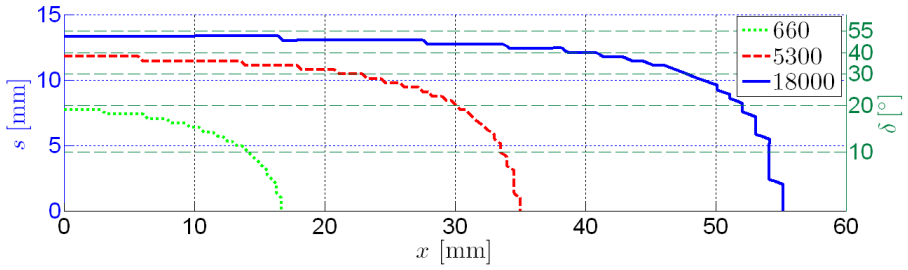
Figure 5.52. Results of frictionless compression test case “circ. l.” calculated with FEM.



(a) Total contact angle variation $\Delta\delta$ and lateral dimension of contact patch b.

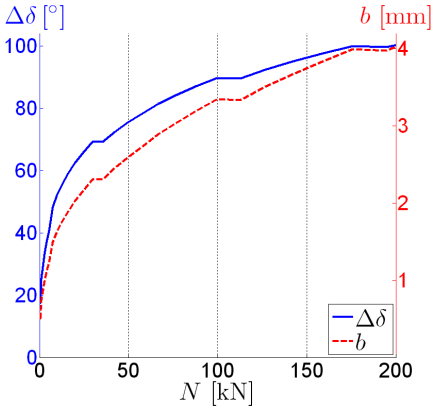


(b) Maximum normal pressure $p_{n,max}$ and a/b ratio.

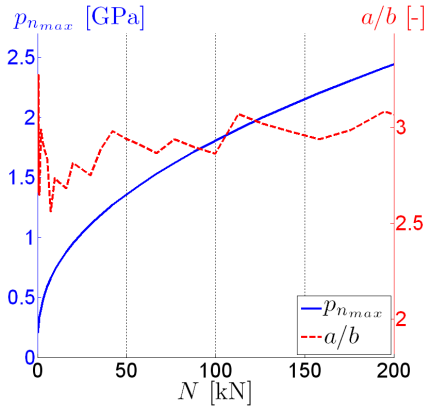


(c) Contact patch contours (symmetric about x = 0 and s = 0) for N = 660, 5300 and 18000 kN corresponding to $\Delta\delta \approx 40, 80, 100^{o}$.

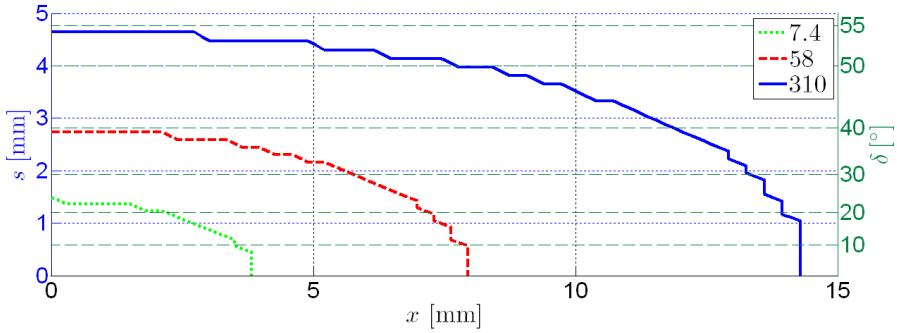
Figure 5.53. Results of frictionless compression test case “el. wide” calculated with FEM.



(a) Total contact angle variation $\Delta\delta$ and lateral dimension of contact patch b.



(b) Maximum normal pressure $p_{n,max}$ and a/b ratio.



(c) Contact patch contours (symmetric about $x = 0$ and $s = 0$) for N = 7.4, 58 and 310 kN corresponding to $\Delta\delta \approx 50, 80, 110$º.

Figure 5.54. Results of frictionless compression test case "el. tall." calculated with FEM.

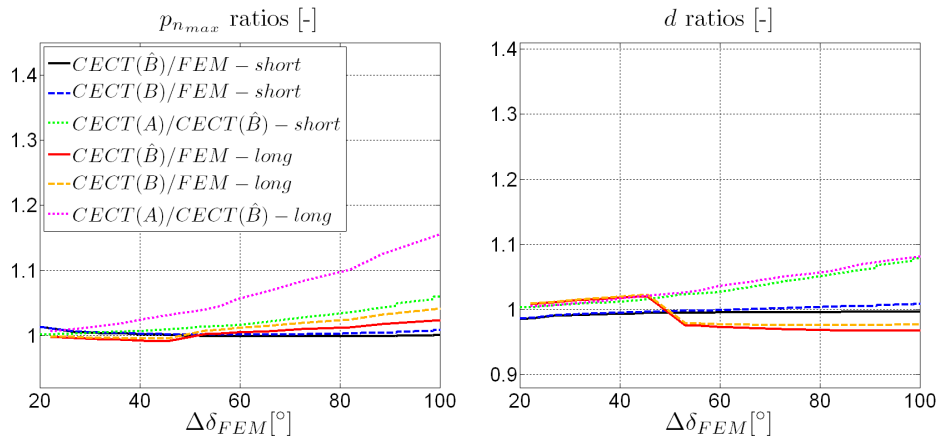
Comparison of FEM with approximate ICs

Here a comparison is presented of results for the normal contact problem computed with FEM and with CECT using A_{nn}, B_{nn} and $\hat{B}_{nn}$ ICs. The ratios of different results are plotted together for test cases "circ.*" in Figure 5.55 on one hand, and for test cases "el.*" in Figure 5.56 on the other.

Using the total contact angle variation of the solutions computed with FEM $\Delta\delta_{FEM}$ in the abscissa axes, the cases "circ.*" on the one hand and the cases "el.*" on the other are compared at equal angle variations rather than at equal normal loads N. For each $\Delta\delta_{FEM}$, the results of different runs (FEM and CECT

with different ICs) for one test case correspond to a same normal load N, different from the load N used in the other test cases. That is, there is a different $\Delta\delta_{FEM}$ versus N correspondence for each test case. The ratio values show the agreement between different runs, with a ratio value of 1 for precise agreement. The FEM results are considered as the reference, although their precision (mostly that of the dimensions of the contact patch) is limited by the mesh resolution for small $\Delta\delta_{FEM}$ values. Also, the approach values computed with CECT do not converge to the FEM values for low $\Delta\delta_{FEM}$ values in all cases, due to the global or "rigid body" displacements contained in the FEM results.

Referring to test cases "circ.*", the curves *CECT(A)*/*CECT(* $\hat{B}$ *)* of Figure 5.55, with the ratios of the results computed with CECT using the A_{nn} and $\hat{B}_{nn}$ ICs, show that the errors in the IC_{nn} propagate more strongly to contact results in case "circ. long". At the same time, the maximum normal pressure and contact area values obtained with the B_{nn} ICs have more discrepancies with the FEM results in case "long" than in case "short". The reason for the increased errors in case "long" compared to case "short" is discussed in the next subsection. The step change around 45–55º in curves *CECT(* $\hat{B}$ *)*/*FEM* and *CECT(B)*/*FEM* of the approach for case "circ. long" in Figure 5.55b is due to performing the calculations with different FE models for low and for high $\Delta\delta_{FEM}$ in this case, which changed the FEM global displacements slightly.

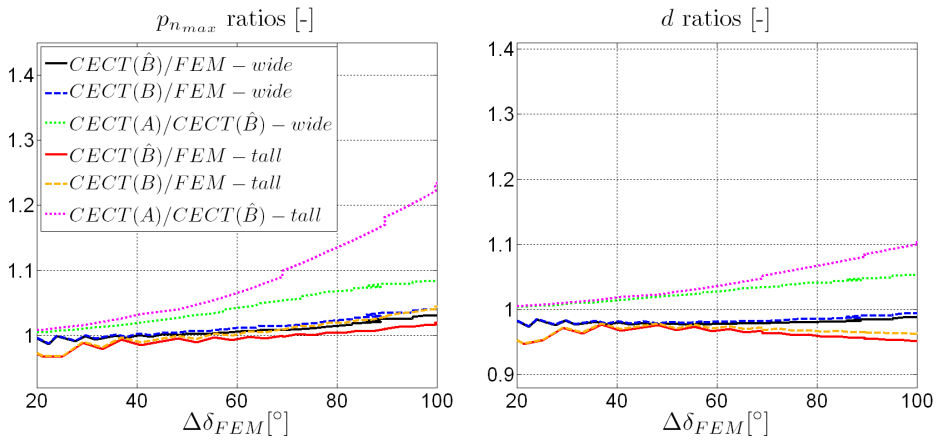


(a) $p_{n,max}$. (b) Approach d.

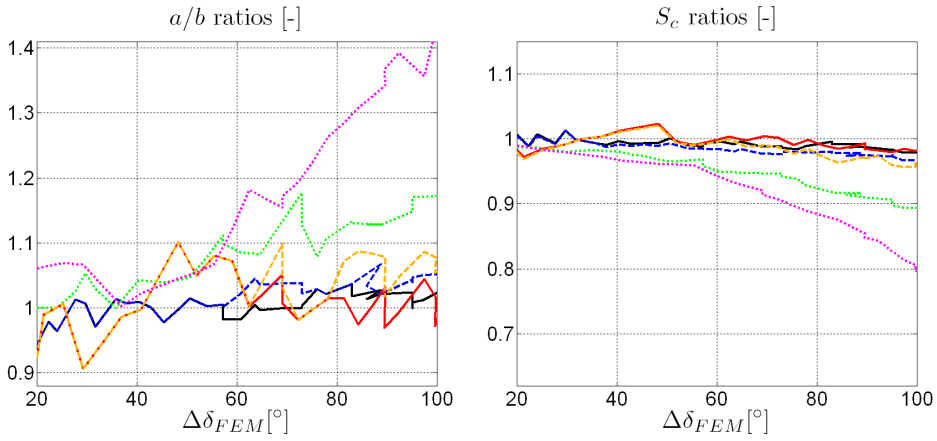
(c) a/b. (d) Contact area S_c.

Figure 5.55. Ratios of results of normal contact problem for test cases “circ.*”.

With regard to test cases “el.*”, the curves *CECT(A)/CECT(* $\hat{B}$ *)* of Figure 5.56 indicate that the use of the A_{nn} ICs lead to higher errors in case “tall”, where the contact angle variation is faster around the central contact point.



(a) $p_{n,max}$. (b) Approach d.



(c) a/b. (d) Contact area S_c.

Figure 5.56. Ratios of results of normal contact problem for test cases “el.*”.

For both the cases “circ.*” and “el.*”, the A_{nn} provide good accuracy in the normal contact problem for $\Delta\delta$ values below 40º, with errors limited below 5% in the computed representative quantities. An exception concerns the aspect ratios a/b, where the accuracy of the results is limited more by the mesh resolution. The good accuracy provided by the A_{nn} for low $\Delta\delta$ is in agreement with the findings of several authors. In [Chen 1988] and [Alonso 2008], this conclusion was reached for the case of frictionless 2D contact, and in [Fagan 2001] for frictionless 2D cylindrical and spherical contact. In [Paul 1979] and [Li Z 2002], it is stated that using the Boussinesq IF of the half-space for the

normal contact problem is a reasonable good approximation for the study of wheel–rail conformal contact problems.

According to the *CECT(*$\hat{B}$*)/FEM* and *CECT(B)/FEM* curves in Figure 5.55 and Figure 5.56, the B_{nn} and $\hat{B}_{nn}$ ICs provide good accuracy in all the test cases, at least up to $\Delta\delta$ values of 100º. The $\hat{B}_{nn}$ are more accurate in these test cases, except for the approach values computed in cases "circ. long" and "el. tall", for which the values obtained with $\hat{B}_{nn}$ ICs diverge slightly more from those computed with FEM than the values obtained with B_{nn} ICs.

In addition to the contact patch a/b ratio and the form of the contact angle variation, the pressure distribution across the contact patch is another factor that influences the performance of the approximated ICs in the solution of contact problems. This is discussed in a later subsection.

Effects of the contact patch length/width ratio

The sensitivity of the contact results to errors in the ICs tends to be higher with more elongated contact patches, i.e. with higher a/b ratios. A set of non-conformal, Hertzian cases is studied here to explain this. The resulting contact patches of these cases have different a/b ratios, ranging from 0.2 to 5.0. In each case, the contact patch is divided in longitudinal strips, and the contribution of the normal pressures acting on each strip to the elastic normal displacement difference u_n of a point P is computed. The contribution of a strip S located at lateral position s is denoted as u_{PnSn}. Figure 5.57 shows the results for the point P located at position $(x, s) = (0, -b)$ on the border of the contact ellipse, for contact ellipses with different a/b ratios. The unitary contributions are plotted in normalized form so that the area beneath each curve is constant.

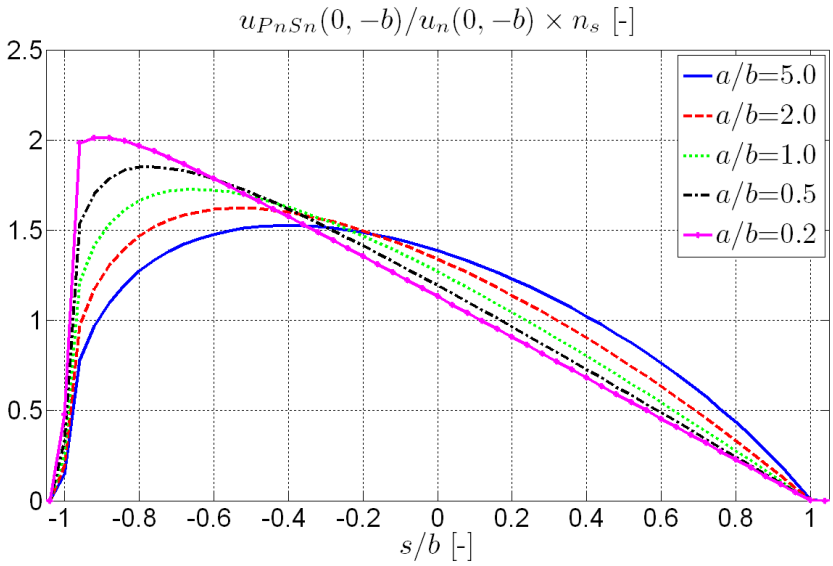


Figure 5.57. Unitary contributions of the longitudinal strips of the contact patch to the normal displacement difference u_n at point (x, s) = (0, $-b$), for Hertzian cases with different a/b semi-axis ratios.

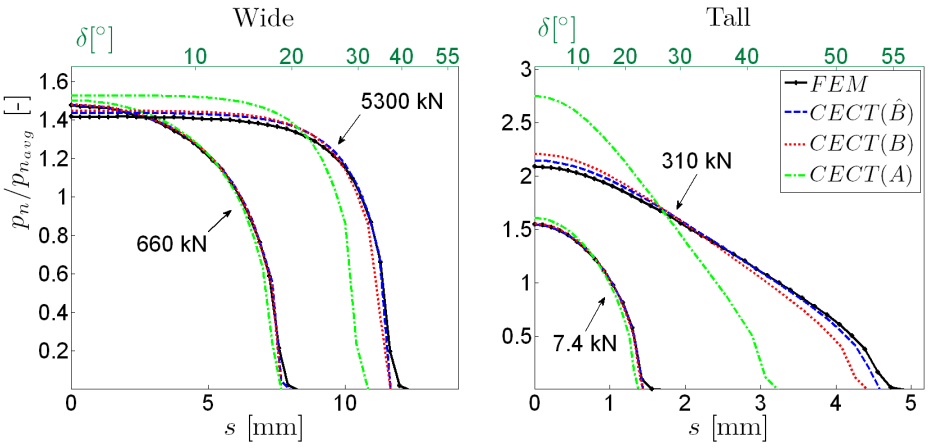
As may be seen in Figure 5.57, the unitary contributions of the different strips are more spread across the contact patch in the case of contact patches more elongated in the longitudinal direction. Similar observations may be made with points P at other positions. The deviations of the true ICs of non-planar solids from those of the half-space tend to be stronger in the lateral direction, with the change in the surface orientation, as illustrated with the different variants of the B approximated ICs. According to this, it is understood that having the contributions to the elastic field more spread in the lateral direction in the contact patch tends to lead to higher errors in the elastic field and hence in the contact solution computed with approximate ICs, by having more weight ICs with higher changes in surface orientation, i.e. higher non-flatness.

Effects of the pressure distribution

The sensitivity of contact results to errors in the ICs also depends on the stress distributions in the contact patch. The normal pressure profiles at x = 0 are plotted in Figure 5.58 for the two "el.*" test cases for two normal loads. The curves are non-dimensionalised with a measure of the average normal pressure

for each test case and load, computed as the load N divided by the contact area obtained from the FEM analysis.

In Figure 5.58 the pressure profiles are plotted in the lateral direction, in which they deviate from the Hertzian distribution. The distribution evolves into a more uniform distribution for test case "wide", and to a more peaky distribution for test case "tall". This is attributed to the geometries of the two cases, deviating from circular cross-sections. For the "wide" test case, the undeformed distance varies less at the centre of the contact and more rapidly towards the sides. The opposite occurs in the "tall" test case. The pressure profiles maintain shapes close to Hertzian in the longitudinal direction (not shown), in these two test cases with zero yaw angle, becoming just slightly more peaky as the load increases.



(a) Test case "el. wide". (b) Test case "el. tall".

Figure 5.58. Non-dimensional normal pressure profiles along s at $x = 0$ for test cases "el.*", for normal loads corresponding to $\Delta\delta \approx 40$, 80º in test case "wide" and 40, 110º in test case "tall". Curves symmetric about $s = 0$.

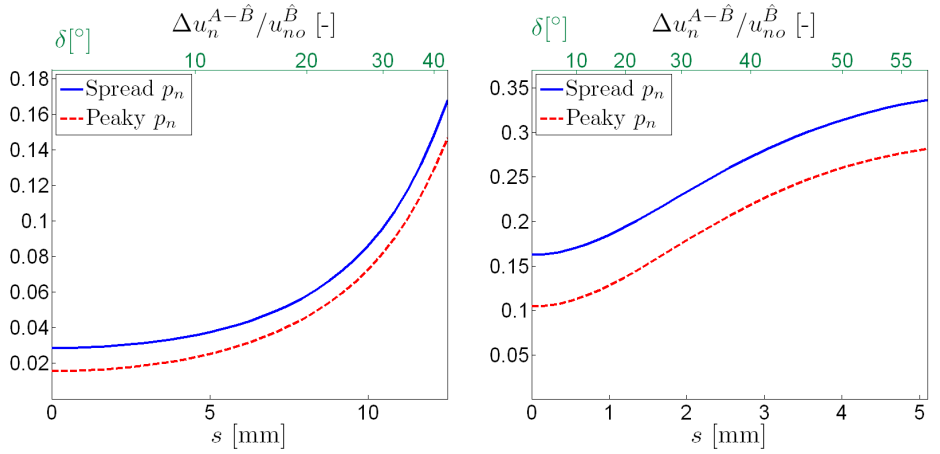
The greater spreading of normal pressures in the contact patch in test case "wide" causes an increased sensitivity to errors in the ICs. To show this, normal elastic displacement differences u_n are computed with the A_{nn} and $\hat{B}_{nn}$ ICs and with a common pressure profile for each test case. The two pressure distributions that are considered are those computed with CECT and $\hat{B}_{nn}$ ICs, corresponding to the highest load levels plotted in Figure 5.58. These are

labelled as "spread p_n" for test case "wide", and "peaky p_n" for test case "tall". The pressure profiles are mapped into the contact patches computed in the same way with CECT and $\hat{B}_{nn}$ ICs for test case "wide" and N = 5300 kN ($\Delta\delta \approx 80^o$) on the one hand, and for test case "tall" and N = 310 kN ($\Delta\delta \approx 110^o$) on the other.

Figure 5.59 shows a sample of the differences between the u_n-s computed with two pressure distributions on the geometry of each test case with two sets of IC_{nn} ICs. These differences between the u_n-s are measured in non-dimensionalised form as

$$\Delta u_n^{A-\hat{B}}(x,s) = \left(u_n^{A}(x,s) - u_n^{\hat{B}}(x,s)\right) \Big/ \max_{x,s} u_n^{\hat{B}}(x,s) \tag{5.81}$$

The superindices in Eq. (5.81) refer to the IC_{nn} with which each u_n are computed.



(a) Test case "el. wide". (b) Test case "el. tall".

Figure 5.59. Non-dimensional differences of normal displacements along s at x = 0 computed with A_{nn} and $\hat{B}_{nn}$ ICs and different normal pressure distributions. Curves symmetric about s = 0.

Comparing the test cases, bigger differences $\Delta u_n^{A-\hat{B}}$ are observed for case "tall", with a more rapid contact angle variation, than for case "wide". This is in line with the higher errors with the use of the A_{nn} evidenced in Figure 5.56 for case

"tall". The *s*-values considered also encompass a bigger $\Delta\delta$ for case "tall" than for case "wide" ($\Delta\delta \approx 110^o$ resp. 80^o). Coming to the point of this comparison, further differences are found comparing the pressure distributions in each subfigure. In each geometry, the differences $\Delta u_n^{A-\hat{B}}$ are higher with the more spread pressure profile. This proves that for the same errors in the ICs, more spread pressure distributions lead to higher errors in the elastic displacements and hence in the computed contact solution. This is easily understood as with more spread pressure profiles the contributions to the elastic field are more spread in the lateral direction of the contact patch, as it happens with more elongated contact patches, cf. Figure 5.57.

As a conclusion, in this section it has been seen how the geometry of the contact surface, the *a*/*b* ratio and the pressure distribution across the contact patch may affect the precision of the contact results computed with approximate ICs. The errors contained in the A_{nn} and B_{nn} are higher in zones with faster angle variations, and consequently the errors obtained in the solution of the contact problem are higher with geometries in which the contact angle variation is faster around the central contact point (i.e., the zone in the contact patch with the greatest contributions to the total load). It has been seen as well that a more elongated contact patch, and a more uniform pressure distribution across the contact patch increases the sensitivity to errors in the ICs, by putting more weight in the contributions of ICs associated with higher changes in surface orientation, which normally contain a higher error.

5.2.3.7.2 Tangential contact

To assess the precision of different variants of approximated ICs in the tangential part of the contact problem, the frictional contact of a concave wheel on a convex rail is considered here. Three different values of constant coefficients of friction are used, namely μ = 0.20, 0.40 and 0.60. A geometry similar to that described in Figure 5.25 and Table 5.4 is considered (equal to the "circ. s" geometry described in Table 5.9), except that the circular cross-sectional profile of the wheel is replaced with an elliptical one with semi-axes b_{el} and d_{el} of 10.15 and 10.25 respectively, close to the original. This modifies the shape of the wheel surface only very slightly, such that the C_{ij} ICs presented in §5.2.3.6 continue to be valid. The change does however alter significantly the

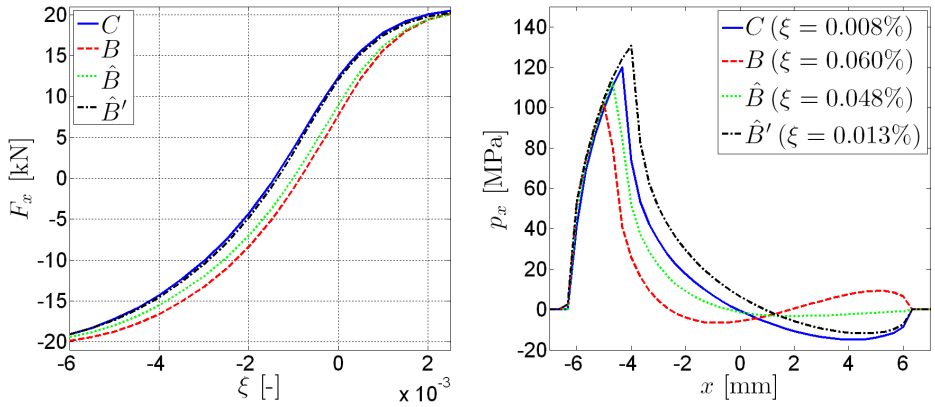
normal undeformed distance to the rail, making the normal pressure distribution more evenly distributed in the lateral direction.

Figure 5.60 summarizes results obtained with μ = 0.20 for the laterally symmetric rolling contact with a total normal load N of 100 kN, zero yaw angle and zero lateral rigid slip at the central contact point. Results are shown for B, $\hat{B}$ and $\hat{B}'$ approximated ICs, along with those obtained with the C ICs for geometry 2 (cf. Figure 5.32 and Table 5.6) that are considered the reference. These results may be compared with the ones shown in Fig. 14 of [Blanco-Lorenzo 2018], for a similar geometry and different N and μ.

Figure 5.60a shows the longitudinal creep force curves obtained with the different sets of ICs. Significant offsets are shown for the B and $\hat{B}$ ICs, which are attributed to the differential stiffness effect between the convex and concave bodies. This is discussed further in §6.2.2.2. This offset is well compensated by the adjustment between the $\hat{B}'_{xn}$ and the C_{xn}, cf. Figure 5.42. The adjustment parameters for the $\hat{B}'_{xn}$ used here are those listed in Table 5.8 for geometry 2. These yield a creep force curve that is in close correspondence to the reference obtained with the C ICs.

The patterns of longitudinal tangential stress p_x computed with the $\hat{B}'$ ICs are also improved, as illustrated in Figure 5.60b. This figure shows the p_x profiles along the centreline of the contact patch obtained with different sets of ICs, for the same value of longitudinal traction ratio $f_x = F_x / \mu N$. The longitudinal creepage values ξ at which this f_x is obtained are indicated in the legend of the graph for each set of ICs.

For higher values of μ, the differences between the results obtained in the tangential part of the contact problem with the different sets of ICs for this rolling contact case diminish, as may be seen in Figure 5.63 for μ = 0.40. This happens because the influence of the normal pressures on the tangential contact problem is proportionally lower with higher μ and tangential stresses, and hence the role of the IC_{xn} gets less important.



(a) Longitudinal creepage–creep force curve. (b) p_x along x at $s = 0$, for $f_x = 0.65$.

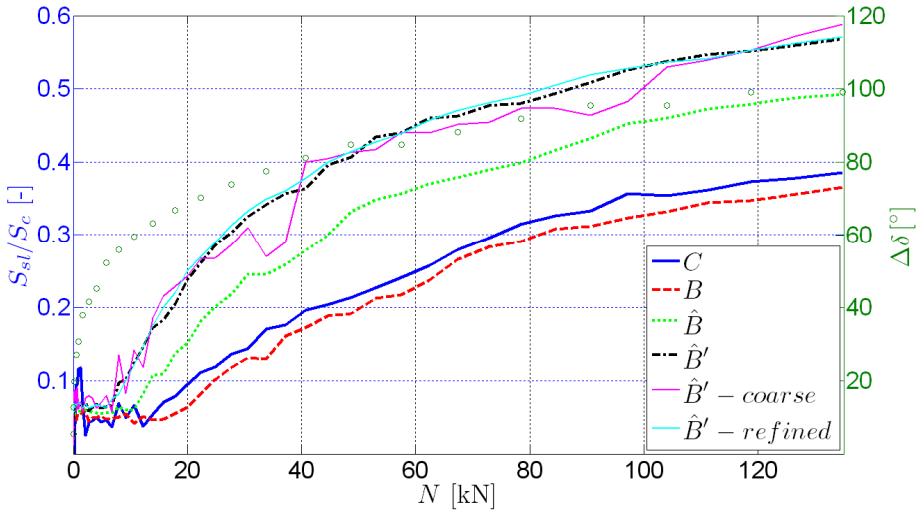
Figure 5.60. Performance of approximated ICs in symmetric conformal rolling contact case with $\mu = 0.20$.

Previously, Figure 5.46 showed that the $\hat{B}$ and $\hat{B}$' ICs provide a worse approximation of the C_{sn} than the B ICs. As the C_{sn} are greater in magnitude than the C_{xn}, this may affect the tangential part of the contact problem significantly, especially in problems with larger rigid slips in s direction in relation to those in x direction. A problem where this happens is the static normal compression between a convex and a concave body with zero shift in longitudinal direction. This is computed for the same geometry and coefficients of friction as in the previous rolling contact case.

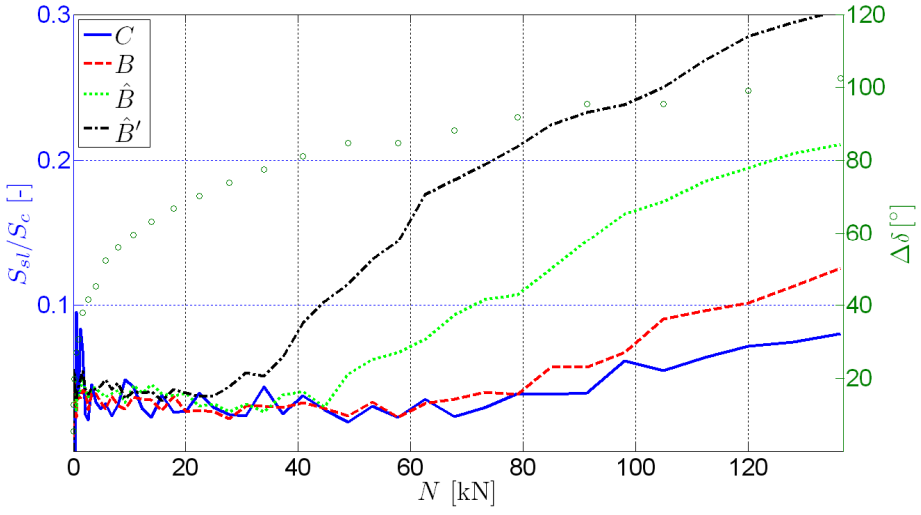
The result that is shown in Figure 5.61 is the ratio of area in sliding S_{sl} to total contact area S_c, as a function of the total compressive load N (sum of the resultants of the normal and tangential stresses in the contact patch), monotonically increasing, for $\mu = 0.40$ and $\mu = 0.60$. With $\mu = 0.20$, the differences between the sliding areas obtained with the different sets of ICs are attenuated in this case due to the higher saturation level.

The $\Delta\delta$ spanned by the contact patch as a function of the load (as computed with the C ICs) is shown in each subfigure with circle markers. Clearly, the B ICs provide the results closest to the C ICs, while the adjusted $\hat{B}$' ICs provide the worst results. This suggests that the presented adjustment for the differential stiffness effect, intended for the improvement of the B_{xn}, should not be applied in the B_{sn}. Furthermore, the B_{sn} provide a better approximation of the C_{sn} than

the $\hat{B}_{sn}$, and consequently more precise results are obtained in the tangential part of this contact problem with the B_{sn} than with the $\hat{B}_{sn}$.



(a) μ = 0.40.



(b) μ = 0.60.

Figure 5.61. Comparison of ratios of sliding areas S_{sl} to total contact areas S_c obtained with *B* and *C* ICs in a symmetric conformal static compression contact case.

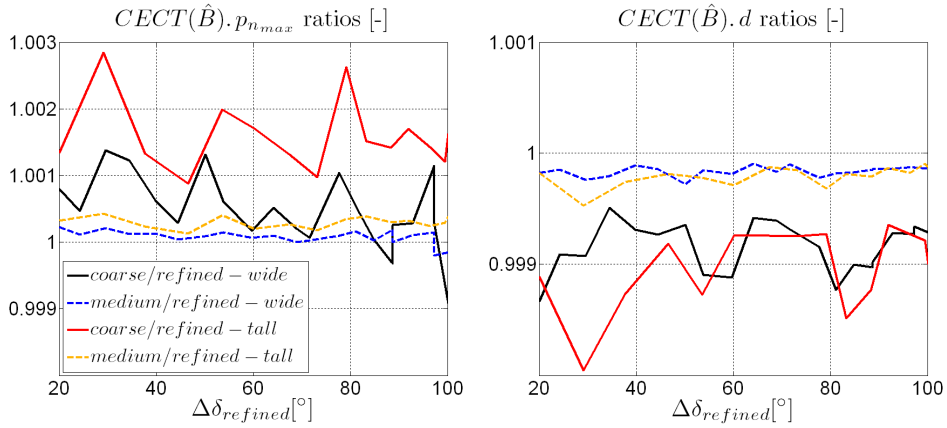
5.2.3.7.3 Effect of mesh refinement on the numerical accuracy of FEM results

Refined meshes have been used in the computations shown previously in this section, in order to avoid the errors associated to the mesh resolution being significant in relation to the differences between the different sets of ICs. The contact calculations with CECT have been performed with meshes encompassing about 50 elements inside the contact in the lateral direction, and 40 in the longitudinal, except for the compression cases at lower loads. The FEM contact models have been constructed with finer meshes.

Some of the computations with CECT have been repeated with different mesh refinements, in order to demonstrate numerical accuracy of the presented results with grid refinement. The numerical accuracy of the FEM models has been discussed in §4.2.2. The next figures show some comparisons of results obtained with CECT with different meshes. The results termed as "coarse" and "refined" in these figures have been obtained with meshes having about half and twice as many elements in each of the two surface directions as those of the original calculations. The original meshes are termed as "medium". In all the results shown below, it is verified that the numerical errors related to mesh refinement are small in relation to the differences between the different sets of ICs.

Figure 5.62 shows the ratios of results of the normal contact problem for test cases "el.*" computed with CECT and *B* ICs with different mesh refinements. The quantity $\Delta\delta_{refined}$ of the abscissa axes of this figure is the total contact angle variation of the solutions computed with the most refined mesh in each case. The ratios shown here may be compared with the ratios shown in Figure 5.56 for results obtained with FEM and CECT with different sets of ICs. As expected, the most affected quantities by the mesh refinement are the ones that have their accuracy directly determined by the number of elements inside the contact in each direction, namely the contact area S_c and mostly the a/b ratio. The noise seen in the curves "medium/refined" in Figure 5.62 c and d has roughly the same amplitude as the noise that can be seen in the curves of Figure 5.56 c and d, but their mean is around unity and do not show any appreciable trend which could distort the comparisons between the results obtained with the different sets of ICs. In the other results shown, the differences between the

different meshes are much smaller. The errors produced with the coarser meshes are lower in case “wide” than in case “tall”.



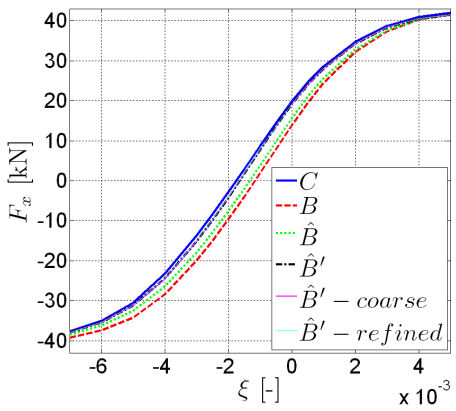
(a) $p_{n,max}$. (b) Approach d.

(c) a/b. (d) Contact area S_c.

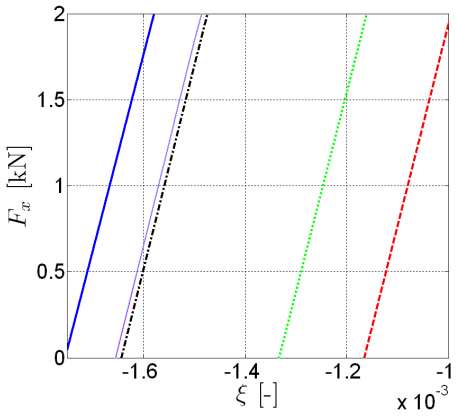
Figure 5.62. Ratios of results of normal contact problem for test cases “el.*” computed with CECT and $\hat{B}$ ICs with different mesh refinements.

Figure 5.63 shows mesh sensitivity results for the tangential part of the contact problem in a similar conformal rolling contact case as that considered in Figure 5.60. The only difference from that case is the value of μ, which is now set to 0.40. Here results using the $\hat{B}'$ ICs have been computed with different mesh densities. As can be seen in the different graphs, the differences between the results obtained with different meshes and the same set of $\hat{B}'$ ICs are much

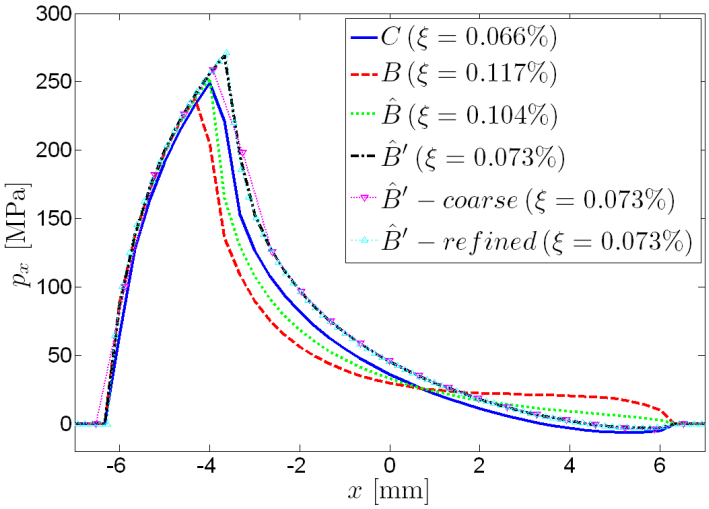
lower than the differences between the different sets of ICs. This is also verified in the frictional static compression case dealt with in Figure 5.61a, where also results computed with the same set of $\hat{B}'$ ICs and different mesh densities have been included. Here the differences between the curve corresponding to the coarse mesh and the other two are due to the lower mesh resolution of the coarse mesh, like the noise in the a/b and S_c curves in Figure 5.62 c and d.



(a) Longitudinal creepage–creep force curve.



(b) Zoomed view of (a).



(c) p_x along x at $s = 0$, for $f_x = 0.65$

Figure 5.63. Performance of approximated ICs in symmetric conformal rolling contact case with $\mu = 0.40$. Included results with different mesh refinements.

In all the cases, the differences between the results obtained with the medium and refined meshes are very small both in the normal and in the tangential parts of the contact problem.

Chapter 6

Characteristics of wheel–rail conformal contact

Chapter summary

This chapter is devoted to the analysis of the particular characteristics of wheel–rail conformal contact. This is based mostly on numerical analyses with the detailed contact models presented in Chapter 4, namely CECT and FEM contact models. The solutions obtained with CECT and FEM are contrasted with each other in different cases to validate the models for conformal wheel–rail contact analysis. A geometrical analysis of the undeformed distance between two non-planar conforming surfaces is also presented. Different features that conformity brings about are pinpointed and contrasted with the characteristics of non-conformal contact. The contact mechanics analyses cover 3D frictional contact. In contrast, most of the past published work on conformal contact mechanics deals with 2D contact, and focuses on the normal part of the contact problem.

6.1 Analysis of the undeformed distance

In this section a geometrical analysis is carried out of the undeformed distance h between the surfaces of a prismatic body (designated as the rail) and a body of revolution (designated as the wheel), which touch at one or more rigid contact points, providing some analytical approximations for h. The analysis can be easily extended to consider both bodies as solids of revolution. Both the case of having the axis of revolution of the wheel in the YZ plane (i.e. with zero yaw angle), and the case of non-zero yaw angle will be considered. With this analysis, it is intended to provide insight on the geometrical effects of the non-flatness of the contact surface on the undeformed distance, which are manifested in conformal contact, and to show the influencing parameters that determine their relevance.

Figure 6.1 depicts the profiles of a convex rail and a concave wheel, contacting at a single point. This configuration is used as the baseline for the analyses presented in this section. A local contact Cartesian coordinate system is defined, with its origin in the contact point, and with its three axes, (x_c, y_c, z_c), defined as follows:

- x_c: perpendicular to the plane of the figure, oriented along the longitudinal direction.
- y_c: tangent to the surfaces of both bodies at the contact point, contained in the plane perpendicular to the longitudinal direction. It defines the lateral direction.
- z_c: perpendicular to x_c and y_c, and normal to the surfaces of both contacting bodies at the contact point.

This system is analogous to the {x, y, z} system used in planar contact (cf. §1.2). Additionally, the (x, s, n) local contact curvilinear coordinate system is used, as defined in §4.1.1.1 (cf. Figure 4.2). The x and x_c axes are parallel, and at the contact point, the s, n axes correspond to the y_c, z_c axes, respectively.

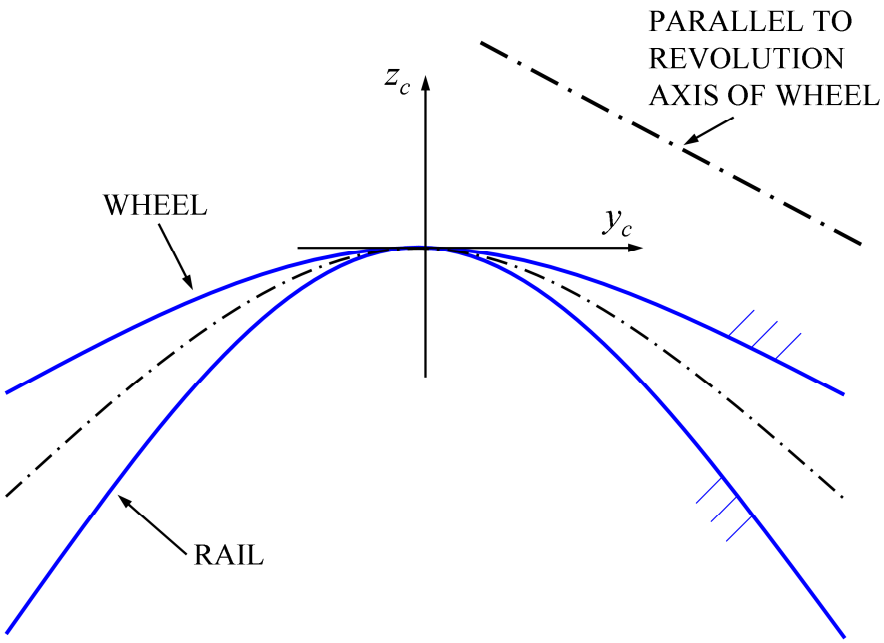


Figure 6.1. Non-planar contact surfaces and (x_c, y_c, z_c) coordinate system.

Figure 6.1 shows also the non-planar potential contact surface, in chain line between the wheel and rail profiles, and a parallel to the wheel axis of revolution. The latter is contained in the plane of the figure in the case of zero yaw angle only, and is not aligned in general with the y_c nor with the z_c axis. The wheel roll angle is assumed to be zero, without loss of generality.

The following assumptions are made:

- The distances between the surfaces of both bodies are much smaller than the dimensions in the x_c and y_c directions of the analysis domain.
- The surfaces of both bodies are smooth.
- The surfaces of both contacting bodies are not conforming in the longitudinal direction. However, they may be closely conforming in the y_c-z_c plane. Consequently, the surfaces of both bodies cannot be assumed to be nearly planar in the analysis domain around the contact point, as they could in punctual contact.

- The variation of the rolling radius of the wheel in the analysis domain is much smaller than the rolling radius of the wheel. Therefore, the wheel rolling radius can be taken as constant in this domain.

6.1.1 Zero yaw angle

Figure 6.2 shows a concave wheel on a convex rail, with zero yaw angle between them. The cross-section of the rail is shown, together with two wheel cross-sections with planes perpendicular to x_c, at $x_c = 0$ and $x_c = x$. The following nomenclature is used:

- Δz_0: profile of the distance between both bodies along the z_c axis at $x_c = 0$.
- h_0: profile of the normal distance between both bodies at $x_c = 0$.
- δ, δ_0: contact angle, with subindex 0 referring to the contact point.
- R_{roll}: rolling radius of the wheel at the contact point.

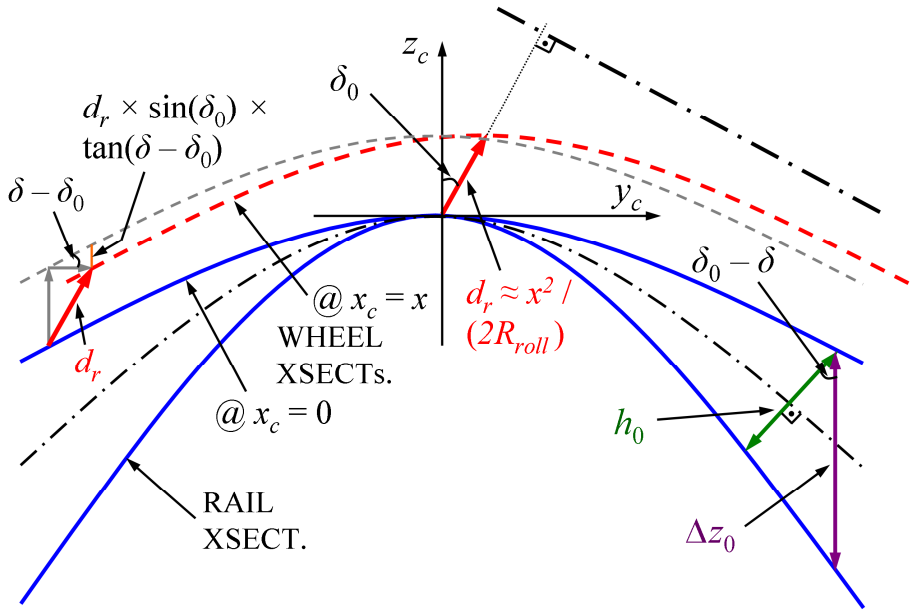


Figure 6.2. Analysis of the wheel–rail undeformed distance with zero yaw angle.

For small x_c / R_{roll} values, the shape of the different cross-sections of the wheel do not change much. So, the cross-sections of the wheel at different $x_c = x$

positions can be obtained by simply translating the principal cross-section at $x_c = 0$ by a distance d_r, almost equal to $x^2 / 2R_{roll}$. This translation, which takes place in the radial wheel direction (i.e. perpendicular to the wheel axis of revolution), can be decomposed into its components along the y_c and z_c axes, $d_y = d_r \sin \delta_0$ and $d_z = d_r \cos \delta_0$, as indicated by the triangle formed by the grey and red arrows in the left side of Figure 6.2. In this way, this translation can be thought of as being composed of two steps: the first one along the z_c axis, from the wheel cross-section at $x_c = 0$ (in blue solid line in Figure 6.2) to the profile in dashed grey line in the figure, and the second one along the y_c axis, from this profile to the final position of the wheel cross-section at $x_c = x$ (dashed red line in the figure). Therefore, if zw_0 (y_c) is the function of the wheel profile at $x_c = 0$, the wheel cross-section at different x_c positions can be approximated by the function zw given in Eq. (6.1). zw_0 and zw are measured along the z_c axis.

$$zw\left(x_c, y_c\right) = zw_0\left(y_c - \frac{\sin\delta_0}{2R_{roll}} x_c^{\,2}\right) + \frac{\cos\delta_0}{2R_{roll}} x_c^{\,2} \tag{6.1}$$

This already defines the surface of the wheel in (x_c, y_c) coordinates. It is possible to compute the surface of a yawed wheel in a similar way, with an appropriate coordinate transformation. On the other hand, the surface of the rail is determined by its cross section zr (y_c), which does not vary along x_c. A first way of computing Δz, the distance along the z_c axis between both bodies, can be directly subtracting the z_c coordinates of both surfaces:

$$\Delta z_1 (x_c, y_c) = zw (x_c, y_c) - zr (y_c) \tag{6.2}$$

Next, the contributions to Δz of the components along z_c and y_c of the translation of the wheel cross-section from its position at $x_c = 0$ to that at $x_c = x$ will be considered. The contribution of the component along z_c is uniform all across the y_c direction, and is equal to the projection of d_r on z_c. On the contrary, the contribution of the component along y_c is variable in that direction, and is equal to the projection of d_r on y_c multiplied by the tangent of the difference of contact angles δ between each lateral position of the wheel principal cross-section and the contact point. This is indicated by the vertical orange line in the small triangle in the left side of Figure 6.2. A second way to compute Δz is to

start from Δz_0, and add the contributions of these components. This leads to Eq. (6.3):

$$\Delta z_2\left(x_c, y_c\right) = \Delta z_0\left(y_c\right) + \frac{\cos\delta_0}{2R_{roll}} x_c^{\,2} - \frac{\sin\delta_0}{2R_{roll}} x_c^{\,2} \tan\left(\delta\left(y_c\right) - \delta_0\right) \tag{6.3}$$

Eq. (6.3) brings to light the role of the varying contact angle δ with a non-zero δ_0, even when computing distances Δz in a constant direction along z_c. This highlights the importance of carrying out a non-planar geometric analysis for the precise computation of undeformed surface distances in non-planar contacts, especially with high δ_0.

The assumption of a constant R_{roll} may be dropped in the above analysis, by taking the corresponding value R_{roll} (y_c) at each lateral position of the wheel principal cross-section. If this is done, the change of the wheel cross-section from $x_c = 0$ to $x_c = x$ will no longer be a simple translation: the shape of the cross-section will also change, as each point in the wheel profile will translate a different distance d_r (y_c) along the radial direction. Anyway, Eqs. (6.1) and (6.3) remain valid, taking into account the smallness of the distances between the two bodies and the smoothness of their surfaces.

Lastly, the computation of the normal distance h is considered; that is, the distance along the local normal direction at each position. As it is well known, in punctual wheel–rail contact, with a constant contact angle δ_0 in the analysis domain, Eq. (6.4) is applicable:

$$h\left(x_c, y_c\right) = h_0\left(y_c\right) + \frac{\cos\delta_0}{2R_{roll}} x_c^{\,2} = \Delta z\left(x_c, y_c\right) \tag{6.4}$$

According to Eq. (6.4), the effective radius of curvature of the wheel surface in the longitudinal direction, R_{xw}, is equal to R_{roll} / cos δ_0. For surfaces which are closely conforming in the y_c-z_c plane, the same equation may be applied, just changing the contact angle δ_0 by the variable local contact angle δ at each lateral position. This may be most easily visualized by considering two linearized sections of the wheel and rail cross-sections, at two points with different contact angles δ_1 and δ_2, as depicted in Figure 6.3. Therefore, h is computed according to Eq. (6.5):

$$h\left(x_c, y_c\right) = h_0\left(y_c\right) + \frac{\cos\delta}{2R_{roll}} x_c^{\,2} = h_0\left(y_c\right) + \Delta h \tag{6.5}$$

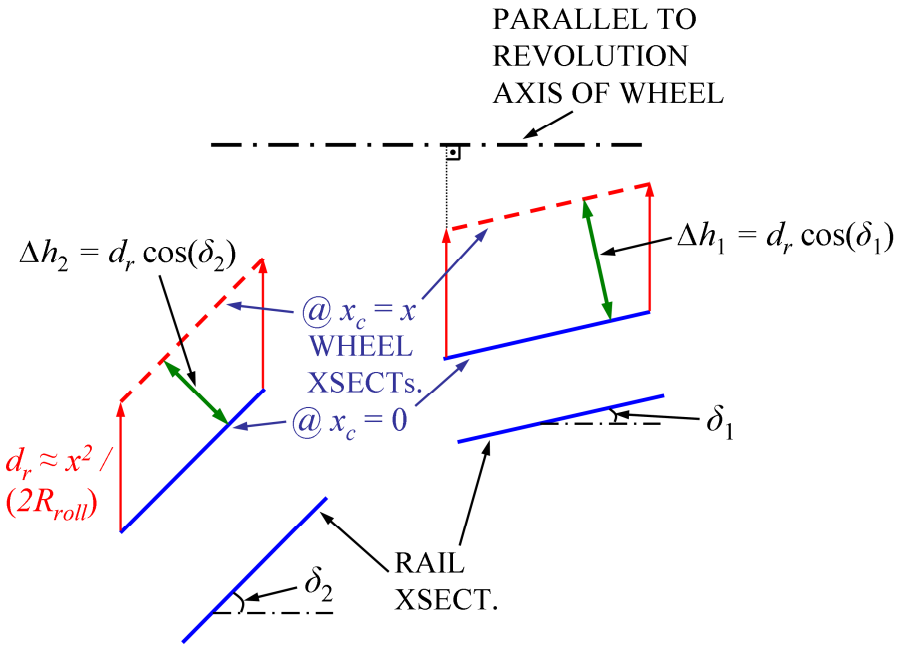


Figure 6.3. Variation of the effective longitudinal curvature of the wheel with the contact angle.

The differences between the vertical and normal distances, Δz and h, may be quantified comparing Eqs. (6.3) and (6.5). On the one hand, the ratio between the first term in the right hand side of both equations is equal to $\cos(\delta_0 - \delta)$, as may be seen in the triangle in the bottom right side of Figure 6.2 containing h_0 and Δz_0 as two of its sides. The ratio between the remaining terms of both equations is the same:

$$\cos\delta_0 - \sin\delta_0 \tan\left(\delta - \delta_0\right) = \frac{\cos\delta_0 \cos\left(\delta - \delta_0\right) - \sin\delta_0 \sin\left(\delta - \delta_0\right)}{\cos\left(\delta - \delta_0\right)}$$

$$= \frac{\cos\delta}{\cos\left(\delta - \delta_0\right)} \rightarrow \frac{\cos\delta}{\cos\delta_0 - \sin\delta_0 \tan\left(\delta - \delta_0\right)} = \cos\left(\delta - \delta_0\right) \tag{6.6}$$

Therefore, the ratio between h and Δz is equal to the cosine of the difference of contact angles[26] between each lateral position of the cross-section of the PCS and the reference δ_0, independently of δ_0. I.e., $h / \Delta z = \cos(\delta - \delta_0)$. For example, for a contact angle difference $\delta - \delta_0 = 20^o$, the difference between h and Δz is just about 6%.

It has to be noted that h values cannot be computed properly from Δz values at $\delta - \delta_0$ close to 90º, due to the $\cos 90^o = 0$ singularity. This problem is avoided in the method for the calculation of h implemented in CECT explained in §4.1.1.2, by calculating the projection on the local n direction of the vector between each pair of wheel–rail homologous points, which is determined with both the Y and Z coordinates of each point. Also, homologous points for the calculation of Δz are defined differently here, as those with the same y_c coordinate instead of s coordinate. This again is valid for $\delta - \delta_0$ values not too close to 90º. Taking into account the smallness of h with respect to the contact dimensions and the smoothness of the surfaces, the difference between both definitions of homologous points is negligible. For the particular case of having both wheel and rail surfaces parallel around a pair of homologous points (as depicted in Figure 6.3), both definitions lead to the same distances.

Eqs. (6.3) and (6.5) use a linearized approximation of the wheel profile at each position y_c, in the range $(y_c - d_r \sin(\delta_0), y_c)$. When the lateral displacement of the wheel profile, $d_r \sin(\delta_0)$, is not small in relation to the lateral curvature radii of the wheel profile R_{sw}, an error due to this linearization becomes appreciable. This error is represented in Figure 6.4 as ε_z at $y_c = 0$. The same three cross-sections of the wheel shown in Figure 6.2 are depicted in Figure 6.4: the cross-section at $x_c = 0$ in solid blue line, the final cross-section at $x_c = x$ (after the d_r

[26] Note that in Figure 6.2 the angular difference $(\delta - \delta_0)$ has been defined both with reference to the wheel profile in the left side of the figure, and to the profile of the PCS in the right side. In order for the above statement to be applicable, the wheel and rail surfaces have to be closely conforming in the y_c–z_c plane, so that the profiles of the wheel, the rail, and the reference contact surfaces are almost merging with one another. In the situation of having a significant contact angle variation in the analysis domain, this does not introduce any new restriction in the analysis with respect to that of the smallness of the distances already considered.

translation) in dashed red line, and an intermediate cross-section at $x_c = x$, after the d_z translation along the z_c axis.

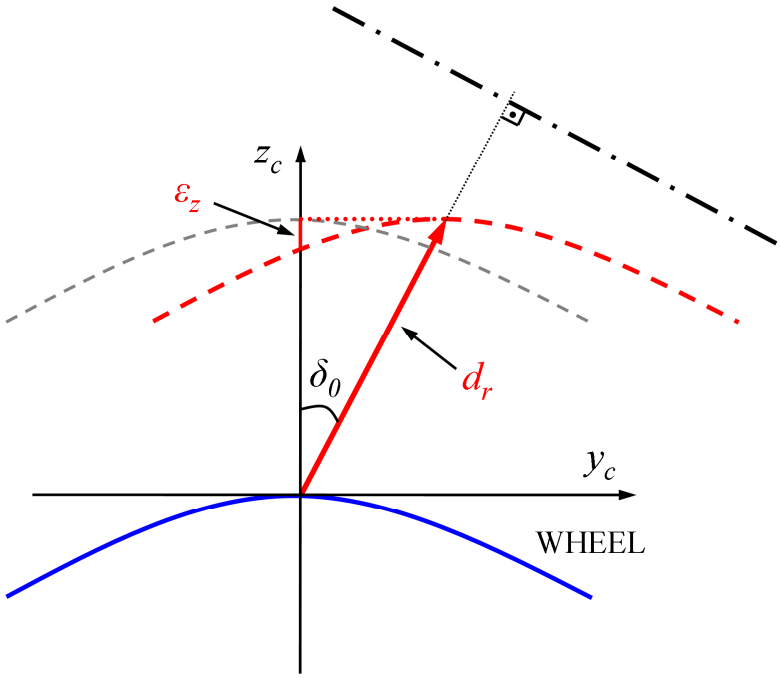


Figure 6.4. Error in linearization of wheel profile.

ε_z may be approximated according to Eq. (6.7). This uses the relation $C_{sw} = 1 / R_{sw} = zw_0'' / (1 + zw_0'^2)^{3/2}$ between the lateral curvature C_{sw} of the wheel profile zw_0 and its derivatives with respect to y_c, and a quadratic approximation of the wheel profile. The first derivative zw_0' is equal to $\tan(\delta - \delta_0)$.

$$\varepsilon_z \approx \frac{C_{sw}}{2}\left(1+\tan^2(\delta-\delta_0)\right)^{3/2}\left(\frac{x_c^2}{2R_{roll}}\sin\delta_0\right)^2 \tag{6.7}$$

The significance of ε_z may be assessed by comparing it with the overall vertical displacement of the wheel cross-section at x_c, $d_z = x_c^2 / 2R_{roll} \times \cos\delta_0$. Its ratio is given in Eq. (6.8). According to this, and assuming that the relevant x_c range changes in proportion to $R_{xw} = R_{roll} / \cos\delta_0$, ε_z may become more relevant for higher δ_0, x_c, and $|R_{xw} / R_{sw}|$ values.

$$\frac{\varepsilon_z}{d_z} = \frac{C_{sw}}{2}\left(1+\tan^2(\delta-\delta_0)\right)^{3/2}\sin\delta_0\tan\delta_0\frac{x_c^{\,2}}{2R_{roll}} \tag{6.8}$$

Next, the results obtained with the described formulae for a particular case are illustrated. The "circ. a." case is considered, with geometric parameters defined in Table 6.1. In this case, the wheel and rail cross-sections are circular, with lateral radii of curvature R_{sw} = –10.5 mm and R_{sr} = 10.0 mm, and δ_0 = 45º. Figure 6.5 shows the contours of vertical and normal distances calculated with the previous formulae, together with the contours of normal undeformed distances calculated with CECT following the procedure explained in §4.1.1.2, which are taken here as the reference result. The contours are shown at levels of 0.05 mm and 0.1 mm for each case.

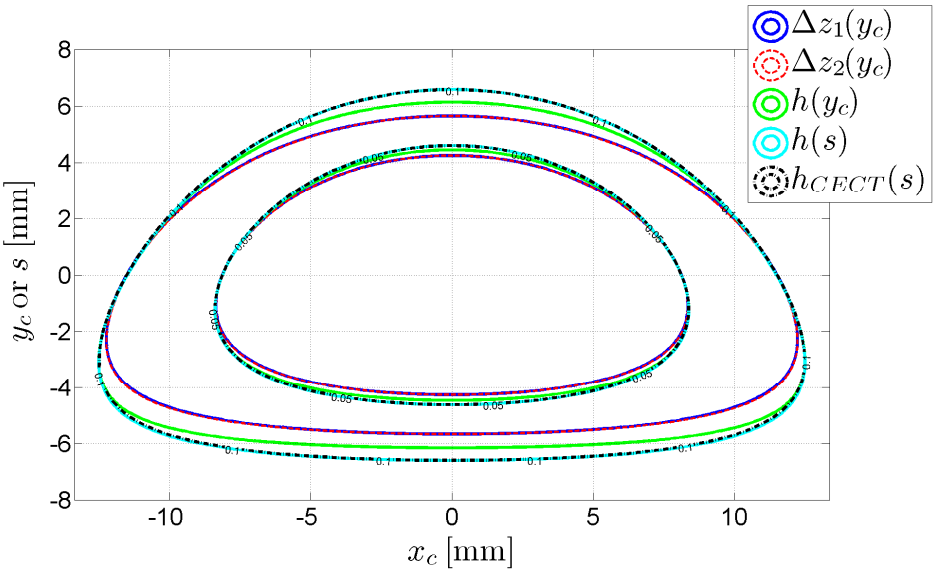


Figure 6.5. Contours of vertical and normal distances for "circ. a." case, with circular conforming sections and contact angle δ_0 of 45º.

As can be seen in the figure, the distance contours are asymmetric in the lateral direction, which is due to the variation of R_{xw} with the contact angle $\delta(y_c)$ (cf. §6.2.1.1). The contours in solid blue and dashed red lines are the vertical distances Δz_1 and Δz_2 computed with Eqs. (6.2) and (6.3), respectively, and are seen to be nearly the same. The contours in solid green lines represent the

normal distances computed with Eq. (6.5) plotted on the x_c–y_c plane. The contour of these normal distances at h = 0.05 mm is quite close to the corresponding contours of vertical distances. Obviously, as the values of $\delta - \delta_0$ increase, it is more important to make the distinction between y_c and s, and between vertical and normal distances. When plotted on the curvilinear coordinate s, the normal distances computed with Eq. (6.5), represented in solid cyan lines in the figure, are seen to be very close to those computed in CECT, represented in black dash-dotted lines.

6.1.2 Non-zero yaw angle

The wheel cross-section translation approach shown for the case of zero yaw angle may also be applied with non-zero, small yaw angles, though the error in the resulting undeformed distance profiles may be appreciable with very closely conforming surfaces or with not so small yaw angles (e.g. ~50 mrad). Figure 6.6 shows the cross-sectional geometry for this case, as a generalization of Figure 6.2. In this case, the final wheel cross-section at $x_c = x$ is shown in orange dashed line. To obtain it, first a radial translation d_r is applied, arriving at the cross-section shown in dashed red line (the same as the final wheel cross-section shown in Figure 6.2, with zero yaw angle), and from it a translation of ψ x_c is applied, being ψ the yaw angle. It is assumed that the yaw rotation of the wheel takes place around an axis parallel to the direction of d_r, so the ψ x_c translation is perpendicular to d_r. The direction of the ψ x_c translation shown in Figure 6.6 corresponds to a positive ψ rotation and x_c coordinate.

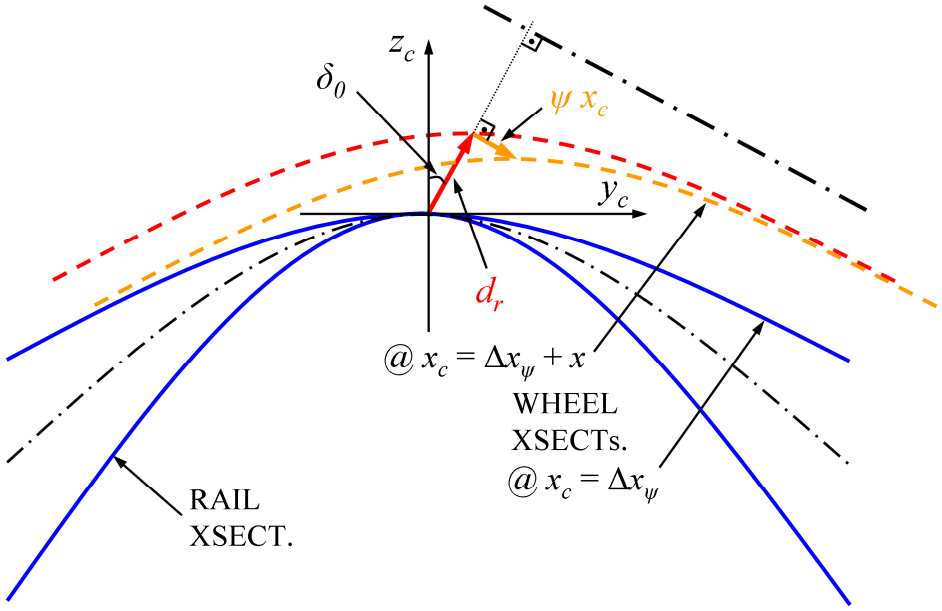


Figure 6.6. Analysis of the wheel–rail undeformed distance with non-zero, small yaw angle.

Adding the components of $\psi\ x_c$ along y_c and z_c to the corresponding components of d_r, Eqs. (6.9), (6.10) and (6.11) are obtained for zw, Δz_2 and h, as an extension of Eqs. (6.1), (6.3) and (6.5), respectively.

$$zw\left(x_c - \Delta x_\psi, y_c\right) = zw_0\left(y_c - \frac{\sin\delta_0}{2R_{roll}}x_c^{\,2} - \psi\ x_c\cos\delta_0 - \Delta y_\psi\right)$$

$$+\frac{\cos\delta_0}{2R_{roll}}x_c^{\,2} - \psi\ x_c\sin\delta_0 + \Delta z_\psi \qquad (6.9)$$

$$\Delta z_2\left(x_c - \Delta x_\psi, y_c\right) = \Delta z_0^{\,*}\left(y_c\right) + \frac{\cos\delta_0}{2R_{roll}}x_c^{\,2} - \psi\ x_c\sin\delta_0$$

$$-\left(\frac{\sin\delta_0}{2R_{roll}}x_c^{\,2} + \psi\ x_c\cos\delta_0\right)\tan\left(\delta\left(y_c\right) - \delta_0\right) + \Delta z_\psi \qquad (6.10)$$

$$h\left(x_c - \Delta x_\psi, y_c\right) = h_0^*\left(y_c\right) + \frac{\cos\left(\delta\left(y_c\right)\right)}{2R_{roll}} x_c^{\,2}$$

$$- \psi\, x_c \sin\left(\delta\left(y_c\right)\right) + \Delta z_\psi \cos\left(\delta - \delta_0\right) \tag{6.11}$$

In these equations, additional terms Δx_ψ, Δy_ψ, Δz_ψ, and modified terms $\Delta z_0^*(y_c)$ and $h_0^*(y_c)$ appear, as a result of the shift of the rigid contact point due to the yaw angle. The longitudinal shift of the rigid contact point is taken into account by using $(x_c - \Delta x_\psi)$ instead of x_c as the first argument of the functions in the left-hand side of the equations. Thus, the functions *zw*, Δz_2 and *h* of (x_c, y_c) may be obtained by replacing x_c with $x_c + \Delta x_\psi$ in the right hand side of Eqs. (6.9)–(6.11). In the functions modified in this way, the rigid contact point is at $(x_c, y_c) = (0, 0)$, while in the functions as given in Eqs. (6.9)–(6.11), it is at $(x_c, y_c) = (\Delta x_\psi, 0)$, in accordance with the position of the blue cross-section of the wheel annotated in Figure 6.6. The terms Δy_ψ and Δz_ψ stand for the necessary shift of the wheel cross-section in the y_c-z_c plane, so that its position at the section where the rigid contact point is located coincides with that at $x_c = 0$ with zero yaw angle. $\Delta z_0^*(y_c)$ and $h_0^*(y_c)$ are modified reference distance profiles calculated at $x_c = 0$ taking into account the horizontal shift Δy_ψ of the wheel profile. Neglecting the error shown in Figure 6.4, these may be approached according to Eqs. (6.12) and (6.13) (cf. the vertical orange line in the small triangle in the left side of Figure 6.2 for the term $\Delta y_\psi \tan(\delta - \delta_0)$):

$$\Delta z_0^*\left(y_c\right) \approx \Delta z_0\left(y_c\right) - \Delta y_\psi \tan\left(\delta\left(y_c\right) - \delta_0\right) \tag{6.12}$$

$$h_0^*\left(y_c\right) \approx h_0\left(y_c\right) - \Delta y_\psi \tan\left(\delta\left(y_c\right) - \delta_0\right) \cos\left(\delta\left(y_c\right) - \delta_0\right)$$

$$= h_0\left(y_c\right) - \Delta y_\psi \sin\left(\delta\left(y_c\right) - \delta_0\right) \tag{6.13}$$

The quantity Δx_ψ is sought next. $h\,(x_c, y_c)$ is minimum at the rigid contact point. Therefore, its partial derivatives with respect to x_c and y_c are 0 at this point. Δx_ψ may be found taking the derivative of Eq. (6.11) with respect to x_c, according to Eqs. (6.14) and (6.15):

$$\frac{\partial h\left(x_c - \Delta x_\psi\right)}{\partial x_c} \approx \frac{\cos\delta}{R_{roll}} x_c - \psi \sin\delta \tag{6.14}$$

$$\frac{\partial h\left(x_c - \Delta x_\psi\right)}{\partial x_c} = 0 \text{ at } x_c = \Delta x_\psi \rightarrow \Delta x_\psi = R_{roll}\, \psi \tan\delta \tag{6.15}$$

The approximation $\partial\delta / \partial x_c \approx 0$ has been used in Eq. (6.14). As the rigid contact point is known to be located at the δ_0 angle, δ may be replaced by δ_0 in Eqs. (6.14) and (6.15). Δx_ψ may also be calculated applying the condition $n_X = 0$ in Eq. (4.85). For zero roll angle, this is solved directly, obtaining $\sin(\theta) = \tan(\delta_0) \tan(\psi)$. θ is the azimuthal coordinate of the wheel shown in Figure 4.46. So, $\Delta x_\psi = R_{roll} \sin(\theta) = R_{roll} \tan(\delta_0) \tan(\psi)$. Eq. (6.15) is a good approximation of Δx_ψ for small ψ, and δ_0 not too close to 90º (e.g., $\psi < 50$ mrad and $\delta_0 < 80$º, which is usually fulfilled in practical wheel–rail contact cases).

Δy_ψ and Δz_ψ are determined compensating the shift in the y_c-z_c plane of the wheel cross-section, so that it lies in the same position in the y_c-z_c plane at $x_c = \Delta x_\psi$ as with zero yaw angle at $x_c = 0$. So, considering Eq. (6.9):

$$\Delta y_\psi = -\left(\frac{\sin\delta_0}{2R_{roll}} \Delta x_\psi{}^2 + \psi\, \Delta x_\psi \cos\delta_0 \right) \tag{6.16}$$

$$\Delta z_\psi = \psi\, \Delta x_\psi \sin\delta_0 - \frac{\cos\delta_0}{2R_{roll}} \Delta x_\psi{}^2 \tag{6.17}$$

The ratio $h / \Delta z = \cos(\delta - \delta_0)$ holds in this case as well. To show it, Eq. (6.18) demonstrates that the ratio between the added terms for h in Eq. (6.11) and for Δz_2 in Eq. (6.10) due to the $\psi\, x_c$ translation, is $\cos(\delta - \delta_0)$, equal to the value found in §6.1.1 for the rest of the terms of h and Δz_2.

$$\sin\delta_0 + \cos\delta_0 \tan\left(\delta - \delta_0\right) = \frac{\sin\delta_0 \cos\left(\delta - \delta_0\right) + \cos\delta_0 \sin\left(\delta - \delta_0\right)}{\cos\left(\delta - \delta_0\right)}$$

$$= \frac{\sin\delta}{\cos\left(\delta - \delta_0\right)} \rightarrow \frac{\sin\delta}{\sin\delta_0 + \cos\delta_0 \tan\left(\delta - \delta_0\right)} = \cos\left(\delta - \delta_0\right) \tag{6.18}$$

The results obtained with the described formulae are illustrated in Figure 6.7 for a case with the same geometry as the one considered in Figure 6.5 (i.e. case "circ. a."), except that now ψ is set to –3.5364 mrad. Figure 6.7 compares the contours of vertical and normal distances computed in different ways, with the formulae shown here and with CECT, as in Figure 6.5. The contours are shown at levels of 0.04 mm and 0.08 mm for each case. The figure also includes the wheel contact locus, plotted on $x_c - \Delta x_\psi$ vs. s axes in dashed magenta line.

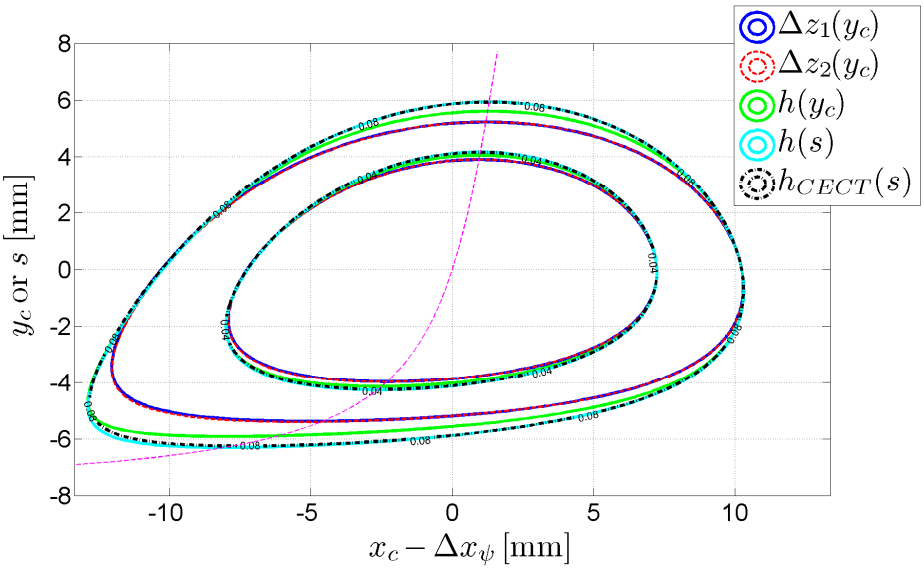


Figure 6.7. Contours of vertical and normal distances for "circ. a." case, with circular conforming sections, contact angle δ_0 of 45º and yaw angle ψ of –3.5364 mrad.

As can be seen in the figure, now the distance contours are asymmetric in both lateral and longitudinal directions. The asymmetry in the longitudinal direction is caused by the terms related to the ψ x_c displacement, which are identified as the ψ x_c sin δ term of Eq. (6.11), or the ψ x_c (sin δ_0 + cos δ_0 tan($\delta - \delta_0$)) terms of Eq. (6.10). For a given yaw angle, these terms are more important with higher δ and x_c values. Considering that the maximum relevant x_c values increase with R_{roll}, it may also be said that these terms tend to be more important with higher R_{roll} values. Evidently, as the conformity in the y_c-z_c plane between wheel and rail profiles decreases, both the terms causing longitudinal and lateral asymmetries (i.e., the terms related to the component along y_c of the d_r and ψ x_c displacements) are less important, because the original undeformed distance

profile h_0 has higher values, and because $\tan(\delta - \delta_0)$ decreases. This may be easily visualized considering that lateral shifts of the wheel profile as those shown in Figure 6.6 are less prone to affect the original h_0 profile with less conforming wheel and rail profiles.

Regarding the comparison between the different distance calculation methods in Figure 6.7, similar comments may be made as for Figure 6.5. In this case, slight differences are observed at the lower left zone of the contours at h = 0.08 mm of the normal distances calculated with Eq. (6.11) and plotted on the s coordinate, and of those calculated in CECT. In this zone, $\delta \approx 80^{\text{o}}$. Anyway, the agreement between both calculations is still very good.

It is instructive to look at the form of h in the longitudinal direction. The method developed in [Baeza 2018] for the calculation of wheel–rail interpenetration areas shows that the longitudinal variation of h in each contact strip may be approached as the distance between an arc and a plane, i.e. as a quadratic function for small x_c / R_{roll} values. Following this, an alternative analytical formula for h is set forth next, with the following premises:

- h is quadratic along x_c. The wheel longitudinal radius of curvature R_{xw} in each strip is equal to R_{roll} / cos(δ), taken as constant along x_c for small yaw angles.
- The minimum of h in each strip is located at the longitudinal position of the possible contact point of the wheel with the rail, i.e. the wheel contact locus.
- The undeformed distance profile h_0 (s) at the longitudinal position where the rigid contact point is located (at $x_c = \Delta x_\psi$) is equal to the reference profile h_0 (s) at $x_c = 0$ with $\psi = 0$.

According to these points, Eq. (6.19) gives an alternative formula for h.

$$h\left(x_c - \Delta x_\psi, y_c\right) = h_0\left(y_c\right) + \frac{\cos\delta}{2R_{roll}}\left[\left(x_c - x_0\right)^2 - \left(x_0 - \Delta x_\psi\right)^2\right] \tag{6.19}$$

In this equation, x_0 is the x_c coordinate of the wheel contact locus at each lateral position, given by Eq. (6.20). At $y_c = 0$, $x_0 = \Delta x_\psi$. The term $-\psi\, y_c \cos\delta_0$ included in Eq. (6.20) is due to the longitudinal translation of the points on the wheel due

to the yaw rotation. This effect is not included in the previous equations developed in this section. The term $-\psi\, y_c \cos\delta_0$ is negligible with respect to $R_{roll}\, \psi \tan\delta$ for $R_{roll} / |R_{sw}| >> 1$ or for high δ values, except where $\delta \approx 0$ and $y_c \neq 0$, with non-zero δ_0.

$$x_0(y_c) = \psi \left(R_{roll}(y_c) \tan\delta - y_c \cos\delta_0 \right) \tag{6.20}$$

Figure 6.8 illustrates the approximation of h along a longitudinal strip of the contact surface according to Eq. (6.19). The contours of h plotted in Figure 6.7 (calculated with Eq. (6.11)) are nearly coincident with the contours of h calculated with Eq. (6.19) (not shown).

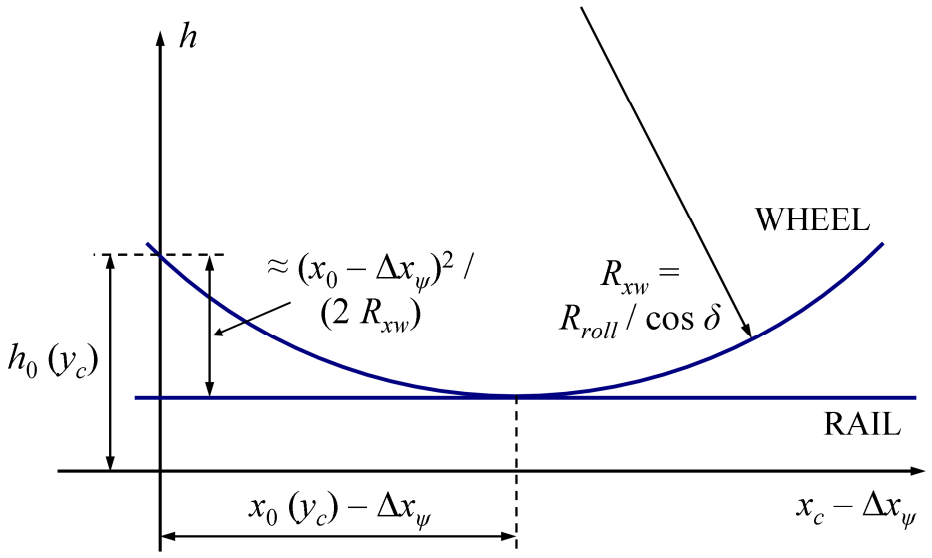


Figure 6.8. Quadratic approximation of h along x_c.

By now, it is clear that in this approximation of h, the wheel contact locus, plotted in dashed magenta line in Figure 6.7, coincides with the point of minimum h in each strip, and this point is midway between the two points of the strip at a given h level above that minimum. Thus, the wheel contact locus may be taken as an indicator of the degree of asymmetry of the contact patch in longitudinal direction. Contact loci with significant $x_0 - \Delta x_\psi$ values in relation to the contact patch dimensions and substantially curved indicate more asymmetrical contact patches. In these cases, semi-Hertzian representations of h

(with $x_0 - \Delta x_\psi = 0$) as those found in simplified wheel–rail contact models as those of [Ayasse 2005], [Quost 2006] and [Sichani 2014] are not accurate.

Neglecting the $-\psi\ y_c \cos\delta_0$ term in Eq. (6.20) and approximating R_{roll} as constant in the whole contact surface, $x_0 - \Delta x_\psi \approx \psi\ R_{roll}\ (\tan\delta - \tan\delta_0)$. So, the quantity $R_{roll}\ (\tan\delta - \tan\delta_0)$ is a measure of the sensitivity of h to the yaw angle ψ. Here it is seen again that h becomes more sensitive to ψ with higher R_{roll} and δ. With reference to the role of R_{roll}, it may be thought of the extreme case of very conforming prismatic surfaces (with $R_{roll} = \infty$). In this case, a slight misalignment between both surfaces changes considerably the contact conditions, from a single line contact to two separate point contacts at opposite extremes of the bodies. The sensitivity of h to ψ is further discussed in §6.2.1.2.

As with zero yaw angle, the expressions for Δz_2 and h developed in this section are affected by the second-order error shown in Figure 6.4.

6.2 Effects of conformity on wheel–rail contact mechanics

This section discusses the effects that conformity brings to wheel–rail contact mechanics, which distinguish conformal and non-conformal wheel–rail contact. Effects on the geometry, on the influence coefficients (ICs) and on contact subsurface stresses are discussed.

With the aim of making the explanation as clear as possible, the different effects exposed are illustrated with contact cases with simple geometries, with constant lateral curvature radii R_{sw} and R_{sr} in the contact zone. The parameters of three of the considered cases are listed in Table 6.1. In these three cases, a concave wheel on a convex rail is considered, as can be observed with the sign of R_{sw} and R_{sr}. Cases “circ. s.” and “circ. l.” are defined in §5.2.3.7.1 (cf. Table 5.9), and their relevant geometrical parameters are indicated again in Table 6.1 for convenience. The geometries of these two cases are symmetric both in longitudinal and lateral directions for zero yaw angle. Case “circ. a.” (meaning “circ. asymmetric”) has the same reference curvatures as case “circ. l.”, but a non-zero mean contact angle δ_0. In all the cases, the indicated values for R_{xw} and δ_0 correspond to the initial contact point. Other geometries considered are described later.

Table 6.1 Parameters of conformal contact cases with circular wheel and rail profiles.

	Values		
Parameter	**"Circ. s."**	**"Circ. l."**	**"Circ. a."**
R_{xw} [mm]	500	654	654
R_{sw} [mm]	–10.1	–10.5	–10.5
R_{sr} [mm]	10.0	10.0	10.0
δ_0 [°]	0	0	45

In all the examples steel is considered as the material of both contacting bodies, with a Young's modulus E of 210 GPa and a coefficient of Poisson ν of 0.30. Unless otherwise noted, the results shown in this section have been obtained with FEM, making nonlinear static stepwise simulations as described in §4.2.1, to simulate frictional contact cases, static or rolling. Coulomb's law is considered for frictional contact cases, with a constant coefficient of friction μ of 0.30. In other cases, frictionless contact is considered. A contact normal load N of 80 kN is considered in the rolling contact cases. All the examples shown have been solved also with CECT, obtaining a good correspondence with the FEM results when the ICs used in CECT have been sufficiently accurate. Some comparisons between FEM and CECT results for different conformal contact cases are shown in [Blanco-Lorenzo 2016] and [Blanco-Lorenzo 2018].

A parameter used to characterize a conformal wheel–rail contact, namely the total contact angle variation in the contact patch, $\Delta\delta$, is illustrated in Figure 6.9. The figure shows a section view of a concave wheel on a convex rail, perpendicular to the rolling direction x. The contact patch profile is shown in magenta line. $\Delta\delta$ is the range of inclination angles in the YZ plane spanned by the contact surface, and is a measure of the degree of the contact conformity. The figure also shows the mean contact angle δ_0, previously introduced, and the s and n directions of the local contact curvilinear $\{x, s, n\}$ system at a given lateral position. The origin of this coordinate system may be arbitrarily defined. Normally, $s = 0$ is set at the initial contact point or at the position defined by δ_0. δ_0 is an intermediate value within the range of δ values spanned by the contact patch. It may be defined in different ways, for example as the average of the δ values in the two lateral extremes of the contact patch, or as the local δ at the initial or rigid contact point. In [Vollebregt 2021] other alternatives are explained.

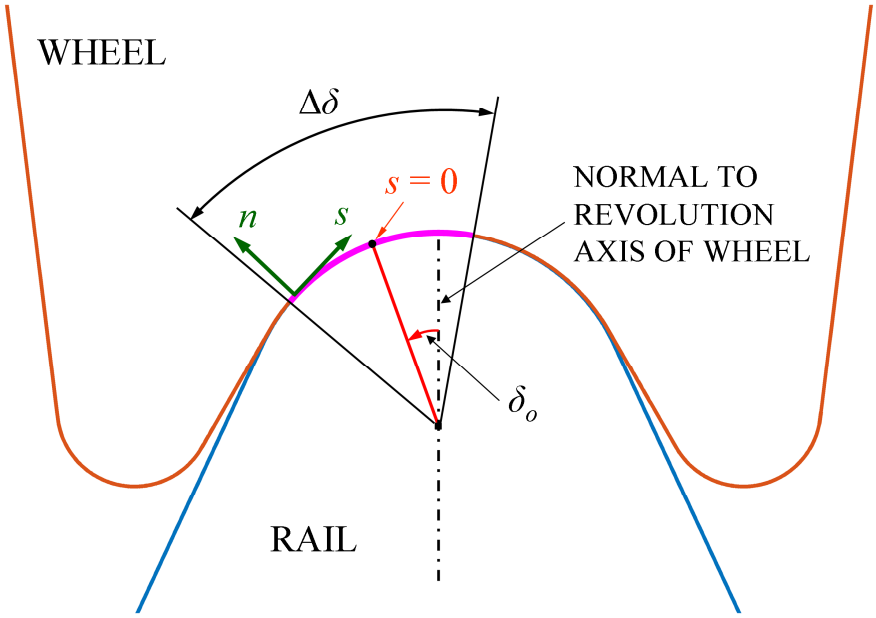


Figure 6.9. Conformal contact geometry with definition of δ_0 and $\Delta\delta$.

6.2.1 Geometrical effects

The geometrical effects of conformity are derived from the non-flatness and from the closeness of the contacting surfaces. The geometrical effects treated in this section are referred particularly to the normal undeformed distances and to the rigid slip velocities. These are the fundamental input variables for the normal and the tangential contact problems, respectively, as explained in §4.1.1.2 and §4.1.1.3. With the aim of capturing adequately these effects particular of conformal contact, it is necessary to consider in the analysis the mentioned non-flatness of the surfaces, abandoning commonly adopted simplifications in non-conformal contact.

6.2.1.1 Change of wheel longitudinal curvature radius

In wheel–rail contact, the wheel rolling radius R_{roll} is much higher than the contact patch dimensions, and by extension much higher also than its variation within the contact patch. On the contrary, the longitudinal curvature radius of the wheel, R_{xw}, equal to R_{roll} / cos δ, is variable, as δ is variable. R_{xw} tends to

increase with δ, and its variation is higher with higher mean contact angles δ_0. Consequently, the longitudinal dimension of the contact patch strips tends to increase in the zones with higher contact angles.

This is illustrated in Figure 6.10 with case "circ. a.", which has a relatively high δ_0. The input geometrical parameters of this case are indicated in the upper part of the figure as well. With N = 80 kN, $\Delta\delta$ is approximately 42º with ψ = 0. The contact patch contours obtained with different analysis types are shown in the figure, without and with the effects of conformity included. The non-conformal analysis has been carried out with the exact contact theory, without including the necessary extensions for conformal contact, and calculating the undeformed distance with the usual simplification in non-conformal contact of constant longitudinal curvature radius in the whole contact patch. On the other hand, the conformal analyses have been carried out with FEM. The figure also includes the contours of the adhesion and slip zones, for the steady rolling contact with μ = 0.30 (constant) and creepages on the initial contact point given by $\xi_0 = 0$; $\eta_0 = -\psi / \cos \delta_0$, and $\varphi_0 = -\sin \delta_0 / R_{roll}$.

The contact patch contour obtained by means of the non-conformal analysis, represented in blue and designated as "Non-conf.", is symmetric in longitudinal and lateral directions. Each of the other two contact patches correspond to a different ψ. The one corresponding to zero ψ, represented in red and designated as "ψ = 0", is not symmetric laterally, being more elongated in the side with higher contact angles (corresponding to the negative axis of the s coordinate in the figure), due to the described variation of R_{xw} with δ. The contact patch "ψ = 0" corresponds to the case shown in Figure 6.5.

With low δ_0 values the effect is less notable, but is manifested as well, and as a consequence the contact patches tend to change from an elliptical to a more rectangular shape as $\Delta\delta$ increases. This may be appreciated in Figure 5.51c and Figure 5.52c, which correspond to cases "circ. s." and "circ. l.", respectively. The same contact patch contours plotted in those figures are plotted in normalized axes in Figure 6.11, for a better visualization of the effect. The longitudinal x and lateral s coordinates are normalized in this figure dividing them by the corresponding maximum coordinates of the contact patch in each direction, x_{max} and s_{max}. The reference elliptic contour for the non-conformal case is also shown in each graph in black dotted line. This effect may not be

present when R_{roll} is not constant with convex lateral R_{sw} radii, as may happen in rolling bearings. For example, in the case of a ball in a conforming groove, R_{xw} clearly does not change across the contact patch, and therefore the contact patch maintains a nearly elliptical shape regardless of $\Delta\delta$.

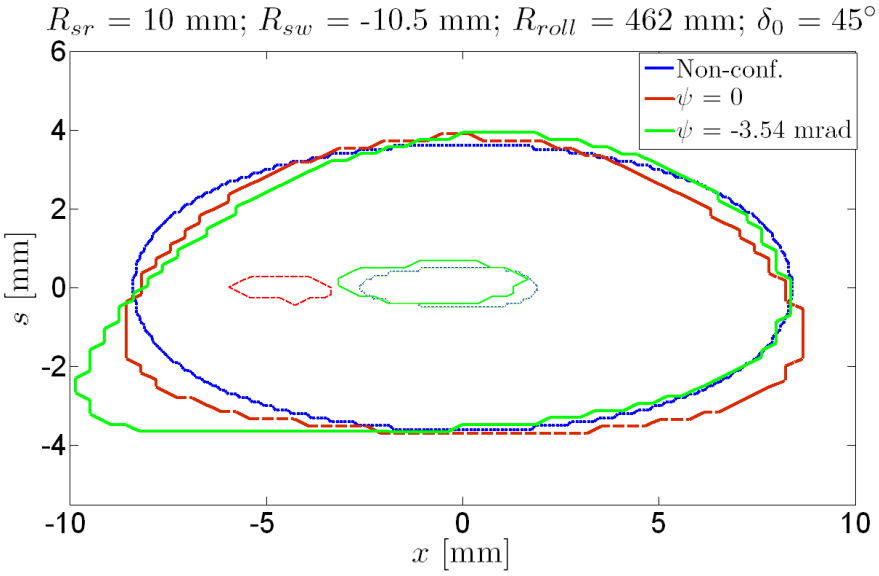
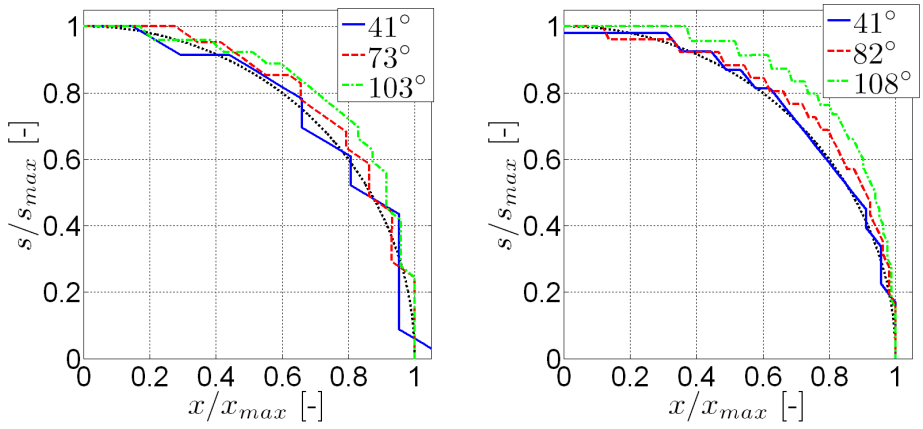


Figure 6.10. Contact patch contours in a conformal contact case, illustrating the effects of conformity in the geometry of the normal undeformed distance.



(a) Case “circ. s.”. (b) Case “circ. l.”.

Figure 6.11. One quarter of the contact patch contours (symmetric in the lateral and in the longitudinal directions about s = 0 and x = 0, respectively) obtained in frictionless

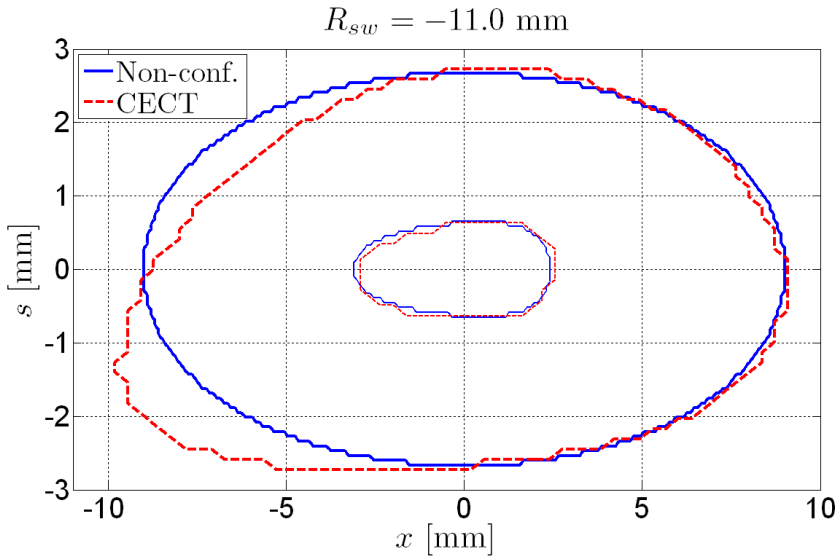
conformal contact cases with $\delta_0 = 0$, for different $\Delta\delta$-s as shown in the legend of each figure, plotted in non-dimensional axes.

6.2.1.2 Sensitivity of the undeformed distance to the yaw angle

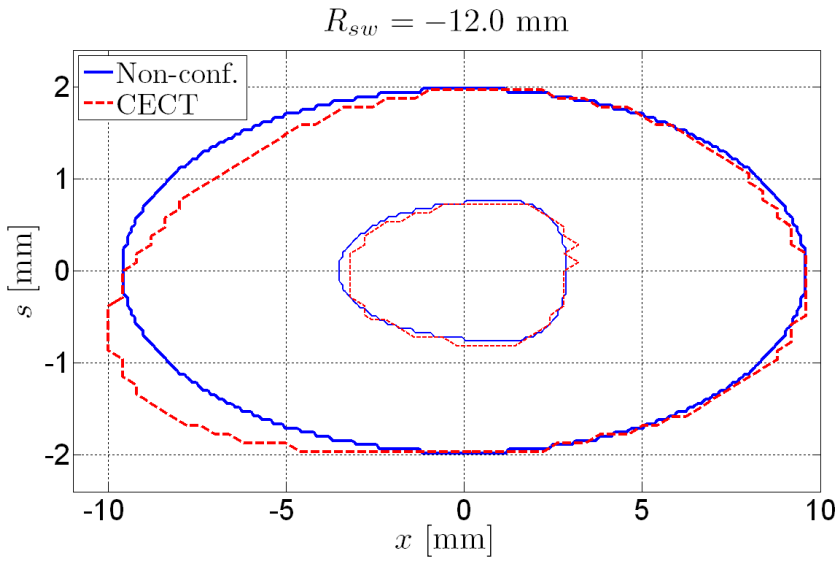
Figure 6.10 includes the contact patch for case "circ. a." obtained with a conformal analysis with a yaw angle ψ of –3.5364 mrad, in green and designated as "ψ = –3.54 mrad", together with the previously described one for $\psi = 0$. This corresponds to the same case considered in Figure 6.7. While ψ is not large, it may be seen that the difference with the contact patch for $\psi = 0$ is considerable. This is a consequence of the amplification of the rotation of the principal axes of h that is produced in conformal contact, particularly when R_{roll} is much higher than the lateral curvature radii R_{sw} and R_{sr}.

The contact patches with non-zero ψ and high δ_0 in conformal contact are asymmetric in longitudinal as well as in lateral directions, tending to become cam-shaped as the contact patch designated as "ψ = –3.54 mrad" in Figure 6.10 (cf. also the h contours for the same case in Figure 6.7). This is due to the combined effect of the variation of R_{xw} with δ on the one hand, and of the sensitivity of the orientation of the principal directions of h to ψ on the other.

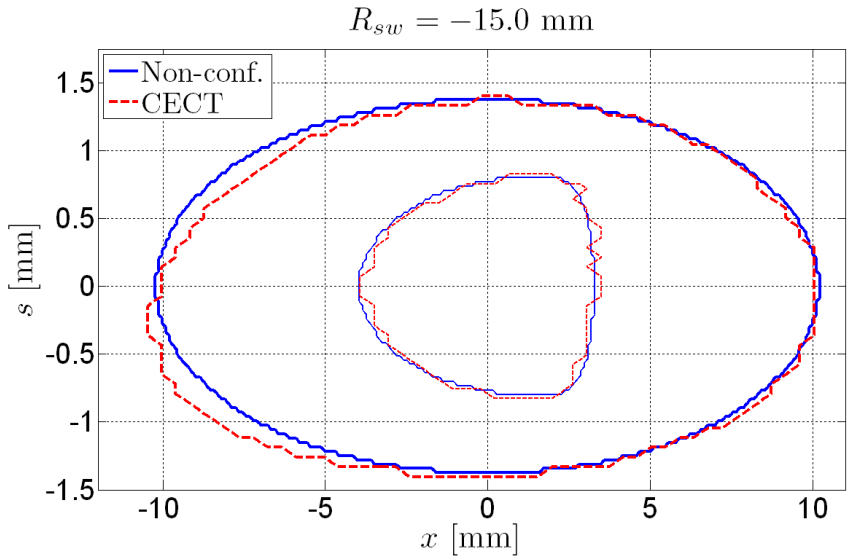
Figure 6.12 shows the contact patch contours for other cases similar to "circ. a." except for R_{sw}, which is given the value indicated in the figure in each case. In all the cases shown in this figure, ψ = –3.5364 mrad, as for the cam-shaped contact patch shown in Figure 6.10. Each graph includes the contact patch calculated without taking into account the effects of conformity, as in Figure 6.10. In this case, the conformal contact calculation is carried out with CECT, with the B approximated ICs of §5.2.3.1. The figure includes the contours of the adhesion and slip zones in each case, calculated with similar conditions as explained before for Figure 6.10. It has to be noted that the x and s axes of each of these graphs have different scales. As R_{sw} increases and the conformity level decreases, the effects of the variation of R_{xw} with δ and of the sensitivity of the orientation of the principal directions of h to ψ become smaller. Consequently, the contact patches become less asymmetrical in lateral and longitudinal directions, as can be seen in the figure. The $\Delta\delta$ values for the cases shown in the figure, with R_{sw} = –11.0, –12.0 and –15.0 mm, are 30º, 21º and 13.5º, respectively.



(a) R_{sw} = –11.0 mm.



(b) R_{sw} = –12.0 mm.



(c) R_{sw} = –15.0 mm.

Figure 6.12. Contact patch contours for cases similar to "circ. a.", with ψ = –3.5364 mrad and different R_{sw}.

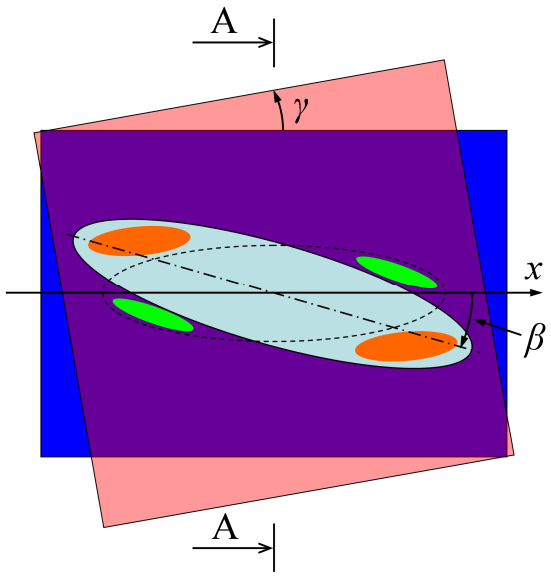
The aforementioned amplification effect of the rotation of the principal axes of h is present also for planar contacts (as in the test case of §4.3.1), and it may be predicted with a geometrical analysis of the combined undeformed distance h of both contacting solids, as done in Appendix 2 of [Johnson 1987] for two quadratic surfaces in non-conformal contact[27]. Following that analysis, Eqs. (2.20)–(2.23) are obtained, which define the principal curvatures of the resulting combined undeformed distance between the two surfaces, and the orientation of the planes containing them.

Applying Eqs. (2.21)–(2.23) to the geometry of the example considered in Figure 6.10 with ψ = –3.5364 mrad, the obtained β angle is about 30 times higher than γ (cf. Figure 2.3). For a concave wheel on a convex rail, when the

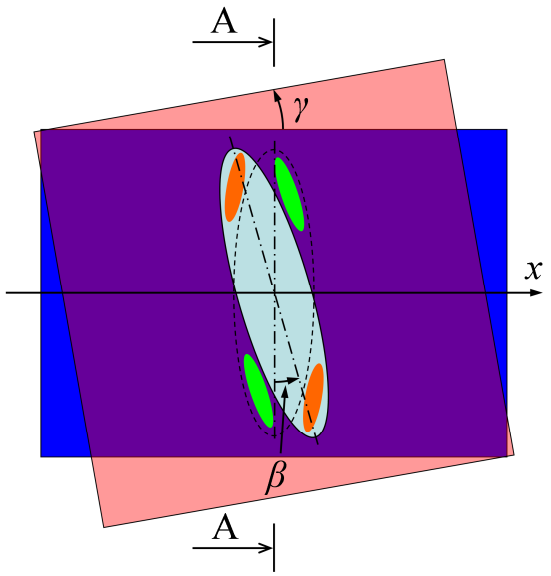
[27] Despite the referenced appendix is entitled "Geometry of smooth non-conforming surfaces in contact", the geometric analysis presented there is applicable also to conforming surfaces, in a sufficiently small (approximately planar) region of the surfaces.

longitudinal dimension of the contact patch (a) is bigger than the lateral one (b), the rotation of the principal directions of h is produced in the opposite direction to the γ rotation; and in the same direction when $a < b$. The contrary happens for a convex wheel on a concave rail. This effect may be easily visualized by considering the foreseeable locations of material interference and separation regions for a pair of conforming wheel and rail sections with e.g. $\delta_0 = 0$, for non-zero γ. This is illustrated in Figure 6.13a for a contact patch with $a > b$, and in Figure 6.13b for a contact patch with $a < b$. In both cases, a concave wheel on a convex rail is considered, as shown in Figure 6.13c.

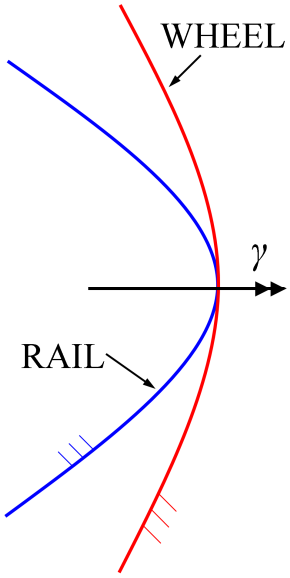
Figure 6.13 a and b are plan views of a concave wheel, represented in semi-transparent red colour, on a convex rail, represented in blue colour. The wheel is yawed at an angle γ with respect to the rail, as shown in these figures. In the cross-section of Figure 6.13c, the wheel profile is represented in red line and the rail profile in blue line. Both in Figure 6.13 a and b, the contact patch in the final position is represented in light blue fill colour. The contour of the contact patch for $\gamma = 0$ is shown as well in dashed line in each figure. Due to the γ rotation of the wheel, the wheel material is displaced in planes parallel to the plane of Figure 6.13 a and b, coming closer to the rail surface in some parts, and separating from it in other parts. According to the considered geometry and direction of γ rotation, it may be seen that the interference regions (i.e., the regions in which the wheel surface approaches the rail surface) are in the left upper and right lower parts of the contact patches, as seen in Figure 6.13 a and b. These regions are represented in orange colour. On the other hand, the separation regions are in the right upper and left lower parts of the contact patches, and represented in green colour. Combining the effect of the interference and separation regions with the contact patch contour for $\gamma = 0$, the final (rotated) contact patch is obtained. In this way, it is understood that the β rotation of the contact patch is produced in the opposite direction to the γ rotation of the wheel for the contact patch with $a > b$ (Figure 6.13a), and in the same direction for $a < b$ (Figure 6.13b).



(a) Plan view, $a > b$.



(b) Plan view, $a < b$.



(c) Section A–A.

Figure 6.13. Direction of β rotation depending on the contact patch a / b ratio.

As commented in §6.1.2, the wheel contact locus provides information on the sensitivity of h to the yaw angle ψ. According to the analysis presented there,

contact loci with larger $x_0 - \Delta x_\psi$ values indicate higher sensitivities of h to ψ. However, to make comparisons between different cases, the shape of h or the contact patch has to be taken into account as well. To illustrate this, Figure 6.14 shows the contours of two different h surfaces at a given h level, which share the same wheel contact locus, shown in the figure in dashed magenta line. The principal axes of each contour are shown as well, in lighter line. Both h surfaces are quadratic, and the wheel contact locus is obtained as the line joining the points with minimum and maximum s coordinates of their elliptical contours. It may also be considered that the wheel contact locus depends only on the wheel surface, and different h contours may be obtained by changing the rail surface. As may be observed, one of the contours, which is more elongated in longitudinal direction, has a much lower rotation, and is nearly symmetric in the x and s directions, which indicates lower sensitivity of h to ψ.

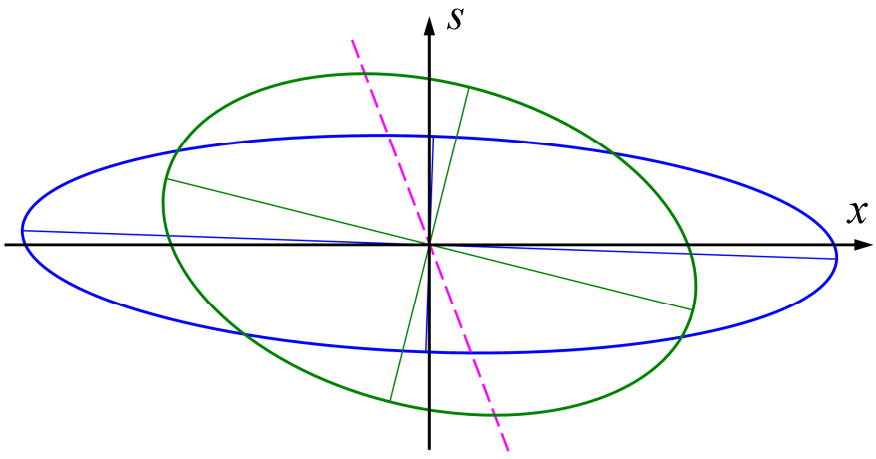


Figure 6.14. h contours with different shapes and rotations, with the same wheel contact locus.

6.2.1.3 Factoring of the approach with cosine term in the undeformed distance

In non-conformal contact, the effect on h of the approach d between the contacting solids is uniform in the whole contact patch. On the contrary, in conformal contact the effect is variable with the inclination of the contact surface in the YZ plane. This was explained in §4.1.1.2 (cf. Figure 4.10 and Eq.

(4.15)). The inclination of the contact surface in the *YZ* plane may be characterized either with the α angles (cf. §4.1.1.2) or with the contact angles δ.

The term $d \cos(\Delta\alpha)$ introduces an additional variation of h in lateral direction, increasing its lateral effective curvature. This is illustrated considering a simple geometry with constant curvature radii and zero ψ and δ_0 angles, as may be the one represented in Figure 4.10. In this case, h may be expressed in first approximation as $h = A \cos\delta\, x^2 + B\, s^2$, taking the origin of the x, s coordinates in the initial contact point. This approximation is valid even for considerable $\Delta\delta$ angles. A and B are obtained from the curvatures of the contacting surfaces as $A = (C_{xr} + C_{xw}) / 2$ and $B = (C_{sr} + C_{sw}) / 2$, as in non-conformal contact. The convex curvatures are positive, and the concave ones negative. On the other hand, taking the series expansion of the cosine function, the term $d \cos(\Delta\alpha) = d \cos\delta$ (δ being equal to $\Delta\alpha$ for $\delta_0 = 0$ and d applied along the direction defined by δ_0) is approximated as indicated in Eq. (6.21). In this equation r is the mean lateral curvature radius of the contact surface, which for elastically similar materials may be approached as the average of the absolute values of the lateral curvature radii of the two undeformed contacting surfaces.

$$d\cos\delta \approx d\left(1-\frac{\delta^2}{2}\right) = d\left(1-\frac{(s/r)^2}{2}\right) \tag{6.21}$$

Substituting the approached equation for h given above and Eq. (6.21) in Eq. (4.15), and reordering terms, Eq. (6.22) is obtained. The effective lateral curvature B_e may be identified in this equation as $B_e = B + d / (2\, r^2)$. In non-conformal contact, the term $d / (2\, r^2)$ is very small with respect to B, and may be neglected. On the contrary, in conformal contact it may easily reach the same order of magnitude as B or higher.

$$h' \approx A\cos\delta\, x^2 - d + \left(B + \frac{d}{2r^2}\right) s^2 \tag{6.22}$$

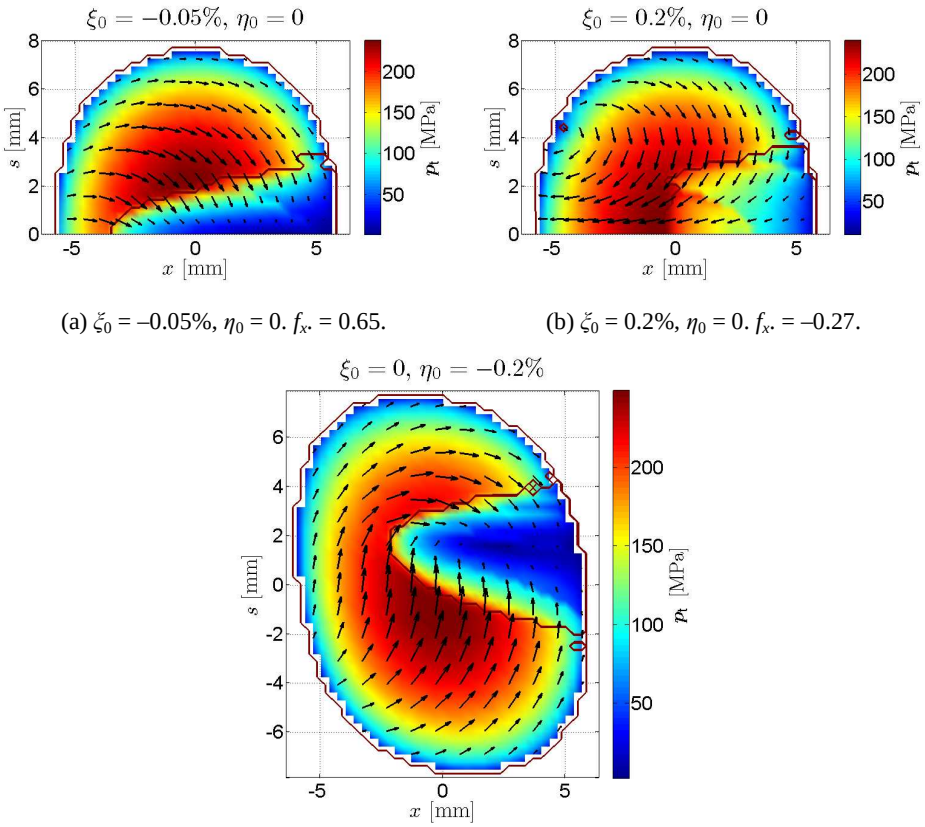
As a consequence of the increase of B_e with d (and hence with the normal load), the a/b ratio between the longitudinal and lateral dimensions of the contact patch tends to increase with the normal load in conformal contact, obtaining more elongated contact patches in the rolling direction for higher loads. This is

verified in the cases with circular profiles “circ. s.” and “circ. l.” studied in §5.2.3.7, as may be observed in Figure 5.51b and Figure 5.52b. For cases with non-constant transverse curvatures, this effect may be attenuated or accentuated, depending on the lateral curvature variation. For example, in the case “el. tall” studied in §5.2.3.7 (cf. Figure 5.50c), in which the lateral curvatures decrease in lateral direction from the initial contact point, the effect is attenuated.

6.2.1.4 Asymmetries in the rigid slip velocity field

In conformal contact, with a geometric spin variable in lateral direction (equal to $-\sin \delta / R_{roll}$), asymmetries in the fields of rigid slip velocity $\mathbf{w}$ are generated which are not seen in non-conformal contact. To illustrate some of these asymmetries, the “circ. s.” case is considered. This case presents a high conformity level, with a $\Delta\delta$ of approx. 88º. Figure 6.15 shows the fields of tangential contact stresses $\mathbf{p_t}$ obtained in this case in steady-state rolling contact situations with different creepages, together with the contours of adhesion and slip zones. The creepages imposed in each case are indicated in the upper part of each subfigure. These creepages, with subindex 0, are referred to the initial contact point. This same notation is used to define the creepages in different rolling contact cases throughout this thesis.

Two laterally symmetric rolling contact cases are considered in subfigures a and b of Figure 6.15, with zero ψ and different ξ_0. As there is lateral symmetry with respect to the contact centreline at $s = 0$, only half of the contact patch for $s > 0$ is shown in these subfigures. In subfigure a ξ_0 is negative (i.e., tractive), and in subfigure b ξ_0 is positive (i.e., braking). The traction coefficient f_x is indicated in the caption of each subfigure. It is equal to the resultant longitudinal contact force F_x acting on the wheel divided by μN.



(a) ξ_0 = –0.05%, η_0 = 0. f_x. = 0.65.

(b) ξ_0 = 0.2%, η_0 = 0. f_x. = –0.27.

(c) ξ_0 = 0, η_0 = –0.2%.

Figure 6.15. Tangential contact stresses in conformal contact cases in steady rolling with the geometry “circ. s.”.

As may be seen, the patterns of tangential stresses and contours of adhesion and slip zones are quite different in the tractive and in the braking cases. In the tractive case, the effects of the tractive longitudinal creepage at $s = 0$ and of the increasing rolling radius for increasing $|s|$ are in the same sense, while in the braking case they oppose each other. In this way, the magnitude of the $\mathbf{w}$ velocities increases faster with $|s|$ in the tractive case. In fact, in the braking case $\|\mathbf{w}\|$ initially decreases from $s = 0$, changing from positive longitudinal component w_x at $s = 0$ to negative at the lateral contact zones, with higher rolling radii. This is shown in Figure 6.16, where $\|\mathbf{w}\|$ divided by the rolling velocity V is plotted for the cases of Figure 6.15 a and b along s at $x = 0$. As a result of these variations of the rigid slip velocities, the adhesion zone tends to

become more extended in the lateral direction in the braking case, and it becomes more acute shaped in the tractive case.

In both the tractive and braking cases, the lateral zones of the contact patch, with a high contact angle and high associated geometrical spin, are sliding, and remain so not only in the two particular situations shown here, but in the whole range of longitudinal creepages within the tractive and braking saturation limits. It is observed as well that the adhesion zone is less extended in the longitudinal direction in the braking case than in the tractive one, even though the resultant longitudinal force is smaller in the braking case. This is because the absolute value of the creepage is higher in the braking case. Despite the higher absolute value of the creepage in the braking case, $|f_x|$ is smaller than in the tractive case, due to the contribution of the most lateral longitudinal strips of the contact patch, which, as a result of the higher rolling radii at these zones, is tractive.

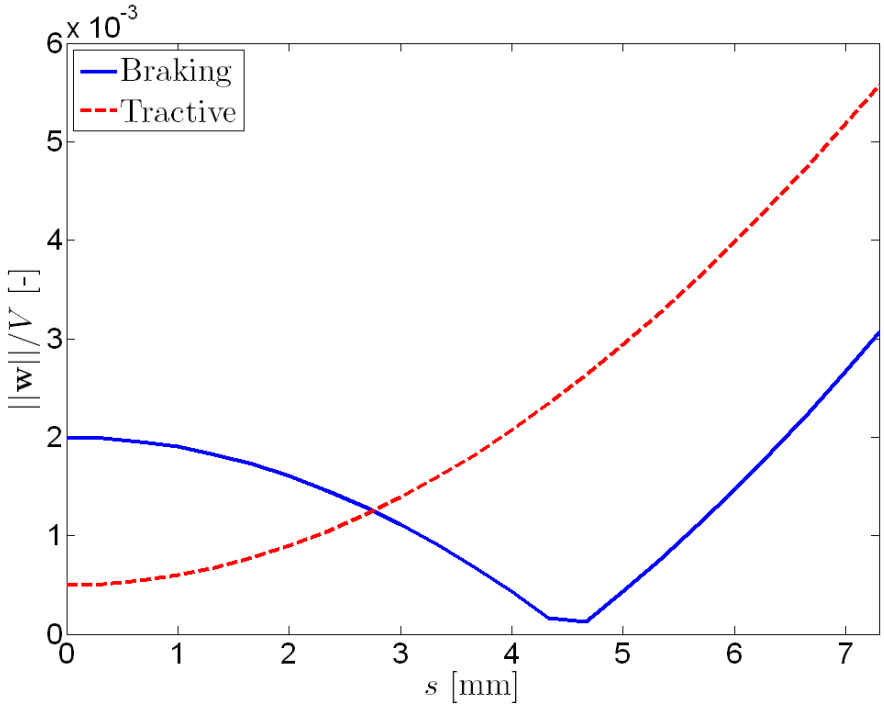


Figure 6.16. Magnitude of the rigid slip velocities across the central cross-section of the contact patch at $x = 0$ for the tractive and braking rolling cases of Figure 6.15 a and b. Results symmetric about $s = 0$.

The described characteristics for the tractive and braking regimes of Figure 6.15 a and b correspond to a concave wheel rolling on a convex rail. In the case of a convex wheel on a concave rail, the behaviour for the two regimes is interchanged. The dissimilar characteristics described for the tractive and braking regimes of Figure 6.15 a and b contrast with what happens in non-conformal rolling contact with elastically similar materials: in these conditions, when changing the sign of the creepages, similar contact solutions are obtained, changing just the sign of the tangential stresses. A further difference with respect to non-conformal contact is that the field of **w** in general does not have a single pole, as a consequence of the variable geometric spin. As a result of the nonlinear variation of the rolling radii across the contact patch, wherein there may be a maximum or minimum of the rolling radius in the central part of the contact patch, the longitudinal rigid slip may change sign more than once across the contact patch. This leads to the duplicity of the so-called zero-slip lines in ball bearings and in roller bearings with crowned rollers, where the contact patches are very short in longitudinal direction and the rigid slip velocities are usually mostly aligned with the longitudinal direction (cf. Chapter 7).

In the case of Figure 6.15c, there is a η_0 of –0.2%, imposed through a yaw angle of 2 mrad. In this case, the solution to the tangential problem is not symmetric laterally, in contrast with what would happen in non-conformal contact, due to the combination of the lateral creepage with the geometric spin, of different sign in each contact patch side. In the normal part, it may be observed the considerable rotation of the contact patch produced by the small ψ of 2 mrad. This is a consequence of the previously described high sensitivity of the orientation of the principal directions of h to ψ.

6.2.1.4.1 Lateral distributions of frictional work

The **w** field conditions the fields of slip velocities and frictional work densities in the contact patch. The lateral distributions of frictional work in the rail after a wheel passage, W_{fric}, are examined next for different conformal steady rolling contact cases.

Figure 6.17 depicts the lateral distributions of W_{fric} for several cases with the geometry of case “circ. s.”, with $\psi = 0$ and $\eta_0 = 0$ as in Figure 6.15 a and b, for varying longitudinal creepages. These results have been computed with FEM,

as the ones shown in Figure 6.15. The abscissa axis of the graph of the figure spans only one half of the contact patch width, as the results are symmetric about s = 0. Starting with the case with the highest braking (i.e., positive) ξ_0 shown in the figure, the maximum W_{fric} is clearly located in the centreline of the contact patch, due to the fact that both the highest tractions and slip velocities are located in this part. On the other hand, at not too high braking ξ_0, a relative maximum of W_{fric} begins to appear in the sides of the contact area, associated to the increasing slip velocities in those parts.

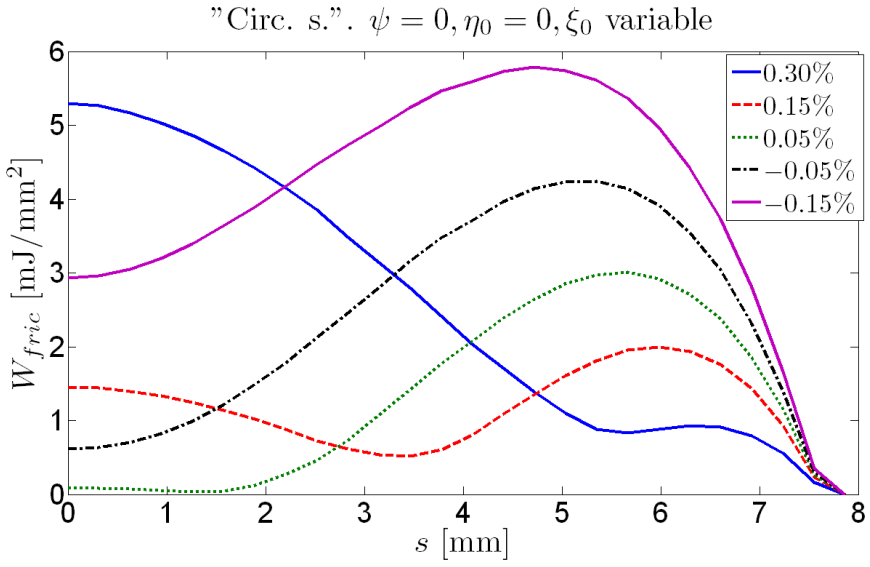


Figure 6.17. Lateral distribution of W_{fric} in conformal steady rolling contact with the geometry "circ. s.", with ψ = 0 and η_0 = 0, for different ξ_0 as shown in the legend of the figure. Results symmetric about s = 0.

A case with non-zero η_0 is considered next. It has the "circ. l." geometry, with creepages ξ_0 = 0 and η_0 = –0.2%. The lateral distribution of W_{fric} for this case is shown in Figure 6.18. Figure 6.36b and Figure 6.37b show the distributions of tangential stresses and relative slip velocities for this case. The details of the different calculation models shown in the legend of the figure are explained in §6.2.4.

In this case, the tangential stress distributions and adhesion and slip areas are not symmetrical around the longitudinal axis of the contact patch, as can be seen in Figure 6.36b and Figure 6.37b. This is due to the combination of the

lateral creepage with the geometric spin, which is variable across the contact patch. However, the distributions of W_{fric} obtained with non-conformal as well as with conformal analyses are nearly symmetrical in the lateral direction, as can be seen in Figure 6.18. This results from two opposing effects:

- The larger tangential stresses and slip area in the side towards which the wheel pushes the rail (i.e., towards decreasing *s*-coordinates), as is seen in Figure 6.36b.
- The larger slip velocities in the other side, as is shown in Figure 6.37b.

Both effects are due to the combined effect of the rigid slip velocities resulting from the applied lateral creepage on the one hand, and those resulting from the geometric spin on the other hand. In the rear part of the contact patch (i.e., in the side with decreasing *x*-coordinates), where the slip area is located, these have the same sense and therefore they sum up in the side with increasing *s*-coordinates, and in the other side they have opposing senses. The CECT(*C*) and FEM results agree quite well, and render lower W_{fric} values than the non-conformal calculation. The differences between the CECT($\hat{B}$) and CECT(*C*) results are explained in §6.2.4.2, in the discussion of Figure 6.35b.

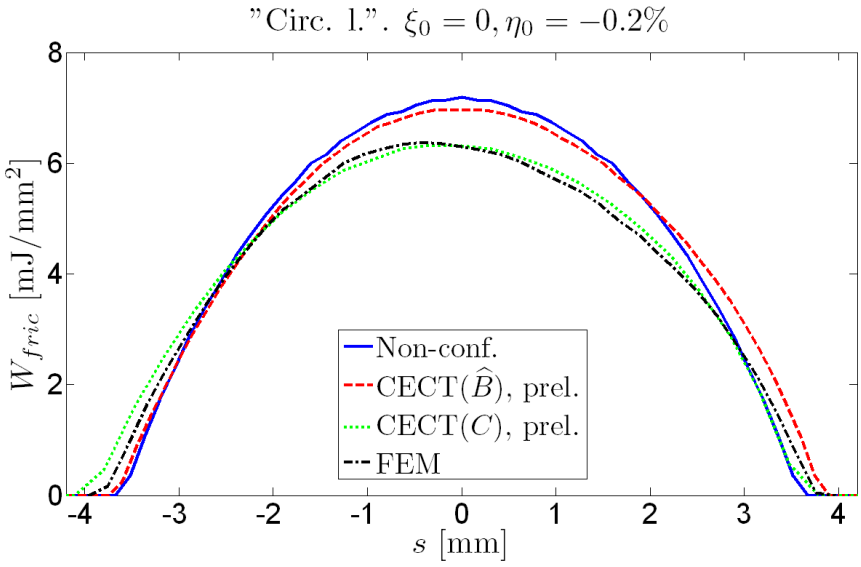


Figure 6.18. Lateral distribution of W_{fric} for case "circ. l." in steady rolling with $\xi_0 = 0$ and $\eta_0 = -0.2\%$.

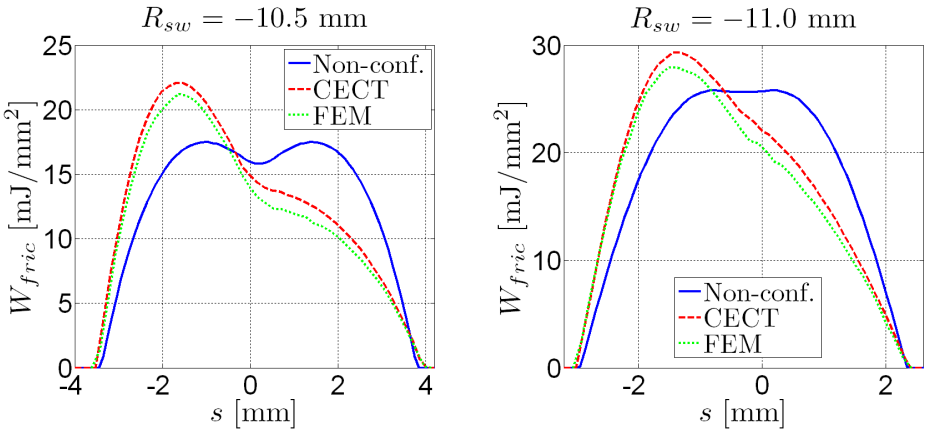
Figure 6.19 shows the W_{fric} distributions for different cases with geometry similar to case "circ. a." except for R_{sw}, which is given the value indicated in the figure in each case, and with approximate δ_0 of 45º. In all the cases shown in this figure, $\xi_0 = \eta_0 = 0$, and $\psi = 0$. The results of different non-conformal and conformal contact calculations are included in the figure. These are the results obtained with the exact contact theory without taking into account the effects of conformity, marked as "Non-conf.", the results obtained with CECT and approximate $\hat{B}$ ICs, and the results obtained with FEM.

While the W_{fric} distributions obtained with the non-conformal calculation are symmetric laterally, the W_{fric} distributions obtained with the conformal analyses are clearly biased towards the side with decreasing s coordinates, which corresponds to the side with increasing contact angles, and typically to the gauge side of the track. This is caused by the following factors:

- The higher slip velocities in this side (in this case with $\xi_0 = 0$), resulting from the non-linear variation of rolling radii, with faster variation at higher contact angles. These higher slip velocities can give rise to more severe wear regimes, see e.g. [Lim 1987], [Lewis 2004] and [Vuong 2011], and the resulting lateral distributions of wear may be even more biased than the frictional work distributions.
- The larger dimensions of the contact patches in this side, due to the lower wheel longitudinal curvatures there (cf. §6.2.1.1).

As the level of conformity decreases with increasing R_{sw} in these cases, the lateral distributions of W_{fric} obtained with the conformal analyses tend to converge to those obtained with the non-conformal analyses as expected, becoming more symmetric across the width of the contact. Still, a noticeable difference remains between the non-conformal and conformal analyses even in the case with $R_{sw} = -15.0$ mm as may be seen in Figure 6.19d, this case having a $\Delta\delta$ of just about 13.5º.

A further point to note is that a good agreement is found between the results obtained with CECT and FEM, although the latter yield somewhat lower levels of W_{fric}.



(a) R_{sw} = –10.5 mm. (b) R_{sw} = –11.0 mm.

(c) R_{sw} = –12.0 mm. (d) R_{sw} = –15.0 mm.

Figure 6.19. Lateral distributions of W_{fric} for cases similar to "circ. a." in steady rolling, with $\xi_0 = \eta_0 = 0$, $\psi = 0$ and different R_{sw}.

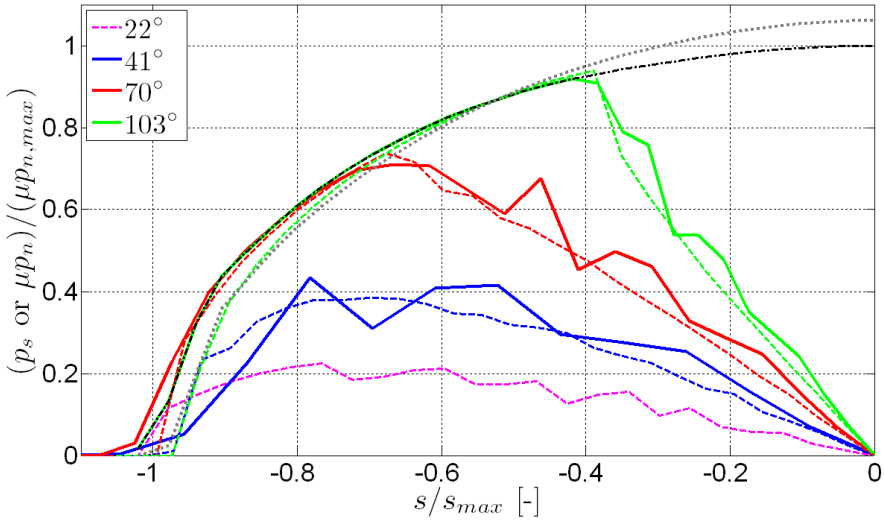
6.2.1.5 Contribution of the approach to the lateral rigid slip

Besides the previously described factoring of d with the $\cos(\Delta\alpha)$ term, the variation of the normal n and tangential s directions in the contact surface gives rise to a component of d along s, equal to $\Delta d \sin(\Delta\alpha)$, as explained in §4.1.1.3 (cf. Figure 4.12 and Eq. (4.24)). This component contributes to the rigid slip in s direction, or to its velocity w_s. By way of example, Figure 6.20 to Figure 6.22 show the tangential stresses and contours of adhesion and slip zones obtained in static compression cases with the "circ. s." and "circ. l." geometries, with $\psi = 0$,

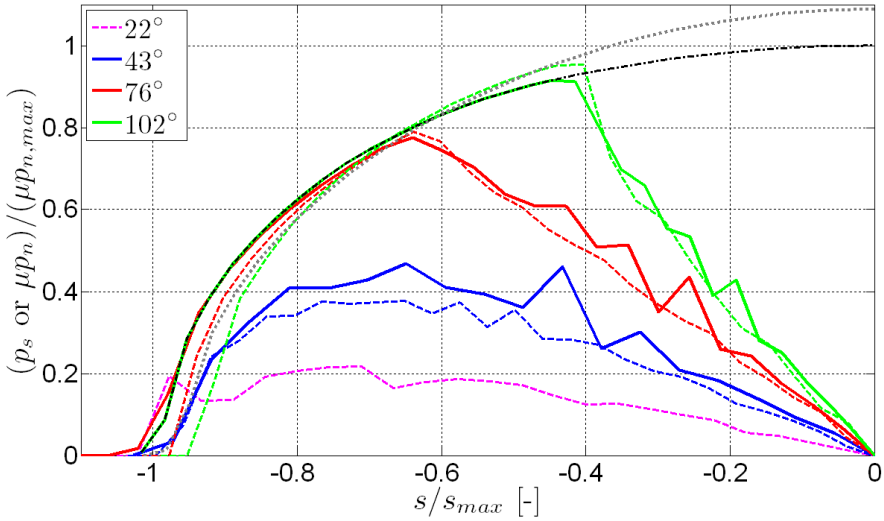
symmetric in longitudinal and lateral directions. The value of μ in these cases is 0.30, as in the previous examples of frictional contact. The load N is applied in a monotonously increasing way, from an initial zero value. Evidently, a purely non-conformal contact model would not capture these tangential stresses.

Figure 6.20 shows the lateral distributions of lateral tangential stress p_s across the middle section of the contact patch, for different $\Delta\delta$-s encompassed by the contact patch, computed both with FEM, and with CECT with approximate B ICs. Only one-half of the section is shown, as the tangential stresses are anti-symmetric with respect to $s = 0$. The relationship between $\Delta\delta$ and load N for the frictionless compression cases with these geometries are shown in Figure 5.51a and Figure 5.52a.

The results are presented in non-dimensional form in Figure 6.20. The tangential stresses are non-dimensionalized dividing them with the maximum traction bound in each case, i.e., the maximum normal pressure $p_{n,max}$ multiplied by μ. The non-dimensional values of the abscissa axes of the graphs are the lateral s coordinates divided by the maximum s coordinate in the contact patch in each case. The results used to non-dimensionalize the variables are those computed with FEM, except for the cases with lowest $\Delta\delta$-s, where the results computed with CECT have been used. In addition to the p_s curves, the curves of non-dimensional normal pressures at the same cross-section computed with FEM are also shown in the figure, for the situations with the highest $\Delta\delta$-s in each case, with an approximate $\Delta\delta$ of 100º. The normal pressures are non-dimensionalized dividing them by the maximum normal pressure in each case. The normal pressure curves for both the described frictional compression cases and for equivalent frictionless cases are shown.



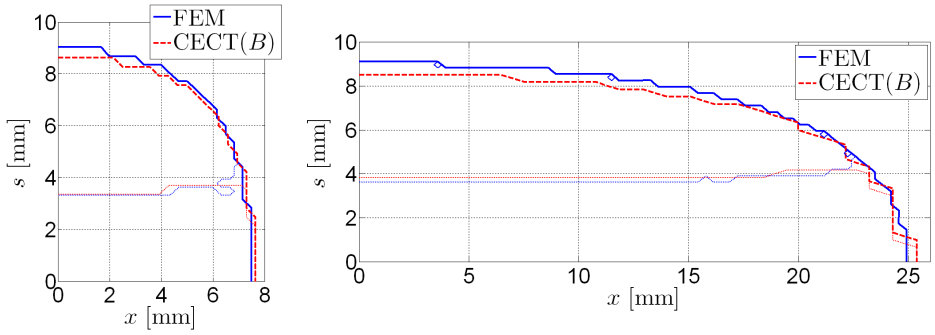
(a) Case “circ. s.”.



(b) Case “circ. l.”.

Figure 6.20. Non-dimensional lateral tangential stresses across the middle section of the contact patch in frictional static compression cases with monotonically increasing load. Results computed with FEM in solid lines and results computed with CECT in dashed lines. Non-dimensional normal pressure curves are included for $\Delta\delta \approx 100$º, with friction in black dash-dotted lines, and without friction in grey dotted lines.

The relative width of the adhesion zone decreases, and the level of the p_s stresses increases with $\Delta\delta$, as can be seen in Figure 6.20. Qualitatively the results are similar for both cases "circ. s." and "circ. l.". There is also a good agreement in the p_s curves obtained with FEM and CECT in each case. This good agreement is maintained in the contours of the adhesion and slip zones, depicted in Figure 6.21 for the situations with the highest $\Delta\delta$ in each case. As can be seen in the Figure 6.21, the width of the adhesion zone, located in the centre of the contact patch, is nearly constant along the contact patch, with the boundary between the adhesion and slip zones being nearly parallel to the longitudinal direction both in case "circ. s." and "circ. l.".



(a) Case "circ. s.". (b) Case "circ. l.".

Figure 6.21. One quarter of the contact patches (symmetric in the lateral and longitudinal directions) with contours of adhesion and slip zones in frictional static compression cases with $\Delta\delta \approx 100^{o}$.

Figure 6.22 shows the distribution of tangential stresses obtained in the compression case with the "circ. l." geometry for a $\Delta\delta$ of 43º, corresponding to a load N of 80 kN, computed with FEM.

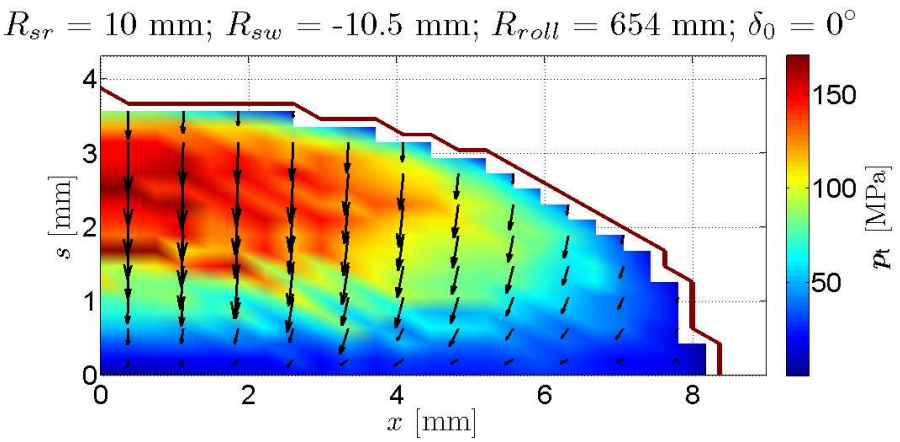


Figure 6.22. Tangential stresses in conformal static compression generated by the approach *d*. Case "circ. l." with N = 80 kN.

The tangential stresses shown in these figures are a direct consequence of the components of d in s direction. d has a final value of 89 μm for the case of Figure 6.22. In a similar way, relative displacements of the contacting solids in other directions in the YZ plane generate variations in the lateral profile of h as well. These effects may be considered a form of coupling between the normal and tangential contact problems.

Regarding the normal pressures, there are appreciable differences between the distributions of normal pressures obtained with and without friction, as can be observed in Figure 6.20. The maximum normal pressures tend to decrease with friction in the interface. This is due to the coupling effect between the normal and the tangential parts of the contact problem that is explained in §6.2.2.3.

6.2.2 Influence coefficients

The effect of the non-flatness of the contacting surfaces on the ICs is treated in depth in Chapter 5. Here the resulting effects on the contact solutions are discussed.

6.2.2.1 Decrease of ICnn-s with surface orientation change

The combined IC_{nn} ICs decrease along s more rapidly in curved than in flat surfaces as shown in §5.2.3.6.1. In other words, the influence of the normal pressures is more localized laterally with curved contact surfaces than with

planar surfaces. The resulting effect is an increase of the normal contact pressure levels in the lateral extremes of the contact patch, and hence a widening of the contact patch, to satisfy the non-overlap condition between the contacting surfaces in these zones.

This effect is quantified in Figure 6.23 for the frictionless compression of two solids with the "circ. s." geometry. For this purpose, the figure compares the contact patch width as a function of N calculated with FEM on the one hand, and with CECT on the other, using in the latter case the half-space A_{nn} ICs instead of the real IC_{nn}-s of the non-planar solids. The reference results are the ones computed with FEM. As may be seen in the figure, with $\Delta\delta$ lower than aprox. 40º good results are obtained with the A_{nn}. With higher $\Delta\delta$, the contact patches obtained with the A_{nn} are narrower than the real ones. The figures also show the solutions obtained with CECT with other two sets of ICs: on the one hand with the approximated B_{nn} ICs according to Eq. (5.72), and on the other hand with the C_{nn} ICs of the non-planar solids of this case calculated numerically with FEM. A substantial improvement in the results obtained with the B_{nn} is verified with respect to the results obtained with the A_{nn}. It is verified as well that the results obtained with the C_{nn} are nearly coincident with the FEM results.

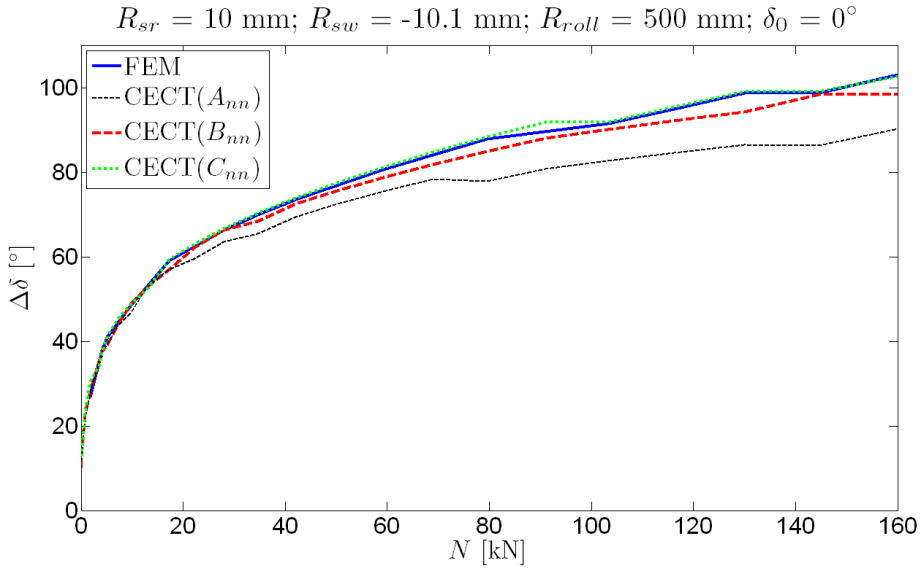


Figure 6.23. Effect of the IC_{nn}-s in the contact patch width.

6.2.2.2 Normal–tangential coupling: effects in the tangential part

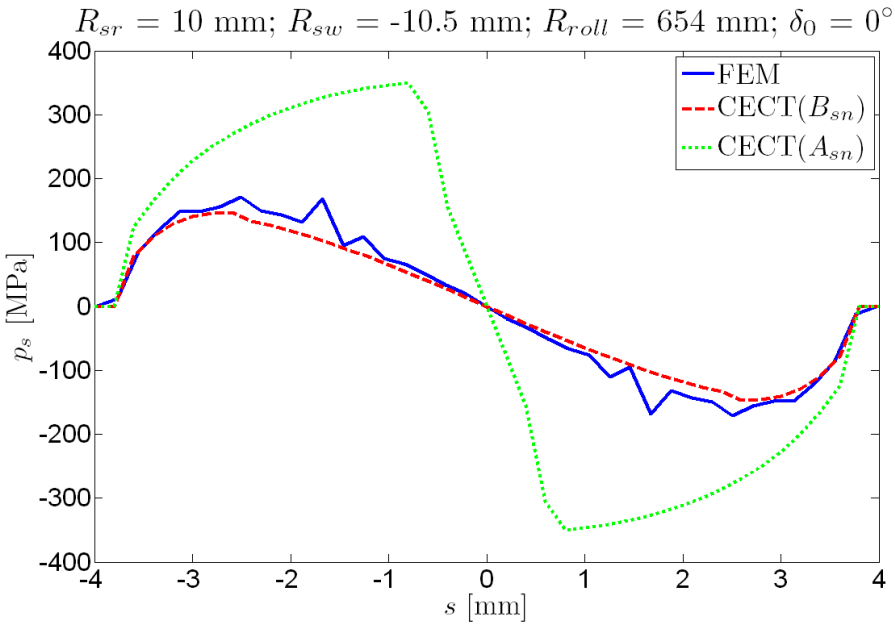
A notable feature of conformal contact derived from the variation of the ICs with respect to those of the half-space is the coupling between the normal and tangential contact problems. As the contacting surfaces are not flat, elastic quasiidentity no longer applies with elastically similar contacting materials. In other words, the combined IC_{sn} and IC_{xn} ICs as well as their reciprocals are not zero even when the elastic mismatch constant K (cf. Eq. (2.115)) is zero. The effects of this coupling are normally higher in the tangential part, because the normal pressures are normally higher than the tangential stresses in steel on steel contact (the more so in rolling contact). Figure 6.24 illustrates two of these effects; one associated to the IC_{sn} ICs (Figure 6.24a) and another to the IC_{xn} ICs (Figure 6.24b).

Figure 6.24a shows the profile of the lateral tangential contact stresses p_s in the central cross-section of the contact patch (at $x = 0$) obtained in the static monotonous compression case with the "circ. l." geometry and final load of 80 kN. The solution calculated with FEM is shown on the one hand, considered as the reference. On the other hand, two solutions obtained with CECT are shown. In one of the CECT solutions the half-space A_{sn} ICs have been used, which in this case are zero as the materials of both contacting bodies are elastically similar. In the other CECT solution, the approximated B_{sn} ICs have been used, calculated according to Eq. (5.72). To measure the influence of the IC_{sn}-s in this case, this is the only difference between both CECT solutions, having used for the rest of ICs the same approximated B_{ij}-s in both of them (with `idgt` = 0, cf. §4.1.2.1.1).

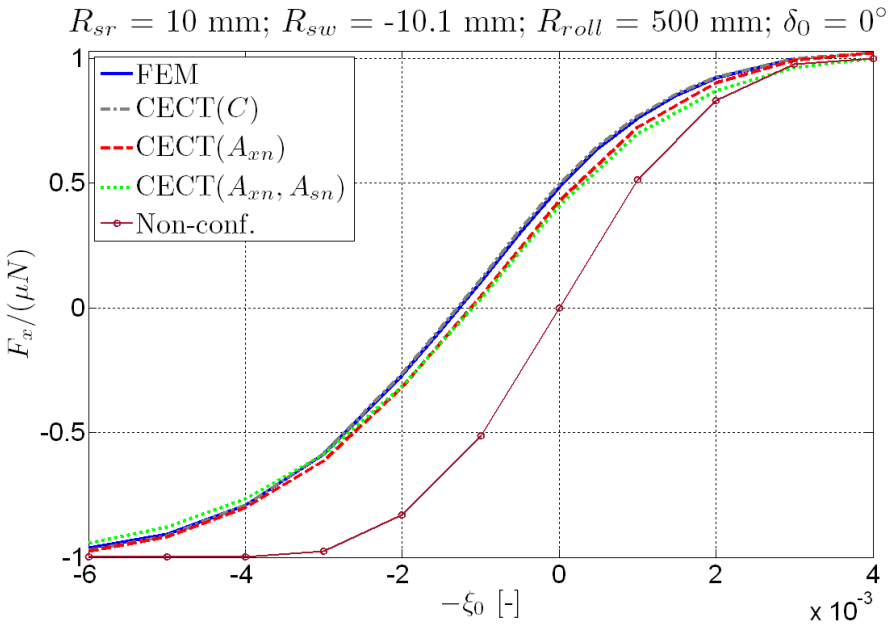
The IC_{sn}-s are the ICs which present the greatest deviations with respect to the half-space, as shown in §5.2.3.6.2, and their effect in the tangential contact is very considerable, as may be seen in Figure 6.24a. The normal contact pressures, through the IC_{sn}-s, generate elastic displacement differences in lateral direction u_s which cancel out to a great extent the rigid displacements in that direction associated to d, discussed in §6.2.1.5. As a result, the calculated p_s using zero IC_{sn}-s (that is, neglecting the influence of the normal pressures on the tangential contact) are much higher than the values calculated with FEM, being saturated in most of the contact width. The differences are noteworthy already in cases with conformity levels quite lower than the considered one. On the

other hand, it may be seen that the solution obtained with CECT and the B_{sn} ICs approaches quite well the FEM solution, which is indicative of the quality of the approximation of the B_{sn}.

Applying the rationale of the surface orientation based approximation of the ICs exposed in §5.2.3.1, it may be understood how the normal pressures generate the mentioned u_s. For example, considering the effect of the normal pressures of the central part of the contact (the greatest in magnitude), acting in opposite sense in each of the contacting solids, it may be seen that in positions away from the contact centre they give rise to components along s which tend to push towards the centre the points of the concave surface, and away from the centre the points of the convex surface.



(a) Effect of the IC_{sn}-s counteracting the rigid shifts associated to d in static compression case.



(b) Effect of the IC_{xn}-s offsetting the longitudinal creepage–creep force curve.

Figure 6.24. Effects of the normal–tangential coupling in the tangential contact.

Figure 6.24b shows the longitudinal creepage–creep force curve for steady rolling with the "circ. s." geometry, with zero ψ and η_0, and the same normal load and coefficient of friction as in the cases of Figure 6.15 (N = 80 kN and μ = 0.30). The curve calculated with FEM is shown on the one hand, and on the other three curves calculated with CECT and three different sets of ICs, and one curve calculated with the exact contact theory without taking into account any effect of conformity. The latter curve, designated as "Non-conf.", has been calculated assuming planar contact, considering semi-Hertzian geometry with the vertical undeformed distance profile of the real surfaces in the central cross-section of the contact (at x = 0), and using the creepages of the initial contact point in the whole contact patch. One of the CECT curves, designated as "CECT(C)", has been obtained with the C_{ij} ICs calculated numerically with FEM for the non-planar solids of this case. Another of the CECT curves, designated as "CECT(A_{xn})", has been obtained with the same set of C_{ij} ICs except for the IC_{xn}-s, which have been set to 0, making them equal to the A_{xn} ICs for this case with elastically similar materials. In the third CECT curve,

designated as "CECT(A_{xn}, A_{sn})", in addition to the IC_{xn}-s, the IC_{sn}-s have been set to 0 as well, thus completely disabling the influence of the normal part on the tangential part of the contact problem. Similar longitudinal creepage–creep force curves for other rolling contact cases are shown in Figure 5.60a and Figure 5.63a, with CECT results obtained with different variants of approximated B_{ij} ICs. The approximated B_{ij} ICs are not included in this comparison, as they are not a good approximation for the IC_{xn}, as shown in §5.2.3.6.2 (except for empirically adjusted $\hat{B}'_{xn}$ ICs).

The shown longitudinal creepage–creep force curves (except for the "Non-conf." curve) do not pass through the (0, 0) point in the f_x–ξ_0 graph due to the convention adopted to define ξ_0, in the initial contact point where the wheel rolling radius is the minimum of the contact patch in this case. As could be expected, the correspondence of the CECT curve obtained with the complete C_{ij} set with the FEM curve is very good.

On the other hand, the "CECT(A_{xn})" curve presents an offset towards the right in the f_x–ξ_0 graph. Consequently, for a given ξ_0, the f_x of this curve is lower. The reason of this effect is the differential stiffness of both contacting bodies, treated in §5.2.3.5. A similar effect is produced in non-conformal contact when both contacting bodies have dissimilar materials with different stiffness. The contact compressive normal pressures push the points of the surface towards where they are applied, and these displacements are bigger in the rail (the convex body in this case) than in the wheel. Therefore, due to the action of the normal pressures, in the front part of the contact patch the particles of the wheel surface are advanced (in the rolling direction) with respect to the homologous rail particles, and the opposite happens in the rear part of the contact patch. The net effect of the normal pressures is similar to having an additional tractive longitudinal creepage of the wheel particles with respect to the rail particles, during their traversal of the contact patch from the leading to the trailing edge. This is illustrated in Figure 6.25, where a pair of homologous wheel and rail contacting particles are represented at 3 instants $t_1 < t_2 < t_3$ in different positions of the contact patch, for a wheel advancing in the positive X axis. The sense of the rotation velocity $\boldsymbol{\omega}$ of the wheel is also represented in the figure.

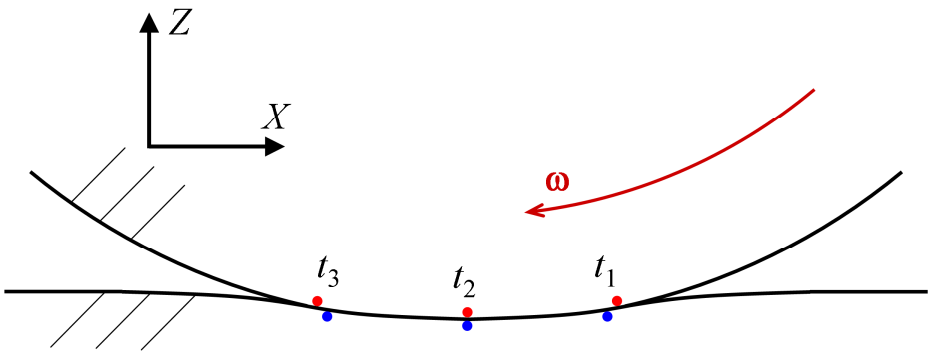


Figure 6.25. Longitudinal creepage effect induced by differential longitudinal stiffness between wheel and rail.

This reasoning is valid for planar contacts. In non-planar contact, the picture is not so clear, because the IC_{xn}-s reverse their sign not only at $x = 0$, depending on the change of orientation of the surface from the point of load application (cf. e.g. the C^1 and C^2 curves of Figure 5.41a and Figure 5.42a). Nevertheless, in the considered case, the effect is in the same sense as anticipated with the reasoning for planar geometries in the majority of the contact patch, and only in the most lateral parts of the contact patch this effect is reversed.

It is noted that also in this case a concave wheel rolling on a convex rail has been considered. In the case of a convex wheel rolling on a concave rail, the rail would be stiffer than the wheel in x direction, and the offset between the longitudinal creepage–creep force curves illustrated in Figure 6.24b would happen in the opposite sense.

To further assess the role of the IC_{xn} ICs, the longitudinal elastic displacement differences u_x are decomposed into the contributions from the p_x, p_s and p_n stresses, according to Eq. (6.23). In this equation, $\mathbf{\{u_{xj}\}}$ are the contributions to $\mathbf{\{u_x\}}$ from the stresses acting in j direction, with $j \in \{x, s, n\}$. The different $\mathbf{\{\cdot\}}$ vectors of the equation collect the values of all the APs of the contact mesh, and the $\mathbf{[IC_{xj}]}$ matrices are referred in the same way to the contact mesh. In the equation the $\mathbf{\{u_{xj}\}}$ elastic displacement differences are defined as the multiplication of the $\mathbf{[IC_{xj}]}$ matrices with the $\mathbf{\{p_j\}}$ stress vectors in the space domain for clarity, though in practice their computation is carried out more efficiently in the Fourier domain, as explained in §4.1.1.4.2.

$$\begin{aligned}\{\mathbf{u_x}\} &= \{\mathbf{u_{xx}}\} + \{\mathbf{u_{xs}}\} + \{\mathbf{u_{xn}}\} \\ &= [\mathbf{IC_{xx}}]\,\{\mathbf{p_x}\} + [\mathbf{IC_{xs}}]\,\{\mathbf{p_s}\} + [\mathbf{IC_{xn}}]\,\{\mathbf{p_n}\}\end{aligned} \tag{6.23}$$

Figure 6.26 shows the gradients along the longitudinal direction of the different contributions to the u_x displacements, decomposed according to Eq. (6.23), for the tractive rolling contact case of Figure 6.15a (with $\xi_0 = -0.05\%$), along the contact patch centreline (at $s = 0$). The longitudinal gradients $\nabla_x u_{xj} = \partial u_{xj} / \partial x$ are shown instead of the absolute u_{xj} displacements, because the longitudinal gradients enter directly in the basic kinematic equation of the tangential part of the rolling contact problem (Eq. (1.7)).

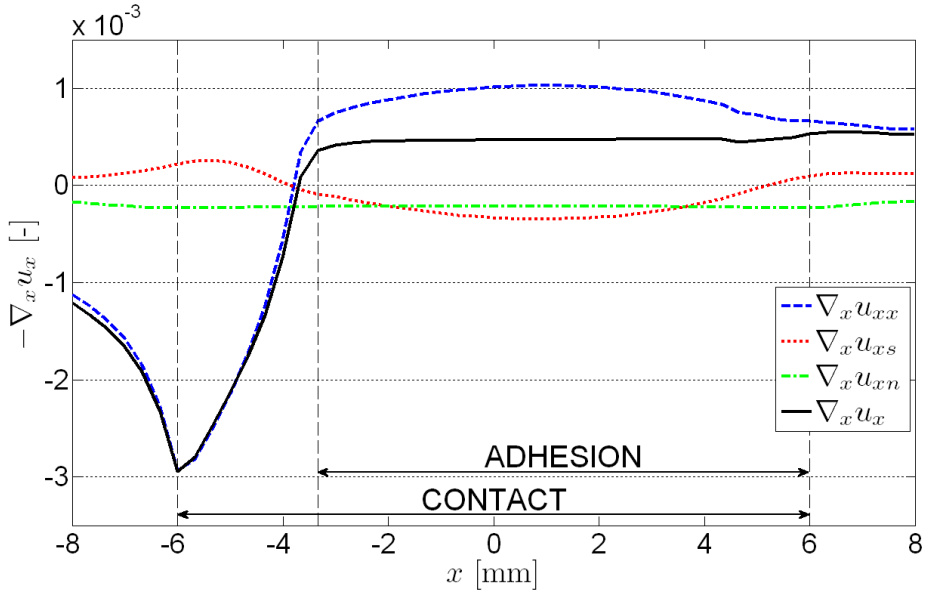


Figure 6.26. Longitudinal gradients of u_{xj} displacement differences along the contact centreline for the tractive rolling contact case of Figure 6.15a.

The whole contact length and the adhesion zone are marked in Figure 6.26. As can be seen in the figure, the total longitudinal gradient $\nabla_x u_x$ (with minus sign) is equal to the imposed rigid longitudinal creepage in the adhesion region, which has the constant value of $\xi_0 = -0.05\%$ in the contact centreline, except for small numerical errors. The p_x stresses contribute the most to this gradient, as could be expected. And not only that, but the $\nabla_x u_{xx}$ contribution due to the p_x

stresses reaches almost twice the value of the imposed rigid longitudinal creepage in the central part of the adhesion area, to compensate the opposing effect of the p_s and p_n stresses. The $\nabla_x u_{xn}$ contribution is roughly constant in the whole contact patch centreline, while the $\nabla_x u_{xs}$ contribution reaches its maximum in the central part of the adhesion zone. Both of these contributions are significant in relation to the imposed reference ξ_0.

Lastly, the slopes of the different creepage–creep force curves in Figure 6.24b are examined. The "CECT(A_{xn}, A_{sn})" curve presents slightly smaller slopes than the "CECT(A_{xn})" curve. The reason is the increase of the p_s associated to the use of the A_{sn}-s illustrated in Figure 6.24a, consuming a bigger part of the frictional capacity and limiting the remaining capacity for the p_x. The FEM and CECT curves present considerably smaller slopes than the "Non-conf." curve for the same reason: in the non-conformal calculation, there are no lateral rigid slip velocities, and consequently the tangential stresses are nearly aligned with the longitudinal direction, meaning that the whole frictional capacity is acting in this direction. The decrease of the slopes of the creepage–creep force curves with respect to the non-conformal calculation could be pointed out as a geometrical effect of conformity on the rigid slip velocity field. This is prone to be noted mainly in cases with low mean contact angles and low spin reference creepage. Other creepage–creep force curves are examined in §6.2.4.

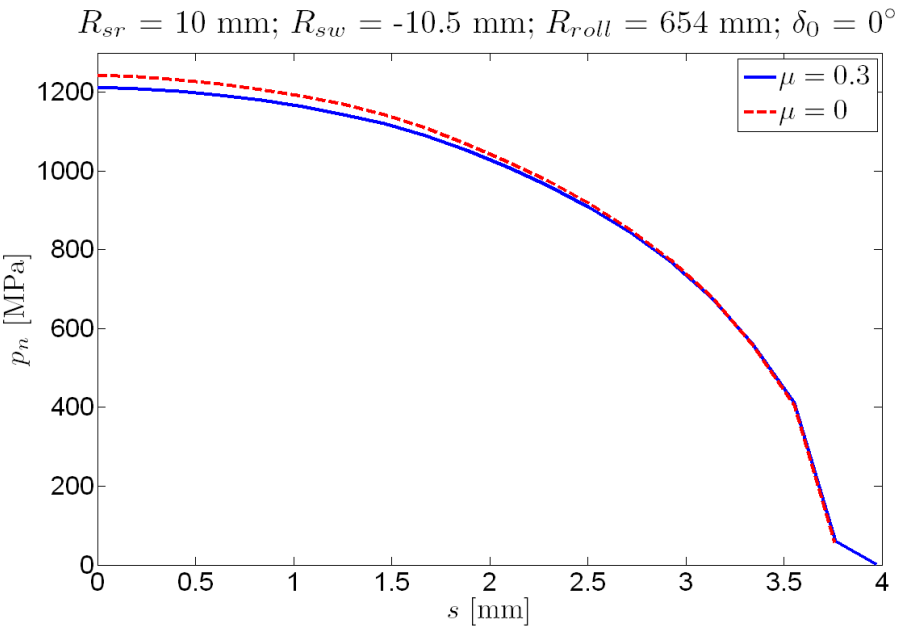
6.2.2.3 Normal–tangential coupling: effects in the normal part

Figure 6.27 illustrates two effects of the normal–tangential coupling on the normal part of the contact problem, associated to the IC_{ns} ICs.

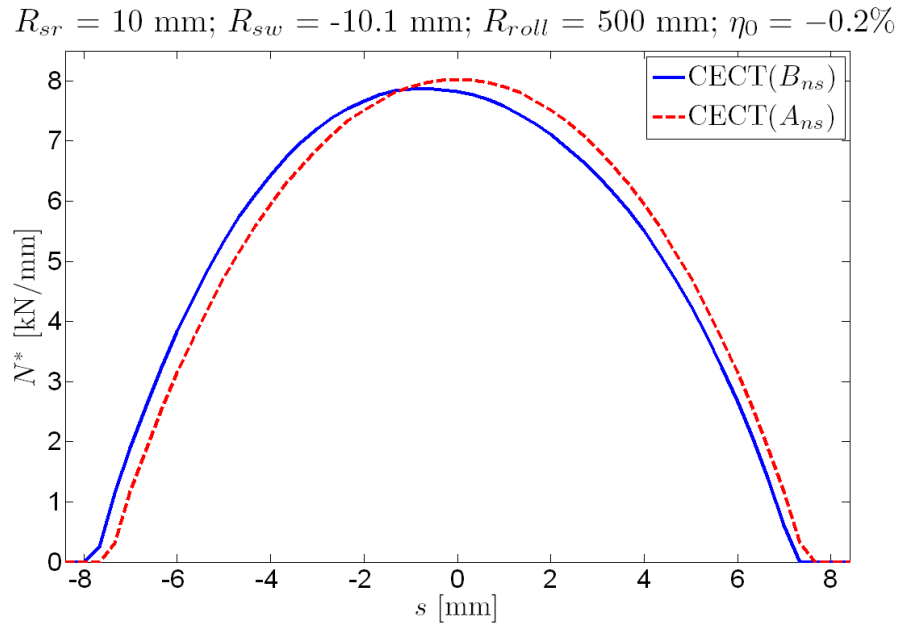
Figure 6.27a shows the profile of the normal pressures p_n in the central cross-section of the contact patch (at $x = 0$) obtained in normal compression with the "circ. l." geometry, $\mu = 0.3$, and a final load of 80 kN, together with the profile corresponding to the compression case with the same geometry and load but without friction. The tangential stress distribution for the frictional case is represented in Figure 6.22. The p_n profile obtained in the case with friction ($\mu = 0.3$) is flatter than the one obtained without friction ($\mu = 0$), and the maximum p_n is aprox. 2.5% higher in the frictionless case. Part of the difference is because the tangential stresses p_s sustain part of the normal load N in the case with friction. Specifically, in this case the p_s sustain 1.86% of N. Assuming as first

approximation Hertzian proportionality between the resultant force of p_n and the maximum p_n, this explains aprox. a 0.6% of the difference between the maximum p_n values in the calculations with and without friction. The rest is due to the normal elastic displacements u_n produced by the p_s stresses. In the central part of the contact, the p_s at both sides of the contact patch, acting in opposite senses in each contacting body, render an inwards normal component both in the concave and in the convex body, altering the h profile in such a way that the p_n in the central part (at $s = 0$) are reduced. In the considered case of symmetric compression, the effects of the p_s at both sides of the contact patch are summed up. In other cases of rolling contact calculated for the same geometry with different φ_0 and η_0 values, this effect has been more limited, even with higher levels of tangential stress.

Figure 6.20 shows the p_n profiles for cases with higher $\Delta\delta$, and with higher resulting differences between the frictional and frictionless cases due to the normal–tangential coupling effect described above. The difference between the maximum p_n obtained in the frictional and frictionless cases reaches 6.3% in case “circ. s.” (Figure 6.20a), and 8.9% in case “circ. l.” (Figure 6.20b). The coupling effect is stronger in case “circ. l.” due to the higher cross-influence between different longitudinal strips of the contact patch with more elongated contact patches, as depicted in Figure 5.57. In these figures, the normal loads have been adjusted in the frictionless cases so that the resultants of the normal pressures are the same as in the corresponding frictional cases, to make a more direct comparison between the cases with and without friction. I.e., the total normal loads are higher in the frictional cases, due to the contribution of the tangential stresses. Therefore, the differences in the p_n profiles depicted in Figure 6.20 are entirely due to the normal–tangential coupling effect via the IC_{ns} ICs. In the frictional cases with $\Delta\delta \approx 100^{\circ}$, the proportion of the normal load sustained by the tangential stresses amounts to 8.9% of the total normal load in case “circ. s.”, and to 8.5% in case “circ. l.”.



(a) Flattening of the lateral profile of normal pressures in symmetric static compression case.



(b) Lateral shift of the contact with net lateral contact force.

Figure 6.27. Effects of the normal–tangential coupling in the normal contact.

Figure 6.27b graphs the profile of normal load per unit contact patch width in s direction, N^*, for the rolling contact case with yaw angle shown in Figure 6.15c. This case is not symmetric in lateral direction, having a net lateral force transmitted in the contact. The figure shows two curves, both computed with CECT and the approximated B_{ij} ICs for non-planar solids, except for the IC_{ns} ICs which have been set to zero in the curve designated as "CECT(A_{ns})". There is a lateral shift between both curves shown in the figure, as a result of the action of the p_s via the IC_{ns}-s, being the curve obtained with zero IC_{ns}-s (as would correspond to non-conformal contact) nearly symmetric in lateral direction with respect to $s = 0$. Here the FEM results are not included, as they may contain a lateral shift of the contact patch not directly comparable with the CECT results, due to the global deflections present in the FE model (cf. Figure 4.39).

The observed effect of the IC_{ns}-s in this case may be explained again with the surface orientation reasoning in which the B approximation of ICs is based. In the side of the contact patch towards which the concave body pushes the convex one, the p_s stresses of most of the contact width have an outwards normal component in both bodies, and the opposite happens in the other side of the contact patch. In this way, when the p_s have a predominant direction in the contact (that is, when there is a net lateral contact force), they tend to raise the surfaces of both bodies in the side of the contact patch towards which the concave body pushes the convex one, and to sink them in the other side. Due to the described deformations, a lateral shift of the contact is produced towards the side the concave body pushes the convex one. The described u_n are represented schematically in Figure 6.28 for a curved surface subject to the action of some p_s in the same direction in the whole contact width. The undeformed surface is represented in solid black line, and the deformed surface in dashed red line. It must be pointed out that here reference is made to the deformations produced exclusively by the "normal components" of p_s (analogous to the P_{nI} component of the force represented in Figure 5.28), and not to the ones due to the global deflections of the solids due to the remaining "tangential components" of p_s for example.

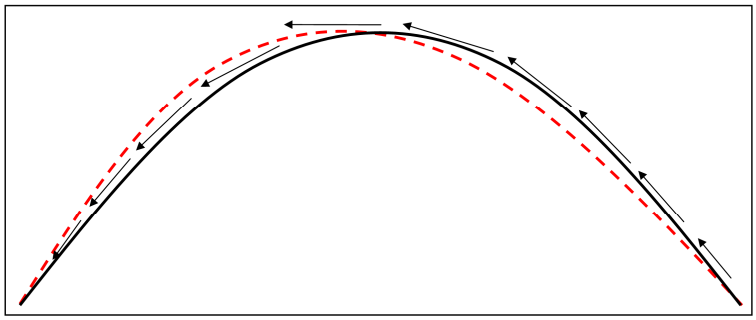


Figure 6.28. Normal elastic displacements in a curved surface caused by the "normal components" of the p_s stresses. View of *YZ* plane.

An effect analogous to the one illustrated in Figure 6.27b and Figure 6.28 may be anticipated in *x* direction, associated to the IC_{nx}-s. However, as the IC_{nx}-s are smaller in magnitude than the IC_{ns}-s with similar materials, this effect is smaller than the one shown in Figure 6.27b, and higher coefficients of friction are needed for it to become appreciable. Figure 6.29 illustrates this effect comparing the p_x distributions obtained along the contact patch centreline (at s = 0) in braking and tractive rolling contact situations for the same geometry, N and μ considered in Figure 6.15. The ξ_0 values of the braking and tractive situations are 0.6% and –0.4%, respectively, and $\eta_0 = 0$ in both cases. With these creepages, F_x is almost at its negative saturation value in the braking situation, and at its positive saturation value in the tractive situation. The results shown in Figure 6.29 have been obtained with CECT and numerically calculated C_{ij} ICs. The contact patch shifts in *x* direction towards where the convex body (the more compliant in longitudinal direction) pushes the concave one. In addition, the contact patch develops an asymmetry in *x* direction, becoming narrower in the side towards which it shifts, and wider in the other side. But with similar materials, this asymmetry becomes noticeable only for very high coefficients of friction (say $\mu >> 1$), which may not be realistic for steel on steel contact.

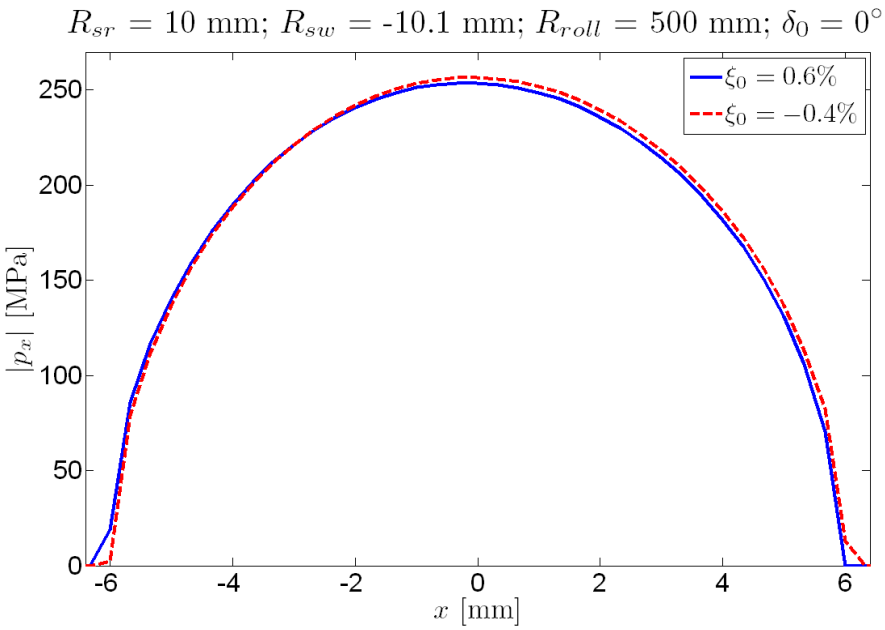
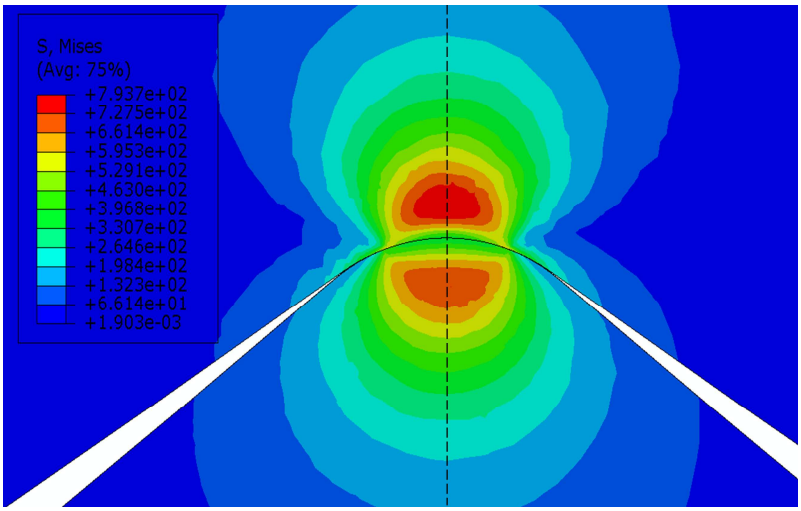


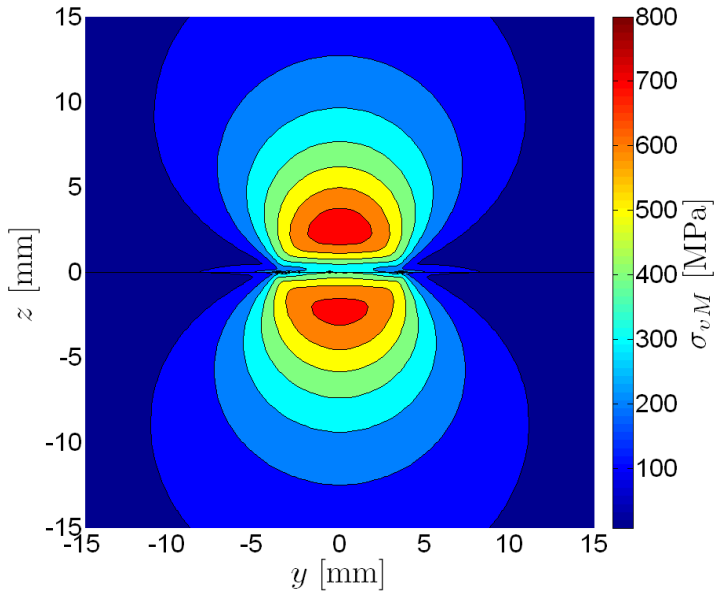
Figure 6.29. Effect of the normal–tangential coupling in the normal contact due to the IC_{nx} ICs. Longitudinal shift of the contact with net longitudinal contact force.

6.2.3 Subsurface stresses

The contact stresses affect differently to the concave and convex contacting solids, tending to be more unfavourable the subsurface stress state in the concave solid. As an example, Figure 6.30 and Figure 6.31 show the subsurface stresses in the symmetric static compression case considered in Figure 6.22. Figure 6.30 shows the von Mises stresses σ_{vM} in the central cross-section of the contact patch parallel to the *YZ* plane at $x = 0$, obtained with FEM on the one hand (subfigure a), and calculated analytically for the half-space on the other hand (subfigure b). For this latter analytical calculation, the contact stresses obtained for the same case with CECT are used, applying them to the flat surface of the half-space that is taken to represent each of the non-planar contacting solids. Figure 6.31 shows the von Mises stress curves along the normal direction in the central contact point (i.e., along the dashed line of Figure 6.30a), corresponding to each of the graphs shown in Figure 6.30.



(a) FEM results.



(b) Results for the half-space with the contact stresses computed for the same conformal contact case with CECT.

Figure 6.30. Subsurface von Mises stresses in conformal compression with elastically similar materials, in the central cross-section of the contact patch. Stresses in MPa.

Qualitatively, the subsurface von Mises stress field around the contact obtained for the non-planar solids with FEM, Figure 6.30a, resembles the field which

would be obtained for half-spaces in non-conformal contact (similar to the field shown in Figure 6.30b), despite the contact interface is different, changing from a plane in non-conformal contact to a curved surface in conformal contact. However, the subsurface stresses are higher in the concave solid.

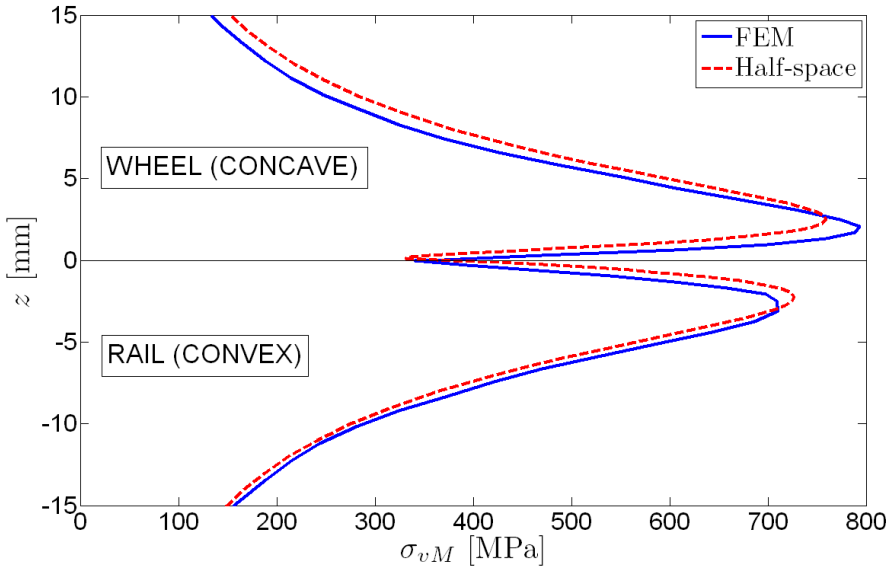


Figure 6.31. Subsurface von Mises stresses along the normal direction in the central contact point in the case shown in Figure 6.30.

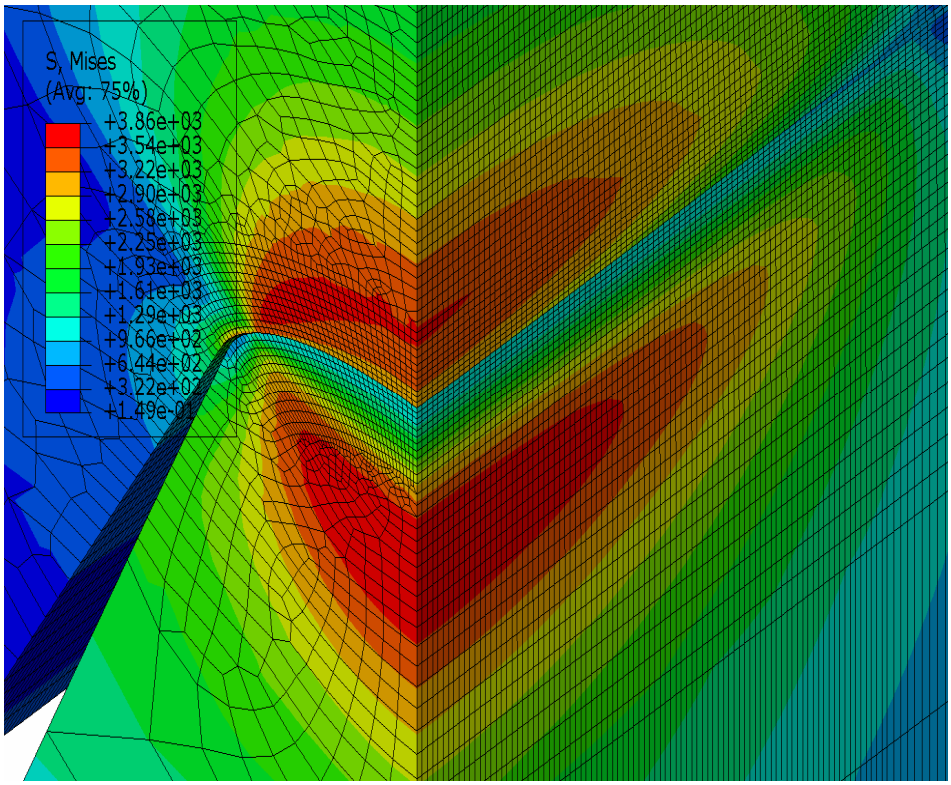
Looking at the stress curves for the half-space show in Figure 6.31, it may be seen that part of the differences between the stresses in each of the solids is explained by the combined effect of the normal pressures with the tangential stresses, with different sign in each solid. The tangential stresses acting on the concave surface, combined with the normal pressures, turn out to be more unfavourable than the tangential stresses of opposite sign acting over the convex surface combined with the same normal pressures. On the other hand, the stress curves obtained with FEM are more asymmetric than the ones obtained for the half-space with the same contact stresses. A greater stress concentration is produced in the concave solid, and the opposite happens in the convex solid. As a consequence, the subsurface stresses in the concave solid are higher than the ones corresponding to the half-space, and those of the convex solid are lower. The location of the maximum stress points is also different in the non-planar

solids, tending to be located closer to the surface in the concave solid, and at a greater depth in the convex solid.

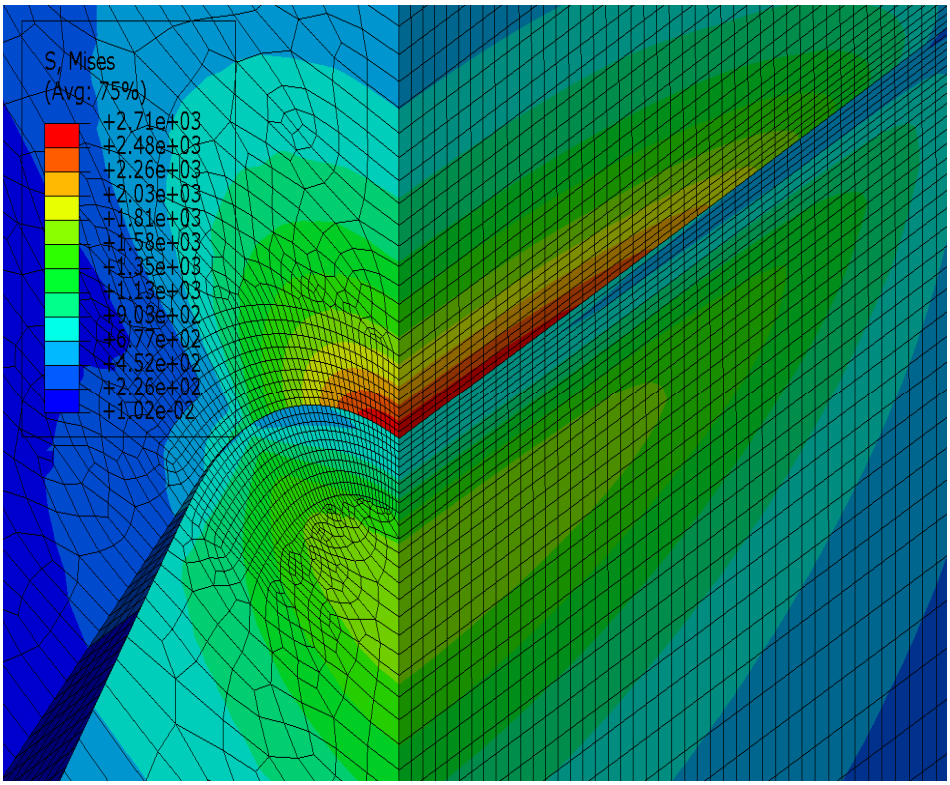
It has to be mentioned that the tangential stresses in the compression case considered here are relatively low in relation to the normal pressures. As the tangential stresses increase, the points with maximum von Mises stresses move towards the surface, and the maximum stress values at each of the contacting bodies tend to approach each other.

The subsurface von Mises stresses for two other conformal compression cases are shown in Figure 6.32, with different non-planar surface and contact pressure profiles. These are the "el. wide" and "el. tall" frictionless compression cases studied in §5.2.3.7.1, with elliptical cross-sections. Their geometrical parameters are listed in Table 5.9, and some pressure profiles at different load levels for each case are plotted in Figure 5.58. The "el. wide" case is shown in Figure 6.32a, and the "el. tall" case in Figure 6.32b. The normal load *N* applied in each case is indicated in the caption of each subfigure.

In the cases shown in Figure 6.32, the highest stresses also occur in the concave body. The differences between the stress levels in both bodies are higher in the "el. tall" case (Figure 6.32b). The "*s* components" of the normal pressures, with tensile action in the concave body, alter the direct stresses in *s* direction. As a result, in the concave body the stress state in the surface becomes less hydrostatic than in planar contact, and the point of maximum von Mises stress moves towards the surface. In case "el. tall", with a more peaky pressure profile (cf. Figure 5.58), the maximum von Mises stress is located at the centre, at the location of maximum pressure. Additionally, due to the faster contact angle variation around the root of the concave section in that case, the *s* direct stresses in that zone are altered more (to the extent that they become tensile), contributing to increasing the von Mises stress for the same pressure. In case "el. wide", with a flatter pressure profile, the stress state at the centre is comparatively milder, and the worst state is found at the contact edge, where tensile radial (i.e. directed along surface lines emanating from the contact centre) stresses exist (cf. Figure 7.8 of [Hills 1993]). These are also increased with respect to planar contact in the concave body in the *s*-*n* plane due to the aforementioned "*s* components" of the normal pressures.



(a) Case “el. wide” for N = 5300 kN.



(b) Case “el. tall” for N = 310 kN.

Figure 6.32. Subsurface von Mises stresses in frictionless conformal contact between wheels and rails with elliptical cross-section. Isometric view of cut at contact centre along longitudinal and lateral directions. Stresses in MPa.

6.2.4 Creepage–creep force curves

Here longitudinal and lateral creepage–creep force curves are examined for cases “circ. s.”, “circ. l.” and “circ. a.” described in Table 6.1. Normalized longitudinal and lateral contact resultant forces, f_x and f_s, are shown in the curves. These are defined as the corresponding contact resultant force divided by μN. Results obtained with FEM and with different versions of the exact contact theory are shown in each case. The ICs used with CECT are indicated inside parentheses in the legend of the graphs.

For the lateral creepage–creep force curves, the CECT curves also include the indication of whether a lateral preload is applied, equivalent to that applied in

the corresponding FEM calculation. The CECT cases without lateral preload are indicated with “no pr.”, and the ones with lateral preload with “prel.”. In the FEM calculations for the cases with lateral creepage, the lateral preload applied before the rolling phase is approximately half of the final lateral creep force in each case, previously estimated with CECT. This lateral preload has an effect on the lateral position of the contact patch as explained in §4.2.1.5 (cf. Figure 4.39), and hence on its mean contact angle, spin creepage and resultant contact forces. With a preload of half of the final lateral creep force, the contact patch remains approximately centred in its initial position. Without preload, the contact patch shifts laterally. For a concave wheel on a convex rail, this lateral shift is produced in the opposite direction of the resultant lateral force acting on the wheel. Lastly, the results obtained with the exact contact theory without taking into account any effect of conformity, as explained in §6.2.2.2, are indicated as “Non-conf.”.

6.2.4.1 Case “circ. s.”

Figure 6.33 shows the lateral creepage–creep force curves obtained for case “circ. s.” with zero ξ_0. Longitudinal creepage–creep force curves for this case are shown in Figure 6.24b, and tangential stress distributions for different creepages in Figure 6.15.

Similar comments as for Figure 6.35b in §6.2.4.2 may be made in this case, with higher differences between the different calculations being observed here, due to the higher conformity level. In this case, the *C* ICs used in the calculations with CECT correspond to the geometry of the FE model, which is the one plotted in Figure 5.32b. So the “CECT(*C*), prel.” and FEM curves match better than in Figure 6.35b.

To illustrate the sensitivity of the contact patch lateral shift to the lateral preload, Figure 6.34 shows the contact patch contours obtained with CECT and *C* ICs for the case with $\eta_0 = -0.3\%$, with and without lateral preload, together with the contact patch contour obtained with FEM and the same lateral preload as in the “CECT(*C*), prel.” calculation.

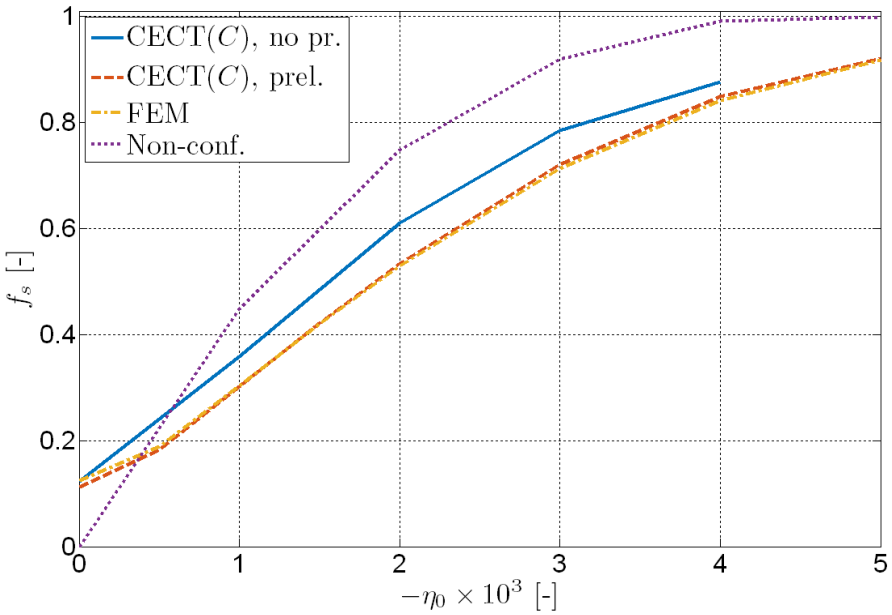


Figure 6.33. Lateral creepage–creep force curve for case "circ. s.". ξ_0 = 0.

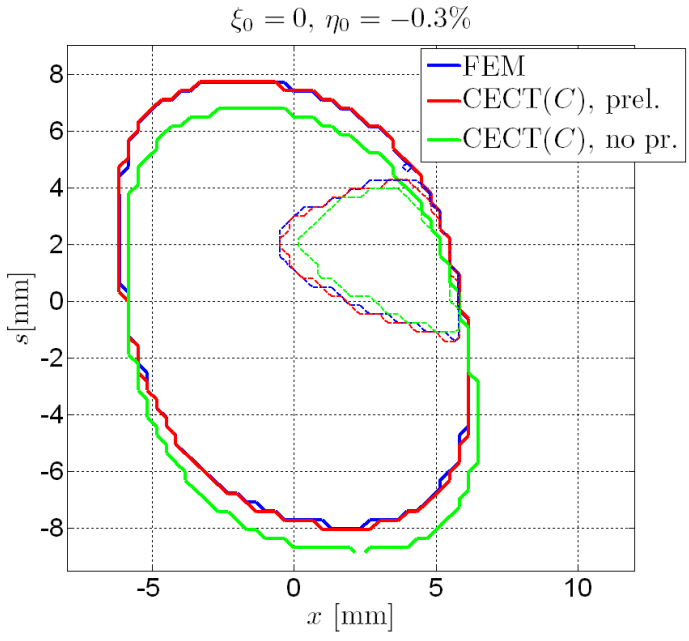
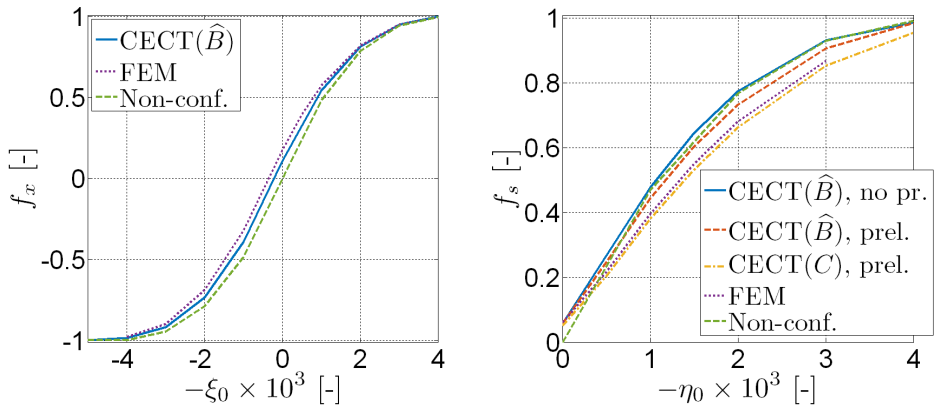


Figure 6.34. Contact patch contours obtained for case "circ. s." with ξ_0 = 0 and η_0 = –0.3%, with and without lateral preload.

6.2.4.2 Case “circ. l.”

Figure 6.35 shows longitudinal and lateral creepage–creep force curves for case “circ. l.”. Figure 6.36 shows the contact patch contours and tangential stress distributions obtained for two different creepage situations in this case, and Figure 6.37 the corresponding relative slip velocity distributions.



(a) Longitudinal. $\eta_0 = 0$. (b) Lateral. $\xi_0 = 0$.

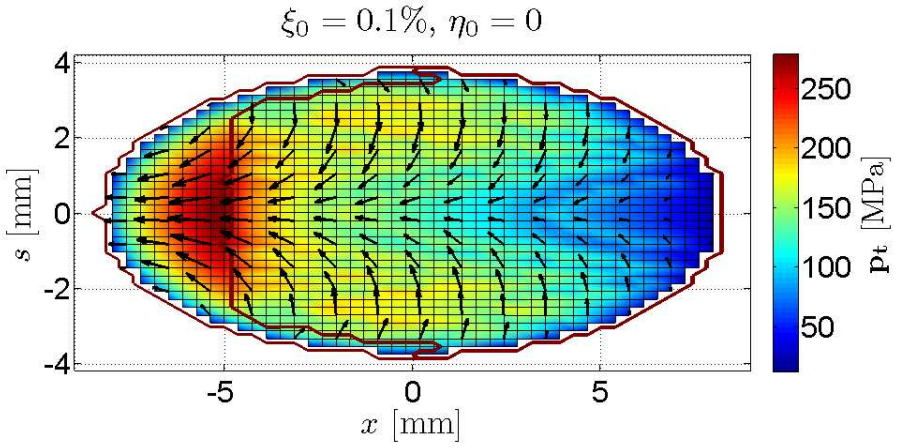
Figure 6.35. Creepage–creep force curves for case “circ. l.”.

Figure 6.35a shows longitudinal creepage–creep force curves obtained with zero η_0. Similar comments as for Figure 6.24b may be made in this case, with the difference that now the conformity level is smaller and consequently the differences between the different results are smaller. The FEM curve is slightly above the CECT($\hat{B}$) curve due to the differential stiffness between the contacting bodies. Both the FEM and CECT curves are above the curve obtained in the non-conformal calculation due to the convention adopted for the definition of ξ_0, as explained in §6.2.2.2. In this case, the differences in the slopes of the different curves are not significant.

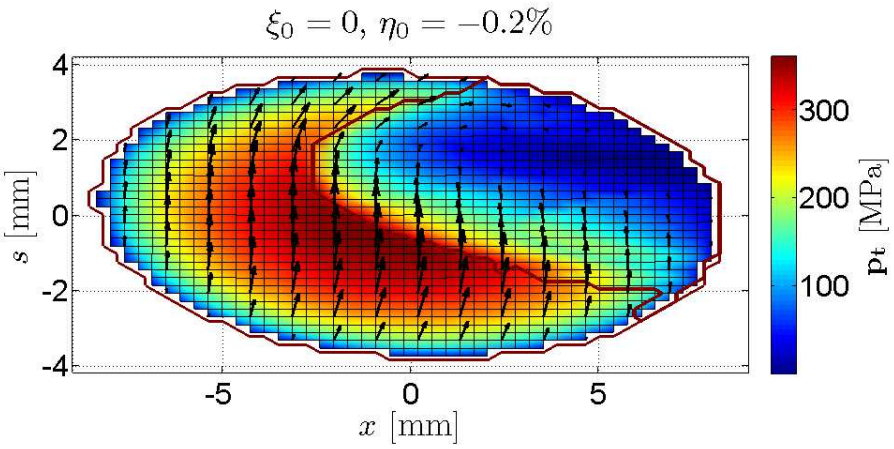
On the contrary, the lateral creepage–creep force curves shown in Figure 6.35b present significant differences in their slopes. The CECT curve with no preload has higher lateral forces for a given η_0 than with preload, due to the lateral contact shift and positive spin contribution. The “CECT($\hat{B}$), prel.” curve has significantly higher slopes than the FEM curve, even though the contact patches obtained for both curves are approximately aligned. This is attributed to the

greater lateral flexibility than predicted with the approximated *B* ICs, with the C_{ss} ICs having higher gradients in longitudinal direction (cf. Figure 5.37). The "CECT(*C*), prel." curve approaches better the FEM curve. The remaining differences between these two curves are due to the fact that the *C* ICs do not correspond to the cross-sectional geometry of the FE model used in this case. The *C* ICs correspond to the geometry shown in Figure 5.32b, with rail and wheel surface inclination angles δ_r and δ_w of 65º and 60º, respectively (cf. Table 5.6). In the FEM model used in this case, the values of δ_r and δ_w are 40º and 35º, respectively. The slopes of the "Non-conf." curve are significantly higher than in the conformal calculations.

The curves obtained with the conformal calculations do not pass through the (η_0, f_s) = (0, 0) point, because there is a non-zero contact resultant force in the *YZ* plane F_{lat}, which has the same direction as the load *N* for η_0 = 0, and approaches the perpendicular direction (which may be regarded as the contact lateral direction) as η_0 increases.

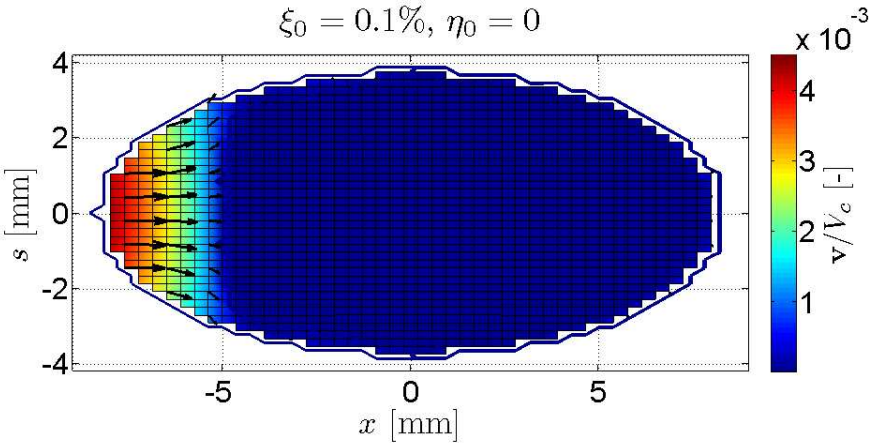


(a) ξ_0 = 0.1%, η_0 = 0.

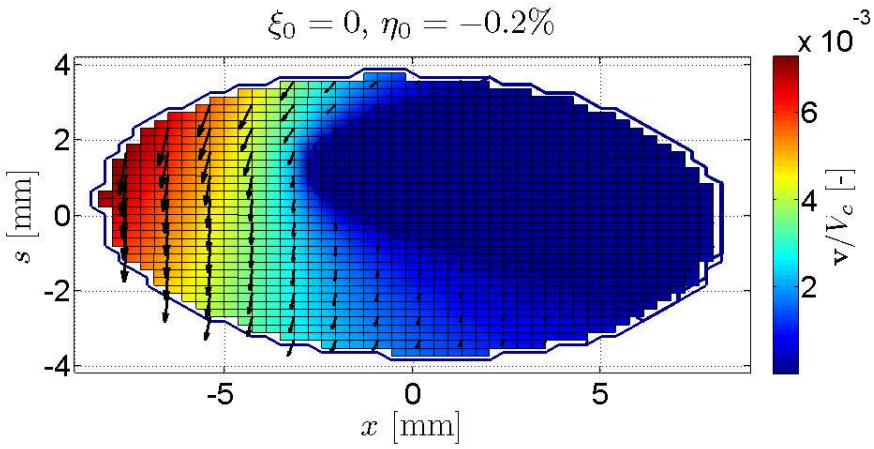


(b) ξ_0= 0, η_0 = –0.2%.

Figure 6.36. Tangential stress distributions in different creepage situations for case "circ. l." obtained with FEM.



(a) ξ_0 = 0.1%, η_0 = 0.



(b) ξ_0= 0, η_0 = –0.2%.

Figure 6.37. Relative slip velocity distributions in different creepage situations for case "circ. l." obtained with FEM.

6.2.4.3 Case "circ. a."

Figure 6.38 shows the lateral creepage–creep force curve for case "circ. a." with ξ_0 = 0. The FEM curve is slightly offset from the CECT($\hat{B}$) curve towards higher positive lateral forces. The curve obtained with the non-conformal calculation falls between the former two. The differences between the different curves are relatively small, with better agreement being found with negative η_0. The high spin coming from the high mean contact angle is the most determinant factor for this lateral creepage–creep force curve, and this is captured in the non-conformal as well as in the conformal calculations.

Most of the contact patch is in slip due to the high spin, as can be seen in Figure 6.39 for η_0 = 0.5%. The contact patch contour for this case is shown in Figure 6.10 marked as "ψ = –3.54 mrad", and the contours of undeformed distance in Figure 6.7. Another consequence of the high spin is a decreased sensitivity of the lateral resultant contact force to the preload before the rolling phase in the FEM calculation. This is shown by the small black point at (η_0, f_s) = (0, 0.55) in the graph of Figure 6.38, which marks the resultant contact force obtained with FEM omitting the preload. As can be seen, it is very close to the FEM curve in the graph, which has been obtained with preload.

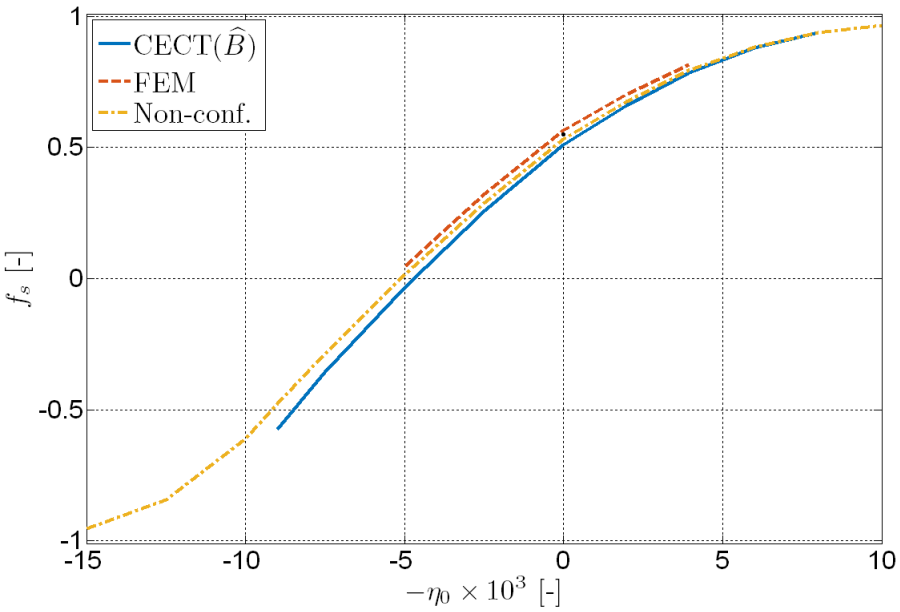


Figure 6.38. Lateral creepage–creep force curve for case "circ. a.". $\xi_0 = 0$.

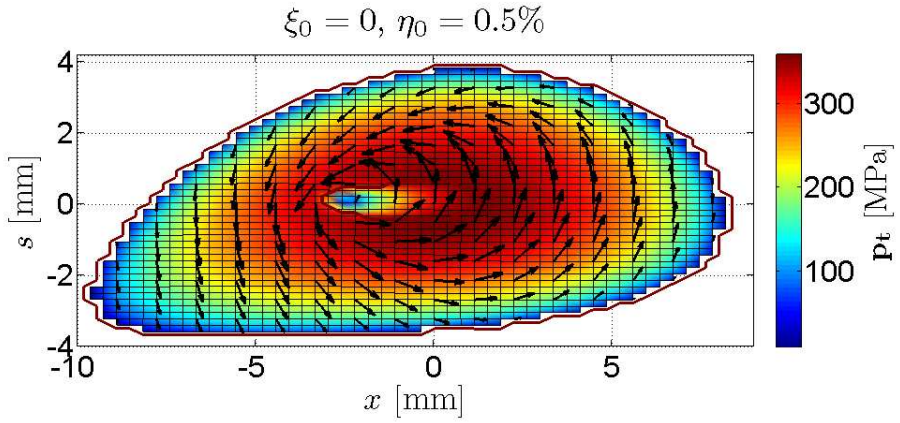
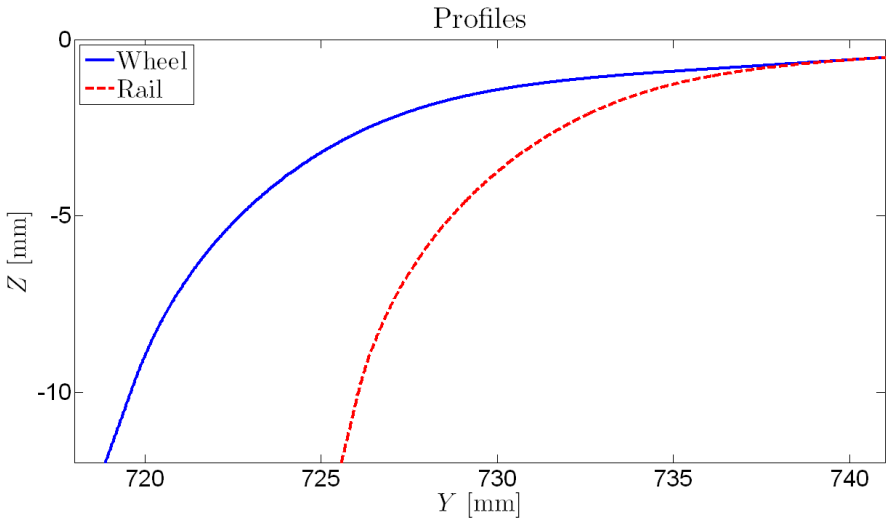


Figure 6.39. Tangential stress distribution in case "circ. a." with $\xi_0 = 0$ and $\eta_0 = -0.5\%$ obtained with FEM.

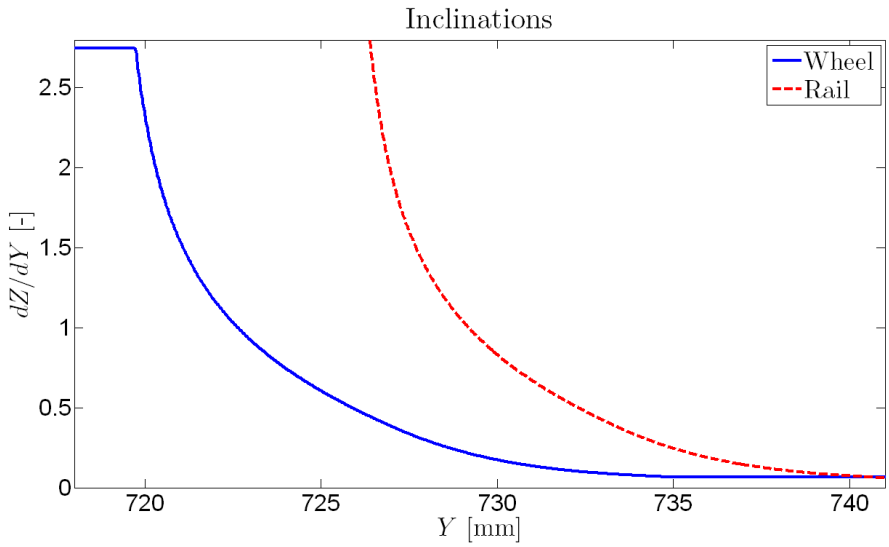
6.3 Conformal wheel–rail contact case studies

In this section, CECT is applied for the solution of two conformal contact cases in steady rolling, which may be representative in the wheel–rail application. The studied cases are designated as A and B. In case A the contact is produced closer to the wheel flange, with higher contact and yaw angles than in case B. In case B the contact is produced closer to the wheel tread, and the yaw angle is negative; i.e., the wheel is oriented inwards of the track in the running direction. Next, the relevant input data common for both cases are listed.

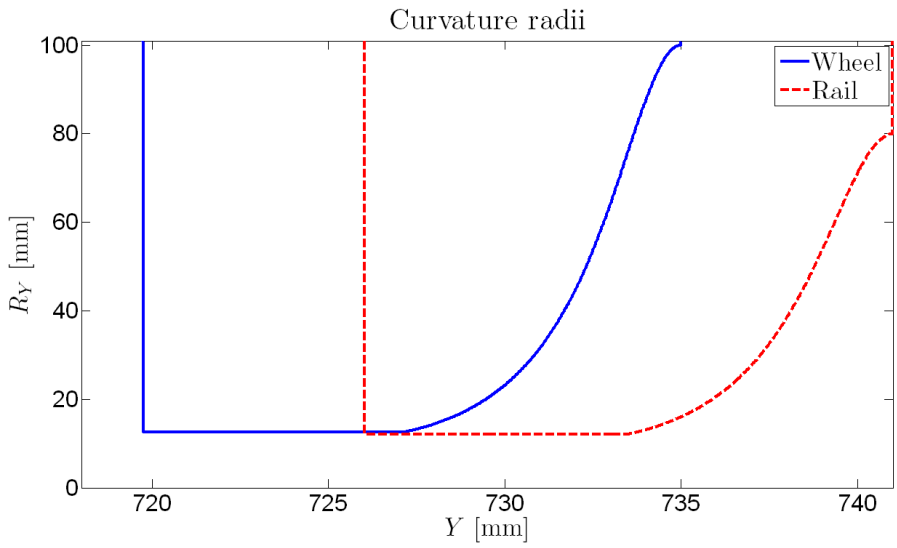
- Wheel and rail profiles: synthetic profiles are used, with lateral curvatures relatively close to each other in the transition zone between the nominal rolling points in the profiles, and the wheel flange root and railhead corner. Figure 6.40 shows the geometry of the used profiles as a function of the track lateral coordinate Y (cf. Figure 4.1b), being the wheelset centred on the track. The profiles of both rails of the track are positioned with an inclination of 1/40 towards the track centre. With these profiles, conformal contact situations are simulated which could take place with worn profiles.
- Wheelset and track half-widths: 750 and 753 mm, respectively. These half-widths are measured from the wheelset and track centre to the origin points of the local coordinate systems of the wheel and rail profiles, located in the wheel tread and in the railhead centre.
- Nominal wheel rolling radius R_{roll}: 460 mm.
- Material elastic properties (the same for both contacting solids): steel is considered, with E = 210 GPa and ν = 0.30.
- Coefficient of friction μ: variable in lateral (s) direction of the PCS. Figure 6.41 shows its variation on the profile of the PCS, which is linear with the profile length from a minimum value of 0.10 in the point with an inclination of 40º, to a maximum value of 0.50 in the point with an inclination of 10º. This profile of the coefficient of friction is intended to emulate the situation with different lubrication conditions that may take place in the transition between the wheel tread and flange, or between the corresponding contact zones of the rail.



(a) Profiles.



(b) Inclinations.



(c) Curvature radii.

Figure 6.40. Geometry of wheel and rail profiles used for the case studies of wheel–rail conformal contact.

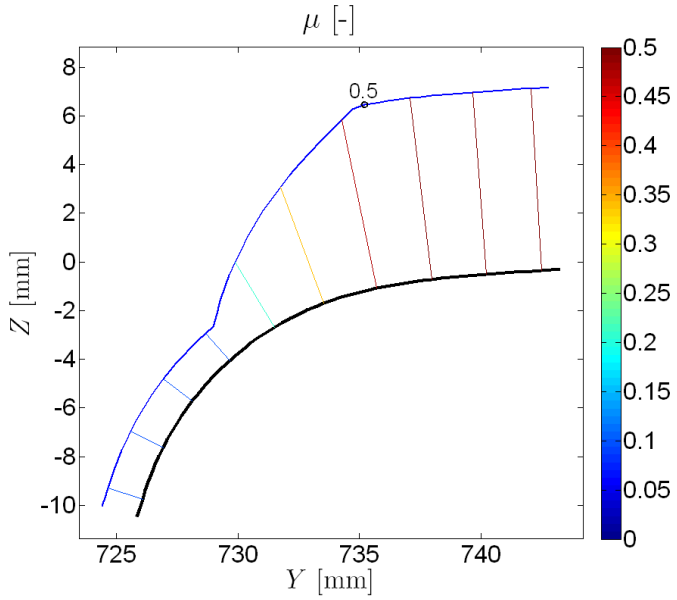


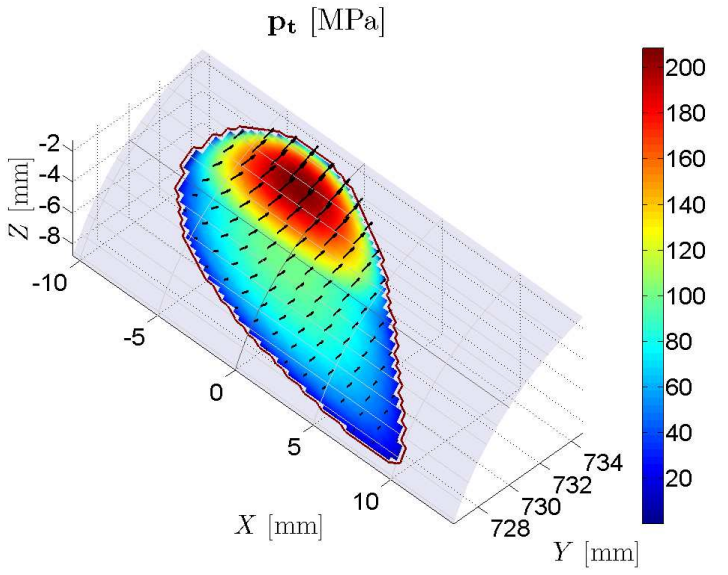
Figure 6.41. Considered spatial variation of the coefficient of friction on the profile of the potential contact surface.

The particular data for each case are listed in Table 6.2. The independent input data are the normal force F_N (more precisely, the projection of the resultant of normal pressures along the direction of the approach of the solids), the lateral displacement y and the yaw angle ψ of the wheelset with respect to the track, and the longitudinal creepage ξ_0 in the initial contact point. The latter is defined with the value $[(2\ r_{0,+}) / (r_{0,+} + r_{0,-}) - 1]$ in each case, being $r_{0,+}$ and $r_{0,-}$ the local rolling radii of the wheels in the initial contact points with each rail. The wheel of which the contact is studied is the one having a local rolling radius of $r_{0,+}$, located in the positive side of the lateral Y axis with respect to the track centre, in the same direction as the lateral displacement y applied to the wheelset. The dependent DOFs of the wheelset (the vertical displacement z and the roll angle ϕ) are obtained solving the geometric contact problem of the wheelset on the track, taking into account both wheel–rail contact pairs, as explained in §4.2.3.1. The contact angle in the initial contact point δ_0 is also obtained after solving the geometric wheelset–track contact problem, obtaining the positions of the initial contact points in each wheel–rail pair. The direction of the approach between the contacting solids is defined with this δ_0 angle.

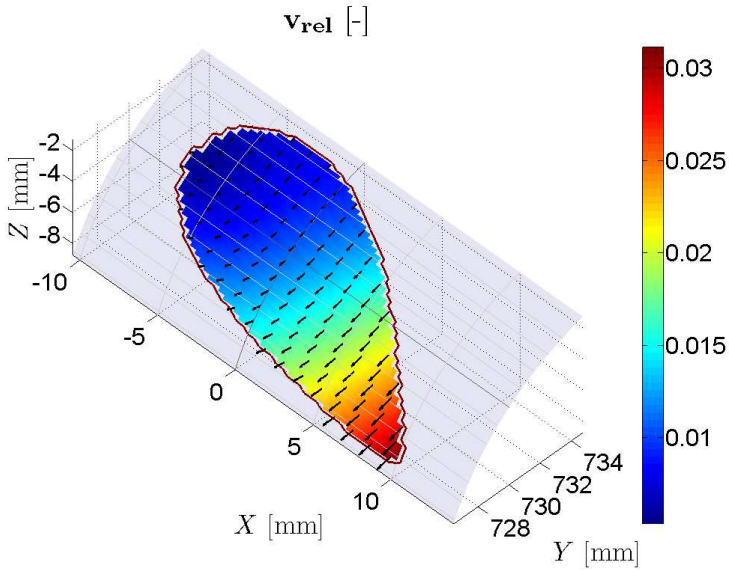
Table 6.2 Input data for wheel–rail conformal contact case studies.

Parameter	Case A	Case B
F_N [kN]	70	40
y [mm]	6.21	6.06
ψ [mrad]	8	–2
z [mm]	0.13	0.093
ϕ [mrad]	0.64	0.573
δ_0 [º]	40	15.6
ξ_0 [%]	0.454	0.184

Figure 6.42 shows the fields of tangential stresses and relative slip velocities obtained in case A with CECT, and Figure 6.43 the ones corresponding to case B, on the curved contact surfaces. The vectors shown are those corresponding to the wheel. I.e., the tangential stresses are those applied on the wheel, and the slip velocities the ones of the wheel with respect to the rail, as in other figures with tangential stresses and slip velocities shown in the thesis.

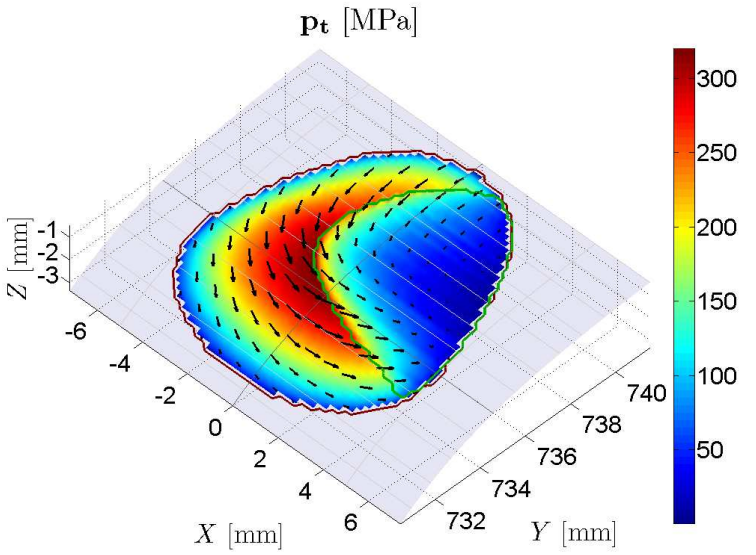


(a) Tangential stresses $\mathbf{p_t}$.

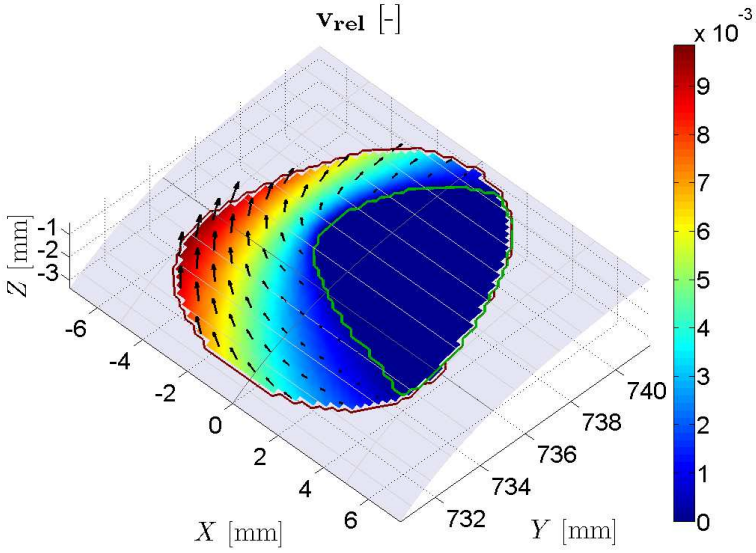


(b) Relative slip velocities $\mathbf{v_{rel}} = \mathbf{v} / V_c$.

Figure 6.42. Tangential problem solution of case A.



(a) Tangential stresses $\mathbf{p}_t$.



(b) Relative slip velocities $\mathbf{v}_{rel} = \mathbf{v} / V_c$.

Figure 6.43. Tangential problem solution of case B.

The contact patch of case A (Figure 6.42) presents the cam shape described in §6.2.1.2, substantially asymmetric in lateral as well as in longitudinal directions. The contact patch of case B (Figure 6.43) is also asymmetric in both directions, but much less than in case A, due to the lower magnitude of the contact and yaw angles. The lateral asymmetry of the contact patch of case B is largely due to the variation of the lateral curvatures of the wheel and rail profiles. On the contrary, in case A the profiles are circular in most of the contact patch.

In both cases there are considerable slip levels, and the magnitude of the tangential stresses is largely conditioned by the adhesion limit. In this way, there are higher tangential stress levels in case B (Figure 6.43a), in which the contact is produced in a zone with a higher coefficient of friction, despite the saturation level of the tangential stresses is lower than in case A (Figure 6.42a). The figures also show the adhesion and slip zones in the contact patch. In case B there is an adhesion zone in the leading part of the contact patch as may be seen in Figure 6.43, while in case A the whole contact patch is in slip. The slip velocities are much higher in case A, due to the higher spin and lateral creepage levels, consequence of the higher contact and yaw angles, respectively. The greatest slip levels are found in the leading part in case A (Figure 6.42b), and in the trailing part in case B (Figure 6.43b), due to the combination of the spin with the lateral creepage associated to the yaw angle, with different sign in each case, as it happens in non-conformal contact.

Table 6.3 lists some relevant results of the contact solution for each case. These include the following:

- The total contact angle variation in the contact patch $\Delta\delta$.
- The longitudinal $2a$ and lateral $2b$ dimensions of the contact patch, as well as its area A_c. $2a$ is calculated as the difference between the maximum and minimum x coordinates of the contact patch, and $2b$ is calculated similarly with the lateral coordinates s.
- The longitudinal and lateral components of the resultant force of the tangential stresses, F_x and F_{lat}, respectively.
- The direction of F_{lat} in the YZ plane, dir(F_{lat}) (cf. the `dir_Flat` output in §4.1.2.2).

- M_{sp}, the spin moment generated by the $\mathbf{p_t}$, calculated with respect to the position of the initial contact point in each case.

Table 6.3 Results for wheel–rail conformal contact cases.

	Case A	**Case B**
$\Delta\delta$ [º]	46.7	28.8
$2a$ [mm]	17.6	10.2
$2b$ [mm]	10.0	10.4
A_c [mm^2]	110	80.5
F_x [kN]	3.41	7.57
F_{lat} [kN]	8.99	4.41
dir(F_{lat}) [º]	37.4	–170
M_{sp} [N.m]	5.16	14.0

Case A has the highest $\Delta\delta$. The contact patch of case A is also the one with the highest longitudinal dimension. The higher normal load in this case contributes to this, and also the higher longitudinal curvature radii of the wheel resulting from the higher contact angles.

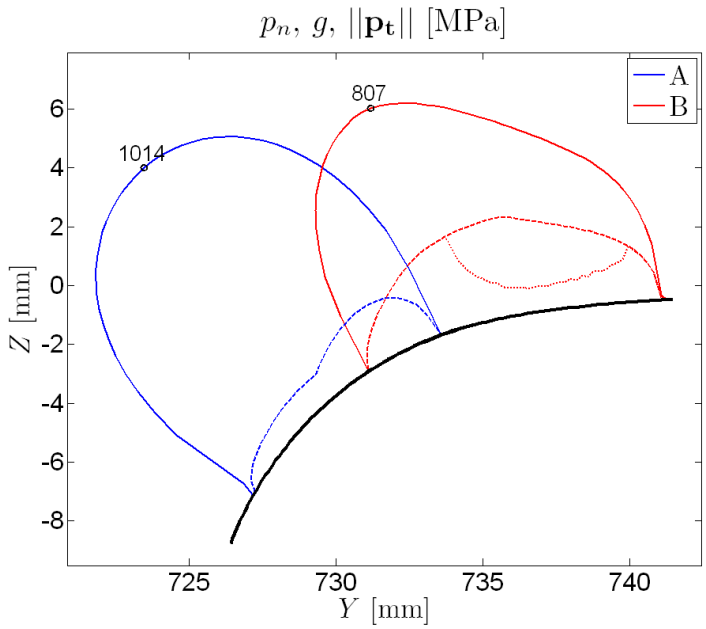
Regarding the resultant forces, case B has a higher F_x, despite the lower longitudinal rigid slip velocities w_x in this case (see as a reference the ξ_0 values of Table 6.2), due to the higher coefficients of friction and the lower saturation of F_{lat}. In case A, as the effects of the spin and of the lateral creepage resulting from the positive yaw angle are summed up, the magnitude of F_{lat} is higher. The direction of F_{lat} is towards the field side in case A, and towards the track centre in case B. Its direction does not coincide exactly with the one defined by the δ_0 of each case (cf. Table 6.2), due to the non-flatness of the contact patch and therefore not having the $\mathbf{p_t}$ throughout the contact patch acting in the same plane. Similarly, in the normal part, in general the direction of the resultant of the normal pressures does not coincide with the direction of the normal approach between the contacting solids.

The fact that the resultant forces of the normal contact stresses on the one hand and of the tangential contact stresses on the other are no longer perpendicular to each other, no longer existing unique “normal” and “tangential” contact directions associated to each one, leads to a coupling of the normal and tangential problems at equilibrium of forces level, besides the coupling at contact mechanics level described in §6.2.2. Due as well to the non-flatness of the contact surface, the M_{sp} is not oriented in the normal direction to the mean

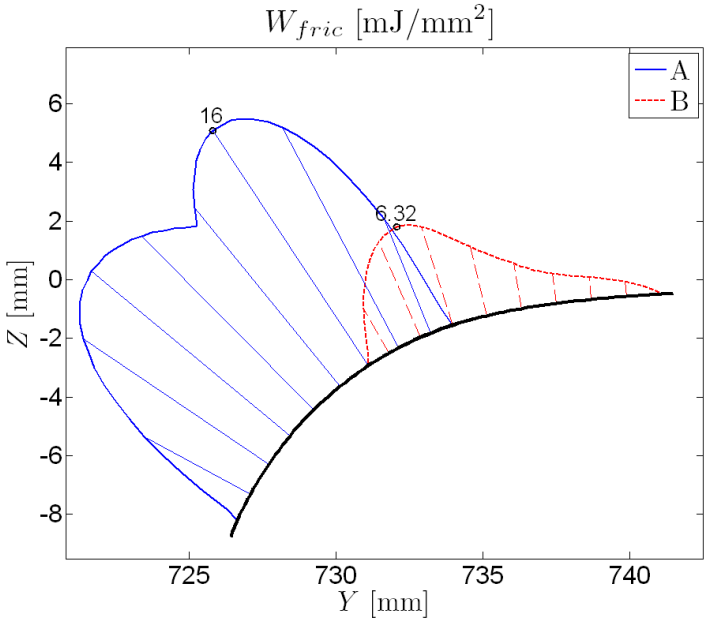
contact plane. The deviation is important especially in case A, with a contact patch with higher $\Delta\delta$ and also more asymmetric. In this case, the direction of M_{sp} is deviated in about 60º with respect to the normal to the mean contact plane, and there is a considerable component of M_{sp} in longitudinal direction, of approximately 40% of its magnitude.

Figure 6.44 shows some representative contact magnitudes on the same profile of the contact surface together for cases A and B. Figure 6.44a shows the normal pressures p_n in solid line, the adhesion limits g in dashed line, and the magnitude of the tangential stresses $\|\mathbf{p}_t\|$ in dotted line, in the cross-section passing through the initial contact point in each case. The position of the initial contact point is marked in Figure 6.42 and in Figure 6.43 with thicker lines along the two principal directions of the contact surface. Figure 6.44b shows the profiles of integrated frictional work after a wheel passage, W_{fric}.

The transition zone of μ may be clearly appreciated in Figure 6.44a, within the points with discontinuous slope in the curves of g. The beginning of the increase of μ from its minimum value of 0.1 in the most vertical zone of the profile marks as well a clear tendency change in the curve of W_{fric} of case A (Figure 6.44b). From the most vertical zone of the surface, after reaching W_{fric} its first maximum in case A, initially it begins to decrease towards the most horizontal zone, due to the reduction of the slip velocities. Upon entering in the zone of increasing μ, the magnitude of the tangential stresses increases, and the trend of W_{fric} changes, increasing again. This is illustrative of the decisive influence of lubrication to reduce wear, which is closely related to frictional work, in zones with high slip. In case B, the values of W_{fric} are much lower than in case A despite the higher coefficient of friction, due to the lower slip velocities in this case (compare Figure 6.42b and Figure 6.43b).



(a) p_n, g and $\|\mathbf{p}_t\|$ at $x = 0$. p_n in solid line, g in dashed line and $\|\mathbf{p}_t\|$ in dotted line.



(b) Frictional work after a wheel passage, W_{fric}.

Figure 6.44. Results of cases A and B on the profile of the potential contact surface.

Chapter 7

Contact analysis in rolling bearings

Chapter summary

The aim of this chapter is to provide tools for detailed rolling element–raceway frictional contact mechanics analyses in rolling bearings. For this purpose, the numerical contact models presented in Chapter 4 (i.e., CECT and FE contact models) are used in the rolling bearing application, and a tool based on strip theory is developed. The extensions needed in Kalker's exact contact theory, originally developed for the wheel–rail case, for its application in rolling bearings, are explained in this chapter. The use of partial slip contact models in this application contrasts with the usual simplification of using the full slip assumption in the tangential part of the contact problem.

The different contact models are applied in a case study of a spherical roller bearing. The adequacy of the more simplified contact solutions is assessed by contrasting them with the solutions obtained with the more comprehensive models. Situations with different normal loads and friction levels are analyzed, and two distinct equilibrium configurations of the roller are identified.

7.1 Rolling bearing geometry and kinematics

Some geometric and kinematic relationships in rolling bearings are given in this section, which are necessary for the correct definition of the rolling element–raceway contact problems.

Figure 7.1 shows a sectional view of a rolling bearing (in particular, a spherical roller thrust bearing), including a roller and the inner and outer rings. Each body is identified with a letter. *I* is used for the inner ring, *R* for the rolling element (sometimes referred to as the roller here) and *E* for the outer or exterior ring. The following relevant internal dimensions are shown in the figure:

- r_I, r_E: groove curvature radii of the inner and outer races. Circular profiles are considered.
- R: roller crown radius. The roller cross-sectional profile is considered circular, as the races.
- D: roller nominal diameter. It is the diameter of the nominal rolling circle of the roller, marked in dashed red line in the figure.
- α_o: nominal angle between the roller axis and the bearing axis.
- α_R: roller angle. It is the half-angle of the nominal rolling cone of the roller.
- R_p: pitch radius. It is the radius of the circumference described by the centre of the nominal rolling circle of the roller around the bearing axis. In spherical roller bearings and spherical roller thrust bearings, R_p is related with r_E, D, α_o and α_R through Eq. (7.1):

$$R_p = r_E \cos(\alpha_o + \alpha_R) - D / 2 \cos(\alpha_o) \tag{7.1}$$

  In roller bearings with non-zero nominal α_R, a different pitch radius may be defined, considering the centre of the rolling circle of the roller with zero α_R, in which the tangent plane to the rolling element surface is parallel to its axis (for rolling elements with non-zero nominal α_R, this axial position will normally fall beyond the physical limits of the rolling element). The pitch radius corresponding to this position is designated here as $R_{p,o}$, and is given by Eq. (7.2).

$$R_{p,o} = R_p + R \sin(\alpha_R) \sin(\alpha_o) \tag{7.2}$$

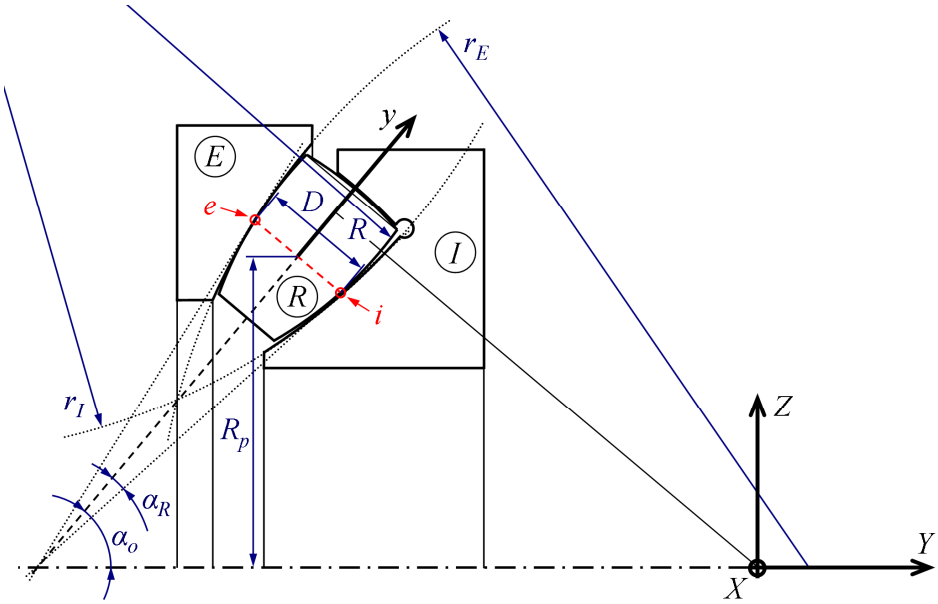


Figure 7.1. Sectional view of a rolling bearing, with some relevant dimensions.

The following coordinate systems are defined:

- Bearing $\{X, Y, Z\}$: Cartesian coordinate system that rotates with orbital rotation velocity ω_m following the radial bearing plane containing the central point of the considered roller. Its origin is on the common rotation axis of the inner and outer rings, in the intersection point between this axis and the perpendicular to the y axis of the rolling element from the position corresponding to $R_{p,o}$. The YZ plane contains the rotation axis of the rings and the roller centre. The Y axis is aligned with the rotation axis of the rings, the Z axis points vertically upwards, and the X axis completes the right-handed coordinate system.
- Global or bearing reference $\{X_o, Y_o, Z_o\}$: fixed Cartesian coordinate system, with the same origin as the $\{X, Y, Z\}$ system, and which has the orientation of this system at a given reference instant. This system is used in the FE models described in §7.2.3.

- Roller reference $\{x, y, z\}$: Cartesian coordinate system with origin in the nominal location of the nominal rolling circle of the roller in the *YZ* plane, with the *y* axis aligned with the nominal roller axis of revolution. It is obtained from the bearing $\{X, Y, Z\}$ system with a rotation α_o around the *X* axis and displacements along the *Y* and *Z* axes, being the displacement along the *Z* axis equal to R_p.
- Roller $\{u, v, w\}$: Cartesian coordinate system with origin in the centre of the nominal rolling circle of the roller, which moves solidary to the roller. It is obtained from the $\{x, y, z\}$ system with the following sequence of rotations and displacements:
    - β tilt rotation around the *x* axis. After this rotation, a z' axis is obtained from the *z* axis.
    - γ yaw or skew rotation around the z' axis. After this rotation, the orientation of the *v* axis is defined, which coincides with the roller axis.
    - θ_R pitch rotation around the *v* axis.
    - d_y and d_z displacements along the *y* and *z* axes.

The tilt and skew rotations of the roller are illustrated in Figure 7.2. According to these rotations, and the nominal orientation of the roller axis in the *YZ* plane defined with the α_o angle, the cosine directors (l_x, l_y, l_z) of the roller axis *v* in the $\{X, Y, Z\}$ coordinate system are $l_x = -\sin(\gamma)$, $l_y = \cos(\gamma)\cos(\alpha_o + \beta)$, and $l_z = \cos(\gamma)\sin(\alpha_o + \beta)$. β and γ are related to the roll ϕ and yaw ψ rotations introduced in Figure 4.1b through Eqs. (7.3) and (7.4):

$$\psi = \operatorname{atan}\left(\frac{-l_x}{l_y}\right) \tag{7.3}$$

$$\phi = \operatorname{sgn}(l_z)\left(\pi\frac{1-\operatorname{sgn}(l_y)}{2} + \operatorname{sgn}(l_y)\,\operatorname{asin}|l_z|\right) \tag{7.4}$$

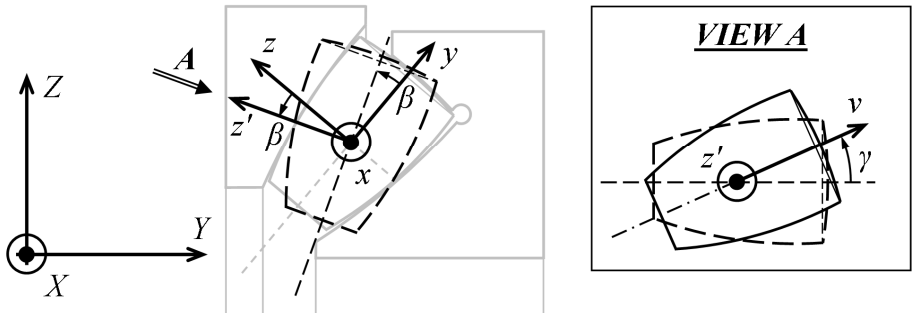


Figure 7.2. Tilt and skew rotations of the rolling element.

The nominal contact points of the roller with the inner and outer races are identified as "*i*" and "*e*" in Figure 7.1. The nominal rolling radii of the rings, $r_{\text{roll,I,i}}$ and $r_{\text{roll,E,e}}$, are given by Eq. (7.5). These are distances in the *Z* axis from the bearing axis to the inner and outer nominal contact points. The longitudinal curvature radius of each ring at the nominal contact point is equal to the rolling radius divided by cos(*α*), with *α*, the contact angle, being nominally equal to $\alpha_o - \alpha_R$ for the inner contact and $\alpha_o + \alpha_R$ for the outer contact.

$$r_{\text{roll,I,i}} = R_p - \frac{D}{2}\cos(\alpha_o);\ r_{\text{roll,E,e}} = R_p + \frac{D}{2}\cos(\alpha_o) \tag{7.5}$$

In steady rolling, the motion of the roller is composed of an orbital rotation around the bearing *Y* axis with angular velocity ω_m and a pitch rotation around its own *v* axis with angular velocity ω_R. The theoretical angular velocities for perfect rolling on the nominal contact points are given by Eqs. (7.6) and (7.7). In these equations, ω_I and ω_E are the rotation velocities of the inner and outer rings (around the bearing *Y* axis). These equations may be worked out by simple kinematic analysis, setting out the velocities of the inner and outer nominal contact points as belonging to the corresponding ring and to the roller.

$$\omega_{m,o} = \frac{r_{\text{roll,I,i}}\ \omega_I + r_{\text{roll,E,e}}\ \omega_E}{r_{\text{roll,I,i}} + r_{\text{roll,E,e}}} \tag{7.6}$$

$$\omega_{R,o} = \frac{-2 r_{\text{roll,I,i}}\ r_{\text{roll,E,e}}\left(\omega_I - \omega_E\right)}{D\left(r_{\text{roll,I,i}} + r_{\text{roll,E,e}}\right)} \tag{7.7}$$

The components in the $\{X, Y, Z\}$ coordinate system of the total rotation velocity vector of the rolling element with respect to each ring, $\boldsymbol{\omega}_{\mathbf{RT/j}}$, $j \in \{I, E\}$, are given in Eqs. (7.8)–(7.10), with $\omega_{m/j} = \omega_m - \omega_j$.

$$\omega_{RT/j,X} = \omega_R \, l_x \tag{7.8}$$

$$\omega_{RT/j,Y} = \omega_R \, l_y + \left(\omega_m - \omega_j\right) \tag{7.9}$$

$$\omega_{RT/j,Z} = \omega_R \, l_z \tag{7.10}$$

7.1.1 Kinematics of finite rotations

The formulation treated in this section is useful to work with the FE models described in §7.2.3. Unlike in the wheel–rail contact FE models introduced in §4.2, the orientation of the axis of revolution of the rolling element is variable during the rolling phase of the simulation in these rolling bearing FE models. This makes specifying and interpreting input and output orientation angles of the rolling element not straightforward. It has to be considered that compound finite rotations are not additive, as shown in the example of §1.2.2 Conventions – Finite rotations – Compound rotations in the Abaqus Analysis User's Manual [Abaqus doc 2012].

The following known definitions and formulas are introduced:

- Dyadic product: given two 3-element vectors $\mathbf{u} = \{u_1, u_2, u_3\}^{\mathrm{T}}$ and $\mathbf{v} = \{v_1, v_2, v_3\}^{\mathrm{T}}$, their dyad or dyadic product $\mathbf{uv}$ is a second-order tensor equal to $\mathbf{u} \otimes \mathbf{v}^{\mathrm{T}}$, where $\otimes$ stands for the Kronecker product (also known as the tensor product or the direct product). The result is expanded in Eq. (7.11).

$$\mathbf{uv} = \mathbf{u} \otimes \mathbf{v}^{\mathrm{T}} = \begin{bmatrix} u_1\mathbf{v}^{\mathrm{T}} \\ u_2\mathbf{v}^{\mathrm{T}} \\ u_3\mathbf{v}^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} u_1v_1 & u_1v_2 & u_1v_3 \\ u_2v_1 & u_2v_2 & u_2v_3 \\ u_3v_1 & u_3v_2 & u_3v_3 \end{bmatrix} \tag{7.11}$$

- Axial rotation vector: a finite rotation may be defined by means of its axial vector $\theta\mathbf{a}$. $\mathbf{a}$ is a 3-element unit vector aligned with the axis of

rotation, and the angle of rotation is equal to θ. There is a skew-symmetric matrix **[ã]** associated with vector **a**, which is given by Eq. (7.12) for $\mathbf{a} = \{a_1, a_2, a_3\}^T$:

$$[\tilde{\mathbf{a}}] = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \tag{7.12}$$

- Rotation matrix. A rotation may be defined with a rotation matrix as well. The following two notations are used here to define a rotation matrix:
  - **[$\mathbf{R_a}(\gamma)$]** stands for the rotation matrix corresponding to a rotation of angle γ around axis **a**. It is given in the same coordinate system as **a**.
  - **[$\mathbf{R}_{B/A}$]** stands for the rotation matrix to get a system of axes B from a system of axes A, given in the system of axes A.

  In what follows a rotation matrix designated as **[A]** is considered instead for notational simplicity. This is a 3×3 orthonormal matrix, which, when multiplied by a vector **v**, returns the rotated vector **v'**. I.e., **v'** = **[A] v**. If an orthonormal frame with unit vectors {**i**, **j**, **k**} is attached to **v**, which after rotation, becomes {**i'**, **j'**, **k'**}, **[A]** may be calculated as the sum of the dyads of the rotated unit vectors with the corresponding unit vectors before rotation, as indicated in Eq. (7.13). In this equation, $i_l, j_l, \ldots, l \in \{1, 2, 3\}$ are the components of the different unit vectors, given in the same coordinate system for the six vectors involved in the equation. The resulting matrix **[A]** is in this same coordinate system. It is easy to prove this equation by scalar multiplication with each of the unit vectors {**i**, **j**, **k**}. As the scalar multiplication of each unit vector with itself is 1, and with the other two unit vectors is 0, it is straightforward to see for example that **[A] i** = **i'**.

$$[\mathbf{A}] = \mathbf{i}'\mathbf{i} + \mathbf{j}'\mathbf{j} + \mathbf{k}'\mathbf{k} =$$

$$\begin{bmatrix} i'_1 i_1 + j'_1 j_1 + k'_1 k_1 & i'_1 i_2 + j'_1 j_2 + k'_1 k_2 & i'_1 i_3 + j'_1 j_3 + k'_1 k_3 \\ i'_2 i_1 + j'_2 j_1 + k'_2 k_1 & i'_2 i_2 + j'_2 j_2 + k'_2 k_2 & i'_2 i_3 + j'_2 j_3 + k'_2 k_3 \\ i'_3 i_1 + j'_3 j_1 + k'_3 k_1 & i'_3 i_2 + j'_3 j_2 + k'_3 k_2 & i'_3 i_3 + j'_3 j_3 + k'_3 k_3 \end{bmatrix} \quad (7.13)$$

As it is known, the columns of **[A]** contain the cosine directors of the rotated unit vectors {**i'**, **j'**, **k'**} on the non-rotated basis {**i**, **j**, **k**}, if {**i**, **j**, **k**} is the basis in which **[A]** is given, i.e., if **[i j k]** = **[I]**. So, if **v** is given in the basis {**i'**, **j'**, **k'**}, the multiplication **[A] v** may be also interpreted as the transformation of **v** from the basis {**i'**, **j'**, **k'**} to the basis {**i**, **j**, **k**}. The inverse rotation (or transformation) is achieved with the transposed matrix $[\mathbf{A}]^{\mathrm{T}}$, as $[\mathbf{A}]^{-1} = [\mathbf{A}]^{\mathrm{T}}$.

[A] may be also obtained from the axial vector $\theta\mathbf{a}$ of the rotation through Eq. (7.14). This is known as the Rodriguez formula. Its proof may be found in §2.1 of [Shabana 2005]. Another form of the Rodriguez formula is given in Eq. (7.15), wherein **[A]** is expressed as the exponential of θ**[ã]**. Details on the equivalence of both equations may be found in §2.3 [Shabana 2005].

$$[\mathbf{A}] = [\mathbf{I}] + \sin\theta\,[\tilde{\mathbf{a}}] + 2\sin^2\frac{\theta}{2}([\tilde{\mathbf{a}}])^2 \quad (7.14)$$

$$[\mathbf{A}] = [\mathbf{I}] + \sin\theta\,[\tilde{\mathbf{a}}] + (1-\cos\theta)([\tilde{\mathbf{a}}])^2 = \exp(\theta[\tilde{\mathbf{a}}]) \quad (7.15)$$

Yet another form of the Rodriguez formula is given in Eq. (7.16), cf. §1.3.1 Rotation variables in the Abaqus Theory Manual [Abaqus doc 2012]. This equation is given in component form in Eq. (7.17), where the summation convention is used. In this equation, δ_{ij} is the Kronecker delta, and ε_{ijk} is the alternator tensor, defined by $\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = 1$, $\varepsilon_{132} = \varepsilon_{213} = \varepsilon_{321} = -1$, and all other $\varepsilon_{ijk} = 0$.

$$[\mathbf{A}] = \cos\theta\,[\mathbf{I}] + \sin\theta\,[\tilde{\mathbf{a}}] + (1-\cos\theta)\,\mathbf{a}\otimes\mathbf{a} \quad (7.16)$$

$$A_{ij} = \cos\theta\ \delta_{ij} + \sin\theta\ \varepsilon_{ijk} a_k + (1-\cos\theta)\ a_i a_j \tag{7.17}$$

Lastly, for a given **[A]** rotation matrix, its associated axial vector **a** may be obtained through Eqs. (7.18) and (7.19). Tr(**[A]**) in Eq. (7.18) stands for the trace of **[A]**.

$$\mathrm{Tr}([\mathbf{A}]) = 1 + 2\cos\theta \tag{7.18}$$

$$2\sin\theta\ [\tilde{\mathbf{a}}] = [\mathbf{A}] - [\mathbf{A}]^{\mathrm{T}} \tag{7.19}$$

7.1.1.1 Composition of a sequence of rotations into a single rotation

The composition of a sequence of finite rotations into a single rotation is considered next. This is useful to define the initial positioning of the rolling element in the FE models described in §7.2.3, for instance.

The global fixed {X_o, Y_o, Z_o} system and the {u, v, w} system attached to the rolling element are considered. The unit vectors corresponding to the X_o, Y_o, Z_o axes are designated as **i**, **j**, **k**, respectively, and those corresponding to the u, v, w axes as **u**, **v**, **w**. Different punctuations are used to denote different positions of **u**, **v** and **w**, as explained below.

The rolling element is initially positioned with its centre located in the global Y_o-Z_o plane, at $Z_o = R_p$, with its v axis inclined at an angle α_o with respect to the Y_o axis, and with zero skew and pitch rotations. In this initial orientation, the unit vectors of the rolling element axes are $\mathbf{u_o}$, $\mathbf{v_o}$ and $\mathbf{w_o}$. The rotation matrix which gives the components of $\mathbf{u_o}$, $\mathbf{v_o}$ and $\mathbf{w_o}$ in the {**i**, **j**, **k**} basis is $[\mathbf{R}_\mathbf{X}(\alpha_o)]$, see Eq. (7.20).

$$[\mathbf{R}_\mathbf{X}(\alpha_o)] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{\alpha_o} & -s_{\alpha_o} \\ 0 & s_{\alpha_o} & c_{\alpha_o} \end{bmatrix} \tag{7.20}$$

The following sequence of rotations and displacements is considered, illustrated in Figure 7.3:

- Tilt rotation β. After this rotation, the unit vectors of the rolling element axes are **u'**, **v'** and **w'**. The rotation matrix which gives the components of **u'**, **v'** and **w'** in the {**i**, **j**, **k**} basis is $\mathbf{[R_X}(\alpha_o+\beta)\mathbf{]}$. The angle $\alpha_o+\beta$ will be designated as α in what follows. Figure 7.3 shows in black dashed line the circular trajectory in which the tips of the unit vectors of axes $\mathbf{v_o}$ and $\mathbf{w_o}$ is enclosed during this rotation.

- Orbital rotation θ_m and corresponding displacement in the pitch circle. After this rotation, the unit vectors of the rolling element axes are **u''**, **v''** and **w''**. The rotation matrix to get **u''**, **v''** and **w''** from **u'**, **v'** and **w'** is $\mathbf{[R_Y}(\theta_m)\mathbf{]}$. Therefore, the rotation matrix which gives the components of **u''**, **v''** and **w''** in the {**i**, **j**, **k**} basis is $\mathbf{[R_Y}(\theta_m)\mathbf{][R_X}(\alpha)\mathbf{]}$. The components of this rotation matrix are given in Eq. (7.21), where s_δ and c_δ stand for sin(δ) and cos(δ), respectively (and s_m and c_m stand for sin(θ_m) and cos(θ_m), respectively). For example, the first column of this matrix, $\{c_m, 0, -s_m\}^T$, gives the components of **u''** in the {**i**, **j**, **k**} basis, i.e., $\mathbf{u''} = c_m\,\mathbf{i} - s_m\,\mathbf{k}$. To further illustrate this, the relationship between the {**i**, **j**, **k**} and the {**u''**, **v''**, **w''**} bases is given in Eq. (7.22).

  Note that Eq. (7.22) collects 3 linear combinations of vectors **u''**, **v''** and **w''** (one with each row of the rotation matrix), the first to obtain **i**, the second **j**, and the third **k**. Eqs. (4.18), (4.20), (7.24) and (7.25) are to be interpreted analogously. Eq. (7.22) is equivalent to performing the matrix-vector multiplication of the rotation matrix with each of the **i**, **j**, and **k** vectors, to get the **u''**, **v''** and **w''** vectors, respectively, being all the vectors given by their components in the {**i**, **j**, **k**} basis, which is the basis of this rotation matrix. As an example, Eq. (7.23) gives the multiplication to rotate **i** into **u''**. This is a multiplication of the form **v'** = **[A] v** indicated before Eq. (7.13).

$$\left[\mathbf{R}_\mathbf{Y}(\theta_m)\right]\left[\mathbf{R}_\mathbf{X}(\alpha)\right]=\begin{bmatrix} c_m & s_m s_\alpha & s_m c_\alpha \\ 0 & c_\alpha & -s_\alpha \\ -s_m & c_m s_\alpha & c_m c_\alpha \end{bmatrix} \tag{7.21}$$

$$\begin{Bmatrix}\mathbf{i}\\ \mathbf{j}\\ \mathbf{k}\end{Bmatrix} = [\mathbf{R}_{\mathbf{Y}}(\theta_m)]\,[\mathbf{R}_{\mathbf{X}}(\alpha)] \begin{Bmatrix}\mathbf{u''}\\ \mathbf{v''}\\ \mathbf{w''}\end{Bmatrix} \tag{7.22}$$

$[\mathbf{R}_{\mathbf{Y}}(\theta_m)]\,[\mathbf{R}_{\mathbf{X}}(\alpha)]\,\mathbf{i}$

$$= [\mathbf{R}_{\mathbf{Y}}(\theta_m)]\,[\mathbf{R}_{\mathbf{X}}(\alpha)] \begin{Bmatrix}1\\0\\0\end{Bmatrix} = \mathbf{u''} = \begin{Bmatrix} c_m\\ 0\\ -s_m\end{Bmatrix} \tag{7.23}$$

In addition to the described orientation change, after the orbital rotation the centre of the rolling element has displacements along the X_o and Z_o axes of $R_p\ s_m$ and $R_p\ (c_m - 1)$, respectively. Figure 7.3 shows in green dashed line the circular trajectory in which the centre of the rolling element moves during this displacement, parallel to the XZ plane.

- Pitch rotation θ_R. After this rotation, the rolling element axes attain their final orientation, with unit vectors **u**, **v** and **w**. These vectors are coloured in red in Figure 7.3, and the circular trajectory which they follow during this pitch rotation, in a plane perpendicular to **v**, is shown in orange dotted line in the figure. The rotation matrix to get **u**, **v** and **w** from **u''**, **v''** and **w''** is $\mathbf{[R_v(\theta_R)]} = [c_R\ 0\ s_R;\ 0\ 1\ 0;\ -s_R\ 0\ c_R]$ (the semicolons delimit the different rows of the matrix, and s_R and c_R stand for $\sin(\theta_R)$ and $\cos(\theta_R)$, respectively). The relationship between the {**u**, **v**, **w**} and the {**u''**, **v''**, **w''**} bases is given in Eq. (7.24). For example, the 3rd column of $\mathbf{[R_v(\theta_R)]}$, $\{s_R, 0, c_R\}^T$, gives the components of **w** in the {**u''**, **v''**, **w''**} basis, i.e., $\mathbf{w} = s_R\,\mathbf{u''} + c_R\,\mathbf{w''}$. The rotation matrix $\mathbf{[R_{R/G}]}$ to get the unit vectors of the roller (R) axes **u**, **v** and **w** from the unit vectors of the global (G) axes **i**, **j** and **k** is calculated replacing Eq. (7.24) in the right-hand side of Eq. (7.22). It is given in Eq. (7.25) in compact form (in terms of the multiplication of the different rotation matrices involved), and Eq. (7.26) gives the resulting expressions for each component of the matrix.

$$\begin{Bmatrix} \mathbf{u''} \\ \mathbf{v''} \\ \mathbf{w''} \end{Bmatrix} = [\mathbf{R}_{\mathbf{v}}(\theta_R)] \begin{Bmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{Bmatrix} \tag{7.24}$$

$$\begin{Bmatrix} \mathbf{i} \\ \mathbf{j} \\ \mathbf{k} \end{Bmatrix} = [\mathbf{R}_{\mathbf{Y}}(\theta_m)]\,[\mathbf{R}_{\mathbf{X}}(\alpha)]\,[\mathbf{R}_{\mathbf{v}}(\theta_R)] \begin{Bmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{Bmatrix};$$

$$[\mathbf{R}_{\mathrm{R/G}}] = [\mathbf{R}_{\mathbf{Y}}(\theta_m)]\,[\mathbf{R}_{\mathbf{X}}(\alpha)]\,[\mathbf{R}_{\mathbf{v}}(\theta_R)] \tag{7.25}$$

$$[\mathbf{R}_{\mathrm{R/G}}] = \begin{bmatrix} c_m c_R - s_m c_\alpha s_R & s_m s_\alpha & c_m s_R + s_m c_\alpha c_R \\ s_\alpha s_R & c_\alpha & -s_\alpha c_R \\ -s_m c_R - c_m c_\alpha s_R & c_m s_\alpha & -s_m s_R + c_m c_\alpha c_R \end{bmatrix} \tag{7.26}$$

The sought rotation matrix **[A]** between the initial and final roller orientations, i.e., from {$\mathbf{u_o}$, $\mathbf{v_o}$, $\mathbf{w_o}$} to {**u**, **v**, **w**}, is obtained applying Eq. (7.13), where the rotated vectors are **u**, **v** and **w** and the initial ones $\mathbf{u_o}$, $\mathbf{v_o}$ and $\mathbf{w_o}$, respectively. The components in the global {X_o, Y_o, Z_o} coordinate system of vectors **u**, **v** and **w** on the one hand, and $\mathbf{u_o}$, $\mathbf{v_o}$ and $\mathbf{w_o}$ on the other, are given in the columns of matrices $\mathbf{[R_{R/G}]}$ (Eq. (7.26)) and $\mathbf{[R_X(\alpha_o)]}$ (Eq. (7.20)), respectively. The axial rotation vector $\theta\mathbf{a}$ of **[A]** may then be obtained with Eqs. (7.18) and (7.19).

Given inner and outer ring rotations θ_I and θ_E, adequate values for the orbital θ_m and pitch θ_R rotations of the rolling element may be obtained through Eqs. (7.6) and (7.7), replacing in these equations the rotation velocities ω_I and ω_E with the rotation angles θ_I and θ_E. The rotations obtained in this way would correspond to the theoretical motion for perfect rolling. If necessary, additional offsets could be added to the theoretical θ_m and θ_R angles for perfect rolling to adjust the position of the rolling element as necessary.

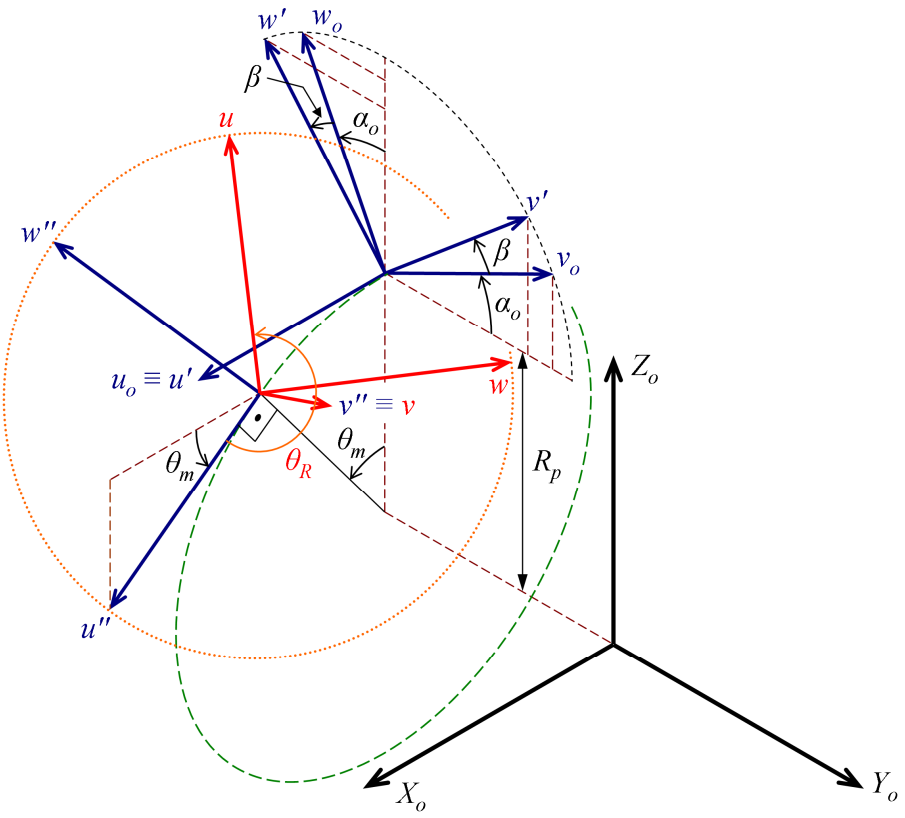


Figure 7.3. Sequence of positioning rotations and displacements of the rolling element. Isometric view.

7.1.1.2 Decomposition of a rotation into rolling element reference rotations

Now, a single rotation is to be decomposed into several reference rotations of the rolling element. This may be regarded as the inverse of the composition of a sequence of rotations considered in §7.1.1.1. It may be useful for correctly interpreting the output rotation quantities of the rolling element in the FEM simulations described in §7.2.3 for instance, decomposing the total rotational motion into the different rotation components relevant for the analysis.

The following inputs are considered, given in the $\{X_o, Y_o, Z_o\}$ coordinate system:

- The initial orientation of the rolling element. Apart from the α_o angle, initial non-zero tilt and skew angles are considered, β_o and γ_o, generalizing the development of §7.1.1.1.
- The total rotation of the rolling element from its initial configuration, defined by its axial rotation vector $\theta\mathbf{a}$. With this, the rotation matrix **[A]** from the initial to the final orientation of the rolling element is calculated applying Eq. (7.17).
- The position of the rolling element centre $\mathbf{p} = \{p_1, p_2, p_3\}^T$.

The following angles or rotations are to be calculated, which determine the orientation and rotation velocity of the rolling element: θ_m, θ_R, β and γ. β and γ are angles with respect to the theoretical rolling element orientation at each instant, and θ_m and θ_R are rotations from the initial position. The β and γ rotations of the rolling element are illustrated in Figure 7.2, and the θ_m and θ_R rotations in Figure 7.3. The β rotation is also illustrated in Figure 7.3.

Apart from the orientation and rotation velocity of the rolling element, its forward velocity and position in the bearing *YZ* radial plane are also of interest. The determination of these from the **p** position vector is straightforward.

To calculate the orientation and rotation velocity of the rolling element, first the following matrices are calculated:

- $\mathbf{[R_X}(\alpha_o)\mathbf{]}$ (Eq. (7.20)).
- $\mathbf{[R}_{R/G,o}\mathbf{]}$, the rotation matrix to get the vectors $\mathbf{u_o}$, $\mathbf{v_o}$ and $\mathbf{w_o}$ in the initial position of the rolling element, from the global vectors **i**, **j** and **k**. In the initial position, $\theta_m = 0$ and $\theta_R = 0$. On the other hand, non-zero initial β_o and γ_o angles are considered, unlike in §7.1.1.1. Therefore, $\mathbf{[R}_{R/G,o}\mathbf{]}$ may not be calculated using Eqs. (7.25) and (7.26) with $\theta_m = 0$ and $\theta_R = 0$, i.e. simply as $\mathbf{[R_X}(\alpha_o)\mathbf{]}$. Instead, the $\mathbf{[R}_{R/Ref,o}\mathbf{]}$ matrix is calculated first, which gives vectors $\mathbf{u_o}$, $\mathbf{v_o}$ and $\mathbf{w_o}$ in the $\{x_o, y_o, z_o\}$ coordinate system, which is the roller reference $\{x, y, z\}$ reference system in the initial position. $\mathbf{[R}_{R/Ref,o}\mathbf{]}$, which represents the resulting rotation from the initial β_o and γ_o rotations applied in this order, is given in Eq. (7.27). Here it is taken into account that in the initial position, with $\theta_R = 0$, the

z' axis depicted in Figure 7.2 is equal to the w axis of the roller coordinate system.

$$\left[\mathbf{R}_{\mathrm{R/Ref,0}}\right]=\left[\mathbf{R}_{\mathbf{x}}(\beta_o)\right]\left[\mathbf{R}_{\mathbf{w}}(\gamma_o)\right]=\begin{bmatrix} c_{\gamma_o} & -s_{\gamma_o} & 0 \\ c_{\beta_o}s_{\gamma_o} & c_{\beta_o}c_{\gamma_o} & -s_{\beta_o} \\ s_{\beta_o}s_{\gamma_o} & s_{\beta_o}c_{\gamma_o} & c_{\beta_o} \end{bmatrix} \tag{7.27}$$

$\mathbf{[R}_{\mathrm{R/G,o}}\mathbf{]}$ is then obtained pre-multiplying $\mathbf{[R}_{\mathrm{R/Ref,o}}\mathbf{]}$ by $\mathbf{[R}_{\mathbf{X}}(\alpha_o)\mathbf{]}$, taking into account that $\mathbf{[R}_{\mathbf{X}}(\alpha_o)\mathbf{]}$ rotates the axes $\{X_o, Y_o, Z_o\}$ to the axes $\{x_o, y_o, z_o\}$. This is expressed in Eq. (7.28).

$$\left[\mathbf{R}_{\mathrm{R/G,0}}\right]=\left[\mathbf{R}_{\mathbf{X}}(\alpha_o)\right]\left[\mathbf{R}_{\mathbf{x}}(\beta_o)\right]\left[\mathbf{R}_{\mathbf{w}}(\gamma_o)\right] \tag{7.28}$$

The θ_m rotation is determined as atan(p_1 / p_3). Specifically, the atan2 function is used, which returns an angle in the ($-\pi$, π) interval. The $\{X_o, Y_o, Z_o\}$ system is defined so that $\theta_m = 0$ in the initial position, as illustrated in Figure 7.3, and the origin of θ_m is in the positive Z_o axis. Then the rotation matrix $\mathbf{[R}_{\mathrm{Ref/G}}\mathbf{]}$ is calculated according to Eq. (7.29). This matrix rotates the global axes $\{X_o, Y_o, Z_o\}$ to the roller reference axes $\{x, y, z\}$ in the final position.

$$\begin{aligned}\left[\mathbf{R}_{\mathrm{Ref/G}}\right]&=\left[\mathbf{R}_{\mathbf{Y}}(\theta_m)\right]\left[\mathbf{R}_{\mathbf{X}}(\alpha_o)\right]\\ &=\begin{bmatrix} c_m & 0 & s_m \\ 0 & 1 & 0 \\ -s_m & 0 & c_m \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{\alpha_o} & -s_{\alpha_o} \\ 0 & s_{\alpha_o} & c_{\alpha_o} \end{bmatrix}\\ &=\begin{bmatrix} c_m & s_m s_{\alpha_o} & s_m c_{\alpha_o} \\ 0 & c_{\alpha_o} & -s_{\alpha_o} \\ -s_m & c_m s_{\alpha_o} & c_m c_{\alpha_o} \end{bmatrix}\end{aligned} \tag{7.29}$$

On the other hand, the **u**, **v** and **w** vectors corresponding to the final orientation of the rolling element are calculated by multiplying **[A]** with $\mathbf{u_o}$, $\mathbf{v_o}$ and $\mathbf{w_o}$, respectively. $\mathbf{u_o}$, $\mathbf{v_o}$ and $\mathbf{w_o}$ are given by the columns of the $\mathbf{[R}_{\mathrm{R/G,o}}\mathbf{]}$ matrix calculated in Eq. (7.28).

The **u**, **v** and **w** vectors calculated in this way are in the $\{X_o, Y_o, Z_o\}$ system. Now, they are to be transformed to the final $\{x, y, z\}$ system. For this purpose, they are pre-multiplied with $\mathbf{[R}_{\mathrm{Ref/G}}\mathbf{]}^{\mathrm{T}}$. The components of **u** in the $\{x, y, z\}$ system calculated in this way are denoted as u_x, u_y, u_z, and analogously for the components of **v** and **w**.

Finally, the remaining angles β, γ and θ_R that characterize the orientation of the rolling element are calculated according to Eqs. (7.30)–(7.32). These equations are derived taking into account the defined relationship between the $\{x, y, z\}$ and the $\{u, v, w\}$ coordinate systems (see the $\mathbf{[R}_{\mathrm{R/Ref,o}}\mathbf{]}$ matrix given in Eq. (7.27) as a reference).

$$\beta = \operatorname{atan}\left(\frac{v_z}{v_y}\right) \tag{7.30}$$

$$\gamma = -\operatorname{asin}\left(v_x\right) \tag{7.31}$$

$$\theta_R = \operatorname{sgn}\left(w_x \cos\gamma\right) \operatorname{acos}\left(-w_y \sin\beta + w_z \cos\beta\right) \tag{7.32}$$

7.2 Description of contact models

It is aimed to obtain detailed contact solutions for each rolling element–raceway contact of a given rolling element of the bearing. To this end, different partial slip contact models are considered. Each of them is described in the following subsections.

The following common assumptions are considered, also used in previous chapters of this thesis:

- The bodies are homogeneous, and the material's behaviour is linear elastic and isotropic.
- Coulomb's friction law is used, with a constant coefficient of friction. This is commonly applied in dry or in boundary lubricated contacts.

The motions of the inner and outer rings are prescribed, as well as the approach between them. The precise motion and position of the rolling element are to be determined as part of the solution so that the rolling element complies with

steady equilibrium. More precisely, quasi-steady equilibrium configurations of the rolling element are computed, since the load on it changes continuously during its orbital rotation around the bearing in other than purely axially loaded bearings.

The approach defined for the particular rolling element considered includes possible global deflections of the rings. The problem of load distribution among the different rolling elements of the bearing is assumed to have been previously solved, and is out of the scope of this chapter. Commonly used methods for the solution of this problem are explained in [Harris 2001]. Factors such as structural flexibility of the rings and manufacturing errors may considerably influence the resulting load distribution, and this is a field of open research (see e.g. [Amasorrain 2003], [Aguirrebeitia 2010], [Aithal 2015], [Heras 2018]).

7.2.1 Strip theory – ContRod2D program

A rolling element with two approximately elliptical contact patches is considered, one with the inner raceway and another with the outer raceway. The strip theory (cf. §2.1.2.3) is applied to solve the tangential contact problem in each contact. The necessary geometric and kinematic inputs for the definition of the rolling contact problem are obtained under the following assumptions:

- The rolling element is considered in its nominal orientation, with neither tilt nor skew.
- The inner and outer rolling element–race contact centres are considered to be located in the same rolling circle of the rolling element.
- The contact centres, which are prescribed, may be different from the nominal contact points.

The normal contact problem is solved first using Hertz theory. The same normal load N is prescribed for the inner and outer contacts. The crown radius of the rolling element and groove radii of the races are used in the lateral direction, and the longitudinal curvature radii of the rolling element and races at the contact centres in the longitudinal direction.

The tangential contact problem is solved by dividing the contact patch into parallel strips aligned with the x direction and applying Carter / Fromm's theory

(cf. §2.1.2.1) independently in each of them. The interaction between different strips is neglected. A pure longitudinal creepage is considered, constant in each strip. The longitudinal rigid slip velocities w_x are calculated as the difference of the relative velocities of the contact patch over each contacting surface. The contact relative velocities $V_{c,j}$ over the surface of each race j, $j \in \{I, E\}$, and over the surface of the rolling element, $V_{c,R}$, are given in Eqs. (7.33) and (7.34). These velocities are identical for perfect rolling on the nominal contact points of the inner and outer contacts.

$$V_{c,j} = |\omega_{m/j}|\, r_{\mathrm{roll,j}} = |\omega_m - \omega_j|\, r_{\mathrm{roll,j}} \tag{7.33}$$

$$V_{c,R} = |\omega_R|\, r_{\mathrm{roll,R}} \tag{7.34}$$

In these equations, $r_{\mathrm{roll,j}}$ are the rolling radii of each race (measured along the bearing Z axis), and $r_{\mathrm{roll,R}}$ is the rolling radii of the rolling element (measured along the rolling element z axis). These are variable in the lateral direction of the contact patch. The longitudinal creepage ξ is thus calculated according to Eq. (7.35). The contact velocity V_c is taken as the average of $V_{c,j}$ and $V_{c,R}$.

$$\xi = \frac{V_{c,R} - V_{c,j}}{V_c} = \frac{2\left(V_{c,R} - V_{c,j}\right)}{V_{c,R} + V_{c,j}} \tag{7.35}$$

For given ring rotation velocities ω_I and ω_E, ω_m and ω_R are iterated until the equilibrium of longitudinal forces and moments around the y axis are satisfied in the rolling element with the inner and outer contact resultants. Only the tangential contact forces contribute to these equilibrium conditions. This search for the rolling element equilibrium is a simplified version of the procedure explained in §7.2.2.2, and has been implemented in a program called ContRod2D. During this search, the position of the rolling element and contact patches remain fixed, and therefore the normal contact problem is calculated only once for each contact patch, while the tangential contact is recomputed in each iteration according to the changing ω_m and ω_R velocities and resulting creepages.

The inputs and outputs of ContRod2D are explained in the next subsections.

7.2.1.1 ContRod2D input

The input variables for the ContRod2D program are specified next.

- *GEO*. Vector with 7 real numbers with the basic geometric data of the rings–rolling element assembly, in the following order (cf. Figure 7.1): [R_p, D, r_I, r_E, R, α_o, α_R].

  Units: R_p, D, r_I, r_E and R are given in mm, and α_o, α_R in rad.

  Notes:

  - The location of the centres of the contact patches is determined with R_p, D, α_0 and α_R. It is assumed that the centres of the contact patches of the rolling element with the inner and outer races are located in the same rolling circle, though they need not be the nominal contact points for the given geometry. To have the centres of the contact patches displaced a distance s_0 in the lateral (s) direction of the contact with respect to their nominal position, the modified pitch radius R'_p, roller nominal diameter D' and roller angle α'_R given in Eqs. (7.36)–(7.38) are input, instead of their nominal values R_p, D and α_R. These equations are deduced from the geometry of Figure 7.1. In these equations, r_m is the average of the transverse radii of the inner and outer contact surface profiles. It may be calculated as $r_m = r_E / 4 + r_I / 4 + R / 2$.

$$\alpha'_R = \alpha_R + \operatorname{sgn}(\alpha_o)\ \operatorname{sgn}(\omega_I)\ \frac{s_0}{r_m} \tag{7.36}$$

$$D' = D - 2r_m(\cos\alpha_R - \cos\alpha'_R) \tag{7.37}$$

$$R'_p = R_p - r_m(\sin\alpha'_R - \sin\alpha_R)\ \sin|\alpha_o| \tag{7.38}$$

  - The profiles of the raceways are assumed to be concave, and that of the rolling element convex. R must be lower in magnitude than r_I and r_E.
  - Both α_o and α_R must be in the $(-\pi/2, \pi/2)$ interval.

- α_o is measured from the positive *Y* global bearing axis to the positive local rolling element *y* axis, being positive anti-clockwise as viewed towards the negative *X* direction.
- α_R is positive when the vertex of the rolling cone points towards the intersection of its axis with the axis of the rings.

  Nominally non-zero α_R should be positive, so that the vertex of the equivalent rolling cone may be on the rings' axis. In this way, an adequate kinematics may be obtained, in theory with perfect rolling for point contacts. On the other hand, small negative α_R values may appear in cases with zero nominal α_R, to equilibrate axial resultants of the contact frictional forces or centrifugal forces for example. This is the case of the spherical roller bearing (SRB) studied in §7.3.

  With nominally non-zero α_R (e.g. in tapered roller bearings), a third contact is necessary (the roller end–flange contact) to provide a force in the *y* axis to equilibrate the pair of (not aligned) normal forces between the rolling element and the inner and outer raceways. The normal force in that third contact has no influence on the calculation carried out by this program. Further, it is assumed that the friction force in this third contact is zero.

- *OM*. Vector with rotation velocities. It may contain either one or three real numbers, in the following order: [ω_I, (ω_m), (ω_R)].

  Units: rad/s.

  Notes:

  - It is assumed that $\omega_E = 0$, without loss of generality.
  - ω_I is mandatory, and ω_m and ω_R optional.

    If only ω_I is given, the program will iterate to obtain the ω_m and ω_R rotation velocities with which the equilibrium of forces is satisfied in the rolling element.

If ω_I, ω_m and ω_R are given, no iteration is carried out for the rolling element equilibrium. In this case, the contact solutions are computed for the given ω_I, ω_m and ω_R, and in general they will not comply equilibrium in the rolling element.

- *MAT*. Vector with two real numbers defining the individual elastic constants of each body of the rings–rolling element assembly, in the following order: [E_k, ν_k]. E_k and ν_k are, respectively, the individual Young's modulus and the coefficient of Poisson of each body. E_k is given in GPa. It is assumed that the inner and outer rings and the rolling element have the same elastic constants.
- *N*. Three-element vector with data for the normal problem: [N_i, N_e, B_dgt].
    - N_i, N_e: contact normal forces between the rolling element and the interior (i) and exterior (e) races, respectively. [kN]
    - B_dgt: bit to deactivate (0) or to activate the correction of the B parameter of effective lateral curvature according to Eq. (6.22).
- *f*. Two-element vector with the coefficients of friction [μ_i, μ_e] in the contacts between the rolling element and the interior and exterior races.
- *DISCR*. Vector with two positive integer numbers, [n_x, n_s], defining the number of elements in which the inner and outer contact ellipses are discretized in longitudinal (n_x) and lateral (n_s) directions.

  The discretization is similar to that of FASTSIM (cf. Figure 2.15): the contact ellipse is divided in n_s longitudinal strips of the same width, and each strip is divided in turn in n_x elements in longitudinal direction, equal in each strip, but different in different strips.

  The same discretization is used for both contact ellipses.

7.2.1.2 ContRod2D output

ContRod2D returns the following variables:

- *cont_i*, *cont_e*. Structures with the results of the interior (_i; between the rolling element and the inner race) and the exterior (_e; between the rolling element and the outer race) contact, respectively.

  Both structures have the same fields, being similar to the `sol` structure output by CECT (cf. §4.1.2.2). A difference is in the mesh structure, the mesh of CECT being regular with equal rectangular elements, and the mesh of ContRod2D being similar to that of FASTSIM, as indicated in §7.2.1.1.

  Fields of the *cont_i* and *cont_e* structures that are not present in the `sol` output structure of CECT or that are different are listed below. Scalar fields are written in italics, vector fields inside curly brackets ({}), and matrix fields inside square brackets ([]). The vectors below have $n_s + 2$ elements (each element corresponding to a strip of the contact patch discretization, plus an additional exterior strip in each side) and the matrices have $(n_x + 2) \times (n_s + 2)$ elements (each element corresponding to an element of the contact patch discretization, plus an additional exterior row and column in each side, adjacent to the contact patch limits). The ordering of elements in these output vectors and matrices is similar to that in CECT (cf. §4.1.2), i.e., in ascending *x* and *s* directions.

  - *a*, *b*: semiaxes of the contact ellipse, in longitudinal (*x*) and lateral (*s* or *y*) directions, respectively. [mm]
  - *d*: normal approach between the contacting bodies. [mm]
  - $p_{n,o}$: maximum contact normal pressure, which takes place at the central point of the contact ellipse. [MPa]
  - F_x: resultant contact force along local contact *x* direction (analogous to CECT `Fx` output variable). [kN]
  - M_y: resultant contact moment along local rolling element *y* axis. It is caused entirely by the p_x stresses. [N.m]
  - **{ξ}**: longitudinal creepages. They are assumed constant in each contact strip. [-]

Sign criteria: positive when the relative velocity of the rolling element surface with respect to the race surface is in the same direction as the contact patch travelling velocity in the relative movement with respect to the corresponding race. That is, a positive creepage corresponds to a braking creepage of the rolling element with respect to the race.

- **{ξ^*}**: normalized longitudinal creepages (cf. §2.1.2.1). [-]
- **{c'}**: length of the adhesion zone in each strip divided by the strip length. [-]
- **{N^*}**, **{F_x^*}**: normal and longitudinal forces per unit width in each contact strip. [N/mm]
- **[x]**, **[s]**: longitudinal (*x*) and lateral (*s*) local contact coordinates of the element centres forming the discretization of the contact ellipse. [mm]
- **[Y]**, **[Z]**: lateral (bearing axial) *Y* and vertical (bearing radial) *Z* coordinates of the discretization element centres, in the *YZ* radial plane of the bearing assembly. [mm]
- **{α}**: inclinations of the contact surface profile in the *YZ* plane, with respect to the axis of the bearing assembly. Analogous to the CECT `ang_prf_pcs` output variable (cf. §4.1.2.2). [rad]
- **[s_{rel}]**: slip velocities divided by the forward velocity of the central contact patch strip with respect to the race. [-]

  [w_x], **[w_y]**: rigid slip velocities in longitudinal (*x*) and lateral (*y*) directions. **[w_y]** = **[0]**. [mm/s]

  Analogous to `wx` and `ws` fields of CECT `sol` output structure, except for the units.

- **[v_x]**, **[v_y]**: slip velocities in longitudinal (*x*) and lateral (*y*) directions. **[v_y]** = **[0]**. [mm/s]
- **[P_{fric}]**: frictional power density, equal to $v_x \times p_x$ in each element. [mW/mm^2]

- **{W_{fric}}**: integrated frictional work in each contact strip after a full contact passage; cf. Eqs. (4.88) and (4.89). [mJ/mm^2]

In addition to the fields listed above, the following fields are present in the *cont_i* and *cont_e* structures: **[eldiv]**, **[p_n]** and **[p_x]**. These are analogous to the corresponding fields of the `sol` output structure of CECT.

Some of the fields are either redundant or trivial (with zero values), and are included for compatibility with post-processing scripts prepared for outputs from other contact models.

- ROT. Structure with the rotation velocities of the rolling element, with the following fields: $\omega_{m,o}$, $\omega_{R,o}$, ω_m, ω_R. Each field contains a real number, as described next.

  - $\omega_{m,o}$ and $\omega_{R,o}$ are the theoretical values of ω_m and ω_R, respectively, considering perfect rolling on the initial contact points. [rad/s]

  - ω_m and ω_R are the orbital and pitch rotation velocities of the rolling element. ω_m is aligned with the bearing Y axis, and ω_R with the rolling element v axis. [rad/s]

    When ω_m and ω_R are not given in the OM input, if a prescribed maximum number of iterations is reached without finding an equilibrium solution, zeros are returned.

7.2.2 Exact contact theory

The formulation of CECT has been detailed in Chapter 4, and its application to wheel–rail contact problems has been demonstrated in Chapter 6. Here, CECT is extended to contact problems in rolling bearings.

In the type of contact mechanics problems treated here, the CEM allows the construction of models with much fewer DOFs than the FEM described in §7.2.3 for a similar level of detail in the contact solution. Further advantages of the exact contact theory are the use of an Eulerian mesh, by which the mesh is restricted to a region only slightly larger than the contact patch itself, and that it is possible to solve steady rolling contact problems directly, as opposed to the

incremental or multi-stepping technique followed with the FE models described in §7.2.3. All this translates into much lower computational demands than the FEM simulation methodology for contact mechanics problems used in this thesis.

A starting point of the contact analysis is the definition of the geometry of the PCS, in which the contact patch of the rolling element with each raceway is contained. This is defined as a prismatic surface extruded along the *X* direction. Its profile is between the profiles of the rolling element and the raceway. Each contact is analyzed in the same local curvilinear {*x*, *s*, *n*} system associated to the PCS described in §4.1.1.1. It moves with the contact patch, and its origin is defined to be in the corresponding nominal contact point. The *s* and *n* axes, of variable direction, are contained in the radial *YZ* plane. The *s* axis is tangent to the profile of the contact surface, and the *n* axis is normal to it, pointing into the rolling element. The *x* axis is perpendicular to the radial *YZ* plane, pointing towards the direction of the relative movement of the rolling element with respect to each raceway. Figure 7.4 depicts the {*x*, *s*, *n*} systems for the inner and outer contacts in different lateral positions. The cross-section of a SRB is represented, which is studied in §7.3. The profiles of the potential contact surfaces are also represented, in dotted lines. The *X* and *Y* axes of the {*X*, *Y*, *Z*} system[28] are flipped for the analysis of each contact so that the *X* and *x* axes point in the same direction, to facilitate transformations between both systems.

The inner and outer contact problems are specified and solved independently, considering the relative motion of the rolling element with respect to each raceway. I.e., the raceway is taken as the rail, and is assumed to be fixed (with no movement), and the rolling element is taken as the wheel.

[28] In this chapter, the {*X*, *Y*, *Z*} system illustrated in Figure 7.1 is used for convenience. The orientation of this system is the same as the one of the global {*X*, *Y*, *Z*} system used in CECT, explained in Chapter 4, but its origin is different. The *Y* axis of the system illustrated in Figure 7.1 is at the *Z* = –`r_nom_r` coordinate of the global CECT system (cf. §4.1.2.1.1).

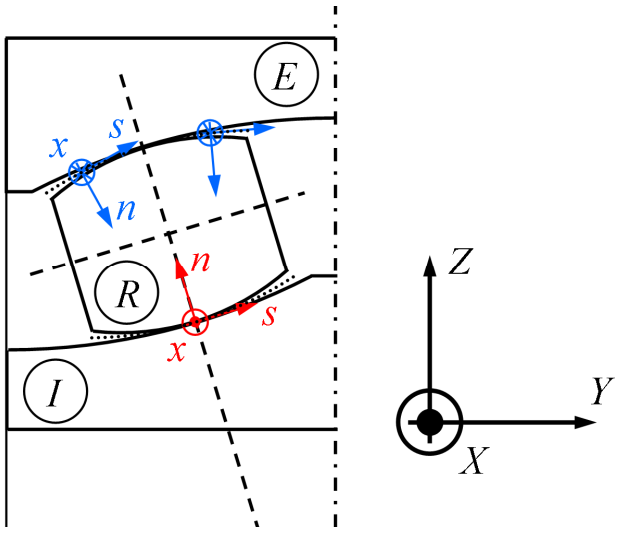


Figure 7.4. Potential contact surfaces and local contact coordinate systems in roller–raceway contacts for a rolling element.

7.2.2.1 Extensions in CECT for rolling bearings

The application of the exact contact theory to rolling bearings is affected by the following two fundamental geometric differences between rolling element–raceway and wheel–rail contact:

- The raceway is not straight in the rolling direction.
- The rolling radii of the contacting surfaces of the rolling element and raceway are not large with respect to their variations in the contact patch.

Each of these points is addressed in the following subsections.

7.2.2.1.1 Curvature of the raceway in the rolling direction

The raceways are taken as bodies of revolution, like the rolling element. The axes of revolution of each pair of contacting bodies (inner raceway–roller and outer raceway–roller) may not be aligned nor contained in the same plane. The axis of revolution of the raceway is arranged to be aligned with the *Y* axis in each contact.

The inner raceway–rolling element contact is counterformal in the rolling direction. Additionally, the longitudinal curvature radii of the outer raceway are considered to be much higher than those of the rolling element. Within linear elasticity, this implies that the longitudinal dimensions of the contact patches are much smaller than these radii. So, the hypothesis of a prismatic contact surface previously described is retained for both the inner and outer contacts.

On the other hand, the longitudinal curvatures of the raceways affect the rolling element–raceway normal undeformed distances. The longitudinal dimension of the inner contact patch decreases as a result of the convex longitudinal curvature of the inner raceway, and that of the outer contact patch increases as a result of the concave longitudinal curvature of the outer raceway. The normal undeformed distances are computed following mostly the same procedure explained in §4.1.1.2, which is based on computing intersections of the contacting surfaces with planes perpendicular to *x*. Only the computation of the intersections of the raceway surface is changed.

The intersections of the rolling element surface are computed in the same way as in the wheel–rail case, using Eqs. (4.13) and (4.14). The intersections of the raceway surface are no longer constant in the *x* direction as with a straight rail. These are computed with the same equations as for the rolling element, particularized for having the raceway axis of revolution parallel to the *Y* axis. In this way, the *Z* coordinates of the raceway surface intersection are calculated as $\mathrm{sgn}\left(r_{roll,j}\right)\times\sqrt{r_{roll,j}{}^2 - x^2}$, with $j \in \{I, E\}$. In this expression, $\mathrm{sgn}(r_{\mathrm{roll,j}}) = 1$ if the raceway axis of revolution is below the point $(x, s) = (0, 0)$ of the contact patch along the *Z* axis, and –1 otherwise (in accordance with the sign criteria for `r_nom_r` indicated in §4.1.2.1.1). $r_{\mathrm{roll,j}}$ is equal to $|Z_r|$, being Z_r the *Z* coordinates of the raceway profile. The raceway axis of revolution is defined to be at $x = 0$, and the origin of the *Z* coordinate is on this axis (cf. Figure 7.1). The *Y* coordinates of the intersection are the same as those of the raceway profile on the *YZ* plane. For correct results, the raceway profile should not cross the $Z = 0$ line, and the coordinates Z_r have the same sign as `r_nom_r`.

The longitudinal curvature of the raceway also implies that in steady rolling, the rotation velocity vector of the rolling element in its relative motion with respect to the raceway *j*, $\boldsymbol{\omega}_{\mathbf{R/j}}$, is not aligned with the axis of revolution of the rolling

element *v*, unless *v* is parallel to the bearing *Y* axis. This is because $\boldsymbol{\omega}_{\mathbf{R/j}}$ is the vectorial sum of $\omega_{m/j}$ and ω_R. It is allowed for through the CECT input parameter `bit_l6` and the parameters of the 6th line of the particular data for each case in the CECT input file, cf. §4.1.2.1.1. The rigid slip velocities are computed as the projections of the relative velocity vectors over the PCS, following the procedure explained in §4.1.1.3.

7.2.2.1.2 Varying rolling radii of the contacting surfaces

The relation of velocities of Eq. (1.7), reproduced here for convenience, holds in the tangential contact problem:

$$\mathbf{v} = \mathbf{w} + \frac{\mathrm{D}\mathbf{u}}{\mathrm{D}t} \approx \mathbf{w} - V_c \frac{\partial \mathbf{u}}{\partial x} + \frac{\partial \mathbf{u}}{\partial t} \tag{7.39}$$

In Eq. (7.39), **v** is the slip velocity, **w** is the rigid slip velocity and **u** is the tangential elastic displacement difference of the contacting surfaces. They are two-dimensional vectors with components in the (*x*, *s*) tangent directions of the contact surface. These vectors are functions of the (*x*, *s*) surface position in the contact reference frame and in non-stationary problems of time as well. The uppercase time derivative in the central side of the equation denotes a Lagrangian derivation associated with the material particles of the contacting surfaces. V_c is the rolling or contact velocity, approximately the velocity at which the material particles of the bodies flow through the Eulerian {*x*, *s*, *n*} coordinate system attached to the contact patch along the negative *x* direction; or in other words, the velocity of the contact over the contacting surfaces in the relative motion with respect to each contacting surface. In wheel–rail contact, V_c may be taken as uniform throughout the whole contact. In rolling bearings, it may be variable because the rolling radii of the bodies may be variable. This affects the computation of the surface velocities associated with the elastic deformations of the contact surfaces, i.e. the second term in the central side of Eq. (7.39).

The discretization in time of Eq. (7.39) leads to:

$$\mathbf{v} = \mathbf{w} + \frac{\mathbf{u} - \mathbf{u}'}{\Delta t} \tag{7.40}$$

In Eq. (7.40), Δt is the chosen increment for the time discretization, **u**, **v**, and **w** are field magnitudes in the current time instant t, and **u'** is the elastic displacement difference in the previous time instant and position. For the contacting particles occupying position (x, s) in the current time instant t, $\mathbf{u}' = \mathbf{u}'(x, s, t) = \mathbf{u}(x + V_c\,\Delta t, s, t - \Delta t)$. The quantity $V_c\,\Delta t$ is designated as Δq, as indicated in §2.1.3. Δt is fixed for the whole contact patch, so with a variable V_c, Δq is variable in the contact patch. A reference Δq, Δq_o, is chosen in the contact patch. This may be the value of Δq at $(x, s) = (0, 0)$, for example. **u'** is obtained in different ways in transient and in steady-state rolling contact problems.

In transient rolling contact problems, the elastic displacement field in the previous time instant is known. It is computed with the contact stresses of the previous time instant, $\mathbf{p}' = \mathbf{p}(x, s, t - \Delta t)$, according to Eq. (7.41).

$$\mathbf{u'_{Ii}} = \left[\mathbf{IC'_{IiJj}}\right] \mathbf{p'_{Jj}} \tag{7.41}$$

Eq. (7.41) expresses the discretized relationship between the contact surface stresses and elastic displacements at $t - \Delta t$. A similar relationship may be written for the current time t (cf. Eq. (2.100)). $\mathbf{p'_{Jj}}$ is the vector of contact stresses at time $t - \Delta t$, with element p'_{Jj} of this vector being the contact stress along direction j on element J of the contact mesh. The convention of summation over repeated indices is used in Eq. (7.41), with index J extending over all the elements of the mesh entering into contact, and index $j \in \{x, s, n\}$. $\mathbf{[IC'_{IiJj}]}$ is a matrix of ICs of the previous time instant t'. As explained in §2.1.3, element IC'_{IiJj} of this matrix gives the elastic displacement difference along direction $i \in \{x, s\}$ on element I due to a unit stress along direction j on element J, at the position $(x_I + \Delta q, s_I)$ occupied by the particles of element I in a previous instant. Here it is generally desirable a common Δq_o for the whole contact mesh for efficiency in the computation of ICs and their convolution[29] with the contact

[29] With a regular mesh, the convolution of the ICs with the contact stresses may be performed efficiently in the Fourier domain with FFTs, and without forming complete $NE \times NE$ $\mathbf{[IC'_{IiJj}]}$ matrices for a 2D mesh with NE elements, as explained in §4.1.1.4.2. This, however, is not relevant to the subject discussed here, and the variables and equations in this section are expressed in the space domain for convenience.

stresses. To get the desired displacements at ($x_I + \Delta q$, s_I, $t - \Delta t$), the initially obtained displacements at ($x_I + \Delta q_o$, s_I, $t - \Delta t$) are interpolated in x direction (or extrapolated as necessary at extreme positions of the mesh).

In steady rolling contact problems, the transient term $\partial\mathbf{u}/\partial t$ in the right-hand side of Eq. (7.39) vanishes, and the second term of the right-hand side of Eq. (7.40) becomes a purely spatial gradient. Therefore, the difference $\mathbf{u} - \mathbf{u}'$ in this term may be scaled proportionally to V_c in each mesh position.

($\mathbf{u} - \mathbf{u}'$) is split into the contributions due to the normal pressures and the tangential stresses, $(\mathbf{u} - \mathbf{u}')_n$ and $(\mathbf{u} - \mathbf{u}')_t$, with $(\mathbf{u} - \mathbf{u}') = (\mathbf{u} - \mathbf{u}')_n + (\mathbf{u} - \mathbf{u}')_t$. Following the Panagiotopoulos process to solve the generally coupled normal-tangential contact problem, $(\mathbf{u} - \mathbf{u}')_t$ is initially unknown, and $(\mathbf{u} - \mathbf{u}')_n$ is known in the tangential problem. If the available $(\mathbf{u} - \mathbf{u}')_n$ corresponds to a common Δq_o in the whole contact patch, the value for the applicable Δq in each position is obtained by scaling $(\mathbf{u} - \mathbf{u}')_n$ with the ratio $\Delta q / \Delta q_o$.

For $(\mathbf{u} - \mathbf{u}')_t$, this scaling is realized by modifying the elements of the matrices of ICs with which this difference is computed within the solver used for the tangential problem. This may be calculated according to Eq. (7.42) or to Eq. (7.43). In either case, the fact that in steady rolling $\mathbf{p}' = \mathbf{p}(x, s, t)$ is used. The subindices *IiJj* used in Eq. (7.41) are dropped in these equations for brevity, and here $j \in \{x, s\}$.

$$(\mathbf{u} - \mathbf{u}')_t = ([\mathbf{IC}] - [\mathbf{IC}'])\,\mathbf{p} = [\mathbf{dIC}]\,\mathbf{p} \tag{7.42}$$

$$(\mathbf{u} - \mathbf{u}')_t = [\mathbf{IC}](\mathbf{p}(x, s, t) - \mathbf{p}(x + \Delta x, s, t)) = [\mathbf{IC}]\,\mathbf{dp} \tag{7.43}$$

Referring to Eq. (7.42), if the initially computed difference matrix **[dIC]** corresponds to a common Δq_o in the whole contact patch (i.e. if **[dIC]** = **[dIC**(Δq_o)**]**), a scaled difference matrix **[dIC$_{sc}$**(Δq)**]** is used within the tangential solver instead of **[dIC**(Δq_o)**]**. Each element of **[dIC$_{sc}$**(Δq)**]** is computed as $dIC_{sc,IiJj}(\Delta q) = dIC_{IiJj}(\Delta q_o) \times \Delta q_I / \Delta q_o$, being Δq_I the value of Δq at the position of element I. This is done within the version of the ConvexGS solver implemented in CECT. Regarding Eq. (7.43), which is used in the SteadyGS tangential solver, Δx is the element size in the longitudinal direction, common in the whole contact mesh. In this case, the scaling of **[IC]** is performed

analogously as explained for **[dIC]** in Eq. (7.42). In either case it has to be borne in mind that this scaling is valid only for computing differences $(\mathbf{u} - \mathbf{u}')_t$, but not values **u** or **u'**. Eqs. (7.42) and (7.43) become:

$$\left(\mathbf{u}-\mathbf{u}'\right)_t = \left[\mathbf{dIC_{sc}}\left(\Delta q\right)\right]\,\mathbf{p} \tag{7.44}$$

$$\left(\mathbf{u}-\mathbf{u}'\right)_t = \left[\mathbf{IC_{sc}}\left(\Delta q\right)\right]\,\mathbf{dp} \tag{7.45}$$

The calculation of V_c is carried out in a similar way as described in §7.2.1, computing first contact velocities $V_{c,j}$ and $V_{c,R}$ over the surfaces of the raceway and the rolling element. For this purpose, the input rotation velocity of the wheel $\boldsymbol{\omega}_{\mathbf{R/j}}$ has to be decomposed into orbital and pitch rotation velocities. The contact velocities may be approximated as constant in x due to the short longitudinal dimension of the contact patch. V_c and Δq may therefore be taken as variable only in the s direction. Taking this into account, a vector **{Δq}** is obtained as outcome from this calculation, which contains a Δq value for each lateral position of the mesh of the PCS. The main steps of this calculation are as follows:

- Location of the point A of the rolling element axis of revolution which intersects the radial plane YZ_0, parallel to the bearing YZ plane and passing through the origin of the local contact coordinate system. The coordinates of point A in the global coordinate system are (X_A, Y_A, Z_A).

  For this purpose, first the axial coordinate of the origin of the local contact coordinate system is determined in the rolling element (wheelset) coordinate system, $y_{w,C}$. With reference to Figure 4.3, assuming that point C is the origin or the local contact coordinate system, point O is the point on the rolling element axis having the same y_w coordinate. The coordinates of this point on the rolling element coordinate system are then $(x_{w,O}, y_{w,O}, z_{w,O}) = (0, y_{w,C}, 0)$, and the transformation to its coordinates (X_O, Y_O, Z_O) in the $\{X, Y, Z\}$ system is given by Eqs. (7.46)–(7.48). In these equations, c_γ and s_γ stand, respectively, for the cosine and the sine of γ, and (Y_R, Z_R) are the (Y, Z) coordinates of the centre of the rolling element axis. ψ and ϕ are the yaw and roll angles of the rolling element axis, cf. Figure 4.3.

$$X_O = -y_{w,C}\, c_\phi\, s_\psi \tag{7.46}$$

$$Y_O = y_{w,C}\, c_\phi\, c_\psi + Y_R \tag{7.47}$$

$$Z_O = y_{w,C}\, s_\phi + Z_R \tag{7.48}$$

If the rolling element axis of revolution is parallel to the radial plane, point A is in the same position as point O, and therefore $(Y_A, Z_A) = (Y_O, Z_O)$. Otherwise, Y_A and Z_A are calculated with Eqs. (7.49) and (7.50). X_C is defined by the input `x_0`, and t_γ stands for the tangent of γ.

$$Y_A = Y_O + (X_O - X_C) / t_\psi \tag{7.49}$$

$$Z_A = Z_O + (X_O - X_C)\, t_\phi / s_\psi \tag{7.50}$$

- Calculation of the longitudinal component of the velocity of A, $v_{A,x}$. The velocity of A, $\mathbf{v_A}$, is calculated by applying $\mathbf{v_A} = \mathbf{v_{P/j}} + \boldsymbol{\omega}_{\mathbf{R/j}} \times \mathbf{r}_{\mathbf{A\text{-}P}}$, being $\|\mathbf{v_{P/j}}\|$ given by the input parameter `sftx_o_vel`, and its cosine directors by `dir_vx`, `dir_vy` and `dir_vz`.
- The orbital radius of A, $r_{m,A}$, is equal to $|Z_A|$ if $|$`r_nom_r`$| > 0$. Otherwise (for a straight raceway), $r_{m,A} = \infty$.
- The orbital rotation velocity of the rolling element, $\omega_{m/j}$, is calculated as $\omega_{m/j} = v_{A,x} / r_{m,A}$ if $|$`r_nom_r`$| > 0$. For a straight raceway, $\omega_{m/j} = 0$. ω_m is also the rotation velocity of the YZ_0 plane. The rotation velocity of the rolling element with respect to the YZ_0 plane (i.e., the pitch velocity of the rolling element) is equal to $\boldsymbol{\omega}_{\mathbf{R/j}} - \omega_{m/j}\, \mathbf{j}$, being $\mathbf{j}$ the unit vector aligned with the Y axis.
- For a curved raceway, $V_{c,j}$ is calculated for each lateral position of the mesh of the PCS applying Eq. (7.33). For a straight rail, $V_{c,j}$ is taken as equal to $v_{A,x}$. $V_{c,R}$ is also calculated at each lateral position of the mesh, at the $x = 0$ section. It is calculated using the formulation developed in §4.1.1.3, considering zero forward velocity of point A, and the rotation velocity of the rolling element with respect to the YZ_0 plane calculated

in the previous step. $V_{c,R}$ is taken as the longitudinal component of the velocity calculated in this way, with changed sign.

- For the general case of contacting bodies with dissimilar elastic constants, V_c is computed as the weighted average of $V_{c,j}$ and $V_{c,R}$, using the elastic constant $E_B^* = E_B / (1 - \nu_B^2)$ of each body B as weight. The values for all the lateral positions of the contact mesh are collected in a vector $\mathbf{\{V_c\}}$. The value of V_c corresponding to the $s = 0$ position is taken as the reference contact velocity.
- The $\mathbf{\{\Delta q\}}$ vector is obtained as $\Delta q_o \times \mathbf{\{V_c\}} / V_c(s{=}0)$. Δq_o is the `dq` parameter given as input.

A possible refinement on the above procedure could be to retain the velocities $V_{c,j}$ and $V_{c,R}$ to compute the velocities due to elastic deformations of each surface separately instead of using a single, averaged V_c. This would require doubling the amount of IC matrices to be handled. However, if the difference between these velocities is not small, there will be large rigid slip velocities, which in linear elasticity will dominate the resultant slip velocities **v** as Kalker pointed out [Kalker 1990], [Kalker 2001]. In this case, a small improvement in the computation of the term $(\mathbf{u} - \mathbf{u}') / \Delta t$ of Eq. (7.40) is not relevant.

In the SRB studied in §7.3, it has been observed that using a variable contact velocity across the contact patch to compute the velocities $(\mathbf{u} - \mathbf{u}') / \Delta t$ has little influence on the tangential problem solution, because the variations of the rolling radii are relatively modest and because there are large rigid slip velocities in most of the contact patch as a result of the curved profiles of the roller and the raceways. However, having a curved raceway in the rolling direction makes possible to have large variations in the rolling radii across the contact patch without having large rigid slip velocities. For example, in the case of steep-angle tapered roller bearings, this effect may be more relevant.

7.2.2.2 Computation of equilibrium configurations of the rolling element – equil_roller program

A Newton-Raphson iteration procedure is carried out to obtain the position, orientation and rotation velocities of the rolling element with which its steady

rolling equilibrium is satisfied. The main steps of the process are illustrated in the flowchart of Figure 7.5.

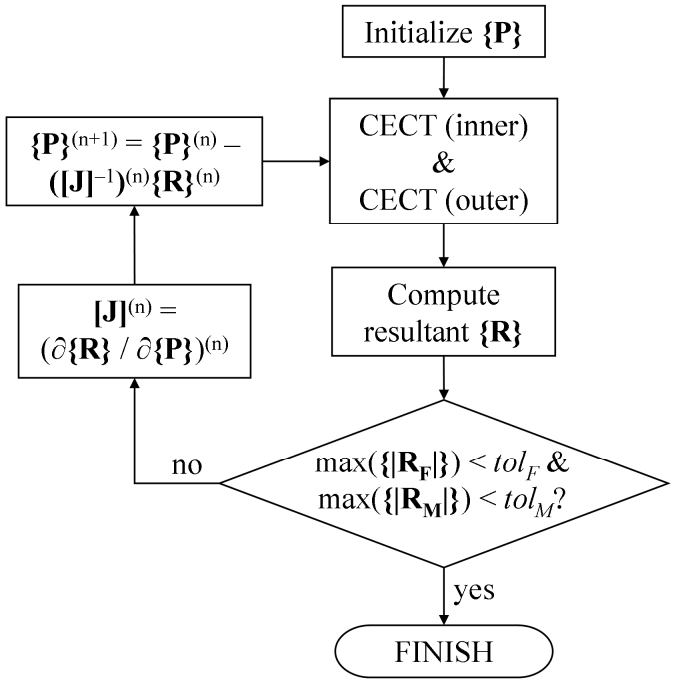


Figure 7.5. Flowchart of the search for the steady equilibrium configuration of the rolling element with contact solutions from CECT.

In the steady equilibrium configuration, the sum of all the forces and moments acting on the rolling element, including the inertia forces, is zero, while the rolling element is in a steady configuration in the radial *YZ* plane. The rolling element configuration is defined with the following variables, which are arranged in a vector **{P}**:

- Position, defined by displacements d_Y and d_Z of the rolling element centre in the *YZ* plane with respect to its nominal position.
- Rotation velocities ω_m and ω_R.

- Orientation, defined by rolling element tilt and skew angles β and γ[30].

Apart from the forces in the contacts with the raceway, an additional resultant force and moment may be considered acting on the rolling element. These may be due for example to inertia effects, contact with the cage or lubricant churning. The resultant forces and moments due to the contact and additional force and moment acting on the rolling element are summed in a vector **{R}**.

The process illustrated in Figure 7.5 for the determination of the steady equilibrium configuration of the rolling element has been implemented in a program named equil_roller. The main steps of this process are as follows:

1. **{P}** is initialized. This may be done with the nominal rolling element position, or with a previously obtained equilibrium solution for a similar problem if one is available.

2. The inner and outer rolling element–raceway contacts are solved with CECT for the rolling element configuration defined in **{P}**.

   Although here cases with specifically two contacts in the rolling element are considered, the extension of this procedure to a different number of contacts would be conceptually straightforward. Also, a workaround could be applied to allow for cases with more contact patches in the rolling element (e.g. four-point contact slewing bearings, or tapered roller bearings) within the two-contact procedure. This would consist on extending the PCS of the inner and/or the outer contacts as necessary, so that more than one contact patch (between the same two pair of bodies) is accommodated in the same contact calculation (e.g., the roller end–flange contact, in addition to the corresponding roller–raceway contact). This would be possible if the separation between the contacting surfaces remains small between the different contact patches

[30] In the case of ball bearings, the possibility of what could be considered non-zero tilt or gyroscopic rotation velocity is covered by a non-zero skew angle (though regarded here as a component of the pitch rotation velocity, according to the described DOFs of the rolling element configuration).

(recall the limitation illustrated in Figure 4.9a for profiles that separate and join more than once).

3. **{R}** is calculated considering the contact solutions obtained in the previous step. **{R}** is composed of subvectors $\mathbf{\{R_F\}}$ and $\mathbf{\{R_M\}}$ of forces and moments, respectively, each of them of 3 elements.

4. The condition for steady equilibrium is **{R}** = **{0}**. To check this condition, different tolerances are defined for $\mathbf{\{R_F\}}$ and $\mathbf{\{R_M\}}$, namely tol_F and tol_M. If all the elements of $\mathbf{\{R_F\}}$ and $\mathbf{\{R_M\}}$ are below their respective tolerance (in absolute value), the equilibrium configuration of the roller has been found. Otherwise, a different **{P}** is tried. **{P}** is updated according to a Newton-Raphson algorithm as defined in the following steps.

5. **[J]** is calculated, the 6×6 Jacobian matrix that contains the sensitivities of each element of **{R}** to each element of **{P}**. This involves the solution of 6 pairs of inner and outer roller–raceway contact problems with CECT, one for each element of **{P}**, P_i. Each of these pairs of contact problems is solved considering a modified $\mathbf{\{P'_i\}}$ vector, in which P_i is given a small increment δP_i. (i.e., P_i is substituted by $P_i + \delta P_i$), and the rest of the elements are the same as in **{P}**. With each of these pairs of contact solutions, a modified resultant is obtained on the roller, $\mathbf{\{R'_i\}}$. The column of **[J]** corresponding to P_i is calculated as $\partial\mathbf{\{R\}} / \partial P_i \approx (\mathbf{\{R'_i\}} - \mathbf{\{R_i\}}) / \delta P_i$.

6. **{P}** is updated as $\mathbf{\{P\}}^{(n+1)} = \mathbf{\{P\}}^{(n)} - (\mathbf{[J]}^{-1})^{(n)}\ \mathbf{\{R\}}^{(n)}$. The superindices in parentheses (n) and (n+1) denote the iteration number.

7. With the updated **{P}**, the process returns to step 2. Steps 2–6 are looped until **{R}** is sufficiently close to **{0}**, as defined by the tolerances tol_F and tol_M.

The contact problems in step 2 above on the one hand, and in step 5 on the other, may be solved with different meshes, using coarser meshes in step 5 to save computation time. Meshes of roughly 50×50 elements have been used in the study of §7.3.4 for the contact problems of step 2.

The matrix **[J]** presents a significant noise level due to discretization error when some element of the contact mesh changes from inside to outside contact or vice versa in some of the contact solutions computed in step 5. This poses a considerable difficulty for the convergence of the Newton-Raphson process described above, and in some cases leads to divergence. This problem has been tackled by using very small increments δP_i, to avoid changes in the set of elements making up the contact patch. But excessively small δP_i increments could lead to being left with too few significant digits when computing the differences (**{R'**$_i$**}** – **{R**$_i$**}**), so some testing is necessary to arrive at adequate δP_i values for the described numerical computation of **[J]**.

The following subsections describe the input and output of the equil_roller program.

7.2.2.2.1 equil_roller input

The input for the equil_roller program is described next.

The input is supplied in a file with a specified name *finp*. The program creates .txt input files for CECT with the *finp* root and suffixes *_i*-int and *_i*-ext for the inner and outer contacts, respectively, being *i* the number of equilibrium iteration padded if necessary with zeros up to 3 digits. A .log file is also created, with the same root and with suffix -eq_ro, with information of the equilibrium iterations and the intermediate contact solutions obtained.

The variables that must be defined in the *finp* input file are as follows:

- *GEO*. Structure with the basic geometric data of the rings–rolling element assembly, with the following fields: R_p, D, prf_I, prf_E, prf_R, R, α_o, α_R].

  R_p, D, R, α_o and α_R are as defined in Figure 7.1.

  prf_I, prf_E and prf_R are the text input files in which the profiles of the inner ring (*I*), outer ring (*E*), and rolling element (*R*) are defined. They have the same format as the `prf_w` and `prf_r` CECT input files, cf. §4.1.2.1.3.

  Units: R_p, D, R and the profiles defined in the files prf_I, prf_E and prf_R are given in mm. α_o and α_R are given in rad.

Notes:

- The rolling element profile defined in prf_R must be a convex circular profile of radius R. R must be lower in magnitude than the transverse curvature radii of the inner and outer raceway profiles.

- The definition of the profiles in their local (y_L, z_L) plane must be as follows. All the profiles must be tangent to the $z_L = 0$ horizontal at the (0, 0) point. The rolling element profile must be defined in $z_L > 0$, and therefore with its centre located at (0, R).

  According to the positioning of the profiles described in §7.2.2.2.2, the nominal contact point on each race profile in their local (y_L, z_L) plane is at its (0, 0) point, and at the ($-R \sin(\alpha_R)$, $R\,(1 - \cos(\alpha_R))$) point on the rolling element profile.

  The rolling element profile defined in prf_R is used for both the inner and outer contacts, with a nominal roll angle of α_o for the inner contact, and $\pi - \alpha_o$ for the outer contact, as indicated in §7.2.2.2.2. Taking into account as well that the X and Y axes of the inner and outer contacts have opposite directions, the direction of the s coordinate of the rolling element profile is different in the inner and outer contacts. This has to be taken into account in the definition of the inner and outer raceway profiles, so that their directions are consistent with that of the rolling element profile in each contact (cf. Figure 4.27). For example, if the s coordinate on the rolling element profile goes from left to right in its local (y_L, z_L) plane at the (0, 0) point, the s coordinate on the profile of the inner raceway must go also from left to right at (0, 0), and that of the profile of the outer raceway from right to left. In this way, the orientation of the local contact coordinate systems depicted in Figure 7.4 is achieved.

- *OM*. Structure with the absolute rotation velocities of the inner and outer rings, with fields *I* for the inner ring and *E* for the outer ring. The axes of revolution and rotation of both rings coincide with the *Y* axis.

  Units: rad/s.

- *MAT*. Structure with the elastic properties of the materials, with the following fields:

  - E_R, E_I, E_E: Young's moduli of the rolling element (*R*), inner ring (*I*) and outer ring (*E*). [GPa]
  - ν_R, ν_I, ν_E: coefficients of Poisson of the rolling element (*R*), inner ring (*I*) and outer ring (*E*). [-]
  - `mat_dgt`: digit with which spatially variable material properties may be specified. Cf. §4.1.2.1.1. In relation to this, the following fields are given: $[f]_{[B],inf}$, $r_{inf,[f],[B]}$, $exp1_{[f],[B]}$ and $exp2_{[f],[B]}$. [*f*] stands for an elastic property ($[f] \in \{E, \nu\}$) and [*B*] for a contacting body ($[B] \in \{R, I, E\}$). These fields are analogous to the corresponding fields included as common data in the CECT main input file for `mat_dgt` = 2 (cf. Table 4.3).

- *CAR*. Structure with the following fields that determine the load in the rolling element:

  - *d*: approach between the inner and outer rings of the bearing (>0). [mm]
  - dir_d: direction of *d* in the *YZ* plane. With origin in the positive *Z* axis and positive counter clock-wise while looking towards the negative *X* direction, considering that the approach *d* is applied to the inner ring, being the outer one fixed. [rad]
  - m_R: mass of the rolling element. It is used to calculate the centrifugal force F_c on it, which acts along the *Z* axis. [kg]
  - $I_{xx,R}$: inertia moment of the rolling element about an axis perpendicular to its axis of revolution and passing through its centre of gravity. It is used to calculate the gyroscopic moment M_g on it, which acts along the *X* axis. [kg·m^2]

NOTE: for the calculation of the centrifugal force and gyroscopic moment, it is assumed that centre of gravity of the rolling element is located in the centre of its nominal rolling circle. The inertia forces of the rolling element are updated in each equilibrium iteration, according to the position of the rolling element in the *YZ* plane. They are calculated as follows:

$$F_c = m_R \, \omega_m^{\,2} \, R_{p,act} \tag{7.51}$$

$$M_g = -I_{xx,R} \, \omega_R \, \omega_m \sin(\alpha_o + \beta) \tag{7.52}$$

$R_{p,act}$ in Eq. (7.51) is the current pitch radius, equal to $R_{p,o} + d_Z - R \sin(\alpha_R) \sin(\alpha_o + \beta)$.

- F_{ad}, M_{ad}: additional force (F_{ad}) and moment (M_{ad}). They are 3-element vectors, with the components in the {*X*, *Y*, *Z*} system of the additional force and moment acting on the rolling element (apart from the forces and moments from the inner and outer contacts computed in CECT, and from inertia effects).

  Units: F_{ad} is given in kN, and M_{ad} in N.m.

- *f*. Structure with *i* (interior) and *e* (exterior) fields, with the data of the coefficients of friction for each rolling element–raceway contact. Each field may be a scalar value (with the constant coefficient of friction for the whole contact), or a character string specifying the file name where the coefficient of friction data is defined, which may be variable in space. The format of the file for the latter case is described in §4.1.2.1.4.

- *DISCR*. This structure defines the mesh and ICs. It is composed of structures *i* and *e*, corresponding to each rolling element–raceway contact. Each of these structures in turn has the following fields:

  - `c_o_n`, `ics`, `x_0`, `s_0`, `x_inf`, `x_sup`, `s_inf`, `s_sup`, `nx`, `ns`: as defined for CECT, see §4.1.2.1.1. All these fields are mandatory.

The remaining fields listed below are optional:

- n_{xj}, n_{sj}: fields with analogous meaning to `nx` and `ns`. If they are given, they are used in the contact solutions for the calculation of the Jacobian of the equilibrium equations (step 5 in the equilibrium calculation procedure explained above), and `nx` and `ns` are used only for the solution of the rolling element–raceway contacts for the current rolling element configuration (step 2). If they are not given, `nx` and `ns` are used in both steps 2 and 5.

- aj_x, aj_s: bits to activate automatic adjustment of the limits of the PCS in x and s directions. Each bit applies for both the inner and outer contacts. If a bit is given with a value of 1, the limits of the PCS are adjusted in the corresponding direction (x or s) in each equilibrium iteration. The adjustment is carried out based on the overlap of the undeformed contact surfaces, and using the f_x and f_s factors defined below. In any case, the limits of the mesh are bounded by the values `x_inf`, `x_sup`, `s_inf` and `s_sup` given as input. `x_inf`, `x_sup`, `s_inf` and `s_sup` have to be defined with adequate margin, so that the contact patches fall entirely within their limits in all the equilibrium iterations.

  Default value: 0.

- f_x, f_s: scaling factors to calculate the limits of the PCS in x and s directions, when the automatic adjustment of these limits is carried out (i.e. when aj_x or aj_s = 1). Each f_{xs} ($xs \in \{x, s\}$) factor is taken into account only when the corresponding aj_i bit is equal to 1. Each factor applies for both the inner and outer contacts.

  To adjust automatically the limits of the PCS, first the overlap of the undeformed surfaces is calculated. This is identified as the region in which the undeformed distance h is negative. h is calculated following the procedure developed in §4.1.1.2, in the mesh defined by the mandatory inputs of *DISCR* listed above. The overlap region is bounded by the coordinates $xs_{\mathrm{ol,min}}$, $xs_{\mathrm{ol,max}}$

in each *xs* direction. The corresponding range is calculated as $\Delta xs_{ol} = xs_{ol,max} - xs_{ol,min}$, and the centre as $xs_{ol,med} = (xs_{ol,max} + xs_{ol,min}) / 2$. Each f_{xs} factor is used to scale the corresponding range Δxs_{ol}. The adjusted limits are thus calculated as $xs_{min} = \max(xs_{inf}, xs_{ol,med} - f_{xs}\,\Delta xs_{ol} / 2)$, and $xs_{max} = \min(xs_{sup}, xs_{ol,med} + f_{xs}\,\Delta xs_{ol} / 2)$, being xs_{inf} equal to `x_inf` or `s_inf`, and xs_{sup} equal to `x_sup` or `s_sup`, as applicable. The number of elements in each direction are defined with the inputs `nx` and `ns`. A final adjustment of the limits xs_{min} and xs_{max} is carried out so that one of the APs falls at $xs = 0$.

Default value: 0.

Recommended values: for circular contacts, around 0.6; for very elongated contacts, 1.1 in the long direction, and 0.6 in the short direction.

- *NUM*. Structure with numerical parameters, with the following field and structures:
    - *maxite*: maximum number of allowed equilibrium iterations. The execution of the program is stopped when this number of equilibrium iterations is reached, printing a warning message if equilibrium has not been achieved within the specified tolerances.
    - *tol*: structure with *F* and *M* fields, with the allowed equilibrium tolerances for forces (*F*) and moments (*M*) in absolute value in each direction of the {*X*, *Y*, *Z*} system. These are the tol_F and tol_M tolerances depicted in the flowchart of Figure 7.5.

      Units: *F* in kN, and *M* in N.m.
    - *djac*: structure with fields $\Delta d_{Y,u}$, $\Delta d_{Z,u}$, $\Delta \omega_{m,u}$, $\Delta \omega_{R,u}$, $\Delta\beta$ and $\Delta\gamma$, with increments corresponding to each of the previously described DOFs defining the configuration of the rolling element **{P}**, for the numerical calculation of the Jacobian **[J]** of the equilibrium equations. These are the previously referred δP_i

increments, normalized in the cases with "*,u*" subindex as indicated next.

Units:

- $\Delta d_{Y,u}$ and $\Delta d_{Z,u}$ in mm/mm. These are normalized increments, per unit of rigid approach *CAR.d*.
- $\Delta\omega_{m,u}$ and $\Delta\omega_{R,u}$ are dimensionless. These are normalized increments, for ring rotation velocities ω_I = –1 rad/s and ω_E = 0.
- $\Delta\beta$ and $\Delta\gamma$ are in rad.

- *cfg_ini* (optional): structure with fields d_Y, d_Z, $\omega_{m,u}$, $\omega_{R,u}$, β and γ, defining the initial configuration of the rolling element. $\omega_{m,u}$, $\omega_{R,u}$ are normalized orbital and pitch rotation velocities of the rolling element, corresponding to ring rotation velocities ω_I = –1 rad/s and ω_E = 0. The rest of the variables are the previously described DOFs defining the position of the rolling element in the radial plane.

  Some or all of the fields of *cfg_ini* may be omitted. The default values are 0 for d_Y, d_Z, β and γ, and the theoretical unitary rotation velocities for perfect rolling for $\omega_{m,u}$ and $\omega_{R,u}$ (cf. Eqs. (7.6) and (7.7)).

  Units: d_Y and d_Z are in mm. $\omega_{m,u}$ and $\omega_{R,u}$ are dimensionless. β and γ are in rad.

- *ivar_slc* (optional): vector with up to 6 integers from 1 to 6, with the indexes of the DOFs of the rolling element with which to iterate in the search for equilibrium. If not all the DOFs are chosen, the equilibrium is checked only for some of the components of the resultant force and moment in the rolling element. The not chosen DOFs are left fixed at their initial values. The coding of the DOFs and correspondence with the resultant force and moment components that are checked for equilibrium is indicated in Table 7.1.

The default behaviour is to iterate with all the rolling element configuration variables / components of residuals.

Table 7.1. Coding of rolling element degrees of freedom defined in *ivar_slc*, and correspondence with components of residuals.

ID	Degree of freedom	{R} component
1	d_Y	F_Y
2	d_Z	F_Z
3	$\omega_{m,u}$	F_X
4	$\omega_{R,u}$	M_Y
5	β	M_X
6	γ	M_Z

- *lim* (optional): structure with lower and upper bounds for each of the DOFs of the rolling element, with fields d_Y, d_Z, $\omega_{m,u}$, $\omega_{R,u}$, β and γ. Each field is a two-element vector, being the first the lower saturation bound and the second the upper saturation bound. The saturations are applied in all the equilibrium iterations. This may aid to avoid divergence if a bad **[J]** matrix has been calculated in some cycle (as a result of changing the set of elements in the contact patch).

  Some or all of the fields of *lim* may be omitted. The default behaviour is not to apply any saturation bound. Only the fields of *lim* corresponding to the chosen DOFs (cf. *ivar_slc* input) are taken into account.

  Units: as in the corresponding fields of *djac*.

- *FILES* (optional). Structure with *in*, *inj*, *out* and *outj* fields, each of them with a bit with which it may be chosen to delete (0) or preserve (1) the CECT input and output files generated during the execution of equil_roller.

  Each field is referred to the following files. In each case, the total number of files involved is given in parenthesis, being *ite* the number of equilibrium iterations until convergence, and considering that there are two contacts involved (inner and outer).

- *in*: main CECT input files corresponding to the rolling element configuration **{P}** (2×*ite*).
- *out*: .mat, .prf and .log CECT output files corresponding to the rolling element configuration **{P}** (2×3×*ite*).
- *inj*: main CECT input files for the calculation of **[J]** (2×n_{pj}×*ite*).
- *outj*: .mat, .prf and .log CECT output files corresponding to the calculation of **[J]** (2×n_{pj}×3×*ite*).

  NOTES for *inj* and *outj*:

  - If *DISCR*.n_{xj} = *DISCR*.nx and *DISCR*.n_{sj} = *DISCR*.ns, n_{pj} is the number of DOFs with which it is iterated in the equilibrium search (cf. *ivar_slc* input). Otherwise, it is this number of DOFs plus one. In the latter case, this is because in addition to solving the contact problems corresponding to the perturbed configurations **{P'$_i$}** for each of the selected DOFs, the contact problems for the unperturbed current configuration **{P}**, already solved on the nx×ns mesh, are solved as well on the mesh with n_{xj}×n_{sj} elements. This is necessary to ensure that the residual variations with which **[J]** is computed contain only the influence of the variations of the rolling element DOFs, and not numerical differences due to differences of the mesh.
  - The structure of the name root of the generated CECT files corresponding to the calculation of **[J]** is as follows: *finp*_*i*_j*n*-*ie*. Being:
    - *finp*: filename of the input file for equil_roller.
    - *i*: number of the equilibrium iteration, padded if necessary with zeros up to 3 digits.
    - *n*: code of the perturbed DOF, according to Table 7.1. *n* = 0 for the unperturbed current configuration (calculated within the calculation

step of **[J]** only if *DISCR*.$n_{xj} \neq$ *DISCR*.nx or *DISCR*.$n_{sj} \neq$ *DISCR*.ns, as explained before).

- *ie*: "int" for inner contact and "ext" for outer contact.

The default behaviour is to preserve all the files.

NOTE: the CECT input and output files of the last equilibrium iteration are always preserved, regardless of the values in the *in* and *out* fields.

7.2.2.2.2 equil_roller output

The following variables are returned by equil_roller:

- *PAR*. Structure with nominal position, geometry, orientation and rotation velocity parameters of the rolling element, with the following fields:
  - *i*, *e*: similar structures for the inner (*i*) and outer (*e*) contacts, each of them with the following fields with relevant CECT input parameters described in §4.1.2.1.1:
    - incl_r: tangent of the inclination angle of the rail (raceway) profile in the *YZ* plane. [-]

      It is fixed at $\tan(\alpha_o - \alpha_R)$ for the inner contact, and $-\tan(\alpha_o + \alpha_R)$ for the outer.
    - r_nom_r: nominal rolling radius of each raceway, cf. Eq. (7.5). [mm]
    - despl_y, despl_z: displacements along the *Y* and *Z* axes with which the nominal contact point is located at the local (0, 0) point of the local system of the raceway profile, which is also the (0, 0) point of the track {*X*, *Y*, *Z*} system used by CECT (the latter being different from the {*X*, *Y*, *Z*} system defined in Figure 7.1). [mm]

      Taking into account the previously explained definition of the raceway and rolling element profiles, and that the track and wheelset half-widths (hw_tr and hw_ws; cf.

§4.1.2.1.1) are set to 0, these displacements are defined as follows. The ± and ∓ signs in Eqs. (7.53) and (7.54) are applied for the inner and outer contacts, in this order.

$$\texttt{despl\_y} = \pm R \sin(\alpha_R) \cos(\alpha_o) + (R\,(1 - \cos(\alpha_R)) - \texttt{r\_nom\_w}) \sin(\alpha_o) \quad (7.53)$$

$$\texttt{despl\_z} = R \sin(\alpha_R) \sin(\alpha_o) - \texttt{r\_nom\_w} \mp (R\,(1 - \cos(\alpha_R)) - \texttt{r\_nom\_w}) \cos(\alpha_o) \quad (7.54)$$

- `phi`: roll angle of the rolling element. Equal to α_o for the inner contact, and to $\text{sgn}(\alpha_o)\,(\pi - |\alpha_o|)$ for the outer. [rad]
- `vel`: forward velocity of the centre of the rolling circle of the rolling element corresponding to the $R_{p,o}$ pitch radius, in the relative movement with respect to each ring. [m/s]

  It corresponds to the `sftx_o_vel` CECT input.

  It is equal to $R_{p,o}\,|\omega_m - \omega_I|$ for the inner contact, and to $R_{p,o}\,|\omega_m - \omega_E|$ for the outer.
- `omega`, `dir_om`: modulus and cosine directors (in the global coordinate system) of the total rotation velocity of the rolling element with respect to each ring. See Eqs. (7.8)–(7.10) (note that the signs of the components in X and Y directions given in those equations have to be reversed for the outer contact, due to the reversal of the axes X and Y axes described at the beginning of §7.2.2). `dir_om` is a 3-element vector with the `dir_omx`, `dir_omy` and `dir_omz` CECT inputs.

  Units: `omega` in rad/s, `dir_om` dimensionless.

  - o `r_nom_w`: rolling radius of the rolling element at the (y_L, z_L) = (0, 0) point of its profile. It is equal to D / 2 + R (1 – cos(α_R)). [mm]
  - o $\omega_{m,o}$: theoretical orbital rotation velocity of the rolling element for perfect rolling (cf. Eq. (7.6)). [rad/s]
  - o $\omega_{R,o}$: theoretical pitch rotation velocity of the rolling element for perfect rolling (cf. Eq. (7.7)). [rad/s]

- *eq*. Structure with the equilibrium DOFs of the rolling element, with the following fields: d_Y [mm], d_Z [mm], $\omega_{m,u}$ [-],$\omega_{R,u}$ [-], β [rad] and γ [rad]. Each of these fields is a vector with *NUM.maxite* elements.
- *rs*. Structure with the *F* and *M* fields, with the residual force (*F*) and moment (*M*) vectors in each equilibrium iteration. Each of these fields is a matrix with *NUM.maxite* × 3 elements.

  Units: *F* in kN, and *M* in N.m.

In the case of the *eq* and *rs* output structures, row *i* of each field contains the corresponding data in equilibrium iteration *i*. Rows for iterations higher than the total number of needed iterations to achieve equilibrium contain zeros.

In addition, besides the CECT files, a .log file is printed, named *finp*–eq_ro.log as mentioned at the beginning of §7.2.2.2.1. This file includes information about the equilibrium iterations and the intermediate solutions obtained. The information included is listed next.

- Header of the file, including name of the equil_roller input file, version of the program and date of execution.
- If the equil_roller input file is not found, an error message is printed and execution is stopped.
- Identifiers of the DOFs of the rolling element configuration with which it is iterated in the search for equilibrium, together with the objective components of force and moment residuals (cf. *NUM.ivar_slc* input in §7.2.2.2.1).
- For each equilibrium iteration, the following information is included:

- Current value of each DOF of the rolling element configuration.
- Writing of the CECT input file for each contact, including the name of each file.
- End of calculation of each contact with CECT, including the elapsed time after each CECT execution.
- Force and moment residual vectors, together with the objective value for each component.
- Progress of the contact solutions related to the calculation of the Jacobian, including the name of the written CECT input files, and elapsed times after each CECT execution.
- Jacobian matrix, including headers in each of its rows and columns to identify the related residual component and DOF.
- Increments of the DOFs of the rolling element configuration for the next iteration (including only the DOFs with which it is iterated). If, after the application of these increments, the upper or lower bound for each DOF defined with the *NUM.lim* input are exceeded, the required saturations are applied, and a warning message is printed for each saturated DOF.

- Outcome of the equilibrium search process. It is a message indicating whether equilibrium in the rolling element has been achieved or not (within the specified tolerances). For the case in which not all of the components of the force and moment residuals are checked for equilibrium according to the *NUM.ivar_slc* input, it is also indicated if the not checked components have converged to the specified tolerances.
- Summary of results of each iteration. Two lists are printed with results of each equilibrium iteration. The first one includes the six DOFs defining the position **{P}** of the rolling element, and the second one the six components of force and moment residuals.

Regarding the writing of the different files, the following applies:

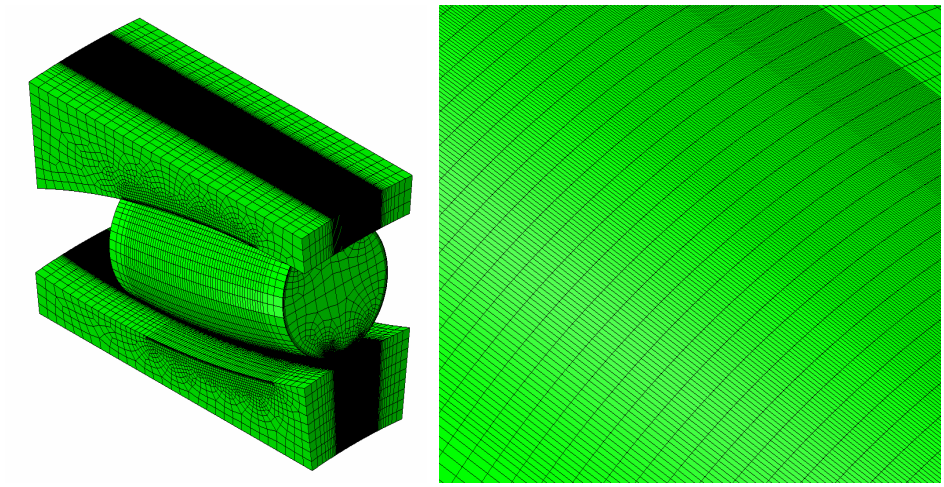
- The structure of the roots of the names of the written files are *finp_i-ie* for the CECT files related to the main contact problems, and *finp_i*_j*n-*

ie for the CECT files related to the Jacobian calculation, as previously explained.

- If a CECT input file with the same name as a new file intended to be created exists in the current working directory (CWD), it is attempted to rename the existing file, together with each associated output file (.mat, .prf and .log) if they exist, appending the current date to its name. If a file with this modified name also exists in the CWD, an error message is printed and equil_roller execution is stopped. Otherwise, the existing file is renamed, and a warning message is printed. The behaviour is different for the *finp*–eq_ro.log file: if a file with this name exists, the name of the newly created .log file is modified, appending the current date to its name. If some file with this modified name also exists, it is overwritten.

7.2.3 Finite element models

Each FE model for the calculation of the rolling element–raceway contacts comprises one rolling element and a portion of the inner and outer rings in contact with it. A general view of one of the models built for the SRB studied in §7.3 is shown in Figure 7.6a. This model has been built in Abaqus/Standard [Abaqus doc 2012]. The contact regions of the roller and the raceways are finely meshed with solid linear hexahedron C3D8 elements. Figure 7.6b shows a detail of the mesh of the roller surface in the most finely meshed region. For the model shown in particular, the minimum element size is approx. 0.024×0.37×0.06 mm (in longitudinal, lateral and depth directions, respectively), and the model has a total of about 740000 nodes. Other models have been built with coarser and with finer meshes, observing that with the quoted element size accurate contact surface results are obtained. Frictional contact pairs are defined between the rolling surface of the roller and the inner and outer raceways. A surface-to-surface contact detection method is used, and the contact constraints are enforced with the penalty method, both for the normal and tangential parts, as in the wheel–rail contact FE models described in §4.2.1.

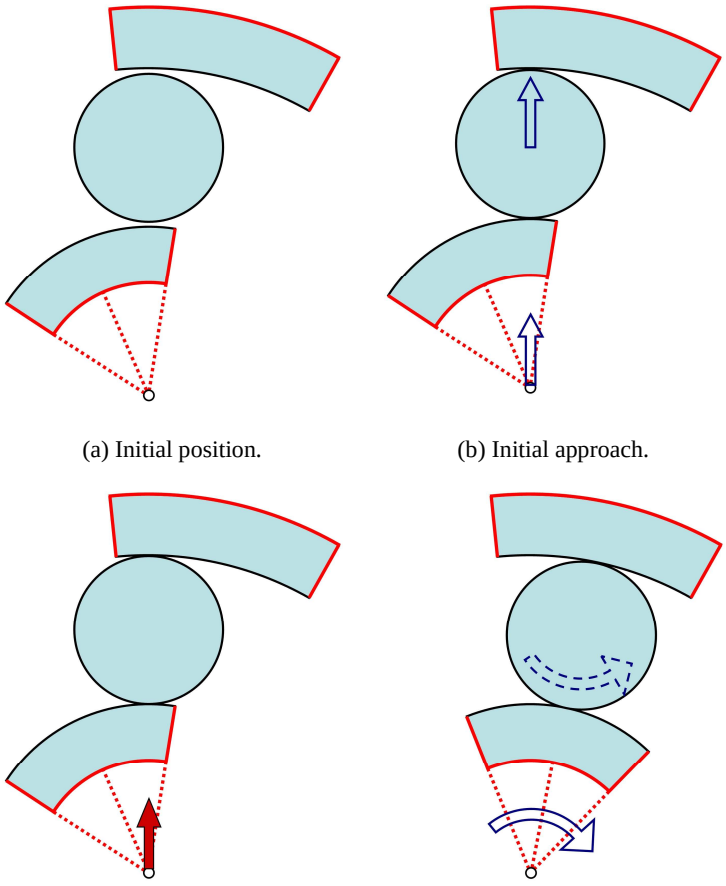


(a) General view. (b) Mesh detail.

Figure 7.6. 3D FE model for contact analysis of rolling element–raceway contacts in rolling bearings. The model shown corresponds to the SRB studied in §7.3.

The rolling contact between the roller and the inner and outer rings is simulated in a Lagrangian, multi-step static simulation. A central reference node is defined for each ring, through which their motion is prescribed. The main phases of the FEM simulation are the initial positioning and normal loading of the roller, and the rolling stage. These are illustrated in Figure 7.7. The figure shows a schematic view of the two rings and the rolling element perpendicular to the bearing Y axis during the different simulation phases. The central node of the outer ring is fully constrained during the whole simulation.

All the bodies are initially positioned in such a way that the inner and outer contacts fall into the most finely meshed regions of the contact surfaces at the end of the simulation. The axial position of the roller also has to be adjusted at the beginning of the simulation, as explained in §7.3.3.



(a) Initial position. (b) Initial approach.

(c) Normal load. (d) Rolling.

Figure 7.7. Main phases of the FEM rolling contact simulations of a rolling bearing.

The normal loading phase is split into two steps, depicted in Figure 7.7b and c. In the first of these steps, small radial displacements are imposed on the roller and inner ring to achieve the initial closure of the inner and outer contacts. This is indicated by the blue vertical arrows on the central nodes of the inner ring and the roller. This displacement-controlled step is necessary because before achieving the initial contact closure the system is singular if the rigid body motion of the bodies is not constrained. In the second step of the normal loading phase (Figure 7.7c), the radial displacement constraints of the roller and the inner ring are released, and the desired radial load is imposed on the central node of the inner ring. This is indicated by the red vertical arrow on this node.

At the end of this step, the roller is in equilibrium under the inner and outer contact forces and the inertia forces. The latter are applied through the central node of the roller as prescribed concentrated forces and moments. The resulting normal approach of the central node of the inner ring towards the outer ring at the end of this phase is held constant during the subsequent rolling stage.

During the rolling stage (Figure 7.7d), small axial rotation increments are imposed on the central node of the inner ring. Thus, the motions of the central nodes of the rings are entirely prescribed, while the movement of the roller is driven by the contacts with the inner and outer raceways. The rolling phase of the simulation is also split into several steps. The applied axial rotation increments to the inner ring are adjusted according to the mesh density of the surface sections traversed by the contact patches in each of these steps, for the displacement increments of the contact patches in each substep to be similar to the mesh element size in the longitudinal direction, as in the developed wheel–rail contact FE models (cf. §4.2.2.2).

A quasi-stationary rolling contact state is established at the contact level within rolling distances of the order of a few times the longitudinal contact patch dimension (1–2 times as shown by experience with the exact contact theory, as quoted in [Kalker 1979-b]). On the other hand, in the simulations carried out, it has been observed that larger rolling distances are required for the roller to reach a steady configuration, as is shown in §7.3.3. For the SRB studied in §7.3, simulations with inner ring rotated angles of up to 170 mrad have been carried out in this work. This corresponds to a roller–raceway rolled distance of about 7.6 mm, or about 17 times the total longitudinal dimension of the outer contact patch with a normal load of 5.5 kN between the roller and each raceway. In contrast, much lower rolled distances have been seen to be necessary in other applications of detailed FEM for rolling contact analyses in rolling bearings. E.g., a rolled distance of one time the longitudinal contact patch dimension was applied in [Heras 2017b], observing that this was sufficient for the friction torque to stabilize.

Taking into account the large size of the model, which includes two finely meshed contact regions, and the long rolling distances, the mesh construction and simulation setup have to be properly optimized to get affordable FEM simulations. With this aim, a variable mesh density has been used in the rolling

direction, as may be appreciated in Figure 7.6b, providing the finer mesh in the final part of the simulated rolling path in each contact surface. Compared to wheel–rail rolling contact FEM simulations shown in Chapters 4 and 6, these roller–raceway rolling contact simulations have presented greater difficulty. The fact of having the roller driven just by the contacts with the rings rather than with prescribed motion, as well as the increased model size, contribute to the added complexity.

7.3 Case study: spherical roller bearing

The partial slip contact models described in §7.2 have been used to compute roller–raceway frictional contact solutions for a SRB. The considered SRB is described in §7.3.1. The performance of the considered contact models is assessed in §7.3.2. In §7.3.3 the quality of the contact solutions obtained with FEM is discussed, examining the evolution of relevant output quantities along the simulations. Lastly, §7.3.4 presents a study of the steady rolling equilibrium attitudes of the roller in a range of normal loads and friction levels.

7.3.1 Description of the studied spherical roller bearing

A double-row SRB with symmetrical barrel-shaped rollers is considered. This is a type of bearing commonly used in railway axles. These bearings have self-aligning and high load-carrying capacity, and can carry combined radial and thrust loads [Harris 2001], [Kleckner 1982]. Figure 7.8 shows a sectional view of the bearing (not to scale), and Table 7.2 lists its main internal dimensions relevant to the contact analyses. One-half of the bearing section is shown in the figure, which is symmetrical with respect to the *XZ* plane. The bearing material is steel, with a Young's modulus E of 207 GPa and a coefficient of Poisson ν of 0.30.

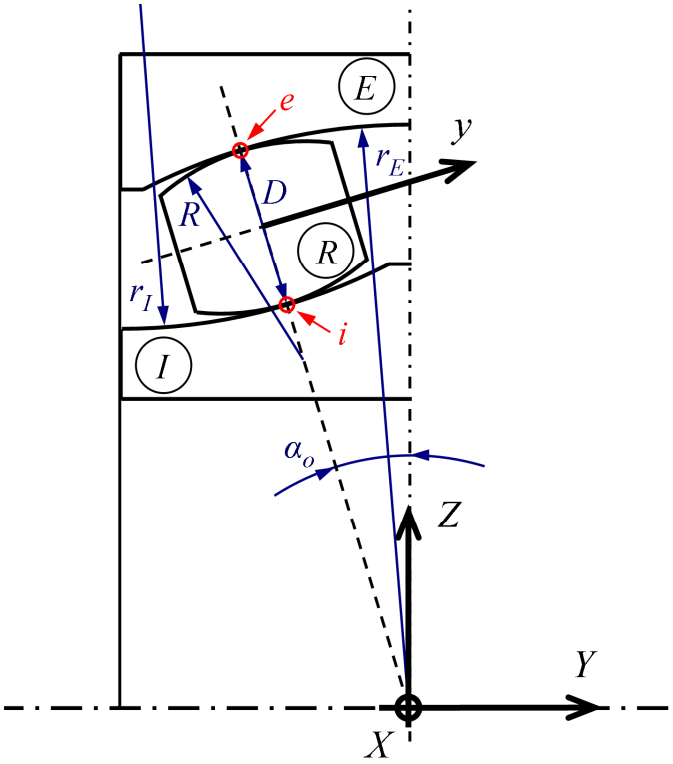


Figure 7.8. Sectional view of the studied spherical roller bearing.

Table 7.2. Geometric parameters of the studied spherical roller bearing.

Parameter	Description	Value
D [mm]	Roller nominal diameter	19.8
r_E [mm]	Outer race groove curvature radius	102.4
r_I [mm]	Inner race groove curvature radius	103.0
R [mm]	Roller crown radius	100.3
α_o [º]	Nominal contact angle	11.5
α_R [º]	Roller angle	0.0

The bearing is considered to be radially loaded. The resulting roller–raceway contact patches have a high width-to-length ratio (approx. 40) and modest total contact angle variations in the lateral direction, $\Delta\delta$. The length refers to the dimension in the rolling direction and the width to the dimension in the lateral direction. $\Delta\delta \approx 11^{\circ}$ with a normal load N on the roller of 5.5 kN. This load has been computed in the most loaded roller for a total radial load of 28.7 kN in the bearing, considering a radial bearing clearance of 0.1 mm.

The rotation velocities of the inner and outer rings of the bearing (around the global Y axis) are ω_I = 476.2 rpm and ω_E = 0. With these velocities and the considered bearing geometry (radial clearance included), the theoretical orbital and pitch rotation velocities of the roller for perfect rolling on the nominal contact points are $\omega_{m,o}$ = 22.25 rad/s and $\omega_{R,o}$ = 225.5 rad/s. Without radial clearance[31], these velocities change to $\omega_{m,o}$ = 22.27 rad/s and $\omega_{R,o}$ = 225.7 rad/s.

Each roller as a mass m_R of 65.6 g and inertia moment about the u axis $I_{R,uu}$ of 5.87 kg·mm^2. Together with the kinematic parameters considered, this results in a centrifugal force and a gyroscopic moment of approx. 3 N (along the positive Z axis) and 6 N.mm (along the positive X axis for the roller depicted in Figure 7.8), respectively. In the cases analyzed in this chapter, these roller inertia forces are much lower in magnitude than the roller–raceway contact forces, and their effect on the contact solutions is negligible. Apart from these inertia forces, no other additional forces are considered acting on the roller in this study.

7.3.2 Comparison of strip theory, CECT and FEM

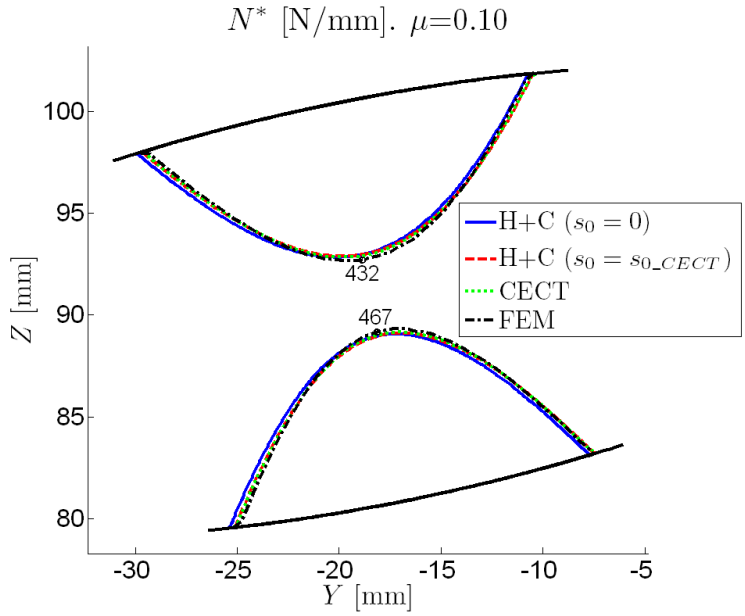
In this section, the roller–raceway contact solutions obtained with the contact models described in §7.2 are compared, to assess the ability of each model to produce representative results. For this purpose, the steady rolling positions of the roller with a normal load N of 5.5 kN and two different values of the coefficient of friction μ, namely μ = 0.10 and 0.30, are considered. Besides comparing the results of the different models, the main features of the contact solutions are illustrated for different friction levels. With μ = 0.10 most of the contact patch is sliding. With the higher μ, a bigger adhesion zone is obtained, and the elasticity of the bodies plays a more important role in the tangential

[31] Once the normal load in the considered roller is fixed, the radial clearance has a negligible influence on the roller local contact solution. However, it has to be taken into account in the calculation of the theoretical rotation velocities, for proper comparisons between equilibrium rotation velocities obtained with different calculation methods, if different clearances are considered. The results presented in this chapter have been obtained considering a radial clearance of 0.1 mm only in the case of the FEM calculations. In the rest of the calculations, no radial clearance has been considered.

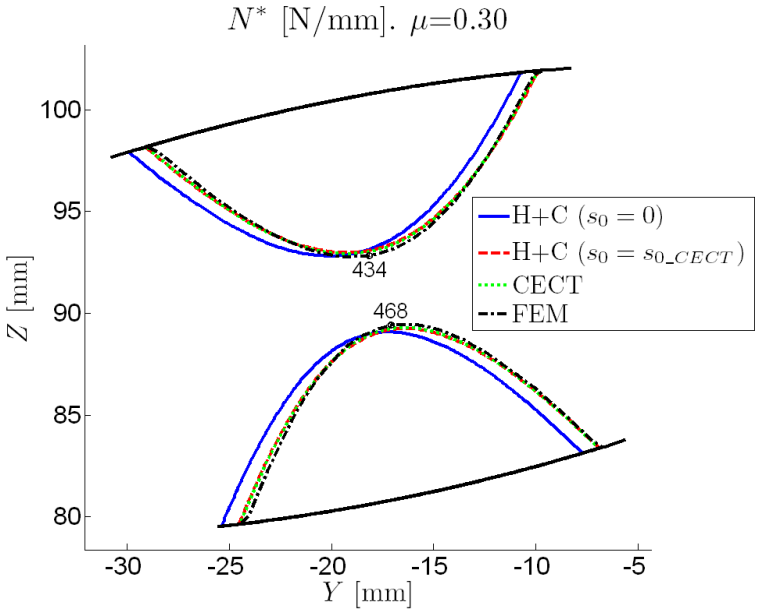
contact. A partial slip contact model may provide a greater precision improvement in the contact solutions in the latter case, with respect to a full slip contact model.

Two different solutions computed with strip theory are included in each case, labelled as "H+C" (Hertz + Carter). The difference between both solutions is the position of the roller–raceway contact centres. In one of them, each roller–raceway contact centre is located in its nominal position. This is identified with "$s_0 = 0$". In the other, identified with "$s_0 = s_{0\_CECT}$", the roller–raceway contact centres are offset from their nominal positions in the contact lateral (s) direction. The same offset is applied for the inner and outer contacts, computed as the average of the offsets for each contact of the corresponding equilibrium solution computed with CECT.

Figure 7.9 and Figure 7.10 show the lateral distributions of normal load and longitudinal tractions per unit contact width, N^* and F_x^*, across the inner and outer contact patches. The distributions are plotted on the profiles of the roller–raceway potential contact surfaces, on the bearing YZ plane. The contact patches appear offset from their nominal position towards the positive y direction, as can be seen comparing the CECT and FEM solutions with the H+C ($s_0 = 0$) solution. This offset is higher for higher μ. The reason for this offset is explained in §7.3.4. The FEM solutions give a slightly more concentrated load distribution in the lateral direction, as seen in Figure 7.9. Apart from this, there are no remarkable differences in the normal contact solutions obtained with H+C ($s_0 = s_{0\_CECT}$), CECT and FEM.

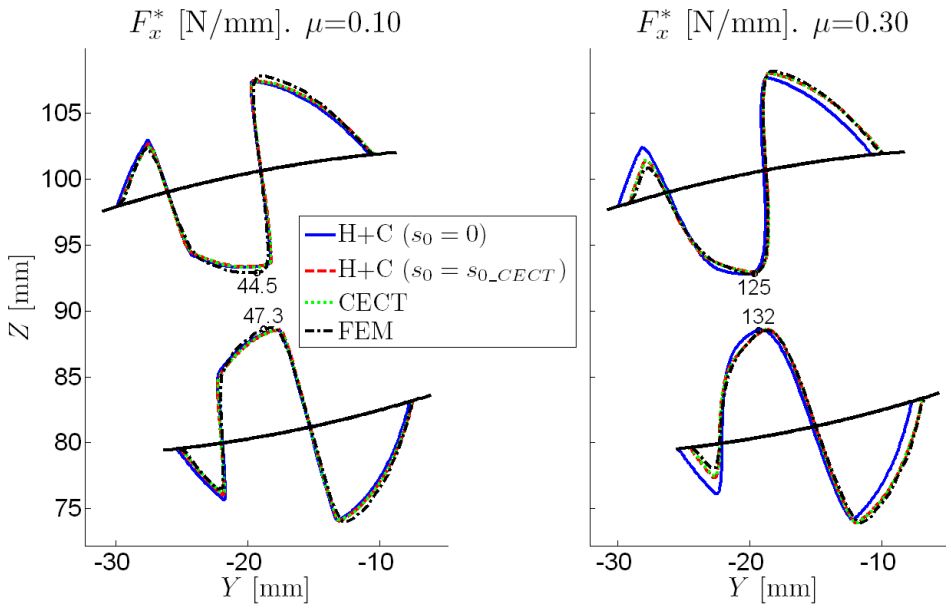


(a) μ = 0.10.



(b) μ = 0.30.

Figure 7.9. Lateral distribution of normal load in roller–raceway contacts with N = 5.5 kN.

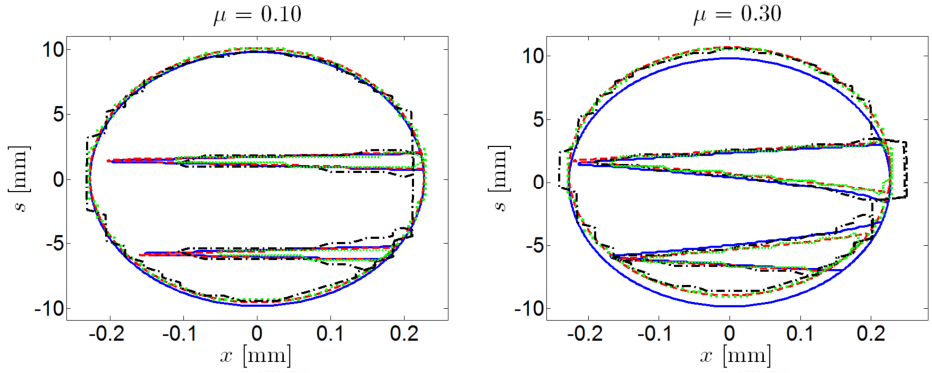


(a) μ = 0.10. (b) μ = 0.30.

Figure 7.10. Lateral distribution of traction in roller–raceway contacts with N = 5.5 kN.

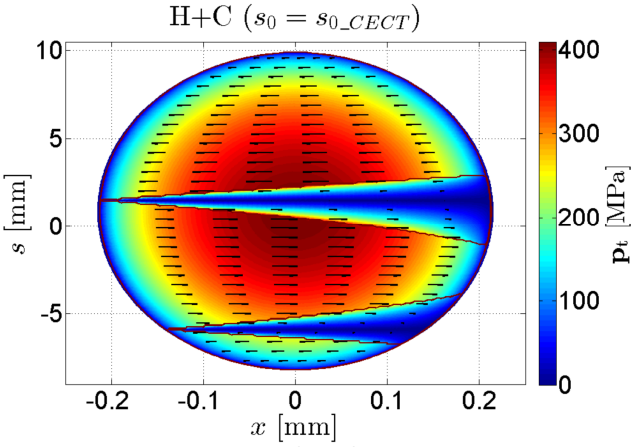
Figure 7.11 and Figure 7.12 show the contact patch contours and the tangential stresses. As is characteristic with crowned rollers, the tangential contact solutions are dominated by high longitudinal creepages due to the variable rolling radii of the roller across the contact with the raceways, with two zero-slip bands in which they change sign. This may also be seen in the longitudinal traction distributions of Figure 7.10. A good match is observed between the different solutions in the tangential contact as well. On the other hand, mostly with high μ notable differences are seen between the H+C solutions obtained with and without adjusted contact centre positions. These differences may be clearly appreciated in the lateral distributions of frictional work shown in Figure 7.13. This shows that not making the lateral contact position adjustment may lead to some inaccuracy in the tangential contact solution. While quite accurate results may be obtained with H+C, it is not possible to work out the necessary equilibrium condition in the lateral direction to perform this position adjustment taking into account only the longitudinal tangential stresses in the calculation. The differences between both H+C solutions are higher with higher μ because

the roller shift between its equilibrium and nominal position increases for higher μ, as explained in §7.3.4.

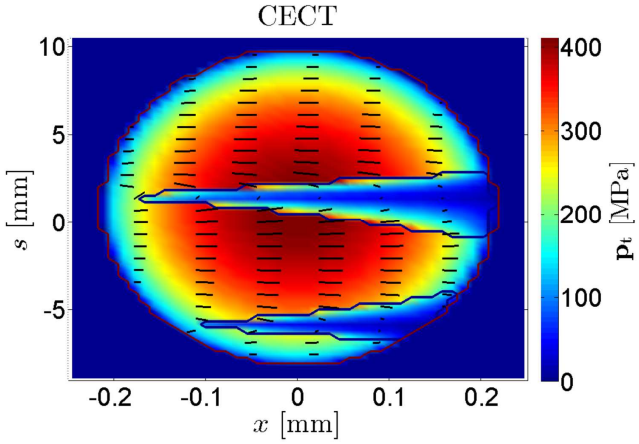


(a) μ = 0.10. (b) μ = 0.30.

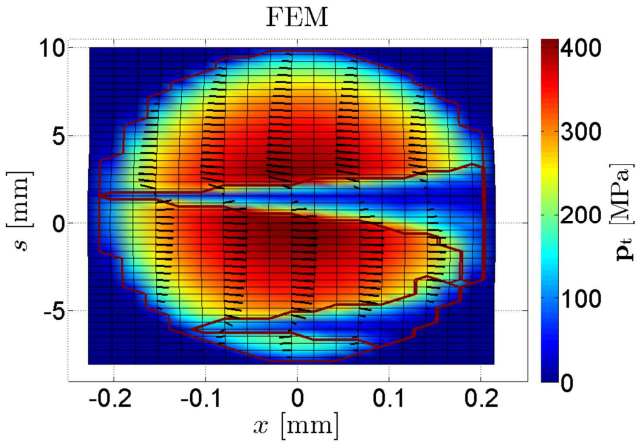
Figure 7.11. Contours of adhesion and slip zones in the outer roller–raceway contact patch with N = 5.5 kN. Solid blue: H+C (s_0 = 0); dashed red: H+C (s_0 = s_{0_CECT}); dotted green: CECT; chain black: FEM.



(a) H+C (s_0 = s_{0_CECT}).

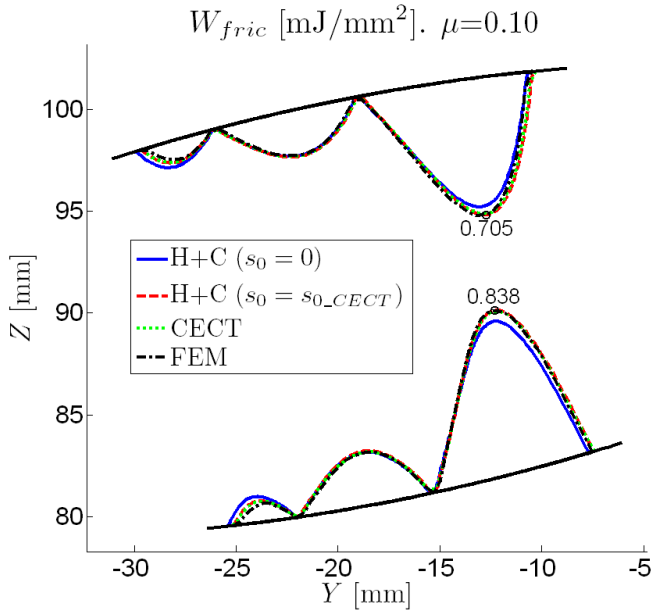


(b) CECT.

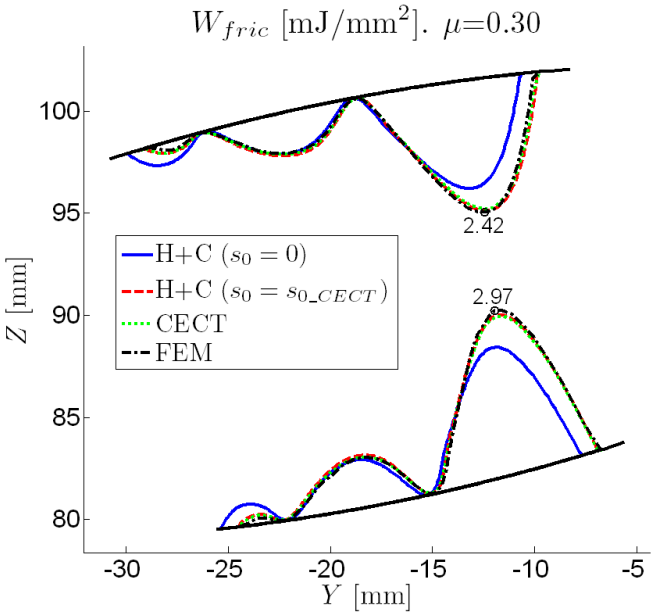


(c) FEM.

Figure 7.12. Magnitude and direction of tangential stresses and contours of adhesion and slip zones in the inner roller–raceway contact patch with N = 5.5 kN and μ = 0.30.



(a) μ = 0.10.



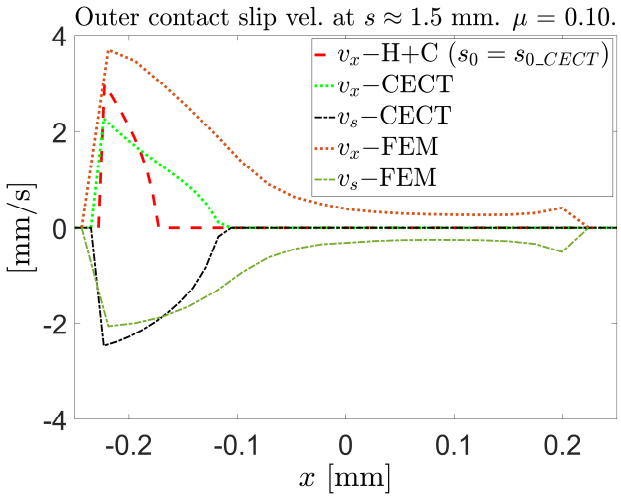
(b) μ = 0.30.

Figure 7.13. Lateral distribution of frictional work in the inner and outer races after one roller passage with N = 5.5 kN.

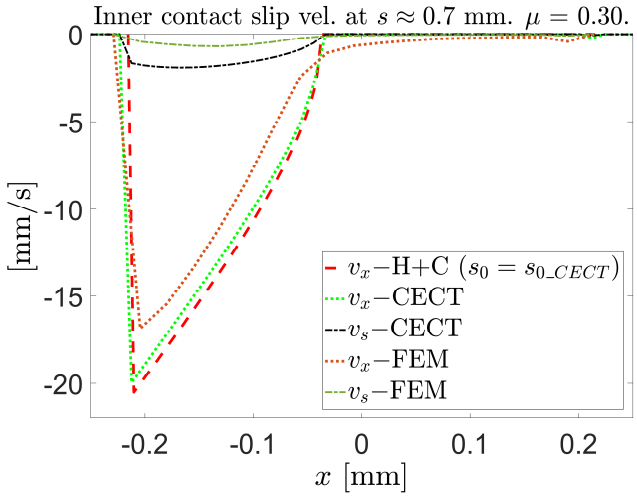
The slip velocities and frictional work levels are higher towards the bearing *XZ* centre plane, as can be seen in Figure 7.13. Here, the longitudinal creepage due to the rolling radii difference of the roller surface with the rolling cone is superposed with the spin creepage. The lateral contact shift is produced towards this side, and consequently, the maximum frictional power densities in the contact patch increase.

Due to the short longitudinal dimension of the contact patches, the slip velocities are mostly aligned in the rolling direction, except around the zero-slip bands. Figure 7.14 shows longitudinal distributions of slip velocities along slices close to one of these regions. In the case with low μ, notable differences are observed between the H+C, CECT and FEM results around the zero-slip bands, where the tangential stresses change from their positive to their negative saturation bounds in a relatively small distance in the lateral direction. In this region, significant lateral components of stress and slip velocities appear, as seen in Figure 7.14a, and the 2D simplification is inadequate. Part of the differences between the CECT and FEM results seen in Figure 7.14a may be attributed to the fact that the lateral positions in which the results are taken in each model are not exactly the same. The results obtained with CECT and FEM indicate higher levels of tangential stresses and slip than predicted with the strip theory neglecting the interaction between adjacent strips. Nevertheless, these regions occupy a small part of the whole contact patch.

In the case with high μ, the changes of ξ^* are slower in the lateral direction, which implies lower interaction between adjacent strips. Consequently, the lateral components of stress and slip velocities are comparatively smaller, and a better agreement is found between the tangential stresses (not shown in the figure) and slip velocities obtained with H+C, CECT and FEM, as shown in Figure 7.14b. Figure 7.15 shows the distributions of longitudinal rigid slip in the outer contact for μ = 0.10 and 0.30, both in terms of longitudinal rigid slip velocities w_x and normalized longitudinal creepages ξ^*. The graphs for the inner contact are qualitatively similar. While the profiles of w_x are similar for both values of μ, the profiles of ξ^* are different mainly due to the different μ. Here it is also shown for each μ how ξ^* changes as a result of the lateral shift of the contact, mainly for high μ where the shift is higher.

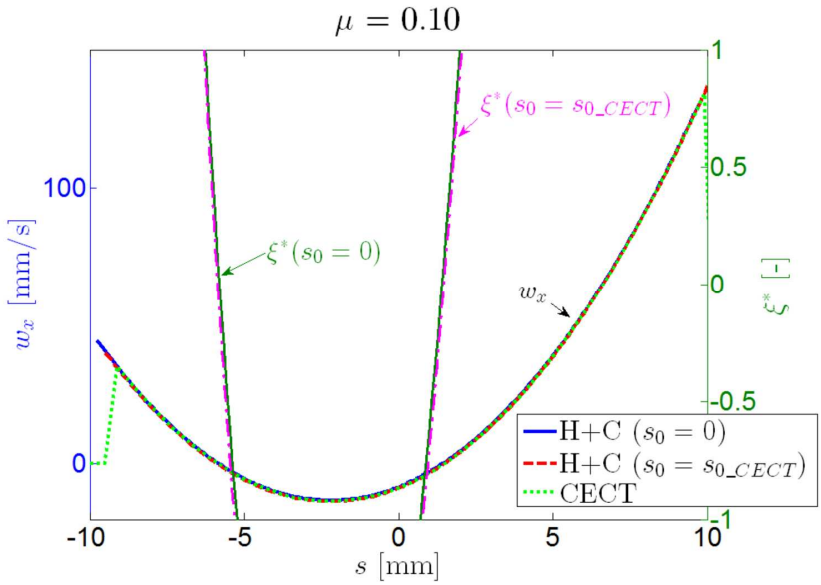


(a) μ = 0.10. Outer contact —at s = 1.54 mm in H+C results, 1.52 mm in CECT results and 1.60 mm in FEM results.

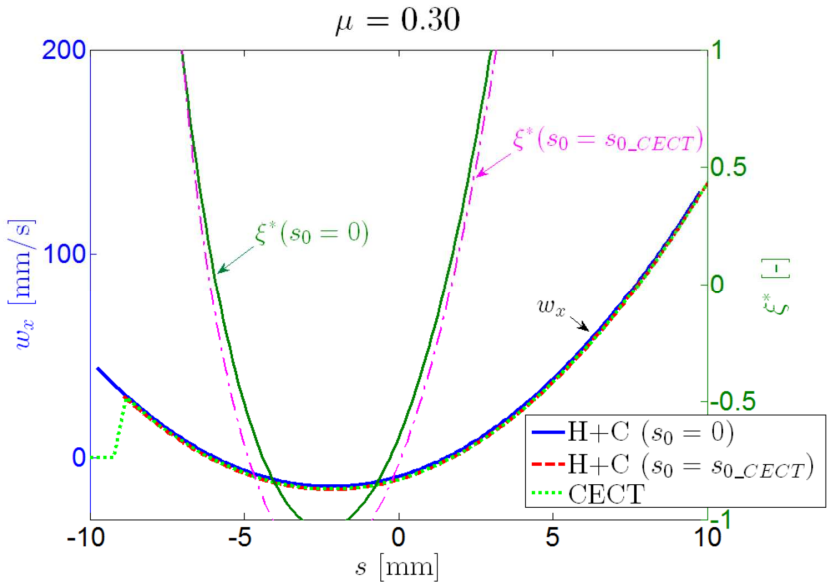


(b) μ = 0.30. Inner contact —at s = 0.68 mm in H+C and CECT results and 0.80 mm in FEM results.

Figure 7.14. Longitudinal distribution of slip velocities at lateral positions of the roller–raceway contacts partly in adhesion. N = 5.5 kN.



(a) μ = 0.10.



(b) μ = 0.30.

Figure 7.15. Lateral distribution of longitudinal rigid slip in the outer roller–raceway contact with N = 5.5 kN.

7.3.3 Assessment of stationarity of FEM solutions

Figure 7.16 shows the evolution along a FEM simulation of different parameters that are checked to determine if the obtained FEM results correspond to stabilized equilibrium positions. The results shown correspond to the case of §7.3.2 with $\mu = 0.30$. Similar trends are observed for the case with $\mu = 0.10$, with some of the parameters stabilizing earlier.

Figure 7.16a shows the resultant tangential forces in each contact. The lateral tangential forces become quite stabilized within the first half of these simulations after a rolled distance of about eight times the longitudinal dimension of the contact patch (a rolled distance equal to the longitudinal dimension of the outer contact patch corresponds to a rotated angle of the inner ring θ_I of about ten mrad, cf. §7.2.3. The longitudinal forces reach stabilization earlier.

Figure 7.16b shows the lateral locations of the centres of the normal pressures of the inner and outer contacts, $Y_{c,i}$ and $Y_{c,e}$. Their evolution is determined primarily by roller tilting in this simulation. The roller also has a slight lateral motion, also shown in these figures in non-dimensional form η as the axial velocity of the roller centre $v_{R,y}$ divided by the rolling velocity along the outer contact centreline. This does not represent exactly the lateral creepage in the roller–raceway contact, as it does not include the part due to roller skew.

The lateral motion of the roller causes variations in the contact lateral resultant forces enough to comply quasistatic equilibrium of the roller in significantly different lateral positions. The lateral position of the roller influences the lateral shift of the contact patches, which impacts on the tangential solution, as shown in §7.3.2. The actual steady equilibrium position of the roller is the one with zero η. Reaching this position from an arbitrary initial position would require impractically long rolling distances for this FE model. In this work, the correct equilibrium lateral position of the roller is found iterating with different initial lateral positions. In this way, plots of η versus roller lateral position, such as the ones shown in Figure 7.17, are produced. In this figure, d_y designates the axial offset of the roller centre from its nominal position along the y axis. The changes of d_y are on the order of just one micron during each of these simulations. This shows that an accurate analysis of the motions of the roller is

necessary to interpret the FEM results correctly and to obtain the steady rolling solution. Given the slow lateral movement of the roller in lateral positions different from the one corresponding to steady equilibrium, the roller may not reach its steady state equilibrium position during its whole orbiting motion around the bearing in some cases.

The trends of roller tilt and skew shown in Figure 7.16c indicate that the tilt is stabilized during the simulation, but not the skew. Nevertheless, the skew angle variations are not very high, and the values obtained in the final phase of this simulation are close to the value computed with CECT for steady equilibrium, with a difference of just about 0.2 mrad. The differences in the tilt angles obtained with CECT and FEM are of the same order. It is considered that these angles are small enough to be able to take the FEM solution as a representative steady equilibrium solution.

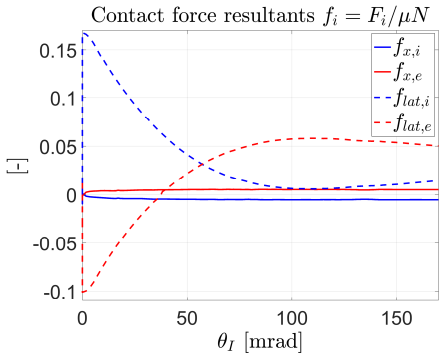
Calculations with CECT indicate that the individual contact lateral resultant forces in the inner and outer contacts are relatively sensitive to the tilt and skew angles. Higher negative tilt and skew angles lead to higher lateral force in the outer contact and lower in the inner contact, and higher negative skew also leads to slightly higher total lateral force. Indeed, the individual lateral contact forces computed with CECT and FEM do not match well. But other results, such as the total lateral resultant force on the roller from the inner and outer contacts, the roller axial displacement d_y, and the resultant longitudinal forces, are not so sensitive to these angles, and a relatively good agreement is found between the CECT and FEM results, as shown in Table 7.3 for the cases of §7.3.2. The table includes the longitudinal and lateral frictional forces in each contact, F_x and F_{lat}, and the sum of the lateral forces in the inner and outer contacts $\sum F_{lat}$. All the forces are divided by μN.

Table 7.3. Frictional contact resultant forces on the roller and roller axial offset computed with CECT and FEM for N = 5.5 kN and μ = 0.10 and 0.30.

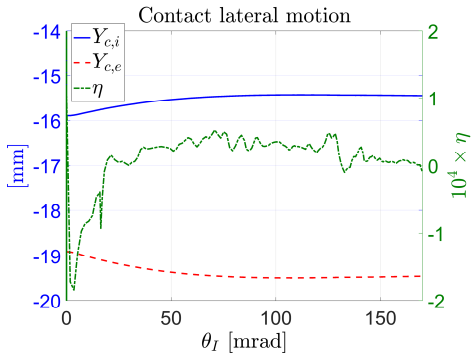
	μ = 0.10			μ = 0.30		
	Full slip	**CECT**	**FEM**	**Full slip**	**CECT**	**FEM**
$f_{x,i}$ [-]	4.36E–03	–4.96E–03	–4.8E–03	2.58E–02	–5.33E–03	–5.2E–03
$f_{x,e}$ [-]	7.31E–03	4.99E–03	4.8E–03	–1.28E–02	5.33E–03	5.2E–03
$f_{lat,i}$ [-]	–5.20E–03	2.23E–02	5.5E–04	–1.08E–02	2.34E–02	1.3E–02
$f_{lat,e}$ [-]	5.63E–03	2.87E–02	5.7E–02	1.20E–02	3.22E–02	5.2E–02
$\sum f_{lat}$ [-]	4.54E–04	5.09E–02	5.8E–02	1.17E–03	5.55E–02	6.5E–02
d_y [μm]	-	6.3	7.2	-	20.4	23.5

Apart from the CECT and FEM results, Table 7.3 also shows the results of a full slip solution for each friction level. These full slip solutions have been derived from the CECT equilibrium solution for each case, considering the corresponding normal pressure distributions and rigid slip velocities computed with CECT, as follows. At each point of the contact patch, the magnitude of the tangential stress is defined as the normal pressure times the coefficient of friction, and its direction is determined with the direction of the rigid slip. The resultant contact forces are then obtained as the vector sum of the contributions of each element of the mesh of the contact surface, as in the other models. The full slip solutions obtained in this way are not equilibrium solutions; these would have to be computed for other roller (equilibrium) configurations through a procedure such as that shown in Figure 7.5. It is observed that the lateral forces, and mostly the values of their sum $\sum f_{lat}$ for the full slip solutions, are much lower than the corresponding values obtained with CECT and with FEM. This was to be expected, as the lateral force due to spin (the camber thrust) is absent in symmetric contact patches with the full slip model. Consequently, the equilibrium configurations obtained with the full slip solutions would surely have much lower d_y displacements than the ones computed with CECT and FEM, cf. §7.3.4.

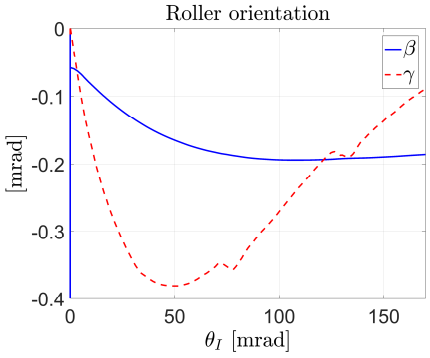
Figure 7.16d shows the trends of the rotation velocities ω_m and ω_R of the roller relative to their theoretical values corresponding to perfect rolling on the nominal contact points. It may be seen that the roller rotation velocities are also stabilized by the end of these simulations. The perturbations seen in the trends of these velocities, η and γ at about 75 and 125 mrad of inner ring rotated angle (θ_I) correspond to contacts changing to zones of different mesh densities in the rolling direction, cf. Figure 7.6b.



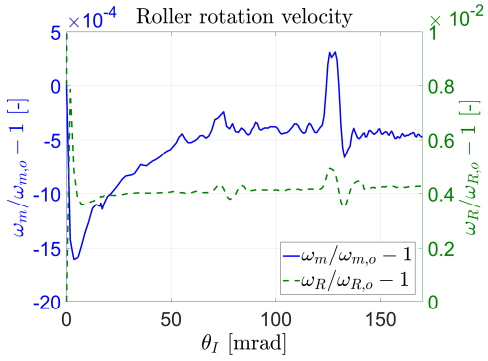
(a) Frictional contact resultant forces $f_i = F_i / \mu N$.



(b) Axial positions in bearing of centres of contact normal pressures, and lateral displacement of roller.



(c) Roller orientation (tilt β and skew γ).



(d) Roller rotation velocities ω_m and ω_R.

Figure 7.16. Indicators of stationarity of FE rolling contact solution for N = 5.5 kN and μ = 0.30 along the rolling stage of simulation.

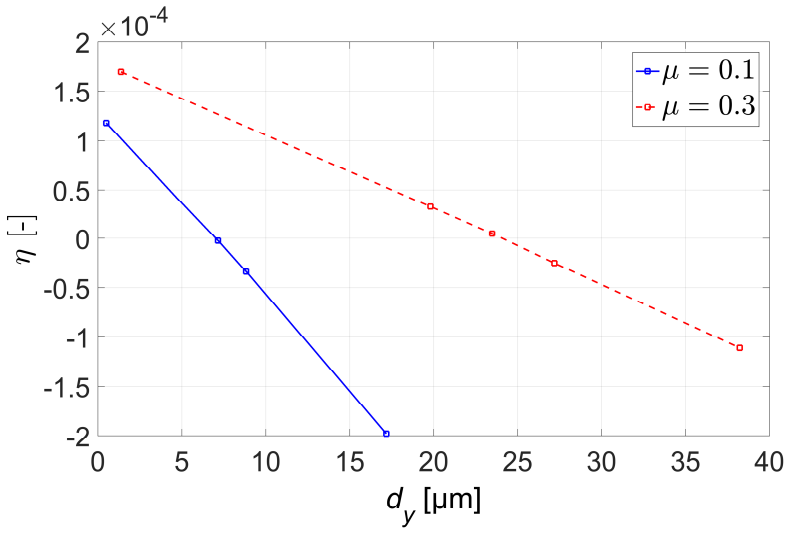


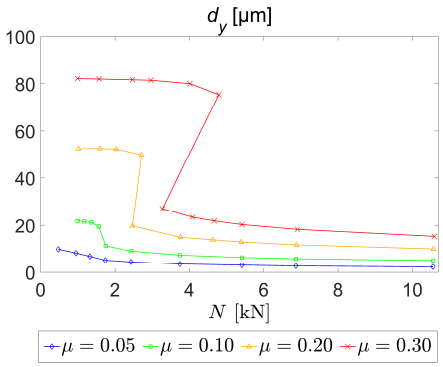
Figure 7.17. Non-dimensional axial velocity of roller η as a function of its initial lateral position in FEM rolling contact simulations.

7.3.4 Equilibrium solutions for different operating conditions

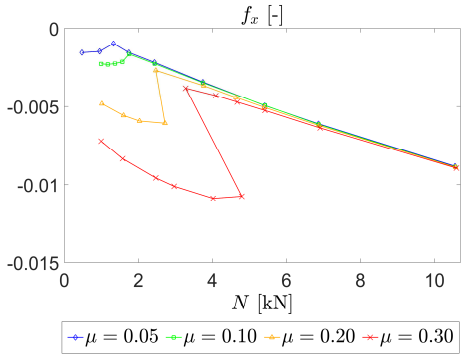
Steady rolling equilibrium configurations of the roller have been computed with CECT and equil_roller (cf. §7.2.2.2) for a range of normal loads and friction levels. The results are summarized in Figure 7.18. Two distinct equilibrium configurations of the roller are observed. There is a first configuration, which is closer to the nominal roller position, feasible at higher loads. This will be identified here as configuration 1. The results shown in previous sections correspond to this configuration. On the other hand, there is a second configuration, feasible at lower loads. This will be identified here as configuration 2. With higher coefficients of friction μ, there is a range of intermediate loads in which both configurations are feasible.

The second configuration is farther from the nominal roller position, with higher axial offsets (Figure 7.18a), tilt and skew angles of the roller. This configuration also involves higher rolling torques and associated resultant longitudinal forces in each contact (Figure 7.18b) and higher frictional powers (Figure 7.18c). The transition loads from configuration 1 to configuration 2

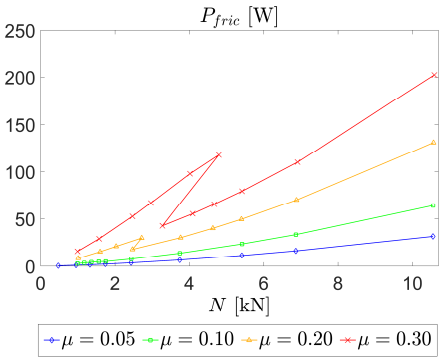
increase as μ increases. Also, the leap between both configurations is higher for higher μ.



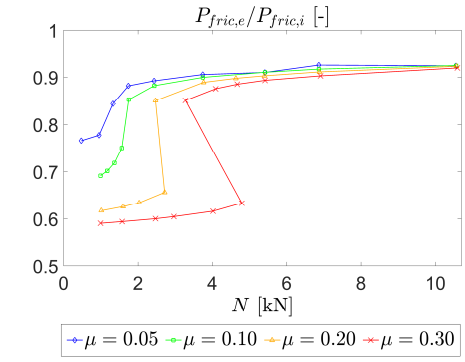
(a) Axial offset d_y of the roller from its nominal position.



(b) Longitudinal force in roller–race contacts $f_x = F_x / \mu N$.



(c) Total dissipated power by friction in roller–race contacts, P_{fric}.



(d) Ratio of frictional power in contact of roller with outer to inner race, $P_{fric,e} / P_{fric,i}$.

Figure 7.18. Steady rolling contact configurations of the roller computed with contact solutions of CECT for different normal loads and coefficients of friction.

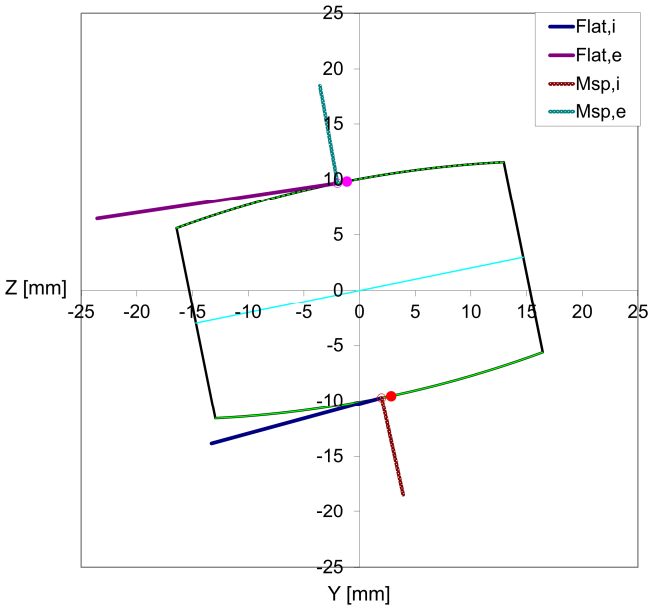
Figure 7.19 and Figure 7.20 illustrate a case of each configuration. The figures depict free-body diagrams of the roller, including the resultant tangential frictional forces and spin moments in each contact and the centres of the normal pressures marked with filled circles. The nominal contact positions are also shown with unfilled red circles. The frictional forces and moments are represented acting on these points on the roller.

In equilibrium configuration 1, the resultant frictional lateral forces transmitted to the roller in both contacts have the same direction, as may be observed in

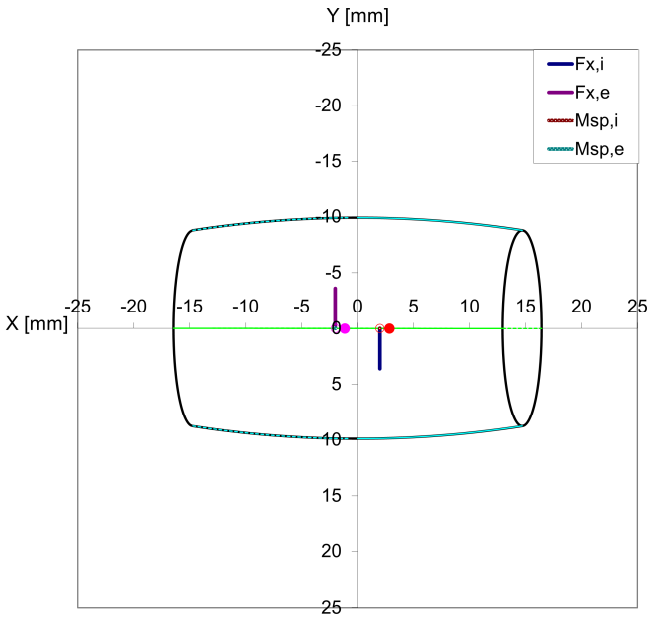
Figure 7.19a. The roller has a small axial displacement in the direction of these forces so that the contact centres are shifted in this direction, and an axial component of the normal forces appears to balance the lateral frictional forces. The lateral frictional forces are higher for higher μ. As a result, the axial displacement of the roller in the equilibrium position is higher for higher μ, and the rolling torques and frictional power increase somewhat more than proportionally with μ.

For a given μ, the axial displacement of the roller is higher at lower normal loads, cf. Figure 7.18a. This is because the ratio of camber thrust to normal load increases at lower loads in this case. In non-conformal contact, this ratio remains constant for a given ellipticity of the contact patch, coefficient of friction and spin. In this case, as the load increases, this ratio becomes smaller due to the higher friction saturation in the longitudinal direction, resulting from higher longitudinal rigid slip velocities. This effect is notable despite the modest total contact angle variation in the contact patch (approx. 11º with N = 5.5 kN) because the variation of the rolling radii is relatively high. This is a distinctive feature of conformal contact in rolling bearings with respect to conformal wheel–rail contact, as a result of the much higher ratios of lateral profile radii over rolling radii seen in this application.

Continuing with equilibrium configuration 1, the roller also adopts a small negative tilt, on the order of –0.05 mrad according to the computations with CECT (higher with higher μ). The tilt values computed with FEM are higher, cf. Figure 7.16, possibly due to global defections in the FE model. As a result of the negative tilt of the roller, the offset from the nominal position is slightly higher in the inner contact, and the inner normal load causes a higher moment along x than the outer. This is balanced by the difference in the lateral frictional forces of both contacts, being higher the outer one. The lateral frictional forces are greater than the longitudinal ones, as may be seen comparing Figure 7.19a and b, in which the forces are represented at the same scale.

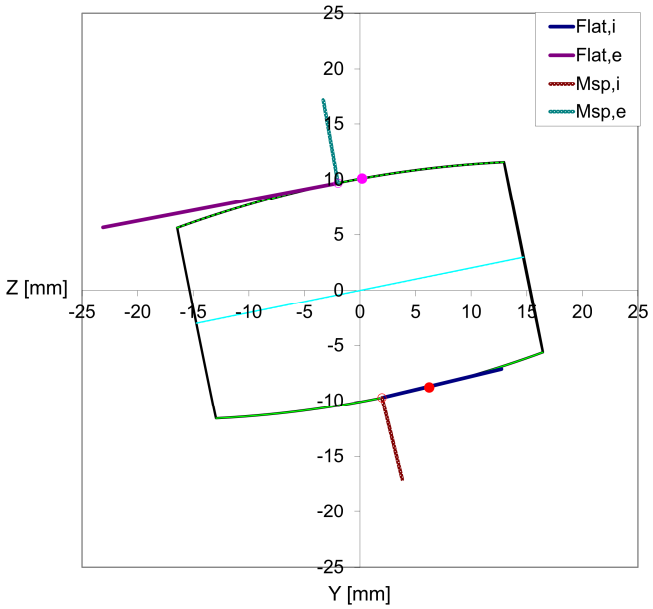


(a) Front view.

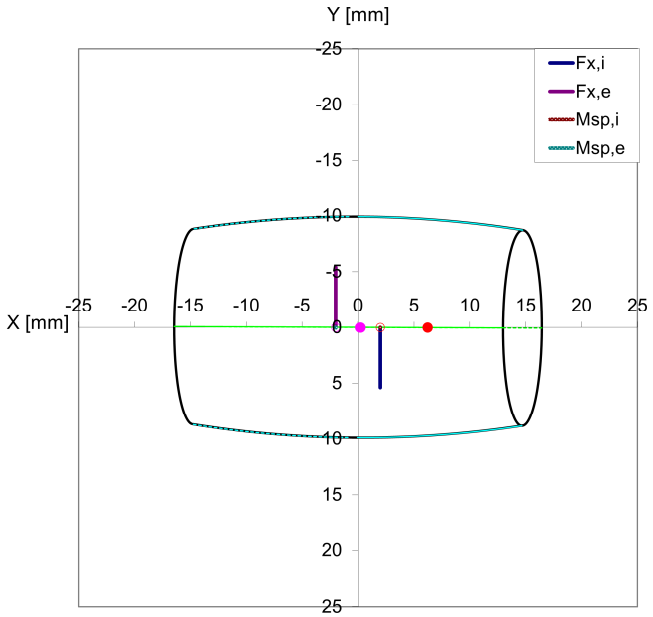


(b) Plan view.

Figure 7.19. Roller free-body diagram in equilibrium configuration 1 with N = 5.5 kN and μ = 0.30.



(a) Front view. Note: the lateral forces have been scaled down by a factor of 10 in this view.



(b) Plan view.

Figure 7.20. Roller free-body diagram in equilibrium configuration 2 with N = 4 kN and μ = 0.30.

The roller axial offsets, tilt and skew angles are much higher in equilibrium configuration 2. The skew angles increase up to –4 mrad, and the tilt angles increase up to about –0.4 mrad with $\mu = 0.3$. The offset of the inner contact from its nominal position is considerably higher than that of the outer contact, as may be seen in Figure 7.20a, as a result of this increased tilt. The dissipated power in the inner contact also increases more than in the outer contact in this configuration, cf. Figure 7.18d. The lateral frictional forces are much higher than in configuration 1, as may be seen comparing Figure 7.20a with Figure 7.19a. Note that the lateral forces have been scaled down by a factor of 10 in Figure 7.20a; the scale of the rest of the forces and moments is the same in Figure 7.19 and Figure 7.20. The direction of the lateral frictional force of the inner contact is also reversed.

It has to be mentioned that in some cases in which equilibrium configuration 2 is found as the only one complying steady equilibrium exactly, the force residuals with roller configurations which may be assimilated to equilibrium configuration 1 are not high. Also, the slow lateral movement of the roller when being out of its steady equilibrium position is recalled. Further research is necessary on the practical significance of equilibrium configuration 2.

The effects described here, as well as the lateral contact shifts shown in §7.3.2, are more relevant with higher coefficients of friction. A 3D partial slip contact model is necessary to capture them, with the ability to predict lateral frictional forces in addition to the longitudinal ones, as well as moments related to both longitudinal and lateral contact stresses. The lateral frictional forces referred to here are the ones coming e.g. from spin in a contact with no lateral creepage, which full slip and 1D models fail to predict. At lower friction levels, such as may be found in well-lubricated conditions, the offsets of the roller from its nominal position are small, as may be seen in Figure 7.18a. In this case, it may be satisfactory to assume nominal roller position and use a simpler contact mechanics model, e.g. the strip theory or the full slip solution.

The computation of the equilibrium configuration(s) of the roller for each combination of N and μ has taken on the order of minutes to tens of minutes with CECT. Each of these equilibriums involves the computation of several pairs of inner/outer contact solutions, iterating with the six parameters defining the roller configuration as explained in §7.2.2.2. The FEM described in §7.2.3

has much greater computational demands, with a single run taking about tens of hours in a high-performance computer. The application of this FEM in this study would be much less practical due to the high computational costs and the need to start from positions close to the sought equilibrium configuration, as illustrated in §7.3.3.

Chapter 8

Closure

Chapter summary

This chapter compiles the conclusions of the work carried out in this thesis and outlines perspectives for future works. Finally, the publications that have resulted from this work are listed.

8.1 Conclusions

8.1.1 Models of vehicle–track dynamic interaction

Different types of vehicle–track models have been constructed to study the vertical dynamic vehicle–track interaction in the time domain. A comprehensive representation of the whole vehicle–track system, necessary for the study of dynamic phenomena at high frequencies, has been achieved making use of systematic methodologies and standard tools offered in a commercial multibody (MBS) and in a commercial Finite Element (FE) analysis package.

The most comprehensive models comprise MBS representations of the vehicle and the track, being 'moving mass' type models, in which the vehicle moves with respect to the track. These models allow the definition of non-linear characteristics for both the vehicle and the track, and consideration of interaction between different wheelsets. The flexibility of the continuous elements of the track (i.e., the rails, and also the slab in the case of slab tracks) is incorporated in these models via the normal modes of vibration previously calculated with FE models of these components. On the other hand, simpler vehicle–track interaction models have been constructed. One of these simpler models uses a MBS representation of the track by means of assemblies of rigid bodies travelling beneath each of the vehicle's axles. The other simplified model is a 'moving irregularity' type linear FE model, in which the vehicle is represented as an assembly of lumped masses stationary on a given position in the track.

With the constructed vehicle–track models, the dynamic performance of a high speed ballasted track and three different types of slab tracks has been studied by means of numerical simulations in the time domain of a vehicle running on a straight section of track with vertical irregularities at high speed. All the studied tracks feature discretely supported rails. The validity of the results obtained with the different models developed has also been assessed, as well as the most adequate configuration of the models in different situations.

The conclusions obtained are summarized below. As mentioned in Chapter 3, the different results have been obtained considering a single set of typical physical parameters for each of the tracks. However, some of the parameters,

e.g. the pad stiffness, and consequently the dynamic properties of the tracks can change within a certain range. Consequently, the conclusions obtained must not be generalized straightforwardly to other possible configurations of each type of track.

- The wheel–rail contact forces have a lower degree of fluctuations in the ballasted track and in the STEDEF track, than in the other two types of slab tracks studied (RHEDA 2000 and floating slab). In the case of the floating slab track, although vibration transmission to the ground is expected to be considerably reduced, vibration levels between the loading source and the slab may be enlarged, as pointed out in [Kuo 2008].
- The pad forces in the floating slab track are much higher than in the other tracks studied due to the low degree of rail bending allowed by the stiffness of both pad and slab, which prevents the load of the vehicle from being more effectively spread between more pads along the track.
- In the studied design of floating slab track, another problem is seen for high-speed applications: the effect of the parametric excitation is much higher than in the other studied track types, implying that even for smooth railhead and wheel surfaces, a considerable degree of fluctuations is already seen in the wheel–rail contact forces.
- At low and mid frequencies, up to a few hundred Hertz, and when the effects of the parametric excitation are not important, very simple, full FE linear models of the vehicle–track system, of the ‘moving irregularity’ type, and with the wheel–rail contact condition simplified to a linear spring, are seen to be entirely satisfactory for the study of vertical vehicle–track dynamics.
- Additionally, in the case of tracks without a floating slab or tracks with a floating slab in which the dominant wheel–rail vibration takes place at frequencies much higher than the slab cut-on frequency, the interaction between different wheelsets of the vehicle is seen to be very small at frequencies up to a few hundred Hertz. In these cases, the track models may be further simplified to discrete spring–damper–mass assemblies beneath each wheelset with very few DOFs. The results obtained with

these simple models match very well with those obtained with more complex track models, if the physical parameters of mass, stiffness, and damping of the discrete track elements are carefully chosen adjusting the receptance of the simplified (RMB) track model to that of the more comprehensive track model. Also, a very simple model of the vehicle may be used for the study of wheel–rail contact forces in these cases, which consists of a single wheelset with a static load applied on it.

- Modelling the elasticity of the slab in slab tracks may significantly increase the complexity of vehicle–track models. Although mathematical procedures have been developed that enable a very detailed modelling of the slab as an elastic solid, in this work, it is found that a much simpler modelling of the slab by means of 1D beam elements is entirely satisfactory to study dynamic phenomena even at high frequencies, as it has been shown that the slab has fundamentally a 1D vibration in the vertical plane. With the simpler 1D beam model of the slab, the computational costs of the model are greatly reduced with respect to the models in which the slab is represented as a 3D solid. The convenience of either types of model for the slab should be assessed in different cases depending on the physical characteristics of the slab.

8.1.2 Numerical analysis of wheel–rail conformal contact

Comprehensive 3D numerical models have been developed for the analysis of wheel–rail elastic conformal contact. These have included a version of Kalker's exact contact theory extended for conformal contact that has been named CECT (Conformal Exact Contact Theory) and FE models properly tuned for the intended contact mechanics analyses.

A detailed description has been provided of the extensions for conformal contact implemented in the developed version of the exact contact theory, and of the setup and tuning of the FE models. The extensions for conformal contact implemented in CECT are related to the geometric analysis for the computation of the normal undeformed distances and rigid slip velocities, and to the determination, arrangement and computations with the influence coefficients (ICs) of solids with non-planar surfaces. This has included a new method for the

precise calculation of distances between non-planar surfaces of revolution with non-coplanar axes.

CECT is able to cope with cases with high contact and yaw angles, besides conformal contact. The FE models represent with a high degree of detail the wheel–rail contact zone, and are static models. The steady rolling contact is achieved via a step-by-step technique, in which prescribed displacements and rotations are applied to the wheel in successive increments, in a non-linear static simulation.

The computational costs have been found to be roughly three orders of magnitude higher with the FE models than with CECT, being on the order of tens of hours with the FE models, and tens of seconds with CECT for a steady-state rolling case. These figures are to be considered only as orientative, as the run times vary considerably from case to case and with the discretization, and may also depend on a number of circumstances other than the physical models themselves, such as convergence tolerance settings or the particular algorithmic implementation of the models. The exact contact theory, while having much lower computational costs than the FE models, has been proved to produce accurate results with high levels of conformity, being capable to capture the particular characteristics of conformal contact both in the normal and in the tangential parts of the contact problem.

8.1.3 Influence coefficients of the elastic solid

New results have been developed for the ICs of the half-space, working out the analytical expressions of the ICs of surface displacements for rectangular elements with bilinear stress distribution. The bilinear elements have been found to produce no relevant precision improvement over the usual uniform elements in the numerical solution of contact problems with the exact contact theory.

The determination of ICs for non-planar solids has been investigated, by means of analytical approximation and numerical calculation. Precise ICs are not readily available for general non-planar solids, and they must be generally obtained through numerical calculation. This involves a notable difficulty for the application of the exact contact theory in conformal contact problems, especially when precise results are sought with moderate or high conformity

levels considering the tangential part of the contact problem. Different possibilities have been explored for approximate analytical computation, by which this difficulty may be alleviated, and numerical calculation of ICs can be avoided as much as possible.

The numerical calculation has been tackled with linear 3D FE models, and it has been shown how the results obtained around the loaded element may be extrapolated, to avoid the need for local refinement of the loaded area. This facilitates the automatization of the process for the numerical calculation of ICs.

For the analytical approximation, a surface orientation based approach has been set out, consisting on combining the half-space ICs according to the surface orientation change between the point where the load is applied and the point where the displacements are observed. Different variants of the approximation have been studied. This approximation has been proved to capture well the main differences of the ICs of non-planar solids with respect to those of the half-space, and with a simple reasoning it provides insight into the way in which these differences occur.

On the other hand, some limitations of the surface orientation based approach have been observed. An extension has been set out to incorporate the effect of the geometric differential stiffness between convex and concave conforming solids, based on the definition of spatially variable modified elastic properties of the solids. This extension renders a considerable improvement in the approximation of the IC_{xn}. However, the associated parameter adjustment is empirical, and necessitates some numerically calculated C_{xn} ICs as reference. Additionally, this extension has currently been verified to work well only with one of the variants of the surface orientation based approximation, designated as $\hat{B}$. Despite their low magnitude, the IC_{xn} have an appreciable influence on the longitudinal tangential stresses, and use of inaccurate IC_{xn} may lead to a significant offset in the longitudinal creep force curve.

A fundamental assumption of the studied approximations is that the contacting solids may be considered approximately prismatic around the contact, as it happens in relevant rolling contact applications like the wheel–rail case or in rolling bearings. The different approximated ICs have been compared with two

sets of numerically calculated ICs, and their performance has been assessed in the normal and tangential parts of the contact problem.

Regarding the propagation of errors from the ICs to the solution of contact problems, it has been shown that geometries with faster contact angle variations are more prone to errors in the normal contact solution due to the higher errors contained in the approximated ICs. Also, the sensitivity to errors in the ICs increases when the contact pressures are more spread out in the lateral direction of the contact patch, and in contact patches with higher length to width ratios. For the normal part of the contact problem, it has been verified that the use of the half-space A_{nn} IC is satisfactory for low to moderate conformity levels (with total contact angles less than about 45º), as found by other authors before. For the tangential part of the contact problem, it has been shown that relatively small errors in the IC_{sn} and IC_{xn} ICs may lead to appreciable differences in the results. The need to properly approximate the IC_{sn} in non-planar conformal contact is highlighted, as high errors may result in the tangential contact problem if the half-space A_{sn}-s are used directly. The proposed surface orientation based approach of ICs substantially improves the results obtained with the direct use of the half-space ICs.

The best choice of approximated ICs could be to use the $\hat{B}$ ICs, except for the IC_{sn}, where the B ICs seem to perform better. The $\hat{B}_{xn}$ should be adjusted for the differential stiffness effect whenever possible; failing this, the direct use of the half-space A_{xn} is preferred over the different variants of B_{xn}. The IC_{ns} and IC_{nx} ICs can be defined according to the selected IC_{sn} and IC_{xn} ICs to fulfil reciprocity exactly. These ICs are of less importance because the normal displacements are generally dominated by the normal pressures.

The study has focused on contact between bodies with elastically similar materials. Significant differences are noted between the individual ICs of separate non-planar solids, IC^k_{xx} and especially IC^k_{ss}, that are of primary importance in the tangential contact problem. The errors in these individual ICs are of opposite signs in the convex and concave bodies, and tend to be cancelled out in the combined ICs. The extent of this compensation depends on the relative stiffness of both bodies, and may not be sufficiently favourable

especially when the convex body is the one with the more flexible material. In that case, the approximation of these ICs should be improved further.

8.1.4 Characteristics of wheel–rail conformal contact

The characteristics of wheel–rail conformal contact and its differences with respect to non-conformal contact have been studied in detail, by means of numerical analyses with the developed contact mechanics models. The study has focused in steel-on-steel contact, present in important industrial applications in addition to the wheel–rail case, as rolling bearings. 3D frictional conformal contact has been tackled, covering both the normal and tangential parts of static (compression, shift) and rolling contact problems. In contrast, most of the past published literature in conformal contact has focused on 2D contact or on frictionless contact.

Different effects derived from conformity in wheel–rail contact mechanics have been illustrated making use of simple geometries, with constant curvatures. The effects of conformity begin to be appreciable in the tangential part of the contact problem first, already with relatively low contact angle variations. At geometric level, the most relevant effects are related to the variation of the longitudinal curvature of the wheel and to the variation of the geometric spin, which in turn are originated by the variation in the contact angle. It is noteworthy as well the high sensitivity of the geometry of the normal undeformed distance to the relative orientation between the contacting solids, and particularly to the yaw angle. Some of these effects are listed next. In general, these effects become more appreciable with higher contact angles.

- The contact patches become asymmetric both in longitudinal and lateral directions, even in cases with constant lateral curvatures of the contact surfaces. The contact patches tend to become more elongated in the side with higher contact angles. Nevertheless, the overall contact patch dimensions and maximum normal pressures do not change significantly with respect to those obtained with a planar contact analysis.
- The contact patches tend to change from elliptical to rectangular shapes as the total contact angle increases. For a given geometry, they also tend to acquire shapes with higher length-to-width ratios as the load increases, becoming more elongated in the rolling direction.

- The frictional work distribution inside the contact patch tends to be biased towards the side with increasing contact angles, due to the larger longitudinal dimension of the contact patch and slip velocities in this zone.
- Different patterns are developed in the rigid slip velocity fields depending on the sign of the creepages. Considering the rolling of a concave wheel on a convex rail, in situations with zero mean contact angles and lateral creepages, the adhesion area in the contact patch tends to be more acute shaped, and more extended in the longitudinal direction, in tractive rolling than in braking rolling. The frictional work density tends to be more evenly distributed across the width of the contact patch in tractive rolling, and to be more concentrated in the central region in braking rolling.
- The creepage–creep force curves in cases with pure creepage or low spin have lower slopes than in planar contact, because the tangential stresses are not aligned in a single direction. Additionally, the longitudinal creepage–creep force curves are offset from the curves obtained with a planar analysis. This offset comes on the one hand from the variation in rolling radii and consequent variation in longitudinal rigid slip velocities, and may vary depending on the convention adopted to define the reference creepages. On the other hand, there is the offset due to the previously mentioned geometric differential stiffness, acting through the IC_{xn} IC.

The presented results evidence that it is essential to use proper representations of the normal undeformed distance and rigid slip velocity fields for precise contact mechanics analyses in conformal contact. These may deviate considerably from those calculated with a planar analysis. The precise determination of the normal undeformed distances and rigid slip velocities requires more detailed geometric analyses than those commonly used in non-conformal contact, taking into account the curved contact geometry, which implies different contact angles and rolling radii at each point.

Regarding the mechanical behaviour of the non-planar contacting solids at the local contact level, some relevant effects derived from its alteration from that of the half-space include the following:

- Tendency for more lateral spreading of the contact patch, due to the decrease of the IC_{nn} IC with the variation of the contact angle.
- The coupling between the normal and tangential contact problems, mainly through the IC_{sn} IC. As a result, quasiidentity is not applicable even with elastically similar materials. Neverless, the coupling is relatively mild with the coefficients of friction commonly found in wheel–rail contact, usually below 0.50. The effects of the normal–tangential coupling are more notable in the tangential part of the contact problem, due to the higher level of the normal pressures compared to the tangential stresses. Apart from this coupling at the contact mechanics level, there is also a coupling at the equilibrium of forces level.

The subsurface stresses at the contact level are also affected by the non-planar geometry. The maximum von Mises stress tends to increase in the concave body, and its location tends to move towards the surface, while the tendency is the opposite in the convex body.

The application of CECT has been demonstrated in different wheel–rail contact cases, with different levels of conformity and friction. The coefficient of friction conditions to a great extent the tangential problem solution in situations with high levels of rigid slip, as the ones that may occur in the contact between the rail gauge corner and the wheel flange root, and may have important variations depending on the lubrication conditions among other factors.

8.1.5 Frictional contact analysis in a spherical roller bearing

The roller–raceway frictional contact has been studied in a spherical roller bearing. For this purpose, different partial slip contact models have been used, as opposed to the usual full slip approach: the strip theory, CECT, and FEM. In this way, the use of the numerical models developed for the study of wheel–rail conformal contact has been demonstrated in the rolling bearing application. This has involved the need for some adaptations.

Regarding the application of CECT in rolling bearings, the use of a variable discretization size Δq is highlighted, due to the variable rolling radii of the

contact surfaces. Regarding the FEM, it has been shown that a precise analysis of the motions of the roller is necessary to arrive at the solution corresponding to the steady rolling configuration of the roller. Also, in the presented FEM analyses, relatively long rolling distances, of the order of ten times the total longitudinal dimension of the contact patch, have been necessary to achieve reasonably stabilized solutions. These contrast with the much lower distances necessary to reach a quasi-steady state at the local contact level. The required rolling distances could be different in other cases or with other analysis set-ups.

In the studied spherical roller bearing case, the contact patches present a high width-to-length ratio, modest contact angle variations, and rigid slip velocities mostly aligned in the rolling direction. These features point to 2D-like contact solutions and anticipate that representative results may be obtained with strip theory. Indeed, the results obtained with the strip theory agree quite well with those obtained with the more comprehensive 3D contact models, as long as the contact patch position is appropriately adjusted. However, a 3D partial slip contact model is needed to perform this adjustment, with a proper description of the rigid slip velocities in the roller–raceway contacts according to the geometry of the non-planar contact surfaces, and the capability to compute lateral frictional forces in addition to the longitudinal ones, as well as associated moments. While the lateral frictional forces are relatively small in the studied short contact patches (as compared to the normal loads), with high friction, they may be high enough to cause a noticeable lateral shift of the contact patch with a roller having zero nominal roller angle. This lateral shift, in turn, influences the tangential contact solution. Al lower friction levels, the frictional forces are smaller, and the contact shifts from their nominal positions have been shown to decrease in the studied bearing. Still, the relevance of possible contact shifts should be evaluated for each bearing geometry and operating conditions. These could come for reasons other than frictional forces, e.g., in the presence of internal clearances, axial loading, and misalignment.

The advantages and disadvantages of the different contact models considered in this study are summarized in the following points:

- The strip theory may be used to get detailed steady-state contact solutions efficiently in cases with wide contact patches and slip

velocities aligned in one direction, in which the contact location may be estimated beforehand, aside from the contact frictional forces.

- The exact contact theory is a comprehensive 3D partial slip contact model that may be applied to get precise frictional contact solutions within the framework of linear elastostatics, providing a realistic representation of the effect of the contact frictional forces on the roller equilibrium. Its computational costs are between those of strip theory and FEM, while allowing a much more general application than strip theory.
- The main difficulty for the practical application of FEM in the type of contact problems treated here is its high computational costs. Its use should be reserved for studying phenomena not allowed by simpler contact models, such as non-linear material behaviour.

The steady rolling configurations of the roller have been computed for different normal loads and friction levels with CECT. Two different steady equilibrium configurations of the roller have been found. There is a first configuration, which may be described as the expected one, in which the roller has low or moderate offsets from its nominal position in the bearing radial plane. This configuration is feasible at higher loads. The second configuration, which is feasible at lower loads, presents higher roller offsets and dissipated powers. The leap between both configurations increases with higher friction levels. The finding of these two configurations in a numerical analysis requires using a 3D partial slip contact model. Further research is necessary on the significance of the second steady equilibrium configuration of the roller. For this purpose, it may be interesting to investigate the transient evolution of the roller in its orbital motion around the bearing.

8.2 Future works

The following subsections suggest possible directions of further research related to the work carried out in this thesis.

8.2.1 Vehicle–track dynamic interaction

The scope of the developed MBS–FE vehicle–track dynamic interaction models could be extended for lateral dynamics. This would allow using the models for the study of phenomena that may happen during inscription of the vehicle in tight curves for example.

The extension of the models for lateral dynamics would involve an increase in their complexity. It would be necessary to model the rail cross-section in more detail, to take into account its distortion in addition to the bending and torsional deformations of the rail. Additionally, the flexibility of the wheels should be considered. While it may be reasonable to consider the wheelsets as rigid for vertical dynamics in certain frequency ranges as has been done in this work, for lateral dynamics it becomes more imperative to consider them as flexible bodies. All this would lead to a higher number of degrees of freedom and modes of vibration to be included in the models.

8.2.2 Conformal contact

Directions for further research on conformal contact are listed below. These are set out with regards to the wheel–rail application, while many are extensible to the rolling bearing case.

- Development of simplified conformal rolling contact models for use in vehicle dynamics simulations. The comprehensive numerical models developed in this thesis are aimed for detailed contact mechanics studies. For rail vehicle dynamic simulations, more emphasis has to be placed on computational efficiency (cf. §2.1). At the same time, the simplified models should retain the ability to capture the inherent features of conformal contact relevant for the analyses to be carried out.

  The dynamic and contact mechanics simulations could also be undertaken separately, getting first the contact parameters from the dynamic simulations, and calculating with these a limited number of detailed contact solutions in a second step with a comprehensive contact mechanics model. But contact conformity could influence the vehicle

dynamics, calling for coupled simulation, as done in [Bashkar 1997] and [Pascal 2019] with simplified conformal contact models.

- Application to damage analysis. The conformal contact models could be applied to damage analyses such as wear evolution predictions in identified critical scenarios. The results of the more comprehensive models could then be compared with those of more simplified models, and if possible, with field measurements.

- Temperature effects in the contact. Changes in the profiles of the contacting surfaces resulting from non-uniform temperature distributions across them could alter significantly the undeformed distances in closely conforming surfaces. This could have an effect both in the normal and in the tangential parts of the contact problem. Further, higher slip levels are usually encountered in conformal contact situations compared to planar contact, which may promote conditions with higher temperatures and temperature gradients inside the contact.

- Complex frictional behaviour. This includes the consideration of 3rd body layer, and variable coefficient of friction with the slip velocity. The exact contact theory allows the consideration of these features, and are actually implemented in CONTACT (cf. [CONTACT UG 2013], [Vollebregt 2014d]). In this way, conformal contacts could be studied considering these interfacial behaviours. This could be undertaken either with CONTACT, or adding these capabilities in a future version of CECT.

  The impact of these features in the real contact behaviours may be much higher than the inaccuracies coming from more simplified contact models, such as the ones described in §2.1.2; cf. [Hobbs 1967], [Vollebregt 2014d], and [Magel 2017]. In this regard, the main difficulty lies in the proper characterization of the real constitutive behaviours and involved physical parameters, rather than in the solution of the contact problem itself.

- Plasticity. Particular cases of planar contact have been studied in detail, e.g. developing shakedown maps for Hertzian cases in partial slip (cf. [Johnson 2001]). Numerical studies of planar contacts in the wheel–rail application have been carried out as well with detailed 3D FE models, as in [Zhao 2015]. In the rolling bearing application, [Tonazzi 2017] is an example of a case study of conformal contact with plasticity. Still, there is a lack in the literature of detailed studies of 3D frictional conformal contact with plasticity in the wheel–rail application. A higher complexity than in elastic contacts is involved, as multiple cycles need to be simulated, and more detailed material data is needed, being necessary to properly characterize the cyclic plastic behaviour of the materials.

- Layered solids. Rolling contact surfaces are sometimes treated to enhance their wear, RCF and/or corrosion resistance. Sometimes, the treatments involve coatings or claddings with different elastic properties from the substrate material. Contact problems with layered solids could be treated with the contact models developed in this thesis.

  For the application of CECT, it would be necessary to numerically calculate ICs. A further development could be to extend the calculation of approximated ICs for layered solids, to avoid their calculation with dedicated FE models. A difficulty here is that in the case of layered half-spaces, ICs are available in analytical form only in the Fourier domain, being necessary to obtain the values in the space domain by numerical integration.

8.3 Publications directly derived from the thesis

Part of the work developed in this thesis has been published in the references listed in the following subsections.

8.3.1 JCR papers

- Blanco-Lorenzo J, Santamaria J, Vadillo EG, Oyarzabal O. (2011). Dynamic comparison of different types of slab track and ballasted track

using a flexible track model. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 225(6), 574–92. https://doi.org/10.1177/0954409711401516.

- Blanco-Lorenzo J, Santamaria J, Vadillo EG, Correa N. (2016). On the influence of conformity on wheel–rail rolling contact mechanics. Tribology International 103, 647–67. https://dx.doi.org/10.1016/j.triboint.2016.07.017.

- Blanco-Lorenzo J, Santamaria J, Vadillo EG, Correa N. (2018). A contact mechanics study of 3D frictional conformal contact. Tribology International 119, 143–56. https://doi.org/10.1016/j.triboint.2017.10.022.

- Blanco-Lorenzo J, Vollebregt EAH, Santamaria J, Vadillo EG. (2021). Approximating the influence coefficients of non-planar elastic solids for conformal contact analysis. Tribology International 154, 106671. https://doi.org/10.1016/j.triboint.2020.106671.

- Blanco-Lorenzo J, Liu S, Santamaria J, Meehan PA, Vadillo EG. (2023). Frictional contact analysis in a spherical roller bearing. Journal of Computational Design and Engineering 10(1), 139–59. https://doi.org/10.1093/jcde/qwac126.

8.3.2 Conferences

- Blanco-Lorenzo J, Santamaria J, Vadillo EG, Correa N. Finite element study of wheel-rail conformal rolling contact. In: Bernasch J, Rosenberger M (Eds.), Proceedings of the 24th International Symposium on Dynamics of Vehicles on Roads and Tracks. Graz, Austria, 17–21 August, 2015.
  Best poster award in the category of Rail.

- Blanco-Lorenzo J, Santamaria J, Vadillo EG, Correa N. Characteristics of wheel–rail conformal rolling contact. In: Ambrósio J, Schiehlen W, Pombo J (Eds.), Proceedings of the EUROMECH Colloquium 578:

Rolling Contact Mechanics for Multibody System Dynamics. Funchal, Madeira, Portugal, 10–13 April, 2017.

- Blanco-Lorenzo J, Santamaria J, Vadillo EG, Oyarzabal O, Robles R. Análisis del contacto conforme rueda-carril bajo diferentes condiciones de fricción. XXIII Congreso Nacional de Ingeniería Mecánica. Jaén, 20-22 octubre, 2021.

8.3.3 Book chapter

- Viñolas J, Vadillo EG, San Emeterio A, Vera E, Alonso A, Nieto J, Santamaria J, Oyarzabal O, Correa N, Gómez J, Blanco-Lorenzo J, Herreros J. (2011). Ch. 4: Interacción vehículo-vía. In: Estudio del comportamiento a medio y largo plazo de las estructuras ferroviarias de balasto y placa. CEDEX, Ministerio de Fomento. ISBN: 978-84-7790-528-8.

Appendix

Integrals for half-space ICs of bilinear rectangular elements

Details of the calculation of the J_1, J_2, J_3, and J_5 integrals defined in Eqs. (5.1), (5.2), (5.3), and (5.5) for rectangular elements in the half-space surface with bilinear load distribution are given here. Specifically, the contributions to these integrals of element number 1 shadowed in blue in Figure A.1 are considered, denoted as J_1^1, J_2^1, J_3^1 and J_5^1. The resulting expressions for J_2^1, J_3^1 and J_5^1 are given in Eqs. (5.24), (5.25) and (5.26), respectively.

The same Cartesian coordinate system as in §5.1 is used. (x, y) are the coordinates of the node shared by the four loaded elements in the half-space surface as illustrated in Figure A.1. The integrals for the different J_i^1 are carried out over the surface of element 1, i.e. on (X, Y) where $x < X < x + a$ and $y < Y < y + b$, first along X and then along Y. Letter D is used for the expressions arising after the integration in X, and letter E for the ones arising after the subsequent integration in Y.

The bilinear load distribution $T(X, Y)$ in element 1 is given in Eq. (5.23), and reproduced in Eq. (A.1) for convenience.

$$T(X,Y)=\left(1-\frac{X-x}{a}\right)\left(1-\frac{Y-y}{b}\right) \tag{A.1}$$

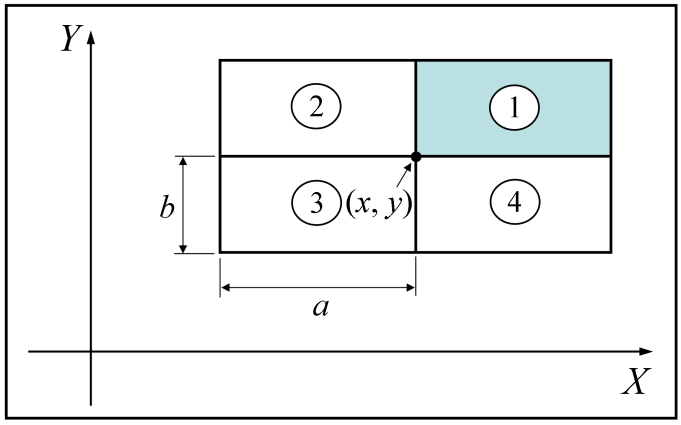


Figure A.1. Coordinate system and element dimensions for the calculation of the ICs for rectangular elements with bilinearly varying load distributions.

The double brackets in the equations given in the following subsections stand for the evaluation of the integral expression *f* inside them at the limit points (*x*, *x*+*a*) for *X* and (*y*, *y*+*b*) for *Y*, as expressed in Eq. (A.2), reproduced from Eq. (5.27). ρ is equal to the distance from the origin to the (*X*, *Y*) point, i.e. $\rho=\sqrt{X^2+Y^2}$.

$$[[\,f\,]] = f(x+a, y+b) - f(x, y+b) - f(x+a, y) + f(x, y) \tag{A.2}$$

A.1 J_1

The integral to be calculated is given in Eq. (A.3). This integral may be calculated as $J_2^{\,1} + J_4^{\,1}$, cf. Eq. (5.1). Here, the integration of Eq. (A.3) is shown.

$$J_1^{\,1} = \int_y^{y+b}\int_x^{x+a}\left(1-\frac{X-x}{a}\right)\left(1-\frac{Y-y}{b}\right)\frac{1}{\rho}\,dX\,dY \tag{A.3}$$

The integrals of Eqs. (A.5) and (A.6) are carried out for the integration in *X* of Eq. (A.3) (expressed in Eq. (A.4)), and are combined according to Eq. (A.7).

$$D_1 = \int_x^{x+a} \left(1 - \frac{X-x}{a}\right)\left(1 - \frac{Y-y}{b}\right) \frac{1}{\sqrt{X^2+Y^2}}\, dX \tag{A.4}$$

$$D_{1,1} = \int_x^{x+a} \frac{dX}{\sqrt{X^2+Y^2}} = \left[\ln\left(X + \sqrt{X^2+Y^2}\right)\right]_x^{x+a}$$

$$= \ln\left(\frac{x+a+\sqrt{(x+a)^2+Y^2}}{x+\sqrt{x^2+Y^2}}\right) = \operatorname{asinh}\frac{X}{Y}\Big]_x^{x+a}$$

$$= \left|\operatorname{asinh}\frac{x+a}{Y} - \operatorname{asinh}\frac{x}{Y}\right| = \left[\ln\left(X + \sqrt{X^2+Y^2}\right)\right]_x^{x+a} \tag{A.5}$$

$$D_{1,2} = \int_x^{x+a} \frac{X}{\sqrt{X^2+Y^2}}\, dX = \sqrt{X^2+Y^2}\Big]_x^{x+a} \tag{A.6}$$

$$D_1 = \left(1 - \frac{Y-y}{b}\right)\left[\left(1+\frac{x}{a}\right) D_{1,1} - \frac{D_{1,2}}{a}\right] \tag{A.7}$$

Next, the result of Eq. (A.7) is to be integrated in Y. For this purpose, the integrals of Eqs. (A.8)–(A.11) are carried out. For the integral of Eq. (A.8), the substitution $u = X / Y$ is used.

$$E_{1,1} = \int_y^{y+b} \left|\operatorname{asinh}\frac{x+a}{Y} - \operatorname{asin}\frac{x}{Y}\right| dY$$

$$= \left[\left[X \operatorname{sgn}(Y) \left| \frac{Y}{X} \operatorname{asinh}\frac{X}{Y} + \ln\left|\frac{1+\sqrt{1+(X/Y)^2}}{X/Y}\right| \right| \right]\right]$$

$$= \left[\left[Y \ln(X+\rho) + X \ln\left|\frac{Y+\rho}{X}\right| \right]\right] \tag{A.8}$$

$$E_{1,2} = \int_y^{y+b} Y \left| \operatorname{asinh}\frac{x+a}{Y} - \operatorname{asin}\frac{x}{Y} \right| dY$$

$$= \frac{1}{2}\left[\left[X^2 \left(\frac{\operatorname{asinh}\dfrac{X}{|Y|}}{\left(\dfrac{X}{Y}\right)^2} + \frac{\sqrt{1+\left(\dfrac{X}{Y}\right)^2}}{\dfrac{X}{|Y|}} \right)\right]\right]$$

$$= \frac{1}{2}\left[\left[Y^2 \ln(X+\rho) + X\rho \right]\right] \tag{A.9}$$

$$E_{1,3} = \int_y^{y+b} \left(\sqrt{(x+a)^2 + Y^2} - \sqrt{x^2 + Y^2} \right) dY$$

$$= \frac{1}{2}\left[\left[\operatorname{sgn}(X) \left(\operatorname{sgn}(X)\, Y\rho + X^2 \operatorname{asinh}\frac{Y}{X} \right)\right]\right]$$

$$= \frac{1}{2}\left[\left[X^2 \ln(Y+\rho) + Y\rho \right]\right] \tag{A.10}$$

$$E_{1,4} = \int_y^{y+b} Y \left(\sqrt{(x+a)^2 + Y^2} - \sqrt{x^2 + Y^2} \right) dY$$

$$= \frac{1}{3}\left[\left[\rho^3 \right]\right] \tag{A.11}$$

$J_1^{\,1}$ is calculated from the previous integrals according to Eq. (A.12).

$$J_1^{\,1} = \left(1+\frac{y}{b}\right)\left(1+\frac{x}{a}\right) E_{1,1} - \frac{1}{b}\left(1+\frac{x}{a}\right) E_{1,2}$$

$$- \frac{1}{a}\left(1+\frac{y}{b}\right) E_{1,3} + \frac{1}{ab} E_{1,4} \tag{A.12}$$

After operating, the expression of Eq. (A.13) is obtained.

$$J_1^{\,1} = \left[\left[\left(1+\frac{x}{a}\right)\left\{\left(1+\frac{y}{b}\right)X\,\mathrm{sgn}(Y)\left|\frac{Y}{X}\,\mathrm{asinh}\frac{X}{Y}+\ln\left|\frac{1+\sqrt{1+(X/Y)^2}}{X/Y}\right|\right|\right.\right.$$

$$\left.-\frac{X^2}{2b}\left(\frac{\mathrm{asinh}\dfrac{X}{|Y|}}{\left(\dfrac{X}{Y}\right)^2}+\frac{\sqrt{1+\left(\dfrac{X}{Y}\right)}}{\dfrac{X}{|Y|}}\right)\right\}$$

$$-\frac{1}{2a}\left(1+\frac{y}{b}\right)\mathrm{sgn}(X)\left(\mathrm{sgn}(X)Y\rho+X^2\,\mathrm{asinh}\frac{Y}{X}\right)$$

$$\left.\left.+\frac{1}{3ab}\rho^3\right]\right] \tag{A.13}$$

A.2 J_2

The integral to be calculated is given in Eq. (A.14).

$$J_2^{\,1} = \int_y^{y+b}\int_x^{x+a}\left(1-\frac{X-x}{a}\right)\left(1-\frac{Y-y}{b}\right)\frac{X^2}{\rho^3}\,dX\,dY \tag{A.14}$$

The integrals of Eqs. (A.16) and (A.17) are carried out for the integration in X of Eq. (A.14) (expressed in Eq. (A.15)), and are combined according to Eq. (A.18).

$$D_2 = \int_x^{x+a}\left(1-\frac{X-x}{a}\right)\left(1-\frac{Y-y}{b}\right)\frac{X^2}{\left(X^2+Y^2\right)^{3/2}}\,dX \tag{A.15}$$

$$D_{2,1} = \int_x^{x+a} \frac{X^2}{\left(X^2+Y^2\right)^{3/2}} dX = \left[\frac{-X}{\sqrt{X^2+Y^2}} + \operatorname{asinh}\frac{X}{Y} \right]_x^{x+a}$$

$$= \left[\frac{-X}{\sqrt{X^2+Y^2}} \right]_x^{x+a} + \left| \operatorname{asinh}\frac{x+a}{Y} - \operatorname{asinh}\frac{x}{Y} \right|$$

$$= \left[\frac{-X}{\sqrt{X^2+Y^2}} + \ln\left(X+\sqrt{X^2+Y^2}\right) \right]_x^{x+a} \tag{A.16}$$

$$D_{2,2} = \int_x^{x+a} \frac{X^3}{\left(X^2+Y^2\right)^{3/2}} dX = \left[\sqrt{X^2+Y^2} + \frac{Y^2}{\sqrt{X^2+Y^2}} \right]_x^{x+a} \tag{A.17}$$

$$D_2 = \left(1-\frac{Y-y}{b}\right)\left[\left(1+\frac{x}{a}\right) D_{2,1} - \frac{D_{2,2}}{a}\right] \tag{A.18}$$

Next, the result of Eq. (A.18) is to be integrated in Y. For this purpose, the integrals of Eqs. (A.19)–(A.22) are carried out, together with the integrals $E_{1,1}$, $E_{1,2}$, $E_{1,3}$, and $E_{1,4}$ defined in Eqs. (A.8)–(A.11).

$$E_{2,5} = \int_y^{y+b} \left(\frac{x+a}{\sqrt{(x+a)^2+Y^2}} - \frac{x}{\sqrt{x^2+Y^2}} \right) dY$$

$$= \left[X \left| \operatorname{asinh}\frac{y+b}{X} - \operatorname{asinh}\frac{y}{X} \right| \right]_x^{x+a} = [[X \ln(Y+\rho)]] \tag{A.19}$$

$$E_{2,6} = \int_y^{y+b} \left(\frac{(x+a)Y}{\sqrt{(x+a)^2+Y^2}} - \frac{xY}{\sqrt{x^2+Y^2}} \right) dY = [[X\rho]] \tag{A.20}$$

$$E_{2,7} = \int_y^{y+b} \left(\frac{Y^2}{\sqrt{(x+a)^2 + Y^2}} - \frac{Y^2}{\sqrt{x^2 + Y^2}} \right) dY$$

$$= \frac{1}{2} \left[\!\left[\; Y\rho - X^2 \ln(Y + \rho) \; \right]\!\right] \tag{A.21}$$

$$E_{2,8} = \int_y^{y+b} \left(\frac{Y^3}{\sqrt{(x+a)^2 + Y^2}} - \frac{Y^3}{\sqrt{x^2 + Y^2}} \right) dY$$

$$= \left[\left[\frac{1}{3}\rho^3 - X^2 \rho \; \right]\right] \tag{A.22}$$

J_2^1 is calculated from the previous integrals according to Eq. (A.23).

$$J_2^{\,1} = \left(1 + \frac{y}{b}\right)\left(1 + \frac{x}{a}\right)\left(E_{1,1} - E_{2,5}\right) - \frac{1}{b}\left(1 + \frac{x}{a}\right)\left(E_{1,2} - \frac{E_{2,6}}{2}\right)$$

$$- \frac{1}{a}\left(1 + \frac{y}{b}\right)\left(E_{1,3} + E_{2,7}\right) + \frac{1}{ab}\left(E_{1,4} + E_{2,8}\right) \tag{A.23}$$

After operating, the expression of Eq. (5.24) is obtained for J_2^1. In doing this, there is a $(y + b)\, X \ln|1/X|$ term that cancels out in the four evaluations and is therefore eliminated from the resulting primitive function.

A.3 J_3

The integral to be calculated is given in Eq. (A.24).

$$J_3^{\,1} = \int_y^{y+b} \int_x^{x+a} \left(1 - \frac{X - x}{a}\right)\left(1 - \frac{Y - y}{b}\right) \frac{XY}{\rho^3} \, dX dY \tag{A.24}$$

For the integration in X of Eq. (A.24) (expressed in Eq. (A.25)), the integral of Eq. (A.26) is carried out together with the integral $D_{2,1}$ defined in Eq. (A.16). These are then combined according to Eq. (A.27).

$$D_3 = \int_x^{x+a} \left(1 - \frac{X-x}{a}\right)\left(1 - \frac{Y-y}{b}\right) \frac{XY}{\left(X^2+Y^2\right)^{3/2}}\, dX \tag{A.25}$$

$$D_{3,1} = \int_x^{x+a} \frac{X}{\left(X^2+Y^2\right)^{3/2}} dX = \left[\frac{-1}{\sqrt{X^2+Y^2}}\right]_x^{x+a} \tag{A.26}$$

$$D_3 = \left(1 - \frac{Y-y}{b}\right) Y \left[\left(1 + \frac{x}{a}\right) D_{3,1} - \frac{D_{2,1}}{a}\right] \tag{A.27}$$

Next, the result of Eq. (A.27) is to be integrated in Y. For this purpose, the integrals of Eqs. (A.28)–(A.30) are carried out, together with the integrals $E_{1,2}$, $E_{2,6}$, and $E_{2,7}$ defined in Eqs. (A.9), (A.20) and (A.21), respectively.

$$E_{3,1} = \int_y^{y+b} \left(\frac{Y}{\sqrt{(x+a)^2+Y^2}} - \frac{Y}{\sqrt{x^2+Y^2}}\right) dY = [[\rho]] \tag{A.28}$$

$$E_{3,4} = \int_y^{y+b} \left(\frac{Y^2(x+a)}{\sqrt{(x+a)^2+Y^2}} - \frac{Y^2 x}{\sqrt{x^2+Y^2}}\right) dY$$

$$= \frac{1}{2}\left[\left[\; XY\rho - X^3 \ln(Y+\rho)\;\right]\right] \tag{A.29}$$

$$E_{3,6} = \int_y^{y+b} Y^2 \left|\operatorname{asinh}\frac{x+a}{Y} - \operatorname{asinh}\frac{x}{Y}\right| dY$$

$$= \frac{1}{3}\left[\left[Y^3 \ln(X+\rho) - \frac{X^3}{2}\ln(Y+\rho) + \frac{1}{2}XY\rho\right]\right] \tag{A.30}$$

For the calculation of $E_{3,6}$, the result given in Eq. (A.31) is used. In this equation, the substitution $u = x / Y$ is used.

$$\int_y^{y+b} Y^2 \operatorname{asinh}\frac{x}{Y}\,dY = \int_{x/(y+b)}^{x/y} \frac{x^2}{u^2}\operatorname{asinh}(u)\left(\frac{-x}{u^2}\right)du$$

$$= -x^3 \int_{x/(y+b)}^{x/y} \frac{1}{u^4}\operatorname{asinh}(u)\,du = -x^3\left[\frac{-1}{3u^3}\operatorname{asinh}(u)\right.$$

$$\left. + \frac{1}{6}\operatorname{asinh}\left(\frac{1}{u}\right) - \frac{1}{6u^2}\sqrt{1+u^2}\right]_{x/(y+b)}^{x/y}$$

$$= -x^3\left(\frac{-(y+b)^3}{3x^3}\operatorname{asinh}\frac{x}{y+b} + \frac{1}{6}\operatorname{asinh}\frac{y+b}{x}\right.$$

$$- \frac{(y+b)^2}{6x^2}\sqrt{1+\left(\frac{x}{y+b}\right)^2} + \frac{y^3}{3x^3}\operatorname{asinh}\frac{x}{y}$$

$$\left. - \frac{1}{6}\operatorname{asinh}\frac{y}{x} + \frac{y^2}{6x^2}\sqrt{1+\left(\frac{x}{y}\right)^2}\right)$$

$$= \left[\frac{Y^3}{3}\operatorname{asinh}\frac{x}{Y} - \frac{x^3}{6}\operatorname{asinh}\frac{Y}{x} + \frac{xY}{6}\sqrt{x^2+Y^2}\right]_y^{y+b}$$

$$= \frac{1}{3}\left[Y^3 \ln\left(x+\sqrt{x^2+Y^2}\right) - \frac{x^3}{2}\ln\left(Y+\sqrt{x^2+Y^2}\right)\right.$$

$$\left. + \frac{xY}{2}\sqrt{x^2+Y^2}\right]_y^{y+b} \tag{A.31}$$

J_3^1 is calculated from the integrals $E_{1,2}$, $E_{2,6}$, $E_{2,7}$, $E_{3,1}$, $E_{3,4}$, and $E_{3,6}$ according to Eq. (A.32).

$$J_3^{\,1} = -\left(1+\frac{y}{b}\right)\left(1+\frac{x}{a}\right) E_{3,1} + \frac{1}{b}\left(1+\frac{x}{a}\right) E_{2,7}$$

$$+\frac{1}{a}\left(1+\frac{y}{b}\right)\left(E_{2,6}-E_{1,2}\right)+\frac{1}{ab}\left(-E_{3,4}+E_{3,6}\right) \tag{A.32}$$

After operating, the expression of Eq. (5.25) is obtained for $J_3^{\,1}$.

A.4 J_5

The integral to be calculated is given in Eq. (A.33).

$$J_5^{\,1} = \int_y^{y+b}\int_x^{x+a}\left(1-\frac{X-x}{a}\right)\left(1-\frac{Y-y}{b}\right)\frac{X}{\rho^2}\,dX\,dY \tag{A.33}$$

The integrals of Eqs. (A.35) and (A.36) are carried out for the integration in X of Eq. (A.33) (expressed in Eq. (A.34)), and are combined according to Eq. (A.37).

$$D_5 = \int_x^{x+a}\left(1-\frac{X-x}{a}\right)\left(1-\frac{Y-y}{b}\right)\frac{X^2}{\left(X^2+Y^2\right)^{3/2}}\,dX \tag{A.34}$$

$$D_{5,1} = \int_x^{x+a}\frac{X}{X^2+Y^2}dX = \frac{1}{2}\left[\ln\left(X^2+Y^2\right)\right]_x^{x+a} \tag{A.35}$$

$$D_{5,2} = \int_x^{x+a}\frac{X^2}{X^2+Y^2}dX = \left[X - Y\,\mathrm{atan}\frac{X}{Y}\right]_x^{x+a} \tag{A.36}$$

$$D_5 = \left(1-\frac{Y-y}{b}\right)\left[\left(1+\frac{x}{a}\right) D_{5,1} - \frac{D_{5,2}}{a}\right] \tag{A.37}$$

Next, the result of Eq. (A.37) is to be integrated in Y. For this purpose, the integrals of Eqs. (A.38)–(A.43) are carried out.

$$E_{5,1} = \int_y^{y+b} \left(\ln\left((x+a)^2 + Y^2 \right) - \ln\left(x^2 + Y^2 \right) \right) dY$$

$$= \left[\left[Y \ln(\rho) - 2Y + 2X \ \operatorname{atan} \frac{Y}{X} \right] \right] \tag{A.38}$$

$$E_{5,2} = \int_y^{y+b} Y \left(\ln\left((x+a)^2 + Y^2 \right) - \ln\left(x^2 + Y^2 \right) \right) dY$$

$$= \frac{1}{2} \left[\left[\rho \ln(\rho) - Y^2 \right] \right] \tag{A.39}$$

$$E_{5,3} = \int_y^{y+b} dY = b \tag{A.40}$$

$$E_{5,4} = \int_y^{y+b} Y \, dY = \frac{2y+b}{2} \tag{A.41}$$

$$E_{5,5} = \int_y^{y+b} Y \left(\operatorname{atan} \frac{x+a}{Y} - \operatorname{atan} \frac{x}{Y} \right) dY$$

$$= \frac{1}{2} \left[\left[X^2 \left\{ \left(1 + \frac{Y^2}{X^2} \right) \operatorname{atan} \frac{X}{Y} + \frac{Y}{X} \right\} \right] \right] \tag{A.42}$$

$$E_{5,6} = \int_y^{y+b} Y^2 \left(\operatorname{atan} \frac{x+a}{Y} - \operatorname{atan} \frac{x}{Y} \right) dY$$

$$= \frac{1}{3} \left[\left[X^3 \left\{ \left(\frac{Y}{X} \right)^3 \operatorname{atan} \frac{X}{Y} + \frac{1}{2} \left(\frac{Y}{X} \right)^2 + \frac{1}{2} \ln\left(\frac{X^2}{\rho^2} \right) \right\} \right] \right] \tag{A.43}$$

For the calculation of $E_{5,5}$ and $E_{5,6}$, the results given in Eqs. (A.44) and (A.45) are used, respectively. In these equations, the substitution $u = x / Y$ is used.

$$\int_y^{y+b} Y \operatorname{atan}\frac{x}{Y}\,dY = -x^2 \int_{\frac{x}{y}}^{\frac{x}{y+b}} \frac{1}{u^3}\ \operatorname{atan}(u)\ du$$

$$= \frac{x^2}{2}\left[\left(1+\frac{1}{u^2}\right)\operatorname{atan}(u)+\frac{1}{u}\right]_{\frac{x}{y}}^{\frac{x}{y+b}} \tag{A.44}$$

$$\int_y^{y+b} Y^2 \operatorname{atan}\frac{x}{Y}\,dY = -x^3 \int_{\frac{x}{y}}^{\frac{x}{y+b}} \frac{1}{u^4}\ \operatorname{atan}(u)\ du$$

$$= -x^3\left\{\left[\frac{-1}{3u^3}\operatorname{atan}(u)\right]_{\frac{x}{y}}^{\frac{x}{y+b}} + \frac{1}{3}\int_{\frac{x}{y}}^{\frac{x}{y+b}} \frac{1}{u^3\left(1+u^2\right)}\ du\right\}$$

$$= -x^3\left[\frac{-1}{3u^3}\operatorname{atan}(u) - \frac{1}{6}\left(\frac{1}{u^2}+\ln\left(\frac{u^2}{1+u^2}\right)\right)\right]_{\frac{x}{y}}^{\frac{x}{y+b}}$$

$$= \frac{x^3}{3}\left[\frac{1}{u^3}\operatorname{atan}(u)+\frac{1}{2u^2}+\frac{1}{2}\ln\left(\frac{u^2}{1+u^2}\right)\right]_{\frac{x}{y}}^{\frac{x}{y+b}} \tag{A.45}$$

$J_5^{\,1}$ is calculated from the previous integrals according to Eq. (A.46).

$$J_5^{\,1} = \left(1+\frac{y}{b}\right)\left(1+\frac{x}{a}\right)\ \frac{E_{5,1}}{2} - \frac{1}{2b}\left(1+\frac{x}{a}\right)\ E_{5,2} - \left(1+\frac{y}{b}\right)\ E_{5,3}$$

$$+\frac{E_{5,4}}{b}+\frac{1}{a}\left(1+\frac{y}{b}\right)\ E_{5,5} - \frac{E_{5,6}}{ab} \tag{A.46}$$

After operating, the expression of Eq. (5.26) is obtained for $J_5^{\,1}$. In doing this, there is a Y term and a $Y^2 / 4$ term that cancel out in the four evaluations, and are therefore eliminated from the resulting primitive function. The last term in

Eq. (5.26), $(-(y + b) / 2 \times \mathrm{sgn}(XY) \times X^2 \times \pi / 2)$, has been added to compensate the discontinuity of the $((y + b) / 2 \times R^2 \times \mathrm{atan}(X / Y))$ term when Y passes through 0.

Bibliography

[Abaqus doc 2012] Abaqus. Documentation, version 6.12. Dassault Systèmes Simulia Corp.; 2012. http://www.simulia.com.

[Abascal 2007] Abascal R, Rodríguez-Tembleque L. (2007). Steady-state 3D rolling contact using boundary elements. Communications in Numerical Methods in Engineering 23, 905–20. https://doi.org/10.1002/cnm.931.

[Abdullah 2020] Abdullah MU, Khan ZA, Kruhoeffer W, Blass T. (2020). A 3D finite element model of rolling contact fatigue for evolved material response and residual stress estimation. Tribology Letters 68, 122. https://doi.org/10.1007/s11249-020-01359-w.

[Aguirrebeitia 2010] Aguirrebeitia J, Avilés R, Fernández de Bustos I, Abasolo M. (2010). Calculation of general static load-carrying capacity for the design of four-contact-point slewing bearings. Journal of Mechanical Design 132(6), 064501. https://doi.org/10.1115/1.4001600.

[Aithal 2015] Aithal S, Siva Prasad N, Shunmugan MS, Chellapandi P. (2015). Effect of manufacturing errors on load distribution in large diameter slewing bearings of fast breeder reactor rotatable plugs. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 230(9), 1449–60. https://doi.org/10.1177/0954406215579947.

[Alonso 2007] Alonso A, Giménez JG. (2007). Tangential problem solution for non-elliptical contact areas with the FastSim algorithm. Vehicle System Dynamics 45(4), 341–57. http://dx.doi.org/10.1080/00423110600999763.

[Alonso 2008] Alonso A, Giménez JG. (2008). Wheel–rail contact: roughness, heat generation and conforming contact influence. Tribology International 41, 755–68. https://doi.org/10.1016/j.triboint.2008.01.004.

[Alves Costa 2012] Alves Costa P, Calçada R, Silva Cardoso A. (2012). Track–ground vibrations induced by railway traffic: In-situ measurements and validation of a 2.5D FEM-BEM model. Soil Dynamics and Earthquake Engineering 32, 111–28. https://dx.doi.org/10.1016/j.soildyn.2011.09.002.

[Amasorrain 2003] Amasorrain JI, Sagartzazu X, Damián J. (2003). Load distribution in a four contact-point slewing bearing. Mechanism and Machine Theory 38(6), 479–96. https://doi.org/10.1016/S0094-114X(03)00003-X.

[ANSYS doc 2010] ANSYS Mechanical. Documentation, release 13.0. ANSYS Inc.; 2010. http://www.ansys.com.

[Antes 1992] Antes H, Panagiotopoulos PD. The Boundary Integral Approach to Static and Dynamic Contact Problems. Equality and Inequality Methods. Springer Basel AG; 1992. ISBN: 978-0-8176-2592-4.

[Ayasse 2005] Ayasse JB, Chollet H. (2005). Determination of the wheel rail contact patch in semi-Hertzian conditions. Vehicle System Dynamics 43, 161–72. http://dx.doi.org/10.1080/00423110412331327193.

[Ayasse 2006] Ayasse JB, Chollet H. Wheel–rail Contact. In: Iwnicki S (Ed), Handbook of Railway Vehicle Dynamics. Boca Raton, USA: Taylor & Francis; 2006. https://doi.org/10.1201/9781420004892.

[Baeza 2006] Baeza L, Roda A, Nielsen JCO. (2006). Railway vehicle/track interaction analysis using a modal substructuring approach. Journal of Sound and Vibration 293, 112–24. https://doi.org/10.1016/j.jsv.2005.09.006.

[Baeza 2007] Baeza L, Fuenmayor FJ, Carballeira J, Roda A. (2007). Influence of the wheel–rail contact instationary process on contact parameters. Journal of Strain Analysis 42, 377–87. https://doi.org/10.1243/03093247JSA247.

[Baeza 2011] Baeza L, Vila P, Xie G, Iwnicki SD. (2011). Prediction of rail corrugation using a rotating flexible wheelset coupled with a flexible track model and a non-Hertzian/non-steady contact model. Journal of Sound and Vibration 330, 4493–507. http://dx.doi.org/10.1016/j.jsv.2011.03.032.

[Baeza 2018] Baeza L, Thompson DJ, Squicciarini G, Denia FD. (2018). Method for obtaining the wheel–rail contact location and its application to the normal problem calculation through 'CONTACT'. Vehicle System Dynamics 56(11), 1734–46. https://doi.org/10.1080/00423114.2018.1439178.

[Barber 2002] Barber JR. Elasticity. 2nd edition. Kluwer Academic Publishers; 2002. https://dx.doi.org/10.1007/0-306-48395-5.

[Bashkar 1997] Bashkar A, Johnson KL, Wood GD, Woodhouse J. Wheel–rail dynamics with closely conformal contact. Part 1: dynamic modelling and stability analysis. (1997). Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 211(1), 11–26. https://doi.org/10.1243/0954409971530860.

[Bass 1987] Bass JM. (1987). Three-dimensional finite deformation, rolling contact of a hyperelastic cylinder: formulation of the problem and computational results. Computers and Structures 26(6), 991–1004. https://doi.org/10.1016/0045-7949(87)90116-7.

[Behnke 2015] Behnke R, Kaliske M. (2015). Thermo-mechanically coupled investigation of steady state rolling tires by numerical simulation and experiment. International Journal of Non-Linear Mechanics 68, 101–31. https://doi.org/10.1016/j.ijnonlinmec.2014.06.014.

[Bezin 2009] Bezin Y, Iwnicki SD, Cavalletti M, de Vries E, Shahzad F, Evans G. (2009). An investigation of sleeper voids using a flexible track model integrated with railway multi-body dynamics. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 223, 597–607. https://doi.org/10.1243/09544097JRRT276.

[Bezin 2010] Bezin Y, Farrington D, Penny C, Temple B, Iwnicki SD. (2010). The dynamic response of slab track constructions and their benefit with respect to conventional ballasted track. Vehicle System Dynamics 48(suppl.), 175–93. https://doi.org/10.1080/00423111003693201.

[Bian 2013] Bian J, Gu Y, Murray MH. (2013). A dynamic wheel–rail impact analysis of railway track under wheel flat by finite element analysis. Vehicle System Dynamics 51, 784–97. http://dx.doi.org/10.1080/00423114.2013.774031.

[Blanco-Lorenzo 2016] Blanco-Lorenzo J, Santamaria J, Vadillo EG, Correa N. (2016). On the influence of conformity on wheel-rail rolling contact mechanics. Tribology International 103, 647–67. https://dx.doi.org/10.1016/j.triboint.2016.07.017.

[Blanco-Lorenzo 2018] Blanco-Lorenzo J, Santamaria J, Vadillo EG, Correa N. (2018). A contact mechanics study of 3D frictional conformal contact. Tribology International 119, 143–56. https://doi.org/10.1016/j.triboint.2017.10.022.

[Blanco-Lorenzo 2021] Blanco-Lorenzo J, Vollebregt EAH, Santamaria J, Vadillo EG. (2021). Approximating the influence coefficients of non-planar elastic solids for conformal contact analysis. Tribology International 154, 106671. https://doi.org/10.1016/j.triboint.2020.106671.

[Boccini 2016] Boccini E, Marini L, Meli E, Rindi A, Romani M. Development of an innovative wheel–rail conformal contact model. In: Pombo J (Ed.), Proceedings of the Third International Conference on Railway Technology: Research, Development and Maintenance. Cagliari, Italy, 5–8 April, 2016.

[Bomidi 2014] Bomidi JAR, Sadeghi F. (2014). Three-dimensional finite element elastic-plastic model for subsurface initiated spalling in rolling contacts. Journal of Tribology 136, 011402. https://doi.org/10.1115/1.4025841.

[Boussinesq 1885] Boussinesq J. Application des potentiels à l'étude de l'équilibre et du mouvement des solides élastiques. Paris, Gauthier, Villars; 1885. ISBN: 978-0243352630.

[Brandt 1990] Brandt A, Lubrecht AA. (1990). Multilevel matrix multiplication and fast solution of integral equations. Journal of Computational Physics 90, 348–70. https://doi.org/10.1016/0021-9991(90)90171-V.

[Brigham 1988] Brigham EO. The fast Fourier transform and its applications. New Jersey, USA: Prentice-Hall; 1988. ISBN: 0-13-307505-2.

[Burgelman 2014] Burgelman N, Li Z, Dollevoet R. (2014). A new rolling contact method applied to conformal contact and the train–turnout interaction. Wear 321, 94–105. https://doi.org/10.1016/j.wear.2014.10.008.

[Burmister 1945] Burmister DM. (1945). The general theory of stresses and displacements in layered systems. I–III. Journal of Applied Physics 16, 89–94; 126–7; 296–302. https://doi.org/10.1063/1.1707558; https://doi.org/10.1063/1.1707562; https://doi.org/10.1063/1.1707590.

[Carter 1926] Carter FW. (1926). On the action of a locomotive driving wheel. Proceedings of the Royal Society of London, Series A 112, 151–7. https://doi.org/10.1098/rspa.1926.0100.

[Cattaneo 1938] Cattaneo C. (1938). Sul contatto di due corpi elastici, distribuzione locale degli sforzi. I–III. Rend. Accad. Naz. dei Lincei, Ser. 6, fol. 27; 342–8; 434–6; 474–8.

[Cerruti 1882] Cerruti V. (1882). Roma, Acc. Lincei, Mem. fis. mat.

[Chen 1988] Chen HH, Marshek KM. (1988). Effect of clearance and material property on contact pressure in two-dimensional conforming cylinders. Mechanism and Machine Theory 23(1), 55–62. https://doi.org/10.1016/0094-114X(88)90009-2.

[Chen 1971] Chen WT. (1971). Computation of stresses and displacements in a layered elastic medium. International Journal of Engineering Science 9, 775–800. https://doi.org/10.1016/0020-7225(71)90072-3.

[Chiu 1963] Chiu YP, Weinstein AS, Zorowski CF. (1963). The stresses, deformations and bending of a simply supported cylinder subjected to a symmetric normal surface loading. International Journal of Mechanical Sciences 5, 119–36. https://doi.org/10.1016/0020-7403(63)90017-1.

[Chiu 1964] Chiu YP, Weinstein AS, Zorowski CF. (1964). Elastic behaviour of a cylinder subjected to a biaxially symmetric normal surface loading. Quarterly Journal of Mechanics and Applied Mathematics 17(2), 199–215. https://doi.org/10.1093/qjmam/17.2.199.

[Chiu 1987] Chiu YP, Hartnett MJ. (1987). A numerical solution for the contact problem involving bodies with cylindrical surface considering cylinder effect. Journal of Tribology 109, 479–85. https://doi.org/10.1115/1.3261478.

[Chongyi 2010] Chongyi C, Chengguo W, Ying J. (2010). Study on numerical method to predict wheel/rail profile evolution due to wear. Wear 269, 167–73. http://dx.doi.org/ 10.1016/j.wear.2009.12.031.

[Chu 2000] Chu E, George A. Inside the FFT black box. Serial and parallel fast Fourier transform algorithms. Boca Raton, USA: CRC Press; 2000. ISBN: 0-8493-0270-6.

[Ciavarella 2001a] Ciavarella M, Decuzzi P. (2001). The state of stress induced by the plane frictionless cylindrical contact. I. The case of elastic similarity. International Journal of Solids and Structures 38, 4507–23. https://doi.org/10.1016/S0020-7683(00)00289-4.

[Ciavarella 2001b] Ciavarella M, Decuzzi P. (2001). The state of stress induced by the plane frictionless cylindrical contact. II. The general case (elastic dissimilarity). International Journal of Solids and Structures 38, 4525–33. https://doi.org/10.1016/S0020-7683(00)00290-0.

[Ciavarella 2006] Ciavarella M, Baldini A, Barber JR, Strozzi A. (2006). Reduced dependence on loading parameters in almost conforming contacts. International Journal of Mechanical Sciences 48, 917–25. https://doi.org/10.1016/j.ijmecsci.2006.03.016.

[CONTACT UG 2013] Vollebregt EAH. User guide for CONTACT version 13.1. Delft: VORtech BV; 2013. http://www.kalkersoftware.org/.

[Cooper 1968] Cooper DH. (1968). Tables of Hertzian contact-stress coefficients. Report R-387. Coordinated Science Laboratory. University of Illinois, Urbana.

[Correa 2012] Correa N, Vadillo EG, Santamaria J, Gómez J. (2012). A rational fraction polynomials model to study vertical dynamic wheel–rail interaction. Journal of Sound and Vibration 331, 1844–58. https://doi.org/10.1016/j.jsv.2011.12.012.

[Correa 2015] Correa N, Vadillo EG, Santamaria J, Blanco-Lorenzo J. A comprehensive study of the influence of rail welds and vehicle speed on wheel–rail forces. In: Proceedings of the 10th International Conference on Contact Mechanics and Wear of Rail/Wheel Systems (CM2015), Colorado Springs, USA, 30 August – 3 September, 2015.

[Correa 2018a] Correa N, Vadillo EG, Santamaria J, Blanco-Lorenzo J. (2018). On the non-proportionality between wheel/rail contact forces and speed during wheelset passage over specific welds. Journal of Sound and Vibration 413, 79–100. https://doi.org/10.1016/j.jsv.2017.10.031.

[Correa 2018b] Correa N, Vadillo EG, Santamaria J, Blanco-Lorenzo J. Una explicación a la disminución de las fuerzas de contacto rueda/carril con la velocidad en los defectos de soldadura. XXII Congreso Nacional de Ingeniería Mecánica, Madrid, 19–21 Septiembre 2018.

[Croft 2011] Croft B, Jones C, Thompson D. (2011). Velocity-dependent friction in a model of wheel–rail rolling contact and wear. Vehicle System Dynamics 49(11), 1791–802. https://dx.doi.org/10.1080/00423114.2010.543138.

[Damme 2003] Damme S, Nackenhorst U, Wetzel A, Zastrau BW. On the numerical analysis of the wheel–rail system in rolling contact. In: Popp K, Schielen W (Eds.), System Dynamics and Long-term Behavior of Railway Vehicles, Track and Subgrade, Springer, Berlin, Heidelberg, 2003, 155–74. https://doi.org/10.1007/978-3-540-45476-2_10.

[Di Gialleonardo 2012] Di Gialleonardo E, Braghin F, Bruni S. (2012). The influence of track modelling options on the simulation of rail vehicle dynamics. Journal of Sound and Vibration 331, 4246–58. https://dx.doi.org/10.1016/j.jsv.2012.04.024.

[Dietz 2002] Dietz S, Hippman G, Schupp G. (2002). Interaction of vehicle and flexible tracks by co-simulation of multibody vehicle systems and finite element track models. Vehicle System Dynamics 37, 372–84. https://doi.org/10.1080/00423114.2002.11666247.

[Donea 2004] Donea J, Huerta A, Ponthot JP, Rodríguez Ferran A. Arbitrary Lagrangian–Eulerian Methods. In: Stein E, Borst R, Hughest T (Eds.), Encyclopedia of computational mechanics, vol. 1: fundamentals. Wiley, New York, 2004, 413–37. https://doi.org/10.1002/0470091355.ecm009.

[Draganis 2015] Draganis A, Larsson F, Ekberg A. (2015). Finite element modelling of frictional thermomechanical rolling/sliding contact using an arbitrary Lagrangian–Eulerian formulation. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology 229(7), 870–80. https://doi.org/10.1177/1350650115572197.

[Draganis 2017] Draganis A. (2017). Finite element modelling of transient thermomechanical rolling contact featuring mixed control of the rigid body motion. Journal of Tribology 139(1), 011503. https://doi.org/10.1115/1.4033048.

[Dundurs 1970] Dundurs J, Stippes M. (1970). Role of elastic constants in certain contact problems. Journal of Applied Mechanics 37(4), 965–70. https://doi.org/10.1115/1.3408725.

[Dydo 1995] Dydo JR, Busby HR. (1995). Elasticity solutions for constant and linearly varying loads applied to a rectangular surface patch on the elastic half-space. Journal of Elasticity 38, 153–63. https://doi.org/10.1007/BF00042496.

[Ekberg 2002] Ekberg A, Kabo E, Anderson H. (2002). An engineering model for prediction of rolling contact fatigue of railway wheels. Fatigue and Fracture of Engineering Materials and Structures 25, 899–909. https://doi.org/10.1046/j.1460-2695.2002.00535.x.

[Ekberg 2005] Ekberg A, Kabo E. (2005). Fatigue of railway wheels and rails under rolling contact and thermal loading–an overview. Wear 258, 1288–300. https://doi.org/10.1016/j.wear.2004.03.039.

[EN 13001-3-3 2014] EN 13001-3-3. (2014). Cranes. General design. Part 3-3: Limit states and proof of competence of wheel/rail contacts. CEN.

[EN 14363 2017] EN 14363. (2017). Railway applications. Testing and simulation for the acceptance of running characteristics of railway vehicles. Running Behaviour and stationary tests. CEN.

[ERRI B 176/3 1993] ERRI B 176/3. Benchmark problem results and assessment. B176/DT290. UIC Railway Publications, Utrecht, The Netherlands, 1993.

[Fagan 2001] Fagan MJ, McConnachie J. (2001). A review and detailed examination of non-layered conformal contact by finite element analysis. Journal of Strain Analysis 36(2), 177–95. https://dx.doi.org/10.1243/0309324011512739.

[Fallahnezhad 2019] Fallahnezhad K, Liu S, Brinji O, Marker M, Meehan PA. (2019). Monitoring and modelling of false brinelling for railway bearings. Wear 424–5, 151–64. https://doi.org/10.1016/j.wear.2019.02.004.

[Faria 1989] Faria LO, Bass JM, Oden JT, Becker EB. (1989). A three-dimensional rolling contact model of a reinforced rubber tire. Tire Science and Technology 17(3), 217–33. https://doi.org/10.2346/1.2141686.

[Faria 1992] Faria LO, Oden JT, Yavari B, Tworzydlo W, Bass JM, Becker EB. (1992). Tire modeling by finite elements. Tire Science and Technology 20(1), 33–56. https://doi.org/10.2346/1.2139507.

[Flamant 1892] Flamant. Compt. Rendus, 114, 1465, Paris. 1892.

[Fromm 1926] Fromm H. (1926–27). Berechnung des Schlupfes beim Rollen deformierbaren Scheiben (English title: Analysis of creep of deformable discs during rolling), Zeitschrift für Angewandte Mathematik und Mechanik 7, 27–58 (Dissertation TH Berlin, 1926).

[Galvín 2010a] Galvín P, Romero A, Domínguez J. (2010). Fully three-dimensional analysis of high-speed train-track-soil-structure dynamic interaction. Journal of Sound and Vibration 329, 5147–63. https://dx.doi.org/10.1016/j.jsv.2010.06.016.

[Galvín 2010b] Galvín P, Romero A, Domínguez J. (2010). Vibrations induced by HST passage on ballast and non-ballast tracks. Soil Dynamics and Earthquake Engineering 30, 862–73. https://dx.doi.org/10.1016/j.soildyn.2010.02.004.

[Gao 2022] Gao Y, Xu JM, Liu YB, Dong ZG, Wang P, Jiang ZQ. (2022). An investigation into transient frictional rolling contact behaviour in a switch panel: validation an numerical simulation. Vehicle System Dynamics 60(1), 114–31. https://doi.org/10.1080/00423114.2020.1802492.

[Garg 1984] Garg VK, Dukkipati RV. Dynamics of railway vehicle systems. Toronto, Canada: Academic Press; 1984. ISBN: 0-12-275950-8.

[Giménez 2005] Giménez JG, Alonso A, Gómez E. (2005). Introduction of a friction coefficient dependent on the slip in the FastSim algorithm. Vehicle System Dynamics 43(4), 233–44. https://dx.doi.org/10.1080/00423110412331282913.

[Gonzalez 2008] Gonzalez FJ, Suarez B, Paulin J, Fernandez I. (2008). Safety assessment of underground vehicles passing over highly resilient straight track in the presence of a broken rail. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 222, 69–84. https://doi.org/10.1243/09544097JRRT128.

[González 1998] González JA, Abascal R. (1998). Using the boundary element method to solve rolling contact problems. Engineering Analysis with Boundary Elements 21, 385–92. https://doi.org/10.1016/S0955-7997(98)00026-5.

[González 2000] González JA, Abascal R. (2000). An algorithm to solve coupled 2D rolling contact problems. International Journal for Numerical Methods in Engineering 49, 1143–67. https://doi.org/10.1002/1097-0207(20001130)49:9<1143::AID-NME991>3.0.CO;2-B.

[Goodman 1965] Goodman LE, Keer LM. (1965). The contact stress problem for an elastic sphere indenting an elastic cavity. International Journal of Solids and Structures 1(4), 407–15. https://doi.org/10.1016/0020-7683(65)90005-3.

[Grassie 2009] Grassie SL. (2009). Rail corrugation: characteristics, causes, and treatments. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 264, 581–96. https://doi.org/10.1243/09544097JRRT264.

[Guiral 2013] Guiral A, Alonso A, Baeza L, Giménez JG. (2013). Non-steady state modelling of wheel–rail contact problem. Vehicle System Dynamics 51(1), 91-108. https://dx.doi.org/10.1080/00423114.2012.713499.

[Guyan 1965] Guyan RJ. (1965). Reduction of stiffness and mass matrices. AIAA Journal 3(2), 381. https://doi.org/10.2514/3.2874.

[Haines 1963] Haines DJ, Ollerton E. (1963). Contact stress distributions on elliptical contact surfaces subjected to radial and tangential forces. Proceedings of the Institution of Mechanical Engineers 177(4), 95–114. https://doi.org/10.1243/PIME_PROC_1963_177_014_02.

[Halling 1964] Halling J. (1964). Microslip between a rolling element and its track arising from geometric conformity and applied surface tractions. Journal of Mechanical Engineering Science 6(1), 64–73. https://doi.org/10.1243/JMES_JOUR_1964_006_012_02.

[Halling 1966–67] Halling J. (1966–67). Analysis of spin/roll conditions and the frictional-energy dissipation in angular-contact thrust ball bearings. Proceedings of the Institution of Mechanical Engineers 181(16), 349–62. https://doi.org/10.1243/PIME_PROC_1966_181_035_02.

[Harris 2001] Harris TA. Rolling bearing analysis. 4th edition. New York, USA: John Wiley & Sons, Inc.; 2001. ISBN: 0-471-35457-0.

[Hartnett 1979] Hartnett MJ. (1979). The analysis of contact stresses in rolling element bearings. Journal of Lubrication Technology 101, 105–9. https://doi.org/10.1115/1.3453270.

[Heras 2017a] Heras I, Aguirrebeitia J, Abasolo M, Plaza J. (2017). Friction torque in four-point contact slewing bearings: Applicability and limitations of current analytical formulations. Tribology International 115, 59–69. https://doi.org/10.1016/j.triboint.2017.05.011.

[Heras 2017b] Heras I, Aguirrebeitia J, Abasolo M. Friction torque in four contact point slewing bearings: Effect of manufacturing errors and ring stiffness. Mechanism and Machine Theory 112, 145–54. https://doi.org/10.1016/j.mechmachtheory.2017.02.009.

[Heras 2018] Heras I. Four-point contact slewing bearings for wind turbines: advances in structural modelling and friction torque calculation. Ph.D. thesis, University of the Basque Country, Bilbao, Spain. 2018.

[Herting 2004] Herting DN. The MSC/NASTRAN Advanced Dynamic Analysis User's Guide, version 70. 2004.

[Hertz 1882] Hertz H. (1882). Über die Berührung fester elastischer Körper (On the contact of elastic solids). J für reine und angewandte Mathematik 92, 156–71. (English translation in Miscellaneous Papers by H. Hertz, Eds. Jones and Schott, London: Macmillan, 1896.) https://doi.org/10.1515/crll.1882.92.156.

[Hetényi 1970] Hetényi M. (1970). A general solution for the elastic quarter space. Journal of Applied Mechanics 37(1), 70–6. https://doi.org/10.1115/1.3408492.

[Hills 1993] Hills DA, Nowell D, Sackfield A. Mechanics of elastic contacts. Oxford, UK: Butterworth-Heinemann; 1993. ISBN: 978-0-7506-0540-3. https://doi.org/10.1016/C2009-0-24029-3.

[Hobbs 1967] Hobbs AEH. (1967). A survey of creep. Brit Rail Res Dept. Rept DYN 52, Derby (UK).

[Hou 2001] Hou JP, Hills DA. (2001). Contact between a pin and a plate with a hole under interference-fit and clearance-fit conditions. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 215(6), 629–39. https://doi.org/10.1243/0954406011524009.

[Houpert 1999] Houpert L. Numerical and analytical calculations in ball bearings. In: Proceedings of the 8th European Space Mechanisms and Tribology Symposium. Toulouse, France, 1999.

[Houpert 2002] Houpert L. (2002). Ball bearing and tapered roller bearing torque: analytical, numerical and experimental results. Tribology Transactions 45(3), 345–53. https://doi.org/10.1080/10402000208982559.

[Hu 2002] Hu H, Wriggers P. (2002). On the adaptive finite element method of steady-state rolling contact for hyperelasticity in finite deformations. Computer Methods in Applied Mechanics and Engineering 191:1333–48. https://doi.org/10.1016/S0045-7825(01)00326-7.

[Hussein 2006] Hussein MFM, Hunt HEM. (2006). Modelling of floating-slab tracks with continuous slabs under oscillating moving loads. Journal of Sound and Vibration 297, 37–54. https://doi.org/10.1016/j.jsv.2006.03.026.

[Jeffrey 2008] Jeffrey A, Dai H-H. Handbook of Mathematical Formulas and Integrals. 4th edition. USA: Elsevier Academic Press; 2008. ISBN: 978-0-12-374288-9.

[Jin 2005] Jin X, Wen Z, Zhang W, Shen Z. (2005). Numerical simulation of rail corrugation on a curved track. Computers and Structures 83, 2052–65. http://dx.doi.org/10.1016/j.compstruc.2005.03.012.

[Jin 2008] Jin X, Xiao X, Wen Z, Zhou Z. (2008). Effect of sleeper pitch on rail corrugation at a tangent track in vehicle hunting. Wear 265, 1163–75. https://doi.org/10.1016/j.wear.2008.01.028.

[Johnson 1958a] Johnson KL. (1958). The effect of spin upon the rolling motion of an elastic sphere on a plane. Journal of Applied Mechanics 25(3), 332–8. https://doi.org/10.1115/1.4011822.

[Johnson 1958b] Johnson KL. (1958). The effect of a tangential contact force upon the rolling motion of an elastic sphere on a plane. Journal of Applied Mechanics 25(3), 339–46. https://doi.org/10.1115/1.4011823.

[Johnson 1987] Johnson KL. Contact mechanics. Cambridge, UK: Cambridge University Press; 1987. https://doi.org/10.1017/CBO9781139171731.

[Johnson 2001] Johnson KL. Plastic deformation in rolling contact. In: Jacobson B & Kalker JJ (Eds.). Rolling Contact Phenomena, CISM Courses and Lectures, No. 411, Chapter 3, pp. 163–202. Wien, New York: Springer; 2001. https://doi.org/10.1007/978-3-7091-2782-7_3.

[Jones 1959] Jones AB. (1959). Ball motion and sliding friction in ball bearings. ASME Journal of Basic Engineering 81(1), 1–12. https://doi.org/10.1115/1.4008346.

[Jones 2000] Jones CJC, Sheng X, Petyt M. (2000). Simulations of ground vibration from a moving harmonic load on a railway track. Journal of Sound and Vibration 231(3), 739–51. https://doi.org/10.1006/jsvi.1999.2559.

[Ju 2000] Ju SH, Horng TL, Cha KC. (2000). Comparisons of contact pressures of crowned rollers. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology 214, 147–56. https://doi.org/10.1243/1350650001543061.

[Kabe 2000] Kabe K, Koishi M. (2000). Tire cornering simulation using finite element analysis. Journal of Applied Polymer Science 78, 1566–72. https://doi.org/10.1002/1097-4628(20001121)78:8<1566::AID-APP140>3.0.CO;2-I.

[Kaiser 2012] Kaiser I. (2012). Refining the modelling of vehicle-track interaction. Vehicle System Dynamics 50, 229-43. http://dx.doi.org/10.1080/00423114.2012.671948.

[Kalker 1967a] Kalker JJ. On the rolling contact of two elastic bodies in the presence of dry friction. Ph.D. dissertation, Technische Universiteit Delft, The Netherlands. 1967. http://resolver.tudelft.nl/uuid:aa44829b-c75c-4abd-9a03-fec17e121132.

[Kalker 1967b] Kalker JJ. (1967). A strip theory for rolling with slip and spin. Proc Kon Ned Akad van Wetenschappen; Mechanics, Series B,70(1), 10–62.

[Kalker 1972] Kalker JJ, Van Randen Y. (1972). A minimum principle for frictionless elastic contact with application to non-Hertzian half-space contact problems. Journal of Engineering Mathematics 6(2), 193-206. https://doi.org/10.1007/BF01535102.

[Kalker 1979a] Kalker JJ. (1979). Survey of wheel–rail rolling contact theory. Vehicle System Dynamics 5, 317–58. https://doi.org/10.1080/00423117908968610.

[Kalker 1979b] Kalker JJ. (1979). The computation of three-dimensional rolling contact with dry friction. International Journal for Numerical Methods in Engineering 14, 1293–307. https://doi.org/10.1002/nme.1620140904.

[Kalker 1982] Kalker JJ. (1982). A fast algorithm for the simplified theory of rolling contact. Vehicle System Dynamics 11, 1–13. https://doi.org/10.1080/00423118208968684.

[Kalker 1990] Kalker JJ. Three-dimensional elastic bodies in rolling contact, volume 2 of Solids Mechanics and its Applications. Dordrecht, The Netherlands: Kluwer Academic Publishers; 1990. ISBN: 0-7923-0712-7.

[Kalker 1991] Kalker JJ. (1991). Wheel–rail rolling contact theory. Wear 144, 243–61. https://doi.org/10.1016/0043-1648(91)90018-P.

[Kalker 1996] Kalker JJ. (1996). Book of tables for the Hertzian creep–force law. Reports of the Faculty of Technical Mathematics and Informatics 96(61). ISSN: 0922-5641.

[Kalker 2001] Kalker JJ. Rolling contact phenomena - linear elasticity. In: Jacobson B & Kalker JJ (Eds.). Rolling Contact Phenomena, CISM Courses and Lectures, No. 411, Chapter 1, pp. 1–84. Wien, New York: Springer; 2001. https://doi.org/10.1007/978-3-7091-2782-7_1.

[Karapetyan 2009] Karapetyan AV. (2009). A two-parameter friction model. Journal of Applied Mathematics and Mechanics 73, 367–70. https://doi.org/10.1016/j.jappmathmech.2009.08.016.

[Kik 1996] Kik W, Piotrowski J. A fast, approximate method to calculate normal load at contact between wheel and rail and creep forces during rolling. In: Zobory I (Ed.), Proceedings of the 2nd Mini Conf on Contact Mechanics and Wear of Rail/Wheel Systems, pp. 11–20. Budapest, Hungary, 1996.

[Kikuchi 1988] Kikuchi N, Oden JT. Contact problems in elasticity. A study of variational inequalities and finite element methods. Philadelphia, USA: SIAM; 1988. ISBN: 0-89871-468-0.

[Kireenkov 2008] Kireenkov AA. (2008). Combined model of sliding and rolling friction in dynamics of bodies on a rough plane. Mechanics of Solids 43(3), 412–25. https://doi.org/10.3103/S0025654408030138.

[Kleckner 1982] Kleckner RJ, Pirvics J. (1982). Spherical roller bearing analysis. Journal of Lubrication Technology 104, 99–108. https://doi.org/10.1115/1.3253170.

[Klingel 1883] Klingel. (1883). Uber den Lauf der Eisenbahnwagen auf Gerarder Bahn. Organ Fortsch. Eisenb-wes. 38, 113–23.

[Knothe 1993] Knothe KL, Grassie SL. (1993). Modeling of railway track and vehicle/track interaction at high frequencies. Vehicle System Dynamics 22, 209–62. https://doi.org/10.1080/00423119308969027.

[Knothe 2008] Knothe K. (2008). History of wheel/rail contact mechanics: from Redtenbacher to Kalker. Vehicle System Dynamics 46(1–2), 9–26. https://dx.doi.org/10.1080/00423110701586469.

[Kudra 2013] Kudra G, Awrejcewicz J. (2013). Approximate modelling of resulting dry friction forces and rolling resistance for elliptic contact shape. European Journal of Mechanics A/Solids 42, 358–75. https://doi.org/10.1016/j.euromechsol.2013.07.005.

[Kuo 2008] Kuo C, Huang C, Chen Y. (2008). Vibration characteristics of floating slab track. Journal of Sound and Vibration 317, 1017–34. https://doi.org/10.1016/j.jsv.2008.03.051.

[Landau 1986] Landau LD, Lifshitz EM. Theory of elasticity. 3rd edition. Pergamon Press; 1986. ISBN: 0-08-033916-6.

[Laursen 2003] Laursen TA. Computational contact and impact mechanics. Berlin, Germany: Springer-Verlag; 2003. ISBN: 978-3-662-04864-1. https://doi.org/10.1007/978-3-662-04864-1.

[Laursen 2006] Laursen TA. Stanciulescu I. (2006). An algorithm for incorporation of frictional sliding conditions within a steady state rolling framework. Communications in Numerical Methods in Engineering 22, 301–18. https://doi.org/10.1002/cnm.815.

[Leblanc 2007] Leblanc A, Nelias D. (2007). Ball motion and sliding friction in a four-contact-point ball bearing. Journal of Tribology 129, 801–8. https://doi.org/10.1115/1.2768079.

[Legrand 1995] Legrand E, Mondier JB. Estimation of friction torque in solid lubricated ball bearings. In: Proceedings of the 6th European Space Mechanisms and Tribology Symposium. Zürich, Switzerland, 4–6 October, 1995.

[Leine 2003] Leine RI, Glocker Ch. (2003). A set-valued force law for spatial Coulomb–Contensou friction. European Journal of Mechanics A/Solids 22, 193–216. https://doi.org/10.1016/S0997-7538(03)00025-1.

[Lewis 2004] Lewis R, Olofsson U. (2004). Mapping rail wear regimes and transitions. Wear 257, 721–9. https://dx.doi.org/10.1016/j.wear.2004.03.019.

[Le Tallec 1994] Le Tallec P, Rahier C. (1994). Numerical models of steady rolling for non-linear viscoelastic structures in finite deformations. International Journal for Numerical Methods in Engineering 37, 1159–86. https://doi.org/10.1002/nme.1620370705.

[Li F 2018] Li F, Hu W, Meng Q, Zhan Z, Shen F. (2018). A new damage-mechanics-based model for rolling contact fatigue analysis of cylindrical roller bearing. Tribology International 120, 105–14. https://doi.org/10.1016/j.triboint.2017.12.001.

[Li J 2001] Li J, Berger EJ. (2001). A Boussinesq-Cerruti solution set for constant and linear distribution of normal and tangential load over a triangular area. Journal of Elasticity 63, 137–51. http://dx.doi.org/10.1023/A:1014013425423.

[Li Z 2002] Li Z. Wheel-Rail Rolling Contact and Its Application to Wear Simulation. Ph.D. dissertation, Technische Universiteit Delft, The Netherlands. 2002.

[Li Z 2008] Li Z, Zhao X, Esveld C, Dollevoet R, Molodova M. (2008). An investigation into the causes of squats—Correlation analysis and numerical modelling. Wear 265, 1349–55. https://doi.org/10.1016/j.wear.2008.02.037.

[Lim 1987] Lim S, Ashby M. (1987). Wear-mechanism maps. Acta Metallurgica 35(1), 1–24. https://doi.org/10.1016/0001-6160(87)90209-4.

[Lin 2021] Lin C-L, Meehan PA. (2021). Morphological and elemental analysis of wear debris naturally formed in grease lubricated railway axle bearings. Wear 484–5, 203994. https://doi.org/10.1016/j.wear.2021.203994.

[Liu 2000] Liu S, Wang Q, Liu G. (2000). A versatile method of discrete convolution and FFT (DC-FFT) for contact analyses. Wear 243, 101–11. https://doi.org/10.1016/S0043-1648(00)00427-0.

[Liu 2012] Liu S, Chen WW. (2012). Two-dimensional numerical analyses of double conforming contacts with effect of curvature. International Journal of Solids and Structures 49, 1365–74. https://doi.org/10.1016/j.ijsolstr.2012.02.019.

[Liu 2013] Liu S. Numerical simulation of double conformal contacts involving both interference and clearance. (2013). Tribology Transactions 56(5), 867–78. https://doi.org/10.1080/10402004.2013.806686.

[Lombaert 2006] Lombaert G, Degrande G, Vanhauwere B, Vandeborght B, François S. (2006). The control of ground-borne vibrations from railway traffic by means of continuous floating slabs. Journal of Sound and Vibration 297, 946–61. https://doi.org/10.1016/j.jsv.2006.05.013.

[Love 1906] Love AEH. A treatise on the mathematical theory of elasticity. 2nd edition. Cambridge University Press; 1906.

[Magel 2017] Magel EE. (2017). A survey of wheel/rail friction. National Research Council, Canada. Report no DOT/FRA/ORD-17/21.

[Marini 2022] Marini L, Meli E, Rindi A. Development of a conformal contact model for railway application. In: The Fifth International Conference on Railway Technology: Research, Development and Maintenance. Montpellier, France, 22–25 August, 2022.

[Marques 2018] Marques F, Magalhães H, Liu B, Pombo J, Flores P, Ambrósio J, Bruni S. A new simplified approach to deal with conformal contact in railway dynamics. In: The 5th Joint International Conference on Multibody System Dynamics. Lisbon, Portugal, 24–28 June, 2018.

[Marshek 1984] Marshek KM, Chen HH. Numerical solution for contact pressure distributions in misaligned self-lubricated journal bearings. SAE Technical Paper Series, 841125; International Off-Highway & Powerplant Congress & Exposition, Milwaukee, Wisconsin. September 10-13, 1984. https://doi.org/10.4271/841125.

[MATLAB www] https://www.mathworks.com

[Meehan 2017] Meehan PA, Milne CD, Liu S. Investigation of wear degradation of railway spherical roller bearings. In Proceedings of the 25th International Symposium on Dynamics of Vehicles on Roads and Tracks (IAVSD 2017). Rockhampton, Australia, 14–18 August, 2017. https://doi.org/10.1201/9781351057189.

[Melis 2008] Melis M. Apuntes de introducción a la dinámica vertical de la vía y a las señales digitales en ferrocarriles. Escuela de Ingenieros de Caminos, Canales y Puertos, Universidad Politécnica de Madrid. 2008. ISBN: 9788461276868.

[Meymand 2016] Meymand SZ, Keylin A, Ahmadian M. (2016). A survey of wheel–rail contact models for rail vehicles. Vehicle System Dynamics 54(3), 386–428. http://dx.doi.org/10.1080/00423114.2015.1137956.

[Muskhelishvili 1977] Muskhelishvili NI. Some basic problems of the mathematical theory of elasticity. Reprint of the second edition of English translation by J.R.M. Radok. Dordrecht, The Netherlands: Noordhoff; 1977. ISBN: 978-94-017-3034-1. https://doi.org/10.1007/978-94-017-3034-1.

[Nackenhorst 1993] Nackenhorst U. On the finite element analysis of steady state rolling contact. In: Aliabadi MH, Brebbia CA (Eds.), Contact Mechanics—Computational Techniques, Computational Mechanics Publication, Southampton, Boston, 1993, 241–8.

[Nackenhorst 2000] Nackenhorst U, Zastrau BW, Jarewski J. (2000). Finite element modelling of 3D elastic–elastic rolling contact. Z Angew Math Mech 80(suppl. 1), 57–60. https://doi.org/10.1002/zamm.20000801315.

[Nackenhorst 2004] Nackenhorst U. (2004). The ALE-formulation of bodies in rolling contact. Theoretical foundations and finite element approach. Computer Methods in Applied Mechanics and Engineering 193, 4299–322. https://doi.org/10.1016/j.cma.2004.01.033.

[Nielsen 2003] Nielsen JCO, Lunden R, Johansson A, Vernersson T. (2003). Train–track interaction and mechanisms of irregular wear on wheel and rail surfaces. Vehicle System Dynamics 40(1), 3–54. https://dx.doi.org/10.1076/vesd.40.1.3.15874.

[Nogi 1997] Nogi T, Kato T. (1997). Influence of a hard surface layer on the limit of elastic contact−Part 1: analysis using a real surface model. Journal of Tribology 119, 493–500. https://doi.org/10.1115/1.2833525.

[Nyqvist 2012] Nyqvist J. Three-dimensional analysis of multilayered rough surface contact. Ph.D. dissertation, Imperial College, London, UK. 2012.

[Nyqvist 2015] Nyqvist J, Kadiric A, Ioannides S, Sayles R. (2015). Semi-analytical model for rough multilayered contacts. Tribology International 87, 98–112. https://doi.org/10.1016/j.triboint.2015.01.006.

[Oden 1986] Oden JT, Lin TL. (1986). On the general rolling contact problem for finite deformations of a viscoelastic cylinder. Computer Methods in Applied Mechanics and Engineering 57, 297–367. https://doi.org/10.1016/0045-7825(86)90143-X.

[Oden 1988] Oden JT, Lin TL, Bass JM. A finite element analysis of the general rolling contact problem for viscoelastic rubber cylinder. Tire Science and Technology 16(1), 18–43. https://doi.org/10.2346/1.2148795.

[Oh 1976] Oh KP, Trachman EG. (1976). A numerical procedure for designing profiled rolling elements. Journal of Lubrication Technology 98, 547–51. https://doi.org/10.1115/1.3452922.

[Oloffson 2000] Oloffson U, Andersson S, Björklund S. (2000). Simulation of mild wear in boundary lubricated spherical roller thrust bearings. Wear 241, 180–5. https://doi.org/10.1016/S0043-1648(00)00373-2.

[Otero 2009] Otero J, Martínez J, de los Santos MA. Evaluación de las vibraciones generadas al paso de un tren considerando diferentes tipologías de vía. In: 9º Congreso Iberoamericano de Ingeniería Mecánica, 2009.

[O'Sullivan 1988] O'Sullivan TC, King RB. (1988). Sliding contact stress field due to a spherical indenter on a layered elastic half-space. Journal of Tribology 110, 235–40. https://doi.org/10.1115/1.3261591.

[Oyarzabal 2009] Oyarzabal O, Gomez J, Santamaria J, Vadillo EG. (2009). Dynamic optimization of track components to minimize rail corrugation. Journal of Sound and Vibration 319, 904–17. https://doi.org/10.1016/j.jsv.2008.06.020.

[Padovan 1984] Padovan J, Tovichakchaikul S, Zeid I. (1984). Finite element analysis of steadily moving contact fields. Computers and Structures 18(2), 191–200. https://doi.org/10.1016/0045-7949(84)90119-6.

[Padovan 1987] Padovan J. (1987). Finite element analysis of steady and transiently moving/rolling nonlinear viscoelastic structure—I. Theory. Computers and Structures 27(2), 249–57. https://doi.org/10.1016/0045-7949(87)90093-9.

[Panagiotopoulos 1975] Panagiotopoulos PD. (1975). A nonlinear programming approach to the unilateral contact-, and friction-boundary value problem in the theory of elasticity. Ingenieur-Archiv 44, 421–32. https://doi.org/10.1007/BF00534623.

[Pascal 2007a] Pascal J-P, Jourdan F. The “rigid multi-Hertzian method” as applied to conformal contacts. In: Proceedings of the ASME 2007 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, pp. 1811–25. Las Vegas, USA, 4–7 September, 2007. https://doi.org/10.1115/DETC2007-34379.

[Pascal 2007b] Pascal J-P, Zaazaa KE. A study of the effect of *m* and *n* coefficients of the Hertzian contact theory on railroad vehicle dynamics. In: Proceedings of the ASME 2007 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, pp. 1893–901. Las Vegas, USA, 4–7 September, 2007. https://doi.org/10.1115/DETC2007-34972.

[Pascal 2016] Pascal J-P, Soua B. (2016). Solving conformal contacts using multi-Hertzian techniques. Vehicle System Dynamics 54(6), 784–813. https://doi.org/10.1080/00423114.2016.1161201.

[Pascal 2019] Pascal J-P, Sani JR. (2019). Dynamics of an isolated railway wheelset with conformal wheel–rail interactions. Vehicle System Dynamics 57(12), 1947–69. https://doi.org/10.1080/00423114.2018.1557704.

[Paul 1979] Paul B, Hashemi J. (1979). Numerical determination of contact pressure between closely conforming wheels and rails. Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania. Report no FRA/ORD-79-41.

[Paul 1981] Paul B, Hashemi J. (1981). Contact pressures on closely conforming elastic bodies. Journal of Applied Mechanics 48, 543–48. https://doi.org/10.1115/1.3157671.

[Persson 1964] Persson A. On the stress distribution of cylindrical elastic bodies in contact. Ph.D. dissertation, Chalmers, Tekniska, Goteborg, Sweden. 1964. ISBN: 99-0510745-2.

[Pilkey 2005] Pilkey WD. Formulas for stress, strain, and structural matrices. 2nd edition. New Jersey, USA: John Wiley & Sons; 2005. ISBN 0-471-03221-2.

[Piotrowski 1999] Piotrowski J, Kik W. (1999). The influence of spin on creep forces for non-flat contact area. Vehicle System Dynamics 31(suppl.), 158–77.

[Piotrowski 2005] Piotrowski J, Chollet H. (2005). Wheel–rail contact models for vehicle system dynamics including multi-point contact. Vehicle System Dynamics 43(6–7), 455–83. https://dx.doi.org/10.1080/00423110500141144.

[Piotrowski 2008] Piotrowski J, Kik W. (2008). A simplified model of wheel/rail contact mechanics for non-Hertzian problems and its application in rail vehicle dynamic simulations. Vehicle System Dynamics 46(1–2), 27–48. http://dx.doi.org/10.1080/00423110701586444.

[Piotrowski 2017] Piotrowski J, Liu B, Bruni S. (2017). The Kalker book of tables for non-Hertzian contact of wheel and rail. Vehicle System Dynamics 55(6), 875–901. https://dx.doi.org/10.1080/00423114.2017.1291980.

[Pletz 2014a] Pletz M, Daves W, Yao W, Kubin W, Scheriau S. (2014). Multi-scale finite element modeling to describe rolling contact fatigue in a wheel–rail test rig. Tribology International 80, 147–55. http://dx.doi.org/10.1016/j.triboint.2014.07.005.

[Pletz 2014b] Pletz M, Daves W, Yao W, Ossberger H. (2014). Rolling contact fatigue of three crossing nose materials—Multiscale FE approach. Wear 314(1–2), 69–77. https://doi.org/10.1016/j.wear.2013.11.013.

[Polach 1999] Polach O. (1999). A fast wheel–rail forces calculation computer code. Vehicle System Dynamics 33(suppl.), 728–39. https://doi.org/10.1080/00423114.1999.12063125.

[Polach 2005] Polach O. (2005). Creep forces in simulations of traction vehicles running on adhesion limit. Wear 258, 992–1000. https://doi.org/10.1016/j.wear.2004.03.046.

[Polonsky 1999] Polonsky IA, Keer LM. (1999). A numerical method for solving rough contact problems based on the multi-level multi-summation and conjugate gradient techniques. Wear 231, 206–19. https://doi.org/10.1016/S0043-1648(99)00113-1.

[Polonsky 2000] Polonsky IA, Keer LM. (2000). Fast methods for solving rough contact problems: a comparative study. Journal of Tribology 122, 36–41. https://doi.org/10.1115/1.555326.

[Popov 2010] Popov VL. Contact mechanics and friction. Physical principles and applications. Berlin, Germany: Springer-Verlag; 2010. ISBN: 978-3-642-10803-7. https://doi.org/10.1007/978-3-642-10803-7.

[Popp 1999] Popp K, Kruse H, Kaiser I. (1999). Vehicle–track dynamics in the mid-frequency range. Vehicle System Dynamics 31, 423–64. https://dx.doi.org/10.1076/vesd.31.5.423.8363.

[Popp 2005] Popp K, Knothe K, Popper P. (2005). System dynamics and long-term behaviour of railway vehicles, track and subgrade: report on the DFG priority programme in Germany and subsequent research. Vehicle System Dynamics 43(6–7), 485–538. https://dx.doi.org/10.1080/00423110500143728.

[Pozzebon 2020] Pozzebon ML, Lin C-L, Meehan PA. (2020). On the modeling of wear in grease-lubricated spherical roller bearings. Tribology Transactions 63, 806–19. https://doi.org/10.1080/10402004.2020.1743400.

[Press 2007] Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical Recipes. The art of scientific computing. 3rd edition. New York, USA: Cambridge University Press; 2007. ISBN: 978-0-511-33555-6.

[Quost 2006] Quost X, Sebes M, Eddhahak A, Ayasse JB, Chollet H, Gautier P, Thouverez F. (2006). Assessment of a semi-Hertzian method for determination of wheel–rail contact patch. Vehicle System Dynamics 44, 789–814. https://doi.org/10.1080/00423110600677948.

[Rafei 2018] Rafei M, Ghoreishy MHR, Naderi G. (2018). Computer simulation of tire rolling resistance using finite element method: Effect of linear and nonlinear viscoelastic models. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering 233(11), 1–15. https://doi.org/10.1177/0954407018804117.

[Redtenbacher 1855] Redtenbacher FJ. Die Gesetze des Lokomotiv-Baues. Verlag von Friedrich Bassermann, Mannheim, 1855, p. 22.

[Reynolds 1875] Reynolds O. (1875). On rolling friction. Phil. Trans. Royal Society 166, 155–74. https://doi.org/10.1098/rstl.1876.0006.

[Roark's 2002] Young WC, Budynas RG. Roark's formulas for stress and strain. 7th edition. New York, USA: McGraw-Hill; 2002. ISBN: 0-07-072542-X.

[Robles 2022] Robles R, Correa N, Vadillo EG, Blanco-Lorenzo J. A new fast non-linear vertical and lateral model to predict rail corrugation. In: Proceedings of the 12th International Conference on Contact Mechanics and Wear of Rail/Wheel Systems (CM2022). Melbourne, Australia, 4–7 September, 2022.

[Robles 2023a] Robles R, Correa N, Vadillo EG, Blanco-Lorenzo J. (2023). Comprehensive efficient vertical and lateral track dynamic model to study the evolution of rail corrugation in sharp curves. Journal of Sound and Vibration 545C, 117448. https://doi.org/10.1016/j.jsv.2022.117448.

[Robles 2023b] Robles R, Correa N, Vadillo EG, Blanco-Lorenzo J. (2023). Predicting rail corrugation in a real line by means of a fast non-linear vertical and lateral model. Wear 524–5, 204896. https://doi.org/10.1016/j.wear.2023.204896.

[Rodríguez-Tembleque 2010] Rodríguez-Tembleque L, Abascal R. (2010). A 3D FEM–BEM rolling contact formulation for unstructured meshes. International Journal of Solids and Structures 47, 330–53. https://doi.org/10.1016/j.ijsolstr.2009.10.008.

[Rothman 1950] Rothman M. (1950). Isolated force problems in two-dimensional elasticity (I). Quarterly Journal of Mechanics and Applied Mathematics 3(3), 279–96. https://doi.org/10.1093/qjmam/3.3.279.

[Sainsot 2011] Sainsot P, Lubrecht AA. (2011). Efficient solution of the dry contact of rough surfaces: a comparison of fast Fourier transform and multigrid methods. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology 225, 441–7. https://doi.org/10.1177/1350650111401535.

[Sato 2002] Sato Y, Matsumoto A, Knothe K. (2002). Review on rail corrugation studies. Wear 253, 130–9. https://doi.org/10.1016/S0043-1648(02)00092-3.

[Schwack 2018] Schwack F, Prigge F, Poll G. (2018). Finite element simulation and experimental analysis of false brinelling and fretting corrosion. Tribology International 126, 352–62. https://doi.org/10.1016/j.triboint.2018.05.013.

[Schwack 2021] Schwack F, Schneider V, Wandel S, de la Presilla RJ, Poll G, Glavatskih S. (2021). On the critical amplitude in oscillating rolling element bearings. Tribology International 163, 107154. https://doi.org/10.1016/j.triboint.2021.107154.

[Shabana 2005] Shabana AA. Dynamics of Multibody Systems. 3rd edition. Cambridge, UK: Cambridge University Press; 2005. https://doi.org/10.1017/CBO9780511610523.

[Shen 1983] Shen ZY, Hedrick JK, Elkins JA. (1983). A comparison of alternative creep force models for rail vehicle dynamic analysis. Vehicle System Dynamics 12(1–3), 79–83. https://doi.org/10.1080/00423118308968725.

[Sichani 2014] Sichani MSh, Enblom R, Berg M. (2014). A novel method to model wheel–rail normal contact in vehicle dynamics simulation. Vehicle System Dynamics 52(12), 1752–64. https://doi.org/10.1080/00423114.2014.961932.

[Sichani 2016a] Sichani MSh, Enblom R, Berg M. (2016). An alternative to FASTSIM for tangential solution of the wheel–rail contact. Vehicle System Dynamics 54(6), 748–64. https://dx.doi.org/10.1080/00423114.2016.1156135.

[Sichani 2016b] Sichani MSh, Enblom R, Berg M. (2016b). A fast wheel–rail contact model for application to damage analysis in vehicle dynamics simulation. Wear 366–7, 123–30. https://doi.org/10.1016/j.wear.2016.06.015.

[SIMPACK Wheel Rail doc 2007] SIMPACK GmbH. SIMPACK Wheel Rail Element Reference, 2007.

[Slack 2010] Slack T, Sadeghi F. (2010). Explicit finite element modelling of subsurface initiated spalling in rolling contacts. Tribology International 43, 1693–702. https://doi.org/10.1016/j.triboint.2010.03.019.

[Spence 1975] Spence DA. (1975). The Hertz contact problem with finite friction. Journal of Elasticity 5(3–4), 297–319. https://doi.org/10.1007/BF00126993.

[Spiegel 2009] Spiegel MR, Lipschutz S, Liu J. Mathematical Handbook of Formulas and Tables. 3rd edition. McGraw-Hill; 2009.

[Steenbergen 2008] Steenbergen MJMM. (2008). Quantification of dynamic wheel–rail contact forces at short rail irregularities and application to measured rail welds. Journal of Sound and Vibration 312, 606–29. https://doi.org/10.1016/j.jsv.2007.11.004.

[Sternberg 1952] Sternberg E., Rosenthal F. (1952). The elastic sphere under concentrated loads. Journal of Applied Mechanics 19, 413–21.

[Steuermann 1940] Steuermann E. (1940). Local deformations in elastic circular cylinders with nearly equal radii under pressure. CR Acad Sci URSS 29(2), 182–4.

[Sundaram 2010a] Sundaram N, Farris TN. (2010). The generalized advancing conformal contact problem with friction, pin loads and remote loading – Case of rigid pin. International Journal of Solids and Structures 47, 801–15. https://doi.org/10.1016/j.ijsolstr.2009.11.019.

[Sundaram 2010b] Sundaram N, Farris TN. (2010). Mechanics of advancing pin-loaded contacts with friction. Journal of the Mechanics and Physics of Solids 58, 1819–33. https://doi.org/10.1016/j.jmps.2010.08.004.

[Suwannachit 2013] Suwannachit A, Nackenhorst U. (2013). A novel approach for thermomechanical analysis of stationary rolling tires within an ALE-kinematic framework. Tire Science and Technology 41(3), 174–95. https://doi.org/10.2346/tire.13.410304.

[Svec 1971] Svec OJ, Gladwell GML. (1971). An explicit boussinesq solution for a polynomial distribution of pressure over a triangular region. Journal of Elasticity 1(2), 167–70. https://doi.org/10.1007/BF00046468.

[Telliskivi 2001] Telliskivi T, Olofsson U. (2001). Contact mechanics analysis of measured wheel–rail profiles using the finite element method. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 15(1), 65–72. https://doi.org/10.1243/0954409011531404.

[Thompson 2000] Thompson DJ, Jones CJC. (2000). A review of the modelling of wheel/rail noise generation. Journal of Sound and Vibration 231(3), 519–36. https://doi.org/10.1006/jsvi.1999.2542.

[Thompson 2009] Thompson D. Railway noise and vibration. Oxford, UK: Elsevier; 2009. ISBN: 978-0-08-045147-3.

[Timoshenko 1951] Timoshenko S, Goodier JN. Theory of Elasticity. 2nd edition. New York, USA: McGraw-Hill; 1951.

[Titovich 2012] Titovich AS, Norris AN. (2012). Green's function for symmetric loading of an elastic sphere with application to contact problems. Journal of Mechanics of Materials and Structures 7(7), 701–19. https://doi.org/10.2140/jomms.2012.7.701.

[Tonazzi 2017] Tonazzi D, Komba EH, Massi F, Le Jeune G, Coudert JB, Maheo Y, Berthier Y. (2017). Numerical analysis of contact stress and strain distributions for greased and ungreased high loaded oscillating bearings. Wear 376–7, 1164–75. https://doi.org/10.1016/j.wear.2016.11.037.

[Toumi 2016] Toumi M, Chollet H, Yin H. (2016). Finite element analysis of the frictional wheel–rail rolling contact using explicit and implicit methods. Wear 366–7, 157–66. https://doi.org/10.1016/j.wear.2016.06.008.

[Venner 2000] Venner CH, Lubrecht AA. Multilevel Methods in Lubrication. Amsterdam, The Netherlands: Elsevier Science BV; 2000. ISBN: 0 444 50503 2.

[Vermeulen 1964] Vermeulen PJ, Johnson KL. (1964). Contact of nonspherical elastic bodies transmitting tangential forces. Journal of Applied Mechanics 31(2), 338–40. https://doi.org/10.1115/1.3629610.

[VI-grade www] https://www.vi-grade.com. Accessed on 05/04/2023.

[Vincent 2000] Vincent N. (2000). Rolling noise control at source: state-of-the-art survey. Journal of Sound and Vibration 231(3), 865–76. https://doi.org/10.1006/jsvi.1999.2646.

[Vo 2014] Vo KD, Tieu AK, Zhu HT, Kosasih PB. (2014). A 3D dynamic model to investigate wheel–rail contact under high and low adhesion. International Journal of Mechanical Sciences 85, 63–75. https://doi.org/10.1016/j.ijmecsci.2014.05.007.

[Vo 2015] Vo KD, Zhu HT, Tieu AK, Kosasih PB. (2015). FE method to predict damage formation on curved track for various worn status of wheel/rail profiles. Wear 322–3, 61–75. http://dx.doi.org/10.1016/j.wear.2014.10.015.

[Vollebregt 1995] Vollebregt EAH. (1995). A Gauss-Seidel type solver for special convex programs, with application to frictional contact mechanics. Journal of Optimization Theory and Applications 87, 47–67. https://doi.org/10.1007/BF02192041.

[Vollebregt 2009] Vollebregt EAH. Refinement of Kalker's rolling contact model. In: Proceedings of the 8th International Conference on Contact Mechanics and Wear of Rail/Wheel Systems (CM2009). Firenze, Italy, 15–18 September, 2009.

[Vollebregt 2010] Vollebregt EAH. Improving the speed and accuracy of the frictional rolling contact model "CONTACT". In: Proceedings of the 10th International Conference on Computational Structures Technology, Topping BHV, Adam JM, Pallarés FJ, Bru R, Romero ML (Eds.), Civil-Comp Press, Stirlingshire, Scotland, 2010.

[Vollebregt 2012] Vollebregt EAH. 100-fold speed-up of the normal contact problem and other recent developments in "CONTACT". In: Proceedings of the 9th International Conference on Contact Mechanics and Wear of Rail/Wheel Systems (CM2012). Chengdu, China, 27–30 August, 2012.

[Vollebregt 2014a] Vollebregt EAH, Segal G. (2014). Solving conformal wheel–rail rolling contact problems. Vehicle System Dynamics 52(suppl. 1), 455–68. http://dx.doi.org/10.1080/00423114.2014.906634.

[Vollebregt 2014b] Vollebregt EAH. (2014). A new solver for the elastic normal contact problem using conjugate gradients, deflation, and an FFT-based preconditioner. Journal of Computational Physics 257, 333–51. https://doi.org/10.1016/j.jcp.2013.10.005.

[Vollebregt 2014c] Vollebregt EAH. (2014). The bound-constrained conjugate gradient method for non-negative matrices. Journal of Optimization Theory and Applications 162, 931–53. https://doi.org/10.1007/s10957-013-0499-x.

[Vollebregt 2014d] Vollebregt EAH. (2014). Numerical modelling of measured railway creep versus creep-force curves with CONTACT. Wear 314(1–2), 87–95. https://doi.org/10.1016/j.wear.2013.11.030.

[Vollebregt 2018] Vollebregt EAH. (2018). Conformal contact: corrections and new results. Vehicle System Dynamics 56(10), 1622–32. https://doi.org/10.1080/00423114.2018.1424917.

[Vollebregt 2021] Vollebregt E. (2021). Detailed wheel/rail geometry processing with the conformal contact approach. Multibody System Dynamics 52, 135–67. https://doi.org/10.1007/s11044-020-09762-w.

[Vollebregt 2022] Vollebregt E. (2022). Detailed wheel/rail geometry processing using the planar contact approach. Vehicle System Dynamics 60(4), 1253–91. https://doi.org/10.1080/00423114.2020.1853180.

[VOSSLOH www] Vossloh Fastening Systems. www.vossloh-fastening-systems.de.

[Vuong 2011] Vuong TT, Meehan PA, Eadie DT, Oldknow K, Elvidge D, Bellette PA, Daniel WJT. (2011). Investigation of a transitional wear model for wear and wear-type rail corrugation prediction. Wear 271, 287–98. http://dx.doi.org/10.1016/j.wear.2010.10.008.

[Wang G 1989] Wang G, Knothe K. (1989). The influence of inertia forces on steady-state rolling contact. Acta Mechanica 79, 221–32. https://doi.org/10.1007/BF01187264.

[Wang G 1993] Wang G, Knothe K. (1993). Stress analysis for rolling contact between two viscoelastic cylinders. Journal of Applied Mechanics 60, 310–7. https://doi.org/10.1115/1.2900794.

[Wang K 1984] Wang K. (1984). The computation of wheel–rail contact locus and wheel–rail geometrical parameters (in Chinese). Journal of Southwest Jiaotong University 1, 89–98.

[Wang Z 2012] Wang Z, Jin X, Keer LM, Wang Q. (2012). A numerical approach for analyzing three-dimensional steady-state rolling contact including creep using a fast semi-analytical method. Tribology Transactions 55(4), 446–57. http://dx.doi.org/10.1080/10402004.2012.667518.

[Wang Z-J 2010] Wang Z-J, Wang W-Z, Wang H, Zhu D, Hu Y-Z. (2010). Partial slip contact analysis on three-dimensional elastic layered half space. Journal of Tribology 132(2), 021403. https://doi.org/10.1115/1.4001011.

[Wickens 2003] Wickens AH. Fundamentals of rail vehicle dynamics. Guidance and stability. Lisse, The Netherlands: Swets & Zeitlinger B.V.; 2003. ISBN: 90 265 1946 X.

[van der Wekken 2019] van der Wekken CD, Vollebregt EAH. (2019). Numerical calculation of the elastic field in a half-space using bilinear elements. Mathematics and Mechanics of Solids 24(11), 3537–53. https://doi.org/10.1177/1081286519848822.

[Wikipedia Maglev www] https://en.wikipedia.org/wiki/Maglev. Accessed on 03/04/2023.

[Wollny 2016] Wollny I, Hartung F, Kaliske M. (2016). Numerical modeling of inelastic structures at loading of steady state rolling. Thermo-mechanical asphalt pavement computation. Computational Mechanics 57, 867–86. https://doi.org/10.1007/s00466-016-1266-2.

[Woodward 1976] Woodward W, Paul B. (1976). Contact stresses for closely conforming bodies–application to cylinders and spheres. Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania. Report no DOT-TST-77-48.

[Wriggers 1995] Wriggers P. (1995). Finite element algorithms for contact problems. Archives of Computational Methods in Engineering 2(4), 1–49. https://doi.org/10.1007/BF02736195.

[Wriggers 2001] Wriggers P. Finite element methods for rolling contact. Ch 2. In: Jacobson B & Kalker JJ (Eds.). Rolling Contact Phenomena, CISM Courses and Lectures, No. 411, Chapter 2, pp. 85–162. Wien, New York: Springer; 2001. https://doi.org/10.1007/978-3-7091-2782-7_2.

[Wriggers 2006] Wriggers P. Computational contact mechanics. 2nd edition. Berlin, Germany: Springer-Verlag; 2006. ISBN: 978-3-540-32608-3.

[Wu 2004] Wu TX, Thompson DJ. (2004). On the parametric excitation of the wheel/track system. Journal of Sound and Vibration 278(4–5), 725–47. https://doi.org/10.1016/j.jsv.2003.10.047.

[Xia 2010] Xia H, Cao YM, De Roeck G. (2010). Theoretical modeling and characteristic analysis of moving-train induced ground vibrations. Journal of Sound and Vibration 329(7), 819–32. https://doi.org/10.1016/j.jsv.2009.10.007.

[Xie 2008] Xie G, Iwnicki SD. (2008). Simulation of wear on a rough rail using a time-domain wheel–track interaction model. Wear 265(11–12), 1572–83. https://doi.org/10.1016/j.wear.2008.03.016.

[Yang 2016] Yang Z, Li Z, Dollevoet R. (2016). Modelling of non-steady-state transition form single-point to two-point rolling contact. Tribology International 101, 152–63. http://dx.doi.org/10.1016/j.triboint.2016.04.023.

[Yastrebov 2011] Yastrebov VA. Computational Contact Mechanics: geometry, detection and numerical techniques. Ph.D. dissertation, l'École Nationale Supérieure des Mines de Paris, France. 2011.

[Zeid 1981] Zeid I, Padovan J. (1981). Finite element modelling of rolling contact. Computers and Structures 14(1–2), 163–70. https://doi.org/10.1016/0045-7949(81)90098-5.

[Zhai 2009] Zhai W, Wang K, Cai C. (2009). Fundamentals of vehicle–track coupled dynamics. Vehicle System Dynamics 47(11), 1349–76. https://doi.org/10.1080/00423110802621561.

[Zhang 2013] Zhang ZM, Wang W, Wong PL. (2013). An explicit solution for the elastic quarter-space problem in matrix formulation. International Journal of Solids and Structures 50(6), 976–80. https://doi.org/10.1016/j.ijsolstr.2012.12.001.

[Zhao J 2015] Zhao J, Vollebregt EAH, Oosterlee CW. A fast nonlinear conjugate gradient based method for 3D concentrated frictional contact problems. (2015). Journal of Computational Physics 288, 86–100. https://dx.doi.org/10.1016/j.jcp.2015.02.016.

[Zhao J 2016] Zhao J, Vollebregt EAH, Oosterlee CW. (2016). Extending the BEM for elastic contact problems beyond the half-space approach. Mathematical Modelling and Analysis 21(1), 119–41. https://doi.org/10.3846/13926292.2016.1138418.

[Zhao X 2011] Zhao X, Li Z. (2011). The solution of frictional wheel–rail rolling contact with a 3D transient finite element model: validation and error analysis. Wear 271(1–2), 444–52. https://doi.org/10.1016/j.wear.2010.10.007.

[Zhao X 2014] Zhao X, Wen Z, Zhu M, Jin X. (2014). A study on high-speed rolling contact between a wheel and a contaminated rail. Vehicle System Dynamics 52(10), 1270–87. http://dx.doi.org/10.1080/00423114.2014.934845.

[Zhao X 2015] Zhao X, Li Z. (2015). A three-dimensional finite element solution of frictional wheel–rail rolling contact in elasto-plasticity. Journal of Engineering Tribology 229(1), 86–100. https://doi.org/10.1177/1350650114543717.

[Zhuravlev 1998] Zhuravlev VPh. (1998). The model of dry friction in the problem of the rolling or rigid bodies. Journal of Applied Mathematics and Mechanics 62(5), 705–10. https://doi.org/10.1016/S0021-8928(98)00090-2.

[Ziefle 2008] Ziefle M, Nackenhorst U. (2008). Numerical techniques for rolling rubber wheels: treatment of inelastic material properties and frictional contact. Computational Mechanics 42, 337–56. https://doi.org/10.1007/s00466-008-0243-9.

[Zobova 2019] Zobova AA. (2019). Dry friction distributed over a contact patch between a rigid body and a visco-elastic plane. Multibody System Dynamics 45, 203–22. https://doi.org/10.1007/s11044-018-09637-1.