CISM International Centre for Mechanical Sciences 585
Courses and Lectures

Alexander Popp
Peter Wriggers *Editors*

Contact Modeling for Solids and Particles

International Centre for Mechanical Sciences

Springer

CISM International Centre for Mechanical Sciences

Courses and Lectures

Volume 585

The series presents lecture notes, monographs, edited works and proceedings in the field of Mechanics, Engineering, Computer Science and Applied Mathematics. Purpose of the series is to make known in the international scientific and technical community results obtained in some of the activities organized by CISM, the International Centre for Mechanical Sciences.

More information about this series at http://www.springer.com/series/76

Alexander Popp · Peter Wriggers
Editors

Contact Modeling for Solids and Particles

Springer

Editors
Alexander Popp
Institute for Mathematics and
Computer-Based Simulation
University of the Bundeswehr Munich
Neubiberg
Germany

Peter Wriggers
Institute of Continuum Mechanics
Leibniz University Hannover
Hannover
Germany

ISSN 0254-1971 ISSN 2309-3706 (electronic)
CISM International Centre for Mechanical Sciences
ISBN 978-3-319-90154-1 ISBN 978-3-319-90155-8 (eBook)
https://doi.org/10.1007/978-3-319-90155-8

Library of Congress Control Number: 2018939021

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG
part of Springer Nature
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The chapters contained in this volume correspond to lectures given during the course "Computational Contact and Interface Mechanics" that was held at the CISM in Udine (Italy), October 3–7, 2016.

Contact is virtually omnipresent in many technical and biological systems, which motivates a broad dissemination of profound knowledge regarding the computational modeling of contact phenomena in the context of both solid mechanics and particle mechanics. In this spirit, the main objective of the CISM course and, consequently, of this volume is to convey modern techniques and the latest state-of-the-art with regard to the most fundamental aspects of computational contact modeling for solids and particles. However, since contact can readily be interpreted as a special type of interface problem, it seems advisable not to isolate contact mechanics, but rather to address it in the context of a broader class of problems denoted as computational interface mechanics. Apart from the computational treatment of contact interaction and friction, computational interface mechanics also comprises other related physical phenomena such as wear, lubrication, and thermomechanical interface effects. Put in short terms, computational contact and interface mechanics are concerned with the treatment of complex interface effects at different length scales ranging from atomistic models to micro- and mesoscale models and further to continuum models at the macroscale. The nature of many interface phenomena even requires a multi-scale perspective and associated models to bridge the spectrum of relevant length scales.

Therefore, the aforementioned aim of the volume has been expanded toward firstly conveying a clear understanding of the underlying physics of interfaces, and secondly giving a comprehensive insight into the current state-of-the-art and selected cutting-edge research directions in the computational treatment of interface effects. With regard to the first aim, the carefully prepared chapters of this volume will focus on the modeling of friction, wear, lubrication, thermomechanics, and particle contact (e.g., granular media). In view of the second objective, the most important computational aspects will be addressed, including discretization techniques for finite deformations, solution algorithms for single- and multi-processor computing environments, multi-scale approaches, and multi-physics problems

including contact and interface constraints. Among the computational techniques covered in this volume are finite element methods (FEM), mortar methods, isogeometric analysis (IGA), virtual element methods (VEM), contact domain and third medium methods as well as discrete element methods (DEM).

Each chapter will start from the respective basics of physical modeling and computational techniques and will then move on to an in-depth treatment of cutting-edge research topics. The volume starts with a general introduction to FEM-based contact modeling in nonlinear solid mechanics by Alexander Popp, followed by an overview of several emerging, non-conventional discretization techniques for contact problems by Peter Wriggers. Subsequently, the focus is shifted more toward more complex interface phenomena such as wear and soft elastohydrodynamic lubrication by Stanisław Stupkiewicz. Finally, the fundamentals of contact modeling for particles using DEM are reviewed by Jerzy Rojek. The content is primarily designed for doctoral students and postdoctoral researchers in applied mathematics, mechanics, engineering, and physics with a strong interest in contact and interface phenomena. However, the book is equally suited for academic and industrial researchers, who have only little experience with regard to the computational treatment of interface effects and who would like to gain a compact yet comprehensive overview of the field, as well as for practicing computational engineers working on high-level industrial applications.

The editors wish to express their sincere gratitude to all contributors to this volume, to the remaining lecturers of the corresponding CISM course, to all course participants who created an enthusiastic and truly inspiring atmosphere, and last but not least to all members of CISM and CISM Springer for their support and thoughtful suggestions during the course and during the preparation of this volume.

Neubiberg, Germany Alexander Popp
Hannover, Germany Peter Wriggers

Contents

State-of-the-Art Computational Methods for Finite Deformation Contact Modeling of Solids and Structures

Alexander Popp

Abstract In this contribution, we review mortar finite element methods (FEM), which are nowadays the most well-established computational technique for contact modeling of solids and structures in the context of finite deformations and frictional sliding. Based on some concepts of nonlinear continuum mechanics, the mortar approach is first presented for the more accessible case of mesh tying (also referred to as tied contact). Mortar methods for unilateral contact then follow in a rather straightforward manner, despite the fact that several complexities, such as inequality constraints, are added to the problem formulation. A special focus is set on practical aspects of the implementation of mortar methods within a fully nonlinear, 3D finite element environment. Specifically, the choice of suitable discrete Lagrange multiplier bases, aspects of high performance computing (HPC), numerical integration procedures and new discretization techniques such as isogeometric analysis (IGA) using NURBS are discussed. Eventually, the great potential of mortar methods in the more general field of computational interface mechanics is exemplified through applications such as wear modeling and coupled thermo-mechanical interfaces.

1 Introduction and Motivation

Contact phenomena are virtually omnipresent in nature and biological systems. The associated length and time scales cover the entire spectrum from the nanoscale to the macroscopic level and from hypervelocity impact to quasi-static contact interaction, respectively. For example, the plate tectonics process of the continental drift, the simple motion sequence when walking or the flow of red blood cells (erythrocytes) through blood vessels are all representatives of processes largely dominated by contact and associated physical effects. Beyond that, science and engineering have exploited the principles of contact mechanics to develop processes, such as

A. Popp (✉)
Institute for Mathematics and Computer-Based Simulation,
University of the Bundeswehr Munich, Neubiberg, Germany
e-mail: alexander.popp@unibw.de

A. Popp and P. Wriggers (eds.), *Contact Modeling for Solids and Particles*,
CISM International Centre for Mechanical Sciences 585,
https://doi.org/10.1007/978-3-319-90155-8_1

deep-drawing or extrusion-molding, as well as technical systems and machine parts, including car tires, fluid bearings, gears, shafts and splines or elastomeric seals.

Contact mechanics can be looked at from several different perspectives. For some scenarios, e.g. in nanotribology, it is helpful or even mandatory to investigate contact interaction at an atomistic level. For many contact applications, however, a purely macroscopic viewpoint based on classical continuum assumptions is sufficient. Throughout this chapter, a continuum approach will be followed, mainly considering contact mechanics as a particularly challenging subclass of solid and structural mechanics. The geometrical constraint of non-penetration of different solid bodies can then easily be identified as the most important underlying principle of contact interaction. In addition, the overall contact phenomenon is commonly also influenced by one or several closely related interface effects, for example sticking and sliding friction, adhesion, elastohydrodynamic lubrication and wear. Altogether, contact and its associated phenomena introduce strong additional nonlinearities into solid mechanics problems, where contact itself can basically be interpreted as a set of complex boundary conditions, possibly changing over time. Together with the already typical nonlinearities inherent in general solid mechanics, i.e. large deformations and nonlinear constitutive (material) behavior, this evinces the challenges and difficulties of mathematically describing and solving contact interactions, even if the given problem setup is quite simple. Due to this complexity, only very few contact problem settings exist, where analytical solution techniques are actually applicable. The early work conducted by Hertz (1882) on pressure distributions between contacting elastic bodies more than a century ago, is commonly considered to be the origin of modern contact analysis. A comprehensive overview of the basic principles of contact mechanics, together with the most important analytical solution techniques can be found in the textbooks by Johnson (1985) and Timoshenko and Goodier (1970).

With general contact problems being hardly accessible for mathematical analysis, experimental procedures and numerical modeling are naturally becoming the focus of attention. Physical experiments are a convenient way of gaining information about certain aspects of contact mechanics, e.g. for determining coefficients of friction related to different material pairings. However, for the majority of contact scenarios, the applicability of experimental procedures is either limited or practically impossible. As a prominent example, experimental crashworthiness assessment, in accordance with safety regulations and consumer protection tests, causes considerable costs in the automotive industry. Complex contact phenomena in patient-specific surgery planning or during the design of medical devices, e.g. guaranteeing the optimal placement and minimum leakage of arterial stents, do not even allow for meaningful experimental tests at all. Thus, combining the aforementioned exemplary arguments, it becomes obvious that there is a very high and ever-growing demand for powerful numerical modeling and simulation techniques in the field of contact mechanics. What makes improved contact simulation approaches even more promising and likely to generate significant impact is the fact that the resulting numerical algorithms can typically be employed for a very broad range of scientific and technical interests. In fundamental physical, chemical or biological research, as well as in the applied sciences, novel methods and tools of computational contact mechanics

allow for a better understanding of complex systems, which are influenced by contact phenomena. On the other hand, many aspects of engineering practice and product development (e.g. minimizing the frictional loss in gear transmissions, optimizing the structural integrity of car bodies in crash situations) also heavily benefit from improvements in contact modeling and simulation.

2 Contact Mechanics and FEM

All ideas and methods of computational contact mechanics will be exclusively discussed in the context of the finite element method (FEM) throughout this chapter. Since the 1960s, the FEM has gradually evolved as the dominating numerical approximation technique for the solution of partial differential equations (PDEs) in various fields, especially solid and structural mechanics including contact mechanics, but also in fluid mechanics, thermodynamics and for the treatment of coupled problems. The general FEM literature is abundant, exemplarily the interested reader is referred to the monographs by Bathe (1996), Hughes (2000), Belytschko et al. (2000), Reddy (2004), Zienkiewicz et al. (2005) and Zienkiewicz and Taylor (2005). Other approaches for the numerical simulation of contact mechanics are only mentioned very briefly here for the sake of completeness. Multibody dynamics are a fitting tool when analyzing contact and impact phenomena of rigid bodies, with possible extensions to elastic multibody dynamics allowing for a certain degree of deformation of the contacting bodies. Moreover, particle methods such as the discrete element method (DEM) are frequently used for investigating granular and particulate materials, whose mechanical behavior is largely dominated by contact interaction. While finite elements would not be the method of choice for such applications, this chapter is mainly related to contact of elastic solid bodies, possibly including very large deformations. In this context, the FEM undoubtedly provides a very convenient framework for numerical modeling and simulation. Furthermore, there is an increasing interest in the interplay of contact mechanics with other physical phenomena, such as thermomechanics, wear and the lubrication behavior of thin fluid films, where finite elements are also an eligible approach, e.g. due to their generality and geometrical flexibility.

First contributions to the treatment of contact mechanics within the FEM can be traced back to the 1970s and 1980s. In Francavilla and Zienkiewicz (1975) and Hughes et al. (1976), contact conditions are formulated based on a very simple, purely node-based approach, which requires node-matching finite element meshes at the contact interface and is restricted to small deformations. Subsequently, a different idea was expedited, typically denoted as node-to-surface or node-to-segment (NTS) approach and characterized by a discrete, point-wise enforcement of the non-penetration condition at the finite element nodes. This NTS approach could readily be applied to the case of finite deformations and large sliding motions, therefore soon becoming the standard procedure in computational contact mechanics. Without claiming that the following listing is exhaustive, the reader is referred to Bathe and

Chaudhary (1985), Hallquist et al. (1985), Benson and Hallquist (1990), Simo and Laursen (1992), Laursen (1992), Laursen and Simo (1993) and Wriggers et al. (1990) for a comprehensive overview. An important basis for the methods to be proposed in this chapter is formed by the first investigations on the so-called segment-to-segment (STS) approach in Papadopoulos and Taylor (1992) and Simo et al. (1985). In contrast to the purely point-wise procedure typical of NTS methods, the STS approach is based on a thorough sub-division of the contact surface into individual segments for numerical integration together with an independent approximation of the contact pressure. Thereby, the STS approach can be interpreted as precursor of mortar finite element methods for computational contact mechanics, which will be the main topic here.

Before reviewing the literature on mortar methods, however, an overview of other important aspects of computational contact mechanics aside from the discretization approach (NTS, STS, mortar) is given. One main focus of attention has been set on different procedures for the enforcement of contact constraints, with the most prominent representatives being penalty methods, Lagrange multiplier methods and Augmented Lagrange methods, see Alart and Curnier (1991) for an excellent overview and discussion. Further questions related to contact modeling within a finite element framework comprise efficient search algorithms (Williams and O'Connor 1999), mesh adaptivity (Wriggers and Scherf 1995; Carstensen et al. 1999; Hüeber and Wohlmuth 2012), covariant surface description (Laursen and Simo 1993; Schweizerhof and Konyukhov 2005), surface smoothing (Wriggers et al. 2001; Puso and Laursen 2002), the treatment of contact on enriched and embedded interfaces (Laursen et al. 2012), modeling of interface effects other than friction (Yang and Laursen 2009; Sauer 2011), beam contact (Wriggers and Zavarise 1997; Zavarise and Wriggers 2000) and energy conservation in the context of contact dynamics (Laursen and Chawla 1997; Laursen and Love 2002; Hager et al. 2008; Hesch and Betsch 2009), among others. Apart from numerous original papers, a comprehensive introduction to most of these topics can be found in the textbooks by Laursen (2002) and Wriggers (2006).

Nevertheless, novel robust discretization techniques for finite deformation contact problems, and especially mortar finite elements adapted for this purpose, have arguably received most attention in the field of computational contact mechanics in recent years. Mortar methods, which were originally introduced as an abstract domain decomposition technique (Bernardi et al. 1994; Ben Belgacem 1999; Seshaiyer and Suri 2000), are characterized by an imposition of the occurring interface constraints in a weak sense and by the possibility to prove their mathematical optimality. In the context of contact analysis, this allows for a variationally consistent treatment of non-penetration and frictional sliding conditions despite the inevitably non-matching interface meshes for finite deformations and large sliding motions. Early applications of mortar finite element methods for contact mechanics can, for example, be found in Ben Belgacem et al. (1998), Hild (2000) and McDevitt and Laursen (2000), though limited to small deformations. Gradually, restrictions of mortar-based contact formulations with respect to nonlinear kinematics have been removed, leading to the implementations given in Puso and Laursen (2004a, b), Fischer and Wriggers (2005),

Fischer and Wriggers (2006), Hesch and Betsch (2009), Tur et al. (2009) and Hesch and Betsch (2011).

An alternative choice for the discrete Lagrange multiplier space, so-called dual Lagrange multipliers, was proposed in Wohlmuth (2000, 2001) and, in contrast to the standard mortar approach, generates interface coupling conditions that are much easier to realize without impinging upon the optimality of the method. Applications of this approach to small deformation contact problems can be found in Hüeber and Wohlmuth (2005), Flemisch and Wohlmuth (2007), Brunssen et al. (2007) and Hüeber et al. (2008), and first steps towards finite deformations have been undertaken in Hartmann (2007) and Hartmann et al. (2007). A fully nonlinear extension of the dual mortar approach including consistent linearization of all deformation-dependent quantities has been proposed in Popp et al. (2009, 2010), with extensions to frictional sliding, second-order finite elements and a consistent treatment of dropping-edge problems following shortly afterwards (Cichosz and Bischoff 2011; Popp et al. 2012; Wohlmuth et al. 2012; Popp et al. 2013; Popp and Wall 2014). Another interesting feature of dual Lagrange multiplier interpolation is that it naturally fits together with so-called primal-dual active set strategies for constraint enforcement. It is well-known from the mathematical literature on constrained optimization problems and also from applications in computational contact mechanics, that primal-dual active set strategies can equivalently be interpreted as semi-smooth Newton methods (Alart and Curnier 1991; Qi and Sun 1993; Christensen et al. 1998; Christensen 2002; Hintermüller et al. 2002), thus allowing for the design of very efficient global solution algorithms, especially in the context of nonlinear material behavior and finite deformations.

Recent developments in the meanwhile rather broad field of mortar finite element methods for computational contact mechanics include, without being complete, the following topics: smoothing techniques (Tur et al. 2012), isogeometric analysis using NURBS (Temizer et al. 2011, 2012; De Lorenzis et al. 2014; Brivadis et al. 2015), improved numerical integration schemes (Farah et al. 2015), complex interface models such as wear (Cavalieri and Cardona 2013; Farah et al. 2016, 2017), treatment of embedded interfaces (Laursen et al. 2012) as well as aspects of adaptivity and high performance computing (Popp and Wall 2014; Kindo et al. 2014). While a few different discretization approaches have been suggested, see e.g. the contact domain method proposed in Hartmann et al. (2009) and Oliver et al. (2009), and while NTS methods are still very popular in engineering practice, mortar-based contact formulations have become quite well-established in the meantime and can arguably be seen as state-of-the-art method for computational contact mechanics.

3 Overview of Nonlinear Continuum Mechanics

In this section, the basic concepts of nonlinear continuum mechanics are reviewed with a focus on the governing equations for solid dynamics and contact interaction required later. These remarks are not intended to give an exhaustive overview of

the topic, but are rather geared towards outlining the necessary basics for contact mechanics. For more extensive reviews in the field of solid and structural dynamics, the reader is referred to the corresponding literature, e.g., Gurtin (1981), Marsden and Hughes (1994), Ogden (1997), Bonet and Wood (1997), Holzapfel (2000) and Simo and Hughes (1998). Large parts of this section are based on the author's previously published work (Popp 2012).

3.1 Kinematics

In this section, the fundamental kinematic relationships describing the deformation of a homogeneous body are presented. The classical (Boltzmann) continuum model in a three-dimensional Euclidean space description is assumed. Two distinct observer frames are defined: the reference configuration $\Omega_0 \subset \mathbb{R}^3$ denotes the domain occupied by all material points $\boldsymbol{X}$ at time $t = 0$, while the current configuration $\Omega_t \subset \mathbb{R}^3$ describes the changed positions $\boldsymbol{x}$ at a certain time t. The motion and deformation from reference to current configuration are tracked with the bijective nonlinear deformation map

$$\Phi_t : \begin{cases} \Omega_0 \rightarrow \Omega_t \\ \boldsymbol{X} \mapsto \boldsymbol{x} \end{cases}, \tag{1}$$

which also allows for the notations $\boldsymbol{x} = \Phi_t(\boldsymbol{X}, t)$ and $\boldsymbol{X} = \Phi_t^{-1}(\boldsymbol{x}, t)$. The absolute displacement of a material point (see again Fig. 1) is then described as

$$\boldsymbol{u}(\boldsymbol{X}, t) = \boldsymbol{x}(\boldsymbol{X}, t) - \boldsymbol{X}. \tag{2}$$

Within the total Lagrangian approach, kinematic relations and all derived quantities are described with respect to the material points in the reference configuration Ω_0. Thus, the material point position $\boldsymbol{X}$ plays the role of an independent variable for the problem formulation, while the primary unknown to be solved for

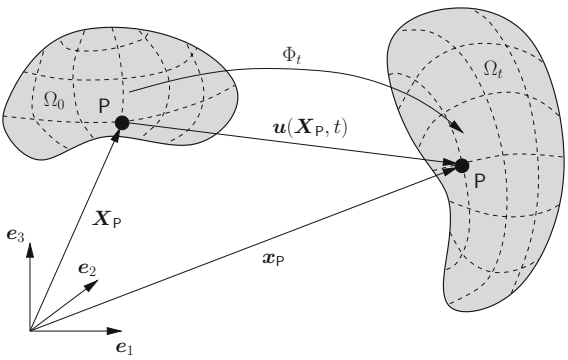

Fig. 1 Cartesian coordinate system, reference configuration and current configuration for a total Lagrangian description of motion

is the time-dependent deformation map $\Phi_t(\boldsymbol{X}, t)$, or equivalently the displacement vector $\boldsymbol{u}(\boldsymbol{X}, t)$.

A fundamental measure for deformation and strain in the context of finite deformation solid mechanics is given by the deformation gradient $\boldsymbol{F}$, defined as partial derivative of the current configuration with respect to the reference configuration:

$$\boldsymbol{F} = \frac{\partial \boldsymbol{x}(\boldsymbol{X}, t)}{\partial \boldsymbol{X}} = \boldsymbol{I} + \frac{\partial \boldsymbol{u}(\boldsymbol{X}, t)}{\partial \boldsymbol{X}}, \tag{3}$$

where $\boldsymbol{I}$ is the second-order identity tensor. Assuming as usual bijectivity and smoothness of the deformation map Φ_t, the inverse deformation gradient $\boldsymbol{F}^{-1} = \partial \boldsymbol{X}/\partial \boldsymbol{x}$ is also well-defined, therefore guaranteeing a positive determinant $J = \det \boldsymbol{F} > 0$. This quantity, also commonly denoted as Jacobian determinant of the deformation, represents the transformation of an infinitesimal volume element between the two configurations:

$$\mathrm{d}V = \det \boldsymbol{F}\, \mathrm{d}V_0 = J\, \mathrm{d}V_0. \tag{4}$$

The deformation gradient also allows for the mapping of an infinitesimal, oriented area element from reference to current configuration, yielding

$$\mathrm{d}\boldsymbol{A} = J\, \boldsymbol{F}^{-\mathsf{T}} \cdot \mathrm{d}\boldsymbol{A}_0, \tag{5}$$

which is commonly referred to as Nanson's formula. Herein, the infinitesimal area elements are interpreted as vectors $\mathrm{d}\boldsymbol{A}_0 = \mathrm{d}A_0\, \boldsymbol{N}$ and $\mathrm{d}\boldsymbol{A} = \mathrm{d}A\, \boldsymbol{n}$, where $\boldsymbol{N}$ and $\boldsymbol{n}$ denote unit normal vectors of the area element in the reference and current configuration, respectively.

An apparent choice for a suitable nonlinear strain measure is the so-called Green–Lagrange strain tensor $\boldsymbol{E}$ defined in the material configuration as

$$\boldsymbol{E} = \frac{1}{2}(\boldsymbol{F}^{\mathsf{T}} \cdot \boldsymbol{F} - \boldsymbol{I}) = \frac{1}{2}(\boldsymbol{C} - \boldsymbol{I}). \tag{6}$$

Although strain measures are never unique, the Green–Lagrange strain tensor is a very common choice in nonlinear solid mechanics, and can be considered particularly convenient if large deformations occur but only a moderate amount of stretch and compression.

The first and second time derivatives of the displacement vector $\boldsymbol{u}(\boldsymbol{X}, t)$ in material description, i.e. velocities $\dot{\boldsymbol{u}}(\boldsymbol{X}, t)$ and accelerations $\ddot{\boldsymbol{u}}(\boldsymbol{X}, t)$, are defined as follows:

$$\dot{\boldsymbol{u}}(\boldsymbol{X}, t) = \left.\frac{\partial \boldsymbol{u}(\boldsymbol{X}, t)}{\partial t}\right|_{\boldsymbol{X}} = \frac{\mathrm{d}\boldsymbol{u}(\boldsymbol{X}, t)}{\mathrm{d}t}, \tag{7}$$

$$\ddot{\boldsymbol{u}}(\boldsymbol{X}, t) = \left.\frac{\partial \dot{\boldsymbol{u}}(\boldsymbol{X}, t)}{\partial t}\right|_{\boldsymbol{X}} = \frac{\mathrm{d}\dot{\boldsymbol{u}}(\boldsymbol{X}, t)}{\mathrm{d}t} = \frac{\mathrm{d}^2\boldsymbol{u}(\boldsymbol{X}, t)}{\mathrm{d}t^2}. \tag{8}$$

Corresponding rate forms (i.e. time derivatives) of the deformation measures, such as the material velocity gradient $\boldsymbol{L} = \dot{\boldsymbol{F}}$ or the material strain rate tensor $\dot{\boldsymbol{E}} = \frac{1}{2}(\dot{\boldsymbol{F}}^{\mathsf{T}} \cdot \boldsymbol{F} + \boldsymbol{F}^{\mathsf{T}} \cdot \dot{\boldsymbol{F}}) = \frac{1}{2}\dot{\boldsymbol{C}}$ are readily defined, too.

3.2 Stresses and Constitutive Laws

The motion and deformation of an elastic body effects internal stresses. This is readily described by the traction vector $\boldsymbol{t}$ in the current configuration:

$$\boldsymbol{t}(\boldsymbol{n}, \boldsymbol{x}, t) = \lim_{\Delta A \to 0} \frac{\Delta \boldsymbol{f}}{\Delta A}, \tag{9}$$

yielding the limit value of the resulting force $\boldsymbol{f}$ acting on an arbitrary surface area ΔA characterized by its unit surface normal vector $\boldsymbol{n}$. The Cauchy theorem then correlates tractions and stresses via

$$\boldsymbol{t} = \boldsymbol{\sigma} \cdot \boldsymbol{n}. \tag{10}$$

Herein, the symmetric Cauchy stress tensor $\boldsymbol{\sigma}$ represents the true internal stress state within a body in its a priori unknown current configuration, with diagonal and off-diagonal components components being interpretable as normal stresses and shear stresses, respectively. A multitude of alternative stress definitions is also prevailing in nonlinear continuum mechanics. Exemplarily, the first Piola–Kirchhoff stress tensor $\boldsymbol{P}$ maps the *material* surface element $\mathrm{d}\boldsymbol{A}_0 = \mathrm{d}A_0\boldsymbol{N}$ onto the spatial resulting force $\boldsymbol{f}$. Its definition is obtained from the Cauchy stress tensor $\boldsymbol{\sigma}$ by applying Nanson's formula (5), yielding

$$\boldsymbol{P} = J\, \boldsymbol{\sigma} \cdot \boldsymbol{F}^{-\mathsf{T}}. \tag{11}$$

Consequently, it is possible to construct a stress tensor purely based on quantities in the reference configuration, too. By also transforming the resulting force vector $\boldsymbol{f}$ accordingly, the symmetric second Piola–Kirchhoff stress tensor $\boldsymbol{S}$ emerges as

$$\boldsymbol{S} = \boldsymbol{F}^{-1} \cdot \boldsymbol{P} = J\, \boldsymbol{F}^{-1} \cdot \boldsymbol{\sigma} \cdot \boldsymbol{F}^{-\mathsf{T}}. \tag{12}$$

With typical measures for both strains and stresses being established, constitutive relations provide the missing link between kinematics and material response. Throughout this chapter, only homogeneous bodies undergoing purely elastic deformation processes without internal dissipation are considered. Moreover, the existence of a so-called strain energy function or elastic potential $\Psi(\boldsymbol{F})$ is assumed, which only depends upon the current state of deformation (*hyperelastic* material behavior). The requirement of objectivity implies that Ψ remains unchanged when an arbitrary rigid

body rotation is applied to the current configuration. A common formulation of hyperelastic materials in the reference frame then follows as

$$\boldsymbol{S} = \frac{\partial \Psi}{\partial \boldsymbol{E}}. \tag{13}$$

The relation between $\boldsymbol{S}$ and $\boldsymbol{E}$ given by (13) will in general be nonlinear. Thus, it is possible (and necessary within typical finite element procedures) to determine the fourth-order material elasticity tensor $\mathscr{C}_{\mathsf{m}}$ via repeated derivation, yielding

$$\mathscr{C}_{\mathsf{m}} = \frac{\partial \boldsymbol{S}}{\partial \boldsymbol{E}} = \frac{\partial^2 \Psi}{\partial \boldsymbol{E}\, \partial \boldsymbol{E}}. \tag{14}$$

Exemplarily, only one prevailing constitutive model is presented here: the St.-Venant–Kirchhoff material model is an isotropic, hyperelastic model based on a quadratic strain energy function

$$\Psi_{\mathsf{SVK}} = \frac{\lambda}{2}(\mathrm{tr}\, \boldsymbol{E})^2 + \mu \boldsymbol{E} : \boldsymbol{E}. \tag{15}$$

In this context, λ and μ represent the so-called Lamé parameters, which are correlated with the more common Young's modulus E and Poisson's ratio ν via

$$\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)}, \quad \mu = \frac{E}{2(1+\nu)}. \tag{16}$$

Inserting (15) into (13) and (14), it can easily be observed that the St.-Venant–Kirchhoff material model defines a linear relationship between Green–Lagrange strains $\boldsymbol{E}$ and second Piola–Kirchhoff stresses $\boldsymbol{S}$, and can therefore be interpreted as an objective generalization of Hooke's law to the geometrically nonlinear realm. Many other constitutive laws exist for miscellaneous applications (e.g. the well-known Neo–Hookean, Mooney–Rivlin or Ogden models for rubber materials). However, with the focus of this chapter being on contact interaction rather than constitutive modeling, the interested reader is referred to the abundant literature on hyperelasticity, viscoelasticity or elastoplasticity for further details, e.g. in Holzapfel (2000), Ogden (1997) and Simo and Hughes (1998).

3.3 Initial Boundary Value Problem

Exemplarily, the IBVP will be presented in the reference configuration here, however the spatial description is derived analogously. For the definition of suitable boundary conditions, $\partial\Omega_0$ is decomposed into two complementary sets in the absence of contact: Γ_σ represents the Neumann boundary, where the tractions $\hat{\boldsymbol{t}}_0$ are given, and Γ_{u} denotes the Dirichlet boundary, where displacements $\hat{\boldsymbol{u}}$ are prescribed. Neumann and Dirichlet boundaries are disjoint sets, i.e.

$$\Gamma_\sigma \cup \Gamma_\mathsf{u} = \partial\Omega_0, \quad \Gamma_\sigma \cap \Gamma_\mathsf{u} = \emptyset. \tag{17}$$

The initial boundary value problem in material description can be summarized as follows :

$$\mathrm{Div}\boldsymbol{P} + \hat{\boldsymbol{b}}_0 = \rho_0 \ddot{\boldsymbol{u}} \qquad \text{in } \Omega_0 \times [0, T], \tag{18}$$

$$\boldsymbol{u} = \hat{\boldsymbol{u}} \qquad \text{on } \Gamma_\mathsf{u} \times [0, T], \tag{19}$$

$$\boldsymbol{P} \cdot \boldsymbol{N} = \hat{\boldsymbol{t}}_0 \qquad \text{on } \Gamma_\sigma \times [0, T]. \tag{20}$$

Herein, T denotes the end of the considered time interval. Due to the time dependency within the balance of linear momentum in (18), which contains second derivatives with respect to time t, suitable initial conditions for the displacements $\hat{\boldsymbol{u}}_0(\boldsymbol{X})$ and velocities $\hat{\dot{\boldsymbol{u}}}_0(\boldsymbol{X})$ at time $t = 0$ are needed, viz.

$$\boldsymbol{u}(\boldsymbol{X}, 0) = \hat{\boldsymbol{u}}_0(\boldsymbol{X}) \qquad \text{in } \Omega_0, \tag{21}$$

$$\dot{\boldsymbol{u}}(\boldsymbol{X}, 0) = \hat{\dot{\boldsymbol{u}}}_0(\boldsymbol{X}) \qquad \text{in } \Omega_0. \tag{22}$$

The definition of a material model, such as for instance the one given in (15), eventually rounds off the initial boundary value problem of finite deformation solid mechanics. The IBVP is also commonly referred to as *strong* formulation of nonlinear solid mechanics, as Eqs. (18)–(22) are enforced at each individual point within the domain Ω_0.

3.4 Contact Kinematics

From the viewpoint of mathematical problem formulation, contact and impact procedures can be classified into several different categories. A problem setup consisting of one single deformable body and a rigid obstacle is commonly referred to as Signorini contact, while the typical general problem formulation rests upon the assumption of two deformable bodies undergoing contact interaction. Moreover, self contact and contact involving multiple bodies represent well-known special cases. While it is usually advantageous or even essential to design specific numerical algorithms for the aforementioned special cases, all mathematical basics concerning contact kinematics and contact constraints can yet be perfectly derived for the case of two deformable bodies.

Hence, deformable-deformable contact of two bodies undergoing finite deformations, as illustrated in Fig. 2, serves as prototype exclusively considered here. Let the open sets $\Omega_0^{(1)}, \Omega_0^{(2)} \subset \mathbb{R}^3$ and $\Omega_t^{(1)}, \Omega_t^{(2)} \subset \mathbb{R}^3$ represent two bodies in the reference and current configuration, respectively. As the two bodies approach each other and may potentially come into contact on parts of their boundaries, the surfaces $\partial\Omega_0^{(i)}$, $i = 1, 2$, are now divided into three disjoint subsets, viz.

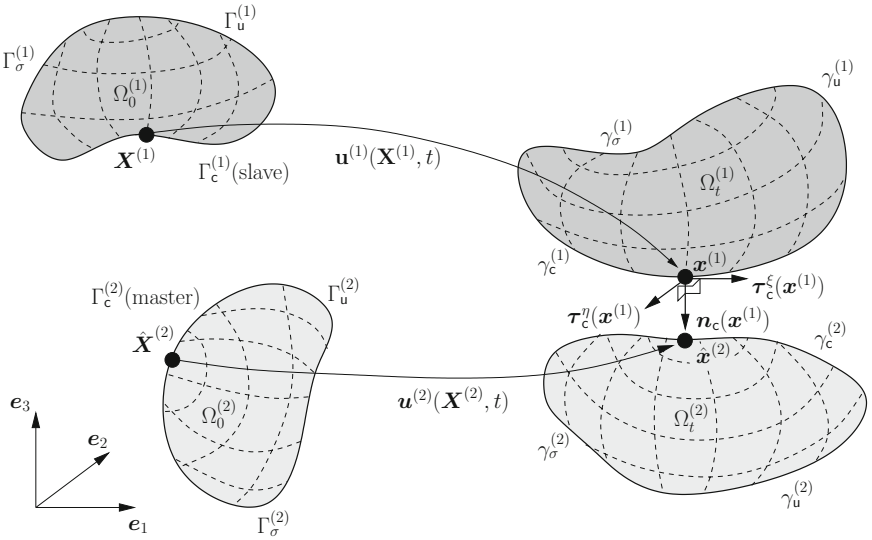

Fig. 2 Kinematics and basic notation for a two body unilateral contact problem in 3D

$$\begin{aligned} \partial\Omega_0^{(i)} &= \Gamma_{\mathsf{u}}^{(i)} \cup \Gamma_{\sigma}^{(i)} \cup \Gamma_{\mathsf{c}}^{(i)}, \\ \Gamma_{\mathsf{u}}^{(i)} \cap \Gamma_{\sigma}^{(i)} &= \Gamma_{\mathsf{u}}^{(i)} \cap \Gamma_{\mathsf{c}}^{(i)} = \Gamma_{\sigma}^{(i)} \cap \Gamma_{\mathsf{c}}^{(i)} = \emptyset, \end{aligned} \tag{23}$$

where $\Gamma_{\mathsf{u}}^{(i)}$ and $\Gamma_{\sigma}^{(i)}$ are the well-known Dirichlet and Neumann boundaries, and $\Gamma_{\mathsf{c}}^{(i)}$ represents the *potential* contact surface. The counterparts in the current configuration are denoted as $\gamma_{\mathsf{u}}^{(i)}$, $\gamma_{\sigma}^{(i)}$ and $\gamma_{\mathsf{c}}^{(i)}$. It is characteristic of contact problems that the *actual*, so-called active contact surface $\Gamma_{\mathsf{a}}^{(i)} \subseteq \Gamma_{\mathsf{c}}^{(i)}$ is unknown, possibly continuously changing over time and thus has to be determined as part of the nonlinear solution process. For the sake of completeness, and to be mathematically precise, the currently inactive contact surface $\Gamma_{\mathsf{i}}^{(i)} = \Gamma_{\mathsf{c}}^{(i)} \setminus \Gamma_{\mathsf{a}}^{(i)}$ should technically be interpreted as part of the Neumann boundary $\Gamma_{\sigma}^{(i)}$.

A classical nomenclature in contact mechanics is retained throughout this chapter by referring to $\Gamma_{\mathsf{c}}^{(1)}$ as the *slave* surface and to $\Gamma_{\mathsf{c}}^{(2)}$ as the *master* surface, although the master-slave concept actually only makes sense in the context of finite element discretization and although its traditional meaning will not be entirely conveyed to the mortar FE approach presented later on.

Both bodies are required to satisfy the initial boundary value problem previously presented in Sect. 3.3, with the motion and deformation being described by the absolute displacement vectors $\boldsymbol{u}^{(i)} = \boldsymbol{x}^{(i)} - \boldsymbol{X}^{(i)}$. Moreover, a new fundamental geometric measure for proximity, potential contact and penetration of the two bodies is introduced with the so-called gap function $\boldsymbol{g}_{\mathsf{n}}(\boldsymbol{X}, t)$ in the current configuration. It is evident that the gap function and other contact-related quantities need to be examined in a spatial description, even though the IBVP may still be formulated with

respect to the reference configuration. The gap function is defined as

$$g_{\mathsf{n}}(\boldsymbol{X},t) = -\boldsymbol{n}_{\mathsf{c}} \cdot \left[\boldsymbol{x}^{(1)}(\boldsymbol{X}^{(1)},t) - \hat{\boldsymbol{x}}^{(2)}(\widehat{\boldsymbol{X}}^{(2)}(\boldsymbol{X}^{(1)},t),t) \right], \tag{24}$$

where some alternatives exist for the identification of the contact point $\hat{\boldsymbol{x}}^{(2)}$ on the master surface associated with each point $\boldsymbol{x}^{(1)}$ on the slave surface and also for the corresponding contact normal vector $\boldsymbol{n}_{\mathsf{c}}$. The classical and perhaps most intuitive choice in contact mechanics is based on the so-called closest point projection (CPP), which determines $\hat{\boldsymbol{x}}^{(2)}$ as

$$\hat{\boldsymbol{x}}^{(2)} = \arg \min_{\boldsymbol{x}^{(2)} \in \gamma_{\mathsf{c}}^{(2)}} \|\boldsymbol{x}^{(1)} - \boldsymbol{x}^{(2)}\|. \tag{25}$$

Consequently, $\boldsymbol{n}_{\mathsf{c}}$ is then chosen to be the outward unit normal to the current master surface $\gamma_{\mathsf{c}}^{(2)}$ in $\hat{\boldsymbol{x}}^{(2)}$. A very comprehensive overview of the closest point projection, its mathematical properties and possible pitfalls due to non-uniqueness and certain pathological cases can be found in Konyukhov and Schweizerhof (2008). However, a slightly different approach is followed here, with the outward unit normal to the current *slave* surface $\gamma_{\mathsf{c}}^{(1)}$ being considered as contact normal $\boldsymbol{n}_{\mathsf{c}}$. Hence, the master side contact point $\hat{\boldsymbol{x}}^{(2)}$ is the result of a smooth interface mapping $\chi : \gamma_{\mathsf{c}}^{(1)} \rightarrow \gamma_{\mathsf{c}}^{(2)}$ of $\boldsymbol{x}^{(1)}$ onto the master surface $\gamma_{\mathsf{c}}^{(2)}$ along $\boldsymbol{n}_{\mathsf{c}}$, see Fig. 2. Especially in the context of mortar finite element discretization, this choice has some practical advantages over the classical closest point projection common for node-to-segment discretization.

Together with two vectors $\boldsymbol{\tau}_{\mathsf{c}}^{\xi}$ and $\boldsymbol{\tau}_{\mathsf{c}}^{\eta}$ taken from the tangential plane, $\boldsymbol{n}_{\mathsf{c}}$ forms a set of orthonormal basis vectors in the slave surface point $\boldsymbol{x}^{(1)}$. As these basis vector are attached to $\boldsymbol{x}^{(1)}$ and also move accordingly, they are commonly referred to as slip advected basis vectors. In this context, it is worth noting that the contact surface $\gamma_{\mathsf{c}}^{(1)}$ is a two-dimensional manifold, which means that the tangential plane in each point $\boldsymbol{x}^{(1)}$ locally defines an $\mathbb{R}^2$ space embedded into the global $\mathbb{R}^3$. Therefore, any quantity on $\gamma_{\mathsf{c}}^{(1)}$ is readily parametrized with the two local coordinates $\xi(\boldsymbol{X}^{(1)},t)$ and $\eta(\boldsymbol{X}^{(1)},t)$. While the gap function characterizes contact interaction in normal direction, the primary kinematic variable for frictional sliding in tangential direction is given by the relative tangential velocity

$$\boldsymbol{v}_{\tau,\mathsf{rel}} = (\boldsymbol{I} - \boldsymbol{n}_{\mathsf{c}} \otimes \boldsymbol{n}_{\mathsf{c}}) \cdot \left[\dot{\boldsymbol{x}}^{(1)}(\boldsymbol{X}^{(1)},t) - \dot{\hat{\boldsymbol{x}}}^{(2)}(\widehat{\boldsymbol{X}}^{(2)}(\boldsymbol{X}^{(1)},t),t) \right]. \tag{26}$$

Note that this expression for $\boldsymbol{v}_{\tau,\mathsf{rel}}$ is only exact in the case of perfect sliding and persistent contact, i.e. assuming $g_{\mathsf{n}} = \dot{g}_{\mathsf{n}} = 0$. Nevertheless, it is typically employed for quantifying the relative tangential movement of contacting bodies in all cases, even if the described prerequisites are not met exactly. To clarify the notation in (26), it is pointed out that $\dot{\hat{\boldsymbol{x}}}^{(2)}$ represents the current velocity of the material point $\widehat{\boldsymbol{X}}^{(2)}$, viz. the material contact point associated with $\boldsymbol{X}^{(1)}$ at time t. Therefore, it does *not* include a change of the material contact point $\widehat{\boldsymbol{X}}^{(2)}$ itself, or in other words, it does

not include a change of the CPP of slave point $\boldsymbol{x}^{(1)}$. Based on the tangential plane defined above, $\boldsymbol{v}_{\tau,\mathsf{rel}}$ can be decomposed into

$$\boldsymbol{v}_{\tau,\mathsf{rel}} = v_\tau^\xi \boldsymbol{\tau}_\mathsf{c}^\xi + v_\tau^\eta \boldsymbol{\tau}_\mathsf{c}^\eta. \tag{27}$$

As already mentioned, the definition of the relative tangential velocity given above is only frame-indifferent when perfect sliding occurs ($\boldsymbol{g}_\mathsf{n} = 0$), see e.g. Laursen (2002). However, since an objective measure of the slip rate is essential for formulating frictional contact conditions in finite deformation formulations, an appropriate algorithmic modification of the slip rate is typically carried out later in the course of finite element discretization.

Similar to the kinematic measures $\boldsymbol{g}_\mathsf{n}$ and $\boldsymbol{v}_{\tau,\mathsf{rel}}$, the contact traction $\boldsymbol{t}_\mathsf{c}^{(1)}$ on the slave surface $\gamma_\mathsf{c}^{(1)}$ can be split into normal and tangential components, yielding

$$\boldsymbol{t}_\mathsf{c}^{(1)} = p_\mathsf{n} \boldsymbol{n}_\mathsf{c} + \boldsymbol{t}_\tau = p_\mathsf{n} \boldsymbol{n}_\mathsf{c} + t_\tau^\xi \boldsymbol{\tau}_\mathsf{c}^\xi + t_\tau^\eta \boldsymbol{\tau}_\mathsf{c}^\eta. \tag{28}$$

Moreover, due to the balance of linear momentum on the contact interface, the traction vectors on slave side $\gamma_\mathsf{c}^{(1)}$ and master side $\gamma_\mathsf{c}^{(2)}$ are identical except for opposite signs, i.e.

$$\boldsymbol{t}_\mathsf{c}^{(1)} = -\boldsymbol{t}_\mathsf{c}^{(2)}. \tag{29}$$

For further details on these topics, the interested reader is referred to classical textbooks on contact mechanics, e.g. Johnson (1985) and Kikuchi and Oden (1988), or to more recent monographs on computational methods for contact mechanics, e.g. Laursen (2002) and Wriggers (2006).

3.5 Tied Contact Constraints

While the main focus of this chapter is on unilateral contact problems, the integration of mesh tying or tied contact problems for connecting dissimilar meshes suggests itself due to the numerous conceptual similarities. Mesh tying applications are also closely connected to the notion of domain decomposition. Thus, in Sect. 5, mesh tying serves as simplified model problem through which many methodological and later also implementational aspects of computational contact mechanics can be clearly illustrated.

As will be seen in the following, mesh tying (or tied contact) perfectly fits into the framework of contact kinematics defined above and can simply be interpreted as a special case from now on. The fundamental kinematic measure for mesh tying is simply the relative displacement between the two bodies, sometimes also referred to as gap vector $\boldsymbol{g}(\boldsymbol{X}, t)$, viz.

$$\boldsymbol{g}(\boldsymbol{X}, t) = \boldsymbol{u}^{(1)}(\boldsymbol{X}^{(1)}, t) - \hat{\boldsymbol{u}}^{(2)}(\widehat{\boldsymbol{X}}^{(2)}(\boldsymbol{X}^{(1)}, t), t). \tag{30}$$

Since it is typically assumed that the two bodies to be tied together share a common interface $\Gamma_{\mathrm{c}}^{(1)} \equiv \Gamma_{\mathrm{c}}^{(2)} \equiv \Gamma_{\mathrm{c}}$ in the reference configuration, the gap vector is equivalently expressed as

$$\boldsymbol{g}(\boldsymbol{X}, t) = \boldsymbol{x}^{(1)}(\boldsymbol{X}^{(1)}, t) - \hat{\boldsymbol{x}}^{(2)}(\widehat{\boldsymbol{X}}^{(2)}(\boldsymbol{X}^{(1)}, t), t), \tag{31}$$

thus demonstrating the similarity with the scalar gap function $g_{\mathrm{n}}(\boldsymbol{X}, t)$ for unilateral contact defined in (24) even more clearly. As compared with unilateral contact, mesh tying firstly requires no distinction between normal and tangential directions at the interface, and secondly results in a simple vector-valued equality constraint:

$$\boldsymbol{g}(\boldsymbol{X}, t) = \boldsymbol{0}. \tag{32}$$

3.6 Normal Contact Constraints

After the short interlude on mesh tying, the focus in now again set on unilateral contact conditions. Examining the gap function defined in (24) in more detail, it becomes obvious that a positive value $g_{\mathrm{n}}(\boldsymbol{X}, t) > 0$ characterizes points currently not in contact, while a negative value $g_{\mathrm{n}}(\boldsymbol{X}, t) < 0$ denotes the (physically non-admissible) state of penetration. Therefore, the classical set of Karush–Kuhn–Tucker (KKT) conditions, commonly also referred to as Hertz–Signorini–Moreau (HSM) conditions for frictionless contact on the contact boundary can be stated as

$$g_{\mathrm{n}}(\boldsymbol{X}, t) \geq 0, \quad p_{\mathrm{n}}(\boldsymbol{X}, t) \leq 0, \quad p_{\mathrm{n}}(\boldsymbol{X}, t)\, g_{\mathrm{n}}(\boldsymbol{X}, t) = 0. \tag{33}$$

As can be seen from Fig. 3, the KKT conditions not only define a non-smooth and nonlinear contact law, but one that is multi-valued at $g_{\mathrm{n}}(\boldsymbol{X}, t) = 0$. However, this set of inequality conditions also allows for a very intuitive physical interpretation. Due

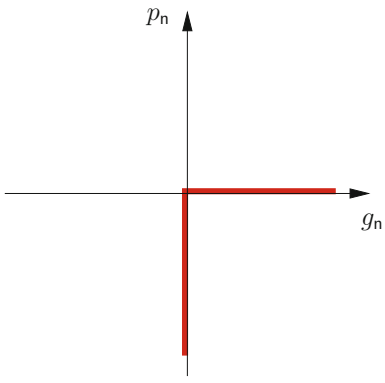

Fig. 3 Karush–Kuhn–Tucker (KKT) conditions of non-penetration

to the sign convention of the gap function introduced here, the first KKT condition simply represents the geometric constraint of non-penetration, whereas the second KKT condition implies that no adhesive stresses are allowed in the contact zone. Finally, the third KKT condition, well-known as complementarity condition, forces the gap to be closed when non-zero contact pressure occurs (contact) and the contact pressure to be zero when the gap is open (no contact). Note, that the type of KKT conditions defined in (33) also arise in many other problem classes of constrained optimization, and thus standard solution techniques (e.g. based on Lagrange multiplier methods and active set strategies) from optimization theory can readily be adapted for contact mechanics.

For the sake of completeness, the so-called persistency condition is also mentioned here. In the context of contact *dynamics*, the persistency condition is sometimes considered as an additional contact condition, requiring that

$$p_{\mathsf{n}}(\boldsymbol{X}, t)\, \dot{g}_{\mathsf{n}}(\boldsymbol{X}, t) = 0. \tag{34}$$

Herein, $\dot{g}_{\mathsf{n}}(\boldsymbol{X}, t)$ represents the material time derivative of the gap function. Therefore, the persistency condition in combination with the KKT conditions in (33) basically demands that the contact pressure is only non-zero when the bodies are in contact and also remain so (persistent contact). On the contrary, the contact pressure is zero in the instant of bodies coming into contact and in the instant of separation. The persistency condition plays an important role in the design of energy conserving numerical algorithms for contact dynamics, see e.g. Laursen and Chawla (1997), Laursen and Love (2002), and bears a certain resemblance to the consistency condition in plasticity, see e.g. Simo and Hughes (1998).

3.7 Frictional Contact Constraints

While frictionless response (i.e. $\boldsymbol{t}_\tau = \boldsymbol{0}$) is a common modeling assumption, and especially helpful for a thorough development of computational methods for contact mechanics, the real contact behavior of many technical systems is determined by the frictional response to tangential loading. The associated scientific field of tribology is extremely broad, also encompassing physical phenomena such as adhesion, wear or elastohydrodynamic lubrication. The following overview is restricted to a purely macroscopic observation of dry friction, classically described by Coulomb's law. One possible and widely used notation of Coulomb friction is given by

$$\Phi := \|\boldsymbol{t}_\tau\| - \mathfrak{F}|p_{\mathsf{n}}| \leq 0, \quad \boldsymbol{v}_{\tau,\mathsf{rel}} + \beta \boldsymbol{t}_\tau = \boldsymbol{0}, \quad \beta \geq 0, \quad \Phi\beta = 0. \tag{35}$$

Herein, $\|\cdot\|$ denotes the L^2-norm in $\mathbb{R}^3$, $\mathfrak{F} \geq 0$ is the friction coefficient and $\beta \geq 0$ is a scalar parameter. An intuitive physical interpretation of Coulomb's law as described in (35) is readily available, too. The first (inequality) condition, commonly referred to as slip condition, requires that the magnitude of the tangential stress $\boldsymbol{t}_\tau$ does not

exceed a threshold defined by the coefficient of friction $\mathfrak{F}$ and the normal contact pressure p_{n}. The frictional response is then characterized by two physically distinct situations. The stick state, defined by $\beta = 0$, does not allow for any relative tangential movement in the contact zone, i.e. $\boldsymbol{v}_{\tau,\mathsf{rel}} = \boldsymbol{0}$. In contrast, the slip state, defined by $\beta > 0$, implicates relative tangential sliding of the two bodies in accordance with the so-called slip rule given as second equation in (35). The last equation in (35) is again a complementarity condition, here separating the two independent solution branches of stick and slip. A commonly cited similarity of Coulomb's law exists with the most simple formulations of elastoplasticity, see e.g. Simo and Hughes (1998). This similarity is especially interesting in the course of developing numerical algorithms for friction, which usually reuse well-known methodologies from computational inelasticity.

Finally, it is pointed out that frictional response in contact is a path-dependent process, thus introducing mechanical dissipation and making a system representation based on elastic potentials infeasible. Path-dependency can easily be observed in the fact that the tangential contact traction $\boldsymbol{t}_\tau$ depends on the velocity $\boldsymbol{v}_{\tau,\mathsf{rel}}$ or on the rate of change of the tangential displacement if interpreted incrementally.

4 Overview of Nonlinear FEM

This section provides a brief introduction to the numerical treatment of nonlinear solid mechanics problems with finite element methods. Based on a weak formulation of the previously derived IBVP, the FEM for space discretization as well as typical implicit time stepping schemes for time discretization are presented. Large parts of this section are based on the author's previously published work (Popp 2012).

4.1 From Strong Formulation to Weak Formulation

Many numerical methods for the solution of partial differential equations, and finite element methods in particular, require a transformation of the IBVP defined in (18)–(22) within a so-called weak or variational formulation. Although other variational principles exist, the well-known *principle of virtual work (PVW)* is derived exclusively here, with the starting point being a weighted residual notation of the balance equation (18) and the traction boundary condition (20), i.e.

$$\int_{\Omega_0} (\rho_0 \ddot{\boldsymbol{u}} - \mathrm{Div}\boldsymbol{P} - \hat{\boldsymbol{b}}_0) \cdot \boldsymbol{w} \, \mathrm{d}V_0 + \int_{\Gamma_\sigma} (\boldsymbol{P} \cdot \boldsymbol{N} - \hat{\boldsymbol{t}}_0) \cdot \boldsymbol{w} \, \mathrm{d}A_0 = 0. \tag{36}$$

Herein, the weighting or test functions $\boldsymbol{w}$ are initially arbitrary and can be interpreted as virtual displacements, i.e. $\boldsymbol{w} = \delta\boldsymbol{u}$. Since the solution for the displacements is known on the Dirichlet boundary Γ_{u}, it is required that

$$\boldsymbol{w} = \boldsymbol{0} \quad \text{on } \Gamma_{\mathsf{u}} \times [0, T]. \tag{37}$$

Applying Gauss divergence theorem and inserting (37) and (12) yields

$$\underbrace{\int_{\Omega_0} \rho_0 \ddot{\boldsymbol{u}} \cdot \delta \boldsymbol{u} \, \mathrm{d}V_0}_{-\delta \mathcal{W}_{\mathsf{kin}}} + \underbrace{\int_{\Omega_0} \boldsymbol{S} : \delta \boldsymbol{E} \, \mathrm{d}V_0}_{-\delta \mathcal{W}_{\mathsf{int}}} \underbrace{- \int_{\Omega_0} \hat{\boldsymbol{b}}_0 \cdot \delta \boldsymbol{u} \, \mathrm{d}V_0 - \int_{\Gamma_\sigma} \hat{\boldsymbol{t}}_0 \cdot \delta \boldsymbol{u} \, \mathrm{d}A_0}_{-\delta \mathcal{W}_{\mathsf{ext}}} = 0. \tag{38}$$

Three distinct contributions to the PVW can be identified. The first term in (38) represents the kinetic virtual work contribution $\delta \mathcal{W}_{\mathsf{kin}}$, the second term denotes the internal virtual work contribution $\delta \mathcal{W}_{\mathsf{int}}$, and the third and fourth term together form the virtual work of external loads $\delta \mathcal{W}_{\mathsf{ext}}$. The PVW emerges as a very general principle of solid mechanics, as it does not require the existence of an associated potential $\mathcal{W}$. As an example, no constitutive assumptions whatsoever enter the weak formulation in (38), thus making it also valid and applicable for problems such as elastoplasticity, frictional sliding or non-conservative loading.

It can easily be shown that solutions of the IBVP (i.e. of the strong formulation) also satisfy the weak formulation (38). As long as no restrictions are set on the choice of the weighting functions $\delta \boldsymbol{u}$, the two are formally identical, see e.g. Hughes (2000). However, due to the manipulations introduced above, the weak formulation poses weaker differentiability requirements to the solution functions $\boldsymbol{u}$, because only first derivatives of $\boldsymbol{u}$ with respect to $\boldsymbol{X}$ appear in (38) instead of second derivatives as in (18). Thus, the following solution and weighting spaces can be defined:

$$\mathcal{U} = \left\{ \boldsymbol{u} \in H^1(\Omega) \mid \boldsymbol{u}(\boldsymbol{X}, t) = \hat{\boldsymbol{u}}(\boldsymbol{X}, t) \text{ on } \Gamma_{\mathsf{u}} \right\}, \tag{39}$$

$$\mathcal{V} = \left\{ \delta \boldsymbol{u} \in H^1(\Omega) \mid \delta \boldsymbol{u}(\boldsymbol{X}) = \boldsymbol{0} \text{ on } \Gamma_{\mathsf{u}} \right\}. \tag{40}$$

Herein, $H^1(\Omega)$ denotes the Sobolev space of functions with square integrable values and first derivatives. While the solution space $\mathcal{U}$ may in general depend on the time t due to a possible time dependency of the Dirichlet boundary conditions, the weighting space $\mathcal{V}$ does not depend on the time t in any way. In conclusion, the weak formulation of the nonlinear solid mechanics problems at hand can be restated as follows: Find $\boldsymbol{u} \in \mathcal{U}$ such that

$$\delta \mathcal{W} = 0 \quad \forall \, \delta \boldsymbol{u} \in \mathcal{V}. \tag{41}$$

4.2 Space Discretization

Space discretization is exclusively considered in the context of finite element methods here. However, as a detailed introduction to all important aspects of the FEM is beyond the scope of this chapter, only the basic ideas and notation will be highlighted. For a more elaborate survey of finite element methods, the reader is again referred to

the corresponding literature, e.g. in Bathe (1996), Hughes (2000), Belytschko et al. (2000), Reddy (2004), Zienkiewicz and Taylor (2005) and Zienkiewicz et al. (2005).

Simply speaking, the concept of finite element discretization in this context is based on finding a numerical solution to (41) at discrete points, commonly referred to as nodes. The nodes are connected to form elements, which allows to formulate the following approximate partitioning of the domain Ω_0 into nele element subdomains:

$$\Omega_0 \approx \bigcup_{e=1}^{\mathsf{nele}} \Omega_0^{(e)}. \tag{42}$$

The displacement solution $\boldsymbol{u}^{(e)}$ on element e is then typically approximated by local interpolation functions $N_k(\boldsymbol{X})$, yielding

$$\boldsymbol{u}^{(e)}(\boldsymbol{X},t) \approx \boldsymbol{u}_h^{(e)}(\boldsymbol{X},t) = \sum_{k=1}^{\mathsf{nnod}^{(e)}} N_k(\boldsymbol{X})\mathbf{d}_k(t), \tag{43}$$

where the discrete nodal values of the displacements $\mathbf{d}_k(t)$ have been introduced. Furthermore, the subscript $\cdot_h$ signifies a spatially discretized quantity throughout this chapter and $\mathsf{nnod}^{(e)}$ represents the number of nodes associated with the element e. The interpolation functions $N_k(\boldsymbol{X})$, commonly referred to as shape functions, are typically (but not exclusively) low-order polynomials, e.g. Lagrange polynomials, thus meeting the differentiability requirements of the weak form. Based on the so-called isoparametric concept, the element geometry in the reference configuration $\boldsymbol{X}^{(e)}$ and current configuration $\boldsymbol{x}^{(e)}$ is approximated using the same shape functions. Typically, $\Omega_0^{(e)}$ is mapped to a reference element geometry or parameter space $\boldsymbol{\xi} = (\xi, \eta, \zeta)$, e.g. the cube $[-1, 1] \times [-1, 1] \times [-1, 1]$, which defines an element Jacobian matrix $\boldsymbol{J}^{(e)} = \partial \boldsymbol{X}^{(e)} / \partial \boldsymbol{\xi}$. Thus, the interpolation of displacements, current geometry and reference geometry at the element level is alternatively expressed as

$$\boldsymbol{u}_h^{(e)}(\boldsymbol{\xi},t) = \sum_{k=1}^{\mathsf{nnod}^{(e)}} N_k(\boldsymbol{\xi})\mathbf{d}_k(t), \tag{44}$$

$$\boldsymbol{x}_h^{(e)}(\boldsymbol{\xi},t) = \sum_{k=1}^{\mathsf{nnod}^{(e)}} N_k(\boldsymbol{\xi})\mathbf{x}_k(t), \tag{45}$$

$$\boldsymbol{X}_h^{(e)}(\boldsymbol{\xi}) = \sum_{k=1}^{\mathsf{nnod}^{(e)}} N_k(\boldsymbol{\xi})\mathbf{X}_k, \tag{46}$$

with nodal positions $\mathbf{X}_k$ and $\mathbf{x}_k(t)$ in the reference and current configuration, respectively. Finally, time derivatives of the displacements, e.g. the accelerations $\ddot{\boldsymbol{u}}$, and the weighting functions $\delta\boldsymbol{u}$ are also interpolated using the same shape functions. The latter convention is commonly referred to as Bubnov–Galerkin approach, as compared

with a Petrov–Galerkin approach, where an independent set of shape functions is chosen for interpolating the weighting functions.

Examining (44) more closely, it becomes obvious that the finite element method basically introduces restrictions on the solution and weighting spaces defined in (39) and (40). In the discrete setting, these spaces only contain a finite number of solution and weighting functions, respectively, which is expressed mathematically in terms of finite dimensional subspaces $\mathcal{U}_h \subset \mathcal{U}$ and $\mathcal{V}_h \subset \mathcal{V}$. The limited selection of solution and weighting functions then serves as a basis for the numerical solution, i.e. the weak formulation is recast into a discrete form, which is no longer equivalent to strong and weak formulation, but rather represents an approximation.

The individual contributions to the discretized weak form are integrated element-by-element using Gauss quadrature and then sorted into global vectors based on the so-called assembly operator, which governs the arrangement of local vectorial quantities into global vectors. After inserting the interpolations given by (44) into the weak formulation (38), the final spatially discretized formulation emerges as

$$\delta\mathbf{d}^{\mathsf{T}}(\mathbf{M}\ddot{\mathbf{d}} + \mathbf{f}_{\mathsf{int}}(\mathbf{d}) - \mathbf{f}_{\mathsf{ext}}) = 0, \tag{47}$$

with the global mass matrix $\mathbf{M}$, the global vector of nonlinear internal forces $\mathbf{f}_{\mathsf{int}}$ and the global vector of external forces $\mathbf{f}_{\mathsf{ext}}$. Moreover, $\delta\mathbf{d}$, $\ddot{\mathbf{d}}$ and $\mathbf{d}$ are global vectors comprising all discrete nodal values of virtual displacements, accelerations and displacements. Due to the interpolation introduced above, all vectors in (47) are of the size $\mathsf{ndof} = \mathsf{ndim} \cdot \mathsf{nnod}$, where nnod is the total number of nodes in the entire domain and ndim is the number of spatial dimensions. The variable name ndof refers to the fact that the discrete values of the nodal displacements $\mathbf{d}$ are also denoted as degrees of freedom. Since (47) must hold for arbitrary virtual displacements $\delta\mathbf{d}$, it can equivalently be written as

$$\mathbf{M}\ddot{\mathbf{d}} + \mathbf{f}_{\mathsf{int}}(\mathbf{d}) - \mathbf{f}_{\mathsf{ext}} = \mathbf{0}. \tag{48}$$

This defines a system of ndof ordinary differential equations (ODEs), commonly referred to as semi-discrete equations of motion. So far, only space discretization with the finite element method has been established, but the system is still continuous with respect to time.

4.3 Time Discretization

There exists a large variety of finite difference methods suitable for time discretization of the semi-discrete equations of motion (48). In doing so, time derivatives are approximated by their discrete counterparts, the difference quotients. Based on the introduction of a constant time step size Δt, the time interval of interest $t \in [0, T]$ is subdivided into several intervals $[t_n, t_{n+1}]$, where $n \in \mathbb{N}_0$ is the time step index,

and thus the spatially discretized displacement solution $\mathbf{d}(t)$ is computed at a series of discrete points in time.

In principle, time integration methods can be divided into implicit and explicit schemes. While implicit methods lead to a fully coupled system of ndof nonlinear discrete algebraic equations for the unknown displacements $\mathbf{d}_{n+1} := \mathbf{d}(t_{n+1})$, explicit methods allow for a direct extrapolation towards $\mathbf{d}_{n+1}$ without requiring a solution step. Here, only implicit schemes will be considered. They represent the method of choice for problems dominated by a low frequency response, while explicit methods are widely used in the context of high frequency responses and wave-like phenomena, e.g. in high velocity impact situations. In general, implicit time integration methods can be shown to be unconditionally stable, thus allowing for relatively large time step sizes as compared with explicit schemes. However, the implementation of implicit methods is more challenging due to the fact that nonlinear solution methods (see Sect. 4.4) including a linearization of the entire finite element formulation are required.

Here, the presentation is restricted to one exemplary and widely used implicit time integration scheme, viz. the generalized-α method introduced by Chung and Hulbert (1993). This one-step time integration scheme is based on the well-known Newmark method, which allows for expressing the approximate discrete velocities $\mathbf{v}_{n+1} \approx \dot{\mathbf{d}}(t_{n+1})$ and accelerations $\mathbf{a}_{n+1} \approx \ddot{\mathbf{d}}(t_{n+1})$ at the end of the considered time interval $[t_n, t_{n+1}]$ solely in terms of already known quantities at time t_n and the unknown displacements $\mathbf{d}_{n+1}$, i.e.

$$\mathbf{v}_{n+1}(\mathbf{d}_{n+1}) = \frac{\gamma}{\beta \Delta t}(\mathbf{d}_{n+1} - \mathbf{d}_n) - \frac{\gamma - \beta}{\beta}\mathbf{v}_n - \frac{\gamma - 2\beta}{2\beta}\Delta t \mathbf{a}_n, \tag{49}$$

$$\mathbf{a}_{n+1}(\mathbf{d}_{n+1}) = \frac{1}{\beta \Delta t^2}(\mathbf{d}_{n+1} - \mathbf{d}_n) - \frac{1}{\beta \Delta t}\mathbf{v}_n - \frac{1 - 2\beta}{2\beta}\Delta t \mathbf{a}_n, \tag{50}$$

where $\beta \in [0, 1/2]$ and $\gamma \in [0, 1]$ are two parameters characterizing the behavior of the method. The generalized-α method introduces generalized mid-points $t_{n+1-\alpha_\mathsf{m}}$ and $t_{n+1-\alpha_\mathsf{f}}$ and shifts the evaluation of the individual terms in (48) from t_{n+1} to these midpoints. The following linear interpolation rules are commonly established for the generalized-α method:

$$\mathbf{d}_{n+1-\alpha_\mathsf{f}} = (1 - \alpha_\mathsf{f})\,\mathbf{d}_{n+1} + \alpha_\mathsf{f}\,\mathbf{d}_n, \tag{51}$$

$$\mathbf{v}_{n+1-\alpha_\mathsf{f}} = (1 - \alpha_\mathsf{f})\,\mathbf{v}_{n+1} + \alpha_\mathsf{f}\,\mathbf{v}_n, \tag{52}$$

$$\mathbf{a}_{n+1-\alpha_\mathsf{m}} = (1 - \alpha_\mathsf{m})\,\mathbf{a}_{n+1} + \alpha_\mathsf{m}\,\mathbf{a}_n, \tag{53}$$

$$\mathbf{f}_{\mathsf{ext},n+1-\alpha_\mathsf{f}} = (1 - \alpha_\mathsf{f})\,\mathbf{f}_{\mathsf{ext},n+1} + \alpha_\mathsf{f}\,\mathbf{f}_{\mathsf{ext},n}. \tag{54}$$

Eventually, the fully (i.e. space and time) discretized finite element formulation of nonlinear solid mechanics, also referred to as discrete linear momentum balance, is obtained as

$$\mathbf{M}\mathbf{a}_{n+1-\alpha_\mathsf{m}} + \mathbf{C}\mathbf{v}_{n+1-\alpha_\mathsf{f}} + \mathbf{f}_{\mathsf{int}}(\mathbf{d}_{n+1-\alpha_\mathsf{f}}) - \mathbf{f}_{\mathsf{ext},n+1-\alpha_\mathsf{f}} = \mathbf{0}. \tag{55}$$

One important advantage of the generalized-α method is that it allows for introducing controllable numerical dissipation into the considered system, while at the same time retaining the important properties of unconditional stability and second-order accuracy. Controllable numerical dissipation in this context means that the parameters β, γ, α_{m} and α_{f} can be harmonized such that the desired damping effect is only achieved in the spurious high frequency modes, while damping in the low frequency domain is kept at a minimum. This procedure is usually united in the notion of a spectral radius ρ_∞ as the sole free parameter to choose for a generalized-α method. The other parameters then follow directly from the requirements of unconditional stability, second-order accuracy and optimized numerical dissipation as

$$\alpha_{\mathsf{m}} = \frac{2\rho_\infty - 1}{\rho_\infty + 1}, \quad \alpha_{\mathsf{f}} = \frac{\rho_\infty}{\rho_\infty + 1}, \quad \beta = \frac{1}{4}(1 - \alpha_{\mathsf{m}} + \alpha_{\mathsf{f}})^2, \quad \gamma = \frac{1}{2} - \alpha_{\mathsf{m}} + \alpha_{\mathsf{f}}. \tag{56}$$

Note that no numerical dissipation is introduced into the system for the choice $\rho_\infty = 1$. Moreover, the generalized-α method also contains the classical Newmark method as a special case by setting $\alpha_{\mathsf{m}} = \alpha_{\mathsf{f}} = 0$.

For the sake of completeness, it is pointed out that quasistatic problems, i.e. neglecting inertia effects, are also considered in the following. In that case, the time parameter t only plays the role of a pseudo-time and no time integration method is needed, but the quasistatic solution is rather computed as a series of static equilibrium states.

4.4 Linearization and Solution Techniques for Nonlinear Equations

Within each time step, the system of ndof nonlinear discrete algebraic Eq. (55) needs to be solved for the unknown displacements $\mathbf{d}_{n+1}$. Throughout this contribution, the Newton–Raphson method is employed as an iterative nonlinear solution technique. Within each iteration step i, the residual of the discrete linear momentum balance can be defined as

$$\mathbf{r}_{\mathsf{effdyn}}(\mathbf{d}^i_{n+1}) = \mathbf{M}\mathbf{a}^i_{n+1-\alpha_{\mathsf{m}}} + \mathbf{C}\mathbf{v}^i_{n+1-\alpha_{\mathsf{f}}} + \mathbf{f}_{\mathsf{int}}(\mathbf{d}^i_{n+1-\alpha_{\mathsf{f}}}) - \mathbf{f}_{\mathsf{ext},n+1-\alpha_{\mathsf{f}}}. \tag{57}$$

The Newton–Raphson method is based on repeated linearization of the residual in (57), solution of the resulting linearized system of equations and incremental update of the unknown displacements until a user-defined convergence criterion is met. At first, the linearization is obtained from the truncated Taylor expansion of (57), viz.

$$\mathrm{Lin}\, \mathbf{r}_{\mathsf{effdyn}}(\mathbf{d}^i_{n+1}) = \mathbf{r}_{\mathsf{effdyn}}(\mathbf{d}^i_{n+1}) + \underbrace{\left.\frac{\partial \mathbf{r}_{\mathsf{effdyn}}(\mathbf{d}_{n+1})}{\partial \mathbf{d}_{n+1}}\right|^i}_{\mathbf{K}_{\mathsf{effdyn}}(\mathbf{d}^i_{n+1})} \Delta\mathbf{d}^{i+1}_{n+1}, \tag{58}$$

where the partial derivative of $\mathbf{r}_{\mathrm{effdyn}}(\mathbf{d}_{n+1}^{i})$ with respect to the displacements is commonly referred to as dynamic effective tangential stiffness matrix $\mathbf{K}_{\mathrm{effdyn}}(\mathbf{d}_{n+1}^{i})$ of size $\mathsf{ndof} \times \mathsf{ndof}$. In the context of the generalized-α method, the dynamic effective tangential stiffness matrix can be determined based on Newmark's approximation given in (49) and (50) and the generalized midpoints defined in (51)–(54), yielding

$$\begin{aligned}\mathbf{K}_{\mathrm{effdyn}}(\mathbf{d}_{n+1}^{i}) &= \left.\frac{\partial \mathbf{r}_{\mathrm{effdyn}}(\mathbf{d}_{n+1})}{\partial \mathbf{d}_{n+1}}\right|^{i} = \\ &= \left[\frac{1-\alpha_{\mathsf{m}}}{\beta\Delta t^2}\mathbf{M} + \frac{(1-\alpha_{\mathsf{f}})\gamma}{\beta\Delta t}\mathbf{C} + (1-\alpha_{\mathsf{f}})\mathbf{K}_{\mathsf{T}}(\mathbf{d}_{n+1-\alpha_{\mathsf{f}}})\right]^{i}, \end{aligned} \tag{59}$$

where $\mathbf{K}_{\mathsf{T}}(\mathbf{d}_{n+1-\alpha_{\mathsf{f}}})$ is the tangential stiffness matrix associated with the internal forces as

$$\mathbf{K}_{\mathsf{T}}(\mathbf{d}_{n+1-\alpha_{\mathsf{f}}}) = \frac{\partial \mathbf{f}_{\mathsf{int}}(\mathbf{d}_{n+1-\alpha_{\mathsf{f}}})}{\partial \mathbf{d}_{n+1-\alpha_{\mathsf{f}}}}. \tag{60}$$

To sum up, the Newton–Raphson method provides an iterative procedure for finding the unknown solution $\mathbf{d}_{n+1}$ for which the residual $\mathbf{r}_{\mathrm{effdyn}}(\mathbf{d}_{n+1})$ vanishes. Within each iteration, it is required that

$$\mathrm{Lin}\,\mathbf{r}_{\mathrm{effdyn}}(\mathbf{d}_{n+1}^{i}) \stackrel{!}{=} \mathbf{0}, \tag{61}$$

or in other words, the following linear system of equations has to be solved:

$$\mathbf{K}_{\mathrm{effdyn}}(\mathbf{d}_{n+1}^{i})\Delta\mathbf{d}_{n+1}^{i+1} = -\mathbf{r}_{\mathrm{effdyn}}(\mathbf{d}_{n+1}^{i}). \tag{62}$$

Having solved (62), the displacements $\mathbf{d}_{n+1}^{i+1}$ at the end of the time step can be updated via

$$\mathbf{d}_{n+1}^{i+1} = \mathbf{d}_{n+1}^{i} + \Delta\mathbf{d}_{n+1}^{i+1}, \tag{63}$$

and the iteration counter is increased by one, i.e. $i \rightarrow i+1$. The procedure in (62) and (63) is repeated until a certain user-defined convergence criterion, usually with regard to the L^2-norm of the residual $\|\mathbf{r}_{\mathrm{effdyn}}(\mathbf{d}_{n+1}^{i})\|$, is met. The most advantageous property of the Newton–Raphson method is its local quadratic convergence. This means that if the start solution estimate $\mathbf{d}_{n+1}^{0}$ is sufficiently close to the actual solution $\mathbf{d}_{n+1}$, i.e. within the problem-dependent convergence radius, then the residual norm approaches zero with a quadratic convergence rate.

In this contribution, only exact Newton–Raphson methods are considered as described above or later also their semi-smooth variants for the inclusion of contact constraints. However, the computational cost associated with such an approach can be considerable for nonlinear solid mechanics problems, bearing in mind that it requires a consistent linearization and thus a determination of the tangential stiffness matrix $\mathbf{K}_{\mathsf{T}}(\mathbf{d}_{n+1-\alpha_{\mathrm{f}}})$ within *each* iteration step. In practice, this often leads to the application of quasi-Newton methods or modified Newton methods, which are based on a computationally cheaper approximation of the stiffness matrix (e.g. via secants), but sacrifice optimal convergence behavior. Apart from that, many extensions of the Newton–Raphson method aim at enlarging its local convergence radius. Popular examples of such globalization strategies are line search methods and the pseudo-transient continuation (PTC) technique, see e.g. Gee et al. (2009) and references therein.

5 Mortar Methods for Tied Contact

Mesh tying (also referred to as tied contact) serves as a model problem for the introduction to mortar finite element methods here. The basic motivation for such mortar mesh tying algorithms is to connect dissimilar meshes in nonlinear solid mechanics in a variationally consistent manner. Reasons for the occurrence of non-matching meshes can be manifold and range from different resolution requirements in the individual subdomains over the use of different types of finite element interpolation to the rather practical experience that the submodels to be connected are commonly meshed independently. Further details and a full derivation of all formulations can be found in the author's original work (Popp 2012).

5.1 Strong Formulation

Without loss of generality, only the case of a body with one sole tied contact interface is considered. On each subdomain $\Omega_0^{(i)}$, the initial boundary value problem of finite deformation elastodynamics needs to be satisfied, viz.

$$\mathrm{Div}\boldsymbol{P}^{(i)} + \hat{\boldsymbol{b}}_0^{(i)} = \rho_0^{(i)}\ddot{\boldsymbol{u}}^{(i)} \qquad \text{in } \Omega_0^{(i)} \times [0,T], \tag{64}$$

$$\boldsymbol{u}^{(i)} = \hat{\boldsymbol{u}}^{(i)} \qquad \text{on } \Gamma_{\mathsf{u}}^{(i)} \times [0,T], \tag{65}$$

$$\boldsymbol{P}^{(i)} \cdot \boldsymbol{N}^{(i)} = \hat{\boldsymbol{t}}_0^{(i)} \qquad \text{on } \Gamma_{\sigma}^{(i)} \times [0,T], \tag{66}$$

$$\boldsymbol{u}^{(i)}(\boldsymbol{X}^{(i)},0) = \hat{\boldsymbol{u}}_0^{(i)}(\boldsymbol{X}^{(i)}) \qquad \text{in } \Omega_0^{(i)}, \tag{67}$$

$$\dot{\boldsymbol{u}}^{(i)}(\boldsymbol{X}^{(i)},0) = \hat{\dot{\boldsymbol{u}}}_0^{(i)}(\boldsymbol{X}^{(i)}) \qquad \text{in } \Omega_0^{(i)}. \tag{68}$$

The tied contact constraint, also formulated in the reference configuration, is given as

$$\boldsymbol{u}^{(1)} = \boldsymbol{u}^{(2)} \qquad \text{on } \Gamma_{\mathsf{c}} \times [0, T]. \tag{69}$$

Equations (64)–(69) represent the final strong form of a mesh tying problem in nonlinear solid mechanics. In the course of deriving a weak formulation (see next paragraph), the balance of linear momentum at the mesh tying interface Γ_{c} is typically exploited and a Lagrange multiplier vector field $\boldsymbol{\lambda}$ is introduced, thus setting the basis for a mixed variational approach.

5.2 Weak Formulation

To start the derivation of a weak formulation of (64)–(69), appropriate solution spaces $\boldsymbol{\mathcal{U}}^{(i)}$ and weighting spaces $\boldsymbol{\mathcal{V}}^{(i)}$ need to be defined as

$$\boldsymbol{\mathcal{U}}^{(i)} = \left\{ \boldsymbol{u}^{(i)} \in H^1(\Omega) \mid \boldsymbol{u}^{(i)} = \hat{\boldsymbol{u}}^{(i)} \text{ on } \Gamma_{\mathsf{u}} \right\}, \tag{70}$$

$$\boldsymbol{\mathcal{V}}^{(i)} = \left\{ \delta\boldsymbol{u}^{(i)} \in H^1(\Omega) \mid \delta\boldsymbol{u}^{(i)} = \boldsymbol{0} \text{ on } \Gamma_{\mathsf{u}} \right\}. \tag{71}$$

Moreover, the Lagrange multiplier vector $\boldsymbol{\lambda} = -\boldsymbol{t}_{\mathsf{c}}^{(1)}$, which represents the *negative* slave side contact traction $\boldsymbol{t}_{\mathsf{c}}^{(1)}$ and is supposed to enforce the mesh tying constraint (69), is chosen from a corresponding solution space denoted as $\boldsymbol{\mathcal{M}}$. In terms of its classification in functional analysis, this space represents the dual space of the trace space $\boldsymbol{\mathcal{W}}^{(1)}$ of $\boldsymbol{\mathcal{V}}^{(1)}$. In the given context, this means that $\mathcal{M} = H^{-1/2}(\Gamma_{\mathsf{c}})$ and $\mathcal{W}^{(1)} = H^{1/2}(\Gamma_{\mathsf{c}})$, where $\mathcal{M}$ and $\mathcal{W}^{(1)}$ denote single scalar components of the corresponding vector-valued spaces $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\mathcal{W}}$.

Based on these considerations, a saddle point type weak formulation is derived next. Basically, this can be done by extending the standard weak formulation of nonlinear solid mechanics as defined in (38) to two subdomains and combining it with Lagrange multiplier coupling terms. Find $\boldsymbol{u}^{(i)} \in \boldsymbol{\mathcal{U}}^{(i)}$ and $\boldsymbol{\lambda} \in \boldsymbol{\mathcal{M}}$ such that

$$-\delta\mathcal{W}_{\mathsf{kin,int,ext}}(\boldsymbol{u}^{(i)}, \delta\boldsymbol{u}^{(i)}) - \delta\mathcal{W}_{\mathsf{mt}}(\boldsymbol{\lambda}, \delta\boldsymbol{u}^{(i)}) = 0 \qquad \forall\, \delta\boldsymbol{u}^{(i)} \in \boldsymbol{\mathcal{V}}^{(i)}, \tag{72}$$

$$\delta\mathcal{W}_{\lambda}(\boldsymbol{u}^{(i)}, \delta\boldsymbol{\lambda}) = 0 \qquad \forall\, \delta\boldsymbol{\lambda} \in \boldsymbol{\mathcal{M}}. \tag{73}$$

Herein, the kinetic contribution $\delta\mathcal{W}_{\mathsf{kin}}$, the internal and external contributions $\delta\mathcal{W}_{\mathsf{int,ext}}$ and the mesh tying interface contribution $\delta\mathcal{W}_{\mathsf{mt}}$ to the overall virtual work on the two subdomains, as well as the weak form of the mesh tying constraint $\delta\mathcal{W}_{\lambda}$, have been abbreviated as

$$-\delta \mathcal{W}_{\text{kin}} = \sum_{i=1}^{2} \left[\int_{\Omega_0^{(i)}} \rho_0^{(i)} \ddot{\boldsymbol{u}}^{(i)} \cdot \delta \boldsymbol{u}^{(i)} \, \mathrm{d}V_0 \right], \tag{74}$$

$$-\delta \mathcal{W}_{\text{int,ext}} = \sum_{i=1}^{2} \left[\int_{\Omega_0^{(i)}} \left(\boldsymbol{S}^{(i)} : \delta \boldsymbol{E}^{(i)} - \hat{\boldsymbol{b}}_0^{(i)} \cdot \delta \boldsymbol{u}^{(i)} \right) \mathrm{d}V_0 - \int_{\Gamma_\sigma^{(i)}} \hat{\boldsymbol{t}}_0^{(i)} \cdot \delta \boldsymbol{u}^{(i)} \, \mathrm{d}A_0 \right], \tag{75}$$

$$-\delta \mathcal{W}_{\text{mt}} = \int_{\Gamma_{\text{c}}} \boldsymbol{\lambda} \cdot (\delta \boldsymbol{u}^{(1)} - \delta \boldsymbol{u}^{(2)}) \, \mathrm{d}A_0, \tag{76}$$

$$\delta \mathcal{W}_{\lambda} = \int_{\Gamma_{\text{c}}} \delta \boldsymbol{\lambda} \cdot (\boldsymbol{u}^{(1)} - \boldsymbol{u}^{(2)}) \, \mathrm{d}A_0. \tag{77}$$

It is important to point out that, strictly speaking, the coupling bilinear forms $\delta \mathcal{W}_{\text{mt}}$ and $\delta \mathcal{W}_{\lambda}$ cannot be represented by integrals, because the involved spaces $H^{1/2}(\Gamma_{\text{c}})$ and $H^{-1/2}(\Gamma_{\text{c}})$ do not satisfy the requirements for a proper integral definition. Instead, a mathematically correct notation would use so-called duality pairings $\langle \boldsymbol{\lambda}, (\delta \boldsymbol{u}^{(1)} - \delta \boldsymbol{u}^{(2)}) \rangle_{\Gamma_{\text{c}}}$ and $\langle \delta \boldsymbol{\lambda}, (\boldsymbol{u}^{(1)} - \boldsymbol{u}^{(2)}) \rangle_{\Gamma_{\text{c}}}$, see e.g. Wohlmuth (2000). However, during finite element discretization the solution spaces are restricted to discrete subsets of $L^2(\Gamma_{\text{c}})$ functions, and by then at the latest the coupling terms may be formulated as surface integrals. Moreover, even in the mathematical literature the distinction between duality pairing and integral is not treated consistently, and thus the slightly inaccurate formulation in (76) and (77) is preferred here due to readability.

The coupling terms on Γ_{c} also allow for a direct interpretation in terms of variational formulations and the principle of virtual work. Whereas the contribution in (76) represents the virtual work of the unknown interface tractions $\boldsymbol{\lambda} = -\boldsymbol{t}_{\text{c}}^{(1)} = \boldsymbol{t}_{\text{c}}^{(2)}$, the contribution in (77) ensures a weak, variationally consistent enforcement of the tied contact constraint (69). Unlike for unilateral contact with inequality constraints, there exist no further restrictions on the Lagrange multiplier space $\boldsymbol{\mathcal{M}}$ here (such as e.g. positivity). Nevertheless, the concrete choice of the discrete Lagrange multiplier space $\boldsymbol{\mathcal{M}}_h$ in the context of mortar finite element discretizations is decisive for the stability of the method and for optimal a priori error bounds, cf. Sect. 7.1. Finally, it is pointed out that the weak formulation (72) and (73) possesses all characteristics of saddle point problems and Lagrange multiplier methods.

5.3 Finite Element Discretization

For the spatial discretization of the tied contact problem (72) and (73), standard isoparametric finite elements are employed. This defines the usual finite dimensional subspaces $\boldsymbol{\mathcal{U}}_h^{(i)}$ and $\boldsymbol{\mathcal{V}}_h^{(i)}$ being approximations of $\boldsymbol{\mathcal{U}}^{(i)}$ and $\boldsymbol{\mathcal{V}}^{(i)}$, respectively. Throughout this chapter, both first-order and second-order interpolation is considered with finite element meshes typically consisting of 3-node triangular (*tri3*), 4-node quadri-

lateral (*quad4*), 6-node triangular (*tri6*), 8-node quadrilateral (*quad8*) and 9-node quadrilateral (*quad9*) elements in 2D, and of 4-node tetrahedral (*tet4*), 8-node hexahedral (*hex8*), 10-node tetrahedral (*tet10*), 20-node hexahedral (*hex20*) and 27-node hexahedral (*hex27*) elements in 3D.

With the focus being on the finite element discretization of the coupling terms here, only the geometry, displacement and Lagrange multiplier interpolations on $\Gamma_{\mathrm{c},h}^{(i)}$ will be considered in the following. Discretization of the remaining contributions to (72) is not discussed, but the reader is instead referred to the abundant literature. As explained in Sect. 4.2, the subscript $\cdot_h$ refers to a spatially discretized quantity. Obviously, there exists a connection between the employed finite elements in the domains $\Omega_{0,h}^{(i)}$ and the resulting surface facets on the mesh tying interfaces $\Gamma_{\mathrm{c},h}^{(i)}$. For example, a mixed 3D finite element mesh composed of *tet4* and *hex8* elements yields *tri3* and *quad4* facets on the surface of tied contact. Consequently, the following general form of geometry and displacement interpolation on the discrete mesh tying surfaces holds:

$$\boldsymbol{x}_h^{(1)}|_{\Gamma_{\mathrm{c},h}^{(1)}} = \sum_{k=1}^{n^{(1)}} N_k^{(1)}(\xi^{(1)}, \eta^{(1)})\, \mathbf{x}_k^{(1)}, \qquad \boldsymbol{x}_h^{(2)}|_{\Gamma_{\mathrm{c},h}^{(2)}} = \sum_{l=1}^{n^{(2)}} N_l^{(2)}(\xi^{(2)}, \eta^{(2)})\, \mathbf{x}_l^{(2)}, \tag{78}$$

$$\boldsymbol{u}_h^{(1)}|_{\Gamma_{\mathrm{c},h}^{(1)}} = \sum_{k=1}^{n^{(1)}} N_k^{(1)}(\xi^{(1)}, \eta^{(1)})\, \mathbf{d}_k^{(1)}, \qquad \boldsymbol{u}_h^{(2)}|_{\Gamma_{\mathrm{c},h}^{(2)}} = \sum_{l=1}^{n^{(2)}} N_l^{(2)}(\xi^{(2)}, \eta^{(2)})\, \mathbf{d}_l^{(2)}. \tag{79}$$

The total number of slave nodes on $\Gamma_{\mathrm{c},h}^{(1)}$ is $n^{(1)}$, and the total number of master nodes on $\Gamma_{\mathrm{c},h}^{(2)}$ is $n^{(2)}$. Discrete nodal positions and discrete nodal displacements are given by $\mathbf{x}_k^{(1)}$, $\mathbf{x}_l^{(2)}$, $\mathbf{d}_k^{(1)}$ and $\mathbf{d}_l^{(2)}$. The shape functions $N_k^{(1)}$ and $N_l^{(2)}$ are defined with respect to the usual finite element parameter space, commonly denoted as $\xi^{(i)}$ for two-dimensional problems (i.e. 1D mesh tying interfaces) and as $\boldsymbol{\xi}^{(i)} = (\xi^{(i)}, \eta^{(i)})$ for three-dimensional problems (i.e. 2D mesh tying interfaces). As mentioned above, the shape functions are derived from the underlying bulk discretization. Although not studied here, the proposed algorithms can in principle be transferred to higher-order interpolation and alternative shape functions, such as non-uniform rational B-splines (NURBS), see e.g. Cottrell et al. (2009), De Lorenzis et al. (2011) and Temizer et al. (2011, 2012).

In addition, an adequate discretization of the Lagrange multiplier vector $\boldsymbol{\lambda}$ is needed, too, and will be based on a discrete Lagrange multiplier space $\boldsymbol{\mathcal{M}}_h$ being an approximation of $\boldsymbol{\mathcal{M}}$. Some details concerning the choice of $\boldsymbol{\mathcal{M}}_h$, and especially concerning the two possible families of standard and dual Lagrange multipliers, will follow in Sect. 7.1. Thus, only a very general notation is given at this point:

$$\boldsymbol{\lambda}_h = \sum_{j=1}^{m^{(1)}} \Phi_j(\xi^{(1)}, \eta^{(1)})\, \boldsymbol{\lambda}_j, \tag{80}$$

with the (still to be defined) shape functions Φ_j and the discrete nodal Lagrange multipliers $\boldsymbol{\lambda}_j$. The total number of slave nodes carrying additional Lagrange multiplier degrees of freedom is $m^{(1)}$. Typically for mortar methods, every slave node also serves as coupling node, and thus in the majority of cases $m^{(1)} = n^{(1)}$ will hold. However, in the context of second-order finite elements, it will be favorable to chose $m^{(1)} < n^{(1)}$ in certain cases. Substituting (78) and (80) into the interface virtual work $\delta\mathcal{W}_{\mathsf{mt}}$ in (72) yields

$$\begin{aligned} -\delta\mathcal{W}_{\mathsf{mt},h} = &\sum_{j=1}^{m^{(1)}} \sum_{k=1}^{n^{(1)}} \boldsymbol{\lambda}_j^{\mathsf{T}} \left(\int_{\Gamma_{\mathsf{c},h}^{(1)}} \Phi_j \, N_k^{(1)} \, \mathrm{d}A_0 \right) \delta \mathbf{d}_k^{(1)} \\ &- \sum_{j=1}^{m^{(1)}} \sum_{l=1}^{n^{(2)}} \boldsymbol{\lambda}_j^{\mathsf{T}} \left(\int_{\Gamma_{\mathsf{c},h}^{(1)}} \Phi_j \, (N_l^{(2)} \circ \chi_h) \, \mathrm{d}A_0 \right) \delta \mathbf{d}_l^{(2)}, \end{aligned} \tag{81}$$

where $\chi_h : \Gamma_{\mathsf{c},h}^{(1)} \to \Gamma_{\mathsf{c},h}^{(2)}$ defines a suitable discrete mapping from the slave to the master side of the mesh tying interface. Such a mapping (or projection) becomes necessary due to the fact that the discretized coupling surfaces $\Gamma_{\mathsf{c},h}^{(1)}$ and $\Gamma_{\mathsf{c},h}^{(2)}$ are, in general, no longer geometrically coincident. This becomes very clear when thinking of a curved mesh tying interface with non-matching finite element meshes on the two different sides. As illustrated in Fig. 4, tiny gaps and overlaps may be generated in the discretized setting, although the surfaces had still been coincident in the continuum framework. Throughout this contribution, numerical integration of the mortar coupling terms is exclusively performed on the slave side $\Gamma_{\mathsf{c},h}^{(1)}$ of the interface. In (81), nodal blocks of the two mortar integral matrices commonly denoted as $\mathbf{D}$ and $\mathbf{M}$ can be identified. This leads to the following definitions:

$$\mathbf{D}[j,k] = D_{jk} \, \mathbf{I}_{\mathsf{ndim}} = \int_{\Gamma_{\mathsf{c},h}^{(1)}} \Phi_j N_k^{(1)} \mathrm{d}A_0 \, \mathbf{I}_{\mathsf{ndim}}, \tag{82}$$

$$\mathbf{M}[j,l] = M_{jl} \, \mathbf{I}_{\mathsf{ndim}} = \int_{\Gamma_{\mathsf{c},h}^{(1)}} \Phi_j (N_l^{(2)} \circ \chi_h) \, \mathrm{d}A_0 \, \mathbf{I}_{\mathsf{ndim}}, \tag{83}$$

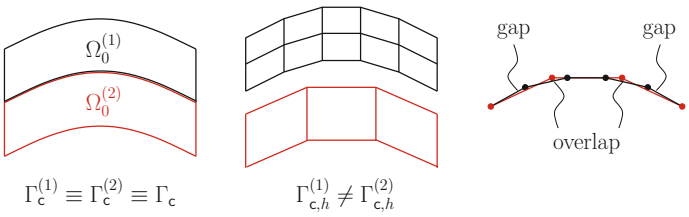

Fig. 4 Gaps and overlaps in a curved mesh tying interface with non-matching FE meshes

where $j = 1, \ldots, m^{(1)},\ k = 1, \ldots, n^{(1)},\ l = 1, \ldots, n^{(2)}$. Note that $\mathbf{I}_{\mathsf{ndim}} \in \mathbb{R}^{\mathsf{ndim} \times \mathsf{ndim}}$ is an identity matrix whose size is determined by the global problem dimension ndim, viz. either $\mathsf{ndim} = 2$ or $\mathsf{ndim} = 3$. In general, both mortar matrices $\mathbf{D}$ and $\mathbf{M}$ have a rectangular shape. However, $\mathbf{D}$ becomes a square matrix for the common choice $m^{(1)} = n^{(1)}$. More details concerning the actual numerical integration of the mass matrix type of entries in $\mathbf{D}$ and $\mathbf{M}$ as well as the implementation of the interface mapping χ_h for 3D will be given in Sects. 5.4 and 7.3.

For the ease of notation, all nodes of the two subdomains $\Omega_0^{(1)}$ and $\Omega_0^{(2)}$, and correspondingly all degrees of freedom (DOFs) in the global discrete displacement vector $\mathbf{d}$, are sorted into three groups: a group $\mathcal{S}$ containing all slave interface quantities, a group $\mathcal{M}$ of all master quantities and a group denoted as $\mathcal{N}$, which comprises all remaining nodes or DOFs. The global discrete displacement vector can be sorted accordingly, yielding $\mathbf{d} = (\mathbf{d}_{\mathcal{N}}, \mathbf{d}_{\mathcal{M}}, \mathbf{d}_{\mathcal{S}})$. Going back to (81), this allows for the following definition:

$$-\delta \mathcal{W}_{\mathsf{mt},h} = \delta \mathbf{d}_{\mathcal{S}}^{\mathsf{T}} \mathbf{D}^{\mathsf{T}} \boldsymbol{\lambda} - \delta \mathbf{d}_{\mathcal{M}}^{\mathsf{T}} \mathbf{M}^{\mathsf{T}} \boldsymbol{\lambda} = \delta \mathbf{d}^{\mathsf{T}} \underbrace{\begin{bmatrix} \mathbf{0} \\ -\mathbf{M}^{\mathsf{T}} \\ \mathbf{D}^{\mathsf{T}} \end{bmatrix}}_{\mathbf{B}_{\mathsf{mt}}^{\mathsf{T}}} \boldsymbol{\lambda} = \delta \mathbf{d}^{\mathsf{T}} \mathbf{f}_{\mathsf{mt}}(\boldsymbol{\lambda}). \tag{84}$$

Herein, the discrete mortar mesh tying operator $\mathbf{B}_{\mathsf{mt}}$ and the resulting discrete vector of mesh tying forces $\mathbf{f}_{\mathsf{mt}}(\boldsymbol{\lambda}) = \mathbf{B}_{\mathsf{mt}}^{\mathsf{T}} \boldsymbol{\lambda}$ acting on the slave and the master side of the interface are introduced. To finalize the discretization of the considered mesh tying problem, a closer look needs to be taken at the weak constraint contribution $\delta \mathcal{W}_{\lambda}$ in (73). Due to the saddle point characteristics and resulting symmetry of the mixed variational formulation in (72) and (73), all discrete components of $\delta \mathcal{W}_{\lambda}$ have already been introduced and the final formulation is given as

$$\delta \mathcal{W}_{\lambda,h} = \delta \boldsymbol{\lambda}^{\mathsf{T}} \mathbf{D} \mathbf{d}_{\mathcal{S}} - \delta \boldsymbol{\lambda}^{\mathsf{T}} \mathbf{M} \mathbf{d}_{\mathcal{M}} = \delta \boldsymbol{\lambda}^{\mathsf{T}} \mathbf{B}_{\mathsf{mt}} \mathbf{d} = \delta \boldsymbol{\lambda}^{\mathsf{T}} \mathbf{g}_{\mathsf{mt}}(\mathbf{d}), \tag{85}$$

with $\mathbf{g}_{\mathsf{mt}}(\mathbf{d}) = \mathbf{B}_{\mathsf{mt}} \mathbf{d}$ representing the discrete mesh tying constraint at the coupling interface. Taking into account the typical finite element discretization of all remaining contributions to the first part of the weak formulation (72), as previously outlined in Sect. 4.2, the semi-discrete equations of motion including tied contact forces and the constraint equations emerge as

$$\mathbf{M} \ddot{\mathbf{d}} + \mathbf{C} \dot{\mathbf{d}} + \mathbf{f}_{\mathsf{int}}(\mathbf{d}) + \mathbf{f}_{\mathsf{mt}}(\boldsymbol{\lambda}) - \mathbf{f}_{\mathsf{ext}} = \mathbf{0}, \tag{86}$$

$$\mathbf{g}_{\mathsf{mt}}(\mathbf{d}) = \mathbf{0}. \tag{87}$$

Mass matrix $\mathbf{M}$, damping matrix $\mathbf{C}$, internal forces $\mathbf{f}_{\mathsf{int}}(\mathbf{d})$ and external forces $\mathbf{f}_{\mathsf{ext}}$ result from standard FE discretization. It is important to point out that the actual mortar-based interface coupling described here is completely independent of the concrete choice of the underlying finite element formulation. The same also holds true for the question which particular material model is applied. As both topics,

i.e. nonlinear finite elements for continua and complex material models, are discussed at length in the literature, details will not be repeated here but the focus will remain solely on the mesh tying terms $\mathbf{f}_{\text{mt}}(\boldsymbol{\lambda})$ and $\mathbf{g}_{\text{mt}}(\mathbf{d})$.

Examining the semi-discrete problem statement in (86) and (87) in more detail, the well-known nonlinearity of the internal forces $\mathbf{f}_{\text{int}}(\mathbf{d})$ due to the consideration of finite deformation kinematics and nonlinear material behavior becomes apparent. However, neither the discrete interface forces $\mathbf{f}_{\text{mt}}(\boldsymbol{\lambda})$ nor the mesh tying constraints $\mathbf{g}_{\text{mt}}(\mathbf{d})$ introduce an additional nonlinearity into the system. This is due to the fact that no relative movement of the subdomains is permitted in mesh tying problems. Therefore, the mortar integral matrices $\mathbf{D}$ and $\mathbf{M}$ and hence also the discrete mesh tying operator $\mathbf{B}_{\text{mt}}$ only need to be evaluated *once* at problem initialization and thus do not depend on the actual displacements, even if finite deformations of the considered body are involved. With respect to numerical efficiency, this means that evaluating the mortar coupling terms for tied contact problems is a one-time cost, which can usually be neglected as compared with the remaining computational costs. Only for the unilateral contact case discussed in Sect. 6, this will no longer be the case. The question how to numerically evaluate the entries of $\mathbf{B}_{\text{mt}}$ in 3D problems is discussed in the following paragraph.

5.4 Evaluation of Mortar Integrals in 3D

All general concepts of the evaluation of mortar integrals in 3D can also be transferred back to the simple 2D case. The integral entries of both matrices $\mathbf{D}$ and $\mathbf{M}$ will be computed based on so-called mortar segments in order to achieve the maximum possible accuracy of Gauss quadrature and to guarantee linear momentum conservation in the semi-discrete setting. Projection operations between slave surface $\Gamma_{\text{c},h}^{(1)}$ and master surface $\Gamma_{\text{c},h}^{(2)}$, which consist of two-dimensional facets, are based on nodal averaging and a C^0-continuous field of normal vectors, cf. Fig. 17. For 3D situations, the averaged nodal normal vector $\mathbf{n}_k$ is given as

$$\mathbf{n}_k = \frac{\sum_{e=1}^{n_k^{\text{adj}}} \mathbf{n}_k^{(e)}}{\| \sum_{e=1}^{n_k^{\text{adj}}} \mathbf{n}_k^{(e)} \|}, \tag{88}$$

where the total number of slave facets n_k^{adj} adjacent to slave node k may vary within a much wider range than in 2D (for instance $n_k^{\text{adj}} = 4$ in Fig. 17). In anticipation of unilateral contact formulations, (88) also defines a tangential plane at slave node k, from which the two unit tangent vectors $\boldsymbol{\tau}_k^{\xi}$ and $\boldsymbol{\tau}_k^{\eta}$ can be chosen to form an orthonormal basis together with $\mathbf{n}_k$ as

$$\mathbf{n}_k \cdot \boldsymbol{\tau}_k^{\xi} = 0, \quad \boldsymbol{\tau}_k^{\eta} = \mathbf{n}_k \times \boldsymbol{\tau}_k^{\xi}. \tag{89}$$

Mortar segments must be defined such that the shape function integrands in (82) and (83) are C^1-continuous on these surface subsets. However, it is quite obvious that this task is much more complex in three dimensions than it would be in two dimensions, because mortar segments are arbitrarily shaped polygons as compared with line segments in the 2D case. Beyond that, the choice of an adequate mortar integration surface itself is quite difficult. In the 2D mortar mesh tying formulation that is not discussed here, integration is performed directly on the slave surface $\Gamma_{\mathsf{c},h}^{(1)}$. Unfortunately, it is not trivial to directly transfer this approach to three dimensions, because of the possible warping of surface facets.

The general topic of numerical integration, and an overview of the available (segment-based and element-based) integration schemes for this purpose is given in Sect. 7.3

5.5 Solution Methods

The attention is now turned back to the actual mortar finite element approach for tied contact derived in Sect. 5.3, and in particular to the final fully discretized version (i.e. after time discretization with the generalized-α method previously discussed in Sect. 4.3) of (86) and (87). All solution methods for this system of $\mathsf{ndof} + \mathsf{nco}$ nonlinear discrete algebraic equations, where the global number of constraints is given by $\mathsf{nco} = \mathsf{ndim} \cdot m^{(1)}$, are based on a standard Newton–Raphson iteration as introduced in Sect. 4.4. With only equality constraints being present, no active set strategies are needed for mesh tying systems, but the iterative solution techniques can be applied directly, thus yielding standard (or smooth) Newton methods. Primal-dual active set strategies and the associated notion of semi-smooth Newton methods only become important in the context of unilateral contact considered in Sect. 6.

As explained in Sect. 4.4, the Newton–Raphson method (or Newton's method) is based on a subsequent linearization of the residual, here defined by the discrete balance of linear momentum and the discrete mesh tying constraints in the time-discretized versions of (86) and (87). Each nonlinear solution step (iteration index i) then consists of solving the resulting linearized system of equations and an incremental update of the unknown displacements $\mathbf{d}_{n+1}$ and Lagrange multipliers $\boldsymbol{\lambda}_{n+1-\alpha_\mathsf{f}}$ until a user-defined convergence criterion is met. Taking into account that the discrete mesh tying operator $\mathbf{B}_{\mathsf{mt}}$ defined in (84) does not depend on the displacements, consistent linearization in iteration step i yields:

$$\mathbf{K}_{\mathsf{effdyn}}(\mathbf{d}_{n+1}^{i})\ \Delta\mathbf{d}_{n+1}^{i+1} + \mathbf{B}_{\mathsf{mt}}\boldsymbol{\lambda}_{n+1-\alpha_\mathsf{f}}^{i} = -\mathbf{r}_{\mathsf{effdyn}}(\mathbf{d}_{n+1}^{i}), \tag{90}$$

$$\left.\frac{\partial \mathbf{g}_{\mathsf{mt}}(\mathbf{d}_{n+1})}{\partial \mathbf{d}_{n+1}}\right|^{i} \Delta\mathbf{d}_{n+1}^{i+1} = -\mathbf{g}_{\mathsf{mt}}(\mathbf{d}_{n+1}^{i}). \tag{91}$$

Herein, the fact that the Lagrange multipliers only enter the discrete mesh tying in a linear fashion has been made use of. Due to this linearity, it is possible to solve

directly for the unknown Lagrange multipliers $\boldsymbol{\lambda}^{i}_{n+1-\alpha_\mathrm{f}}$ in each iteration step instead of an incremental formulation. Moreover, as mentioned in Sect. 4.4, all discrete force terms (inertia, damping, internal and external forces) except for the additional mesh tying forces $\mathbf{f}_{\mathrm{mt}}(\boldsymbol{\lambda}^{i}_{n+1-\alpha_\mathrm{f}})$ are summarized in the residual $\mathbf{r}_{\mathrm{effdyn}}(\mathbf{d}^{i}_{n+1})$ and the partial derivative of $\mathbf{r}_{\mathrm{effdyn}}(\mathbf{d}^{i}_{n+1})$ with respect to the displacements $\mathbf{d}$ is commonly referred to as dynamic effective tangential stiffness matrix $\mathbf{K}_{\mathrm{effdyn}}(\mathbf{d}^{i}_{n+1})$, as introduced in (58). Finally, it is pointed out that the constraints $\mathbf{g}_{\mathrm{mt}}(\mathbf{d}_{n+1}) = \mathbf{0}$ are already enforced at time $t = 0$ to assure angular momentum conservation. Thus, the right-hand side of the linearized constraint equation in (91) simply reduces to zero.

The linearized statement in (90) and (91) already gives a hint as to the typical saddle point structure of the resulting Lagrange multiplier system. Analyzing the linearized mesh tying system (90) in more detail and splitting the global displacement vector $\mathbf{d} = (\mathbf{d}_{\mathcal{N}}, \mathbf{d}_{\mathcal{M}}, \mathbf{d}_{\mathcal{S}})$ as well as all other involved quantities into three subsets as defined in Sect. 5.3 leads to the following notation in matrix-vector notation:

$$\begin{bmatrix} \mathbf{K}_{\mathcal{N}\mathcal{N}} & \mathbf{K}_{\mathcal{N}\mathcal{M}} & \mathbf{K}_{\mathcal{N}\mathcal{S}} & \mathbf{0} \\ \mathbf{K}_{\mathcal{M}\mathcal{N}} & \mathbf{K}_{\mathcal{M}\mathcal{M}} & \mathbf{0} & -\mathbf{M}^{\mathsf{T}} \\ \mathbf{K}_{\mathcal{S}\mathcal{N}} & \mathbf{0} & \mathbf{K}_{\mathcal{S}\mathcal{S}} & \mathbf{D}^{\mathsf{T}} \\ \mathbf{0} & -\mathbf{M} & \mathbf{D} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{d}_{n+1,\,\mathcal{N}} \\ \Delta\mathbf{d}_{n+1,\,\mathcal{M}} \\ \Delta\mathbf{d}_{n+1,\,\mathcal{S}} \\ \boldsymbol{\lambda}_{n+1-\alpha_\mathrm{f}} \end{bmatrix} = - \begin{bmatrix} \mathbf{r}_{\mathcal{N}} \\ \mathbf{r}_{\mathcal{M}} \\ \mathbf{r}_{\mathcal{S}} \\ \mathbf{0} \end{bmatrix}. \tag{92}$$

Herein, the nonlinear iteration index i and the subscript $\cdot_{\mathrm{effdyn}}$ of the residual vector $\mathbf{r}_{\mathrm{effdyn}}$ and the tangential stiffness matrix $\mathbf{K}_{\mathrm{effdyn}}$ have been omitted for the ease of notation. Note that no matrix blocks $\mathbf{K}_{\mathcal{M}\mathcal{S}}$ and $\mathbf{K}_{\mathcal{S}\mathcal{M}}$ exist, because slave and master side degrees of freedom are only coupled via the mortar approach. Due to the inherent symmetry of $\mathbf{K}_{\mathrm{effdyn}}$, the global linearized mesh tying system (92) is also *symmetric* and has the typical saddle point structure with a zero matrix block associated with the Lagrange multipliers $\boldsymbol{\lambda}_{n+1-\alpha_\mathrm{f}}$ on the main diagonal. Thus, while a conforming discretization would yield a positive definite system, the coupled mesh tying system considered here becomes indefinite with both positive and negative eigenvalues due to the saddle point characteristics of the Lagrange multiplier method.

The linear system (92) needs to be solved within each nonlinear iteration step. Unfortunately, efficient iterative solution techniques and especially the associated preconditioners usually perform very poorly for such indefinite systems or are not applicable at all. The main reason for this lies in the fact that common preconditioning techniques, e.g. the Jacobi and Gauss–Seidel methods, fail for zero diagonal matrix entries as occurring in (92). Nevertheless, there exist some specific solution methods for this type of saddle point matrix block system, which are both well-established and quite efficient. One popular representative, also employed as preconditioner in this contribution whenever large mesh tying and contact systems are considered with a standard Lagrange multiplier approach, is given by the so-called semi-implicit method for pressure-linked equations (SIMPLE) and its many descendants, see e.g. Elman et al. (2008) for a very comprehensive overview in the context of the incompressible Navier–Stokes equations for fluid dynamics.

As will be explained in Sect. 7.1, the dual Lagrange multiplier approach is characterized by its localization of the coupling constraints at the mesh tying interface, and thus algebraically by mortar matrix $\mathbf{D}$ reducing to a diagonal matrix. This makes $\mathbf{D}$ trivial to invert and allows for efficient condensation operations of the slave side degrees of freedom, i.e. both Lagrange multipliers and the discrete slave side displacements. The basis for this condensation is given by the saddle point system in (92), which is of course equally valid for dual Lagrange multiplier interpolation. In preparation of a first condensation step, the third row of (92) is used to express the unknown Lagrange multipliers $\boldsymbol{\lambda}_{n+1-\alpha_\mathrm{f}}$ as

$$\boldsymbol{\lambda}_{n+1-\alpha_\mathrm{f}} = \mathbf{D}^{-\mathsf{T}} \left(-\mathbf{r}_{\mathcal{S}} - \mathbf{K}_{\mathcal{S}\mathcal{N}} \Delta \mathbf{d}_{n+1,\mathcal{N}} - \mathbf{K}_{\mathcal{S}\mathcal{S}} \Delta \mathbf{d}_{n+1,\mathcal{S}} \right). \tag{93}$$

Insertion into the second row of (92) yields the following intermediate system:

$$\begin{bmatrix} \mathbf{K}_{\mathcal{N}\mathcal{N}} & \mathbf{K}_{\mathcal{N}\mathcal{M}} & \mathbf{K}_{\mathcal{N}\mathcal{S}} \\ \mathbf{K}_{\mathcal{M}\mathcal{N}} + \mathbf{P}^{\mathsf{T}}\mathbf{K}_{\mathcal{S}\mathcal{N}} & \mathbf{K}_{\mathcal{M}\mathcal{M}} & \mathbf{P}^{\mathsf{T}}\mathbf{K}_{\mathcal{S}\mathcal{S}} \\ \mathbf{0} & -\mathbf{M} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{d}_{n+1,\mathcal{N}} \\ \Delta \mathbf{d}_{n+1,\mathcal{M}} \\ \Delta \mathbf{d}_{n+1,\mathcal{S}} \end{bmatrix} = - \begin{bmatrix} \mathbf{r}_{\mathcal{N}} \\ \mathbf{r}_{\mathcal{M}} + \mathbf{P}^{\mathsf{T}}\mathbf{r}_{\mathcal{S}} \\ \mathbf{0} \end{bmatrix}, \tag{94}$$

where the mortar projection operator $\mathbf{P} = \mathbf{D}^{-1}\mathbf{M}$ that will formally be introduced in (143) is used to abbreviate the notation. As a second step, the constraint equation in the last row of (94) can be expressed as

$$\Delta \mathbf{d}_{n+1,\mathcal{S}} = \mathbf{D}^{-1}\mathbf{M}\Delta \mathbf{d}_{n+1,\mathcal{M}} = \mathbf{P}\Delta \mathbf{d}_{n+1,\mathcal{M}}. \tag{95}$$

The final condensed system for the dual Lagrange multiplier approach is then obtained by reinserting this result into the first row and second row of the intermediate system, viz.

$$\begin{bmatrix} \mathbf{K}_{\mathcal{N}\mathcal{N}} & \mathbf{K}_{\mathcal{N}\mathcal{M}} + \mathbf{K}_{\mathcal{N}\mathcal{S}}\mathbf{P} \\ \mathbf{K}_{\mathcal{M}\mathcal{N}} + \mathbf{P}^{\mathsf{T}}\mathbf{K}_{\mathcal{S}\mathcal{N}} & \mathbf{K}_{\mathcal{M}\mathcal{M}} + \mathbf{P}^{\mathsf{T}}\mathbf{K}_{\mathcal{S}\mathcal{S}}\mathbf{P} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{d}_{n+1,\mathcal{N}} \\ \Delta \mathbf{d}_{n+1,\mathcal{M}} \end{bmatrix} = - \begin{bmatrix} \mathbf{r}_{\mathcal{N}} \\ \mathbf{r}_{\mathcal{M}} + \mathbf{P}^{\mathsf{T}}\mathbf{r}_{\mathcal{S}} \end{bmatrix}. \tag{96}$$

This final linearized system unifies several beneficial properties as compared with the equivalent saddle point formulation given in (92). Firstly, the discrete Lagrange multiplier degrees of freedom $\boldsymbol{\lambda}_{n+1-\alpha_\mathrm{f}}$ have been removed from the global system and thus the commonly cited disadvantage of an increased system size for Lagrange multiplier methods is resolved. In fact, owing to the second condensation step, which removes the slave side displacement degrees of freedom $\Delta \mathbf{d}_{n+1,\mathcal{S}}$, the final system size is even reduced as compared with a conforming discretization. Secondly, and more importantly, the typical saddle point structure with a zero diagonal matrix block has been completely removed on the way towards the final system (96), which is instead symmetric and positive definite again.

With regard to linear solvers, the dual Lagrange multiplier approach virtually allows for an "out-of-the-box" application of state-of-the-art iterative solution and preconditioning techniques, such as the CG or GMRES approach in combination

with algebraic multigrid (AMG) methods. Simply speaking, all solvers that were optimized for conforming discretizations in nonlinear solid mechanics are equally applicable to the non-conforming mortar formulation with dual Lagrange multipliers in (96) due to similar system properties. The additional computational effort associated with the condensation operations can be considered very low. In a first, naive implementation, setting up the condensed system would simply require some additional matrix-matrix products of interface-sized matrix blocks such as the discrete projection operator $\mathbf{P}$. However, a more elaborate implementation could even do without explicit matrix-matrix products, but would rather introduce modified local assembly procedures for the individual finite element contributions to the tangential stiffness matrix blocks $\mathbf{K}_{\mathcal{NS}}$, $\mathbf{K}_{\mathcal{SN}}$ and $\mathbf{K}_{\mathcal{SS}}$, taking into account the associated local entries of the mortar projection operator $\mathbf{P}$. In any case, the improved properties and the more efficient solvability of (96) as compared with (92) by far outweigh additional computational costs for the condensation, which makes the dual Lagrange multiplier approach the preferred choice throughout this chapter.

For the sake of completeness, two details should be pointed out. Firstly, the described condensation operations are of course also applicable for standard Lagrange multiplier interpolation with a non-diagonal mortar matrix $\mathbf{D}$, at least theoretically. In practice, however, the inverse matrix $\mathbf{D}^{-1}$ would be densely populated in such a case, which forbids the actual computation and storage of $\mathbf{D}^{-1}$ or likewise $\mathbf{P}$ for moderate or even large system sizes. For dual Lagrange multiplier interpolation, on the contrary, inversion of $\mathbf{D}$ and storage of the sparsely populated matrix $\mathbf{P}$ remain easily manageable even for large-scale mortar mesh tying simulations. Secondly, node-matching interface meshes are contained as a special case in the given mortar formulation. This situation basically leads to $\mathbf{P}$ becoming an identity operator, establishing a one-to-one mapping between slave side and master side displacements. Expression (96) then reduces to exactly the same linearized system that is obtained for a conforming mesh.

5.6 Numerical Example

Patch tests are arguably one of the most common validation tools in finite element analysis, typically used as a first important step towards an assessment of the consistency of new element formulations, see e.g. Irons (1966) and Taylor et al. (1986). In the present context of mesh tying and contact mechanics, patch tests are investigated in order to analyze the ability of mortar methods to exactly represent the simplest possible (i.e. constant) stress states across arbitrary non-conforming interfaces. However, it is well-known that collocation-based methods such as the classical node-to-segment (NTS) approach for mesh tying and unilateral contact typically fail the patch test. Mortar finite element methods, with their variationally consistent interpolation of the interface traction via discrete Lagrange multipliers $\boldsymbol{\lambda}$, guarantee the exact satisfaction of typical patch tests by design.

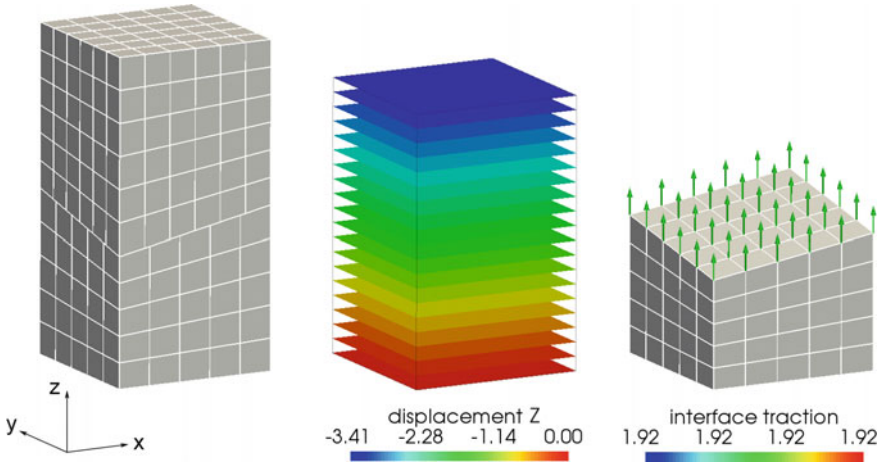

Fig. 5 3D patch test with inclined interface – finite element mesh (left), displacement u_z (middle) and interface tractions represented by the discrete Lagrange multipliers $\boldsymbol{\lambda}$ (right)

As a first test setup, two stacked cubes with an inclined but flat mesh tying interface, as illustrated in Fig. 5, are investigated. This geometric model is obtained by first considering two identical cubes of side length 10 and then moving two opposite corners of the interface by a distance of ± 2 in z-direction. The compressible Neo–Hookean material law introduced in Sect. 3.2 is employed with Young's modulus $E = 10$ and Poisson's ratio $\nu = 0.4$. A constant pressure load $p = -0.2$ is applied to the top surface of the upper block, and the bottom surface of the lower block is supported such that any rigid body movement is precluded, but the bodies are free to expand laterally. The lower block is defined as slave side for mortar coupling and the chosen mesh size ratio of $h^{(1)}/h^{(2)} = 5/6$ generates a non-matching situation at the interface. Figure 5 exemplarily illustrates the displacement solution as well as the Lagrange multiplier (i.e. interface traction) solution in z-direction for a *hex8* discretization. As expected, a linear displacement field and constant interface tractions are obtained. The fact that the patch test is actually passed to machine precision for any first-order or second-order finite element type is emphasized in Fig. 6, where the normal stress component in z-direction of the Cauchy stress tensor $\boldsymbol{\sigma}$ is visualized. While all presented results have been obtained with dual Lagrange multiplier interpolation according to Sect. 7.1, standard Lagrange multipliers would yield identical results.

The second patch test investigated is a 2D rectangular strip (length $l = 8$, width $w = 3$) with five subdomains, each discretized with different first-order and second-order finite elements (i.e. *tri3*, *quad4*, *tri6*, *quad8* and *quad9* elements), see Fig. 7. While this admittedly constitutes a rather academic example, it strikingly demonstrates the mesh generation flexibility offered by mortar methods, and

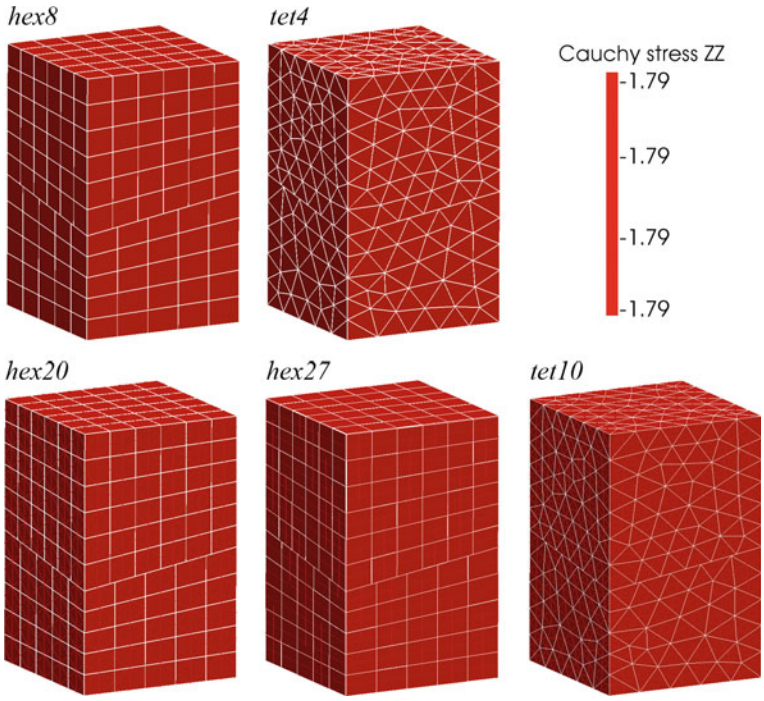

Fig. 6 3D patch test with inclined interface – Cauchy stress σ_{zz} for several different types of first-order and second-order mortar finite element interpolation

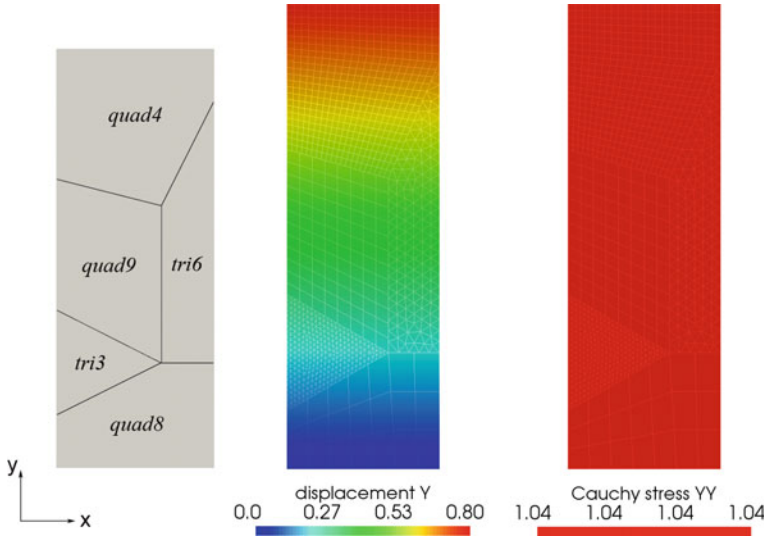

Fig. 7 2D patch test with crosspoints – types of finite element interpolation in the individual subdomains (left), displacement u_y (middle) and Cauchy stress σ_{yy} (right)

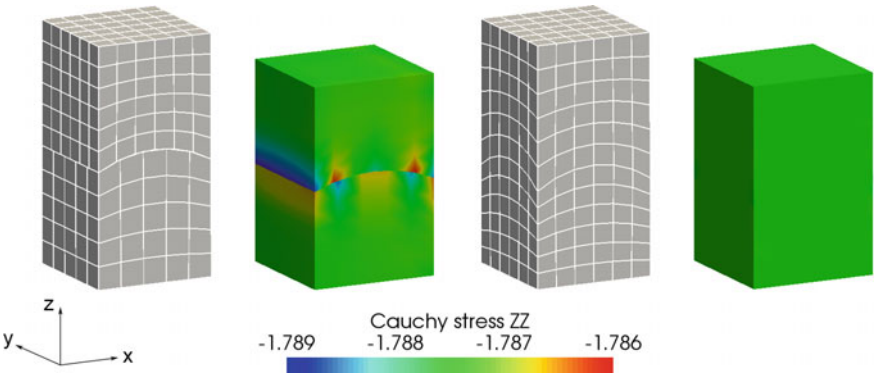

Fig. 8 3D patch test with curved interface – finite element mesh and Cauchy stress σ_{zz} for non-conforming interfaces (left) and for node-matching interfaces (right)

especially also the possibility of a consistent treatment of so-called crosspoints as discussed in Wohlmuth (2001). Again, a compressible Neo–Hookean constitutive model is employed ($E = 10$, $\nu = 0.3$) and the strip is subject to unilateral loading in y-direction. Both displacement and stress solution confirm that this 2D patch test is passed to machine precision. The treatment of crosspoints is readily extended to three dimensions, see e.g. Wohlmuth (2001).

Finally, the first patch test model is reconsidered, but now with a curved mesh tying interface. The exemplary results for a *hex8* mesh in the left part of Fig. 8 illustrate the limits of mortar finite element methods with regard to exact patch test satisfaction. It can be seen quite clearly that the patch test is not satisfied to machine precision in that case, but instead a small error is introduced in the vicinity of the interface. The reason for this result has already been explained in Sect. 5.3 and lies in the fact that the discrete surfaces $\Gamma_{c,h}^{(1)}$ and $\Gamma_{c,h}^{(2)}$ are no longer geometrically coincident for non-matching meshes on curved interfaces, but tiny gaps and overlapping regions appear. Thus, a discrete projection step is needed, which inevitably precludes the constant stress solution to be recovered exactly. This becomes even clearer when analyzing a curved mesh tying interface with node-matching meshes, as visualized in the right part of Fig. 8. In that case, the discrete mesh tying surfaces $\Gamma_{c,h}^{(1)}$ and $\Gamma_{c,h}^{(2)}$ are again coincident, the mortar projection operator $\mathbf{P}$ reduces to an identity mapping and the patch test is satisfied exactly. Nevertheless, it should be pointed out that the error of mortar methods in curved patch tests is only marginal and can factually be neglected from an engineering point of view. Besides, the curved patch test behavior of mortar methods is still significantly better than that of classical NTS schemes, see also Hesch and Betsch (2010).

6 Mortar Methods for Unilateral Contact

Contact interaction in nonlinear solid mechanics and the use of mortar finite element methods in this context are the main focus of interest of this chapter. The goal of all developments presented is to be able to analyze and accurately predict the mechanical response in highly nonlinear unilateral contact scenarios, i.e. including very large deformations and sliding, continuous changes of the active contact area and possibly nonlinear material behavior. From a method development point of view, many aspects of mortar methods already introduced for mesh tying in Sect. 5 can either be re-used directly or in a slightly modified way in order to meet contact-specific demands. For further theoretical considerations and an in-depth analysis of the mathematical foundations of contact mechanics, the comprehensive textbook by Kikuchi and Oden (1988) and the recent review article by Wohlmuth (2011) should be consulted. A full derivation of all formulations reviewed here can be found in the author's original work (Popp 2012).

6.1 Strong Formulation

For the sake of simplicity, only the case of two contacting bodies with one sole contact interface is considered here. However, a generalization to multiple bodies and self contact is rather straightforward and mostly a matter of efficient search algorithms. All necessary notations for the finite deformation unilateral contact problem have already been introduced in Fig. 2, to which the reader is once again referred at this point. The domains $\Omega_0(i) \subset \mathbb{R}^3$ and $\Omega_t^{(i)} \subset \mathbb{R}^3$, $i = 1, 2$, represent two separate bodies in the reference and current configuration, respectively. To allow for the usual Dirichlet and Neumann boundary conditions as well as contact interaction, the surfaces $\partial\Omega_0^{(i)}$ are divided into three disjoint subsets $\Gamma_{\mathsf{u}}^{(i)}$, $\Gamma_{\sigma}^{(i)}$ and $\Gamma_{\mathsf{c}}^{(i)}$, where $\Gamma_{\mathsf{c}}^{(i)}$ represents the *potential* contact surface. Similarly, the spatial surface descriptions $\partial\Omega_t^{(i)}$ are split into $\gamma_{\mathsf{u}}^{(i)}$, $\gamma_{\sigma}^{(i)}$ and $\gamma_{\mathsf{c}}^{(i)}$. Retaining a customary nomenclature in contact mechanics, $\Gamma_{\mathsf{c}}^{(1)}$ is again referred to as slave surface and $\Gamma_{\mathsf{c}}^{(2)}$ as master surface.

On each subdomain $\Omega_0^{(i)}$ the initial boundary value problem of finite deformation elastodynamics needs to be satisfied, viz.

$$\mathrm{Div}\boldsymbol{P}^{(i)} + \hat{\boldsymbol{b}}_0^{(i)} = \rho_0^{(i)}\ddot{\boldsymbol{u}}^{(i)} \qquad \text{in } \Omega_0^{(i)} \times [0,T], \tag{97}$$

$$\boldsymbol{u}^{(i)} = \hat{\boldsymbol{u}}^{(i)} \qquad \text{on } \Gamma_{\mathsf{u}}^{(i)} \times [0,T], \tag{98}$$

$$\boldsymbol{P}^{(i)}\boldsymbol{N}^{(i)} = \hat{\boldsymbol{t}}_0^{(i)} \qquad \text{on } \Gamma_{\sigma}^{(i)} \times [0,T], \tag{99}$$

$$\boldsymbol{u}^{(i)}(\boldsymbol{X}^{(i)},0) = \hat{\boldsymbol{u}}_0^{(i)}(\boldsymbol{X}^{(i)}) \qquad \text{in } \Omega_0^{(i)}, \tag{100}$$

$$\dot{\boldsymbol{u}}^{(i)}(\boldsymbol{X}^{(i)},0) = \hat{\dot{\boldsymbol{u}}}_0^{(i)}(\boldsymbol{X}^{(i)}) \qquad \text{in } \Omega_0^{(i)}. \tag{101}$$

The contact constraints in normal direction are typically given in form of KKT conditions as defined in (33), while frictional sliding according to Coulomb's law has been introduced in (35). For the sake of completeness of the strong formulation, both sets of conditions are repeated:

$$g_{\mathsf{n}} \geq 0, \quad p_{\mathsf{n}} \leq 0, \quad p_{\mathsf{n}}\, g_{\mathsf{n}} = 0 \qquad \text{on } \gamma_{\mathsf{c}}^{(1)} \times [0,T], \tag{102}$$

$$\Phi := \|\boldsymbol{t}_{\tau}\| - \mathfrak{F}|p_{\mathsf{n}}| \leq 0,$$

$$\boldsymbol{v}_{\tau,\mathsf{rel}} + \beta \boldsymbol{t}_{\tau} = \boldsymbol{0}, \quad \beta \geq 0, \quad \Phi\beta = 0 \qquad \text{on } \gamma_{\mathsf{c}}^{(1)} \times [0,T]. \tag{103}$$

Equations (97)–(103) represent the final strong form of a unilateral contact problem in nonlinear solid mechanics. In the course of deriving a weak formulation (see next paragraph), the balance of linear momentum at the contact interface is typically exploited and a Lagrange multiplier vector $\boldsymbol{\lambda}$ is introduced, thus setting the basis for a mixed variational approach. In contrast to the mesh tying case in Sect. 5, it is striking that the unilateral contact constraints are typically formulated (and later also numerically evaluated) in the current configuration.

6.2 Weak Formulation

In the first instance, the most general weak formulation including also Coulomb friction is considered. Similar to the pure solid mechanics case in Sect. 4.1 and the mesh tying case in Sect. 5.2, the well-known solution spaces $\boldsymbol{\mathcal{U}}^{(i)}$ and weighting spaces $\boldsymbol{\mathcal{V}}^{(i)}$ are defined as

$$\boldsymbol{\mathcal{U}}^{(i)} = \left\{ \boldsymbol{u}^{(i)} \in H^1(\Omega) \mid \boldsymbol{u}^{(i)} = \hat{\boldsymbol{u}}^{(i)} \text{ on } \Gamma_{\mathsf{u}} \right\}, \tag{104}$$

$$\boldsymbol{\mathcal{V}}^{(i)} = \left\{ \delta\boldsymbol{u}^{(i)} \in H^1(\Omega) \mid \delta\boldsymbol{u}^{(i)} = \boldsymbol{0} \text{ on } \Gamma_{\mathsf{u}} \right\}. \tag{105}$$

Moreover, the Lagrange multiplier vector $\boldsymbol{\lambda} = -\boldsymbol{t}_{\mathsf{c}}^{(1)}$, which represents the *negative* slave side contact traction $\boldsymbol{t}_{\mathsf{c}}^{(1)}$ and is used to enforce the contact constraints (102) and (103), is chosen from the convex set $\boldsymbol{\mathcal{M}}(\boldsymbol{\lambda}) \subset \boldsymbol{\mathcal{M}}$ given by

$$\boldsymbol{\mathcal{M}}(\boldsymbol{\lambda}) = \left\{ \boldsymbol{\mu} \in \boldsymbol{\mathcal{M}} \mid \langle \boldsymbol{\mu}, \boldsymbol{v} \rangle_{\gamma_{\mathsf{c}}^{(1)}} \leq \langle \mathfrak{F}\lambda_{\mathsf{n}}, \|\boldsymbol{v}_{\tau}\| \rangle_{\gamma_{\mathsf{c}}^{(1)}},\ \boldsymbol{v} \in \boldsymbol{\mathcal{W}},\ v_{\mathsf{n}} \leq 0 \right\}. \tag{106}$$

Herein, $\langle \cdot, \cdot \rangle_{\gamma_{\mathsf{c}}^{(1)}}$ again stands for the scalar or vector-valued duality pairing between $H^{-1/2}$ and $H^{1/2}$ on $\gamma_{\mathsf{c}}^{(1)}$, see also Sect. 5.2. Moreover, $\boldsymbol{\mathcal{M}}$ is the dual space of the trace space $\boldsymbol{\mathcal{W}}^{(1)}$ of $\boldsymbol{\mathcal{V}}^{(1)}$ restricted to $\gamma_{\mathsf{c}}^{(1)}$, i.e. $\mathcal{M} = H^{-1/2}(\gamma_{\mathsf{c}}^{(1)})$ and $\mathcal{W}^{(1)} = H^{1/2}(\gamma_{\mathsf{c}}^{(1)})$, where $\mathcal{M}$ and $\mathcal{W}^{(1)}$ denote single scalar components of the corresponding vector-valued spaces $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\mathcal{W}}$. Thus, the definition of the solution cone for the Lagrange multipliers in (106) satisfies the conditions on $\boldsymbol{\lambda}$ of the Coulomb friction law in a weak sense.

Based on these considerations, the weak saddle point formulation is derived next. Basically, this can be done by extending the standard weak formulation of nonlinear solid mechanics as defined in (38) to two bodies and combining it with contact-specific Lagrange multiplier contributions. Find $\boldsymbol{u}^{(i)} \in \mathcal{U}^{(i)}$ and $\boldsymbol{\lambda} \in \boldsymbol{\mathcal{M}}(\lambda)$ such that

$$-\delta\mathcal{W}_{\mathsf{kin,int,ext}}(\boldsymbol{u}^{(i)}, \delta\boldsymbol{u}^{(i)}) - \delta\mathcal{W}_{\mathsf{co}}(\boldsymbol{\lambda}, \delta\boldsymbol{u}^{(i)}) = 0 \qquad \forall\, \delta\boldsymbol{u}^{(i)} \in \boldsymbol{\mathcal{V}}^{(i)}, \tag{107}$$

$$\delta\mathcal{W}_{\lambda}(\boldsymbol{u}^{(i)}, \delta\boldsymbol{\lambda}) \geq 0 \qquad \forall\, \delta\boldsymbol{\lambda} \in \boldsymbol{\mathcal{M}}(\lambda). \tag{108}$$

Herein, the kinetic contribution $\delta\mathcal{W}_{\mathsf{kin}}$ as well as the internal and external contributions $\delta\mathcal{W}_{\mathsf{int,ext}}$ to the overall virtual work of the two bodies do not change as compared with the mesh tying case in (74) and (75). However, the contact contribution $\delta\mathcal{W}_{\mathsf{co}}$ and the weak constraints $\delta\mathcal{W}_{\lambda}$, including non-penetration and frictional sliding conditions, are given in full length as

$$-\delta\mathcal{W}_{\mathsf{co}} = \int_{\gamma_{\mathsf{c}}^{(1)}} \boldsymbol{\lambda}(\delta\boldsymbol{u}^{(1)} - \delta\boldsymbol{u}^{(2)} \circ \chi)\, \mathrm{d}A, \tag{109}$$

$$\delta\mathcal{W}_{\lambda} = \int_{\gamma_{\mathsf{c}}^{(1)}} (\delta\lambda_{\mathsf{n}} - \lambda_{\mathsf{n}})\, \boldsymbol{g}_{\mathsf{n}}\, \mathrm{d}A - \int_{\gamma_{\mathsf{c}}^{(1)}} (\delta\boldsymbol{\lambda}_{\tau} - \boldsymbol{\lambda}_{\tau})\, \boldsymbol{v}_{\tau,\mathsf{rel}}\, \mathrm{d}A, \tag{110}$$

where $\chi : \gamma_{\mathsf{c}}^{(1)} \to \gamma_{\mathsf{c}}^{(2)}$ defines a suitable mapping from slave to master side of the contact surface, see also Sect. 3.4. In contrast to the mesh tying case, where this mapping only came into play in the discrete setting, $\gamma_{\mathsf{c}}^{(1)}$ and $\gamma_{\mathsf{c}}^{(2)}$ cannot even be guaranteed to be identical in the continuum framework for unilateral contact, because they not only comprise the actual contact surfaces but the potential contact surfaces. As explained in detail in Sect. 5.2, the integral expressions in the coupling bilinear forms $\delta\mathcal{W}_{\mathsf{co}}$ and $\delta\mathcal{W}_{\lambda}$ would need to be replaced by duality pairings $\langle\cdot,\cdot\rangle_{\gamma_{\mathsf{c}}^{(1)}}$ in order to be mathematically concise. However, the integral diction in (74) and (75) is preferred here due to readability. The coupling terms on $\gamma_{\mathsf{c}}^{(1)}$ also allow for a direct interpretation in terms of variational formulations and the principle of virtual work. Whereas the contribution in (109) represents the virtual work of the unknown contact tractions $\boldsymbol{\lambda} = -\boldsymbol{t}_{\mathsf{c}}^{(1)}$, the contribution in (110) ensures a weak, variationally consistent enforcement of the unilateral contact constraints in normal direction as well as the Coulomb friction law. The equivalence of the strong pointwise conditions given in (102) and (103) and the corresponding variational inequalities in (110) can readily be proven, see e.g. Wohlmuth (2011).

The main focus of this chapter is on mortar finite element methods for contact mechanics in general, and on discrete dual Lagrange multiplier spaces in particular, rather than on the physical foundations of frictional sliding or other interface effects. Many scientific questions investigated and answered in the following are completely independent of the precise tangential contact model. Thus, for the sake of simplicity, the weak formulation is restricted to the *frictionless* case from now on, as well as the upcoming derivations concerning finite element discretization. Nevertheless, Coulomb friction is included in the actual implementation originating from

this work, and special remarks on frictional sliding will be given where important, e.g. when considering semi-smooth Newton type active set strategies in Sect. 6.4. Without claiming that this list is exhaustive, details on the mortar finite element discretization of *frictional* contact can be found in Gitterle et al. (2010), Gitterle (2012), Hüeber et al. (2008), Tur et al. (2009), Wohlmuth (2011), Puso and Laursen (2004b) and Yang et al. (2005).

For frictionless sliding, the tangential part $\boldsymbol{t}_\tau$ of the slave side contact traction $\boldsymbol{t}_{\mathsf{c}}^{(1)}$ is supposed to vanish, and thus the set of frictional sliding conditions in (103) is simply replaced by

$$\boldsymbol{t}_\tau = \boldsymbol{0}. \tag{111}$$

Considering appropriate solution spaces, it becomes obvious that frictionless contact allows for a significant simplification of the convex cone of Lagrange multipliers, which is now given as

$$\boldsymbol{\mathcal{M}}^+ = \left\{ \boldsymbol{\mu} \in \boldsymbol{\mathcal{M}} \mid \boldsymbol{\mu}_\tau = \boldsymbol{0},\ \langle \mu_{\mathsf{n}}, w \rangle_{\gamma_{\mathsf{c}}^{(1)}} \geq 0,\ w \in \mathcal{W}^+ \right\}. \tag{112}$$

Herein, $\mathcal{W}^+$ is a closed non-empty convex cone being defined by $\mathcal{W}^+ = \{w \in \mathcal{W},\ w \geq 0\}$. The weak solution of the frictionless contact problem is then obtained from the following saddle point formulation: Find $\boldsymbol{u}^{(i)} \in \boldsymbol{\mathcal{U}}^{(i)}$ and $\boldsymbol{\lambda} \in \boldsymbol{\mathcal{M}}^+$ such that

$$-\delta \mathcal{W}_{\mathsf{kin,int,ext}}(\boldsymbol{u}^{(i)}, \delta \boldsymbol{u}^{(i)}) - \delta \mathcal{W}_{\mathsf{co}}(\boldsymbol{\lambda}, \delta \boldsymbol{u}^{(i)}) = 0 \qquad \forall\, \delta \boldsymbol{u}^{(i)} \in \boldsymbol{\mathcal{V}}^{(i)}, \tag{113}$$

$$\delta \mathcal{W}_\lambda(\boldsymbol{u}^{(i)}, \delta \boldsymbol{\lambda}) \geq 0 \qquad \forall\, \delta \boldsymbol{\lambda} \in \boldsymbol{\mathcal{M}}^+. \tag{114}$$

The contributions $\delta \mathcal{W}_{\mathsf{kin}}$, $\delta \mathcal{W}_{\mathsf{int,ext}}$ and $\delta \mathcal{W}_{\mathsf{co}}$ remain unchanged as previously defined in (74), (75) and (109). However, the weak contact constraints $\delta \mathcal{W}_\lambda$ now reduce to

$$\delta \mathcal{W}_\lambda = \int_{\gamma_{\mathsf{c}}^{(1)}} (\delta \lambda_{\mathsf{n}} - \lambda_{\mathsf{n}})\, g_{\mathsf{n}}\, \mathrm{d}A. \tag{115}$$

Strictly speaking, a scalar Lagrange multiplier λ_{n} would be completely sufficient to enforce the non-penetration condition here. Yet, in view of the more general case of frictional contact, a vector-valued Lagrange multiplier will also be employed for the frictionless case in this contribution, which allows for the nice interpretation of frictionless sliding as a special case of Coulomb's law with $\mathfrak{F} = 0$ and the convex cone of Lagrange multipliers $\boldsymbol{\mathcal{M}}(\boldsymbol{\lambda})$ reducing to $\boldsymbol{\mathcal{M}}^+$. As compared with the mesh tying case in Sect. 5.2, it is noticeable that the weak formulation contains *inequality* conditions for unilateral contact. These require a particular numerical treatment based on active set strategies, as will be explained in Sect. 6.4. As mentioned before, all standard terms (representing kinetic, internal and external virtual work) are formulated in the reference configuration, while the contact virtual work term $\delta \mathcal{W}_{\mathsf{co}}$ and the constraints $\delta \mathcal{W}_\lambda$ are typically formulated in the current configuration for the considered finite deformation contact problems. This is convenient due to the fact

that the contact mapping $\chi : \gamma_{\mathsf{c}}^{(1)} \to \gamma_{\mathsf{c}}^{(2)}$ needs to be evaluated with respect to the deformed geometry, anyway.

6.3 Finite Element Discretization

Similar to the tied contact case, all common types of first-order and second-order finite element interpolations in 2D and 3D are considered here, which again define finite dimensional subspaces $\boldsymbol{\mathcal{U}}_h^{(i)}$ and $\boldsymbol{\mathcal{V}}_h^{(i)}$ being approximations of $\boldsymbol{\mathcal{U}}^{(i)}$ and $\boldsymbol{\mathcal{V}}^{(i)}$, respectively. The general notations of slave and master side displacement interpolation given in (78), as well as the Lagrange multiplier interpolation defined in (80) are still valid. Substituting everything into the contact virtual work expression $\delta\mathcal{W}_{\mathsf{co}}$ in (109) yields

$$\begin{aligned} -\delta\mathcal{W}_{\mathsf{co},h} = &\sum_{j=1}^{m^{(1)}} \sum_{k=1}^{n^{(1)}} \boldsymbol{\lambda}_j^{\mathsf{T}} \left(\int_{\gamma_{\mathsf{c},h}^{(1)}} \Phi_j \, N_k^{(1)} \, \mathrm{d}A \right) \delta \mathbf{d}_k^{(1)} \\ &- \sum_{j=1}^{m^{(1)}} \sum_{l=1}^{n^{(2)}} \boldsymbol{\lambda}_j^{\mathsf{T}} \left(\int_{\gamma_{\mathsf{c},h}^{(1)}} \Phi_j \, (N_l^{(2)} \circ \chi_h) \, \mathrm{d}A \right) \delta \mathbf{d}_l^{(2)}. \end{aligned} \tag{116}$$

Herein, the only two differences to the mesh tying case lie in the integration domain (spatial description $\gamma_{\mathsf{c},h}^{(1)}$ instead of material description $\Gamma_{\mathsf{c},h}^{(1)}$) and in the fact that the discrete contact mapping $\chi_h : \gamma_{\mathsf{c},h}^{(1)} \to \gamma_{\mathsf{c},h}^{(2)}$ now continuously changes due to a relative movement of slave and master surfaces. Thus, as will be seen later on, it is not sufficient to evaluate the mapping only once as for mesh tying, but the mortar matrices $\mathbf{D}$ and $\mathbf{M}$ become deformation-dependent instead. Due to the fundamental importance of the discrete mortar matrices, their blockwise definition is repeated here, although only slightly modified as compared with (82) and (83), i.e.

$$\mathbf{D}[j,k] = D_{jk} \, \mathbf{I}_{\mathsf{ndim}} = \int_{\gamma_{\mathsf{c},h}^{(1)}} \Phi_j N_k^{(1)} \mathrm{d}A \; \mathbf{I}_{\mathsf{ndim}}, \tag{117}$$

$$\mathbf{M}[j,l] = M_{jl} \, \mathbf{I}_{\mathsf{ndim}} = \int_{\gamma_{\mathsf{c},h}^{(1)}} \Phi_j (N_l^{(2)} \circ \chi_h) \, \mathrm{d}A \; \mathbf{I}_{\mathsf{ndim}}, \tag{118}$$

where $j = 1, \ldots, m^{(1)}$, $k = 1, \ldots, n^{(1)}$, $l = 1, \ldots, n^{(2)}$. In analogy to (84), the discrete contact virtual work contribution can be expressed as

$$-\delta\mathcal{W}_{\mathsf{co},h} = \delta\mathbf{d}_{\mathcal{S}}^{\mathsf{T}} \mathbf{D}^{\mathsf{T}} \boldsymbol{\lambda} - \delta\mathbf{d}_{\mathcal{M}}^{\mathsf{T}} \mathbf{M}^{\mathsf{T}} \boldsymbol{\lambda} = \delta\mathbf{d}^{\mathsf{T}} \underbrace{\begin{bmatrix} \mathbf{0} \\ -\mathbf{M}^{\mathsf{T}} \\ \mathbf{D}^{\mathsf{T}} \end{bmatrix}}_{\mathbf{B}_{\mathsf{co}}(\mathbf{d})^{\mathsf{T}}} \boldsymbol{\lambda} = \delta\mathbf{d}^{\mathsf{T}} \mathbf{f}_{\mathsf{co}}(\mathbf{d}, \boldsymbol{\lambda}), \tag{119}$$

where the discrete mortar contact operator $\mathbf{B}_{\mathsf{co}}(\mathbf{d})$ and the resulting discrete vector of contact forces $\mathbf{f}_{\mathsf{co}}(\mathbf{d}, \boldsymbol{\lambda}) = \mathbf{B}_{\mathsf{co}}(\mathbf{d})^{\mathsf{T}}\boldsymbol{\lambda}$ acting on slave and master sides of the interface now depend *nonlinearly* on the current deformation state $\mathbf{d}$.

Next, the focus is shifted towards the weak constraint contribution for frictionless contact defined in (115), where more profound differences to the mesh tying case can be expected. As shown in great detail in Hüeber (2008), the discretized version of the weak formulation in (114) and (115) is equivalent to the following set of pointwise conditions:

$$(\tilde{g}_{\mathsf{n}})_j \geq 0, \quad (\lambda_{\mathsf{n}})_j \geq 0, \quad (\tilde{g}_{\mathsf{n}})_j(\lambda_{\mathsf{n}})_j = 0, \quad j = 1, \ldots, m^{(1)}, \tag{120}$$

where the discrete weighted gap $(\tilde{g}_{\mathsf{n}})_j$ at slave node j is given by

$$(\tilde{g}_{\mathsf{n}})_j = \int_{\gamma_{\mathsf{c}}^{(1)}} \Phi_j \, g_{\mathsf{n},h} \, \mathrm{d}A. \tag{121}$$

Herein, $g_{\mathsf{n},h}$ is the discretized version of the gap function g_{n} introduced in (24). Examining the last two equations in more detail, an interesting analogy becomes apparent. Basically, (120) represents nothing less than a discrete formulation of the original KKT conditions in (102) with an additional weighting based on the Lagrange multiplier shape functions Φ_j. It is worth noting that although a segment-based (mortar) approach has been followed, decoupled constraints at the discrete nodal points are eventually enforced independently, just as it is well-known from traditional NTS schemes. However, the nodal constraints (120) in the mortar formulation convey a substantially increased level of information as compared with the truly nodal constraints in a NTS formulation, owing to the underlying variational approach which is algebraically reflected in the weighted (integral) gap formulation in (121).

For the sake of completeness, it should be pointed out that the nodal decoupling of constraints and thus the final formulation given in (120) is strictly speaking only valid for *dual* Lagrange multiplier interpolation, see Hüeber (2008) for the corresponding mathematical proof, which relies on biorthogonality as defined in (144). In the case of *standard* Lagrange multiplier interpolation, the conversion of (114) and (115) into (120) involves an additional, yet only slight, approximation, see Hüeber (2008). Finally, the frictionless sliding constraint contained in the definition of the convex cone $\mathcal{M}^+$ is readily enforced on a discrete nodal basis, i.e. $(\boldsymbol{\lambda}_\tau)_j = \mathbf{0}$. To sum up, the final space discretized but still time continuous problem formulation, consisting of the semi-discrete equations of motion and the frictionless contact constraints for all slave nodes also carrying discrete Lagrange multiplier degrees of freedom, can be expressed as

$$\mathbf{M}\ddot{\mathbf{d}} + \mathbf{C}\dot{\mathbf{d}} + \mathbf{f}_{\mathsf{int}}(\mathbf{d}) + \mathbf{f}_{\mathsf{co}}(\mathbf{d}, \boldsymbol{\lambda}) - \mathbf{f}_{\mathsf{ext}} = \mathbf{0}, \tag{122}$$

$$(\tilde{g}_{\mathsf{n}})_j \geq 0, \quad (\lambda_{\mathsf{n}})_j \geq 0, \quad (\tilde{g}_{\mathsf{n}})_j(\lambda_{\mathsf{n}})_j = 0, \quad j = 1, \ldots, m^{(1)}, \tag{123}$$

$$(\boldsymbol{\lambda}_\tau)_j = \mathbf{0}, \quad j = 1, \ldots, m^{(1)}. \tag{124}$$

While this finite element formulation has some strong similarities with the mesh tying case in (86) and (87), it also contains three striking additional complexities. Firstly, unilateral contact involves inequality constraints, which require a suitable active set strategy as part of the global solution algorithm (cf. Sect. 6.4). Secondly, normal and tangential contact directions need to be treated separately in order to enforce the different underlying physical principles (non-penetration, frictionless or frictional sliding). Thirdly, and most importantly from the viewpoint of implementation, the contact forces in (122) as well as the contact constraints in (123) and (124) are deformation-dependent. This introduces an additional nonlinearity into the global system and thus demands for an incessant re-evaluation of mortar coupling terms including a consistent linearization for implicit time integration. Corresponding extensions of the numerical integration scheme for the discrete contact operator $\mathbf{B}_{\mathsf{co}}(\mathbf{d})$ and the discrete weighted gaps $(\tilde{g}_{\mathsf{n}})_j$ in both 2D and 3D will be presented in the next three paragraphs.

Finally, a short outlook is also given on the weak constraint contribution for frictional contact according to Coulomb's law as defined in (110), although the frictional part is not in the focus of interest here. Again, it has been shown in great detail in Hüeber (2008) and can be readily understood that the discretized version of the tangential part of the weak formulation in (108) and (110) is equivalent to the following set of pointwise conditions:

$$\begin{aligned} \Phi_j &:= \|(\boldsymbol{\lambda}_\tau)_j\| - \mathfrak{F}|(\lambda_{\mathsf{n}})_j| \leq 0, \\ (\tilde{\mathbf{v}}_{\tau,\mathsf{rel}})_j + \beta_j(\boldsymbol{\lambda}_\tau)_j &= \mathbf{0}, \quad \beta_j \geq 0, \quad \Phi_j\beta_j = 0, \quad j = 1,\ldots,m^{(1)}. \end{aligned} \tag{125}$$

where the discrete relative tangential velocity $(\tilde{\mathbf{v}}_{\tau,\mathsf{rel}})_j$ at slave node j is determined such that it satisfies the requirement of frame indifference, see e.g. Yang et al. (2005) and Gitterle et al. (2010) for further explanations. Similar to the non-penetration condition, it can be observed that (125) basically represents a weak formulation of the original Coulomb friction conditions in (103) with an additional weighting based on the Lagrange multiplier shape functions Φ_j. In the semi-discrete formulation for Coulomb friction, the set of conditions in (125) would simply replace (124), while (122) and (123) would remain unchanged. While by no means exhaustive, the given outlook demonstrates that an extension of the proposed mortar finite element framework towards any tangential constitutive law (e.g. Tresca friction, Coulomb friction) is pretty straightforward. Most importantly, the discrete frictional expressions such as the discrete relative tangential velocity $(\tilde{\mathbf{v}}_{\tau,\mathsf{rel}})_j$ do not require any additional numerical integration efforts, but can rather be constructed from the well-known mortar matrices $\mathbf{D}$ and $\mathbf{M}$ (including history values due to path dependency) and the nodal tangent vectors $\boldsymbol{\tau}_j^{\xi}$ and $\boldsymbol{\tau}_j^{\eta}$ defined in (89).

The main steps for evaluating the entries of the mortar integral matrices $\mathbf{D}$ and $\mathbf{M}$ in 3D will be presented in Sect. 7.3 in the context of tied contact and can be directly transferred to unilateral contact. Concretely, this encompasses the definition of averaged nodal normal vectors and the 3D mortar segmentation algorithm (cf. Fig. 18) with its associated projection, clipping and triangulation procedures.

6.4 Active Set Strategy and Semi-smooth Newton Methods

As mentioned before, the semi-discrete problem statement of unilateral contact in (122)–(124), and in particular its final fully discretized version (i.e. after time discretization with the generalized-α method previously discussed in Sect. 4.3), causes one major additional complexity with regard to global solution schemes as compared with the mesh tying case, namely the contact specific *inequality* constraints, which divide the set of all discrete constraints (i.e. the equivalent of all slave nodes) into two a priori unknown sets of active and inactive constraints. Mathematically speaking, this introduces an additional source of nonlinearity apart from the well-known geometrical and material nonlinearities of nonlinear solid mechanics. To resolve this contact nonlinearity, so-called primal-dual active set strategies (PDASS) will be employed in the solution algorithms developed here.

The idea of any active set strategy in the context of unilateral contact is to find the correct subset of all slave nodes which are in contact with the master surface at the end of the currently considered time interval $[t_n, t_{n+1}]$. As discussed in Sect. 6.3, the contact constraints can be enforced nodally at each slave node $j \in \mathcal{S}$, with $j = 1, \ldots, m^{(1)}$, despite the fact that a segment-based mortar approach is employed here. Consequently, the so-called active set $\mathcal{A} \subseteq \mathcal{S}$ defines a subset of the set of all slave nodes $\mathcal{S}$, and the definition of the inactive set $\mathcal{I} = \mathcal{S} \setminus \mathcal{A}$ is straightforward. Before considering possible formulations of active set strategies, the final KKT conditions defined in (123) are repeated here, with the time index $n+1$ being omitted in the following for the sake of notational simplicity, i.e.

$$\begin{aligned} (\tilde{\boldsymbol{g}}_{\mathsf{n}})_j &\geq 0, \quad \forall\, j \in \mathcal{S} \\ (\lambda_{\mathsf{n}})_j &\geq 0, \quad \forall\, j \in \mathcal{S} \\ (\tilde{\boldsymbol{g}}_{\mathsf{n}})_j (\lambda_{\mathsf{n}})_j &= 0, \quad \forall\, j \in \mathcal{S}. \end{aligned} \tag{126}$$

The aforementioned definitions of the active set and the inactive set in combination with the complementarity condition $(\tilde{\boldsymbol{g}}_{\mathsf{n}})_j (\lambda_{\mathsf{n}})_j = 0$ motivate a first, naive reformulation of the KKT conditions using only equality constraints:

$$\begin{aligned} (\tilde{\boldsymbol{g}}_{\mathsf{n}})_j &= 0, \quad \forall\, j \in \mathcal{A} \\ (\lambda_{\mathsf{n}})_j &= 0, \quad \forall\, j \in \mathcal{I} \\ (\tilde{\boldsymbol{g}}_{\mathsf{n}})_j (\lambda_{\mathsf{n}})_j &= 0, \quad \forall\, j \in \mathcal{S}. \end{aligned} \tag{127}$$

Obviously, the PDASS in (127) suffers from a serious drawback: the contact nonlinearity, i.e. finding the correct active set $\mathcal{A}$ can not be resolved by a Newton–Raphson type approach. This is due to the fact that no directional derivative of the sets themselves with respect to the nodal displacements $\mathbf{d}$ can be extracted from (127). Instead, the given formulation inevitably leads to two nested iterative solution schemes, with the outer (fixed-point type) loop solving for the correct active set and the inner (Newton–Raphson type) loop solving a constrained nonlinear finite element prob-

lem while the active set is *fixed*. Consequently, this approach does not provide the desired efficiency and will not be followed any further in this contribution. Further information on such a fixed-point type treatment of the active set in the context of finite deformation mortar contact can for instance be found in Hartmann et al. (2007) and Hesch and Betsch (2009).

Based on the above considerations, the basic idea of an alternative PDASS formulation is to rearrange the KKT conditions such that a Newton–Raphson type algorithm can be applied not only for geometrical and material nonlinearities, but also for the nonlinearity stemming from contact itself, i.e. the active set search. The resulting primal-dual active set approach is well-known from the general mathematical literature on constrained optimization, see e.g. in Hintermüller et al. (2002) and Qi and Sun (1993), and can equivalently be interpreted as a semi-smooth Newton method. Applications to classical NTS contact formulations can be found in Alart and Curnier (1991), Christensen et al. (1998) and Strömberg et al. (1996), and small deformation mortar contact has been investigated in Hüeber and Wohlmuth (2005). Here, the first successful consistent extension to a finite deformation mortar contact formulation is presented, cf. also Popp et al. (2009, 2010). The main idea is to reformulate the discrete KKT conditions within a so-called nonlinear complementarity (NCP) function, where all details for frictionless and frictional contact are given in the upcoming paragraphs. For the sake of completeness, it should be mentioned that the concept of NCP functions is also applicable to other well-known solid mechanics problems involving inequality constraints such as computational plasticity. For a comprehensive and more general overview, the reader is exemplarily referred to Hager (2010).

The first step for frictionless contact is to reformulate the discrete KKT-conditions in (126) within a complementarity function C_j for each slave node $j \in \mathcal{S}$ as

$$C_j\left(\mathbf{d}, \boldsymbol{\lambda}\right) = (\lambda_{\mathsf{n}})_j - \max\left(0, (\lambda_{\mathsf{n}})_j - c_{\mathsf{n}}(\tilde{g}_{\mathsf{n}})_j\right) = 0, \qquad c_{\mathsf{n}} > 0. \tag{128}$$

This is a nonlinear function of the discrete displacements as both the nodal normal vector $\mathbf{n}_j$ in $(\lambda_{\mathsf{n}})_j = \mathbf{n}_j \cdot \boldsymbol{\lambda}_j$ and the nodal weighted gap $(\tilde{g}_{\mathsf{n}})_j$ defined in (121) depend nonlinearly on $\mathbf{d}$. It can be easily shown that the resulting equality constraint $C_j = 0$ is *equivalent* to the complete set of KKT inequality conditions in (126), and that this equivalence holds for arbitrary positive values of the so-called complementarity parameter c_{n}. The concrete role of c_{n} will be explained later in this paragraph. Figure 9 exemplarily illustrates the nodal complementarity function and emphasizes the equivalence with the KKT conditions.

It is important to see that a distinction between the active set $\mathcal{A}$ and the inactive set $\mathcal{I}$ is implicitly contained in the complementarity function C_j: the max-function is non-smooth and thus consists of two different solution branches. In other words, C_j provides a certain regularization of the non-smooth decision between each slave node being currently active or inactive, yet without introducing any additional approximation. Thus, the resulting PDASS contains derivative information on the sets themselves and allows for the application of a Newton–Raphson type solution scheme also for the nonlinearity stemming from contact. Consequently, all sources of nonlinearities, i.e. finite deformations, nonlinear material behavior and contact itself, can be

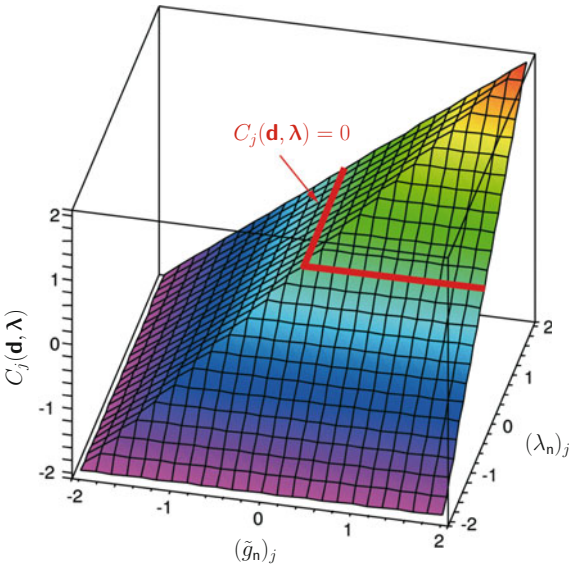

Fig. 9 Exemplary nodal NCP function $C_j(\mathbf{d}, \boldsymbol{\lambda})$ as a function of the nodal weighted gap $(\tilde{\boldsymbol{g}}_\mathsf{n})_j$ and the normal part of the nodal Lagrange multiplier $(\lambda_\mathsf{n})_j$ for a complementarity parameter $c_\mathsf{n} = 1$. The equivalence with the KKT conditions is indicated in red color. Reprinted with permission from Popp et al. (2009), © 2009 John Wiley & Sons, Ltd.

treated within one single iterative scheme. While C_j is a continuous function, it is non-smooth and has no uniquely defined derivative at the positions $(\lambda_\mathsf{n})_j - c_\mathsf{n}(\tilde{\boldsymbol{g}}_\mathsf{n})_j = 0$. Yet, it is well-known from mathematical literature on constrained optimization that the *max*-function can be classified as so-called semi-smooth function, and therefore a semi-smooth (or generalized) Newton method can still be applied. The interested reader is referred to Hintermüller et al. (2002) and Qi and Sun (1993) for more detailed information on semi-smooth Newton methods, for example including a concise proof of their superlinear local convergence behavior. The actual linearization of the NCP function in (128) is based on the concept of generalized derivatives (e.g. the generalized derivative of the max-function) and has been presented in the author's original work in Popp et al. (2009, 2010) along with the remaining parts of the global solution algorithm.

It should be pointed out that the complementarity parameter c_n represents a purely algorithmic parameter. Although quite some similarities appear at first sight, c_n is in stark contrast to a penalty parameter, because it does not influence the accuracy of results. Instead, the weak non-penetration condition in (126) will be satisfied exactly, as can be expected from a Lagrange multiplier method. The choice of c_n only improves or deteriorates convergence of the resulting semi-smooth Newton method. In Hüeber and Wohlmuth (2005), c_n has been suggested to be chosen at the order of Young's modulus E of the contacting bodies to obtain optimal convergence. Numerical investigations for 2D and 3D mortar contact in Popp et al. (2009, 2010), though, have shown very little influence on semi-smooth Newton convergence along a very broad spectrum of values for c_n. Even for relatively large step sizes and fine

contacting meshes, the correct active set is usually found after only a few Newton steps. Once the sets remain constant, of course, quadratic convergence is obtained due to the underlying consistent linearization.

Examining the NCP function for frictionless contact in (128) in more detail allows for an interesting and important observation: there exists a certain similarity between the proposed PDASS with its algorithmic realization as semi-smooth Newton method and the classical Augmented Lagrange method, see also the seminal paper by Alart and Curnier (1991) in this context. Simply speaking, the Augmented Lagrange approach as discussed in Alart and Curnier (1991) aims at a regularized variational formulation, while the PDASS and NCP function concept applies at a later stage with a regularized constraint enforcement. Again, no detailed derivation of the Coulomb friction case is given here, but the interested reader is instead referred to Hüeber et al. (2008), Gitterle et al. (2010), Gitterle (2012) and Wohlmuth (2011) for all details on the semi-smooth Newton approach for frictional contact problems.

6.5 Solution Methods

Again, the final system consists of $\mathsf{ndof} + \mathsf{nco}$ nonlinear discrete algebraic equations, where the number of constraints is $\mathsf{nco} = \mathsf{ndim} \cdot m^{(1)}$. While standard (smooth) Newton–Rapshon methods were the method of choice for mesh tying problems in Sect. 5.5, the active set strategies now require a semi-smooth Newton approach as discussed in the last paragraph. Nevertheless, for frictionless contact this non-smoothness solely affects the contact constraints in normal direction in (123) or to be more precise their reformulation as NCP function in (128). All remaining parts of the nonlinear system, i.e. both the discrete equilibrium of forces in (122) and the frictionless sliding conditions in (124) still show a smooth behavior.

As explained in Sect. 4.4, the Newton–Raphson method is based on a subsequent linearization of the residual, here defined by the discrete balance of linear momentum in (122) and the discrete contact constraints in (124) and (128). Each nonlinear solution step (iteration index i) then consists of solving the resulting linearized system of equations and applying an incremental update of the unknown displacements $\mathbf{d}_{n+1}$ and Lagrange multipliers $\boldsymbol{\lambda}_{n+1}$ until a user-defined convergence criterion is met. Examining the residual in (122) in more detail, an important difference to the mesh tying case becomes apparent: the contact operator $\mathbf{B}_{\mathsf{co}}(\mathbf{d})$ defined in (119), and thus the contact forces $\mathbf{f}_{\mathsf{co}}(\mathbf{d}, \boldsymbol{\lambda})$, depend nonlinearly on the displacements and yield additional contact stiffness blocks when being linearized, i.e.

$$\begin{aligned} \left[\mathbf{K}_{\mathsf{effdyn}}(\mathbf{d}_{n+1}^{i}) + (1-\alpha_{\mathsf{f}})\mathbf{K}_{\mathsf{co}}(\mathbf{d}_{n+1}^{i}, \boldsymbol{\lambda}_{n+1}^{i})\right] \Delta\mathbf{d}_{n+1}^{i+1} + \\ +(1-\alpha_{\mathsf{f}})\mathbf{B}_{\mathsf{co}}(\mathbf{d}_{n+1}^{i})\boldsymbol{\lambda}_{n+1}^{i+1} = -\mathbf{r}_{\mathsf{effdyn}}(\mathbf{d}_{n+1}^{i}) - \alpha_{\mathsf{f}}\mathbf{B}_{\mathsf{co}}(\mathbf{d}_{n})\boldsymbol{\lambda}_{n}. \end{aligned} \tag{129}$$

Herein, the contact stiffness $\mathbf{K}_{\mathrm{co}}$ is defined as

$$\mathbf{K}_{\mathrm{co}}(\mathbf{d}_{n+1}^{i},\boldsymbol{\lambda}_{n+1}^{i}) = \left.\frac{\partial(\mathbf{B}_{\mathrm{co}}(\mathbf{d}_{n+1})\boldsymbol{\lambda}_{n+1})}{\partial\mathbf{d}_{n+1}}\right|^{i}. \tag{130}$$

Moreover, it should be pointed out that contact-related quantities from the last converged time step n appear on the right-hand side of (129) due to the employed generalized-α time integration in combination with a trapezoidal rule interpolation of the contact forces. Similar to the mesh tying case, the interface forces are still linear with respect to the discrete Lagrange multipliers. Consequently, it is possible to solve directly for $\boldsymbol{\lambda}_{n+1}^{i+1}$ in each iteration step and no incremental formulation is needed.

Repeatedly performing semi-smooth Newton steps (iteration index i), each to be solved for the primal-dual pair of discrete variables $(\Delta\mathbf{d}_{n+1}^{i+1},\boldsymbol{\lambda}_{n+1}^{i+1})$, yields the following solution algorithm within the time step $[t_n, t_{n+1}]$:

Algorithm 1

1. Set $i=0$ and initialize the solution $(\mathbf{d}_{n+1}^{0},\boldsymbol{\lambda}_{n+1}^{0})$
2. Initialize $\mathcal{A}_{n+1}^{0}$ and $\mathcal{I}_{n+1}^{0}$ such that $\mathcal{A}_{n+1}^{0}\cup\mathcal{I}_{n+1}^{0}=\mathcal{S}$
3. Find the primal-dual pair $(\Delta\mathbf{d}_{n+1}^{i+1},\boldsymbol{\lambda}_{n+1}^{i+1})$ by solving

$$\tilde{\mathbf{K}}_{\mathrm{effdyn,co}}\Delta\mathbf{d}_{n+1}^{i+1} + (1-\alpha_{\mathrm{f}})\mathbf{B}_{\mathrm{co}}(\mathbf{d}_{n+1}^{i})\boldsymbol{\lambda}_{n+1}^{i+1} = -\tilde{\mathbf{r}}_{\mathrm{effdyn,co}}, \tag{131}$$

$$(\boldsymbol{\lambda}_{j})_{n+1}^{i+1} = \mathbf{0} \quad \forall\, j\in\mathcal{I}_{n+1}^{i}, \tag{132}$$

$$\Delta((\tilde{g}_{\mathsf{n}})_{j})_{n+1}^{i} + ((\tilde{g}_{\mathsf{n}})_{j})_{n+1}^{i} = 0 \quad \forall\, j\in\mathcal{A}_{n+1}^{i}, \tag{133}$$

$$\Delta(\boldsymbol{\tau}_{j}^{\xi})_{n+1}^{i}(\boldsymbol{\lambda}_{j})_{n+1}^{i} + (\boldsymbol{\tau}_{j}^{\xi})_{n+1}^{i}(\boldsymbol{\lambda}_{j})_{n+1}^{i+1} = 0 \quad \forall\, j\in\mathcal{S}, \tag{134}$$

$$\Delta(\boldsymbol{\tau}_{j}^{\eta})_{n+1}^{i}(\boldsymbol{\lambda}_{j})_{n+1}^{i} + (\boldsymbol{\tau}_{j}^{\eta})_{n+1}^{i}(\mathbf{l}_{j})_{n+1}^{i+1} = 0 \quad \forall\, j\in\mathcal{S}. \tag{135}$$

4. Update $\mathbf{d}_{n+1}^{i+1} = \mathbf{d}_{n+1}^{i} + \Delta\mathbf{d}_{n+1}^{i+1}$
5. Set $\mathcal{A}_{n+1}^{i+1}$ and $\mathcal{I}_{n+1}^{i+1}$ to

$$\begin{aligned}\mathcal{I}_{n+1}^{i+1} &:= \left\{ j\in\mathcal{S} \,|\, ((\lambda_{\mathsf{n}})_{j})_{n+1}^{i+1} - c_{\mathsf{n}}((\tilde{g}_{\mathsf{n}})_{j})_{n+1}^{i+1} \geq 0 \right\},\\ \mathcal{A}_{n+1}^{i+1} &:= \left\{ j\in\mathcal{S} \,|\, ((\lambda_{\mathsf{n}})_{j})_{n+1}^{i+1} - c_{\mathsf{n}}((\tilde{g}_{\mathsf{n}})_{j})_{n+1}^{i+1} < 0 \right\}.\end{aligned} \tag{136}$$

6. If $\mathcal{A}_{n+1}^{i+1}=\mathcal{A}_{n+1}^{i}$, $\mathcal{I}_{n+1}^{i+1}=\mathcal{I}_{n+1}^{i}$ and $\|\mathbf{r}_{\mathrm{tot}}\|\leq\epsilon_{\mathrm{r}}$, then stop, else set $i:=i+1$ and go to step (3).

Herein, the following abbreviations have been introduced for notational simplicity:

$$\tilde{\mathbf{K}}_{\mathrm{effdyn,co}} = \mathbf{K}_{\mathrm{effdyn}}(\mathbf{d}_{n+1}^{i}) + (1-\alpha_{\mathrm{f}})\mathbf{K}_{\mathrm{co}}(\mathbf{d}_{n+1}^{i},\boldsymbol{\lambda}_{n+1}^{i}), \tag{137}$$

$$\tilde{\mathbf{r}}_{\mathrm{effdyn,co}} = \mathbf{r}_{\mathrm{effdyn}}(\mathbf{d}_{n+1}^{i}) + \alpha_{\mathrm{f}}\mathbf{B}_{\mathrm{co}}(\mathbf{d}_{n})\boldsymbol{\lambda}_{n}. \tag{138}$$

Moreover, the variable ϵ_{r} denotes an absolute Newton convergence tolerance for the L^2-norm of the total residual vector $\mathbf{r}_{\mathsf{tot}}$, which comprises the force residual and the residual of the contact constraints (132)–(135). All types of nonlinearities including the search for the correct active set are resolved within one single nonlinear solution scheme, with the sets $\mathcal{I}_{n+1}^{i}$ and $\mathcal{A}_{n+1}^{i}$ being updated after each semi-smooth Newton step.

The convergence behavior of the resulting solution scheme is very good. As long as the correct active set is not found, and thus the contact typical non-smoothness is not yet resolved, locally superlinear convergence rates are obtained, see e.g. Hintermüller et al. (2002). Once the sets are fixed, the nonlinear iteration scheme reduces to a standard (smooth) Newton–Raphson method, and thus even locally quadratic convergence rates are achieved in the limit owing to the underlying consistent linearization. While not discussed here, similar observations can also be made for frictional contact according to Coulomb's law and the associated search for the correct stick and slip sets, see e.g. Gitterle et al. (2010), Gitterle (2012) and Hüeber et al. (2008).

In this section, an algebraic representation of the linearized system to be solved within each semi-smooth Newton step is derived and globally assembled matrix notations for the directional derivatives in (131)–(135) are provided. With the assembly procedure itself being rather straightforward in finite element methods, only the final results are given here. The final system to be solved within each semi-smooth Newton step can be expressed as follows:

$$\begin{bmatrix} \tilde{\mathbf{K}}_{\mathcal{N}\mathcal{N}} & \tilde{\mathbf{K}}_{\mathcal{N}\mathcal{M}} & \tilde{\mathbf{K}}_{\mathcal{N}\mathcal{I}} & \tilde{\mathbf{K}}_{\mathcal{N}\mathcal{A}} & \mathbf{0} & \mathbf{0} \\ \tilde{\mathbf{K}}_{\mathcal{M}\mathcal{N}} & \tilde{\mathbf{K}}_{\mathcal{M}\mathcal{M}} & \tilde{\mathbf{K}}_{\mathcal{M}\mathcal{I}} & \tilde{\mathbf{K}}_{\mathcal{M}\mathcal{A}} & -a\mathbf{M}_{\mathcal{I}}^{\mathsf{T}} & -a\mathbf{M}_{\mathcal{A}}^{\mathsf{T}} \\ \tilde{\mathbf{K}}_{\mathcal{I}\mathcal{N}} & \tilde{\mathbf{K}}_{\mathcal{I}\mathcal{M}} & \tilde{\mathbf{K}}_{\mathcal{I}\mathcal{I}} & \tilde{\mathbf{K}}_{\mathcal{I}\mathcal{A}} & a\mathbf{D}_{\mathcal{I}\mathcal{I}}^{\mathsf{T}} & a\mathbf{D}_{\mathcal{I}\mathcal{A}}^{\mathsf{T}} \\ \tilde{\mathbf{K}}_{\mathcal{A}\mathcal{N}} & \tilde{\mathbf{K}}_{\mathcal{A}\mathcal{M}} & \tilde{\mathbf{K}}_{\mathcal{A}\mathcal{I}} & \tilde{\mathbf{K}}_{\mathcal{A}\mathcal{A}} & a\mathbf{D}_{\mathcal{A}\mathcal{I}}^{\mathsf{T}} & a\mathbf{D}_{\mathcal{A}\mathcal{A}}^{\mathsf{T}} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{\mathcal{I}} & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_{\mathcal{M}} & \mathbf{N}_{\mathcal{I}} & \mathbf{N}_{\mathcal{A}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{F}_{\mathcal{I}} & \mathbf{F}_{\mathcal{A}} & \mathbf{0} & \mathbf{T}_{\mathcal{A}} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{d}_{n+1,\mathcal{N}} \\ \Delta\mathbf{d}_{n+1,\mathcal{M}} \\ \Delta\mathbf{d}_{n+1,\mathcal{I}} \\ \Delta\mathbf{d}_{n+1,\mathcal{A}} \\ \boldsymbol{\lambda}_{n+1,\mathcal{I}} \\ \boldsymbol{\lambda}_{n+1,\mathcal{A}} \end{bmatrix} = - \begin{bmatrix} \tilde{\mathbf{r}}_{\mathcal{N}} \\ \tilde{\mathbf{r}}_{\mathcal{M}} \\ \tilde{\mathbf{r}}_{\mathcal{I}} \\ \tilde{\mathbf{r}}_{\mathcal{A}} \\ \mathbf{0} \\ \tilde{\mathbf{g}}_{\mathcal{A}} \\ \mathbf{0} \end{bmatrix} . \quad (139)$$

Herein, the scalar $a := 1 - \alpha_{\mathsf{f}}$ abbreviates the weighting factor introduced by generalized-α time integration. Moreover, the nonlinear iteration index i as well as the subscript $\cdot_{\mathsf{effdyn,co}}$ of the residual vector $\tilde{\mathbf{r}}_{\mathsf{effdyn,co}}$ given in (137) and the effective stiffness matrix $\tilde{\mathbf{K}}_{\mathsf{effdyn,co}}$ defined in (138) have been omitted for the ease of notation.

Again, as has been the case for mesh tying, the *dual* Lagrange multiplier approach can be beneficially exploited to simplify the final linear system of equations. In a first step, the Lagrange multipliers $\boldsymbol{\lambda}_{n+1,\mathcal{I}}$ associated with inactive slave nodes are easily condensed by simply extracting the identity $\boldsymbol{\lambda}_{n+1,\mathcal{I}} = \mathbf{0}$ from the fifth row of (139). This basically removes the fifth row and the fifth column of the original saddle point system. More importantly, based on the fourth row of (139), the Lagrange multipliers $\boldsymbol{\lambda}_{n+1,\mathcal{A}}$ associated with active slave nodes can be expressed as

$$\boldsymbol{\lambda}_{n+1,\mathcal{A}} = \frac{1}{a}\mathbf{D}_{\mathcal{A}\mathcal{A}}^{-\mathsf{T}}\Big(-\tilde{\mathbf{r}}_{\mathcal{A}} - \tilde{\mathbf{K}}_{\mathcal{A}\mathcal{N}}\Delta\mathbf{d}_{n+1,\mathcal{N}} - \tilde{\mathbf{K}}_{\mathcal{A}\mathcal{M}}\Delta\mathbf{d}_{n+1,\mathcal{M}} - \tilde{\mathbf{K}}_{\mathcal{A}\mathcal{I}}\Delta\mathbf{d}_{n+1,\mathcal{I}} - \tilde{\mathbf{K}}_{\mathcal{A}\mathcal{A}}\Delta\mathbf{d}_{n+1,\mathcal{A}}\Big). \tag{140}$$

As will be discussed in Sect. 7.3, the active part of the mortar projection operator $\mathbf{P} = \mathbf{D}^{-1}\mathbf{M}$ can be defined as

$$\mathbf{P}_{\mathcal{A}} = \mathbf{D}_{\mathcal{A}\mathcal{A}}^{-1}\mathbf{M}_{\mathcal{A}}. \tag{141}$$

Inserting (140) into the second and seventh row of (139) yields

$$\begin{bmatrix} \tilde{\mathbf{K}}_{\mathcal{N}\mathcal{N}} & \tilde{\mathbf{K}}_{\mathcal{N}\mathcal{M}} & \tilde{\mathbf{K}}_{\mathcal{N}\mathcal{I}} & \tilde{\mathbf{K}}_{\mathcal{N}\mathcal{A}} \\ \tilde{\mathbf{K}}_{\mathcal{M}\mathcal{N}} + \mathbf{P}_{\mathcal{A}}^{\mathsf{T}}\tilde{\mathbf{K}}_{\mathcal{A}\mathcal{N}} & \tilde{\mathbf{K}}_{\mathcal{M}\mathcal{M}} + \mathbf{P}_{\mathcal{A}}^{\mathsf{T}}\tilde{\mathbf{K}}_{\mathcal{A}\mathcal{M}} & \tilde{\mathbf{K}}_{\mathcal{M}\mathcal{I}} + \mathbf{P}_{\mathcal{A}}^{\mathsf{T}}\tilde{\mathbf{K}}_{\mathcal{A}\mathcal{I}} & \tilde{\mathbf{K}}_{\mathcal{M}\mathcal{A}} + \mathbf{P}_{\mathcal{A}}^{\mathsf{T}}\tilde{\mathbf{K}}_{\mathcal{A}\mathcal{A}} \\ \tilde{\mathbf{K}}_{\mathcal{I}\mathcal{N}} & \tilde{\mathbf{K}}_{\mathcal{I}\mathcal{M}} & \tilde{\mathbf{K}}_{\mathcal{I}\mathcal{I}} & \tilde{\mathbf{K}}_{\mathcal{I}\mathcal{A}} \\ \mathbf{0} & \mathbf{N}_{\mathcal{M}} & \mathbf{N}_{\mathcal{I}} & \mathbf{N}_{\mathcal{A}} \\ a\,\mathbf{T}_{\mathcal{A}}\mathbf{D}_{\mathcal{A}\mathcal{A}}^{-1}\tilde{\mathbf{K}}_{\mathcal{A}\mathcal{N}} & a\,\mathbf{T}_{\mathcal{A}}\mathbf{D}_{\mathcal{A}\mathcal{A}}^{-1}\tilde{\mathbf{K}}_{\mathcal{A}\mathcal{M}} & a\,\mathbf{T}_{\mathcal{A}}\mathbf{D}_{\mathcal{A}\mathcal{A}}^{-1}\tilde{\mathbf{K}}_{\mathcal{A}\mathcal{I}} - \mathbf{F}_{\mathcal{I}} & a\,\mathbf{T}_{\mathcal{A}}\mathbf{D}_{\mathcal{A}\mathcal{A}}^{-1}\tilde{\mathbf{K}}_{\mathcal{A}\mathcal{A}} - \mathbf{F}_{\mathcal{A}} \end{bmatrix} \cdot \begin{bmatrix} \Delta\mathbf{d}_{n+1,\mathcal{N}} \\ \Delta\mathbf{d}_{n+1,\mathcal{M}} \\ \Delta\mathbf{d}_{n+1,\mathcal{I}} \\ \Delta\mathbf{d}_{n+1,\mathcal{A}} \end{bmatrix} = - \begin{bmatrix} \tilde{\mathbf{r}}_{\mathcal{N}} \\ \tilde{\mathbf{r}}_{\mathcal{M}} + \mathbf{P}_{\mathcal{A}}^{\mathsf{T}}\tilde{\mathbf{r}}_{\mathcal{A}} \\ \tilde{\mathbf{r}}_{\mathcal{I}} \\ \tilde{\mathbf{g}}_{\mathcal{A}} \\ a\,\mathbf{T}_{\mathcal{A}}\mathbf{D}_{\mathcal{A}\mathcal{A}}^{-1}\tilde{\mathbf{r}}_{\mathcal{A}} \end{bmatrix}. \tag{142}$$

While inevitable for standard Lagrange multiplier interpolation, the undesirable saddle point structure of (139) with its typical zero diagonal block has successfully been removed. Finally, it should be mentioned that the discrete Lagrange multipliers, and thus their physical interpretation as contact tractions, are recovered from the displacement solution in a variationally consistent way. This recovery can be performed as a pure postprocessing step at the end of each time interval based on the relation given in (140).

6.6 Numerical Example

The numerical example presented in this section demonstrates the applicability of the proposed mortar contact formulations, including the parallel search algorithms and dynamic load balancing strategies to be described in Sect. 7.2, for large-scale simulations on parallel high-performance computing (HPC) systems. The investigated setup, illustrated in Fig. 10, consists of two thin-walled tori with a Neo–Hookean material model ($E = 3000$, $\nu = 0.3$, $\rho_0 = 0.1$). The major and minor radius of the two hollow tori is 76 and 24, respectively, and the wall thickness is 4.5. The lower torus lies in the xy-plane and the upper torus is rotated around the y-axis by 45 degrees. Both the chosen geometry and loading conditions are inspired by a very similar analysis presented in Yang and Laursen (2008) to evaluate contact search strategies. Transient structural dynamics using a generalized-α time integration scheme

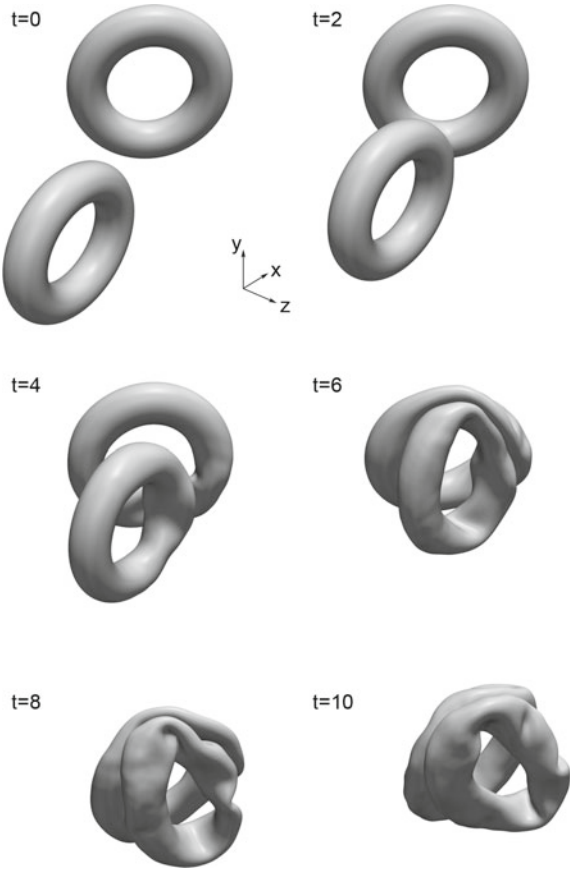

Fig. 10 Two torus impact – stages of deformation

are considered for the solution within 500 time steps and a constant time step size $\Delta t = 0.02$. As can be seen from the exemplary snapshots of deformation in Fig. 10, the lower torus is first accelerated towards the upper torus by a body force and then a very general oblique impact situation with large structural deformations occurs.

The finite element mesh for this 3D impact model involves 4,255,360 first-order hexahedral (*hex8*) elements and 13,994,880 degrees of freedom in total, with both slave and master surfaces consisting of 204,800 contact elements each. The numerical solution is performed in parallel on 120 processors within an overall simulation time of approximately 48 h.

Figures 11 and 12 further illustrate the complexity of the considered simulation model with severe changes of the active contact set and an extremely fine mesh resolution. While there always remains room for improvements of the parallel efficiency (e.g. with respect to efficient linear solvers, see Sect. 9), the results nevertheless strikingly emphasize that the implementation devised within this section is already very mature in this regard.

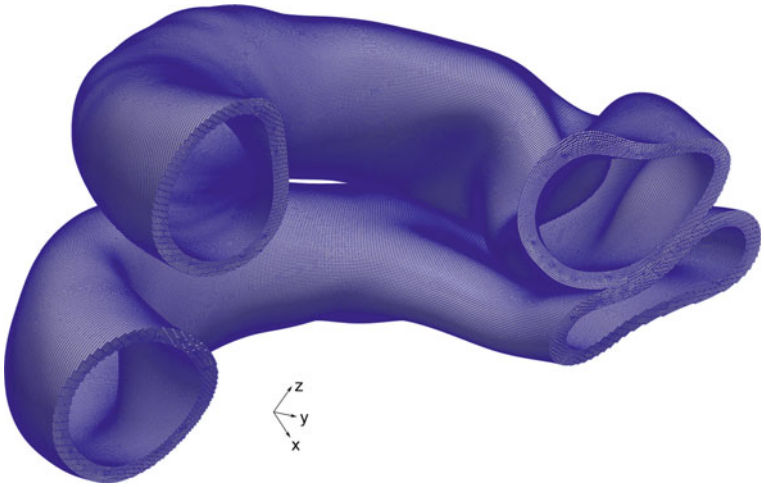

Fig. 11 Two torus impact – exemplary cut through the contact zone at time $t = 4$ and visualization of the finite element mesh

7 Algorithmic Aspects and Extensions

Going beyond the fundamental concepts of mortar finite element methods for mesh tying and unilateral contact (including friction), the following paragraphs shall give an overview of certain important algorithmic aspects that are of utmost importance for the acurate and efficient implementation of such mortar methods within a nonlinear finite element code framework. Specifically, the topics of suitable discrete Lagrange multiplier bases, parallel and high performance computing, numerical integration as well as isogeometric analysis will be highlighted. Further details on each of these topics can be found in the author's original contributions (Popp et al. 2012; Wohlmuth et al. 2012; Popp et al. 2013; Popp and Wall 2014; Farah et al. 2015; Seitz et al. 2016).

7.1 Discrete Lagrange Multipliers

The discrete Lagrange multiplier space $\boldsymbol{\mathcal{M}}_h$ and associated shape functions Φ_j, $j = 1, \ldots, m^{(1)}$, on the slave side of the mesh tying interface were already introduced in Sect. 5.3, although not specified in detail. Yet, this choice of the discrete Lagrange multiplier space is crucial for both the mathematical properties and the numerical efficiency of the resulting mortar approach. There exists a vast amount of literature discussing all relevant characteristics associated with the choice of $\boldsymbol{\mathcal{M}}_h$, such as inf-sup stability of the underlying mixed formulation and optimal a priori error bounds, see e.g. Bernardi et al. (1994), Ben Belgacem (1999), Seshaiyer and Suri (2000) and Wohlmuth (2000). With stability investigations and a priori error estimates not

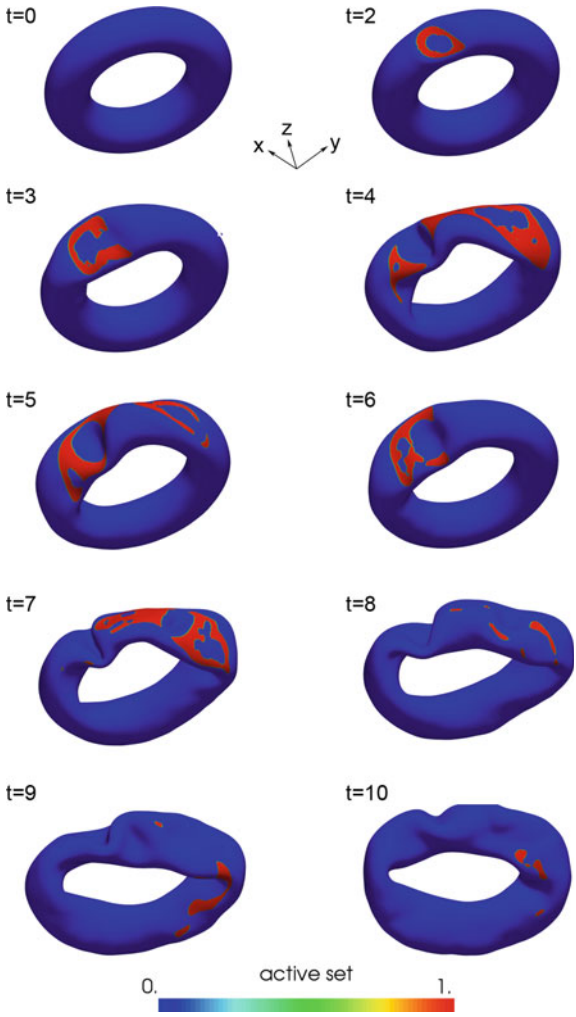

Fig. 12 Two torus impact – active contact set lower torus (1=active)

being in the focus of interest of this contribution, the following considerations rely on the fact that there exists a well-established framework of proofs and rigorous mathematical analyses, which guarantees the applicability of all discrete Lagrange multiplier spaces discussed here to mortar mesh tying problems. For a comprehensive overview, the reader is referred to Wohlmuth (2001) and the references therein.

Throughout this chapter, two different families of discrete Lagrange multipliers, namely *standard* and so-called *dual* Lagrange multipliers, will be distinguished. Standard Lagrange multipliers represent the classical approach for mortar methods (cf. Ben Belgacem 1999; Seshaiyer and Suri 2000) and are usually taken from the finite dimensional subset $\boldsymbol{\mathcal{W}}_h^{(1)} \subset \boldsymbol{\mathcal{W}}^{(1)}$ on the slave side of the interface, where $\boldsymbol{\mathcal{W}}^{(1)}$ is the trace space of $\boldsymbol{\mathcal{V}}^{(1)}$, as explained in Sect. 5.2. Thus, standard mortar

methods typically lead to identical shape functions for Lagrange multiplier and slave displacement interpolation, i.e. $\Phi_j = N_j^{(1)}$.

In contrast, the dual approach is motivated by the observation that the Lagrange multipliers physically represent fluxes (tractions) on the mesh tying interface in the continuous setting. This duality argument is then reflected by constructing dual Lagrange multiplier shape functions based on a so-called biorthogonality condition with the displacements in $\boldsymbol{\mathcal{W}}_h^{(1)}$, see e.g. Wohlmuth (2000). While they are, in general, not continuous and cannot be interpreted as a trace of conforming finite elements, the biorthogonality condition assures that the Lagrange multiplier shape functions Φ_j are again well-defined and satisfy all required approximation properties. One crucial advantage of the dual approach lies in the fact that it heavily facilitates the treatment of typical mortar coupling conditions at the interface, while at the same time preserving the mathematical optimality of the method. Going back to (85), the discrete mesh tying condition can alternatively be expressed as

$$\mathbf{d}_{\mathcal{S}} = \mathbf{D}^{-1}\mathbf{M}\mathbf{d}_{\mathcal{M}} := \mathbf{P}\mathbf{d}_{\mathcal{M}}\,, \tag{143}$$

where $\mathbf{P} = \mathbf{D}^{-1}\mathbf{M}$ represents the discrete interface coupling operator. As will be demonstrated later on for both mesh tying and unilateral contact problems, dual Lagrange multipliers avoid the necessity of solving a mass matrix type of system when evaluating (143), but *localize* the coupling conditions instead. Algebraically, this advantageous property of dual Lagrange multipliers can be observed by the mortar matrix $\mathbf{D}$ in (82) reducing to a diagonal matrix. This allows for very efficient condensation procedures of the discrete Lagrange multiplier degrees of freedom, which completely remove the undesirable saddle point structure of the underlying mesh tying and later unilateral contact systems, see Sects. 5.5 and 6.5.

While the construction of standard Lagrange multiplier bases is absolutely straightforward, the construction of dual Lagrange multiplier bases shall exemplarily be highlighted here for the simple first-order interpolation case in 2D. Details on how to define dual Lagrange multiplier shape functions Φ_j using the so-called biorthogonality relationship with the standard displacement shape functions $N_k^{(1)}$ have first been presented in Scott and Zhang (1990) and Wohlmuth (2000). A common notation of the biorthogonality condition is

$$\int_{\Gamma_{\mathrm{c},h}^{(1)}} \Phi_j N_k^{(1)}\,\mathrm{d}A_0 = \delta_{jk} \int_{\Gamma_{\mathrm{c},h}^{(1)}} N_k^{(1)}\,\mathrm{d}A_0, \qquad j,k = 1,\ldots,m^{(1)}. \tag{144}$$

Herein, δ_{jk} is the Kronecker delta, and the most common choice $m^{(1)} = n^{(1)}$ is assumed. For practical reasons, the biorthogonality condition is typically applied locally on each slave element e, yielding

$$\int_e \Phi_j N_k^{(1)}\,\mathrm{d}e = \delta_{jk} \int_e N_k^{(1)}\,\mathrm{d}e, \qquad j,k = 1,\ldots,m_e^{(1)}, \tag{145}$$

where $m_e^{(1)}$ represents the number of Lagrange multiplier nodes of the considered slave element. Taking into account the assumption that all nodes also carry discrete Lagrange multiplier degrees of freedom, $m_e^{(1)}$ is simply the number of nodes of the current slave facet. Comparing (144) and (82) also clearly reveals why dual shape functions reduce the mortar matrix $\mathbf{D}$ to a diagonal matrix. The dual shape functions resulting from (144), or rather from the elementwise version in (145), have the same polynomial order as the employed standard shape functions, i.e. $p_\lambda = p$. Moreover, it can easily be shown that the biorthogonality condition guarantees a partition of unity property, i.e. $\sum_j \Phi_j = 1, j = 1, \ldots, m_e^{(1)}$, see Flemisch and Wohlmuth (2007) for a proof.

As a simple example, the first-order finite element interpolation case in 2D shall be considered in the following. Obviously, this case leads to *line2* shaped mortar interface segments. With the Jacobian of *line2* segments being constant, the dual Lagrange multiplier shape functions determined by (145) are independent of element distortion, and can be defined a priori instead:

$$\Phi_1(\xi) = \frac{1}{2}(1 - 3\xi), \qquad \Phi_2(\xi) = \frac{1}{2}(1 + 3\xi). \tag{146}$$

Figure 13 illustrates these dual shape functions along with their standard counterparts, i.e. the first-order slave displacement shape functions $N_j^{(1)}$. In contrast to the corresponding standard Lagrange multiplier case, dual Lagrange multiplier shape functions can no longer be positive everywhere in order to fulfill the biorthogonality condition. However, integral positivity is still guaranteed. Moreover, the above defined Φ_j are indeed locally linear polynomials and satisfy a partition of unity property, but nonetheless they represent discontinuous functions.

In general, dual shape functions depend on the actual distortion of the individual underlying finite element, and cannot be defined a priori for non-constant slave element Jacobian determinants. In that regard, the first-order case in 2D illustrated above was a special case. Instead, a local linear mass matrix system of size $m_e^{(1)} \times m_e^{(1)}$ must be solved on each slave element. Details on these quite intricate constructions can for example be found in Wohlmuth (2001), Flemisch and Wohlmuth (2007), Lamichhane et al. (2005), Lamichhane and Wohlmuth (2007), Wohlmuth et al. (2012) and Popp et al. (2012).

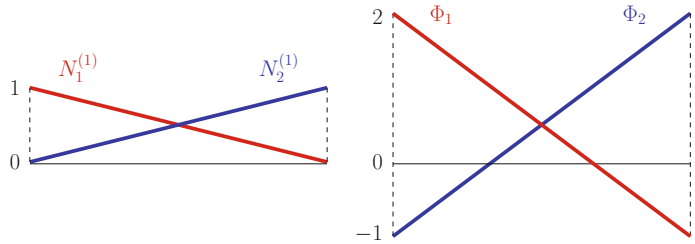

Fig. 13 Slave side displacement shape functions $N_j^{(1)}$ (left) and dual Lagrange multiplier shape functions Φ_j (right) for a *line2* element

7.2 Parallel Computing

The mortar-based mesh tying and contact algorithms developed throughout this contribution are designed for the use on large interconnected computer systems (clusters) with many central processing units (CPUs) and a distributed main memory. Being able to efficiently run large simulations in parallel requires strategies for the partitioning and parallel distribution of the problem data, i.e. finite element meshes (consisting of nodes and elements) as well as global vectors and matrices, into several independent processes, each assigned to a corresponding *processor*. For the sake of simplicity, the term processor refers to an independent processing unit throughout this chapter without implying any specific hardware configuration (such as a single-core or multi-core architecture). Within the finite element based multiphyiscs research code BACI that has been co-developed by the author at the Institute for Computational Mechanics of TUM, this so-called domain (or data) decomposition functionality is provided by the third-party library ParMETIS, see e.g. Karypis and Kumar (1998).

An example of such decompositions is visualized in Fig. 14 for a simple partitioning including only two processors, see also Gee (2004). It can be seen that each node in the mesh is uniquely assigned to one specific processor, and the same holds true for the elements. In addition, some nodes and elements at the transition between different processors must be stored redundantly within all adjacent processors. Therefore,

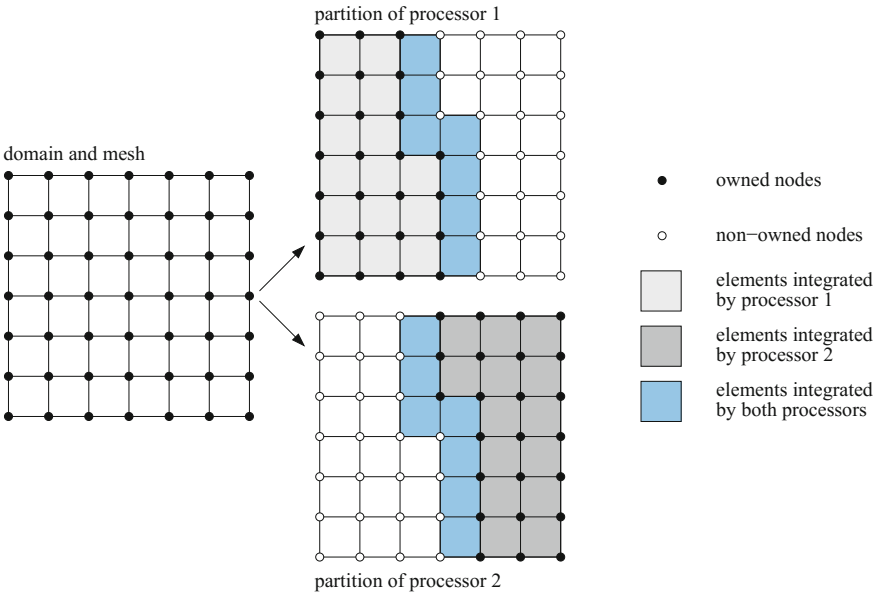

Fig. 14 An example of overlapping domain decomposition and parallel assembly involving two independent processors

this type of partitioning is commonly denoted as *overlapping* decomposition. For the methods developed in this chapter, it is sufficient to consider only the most straightforward case of minimal overlap between the individual partitions, i.e. an overlap of one layer of elements or nodes, respectively. Obviously, this concept of overlapping decomposition fits quite naturally to the typical tasks within a finite element program: first, each processor performs an elementwise integration of its own partition of the computational domain including the (relatively few) elements at the interprocessor boundaries. Then, the resulting quantities (e.g. local element load vectors and stiffness matrices) are assembled into the respective FE nodes of each processor. Thus, overlapping domain decomposition as described above provides a very elegant way of processing finite element integration and assembly, which is completely free of communication due to the distributed storage of the resulting global vector and matrix objects. While this rough introduction is by far not complete or rigorous from the viewpoint of parallel software design, it is sufficient for the following ideas on redistribution and load balancing to be comprehensible. For further details on the C++ based implementation of parallel (i.e. distributed) matrix and vector objects as well as the associated linear algebra, the interested reader is exemplarily referred to the documentation of open-source libraries of the Trilinos Project conducted by Sandia National Laboratories Heroux (2005).

Returning to the efficient parallel treatment of mortar methods and the derived mesh tying and contact algorithms, an exemplary mesh tying problem setup consisting of two cubic bodies as depicted in Fig. 15 is considered now. In total, the FE model contains 681,476 volume elements (with 2,136,177 displacement degrees of freedom) and 15,041 contact interface elements, which are distributed in parallel among several processors. As explained in the last paragraph, this partitioning generated via the ParMETIS library is in a sense optimal for the integration and assembly of the individual volume finite elements of the two bodies, i.e. the corresponding workload is equally distributed among all processors. For both tied and unilateral contact interaction, however, additional (but conceptually similar) tasks have to be performed locally at the interface: as will be explained in detail in Sect. 7.3, computing the interface contributions to the overall discrete problem formulation involves the mortar segmentation process, integration and assembly of the mortar matrices $\mathbf{D}$ and $\mathbf{M}$, to name only the most important tasks. Especially in three dimensions and for large interfaces, these computations may become quite time-consuming, so that they actually carry considerable weight as compared to the remaining time needed for FE evaluation and linear solvers. In contrast to NTS formulations, the high approximation quality of mortar methods comes at a price here. Unfortunately, the parallel distribution of the mortar interface itself is not optimal at all, which can easily be seen in Fig. 15. In this context, it is important to commemorate the slave-master concept typically used for implementing contact algorithms, where the interface-related workload is completely assigned to the slave side (or non-mortar side) whereas the master side (or mortar side) is passive. Thus, in the given example, the slave side of the interface (and thus the entire workload related to mesh tying) is associated with only 4 out of 16 processors.

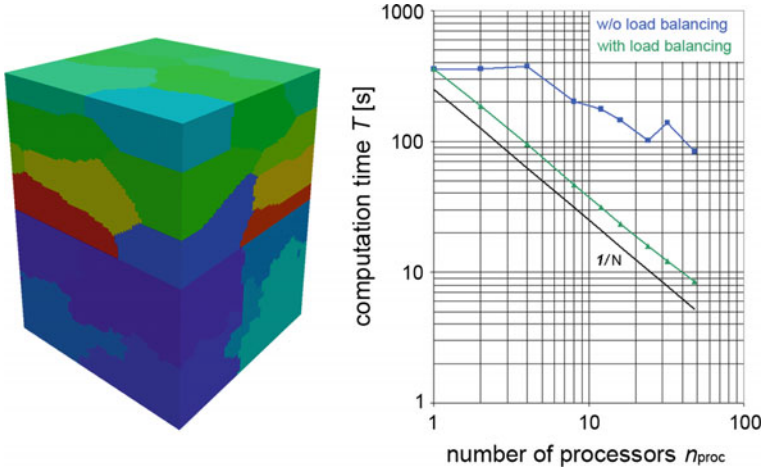

Fig. 15 Parallel redistribution and load balancing – initial partitioning for exemplary mesh tying problem setup using 32 processors (*left*) and strong scaling diagram (*right*)

The right hand side of Fig. 15 illustrates typical results for the parallel efficiency of the presented mortar algorithms in a so-called *strong scaling* diagram. Therein, the computation time for numerical integration and assembly of all interface-related quantities T is plotted against the total number of processors n_{proc} with logarithmic scales applied to both axes. Perfect scalability of the examined numerical algorithm is represented by a straight line with a negative slope of -1, thus representing the evident relation

$$T = \frac{c}{n_{\text{proc}}} \qquad \text{with } c > 0. \tag{147}$$

It can clearly be seen that no perfect scalability is achieved with the presented algorithms without load balancing (blue curve in Fig. 15). This is due to the non-optimal distribution of the slave surface among the participating processors as already described above. The results clearly motivate the development of an efficient parallel redistribution and load balancing strategy for mortar finite element methods. The approach proposed in the following is based on three steps, where the first one is of fundamental importance and is therefore needed for both mesh tying and contact applications. In contrast, the second and third step are purely contact-specific.

The rather simple basic idea of the first step is an *independent* parallel distribution of the finite elements in the domain and the mortar elements at the mesh tying or contact interface in order to achieve optimal parallel scalability of the computational tasks associated with both, i.e. integration and assembly in $\Omega^{(1)}$ and $\Omega^{(2)}$ as well as integration and assembly on $\gamma_c^{(1)}$ and $\gamma_c^{(2)}$. Again using ParMETIS, this redistribution of the interface elements can readily be performed during problem initialization at $t = 0$. Results for the test model introduced above are also visualized (green curve in Fig. 15), thus demonstrating that this simple modification already

allows for excellent parallel scalability within a wide range concerning the number of processors n_{proc}. However, dependent on the considered problem size, parallel redistribution only makes sense up to a certain n_{proc}. It is quite natural that such a limit exists, because there are of course some computational costs associated with the proposed redistribution procedure itself. If too many processors are used in relation to the problem size, these costs (mainly due to communication) become dominant and redistribution is no longer profitable beyond this point.

As already mentioned, this strategy can be further refined for unilateral contact applications. In contrast to mesh tying, contact interfaces are characterized by two additional complexities: the actual contact zone is not known a priori and it may constantly and significantly vary over time. Thus, in a second and third step, the proposed redistribution strategy is adapted such that it accommodates these additional complexities. Concretely, it can be seen from the Hertzian contact example in Fig. 16 that parallel redistribution must be limited to the actual contact area instead of the potential contact area, because the entire computational effort of numerical integration and assembly is connected with the former. Moreover, whenever finite deformations and large sliding motions occur, the described redistribution needs to be performed dynamically, i.e. over and over again. Such a dynamic load balancing

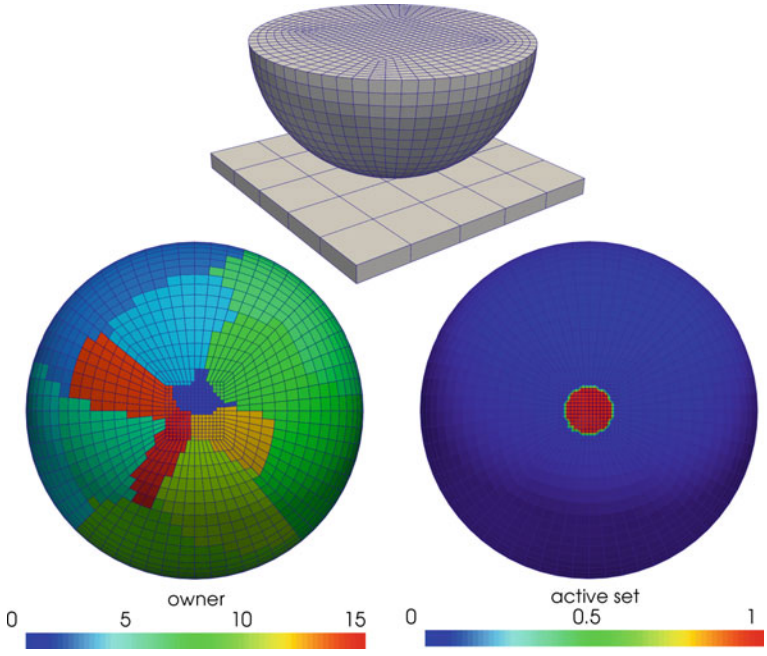

Fig. 16 Motivation for parallel redistribution exemplified with a Hertzian contact example – the active contact region (bottom right) is relatively small as compared with the potential contact surface (i.e. the whole hemisphere). Without redistribution only 6 out of 16 processors would carry the entire workload associated with contact evaluation (bottom left)

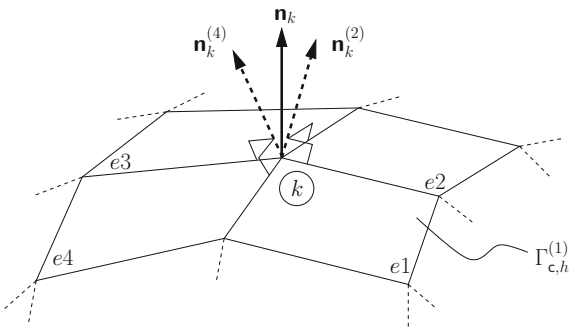

Fig. 17 Nodally averaged normal vector $\mathbf{n}_k$ at a slave node k with four adjacent slave facets $e1$ to $e4$. The element normal vectors $\mathbf{n}_k^{(e)}$ are exemplified for elements $e2$ and $e4$. Reprinted with permission from Popp et al. (2010), © 2010 John Wiley & Sons, Ltd.

strategy is then typically triggered by a suitable measure for the workload of each individual processor. The parallel balance of the workload among all processors is monitored and a simple criterion whether to apply dynamic load balancing within the current time step or not can be formulated as

$$IF\left(\frac{T^{\max}}{T^{\min}} > r\right) \rightsquigarrow \text{redistribute.} \tag{148}$$

Herein, the minimum and maximum computation times of one individual processor in the last time step are denoted as $T^{\min}$ and $T^{\max}$, respectively. The parameter $r > 1$ represents a user-defined tolerance. For example, choosing $r = 1.2$ implies that at most 20% unbalance of the parallel workload distribution are tolerated. Of course, the rather simple condition in (148) can easily be extended to incorporate more sophisticated criteria for dynamic load balancing. However, already the short overview given here shows that redistribution and load balancing provide an efficient tool for increased parallel efficiency of mortar algorithms for mesh tying and contact simulations. Corresponding numerical examples (see e.g. Section 6.6) demonstrate that the proposed approach is actually indispensable when considering large-scale applications.

7.3 Numerical Integration

A very efficient, yet at the same time highly accurate coupling algorithm, which performs integration not on the slave surface $\Gamma_{c,h}^{(1)}$ itself, but on its geometrical approximation with piecewise flat segments, has been proposed in Puso (2004) and will also be employed here. For further details and an in-depth mathematical analysis of this algorithm, the reader is also referred to Puso and Laursen (2004a, b) and Dickopf and Krause (2009). This scheme is referred to as *segment-based* integration scheme in the following.

In Fig. 18, the main steps of the 3D numerical integration algorithm for the mortar integrals in $\mathbf{D}$ and $\mathbf{M}$ are illustrated. In the following, the algorithm is outlined for

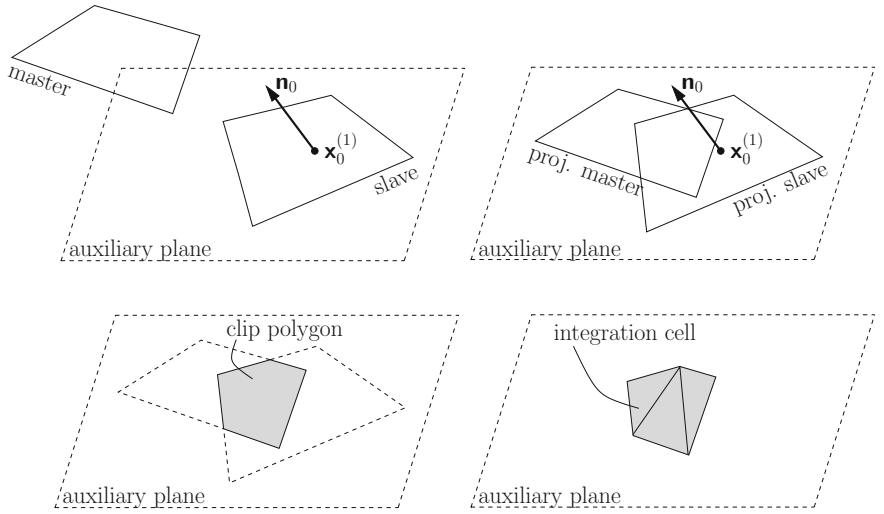

Fig. 18 Main steps of 3D mortar coupling of one slave and master element pair. Construct an auxiliary plane (top left), project slave and master nodes into the auxiliary plane (top right), perform polygon clipping (bottom left), divide clip polygon into triangular integration cells and perform Gauss integration (bottom right)

one pair of slave and master elements (s, m), which are close to each other and thus form an arbitrary overlap.

Algorithm 2

1. Construct an auxiliary plane for numerical integration based on the slave element center $\mathbf{x}_0^{(1)}$ and the corresponding unit normal vector $\mathbf{n}_0$.
2. Project all n_{s}^e slave element nodes $\mathbf{x}_k^{(1)},\ k = 1, \ldots, n_{\mathsf{s}}^e$ onto the auxiliary plane along $\mathbf{n}_0$ to obtain the projected slave nodes $\tilde{\mathbf{x}}_k^{(1)}$. Steps 1 and 2 can also be interpreted as a geometrical approximation of the slave surface removing element warping.
3. Project all n_{m}^e master element nodes $\mathbf{x}_l^{(2)},\ l = 1, \ldots, n_{\mathsf{m}}^e$ onto the auxiliary plane along $\mathbf{n}_0$ to obtain the projected master nodes $\tilde{\mathbf{x}}_l^{(2)}$.
4. Find the clip polygon of the projected slave and master elements in the auxiliary plane by applying a clipping algorithm, see e.g. Foley (1997).
5. Establish n_{cell} triangular integration cells by applying Delaunay triangulation to the clip polygon. Each integration cell consists of three vertices $\tilde{\mathbf{x}}_v^{\mathsf{cell}},\ v = 1, 2, 3$ and is interpolated by standard triangular shape functions on the well-known integration cell parameter space
   $\tilde{\boldsymbol{\eta}} = \left\{(\tilde{\xi}, \tilde{\eta}) | \tilde{\xi} \geq 0,\ \tilde{\eta} \geq 0,\ \tilde{\xi} + \tilde{\eta} \leq 1\right\}$.
6. Define n_{gp} Gauss integration points with coordinates $\tilde{\boldsymbol{\eta}}_g,\ g = 1, \ldots, n_{\mathsf{gp}}$ on each cell and project back along $\mathbf{n}_0$ to slave and master elements to obtain $\boldsymbol{\xi}^{(1)}(\tilde{\boldsymbol{\eta}}_g)$ and $\boldsymbol{\xi}^{(2)}(\tilde{\boldsymbol{\eta}}_g)$.

7. Perform Gauss integration of $D_{jk(\mathsf{s},\mathsf{m})}$ and $M_{jl(\mathsf{s},\mathsf{m})}$, $j,k = 1,\ldots,n_{\mathsf{s}}^e$ and $l = 1,\ldots,n_{\mathsf{m}}^e$ on all integration cells

$$D_{jk(\mathsf{s},\mathsf{m})} = \sum_{c=1}^{n_{\mathsf{cell}}} \left(\sum_{g=1}^{n_{\mathsf{gp}}} w_g \, \Phi_j^{(1)}(\boldsymbol{\xi}^{(1)}(\tilde{\boldsymbol{\eta}}_g)) \, N_k^{(1)}(\boldsymbol{\xi}^{(1)}(\tilde{\boldsymbol{\eta}}_g)) \, J_c \right), \tag{149}$$

$$M_{jl(\mathsf{s},\mathsf{m})} = \sum_{c=1}^{n_{\mathsf{cell}}} \left(\sum_{g=1}^{n_{\mathsf{gp}}} w_g \, \Phi_j^{(1)}(\boldsymbol{\xi}^{(1)}(\tilde{\boldsymbol{\eta}}_g)) \, N_l^{(2)}(\boldsymbol{\xi}^{(2)}(\tilde{\boldsymbol{\eta}}_g)) \, J_c \right). \tag{150}$$

where J_c, $c = 1,\ldots,n_{\mathsf{cell}}$ is the integration cell Jacobian determinant.

Expressions (149) and (150) represent contributions to D_{jk} and M_{jl} given by one slave and master element pair (s,m). Total quantities are obtained by summing up all slave and master element pair contributions. As pointed out in Puso (2004), the above algorithm relies on the fact that the clip polygons of all slave and master element pairs are convex. For further explanations on prerequisites and properties of this numerical integration procedure, the reader is referred to the original paper by Puso (2004).

In this work, seven point integration is used, which allows to exactly integrate polynomials of up to order five. This order of accuracy is sufficient to exactly integrate (149) and (150) for *tri3* surface facets and unwarped *quad4* surface facets. Typical constant stress patch tests on flat interfaces could even be satisfied with much fewer quadrature points. However, it should be pointed out that in the case of surface facet warping, the mapping between slave and master sides introduces rational polynomial functions into the integrands in (149) and (150), and thus the numerical quadrature rule can never reproduce the exact integral value in such cases. However, numerical results including mesh refinement studies on curved mesh tying interfaces demonstrate that the suggested choice of seven Gauss points per integration cell provides a sufficiently accurate quadrature rule. Figure 19 illustrates the generation of integration cells for 3D mortar coupling with a more complex example.

While Algorithm 2 undoubtedly provides the highest achievable accuracy for the numerical integration of D_{jk} and M_{jl} in 3D, some computationally more efficient alternatives have also been suggested in the literature. One prominent example is the simplified integration algorithm proposed in Fischer and Wriggers (2005, 2006) and later reused in De Lorenzis et al. (2011) and Tur et al. (2009), which will be referred to as *element-based* integration scheme in the following. Instead of thoroughly sub-dividing the mesh tying or contact interface into mortar segments, the numerical integration is simply performed element-wise in that approach, deliberately ignoring kinks of the functions to be integrated. Consequently, the devised integration schemes may indeed offer an appealing computational efficiency, but inevitably bring about difficulties with respect to accuracy of numerical integration. Even the exact satisfaction of a simple two-dimensional patch test, as investigated in Fischer and Wriggers (2005), is strongly influenced by the total number of Gauss

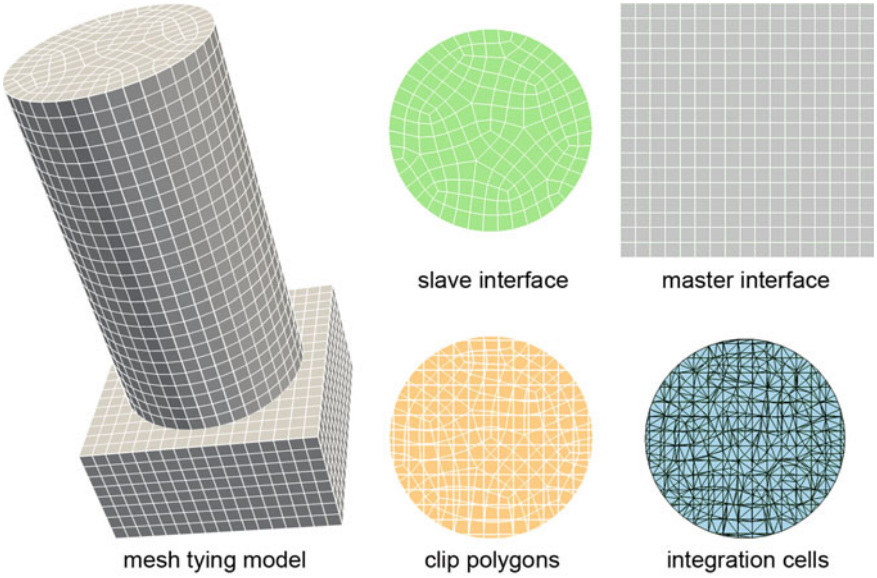

Fig. 19 Main steps of 3D mortar coupling for a representative mesh tying example

points chosen per slave element. An interesting improvement of this approach is suggested in Unger et al. (2007), where adaptive refinement of the integration cells is performed based on a hierarchical quadtree structure. Simply speaking, refinement is only performed close to the kinks of the integrands in (149) and (150) and thus the associated error of numerical integration can be reduced.

In contrast to the 2D case, an extension of the segmentation and integration algorithm to second-order interpolation needs some additional considerations for three-dimensional mortar mesh tying problems. As explained above, the presented method for first-order interpolation is based on the projection of flattened surface elements. This approach has been directly extended to quadratic finite elements in Puso et al. (2008), and is also employed here. The basic idea in Puso et al. (2008) is to subdivide quadratic surface elements into linearly interpolated segments as exemplarily illustrated in Fig. 20 for *quad9* facets. Numerical integration according to Algorithm 2 is then performed on the subsegments. As an example, consider the following mapping between parent element and subsegment space of subsegment *sub3* for the *quad9* element in Fig. 20, which is given by

$$\boldsymbol{\xi}^{\text{sub3}}(\boldsymbol{\xi}^{(1)}) = \begin{bmatrix} 2\xi^{(1)} - 1 \\ 2\eta^{(1)} - 1 \end{bmatrix}. \tag{151}$$

Similar mapping rules can also be readily established for *tri6* and *quad8* surface facets. It is important to point out that the approximation introduced by subdividing mortar elements only affects the integration domain itself, which no longer reflects the quadratic

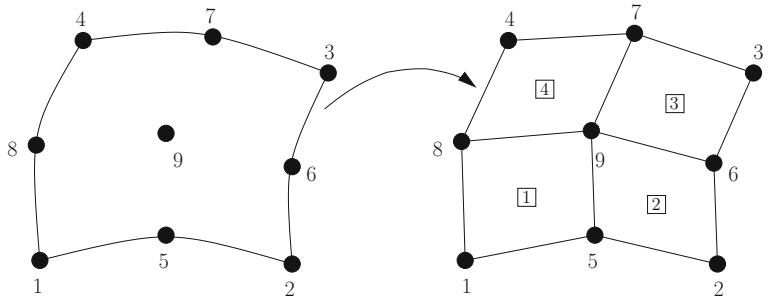

Fig. 20 Subdivision of interface elements with second-order interpolation. Exemplarily, a *quad9* element is split into four *quad4* subsegments *sub1–sub4*, to which the 3D mortar integration algorithm is then applied nearly unchanged. Reprinted with permission from Popp et al. (2010), © 2010 John Wiley & Sons, Ltd.

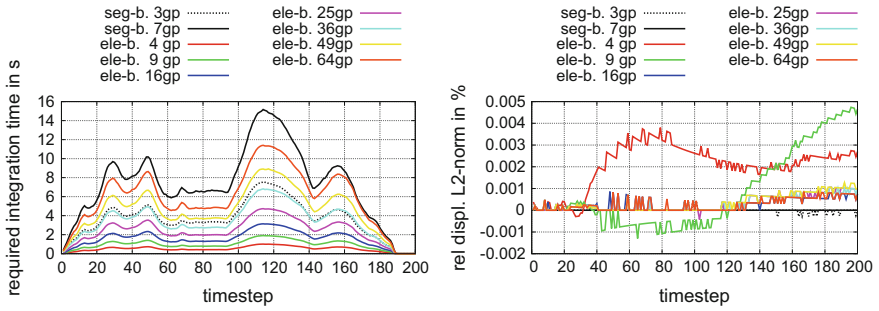

Fig. 21 Tori impact problem – averaged integration time per Newton step (left) and relative error of computed displacement field (right)

finite element surfaces correctly. Yet, by making use of the aforementioned geometric mappings from parent element space to subsegment space and vice versa, one is still able to properly evaluate the higher-order shape function products in (149) and (150).

Numerical integration using the *segment-based* scheme and the *element-based* scheme has been thoroughly compared with regard to accuracy and computational efficiency in Farah et al. (2015). To illustrate the main conclusions that can be drawn from such comparisons, the two tori impact example already introduced in the previous section is revisited here. Therefore, the average required integration times for one Newton step within each time step are plotted in the left subfigure of Fig. 21. In addition, the accuracy of the integration schemes is validated by the right subfigure of Fig. 21, which visualizes the deviations of the relative L^2-norm of the displacements with respect to a reference solution based on *segment-based* integration with 12 Gauss points per integration cell. Using 37 or 64 Gauss points per integration cell does not significantly change the displacement norm compared to 12 Gauss points.

For this example, the *segment-based* integration is tested with 3 and 7 Gauss points per integration cell, and the *element-based* integration method employs 4 to 64 Gauss

points per slave element. For the *segment-based* integration, 3 Gauss points per integration cell is the smallest sensible number of integration points. Thus, it can be seen that compared to the *segment-based* integration, the *element-based* integration method has the ability to significantly reduce the number of integration points. In addition, it is obvious that the required integration time scales linearly with the employed number of integration points, which is why all curves in Fig. 21 have a similar shape. The characteristic shape of the curves depends strongly on the active set. Thus, ups and downs of the curves occur due to time steps with a correspondingly high or low number of nodes being in contact. From time step 190 onwards, the curves are zero-valued due to the fact that the two tori are not in contact any more. Interestingly, the L^2-displacement errors are only marginal and decrease with more and more integration points. Even 4 Gauss points per element are sufficient for the L^2-displacement error being negligible. However, with 4 Gauss points per element, only 7% of integration time of the *segment-based* integration employing 7 integration points per integration cell are required. All in all, it becomes obvious that the *element-based* integration scheme allows for dramatic reductions of the computational costs for practical applications, while still maintaining a sufficient level of accuracy. Further details on this topic can be found in the author's original work in Farah et al. (2015).

7.4 Isogeometric Analysis (IGA)

Robust and accurate contact discretizations for nonlinear finite element analysis have been an active field of research in the past decade and a new class of formulations emerged with the introduction of isogeometric analysis (IGA) (Hughes et al. 2005). IGA is intended to bridge the gap between computer aided design (CAD) and finite element analysis (FEA) by using the smooth non-uniform rational B-splines (NURBS) or T-splines common in CAD also as a basis for the numerical analysis. The use of such smooth basis functions has some advantages over classical Lagrange polynomials for FEA such as a possibly higher accuracy per degree of freedom (Evans et al. 2009; Großmann et al. 2012) and, more importantly, a higher inter-element continuity. While finite elements based on Lagrange polynomials are limited to C^0 inter-element continuity independent of the polynomial order p, NURBS can be constructed with a maximum of C^{p-1} continuity. This high continuity results, amongst others, in a smooth surface representation which makes the application to computational contact mechanics particularly appealing, which has already been anticipated in the original proposition of IGA in Hughes et al. (2005).

As a consequence, in the past five years various discretization techniques have been developed for IGA or transferred from finite element based contact mechanics to IGA, such as node-to-segment (Matzen et al. 2013), Gauss-point-to-segment (Temizer et al. 2011; De Lorenzis et al. 2011; Dimitri et al. 2014; Dimitri 2015; Lu 2011; Sauer and De Lorenzis 2015) and mortar methods (Temizer et al. 2011; De Lorenzis et al. 2011; Temizer et al. 2012; De Lorenzis et al. 2012; Kim and Youn 2012; Dittmann et al. 2014). We refer to the recent review in De Lorenzis et al. (2014) for a comprehensive dis-

cussion of such methods, comparisons to their finite element counterparts and further references. In addition to the mentioned methods based on an isogeometric Galerkin approximation, the higher inter-element continuity of NURBS basis functions allows for the use of collocation methods, see Reali and Hughes (2015) for a general introduction and De Lorenzis et al. (2015), Kruse et al. (2015) for an application to computational contact mechanics. Besides the discretization technique, computational contact algorithms can be distinguished with respect to the underlying solution procedure. While Gauss-point-to-segment approaches are, due to their lack of inf-sup stability (see e.g. Temizer et al. 2011; Dimitri et al. 2014 for numerical investigations), usually combined with a penalty approach, see Dimitri et al. (2014), Dimitri (2015), Lu (2011), Sauer and De Lorenzis (2015), node-to-segment and mortar formulations can be combined with penalty methods (Temizer et al. 2011; De Lorenzis et al. 2011), Uzawa-type algorithms (Temizer et al. 2012), Lagrange multiplier methods (Kim and Youn 2012; Dittmann et al. 2014) or augmented Lagrange methods (De Lorenzis et al. 2012). In contrast to penalty methods, the other mentioned methods fulfill the contact constraints in a discrete sense exactly. In the context of domain decomposition in IGA, optimality and stability of standard mortar methods have only very recently been investigated in Hesch and Betsch (2012), Apostolatos et al. (2014), Dornisch et al. (2015), Brivadis et al. (2015), where also the construction of dual B-spline basis functions has been outlined theoretically.

In this section, the so-called dual mortar method is investigated mainly for contact mechanics using NURBS basis functions. In contrast to standard mortar methods, the use of dual basis functions for the Lagrange multiplier based on the mathematical concept of biorthogonality enables an easy elimination of the additional Lagrange multiplier degrees of freedom from the global system. This condensed system is smaller in size and no longer of saddle point type, but positive definite. A very simple and commonly used element-wise construction of the dual basis functions can directly be transferred to the IGA case. The resulting Lagrange multiplier interpolation satisfies discrete inf-sup stability and biorthogonality, however, the reproduction order is limited to one. In the domain decomposition case, this results in a limitation of the spatial convergence order to $\mathcal{O}(h^{3/2})$ in the energy norm, whereas for unilateral contact, due to the lower regularity of the solution, optimal convergence rates are still met.

Given some still to be defined basis functions Φ as a basis of $\boldsymbol{\mathcal{M}}_h$ and discrete vector-valued Lagrange multipliers $\boldsymbol{\lambda}_j$ at each control point on the potential contact surface, the Lagrange multiplier field on the slave side is approximated by

$$\boldsymbol{\lambda} \approx \boldsymbol{\lambda}_h = \sum_{a=1}^{n_{cp}} \Phi_a \boldsymbol{\lambda}_a. \tag{152}$$

While dual mortar methods are meanwhile well-established in finite elements, the present work, to the author's knowledge, is the first application of dual basis functions in the context of IGA for both domain decomposition and finite deformation frictional contact. Dual basis functions are characterized by fulfilling a biorthogonality condition (Wohlmuth 2000):

$$\int_{\gamma_{c,h}^{(1)}} \Phi_a R_b^{(1)} \, \mathrm{d}\gamma = \delta_{ab} \int_{\gamma_{c,h}^{(1)}} R_b^{(1)} \, \mathrm{d}\gamma \quad , \tag{153}$$

with the Kronecker symbol δ_{ab}. Different methods to construct such dual bases exist, and we want to follow the most simple one, where the dual basis functions have the same support as their primal counterparts, fulfill a partition of unity and are constructed via element-wise linear combinations of the primal shape functions (Flemisch and Wohlmuth 2007; Wohlmuth 2001; Lamichhane and Wohlmuth 2007; Lamichhane et al. 2005). On each element e one readily obtains

$$\Phi_j \big|_e = a_{jk}^e R_k^{(1)} \big|_e, \qquad \boldsymbol{A}_e = [a_{jk}^e] \in \Re^{n_{cpe} \times n_{cpe}} \quad , \tag{154}$$

with the coefficient matrix for each element

$$\begin{aligned} &\boldsymbol{A}_e = \boldsymbol{D}_e \, \boldsymbol{M}_e^{-1}, \\ &\boldsymbol{D}_e = [d_{jk}^e], \quad d_{jk}^e = \delta_{jk} \int_e R_k^{(1)} \, \mathrm{d}e, \\ &\boldsymbol{M}_e = [m_{jk}^e], \quad m_{jk}^e = \int_e R_j^{(1)} \, R_k^{(1)} \, \mathrm{d}e, \qquad j,k = 1, \ldots, n_{cpe}. \end{aligned} \tag{155}$$

In the construction of the coefficient matrix, the local integration for every slave element is only performed on that part of the element domain, for which a feasible projection to the master surface is possible. This is crucial for the consistent treatment of partially projecting elements in complex contact scenarios, as has been investigated for Lagrangian finite elements in Cichosz and Bischoff (2011) for two-dimensional mortar formulations and in Popp et al. (2013) for the general three-dimensional case. To properly detect the integration domain and reduce the integration error to a minimum, a segmentation process for isogeometric contact analysis will be described later on. For a well-defined construction of dual shape functions according to (155), the primal shape functions are required to have a non-zero integral value on the integration domain. Higher-order Lagrange polynomials do, in general, not meet this requirement which necessitates the use of a local basis transformation of the primal basis to obtain well defined dual shape functions, see Popp et al. (2012), Wohlmuth et al. (2012). NURBS, on the other hand, are positive on the entire element, such that the construction (154), (155) is well defined without any further modifications and for any approximation order. For a two-dimensional contact problem, i.e. a one dimensional contact boundary, an exemplary set of primal and dual basis functions of second-order is depicted in Fig. 22.

It should be pointed out that the dual basis functions generated by (154), (155) only guarantee a partition of unity. Consequently, the *global* approximation order is limited to one in the L^2-norm, independent of the *local* approximation. Since the dual NURBS do not posses the optimal reproduction order, optimal convergence rates as proven in Brivadis et al. (2015) cannot be guaranteed. For dual mortar methods based on Lagrange polynomials optimality can be recovered by a transformation of the primal basis (Lamichhane and Wohlmuth 2007) or by extending the support of the basis func-

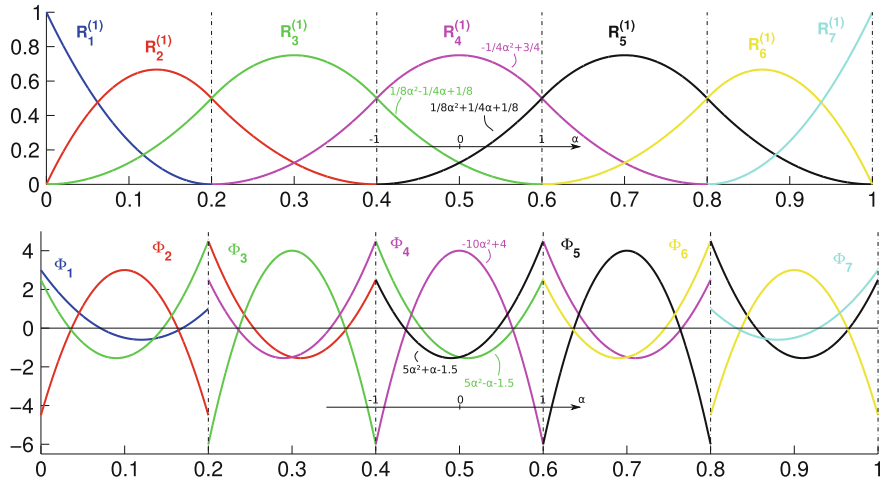

Fig. 22 Primal (top) and dual (bottom) basis functions for a one dimensional B-spline example. The equations for the shape functions in the central element are given to underline the desired biorthogonality and partition of unity. Reprinted with permission from Seitz et al. (2016), © 2016 Elsevier B.V.

tions (Oswald and Wohlmuth 2001). An extension of the latter approach to B-splines is outlined in Brivadis et al. (2015), but in general still unsolved. However, for contact problems the solution is typically in $H^t(\Omega^{(i)})^3$ with $t < 5/2$, such that a priori estimates are already limited by the regularity of the solution. Even this simple construction of dual shape functions meets the requirements in Wohlmuth et al. (2012) for optimal a priori estimates for the displacements in the H^1-norm of order $\mathcal{O}(h^{3/2})$.

Although the presented element-wise construction of dual shape functions yields sub-optimal convergence in domain decomposition applications, they may still be interesting for unilateral contact applications. In this case, the spatial convergence is usually limited by the reduced regularity of the solution, such that even the simple element-wise construction gives optimal convergence in finite element analysis (Wohlmuth et al. 2012). Hence, in our numerical example below, we want to investigate the spatial convergence properties of the isogeometric dual mortar contact algorithm in detail. We therefore use a two dimensional Hertzian-type contact of a cylindrical body (radius R) with a rigid planar surface under plane strain conditions. The two horizontal upper boundaries undergo a prescribed vertical displacement. To avoid singularities in the isogeometric mapping, we introduce a small inner radius (radius r), see Fig. 23 for the geometric setting, the material parameters and the coarsest mesh. Again meshes using second and third-order NURBS basis functions are used as depicted in Fig. 23 for the coarsest level, where different Bézier elements are marked with different shading. In this setup half of the elements on the potential contact surface are located within one ninth of the circumferential length and C^{p-1} continuity is ensured over the entire active contact surface. In the convergence study, uniform mesh refinement via knot insertion is performed on each of the patches resulting in a constant local ele-

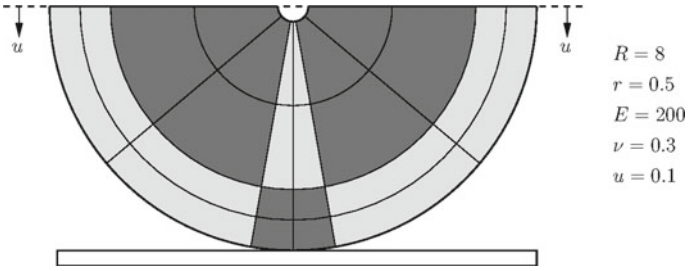

Fig. 23 Hertzian contact - Problem setup and coarsest mesh with Bézier elements in different shading. Reprinted with permission from Seitz et al. (2016), © 2016 Elsevier B.V.

ment aspect ratio. Although only relatively small deformations are to be expected, we use a fully nonlinear description of the continuum using nonlinear kinematics and a Saint–Venant–Kirchhoff material under plane strain assumption as well as the nonlinear contact formulation.

In Tables 1 and 2, we compare different refinement levels and study the convergence behavior in terms of the energy norm. Since no analytical solution is available, we use the finest mesh of level 7 with standard third-order NURBS as a numerical reference solution. Tables 1 and 2 give the error decay over six refinement levels for both a standard and dual Lagrange multiplier interpolation of second and third-order together with the numerical convergence order in each step. In the limit, all methods converge with the expected order of $\mathcal{O}(h^{3/2})$ in the energy norm and also the absolute error values are quantitatively very similar. Only the N^3 standard case gives a slightly higher order in the last step since the next level of this mesh is chosen as the numerical reference solution. In view of these results, the use of dual shape functions for the Lagrange multiplier instead of primal ones does not come at the expense of a reduced accuracy but yields equally accurate results while reducing the total system size to the number of displacement degrees of freedom only. In contrast to the domain decomposition case above, the convergence is now limited by the regularity of the solution, such that both standard and dual interpolations converge with the same order. The use of higher-order NURBS, i.e. third-order in Table 2 or even higher seems questionable from this viewpoint, since no faster convergence is gained from the higher-order interpolation with uniform mesh refinement.

8 Interface Modeling – Wear and Thermomechanics

Since contact can readily be interpreted as a special type of interface problem, it seems advisable not to isolate contact mechanics, but rather to address it in the context of a broader class of problems denoted as computational interface mechanics. Apart from the computational treatment of contact interaction and friction, computational interface mechanics also comprises other related physical phenomena such as wear, thermome-

Table 1 Hertzian contact - spatial convergence for second-order NURBS

Mesh	h	$\|\boldsymbol{u}_{\text{std}}^{N^2} - \boldsymbol{u}_{\text{ref}}\|_{\text{E}}$	$\mathcal{O}(h^x)$	$\|\boldsymbol{u}_{\text{dual}}^{N^2} - \boldsymbol{u}_{\text{ref}}\|_{\text{E}}$	$\mathcal{O}(h^x)$
1	1	2.5817e-1	–	3.1851e-1	–
2	0.5	1.4832e-1	0.80	8.8465e-2	1.85
3	0.25	4.7978e-2	1.63	4.0948e-2	1.11
4	0.125	1.5946e-2	1.59	1.5791e-2	1.37
5	0.0625	5.7112e-3	1.48	5.5474e-3	1.51
6	0.03125	1.9859e-3	1.52	1.9624e-3	1.50

Table 2 Hertzian contact - spatial convergence for third-order NURBS

mesh	h	$\|\boldsymbol{u}_{\text{std}}^{N^3} - \boldsymbol{u}_{\text{ref}}\|_{\text{E}}$	$\mathcal{O}(h^x)$	$\|\boldsymbol{u}_{\text{dual}}^{N^3} - \boldsymbol{u}_{\text{ref}}\|_{\text{E}}$	$\mathcal{O}(h^x)$
1	1	1.6407e-1	–	1.6171e-1	–
2	0.5	8.1487e-2	1.01	1.2201e-1	0.41
3	0.25	2.9319e-2	1.47	5.2129e-2	1.23
4	0.125	1.0849e-2	1.43	1.9212e-2	1.44
5	0.0625	3.9370e-3	1.46	6.8044e-3	1.50
6	0.03125	1.3038e-3	1.59	2.4698e-3	1.46

chanics and phase boundaries. Put in short terms, computational contact and interface mechanics are concerned with the treatment of complex interface effects at different length scales ranging from atomistic models to micro- and meso-scale models and further to classical continuum models at the macro-scale. The nature of many interface phenomena even requires a multi-scale perspective and associated models to bridge the spectrum of relevant length scales. Exemplarily, the following two sections shall highlight the application of the numerical methods discussed above (i.e. in particular mortar finite element methods) to wear modeling and thermo-mechanical interface problems. All details on the resulting schemes can be found in the author's original contributions (Farah et al. 2016, 2017; Seitz et al. 2018).

8.1 Wear Modeling

Contact mechanics including wear is one of the main causes for subsequent failure of machines and component damage and thus highly important for industrial applications. It is a process of material removal associated with frictional effects, which might result in finite shape changes due to the accumulation of wear. Wear is a very complex phenomenon, which relates a geometrical setting including external conditions with tribological material behavior in the contact zone, and therefore correct predictions of

wear effects are quite difficult to make, see Meng and Ludema (1995). The main wear types from the classifications in Popov (2010) and Rabinowicz (1995) are abrasive, adhesive corrosive and fretting wear. Nevertheless, there are many more wear types for different materials and load cases. The formulation predominantly employed for wear calculations is the phenomenological law by Archard (1953), which was firstly proposed by Holm (1946). It relates the worn volume with the normal contact force, a characteristic sliding length and a problem-specific wear parameter. Archards's law is also employed in this contribution as general wear description without discussing microscopical effects of special wear types.

In general, there are two different classes of wear treatment in computational contact mechanics: either only the consideration of very small amounts of wear or finite wear resulting in significant shape changes. The first class is usually treated by tailored modifications of the gap function, which results in slightly overlapping bodies (Farah et al. 2016; Rodríguez-Tembleque et al. 2012; Serre et al. 2001; Strömberg 1996). However, this contribution will focus on the second class, which treats finite wear effects. Standard remeshing procedures are employed in various contributions to prevent bulk elements from degeneration (McColl et al. 2004; Molinari et al. 2001; Öqvist 2001; Põdra and Andersson 1999; Paulin et al. 2008; Sfantos and Aliabadi 2006). An alternative approach to guarantee proper mesh quality is the Arbitrary–Lagrangean–Eulerian formulation, where the mesh movement is considered as pseudo-elasticity problem, see Stupkiewicz (2013). Most of the solution procedures for wear evolution are based on an explicit forward-Euler time integration scheme. Concretely, the standard contact problem is evaluated and only afterwards wear is calculated as a post-processing quantity for the last time step or even for a certain number of time steps. This incremental procedure is widely employed for the finite element method (Lengiewicz and Stupkiewicz 2012; McColl et al. 2004; Öqvist 2001; Põdra and Andersson 1999) and for the boundary element method (Rodríguez-Tembleque et al. 2012; Serre et al. 2001; Sfantos and Aliabadi 2006, 2007). Wear algorithms based on implicit time integration schemes are predominantly available for small amounts of wear and usually introduce additional unknowns into the linearized system of equations, see Ben Dhia and Torkhani (2011), Jourdan and Samida (2009), Strömberg (1996). Up to the authors' knowledge, the algorithm shown in Stupkiewicz (2013) is the only contribution in the context of finite element analysis that treats wear implicitly in a finite deformation and finite wear regime. Yet, it is limited to quasi-steady-state contact scenarios.

Restrictions to periodic cycling and prescribed relative movement of the involved bodies are often made in order to simplify the wear algorithm, see Argatov (2011), Argatov and Tato (2012), Lengiewicz and Stupkiewicz (2013), Páczelt et al. (2012) for reciprocal sliding and Páczelt and Mróz (2005, 2007); Stupkiewicz (2013) for general steady-state simulations. This assumption may be valid for classical tribological test configurations like pin-on-cylinder tests, but it is certainly not applicable to general scenarios.

The underlying contact frameworks for the wear algorithms existing in literature are mostly based on node-to-segment contact formulations, see for example Lengiewicz and Stupkiewicz (2012); Strömberg et al. (1996). Nowadays, the mortar method is undoubtedly the most preferred choice for robust finite element discretizations in com-

putational contact mechanics. Finite deformation mortar algorithms with and without frictional effects can exemplarily be found in Popp et al. (2010), Puso and Laursen (2004b), Puso et al. (2008), Yang et al. (2005). Still, the only wear algorithm based on mortar finite element discretization that can be found in the literature is given in Cavalieri and Cardona (2013), where only small wear effects without shape changes are considered.

The primary aim of this section, which summarizes the author's recent original work in Farah et al. (2017), is to simulate finite wear effects for arbitrary load paths in a fully implicit manner. To prevent element degeneration due to the loss of material, an Arbitrary–Lagrangean–Eulerian formulation with a nonlinear pseudo-elasticity assumption for the mesh motion is employed. The developed implicit partitioned algorithm is based on the configurationally consistent split between a Lagrangean step, where the finite deformation contact problem is solved and a shape evolution step, which realizes the finite configuration change due to wear. The wear equation based on Archard's law is enforced in a weak sense to follow the mortar idea and wear is already included in the Lagrangean step as an additional contribution to the gap function, which leads to an artificial penetration of the involved bodies. Within the shape evolution step, this non-physical overlap is then removed. Additional unknowns due to the Lagrange multiplier approach for contact and due to the wear discretization are eliminated by condensation procedures within the Lagrangean step to guarantee a non-increased system size. Within each time step, the Lagrangean step and the shape evolution step are repeated until convergence of the overall nonlinear coupled problem is obtained.

The numerical example shown here is adapted from Stupkiewicz (2013) to compare the presented implicit wear algorithm with a monolithic steady-state wear algorithm. Steady-state assumptions are valid for periodically repeated contact and frictional sliding problems with many cycles, such as pin-on-disc, reciprocating pin-on-flat, and pin-on-cylinder tribological tests. Usually, these problems are based on splitting the time scale into a fast time of the finite deformation problem and a slow time for the shape evolution due to wear, see Lengiewicz and Stupkiewicz (2012), Lengiewicz and Stupkiewicz (2013), Stupkiewicz (2013). However, within our wear framework, we define a state-independent fixed slip increment per integration point to simulate a steady-state sliding process. Concretely, the 2D pin-on-flat example consists of a hyper-elastic pin, which is pressed into an infinitely long rigid plane, see Fig. 24.

The pin is moved laterally with a constant velocity of $v = 1000 \frac{\text{mm}}{\text{s}}$. Consequently, the absolute value of the integration point slip increment is given as $||\boldsymbol{u}_{\tau,\text{rel}}|| = v \Delta t$. The simulation is performed within 5 pseudo-time steps with $\Delta t = 200\,\text{s}$. Frictionless sliding is assumed, which leads to a formulation of Archard's law in terms of the normal contact pressure. The wear coefficient is assumed constant in the material configuration and defined as $k_\text{w} = 10^{-7}\,\text{MPa}^{-1}$. The pin is loaded at its top edge with a normal force $F = 20 \frac{\text{N}}{\text{mm}}$ acting in negative y-direction. The strain energy function for the hyper-elastic material model is of neo-Hookean type and given as

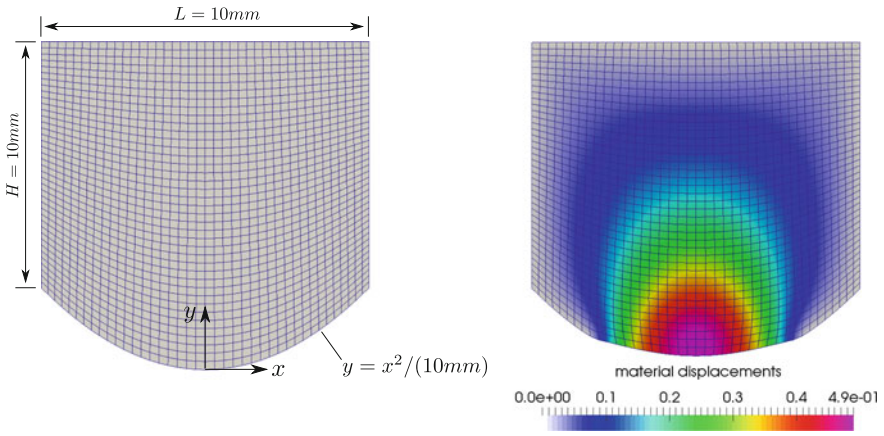

Fig. 24 2D pin-on-flat problem: reference configuration with dimensions (left) and material configuration with material displacements after 5 pseudo time steps (right). Reprinted with permission from Farah et al. (2017), © 2017 John Wiley & Sons, Ltd.

$$\Psi = \frac{\mu}{2}(I_C - 3) - \mu \log(\sqrt{III_C}) + \frac{\lambda}{2}\left((\sqrt{III_C} - 1)\right)^2. \tag{156}$$

Here, I_C and III_C are the invariants of the Cauchy–Green tensor. Furthermore, λ and μ represent the so-called Lamé parameters, which are correlated with the Young's modulus E and the Poisson's ratio ν via

$$\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)} \quad \text{and} \quad \mu = \frac{E}{2(1+\nu)}. \tag{157}$$

The Young's modulus is chosen as $E = 20$ MPa and the Poisson's ratio is $\nu = 0.3$. This 2D simulation is based on a plane-strain assumption and volumetric locking effects are avoided by the F-bar formulation for the employed 4-node quadrilateral elements, see de Souza Neto et al. (1996). The resulting material (i.e. worn) configuration is visualized in the Fig. 24. Here, the material displacements, which connect reference and material configuration, are illustrated. It can be clearly seen that not only nodes attached to the contact boundary are relocated but also inner nodes are properly adapted by our ALE approach. This guarantees a very good mesh quality in the worn configuration. In addition, the evolution of the contact boundary is shown in Fig. 25.

Here, we compare our results with the simulation from Stupkiewicz (2013). Our method matches the results from literature very well, which demonstrates that our wear algorithm can also be applied for such steady-state wear simulations. Further numerical examples as well as the entire background for numerical method development can be found in Farah et al. (2016, 2017).

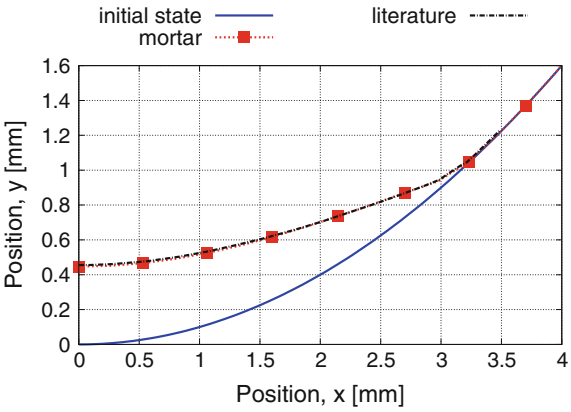

Fig. 25 Worn shape of the pin after 5 pseudo time steps with $\Delta t = 200$ s compared to results from Stupkiewicz (2013). Reprinted with permission from Farah et al. (2017), © 2017 John Wiley & Sons, Ltd.

8.2 Thermomechanics Modeling

In many engineering applications, frictional contact, thermomechanics and elasto-plastic material behavior come hand in hand. Just one class of typical well-known examples are metal forming and impact/crash analysis, where, at high strain rates, thermal effects need to be taken into account. The thermo-mechanical coupling appears in several forms: firstly and most obviously, there is heat conduction across the contact interface. Secondly, the dissipation of frictional work leads to an additional heating at the contact interface. Thirdly, also plastic work within the structure is transformed to heat. Vice versa, the current temperature may influence the elastic and especially the plastic material response. All this necessitates robust and efficient solution algorithms for fully coupled thermo-elasto-plastic contact problems, which has been an active research topic over the past 25 years.

Early implementations of thermo-elastic contact based on well-known node-to-segment (NTS) contact formulations in combination with a penalty constraint enforcement can be found in Johansson and Klarbring (1993), Oancea and Laursen (1997), Wriggers and Miehe (1994), Zavarise et al. (1992), Agelet De Saracibar (1998), Pantuso et al. (2000), Xing and Makinouchi (2002). Within the last decade, more sophisticated variationally consistent contact discretizations based on the mortar method have been developed and applied to thermo-mechanical contact in Hansen (2011), Khoei et al. (2015), Temizer (2014), Dittmann et al. (2014), Hüeber and Wohlmuth (2009). In addition, those algorithms satisfy the contact constraints exactly (at least in a weak sense) by using either Lagrange multipliers or an augmented Lagrangian functional instead of a simple penalty approach. Due to an easier implementation and other benefits like symmetric operators, most of the cited works above employ some sort of partitioned solution scheme for solving the structural problem (at constant temperature) and thermal problem (at constant displacement) sequentially. In thermo-plasticity, those partitioned schemes based on an isothermal split are only conditionally stable (Simo et al. 1992). Only a few researchers have employed monolithic solution schemes, which

solve for displacements and temperatures simultaneously (Zavarise et al. 1992; Pantuso et al. 2000; Dittmann et al. 2014; Hüeber and Wohlmuth 2009).

Numerical algorithms for finite deformation thermo-plasticity go back to the seminal work by Simo et al. (1992), which is based on the isothermal radial return mapping algorithm presented in Simo (1988). Both partitioned and monolithic solution approaches are discussed in Simo et al. (1992). Several extensions to this algorithm have been presented later, e.g. a monolithic formulation in principle axes (Ibrahimbegovic and Chorfi 2002) and a variant including temperature-dependent elastic material properties (Canajija and Brnić 2004). In a different line of work, a variational formulation of thermo-plasticity has been developed in Yang et al. (2006), where the rate of plastic work converted to heat follows from a variational principle instead of being a (constant) material parameter as in Simo et al. (1992). A comparison to experimental results is presented in Stainier and Ortiz (2010) to support this variational form. We point out that both approaches to determine the plastic dissipation, i.e. Simo et al. (1992) and Yang et al. (2006), are applicable within the algorithm for thermo-plasticity that is illustrated here. Besides the mentioned radial return mapping and variational formulations, a different numerical algorithm to isothermal plasticity at finite strains has been developed in Seitz et al. (2015). Based on fundamental ideas from Hager and Wohlmuth (2009), the plastic deformation at every quadrature point is introduced as an additional primary variable and the plastic inequality constraints are reformulated as nonlinear complementarity functions. This allows for a constraint violation during the nonlinear solution procedure, i.e. in the pre-asymptotic range of Newton's method, while ensuring their satisfaction at convergence. As usual in computational plasticity, the material constraints are enforced at each material point independently, such that the additional unknowns can be condensed directly at quadrature point level. It could be shown in Seitz et al. (2015) that due to this less restrictive formulation, a higher robustness can be achieved, which allows for larger time or load steps.

The author's recent original work in Seitz et al. (2018) aims at developing a monolithic solution scheme for the thermo-elasto-plastic frictional contact problem based on a new approach. Mortar finite element methods with dual Lagrange multipliers are applied for the contact treatment using nonlinear complementarity functions to deal with both the inequality constraints arising from frictional contact as well as plasticity in a unified manner. This bears novelty both for the numerical formulation of anisotropic thermo-plasticity within the bulk material as well as for the fully nonlinear thermo-mechanical contact formulation at the interface. Furthermore, full compatibility of the algorithms for thermo-plasticity and thermo-mechanical contact is demonstrated. Concerning plasticity, an extension of Seitz et al. (2015) to coupled thermo-plasticity within a monolithic solution framework is presented. Similar to the isothermal case, the use of Gauss-point-wise decoupled plastic deformation allows for a condensation of the additionally introduced plastic unknowns, where now also thermo-mechanical coupling terms have to be accounted for. The novel thermo-mechanical contact formulation represents a fully nonlinear extension of Hüeber and Wohlmuth (2009) including a consistent linearization with respect to both the displacement and temperature unknowns. Moreover, the use of dual Lagrange multipliers within a mortar contact formulation enables the trivial condensation of the discrete contact Lagrange multipliers

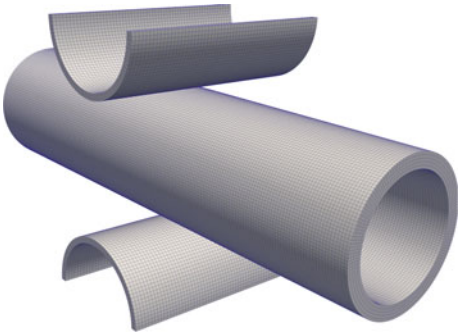

Fig. 26 Squeezed elasto-plastic tube – Initial configuration and mesh

such that the final linearized system to be solved consists of displacement and temperature degrees of freedom only. Our new thermo-mechanical contact formulation is applicable for both classical finite elements based on Lagrange polynomial basis functions as well as isogeometric analysis using NURBS basis functions, for which an appropriate dual basis has recently been proposed in Seitz et al. (2016). Owing to the variational basis of the mortar method, the thermo-mechanical contact patch test on non-matching discretizations is satisfied exactly and optimal convergence rates are achieved (Seitz et al. 2016) (Fig. 26).

While the reader is referred to Seitz et al. (2018) for all details of the formulation, we would at least like to present a fully coupled thermo-elasto-plastic contact example to demonstrate the robustness and efficiency of the developed algorithm. Similar to the example in Seitz et al. (2015) and originally inspired by Hager and Wohlmuth (2009), a squeezed metal tube with an inner and outer radius of 4 cm and 5 cm, respectively, and a length of 40 cm is analyzed. In the middle of the tube it is squeezed by two rigid cylindrical tools with an inner and outer of radius 4.5 cm and 5 cm, respectively, and a length of 16 cm. The material properties are the ones given in Seitz et al. (2018), with plastic isotropy, i.e. $y_{11} = y_0$. Between the tools and the tube, frictional contact with a temperature dependent friction coefficient is assumed with the initial coefficient of friction $\mu_0 = 0.25$, the reference temperature $T_0 = 293$ K and the damage temperature $T_d = 1793$ K. The tools are initially in stress free contact and perform a vertical displacement of $u(t) = (1 - \cos(\frac{t}{1\,\mathrm{s}}\pi)) \cdot 17.5$ cm over time. Figure 27 illustrates the plastic strain and temperature distribution at different times. Due to the symmetry of the problem, only one eighth of the entire model is discretized with about 20.000 elements, and the results are reflected for visualization purposes. First-order hexahedral elements with an F-bar technology are used to avoid volumetric locking, see de Souza Neto et al. (1996) for the original isothermal formulation of this element. In the early deformation stages, plastic deformation and therefore heat generation is mainly located directly beneath the contact zone (see Fig. 27), whereas later the main plastic deformation occurs at the side of the tube, where the highest peak temperatures are reached (see Fig. 27). After contact is released, thermal conduction tends to equilibrate the temperature inhomogeneity, see Fig. 27. To illustrate the efficient nonlinear solution procedure using Newton's method with a consistent linearization, Fig. 28 dis-

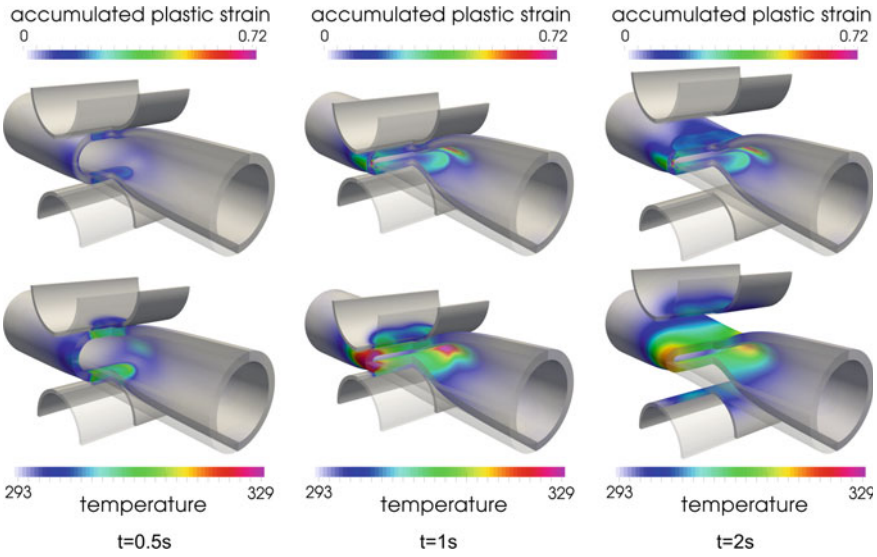

Fig. 27 Squeezed elasto-plastic tube – Deformed configurations at different times including accumulated plastic strain and temperature distribution (results of an eighth model reflected for vizualization)

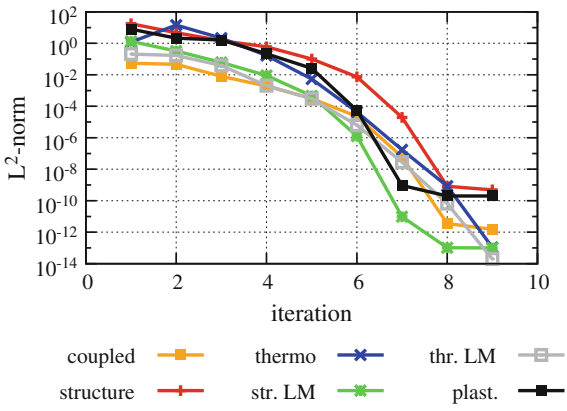

Fig. 28 Squeezed elasto-plastic tube – Convergence of different residuals in Newton's method for $t = 0.5$ s

plays the convergence behavior of different residual contributions in the time step of maximal tool velocity ($t = 0.5$ s). All residuals clearly exhibit a quadratic rate of convergence asymptotically, until they are at some point limited by machine precision.

9 Summary and Outlook

In this contribution, mortar finite element methods have been reviewed in the context of nonlinear solid mechanics, with a special emphasis on unilateral contact and friction as well as more complex interface problems. As a first step, some well-established basic principles of mortar methods have exemplarily been recapitulated for mesh tying (tied contact). The concepts of both standard and dual Lagrange multiplier interpolation were addressed with a focus on the latter. The most important favorable feature of dual Lagrange multiplier techniques is the resulting localization of the occurring interface constraints based on a biorthogonalization procedure. Algebraically, this is reflected in the possibility to easily condense the discrete Lagrange multiplier degrees of freedom (DOFs) associated with the non-matching mortar interfaces from the final linear systems of equations. Moreover, several important algorithmic aspects for an accurate and efficient implementation of mortar methods within a nonlinear finite element code framework have been discussed, including the construction of suitable discrete Lagrange multiplier bases, efficient parallel algorithms for high performance computing, accurate numerical integration procedures and an extension of the mortar approach to isogeometric analysis using NURBS.

In many engineering applications, however, an accurate treatment of non-penetration and Coulomb friction conditions at the contact interfaces is not sufficient to draw all technically relevant conclusions. Stress analysis and lifetime prediction of blade-to-disc joints in aircraft engines is an illustrative example for this statement. Such analyses require a detailed modeling and simulation of the manifold physical phenomena occurring at the contact interfaces. This possibly includes anisotropic friction, the dependency of friction coefficients on state variables (e.g. temperature), heat transmission, dissipation due to frictional sliding and surface degradation due to wear. As an outlook towards such challenging application scenarios in interface mechanics and real-life engineering, recent extensions of mortar finite element methods for wear modeling and thermomechanical contact modeling have been illustrated. For all mentioned applications, mortar methods provide an important algorithmic building block in order to obtain more accurate numerical solutions than possible to date, or even to gain insight into phenomena that have hardly been accessible for computational analysis until now.

References

C. Agelet De Saracibar. Numerical analysis of coupled thermomechanical frictional contact problems. Computational model and applications. *Archives of Computational Methods in Engineering*, 5(3):243–301, 1998.

P. Alart and A. Curnier. A mixed formulation for frictional contact problems prone to Newton like solution methods. *Computer Methods in Applied Mechanics and Engineering*, 92(3):353–375, 1991.

A. Apostolatos, R.t Schmidt, R. Wüchner, and K.-U. Bletzinger. A Nitsche-type formulation and comparison of the most common domain decomposition methods in isogeometric analysis. *International Journal for Numerical Methods in Engineering*, 97(7):473–504, 2014.

J. F. Archard. Contact and rubbing of flat surfaces. *Journal of Applied Physics*, 24(8):981–988, 1953.
I. Argatov. Asymptotic modeling of reciprocating sliding wear with application to local interwire contact. *Wear*, 271(78):1147–1155, 2011.
I. Argatov and W. Tato. Asymptotic modeling of reciprocating sliding wear comparison with finite-element simulations. *European Journal of Mechanics - A/Solids*, 34:1–11, 2012.
K.-J. Bathe. *Finite element procedures*. Prentice Hall, 1996.
K.-J. Bathe and A. Chaudhary. A solution method for planar and axisymmetric contact problems. *International Journal for Numerical Methods in Engineering*, 21(1):65–88, 1985.
T. Belytschko, W. K. Liu, and B. Moran. *Nonlinear finite elements for continua and structures*. Wiley, 2000.
F. Ben Belgacem. The mortar finite element method with Lagrange multipliers. *Numerische Mathematik*, 84(2):173–197, 1999.
F. Ben Belgacem, P. Hild, and P. Laborde. The mortar finite element method for contact problems. *Mathematical and Computer Modelling*, 28(4-8):263–271, 1998.
H. Ben Dhia and M. Torkhani. Modeling and computation of fretting wear of structures under sharp contact. *International Journal for Numerical Methods in Engineering*, 85(1):61–83, 2011.
D. J. Benson and J. O. Hallquist. A single surface contact algorithm for the post-buckling analysis of shell structures. *Computer Methods in Applied Mechanics and Engineering*, 78(2):141–163, 1990.
C. Bernardi, Y. Maday, and A. T. Patera. A new nonconforming approach to domain decomposition: the mortar element method. In H. Brezis and J.L. Lions, editors, *Nonlinear partial differential equations and their applications*, pages 13–51. Pitman/Wiley: London/New York, 1994.
J. Bonet and R. D. Wood. *Nonlinear continuum mechanics for finite element analysis*. Cambridge University Press, 1997.
E. Brivadis, A. Buffa, B. Wohlmuth, and L. Wunderlich. Isogeometric mortar methods. *Computer Methods in Applied Mechanics and Engineering*, 284:292–319, 2015.
S. Brunssen, F. Schmid, M. Schäfer, and B. I. Wohlmuth. A fast and robust iterative solver for nonlinear contact problems using a primal-dual active set strategy and algebraic multigrid. *International Journal for Numerical Methods in Engineering*, 69(3):524–543, 2007.
M. Canajija and J. Brnić. Associative coupled thermoplasticity at finite strain with temperature-dependent material parameters. *International Journal of Plasticity*, 20(10):1851–1874, 2004.
C. Carstensen, O. Scherf, and P. Wriggers. Adaptive finite elements for elastic bodies in contact. *SIAM Journal on Scientific Computing*, 20(5):1605–1626, 1999.
F. J. Cavalieri and A. Cardona. Three-dimensional numerical solution for wear prediction using a mortar contact algorithm. *International Journal for Numerical Methods in Engineering*, 96:467–486, 2013.
P. W. Christensen. A semi-smooth Newton method for elasto-plastic contact problems. *International Journal of Solids and Structures*, 39(8):2323–2341, 2002.
P. W. Christensen, A. Klarbring, J. S. Pang, and N. Strömberg. Formulation and comparison of algorithms for frictional contact problems. *International Journal for Numerical Methods in Engineering*, 42(1):145–173, 1998.
J. Chung and G. M. Hulbert. A time integration algorithm for structural dynamics with improved numerical dissipation: The generalized-α method. *Journal of Applied Mechanics*, 60:371–375, 1993.
T. Cichosz and M. Bischoff. Consistent treatment of boundaries with mortar contact formulations using dual Lagrange multipliers. *Computer Methods in Applied Mechanics and Engineering*, 200(9-12):1317–1332, 2011.
J. A. Cottrell, T. J. R. Hughes, and Y. Bazilevs. *Isogeometric analysis: toward integration of CAD and FEA*. Wiley, 2009.
L. De Lorenzis, I. Temizer, P. Wriggers, and G. Zavarise. A large deformation frictional contact formulation using NURBS-based isogeometric analysis. *International Journal for Numerical Methods in Engineering*, 87(13):1278–1300, 2011.

L De Lorenzis, P Wriggers, and G Zavarise. A mortar formulation for 3D large deformation contact using NURBS-based isogeometric analysis and the augmented Lagrangian method. *Computational Mechanics*, 49(1):1–20, 2012.

L. De Lorenzis, P. Wriggers, and T. J. R. Hughes. Isogeometric contact: a review. *GAMM-Mitteilungen*, 37(1):85–123, 2014.

L. De Lorenzis, J. A. Evans, T. J. R. Hughes, and A. Reali. Isogeometric collocation: Neumann boundary conditions and contact. *Computer Methods in Applied Mechanics and Engineering*, 284:21–54, 2015.

E. A. de Souza Neto, D. Perić, M. Dutko, and D. R. J. Owen. Design of simple low order finite elements for large strain analysis of nearly incompressible solids. *International Journal of Solids and Structures*, 33(20-22):3277–3296, 1996.

T. Dickopf and R. Krause. Efficient simulation of multi-body contact problems on complex geometries: A flexible decomposition approach using constrained minimization. *International Journal for Numerical Methods in Engineering*, 77(13):1834–1862, 2009.

R. Dimitri. Isogeometric treatment of large deformation contact and debonding problems with t-splines: a review. *Curved and Layered Structures*, 2(1), 2015.

R. Dimitri, L. De Lorenzis, M. A. Scott, P. Wriggers, R. L. Taylor, and G. Zavarise. Isogeometric large deformation frictionless contact using T-splines. *Computer Methods in Applied Mechanics and eEgineering*, 269:394–414, 2014.

M. Dittmann, M. Franke, I. Temizer, and C. Hesch. Isogeometric analysis and thermomechanical mortar contact problems. *Computer Methods in Applied Mechanics and Engineering*, 274:192–212, 2014.

W. Dornisch, G. Vitucci, and S. Klinkel. The weak substitution method–an application of the mortar method for patch coupling in NURBS-based isogeometric analysis. *International Journal for Numerical Methods in Engineering*, 103(3):205–234, 2015.

H. Elman, V. E. Howle, J. Shadid, R. Shuttleworth, and R. Tuminaro. A taxonomy and comparison of parallel block multi-level preconditioners for the incompressible Navier–Stokes equations. *Journal of Computational Physics*, 227(3):1790–1808, 2008.

J. A. Evans, Y. Bazilevs, I. Babuška, and T. J. R. Hughes. N-widths, sup–infs, and optimality ratios for the k-version of the isogeometric finite element method. *Computer Methods in Applied Mechanics and Engineering*, 198(21):1726–1741, 2009.

P. Farah, A. Popp, and W. A. Wall. Segment-based vs. element-based integration for mortar methods in computational contact mechanics. *Computational Mechanics*, 55:209–228, 2015.

P. Farah, M. Gitterle, W. A. Wall, and A. Popp. Computational wear and contact modeling for fretting analysis with isogeometric dual mortar methods. *Key Engineering Materials*, 681:1–18, 2016.

P. Farah, W. A. Wall, and A. Popp. An implicit finite wear contact formulation based on dual mortar methods. *International Journal for Numerical Methods in Engineering*, 111:325–353, 2017.

K. A. Fischer and P. Wriggers. Frictionless 2D contact formulations for finite deformations based on the mortar method. *Computational Mechanics*, 36(3):226–244, 2005.

K. A. Fischer and P. Wriggers. Mortar based frictional contact formulation for higher order interpolations using the moving friction cone. *Computer Methods in Applied Mechanics and Engineering*, 195(37-40):5020–5036, 2006.

B. Flemisch and B. I. Wohlmuth. Stable Lagrange multipliers for quadrilateral meshes of curved interfaces in 3D. *Computer Methods in Applied Mechanics and Engineering*, 196(8):1589–1602, 2007.

J. Foley. *Computer graphics: Principles and practice*. Addison-Wesley, 1997.

A. Francavilla and O. C. Zienkiewicz. A note on numerical computation of elastic contact problems. *International Journal for Numerical Methods in Engineering*, 9(4):913–924, 1975.

M. W. Gee. *Effiziente Lösungsstrategien in der nichtlinearen Schalenmechanik*. PhD thesis, Universität Stuttgart, 2004.

M. W. Gee, C. T. Kelley, and R. B. Lehoucq. Pseudo-transient continuation for nonlinear transient elasticity. *International Journal for Numerical Methods in Engineering*, 78(10):1209–1219, 2009.

M. Gitterle. *A dual mortar formulation for finite deformation frictional contact problems including wear and thermal coupling*. PhD thesis, Technische Universität München, 2012.
M. Gitterle, A. Popp, M. W. Gee, and W. A. Wall. Finite deformation frictional mortar contact using a semi-smooth Newton method with consistent linearization. *International Journal for Numerical Methods in Engineering*, 84(5):543–571, 2010.
D. Großmann, B. Jüttler, H. Schlusnus, J. Barner, and A.-V. Vuong. Isogeometric simulation of turbine blades for aircraft engines. *Computer Aided Geometric Design*, 29(7):519–531, 2012.
M. E. Gurtin. *An introduction to continuum mechanics*. Academic Press, 1981.
C. Hager. *Robust numerical algorithms for dynamic frictional contact problems with different time and space scales*. PhD thesis, Universität Stuttgart, 2010.
C. Hager and B. I. Wohlmuth. Nonlinear complementarity functions for plasticity problems with frictional contact. *Computer Methods in Applied Mechanics and Engineering*, 198(41-44):3411–3427, 2009.
C. Hager, S. Hüeber, and B. I. Wohlmuth. A stable energy-conserving approach for frictional contact problems based on quadrature formulas. *International Journal for Numerical Methods in Engineering*, 73(2):205–225, 2008.
J. O. Hallquist, G. L. Goudreau, and D. J. Benson. Sliding interfaces with contact-impact in large-scale Lagrangian computations. *Computer Methods in Applied Mechanics and Engineering*, 51(1-3):107–137, 1985.
G. Hansen. A jacobian-free Newton Krylov method for mortar-discretized thermomechanical contact problems. *Journal of Computational Physics*, 230(17):6546–6562, 2011.
S. Hartmann. *Kontaktanalyse dünnwandiger Strukturen bei großen Deformationen*. PhD thesis, Universität Stuttgart, 2007.
S. Hartmann, S. Brunssen, E. Ramm, and B. I. Wohlmuth. Unilateral non-linear dynamic contact of thin-walled structures using a primal-dual active set strategy. *International Journal for Numerical Methods in Engineering*, 70(8):883–912, 2007.
S. Hartmann, J. Oliver, R. Weyler, J.C. Cante, and J.A. Hernandez. A contact domain method for large deformation frictional contact problems. Part 2: Numerical aspects. *Computer Methods in Applied Mechanics and Engineering*, 198(33-36):2607–2631, 2009.
M. A. Heroux. An overview of the Trilinos project. *ACM Transactions on Mathematical Software*, 31(3):397–423, 2005.
H Hertz. Über die Berührung fester elastischer Körper. *Journal für die reine und angewandte Mathematik*, 92:156–171, 1882.
C. Hesch and P. Betsch. A mortar method for energy-momentum conserving schemes in frictionless dynamic contact problems. *International Journal for Numerical Methods in Engineering*, 77(10):1468–1500, 2009.
C. Hesch and P. Betsch. Transient three-dimensional domain decomposition problems: Frame-indifferent mortar constraints and conserving integration. *International Journal for Numerical Methods in Engineering*, 82(3):329–358, 2010.
C. Hesch and P. Betsch. Transient three-dimensional contact problems: mortar method. Mixed methods and conserving integration. *Computational Mechanics*, 48:461–475, 2011. ISSN 0178-7675.
C. Hesch and P. Betsch. Isogeometric analysis and domain decomposition methods. *Computer Methods in Applied Mechanics and Engineering*, 213:104–112, 2012.
P. Hild. Numerical implementation of two nonconforming finite element methods for unilateral contact. *Computer Methods in Applied Mechanics and Engineering*, 184(1):99 – 123, 2000.
M. Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set strategy as a semi-smooth Newton method. *SIAM Journal on Optimization*, 13(3):865–888, 2002.
R. Holm. *Electric contacts*. Gebers, 1946.
G. A. Holzapfel. *Nonlinear solid mechanics: A continuum approach for engineering*. Wiley, 2000.
S. Hüeber. *Discretization techniques and efficient algorithms for contact problems*. PhD thesis, Universität Stuttgart, 2008.

S. Hüeber and B. I. Wohlmuth. A primal-dual active set strategy for non-linear multibody contact problems. *Computer Methods in Applied Mechanics and Engineering*, 194(27-29):3147–3166, 2005.

S. Hüeber and B. I. Wohlmuth. Thermo-mechanical contact problems on non-matching meshes. *Computer Methods in Applied Mechanics and Engineering*, 198(15–16):1338–1350, 2009.

S. Hüeber and B. I. Wohlmuth. Equilibration techniques for solving contact problems with Coulomb friction. *Computer Methods in Applied Mechanics and Engineering*, 205–208:29–45, 2012.

S. Hüeber, G. Stadler, and B. I. Wohlmuth. A primal-dual active set algorithm for three-dimensional contact problems with Coulomb friction. *SIAM Journal on Scientific Computing*, 30(2):572–596, 2008.

T. J. R. Hughes. *The finite element method: linear static and dynamic finite element analysis*. Dover Publications, 2000.

T. J. R. Hughes, R. L. Taylor, J. L. Sackman, A. Curnier, and W. Kanoknukulchai. A finite element method for a class of contact-impact problems. *Computer Methods in Applied Mechanics and Engineering*, 8(3):249–276, 1976.

T. J. R. Hughes, J. A. Cottrell, and Y. Bazilevs. Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement. *Computer Methods in Applied Mechanics and Engineering*, 194(39):4135–4195, 2005.

A. Ibrahimbegovic and L. Chorfi. Covariant principal axis formulation of associated coupled thermoplasticity at finite strains and its numerical implementation. *International Journal of Solids and Structures*, 39(2):499–528, 2002.

B. M. Irons. Numerical integration applied to finite element methods. In *Proceedings Conference on the Use of Digital Computers in Structural Engineering*. University of Newcastle, 1966.

L. Johansson and A. Klarbring. Thermoelastic frictional contact problems: modelling, finite element approximation and numerical realization. *Computer Methods in Applied Mechanics and Engineering*, 105(2):181–210, 1993.

K. L. Johnson. *Contact mechanics*. Cambridge University Press, 1985.

F. Jourdan and A. Samida. An implicit numerical method for wear modeling applied to a hip joint prosthesis problem. *Computer Methods in Applied Mechanics and Engineering*, 198(27-29):2209 – 2217, 2009.

G. Karypis and V. Kumar. A parallel algorithm for multilevel graph partitioning and sparse matrix ordering. *Journal of Parallel and Distributed Computing*, 48(1):71–95, 1998.

A. R. Khoei, H. Saffar, and M. Eghbalian. An efficient thermo–mechanical contact algorithm for modeling contact–impact problems. *Asian Journal of Civil Engineering*, 16(5):681–708, 2015.

N. Kikuchi and J. T. Oden. *Contact problems in elasticity: A study of variational inequalities and finite element methods*. SIAM, Philadelphia, 1988.

J.-Y. Kim and S.-K. Youn. Isogeometric contact analysis using mortar method. *International Journal for Numerical Methods in Engineering*, 89(12):1559–1581, 2012.

T. M. Kindo, T. A. Laursen, and J. E. Dolbow. Toward robust and accurate contact solvers for large deformation applications: a remapping/adaptivity framework for mortar-based methods. *Computational Mechanics*, 54(1):53–70, 2014.

A. Konyukhov and K. Schweizerhof. On the solvability of closest point projection procedures in contact analysis: Analysis and solution strategy for surfaces of arbitrary geometry. *Computer Methods in Applied Mechanics and Engineering*, 197(33–40):3045–3056, 2008.

R. Kruse, N. Nguyen-Thanh, L. De Lorenzis, and T.J.R. Hughes. Isogeometric collocation for large deformation elasticity and frictional contact problems. *Computer Methods in Applied Mechanics and Engineering*, 296:73–112, 2015.

B. P. Lamichhane and B. I. Wohlmuth. Biorthogonal bases with local support and approximation properties. *Mathematics of Computation*, 76:233–249, 2007.

B. P. Lamichhane, R. P. Stevenson, and B. I. Wohlmuth. Higher order mortar finite element methods in 3D with dual Lagrange multiplier bases. *Numerische Mathematik*, 102(1):93–121, 2005.

T. A. Laursen. *Formulation and treatment of frictional contact problems using finite elements*. PhD thesis, Stanford University, 1992.

T. A. Laursen. *Computational contact and impact mechanics*. Springer-Verlag Berlin Heidelberg, 2002.

T. A. Laursen and V. Chawla. Design of energy conserving algorithms for frictionless dynamic contact problems. *International Journal for Numerical Methods in Engineering*, 40(5):863–886, 1997.

T. A. Laursen and G. R. Love. Improved implicit integrators for transient impact problems - geometric admissibility within the conserving framework. *International Journal for Numerical Methods in Engineering*, 53(2):245–274, 2002.

T. A. Laursen and J. C. Simo. A continuum-based finite element formulation for the implicit solution of multibody, large deformation-frictional contact problems. *International Journal for Numerical Methods in Engineering*, 36(20):3451–3485, 1993.

T. A. Laursen, M. A. Puso, and J. Sanders. Mortar contact formulations for deformable-deformable contact: past contributions and new extensions for enriched and embedded interface formulations. *Computer Methods in Applied Mechanics and Engineering*, 205–208:3–15, 2012.

J. Lengiewicz and S. Stupkiewicz. Continuum framework for finite element modelling of finite wear. *Computer Methods in Applied Mechanics and Engineering*, 205–208:178–188, 2012.

J. Lengiewicz and S. Stupkiewicz. Efficient model of evolution of wear in quasi-steady-state sliding contacts. *Wear*, 303(12):611 – 621, 2013.

J. Lu. Isogeometric contact analysis: Geometric basis and formulation for frictionless contact. *Computer Methods in Applied Mechanics and Engineering*, 200(5):726–741, 2011.

J. E. Marsden and T. J. R. Hughes. *Mathematical foundations of elasticity*. Dover, 1994.

M. E. Matzen, T. Cichosz, and M. Bischoff. A point to segment contact formulation for isogeometric, NURBS based finite elements. *Computer Methods in Applied Mechanics and Engineering*, 255:27–39, 2013.

I. R. McColl, J. Ding, and S. B. Leen. Finite element simulation and experimental validation of fretting wear. *Wear*, 256(11-12):1114 – 1127, 2004.

T. W. McDevitt and T. A. Laursen. A mortar-finite element formulation for frictional contact problems. *International Journal for Numerical Methods in Engineering*, 48(10):1525–1547, 2000.

H. C. Meng and K. C. Ludema. Wear models and predictive equations: their form and content. *Wear*, 181–183:443–457, 1995.

J. F. Molinari, M. Ortiz, R. Radovitzky, and E. A. Repetto. Finite element modeling of dry sliding wear in metals. *Engineering Computations*, 18(3/4):592–610, 2001.

V. G. Oancea and T. A. Laursen. A finite element formulation of thermomechanical rate-dependent frictional sliding. *International Journal for Numerical Methods in Engineering*, 40(23):4275–4311, 1997.

R. W. Ogden. *Non-linear elastic deformations*. Dover Publications, 1997.

J. Oliver, S. Hartmann, J. C. Cante, R. Weyler, and J. A. Hernandez. A contact domain method for large deformation frictional contact problems. Part 1: Theoretical basis. *Computer Methods in Applied Mechanics and Engineering*, 198(33-36):2591–2606, 2009.

M. Öqvist. Numerical simulations of mild wear using updated geometry with different step size approaches. *Wear*, 249(12):6–11, 2001.

P. Oswald and B. Wohlmuth. On polynomial reproduction of dual FE bases. In *Thirteenth International Conference on Domain Decomposition Methods*, pages 85–96, 2001.

P. Põdra and S. Andersson. Simulating sliding wear with finite element method. *Tribology International*, 32(2):71–81, 1999.

I. Páczelt and Z. Mróz. On optimal contact shapes generated by wear. *International Journal for Numerical Methods in Engineering*, 63(9):1250–1287, 2005.

I. Páczelt and Z. Mróz. Optimal shapes of contact interfaces due to sliding wear in the steady relative motion. *International Journal of Solids and Structures*, 44(34):895–925, 2007.

I. Páczelt, S. Kucharski, and Z. Mróz. The experimental and numerical analysis of quasi-steady wear processes for a sliding spherical indenter. *Wear*, 274-275:127–148, 2012.

D. Pantuso, K.-J. Bathe, and P. A. Bouzinov. A finite element procedure for the analysis of thermo-mechanical solids in contact. *Computers & Structures*, 75(6):551–573, 2000.

P. Papadopoulos and R. L. Taylor. A mixed formulation for the finite element solution of contact problems. *Computer Methods in Applied Mechanics and Engineering*, 94(3):373–389, 1992.
C. Paulin, S. Fouvry, and C. Meunier. Finite element modelling of fretting wear surface evolution: Application to a Ti-6A1-4V contact. *Wear*, 264(12):26–36, 2008.
V. Popov. *Contact Mechanics and Friction*. Springer, 2010.
A. Popp. *Mortar methods for computational contact mechanics and general interface problems*. PhD thesis, Technische Universität München, 2012.
A. Popp and W. A. Wall. Dual mortar methods for computational contact mechanics – overview and recent developments. *GAMM-Mitteilungen*, 37(1):66–84, 2014.
A. Popp, M. W. Gee, and W. A. Wall. A finite deformation mortar contact formulation using a primal-dual active set strategy. *International Journal for Numerical Methods in Engineering*, 79(11):1354–1391, 2009.
A. Popp, M. Gitterle, M. W. Gee, and W. A. Wall. A dual mortar approach for 3D finite deformation contact with consistent linearization. *International Journal for Numerical Methods in Engineering*, 83(11):1428–1465, 2010.
A. Popp, B. I. Wohlmuth, M. W. Gee, and W. A. Wall. Dual quadratic mortar finite element methods for 3D finite deformation contact. *SIAM Journal on Scientific Computing*, 34:B421–B446, 2012.
A. Popp, A. Seitz, M. W. Gee, and W. A. Wall. A dual mortar approach for improved robustness and consistency of 3D contact algorithms. *Computer Methods in Applied Mechanics and Engineering*, 264:67–80, 2013.
M. A. Puso. A 3D mortar method for solid mechanics. *International Journal for Numerical Methods in Engineering*, 59(3):315–336, 2004.
M. A. Puso and T. A. Laursen. A 3D contact smoothing method using Gregory patches. *International Journal for Numerical Methods in Engineering*, 54(8):1161–1194, 2002.
M. A. Puso and T. A. Laursen. A mortar segment-to-segment contact method for large deformation solid mechanics. *Computer Methods in Applied Mechanics and Engineering*, 193(6-8):601–629, 2004a.
M. A. Puso and T. A. Laursen. A mortar segment-to-segment frictional contact method for large deformations. *Computer Methods in Applied Mechanics and Engineering*, 193(45-47):4891–4913, 2004b.
M. A. Puso, T. A. Laursen, and J. Solberg. A segment-to-segment mortar contact method for quadratic elements and large deformations. *Computer Methods in Applied Mechanics and Engineering*, 197(6-8):555–566, 2008.
L. Qi and J. Sun. A nonsmooth version of Newton's method. *Mathematical Programming*, 58(1):353–367, 1993.
E. Rabinowicz. *Friction and wear of materials*. Wiley, 1995.
A. Reali and T. J. R. Hughes. An introduction to isogeometric collocation methods. In *Isogeometric Methods for Numerical Simulation*, pages 173–204. Springer, 2015.
J. N. Reddy. *An introduction to nonlinear finite element analysis*. Oxford University Press, 2004.
L. Rodríguez-Tembleque, R. Abascal, and M. H. Aliabadi. Anisotropic wear framework for 3d contact and rolling problems. *Computer Methods in Applied Mechanics and Engineering*, 241-244:1–19, 2012.
R. A. Sauer. Enriched contact finite elements for stable peeling computations. *International Journal for Numerical Methods in Engineering*, 87(6):593–616, 2011.
R. A Sauer and L. De Lorenzis. An unbiased computational contact formulation for 3D friction. *International Journal for Numerical Methods in Engineering*, 101(4):251–280, 2015.
K. Schweizerhof and A. Konyukhov. Covariant description for frictional contact problems. *Computational Mechanics*, 35(3):190–213, 2005.
L. R. Scott and S. Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Mathematics of Computation*, 54:483–492, 1990.
A. Seitz, A. Popp, and W. A Wall. A semi-smooth newton method for orthotropic plasticity and frictional contact at finite strains. *Computer Methods in Applied Mechanics and Engineering*, 285:228–254, 2015.

A. Seitz, P. Farah, J. Kremheller, B. I Wohlmuth, W. A Wall, and A. Popp. Isogeometric dual mortar methods for computational contact mechanics. *Computer Methods in Applied Mechanics and Engineering*, 301:259–280, 2016.

A. Seitz, W. A. Wall, and A. Popp. A computational approach for thermo-elasto-plastic frictional contact based on a monolithic formulation using non-smooth nonlinear complementarity functions. *Advanced Modeling and Simulation in Engineering Sciences,* 5:5, 2018.

I. Serre, M. Bonnet, and R. M. Pradeilles-Duval. Modelling an abrasive wear experiment by the boundary element method. *Comptes Rendus de l'Académie des Sciences-Series IIB-Mechanics*, 329(11):803–808, 2001.

P. Seshaiyer and M. Suri. hp submeshing via non-conforming finite element methods. *Computer Methods in Applied Mechanics and Engineering*, 189(3):1011–1030, 2000.

G. K. Sfantos and M. H. Aliabadi. Wear simulation using an incremental sliding boundary element method. *Wear*, 260(9-10):1119 – 1128, 2006.

G. K. Sfantos and M. H. Aliabadi. A boundary element formulation for three-dimensional sliding wear simulation. *Wear*, 262(5-6):672 – 683, 2007.

J. C. Simo. A framework for finite strain elastoplasticity based on maximum plastic dissipation and the multiplicative decomposition: Part I. Continuum formulation. *Computer Methods in Applied Mechanics and Engineering*, 66(2):199–219, 1988.

J. C. Simo and T. J. R. Hughes. *Computational inelasticity*. Springer, 1998.

J. C. Simo and T. A. Laursen. An augmented Lagrangian treatment of contact problems involving friction. *Computers & Structures*, 42(1):97–116, 1992.

J. C. Simo, P. Wriggers, and R. L. Taylor. A perturbed Lagrangian formulation for the finite element solution of contact problems. *Computer Methods in Applied Mechanics and Engineering*, 50(2):163–180, 1985.

J. C. Simo, N. Tarnow, and K. K. Wong. Exact energy-momentum conserving algorithms and symplectic schemes for nonlinear dynamics. *Computer Methods in Applied Mechanics and Engineering*, 100(1):63–116, 1992.

L. Stainier and M. Ortiz. Study and validation of a variational theory of thermo-mechanical coupling in finite visco-plasticity. *International Journal of Solids and Structures*, 47(5):705–715, 2010.

N. Strömberg. An augmented lagrangian method for fretting problems. *European Journal of Mechanics – A/Solids*, 16:573–593, 1996.

N. Strömberg, L. Johansson, and A. Klarbring. Derivation and analysis of a generalized standard model for contact, friction and wear. *International Journal of Solids and Structures*, 33(13):1817–1836, 1996.

S. Stupkiewicz. An ALE formulation for implicit time integration of quasi-steady-state wear problems. *Computer Methods in Applied Mechanics and Engineering*, 260:130–142, 2013.

R. L. Taylor, J. C. Simo, O. C. Zienkiewicz, and A. C. H. Chan. The patch test: A condition for assessing FEM convergence. *International Journal for Numerical Methods in Engineering*, 22(1):39–62, 1986.

I. Temizer. Multiscale thermomechanical contact: computational homogenization with isogeometric analysis. *International Journal for Numerical Methods in Engineering*, 97(8):582–607, 2014.

I. Temizer, P. Wriggers, and T. J. R. Hughes. Contact treatment in isogeometric analysis with NURBS. *Computer Methods in Applied Mechanics and Engineering*, 200(9–12):1100–1112, 2011.

I. Temizer, P. Wriggers, and T. J. R. Hughes. Three-dimensional mortar-based frictional contact treatment in isogeometric analysis with NURBS. *Computer Methods in Applied Mechanics and Engineering*, 209–212:115–128, 2012.

S. P. Timoshenko and J. N. Goodier. *Theory of elasticity*. McGraw-Hill, 1970.

M. Tur, F. J. Fuenmayor, and P. Wriggers. A mortar-based frictional contact formulation for large deformations using Lagrange multipliers. *Computer Methods in Applied Mechanics and Engineering*, 198(37-40):2860–2873, 2009.

M. Tur, E. Giner, F.J. Fuenmayor, and P. Wriggers. 2d contact smooth formulation based on the mortar method. *Computer Methods in Applied Mechanics and Engineering*, 247-248:1–14, 2012.

R. Unger, M. C. Haupt, and P. Horst. Application of Lagrange multipliers for coupled problems in fluid and structural interactions. *Computers & Structures*, 85(11-14):796–809, 2007.

J. R. Williams and R. O'Connor. Discrete element simulation and the contact problem. *Archives of Computational Methods in Engineering*, 6(4):279–304, 1999.

B. I. Wohlmuth. A mortar finite element method using dual spaces for the Lagrange multiplier. *SIAM Journal on Numerical Analysis*, 38(3):989–1012, 2000.

B. I. Wohlmuth. *Discretization methods and iterative solvers based on domain decomposition*. Springer-Verlag Berlin Heidelberg, 2001.

B. I. Wohlmuth. Variationally consistent discretization schemes and numerical algorithms for contact problems. *Acta Numerica*, 20:569–734, 2011.

B. I. Wohlmuth, A. Popp, M. W. Gee, and W. A. Wall. An abstract framework for a priori estimates for contact problems in 3D with quadratic finite elements. *Computational Mechanics*, 49:735–747, 2012.

P. Wriggers. *Computational contact mechanics*. Springer-Verlag Berlin Heidelberg, 2006.

P. Wriggers and C. Miehe. Contact constraints within coupled thermomechanical analysis - a finite element model. *Computer Methods in Applied Mechanics and Engineering*, 113(3):301–319, 1994.

P. Wriggers and O. Scherf. An adaptive finite element algorithm for contact problems in plasticity. *Computational Mechanics*, 17(1):88–97, 1995.

P. Wriggers and G. Zavarise. On contact between three-dimensional beams undergoing large deflections. *Communications in Numerical Methods in Engineering*, 13(6):429–438, 1997.

P. Wriggers, T. Vu Van, and E. Stein. Finite element formulation of large deformation impact-contact problems with friction. *Computers & Structures*, 37(3):319–331, 1990.

P. Wriggers, L. Krstulovic-Opara, and J. Korelc. Smooth C1-interpolations for two-dimensional frictional contact problems. *International Journal for Numerical Methods in Engineering*, 51(12):1469–1495, 2001.

H. L. Xing and A. Makinouchi. Three dimensional finite element modeling of thermomechanical frictional contact between finite deformation bodies using R-minimum strategy. *Computer Methods in Applied Mechanics and Engineering*, 191(37):4193–4214, 2002.

B. Yang and T. A. Laursen. A contact searching algorithm including bounding volume trees applied to finite sliding mortar formulations. *Computational Mechanics*, 41(2):189–205, 2008.

B. Yang and T. A. Laursen. A mortar-finite element approach to lubricated contact problems. *Computer Methods in Applied Mechanics and Engineering*, 198(47-48):3656–3669, 2009.

B. Yang, Tod A. Laursen, and X. Meng. Two dimensional mortar contact methods for large deformation frictional sliding. *International Journal for Numerical Methods in Engineering*, 62(9):1183–1225, 2005.

Q. Yang, L. Stainier, and M. Ortiz. A variational formulation of the coupled thermo-mechanical boundary-value problem for general dissipative solids. *Journal of the Mechanics and Physics of Solids*, 54(2):401–424, 2006.

G. Zavarise and P. Wriggers. Contact with friction between beams in 3-D space. *International Journal for Numerical Methods in Engineering*, 49(8):977–1006, 2000.

G. Zavarise, P. Wriggers, E. Stein, and B. A. Schrefler. Real contact mechanisms and finite element formulation - a coupled thermomechanical approach. *International Journal for Numerical Methods in Engineering*, 35(4):767–785, 1992.

O. C. Zienkiewicz and R. L. Taylor. *The finite element method for solid and structural mechanics*. Elsevier Butterworth-Heinemann, 2005.

O. C. Zienkiewicz, R. L. Taylor, and J. Z. Zhu. *The finite element method: Its basis & fundamentals*. Elsevier Butterworth-Heinemann, 2005.

Advanced Discretization Methods for Contact Mechanics

Peter Wriggers

Abstract Modeling of contact problems is essential for many problems in engineering in order to predict the behaviour and response of various systems. One can think of pile driving, complex bearings, connections in Civil Engineering, of vehicle road interaction, machines and forming processes in Mechanical Engineering and of MEMS and electrical circuits in Electrical Engineering. All these systems need predictions of the behaviour, durability and efficiency. Hence models are needed that have to be solved by numerical methods due to their complexity. This contribution is aimed at modeling of contact in solid mechanics. Due to the necessity to use numerical methods for the solution of most contact applications this paper will focus mainly on numerical simulation models. Here especially new methodologies are considered that are non-standard and open the possibility for more general application ranges when compared to conventional approaches.

1 Theoretical Background for Contact Mechanics

This chapter will provide a short introduction into the theoretical background of contact mechanics. More sophisticated and elaborate treatments may be found in textbooks, e.g. Johnson (1958), Wriggers (2006) and Laursen (2002).

1.1 Contact Geometry

This section summarizes relations which are necessary to formulate the geometrical contact conditions. In detail the penetration and the relative slip in the contact area are discussed. The first condition also includes the non–penetration condition which is used classically in contact mechanics.

P. Wriggers (✉)
Institute of Continuum Mechanics, Gottfried Wilhelm Leibniz Universität, Hannover, Germany
e-mail: wriggers@ikm.uni-hannover.de

A. Popp and P. Wriggers (eds.), *Contact Modeling for Solids and Particles*,
CISM International Centre for Mechanical Sciences 585,
https://doi.org/10.1007/978-3-319-90155-8_2

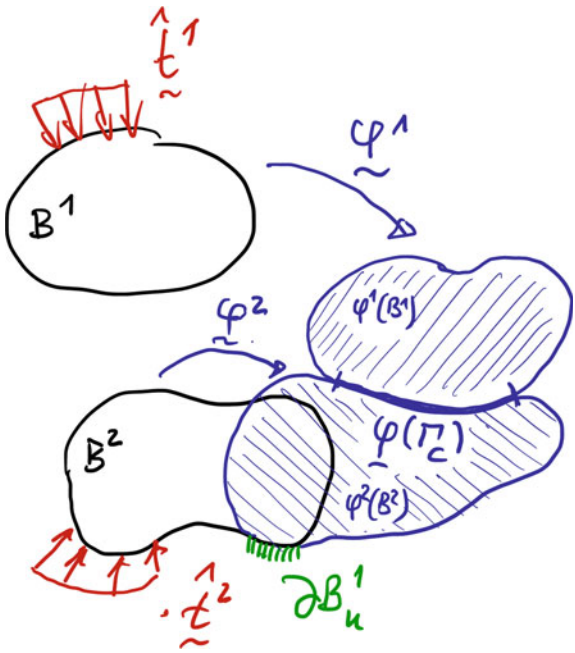

Fig. 1 Contacting continuum bodies

We assume that two bodies which undergo large deformations can come into contact. Let $B^\gamma,\ \gamma = 1, 2$, denote the two bodies of interest and $\boldsymbol{\varphi}^\gamma$ the associated deformation that maps points $\mathbf{X}^\gamma \in B^\gamma$ of the reference configuration onto points $\mathbf{x}^\gamma = \boldsymbol{\varphi}^\gamma(\mathbf{X}^\gamma)$ of the current configuration, see Fig. 1.

For contact a special non–penetration condition has to be formulated. Here we denote by $\Gamma_c \subset \partial B^\gamma$ a possible contact surface of the bodies B^γ, see Fig. 1 for an illustration of this concept. At this contact surface both bodies touch each other in the current configuration. Hence the surface part of body B^1, denoted by $\boldsymbol{\varphi}^1(\Gamma_c^1)$ has the same size and shape as the surface part of body B^2, denoted by $\boldsymbol{\varphi}^2(\Gamma_c^2)$ in the current configuration. The surface part $\boldsymbol{\varphi}^1(\Gamma_c^1)$ is here called slave surface which is in contact with the current master surface $\boldsymbol{\varphi}^2(\Gamma_c^2)$. The latter plays the role of a (moving) reference surface, see Fig. 2.

We parametrise the master surface in its reference and current configuration by the convective coordinates ξ^1, ξ^2 and use $\boldsymbol{\xi} = \{\xi^1, \xi^2\}$ as abbreviation. Thus we consider material curves in the reference and deformed configuration: $\mathbf{X}^2 = \widehat{\mathbf{X}}^2(\boldsymbol{\xi}) \subset \Gamma_c^2$ and $\mathbf{x}^2 = \hat{\mathbf{x}}_t^2(\boldsymbol{\xi}) \subset \boldsymbol{\varphi}_t^2(\Gamma_c^2)$. Then the local deformation gradient of the master surface is given by $\mathbf{F}^2 := \mathbf{a}_\alpha^2 \otimes \mathbf{A}^{2\alpha}$ based on the tangent vectors of the contact surface $\mathbf{a}_\alpha^2 := \hat{\mathbf{x}}_{t,\alpha}^2(\boldsymbol{\xi})$ and the tangent vectors at the same surface point $\mathbf{A}_\alpha^2 := \widehat{\mathbf{X}}_{,\alpha}^2(\boldsymbol{\xi})$ with respect to the surface in the initial configuration.[1]

[1] Here we use the standard relations $\mathbf{a}_\alpha^2 \cdot \mathbf{a}^{2\beta} = \delta_\alpha^\beta$ and $\mathbf{A}_\alpha^2 \cdot \mathbf{A}^{2\beta} = \delta_\alpha^\beta$. Furthermore $(\ )_{,\alpha}$ denotes differentiation with respect to the convective coordinate ξ^α.

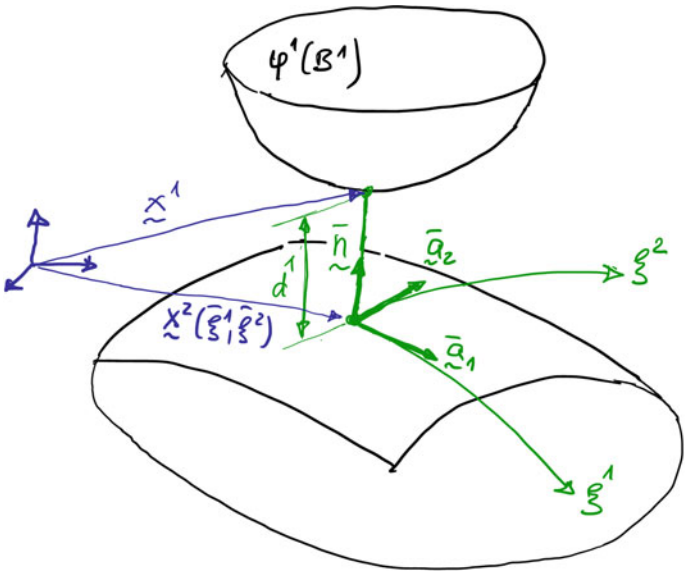

Fig. 2 Contact geometry, closest point projection

Penetration. As the first relevant function for the contact geometry we define a penetration function on the current slave surface $\boldsymbol{\varphi}^1(\Gamma_c^1)$

$$g_{N+} = \| \mathbf{x}^1 - \hat{\mathbf{x}}^2(\bar{\boldsymbol{\xi}}) \| \quad \text{for}[\, \mathbf{x}^1 - \hat{\mathbf{x}}^2(\bar{\boldsymbol{\xi}}) \,] \cdot \bar{\mathbf{n}}^2 < 0 \tag{1}$$

Here $(\bar{\boldsymbol{\xi}})$ is the minimizer of the distance function for a given slave point $\mathbf{x}^1$

$$\hat{d}^1(\boldsymbol{\xi}) = \| \mathbf{x}^1 - \hat{\mathbf{x}}^2(\boldsymbol{\xi}) \| \longrightarrow \text{MIN}\,, \tag{2}$$

see Fig. 2 for the geometrical setup. The values $(\bar{\boldsymbol{\xi}})$ are obtained by writing the necessary condition for the minimum of the distance function (2)

$$\frac{d}{d\xi^\alpha}\hat{d}^1(\boldsymbol{\xi}) = \frac{\mathbf{x}^1 - \hat{\mathbf{x}}^2(\boldsymbol{\xi})}{\| \mathbf{x}^1 - \hat{\mathbf{x}}^2(\boldsymbol{\xi}) \|} \cdot \hat{\mathbf{x}}^2_{,\alpha}(\boldsymbol{\xi}) = 0\,. \tag{3}$$

The solution of (3) requires the orthogonality of the first and second term. Since $\hat{\mathbf{x}}^2_{,\alpha}(\boldsymbol{\xi})$ is the tangent vector $\mathbf{a}^2_\alpha$ the first term must denote the normal $\mathbf{n}^2$. Thus we have the condition $\mathbf{n}^2 \cdot \mathbf{a}^2_\alpha = 0$ which means that the current master point $\hat{\mathbf{x}}^2(\boldsymbol{\xi})$ is the orthogonal projection of a given slave point $\mathbf{x}^1$ onto the current master surface $\boldsymbol{\varphi}^2(\Gamma_c^2)$.

Here and in the following we will denote by a bar over a quantity its evaluation at the minimal distance point $(\bar{\boldsymbol{\xi}})$ which means that these values denote the solution

point of (3). Thus $\bar{\mathbf{n}}^2 := (\bar{\mathbf{a}}_1^2 \times \bar{\mathbf{a}}_2^2) \,/\, \|\bar{\mathbf{a}}_1^2 \times \bar{\mathbf{a}}_2^2\|$ is the outward unit normal on the current master surface at the master point where $\bar{\mathbf{a}}_\alpha^2$ are tangent vectors at $\hat{\mathbf{x}}^2(\bar{\xi^1},\ \bar{\xi^2})$.

The penetration function g_{N+}, see (1), enters as a local kinematical variable the constitutive function for the contact pressure. Contrary

$$[\,\mathbf{x}^1 - \hat{\mathbf{x}}^2(\bar{\boldsymbol{\xi}})\,] \cdot \bar{\mathbf{n}}^2 \geq 0$$

represents the classical non–penetration condition for finite deformations.[2]

The variation of the penetration function in (1) leads with the orthogonality condition $\bar{\mathbf{n}}^2 \cdot \hat{\mathbf{x}}^2_{,\alpha}(\bar{\boldsymbol{\xi}}) \delta\xi^\alpha = 0$ to

$$\delta g_{N+} \;=\; [\,\boldsymbol{\eta}^1 - \hat{\boldsymbol{\eta}}^2(\bar{\boldsymbol{\xi}})\,] \cdot \bar{\mathbf{n}}^2 \tag{5}$$

with $\boldsymbol{\eta}$ being the virtual displacement or test function.

Tangential Relative Velocity and Tangential Relative Slip. The tangential relative slip between two bodies is related to the change of the solution point $(\bar{\boldsymbol{\xi}})$ of the minimal distance problem. Thus one can compute the time derivative of ξ^α from (3)

$$\frac{d}{dt}\{\,[\,\mathbf{x}^1 - \hat{\mathbf{x}}^2(\bar{\boldsymbol{\xi}})\,] \cdot \bar{\mathbf{a}}_\alpha^2\,\} = [\,\mathbf{v}^1 - \hat{\mathbf{v}}^2(\bar{\boldsymbol{\xi}}) - \bar{\mathbf{a}}_\beta^2\, \dot{\bar{\xi}}^\beta\,] \cdot \bar{\mathbf{a}}_\alpha^2 + [\,\mathbf{x}^1 - \hat{\mathbf{x}}^2(\bar{\boldsymbol{\xi}})\,] \cdot \dot{\bar{\mathbf{a}}}_\alpha^2 = 0\,. \tag{6}$$

With the time derivative of the tangent vector $\dot{\bar{\mathbf{a}}}_\alpha^2 = \hat{\mathbf{v}}^2_{,\alpha}(\bar{\boldsymbol{\xi}}) + \hat{\mathbf{x}}^2_{,\alpha\beta}(\bar{\boldsymbol{\xi}})\, \dot{\bar{\xi}}^\beta$ we obtain $\dot{\bar{\xi}}^\beta$ from the following system of equations

$$\bar{H}_{\alpha\beta}\, \dot{\bar{\xi}}^\beta = \bar{R}_\alpha \tag{7}$$

with

$$\bar{H}_{\alpha\beta} = [\,\bar{a}_{\alpha\beta} + g_{N+}\, \bar{b}_{\alpha\beta}\,]\,, \quad \bar{R}_\alpha = [\,\mathbf{v}^1 - \hat{\mathbf{v}}^2(\bar{\boldsymbol{\xi}})\,] \cdot \bar{\mathbf{a}}_\alpha^2 + g_{N+}\, \bar{\mathbf{n}}^2 \cdot \hat{\mathbf{v}}^2_{,\alpha}(\bar{\boldsymbol{\xi}})\,. \tag{8}$$

$\bar{a}_{\alpha\beta}$ and $\bar{b}_{\alpha\beta}$ are the first and second fundamental form of the deformed surface, well known from differential geometry.

With the above computed time derivative of the convective coordinates the tangential relative velocity function on the current slave surface $\boldsymbol{\varphi}_t^1(\Gamma_c^1)$ can be defined by the Lie derivative of the tangential vector

---

[2] For the analysis of small deformation problems the kinematical relation (1) can be linearized which yields

$$\Delta g_{N+} \;=\; [\,\mathbf{u}^1 - \hat{\mathbf{u}}^2(\bar{\boldsymbol{\xi}})\,] \cdot \bar{\mathbf{N}}^2 + g_0\,. \tag{4}$$

$\mathbf{u}^\gamma$ represents the displacement field which is introduced in the kinematically linear case to connect the current and the reference configuration via: $\mathbf{x}^\gamma = \mathbf{X}^\gamma + \mathbf{u}^\gamma$. The variable g_0 denotes the initial gap between the two bodies which is given by $g_0 = [\mathbf{X}^1 - \hat{\mathbf{X}}^2(\bar{\boldsymbol{\xi}})] \cdot \bar{\mathbf{N}}^2$ and the normal $\bar{\mathbf{N}}^2 = (\bar{\mathbf{A}}_1^2 \times \bar{\mathbf{A}}_2^2) \,/\, \|\bar{\mathbf{A}}_1^2 \times \bar{\mathbf{A}}_2^2\|$ is related to the reference configuration.

$$\mathcal{L}_v \, \mathbf{g}_T \; := \; \dot{\bar{\xi}}^{\alpha} \, \bar{\mathbf{a}}_{\alpha}^{2} \, . \tag{9}$$

Equation (9) determines per definition the evolution of the tangential slip $\mathbf{g}_T$ which enters as a local kinematical variable the constitutive function for the contact tangential stress, see next section. The rate $\dot{\bar{\xi}}^{\alpha}$ in (9) at the solution point $(\bar{\xi}^1, \; \bar{\xi}^2)$ has been already computed in (7). It can be shown that (9) provides an objective measure for the tangential slip, see e.g. Wriggers (2006).

In case of no relative movement in tangential direction (stick condition) we have $\mathcal{L}_v \, \mathbf{g}_T = \mathbf{g}_T = \mathbf{0}$.

1.2 Contact Contribution to the Variational Form

In large deformation analysis using numerical solution procedures classically the weak form of the local equilibrium equations is used.

Within the weak form G the constraint equations which govern the contact interaction have to be taken into account. This can be done in many different ways, see e.g. Wriggers (2006). Here we will restrict ourselves to the Lagrangian multiplier and the penalty method. The existence of an a priori unknown contact state yields additional constraints to the continuum problem. Thus the weak form has to be extended by an extra term G_C

$$G = \sum_{\gamma=1}^{2} \left[\int_{B^\gamma} \frac{1}{2} \boldsymbol{S} \cdot \delta \boldsymbol{C} \, dB - \int_{B^\gamma} \rho_0 \, \hat{\boldsymbol{b}} \cdot \boldsymbol{\eta} \, dB - \int_{\Gamma^\gamma} \hat{\boldsymbol{t}} \cdot \boldsymbol{\eta} \, dA \right] + G_C \tag{10}$$

For the continuum contribution a form that is based on the initial configuration B^γ is used with $\boldsymbol{S}$ being the second Piola-Kirchhoff stress tensor. $\boldsymbol{C} = \boldsymbol{F}^T \, \boldsymbol{F}$ is the right Cauchy Green tensor which can be computed using the deformation gradient $\boldsymbol{F} = \text{Grad} \, \boldsymbol{\varphi}$. Its variation is given by $\delta \boldsymbol{C} = \boldsymbol{F}^T \, \text{Grad} \, \boldsymbol{\eta} + \text{Grad} \, \boldsymbol{\eta}^T \, \boldsymbol{F}$ where $\boldsymbol{\eta}$ denotes the variation or test function, also known as $\delta \boldsymbol{u}$. The applied loads are the body forces $\rho_0 \, \hat{\boldsymbol{b}}$ and the surface tractions $\hat{\boldsymbol{t}}$.

For the second Piola Kirchhoff stress one has to formulate a constitutive equation that describes the material behaviour of the solids. Here many different constitutive equations are available for different applications such as elasticity, elasto-plasticity and visco-elasticity to name only a few, for details and applications within the finite element method see Wriggers (2008). In this contribution we will assume hyperelastic materials which lead to a nonlinear functional dependency of the second Piola Kirchhoff stress tensor on the right Cauchy-Green tensor $\boldsymbol{S} = f(\boldsymbol{C})$.

The G_C term that includes the contact constraint can be written with $\delta \mathbf{g}_T = \delta \bar{\xi}^{\alpha} \, \bar{\mathbf{a}}_{\alpha}^{2}$, see (9), as

$$G_C = \int_{\Gamma_c} (t_N \, \delta g_{N+} + t_{T\,\alpha} \, \delta \xi^\alpha) \; d\Gamma . \tag{11}$$

G_C is only valid within the active contact zone Γ_c. It is obvious that the contribution of contact to the weak form is split into a normal and a tangential part. The quantity t_N denotes the normal and $t_{T\,\alpha}$ the tangential contact stress components. In case of no penetration and stick, the contact stresses are known as Lagrangian multipliers. These can only be computed from the general equilibrium equations and thus enter the weak form as unknowns.

The variation of the gap and the relative tangential deformations are described by δg_N, see (5), and $\delta \xi^\alpha$, respectively.

The covariant components of the tangential part of the stress vector in (11) are given by $t_{T\,\alpha} = t_T^\beta \, \bar{a}_{\alpha\beta}$ where $\bar{a}_{\alpha\beta} = \bar{\boldsymbol{a}}_\alpha^2 \cdot \bar{\boldsymbol{a}}_\beta^2$.

In case of a penalty formulation one can assume that linear springs replace the normal and tangential contact stresses. This leads to a contact part of the weak form for the penalty approach at a contact surface Γ_c

$$G_C^P = \int_{\Gamma_C} \left(\epsilon_N \, g_{N+} \delta g_{N+} + \epsilon_T \, \boldsymbol{g}_T \cdot \delta \boldsymbol{g}_T \right) \; d\Gamma \, . \tag{12}$$

Note that the contact formulations above are only valid for the non-penetration and the tangential stick case. Once sliding occurs the tangential stress can only be computed from a constitutive equation which is discussed next.

1.3 Frictional Contact

In case of the tangential contact one has to distinguish between stick and slip state. Here the slip condition is formulated analog to the yield condition in the theory of plasticity, for details see Laursen (2002) and Wriggers (2006).

For the classical Coulomb law with constant friction coefficient μ the slip condition can be written as

$$f^s = ||\boldsymbol{t}_T|| - \mu t_N |g_N| \leq 0 \, . \tag{13}$$

If $f^s < 0$ holds, the bodies stick together and the relative displacement matches the elastic displacement. The tangential stress follows as

$$\boldsymbol{t}_T = \epsilon_T \, (\boldsymbol{g}_T - \boldsymbol{g}_T^s) \, . \tag{14}$$

where $\boldsymbol{g}_T$ is the total relative tangential movement at an incremental step and $\boldsymbol{t}_T^s$ is the relative sliding. Note that the latter can be zero, if no slip has occurred during the total deformation of the bodies.

Frictional sliding is always linked to dissipation. Since the equilibrium state is equivalent to a minimum of remaining power in the system, the dissipated power D during sliding has to be maximal

$$D = \boldsymbol{t}_T \cdot \dot{\boldsymbol{g}}_T^s \geq 0 \,. \tag{15}$$

Including the constraint $f^s \leq 0$ multiplied by a Lagrange parameter $\lambda \geq 0$ this can be written as a Lagrange functional

$$L = \boldsymbol{t}_T \cdot \dot{\boldsymbol{g}}_T^s - \lambda f^s \geq 0 \tag{16}$$

whose minimization leads to

$$\dot{\boldsymbol{g}}_T^s = \lambda \frac{\boldsymbol{t}_T}{||\boldsymbol{t}_T||} \,. \tag{17}$$

Equation (17) denotes the evolution equation for the plastic slip.

The solution of the set of equations which govern the tangential movement is obtained by a radial-return algorithm. The formulation of this algorithm, see e.g. Wriggers (2006), leads to update formulae for tangential stress and tangential gap. It is summarized in Fig. 3 where after a predictor step either no correction occurs in the stick case or a correction takes place in case of slip. Here the evolution of the slip is computed within a time step $\Delta t = t_{n+1} - t_n$ using an Euler backward integration rule. Thus the index n denotes the time at the previous step while $n+1$ is related to the current time.

Predictor step:

$$\begin{aligned} \boldsymbol{t}_T^{tr,n+1} &= \epsilon_T \left(\boldsymbol{g}_T^{n+1} - \boldsymbol{g}_T^{s,n}\right) \\ f^{s,tr} &= ||\boldsymbol{t}_T^{tr,n+1}|| - \mu\, t_N \end{aligned}$$

stick:

$$\begin{aligned} f^{s,tr} &\leq 0 \\ \boldsymbol{g}_T^{s,n+1} &= 0 \\ \boldsymbol{t}_T^{n+1} &= \boldsymbol{t}_T^{tr,n+1} \end{aligned}$$

slip:

$$\begin{aligned} f^{s,tr} &> 0 \\ \boldsymbol{g}_T^{s,n+1} &= \boldsymbol{g}_T^{s,n} + \frac{f^{s,tr}}{\epsilon_T}\, \boldsymbol{n}^{tr} \\ \boldsymbol{t}_T^{n+1} &= \boldsymbol{t}_T^{tr,n+1} - f^{s,tr}\, \boldsymbol{n}^{tr} \\ \boldsymbol{n}^{tr} &= \frac{\boldsymbol{t}_T^{tr,n+1}}{||\boldsymbol{t}_T^{tr,n+1}||} \end{aligned}$$

Fig. 3 Radial-return algorithm for friction

2 Isogeometric Contact Formulations

Isogeometric analysis is a discretization scheme which uses basis functions emanating from computer aided geometric design (CAGD), such as B-Splines, NURBS, T-splines, subdivision surfaces, etc., instead of traditional $\mathcal{C}^0$-continuous Lagrange finite element polynomials. It has been shown in numerous studies that isogeometric analysis provides more precise and efficient geometric representations, high accuracy combined with robustness, simplifies mesh refinement schemes, smooth basis functions with compact support, accurate derivatives and stresses, and the potential to integrate CAGD and analysis processes.

In this approach smooth, compactly-supported basis functions are employed that can improve the modeling of contact problems, especially ones in which sliding is significant. This observation is based on the fact, that $\mathcal{C}^0$-continuous finite element basis functions often create serious convergence problems in the analysis of sliding contact due to non-smooth surface discretization. In order to alleviate these problems various surface smoothing algorithms have been developed in the literature. The formulations are based on either a Hermitian, Spline or BÉZIER interpolation that is used to discretize the master surface defining the normal and tangent vector field. This leads to a C^1 or even C^2 continuous interpolation of the surface. Related work can be found for two- and three dimensional discretizations of deformable solids being in contact in Eterovic and Bathe (1991), Pietrzak and Curnier (1999), Padmanabhan and Laursen (2001), Wriggers et al. (2001) and Krstulovic-Opara et al. (2002). These interpolations lead in general to a more robust behaviour of the iterative solution algorithms for contact since normal and tangent fields are continuous. However they do not increase the order of convergence since the higher order approximations involve only the surface but not the bulk behaviour of the solids.

Here we study contact problems with isogeometric analysis. For this we summarize basic properties of NURBS. Of particular importance is that the NURBS surface discretization will be inherited directly from the NURBS volume discretization, thereby circumventing the need for additional surface smoothing techniques and the difficulties associated with such approaches. The resulting discretization scheme is applied to three-dimensional contact calculations utilizing a penalty-function regularized knot-to-surface (KTS) algorithm. However it is found that a mortar-based KTS approach behaves better than the standard KTS approach.

The comparison studies indicate that NURBS provide better contact pressures than $\mathcal{C}^0$-continuous Lagrange finite elements of the same order. In all cases, the NURBS contact pressures were pointwise-positive whereas higher order $\mathcal{C}^0$-continuous Lagrange finite element results exhibited non-physical negative values. The numerical results for NURBS are encouraging and suggest their use in finite deformation contact analysis.

As will be shown this approach yields highly robust schemes but still needs a high effort when implementing these discretizations, see e.g. Temizer et al. (2011), de Lorenzis et al. (2011), Temizer et al. (2012), de Lorenzis et al. (2012) and Dimitri et al. (2014). The NURBS based methods can be applied to many different problems

and have the advantages that the contact surface remains smooth even when the solids in contact undergo very large deformations.

2.1 Isogeometric Treatment of Contact

Within the numerical contact treatment, B^1 will be identified as a slave body whereas B^2 is the master, see Sect. 1.1. Within this convention, the master surface is parametrized via convective coordinates $\xi^\alpha, \alpha \in \{1, 2\}$, that define covariant tangent vectors $\mathbf{a}_\alpha = \mathbf{x}^{(2)}_{,\alpha}$. Using the metric $a_{\alpha\beta} := \mathbf{a}_\alpha \cdot \mathbf{a}_\beta$ with inverse components $a^{\alpha\beta}$, the contravariant vectors $\mathbf{a}^\alpha := a^{\alpha\beta}\mathbf{a}_\beta$ are induced. The components of the symmetric curvature tensor follow from $b_{\alpha\beta} = \mathbf{a}_{\alpha,\beta} \cdot \boldsymbol{n}^2$.

Two major steps within the numerical implementation are the closest point projection, and the linearization of δg_N, see e.g. Wriggers (2006). The closest point projection defines a residual, see (3),

$$f_\alpha(\xi^1, \xi^2) = \mathbf{a}_\alpha(\xi^1, \xi^2) \cdot \left[\mathbf{x}^1 - \mathbf{x}^2(\xi^1, \xi^2)\right] \tag{18}$$

that vanishes at the projection point corresponding to $\{\bar{\xi}^1, \bar{\xi}^2\}$ for all α: $f_\alpha(\bar{\xi}^1, \bar{\xi}^2) = 0$. Since $f_\alpha(\xi^1, \xi^2)$ is a nonlinear function in ξ^1 and ξ^2 when higher order NURBS are used to discretize the surface an iterative process starting from a guess $\{\xi^1, \xi^2\}$ has to be applied. It requires the tangent

$$K_{\alpha\beta} := f_{\alpha,\beta} = \mathbf{a}_{\alpha,\beta}(\xi^1, \xi^2) \cdot (\mathbf{x}^1 - \mathbf{x}^2(\xi^1, \xi^2)) - a_{\alpha\beta}(\xi^1, \xi^2) \tag{19}$$

that will be used in a local Newton algorithm to find the closest point $\{\bar{\xi}^1, \bar{\xi}^2\}$.

On the other hand, the linearization of δg_N associated with the projection point reads

$$\Delta\delta g_N = \boldsymbol{n}^2 \cdot (\boldsymbol{\eta}^2_{,\alpha}\, \Delta\xi^\alpha + \delta\xi^\alpha\, \Delta\boldsymbol{x}^2_{,\alpha}) + \delta\xi^\alpha\, b_{\alpha\beta}\, \Delta\xi^\beta + g_N \delta\boldsymbol{n}^2 \cdot \Delta\boldsymbol{n}^2 \tag{20}$$

with the variations $\delta\boldsymbol{n}^2 = -(\boldsymbol{\eta}^2_{,\alpha} \cdot \boldsymbol{n}^2 + b_{\alpha\beta}\, \delta\xi^\beta)\mathbf{a}^\alpha$ and $\delta\xi^\beta = H^{\alpha\beta}[(\boldsymbol{\eta}^1 - \boldsymbol{\eta}^2) \cdot \mathbf{a}_\alpha - g_N \mathbf{n}^2 \cdot \boldsymbol{\eta}^2_{,\alpha}]$ where $H^{\alpha\beta}$ are the inverse components of $H^{\alpha\beta}$, given in (8). Similar expressions follow for the increments $\Delta\boldsymbol{\eta}^2$ and $\Delta\xi^\alpha$. In the following, for simplicity the variables associated with the projection point are not denoted explicitly via the $(\bar{\bullet})$ notation.

NURBS Discretization of the Volume. It is clear that both the closest point projection and the linearization of δg_N need the evaluation of $\mathbf{a}_{\alpha,\beta}$. The contact geometry will be a NURBS surface that is directly inherited from the volume NURBS discretization. Therefore, in order to evaluate $\mathbf{a}_{\alpha,\beta}$ and subsequently discretize the contact contributions to the weak form, the NURBS discretization of the volume is

introduced next. In what follows, standard NURBS terminology is employed, see e.g. Piegl and Tiller (1996) and Cottrell et al. (2009) for further details.

Let Ξ^i be the open non-uniform knot vector associated with a patch along the i-th dimension with $m_i = n_i + p_i + 1$ knots:

$$\Xi^i = \{ \underbrace{\xi^i_0, \ldots, \xi^i_{p_i}}_{p_i+1 \text{ equal terms}}, \xi^i_{p_i+1}, \ldots, \xi^i_{n_i}, \underbrace{\xi^i_{n_i+1}, \ldots, \xi^i_{m_i}}_{p_i+1 \text{ equal terms}} \} \quad . \tag{21}$$

Here, p_i is the polynomial order of the accompanying B-spline basis functions, ξ^i_j is the j-th knot and $n_i + 1$ would be the number of accompanying control points in a one-dimensional setting. In a three-dimensional setting, a volume is parametrized by

$$\mathbf{V}(\xi^1, \xi^2, \xi^3) = \sum_{d_1=0}^{c_1} \sum_{d_2=0}^{c_2} \sum_{d_3=0}^{c_3} R_{d_1 d_2 d_3}(\xi^1, \xi^2, \xi^3) \mathbf{P}_{d_1 d_2 d_3} \tag{22}$$

where $\mathbf{P}_{d_1 d_2 d_3}$ are the control points and $R_{d_1 d_2 d_3} \geq 0$ are the rational B-spline (NURBS) basis functions. The latter are defined via a tensor product in a four-dimensional space based on homogeneous coordinates (Piegl and Tiller 1996). The projected form in the three-dimensional space is

$$R_{d_1 d_2 d_3}(\xi^1, \xi^2, \xi^3) = \frac{w_{d_1 d_2 d_3}}{W(\xi^1, \xi^2, \xi^3)} B^1_{d_1}(\xi^1)\, B^2_{d_2}(\xi^2)\, B^3_{d_3}(\xi^3) \tag{23}$$

with $B^i_{d_i}$ as a nonrational B-spline basis function. The normalizing weight W is given in terms of the weights $w_{d_1 d_2 d_3} > 0$ and $B^i_{d_i}$ via

$$W(\xi^1, \xi^2, \xi^3) = \sum_{d_1=0}^{c_1} \sum_{d_2=0}^{c_2} \sum_{d_3=0}^{c_3} w_{d_1 d_2 d_3} B^1_{d_1}(\xi^1)\, B^2_{d_2}(\xi^2)\, B^3_{d_3}(\xi^3). \tag{24}$$

The knot vectors together with the associated control points and the accompanying weights constitute a patch.

The continuity and order of $B^i_{d_i}$ depends on Ξ^i only. If Ξ^i has no repeated interior knot ξ^i_j, $j \in [p_i + 1, n_i]$, then the order-p_i basis function $B^i_{d_i}$ has continuity $\mathcal{C}^{p_i - 1}$. Every repetition of a knot decreases the continuity by one order at this knot. The order of NURBS parametrization will be denoted by $\mathcal{N}^p$ in subsequent sections, while the order of Lagrange polynomials employed will be denoted by $\mathcal{L}^p$. The approximation spaces based on a linear interpolation $\mathcal{N}^1$ and $\mathcal{L}^1$ are identical.

In a finite element setting, all degrees of freedom ($\mathbf{x}$) are discretized via the same NURBS basis functions used for the geometric description. The unique knot spans are conveniently chosen as the integration domains (elements). The counterparts of the h- and p-refinement procedures for FEM discretizations based on Lagrange polynomials are the *knot insertion* and *order elevation* procedures in the NURBS setting. While p-refinement preserves the number of nodes, order elevation leads

to an increase in the number of control points. When the two must be conducted together, the *k-refinement* procedure will be employed where knot refinement precedes order elevation (Cottrell et al. 2009). This has the advantage that a higher degree of smoothness can be achieved within the patch across non-repeated knot entries and the final number of control points is less compared to the case where knot refinement precedes.

For the numerical evaluation of the weak forms emanating from Lagrange or NURBS based discretizations, $2p$ Gauss-Legendre quadrature points will be employed within each element for order-p approximations. This ensures a converged quadrature. See Hughes et al. (2010) for a recent discussion of efficient quadrature schemes appropriate for isogeometric analysis.

NURBS Discretization of the Surface. NURBS surface parametrization is inherited from the volume parametrization in a straightforward fashion. For example, let $\xi^1_- := \xi^1_0$. By construction (Piegl and Tiller 1996)

$$\mathbf{V}(\xi^1_-, \xi^2, \xi^3) = \sum_{d_2=0}^{c_2} \sum_{d_3=0}^{c_3} R^-_{d_2 d_3}(\xi^2, \xi^3) \mathbf{P}^-_{d_2 d_3} \tag{25}$$

where $\mathbf{P}^-_{d_2 d_3} := \mathbf{P}_{0 d_2 d_3}$ and, including the weighting factor,

$$R^-_{d_2 d_3}(\xi^2, \xi^3) := \frac{w_{0 d_2 d_3} B^2_{d_2}(\xi^2)\, B^3_{d_3}(\xi^3)}{\sum_{d_2=0}^{c_2} \sum_{d_3=0}^{c_3} w_{0 d_2 d_3} B^2_{d_2}(\xi^2)\, B^3_{d_3}(\xi^3)} \quad . \tag{26}$$

Hence, only the knowledge of the knot vectors Ξ^2 and Ξ^3 and a reduced set of control points together with the accompanying weights are sufficient to characterize the surface associated with ξ^1_-.

The same principle applies for $\xi^1_+ := \xi^1_{c_1+p_1+1}$ and all other dimensions. Hence, in general, a surface patch (in particular a *contact* patch) is directly inherited from the volume patch and has the same parametrization but only with two dimensions $\alpha \in \{1, 2\}$ that correspond to any two of the three dimensions. The corresponding knot vectors are Ξ^α with associated B-spline basis functions $B^\alpha_{d_\alpha}$ and parametric space coordinates ξ^α that are conveniently chosen as the convective coordinates for contact computations. The surface parametrization is therefore

$$\mathbf{S}(\xi^1, \xi^2) = \sum_{d_1=0}^{c_1} \sum_{d_2=0}^{c_2} R_{d_1 d_2}(\xi^1, \xi^2) \mathbf{P}_{d_1 d_2} \quad . \tag{27}$$

This can now be used within the contact formulation as a discretization of the contacting surfaces.

2.2 Knot-to-Surface Contact Algorithm

The classical node-to-surface (NTS) algorithm of computational contact mechanics cannot be directly employed with NURBS because the control points are not interpolatory. The straightforward extension of NTS to the isogeometric setting corresponds to a *knot-to-surface (KTS)* algorithm. Herein $g_N = 0$ is enforced during contact directly at the quadrature points associated with the projection onto the master surface. Each quadrature point corresponds to a unique value $(\xi^1, \xi^2) \in [0, 1] \times [0, 1]$. Following (Fischer and Wriggers 2005, 2006; Ziefle and Nackenhorst 2008; Franke et al. 2010), the same procedure will be employed for Lagrange polynomial interpolations as well. The number of Gauss-Legendre quadrature points per direction employed for the evaluation of the contact contributions to the weak form will be noted explicitly.

In the following, some aspects of the KTS algorithm are investigated qualitatively. All examples presented in this section employ the penalty method to regularize the contact constraints, see (12).

Figure 4 shows various discretizations for a deformable body compressed against another one in a purely mechanical setting at a coarse discretization. Here the lower body is chosen as the master body. The upper body is ten times stiffer with respect to compression and shear. The $\mathcal{L}^1/\mathcal{N}^1$-discretization displays the typical problem in contact mechanics, namely that at coarse discretizations the interface resolution is unsatisfactory. The $\mathcal{L}^2$-discretization provides a much better resolution with the same number of degrees of freedom. However, the discontinuity in smoothness on the master surface is clearly visible at the contact interface. Such discontinuities lead to convergence difficulties in classical contact algorithms and various smoothing algorithms have been designed to alleviate such difficulties. The $\mathcal{N}^2$-discretization with the same number of elements does not involve any repeated interior knots. Consequently, a qualitatively well-resolved C^1-continuous contact interface is visible. This is a potential advantage that is beneficial particularly for the simulation of frictional contact with large sliding.

The same example in a three-dimensional setting is provided in Fig. 5 employing an $\mathcal{N}^2$-discretization. Here, the slave body is compressed onto the master body (5 loading steps) and subsequently it is rotated through 45° (10 loading steps) and dragged along the diagonal of the master surface (30 loading steps). During the twist

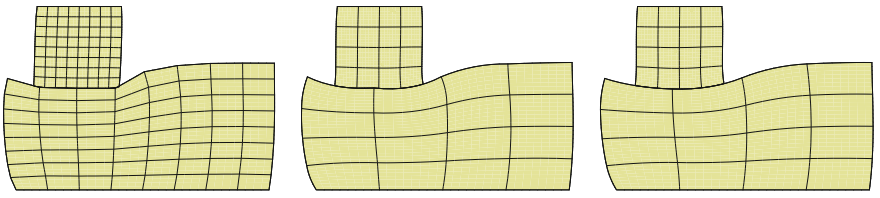

Fig. 4 Two-dimensional contact of two deformable bodies using $\mathcal{L}^1/\mathcal{N}^1$-, $\mathcal{L}^2$- and $\mathcal{N}^2$-discretizations

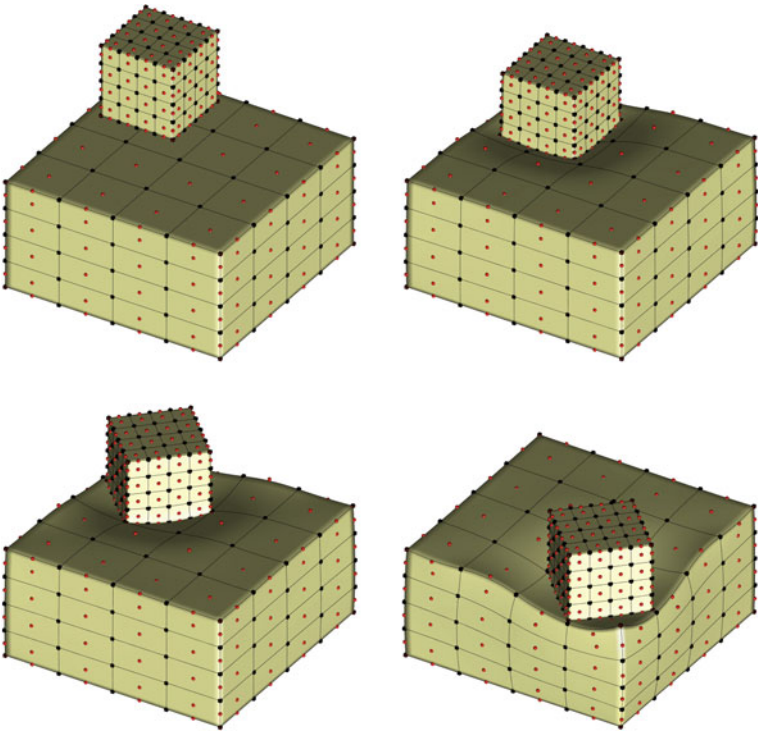

Fig. 5 Three-dimensional contact of two deformable bodies based on an $\mathcal{N}^2$-discretization

and drag stages, the slave surface traverses multiple element boundaries. Again, both the master and slave surfaces are $\mathcal{C}^1$-continuous. In all applications of the KTS algorithm, the active-set search is embedded within the Newton-Raphson iterations for the nonlinear system. The load steps are chosen such that convergence is achieved in at most 10 Newton-Raphson iterations. Within this setup and at the given resolution, one cannot achieve convergence with $\mathcal{L}^1/\mathcal{N}^1$- or $\mathcal{L}^2$-discretizations.

While the straightforward application of the KTS algorithm delivers qualitatively satisfactory results, quantitative investigations are necessary to establish the quality with which the contact constraints are satisfied, for more details see Temizer et al. (2011) and de Lorenzis et al. (2011).

2.3 Relaxation of Contact Constraints: Mortar KTS

Although qualitatively good results were obtained so far, the standard KTS approach delivers an excessively stiff contact constraint enforcement for both Lagrange and NURBS discretizations, see e.g. Temizer et al. (2011). The degree of freedom pro-

vided by the polynomial approximation cannot satisfy the constraints accurately pointwise at the quadrature points. This results e.g. in pressure oscillations near the edge of a contact interface.

In order to relax the over-constrained contact formulation, earlier mortar approaches are applied to the KTS algorithm. For recent applications of mortar-based contact algorithms with further references, the reader is referred to Puso et al. (2008), Hesch and Betsch (2009), Tur et al. (2009), Hüeber and Wohlmuth (2009). In this work, the original KTS contact constraints are relaxed based on an averaging of the normal penetration, as described in Tur et al. (2009) within a Lagrange multiplier setting. The application of integration techniques that accompany various mortar approaches, e.g. Puso et al. (2008) will not be tackled in this contribution, for details see e.g. Temizer et al. (2011) and de Lorenzis et al. (2012).

For conciseness, the surface parametrization in Eq. (27) is expressed as $\mathbf{S} = \sum_I R^I \mathbf{P}^I$ where R^I is the rational basis function associated with the control point at position $\mathbf{P}^I$. Subsequently, the pressure distribution is not defined locally as for standard KTS but through a discretization as for all other degrees of freedom:

$$p_N = \sum_I R^I p_N^I \quad . \tag{28}$$

The pressure degrees of freedom (*control pressures*) are defined within a penalty approach as

$$p_N^I = \varepsilon_N g_N^I \chi^I \tag{29}$$

where g_N^I are averaged control point penetrations:

$$g_N^I = \langle g_N R^I \rangle := \frac{\int_{\mathcal{R}_o^c} g_N R^I \delta A}{\int_{\mathcal{R}_o^c} R^I \delta A} \quad . \tag{30}$$

The active-set is associated with points where $\chi^I = 1$:

$$\chi^I = \begin{cases} 1 \; if \; g_N^I > 0 \\ 0 \; if \; g_N^I \leq 0 \end{cases} \quad . \tag{31}$$

This *mortar KTS* formulation clearly leads to an extended coupling among the contact patch degrees of freedom compared to the standard KTS treatment.

Popp et al. (2009) has suggested that no averaging is necessary for the interior nodes of a Lagrange element. However, an interior node is not a concept that is applicable for NURBS discretizations. Therefore, to retain a full analogy, the same procedure is employed for Lagrange discretizations in a straightforward fashion by replacing R^I with appropriate nodal shape functions. See Fischer and Wriggers (2006) and Puso et al. (2008) for earlier mortar approaches with $\mathcal{L}^2$-discretizations.

To demonstrate the effects of relaxation, the contact pressure and frictional shear stress is computed for an example in which two elastic bodies come into contact,

see Fig. 6a. The upper block is first pressed against a surface and then twisted, for details see Temizer et al. (2012). The geometrical setup and the deformation for this example is depicted in Fig. 6b where a block is first pressed and then rotated by 180° on an another deformable body. For the discretization a quadratic NURBS $\mathcal{N}^2$ interpolation was used. The stress field at the contact surface is shown in Fig. 7. Part (a) depicts the frictional shear stresses t_T after at the final rotation which was achieved in incremental 40 steps. The contact normal pressure t_N is depicted in part (b) of Fig. 7.

As was be observed, the KTS algorithm delivers qualitatively satisfactory results for various two- and three-dimensional finite deformation thermomechanical contact problems. However, quantitative investigations on the classical Hertz contact problem highlight a need for the relaxation of the mechanical contact constraints, see Temizer et al. (2011). Hence the relaxed mortar KTS approach provides superior results in comparison to the standard KTS algorithm for NURBS discretizations as

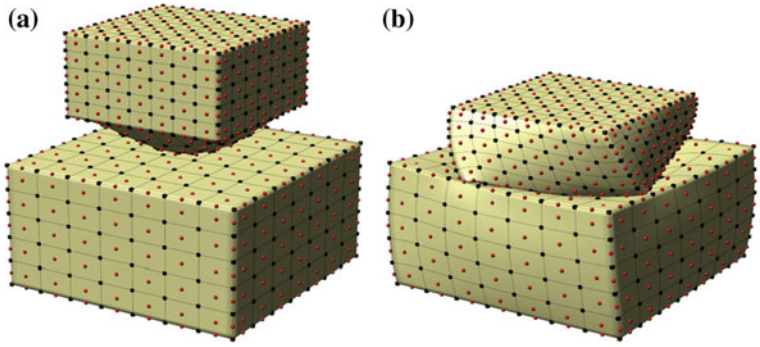

Fig. 6 **a** Initial configuration of the block; **b** deformations after pressing and rotating

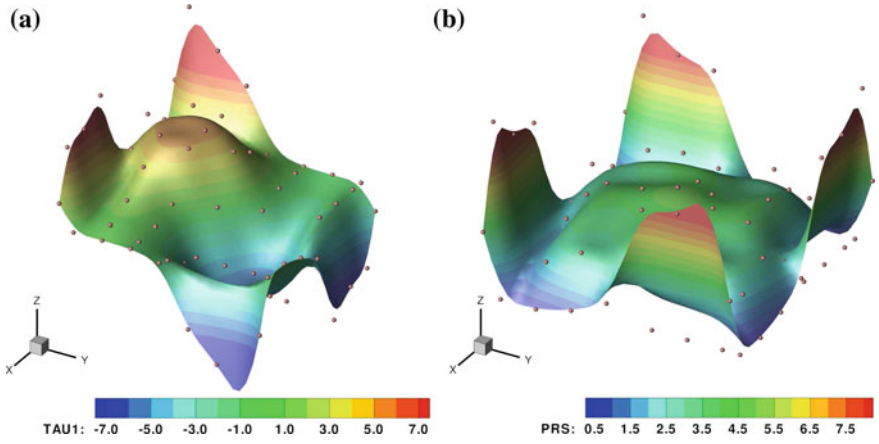

Fig. 7 **a** Initial configuration of the block; **b** deformations after pressing and rotating

well as for its application to Lagrange discretizations. Based on the mortar KTS approach the following conclusion regarding NURBS discretizations for contact problems can be made: (1) at a fixed resolution, NURBS discretizations produce results that are significantly less sensitive to changes in the r-refinement and the order of approximation compared to Lagrange discretizations, (2) for a fixed order, NURBS discretizations deliver monotonically improving results with increasing resolution whereas Lagrange discretizations deliver a non-monotonic improvement, (3) NURBS-based mortar KTS approach guarantees a pointwise-positive contact pressure distribution for all orders whereas $\mathcal{C}^0$-continuous Lagrange finite elements may produce non-physical negative pressures.

In summary, the developed mortar KTS approach delivers robust and accurate results for NURBS-based isogeometric contact treatment. Hence NURBS-based isogeometric analysis is a viable technology for contact problems undergoing finite strains.

3 Virtual Element Method for Contact

The virtual element method (VEM) is a recently developed Galerkin finite element approach that accommodates polygonal or polyhedral elements in two and three dimensions, respectively (see for example Beirão da Veiga et al. (2013), Beirão da Veiga et al. (2014) and, for problems in elasticity, Beirão Da Veiga et al. (2013), Gain et al. (2014)).

The objective here is to present a VEM formulation for contact between elastic bodies. The motivation for using the VEM lies in the ability to impose the contact constraints with the use of arbitrary, that is non-matching, meshes for the two bodies. Thus, there is no need to have the same numbers of nodes on the two candidate contact faces, nor is it necessary to match nodes across the two surfaces. Instead, the algorithm developed here makes provision for the insertion of new nodes on an interface, based on the distribution of nodes on the opposite boundary. This step is easily accommodated in an otherwise standard approach, see e.g. Wriggers et al. (2016).

The contact constraint is imposed using either a Lagrange multiplier or penalty approach, see Sect. 1.2. The formulation and examples are confined to a low-order VEM formulation in two dimensions, in which the shape functions along the element edges are linear. Note however the method is easily extended to higher-order VEM schemes and three dimensions.

3.1 Formulation of the Virtual Element Method

Within the virtual element method one creates a discretization scheme that allows for arbitrarily shaped elements. Even for a linear interpolation these elements can have

an arbitrary number of nodal points that define convex but also non-convex geometries. The domains Ω_e are partitioned into non-overlapping polygonal elements. A lowest-order approach is adopted, so that nodes are placed only at the vertices of the polygonal elements. The construction of such virtual elements is performed in a two step process. In the first step one obtains from the weak form a term that ensures consistency and in a second step one has to design a proper stabilization. This process has some features in common with the development of reduced order finite elements.

Generally the virtual element method relies on the split of the ansatz space into a part $\Pi \boldsymbol{u}_h$ and a remainder

$$\boldsymbol{u}_h = \Pi \boldsymbol{u}_h + (\boldsymbol{u}_h - \Pi \boldsymbol{u}_h)\,. \tag{32}$$

The projection $\Pi \boldsymbol{u}_h$ is defined at element level for a linear interpolation by

$$\Pi \boldsymbol{u}_h = \mathbf{H}\,\mathbf{a} = \begin{bmatrix} 1 & 0 & x & 0 & y & 0 \\ 0 & 1 & 0 & x & 0 & y \end{bmatrix} \begin{Bmatrix} a_1 \\ a_2 \\ \dots \\ a_6 \end{Bmatrix}. \tag{33}$$

The linear ansatz for the deformation along the element edge is given for a boundary segment k of the virtual element, defined by the local nodes (1)–(2) by, see right side of Fig. 8,

$$(\boldsymbol{u}_h)_k \;= (1 - \xi_k)\,\mathbf{u}_1 + \xi_k\,\mathbf{u}_2 = M_{k\,1}\,\mathbf{u}_1 + M_{k\,2}\,\mathbf{u}_2 \quad \text{with } \xi_k = \frac{x_k}{L_k} \tag{34}$$

where, for example, $M_{k\,1}$ is the ansatz function along a segment k related to node (1), ξ_k is the local dimensionless coordinate and $\mathbf{u}_1$ is the nodal value at that node, see Fig. 8.

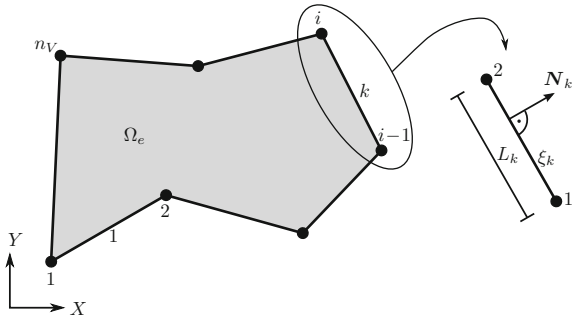

Fig. 8 Virtual element with n_V nodes and local boundary segment of the element

The unknown parameters a_i in (33) that define $\Pi \boldsymbol{u}_h$ can be computed in terms of the unknown nodal displacements $(\boldsymbol{u}_h)_k$ by using the identity,[3] see Beirão Da Veiga et al. (2015),

$$\nabla \Pi \boldsymbol{u}_h|_e = \frac{1}{\Omega_e} \int_{\Omega_e} \text{Grad}\, \boldsymbol{u}_h \, d\Omega = \frac{1}{\Omega_e} \int_{\Gamma_e} \boldsymbol{u}_h \otimes \boldsymbol{N} \; d\Gamma \tag{35}$$

where $\boldsymbol{N}$ is the normal at the boundary Γ_e of the domain Ω_e. From (33), the gradient of the projection is thus given by

$$\nabla \Pi \boldsymbol{u}_h|_e = \begin{bmatrix} a_3 & a_5 \\ a_4 & a_6 \end{bmatrix} \tag{36}$$

which is constant at element level. The right hand side of (35) yields with (34)

$$\frac{1}{\Omega_e} \int_{\Gamma_e} \boldsymbol{u}_h \otimes \boldsymbol{N} \; d\Gamma = \frac{1}{\Omega_e} \sum_{k=1}^{n_V} \int_{\Gamma_k} \begin{bmatrix} u_x(\boldsymbol{x}_k) N_x^k & u_x(\boldsymbol{x}_k) N_y^k \\ u_y(\boldsymbol{x}_k) N_x^k & u_y(\boldsymbol{x}_k) N_y^k \end{bmatrix} L_k \, d\Gamma \tag{37}$$

where we have used $\boldsymbol{N}_k = \{ N_x^k , N_y^k \}^T$ and $\boldsymbol{u} = \{ u_x , u_y \}^T$, and n_V are the number of segments of the element. Now, by inspection, the parameters $a_3 , \ldots , a_6$ can be expressed as functions of the nodal displacements $(\boldsymbol{u}_h)_k$. This projection determines the ansatz $\Pi \boldsymbol{u}_h$ on (33) within an element only up to a constant strain field and has to be supplemented by a further condition to ensure uniqueness. For this purpose we adopt the condition [see for example Beirão Da Veiga et al. (2013)] that the sum of the nodal values of $\boldsymbol{u}_h$ and of its projection $\Pi \boldsymbol{u}_h$ are equal. This yields for each element Ω_e

$$\frac{1}{n_V} \sum_{I=1}^{n_V} \Pi \boldsymbol{u}_h(\boldsymbol{x}_I) = \frac{1}{n_V} \sum_{I=1}^{n_V} \boldsymbol{u}_h(\boldsymbol{x}_I) \, , \tag{38}$$

where $\mathbf{x}_I$ are the coordinates of the nodal point I and the sum includes all boundary nodes. From this relation one can then compute the parameters a_1 and a_2.[4]

With (36) the displacement gradient is known. Thus the deformation gradient related to the projection $\Pi \boldsymbol{u}_h$ can be computed as

$$\boldsymbol{F}|_e = \mathbf{1} + \nabla \Pi \boldsymbol{u}_h|_e \; . \tag{39}$$

The simplest possible formulation for a finite deformation virtual element is a split into a constant part of the deformation gradient and an associated stabilization term. This was performed for the linear case in Beirão Da Veiga et al. (2013). The

[3] Note that this simple and efficient computation of the virtual ansatz space is only valid for linear interpolations. For quadratic interpolations one has to use the weak form (10).

[4] Since the parameters a_i describing the projection never enter the formulation explicitly the name virtual elements was introduced.

same approach can be found in the work of Beirão Da Veiga et al. (2015) and Chi et al. (2017) for the nonlinear case. Here we employ the same approach by starting from the weak form (10). Thus we have, by summing up all element contributions for the n_e virtual elements,

$$G(\boldsymbol{u}) = \mathop{\mathbf{A}}_{e=1}^{n_e} \left[G_\Pi(\Pi \boldsymbol{u}_h|_e) + G_{stab}(\boldsymbol{u}_h|_e - \Pi \boldsymbol{u}_h|_e) \right] . \tag{40}$$

The first part in this equation can be computed by using (39). This yields

$$G_\Pi(\Pi \boldsymbol{u}_h|_e) = \int_{\Omega_e} \left[\boldsymbol{S}(\boldsymbol{F}|_e) \cdot \frac{1}{2} \delta \boldsymbol{C}(\boldsymbol{F}|_e) - \rho_0 \hat{\boldsymbol{b}} \cdot \Pi \boldsymbol{\eta}_h \big|_e \right] d\Omega - \int_{\Gamma_e^\sigma} \hat{\boldsymbol{t}} \cdot \Pi \boldsymbol{\eta}_h \big|_e \, d\Gamma . \tag{41}$$

For more details, we refer to Wriggers et al. (2017) and Wriggers and Hudobivnik (2017).

Several approaches to stabilization can be followed. One way to formulate the stabilization part G_{stab} is to use the split in Eq. (32), see Beirão da Veiga et al. (2013). The stabilization term is then constructed as a sum over all nodes of a positive-definite function involving the remainder of (32)

$$G_{stab} = \frac{\gamma}{2} \sum_{I=1}^{n_V} \left[\boldsymbol{u}_I - \bar{\boldsymbol{H}}(\boldsymbol{x}_I) \boldsymbol{a} \right] \cdot \left[\boldsymbol{\eta}_I - \bar{\boldsymbol{H}}(\boldsymbol{x}_I) \delta \boldsymbol{a} \right] . \tag{42}$$

where γ is a stabilization parameter. In Beirão Da Veiga et al. (2015) this parameter is replaced by a term that depends on the norm of the constitutive tensor $\| \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{F}}(\boldsymbol{F}_e) \|$, and hence on the deformation of the solid. An alternative stabilization parameter $\frac{1}{4} \operatorname{tr}[\frac{\partial \boldsymbol{P}}{\partial \boldsymbol{F}}(\boldsymbol{F}_e)]$, referred to as a trace-based stabilization, was recently proposed in Chi et al. (2017). Both stabilization parameters yield a far better approximation than the constant parameter γ. However, such a term presents a challenge when linearizing in a Newton procedure.

Another stabilization approach was developed in Wriggers et al. (2017) which is based on an approach introduced in Krysl (2015) for stabilized finite elements. The essence of the approach is to introduce a new weak form $\widehat{G}$ and to define the stabilization contribution by

$$G_{stab}(\boldsymbol{u}_h|_e - \Pi \boldsymbol{u}_h|_e) = \hat{G}_\Pi(\boldsymbol{u}_h|_e) - \hat{G}(\Pi \boldsymbol{u}_h|_e) . \tag{43}$$

The second term on the right side ensures the consistency since for $\Omega_e \to 0$ also $G_{stab} \to 0$. The total weak form is now given by

$$G(\boldsymbol{u}_h) = G_\Pi(\Pi \boldsymbol{u}_h) + \widehat{G}(\boldsymbol{u}_h) - \widehat{G}_\Pi(\Pi \boldsymbol{u}_h) , \tag{44}$$

and in which for example

$$\widehat{G}(\boldsymbol{u}_h) = \sum_{e=1}^{n_e} \int_{\Omega_e} \boldsymbol{S}[\boldsymbol{F}(\boldsymbol{u}_h|_e)] \cdot \frac{1}{2}\, \delta \boldsymbol{C}[\boldsymbol{F}(\boldsymbol{u}_h|_e)] d\Omega \,. \tag{45}$$

The terms involving $\Pi \boldsymbol{u}_h$ can be integrated as (41). It remains to devise a procedure for computing the term (45) involving the displacement $\boldsymbol{u}_h|_e$ which can be found in Wriggers et al. (2017).

3.2 Contact Approach Using VEM

Due to the possibility of adding additional nodes within a virtual element contact can be treated in an efficient way. This section provides the basic ideas.

Node insertion algorithm. VEM enables us to add abitrarily new nodes to existing discretizations and elements. Thus it allows for an enforcement of contact constraints at nodal level by using the kinematics described above. The advantage is that the changes made to the element calculation are minimal. Adding a node does not change the basis or integration procedure of the virtual element method as this does not depend on the number of nodes. It only amounts to adding one term to the integration loop. Only the displacement vector and thus the stiffness matrix has to be extended by the additional degrees of freedom. This affects elements with inserted nodes but leaves the surrounding elements unchanged.[5]

Figure 9 illustrate the general procedure used to discretize the contact interface when using the virtual element method. This yield the algorithm:

1. Check nodalwise for contact using the standard search algorithms.
2. For nodes at master or slave surface that come into contact a new node will be introduced at the element boundary (arrows indicate the insertion of new nodes at master or slave surface) where contact was established, see Fig. 9.
3. These nodes are the basis for the ansatz functions needed to formulate the contact constraints using a penalty method or Lagrangian multiplier.

In order find the new node in the contact interface one has to establish the contact conditions the gap between the master and slave elements of the two bodies has to be described and computed. Since we use linear virtual elements this task leads to a closed form solution.

Let us define a projection, see Fig. 10, that denotes the closest distance of a slave node s to the master element m. This projection defines the gap g_s related to a slave node s. This orthogonal projection is computed from

[5] In general the virtual element method leads to stiffness matrices that have the same nodal degrees of freedom as finite elements. Thus VEM fits in the standard FEM framework and hence the VEM can easily be combined with standard finite elements. This can be additionally explored to create a node-to-node contact approach for contact situations with non-matching meshes that is very simple to formulate.

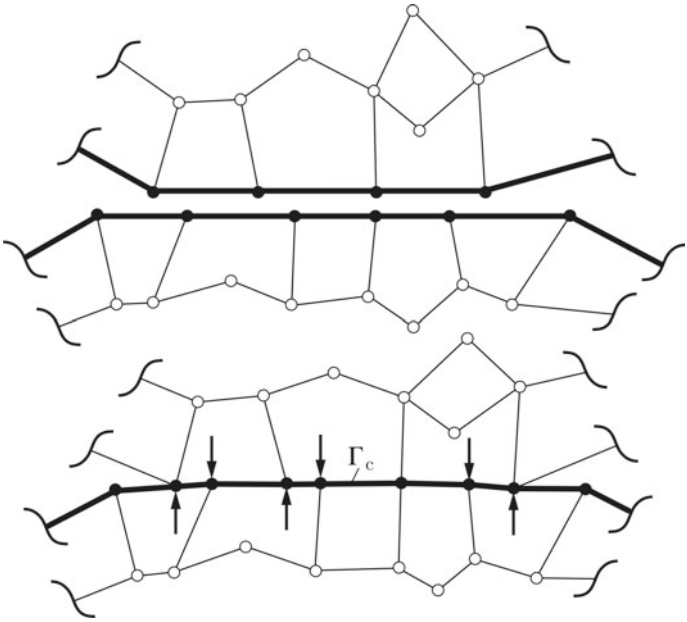

Fig. 9 Non-matching mesh and additional new contact nodes within a VEM discretization

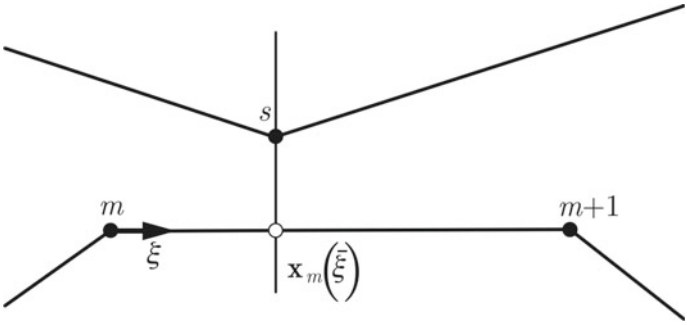

Fig. 10 Kinematics for projection slave to master

$$(\boldsymbol{x}_{m+1} - \boldsymbol{x}_m) \cdot [\boldsymbol{x}_s - \boldsymbol{x}_m(\bar{\xi})] = 0 \tag{46}$$

where $\boldsymbol{x}_i$ are the coordinates of the relevant nodes. For straight contact interfaces this equation has the analytical solution

$$\bar{\xi} = \frac{1}{l_m} \boldsymbol{a}_T^m \cdot (\boldsymbol{x}_s - \boldsymbol{x}_m) \tag{47}$$

with the tangent vector $\boldsymbol{a}_T^m = (\boldsymbol{x}_{m+1} - \boldsymbol{x}_m) / l_m$ and the length of the master segment $l_m = \|\boldsymbol{x}_{m+1} - \boldsymbol{x}_m\|$. With this relation the projection can be performed for any slave node that is possibly in contact. This projection defines a new node at the master element which then is given by

$$\boldsymbol{x}_m(\bar{\xi}) = (1 - \bar{\xi})\, \boldsymbol{x}_m + \bar{\xi}\, \boldsymbol{x}_{m+1} \,. \tag{48}$$

Once the projection is known the gap related to a slave node g_s can be computed from

$$g_s = [\boldsymbol{x}_s - \boldsymbol{x}_m(\bar{\xi})] \cdot \boldsymbol{n}_m = [\boldsymbol{x}_s - \boldsymbol{x}_m] \cdot \boldsymbol{n}_m \tag{49}$$

and used to check for contact. The vector $\boldsymbol{n}_m = \boldsymbol{e}_3 \times \boldsymbol{a}_T^m$ is the normal vector where $\boldsymbol{e}_3$ is a unit vector perpendicular to the two-dimensional plane.

If the projection is extended also to nodes of the master surface being projected onto the slave surface then the local coordinate - here we use η as surface coordinate on the slave surface - is given by

$$\bar{\eta} = \frac{1}{l_s} \boldsymbol{a}_T^s \cdot (\boldsymbol{x}_m - \boldsymbol{x}_s) \tag{50}$$

with the tangent vector $\boldsymbol{a}_T^s = (\boldsymbol{x}_{s+1} - \boldsymbol{x}_s) / l_s$ and the length of the slave segment $l_s = \|\boldsymbol{x}_{s+1} - \boldsymbol{x}_m\|$. Again the projection point leading to a new VEM surface node is then defined by

$$\boldsymbol{x}_s(\bar{\eta}) = (1 - \bar{\eta})\, \boldsymbol{x}_s + \bar{\eta}\, \boldsymbol{x}_{s+1} \,. \tag{51}$$

Once the contact nodes are established by the above process the contact interface can be discretized.[6] For this purpose ansatz functions for the gap as well as for Lagrange multipliers have to be formulated.

Hence the gap function within a segment c (contact element) will be formulated using the interpolation for the displacements at the master surface[7]

$$\boldsymbol{u}_h^1 = \sum_{I=1}^{2} N_I(\xi)\, \boldsymbol{u}_I \quad \text{with} \quad N_1(\xi) = 1 - \xi \quad \text{and} \quad N_2(\xi) = \xi \tag{52}$$

and for the slave surface

$$\boldsymbol{u}_h^2 = \sum_{I=1}^{2} N_I(\eta)\, \boldsymbol{u}_I \quad \text{with} \quad N_1(\eta) = 1 - \eta \quad \text{and} \quad N_2(\eta) = \eta \tag{53}$$

at each side of the contact area. This leads to a continuous function for the gap:

[6] For large sliding, one has to start this update algorithm at every incremental step of the Newton procedure in order to obtain the current local contact connections.

[7] This interpolation is the same as in (34) and thus consistent with the VEM formulation.

$$g_{N\,c} = \left[\sum_{I=1}^{2} N_I(\xi)\, \boldsymbol{u}_I^1 - \sum_{I=1}^{2} N_I(\eta)\, \boldsymbol{u}_I^2 \right] \cdot \boldsymbol{n}_c^1 + g_c(\xi)\,. \tag{54}$$

Here $\boldsymbol{u}_I^\alpha$ are the displacement vectors of bodies Ω^α of the nodal pair associated with the contact segment c, $\boldsymbol{n}_c^1$ is the constant normal vector related to the master surface that is defined as the surface of body Ω^1 within the contact segment c and $g_c(\xi)$ is the initial gap between the surfaces.

Since the initial gap $g_c(\xi)$ is independent of the displacement field, the variation of (54) is given by, see also (4),

$$\delta g_{N\,c} = \left[\sum_{I=1}^{2} N_I(\xi)\, \boldsymbol{\eta}_I^1 - \sum_{I=1}^{2} N_I(\eta)\, \boldsymbol{\eta}_I^2 \right] \cdot \boldsymbol{n}_c^1\,. \tag{55}$$

Lagrange multiplier formulation. To enforce the contact constraints for the frictionless case by a Lagrange multiplier the following term has to be added the weak form, see (11). Here Γ_c is the total contact area that is part of the surface of both bodies, $\lambda_N = t_N$ the Lagrange multiplier and g_N the gap function along the interface between the contacting bodies

$$\int_{\Gamma_c} \lambda_N\, \delta g_N\, d\Gamma + \int_{\Gamma_c} \delta\lambda_N\, g_N\, d\Gamma\,. \tag{56}$$

which has to be added to the weak formulation describing the motion of the two bodies (10) and (44).

Within the Lagrange multiplier method we have to approximate the multipliers along the contact interface. The Lagrange multiplier is defined at the master surface Γ_c^1. Here we use a linear function, leading to the ansatz

$$\lambda_h = \sum_{A=1}^{2} M_A(\xi)\, \lambda_A \quad \text{with} \quad M_1(\xi) = 1 - \xi \quad \text{and} \quad M_2(\xi) = \xi\,. \tag{57}$$

Approximations (55) and (57) can be now inserted in the contact integral (56) appearing in the weak form, leading to

$$\begin{aligned} \int_{\Gamma_c} \lambda_N\, \delta g_N\, d\Gamma \approx \sum_{c=1}^{n_c} \int_0^1 \left(\sum_{A=1}^{2} M_A(\xi)\, \lambda_A \right) \times \\ \left[\sum_{I=1}^{2} N_I(\xi)\, \boldsymbol{\eta}_I^1 - \sum_{I=1}^{2} N_I(\eta)\, \boldsymbol{\eta}_I^2 \right] \cdot \boldsymbol{n}_c^1\, l_c\, d\xi \end{aligned} \tag{58}$$

where n_c denotes all contact segments that are active, ξ and η are the surface coordinates of the contacting solids at the contact element c and l_c is the length of the

contact element. Furthermore the second part of (56) has to be approximated using the same ansatz. This yields

$$\int_{\Gamma_c} \delta\lambda_N \, g_N \, d\Gamma \approx \sum_{c=1}^{n_c} \int_0^1 \left(\sum_{A=1}^{2} M_A(\xi) \, \delta\lambda_A \right) \times \left[\sum_{I=1}^{2} N_I(\xi) \, \boldsymbol{u}_I^1 - \sum_{I=1}^{2} N_I(\eta) \, \boldsymbol{u}_I^2 \right] \cdot \boldsymbol{n}_c^1 \, l_c \, d\xi + g_c(\xi) \tag{59}$$

where n_c are the active contact nodes in Γ_c^h.

Since the surface coordinates ξ and η coincide in this formulation, the integrals in (57) and (58) can be evaluated explicitly for linear ansatz functions. This yields

$$\int_{\Gamma_c} \lambda_N \, \delta g_N \, d\Gamma \approx \langle \lambda_1 \, , \lambda_2 \rangle \, \frac{l_c}{6} \left\{ \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{pmatrix} \eta_{n\,1}^1 \\ \eta_{n\,2}^1 \end{pmatrix} - \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{pmatrix} \eta_{n\,1}^2 \\ \eta_{n\,2}^2 \end{pmatrix} \right\} \tag{60}$$

where $\eta_{n\,I}^{\alpha} = \boldsymbol{\eta}_I^{\alpha} \cdot \boldsymbol{n}_c^1$. With this result the matrix form of (56) can be written for one contact element c as

$$\langle \hat{\boldsymbol{\eta}}^T, \delta\hat{\boldsymbol{\lambda}}^T \rangle \begin{bmatrix} \boldsymbol{0} & \boldsymbol{C}_L \\ \boldsymbol{C}_L^T & \boldsymbol{0} \end{bmatrix} \begin{Bmatrix} \hat{\boldsymbol{u}} \\ \hat{\boldsymbol{\lambda}} \end{Bmatrix} \tag{61}$$

with

$$\hat{\boldsymbol{u}}^T = \langle u_{n\,1}^1 \, , u_{n\,2}^1 \, , u_{n\,1}^2 \, , u_{n\,2}^2 \rangle, \; \hat{\boldsymbol{\lambda}}^T = \langle \lambda_1 \, , \lambda_2 \rangle \text{ and } \boldsymbol{C}_L^T = \frac{l_c}{6} \begin{bmatrix} 2 & 1 & -2 & -1 \\ 1 & 2 & -1 & -2 \end{bmatrix}. \tag{62}$$

Penalty formulation

In the case of the penalty method, we have to discretize the first term in (12) where now $t_N = \epsilon \, g_N$ (ϵ_N is the penalty parameter). This leads with (55) to

$$\int_{\Gamma_c} \epsilon_N \, g_N \, \delta g_N \, d\Gamma = \epsilon_N \sum_{c=1}^{n_c} \int_0^1 \sum_{A=1}^{2} N_A(\xi) \, g_{N\,c\,A}(\xi)) \sum_{I=1}^{2} N_I(\xi) \delta g_{N\,c\,I}(\xi) \, l_c \, d\xi \tag{63}$$

with

$$g_{N\,c\,A} = \left[\boldsymbol{u}_A^1 - \boldsymbol{u}_A^2 \right] \cdot \boldsymbol{n}_c^1 \quad \text{and} \quad \delta g_{N\,c\,I} = \left[\boldsymbol{\eta}_I^1 - \boldsymbol{\eta}_I^2 \right] \cdot \boldsymbol{n}_c^1. \tag{64}$$

Using the notation from above, the ansatz in (63) yields a matrix formulation for one contact element c

$$\hat{\boldsymbol{\eta}}^T \, \boldsymbol{C}_P \, \hat{\boldsymbol{u}} \tag{65}$$

with

$$\boldsymbol{C}_P = \frac{l_c}{6}\begin{bmatrix} 2 & 1 & -2 & -1 \\ 1 & 2 & -1 & -2 \\ -2 & -1 & 2 & 1 \\ -1 & -2 & 1 & 2 \end{bmatrix} \tag{66}$$

3.3 Numerical Examples

In this section numerical studies are presented to demonstrate the robustness, convergence and accuracy of the method. It can be shown that the virtual element contact approach fulfills the patch test.

Example: Contacting Beams. In this example two clamped beams are considered that come into contact during the deformation. The setup is shown in Fig. 11. In this case a linear elastic constitutive relation is used within a small strain assumption.

Both beams have the same size (length $l = 20$, height $h = 1$) and overlap each other by one third of their length. The initial distance between both beams is $g_0 = 0.2$. The right beam ($E = 7000$, $\nu = 0.3$) is loaded with a force in its middle (point A). Due to its deflection it will come into contact with the stiffer beam on the left ($E = 70{,}000$, $\nu = 0.3$). In Fig. 12 the stress σ_{xx} and the deformed meshes are depicted.

The graph in Fig. 13 shows the load-displacement-curve for the loading point A and the end of the left beam (point B). Initially the contact gap is open. With increasing load the contact is established and the stiffness of the system increases.

Note that the beam on the left was discretized using rectangular $n_E = 540$ virtual elements. The beam on the right was discretized with $n_E = 500$ elements in a different

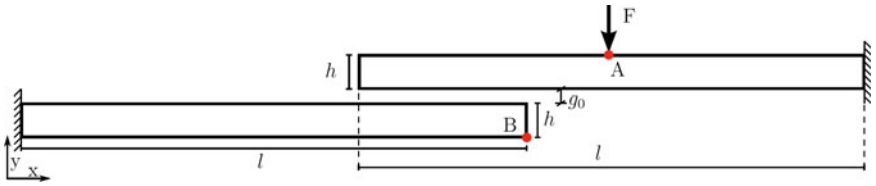

Fig. 11 Setup of two contacting beams

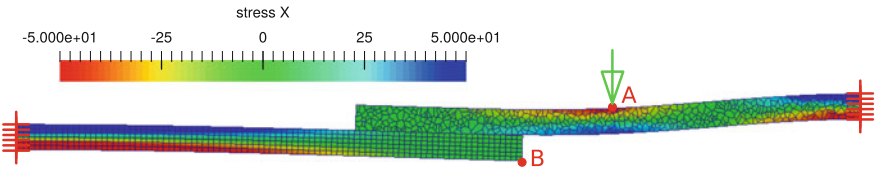

Fig. 12 Stress in x-direction of two contacting clamped beams

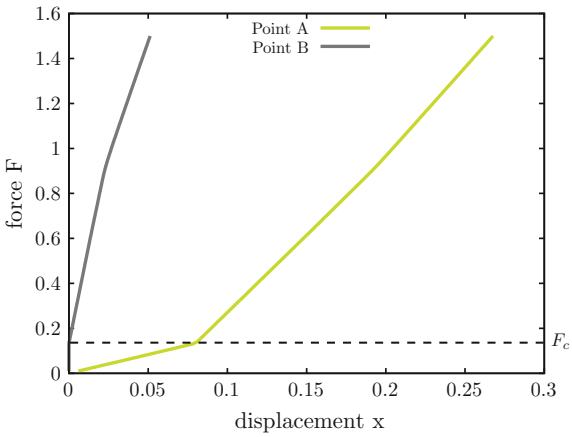

Fig. 13 Load-displacement-curve for the loading point A and the end point B of the contacted beam

fashion using a Voronoi type mesh. This takes full advantage of the possibilities of the virtual element method, allowing for complex and non-standard element shapes. This different meshing does not affect at all the contact procedure.

Example: Hertz - Large deformations. In this example we will show that the algorithm also works for nonlinear problems. Here the classical Hertz problem is investigated where both bodies are made of Neo Hookean material which leads to the constitutive relation for the second Piola-Kirchhoff stress

$$\boldsymbol{S} = \frac{\Lambda}{2}(J^2 - 1)\,\boldsymbol{C}^{-1} + \mu\,(\mathbf{1} - \boldsymbol{C}^{-1})$$

which has to be used in (40). Here the constants Λ and μ are the Lame constants.[8] The two bodies will undergo finite deformations. Only one half of the problem is discretized due to symmetry.

Figure 14 shows the problem where already small deformations have occurred. The lower body is fixed at the bottom while the upper body is loaded by a surface traction at the upper part. At the symmetry line the usual symmetry conditions are formulated which lead to fixing the horizontal displacement components at that line. Load and material parameters are selected such that finite deformations can occur.

In Fig. 15 the final deformation stage is depicted that is reached after eight load steps. The convergence behaviour in the last load step is shown in Fig. 16. It depicts the typical behaviour: in the first four increments linear convergence is observed that is related to the contact search with changing interface connections. Once the contact connection is established quadratic convergence occurs.

From these results it is clear that the virtual element method provides a good basis for treating contact problems as well in the linear as nonlinear range.

[8]The same constitutive relation is used for the stabilization term (45), however with different Lame constants.

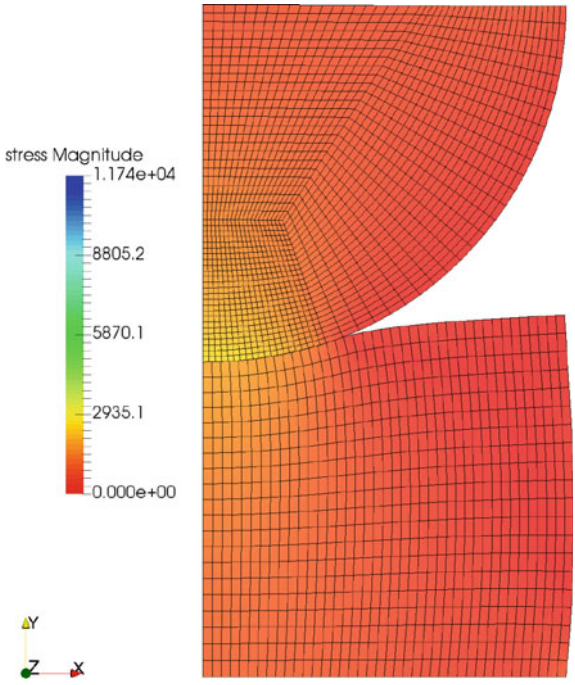

Fig. 14 Hertz problem for finite deformations

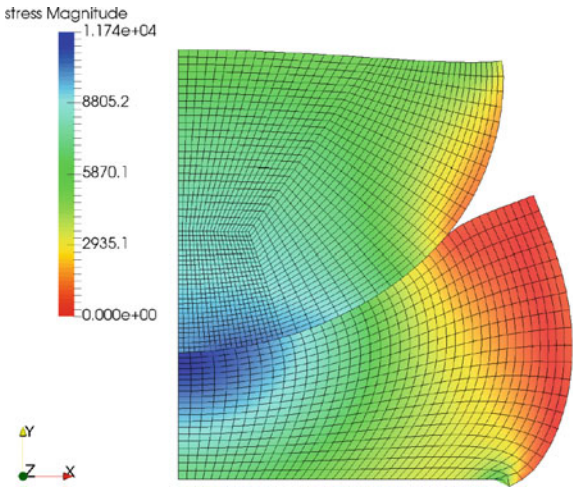

Fig. 15 Final deformation stage for Hertz problem

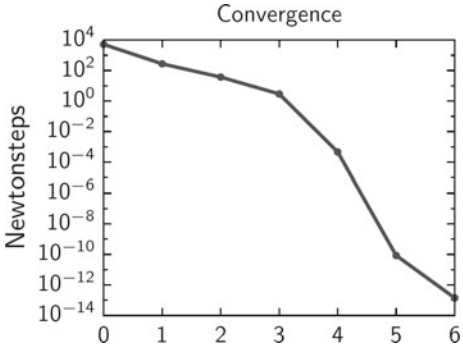

Fig. 16 Final deformation stage for Hertz problem

4 Contact Domain and Third Media Approaches

Another group of contact formulations was developed in the last years by Labra et al. (2008), Onate et al. (2008) and Oliver et al. (2009). These methods are based on a domain method that introduces a contact domain between approaching deformable bodies. In Oliver et al. (2009) and Hartmann et al. (2009) a frictional contact domain method for large deformation problems is presented. The authors formulate the contact constraints within a contact domain, that can be interpreted as a fictive intermediate region which connects potential contact surfaces. This contact domain is discretized by a non-overlapping set of patches that are interpolated from the displacement fields at the contact surfaces. Hence the contact formulation is related to a different, strain-like set of constraints for the normal and tangential gaps. Within this contact domain, classical formulations for contact constraints, like normal and tangential gaps are applied.

Another approach relies on the introduction of a third medium, see Wriggers et al. (2013). It avoids the complexity of the classical schemes related to the exact formulation and enforcement of the contact constraints. The constraints follow as a direct consequence from finite strain measures and thus are implicitly formulated. As can be seen from Fig. 17, the contacting bodies will be imbedded in a medium. This medium will have assigned specific material properties that on one hand approximate rigid body movements before contact and on the other hand account for the contact constraints whenever necessary. In this section we will introduce the basic formulation for such third medium approach that is here formulated for the frictionless case. Since the formulation differs quite a bit from the classical contact formulation we have first to introduce the basics of the continuum formulation that includes also anisotropic behaviour close to contact.

The kinematics and weak form for the contacting bodies B^{α} are already discussed in Sect. 1.2. Thus it remains only to introduce the constitutive relations that are used in the examples of this section.

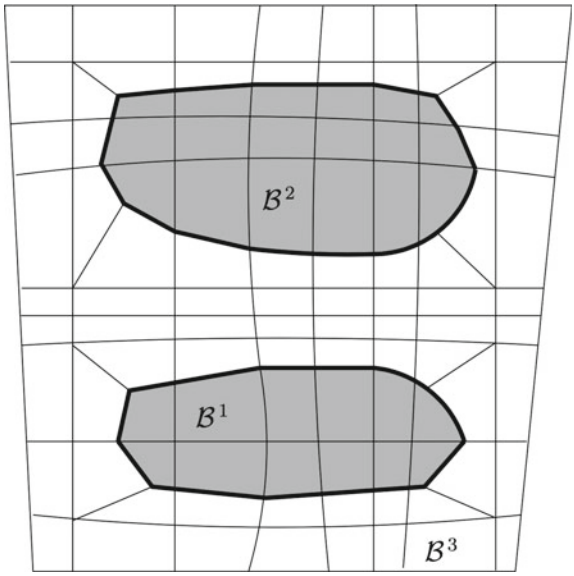

Fig. 17 Contacting bodies B^1 and B^2 imbedded in a medium B^3

4.1 Continuum Mechanics Background

The second Piola-Kirchhoff stress for body B^α follows from the strain energy function by

$$\boldsymbol{S}^\alpha = 2\frac{\partial W^\alpha}{\partial \boldsymbol{C}^\alpha} \,. \tag{67}$$

For simplicity strain energy functions of St. Venant-Kirchhoff or Neo-Hookean type are selected to model the constitutive behaviour of the solids coming in contact. The St. Venant-Kirchhoff energy has the form

$$W^\alpha = \frac{\Lambda^\alpha}{8}(I_1^2 - 6I_1 + 9) + \frac{\mu^\alpha}{4}(I_1^2 - 2I_2 - 2I_1 + 3) \,. \tag{68}$$

Here the Lame constants

$$\Lambda^\alpha = \frac{E^\alpha \nu^\alpha}{(1+\nu^\alpha)(1-2\nu^\alpha)} \qquad \text{and} \qquad \mu^\alpha = \frac{E^\alpha}{2(1+\nu^\alpha)}$$

are given in terms of the Young's modulus E^α and Poisson's ration ν^α for the two bodies. In the case of the Neo-Hookean model we use the strain energy function

$$W^\alpha = \frac{K^\alpha}{2}(J-1)^2 + \frac{\mu^\alpha}{2}(\mathrm{tr}\,\widehat{\boldsymbol{C}} - 3) \tag{69}$$

where K is the bulk modulus and $\widehat{\boldsymbol{C}} = J^{-\frac{2}{3}}\boldsymbol{C}$ is the isotropic part of the right Cauchy-Green deformation tensor which is motivated by a multiplicative split of the deformation gradient $\boldsymbol{F} = J^{\frac{1}{3}}\widehat{\boldsymbol{F}}$ into a volumetric and an unimodular part.

4.2 Continuum Formulation for the Medium

Again the weak form (10) can be formulated. This leads to

$$\int_{\mathcal{B}^M} \boldsymbol{S}^M \cdot \frac{1}{2}\delta\boldsymbol{C}^M \, dV = 0 \tag{70}$$

where $\delta\boldsymbol{C}^M$ denotes the variation of the strains within the third medium. The second Piola-Kirchhoff stress for the third medium $\mathcal{B}^M$ follows from the strain energy function by

$$\boldsymbol{S}^M = 2\frac{\partial W^M}{\partial \boldsymbol{C}^M} \, . \tag{71}$$

Within the third medium the surface tractions were neglected as well as the body forces. Thus the medium can be seen as a nonlinear spring system between the contacting bodies.

Since the properties of the third medium will change depending upon the approach of the bodies a special strain energy function has to be selected. The properties of the strain energy have to approximate the contact constraint. The idea is that a layer of the embedding medium will transport the forces between the contacting bodies and no penetration will be allowed. In case of frictionless contact the material properties have to be highly anisotropic since no tangential forces are allowed at the contact interface. Thus the strain energy of the embedding body has to be able to model

1. the change of the properties from isotropic behaviour to anisotropic behaviour between two bodies when these come close to each other.
2. the increase of the stiffness of the medium when bodies come close to each other to avoid penetration.

In general the strain energy function for the third medium W^M will consist of an isotropic and an anisotropic part where the constitutive parameters can change due to the deformation states of the bodies. Thus the strain energy function consists of an isotropic and an anisotropic part: $W^M = W^M_{iso} + W^M_{aniso}$, see Schröder (2009). Here the the isotropic part is given by

$$W^M_{iso} = a_1 I_1 + a_2 I_2 + a_3 I_3 - d \ln(\sqrt{I_3}) \tag{72}$$

with the material parameters $a_1, a_2, a_3 \geq 0$ and $d = 2a_1 + 4a_2 + 2a_3$. These have to be selected for the medium B^M where the material constants have to be much smaller

than the parameters of the bodies B^α in order to approximate rigid body motions of the bodies coming into contact. Hence the material parameters have to fulfill

$$a_{1,2,3} \ll \Lambda^\alpha \text{ and } \quad a_{1,2,3} \ll \mu^\alpha \,.$$

The anisotropic part of the strain energy function of the medium can be formulated using only one distinct direction of anisotropy in case of frictionless contact. A simple strain energy is given by

$$W^M_{aniso} = c^M \left[\frac{1}{a_4 + 1} \right] (1 - J_4)^{a_4+1} \quad \text{with} \quad J_4 = \text{tr}(\boldsymbol{C}\, \boldsymbol{M}_n) \,. \tag{73}$$

Note that this anisotropic part is only activated when the bodies come close to each other. One criterium is the volume change provided by $J =$ in the third medium. Thus for $J < 0.7$ the anisotropic function (73) is added to the isotropic part of the third medium (72). For $J \geq 0.7$ the anisotropic part is neglected: $W^M_{aniso} = 0$.

In (73) the parameter c^M is associated to the stiffness of the third medium. From assumption 2 above it can be concluded that c^M depends upon the approach (gap) g_N between bodies B^1 and B^2, see Fig. 2, and an additional material parameter a_5 which represents the basic stiffness in the preferred direction. A simple choice is provided by the function

$$c^M(g_N) = \frac{a_5}{g_N^2} \tag{74}$$

which is valid for $g_N \leq d_c$ where d_c is a threshold distance denoting that the approach is close enough to include anisotropic behaviour. For $g_N > d_c$ the constitutive parameter c^M will be selected as $c^M = 0$ and the isotropic part of the third medium will stabilize the solution of the general problem. The definition of the gap g_N is not the same as in (1) for the third medium it is provided in the next Sect. 4.3.

The interface between the two contacting bodies B^1 and B^2 can only transmit tractions in normal direction in case of frictionless contact. This is accounted in the anisotropic strain energy by assuming that the structural tensor $\boldsymbol{M}_n$ can be expressed by $\boldsymbol{M}_n = \boldsymbol{n} \otimes \boldsymbol{n}$. Here $\boldsymbol{n}$ is the contact normal that will be determined next.[9]

4.3 Kinematics at the Interface

As can be seen in Fig. 2, the approach of the bodies is related to the gap g_N and the normal $\boldsymbol{n}$. These geometrical quantities are related to the deformation $\boldsymbol{\varphi}^\alpha$ and can change. Usually the contact constraints are formulated based on the kinematical quantities gap g_N and normal vector $\boldsymbol{n}$. Within the proposed approach, however, we

[9] In case of friction two additional different structural tensors have to be defined that are associated with the tangents at the deformed surface of bodies B^α, e.g, $\boldsymbol{M}^\alpha_1 = \boldsymbol{\varphi}^\alpha_{,1} \otimes \boldsymbol{\varphi}^\alpha_{,1}$ and $\boldsymbol{M}^\alpha_2 = \boldsymbol{\varphi}^\alpha_{,2} \otimes \boldsymbol{\varphi}^\alpha_{,2}$ are defined.

want to avoid the computation of these quantities since they involve rather complex search procedures and complicated geometrical relations, especially in large deformation contact. Thus another method has to be designed that is able to detect contact without explicitly computing the gap g_N.

Since we introduced in our method the third medium that imbeds the bodies coming into contact it is natural to use the deformation state of this intermediate continuum to quantify the gap g_N and compute the normal vector $\boldsymbol{n}$ between the bodies. It is clear that during the approach of two bodies the third medium will be compressed highly. Thus the deformation state of the third medium will be a good indicator for the approach when a certain compression has already occurred in the third medium. The obvious way to obtain directional values is to compute the eigenvalues and eigenvectors of a strain measure. In this case, since we are dealing with large deformations, we solve within the third medium the eigenvalue problem of the right Cauchy Green tensor

$$(\boldsymbol{C} - \omega\,\mathbf{1})\,\boldsymbol{e} = \mathbf{0} \tag{75}$$

where ω is the principal strain and $\boldsymbol{e}$ the associated eigenvector. After solving the eigenvalue problem we can write

$$\boldsymbol{C} = \sum_{i=1}^{3} \omega_i\, \boldsymbol{e}_i \otimes \boldsymbol{e}_i\,. \tag{76}$$

In the representation the minimum eigenvalue $\omega_{min} \in \omega_i$ denotes the highest compression and the associated eigenvector $\boldsymbol{e}_{min}$ the direction of the highest compression. Now this eigenvector defines the normal direction of the interface between the two contacting bodies. Thus $\boldsymbol{n} = \boldsymbol{e}_{min}$ can be used in the anisotropic part of the strain energy function (73).[10]

While the gap is never computed directly the eigenvalue ω_{min} can be used instead to change the stiffness in the anisotropic part of the strain energy function. Thus equation (74) can be rewritten as

$$c^M(g_N) = \frac{a_5}{\omega^i_{min}}\,, \tag{77}$$

where i controls the increase behavior of the stiffness. Furthermore ω^i_{min} can be used as a switch criterium that defines the threshold at which the anisotropic strain energy function is include to model the contact constraints, thus

$$\omega_{min} \geq \varepsilon_{limit} \Rightarrow c^M(g_N) = 0 \quad \text{or} \quad \omega_{min} < \varepsilon_{limit} \Rightarrow c^M(g_N) = \frac{a_5}{\omega^i_{min}}\,. \tag{78}$$

[10]Note that in case of $\omega_{min} \to 0$ the third medium approaches a two dimensional contact interface. Hence the eigenvector $\boldsymbol{e}_{min}$ will be equivalent to the normal vector $\boldsymbol{n}$ of the contact interface which means that the new formulation converges to a classical contact setting for $\omega_{min} \to 0$.

A reasonable choice for ε_{limit} is provided by $\varepsilon_{limit} = 0.5$.

The associated finite element formulation is standard. One can use all types of finite elements for finite deformation analysis. Only the constitutive relations discussed above have to be introduced for the two bodies and the third medium. These special materials were introduced in the program FEAP, see Zienkiewicz and Taylor (2000).

4.4 Comparison to a Standard Contact Formulation

To illustrate the benefits of the proposed method we compare it to a standard contact formulation. Here we use the commercial software program ABAQUS with its predefined master-slave contact definitions to model the contact in the classical way In order to obtain comparable results, we take the Neo-Hookean material model 69 into account, which is also implemented in ABAQUS. The material parameters for the bodies are chosen as $K_1 = K_2 = 3000$ and $\mu_1 = \mu_2 = 3000$. For the medium in our formulation we apply $a_{1,2,3} = 1$, $a_4 = 23$ and $a_5 = 1$. Furthermore, the same switch criterion is given by

$$c^M(g_N) = 0 \quad \text{for} \quad \omega_{min} \geq 0.5 \quad \text{or} \quad c^M(g_N) = \frac{a_5}{\omega_{min}^2} \quad \text{for} \quad \omega_{min} < 0.5\,.$$

The setup of the contact problem and the results for the vertical displacements in the middle of the upper edge of the lower body for the proposed formulation can be found in Fig. 18. In this example, we prescribe displacements $\boldsymbol{u} = [0 \; -2.4]^T$ at the upper half of the ellipse as boundary conditions.

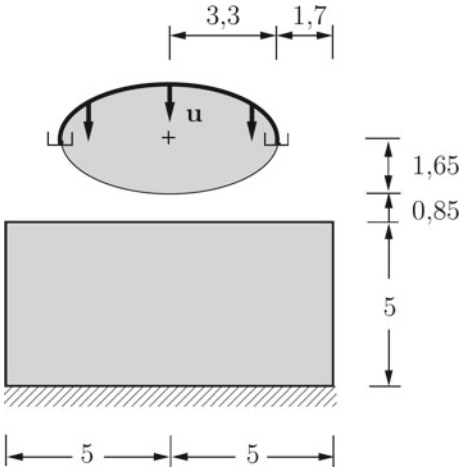

Fig. 18 Original setup of the contact problem

The initial mesh is provided in Fig. 19 and the deformed configuration can be seen in Fig. 20 which shows also the stress distribution of the vertical stresses σ_{22}.

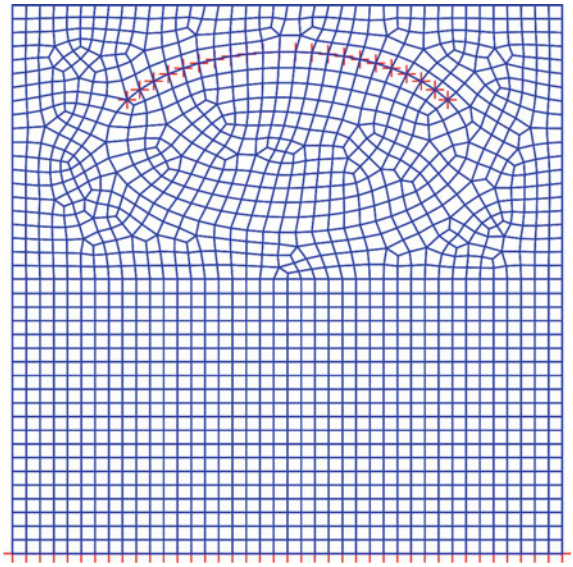

Fig. 19 Mesh in the initial configuration

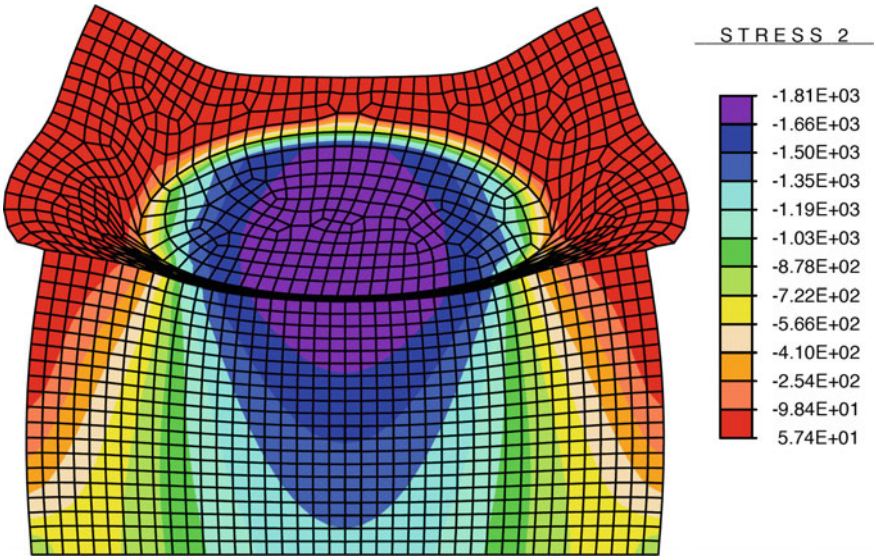

Fig. 20 Deformed configuration and vertical stress field

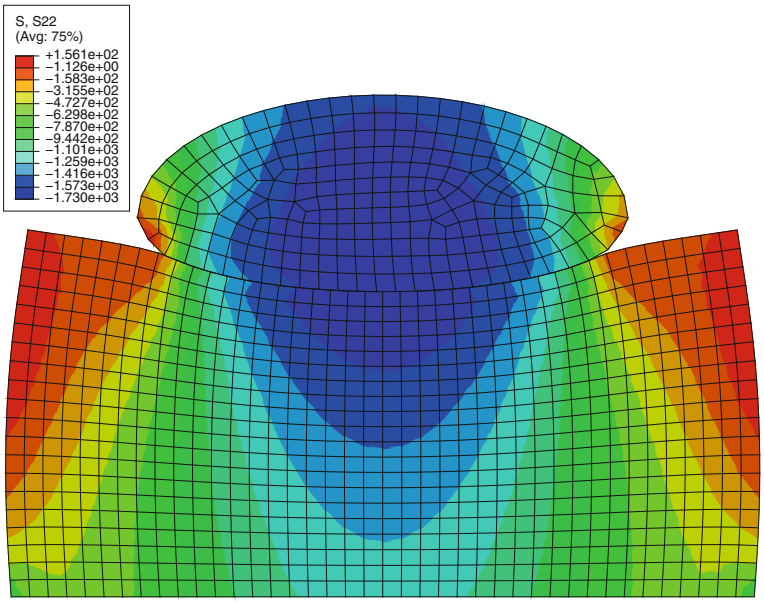

Fig. 21 Plot of the σ_{22} stress distribution for the calculation with ABAQUS

In Fig. 21 the deformed configuration and the σ_{22} stress distribution using the are depicted for the simulation using ABAQUS.

The difference between the ABAQUS simulation and the proposed formulation is around 4.6% for the maximal contact pressure. This deviation is due to the different contact approaches in connection with prescribed displacement boundary conditions. ABAQUS uses a Lagrange multiplier approach where the contact force is determined from the constraint condition. In the third medium approach the bodies do not overlap each other, moreover the medium stays between the bodies – even if it is compressed to a very small size and can therefore cause slightly higher pressures.

Of course, in this example, the mesh for the third medium can be only inserted close to the possible contact interface (when known) which would lead to a more efficient treatment and would avoid not necessary meshing.

The contact formulation using the third medium is based on the eigenvalues of the strain tensor and a constitutive equation of the medium change that changes its properties form isotropic to anisotropic close to contact. The magnitude of the eigenvalue in the medium related to the movement of the bodies define the parameters and directional properties of the intermediate medium. It has been shown that this methodology automatically detects contact within finite deformation analysis without a complex geometrical computation. Further results can be found in Wriggers et al. (2013).

References

Beirão da Veiga L., Brezzi F., Cangiani A., Manzini G., Marini L. and Russo A. Basic principles of virtual element methods. *Mathematical Models and Methods in Applied Sciences*, 23(01):199–214 (2013).

Beirão Da Veiga L., Brezzi F. and Marini L. Virtual Elements for linear elasticity problems. *SIAM Journal on Numerical Analysis*, 51(2):794–812 (2013).

Beirão da Veiga L., Brezzi F., Marini L.D. and Russo A. The Hitchhiker's Guide to the Virtual Element Method. *Mathematical Models and Methods in Applied Sciences*, 24(08):1541–1573 (2014).

Beirão Da Veiga L., Lovadina C. and Mora D. A Virtual Element Method for elastic and inelastic problems on polytope meshes. *Computer Methods in Applied Mechanics and Engineering*, 295:327–346 (2015).

Chi H., Beirão da Veiga L. and Paulino G. Some basic formulations of the virtual element method (VEM) for finite deformations. *Computer Methods in Applied Mechanics and Engineering*, 318:148–192 (2017).

Cottrell J.A., Hughes T.J.R. and Bazilevs Y. *Isogeometric Analysis*. Wiley (2009).

Dimitri R., De Lorenzis L., Scott M., Wriggers P., Taylor R. and Zavarise G. Isogeometric large deformation frictionless contact using t-splines. *Computer Methods in Applied Mechanics and Engineering*, 269:394–414 (2014).

de Lorenzis L., Temizer I., Wriggers P. and Zavarise G. A large deformation frictional contact formulation using NURBS-based isogeometric analysis. *International Journal for Numerical Methods in Engineering*, 87:1278–1300 (2011).

de Lorenzis L., Wriggers P. and Zavarise G. Isogeometric analysis of 3d large deformation contact problems with the augmented lagrangian formulation. *Computational Mechanics*, 49:1–20 (2012).

Eterovic A.L. and Bathe K.J. An interface interpolation scheme for quadratic convergence in the finite element analysis of contact problems. In *Computational Methods in Nonlinear Mechanics*, pages 703–715. Springer-Verlag, Berlin, New York (1991).

Fischer K.A. and Wriggers P. Frictionless 2d contact formulations for finite deformations based on the mortar method. *Computational Mechanics*, 36:226–244 (2005).

Fischer K.A. and Wriggers P. Mortar based frictional contact formulation for higher order interpolations using the moving friction cone. *Comput. Methods Appl. Mech. Engrg.*, 2006:5020–5036 (2006).

Franke D., Düster A., Nübel V. and Rank E. A comparison of the h-, p-, hp-, and rp-version of the FEM for the solution of the 2D Hertzian contact problem. *Computational Mechanics*, 45:513–522 (2010).

Gain A.L., Talischi C. and Paulino G.H. On the Virtual Element Method for three-dimensional linear elasticity problems on arbitrary polyhedral meshes. *Computer Methods in Applied Mechanics and Engineering*, 282:132–160 (2014).

Hartmann S., Oliver J., Weyler R., Cante J. and Hernndez J. A contact domain method for large deformation frictional contact problems. part 2: Numerical aspects. *Computer Methods in Applied Mechanics and Engineering*, 198:2607–2631 (2009).

Hesch C. and Betsch P. A mortar method for energy-momentum conserving schemes in frictionless dynamic contact problems. *Int. J. Numer. Meth. Engng.*, 77:1468–1500 (2009).

Hüeber S. and Wohlmuth B.I. Thermo-mechanical contact problems on non-matching meshes. *Comput. Methods Appl. Mech. Engrg.*, 198:1338–1350 (2009).

Hughes T.J.R., Reali A. and Sangalli G. Efficient quadrature for NURBS-based isogeometric analysis. *Comput. Methods Appl. Mech. Engrg.*, 199:301–313 (2010).

Johnson K.L. *Contact Mechanics*. Cambridge University Press (1985).

Krstulovic-Opara L., Wriggers P. and Korelc J. A C^1-continuous formulation for 3d finite deformation frictional contact. *Computational Mechanics*, 29:27–42 (2002).

Krysl P. Mean-strain eight-node hexahedron with optimized energy-sampling stabilization for large-strain deformation. *International Journal for Numerical Methods in Engineering*, 103:650–670 (2015).

Labra C., Rojek J., Onate E. and Zarate F. Advances in discrete element modelling of underground excavations. *Acta Geotechnica*, 3:317–322 (2008).

Laursen T.A. *Computational Contact and Impact Mechanics*. Springer, Berlin, New York, Heidelberg (2002).

Oliver J., Hartmann S., Cante J., Weyler R. and Hernndez J. A contact domain method for large deformation frictional contact problems. part 1: Theoretical basis. *Computer Methods in Applied Mechanics and Engineering*, 198:2591–2606 (2009).

Onate E., Idelsohn S.R., Celigueta M.A. and Rossi R. Advances in the particle finite element method for the analysis of fluid-multibody interaction and bed erosion in free surface flows. *Computer Methods in Applied Mechanics and Engineering*, 197:1777–1800 (2008).

Padmanabhan V. and Laursen T. A framework for development of surface smoothing procedures in large deformation frictional contact analysis. *Finite Elements in Analysis and Design*, 37:173–198 (2001).

Piegl L. and Tiller W. *The NURBS Book*. Springer, Berlin Heidelberg New York, 2nd edition (1996).

Pietrzak G. and Curnier A. Large deformation frictional contact mechanics: continuum formulation and augmented lagrangean treatment. *Computer Methods in Applied Mechanics and Engineering*, 177:351–381 (1999).

Popp A., Gee M.W. and Wall W.A. A finite deformation mortar contact formulation using a primal-dual active set strategy. *Int. J. Numer. Meth. Engng.*, 79:1354–1391 (2009).

Puso M.A., Laursen T.A. and j. Solberg. A segment-to-segment mortar contact method for quadratic elements and large deformations. *Comput. Methods Appl. Mech. Engrg.*, 197:555–566 (2008).

Schröder J. Anisotropic polyconvex energies. In J. Schröder, editor, *Polyconvex Analysis*, pages 1–53. CISM, Springer, Wien (2009). 62.

Temizer I., Wriggers P. and Hughes T.J.R. Contact treatment in isogeometric analysis with NURBS. *Computer Methods in Applied Mechanics and Engineering*, 200:1100–1112 (2011).

Temizer I., Wriggers P. and Hughes T.J.R. Three-dimensional mortar-based frictional contact treatment in isogeometric analysis with NURBS. *Computer Methods in Applied Mechanics and Engineering*, 209-211:115–128 (2012).

Tur M., Fuenmayor F.J. and Wriggers P. A mortar-based frictional contact formulation for large deformations using Lagrange multipliers. *Comput. Methods Appl. Mech. Engrg.*, 198:2860–2873 (2009).

Wriggers P. *Computational Contact Mechanics*. 2nd ed., Springer, Berlin, Heidelberg, New York (2006).

Wriggers P. *Nonlinear Finite Elements*. Springer, Berlin, Heidelberg, New York (2008).

Wriggers P. and Hudobivnik B. A low order virtual element formulation for finite elasto-plastic deformations. *Computer Methods in Applied Mechanics and Engineering*, 327:459–477 (2017).

Wriggers P., Krstulovic-Opara L. and Korelc J. Smooth C^1- interpolations for two-dimensional frictional contact problems. *International Journal for Numerical Methods in Engineering*, 51:1469–1495 (2001).

Wriggers P., Reddy B., Rust W. and Hudobivnik B. Efficient virtual element formulations for compressible and incompressible finite deformations. *Computational Mechanics*, 60:253–268 (2017).

Wriggers P., Rust W. and Reddy B. A virtual element method for contact. *Computational Mechanics*, 58:1039–1050 (2016).

Wriggers P., Schröder J. and Schwarz A. A finite element method for contact using a third medium. *Computational Mechanics*, 52:837–847 (2013).

Ziefle M. and Nackenhorst U. Numerical techniques for rolling rubber wheels: treatment of inelastic material properties and frictional contact. *Computational Mechanics*, 42:337–356 (2008).

Zienkiewicz O.C. and Taylor R.L. *The Finite Element Method*, volume 1. Butterworth-Heinemann, Oxford, UK, 5th edition (2000).

Finite Wear and Soft Elasto-Hydrodynamic Lubrication: Beyond the Classical Frictional Contact of Soft Solids

Stanisław Stupkiewicz

Abstract Two classes of contact problems are discussed, namely *finite-wear* and *soft-EHL* problems, which go beyond the classical framework of frictional contact problems. The focus is on the finite-deformation effects and on the computational strategies adequate for the modelling of those problems. By finite wear we mean here the class of contact and wear problems in which finite deformations and finite shape changes due to wear are allowed. The soft-EHL regime of hydrodynamic lubrication is encountered in the case of lubricated contact of compliant solids, such as elastomers or soft tissues, when a relatively low hydrodynamic pressure suffices to significantly deform the solid. In each case, the respective continuum formulation is first introduced, followed by the description of the finite-element treatment and by representative numerical examples.

1 Introduction

Contact is usually modelled by considering only two most important interaction modes, i.e. by enforcing the non-penetration condition and by introducing friction forces. The non-penetration condition imposes a unilateral constraint on the relative motion of the contacting bodies in the direction normal to the contact surface, while friction is associated with the relative motion in the tangential direction. Despite the severe complexity of tribological interactions, friction is most frequently modelled by the classical Amonton–Coulomb friction model, which involves only one parameter, the friction coefficient, even if more advanced friction models exist and could, in principle, be applied in relevant situations, provided the model parameters are reliably determined from experiment or from micromechanical considerations.

It seems that continuum formulations of frictional contact problems have already reached a considerable level of maturity, and efficient computational techniques exist for this class of problems (Laursen 2002; Wriggers 2006). In this chapter, we

S. Stupkiewicz (✉)
Institute of Fundamental Technological Research, Polish Academy of Sciences,
Pawińskiego 5B, 02–106, Warsaw, Poland
e-mail: sstupkie@ippt.pan.pl

A. Popp and P. Wriggers (eds.), *Contact Modeling for Solids and Particles*,
CISM International Centre for Mechanical Sciences 585,
https://doi.org/10.1007/978-3-319-90155-8_3

discuss two classes of contact problems that go beyond the classical framework discussed above. Specifically, we focus on *wear* and on *elasto-hydrodynamic lubrication (EHL)*, in both cases with full account of *finite deformations*. Clearly, both wear and EHL have already been subject of intense research, but the finite-deformation effects (including finite shape changes in the case of wear problems) are significantly less recognized.

Finite deformations are typical for *soft solids* such as polymers, including rubber-like materials, soft tissues, some biomaterials, and others. Focusing on contact interactions, we shall treat all those materials as hyperelastic solids, thus neglecting their complex constitutive behaviour that may involve viscoelasticity, history-dependent behaviour, multiphysics couplings, etc. This is admissible because contact formulations and their computational treatment are essentially independent of the bulk material response. Hence, the contact techniques developed can be combined with virtually any solid model, hyperelasticity being the simplest one in the finite deformation regime.

The first part of this chapter is concerned with modelling of *finite wear* problems. Wear is a process of material removal from a surface that is subjected to frictional contact interaction. Wear processes are usually slow, and thus noticeable effects result from repeated contacts and accumulation of wear over a long period. By *finite wear* we mean here a general class of wear problems in which finite deformations are allowed as well as finite shape changes due to wear (Lengiewicz and Stupkiewicz 2012). The approach adopted here for the modelling of progressive wear belongs to the class of incremental solution strategies. An overview and discussion of the relevant computational strategies is presented in Sect. 4.5. Note that wear is a very complex process with several very distinct mechanisms. The activity of the individual wear mechanisms heavily depends on the materials, on surface properties of the contact pair, as well as on the actual contact conditions. The related aspects of constitutive modelling of wear are not discussed here, and we adopt the classical Archard wear law (Archard 1953).

The second part of this chapter is concerned with modelling of *soft-EHL* problems. Contact in the EHL regime occurs when the contacting surfaces are fully separated by the fluid and when the hydrodynamic pressure in the lubricant film is sufficiently high to cause significant elastic deflections of one or both contacting bodies. The EHL theory (Dowson and Higginson 1977; Hamrock et al. 2004) is a well developed theory with classical applications such as gears and rolling-contact bearings, which belong to the class of so-called hard-EHL problems. There is, however, a growing interest in the soft-EHL regime in which the pressure is relatively low, but the elastic deflections are significant because one or both contacting bodies are highly compliant. At the same time, the pressure is not high enough to cause significant increase of lubricant viscosity (on the contrary, the piezoviscous effect is crucial in the hard-EHL problems).

The chapter is organized as follows. The standard formulation of the finite-deformation, finite-slip frictional contact problem is briefly described in Sect. 2 as a background for further developments. In Sect. 3, the Archard-type wear law is consistently formulated in the finite-deformation framework. The continuum formulation

of the finite-wear problem is then provided in Sect. 4 followed by the discussion of the time-integration schemes and computational strategies for modelling of progressive wear. Selected illustrative numerical examples are reported in Sect. 5.

The second part of the chapter, Sects. 6, 7 and 8, is devoted to the soft-EHL problems at finite deformation. Section 6 introduces the Reynolds equation which is the basic tool for the modelling of the lubricant flow in the thin channel between the contacting surfaces. Formulation of the soft-EHL problem is provided in Sect. 7. In particular, the EHL couplings are discussed, including the non-standard coupling that results from the finite-deformation effects, and the finite-element treatment is commented briefly. Finally, illustrative numerical examples are provided in Sect. 8.

2 Finite-Deformation Frictional Contact Problem

Presented below is the standard formulation of the frictional contact problem at finite deformation and finite slip. The formulation is based on the master–slave approach and on the notion of the closest-point projection. For the details and for a broader overview, see the monographs by Laursen (2002) and Wriggers (2006) and references cited therein.

Consider two hyperelastic bodies $\mathcal{B}^{(i)}$, $i = 1, 2$, that occupy domains $\Omega^{(i)}$ in the reference configuration. The boundary of $\Omega^{(i)}$ is divided into three non-overlapping parts: displacements and tractions are prescribed on $\Gamma_u^{(i)}$ and $\Gamma_t^{(i)}$, respectively, while $\Gamma_c^{(i)}$ is the potential contact surface. Deformation of each body is described by the corresponding deformation mapping $\boldsymbol{\varphi}^{(i)}$,

$$\mathbf{x}^{(i)} = \boldsymbol{\varphi}^{(i)}(\mathbf{X}^{(i)}, t), \tag{1}$$

where $\mathbf{X}^{(i)} \in \Omega^{(i)}$, $\mathbf{x}^{(i)} \in \omega^{(i)}$, and $\omega^{(i)} = \boldsymbol{\varphi}^{(i)}(\Omega^{(i)}, t)$ denotes the current configuration.

One of the contact surfaces, say $\Gamma_c^{(1)}$, is selected as the *slave* surface, and the contact pair is defined by projecting a point $\mathbf{x}^{(1)}$ of the deformed slave surface $\gamma_c^{(1)} = \boldsymbol{\varphi}^{(1)}(\Gamma_c^{(1)}, t)$ onto the deformed *master* surface $\gamma_c^{(2)} = \boldsymbol{\varphi}^{(2)}(\Gamma_c^{(2)}, t)$. The projection point is denoted by $\bar{\mathbf{x}}^{(2)}$. Let us introduce parameterization of the master surface $\gamma_c^{(2)}$ by convective coordinates $\boldsymbol{\xi} = \{\xi^1, \xi^2\}$ so that we have $\bar{\mathbf{x}}^{(2)} = \mathbf{x}^{(2)}(\bar{\boldsymbol{\xi}})$, and $\bar{\boldsymbol{\xi}} = \{\bar{\xi}^1, \bar{\xi}^2\}$ are the coordinates of the projection point.

The basic kinematic contact variables are the normal gap g_N and the sliding velocity $\mathbf{v}_T$ that are defined as follows:

$$g_N = (\mathbf{x}^{(1)} - \bar{\mathbf{x}}^{(2)}) \cdot \mathbf{n}, \qquad \mathbf{v}_T = \dot{\bar{\xi}}^{\alpha} \boldsymbol{\tau}_{\alpha}, \tag{2}$$

where $\mathbf{n} = \mathbf{n}^{(2)}$, the unit outer normal to the master surface, is adopted as the normal of the contact pair, $\boldsymbol{\tau}_{\alpha} = \partial \mathbf{x}^{(2)} / \partial \xi^{\alpha}$, $\alpha = 1, 2$, is the tangent basis, and repeated indices are implicitly summed over. Further, the spatial (Cauchy) traction vector $\mathbf{t} = \mathbf{t}^{(2)}$ is

adopted as the contact traction that is decomposed into the normal and tangential components t_N and $\mathbf{t}_T$, respectively,

$$\mathbf{t} = t_N \mathbf{n} + \mathbf{t}_T, \qquad t_N = \mathbf{t} \cdot \mathbf{n}, \qquad \mathbf{t}_T = t_{T\alpha} \boldsymbol{\tau}^\alpha, \tag{3}$$

where $\mathbf{t}^{(2)} = \boldsymbol{\sigma}^{(2)} \mathbf{n}$, $\boldsymbol{\sigma}^{(2)}$ is the Cauchy stress, and $\boldsymbol{\tau}^\alpha$ is the cobasis, such that $\boldsymbol{\tau}^\alpha \cdot \boldsymbol{\tau}_\beta = \delta^\alpha_\beta$, where δ^α_β is the Kronecker delta.

The kinematic variables g_N and $\mathbf{v}_T$ and the contact tractions t_N and $\mathbf{t}_T$ are related by the contact constraints, which can be interpreted as a kind of constitutive equations. Specifically, the normal interaction is governed by the unilateral contact condition,

$$g_N \geq 0, \qquad t_N \leq 0, \qquad g_N t_N = 0, \tag{4}$$

and the tangential interaction is assumed to be governed by the Coulomb friction law,

$$\|\mathbf{t}_T\| + \mu t_N \leq 0, \qquad \|\mathbf{v}_T\| \mathbf{t}_T = \mathbf{v}_T \|\mathbf{t}_T\|, \qquad \|\mathbf{v}_T\| (\|\mathbf{t}_T\| + \mu t_N) = 0. \tag{5}$$

Equilibrium of the two-body system is written in the form of the following virtual work principle,

$$G(\boldsymbol{\varphi}, \delta\boldsymbol{\varphi}) = G_1(\boldsymbol{\varphi}^{(1)}, \delta\boldsymbol{\varphi}^{(1)}) + G_2(\boldsymbol{\varphi}^{(2)}, \delta\boldsymbol{\varphi}^{(2)}) + G_c(\boldsymbol{\varphi}, \delta\boldsymbol{\varphi}) = 0 \qquad \forall\, \delta\boldsymbol{\varphi}, \tag{6}$$

where $\boldsymbol{\varphi} = \{\boldsymbol{\varphi}^{(1)}, \boldsymbol{\varphi}^{(2)}\}$, and the virtual displacements $\delta\boldsymbol{\varphi}^{(i)}$ (test functions) vanish on $\Gamma_u^{(i)}$. Here, G_i is defined individually for each body and denotes the virtual work of internal and external forces, excluding the contact forces, thus

$$G_i(\boldsymbol{\varphi}^{(i)}, \delta\boldsymbol{\varphi}^{(i)}) = \int_{\Omega^{(i)}} \mathbf{P}^{(i)} \cdot \operatorname{Grad} \delta\boldsymbol{\varphi}^{(i)} \mathrm{d}V - \int_{\Gamma_t^{(i)}} \mathbf{T}^{*(i)} \cdot \delta\boldsymbol{\varphi}^{(i)} \mathrm{d}S, \tag{7}$$

where $\mathbf{P}^{(i)}$ is the first Piola–Kirchhoff stress, $\mathbf{T}^{*(i)}$ is the surface traction prescribed on the boundary $\Gamma_t^{(i)}$, and Grad is the gradient operator relative to the reference configuration. The virtual work G_c of the contact forces takes the following form

$$G_c(\boldsymbol{\varphi}, \delta\boldsymbol{\varphi}) = \int_{\Gamma_c^{(1)}} (T_N \delta g_N + T_{T\alpha} \delta\bar{\xi}^\alpha) \mathrm{d}S. \tag{8}$$

The contact contribution is here integrated over the undeformed slave surface $\Gamma_c^{(1)}$. The *nominal* contact tractions T_N and $\mathbf{T}_T$ have thus been introduced,

$$T_N = j^{(1)} t_N, \qquad \mathbf{T}_T = j^{(1)} \mathbf{t}_T, \tag{9}$$

such that T_N and $\mathbf{T}_T$ refer to the unit area in the undeformed configuration $\Omega^{(1)}$ of the slave body. Here, $j^{(1)}$ is the area transformation factor of the slave surface so that $\mathrm{d}s^{(1)} = j^{(1)}\mathrm{d}S^{(1)}$.

The virtual work principle (6) constitutes the basis of the finite-element treatment. It must be complemented by a suitable regularization technique in order to enforce the contact conditions (4) and (5). In the examples reported below, the augmented Lagrangian method is used for that purpose (Alart and Curnier 1991; Pietrzak and Curnier 1999), see Lengiewicz et al. (2011) for the details of the respective AD-based formulation and finite-element implementation.

3 Archard-Type Wear Law

3.1 Nominal and Spatial Wear Rate

Wear is a process of removal of material from a solid surface subjected to a contact interaction. In a continuum description, the wear rate is defined as the volume (or mass) removed per unit area and unit time. Once finite deformations of the contacting bodies are allowed, the notions of volume and area must refer to a specific configuration of the body, and hence a nominal and spatial wear rate can be defined.

The *nominal wear rate* $\dot{W}^{(i)}$ refers to the undeformed configuration $\Omega^{(i)}$ and is defined in terms of the respective elementary volume $\mathrm{d}V^{(i)}$ and surface area $\mathrm{d}S^{(i)}$, thus

$$\dot{W}^{(i)}\mathrm{d}t = \frac{\mathrm{d}V^{(i)}}{\mathrm{d}S^{(i)}} = \frac{1}{\varrho_0^{(i)}}\frac{\mathrm{d}m^{(i)}}{\mathrm{d}S^{(i)}}, \tag{10}$$

where $\varrho_0^{(i)}$ is the mass density in the undeformed configuration. Similarly, the *spatial wear rate* $\dot{w}^{(i)}$ refers to the current configuration $\omega^{(i)}$,

$$\dot{w}^{(i)}\mathrm{d}t = \frac{\mathrm{d}v^{(i)}}{\mathrm{d}s^{(i)}} = \frac{1}{\varrho^{(i)}}\frac{\mathrm{d}m^{(i)}}{\mathrm{d}s^{(i)}}, \tag{11}$$

where $\mathrm{d}v^{(i)}$ and $\mathrm{d}s^{(i)}$ are the corresponding elementary volume and surface area, respectively, and $\varrho^{(i)}$ is the mass density in the current configuration. The following transformation rule applies to the two wear rates,

$$j^{(i)}\dot{w}^{(i)} = J^{(i)}\dot{W}^{(i)}, \tag{12}$$

where $j^{(i)} = \mathrm{d}s^{(i)}/\mathrm{d}S^{(i)}$ is the area transformation factor that follows from the Nanson's formula ($\mathbf{n}\mathrm{d}s = J\mathbf{F}^{-T}\mathbf{N}\mathrm{d}S$), and $J^{(i)} = \mathrm{d}v^{(i)}/\mathrm{d}V^{(i)} = \det\mathbf{F}^{(i)}$ is the determinant of the deformation gradient $\mathbf{F}^{(i)} = \partial\mathbf{x}^{(i)}/\partial\mathbf{X}^{(i)}$.

3.2 Archard Wear Law at Finite Deformation

The classical wear law of Archard (1953) is adopted here, however, the framework is general, and other wear laws can equally be used. In the original form of the Archard law, the wear volume is assumed to be proportional to the normal force and sliding distance. In the continuum formulation, this corresponds to the wear rate being proportional to the normal pressure and sliding velocity. Assuming that the Coulomb friction law holds, the normal contact pressure is proportional to the sliding friction stress, hence the Archard law can be equivalently expressed in terms of the latter.

Now, the product of the (spatial) friction stress $\mathbf{t}_T$ and sliding velocity $\mathbf{v}_T$ is recognized as the frictional dissipation rate density, namely

$$\dot{d} = \mathbf{t}_T \cdot \mathbf{v}_T, \qquad \dot{D}^{(i)} = j^{(i)} \dot{d}. \tag{13}$$

Here, $\dot{d}$ is the spatial density of frictional dissipation rate that is referred to the area in the current configuration $\omega^{(i)}$, while $\dot{D}^{(i)}$ is the nominal density that is referred to the area in the undeformed configuration $\Omega^{(i)}$, hence the area transformation factor in Eq. $(13)_2$.

The Archard wear law can now be formulated in terms of the frictional dissipation rate, thus

$$\dot{W}^{(i)} = K^{(i)} \dot{D}^{(i)}, \tag{14}$$

where $K^{(i)}$ is the wear coefficient. In this formulation, wear volume is proportional to the energy dissipated due to friction, which provides an energetic interpretation of the Archard wear law (cf. Mróz and Stupkiewicz 1994; Fouvry et al. 1996; Ramalho and Miranda 2006). This formulation provides also a natural way to generalize the Archard law to the case of more complex friction laws, including anisotropic friction (e.g., Mróz and Stupkiewicz 1994).

Alternatively, proportionality between the spatial quantities can be postulated, thus leading to the Archard wear law in the following form:

$$\dot{w}^{(i)} = k^{(i)} \dot{d}. \tag{15}$$

By applying the transformation rule (12), the corresponding nominal form is the following:

$$\dot{W}^{(i)} = \frac{1}{J^{(i)}} k^{(i)} \dot{D}^{(i)}, \tag{16}$$

which is *not* equivalent to the spatial one (14). In fact, Eqs. (14) and (16) imply the transformation relationship $k^{(i)} = J^{(i)} K^{(i)}$ between the wear coefficients $K^{(i)}$ and $k^{(i)}$, so that if one of them is a constant then the other one is not a constant, as it depends on the deformation in the subsurface layer. The difference between the two forms vanishes when the material is incompressible ($J^{(i)} = 1$) or when the deformation is small.

Considering that the wear coefficients may, in principle, depend on contact variables, e.g., on the contact pressure, sliding velocity, temperature, etc., the choice between the nominal or spatial form of the Archard wear law is merely free. In particular, there is no experimental evidence that would justify the choice of one or the other form. In the following, the nominal form (14) of the Archard wear law is chosen because, in practice, the wear volume would rather be measured in the undeformed configuration. The nominal wear rate can thus be considered a measurable quantity. Secondly, the computational treatment of the spatial form (16) is more involved because the nominal wear rate $\dot{W}^{(i)}$, which is used in the computational framework, as discussed below, depends then not only on the contact quantities, but also on the deformation in the bulk material (through $J^{(i)}$).

Note that, in an alternative approach (Dragon-Louiset 2001; Peigney 2004; Stolz 2007), the wear criterion is formulated in terms of the thermodynamic driving force for the propagation of a damage interface within the contact subsurface layer. In that case, evaluation of the wear criterion also involves the stresses or strains in the bulk material.

Referring to the finite-element treatment based on the master–slave approach discussed in Sect. 2, we note that the nominal wear rate of the slave surface can be directly determined in terms of the nominal friction traction $\mathbf{T}_T$, namely

$$\dot{W}^{(1)} = K^{(1)}\dot{D}^{(1)}, \qquad \dot{D}^{(1)} = j^{(1)}\dot{d} = j^{(1)}t_{T\alpha}\dot{\bar{\xi}}^{\alpha} = T_{T\alpha}\dot{\bar{\xi}}^{\alpha}. \tag{17}$$

The nominal wear rate of the master surface is then given by

$$\dot{W}^{(2)} = \frac{j^{(2)}}{j^{(1)}}K^{(2)}\dot{D}^{(1)} = \frac{j^{(2)}}{j^{(1)}}\dot{W}^{(2)}_{*}, \qquad \dot{W}^{(2)}_{*} = K^{(2)}\dot{D}^{(1)}, \tag{18}$$

where $\dot{W}^{(2)}_{*}\mathrm{d}t$ can be interpreted as the incremental wear volume of the master surface per unit area of the undeformed slave surface. It is recalled that the wear rates $\dot{W}^{(1)}$ and $\dot{W}^{(2)}$ correspond to the contact pair $(\mathbf{x}^{(1)}, \bar{\mathbf{x}}^{(2)})$ that is defined by the closest-point projection at the current time instant, as discussed in Sect. 2.

4 Finite-Wear Problem

4.1 Finite-Wear Kinematics: Three Configurations

In addition to finite deformation, the two contacting bodies $\mathcal{B}^{(i)}$ are now assumed to undergo finite shape changes due to wear at the contact interface. It is thus convenient to introduce three configurations of the body $\mathcal{B}^{(i)}$: the initial configuration $\hat{\Omega}^{(i)}$, the time-dependent undeformed (worn) configuration $\Omega^{(i)}$, and the current (deformed) configuration $\omega^{(i)}$, cf. Fig. 1. The *shape transformation mapping* $\boldsymbol{\Psi}^{(i)}$ is introduced to describe the shape change due to wear, while, as before, the deformation is described

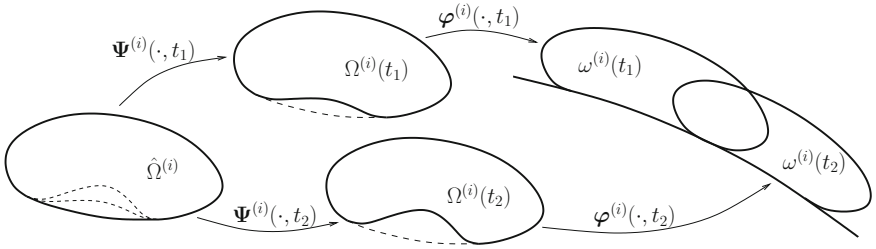

Fig. 1 The three configurations $\hat{\Omega}^{(i)}$, $\Omega^{(i)}$ and $\omega^{(i)}$ at two time instants t_1 and $t_2 > t_1$, shown for one body only (reproduced from Lengiewicz and Stupkiewicz 2012)

by the deformation mapping $\boldsymbol{\varphi}^{(i)}$, thus

$$\mathbf{X}^{(i)} = \boldsymbol{\Psi}^{(i)}(\hat{\mathbf{X}}^{(i)}, t), \qquad \mathbf{x}^{(i)} = \boldsymbol{\varphi}^{(i)}(\mathbf{X}^{(i)}, t), \qquad t \in [0, T], \tag{19}$$

where $\hat{\mathbf{X}}^{(i)} \in \hat{\Omega}^{(i)}$, $\mathbf{X}^{(i)} \in \Omega^{(i)}$ and $\mathbf{x}^{(i)} \in \omega^{(i)}$. For $t > 0$, the initial configuration $\hat{\Omega}^{(i)}$ plays the role of a fixed referential domain for the time-dependent undeformed configuration $\Omega^{(i)}$. If the initial configuration $\hat{\Omega}^{(i)}$ is a stress-free (natural) configuration then the undeformed configuration $\Omega^{(i)}$ is also stress-free.

The finite wear problem at hand comprises two subproblems. The basic unknown in the shape-evolution subproblem is the shape transformation mapping $\boldsymbol{\Psi}^{(i)}$ which is driven by the wear rate. The wear rate results from the deformation subproblem in which the basic unknown is the deformation mapping $\boldsymbol{\varphi}^{(i)}$ that is governed by the equilibrium equation and constitutive relations, along with boundary and contact conditions. As the undeformed configuration $\Omega^{(i)}$, which plays the role of a material reference configuration, evolves in time due to wear, the deformation problem is not a standard one. However, it can be transformed into a standard frictional contact problem, such as that discussed in Sect. 2, by introducing separation of time scales, see Sect. 4.2 below.

As mentioned, the time evolution of the undeformed configuration $\Omega^{(i)}$ results from wear. It is governed by the following relationship, the *shape evolution law*, that links the nominal wear rate $\dot{W}^{(i)}$, see Sect. 3.1, and the time derivative of the shape evolution mapping $\boldsymbol{\Psi}^{(i)}$,

$$\dot{\boldsymbol{\Psi}}^{(i)} \cdot \mathbf{N}^{(i)} = \begin{cases} -\dot{W}^{(i)} & \text{on } \Gamma_c^{(i)}, \\ 0 & \text{on } \partial\Omega^{(i)} \setminus \Gamma_c^{(i)}, \end{cases} \tag{20}$$

where $\partial\Omega^{(i)}$ is the boundary of $\Omega^{(i)}$, $\Gamma_c^{(i)}$ is the potential contact surface, and $\mathbf{N}^{(i)}$ is the unit outer normal in the undeformed configuration $\Omega^{(i)}$. Equation (20) is formulated in the undeformed configuration $\Omega^{(i)}$, i.e., for all $\mathbf{X}^{(i)} \in \partial\Omega^{(i)}$. Accordingly, $\dot{\boldsymbol{\Psi}}^{(i)}$ denotes here $\dot{\boldsymbol{\Psi}}^{(i)} = \dot{\boldsymbol{\Psi}}^{(i)}(\hat{\mathbf{X}}^{(i)}(\mathbf{X}^{(i)}, t), t)$, in view of the one-to-one correspondence between $\mathbf{X}^{(i)}$ and $\hat{\mathbf{X}}^{(i)}$ that is imposed by the shape transformation mapping $\boldsymbol{\Psi}^{(i)}$.

In order to make the notation compact, the above convention is followed below whenever it is convenient and unambiguous.

Note that the shape evolution law (20) prescribes only the normal component of $\dot{\boldsymbol{\Psi}}^{(i)}$ on the boundary $\partial\Omega^{(i)}$. The tangential component is here free, and so is the distribution of $\dot{\boldsymbol{\Psi}}^{(i)}$ in the interior of $\Omega^{(i)}$.

4.2 Separation of Time Scales

The wear process is usually very slow compared to the time scale of the deformation problem. The rate of change of the worn configuration $\Omega^{(i)}$ is thus negligible at this time scale. In other words, the wear rate is negligible compared to the sliding velocity, for instance. It is thus the accumulation of wear over a long time period that leads to a significant change of the worn configuration and to a significant variation of contact conditions.

Let us thus introduce *two time scales*, namely the τ-scale of the deformation problem and the t-scale of the shape changes due to wear. Keeping the shape transformation mapping $\boldsymbol{\Psi}^{(i)}$ unaltered, the deformation mapping $\boldsymbol{\varphi}^{(i)}$ is now expressed in the following form,

$$\mathbf{X}^{(i)} = \boldsymbol{\Psi}^{(i)}(\hat{\mathbf{X}}^{(i)}, t), \quad \mathbf{x}^{(i)} = \boldsymbol{\varphi}_t^{(i)}(\mathbf{X}^{(i)}, \tau), \quad t \in [0, T], \quad \tau \in [t, t + \Delta\tau], \tag{21}$$

where $\mathbf{X}^{(i)} \in \Omega_t^{(i)}$, $\Omega_t^{(i)} = \boldsymbol{\Psi}^{(i)}(\hat{\Omega}^{(i)}, t)$, and $\Delta\tau$ is a characteristic or representative time of the deformation problem, for instance, one cycle of a cyclic loading program. As long as t and τ are varied simultaneously, the deformation mapping $\boldsymbol{\varphi}_t^{(i)}(\mathbf{X}^{(i)}, \tau)$ introduced above is equivalent to that specified by Eq. $(19)_2$.

However, the *separation* of the two time scales can now be assumed such that the deformation problem is analyzed at fixed t and thus for fixed $\Omega_t^{(i)}$. As a result, the deformation problem becomes a standard frictional contact problem, as briefly introduced in Sect. 2, which can be formulated and solved in a standard manner.

Solving the deformation problem at a fixed slow time t yields the deformation mappings $\boldsymbol{\varphi}_t^{(1)}(\mathbf{X}^{(1)}, \tau)$ and $\boldsymbol{\varphi}_t^{(2)}(\mathbf{X}^{(2)}, \tau)$ for $\tau \in [t, t + \Delta\tau]$, and all other quantities involved, such as the contact variables $(t_N, \mathbf{t}_T, \mathbf{v}_T)$. Furthermore, the wear rate $\dot{W}_t^{(i)}(\mathbf{X}^{(i)}, \tau)$ can be computed for all points $\mathbf{X}^{(i)} = \boldsymbol{\Psi}^{(i)}(\hat{\boldsymbol{X}}^{(i)}, t)$ on the potential contact surface $\Gamma_c^{(i)}$ using a suitable wear law, for instance, the Archard law discussed in Sect. 3.2. Computation of the wear rate $\dot{W}_t^{(i)}$ is merely a postprocessing task, as all influential variables are known from the solution of the deformation problem.

The increment $\Delta W_t^{(i)}(\mathbf{X}^{(i)})$ of wear accumulated at $\mathbf{X}^{(i)} \in \Gamma_c^{(i)}$ during the time period $[t, t + \Delta\tau]$ can then be computed by integrating the wear rate $\dot{W}_t^{(i)}$, thus

$$\Delta W_t^{(i)}(\mathbf{X}^{(i)}) = \int_t^{t+\Delta\tau} \dot{W}_t^{(i)}(\mathbf{X}^{(i)}, \tau)\mathrm{d}\tau, \tag{22}$$

and the *average wear rate* at the t-scale, denoted by $\overline{\dot{W}}_t^{(i)}(\mathbf{X}^{(i)})$, can be defined as

$$\overline{\dot{W}}_t^{(i)}(\mathbf{X}^{(i)}) = \frac{\Delta W_t^{(i)}(\mathbf{X}^{(i)})}{\Delta\tau}. \tag{23}$$

Upon adopting the assumption of separation of time scales, it is the average wear rate $\overline{\dot{W}}_t^{(i)}(\mathbf{X}^{(i)})$ that is used in the shape evolution law (20) instead of the instantaneous wear rate $\dot{W}^{(i)}$ in the original finite-wear problem.

In order to apply the assumption of time scale separation, it is required that the shape change associated with the wear increment $\Delta W_t^{(i)}$ accumulated over one deformation cycle is sufficiently small so that its effect on the solution of the deformation problem is negligible. The average wear rate at the t-scale can then be determined as a postprocessing quantity at the τ-scale, and the problem of shape evolution due to wear and the deformation problem are incrementally decoupled. In fact, this decoupling is implicitly assumed in many simulation approaches (e.g., Podra and Andersson 1999; Oqvist 2001; McColl et al. 2004; Hegadekatte et al. 2006; Paulin et al. 2008; Gallego et al. 2010).

Note that, in a general case, specification of the deformation problem at the τ-scale may be nontrivial. In fact, friction is a *path-dependent* phenomenon, and suitable boundary and initial conditions must thus be applied in order to properly describe the complex evolution of stick and slip zones accompanied by the wear-induced shape evolution. In some specific situations, the related problems are easily overcome, for instance, when each loading cycle starts with an open contact (e.g., Paulin et al. 2008; Gallego et al. 2010) or when gross sliding occurs during the loading cycle so that the contact memory is erased.

4.3 Quasi-steady-state Wear Problems

A quasi-steady-state wear problem is defined such that the deformation problem corresponding to a fixed slow time t is a steady-state frictional contact problem once formulated in an appropriate Eulerian frame. Typical examples of quasi-steady-state wear problems are the pin-on-disc tribological test and rolling contact. For instance, in the former case, the reference frame would be attached to the pin, and the disc would be analyzed in an Eulerian frame, or in an arbitrary Lagrangian–Eulerian (ALE) frame in the case of deformable disc.

In a quasi-steady-state wear problem, the deformation subproblem and thus also the deformation mappings $\boldsymbol{\varphi}^{(i)}$ do not depend on the fast time τ, so that we have

$$\mathbf{X}^{(i)} = \boldsymbol{\Psi}^{(i)}(\hat{\mathbf{X}}^{(i)}, t), \qquad \mathbf{x}^{(i)} = \boldsymbol{\varphi}_t^{(i)}(\mathbf{X}^{(i)}), \qquad t \in [0, T], \tag{24}$$

while, as in the general case, the deformation problem is parameterized by the slow time t of the shape evolution problem. The average wear rate at the t-scale is then

simply equal to the actual wear rate at the τ-scale, thus

$$\overline{\dot{W}}_t^{(i)}(\mathbf{X}^{(i)}) = \dot{W}_t^{(i)}(\mathbf{X}^{(i)}), \tag{25}$$

and $\dot{W}_t^{(i)}(\mathbf{X}^{(i)})$ does not depend on τ.

In a steady-state frictional contact problem, the motion is decomposed into a background motion which is a rigid-body motion in the undeformed configuration (treated in an Eulerian description) and deformation (treated in a Lagrangian description). This corresponds to a kind of ALE formulation. As the contacting bodies are here assumed hyperelastic, i.e., their behavior is time- and history-independent, and the inertial effects are neglected, the Eulerian rigid-body motion does not affect the deformation problem, except that relative sliding velocity must be properly defined.

The velocity $\mathbf{v}^{(i)}$ of a material point with the position $\mathbf{x}^{(i)}$ in the deformed configuration $\omega^{(i)}$ results solely from the background motion with velocity $\mathbf{V}^{(i)}$ in the undeformed configuration $\Omega^{(i)}$. Specifically, we have

$$\mathbf{v}^{(i)} = \mathbf{F}^{(i)}\mathbf{V}^{(i)}, \tag{26}$$

where $\mathbf{F}^{(i)} = \partial\mathbf{x}^{(i)}/\partial\mathbf{X}^{(i)}$ is the deformation gradient. The sliding velocity $\mathbf{v}_T$ is then defined as the tangential component of the relative velocity,

$$\mathbf{v}_T = v_T^\alpha \boldsymbol{\tau}_\alpha, \qquad v_T^\alpha = (\mathbf{v}^{(1)} - \bar{\mathbf{v}}^{(2)}) \cdot \boldsymbol{\tau}^\alpha, \tag{27}$$

where $\boldsymbol{\tau}_\alpha$ is the tangent basis, $\boldsymbol{\tau}^\alpha$ is the cobasis, and $\bar{\mathbf{v}}^{(2)} = \mathbf{v}^{(2)}(\bar{\mathbf{x}}^{(2)})$ is the velocity of the projection point $\bar{\mathbf{x}}^{(2)}$, see Sect. 2. With this modification, formulation of the steady-state contact problem is identical to that of the general contact problem discussed in Sect. 2.

4.4 Time Integration of Shape Evolution Problem

In order to arrive at a feasible computational scheme, a time integration scheme must be applied to the time-continuous shape evolution problem (20). First-order explicit and implicit Euler schemes are discussed below. A second-order explicit scheme that employs sensitivity analysis in order to arrive at a more accurate approximation of wear increments has been proposed by Lengiewicz and Stupkiewicz (2012).

In the following, two subsequent discrete time instants t_n and $t_{n+1} = t_n + \Delta t$ are thus considered, and a subscript is used to denote the quantities evaluated at a discrete time instant, e.g., $\boldsymbol{\Psi}_{n+1}^{(i)}(\hat{\mathbf{X}}^{(i)}) = \boldsymbol{\Psi}^{(i)}(\hat{\mathbf{X}}^{(i)}, t_{n+1})$. Application of the Euler time integration scheme to the shape evolution law (20) gives

$$\left(\boldsymbol{\Psi}_{n+1}^{(i)}(\hat{\mathbf{X}}^{(i)}) - \boldsymbol{\Psi}_n^{(i)}(\hat{\mathbf{X}}^{(i)})\right) \cdot \mathbf{N}_{n+\theta}^{(i)}(\hat{\mathbf{X}}^{(i)}) = -\Delta t\, \overline{\dot{W}}_{n+\theta}^{(i)}(\hat{\mathbf{X}}^{(i)}), \tag{28}$$

for $\hat{\mathbf{X}}^{(i)} \in \hat{\Gamma}_c^{(i)}$, while the above condition can be formally extended to the whole boundary $\partial\hat{\Omega}^{(i)}$ by setting $\dot{\overline{W}}_{n+\theta}^{(i)}(\hat{\mathbf{X}}^{(i)}) = 0$ for $\hat{\mathbf{X}}^{(i)} \in \partial\hat{\Omega}^{(i)} \setminus \hat{\Gamma}_c^{(i)}$. The explicit forward-Euler scheme and the implicit backward-Euler scheme are obtained for $\theta = 0$ and $\theta = 1$, respectively.

The incremental shape update scheme (28) prescribes only the normal component of the increment of the shape transformation mapping. In order to completely determine $\boldsymbol{\Psi}_{n+1}^{(i)}$ on the boundary, additional assumptions must be adopted concerning its tangential increment on the boundary $\partial\hat{\Omega}^{(i)}$. For instance, one can assume that the shape transformation is such that the points on the contact boundary $\Gamma_c^{(i)}$ are transformed along the normal direction. This yields the following shape update scheme:

$$\boldsymbol{\Psi}_{n+1}^{(i)}(\hat{\mathbf{X}}^{(i)}) = \boldsymbol{\Psi}_{n}^{(i)}(\hat{\mathbf{X}}^{(i)}) - \Delta t\, \dot{\overline{W}}_{n+\theta}^{(i)}(\hat{\mathbf{X}}^{(i)})\mathbf{N}_{n+\theta}^{(i)}(\hat{\mathbf{X}}^{(i)}). \tag{29}$$

The time increment Δt in the incremental scheme (28) or (29) is in general independent of the characteristic time $\Delta\tau$ of the deformation problem. In fact, Δt can be adopted much larger than $\Delta\tau$, and its value is actually dictated by the desired accuracy of the time integration scheme and possibly by its stability (see below).

Explicit forward-Euler time integration scheme As already mentioned, the explicit scheme is obtained by setting $\theta = 0$. Eq. (29) can be then be rewritten in the following form

$$\mathbf{X}_{n+1}^{(i)} = \mathbf{X}_{n}^{(i)} - \Delta t\, \dot{\overline{W}}_{n}^{(i)}(\mathbf{X}_{n}^{(i)})\mathbf{N}_{n}^{(i)}(\mathbf{X}_{n}^{(i)}), \tag{30}$$

which is the basis of a simple and popular shape update scheme that employs remeshing after the contact problem is solved at each time step (e.g., Podra and Andersson 1999; Oqvist 2001; McColl et al. 2004; Hegadekatte et al. 2006; Paulin et al. 2008). In the context of the finite-element method, the shape update (30) is applied to the boundary nodes. Subsequently, the positions of the interior nodes are determined in a suitable remeshing procedure.

The explicit scheme is simple and easy to implement, but it is only *conditionally stable* so that the time increment must satisfy the stability condition

$$\Delta t < \Delta t_{cr}, \qquad \Delta t_{cr} \sim \frac{h}{E}. \tag{31}$$

As shown by Johansson (1994), the critical time increment Δt_{cr} is proportional to the characteristic mesh size h and inversely proportional to the elastic modulus E. In realistic conditions, the critical time increment may be very small so that the scheme becomes computationally expensive. Conditional stability of the explicit scheme is illustrated by the numerical example reported below.

An approach alternative to the nodal shape update scheme resulting from Eq. (30) has been proposed by Lengiewicz and Stupkiewicz (2012). In that approach, the shape parametrization is independent of the finite-element discretization. Time-dependent shape parameters are introduced for that purpose and the shape evolution law is then enforced approximately through a minimization problem in which evolution of shape parameters is fitted to the wear profile resulting from the finite-element

solution of the deformation subproblem. As illustrated by Lengiewicz and Stupkiewicz (2012), independent shape parametrization with a reduced number of shape parameters improves stability of the explicit scheme. To improve accuracy, it can be combined with a second-order explicit scheme that employs shape sensitivity analysis (Stupkiewicz et al. 2010).

Implicit backward-Euler time integration scheme In the case of the fully implicit time integration scheme ($\theta = 1$), the shape update schemes (28) and (29) involve the normal $\mathbf{N}^{(i)}_{n+1}$ and the wear rate $\dot{\overline{W}}^{(i)}_{n+1}$, both evaluated at t_{n+1}. The former explicitly depends on the unknown shape transformation mapping $\boldsymbol{\Psi}^{(i)}_{n+1}$, the latter depends on the unknown solution $\boldsymbol{\varphi}^{(i)}_{n+1}$ of the deformation subproblem at t_{n+1}, and thus it also implicitly depends on $\boldsymbol{\Psi}^{(i)}_{n+1}$. The two subproblems are thus coupled and the problem must be solved simultaneously for $\boldsymbol{\varphi}^{(i)}_{n+1}$ (displacements) and $\boldsymbol{\Psi}^{(i)}_{n+1}$ (shape transformation). Of course, the size of the problem increases due to additional unknowns.

However, the benefit is that the implicit scheme is *unconditionally stable*, so the time increment is limited only by the desired accuracy and not by the stability condition. In practice, significantly larger time increments can be used as compared to the explicit scheme, thus leading to a computationally efficient scheme.

Illustrative example: time integration schemes A numerical example, taken from Stupkiewicz (2013), is provided here to illustrate accuracy and stability of the explicit and implicit time integration schemes and, in particular, the influence of the elastic modulus and finite-element size on the critical time increment, as predicted by Eq. (31).

Consider a hyperelastic (neo-Hookean) pin in plane-strain conditions of the geometry shown in Fig. 2. Its lateral boundaries are constrained in the lateral direction and are free to move in the vertical direction. The pin is pressed into a moving rigid plane by a constant uniform traction that is applied at the top surface. Frictionless contact is considered in this example, and the wear rate is thus assumed to be proportional, with a constant wear coefficient, to the product of contact pressure and sliding velocity. Details concerning geometry and material and process parameters can be found in Stupkiewicz (2013).

The Young's modulus E is varied between 10 and 640 MPa so that, for a fixed prescribed loading, the deformation and the initial contact area are relatively small for $E = 640$ MPa (small deformation regime in Fig. 2) and they are relatively large for $E = 10$ MPa (finite deformation regime, results corresponding to $E = 20$ MPa are shown in Fig. 2).

Wear-induced evolution of the shape of the contact surface is presented in Fig. 3. The shape evolution problem has been integrated using the explicit and the implicit scheme, both with large time increments ($\Delta t = 200$ s, 5 time steps, solid lines) and with small time increments ($\Delta t = 5$ s, 200 time steps, dashed lines). It can be seen that, in the small-deformation regime, the time increment $\Delta t = 200$ s is higher than the critical one for the explicit scheme, and the corresponding results exhibit numerical instability, while the instability is not observed for the smaller time increment

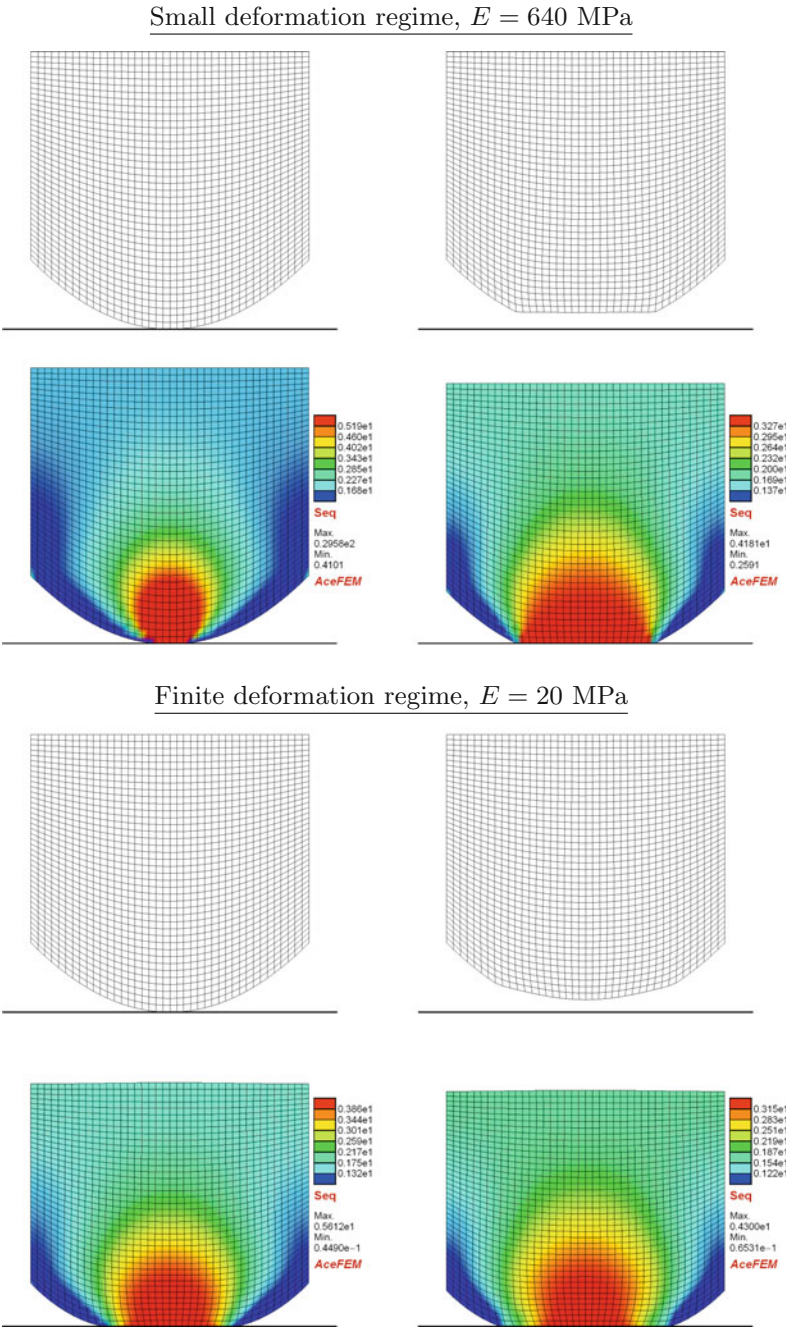

Fig. 2 Two-dimensional pin-on-flat problem in the small-deformation regime (*top*) and in the finite-deformation regime (*bottom*). Finite-element mesh in the undeformed configuration and equivalent stress σ_{eq} in the deformed configuration are shown at the initial time instant $t = 0$ (*left*) and at the final time instant $t = 1000\,\mathrm{s}$ (*right*) (reproduced from Stupkiewicz 2013)

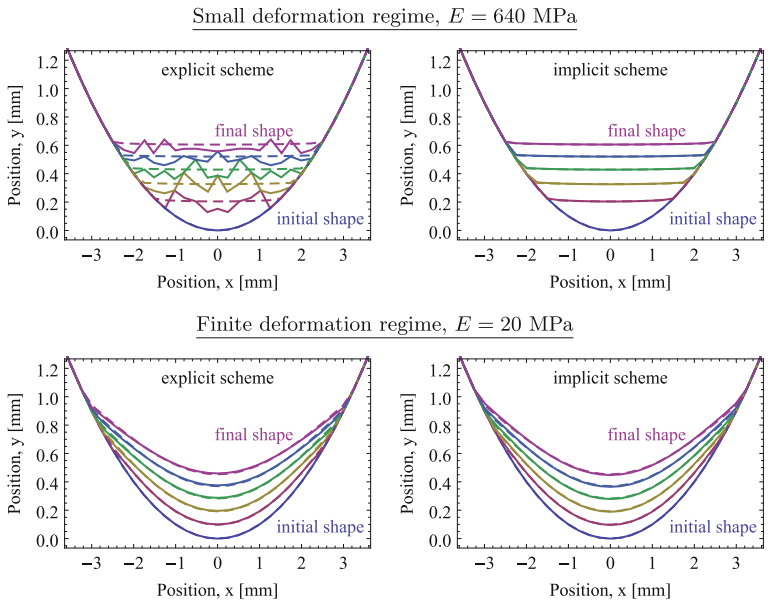

Fig. 3 Two-dimensional pin-on-flat problem: small-deformation (*top*) and finite-deformation (*bottom*) regime. Shape evolution obtained in 5 time steps ($\Delta t = 200$ s, solid lines) and in 200 time steps ($\Delta t = 5$ s, dashed lines) using the explicit (*left*) and implicit (*right*) time integration scheme (reproduced from Stupkiewicz 2013)

$\Delta t = 5$ s. At the same time, the implicit scheme is capable of accurately reproducing significant configuration changes in just 5 time steps. In the finite-deformation regime, the explicit scheme is stable and the accuracy of both schemes is similar. This confirms that the critical time step increases with decreasing elastic modulus, cf. Eq. (31).

A quantitative assessment of accuracy and stability of the two time integration schemes is presented in Fig. 4. The figure shows the solution error as a function of the time increment Δt which has been varied between 1.56 s (640 time steps) and 200 s (5 time steps). Additionally, the Young's modulus E is varied for a fixed mesh density of 80×80 elements, and the mesh density is varied for a fixed Young's modulus $E = 640$ MPa. The solution error has been computed as the Euclidean norm of the difference of the final nodal positions at the contact surface with respect to the reference solution obtained using the time increment of 0.78 s (1280 time steps).

In the case of the implicit scheme, the solution error increases with increasing time increment in an approximately linear manner. This is expected because the Euler scheme is first-order accurate. Similar behavior is observed for the explicit scheme at relatively small time increments. However, a sudden increase of the error is observed at larger time increments. This is related to the instability of the explicit forward-Euler scheme. When the mesh density is increased, for instance, from 20×20 to 40×40 elements so that the element size is decreased by a factor of two, the sudden increase

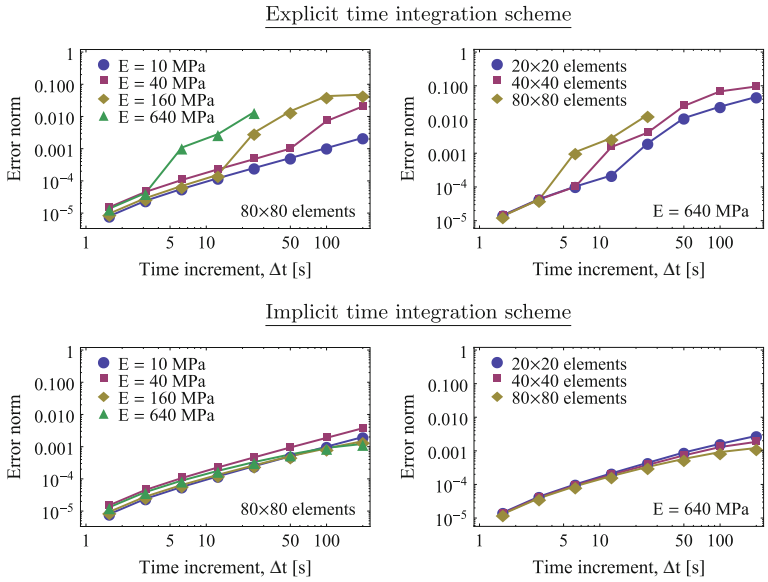

Fig. 4 Two-dimensional pin-on-flat problem: solution error as a function of time increment Δt for the explicit (*top*) and implicit (*bottom*) time integration scheme. Additionally, the elastic modulus E is varied for the mesh of 80×80 elements (*left*) and mesh density is varied for $E = 640$ MPa (*right*) (reproduced from Stupkiewicz 2013)

of the error occurs for a twice smaller time increment. Similar effect is observed when the Young's modulus is varied. Thus, in agreement with the theoretical result of Johansson (1994), see also Eq. (31), the critical time increment is proportional to the element size and inversely proportional to the elastic modulus.

4.5 Shape Update Strategies: Discussion

The incremental solution procedures reported in the literature for the progressive wear problems are usually based on the explicit forward-Euler time integration scheme. This is because computer implementation of the explicit scheme is considerably simpler than implementation of the implicit scheme. The typical explicit procedure amounts to solving the contact problem for the known current shape of the contacting bodies. Knowing the solution of the contact problem, the wear rate can be computed directly as a postprocessing quantity. The wear depth increment is then obtained by multiplying the wear rate by the time increment, the shape is updated accordingly, and the solution proceeds to the next time step. This procedure has been used in combination with the finite-element method (Johansson 1994; Podra and Andersson 1999; Oqvist 2001; McColl et al. 2004; Hegadekatte et al. 2006; Paulin et al. 2008;

Lengiewicz and Stupkiewicz 2012), with the boundary element method (Serre et al. 2001; Sfantos and Aliabadi 2006a; Rodriguez-Tembleque et al. 2012), and with specialized contact solvers (Gallego et al. 2010; Andersson et al. 2011).

When finite configuration changes due to wear are considered, the shape update necessarily involves remeshing (Podra and Andersson 1999; Oqvist 2001; McColl et al. 2004; Hegadekatte et al. 2006; Paulin et al. 2008; Lengiewicz and Stupkiewicz 2012). Alternatively, assuming that the shape changes are small, the shape change can be modelled by simply adding the accumulated wear depth to the initial normal gap (Johansson 1994; Serre et al. 2001; Rodriguez-Tembleque et al. 2012).

As discussed above, the explicit scheme is conditionally stable, and the related instabilities are commonly encountered in computational practice (Johansson 1994; Podra and Andersson 1999; Oqvist 2001; McColl et al. 2004; Sfantos and Aliabadi 2006a; Lengiewicz and Stupkiewicz 2012). The critical time increment decreases with increasing elastic modulus and with decreasing element size, cf. Eq. (31). It follows that mesh refinement increases the computational cost not only due to the increased number of unknowns but also due to the increased number of time steps in view of the stability condition enforced on the time increment. Thus, in problems of practical interest, the solution may be prohibitively expensive.

On the contrary, the implicit backward-Euler scheme is unconditionally stable so that the time increment is constrained only by the desired accuracy of the solution. Application of the implicit scheme requires that the wear increment (or shape transformation resulting from wear) constitutes an additional unknown in the problem. Since the implementation is significantly more involved, the implicit scheme is by far less frequently applied to progressive wear problems (Strömberg 1997; Jourdan and Samida 2009; Ben Dhia and Torkhani 2011; Stupkiewicz 2013; Farah et al. 2017).

In the small-deformation framework adopted by Strömberg (1997), the configuration changes are neglected so that the wear depth could be adopted as an additional unknown to be added to the initial normal gap.

In the approach of Jourdan and Samida (2009), the shape transformation due to wear is restricted to the outer layer of elements only, and the computational treatment is based on the non-smooth contact dynamics method (Jean 1999). The resulting finite-element equations are not fully linearized so that the iterative solution scheme is effectively a modified Newton method. In the implicit scheme developed by Ben Dhia and Torkhani (2011), the coupled wear–deformation problem is not fully linearized either, and the shape transformation is determined using a fixed-point iteration method.

A fully-coupled implicit scheme applicable for quasi-steady-state wear problems has been developed by Stupkiewicz (2013). In this scheme, the shape transformation mapping $\boldsymbol{\Psi}^{(i)}$ and the deformation mapping $\boldsymbol{\varphi}^{(i)}$ constitute the global unknowns of the problem, the former is determined from an auxiliary elasticity problem which is driven by wear increments on the contact boundary. As a result, a kind of an arbitrary Lagrangian–Eulerian (ALE) formulation is obtained in which the shape transformation (i.e., the mesh motion in the finite-element context) is resolved simultaneously with the displacements, all in a fully-implicit monolithic manner. The resulting

finite-element equations are solved using the Newton method, and its quadratic convergence is achieved thanks to full linearization of the governing equations.

A general ALE-like scheme has been recently developed by Farah et al. (2017). The incremental scheme is partitioned into a Lagrangian step and a shape evolution step. The Lagrangian step corresponds to the deformation subproblem, and it is performed for a fixed shape of the contacting bodies. However, wear effects within this step are accounted for by adding the accumulated wear increment to the normal gap. In the shape evolution step, the accumulated wear depth is, in a sense, transferred from the normal gap to the new updated mesh. The shape evolution step employs an auxiliary elasticity problem, similar to that used by Stupkiewicz (2013). This partitioned scheme can be iterated until convergence is achieved. The implicit finite wear framework is combined with a state-of-the-art finite-element contact formulation employing dual mortar methods (Popp et al. 2013).

It is worth mentioning that several asymptotic or simplified approaches are available in the literature, which constitute an alternative to the direct incremental schemes discussed above (e.g., Peigney 2004; Paczelt and Mróz 2005; Sfantos and Aliabadi 2006b; Argatov 2011; Lengiewicz and Stupkiewicz 2013; Menga and Ciavarella 2015). These are not discussed here.

5 Finite Wear: Illustrative Examples

Three illustrative examples are provided in this section. The examples are taken from our earlier work (Lengiewicz and Stupkiewicz 2012; Stupkiewicz 2013), where the details can be found along with an extended discussion of the results.

5.1 Reciprocating Pin-on-flat Problem

A two-dimensional problem is first considered that corresponds, in a simplified manner, to the reciprocating pin-on-flat tribological test. This example is aimed at illustrating the concept of separation of time scales, as discussed in Sect. 4.2.

An elastic pin is pressed into an elastic block and a reciprocating motion is enforced under constant normal force, see Fig. 5 for the geometry and finite-element mesh used in the computations. Both the pin and the block undergo finite deformations and finite configuration changes due to wear. Details concerning geometry, frictional contact and material parameters can be found in Lengiewicz and Stupkiewicz (2012).

The formulation and the computational treatment rely here on the concept of two time scales and time-scale separation. The fast time scale τ corresponds to one loading cycle during which the pin is slid to the right and to the left with a fixed amplitude A, starting at the center of the block. The deformation subproblem that is solved at each slow time scale instant t includes an initial stage during which the

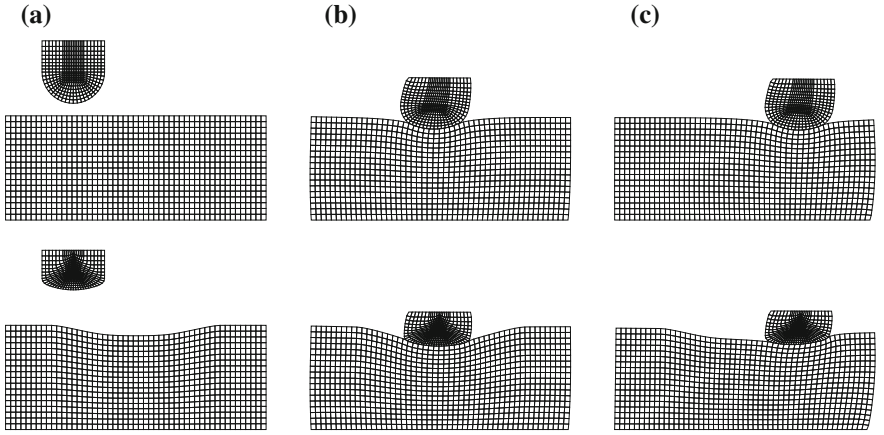

Fig. 5 Reciprocating pin-on-flat problem: initial (unworn, *top*) and final (worn, *bottom*) undeformed and deformed configurations at the beginning of the initial stage (**a**), at the beginning of the actual loading cycle (**b**), and in the right-most position after one quarter of the loading cycle (**c**) (reproduced from Lengiewicz and Stupkiewicz 2012)

pin is brought to contact in the left-most position, Fig. 5a, and then it is slid towards the center of the block, Fig. 5b. This ensures that the actual loading cycle starts in the condition of fully developed frictional sliding. Subsequently, the pin is slid the distance of A to the right, Fig. 5c, then the distance of $2A$ to the left, and finally the distance of A to the right, which completes the loading cycle in a single deformation subproblem. Following the assumption of scale separation, shape changes due to wear are suppressed at the fast time scale. After the deformation subproblem is solved and the corresponding wear increment is computed, the shape evolution subproblem is solved, here using the explicit scheme. The resulting finite shape changes are clearly visible in Fig. 5.

A study of accuracy of the explicit time integration scheme, including also the second-order explicit scheme that employs sensitivity analysis, can be found in Lengiewicz and Stupkiewicz (2012).

5.2 Elastic Ball–Rigid Flat Problem

In this example, an elastic ball is slid against a rigid surface under constant normal load, see Stupkiewicz (2013) for a detailed description. The problem is thus a quasi-steady-state wear problem, and the deformation subproblem is a steady-state frictional contact problem in a frame attached to the ball, cf. Sect. 4.3.

The ball radius is $R = 5\,\text{mm}$, the elastic properties are specified by $E = 100\,\text{GPa}$ and $\nu = 0.3$, and the normal force is $F = 100\,\text{N}$. Considering that the counter-body is rigid, the reduced elastic stiffness of the contact pair is approximately equal to

that of two elastic bodies made of steel. The Hertzian pressure is then $p_0 = 2.1\,\mathrm{GPa}$, and the Hertzian contact radius is $a = 0.15\,\mathrm{mm}$. These are realistic conditions that correspond to the small-deformation regime. A finite-deformation counterpart has also been studied by Stupkiewicz (2013)—the corresponding results are not provided here.

Loading is applied at the mid-plane of the ball so that only one quarter of the ball can be analyzed considering the symmetry with respect to a plane parallel to the sliding direction. As the elastic strains are small, the undeformed and deformed configurations are very close one to the other. Note that frictional contact is here considered. Otherwise the problem would be axially symmetric. The initial and the final shape of the ball are shown in Fig. 6.

The contact pressure at two instants is shown in Fig. 7. The initial pressure at $t = 0$ is not included in Fig. 7 because the finite-element mesh is too coarse to reasonably reproduce the Hertzian pressure distribution (the element size in the contact area is 0.125 mm, while the Hertzian contact radius is $a = 0.15\,\mathrm{mm}$). It can be seen that the pressure is uniform, and its value decreases as the contact area increases due to progressive wear. This response is easily explained by observing that the counter-surface is planar and rigid. As the elastic strains and displacements are small, the worn contact surface is also planar, see Fig. 6b. Accordingly, wear induces a kind of rigid-body motion of the ball in the normal direction. This rigid-body motion is then associated with a uniform wear rate which, through the wear model, induces a uniform contact pressure. Note that the related features of the quasi-steady-state wear problems constitute the basis for simplified asymptotic models (e.g., Paczelt and Mróz 2005; Lengiewicz and Stupkiewicz 2013).

The results reported above were obtained using the implicit scheme with a fixed time increment $\Delta t = 100\,\mathrm{s}$ (actually, at the very beginning of the process, substepping was needed to achieve convergence so that 17 time steps were needed in total to complete the simulation). By numerical experiments, the critical time increment of the explicit time integration scheme has been estimated to be approximately equal to $\Delta t_{cr} \approx 0.1\,\mathrm{s}$. Accordingly, the explicit scheme would require about 10,000 time

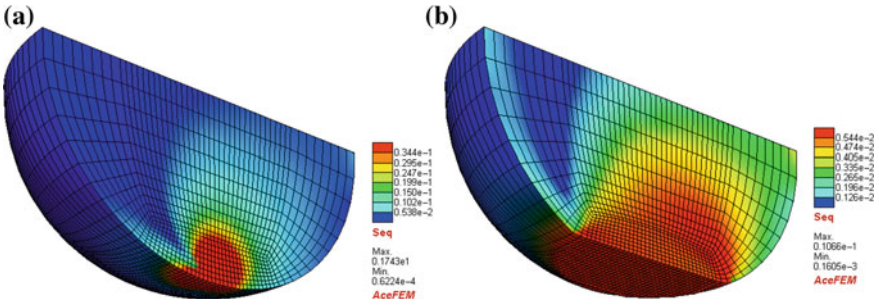

Fig. 6 Elastic ball–rigid flat problem: equivalent stress σ_{eq} in the deformed configuration (the undeformed configuration is nearly identical) at the initial time $t = 0$ (**a**) and at the final time $t = 1000\,\mathrm{s}$ (**b**) (reproduced from Stupkiewicz 2013)

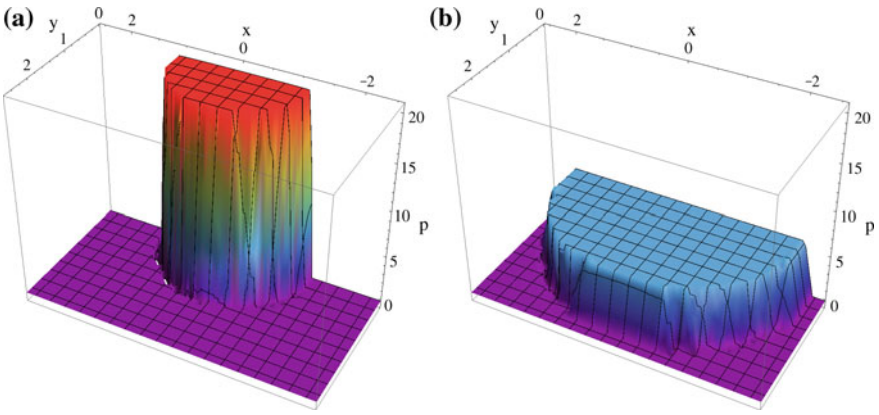

Fig. 7 Elastic ball–rigid flat problem: contact pressure (in MPa) at $t = 100\,\mathrm{s}$ (**a**) and $t = 1000\,\mathrm{s}$ (**b**) (reproduced from Stupkiewicz 2013)

steps to complete the simulation. Even though the computational cost of one time increment in the implicit scheme is higher than that of the explicit scheme due to the increased number of global unknowns, the implicit scheme is more efficient than the explicit one, and the gain in computational cost is significant (about two orders of magnitude).

5.3 Rigid Ball Sliding Against Elastic Half-Space

In this example, a rigid ball is repeatedly slid against an elastic half-space, see Stupkiewicz (2013) for details. Wear due to repeated sliding will thus result in formation of a wear groove on the half-space. This arrangement corresponds, for instance, to the ball-on-disk test, provided the curvature of the sliding path (and wear groove) on the disk is negligible. The problem is a quasi-steady-state problem so that the elastic half-space can be analyzed in an Eulerian frame with the coordinate system attached to the rigid ball. Specifically, an Eulerian description of the rigid-body motion in the undeformed configuration is adopted, while the deformation due to the contact interaction is treated in a Lagrangian manner, see Sect. 4.3. Other relevant tribological tests are illustrated in Fig. 8. In the case of the reciprocating pin-on-flat test, the periodic pin-on-flat arrangement is only an approximation since this problem is actually not a quasi-steady-state problem.

Referring to the pin-on-disk test, the disk wears due to the repeated contact at each revolution of the disk. Hence, the wear rate governing the evolution of the wear groove must be averaged along the sliding path, and the corresponding parameter L, the sliding length per cycle, must be specified. In the case of the pin-on-disk test, L is the circumference of the circular sliding path. This parameter is independent from

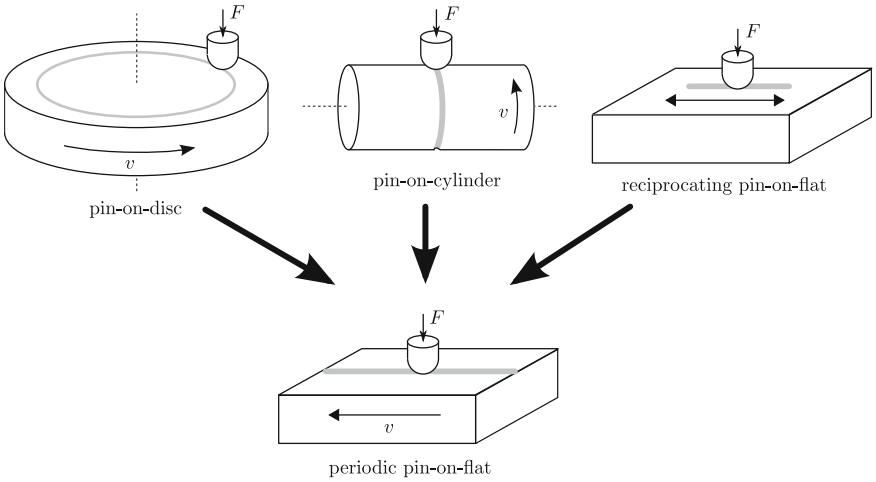

Fig. 8 Periodic pin-on-flat arrangement as an approximation of three tribological tests (reproduced from Lengiewicz and Stupkiewicz 2013)

the actual dimensions of the computational domain, the latter being restricted to the neighbourhood of the contact zone in order to reduce the computational cost.

Two sets of material parameters are considered that correspond to the finite-deformation and the small-deformation regime. Details are provided in Stupkiewicz (2013). The deformed finite-element mesh and the contact pressure at the initial and final time instant are shown in Fig. 9 for the case of the finite-deformation regime.

The wear groove is here uniform along the sliding direction which is a consequence of quasi-steady-state conditions. Accordingly, the shape transformation mapping $\boldsymbol{\Psi}$ is also uniform along the sliding direction, and it is sufficient to prescribe it as a two-dimensional field at one cross-section only. The number of corresponding degrees of freedom in the finite-element model is thus a small fraction of the total number of degrees of freedom, and the additional computational cost of solving the coupled problem of deformation and shape evolution, as referred to the cost of the deformation problem alone, is small. The fully coupled implicit scheme is thus particularly attractive for this class of problems.

In the finite-deformation regime, conditional stability of the explicit scheme is not a crucial issue, as already illustrated in the previous examples. The situation is again very different in the case of the small-deformation regime. Figure 10a presents the corresponding evolution of the contact pressure obtained using the implicit scheme. Initially, the Hertzian pressure distribution corresponds to the initial circular contact zone. Once the wear groove forms, the contact zone becomes elongated and so is the contact pressure distribution. A characteristic feature of the pressure distribution is observed for $t \geq 100\,\mathrm{s}$ which results from the elastic contact interaction of the ball with a nearly cylindrical groove. Specifically, the pressure profile is uniform along the direction perpendicular to the sliding direction, except at the outer edge where

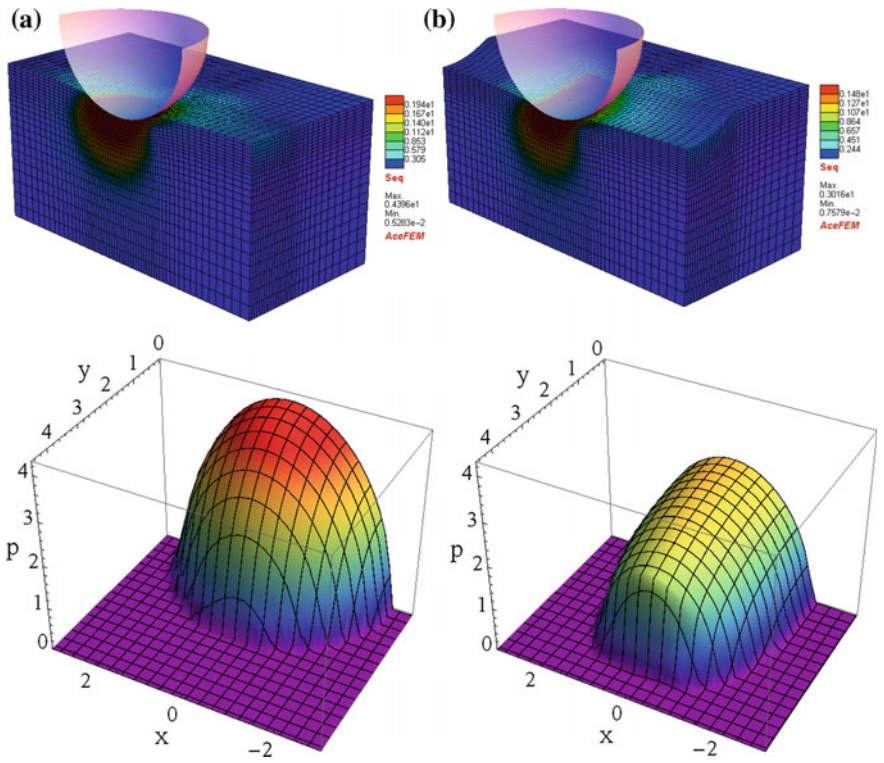

Fig. 9 Rigid ball–elastic half-space problem in the finite-deformation regime: equivalent stress σ_{eq} in the deformed configuration (*top*) and contact pressure (*bottom*) at (**a**) $t = 0$ and (**b**) $t = 1000$ s (reproduced from Stupkiewicz 2013)

a small pressure spike is formed. This pressure distribution is shown in detail at the bottom of Fig. 10a.

In this example, the finite-element mesh (not shown, see Stupkiewicz (2013)) consists of 65,600 hexahedral elements, and the total number of unknowns is 219,432 of which only about 3,400 are the displacement-like quantities corresponding to the shape transformation mapping $\boldsymbol{\Psi}$. The additional computational cost related to the application of the implicit time integration scheme is thus very small. This is because the number of additional unknowns associated with the shape transformation mapping is small compared to the total number of unknowns (less than 2%). The benefit due to stability of the integration scheme is thus obvious. Specifically, it has been checked that the critical time increment of the explicit scheme is here not greater than 0.5–1 s, thus at least 1000–2000 time steps would be needed to obtain a stable solution using the explicit scheme. The corresponding computational cost would thus be approximately two orders of magnitude higher than that of the implicit scheme for which the simulation required only 26 time steps.

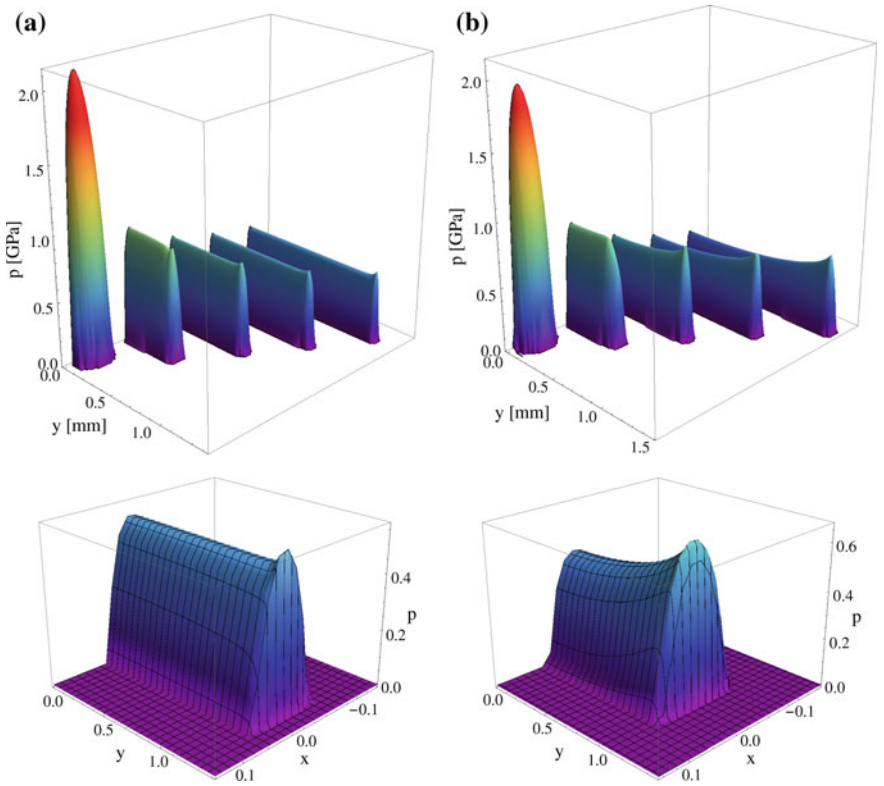

Fig. 10 Rigid ball–elastic half-space problem in the small-deformation regime for the homogeneous (**a**) and inhomogeneous (**b**) half-space. Top figures show evolution of the contact pressure (subsequent graphs correspond to $t = 0, 50, 200, 500, 1000\,\mathrm{s}$) and a detailed view of the contact pressure at $t = 1000\,\mathrm{s}$ is shown in the bottom figures. Results corresponding to the homogeneous half-space are reproduced from Stupkiewicz (2013)

Additional results obtained for an inhomogeneous surface layer are also included in Fig. 10b. Here, it is assumed that the elastic half-space is coated with a functionally graded (FGM) layer in which the elastic modulus and the wear coefficient depend on the depth. Both parameters increase linearly towards the surface starting from their reference values characteristic for the homogeneous substrate. As the material is removed and the wear groove is formed, the material characterized by a lower elastic modulus and a lower wear coefficient is thus gradually exposed. The local wear coefficient is thus non-uniform across the wear groove. As a result, the contact pressure increases towards the groove edges where the wear coefficient is the highest, see Fig. 10b.

5.4 Finite Wear: Summary

In Sects. 3, 4 and 5, a class of non-standard finite-wear contact problems has been discussed, and, in particular, illustrative numerical examples have been presented in Sects. 4.4 and 5.

The distinctive feature of the formulation proposed by Lengiewicz and Stupkiewicz (2012) and described in Sect. 4 is that finite changes of configuration are considered that result from both wear and deformation. This general setting implies that some care must be taken when formulating the problem. For instance, a distinction is made between the nominal and the spatial wear rate, and the wear law, e.g., the classical Archard law, must be adequately formulated, as discussed in Sect. 3. Also, the finite configuration changes have consequences for the finite-element treatment, here discussed in the context of the master-slave approach.

As the second important ingredient, the concept of two time scales has been introduced with the fast time scale of the deformation subproblem and the slow time scale corresponding to the shape evolution problem. Now, assuming separation of the two time scales allows one to partially decouple the two subproblems. Importantly, upon adopting this assumption, the deformation subproblem becomes a standard frictional contact problem which is not affected by the shape changes due to wear and thus can be solved using standard techniques of computational contact mechanics. This has been illustrated by the reciprocating pin-on-flat problem in Sect. 5.1.

Finally, the concept of two time scales provides a framework for developing various computational strategies for time integration of the shape evolution problem. The most common approach is based on applying the explicit forward-Euler scheme which is easy to implement, but which suffers from conditional stability. Alternative strategies include the approach employing the unconditionally-stable implicit backward-Euler scheme that has been studied in Sect. 4.4 and successfully applied to more advanced problems in Sects. 5.2 and 5.3. In particular, in the case of the rigid ball–elastic half-space problem, a high-quality solution of this three-dimensional problem has been obtained at a relatively low computational cost thanks to the application of the implicit scheme. To the best of our knowledge, similar results are not available in the literature, and this is because the application of the usual explicit time-integration scheme would be associated with a very high computational cost due to the severe constraint on the time step, as imposed by the stability criterion.

6 Hydrodynamic Lubrication

6.1 Introduction to Soft-EHL

Hydrodynamic lubrication is a contact regime in which the contacting solids are fully separated by a thin film of fluid (lubricant) such that the load is fully transferred by the hydrodynamic pressure that develops in the fluid. When the hydrodynamic

pressure is sufficiently high to elastically deform the contacting bodies, the thickness of the gap between the bodies changes, which in turn influences the fluid flow. This introduces the elasto-hydrodynamic coupling which is characteristic for the elasto-hydrodynamic lubrication (EHL) regime. Finally, when one or both contacting bodies are highly compliant (or soft), a relatively low pressure suffices to significantly deform the bodies, and this lubrication regime is usually called *soft-EHL*. Alternatively, it is called *elastic-isoviscous* regime, as the effect of the pressure dependence of fluid viscosity is not essential, and the viscosity can be assumed constant, contrary to the elastic-piezoviscous (or hard-EHL) regime in which the related effects are crucial. The focus of the remainder of this chapter is on the soft-EHL problems and, in particular, on the related finite-deformation effects.

Representative examples of soft-EHL problems include elastomeric seals, windscreen wipers, wet tyres, and others. However, the soft-EHL regime is also characteristic for many biotribological systems, such as synovial joints, contact-lens lubrication, eye–eyelid contact, human skin contact, and oral processing of food (e.g., Dowson 1995; de Vicente et al. 2005; Adams et al. 2007; Jones et al. 2008). Clearly, additional difficulties are encountered in the modelling of biotribological systems, which are associated with the modelling of the complex constitutive behaviour of soft tissues and biological fluids. The related effects are not discussed here.

In the context of hydrodynamic lubrication, the fluid flow in the thin channel between the contacting bodies is usually described using the well-known Reynolds equation (Reynolds 1886). The Reynolds equation is obtained from the Navier–Stokes equation by integrating it over the film thickness under several assumptions of which the most important one is that the film is thin so that the flow is laminar and fluid inertia is negligible (Dowson and Higginson 1977; Hamrock et al. 2004). The resulting partial differential equation, introduced in Sect. 6.3 below, relates the hydrodynamic pressure (assumed constant across the film) and the film thickness. In the usual setting, the film thickness is assumed known, and unknown is the pressure. Transition from the Navier–Stokes equation to the Reynolds equation is associated with a dimension reduction so that the Reynolds equation is formulated on the lubrication surface, and the corresponding problem is thus two-dimensional in the general three-dimensional case.

Two phenomena are considered in the modelling of an EHL problem, namely the fluid flow in the thin channel between the contacting bodies and the elastic deflections of the bodies. As mentioned above, the fluid subproblem is conveniently modelled using the Reynolds equation that relates the hydrodynamic pressure and the film thickness. At the same time, the film thickness is influenced by the elastic deflections of the bodies as a result of the action of the hydrodynamic pressure. The two subproblems are thus strongly coupled.

The EHL problem is thus, in fact, a particular kind of a fluid–structure interaction (FSI) problem in which the fluid part is modelled using the Reynolds equation upon adopting the thin-film approximation. In principle, the general FSI approach can alternatively be applied, in which the Navier–Stokes equation is directly used for the fluid part. This general approach offers several advantages by relaxing the assumptions behind the Reynolds equation. At the same time, whenever those assumptions

are reasonably satisfied, the two approaches yield similar results (e.g., Almqvist et al. 2004; Hajishafiee et al. 2017), while the approach based on the Reynolds equation is expected to be significantly more efficient due to the reduced number of degrees of freedom and due to the ease of spatial discretization.

In the classical EHL theory, the solid subproblem is modelled within the linear-elasticity framework. The elasticity problem is then usually formulated for a half-space so that specialized, highly efficient solution methods can be applied (Dowson and Higginson 1977; Hamrock et al. 2004). While both assumptions (linear elasticity, half-space approximation) are well justified in the case of the hard-EHL problems, this is not necessarily so in the case of the soft-EHL problems in which finite deformations may be encountered. Furthermore, the size of the contact zone may be comparable to the size of the contacting bodies so that the half-space approximation is then not adequate.

In realistic conditions, the lubricant film is very thin so that surface roughness may significantly influence the fluid flow and the overall behaviour of the contact pair. However, the related effects are not discussed in the following, and the presentation of the Reynolds equation is restricted to the case of smooth surfaces. Let us only mention here that several approaches are readily available for the modelling of roughness effects through the so-called flow factors introduced first by Patir and Cheng (1978) and reinterpreted later in the rigorous framework of the homogenization theory (Bayada and Faure 1989; Buscaglia and Jai 2000; Bou-Said and Kane 2004; Bayada et al. 2006; Almqvist et al. 2011; Waseem et al. 2017).

6.2 Lubrication Surface and Film Thickness

The Reynolds equation is usually formulated on a planar lubrication surface. Below, a more general form is provided for a non-planar surface. The presentation is here restricted to a time-independent lubrication surface, while the general case of a time-dependent lubrication surface is discussed in detail in the recent work of Temizer and Stupkiewicz (2016).

Let us consider the flow of a fluid (lubricant) in a thin channel between two *physical surfaces* $\gamma_l^{(i)}$, $i = 1, 2$, that represent the contact boundaries of the two contacting bodies in the current configuration, see Fig. 11. The physical surface $\gamma_l^{(i)}$ is parametrized by the convected curvilinear coordinates $\xi^{(i),\alpha}$, $\alpha = 1, 2$, so that the position of a material point in the current configuration is represented by $\mathbf{x}^{(i)} = \mathbf{x}^{(i)}(\boldsymbol{\xi}^{(i)}, t)$ for $\mathbf{x}^{(i)} \in \gamma_l^{(i)}$.

We also introduce the *lubrication surface* $\mathcal{S}$ on which the lubricant flow will be described. It is assumed that the lubrication surface $\mathcal{S}$ is located between the physical surfaces $\gamma_l^{(i)}$. This imposes a constraint on the deformation (motion) of the physical surfaces because $\mathcal{S}$ is here assumed time-independent. Actually, the case of the time-independent lubrication surface is most relevant when one of the physical surfaces is rigid so that the lubrication surface can be identified with this rigid surface.

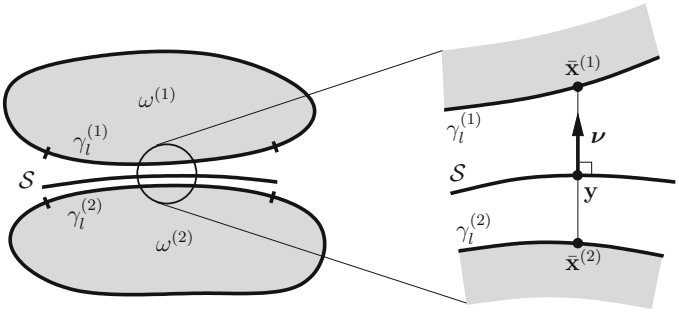

Fig. 11 Physical surfaces $\gamma_l^{(i)}$ and lubrication surface $\mathcal{S}$

Note that there is some ambiguity in the choice of the location of the lubrication surface $\mathcal{S}$ between the physical surfaces $\gamma_l^{(i)}$ when the distance between the physical surfaces is finite (though still sufficiently small so that the Reynolds approximation holds). The results of the computational study reported by Stupkiewicz et al. (2016), see also Fig. 23, show that the solution of the EHL problem is not much afftected by the location of the lubrication surface, and this aspect is not discussed here in detail.

Position of a point on the lubrication surface $\mathcal{S}$ is denoted by $\mathbf{y}$, and parametrization of $\mathcal{S}$ by curvilinear coordinates $\boldsymbol{\eta} = \{\eta^1, \eta^2\}$ is introduced so that $\mathbf{y} = \mathbf{y}(\boldsymbol{\eta})$ for $\mathbf{y} \in \mathcal{S}$. This parametrization introduces the tangent basis $\mathbf{g}_\alpha$ and the co-basis $\mathbf{g}^\alpha$ such that

$$\mathbf{g}_\alpha = \frac{\partial \mathbf{y}}{\partial \eta^\alpha}, \qquad \mathbf{g}^\alpha \cdot \mathbf{g}_\beta = \delta^\alpha_\beta, \qquad \alpha, \beta = 1, 2, \tag{32}$$

where δ^α_β is the Kronecker delta. Let $\boldsymbol{\nu}$ denote the unit normal to $\mathcal{S}$, pointing in the direction from $\gamma_l^{(2)}$ to $\gamma_l^{(1)}$, thus $\mathbf{g}_\alpha \cdot \boldsymbol{\nu} = 0$.

Consider now a scalar field $\phi = \phi(\boldsymbol{\eta})$ defined on $\mathcal{S}$. The *surface gradient* of ϕ can be expressed as

$$\operatorname{grad}_{\mathcal{S}} \phi = \frac{\partial \phi}{\partial \eta^\alpha} \mathbf{g}^\alpha, \tag{33}$$

where the repeated (Greek) indices are implicitly summed over. It follows that the surface gradient of a scalar field is a vector tangent to $\mathcal{S}$. Similarly, the *surface divergence* of a vector field $\boldsymbol{\phi} = \boldsymbol{\phi}(\boldsymbol{\eta})$ defined on $\mathcal{S}$ can be expressed as

$$\operatorname{div}_{\mathcal{S}} \boldsymbol{\phi} = \frac{\partial \boldsymbol{\phi}}{\partial \eta^\alpha} \cdot \mathbf{g}^\alpha. \tag{34}$$

For future use, we also recall the *surface-divergence theorem* for a continuously differentiable *tangential* vector field $\boldsymbol{\phi}$ (thus $\boldsymbol{\phi} \cdot \boldsymbol{\nu} = 0$),

$$\int_{\mathcal{S}} \mathrm{div}_{\mathcal{S}}\, \boldsymbol{\phi}\, \mathrm{d}s = \int_{\partial\mathcal{S}} \boldsymbol{\phi} \cdot \mathbf{m}\, \mathrm{d}l, \tag{35}$$

where $\mathbf{m}$ is a unit vector tangent to $\mathcal{S}$ that is simultaneously an outward normal to $\partial\mathcal{S}$.

One of the basic assumptions in the derivation of the Reynolds equation is that the thickness of the lubricant film, denoted by $h = h(\boldsymbol{\eta}, t)$, is small compared to the dimensions of the lubrication surface $\mathcal{S}$. Even though the present formulation admits finite deformations of the contacting bodies, the above assumption restricts those deformations such that the film thickness is sufficiently small so that the Reynolds approximation holds. In typical conditions, the film thickness is indeed small except in the inlet and outlet zones in which the two contacting surfaces diverge and the film thickness may no longer be small. However, the hydrodynamic pressure buildup is concentrated in the zone where the film thickness is relatively small (note that the Poiseuille term discussed below depends on h^3). Accordingly, even if the film thickness is relatively large in the inlet and outlet zones, and thus the assumption of small film thickness may be violated there, the related effect on the solution in the actual contact zone is not substantial.

Since the fluid film thickness is here considered finite, though small, there is some ambiguity in defining its measure. In the following we adopt the definition based on the *inverse-orthogonal* projection, as illustrated in Fig. 11, but alternative options are also possible (cf. Temizer and Stupkiewicz 2016). The film thickness h is thus decomposed into the measures $h^{(i)}$ of the gap between the physical surfaces $\gamma_l^{(i)}$ and the lubrication surface $\mathcal{S}$,

$$h = h^{(1)} - h^{(2)}, \tag{36}$$

where $h^{(i)}$ is defined by the inverse-orthogonal projection of a point $\mathbf{y} \in \mathcal{S}$ onto $\gamma_l^{(i)}$ along the normal $\boldsymbol{\nu}$,

$$h^{(i)} = (\bar{\mathbf{x}}^{(i)} - \mathbf{y}) \cdot \boldsymbol{\nu}. \tag{37}$$

Here, $\bar{\mathbf{x}}^{(i)} = \mathbf{x}^{(i)}(\bar{\boldsymbol{\xi}}^{(i)}, t)$ is the inverse-orthogonal projection of $\mathbf{y}$ onto $\gamma_l^{(i)}$, defined such that $\bar{\mathbf{x}}^{(i)} = \mathbf{y} + h^{(i)}\boldsymbol{\nu}$, and a bar over the symbol denotes the quantity evaluated at the projection point.

6.3 Reynolds Equation

The Reynolds equation expresses the mass balance of the fluid contained in the thin channel between the physical surfaces $\gamma_l^{(1)}$ and $\gamma_l^{(2)}$ (e.g., Hamrock et al. 2004). For a time-independent non-planar lubrication surface $\mathcal{S}$, the mass balance equation takes the following form:

$$\frac{\partial(\varrho h)}{\partial t} + \mathrm{div}_{\mathcal{S}}(\varrho \mathbf{q}) = 0, \tag{38}$$

where $\mathbf{q}$ is the fluid flux vector tangent to $\mathcal{S}$, to be specified below, and ϱ is the density.

For an *incompressible* fluid (relevant for the soft-EHL), the density ϱ is constant and the Reynolds equation expresses the balance of volume, viz.

$$\frac{\partial h}{\partial t} + \operatorname{div}_{\mathcal{S}} \mathbf{q} = 0. \tag{39}$$

In the Reynolds approximation of the fluid flow within the thin film, the fluid flux $\mathbf{q}$ comprises two components: the Poiseuille term is proportional to the pressure gradient and corresponds to the parabolic profile of the fluid velocity across the film, while the Couette term is proportional to the average tangential velocity, see Fig. 12. Specifically, the fluid flux is given by the following expression,

$$\mathbf{q} = -\frac{h^3}{12\mu} \operatorname{grad}_{\mathcal{S}} p + h\mathbf{v}_T, \tag{40}$$

where p is the pressure, $\mathbf{v}_T$ is the mean tangential velocity of the two physical surfaces, and μ is the fluid viscosity, which is here assumed constant, but its pressure-dependence can be easily accounted for. It is recalled that the Reynolds equation is formulated on the lubrication surface $\mathcal{S}$ so that all quantities depend on the curvilinear coordinates $\boldsymbol{\eta}$ and time, thus, for instance, $p = p(\boldsymbol{\eta}, t)$.

In view of the kinematics introduced by the inverse-orthogonal projection and by the definition of the film thickness h, Eqs. (36)–(37), the mean tangential velocity $\mathbf{v}_T$ is defined as

$$\mathbf{v}_T = \frac{1}{2}\left(\bar{\mathbf{v}}_T^{(1)} + \bar{\mathbf{v}}_T^{(2)}\right), \tag{41}$$

where $\bar{\mathbf{v}}_T^{(i)} = \mathbf{v}^{(i)}(\bar{\boldsymbol{\xi}}^{(i)}, t)$ is the tangential velocity of the physical surface $\gamma_l^{(i)}$ at the projection point $\bar{\mathbf{x}}^{(i)}$,

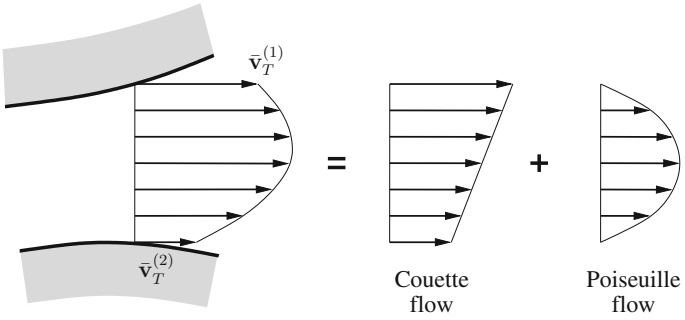

Fig. 12 Fluid velocity across a thin channel: Couette and Poiseuille contributions

$$\mathbf{v}_T^{(i)} = \mathbf{v}^{(i)} - v_N^{(i)}\boldsymbol{\nu}, \qquad \mathbf{v}^{(i)} = \frac{\partial \mathbf{x}^{(i)}}{\partial t}, \qquad v_N^{(i)} = \mathbf{v}^{(i)} \cdot \boldsymbol{\nu}. \tag{42}$$

The Reynolds equation (39)–(40) is accompanied by the standard essential and natural boundary conditions, namely

$$p = p^* \quad \text{on } \partial_p \mathcal{S} \qquad \text{and} \qquad \mathbf{q} \cdot \mathbf{m} = q^* \quad \text{on } \partial_q \mathcal{S}, \tag{43}$$

where p^* is the pressure prescribed on the boundary $\partial_p \mathcal{S}$, q^* is the flux prescribed on the boundary $\partial_q \mathcal{S}$, and $\mathbf{m}$ is a unit vector tangent to $\mathcal{S}$ and normal to $\partial\mathcal{S}$, pointing outwards of $\mathcal{S}$. Note that cavitation, if considered, introduces additional boundary conditions on an unknown cavitation boundary, see Remark 3.1.

The first term in the Reynolds equation (39) is the time derivative of the film thickness $h = h(\boldsymbol{\eta}, t)$. This derivative is evaluated at fixed $\boldsymbol{\eta}$ and, in view of Eqs. (36)–(37), it involves differentiation of the position $\bar{\mathbf{x}}^{(i)}$ of the projection point (for fixed $\boldsymbol{\eta}$). Note that the coordinates $\bar{\boldsymbol{\xi}}^{(i)}$ of the projection point depend on time so that we have $\bar{\mathbf{x}}^{(i)} = \mathbf{x}^{(i)}(\bar{\boldsymbol{\xi}}^{(i)}(\boldsymbol{\eta}, t), t)$. The time derivative of $h^{(i)}$ at fixed $\boldsymbol{\eta}$ is thus equal to

$$\frac{\partial h^{(i)}}{\partial t} = \left(\bar{\mathbf{v}}^{(i)} + \frac{\partial \mathbf{x}^{(i)}}{\partial \xi^{(i),\alpha}} \dot{\bar{\xi}}^{(i),\alpha}\right) \cdot \boldsymbol{\nu} = \bar{v}_N^{(i)} + (\bar{\boldsymbol{\tau}}_\alpha^{(i)} \cdot \boldsymbol{\nu})\, \dot{\bar{\xi}}^{(i),\alpha}, \tag{44}$$

where $\boldsymbol{\tau}_\alpha^{(i)}$ denotes the tangent basis associated with the parametrization of the physical surface $\gamma_l^{(i)}$ by $\boldsymbol{\xi}^{(i)}$, and we have

$$\frac{\partial h}{\partial t} = \bar{v}_N^{(1)} - \bar{v}_N^{(2)} + \left(\bar{\boldsymbol{\tau}}_\alpha^{(1)} \dot{\bar{\xi}}^{(1),\alpha} - \bar{\boldsymbol{\tau}}_\alpha^{(2)} \dot{\bar{\xi}}^{(2),\alpha}\right) \cdot \boldsymbol{\nu}. \tag{45}$$

Note that, in general, $\bar{\boldsymbol{\tau}}_\alpha^{(i)} \cdot \boldsymbol{\nu} \neq 0$ since the physical surfaces are not necessarily parallel to the lubrication surface. It follows from Eq. (45) that the time derivative of the film thickness h involves not only the normal velocities $\bar{v}_N^{(i)}$ of the physical surfaces, but also the term related to the motion of the physical surfaces that are (locally) inclined with respect to the lubrication surface. This is further discussed in Sect. 6.6 in the case of a planar lubrication surface.

For future use, we provide here the formula for the surface traction acting on the physical surface $\gamma_l^{(i)}$:

$$\mathbf{t}^{(i)} = -p\mathbf{n}^{(i)} - \frac{\mu}{h}\left(\bar{\mathbf{v}}_T^{(i)} - \bar{\mathbf{v}}_T^{(i\pm1)}\right) - \frac{h}{2}\,\mathrm{grad}_{\mathcal{S}}\, p, \tag{46}$$

where $\mathbf{n}^{(i)}$ is the unit outward normal to $\gamma_l^{(i)}$. The first term is the normal traction due to the hydrodynamic pressure. The second and the third term are due to the viscous shear stresses in the fluid that act on the solid surfaces (Hamrock et al. 2004, Sect. 7.3). Specifically, the second term results from the Couette flow, and $\bar{\mathbf{v}}_T^{(i\pm1)}$ denotes here the tangential velocity of the countersurface, i.e. $i \pm 1 = 2$ for $i = 1$ and $i \pm 1 = 1$ for $i = 2$. Accordingly, the second term is proportional to the tangential relative

velocity, and it acts on the two surfaces in the opposite direction. The third term results from the Poiseuille flow. This term is proportional to the pressure gradient, and it acts on the two surfaces in the same direction.

6.4 Weak Form of the Reynolds Equation

The weak form of the Reynolds equation (39) is obtained by following the standard procedure. Equation (39) is first multiplied by a test function δp, which vanishes on $\partial_p \mathcal{S}$, and integrated over the lubrication surface $\mathcal{S}$, thus

$$\int_{\mathcal{S}} \left(\frac{\partial h}{\partial t} + \operatorname{div}_{\mathcal{S}} \mathbf{q} \right) \delta p \, \mathrm{d}s = 0 \qquad \forall\, \delta p. \tag{47}$$

By using the identity

$$\operatorname{div}_{\mathcal{S}}(\mathbf{q}\, \delta p) = (\operatorname{div}_{\mathcal{S}} \mathbf{q}) \delta p + \mathbf{q} \cdot \operatorname{grad}_{\mathcal{S}} \delta p \tag{48}$$

and by applying the divergence theorem (34) to the term $\operatorname{div}_{\mathcal{S}}(\mathbf{q}\, \delta p)$, Eq. (47) is then transformed to the following weak form,

$$\int_{\mathcal{S}} \left(\frac{\partial h}{\partial t} \delta p - \mathbf{q} \cdot \operatorname{grad}_{\mathcal{S}} \delta p \right) \mathrm{d}s + \int_{\partial_q \mathcal{S}} q^* \delta p \, \mathrm{d}l = 0 \qquad \forall\, \delta p, \tag{49}$$

where the boundary integral over $\partial_p \mathcal{S}$ vanishes because the test function δp vanishes on $\partial_p \mathcal{S}$.

6.5 Cavitation

Cavitation is an important phenomenon in hydrodynamic lubrication, although, for highly compliant solids, the effect of cavitation on friction is limited (Persson and Scaraggi 2009). The popular mass-conserving cavitation model is briefly presented below. This model is often referred to as the JFO model as it follows the pioneering work of Jakobsson and Floberg (1957) and Olsson (1965). The formulation presented below is based on that developed by Lengiewicz et al. (2014), where more details can found, including the references to the earlier work.

When cavitation occurs, the lubrication surface $\mathcal{S}$ is divided into the full-film region $\mathcal{S}_f$ and the cavitated region $\mathcal{S}_c$ with the cavitation boundary, denoted by Σ, separating the two regions, cf. Fig. 13. The position of the cavitation boundary Σ constitutes a part of the solution hence the cavitation problem belongs to the class of free-boundary problems.

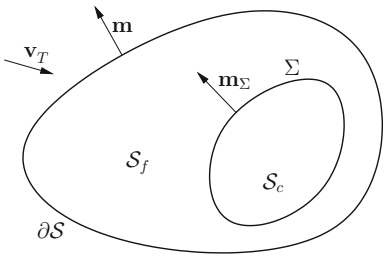

Fig. 13 Cavitation: full-film region $\mathcal{S}_f$ and cavitated region $\mathcal{S}_c$

Cavitation occurs when the hydrodynamic pressure drops to the cavitation pressure p_{cav}, and the pressure is constant and equal to the cavitation pressure $p = p_{cav}$ in the cavitated region, viz.

$$p \geq p_{cav} \ \text{ on } \mathcal{S} \qquad \text{and} \qquad p = p_{cav} \ \text{ on } \mathcal{S}_c. \tag{50}$$

In the cavitated region, the fluid is a mixture of liquid, vapor, and gas, and its density ϱ is thus lower than the density ϱ_0 of the intact fluid,

$$\varrho \leq \varrho_0 \ \text{ on } \mathcal{S} \qquad \text{and} \qquad \varrho = \varrho_0 \ \text{ on } \mathcal{S}_f. \tag{51}$$

Since the focus is here on soft-EHL problems, the intact fluid is assumed incompressible so that the density $\varrho = \varrho_0$ is constant in the full-film region $\mathcal{S}_f$. Introducing the relative density $\bar{\varrho} = \varrho/\varrho_0$ and assuming for simplicity that $p_{cav} = 0$, conditions (50) and (51) can be compactly written in the form of the following complementarity conditions,

$$p \geq 0, \qquad \bar{\varrho} - 1 \leq 0, \qquad p(\bar{\varrho} - 1) = 0 \qquad \text{on } \mathcal{S}, \tag{52}$$

which resemble the unilateral contact conditions (4).

Since the density is no longer constant, the Reynolds equation is now written in the following form, cf. Eq. (38),

$$\frac{\partial(\bar{\varrho}h)}{\partial t} + \operatorname{div}_{\mathcal{S}}(\bar{\varrho}\mathbf{q}) = 0, \tag{53}$$

In the full-film region $\mathcal{S}_f$ the flux is defined by Eq. (40), while in the cavitated region $\mathcal{S}_c$ the flux is assumed to comprise only the Couette term, thus

$$\mathbf{q} = h\mathbf{v}_T \quad \text{on } \mathcal{S}_c, \tag{54}$$

which can be formally obtained from Eq. (40) by putting $\operatorname{grad}_{\mathcal{S}} p = \mathbf{0}$, in agreement with the assumption that the pressure is constant in in the cavitated region.

The mass-balance equation (53) is accompanied by the continuity condition that enforces the mass balance on the cavitation boundary Σ,

$$(\bar{\varrho}^{+}\mathbf{q}^{+} - \bar{\varrho}^{-}\mathbf{q}^{-}) \cdot \mathbf{m}_{\Sigma} = 0 \quad \text{on } \Sigma, \tag{55}$$

where $\mathbf{m}_{\Sigma}$ is the unit vector tangent to $\mathcal{S}$, normal to Σ, and oriented outwards from $\mathcal{S}_c$. The superscripts '+' and '−' denote the limit values of the corresponding quantities as the cavitation boundary Σ is approached from the full-film and cavitated side, respectively.

The weak form of the Reynolds equation combined with the above mass-conserving cavitation model can by obtained by following the procedure outlined in Sect. 6.4, applied to the full-film region $\mathcal{S}_f$ and to the cavitation region $\mathcal{S}_c$ separately, and by exploiting the continuity condition (55). This leads to the following weak form

$$\int_{\mathcal{S}} \left(\frac{\partial(\bar{\varrho}h)}{\partial t}\, \delta p - \bar{\varrho}\mathbf{q} \cdot \operatorname{grad}_{\mathcal{S}} \delta p \right) \mathrm{d}s = 0 \qquad \forall\, \delta p, \tag{56}$$

where it has been additionally assumed that the Dirichlet boundary condition $(43)_1$ is prescribed on the whole boundary $\partial\mathcal{S}$ so that the boundary term vanishes. However, this form is not suitable for the finite-element implementation because it is not compatible with the upwind scheme that is needed to stabilize the advection equation in the cavitated region, see Remark 3.2 below.

An alternative weak form is obtained by applying the divergence theorem only to the pressure-gradient part of the flux $\mathbf{q}$, which yields the following weak form (Lengiewicz et al. 2014, Appendix A)

$$\int_{\mathcal{S}} \left(\frac{\bar{\varrho}h^3}{12\mu} \operatorname{grad}_{\mathcal{S}} p \cdot \operatorname{grad}_{\mathcal{S}} \delta p + \left(\operatorname{div}_{\mathcal{S}}(\bar{\varrho}\mathbf{v}_T h) + \frac{\partial(\bar{\varrho}h)}{\partial t} \right) \delta p \right) \mathrm{d}s + \int_{\Sigma} (\bar{\varrho}^{+} - \bar{\varrho}^{-}) h \mathbf{v}_T \cdot \mathbf{m}_{\Sigma}\, \delta p \, \mathrm{d}l = 0. \tag{57}$$

We note that the last term in the above weak form involves the jump of the relative density $\bar{\varrho}$ along the cavitation boundary Σ. Actually, the relative density $\bar{\varrho}$ is continuous on the rupture boundary and suffers discontinuity on the reformation boundary, see Remark 3.1. Considering that the position of the cavitation boundary Σ is unknown, the jump term in the weak form (57) is an undesired feature from the point of view of finite-element implementation. In the approach developed by Lengiewicz et al. (2014), a continuous finite-element approximation of the relative density $\bar{\varrho}$ (actually, of the void fraction $\lambda = 1 - \bar{\varrho}$) has been introduced so that the jump term vanishes, and a direct finite-element treatment is possible. The jump of $\bar{\varrho}$ on the reformation boundary is then approximated by a continuous solution with a high gradient.

The lubrication and cavitation problem is governed by the weak form (57) and by the complementarity conditions (52). After a small modification, the problem can be formulated as a linear complementarity problem (LCP) that can be solved using the methods available for this class of problems (Giacopini et al. 2010). In the computational scheme developed by Lengiewicz et al. (2014), the complementarity conditions (52) are enforced by introducing a non-smooth constraint function that

relates two independent unknown fields, the pressure p and the relative density $\bar{\varrho}$. This two-field formulation can be transformed to a single-field formulation in which a single variable is used, along with the complementarity conditions (57), to represent the two physical fields, see also Hajjam and Bonneau (2007). A single-field formulation is also employed in the classical Elrod–Adams cavitation algorithm (Elrod and Adams 1974; Elrod 1981) in which the fluid is assumed compressible in the full-film region so that the pressure and the density are related by a one-to-one function. The formulation outlined above is applicable for an incompressible fluid.

Remark 3.1 The mass-flux continuity condition (55) and the cavitation condition (52) imply the well-known boundary conditions of the JFO theory. Specifically, on the film rupture boundary, where $\mathbf{v}_T \cdot \mathbf{m}_\Sigma < 0$, the relative density $\bar{\varrho}$ is continuous so that we have

$$p^+ = 0 \quad \text{and} \quad (\text{grad}_{\mathcal{S}}\, p)^+ \cdot \mathbf{m}_\Sigma = 0. \tag{58}$$

On the reformation boundary, where $\mathbf{v}_T \cdot \mathbf{m}_\Sigma > 0$, the relative density $\bar{\varrho}$ suffers discontinuity which, in view of condition (55), implies discontinuity of the pressure gradient, thus

$$p^+ = 0 \quad \text{and} \quad \frac{\bar{\varrho}^+ h^3}{12\mu}(\text{grad}_{\mathcal{S}}\, p)^+ \cdot \mathbf{m}_\Sigma = (\bar{\varrho}^+ - \bar{\varrho}^-) h \mathbf{v}_T \cdot \mathbf{m}_\Sigma. \tag{59}$$

The formulation developed by Lengiewicz et al. (2014) and outlined above is based on the general continuity condition (55), and the boundary conditions (58) and (59) are not employed directly.

Remark 3.2 In the cavitated region, the lubrication and cavitation problem becomes a pure advection problem because the pressure gradient vanishes in that region in view of the condition $(50)_2$. The standard Galerkin finite-element formulations are not suitable for such problems (Zienkiewicz and Taylor 2000), and upwinding schemes are usually used to stabilize the problem. However, the popular streamline upwind/Petrov–Galerkin (SUPG) method (Brooks and Hughes 1982) is not effective in the case of the weak form (56), because it involves the test-function gradient, which is not affected by the upwind correction when low-order (linear) elements are used. Accordingly, the alternative weak form (57) is needed, for which the SUPG method works well, see Lengiewicz et al. (2014) for details.

6.6 *Traditional Form for a Planar Lubrication Surface*

The Reynolds equation is usually formulated on a planar lubrication surface, which also includes the case of an unwrapped cylindrical surface of a journal bearing. The compact vector notation employed so far in Sect. 6 is not popular either. For completeness, the traditional form of the Reynolds equation is thus included below.

Assume thus that the physical surface $\gamma_l^{(2)}$ is planar and so is the lubrication surface $\mathcal{S}$ which is represented by a domain in the (x, y)-plane so that the normal $\boldsymbol{\nu}$ is aligned with the z-axis. The Reynolds equation (39)–(40) can then be written in the following form:

$$\frac{\partial}{\partial x}\left(\frac{h^3}{12\mu}\frac{\partial p}{\partial x}\right)+\frac{\partial}{\partial y}\left(\frac{h^3}{12\mu}\frac{\partial p}{\partial y}\right) =\frac{\partial}{\partial x}\left(\frac{h}{2}\left(v_x^{(1)}+v_x^{(2)}\right)\right)+\frac{\partial}{\partial y}\left(\frac{h}{2}\left(v_y^{(1)}+v_y^{(2)}\right)\right)+\frac{\partial h}{\partial t}, \tag{60}$$

where $v_x^{(i)}$, $v_y^{(i)}$ and $v_z^{(i)}$ are the components of the velocity $\mathbf{v}^{(i)}$. The time derivative of the film thickness h is now given by the following formula:

$$\frac{\partial h}{\partial t}=v_z^{(1)}-v_z^{(2)}-v_x^{(1)}\frac{\partial h}{\partial x}-v_y^{(1)}\frac{\partial h}{\partial y}, \tag{61}$$

see (Hamrock et al. 2004, Sect. 7.3). As in the general case, cf. Eq. (45), in addition to the contribution of the normal velocity components $v_z^{(i)}$, $\partial h/\partial t$ comprises the term related to the tangential (in-plane) motion of an inclined physical surface $\gamma_l^{(1)}$. Note that $\gamma_l^{(2)}$ is assumed to be planar and parallel to the (x, y)-plane hence it does not contribute to $\partial h/\partial t$.

7 Formulation of the Soft-EHL Problem

This section presents the formulation and the finite-element treatment of the soft-EHL problem. The formulation and its implementation are restricted to steady-state conditions, and one of the contacting bodies is assumed to be rigid and is represented by a rigid surface. Finite deformations of the other body, which is assumed hyperelastic, are fully accounted for.

7.1 Solid Part

Consider first the solid subproblem formulated for the hyperelastic body ($i = 1$, the index is omitted in the following), the other body ($i = 2$) is assumed to be rigid. In the finite-deformation framework, the reference configuration Ω and the current configuration ω are introduced along with the deformation mapping $\mathbf{x} = \boldsymbol{\varphi}(\mathbf{X})$, where $\mathbf{X} \in \Omega$ and $\mathbf{x} \in \omega$. The virtual work principle, i.e. the weak form of equilibrium, has the standard form,

$$G_s(\boldsymbol{\varphi}, \delta\boldsymbol{\varphi};\, p) = \int_\Omega \mathbf{P} \cdot \operatorname{Grad} \delta\boldsymbol{\varphi}\, \mathrm{d}V - \int_{\Gamma_l} \mathbf{T} \cdot \delta\boldsymbol{\varphi}\, \mathrm{d}S = 0 \qquad \forall\, \delta\boldsymbol{\varphi}, \tag{62}$$

where Γ_l is the lubricated contact surface in the reference configuration and $\mathbf{T}$ is the nominal surface traction resulting from the hydrodynamic interaction on Γ_l, which is here, for simplicity, assumed to be the only loading. The formulation can be extended in the standard manner to include other loads, such as prescribed surface traction, body forces, unilateral contact, etc. The pressure field p has been introduced as an additional argument of the functional G_s to indicate that the solid subproblem is coupled to the fluid subproblem through the dependence of the traction $\mathbf{T}$ on p.

The virtual work principle can be written in an equivalent form with the traction term evaluated in the current configuration and expressed in terms of the spatial surface traction $\mathbf{t}$,

$$G_s(\boldsymbol{\varphi}, \delta\boldsymbol{\varphi};\, p) = \int_\Omega \mathbf{P} \cdot \operatorname{Grad} \delta\boldsymbol{\varphi}\, \mathrm{d}V - \int_{\gamma_l} \mathbf{t} \cdot \delta\boldsymbol{\varphi}\, \mathrm{d}s = 0 \qquad \forall\, \delta\boldsymbol{\varphi}, \tag{63}$$

where $\gamma_l = \boldsymbol{\varphi}(\Gamma_l)$ is the lubricated contact surface in the current configuration, and the spatial (Cauchy) traction $\mathbf{t}$ is given by Eq. (46). For a hyperelastic solid, the first Piola–Kirchhoff stress $\mathbf{P}$ is governed by the elastic strain energy function $W(\mathbf{F})$ according to

$$\mathbf{P} = \frac{\partial W}{\partial \mathbf{F}}, \qquad \mathbf{F} = \operatorname{Grad} \boldsymbol{\varphi}, \tag{64}$$

where $\mathbf{F}$ is the deformation gradient.

7.2 Fluid Part

As discussed in Sect. 6, the fluid subproblem is governed by the Reynolds equation that is formulated on the lubrication surface $\mathcal{S}$. When one of the physical surfaces is rigid, as assumed here, it is convenient to define the lubrication surface as the projection of the deformed lubricated contact surface γ_l onto the rigid countersurface, thus $\mathcal{S} = \mathcal{P}(\gamma_l)$, where $\mathcal{P}(\mathbf{x})$ denotes the orthogonal projection of point $\mathbf{x} \in \gamma_l$ onto the rigid countersurface, see Fig. 14. It is stressed that the deformed lubricated contact surface $\gamma_l = \boldsymbol{\varphi}(\Gamma_l)$ depends on the solution of the solid subproblem and so does $\mathcal{S}$.

As a special case of the general weak form (49), the weak form of the steady-state Reynolds equation with the Dirichlet boundary condition prescribed on the whole boundary, i.e. $\partial_p \mathcal{S} = \partial\mathcal{S}$, takes the following simple form:

$$G_f(p, \delta p;\, \boldsymbol{\varphi}) = \int_{\mathcal{S}} \mathbf{q} \cdot \operatorname{grad}_{\mathcal{S}} \delta p\, \mathrm{d}s = 0 \qquad \forall\, \delta p, \tag{65}$$

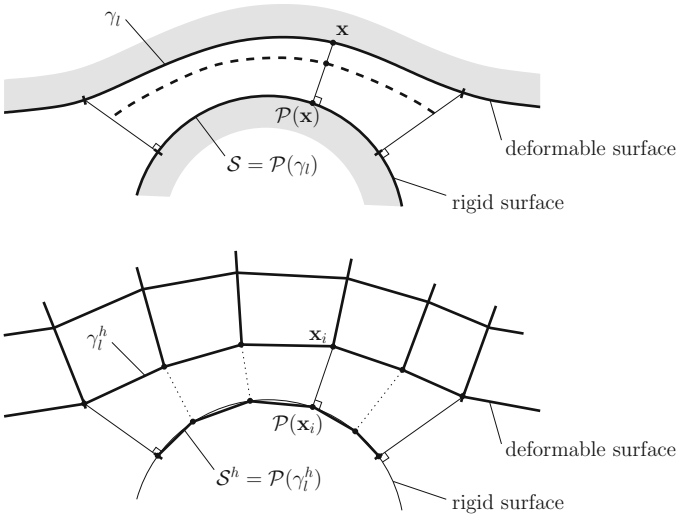

Fig. 14 The solution-dependent lubrication surface $\mathcal{S}$ ($\mathcal{S}^h$) is defined by the projection of the lubricated contact surface γ_l (γ_l^h) onto the rigid countersurface: continuum (*top*) and discretized (*bottom*) setting. The dashed line in the top figure indicates an intermediate surface that could also be adopted as the lubrication surface, see Stupkiewicz et al. (2016) and Temizer and Stupkiewicz (2016)

where $\boldsymbol{\varphi}$ has been introduced as an additional argument of the functional G_f to indicate the respective coupling, e.g., through the film thickness h.

The cavitation is not considered here to make the formulation of the coupled soft-EHL problem possibly simple. In fact, adding the cavitation does not change the general structure of the problem, and it is included in the illustrative examples provided in Sect. 8.

7.3 Elasto-Hydrodynamic Coupling

The strong coupling of the two subproblems introduced above constitutes the essential feature of the EHL problems. Specifically:

(i) the lubricant film thickness h, which influences the lubricant flow through the Reynolds equation, depends on the deformation of the solid, cf. Eqs. (36)–(37);
(ii) the load, i.e. the surface traction $\mathbf{t}$ in the virtual work principle (63), depends on the hydrodynamic pressure p and its gradient, cf. Eq. (46);
(iii) the lubrication surface $\mathcal{S}$, on which the Reynolds equation is formulated, depends on the deformation of the solid through the projection $\mathcal{S} = \mathcal{P}(\gamma_l)$.

The first and the second coupling are characteristic for all EHL problems. The third coupling is solely due to the finite-deformation effects, and it is not present in the

small-deformation framework that is usually adopted in the EHL theory. Note also that the effect of the shear stresses on the deformation of the solid is usually neglected in the EHL theory, and only the pressure loading is considered. However, in soft-EHL problems, this assumption is not necessarily valid, see Stupkiewicz and Marciniszyn (2009). The couplings listed above have already been symbolically indicated in the weak forms (63) and (65) by including p and $\boldsymbol{\varphi}$ as the arguments of the functionals G_s and G_f, respectively.

7.4 Finite-Element Treatment

Following the standard approach, the finite-element approximation of the unknown fields of placement $\boldsymbol{\varphi}$ and pressure p is introduced,

$$\boldsymbol{\varphi}^h = \sum_i N_i^{(\boldsymbol{\varphi})} \boldsymbol{\varphi}_i, \qquad p^h = \sum_i N_i^{(p)} p_i, \tag{66}$$

where $\boldsymbol{\varphi}_i = \mathbf{X}_i + \mathbf{u}_i$ denotes the placement of the i-th node, $N_i^{(\boldsymbol{\varphi})}$ is the corresponding basis function, $\mathbf{X}_i$ is the position of the node in the reference configuration and $\mathbf{u}_i$ is the nodal displacement, which is the actual unknown in the standard finite-element formulation. Similarly, p_i denotes the nodal pressure and $N_i^{(p)}$ the corresponding basis function. Recall that the pressure field is defined on the lubrication surface $\mathcal{S}$ while the displacement is defined in the bulk domain Ω. In the Galerkin method, the test functions are approximated using the same basis functions, thus

$$\delta\boldsymbol{\varphi}^h = \sum_i N_i^{(\boldsymbol{\varphi})} \delta\boldsymbol{\varphi}_i, \qquad \delta p^h = \sum_i N_i^{(p)} \delta p_i. \tag{67}$$

Discretized weak forms are now obtained by introducing the approximations (66)–(67) into the weak forms (63) and (65), viz.

$$G_s^h(\mathbf{U}, \delta\mathbf{U}; \mathbf{P}) = G_s(\boldsymbol{\varphi}^h, \delta\boldsymbol{\varphi}^h; p^h) = 0 \qquad \forall\, \delta\mathbf{U}, \tag{68}$$

and

$$G_f^h(\mathbf{P}, \delta\mathbf{P}; \mathbf{U}) = G_s(p^h, \delta p^h; \boldsymbol{\varphi}^h) = 0 \qquad \forall\, \delta\mathbf{P}, \tag{69}$$

where $\mathbf{U}$ and $\mathbf{P}$ denote the global vectors of unknown nodal displacements and pressures, respectively. Since G_s^h is linear in $\delta\mathbf{U}$ and G_f^h is linear in $\delta\mathbf{P}$ we have

$$G_s^h(\mathbf{U}, \delta\mathbf{U}; \mathbf{P}) = \mathbf{R}_s(\mathbf{U}; \mathbf{P}) \cdot \delta\mathbf{U}, \quad G_f^h(\mathbf{P}, \delta\mathbf{P}; \mathbf{U}) = \mathbf{R}_f(\mathbf{P}; \mathbf{U}) \cdot \delta\mathbf{P}, \tag{70}$$

so that the problem can be written in the residual form,

$$\mathbf{R}_s(\mathbf{U};\mathbf{P}) = \mathbf{0}, \qquad \mathbf{R}_f(\mathbf{P};\mathbf{U}) = \mathbf{0}, \tag{71}$$

or, with a view to applying a monolithic solution scheme, in the following form,

$$\bar{\mathbf{R}}(\bar{\mathbf{U}}) = \mathbf{0}, \qquad \bar{\mathbf{R}} = \{\mathbf{R}_s, \mathbf{R}_f\}, \qquad \bar{\mathbf{U}} = \{\mathbf{U}, \mathbf{P}\}. \tag{72}$$

In the monolithic scheme, the nonlinear equation (72) is solved simultaneously with respect to all unknowns, and the Newton method is used here for that purpose. The tangent matrix required by the Newton method can be efficiently obtained using the automatic differentiation (AD) technique (Korelc 2009; Korelc and Wriggers 2016). Since Eq. (72) governing the coupled problem is highly nonlinear, convergence of the iterative Newton scheme cannot be guaranteed. This problem can be circumvented by applying a kind of continuation method in which the desired solution is obtained by gradually increasing the load (or by varying another influential parameter) and by using the converged solution as the initial guess for the subsequent solution corresponding to the increased load. This approach proved successful in solving a wide range of fully-coupled finite-deformation soft-EHL problems (e.g., Stupkiewicz and Marciniszyn 2009; Stupkiewicz et al. 2016); selected examples are presented in the next section.

8 Soft-EHL: Illustrative Examples

Three illustrative examples of the soft-EHL problems that involve finite deformations of the solid are briefly described in this section. The examples are taken from our earlier work (Stupkiewicz 2009; Stupkiewicz et al. 2016), where more details can be found.

8.1 Reciprocating O-Ring Seal

In this section, selected results are presented for the reciprocating elastomeric O-ring seal that has been studied in detail by Stupkiewicz (2009), see also Stupkiewicz and Marciniszyn (2009). Due to the simple geometry (circular cross section), the O-ring seal is a suitable benchmark problem, and it is frequently studied in various contexts (e.g., Fatu and Hajjam 2011; Yang and Salant 2011; Shin et al. 2016).

The general geometrical setup of a hydraulic seal is shown in Fig. 15. Due to the action of the sealed pressure, the seal is compressed between the housing and the rod, and this is accompanied by visibly finite deformations, see Fig. 16. The reciprocating motion of the rod results in the buildup of the hydrodynamic pressure at the rod–seal interface, and the analysis below is limited to the steady-state hydrodynamic lubrication during the *outstroke* (the rod moves towards the air side, $U > 0$) and *instroke* (the rod moves towards the sealed pressure side, $U < 0$). The unilateral

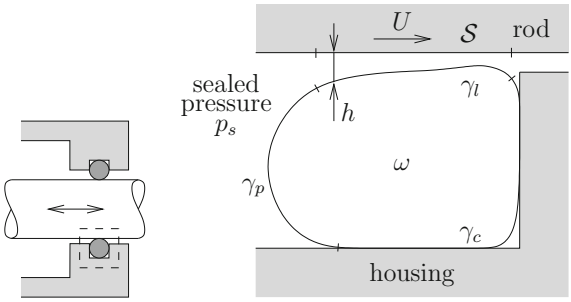

Fig. 15 Reciprocating O-ring seal: schematic of the problem

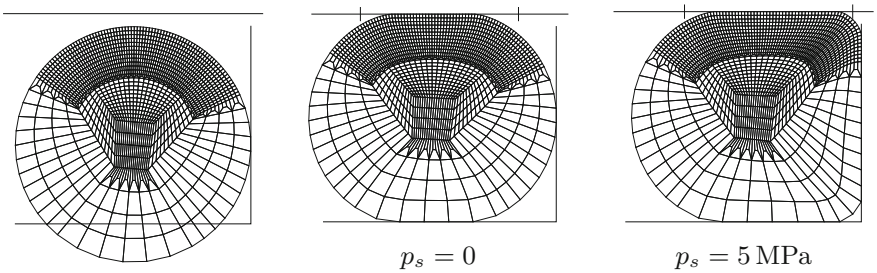

Fig. 16 O-ring seal: finite-element mesh (mesh density 2) in the undeformed configuration (*left*) and in the deformed configuration for two values of the sealed pressure p_s. The housing and the rod are rigid and are represented by solid lines (reproduced from Stupkiewicz 2009)

contact with the housing is also included in the model, and both the rod and the housing are assumed to be rigid. Axial symmetry with respect to the rod axis is assumed so that the problem is two-dimensional, and thus the Reynolds equation becomes one-dimensional.

The seal is assumed to be hyperelastic, governed by the Mooney–Rivlin material model. All the geometrical, material and process parameters can be found in Stupkiewicz (2009). Note that the maximum sealed pressure, $p_s = 5\,\mathrm{MPa}$, exceeds the shear modulus of the seal, $\mu = \mu_1 + \mu_2 = 3.66\,\mathrm{MPa}$. Five mesh densities have been used in the computations with the total number of unknowns ranging from about 1,000 for mesh density 1 to 190,000 for mesh density 16. The mass-conserving cavitation model, cf. Sect. 6.5, is not employed in this example. Instead, the cavitation condition is approximately enforced using the penalty method (Wu 1986).

Figure 17 shows the hydrodynamic pressure and the film thickness for two values of the sealed pressure p_s. It can be seen that a characteristic ridge is formed at outlet, i.e. on the right (left) in the case of the outstroke (instroke). The pressure profiles corresponding to the outstroke and instroke are very similar except for fine details at the inlet and outlet. In particular, a small pressure dimple is observed at the outlet during the instroke, while, at the same location, which corresponds to the inlet during the outstroke, the pressure increases monotonically. A similar dimple is not formed

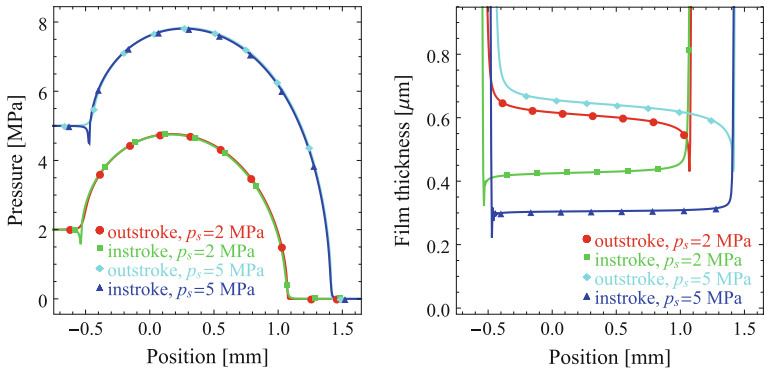

Fig. 17 O-ring seal: the effect of the sealed pressure p_s on the hydrodynamic pressure p (*left*) and film thickness h (*right*) for the rod speed $U = 100$ mm/s (mesh density 16). Position $x = 0$ corresponds to the position of the center of the cross section in the undeformed configuration, see Fig. 16 (reproduced from Stupkiewicz 2009)

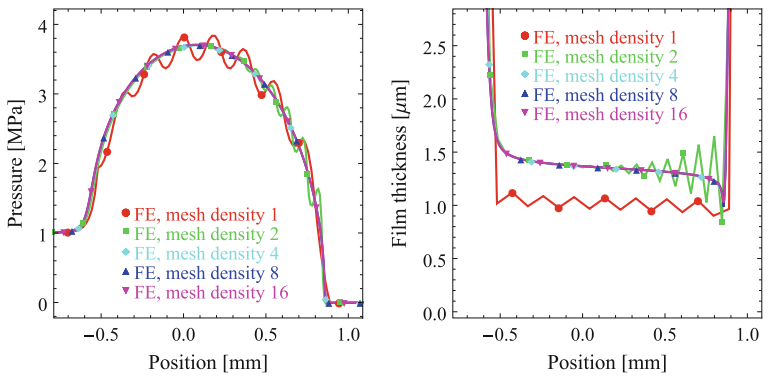

Fig. 18 O-ring seal: convergence with mesh refinement in terms of the pressure p (*left*) and of the film thickness h (*right*) (reproduced from Stupkiewicz 2009)

at the outlet during the outstroke because the pressure would then decrease below zero, which is prevented by the cavitation condition.

As mentioned above, five mesh densities have been used in the computations to check the performance of the computational scheme and convergence with mesh refinement. In particular, it has been observed that spurious oscillations of pressure and film thickness may occur in some situations. This is illustrated in Fig. 18 which shows convergence of the solution with mesh refinement. The oscillations of the film thickness have a zigzag-like appearance which is related to the piecewise-linear finite-element approximation of the displacement field on the boundary. The approximation of the hydrodynamic pressure is here piecewise-polynomial and hence the wavy appearance of the pressure oscillations.

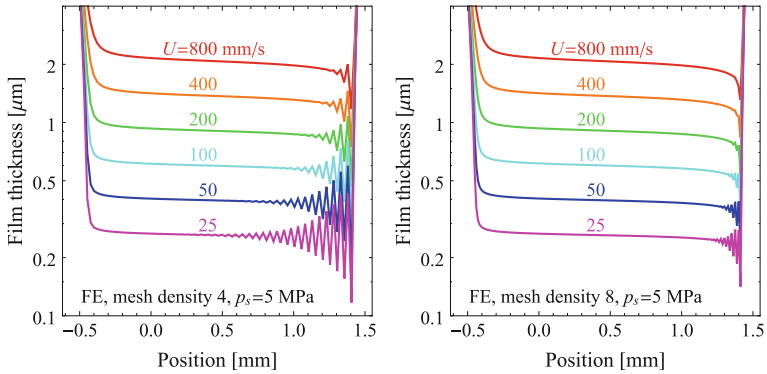

Fig. 19 O-ring seal: effect of the rod speed U and mesh density on the film thickness h for mesh density 4 (*left*) and mesh density 8 (*right*). The more severe lubrication conditions, the finer mesh is needed to avoid spurious oscillations (reproduced from Stupkiewicz 2009)

Different orders of approximation of the pressure have been tested in a quest for a remedy to the spurious oscillations. However, the only remedy found is the mesh refinement, as can be observed in Fig. 18. In the case shown in Fig. 18 (outstroke, $p_s = 1\,\text{MPa}$, $U = 400\,\text{mm/s}$), an oscillation-free solution is obtained for mesh density 4 or higher.

The spurious oscillations are further illustrated in Fig. 19 showing the effect of the rod speed and mesh density. It can be seen that the higher the mesh density, the lower the rod speed at which the oscillations appear. Figure 19 shows the results for mesh densities 4 and 8; further reduction of the oscillations is obtained for mesh density 16, see Fig. 15c in Stupkiewicz (2009). This trend is also visible in the other cases studied.

The general conclusion resulting from the convergence studies is that the spurious oscillations occur in severe lubrication conditions, and the severer the lubrication conditions, the finer mesh is necessary to avoid the oscillations. This applies to the finite-element treatment of the Reynolds equation, as illustrated above, but also to the discontinuous Galerkin method that has also been tested, see Stupkiewicz (2009). The low speed of the rod, which results in a small film thickness, is the main factor that promotes the oscillations. Severe conditions are also associated with higher sealed pressures. Interestingly, a solution with mild oscillations may still provide a reasonable estimation of the actual profile of the pressure and the film thickness once the oscillations are filtered out, see Fig. 18.

8.2 Rigid Cylider Sliding Against a Coated Layer

In this example, a rigid cylinder is sliding against a soft layer with a harder thin coating, see Fig. 20. A hyperelastic neo-Hookean model is adopted for both the layer and

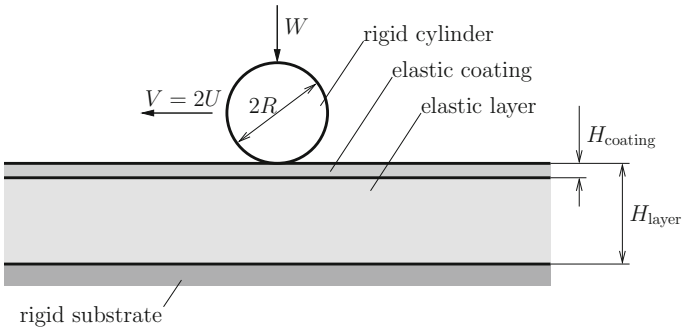

Fig. 20 Rigid cylider sliding against a coated layer (reproduced from Stupkiewicz et al. 2016)

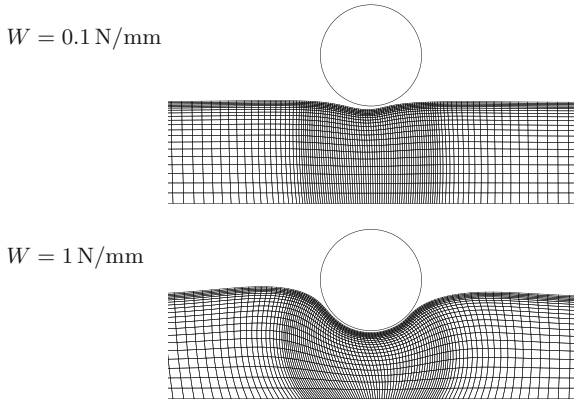

Fig. 21 Coated layer: detail of the deformed finite-element mesh for two values of the load W. A much finer mesh is used in the actual computations (reproduced from Stupkiewicz et al. 2016)

the coating. Steady-state hydrodynamic lubrication is studied in a coordinate system attached to the cylinder. A kind of arbitrary Lagrangian-Eulerian (ALE) description is thus adopted for the layer in which the material flows through the (deformed) finite-element mesh. However, since the layer is elastic, the solid formulation is standard; only the local velocity $\mathbf{v}$ of the material points must be adequately determined in terms of the nominal sliding speed $\mathbf{V}$ according to $\mathbf{v} = \mathbf{F}\mathbf{V}$, where $\mathbf{F}$ is the local deformation gradient. For the material and geometrical parameters used in the computations, the reader is referred to Stupkiewicz et al. (2016).

Figure 21 shows the deformation pattern induced by the lubricated contact. Note that, for better visualization, a coarse mesh is shown in Fig. 21, and a significantly finer mesh has been used in the actual computations. Figure 22 shows the hydrodynamic pressure and the film thickness corresponding to the load W varying between 0.1 and 1 N/mm. To illustrate the effect of finite deformations, which are clearly visible in Fig. 21, the results obtained using the geometrically linear model, i.e. according to the classical EHL theory, are also included in Fig. 22 (indicated by dashed lines). It can be seen that both the pressure and the film thickness are not predicted correctly when the finite-deformation effects are neglected, particularly at the higher loads.

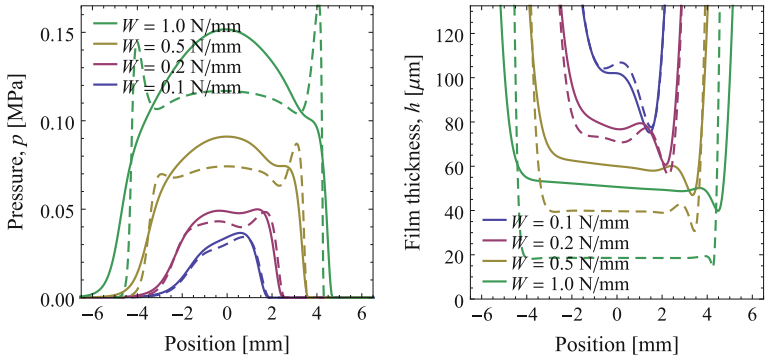

Fig. 22 Coated layer: the hydrodynamic pressure p (*left*) and the film thickness h (*right*) as a function of the load W (nominal entrainment speed $\mu U = 0.1\,\mathrm{N/m}$). Dashed lines indicate the results of the geometrically linear model (reproduced from Stupkiewicz et al. 2016)

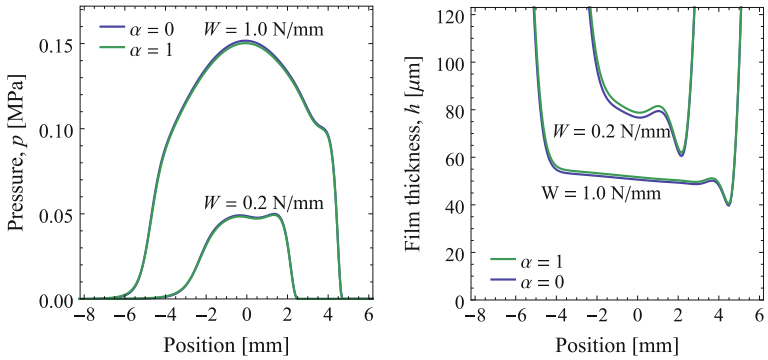

Fig. 23 Coated layer: influence of the position of the lubrication surface $\mathcal{S}$ (specified by parameter α, see text) on the pressure p (*left*) and on the film thickness h (*right*) for $\mu U = 0.1\,\mathrm{N/m}$ (reproduced from Stupkiewicz et al. 2016)

As described in Sect. 7.4, the lubrication surface $\mathcal{S}$ is here defined as the projection on the lubricated contact surface γ_l onto the rigid countersurface, i.e. $\mathcal{S} = \mathcal{P}(\gamma_l)$. Alternatively, surface γ_l itself (i.e. $\mathcal{S} = \gamma_l$) or an intermediate surface (such as that indicated by the dashed line in Fig. 14) could be chosen as the lubrication surface. The effect of this choice is small, as illustrated in Fig. 23. Parameter $\alpha = 0$ corresponds to $\mathcal{S} = \mathcal{P}(\gamma_l)$ and $\alpha = 1$ corresponds to $\mathcal{S} = \gamma_l$. Intermediate values of α are also possible, see Stupkiewicz et al. (2016), but the results are not sensitive to the variation in α, as can already be deduced from Fig. 23.

8.3 Elastic Ball Sliding Against a Rigid Plane

In this last example, a hyperelastic ball is slid against a rigid plane in the steady-state hydrodynamic lubrication regime. A constant normal force W is applied at the horizontal mid plane, which is allowed to move vertically as a rigid plane. Further, the symmetry with respect to the vertical mid plane aligned with the sliding direction is exploited so that only one quarter of the ball is included in the model. The setup corresponds to the ball-on-disc test under pure sliding, assuming that the curvature of the sliding path is neglected. The material and geometrical parameters can be found in Stupkiewicz et al. (2016).

The finite-element mesh, shown in Fig. 24, has been refined in the vicinity of the contact zone, with the finest mesh along the trailing edge of the contact zone. Figure 24 shows also the σ_{zz} component of the Cauchy stress in the deformed configuration. At the highest load $W = 40\,\mathrm{N}$, the radius of the contact zone is equal to approximately 50% of the ball radius. The ball is thus deformed in the finite-strain regime, well beyond the Hertzian contact conditions.

Maps of the lubricant film thickness h are shown in Fig. 25 for selected values of the load W and entrainment speed U, the latter defined as one half of the sliding speed V, thus $U = V/2$. The entrainment speed is here provided in the form of the product μU, μ being the fluid viscosity, since the entrainment speed enters the Reynolds equation only through this product. Hence, the effect of increasing the sliding speed (and the entrainment speed) by the factor of two is the same as if the viscosity was increased by the factor of two.

It can be seen in Fig. 25 that a characteristic ridge is formed along the trailing edge of the contact zone. At lower loads and at higher entrainment speeds, the point of the minimum film thickness is located on the symmetry axis, at the rear of the contact zone. Otherwise, there are two minima located at the side lobes. These qualitative features agree well with the results obtained using the classical EHL theory (Hooke 1995).

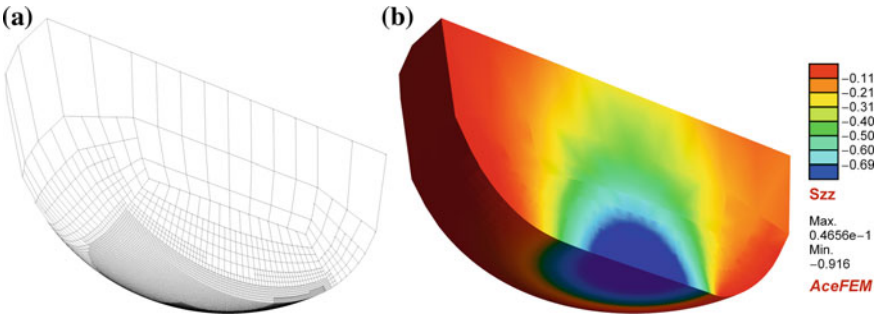

Fig. 24 Elastic ball sliding against a rigid plane: the finite-element mesh (*left*) and the σ_{zz} component of the Cauchy stress (in MPa) in the deformed configuration at the load $W = 40\,\mathrm{N}$ (*right*). The inlet is on the left (reproduced from Stupkiewicz et al. 2016)

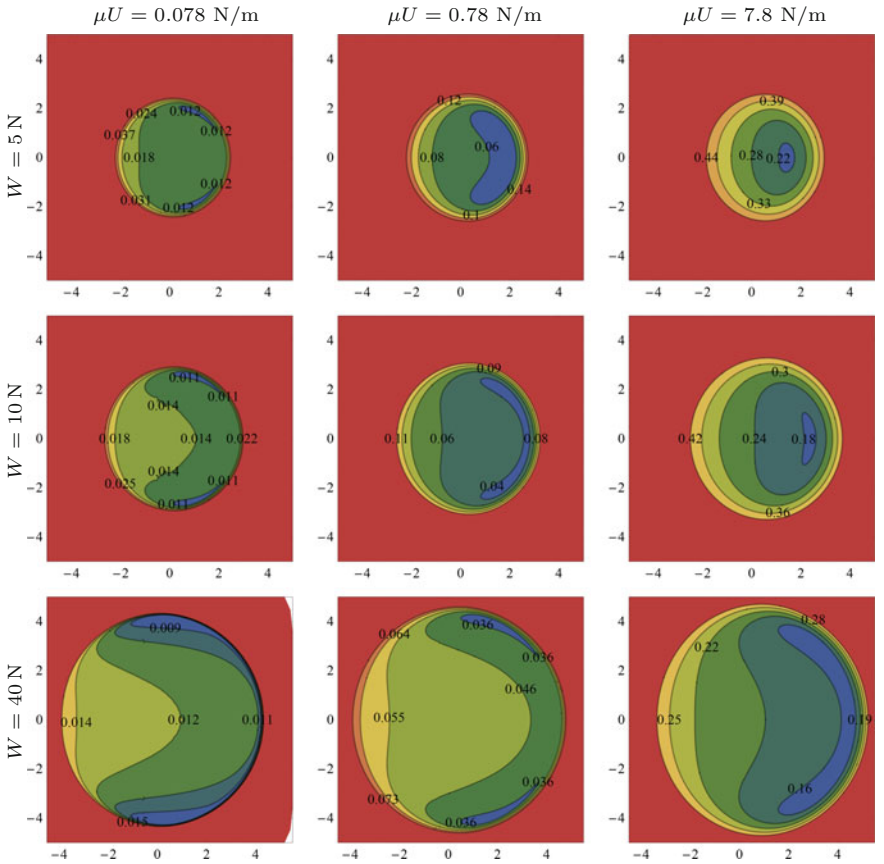

Fig. 25 Elastic ball: maps of the lubricant film thickness h (in mm, position in mm, inlet on the left) (reproduced from Stupkiewicz et al. 2016)

The profiles of the pressure p and the film thickness h along the symmetry axis are shown in Fig. 26. In order to illustrate the finite-deformation effects, the results obtained using the fully nonlinear model (solid lines) are compared to the results obtained using two simplified models. The predictions of the geometrically linear model, in which the configuration changes are neglected as in the classical EHL theory, are denoted by dash-dotted lines. It can be seen that the difference is quite substantial, both in terms of the pressure and the film thickness, for instance, the maximum pressure is 10–15% higher and the minimum film thickness is 10–15% lower in the case of the classical EHL theory. Secondly, the dashed lines depict the results obtained using the geometrically nonlinear model in which the shear (friction) stresses are neglected. It follows that the effect of shear stresses is here negligible. However, as shown by Stupkiewicz et al. (2016), this effect becomes visible, though

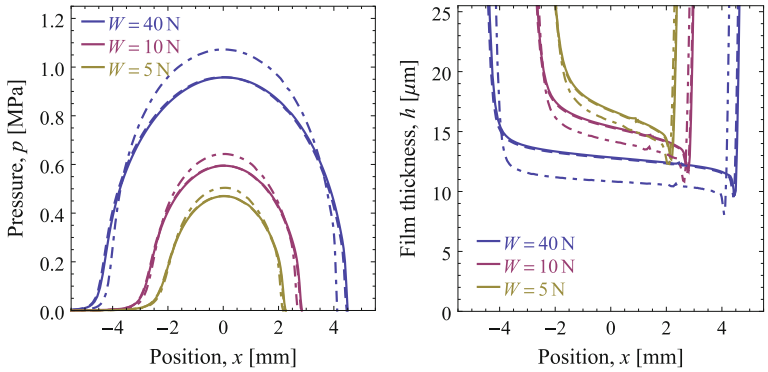

Fig. 26 Elastic ball: pressure p (*left*) and film thickness h (*right*) along the symmetry plane $y = 0$ for $\mu U = 0.078$ N/m. Solid lines denote the nonlinear model, dashed lines denote the nonlinear model without friction stresses, dash-dotted lines denote the geometrically linear model (reproduced from Stupkiewicz et al. 2016)

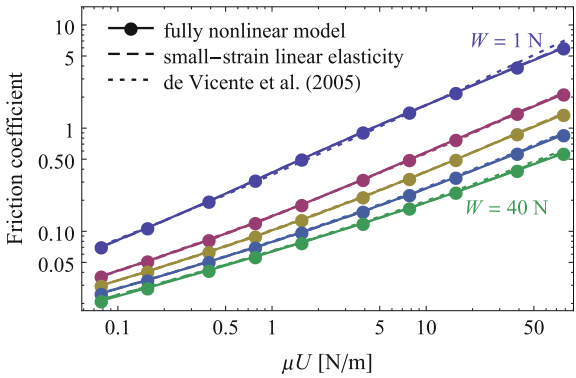

Fig. 27 Elastic ball: friction coefficient as a function of μU and W. The individual lines correspond to the load W equal to 1, 5, 10, 20 and 40 N (from the top to the bottom) (reproduced from Stupkiewicz et al. 2016)

still not substantial, when the entrainment speed μU is increased by one or two orders of magnitude (the respective results are not shown here for brevity).

Interestingly, despite the visible finite deformation effects on the pressure and film thickness, as illustrated in Fig. 26, the friction coefficient is not visibly affected. Figure 27 shows the friction coefficient, defined in the standard manner by dividing the friction force by the normal force, as a function of the entrainment speed μU. It can be seen that the results of the fully nonlinear model and of the geometrically linear model are practically identical while the entrainment speed μU changes by nearly three orders of magnitude and the friction coefficient changes by about two orders of magnitude. The predicted friction coefficient shows also a very good agreement with the regression equation that has been obtained by fitting the predictions of the classical EHL theory (de Vicente et al. 2005).

8.4 Soft-EHL: Summary

In Sects. 6, 7 and 8, we have reviewed the recent progress in the modelling of the soft-EHL problems in the finite-deformation regime. Consistent treatment of the finite-deformation effects in the soft-EHL has several consequences that make the corresponding formulation and its computer implementation distinct from the classical EHL theory.

Finite deformations of one or both contacting bodies may imply finite changes of the lubrication surface, see, e.g., the coated-layer example studied in Sect. 8.2. Accordingly, the Reynolds equation, which is the main tool used for the modelling of the lubricant flow, should be formulated on a non-planar lubrication surface. The corresponding non-classical formulation of the Reynolds equation, provided in Sect. 6, has been here restricted to the case of a time-independent lubrication surface. A detailed discussion of the general case of a time-dependent lubrication surface can be found in Temizer and Stupkiewicz (2016).

At finite deformation, the lubrication surface, as well as its discretization in the computational scheme, depends on the deformation of the solid, and thus it depends on the solution of the EHL problem. This introduces an additional EHL coupling that is not present in the small-strain framework of the classical EHL theory.

Further, an adequate computational method must be used to resolve the finite deformations of the solid. In particular, the elastic half-space approximation, typically used in the classical EHL theory, is not applicable. The finite-element method is here a natural choice, as it provides a general method for treating geometrical and material nonlinearities.

The finite-deformation effects mentioned above are fully accounted for in the computational scheme that is briefly described in Sect. 7.4. So far, the computer implementation is restricted to the case of steady-state problems in which one of the bodies is rigid. Extension to transient lubrication problems for two deformable bodies is a challenging task for future work. The present computational scheme is based on the finite-element method and employs a fully-coupled monolithic solution scheme which proves to perform well.

Representative numerical examples that illustrate the finite-deformation effects have been provided in Sect. 8. Interestingly, in the case of the hyperelastic ball sliding against a rigid plane, the friction coefficient is not affected by the finite deformation of the ball, and the geometrically linear theory delivers accurate predictions of the friction coefficient also at high loads, at which the ball deforms significantly. This result has also been confirmed experimentally, see Fig. 2 in Stupkiewicz et al. (2016).

References

M. J. Adams, B. J. Briscoe, and S. A. Johnson. Friction and lubrication of human skin. *Tribol. Lett.*, 26:239–253, 2007.

P. Alart and A. Curnier. A mixed formulation for frictional contact problems prone to Newton like solution methods. *Comp. Meth. Appl. Mech. Engng.*, 92:353–375, 1991.

A. Almqvist, J. Fabricius, A. Spencer, and P. Wall. Similarities and differences between the flow factor method by Patir and Cheng and homogenization. *Trans. ASME J. Tribol.*, 133:031702, 2011.

T. Almqvist, A. Almqvist, and R. Larsson. A comparison between computational fluid dynamic and Reynolds approaches for simulating transient EHL line contacts. *Tribol. Int.*, 37:61–69, 2004.

J. Andersson, A. Almqvist, and R. Larsson. Numerical simulation of a wear experiment. *Wear*, 271:2947–2952, 2011.

J. F. Archard. Contact and rubbing of flat surfaces. *J. Appl. Phys.*, 24(8):981–988, 1953.

I. I. Argatov. Asymptotic modeling of reciprocating sliding wear with application to local interwire contact. *Wear*, 271:1147–1155, 2011.

G. Bayada and J. B. Faure. A double scale analysis approach of the reynolds roughness comments and application to the journal bearing. *Trans. ASME J. Tribol.*, 111:323–330, 1989.

G. Bayada, I. Ciuperca, and M. Jai. Homogenized elliptic equations and variational inequalities with oscillating parameters. Application to the study of thin flow behavior with rough surfaces. *Nonlinear Anal. Real World Appl.*, 7:950–966, 2006.

H. Ben Dhia and M. Torkhani. Modeling and computation of fretting wear of structures under sharp contact. *Int. J. Num. Meth. Engng.*, 85:61–83, 2011.

B. Bou-Said and M. Kane. Comparison of homogenization and direct techniques for the treatment of roughness in incompressible lubrication. *Trans. ASME J. Tribol.*, 126:733–737, 2004.

A. N. Brooks and T. J. R. Hughes. Streamline upwind/Petrov–Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier–Stokes equations. *Comp. Meth. Appl. Mech. Engng.*, 32:199–259, 1982.

G. C. Buscaglia and M. Jai. A new numerical scheme for non uniform homogenization problems: application to the non linear Reynolds compressible equation. *Math. Probl. Engng.*, 7:355–378, 2000.

J. de Vicente, J. R. Stokes, and H. A. Spikes. The frictional properties of Newtonian fluids in rolling–sliding soft-EHL contact. *Tribol. Lett.*, 20:273–286, 2005.

D. Dowson. Elastohydrodynamic and micro-elastohydrodynamic lubrication. *Wear*, 190:125–138, 1995.

D. Dowson and G. R. Higginson. *Elasto-hydrodynamic Lubrication*. Pergamon Press, 1977.

M. Dragon-Louiset. On a predictive macroscopic contact-sliding wear model based on micromechanical considerations. *Int. J. Sol. Struct.*, 38:1625–1639, 2001.

H. G. Elrod. A cavitation algorithm. *Trans. ASME J. Lubr. Technol.*, 103:350–354, 1981.

H. G. Elrod and M. L. Adams. A computer program for cavitation and starvation problems. In D. Dowson, M. Godet, and C. M. Taylor, editors, *Proc. of the First Leeds–Lyon Symposium on Tribology – Cavitation and Related Phenomena in Lubrication*, pages 37–41. Mechanical Engineering, New York, 1974.

P. Farah, W. A. Wall, and A. Popp. An implicit finite wear contact formulation based on dual mortar methods. *Int. J. Num. Meth. Engng.*, 111:325–353, 2017.

A. Fatu and M. Hajjam. Numerical modelling of hydraulic seals by inverse lubrication theory. *Proc. Instn. Mech. Engrs. Part J: J. Engng. Tribol.*, 225:1159–1173, 2011.

S. Fouvry, P. Kapsa, and L. Vincent. Quantification of fretting damage. *Wear*, 200:186–205, 1996.

L. Gallego, B. Fulleringer, S. Deyber, and D. Nelias. Multiscale computation of fretting wear at the blade/disk interface. *Tribol. Int.*, 43:708–718, 2010.

M. Giacopini, M. T. Fowell, D. Dini, and A. Strozzi. A mass-conserving complementarity formulation to study lubricant films in the presence of cavitation. *Trans. ASME J. Tribol.*, 132:041702–1–12, 2010.

A. Hajishafiee, A. Kadiric, S. Ioannides, and D. Dini. A coupled finite- volume CFD solver for two-dimensional elastohydrodynamic lubrication problems with particular application to rolling element bearings. *Tribol. Int.*, 109:258–273, 2017.

M. Hajjam and D. Bonneau. A transient finite element cavitation algorithm with application to radial lip seals. *Tribol. Int.*, 40:1258–1269, 2007.

B. J. Hamrock, S. R. Schmid, and B. O. Jacobsen. *Fundamentals of Fluid Film Lubrication*. Marcel Dekker, New York, 2 edition, 2004.
V. Hegadekatte, N. Huber, and O. Kraft. Modeling and simulation of wear in pin on disc tribometer. *Tribol. Lett.*, 24:51–60, 2006.
C. J. Hooke. The elastohydrodynamic lubrication of elliptical point contacts operating in the isoviscous region. *Proc. Instn. Mech. Engrs. Part J: J. Engng. Tribol.*, 209:225–234, 1995.
B. Jakobsson and L. Floberg. *The finite journal bearing considering vaporization*. Transactions of Chalmers University Technology, vol. 190. Goteborg, Sweden, 1957.
M. Jean. The non-smooth contact dynamics method. *Comp. Meth. Appl. Mech. Engng.*, 177:235–257, 1999.
L. Johansson. Numerical simulation of contact pressure evolution in fretting. *Trans. ASME J. Tribol.*, 116:247–254, 1994.
M. B. Jones, G. R. Fulford, C. P. Please, D. L. S. McElwain, and M. J. Collins. Elastohydrodynamics of the eyelid wiper. *Bull. Math. Biol.*, 70:323–343, 2008.
F. Jourdan and A. Samida. An implicit numerical method for wear modeling applied to a hip joint prosthesis problem. *Comp. Meth. Appl. Mech. Engng.*, 198:2209–2217, 2009.
J. Korelc. Automation of primal and sensitivity analysis of transient coupled problems. *Comp. Mech.*, 44:631–649, 2009.
J. Korelc and P. Wriggers. *Automation of finite element methods*. Springer International Publishing, Switzerland, 2016.
T. A. Laursen. *Computational Contact and Impact Mechanics*. Springer-Verlag, Berlin, 2002.
J. Lengiewicz and S. Stupkiewicz. Continuum framework for finite element modelling of finite wear. *Comp. Meth. Appl. Mech. Engng.*, 205–208:178–188, 2012.
J. Lengiewicz and S. Stupkiewicz. Efficient model of evolution of wear in quasi-steady-state sliding contacts. *Wear*, 303:611–621, 2013.
J. Lengiewicz, J. Korelc, and S. Stupkiewicz. Automation of finite element formulations for large deformation contact problems. *Int. J. Num. Meth. Engng.*, 85:1252–1279, 2011.
J. Lengiewicz, M. Wichrowski, and S. Stupkiewicz. Mixed formulation and finite element treatment of the mass-conserving cavitation model. *Tribol. Int.*, 72:143–155, 2014.
I. R. McColl, J. Ding, and S. B. Leen. Finite element simulation and experimental validation of fretting wear. *Wear*, 256:1114–1127, 2004.
N. Menga and M. Ciavarella. A Winkler solution for the axisymmetric Hertzian contact problem with wear and finite element method comparison. *J. Strain Anal. Eng. Design*, 50:156–162, 2015.
Z. Mróz and S. Stupkiewicz. An anisotropic friction and wear model. *Int. J. Solids Struct.*, 31:1113–1131, 1994.
K. O. Olsson. *Cavitation in dynamically loaded bearings*. Transactions of Chalmers University Technology, vol. 308. Goteborg, Sweden, 1965.
M. Oqvist. Numerical simulations of mild wear using updated geometry with different step size approaches. *Wear*, 249:6–11, 2001.
I. Paczelt and Z. Mróz. On optimal contact shapes generated by wear. *Int. J. Num. Meth. Engng.*, 63:1250–1287, 2005.
N. Patir and H. S. Cheng. An average flow model for determining effects of three-dimensional roughness on partial hydrodynamic lubrication. *Trans. ASME J. Lubr. Technol.*, 100:12–17, 1978.
C. Paulin, S. Fouvry, and C. Meunier. Finite element modelling of fretting wear surface evolution: Application to a Ti–6Al–4V contact. *Wear*, 264:26–36, 2008.
M. Peigney. Simulating wear under cyclic loading by a minimization approach. *Int. J. Sol. Struct.*, 41:6783–6799, 2004.
B. N. J. Persson and M. Scaraggi. On the transition from boundary lubrication to hydrodynamic lubrication in soft contacts. *J. Phys.: Condens. Matter*, 21:185002, 2009.
G. Pietrzak and A. Curnier. Large deformation frictional contact mechanics: continuum formulation and augmented Lagrangian treatment. *Comp. Meth. Appl. Mech. Engng.*, 177:351–381, 1999.
P. Podra and S. Andersson. Simulating sliding wear with finite element method. *Tribol. Int.*, 32:71–81, 1999.

A. Popp, A. Seitz, M. W. Gee, and W. A. Wall. Improved robustness and consistency of 3D contact algorithms based on a dual mortar approach. *Comp. Meth. Appl. Mech. Engng.*, 264:67–80, 2013.

A. Ramalho and J. C. Miranda. The relationship between wear and dissipated energy in sliding systems. *Wear*, 260:361–367, 2006.

O. Reynolds. On the theory of lubrication and its application to Mr. Beauchamp Towers experiments, including an experimental determination of the viscosity of olive oil. *Phil. Trans. R. Soc. Lond. A*, 177:157–234, 1886.

L. Rodriguez-Tembleque, R. Abascal, and M. H. Aliabadi. Anisotropic wear framework for 3D contact and rolling problems. *Comp. Meth. Appl. Mech. Engng.*, 241–244:1–19, 2012.

I. Serre, M. Bonnet, and R. M. Pradeilles-Duval. Modelling an abrasive wear experiment by the boundary element method. *C. R. Acad. Sci. Paris, Serie II b*, 329:803–808, 2001.

G. K. Sfantos and M. H. Aliabadi. Wear simulation using an incremental sliding Boundary Element Method. *Wear*, 260:1119–1128, 2006a.

G. K. Sfantos and M. H. Aliabadi. Application of BEM and optimization technique to wear problems. *Int. J. Sol. Struct.*, 43:3626–3642, 2006b.

D. C. Shin, J. H. Nam, and D. W. Kim. Experimental interior stress fields of a constantly squeezed O-ring modeling from hybrid transmission photoelasticity. *Exp. Mech.*, 40:59–72, 2016.

C. Stolz. A thermodynamical approach to contact wear as application of moving discontiuities. *Arch. Appl. Mech.*, 77:165–175, 2007.

N. Strömberg. An augmented Lagrangian method for fretting problems. *Eur. J. Mech. A/Solids*, 16:573–593, 1997.

S. Stupkiewicz. Finite element treatment of soft elastohydrodynamic lubrication problems in the finite deformation regime. *Comp. Mech.*, 44:605–619, 2009.

S. Stupkiewicz. An ALE formulation for implicit time inegration of quasi-steady-state problems. *Comp. Meth. Appl. Mech. Engng.*, 260:130–142, 2013.

S. Stupkiewicz and A. Marciniszyn. Elastohydrodynamic lubrication and finite configuration changes in reciprocating elastomeric seals. *Tribol. Int.*, 42:615–627, 2009.

S. Stupkiewicz, J. Lengiewicz, and J. Korelc. Sensitivity analysis for frictional contact problems in the augmented Lagrangian formulation. *Comp. Meth. Appl. Mech. Engng.*, 199:2165–2176, 2010.

S. Stupkiewicz, J. Lengiewicz, P. Sadowski, and S. Kucharski. Finite deformation effects in soft elastohydrodynamic lubrication problems. *Tribol. Int.*, 93:511–522, 2016.

I. Temizer and S. Stupkiewicz. Formulation of the Reynolds equation on a time-dependent lubrication surface. *Proc. Roy. Soc. A*, 472:20160032, 2016.

A. Waseem, J. J. Guilleminot, and I. Temizer. Stochastic multiscale analysis in hydrodynamic lubrication. *Int. J. Num. Meth. Engng.*, 112:1070–1093, 2017.

P. Wriggers. *Computational Contact Mechanics*. Springer, Berlin Heidelberg New York, 2nd edition, 2006.

S. R. Wu. A penalty formulation and numerical approximation of the Reynolds-Hertz problem of elastohydrodynamic lubrication. *Int. J. Engng. Sci.*, 24:1001–1013, 1986.

B. Yang and R. F. Salant. Elastohydrodynamic lubrication simulation of O-ring and U-cup hydraulic seals. *Proc. Instn. Mech. Engrs. Part J: J. Engng. Tribol.*, 225:603–610, 2011.

O. C. Zienkiewicz and R. L. Taylor. *The Finite Element Method*. Butterworth-Heinemann, Oxford, 5th edition, 2000.

Contact Modeling in the Discrete Element Method

Jerzy Rojek

Abstract The discrete element method (DEM) is a wide family of numerical methods for discrete and discontinuous modelling of materials and systems which can be represented by a large collection of particles (discrete elements). The DEM assumes that the discrete elements interact with one another by contact forces. This chapter presents basic aspects of contact modeling in the DEM. The main assumptions, theoretical formulation and numerical algorithm of the DEM are presented. In this work, the DEM formulation employing spherical particles and the soft-contact approach is considered. Basic contact models for the particle interaction are reviewed. Elementary contact mechanisms, including elasticity, plasticity, damping, friction and cohesion are discussed. Selected contact models combining these effects are described. Their performance in modelling single dynamic or quasi-static contact events is analysed. The analysis is focused on the evolution of contact forces during single collisions. Although the force-type interaction is mainly discussed, the moment-type interaction is also introduced. Formulation of the DEM contact taking into account thermal effects as well as thermomechanical coupling finishes this review.

1 Introduction

The term discrete element method (DEM) comprises a family of numerical methods for analysis of discontinuous problems of mechanics of systems of particles (discrete elements) interacting with one another by contact. The DEM is a relatively new numerical method, it was introduced in the 70-s and 80-s of the 20th century in the pioneering works by Cundall (1971), Cundall and Strack (1979) and Walton (1982, 1983). The DEM was further developed in the works of Williams et al. (1985), Bardet and Proubet (1991), Moreau (1994) and many others. It has become a powerful tool for predicting the behaviour of various particulate and non-particulate materials

J. Rojek (✉)
Institute of Fundamental Technological Research,
Polish Academy of Sciences, Warsaw, Poland
e-mail: jrojek@ippt.pan.pl

A. Popp and P. Wriggers (eds.), *Contact Modeling for Solids and Particles*,
CISM International Centre for Mechanical Sciences 585,
https://doi.org/10.1007/978-3-319-90155-8_4

which can be represented by systems of particles, such as soils (Widuliński et al. 2009), powders (Martin et al. 2003), rocks (Cundall 1987; Potyondy and Cundall 2004; Rojek et al. 2011; Zubelewicz and Mroz 1983), concrete (Hentz et al. 2004; Wu et al. 2013), ceramics (Senapati and Zhang 2010) and even metals (Fleissner et al. 2007).

Discrete elements can be of an arbitrary shape (Rothenburg and Bathurst 1992; Tao et al. 2014; Cundall 1988), however, spherical particles are often a preferable choice (Cundall 1987; Widuliński et al. 2009; Plassiard et al. 2009) because of the simplicity of the formulation and the computational efficiency of contact detection algorithms for spherical objects. The present review will deal with the discrete element formulation employing spherical particles.

The contact algorithm plays an essential role in the discrete element method. The contact forces control the motion of the discrete elements and govern the macroscopic behaviour of the particle assembly. Two different approaches to contact treatment in the DEM can be identified, the so-called *soft-contact* approach (Cundall and Strack 1979; Cundall 1987; Potyondy and Cundall 2004) and the *hard-contact* concept (Hong and McLennan 1992; Haff and Werner 1987; Richardson et al. 2011). The soft-contact approach employs regularization of the contact constraints, while the hard-contact approach uses the methods of nonsmooth analysis to solve the problem with unilateral contact constraints.

In the soft-contact DEM formulation, a small overlap of the particles is allowed – the contact non-penetration conditions are satisfied approximately, only. The contact between the particles is assumed to last much longer than the time step, and the contact force evolution is analysed.

In the hard-contact approach, particle penetration is not allowed. The change of the particle momentum due to a collision is determined. The collision time is assumed to be very short and therefore it can be neglected. The contact force variation is not analysed.

In the present work, the soft-contact approach is considered. This approach allows us to adopt a suitable contact model for single particle collisions as well as an adequate model to obtain a required macroscopic behaviour. The present chapter is aimed to present basic concepts of contact modelling in the discrete element method.

The outline of the present chapter is as follows. The formulation of the discrete element method is presented in Sect. 2. Basic assumptions, equations of motion and time integration scheme are briefly described. Section 3 is devoted to contact modelling. Contact conditions are formulated, the penalty regularization of the contact constraints is introduced. Elementary contact mechanisms included in contact models are presented. Selected more complex contact models for force-type interaction are reviewed in Sect. 4. Their formulation and performance in simple problems are presented. The moment-type interaction is introduced in Sect. 5. An extension of the discrete element method on thermal and thermomechanical problems is presented in Sect. 6. Formulation of the thermal and thermomechanical contact is briefly discussed.

2 Discrete Element Method Formulation

2.1 Basic Assumptions

Dynamics of a system of interacting rigid cylindrical (in 2D) or spherical (in 3D) particles will be considered. Both translational and rotational motion of the particles will be taken into account. Initial positions and velocities (both linear and angular ones) of the particles are assumed to be known. The particles are subjected to an external load including point forces and moments, gravity and background damping (due to the interaction of particles with surrounding medium). The particles are assumed to interact by contact with one another and with other obstacles.

The problem to be solved is formulated as an initial-value problem defined by the ordinary differential equations (equations of motion) and appropriate initial conditions supplemented with contact constraints.

2.2 Equations of Motion

The motion of discrete elements (particles) is governed by the standard Newton–Euler equations of rigid body dynamics. The translational and rotational motion of the centre of mass of the i-th spherical or cylindrical element (Fig. 1) is described by the following equations:

$$m_i \ddot{\boldsymbol{u}}_i = \boldsymbol{F}_i \,, \tag{1}$$

$$J_i \dot{\boldsymbol{\omega}}_i = \boldsymbol{T}_i \,, \tag{2}$$

where $\boldsymbol{u}_i$ is the element centroid displacement in a fixed (inertial) coordinate frame $\boldsymbol{X}$:

$$\boldsymbol{u}_i = \boldsymbol{x}_i - \boldsymbol{X}_i \,, \tag{3}$$

$\boldsymbol{\omega}_i$ – the angular velocity, $\boldsymbol{F}_i$ – the resultant force, $\boldsymbol{T}_i$ – the resultant moment about the central axes, m_i – the element mass, and J_i is the moment of inertia which is

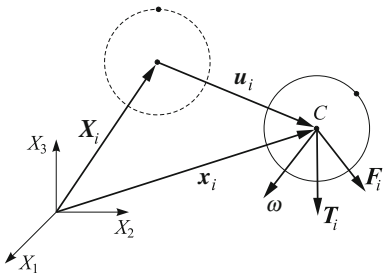

Fig. 1 Motion of a discrete element

given by:

$$J_i = \frac{1}{2} m_i R_i^2 \quad \text{for a cylinder,} \tag{4}$$

$$J_i = \frac{2}{5} m_i R_i^2 \quad \text{for a sphere,} \tag{5}$$

R_i being the i-th particle radius. The form of the rotational Eq. (2) is valid for spheres and cylinders (in 2D) and it is simplified with respect to a general form for an arbitrary rigid body with the rotational inertial properties represented by a second order tensor. The vectors $\boldsymbol{F}_i$ and $\boldsymbol{T}_i$ are sums of:

(i) all forces and moments applied to the i-th element due to an external load, $\boldsymbol{F}_i^{\text{ext}}$ and $\boldsymbol{T}_i^{\text{ext}}$, respectively,
(ii) force- and moment-type contact interactions with neighbouring spheres and all other obstacles, $\boldsymbol{F}_{ij}^{\text{cont}}$ and $\boldsymbol{T}_{ij}^{\text{cont}}$, respectively, $j = 1, \ldots, n_i^{\text{c}}$, where n_i^{c} are the number of elements being in contact with the i-th discrete element,
(iii) forces and moments resulting from external (background) damping, $\boldsymbol{F}_i^{\text{damp}}$ and $\boldsymbol{T}_i^{\text{damp}}$, respectively.

Thus, the vectors $\boldsymbol{F}_i$ and $\boldsymbol{T}_i$ can be written as follows:

$$\boldsymbol{F}_i = \boldsymbol{F}_i^{\text{ext}} + \sum_{j=1}^{n_i^{\text{c}}} \boldsymbol{F}_{ij}^{\text{cont}} + \boldsymbol{F}_i^{\text{damp}}, \tag{6}$$

$$\boldsymbol{T}_i = \boldsymbol{T}_i^{\text{ext}} + \sum_{j=1}^{n_i^{\text{c}}} \boldsymbol{s}_{ij}^{\text{c}} \times \boldsymbol{F}_{ij}^{\text{cont}} + \sum_{j=1}^{n_i^{\text{c}}} \boldsymbol{T}_{ij}^{\text{cont}} + \boldsymbol{T}_i^{\text{damp}}, \tag{7}$$

where $\boldsymbol{s}_{ij}^{\text{c}}$ is the vector connecting the centre of mass of the i-th element with the contact point with the j-th element (Fig. 2).

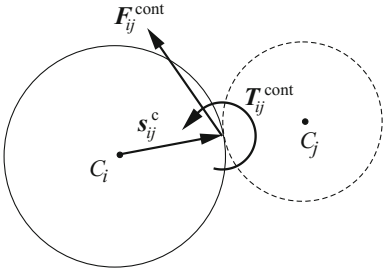

Fig. 2 Contact interaction between two discrete elements

2.3 Time Integration Scheme

Equations of motion (1) and (2) are integrated in time using the explicit central difference scheme. The time integration operator for the translational motion at the n-th time step is as follows:

$$\ddot{\boldsymbol{u}}_i^n = \frac{\boldsymbol{F}_i{}^n}{m_i}\,, \tag{8}$$

$$\dot{\boldsymbol{u}}_i^{n+1/2} = \dot{\boldsymbol{u}}_i^{n-1/2} + \ddot{\boldsymbol{u}}_i^n \Delta t\,, \tag{9}$$

$$\boldsymbol{u}_i^{n+1} = \boldsymbol{u}_i^n + \dot{\boldsymbol{u}}_i^{n+1/2} \Delta t\,. \tag{10}$$

The first two steps in the integration scheme for the rotational motion are identical to those given by Eqs. (8) and (9):

$$\dot{\boldsymbol{\omega}}_i^n = \frac{\boldsymbol{T}_i{}^n}{\boldsymbol{J}_i}\,, \tag{11}$$

$$\boldsymbol{\omega}_i^{n+1/2} = \boldsymbol{\omega}_i^{n-1/2} + \dot{\boldsymbol{\omega}}_i^n \Delta t\,. \tag{12}$$

The vector of incremental rotations $\Delta\boldsymbol{\theta}_i$ is calculated as

$$\Delta\boldsymbol{\theta}_i = \boldsymbol{\omega}_i^{n+1/2} \Delta t\,, \tag{13}$$

Knowledge of the incremental rotation suffices to update the tangential contact forces. If necessary it is also possible to track the total change of rotational position of particles (Argyris 1982). Then, the rotation matrices between the moving frames embedded in the particles and the fixed global frame must be updated incrementally using an adequate multiplicative scheme (Rojek et al. 2001).

Explicit integration in time yields high computational efficiency of the solution for a single step. The disadvantage of the explicit integration scheme is its conditional numerical stability imposing the limitation on the time step Δt. The time step Δt must not be larger than a critical time step Δt_{cr}

$$\Delta t \leq \Delta t_{\mathrm{cr}} \tag{14}$$

determined by the highest natural frequency of the system $\nu_{\max}$

$$\Delta t_{\mathrm{cr}} = \frac{2}{\nu_{\max}}\,. \tag{15}$$

Exact determination of the highest frequency $\nu_{\max}$ would require solution of the eigenvalue problem defined for the whole system of connected rigid particles. The maximum frequency of the whole system can be estimated as the maximum of

natural frequencies ν_i^e of subsets of connected particles surrounding each particle e, cf. Belytschko et al. (1985):

$$\nu_{\max} \leq \nu_{\max}^D, \quad \text{where} \quad \nu_{\max}^D = \max_{i,e} \nu_i^e \tag{16}$$

3 Contact Modelling

3.1 Contact Conditions

The contacting pairs of discrete elements are identified by different search procedures. The contacting particles should satisfy contact constraints. The contact constraints can be expressed in terms of contact interactions and appropriate kinematic parameters.

It is assumed that the contact is concentrated at a point called the contact point and the contact interaction between two particles i and j consists of a concentrated force $\boldsymbol{F}_{ij}^{\text{cont}}$ and a concentrated moment $\boldsymbol{T}_{ij}^{\text{cont}}$ applied at the contact point (Fig. 2). The moment-type resistance will be discussed later on, here, the force interaction will be examined.

The contact force $\boldsymbol{F}_{ij}^{\text{cont}}$ can be decomposed into the normal and tangential components, $(\boldsymbol{F}_{\text{n}}^{\text{cont}})_{ij}$ and $(\boldsymbol{F}_{\text{t}}^{\text{cont}})_{ij}$, respectively

$$\boldsymbol{F}_{ij}^{\text{cont}} = (\boldsymbol{F}_{\text{n}}^{\text{cont}})_{ij} + (\boldsymbol{F}_{\text{t}}^{\text{cont}})_{ij} = (F_{\text{n}}^{\text{cont}})_{ij}\boldsymbol{n}_{ij} + (\boldsymbol{F}_{\text{t}}^{\text{cont}})_{ij}, \tag{17}$$

where $\boldsymbol{n}_{ij}$ is the unit vector normal at the contact point defined as follows

$$\boldsymbol{n}_{ij} = \frac{\boldsymbol{x}_j - \boldsymbol{x}_i}{\|\boldsymbol{x}_j - \boldsymbol{x}_i\|} \tag{18}$$

It is assumed that the normal and tangential contact can be decoupled and can be considered separately. The Signorini conditions for the unilateral (without adhesion) contact in the normal direction can be written as follows[1]:

$$F_{\text{n}}^{\text{cont}} \leq 0, \quad g \geq 0, \quad F_{\text{n}}^{\text{cont}} g = 0, \tag{19}$$

where g is the gap between the particles (see Fig. 3):

$$g = d_{ij} - R_i - R_j \tag{20}$$

d_{ij} being the distance between the particle centres

[1] In the next part of this section indices denoting the elements will be omitted.

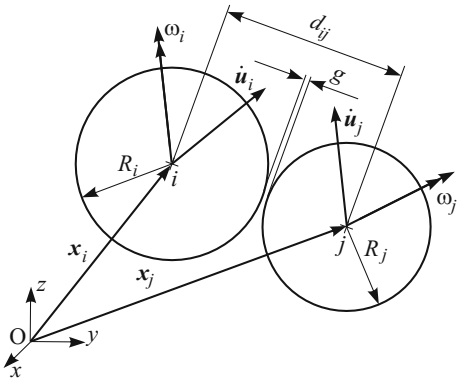

Fig. 3 Definition of geometrical and kinematical parameters in the contact of two particles

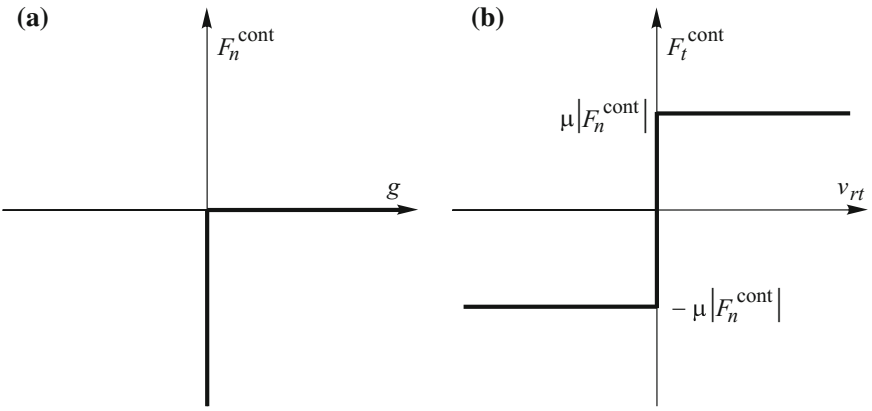

Fig. 4 Graphs of contact laws: **a** unilateral normal contact, **b** tangential Coulomb friction contact

$$d_{ij} = \|\boldsymbol{x}_j - \boldsymbol{x}_i\| \tag{21}$$

The first inequality in Eq. (19) expresses the intensility condition (no tensile force is allowed), the second one specifies the impenetrability condition, and the third condition, called the complementarity condition, enforces the alternative, either $F_{\text{n}}^{\text{cont}} < 0$ and $g = 0$ or $F_{\text{n}}^{\text{cont}} = 0$ and $g > 0$. The unilateral normal contact law is illustrated graphically in Fig. 4a. In the adhesive (bilateral) contact, the tensile contact force is allowed, and the geometric inequality constraint $(19)_2$ is replaced by the equality constraint $g = 0$ (Curnier 1999). A rigorous mathematical treatment of the bilateral contact is much more complicated than that of the unilateral contact.

The tangential interaction is typically caused by interparticle friction. The complementary conditions for the frictional sliding contact in the tangential direction can be written as follows, cf. Klarbring (1999):

$$\phi_t \leq 0\,, \quad \lambda_t \geq 0\,, \quad \phi_t \lambda_t = 0\,, \tag{22}$$

where ϕ_t is the slip criterion and the non-negative parameter λ_t is defined by the slip law:

$$\boldsymbol{v}_{\mathrm{rt}} = \lambda_t \frac{\boldsymbol{F}_{\mathrm{t}}^{\mathrm{cont}}}{\|\boldsymbol{F}_{\mathrm{t}}^{\mathrm{cont}}\|}\,. \tag{23}$$

where $\boldsymbol{v}_{\mathrm{rt}}$ is the relative tangential velocity at the contact point:

$$\boldsymbol{v}_{\mathrm{rt}} = \boldsymbol{v}_{\mathrm{r}} - v_{\mathrm{rn}} \boldsymbol{n}_{ij}\,. \tag{24}$$

with $\boldsymbol{v}_{\mathrm{r}}$ and v_{rn} being the total and normal relative velocities at the contact point given by

$$\boldsymbol{v}_{\mathrm{r}} = (\dot{\boldsymbol{u}}_j + \boldsymbol{\omega}_j \times \boldsymbol{s}_{ji}^c) - (\dot{\boldsymbol{u}}_i + \boldsymbol{\omega}_i \times \boldsymbol{s}_{ij}^c)\,, \tag{25}$$

$$v_{\mathrm{rn}} = \boldsymbol{v}_{\mathrm{r}} \cdot \boldsymbol{n}_{ij}\,, \tag{26}$$

where $\dot{\boldsymbol{u}}_i$ and $\dot{\boldsymbol{u}}_j$ are the translational velocities of the particle mass centres, $\boldsymbol{\omega}_i$ and $\boldsymbol{\omega}_j$ – the angular particle velocities, and $\boldsymbol{s}_{ij}^c$ are $\boldsymbol{s}_{ji}^c$ – the vectors connecting the particle mass centres to the contact points.

There are various models for the threshold of sliding, cf. Raous (1999). The most commonly used model is the Coulomb friction model, for which the slip criterion is given by:

$$\phi_t = \|\boldsymbol{F}_{\mathrm{t}}^{\mathrm{cont}}\| - \mu |F_{\mathrm{n}}^{\mathrm{cont}}| \leq 0 \tag{27}$$

where μ is the Coulomb friction coefficient. The graph corresponding to the Coulomb friction contact is given in Fig. 4b. The Coulomb friction coefficient is usually assumed constant, however, it can also be taken as a variable, for instance, dependent on the sliding velocity.

3.2 Regularization of the Contact Conditions

The discrete element method based on the soft contact approach imposes a penalty-type regularization of the unilateral (normal) and frictional contact constraints. The penalty regularization of the normal contact conditions is accomplished by taking

$$F_{\mathrm{n}}^{\mathrm{cont}} = k_n g\,, \quad \text{if} \quad g < 0\,, \tag{28}$$

where k_n is a certain penalty parameter. The contact conditions $(19)_1$ and $(19)_3$ are still valid. The impenetrability condition $(19)_2$ is satisfied approximately, only. A certain overlap between the contacting particles

$$h = -g > 0 \tag{29}$$

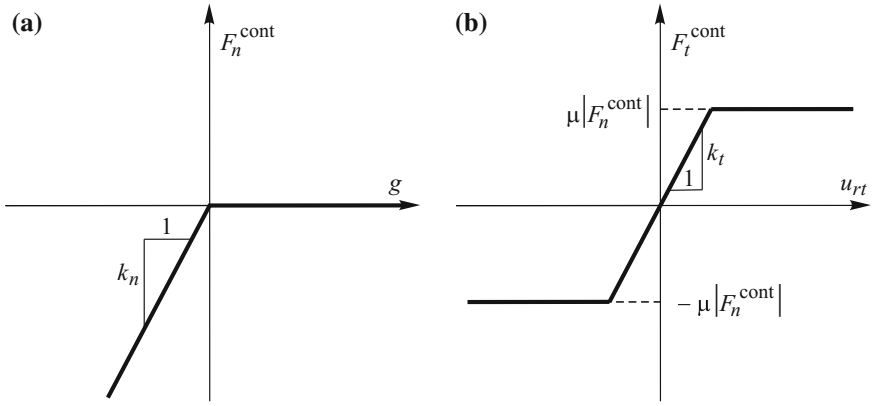

Fig. 5 Graphs of regularized contact laws: **a** normal contact, **b** tangential Coulomb friction contact

is allowed. The penalization is exact if $k_n \to \infty$. We should remember, however, that large penalty values lead to small critical time steps given by Eq. (15). The penalization of the normal contact is illustrated graphically in Fig. 5a.

The regularization of the frictional constraints is carried out by introducing into Eq. (23) a tangential penalty k_t

$$\boldsymbol{v}_{\text{rt}} - \lambda_t \frac{\boldsymbol{F}_{\text{t}}^{\text{cont}}}{\|\boldsymbol{F}_{\text{t}}^{\text{cont}}\|} = \frac{\dot{\boldsymbol{F}}_{\text{t}}^{\text{cont}}}{k_t}\,, \tag{30}$$

Equation (30) shows that the penalty regularization of the Coulomb frictional constraints introduces a decomposition of the total slip velocity $\boldsymbol{v}_{\text{rt}}$ into the reversible and irreversible parts, $\boldsymbol{v}_{\text{rt}}^{\text{r}}$ and $\boldsymbol{v}_{\text{rt}}^{\text{ir}}$, respectively:

$$\boldsymbol{v}_{\text{rt}} = \boldsymbol{v}_{\text{rt}}^{\text{r}} + \boldsymbol{v}_{\text{rt}}^{\text{ir}}\,, \tag{31}$$

where

$$\boldsymbol{v}_{\text{rt}}^{\text{r}} = \frac{\dot{\boldsymbol{F}}_{\text{t}}^{\text{cont}}}{k_t}\,, \tag{32}$$

$$\boldsymbol{v}_{\text{rt}}^{\text{ir}} = \lambda_t \frac{\boldsymbol{F}_{\text{t}}^{\text{cont}}}{\|\boldsymbol{F}_{\text{t}}^{\text{cont}}\|}\,. \tag{33}$$

The graph corresponding to the regularized Coulomb friction model is shown in Fig. 5b, where the tangential contact force $F_{\text{t}}^{\text{cont}}$ has been plotted as a function of the relative tangential displacement evaluated by integrating the relative tangential velocity

$$u_{\text{rt}} = \int v_{\text{rt}}\, \mathrm{d}t\,. \tag{34}$$

3.3 Physical Interpretation of the Penalty Regularization

Penalization of the normal and tangential contact constraints is equivalent to specifying additional constitutive relations on the interface. Thus, the elastic behaviour has been introduced above for compression in the normal contact, and the frictional contact has been reformulated as a problem analogous to that of elastoplasticity.

The overlap of the contacting particles h defined by Eq. (29) is assumed to represent an effect of a local deformation of the particles at the contact point due to the contact interaction (Fig. 6). This assumption gives a possibility to define different force–overlap relationships in order to represent better various deformation mechanisms at the contact zone. Similarly, different force–slip relationships can be defined for the tangential contact.

It must be remarked that it is assumed that the particle deformation due to contact is localized and it does not affect other particle contacts. This assumption is justified if the deformation of real particles is relatively small.

3.4 Elementary Contact Deformation Mechanisms

Contact models may take into account different deformation mechanisms and physical phenomena involved in contact. Typical elementary contact deformation mechanisms and associated effects are summarized below.

Elasticity

The contact force in the normal or tangential direction is given by a linear or nonlinear function of a displacement-type variable u. A linear elastic model defines the force through the relation

$$F = ku\,, \tag{35}$$

where k is a constant stiffness parameter, and the displacement-type variable u can represent the particle overlap or the tangential relative displacement. A nonlinear

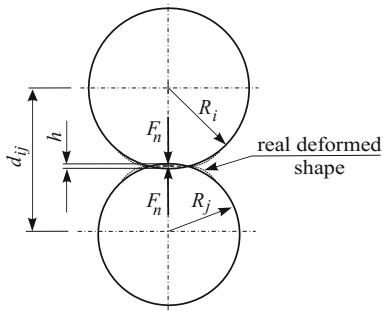

Fig. 6 Overlap of the contacting particles

elastic model is characterized by a variable stiffness. The displacement in an elastic contact model is completely reversible.

Plasticity
A linear or nonlinear force–displacement relationship is obtained from the response of particles undergoing plastic deformation due to contact pressure. The displacement in an ideally plastic contact model is completely irreversible. The theory of plasticity provides a suitable framework for modelling the friction.

Viscosity
A contact force due to a viscous response at the contact interaction is defined by a force–velocity relationship:

$$F = \eta \dot{u} \,, \tag{36}$$

where η represents viscous properties of the contact interface. It can be assumed constant in a linear viscous model or variable in a nonlinear model. A viscous model can be used to represent physical phenomena such as damping or creep.

Friction
Friction being a dissipation mechanism opposing the tangential relative motion of contacting particles is sometimes called dry damping, cf. Zonetti et al. (1999), as opposed to velocity dependent viscous damping mentioned above. The Coulomb model is the most popular model of friction. The graph presenting the contact force in the Coulomb model is presented in Fig. 4b.

Cohesion/adhesion
Modelling of cohesive materials such as rocks or concrete with discrete elements requires accounting for cohesion or adhesion in the contact model. Cohesive/adhesive bonds are introduced between contacting particles. These bonds transfer contact forces opposing the separation of particles in the normal direction as well as the relative motion in the tangential direction.

Damage
Damage represents deterioration of mechanical material properties such as stiffness and strength caused by the development of internal cracks. Accounting for damage effects in the contact allows us to consider a gradual deterioration of mechanical properties represented by cohesive bonds.

Fracture
Accumulated damage can lead to a complete deterioration of the cohesive bonds. The cohesive bonds can also be assumed to be broken in a brittle way when the strength of the cohesive bonds is exceeded. In this way initiation and development of fracture in the material can be modelled in the discrete element method.

Thermal effects
Contact with friction is accompanied by heat generation. The heat generated through friction is absorbed and conducted by the particles. An increasing temperature of contacting particles may affect mechanical contact properties such as the contact stiffness, viscosity or friction coefficient (Shillor et al. 2004). Formulation of the

contact model accounting for thermal effects as well as thermal and thermomechanical formulation of the discrete element method will be presented further on.

Contact models in the discrete element method usually incorporate different mechanisms and effects described above which allow us to model complex behaviour of real materials. Contact models similarly to constitutive material models are often represented graphically by rheological schemes. The rheological schemes are built from rheological elements representing elementary mechanisms. Typical rheological elements are shown in Fig. 7.

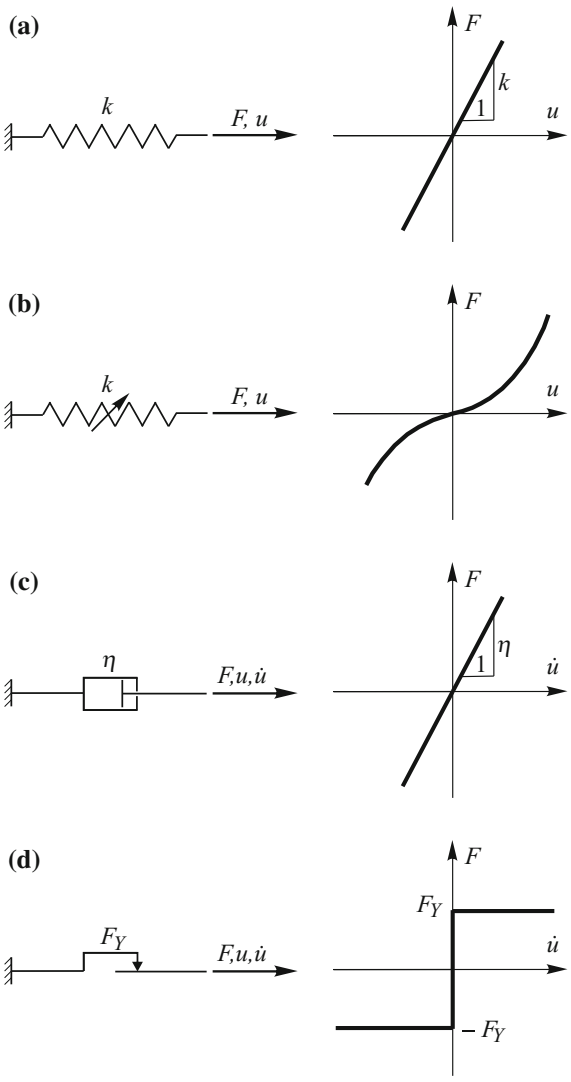

Fig. 7 Rheological elements and plots of the corresponding constitutive relationships: **a** linear spring, **b** nonlinear spring, **c** linear dashpot, **d** slider

The linear and nonlinear springs (Fig. 7a, b) represent elastic properties in a rheological scheme, the linear dashpot (Fig. 7b) corresponds to the viscous effects described by Eq. (36), the slider (Fig. 7d) is used for the friction, the slider alone represents the non-regularized Coulomb friction model, when connected in series with a spring, it can represent the regularized Coulomb friction model.

4 Selected Contact Models

4.1 Linear Viscoelastic Contact Model with Coulomb Friction

Formulation of the model

The model presented here is similar to the model proposed in the pioneering work by Cundall and Strack (1979). The rheological scheme of the model is shown in Fig. 8. The normal contact force is represented by the viscoelastic Kelvin–Voigt element composed of a linear spring connected in parallel with a linear dashpot. The element corresponding to the tangential contact force is constituted by a spring in series with a slider.

The normal contact force $F_{\mathrm{n}}^{\mathrm{cont}}$ transmitted by the Kelvin–Voigt element is composed of the elastic part transferred by the spring $F_{\mathrm{n}}^{\mathrm{e}}$ and the viscous damping part transferred by the dashpot $F_{\mathrm{n}}^{\mathrm{d}}$:

$$F_{\mathrm{n}}^{\mathrm{cont}} = F_{\mathrm{n}}^{\mathrm{e}} + F_{\mathrm{n}}^{\mathrm{d}} \tag{37}$$

The elastic part is evaluated according to the linear relationship analogical to Eq. (28)

$$F_{\mathrm{n}}^{\mathrm{e}} = k_{\mathrm{n}} g \,, \tag{38}$$

where k_{n} is the normal contact stiffness and g is defined by Eq. (20). The formula (38) is used for $g < 0$. A negative value of g denotes a particle overlap. If $g \geq 0$ the elastic contact force is set to zero ($F_{\mathrm{n}}^{\mathrm{e}} = 0$).

The damping part is evaluated according to the linear relationship analogical to Eq. (36)

$$F_{\mathrm{n}}^{\mathrm{d}} = c_{\mathrm{n}} v_{\mathrm{rn}} \,, \tag{39}$$

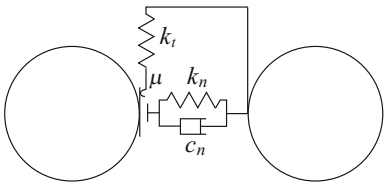

Fig. 8 Rheological scheme of the viscoelastic contact model with the Coulomb friction

where c_{n} is the normal viscous damping parameter and v_{rn} is the normal relative velocity at the contact defined by Eq. (26).

The tangential part of the contact model presented graphically in Fig. 8 corresponds to the regularized Coulomb friction model described in Sect. 3.2. The stiffness of the linear spring k_t corresponds to the penalty parameter introduced in Eq. (30). The slider is blocked until the slip criterion (27) is achieved. The graph of the tangential contact force versus relative tangential displacement is shown in Fig. 5.

The analogy of the regularized frictional contact model to the elastoplasticity commented above, allows us to calculate the friction force employing the radial return algorithm analogous to that used in elastoplasticity. First a trial state is calculated

$$\boldsymbol{F}_t^{\mathrm{trial}} = \boldsymbol{F}_t^{\mathrm{old}} - k_t \boldsymbol{v}_{rt} \Delta t \,, \tag{40}$$

and then the slip condition is checked

$$\phi^{\mathrm{trial}} = \|\boldsymbol{F}_t^{\mathrm{trial}}\| - \mu |F_n| \,. \tag{41}$$

If $\phi^{\mathrm{trial}} \leq 0$, we have the case of stick contact and the friction force is assigned the trial value

$$\boldsymbol{F}_t^{\mathrm{new}} = \boldsymbol{F}_t^{\mathrm{trial}} \,, \tag{42}$$

otherwise (slip contact) a return mapping is performed

$$\boldsymbol{F}_t^{\mathrm{new}} = \mu |F_n| \frac{\boldsymbol{F}_t^{\mathrm{trial}}}{\|\boldsymbol{F}_t^{\mathrm{trial}}\|} \,. \tag{43}$$

Evaluation of the model parameters

There are different approaches to evaluating the contact stiffness k_n in the DEM. It can be taken as uniform in the whole discrete element assembly (Rojek et al. 2005) or it can be calculated locally, usually assuming that it depends on the contacting particle size (Potyondy and Cundall 2004) and it can be given by certain functions of the particle radii R_i and R_j:

$$k_n = f_k(R_i, R_j) \,. \tag{44}$$

Different assumptions for the form of the functions $f_k(R_i, R_j)$ have been discussed by Rojek et al. (2012). Here, one of them will be presented.

The spring modelling contact elasticity can be treated as equivalent to an elastic bar of a non-uniform cross-sectional area (Fig. 9), consisting of two segments, with the lengths

$$L_i = R_i \,, \quad L_j = R_j \tag{45}$$

and the cross-sectional areas

$$A_i = \alpha_i \pi (R_i)^2 \,, \quad A_j = \alpha_j \pi (R_j)^2 \tag{46}$$

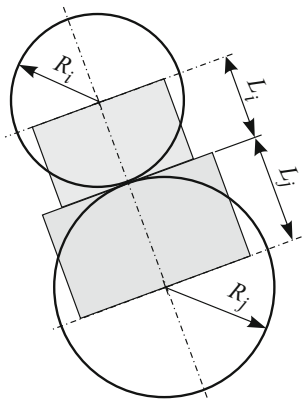

Fig. 9 Schematic connection of two particles

where $0 \leq \alpha_i, \alpha_j \leq 1$ are the coefficients defining the areas of the segments as fractions of the particle cross-sectional area.

The system of the two bar segments can be treated as two springs connected in series. The axial force F^e transferred by the whole system is equal to the forces in the segments i and j, F_i^e and F_j^e:

$$F^e = F_i^e = F_j^e \tag{47}$$

The overall axial deformation of the system, assumed as equal to the overlap g ($g < 0$), can be decomposed to the deformations of both segments, g_i and g_j

$$g = g_i + g_j \tag{48}$$

The force–displacement relationships for each bar can be written in the following form:

$$F_i^e = k_n^i g_i \tag{49}$$

$$F_j^e = k_n^j g_j \tag{50}$$

where and k_n^i and k_n^j are stiffnesses of the segments i and j. Substituting Eqs. (38), (49) and (50) into Eq. (48) and taking into account Eq. (47) we obtain the following equation for the stiffness k_n:

$$\frac{1}{k_n} = \frac{1}{k_n^i} + \frac{1}{k_n^j} \tag{51}$$

which can be transformed to the form

$$k_n = \frac{k_n^i k_n^j}{k_n^i + k_n^j} \tag{52}$$

Expression (52) is identical to that used by Potyondy and Cundall (2004).

Using the assumptions (45) and (46) the stiffness of the segments i and j can be expressed as follows:

$$k_n^i = \frac{E_i A_i}{L_i} = \alpha_i \pi E_i R_i \tag{53}$$

$$k_n^j = \frac{E_j A_j}{L_j} = \alpha_j \pi E_j R_j \tag{54}$$

where E_i and E_j are Young's moduli of the materials of the segments (or of the particles) i and j. Introducing the relationships (53) and (54) into the formula (52) and assuming $E_i = E_j = E$ and $\alpha_i = \alpha_j = \alpha$, we obtain the expression for the equivalent stiffness K in the following form:

$$k_n = \alpha \pi E R^* \tag{55}$$

where R^* is the effective radius defined in terms of the particle radii, R_i and R_j

$$\frac{1}{R^*} = \frac{1}{R_i} + \frac{1}{R_j}. \tag{56}$$

For equal size particles ($R_i = R_j = R$), Eq. (55) takes the form:

$$k_n = \frac{1}{2} \alpha \pi E R \tag{57}$$

The value of the tangential stiffness parameter k_t is, in principle, independent of the normal stiffness parameter, however, it is usually defined with respect to it assuming a certain ratio β of the normal and tangential stiffness

$$\beta = \frac{k_t}{k_n} \tag{58}$$

The ratio β is very important, since it has a large influence on the macroscopic behaviour reproduced by the DEM model. Equivalent macroscopic properties, such as the Young's modulus or the Poisson's ratio can be presented as functions of the ratio k_t/k_n, cf. Marczewska et al. (2016).

The damping coefficient c_n can be related to the critical damping C_cr of the considered system:

$$\zeta = \frac{c_\mathrm{n}}{C_\mathrm{cr}} \tag{59}$$

where ζ is called the damping ratio. It is a non-negative dimensionless parameter ($\zeta \geq 0$). Zero damping ratio, $\zeta = 0$, indicates no damping, $0 < \zeta < 1$ – underdamping, $\zeta = 1$ – critical damping, and $\zeta > 1$ – overdamping. The critical damping C_cr for the system of two rigid bodies with masses m_i and m_j, connected with a spring of the stiffness k_n, cf. Taylor and Preece (1992)

$$C_{\mathrm{cr}} = 2\sqrt{m^* k_{\mathrm{n}}}\,. \tag{60}$$

where the effective mass m^* is defined by

$$\frac{1}{m^*} = \frac{1}{m_i} + \frac{1}{m_j}\,. \tag{61}$$

The damping ratio ζ can be expressed in terms of the coefficient of restitution (COR) e, cf. Nagurka and Huang (2006)

$$\zeta = -\frac{\ln e}{\sqrt{\pi^2 + (\ln e)^2}}\,. \tag{62}$$

The coefficient of restitution e in the normal direction is defined as the ratio of moduli of the relative normal velocities after and before impact, $v_{\mathrm{rn}}^{\mathrm{end}}$ and v_{rn}^{0}, respectively:

$$e = \frac{|v_{\mathrm{rn}}^{\mathrm{end}}|}{|v_{\mathrm{rn}}^{0}|}\,. \tag{63}$$

The relationship (62) is plotted in Fig. (10).

The viscous damping is used in the discrete element method as a mechanism allowing to dissipate energy in particle collisions, and achieve different response of the system to dynamic loading, including quasistatic response if an adequate damping is combined with a slowly applied loading. It must be remarked, however, that the viscous damping introduces certain inconsistencies in the contact model, which will be discussed in the first of numerical examples below.

Numerical examples

Collision of two balls with given initial velocities

The viscoelastic Kelvin–Voigt contact model is used to simulate a collision of two equal balls of radius $R = 10\,\mathrm{mm}$ moving along one line with equal but opposed

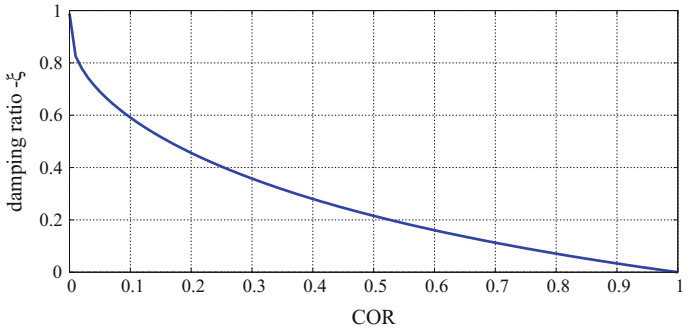

Fig. 10 Relationship between the coefficient of restitution and the damping ratio

velocities $v = 10\,\text{m/s}$. The mass density $\rho = 8000\,\text{kg/m}^3$ has been assumed. The contact stiffness k_n has been evaluated according to Eq. (57) taking the Young's modulus $E = 200\,\text{GPa}$, and the coefficient $\alpha = 0.04$. The effective contact stiffness $k_n = 1.26 \cdot 10^8\,\text{N/m}$ has been evaluated. Different values of damping characterized by coefficients of restitution (COR) $e = 0.1$, 0.5, 0.8 and 1 have been considered. The value $e = 1$ corresponds to an ideally elastic collision.

The gap between the balls and the velocity of one of the balls during the collision are plotted as functions of time in Fig. 11 for different damping. The duration of the collisions corresponds to the interval with negative values of the gap (see Fig. 11a. It can be seen Fig. 11b that after the collision the balls bounce off each other with the velocity dependent on the damping. In the elastic collision, the velocity of the ball after the collision is the same as before the collision. In the inelastic collisions, the velocity of the ball after the collision is lower than before the collision. The higher the damping is (or in other words, the lower the COR is), the lower the rebound velocity is.

Evolution of the total contact force and its components is plotted in Fig. 12 for different damping. The elastic, damping and total contact forces are given as functions of time in Figs. 12a, b and c, respectively. By comparing Figs. 11 and 12, it can be seen that the elastic force is proportional to the gap in agreement with Eq. (38), and

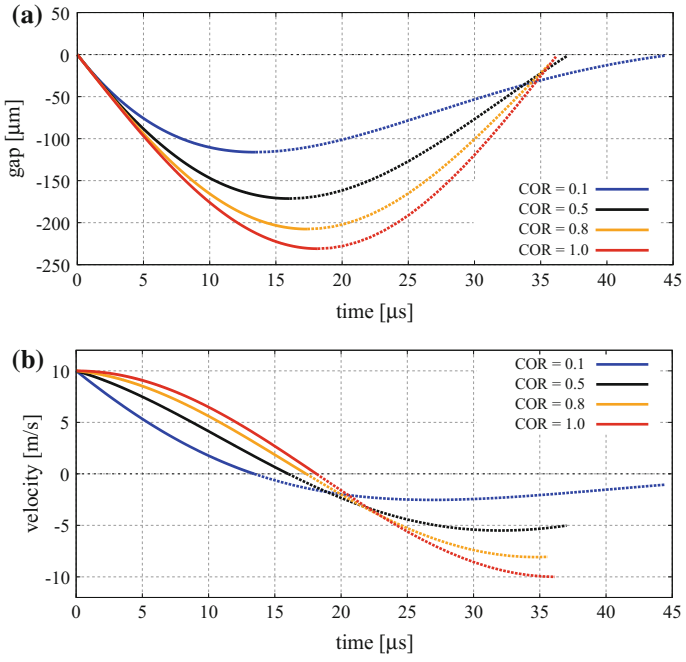

Fig. 11 Evolution of contact kinematic parameters during collision of two balls in the linear Kelvin–Voigt model for different values of COR: **a** gap versus time, **b** velocity versus time (the solid parts of the curves represent loading and the broken ones – unloading)

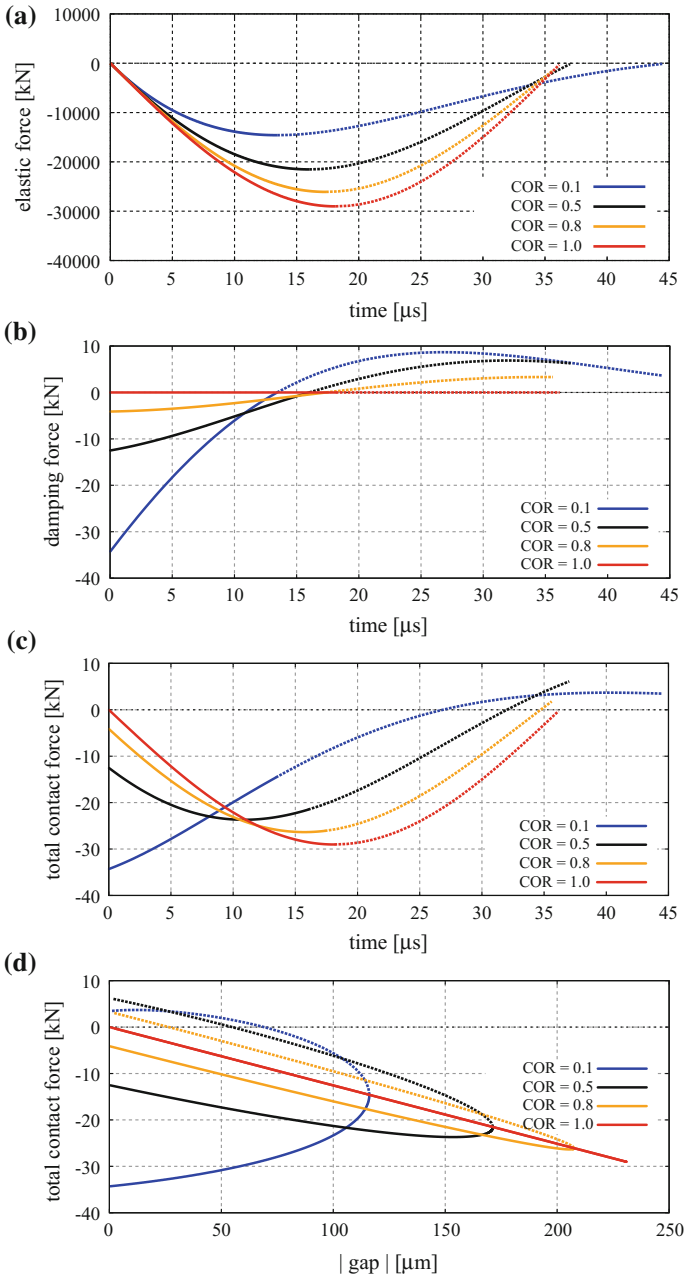

Fig. 12 Evolution of contact forces during collision of two balls in the linear Kelvin–Voigt model for different values of COR (the solid parts of the curves represent loading and the broken ones – unloading)

the damping force is proportional to the velocity in agreement with Eq. (39). It can be observed in Fig. 12b that the damping force acquires a certain non-zero value at the beginning of the collision, and it has a non-zero value at the end of the collision. This, with the zero damping force before and after the collision, leads to discontinuity of the damping component during the analysed period of time. Due to the discontinuity of the viscous damping force, the total contact force $F_{\mathrm{n}}^{\mathrm{cont}}$ displayed in Fig. 12c is also discontinuous at the beginning and end of the collision while real contact forces are continuous.

Moreover, due to the viscous damping the total contact force $F_{\mathrm{n}}^{\mathrm{cont}}$ is cohesive in the final stage of the collision ($g < 0$ and $v_{\mathrm{rn}} > 0$), while the contact forces in cohesionless particle systems are always repulsive.

The inconsistencies of the linear viscoelastic contact model can be mitigated, at least partially, by replacing the linear spring and dashpot with appropriate nonlinear spring and damping elements as proposed by Hunt and Crossley (1975).

The curves of the total force versus gap for different damping are displayed in Fig. 12d. In the ideally elastic case (COR = 1), the loading and unloading force–gap relations coincide. In the damped collisions, the loading and unloading curves do not coincide and form a hysteresis loop. The area within the loop is a measure of energy lost during the collision (Lin and Hui 2002).

Contact of two spheres under step loading

A contact of two equal balls of radius $R = 10\,\mathrm{mm}$ subjected to step compressive loading $F = 20\,\mathrm{kN}$ has been analysed using the linear viscoelastic Kelvin–Voigt model. The same material properties as in the previous example have been assumed. The initial conditions are defined by the zero gap and zero ball velocities. Effect of the damping has been studied taking different values of coefficients of restitution COR = 0.05, 0.5, 0.8 and 1.

Figure 13 shows the time response of the systems in terms of the gap, velocity of one of the balls and total contact force. It can be seen that in the system with zero damping (COR = 1) the balls oscillate with the period which is in perfect agreement with the theoretical value

$$T = 2\pi\sqrt{\frac{m^*}{k_n}} = 72.6\ \mu\mathrm{s}\,.$$

The mean value of the gap (actually the overlap) oscillations coincides with the static state of equilibrium under the applied force

$$g = \frac{F_n}{k_n} = -159.1\ \mu\mathrm{m}\,.$$

The oscillations of the systems with damping are attenuated and the quasistatic state of equilibrium is achieved with the gap calculated above.

This shows a possibility to use the dynamic formulation to solve static problems which is a basic principle of the dynamic relaxation method, employed both in the

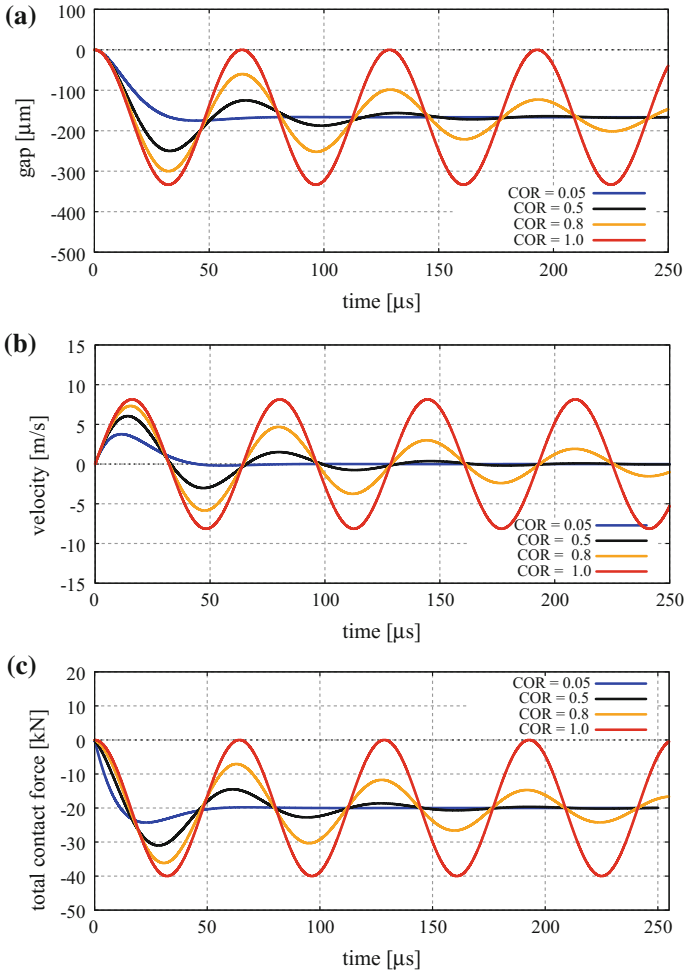

Fig. 13 Evolution of contact kinematic variables and contact forces during contact of two spheres under step loading for different values of COR

discrete element method (Rojek et al. 2013) and the finite element method (Joldes et al. 2011). In the elastic linear problems solved by the dynamic relaxation method, the solution in the transient period is not important, and different values of damping allow us to arrive at the same static solution. The dynamic relaxation method can also be applied with certain cautiousness to path dependent problems.

Contact of two spheres under linearly increasing loading

The system of two balls, the same as in the previous example and with the same initial conditions, has been subjected to a compressive loading increasing linearly from 0 to 20 kN in the time interval from 0 to 0.5 s. Then, for $t > 0.5$ s the loading

has been kept constant. The contact of the balls has been analysed using the linear viscoelastic Kelvin–Voigt model with different values of damping defined by the coefficients of restitution COR = 0.1, 0.5 and 1.

Figure 14 shows the time response of the systems with different damping in terms of the gap, velocity of one of the balls and total contact force. The solution with zero damping (COR = 1) is characterized with oscillations. The oscillations are attenuated in the solutions with damping. For a sufficiently high level of the damping (low values of COR), the response in terms of the gap and contact force is practically linear, in agreement with the linear increase of the applied force. A

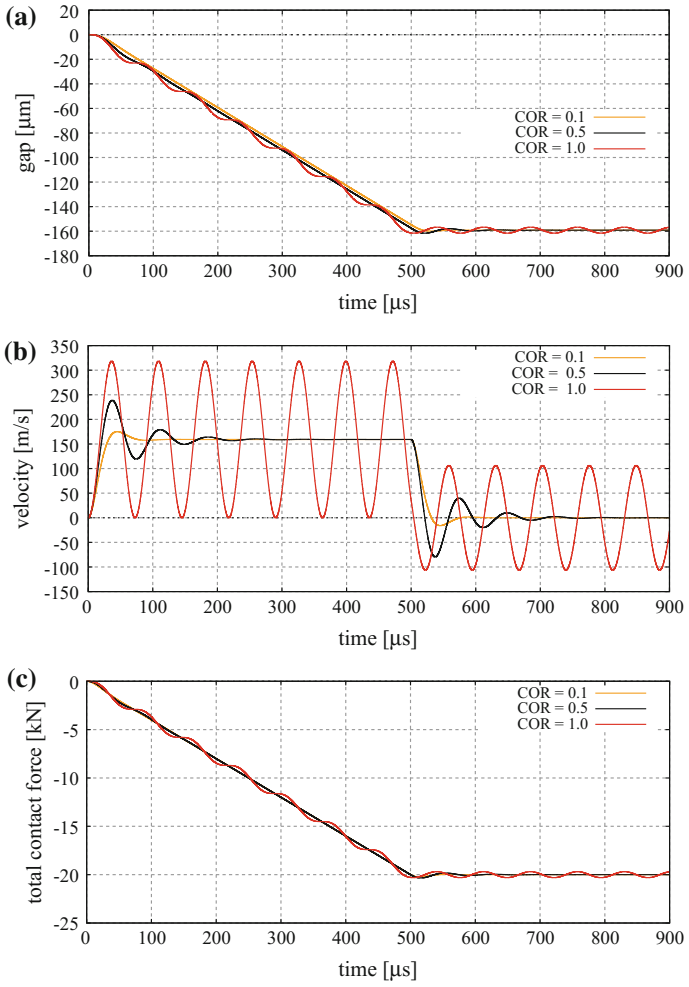

Fig. 14 Evolution of contact kinematic variables and contact forces during contact of two spheres under linearly increasing loading for different values of COR

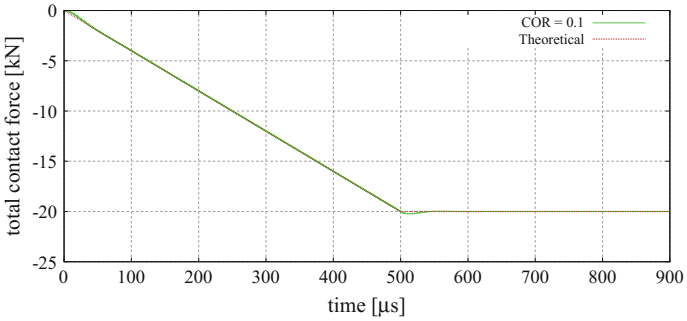

Fig. 15 Comaprison of the numerical and theoretical contact force during contact of two spheres under linearly increasing loading

perfect agreement of the contact force with the applied force can be observed in Fig. 14, which confirms that the loading and response can be considered quasistatic. This demonstrates a possibility to reproduce quasistatic conditions in an incremental form (cf. also Fig. 15), which is important for the analysis of nonlinear and path dependent problems.

4.2 Viscoelastic Hertz–Mindlin–Deresiewicz Model

Formulation of the model

This model combines the Hertz-type viscoelastic model for the normal interaction with the Mindlin–Deresiewicz model of friction acting in the tangential direction. The rheological scheme of the model is similar to that shown in Fig. 8, the difference consisting in replacing the linear springs with nonlinear ones.

The Hertz model employs a nonlinear relationship for the evaluation of the elastic contact force based on the analytical solution of the contact problem between elastic spheres (Hertz 1882; Johnson 1985):

$$F_{\mathrm{n}}^{\mathrm{e}} = -K_{nHz} h^{\frac{3}{2}}, \tag{64}$$

where h ($h = -g$) is the amount the particles' overlap and the contact stiffness parameter K_{nHz} is given by the following formula:

$$K_{nHz} = \frac{4}{3} E^* \sqrt{R^*}, \tag{65}$$

where E^* is the effective modulus of elasticity defined in terms of the Young's moduli, E_i and E_j, and the Poisson's ratios, ν_i and ν_j, of the two contacting particles

$$\frac{1}{E^*} = \frac{1-\nu_i^2}{E_i} + \frac{1-\nu_j^2}{E_j}\,, \tag{66}$$

and R^* is the effective radius defined by Eq. (56). Please note that the contact force has been defined in Eq. (64) in terms of the overlap h instead of the gap g and the minus sign has been introduced in order to keep consistency with the earlier used sign convention treating the compressive contact forces as negative.

In the framework of the DEM, a viscous damping is commonly added to the elastic Hertz force in order to dissipate energy at particle collisions. A linear damping given by Eq. (39) is sometimes taken, cf. Lee (1994). More advanced models, however, use nonlinear damping terms in connection with the Hertzian elastic contact. Hunt and Crossley (1975) have derived the following general form of nonlinear damping

$$F_{\mathrm{n}}^{\mathrm{d}} = \eta_{\mathrm{n}} h^p v_{\mathrm{rn}}^q\,. \tag{67}$$

Tsuji et al. (1992) have proposed the damping term as above with $p = 1/4$ and $q = 1$:

$$F_{\mathrm{n}}^{\mathrm{d}} = \eta_{\mathrm{n}} h^{\frac{1}{4}} v_{\mathrm{rn}}\,. \tag{68}$$

The damping dissipation coefficient η_{n} used in Eq. (68) can be related to the nonlinear spring stiffness K_{nHz} and the coefficient of restitution e as follows (Navarro and de Souza Braun 2013):

$$\eta_{\mathrm{n}} = \sqrt{5}\sqrt{m^* K_{nHz}}\frac{\ln e}{\sqrt{\pi^2 + (\ln e)^2}}\,. \tag{69}$$

Different analytical relationships between the damping ratio and coefficient of restitution in the nonlinear viscoelastic contact model have been derived by (Jankowski 2006). Possibilities of improvements of viscous damping for the Hertz elastic contact are still investigated (Zdancevičius et al. 2017).

The Hertzian normal contact model is commonly combined with the tangential contact model according to Mindlin and Deresiewicz (1953). A full implementation of the Mindlin–Deresiewicz theory leads to complex algorithms, cf. (Renzo and Maio 2004; Kruggel-Emden et al. 2008), therefore different simplifications have been proposed. Employing the Mindlin and Deresiewicz solution for the constant normal force Tsuji et al. (1992) derived the formula for the tangential force:

$$F_t = k_t u_{rt}\,, \tag{70}$$

where the tangential stiffness k_t is calculated as follows:

$$k_t = 8G^*\sqrt{R^* h}\,, \tag{71}$$

the effective shear modulus is defined in terms of the particles shear moduli, G_i and G_j, and the Poisson's ratios, ν_i and ν_j, by the following relationship

$$\frac{1}{G^*} = \frac{2-\nu_i}{G_i} + \frac{2-\nu_j}{G_j}\,, \tag{72}$$

the effective radius R^* is defined by Eq. (56), h is the particle overlap and the relative tangential displacement at the contact point u_{rt} is obtained by integration of the relative tangential velocity:

$$u_{rt} = \int_0^t \boldsymbol{v}_{rt}\,\mathrm{d}t\,. \tag{73}$$

The tangential force F_t is limited by the Coulomb condition

$$F_t \le \mu |F_n|\,. \tag{74}$$

In order to improve an agreement with the full Mindlin–Deresiewicz theory Renzo and Maio (2004); Maio and Renzo (2005) proposed a correction to the model developed by Tsuji et al. (1992) consisting in scaling the stiffness given by Eq. (71) by the factor 2/3.

$$k_t = \frac{2}{3}\left(8G^*\sqrt{R^*h}\right)\,, \tag{75}$$

Numerical example

Collision of two balls with given initial velocities

The viscoelastic Kelvin–Voigt contact model with the nonlinear Hertzian elastic component and the damping component evaluated according to Eqs. (68) and (69) has been used to simulate a collision of two equal balls analysed previously with the linear Kelvin–Voigt contact model. The same ball size (radius $R = 10\,\mathrm{mm}$), properties (mass density $\rho = 8000\,\mathrm{kg/m^3}$, Young's modulus $E = 200\,\mathrm{GPa}$) and initial conditions (velocities $v = 10\,\mathrm{m/s}$) as previously have been assumed. The set of data has been completed with the Poisson's ratio $\nu = 0.3$. Similarly as previously, different values of damping characterized by coefficients of restitution COR = 0.1, 0.5, 0.8 and 1 have been considered.

The evolution of the gap between the balls and the velocity of one of the balls for different damping have been plotted in Fig. 16a and b, respectively. It can be seen that the lower the COR is, the longer the impact and the lower the rebound velocity are.

The elastic, damping and total contact forces are plotted as functions of time in Fig. 17a, b and c, respectively. It can be observed in Fig. 17b that the damping contact force in the present model is no longer discontinuous on the contrary to the damping

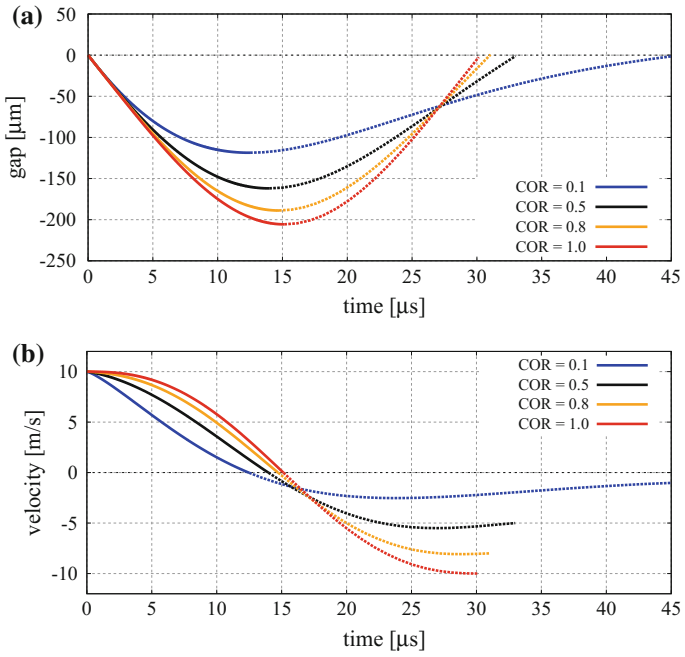

Fig. 16 Evolution of contact kinematic parameters during collision of two balls in the nonlinear Kelvin–Voigt model for different values of COR (the solid parts of the curves represent loading and the broken ones – unloading)

force in the linear Kelvin–Voigt contact model displayed in Fig. 12b. Therefore, the total contact force in Fig. 17c is not discontinuous, either. The other imperfection of the linear Kelvin-Voigt model, manifested in cohesive interaction in the final stage of impact is not eliminated in the nonlinear Kelvin-Voigt model presented here, which can be noticed in Fig. 17c and d. Figure 12d shows the curves of the total force versus the overlap (the negative gap) for different damping. It can be seen that similarly to the linear model, the loading and unloading force–gap relations for the ideally elastic case coincide, and in the damped collisions, the loading and unloading curves form a hysteresis loop. This time, unlike in the linear model, the loading and unloading curves are smooth functions for the zero gap.

4.3 Walton-Braun Elastoplastic Model

Formulation of the model

Viscous damping can be considered as a mechanism representing inelastic particle deformation during particle collision. Plastic deformation is dependent on the dis-

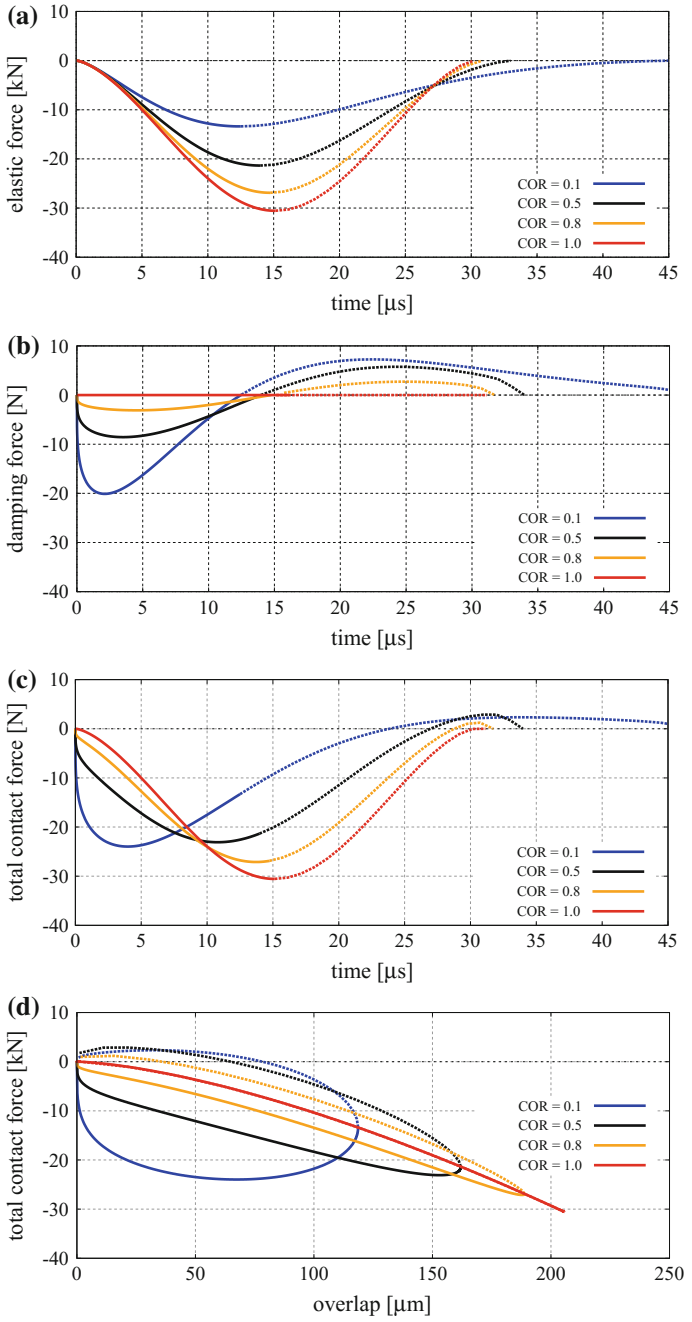

Fig. 17 Evolution of contact kinematic parameters and contact forces during collision of two balls in the nonlinear Kelvin–Voigt model for different values of COR (the solid parts of the curves represent loading and the broken ones – unloading)

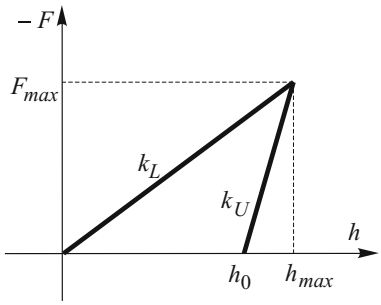

Fig. 18 Force versus particle overlap in the Walton–Braun model

placement type variable, therefore rate-independent hysteretic contact models such as the model proposed by Walton and Braun (1986) seem to be more appropriate for modelling collisions associated with plastic deformations. The Walton–Braun model assumes a linear force–overlap relationship, but the unloading slope (stiffness) is higher than the loading slope (stiffness), which leads to a certain residual irreversible overlap when the force drops to zero. This allows us to treat this model as elastoplastic with elastic unloading. The force as a function of the particle overlap is plotted in Fig. 18. Please note that although the plot is in the first quadrant of the graph, the convention of the contact force sign (compressive contact force – negative) has been kept by taking the negative of force for the vertical axis. The force is given by:

$$F = \begin{cases} -k_L h & \text{if} \quad h \geq h_{max} \quad \text{(loading)}, \\ -k_U (h - h_0) & \text{if} \quad h_0 < h < h_{max} \quad \text{(un-/reloading)}, \\ 0 & \text{if} \quad 0 < h < h_0 \quad \text{(no contact)}. \end{cases} \tag{76}$$

The residual overlap h_0 representing the plastic deformation of the contacting particles can be easily obtained as

$$h_0 = h_{max} \left(1 - \frac{k_L}{k_U} \right) \tag{77}$$

The reloding path follows the unloading path until the maximum overlap is achieved and the loading path is reactivated.

Energy is dissipated due to spring force hysteresis. The coefficient of restitution e, given by

$$e = \sqrt{\frac{k_L}{k_U}}, \tag{78}$$

is independent of the impact velocity, which is in disagreement with experimental observations. A more realistic coefficient of restitution can be obtained using a variable unloading stiffness k_U increasing with the maximum absolute force, F^{max}, or the maximum overlap, h_{max}, achieved before unloading Walton and Braun (1986), so that:

$$k_U = k_L + SF_{max} \tag{79}$$

or

$$k_U = k_L + Bh_{max}\,, \tag{80}$$

where S and B are certain constants. For the Walton-Braun contact model with variable unloading, the coefficient of restitution depends on the relative velocity of approach v_r^0 as follows Walton and Braun, 1986:

$$e = \sqrt{\frac{\omega_0}{Sv_r^0 + \omega_0}}\,, \tag{81}$$

where

$$\omega_0 = \sqrt{\frac{2k_L}{m}}\,. \tag{82}$$

Numerical example

Collision of two balls with given initial velocities

The Walton–Braun contact model has been applied to simulate a collision of two equal balls of radius $R = 10\,\text{mm}$ and of the same density $\rho = 8000\,\text{kg/m}^3$. The loading stiffness has been assumed $k_L = 1.25 \times 10^8\,\text{N/m}$, which is approximately close to the average stiffness in the Hertzian model (in the considered range) used in the numerical example in Sect. 4.2. The variable unloading stiffness has been taken according to Eq. (79) assuming $S = 10\,1/\text{m}$. The problem has been analysed assuming different initial velocities: $v = 10$, 20, 30 and 40 m/s. The evolution of the gap (overlap) between the balls and the velocity of one of the balls in the analysed cases have been plotted in Fig. 19.

The curves have been plotted for the collision time only (until the residual overlap has been achieved during unloading at each case). The coefficients of restitution for the analysed cases evaluated using the general formula (63) and predicted by the specific formula (81) for the Walton–Braun model are given in Table 1. It can be observed that the values obtained in both ways coincide. It can also be observed the values of the coefficients of restitution decrease with an increase of impact velocity, which agrees with experimental observations.

The contact forces for different impact velocities are plotted as functions of the time and overlap in Fig. 20a and b, respectively. It can be observed in Fig. 20a that the highest the impact velocity is, the shorter the collision time is. Figure 20b shows that the loading stiffness for different velocities is the same and the unloading stiffness increase with the impact velocity since higher maximum forces are achieved for higher velocities. It is assumed that the collision ends when the force during unloading decreases to zero. A certain residual overlap corresponds to this instant.

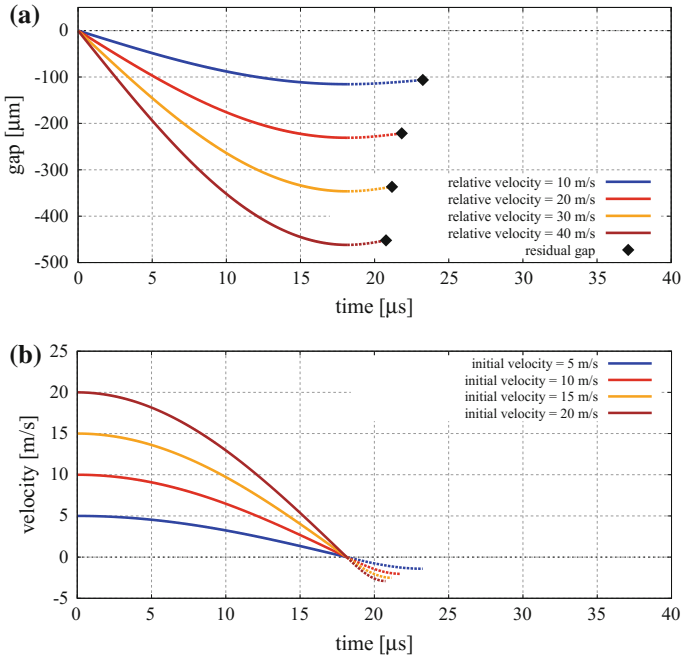

Fig. 19 Evolution of contact kinematic parameters during the collision of two balls in the Walton–Braun elastoplastic model for different values of impact velocities: **a** evolution of the gap (overlap), **b** evolution of the ball velocity (solid parts of the curves represent the loading and broken parts of the curve – the unloading)

Table 1 Coefficients of restitution for the Walton–Braun model with variable unloading

v^0 (m/s)	v_r^0 (m/s)	v_r (m/s)	e, Eq. (63)	e, Eq. (81)
5	10	2.8230	0.28230	0.28231
10	20	4.0744	0.20373	0.20373
15	30	5.0250	0.16750	0.16750
20	40	5.8228	0.14557	0.14557

4.4 Storåkers plastic model

Formulation of the model

The plastic deformation of the contacting spherical particles has been assumed in the model proposed by Storåkers et al. (1997), Storåkers et al. (1999). This model considers a general viscoplastic behaviour combining strain hardening plasticity and creep. Here, a simplified formulation of the model without strain rate effects will be presented. Such a model has been used by Olsson and Larsson (2012) to study powder compaction.

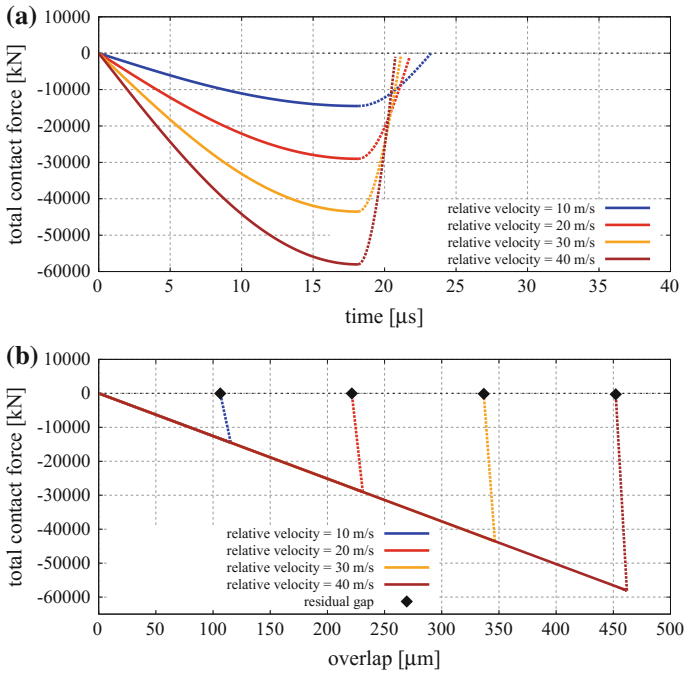

Fig. 20 Evolution of contact forces during the collision of two balls in the Walton–Braun elasto-plastic model for different values of impact velocities: **a** force versus time, **b** force versus overlap (solid parts of the curves represent the loading and broken parts of the curves – the unloading)

Two particles of radii R_i and R_j are considered. The plastic properties of the particles' material are assumed to follow the Hollomon stress-strain relationship

$$\sigma = \sigma_0 \varepsilon^m \tag{83}$$

where σ_0 and m are material constants. The normal interaction force F in the Storåkers model is given by the following equation Olsson and Larsson (2012):

$$F = -2^{1-m/2} 3^{1-m} \pi c^{2+m} \sigma_0 (R^*)^{1-m/2} h^{1+m/2}\,, \tag{84}$$

where R^* is the effective radius defined by Eq. (56), h is the particle overlap, the parameter c depends on the strain hardening exponent m:

$$c = \sqrt{1.43 \exp(-0.97m)}\,. \tag{85}$$

For the ideal plasticity, when $m = 0$ and $\sigma = \sigma_Y$, Eq. (84) is reduced to:

$$F^p = 6\pi c^2 \sigma_Y R^* h\,, \tag{86}$$

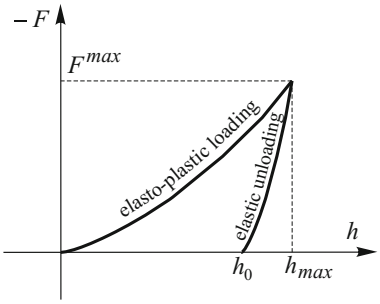

Fig. 21 Force versus particle overlap in the model combining plastic loading according to the Storakers model with the elastic Hertzian unloading

where $c^2 = 1.43$. The linear relationship (86) provides an expression for the stiffness:

$$k = 6\pi c^2 \sigma_Y R^* , \tag{87}$$

which can be used for the loading in the Walton–Braun model.

The Storåkers model has been derived neglecting elastic deformation, cf. Larsson et al. (1996). In such a model, the unloading would be governed by the rigid behaviour (no change of deformation during the unloading). Assuming that the loading curve is valid for an elastoplastic material Olsson and Larsson (2012) combined the Storåkers model with the elastic unloading according to the Hertz model. The contact force versus particle overlap for this model is plotted schematically in Fig. 21. The force during the elastic unloading as well as for the reloading is given by the formula adapted from Eq. (64):

$$F = -\frac{4}{3} E^* \sqrt{R^*} (h - h_0)^{\frac{3}{2}} , \tag{88}$$

where h_0 is obtained from Eq. (88) taking $F^{ep} = F^{max}$ and $h = h_{max}$.

Although the Storåkers model was derived for frictionless contact, it was combined with the regularized Coulomb friction model by Olsson and Larsson (2012).

Numerical example

Contact of two spheres under compressive axial load

Two equal spheres of radius $R = 10$ mm with plastic properties given by Eq. (83) have been considered assuming different values of hardening exponent $m = 0,\ 0.05,\ 0.1$ and 0.2. The yield stress $\sigma_Y = 200$ MPa has been assumed for the ideal plasticity ($m = 0$). The Hollomon constants σ_0 corresponding to the strain hardening exponents m have been evaluated assuming that all the curves pass through the point corresponding to the yield point in the elasto-plastic model: $(\sigma_Y/E, \sigma_Y)$ (taking the Young's modulus $E = 200$ GPa). Thus, the following pairs of the Hollomon constants have been determined: ($m = 0$, $\sigma_0 = 800$ MPa), ($m = 0.05$, $\sigma_0 = 1480$ MPa), ($m = 0.1$, $\sigma_0 = 1375$ MPa) and ($m = 0.2$, $\sigma_0 = 2363$ MPa). The corresponding stress–strain curves are plotted in Fig. 22.

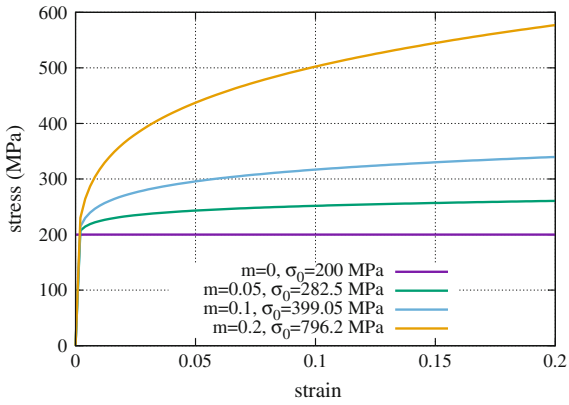

Fig. 22 Stress–strain curves for different strain hardening

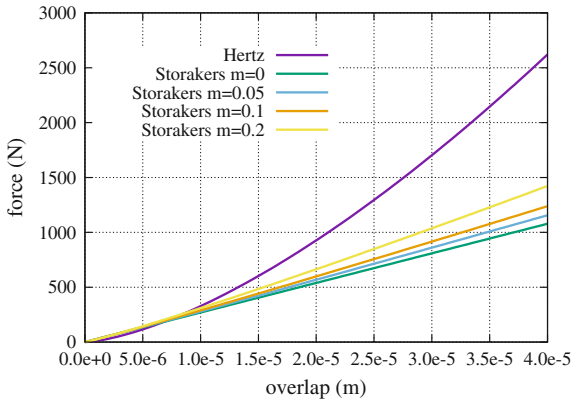

Fig. 23 Force versus particle overlap in the Storakers model in comparison the Hertz model

The compressive loading has been introduced by prescribing the displacements to the particle centres. The evolution of the contact forces F predicted by Storåkers models as functions of the particle overlap h is shown in Fig. 23 in comparison to the Herz model.

4.5 Thornton Elastoplastic Model

The model proposed by Thornton (1997) considers an interaction of two spheres with elastic-perfectly plastic properties. The interaction includes elastic and elastoplastic loading combined with elastic unloading. The force–overlap relationship for loading and unloading is plotted in Fig. 24.

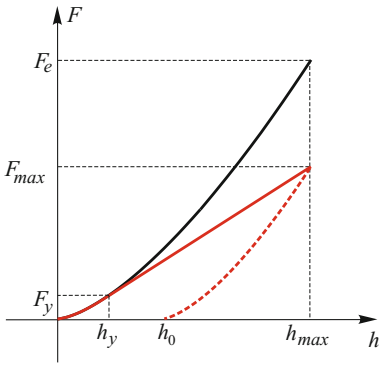

Fig. 24 Force versus particle overlap in the elasto-plastic Thornton model

Elastic loading

The contact force F at the initial stage of loading induces elastic deformation at the contact and it is given by the Hertz law, cf. Eqs. (64) and (65):

$$F = \frac{4}{3} E^* \sqrt{R^*} h^{\frac{3}{2}} , \tag{89}$$

where h is the amount the particles overlap, E^* is the effective modulus of elasticity defined by Eq. (66), and R^* is the effective radius defined by Eq. (56).

The Hertzian contact pressure distribution is given by the following relationship:

$$p = p_0 \left[1 - \left(\frac{r}{a} \right)^2 \right]^{\frac{1}{2}} , \tag{90}$$

where a is the radius of the contact area, r is the distance from the axis of symmetry and

$$p_0 = \frac{3F}{2\pi a^2} . \tag{91}$$

The contact area radius a is related to the particle overlap h as follows:

$$a^2 = R^* h \tag{92}$$

It is assumed that the loading is purely elastic below a certain value F_y (Fig. 24) corresponding to the initial yielding at the contact – when the maximum contact pressure p_0 under an increasing compressive load reaches the yield limit σ_y, called the *c*ontact yield stress (Fig. 25):

$$p_0(a = a_y) = \sigma_y . \tag{93}$$

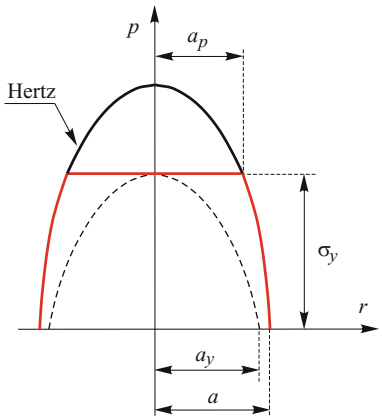

Fig. 25 Contact pressure distribution in the elasto-plastic Thornton model

Combining Eqs. (89), (91), (92) and (93) we can easily obtain the following formulae for the limit elastic contact force F_y and corresponding overlap h_y:

$$F_y = \frac{\pi^3 R^{*2} \sigma_y^3}{6E^{*2}}, \tag{94}$$

$$h_y = \frac{\pi^2 R^* \sigma_y^2}{4E^{*2}}. \tag{95}$$

For the spheres of the same size and with the same properties, Eqs. (94) and (95) have the following form:

$$F_y = \frac{\pi^3 R^2 (1-\nu^2) \sigma_y^3}{6E^2}, \tag{96}$$

$$h_y = \frac{\pi^2 R (1-\nu^2) \sigma_y^2}{2E^2}. \tag{97}$$

It should be remarked that the contact yield stress σ_y should not be identified with the uniaxial yield stress σ_Y. It has been shown by Vu-Quoc et al. (2000) that

$$\sigma_y = A_Y \sigma_Y, \tag{98}$$

where A_Y is a certain parameter dependent on the material properties and yield criterion. For the Huber–Mises criterion and Poisson's ratio $\nu = 0.3$, we obtain $A_Y = 1.61$, for $\nu = 0.4$ we obtain $A_Y = 1.74$.

Elastoplastic loading

After the yielding, the contact pressure distribution with the cut-off corresponding to the contact yield stress σ_y (see Fig. 25) is assumed. Given the contact pressure

distribution shown in Fig. 25 the contact force can be obtained from the following formula:

$$F = F_e - 2\pi \int_0^{a_p} (p(r) - \sigma_y)\, \mathrm{d}r\,, \tag{99}$$

where F_e is the elastic Hertz force corresponding to the contact area a, and the integral term is defined for the area with uniform contact pressure with radius a_p. After integrating Eq. (99) with pressure distribution according to Eq. (90) and performing further transformation the linear force–displacement relationship for the plastic loading is obtained in the following form, cf. Thornton (1997):

$$F = F_y + \pi\sigma_y R^*(h - h_y)\,. \tag{100}$$

Elastic unloading

It is assumed that the unloading is performed according to the Hertzian law, however, due to plastic deformation the contact curvature is smaller, and the unloading is performed assuming a certain curvature defined by the radius R_p^* $1/R_p^*$ $(1/R_p^* < 1/R^*)$. The radius R_p^* is determined from the assumption that with the contact area developed by the actual force F_{max} and the curvature $1/R_p^*$ is the same as it would be obtained with the curvature $1/R^*$ and the equivalent elastic force F_e, which is given by, cf. Fig. 24:

$$F_e = \frac{4}{3} E^* \sqrt{R^*} h_{max}^{\frac{3}{2}}\,, \tag{101}$$

The idea of equivalence of the contact area has been explained in Fig. 26. It can be expressed by the following equivalence:

$$F_{max} R_p^* = F_e R^*\,, \tag{102}$$

Then, the force during unloading is given as follows:

$$F = \frac{4}{3} E^* \sqrt{R_p^*} (h - h_0)^{\frac{3}{2}}\,, \tag{103}$$

The residual overlap h_0 can be determined taking Eq. (103) for h_{max}

$$h_0 = h_{max} - \left(\frac{3F_{max}}{4E^* \sqrt{R_p^*}} \right)^{\frac{2}{3}}\,. \tag{104}$$

Numerical examples

Collision of two balls with given initial velocities

Performance of the Thornton model will be demonstrated in simulations of collisions of two equal balls of radius $R = 10\,\mathrm{mm}$. The following properties have been assumed:

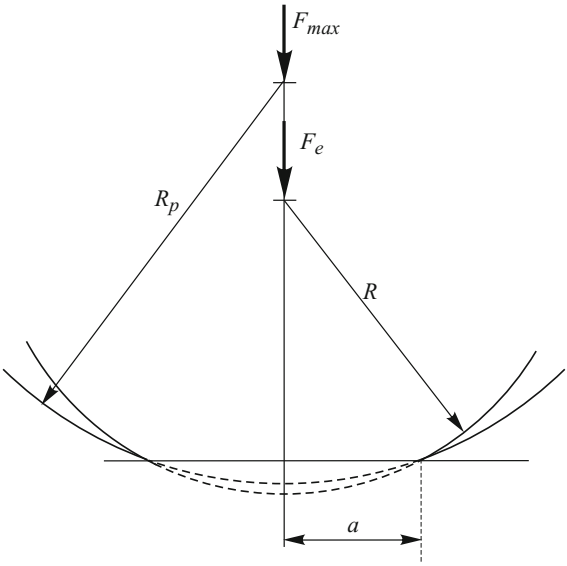

Fig. 26 Definition of the curvature for the elastic unloading in the elasto-plastic Thornton model

mass density $\rho = 8000\,\text{kg/m}^3$, Young's modulus $E = 200\,\text{GPa}$, Poisson's ratio $nu = 0.3$, and the contact yield stress $\sigma_y = 200\,\text{MPa}$. The problem has been analysed for different values of velocities in two ranges – lower velocities: 0.005, 0.010 and 0.015 m/s, and higher velocities: 0.5, 0.1 and 1.5 m/s.

The results for the lower velocities are presented in Figs. 27 and 28, and the results for the higher velocities in Figs. 29 and 30. The evolution of the gap between the balls and the velocity of one of the balls for lower velocities have been plotted in Fig. 27a and b. It can be seen that the lower the impact velocity is, the longer the impact is. The coefficients of restitution for the analysed cases evaluated using the general formula (63) are given in Table 2. It can be observed that the values of the coefficients of restitution decrease with an increase of impact velocity, which agrees with experimental observations.

The contact forces for lower impact velocities are plotted as functions of the time and overlap in Fig. 28a and b, respectively. It can be observed in Fig. 28a that the highest the impact velocity is, the shorter the collision time is. Figure 28b shows that the loading curve for different velocities is the same for different velocities. It can be observed in Fig. 28b that the incipient yielding occurs at a very early stage of loading therefore a very small part of the loading is purely elastic. Most of the loading is characterized by a linear relationship given by Eq. (99). The nonlinear force–displacement relationship is observed for the unloading for lower velocities. The results for higher velocities plotted in Fig. 30 show that practically the whole range of loading is characterized by the linear force–displacement relationship, and the unloading is very close to a linear behaviour. This shows that a linear elastoplastic model such as the Walton–Braun model can be sufficiently accurate for higher impact velocities or higher forces.

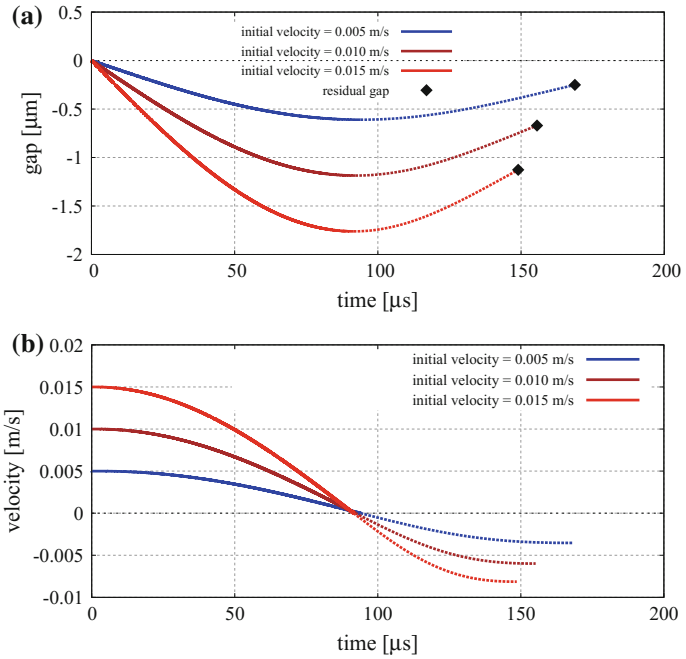

Fig. 27 Evolution of contact kinematic parameters during the collision of two balls in the Thornton elastoplastic model in a range of low impact velocities: **a** evolution of the gap, **b** evolution of the ball velocity (solid parts of the curves represent the loading and broken parts of the curve – the unloading)

4.6 Cohesive Elastic–perfectly Brittle Model

Modelling of cohesive materials such as rocks or concrete requires a model which takes into account a tensile interaction between discrete elements. This model assumes cohesive bonding between neighbouring particles. These bonds can be broken under load allowing us to simulate initiation and propagation of material fracture. After decohesion, standard cohesionless contact conditions are assumed.

Contact laws for the normal and tangential direction for the elastic perfectly brittle model are shown in Fig. 31. When two particles are bonded the contact forces in both normal and tangential directions are calculated from the linear constitutive relationships:

$$F_{\mathrm{n}}^{\mathrm{cont}} = k_{\mathrm{n}} g \,, \tag{105}$$

$$\boldsymbol{F}_{\mathrm{t}}^{\mathrm{cont}} = k_{\mathrm{t}}\, \boldsymbol{u}_{\mathrm{rt}} \,, \tag{106}$$

where: $F_{\mathrm{n}}^{\mathrm{cont}}$ – normal contact force, $\boldsymbol{F}_{\mathrm{t}}^{\mathrm{cont}}$ – tangential contact force, k_{n} – interface stiffness in the normal direction, k_{t} – interface stiffness in the tangential direction,

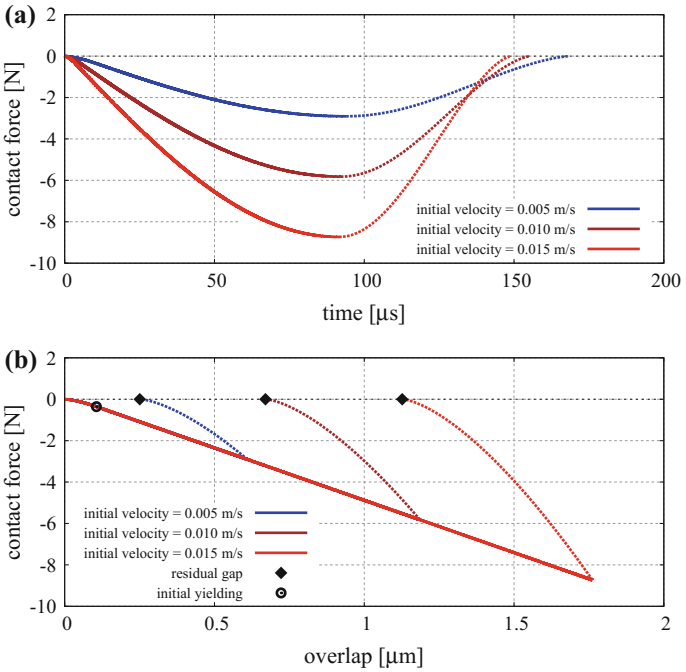

Fig. 28 Evolution of contact forces during the collision of two balls in the Thornton elastoplastic model for a range of low impact velocities: **a** force versus time, **b** force versus overlap (solid parts of the curves represent the loading and broken parts of the curves – the unloading)

g – gap/overlap, $\boldsymbol{u}_{\mathrm{rt}}$ – tangential relative displacement. It should be remarked that unlike Eq. (38), formula (105) is used for both negative and positive values of the gap g.

Cohesive bonds are broken instantaneously when the interface strength is exceeded in the tangential direction by the tangential contact force or in the normal direction by the tensile contact force. The failure (decohesion) criterion can be written as:

$$F_{\mathrm{n}}^{\mathrm{cont}} \leq R_{\mathrm{n}}\,, \tag{107}$$

$$\|\boldsymbol{F}_{\mathrm{t}}^{\mathrm{cont}}\| \leq R_{\mathrm{t}}\,, \tag{108}$$

where: R_{n} — interface strength in the normal direction, R_{t} — interface strength in the tangential direction.

In the absence of cohesion the normal contact force can be compressive only ($R_n \leq 0$) and tangential contact force can be nonzero due to friction if $R_{\mathrm{n}} < 0$ or zero otherwise. The friction force is evaluated according to the regularized Coulomb friction model.

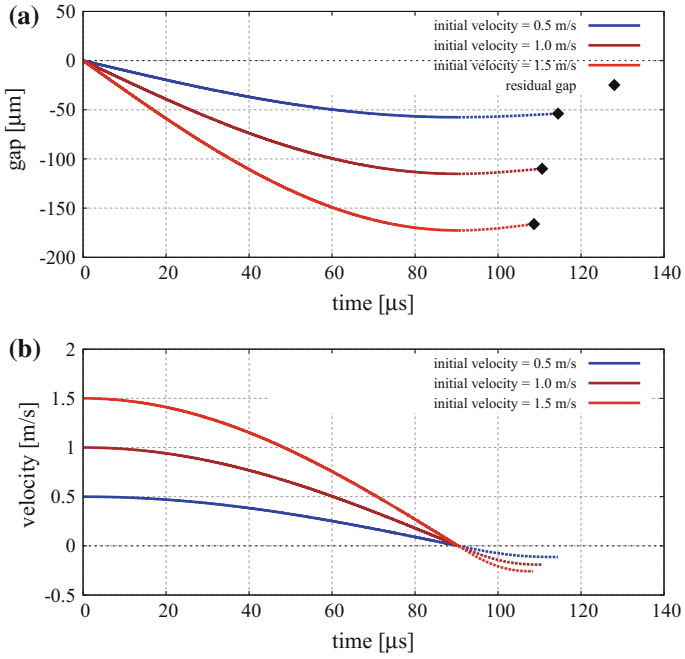

Fig. 29 Evolution of contact kinematic parameters during the collision of two balls in the Thornton elastoplastic model in a range of higher impact velocities: **a** evolution of the gap, **b** evolution of the ball velocity (solid parts of the curves represent the loading and broken parts of the curve – the unloading)

5 Moment Type Interaction

Contact model in the discrete element method except for forces can include a moment type interaction (Wang et al. 2015). In a general case, there can be a moment interaction between bonded and unbonded particles. Here, the moment type interaction between unbonded particles will be presented.

A moment type interaction allows to compensate deficiencies of the discrete element model due to an idealized shape of spherical particles. It provides resistance to a relative rotation of contacting particles (discrete elements). The relative motion of two particles i and j can be described by a relative angular velocity $\boldsymbol{\omega}_{\mathrm{r}}$ given by

$$\boldsymbol{\omega}_{\mathrm{r}} = \boldsymbol{\omega}_i - \boldsymbol{\omega}_j \,. \tag{109}$$

The relative angular velocity $\boldsymbol{\omega}_{\mathrm{r}}$ can be decomposed into the components normal and tangent to the contact plane, $\boldsymbol{\omega}_{\mathrm{rn}}$ and $\boldsymbol{\omega}_{\mathrm{rt}}$, respectively:

$$\boldsymbol{\omega}_{\mathrm{r}} = \boldsymbol{\omega}_{\mathrm{rn}} + \boldsymbol{\omega}_{\mathrm{rs}} = \omega_{\mathrm{rn}} \cdot \boldsymbol{n} + \boldsymbol{\omega}_{\mathrm{rs}} \,. \tag{110}$$

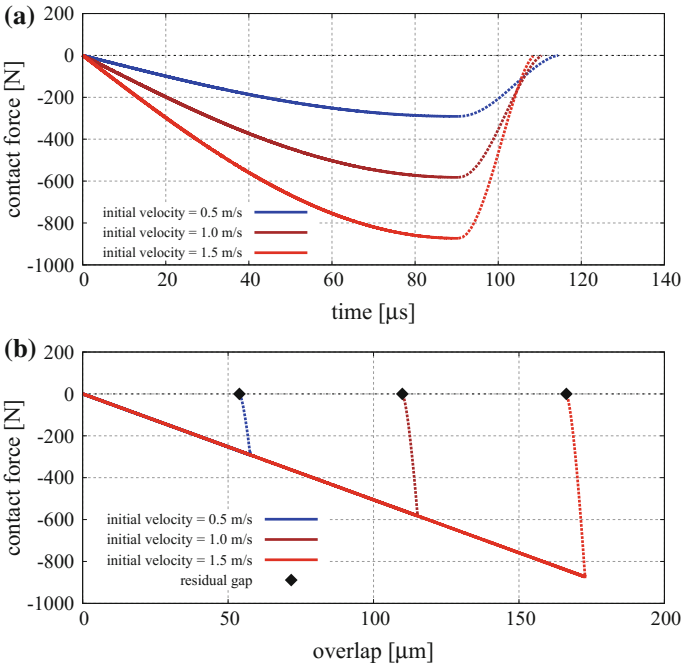

Fig. 30 Evolution of contact forces during the collision of two balls in the Thornton elastoplastic model for a range of higher impact velocities: **a** force versus time, **b** force versus overlap (solid parts of the curves represent the loading and broken parts of the curves – the unloading)

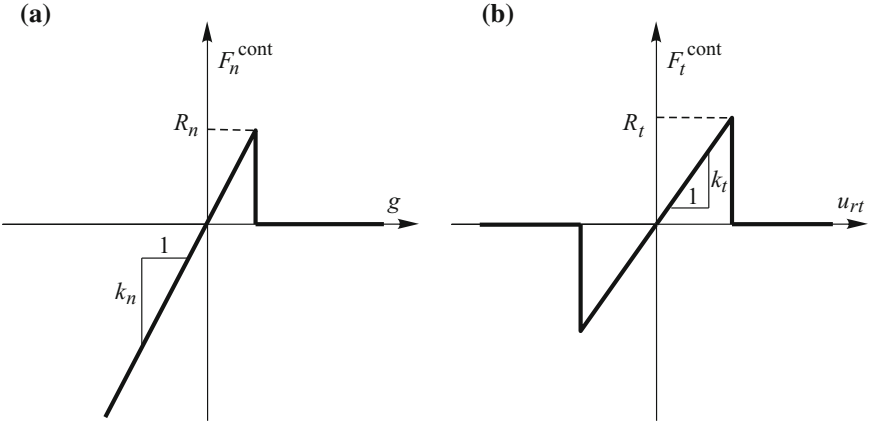

Fig. 31 Force–displacement relationships for the elastic perfectly brittle model: **a** in the normal direction, **b** in the tangential direction

Table 2 Coefficients of restitution for the Thornton model

v^0 (m/s)	v^0_{rel} (m/s)	v_{rel} (m/s)	e, Eq. (63)
0.005	0.010	0.00706	0.7060
0.010	0.020	0.01197	0.5987
0.015	0.030	0.01627	0.5424
0.5	1.0	0.2269	0.2269
1.0	2.0	0.3815	0.1908
1.5	3.0	0.51717	0.1724

The normal component can be obtained by projection of the velocity vector $\boldsymbol{\omega}_\mathrm{r}$ onto the unit normal vector $\boldsymbol{n}$:

$$\omega_\mathrm{rn} = \boldsymbol{\omega}_\mathrm{r} \cdot \boldsymbol{n} \,. \tag{111}$$

Then, the tangent component is obtained from Eq. (110) as follows

$$\boldsymbol{\omega}_\mathrm{rs} = \boldsymbol{\omega}_\mathrm{r} - \omega_{rn}\boldsymbol{n} \,. \tag{112}$$

The motion defined by the tangent component $\boldsymbol{\omega}_\mathrm{rn}$ is called rolling, and that defined by the normal component $\boldsymbol{\omega}_\mathrm{rt}$ is referred to as twisting.

Analogously, the contact interaction moment $\boldsymbol{T}^c$ between the particles can be decomposed into two components – normal and tangential to the contact plane, $\boldsymbol{T}_\mathrm{n}$ i $\boldsymbol{T}_\mathrm{s}$, respectively:

$$\boldsymbol{T}^c = \boldsymbol{T}_\mathrm{n} + \boldsymbol{T}_\mathrm{s} = T_\mathrm{n}\,\boldsymbol{n} + \boldsymbol{T}_\mathrm{s} \,. \tag{113}$$

Models of twisting and rolling resistance can be defined analogously to sliding friction models. A model of twisting resistance can be defined in terms of the angular velocity component $\boldsymbol{\omega}_\mathrm{rn}$ and the component $\boldsymbol{T}_\mathrm{n}$ of the contact moment.

For the model of twisting resistance the Kuhn–Tucker conditions can be written analogously to the conditions (22) as follows:

$$\phi_\mathrm{n} \le 0 \,, \quad \lambda_\mathrm{n} \ge 0 \,, \quad \phi_\mathrm{n}\lambda_\mathrm{n} = 0 \,, \tag{114}$$

where λ_n is defined by the non-associated rolling law:

$$\boldsymbol{\omega}_\mathrm{rn} = \lambda_\mathrm{n} \frac{\boldsymbol{T}_\mathrm{n}}{||\boldsymbol{T}_\mathrm{n}||} \,, \tag{115}$$

and ϕ_n is given by the following equation:

$$\phi_\mathrm{n} = \|\boldsymbol{T}_\mathrm{n}\| - a_\mathrm{n}\mu|F_\mathrm{n}| \,, \tag{116}$$

a_n is a parameter which has dimension of length determining the limit moment of twisting resistance. It has been assumed that this moment is proportional to the normal contact force F_n and Coulomb friction coefficient μ.

Analogously, a model of rolling resistance can be defined. The Kuhn–Tucker conditions for the model of rolling resistance can be written as follows:

$$\phi_\mathrm{s} \leq 0\,, \quad \lambda_\mathrm{s} \geq 0\,, \quad \phi_\mathrm{s}\lambda_\mathrm{s} = 0\,, \tag{117}$$

where λ_s is defined by the non-associated law of rolling:

$$\boldsymbol{\omega}_\mathrm{rs} = \lambda_\mathrm{s} \frac{\boldsymbol{T}_\mathrm{s}}{||\boldsymbol{T}_\mathrm{s}||}\,, \tag{118}$$

and ϕ_s is given by

$$\phi_\mathrm{s} = \|\boldsymbol{T}_\mathrm{s}\| - a_\mathrm{s}|F_\mathrm{n}|\,, \tag{119}$$

where the limit moment of rolling resistance. depends on the normal contact force F_n and the parameter a_s which has dimension of length. This definition is consistent with the concept of rolling friction employed in engineering, where the parameter a_s is called the coefficient of rolling friction.

The conditions (114)–(116) and (117)–(119) can be regularized introducing the penalty coefficients $k_\mathrm{n}^\mathrm{rot}$ and $k_\mathrm{s}^\mathrm{rot}$ into the twisting and rolling laws (115) and (118)

$$\dot{\boldsymbol{T}}_n = k_\mathrm{n}^\mathrm{rot}\left(\boldsymbol{\omega}_\mathrm{rn} - \lambda_\mathrm{n}\frac{\boldsymbol{T}_\mathrm{n}}{||\boldsymbol{T}_\mathrm{n}||}\right)\,, \tag{120}$$

$$\dot{\boldsymbol{T}}_\mathrm{s} = k_\mathrm{s}^\mathrm{rot}\left(\boldsymbol{\omega}_\mathrm{rs} - \lambda_\mathrm{s}\frac{\boldsymbol{T}_\mathrm{s}}{||\boldsymbol{T}_\mathrm{s}||}\right)\,. \tag{121}$$

After regularization the models of twisting and rolling resistance are similar to elasto-plastic models with non-associated plastic flow rules, and the penalty coefficients play roles of the moduli of elasticity.

6 Discrete Element Method for Thermal and Thermomechanical Problems

In many problems, the contact is associated with thermal effects such as heat generation through friction or heat transfer at the contact between particles with different temperatures. The discrete element method can be extended to model thermal and thermomechanical problems.

6.1 Formulation of the Discrete Element Method for Heat Conduction Problem

Thermal formulation of the discrete element method introduced here is based on the assumption that the temperature difference inside particles is negligible and the temperature can be considered uniform within particles. Following this assumption heat conduction inside particles is neglected whereas heat transfer to and from particles through their boundary is considered. Such a simplification of the heat conduction problem is typical for the lumped capacitance model, also called the lumped system analysis (Cengel 2007). This assumption is justified for the discrete element model employing relatively small particles and it is consistent with the formulation of the mechanical problem. Similarly as the mechanical problem is governed by the contact interaction, the heat conduction problem is governed by the conductive heat transfer at the particle contacts.

The schematic of the heat transfer for a single particle is shown in Fig. 32. The thermal model is expressed mathematically by the heat balance equation, which can be written for a single particle in the following form:

$$m_i c_i \dot{\theta}_i = Q_i \,, \tag{122}$$

where: m_i – particle mass, c_i – specific heat, θ_i – particle temperature, Q_i – heat sources or heat fluxes per single particle. Q_i includes externally supplied heat source Q_i^{ext}, heat generated internally within the particle Q_i^{gen}, heat conducted through the contact interface Q_{ij}^{cont}, and convective and radiative heat transfer between particles and environment on the free surface, Q_i^{conv} and Q_i^{rad}

$$Q_i = Q_i^{\text{ext}} + Q_i^{\text{gen}} + \sum_{j=1}^{n_c} Q_{ij}^{\text{cont}} + Q_i^{\text{conv}} + Q_i^{\text{rad}} \tag{123}$$

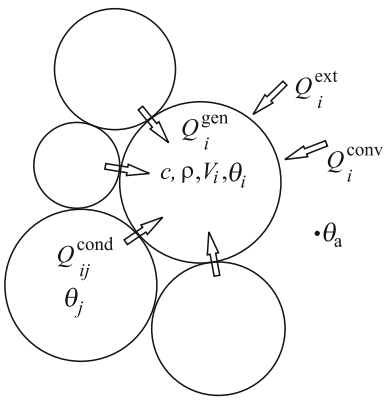

Fig. 32 Schematic of the heat conduction problem for a discrete element (reproduced from Rojek (2014))

where n_c is the number of particles being in contact with the i-th particle.

6.2 Model of Thermal Contact

When two solid bodies of different temperature come into contact heat flows from the body with higher temperature to the body with lower temperature until a thermal equilibrium is achieved. If we take two particles with masses m_i and m_j, heat capacities c_i and c_j, and temperatures θ_i and θ_j, and bring them into contact, if there no heat exchange with exterior, the thermal equilibrium is achieved at the temperature θ given by

$$\theta = \frac{m_i c_i \theta_i + m_j c_j \theta_j}{m_i c_i + m_j c_j} \tag{124}$$

The time necessary to reach equilibrium depends on the heat flux through the contact interface Q_{ij}^{cont} (in J/s). It is commonly assumed that the the heat flux through the contact interface is proportional to the temperature jump $(\theta_i - \theta_j)$, contact area A^{cont}, and a certain coefficient h^{cont} called thermal contact conductance (Cooper et al. 1969):

$$Q_{ij}^{\text{cont}} = -h^{\text{cont}} A^{\text{cont}} (\theta_i - \theta_j) \tag{125}$$

The contact area A^{cont} can be related to the local particle size

$$A^{\text{cont}} = \beta R^{*2} \tag{126}$$

where R^* is the equivalent radius defined by Eq. (56) and β is a certain dimensionless parameter which should be calibrated for a given discrete element model.

It is sometimes convenient to express heat transfer at the contact in terms of one parameter $H^{\text{cont}} = h^{\text{cont}} A^{\text{cont}}$, then Eq. (125) can be rewritten as Zhang et al. (2011):

$$Q_{ij}^{\text{cont}} = -H^{\text{cont}} (\theta_i - \theta_j) \tag{127}$$

In general case of a contact of different bodies or particles, the thermal contact conductance is dependent on the surface roughness, material properties, interface temperature and interface pressure (Cooper et al. 1969. In the discrete element model, however, the thermal contact conductance h^{cont} does not represent the thermal resistance of the interface, only, but it should also take into account the influence of the thermal conductivity λ ot the particle material. It should be treated as a micromechanical parameter which should give required macroscopic properties of the bulk material.

6.3 Time Integration of the Discrete Element Method for Thermal Problem

Heat conduction Eq. (133) can be integrated in time using the explicit forward Euler scheme

$$\theta_i^{n+1} = \theta_i^n + \frac{Q_i^n \Delta t}{m_i c} . \tag{128}$$

The explicit time integration scheme expressed by Eq. (128) is conditionally stable. The time integration step is limited by the critical step $\Delta t_{\text{cr}}^{\text{therm}}$ which can be estimated by the critical value for the one-dimensional heat conduction problem (Hughes 1987)

$$\Delta t_{\text{cr}}^{\text{therm}} \approx \frac{l_{\min}}{2a} , \tag{129}$$

where $l_{\min}$ is the minimum particle centre distance and s is the thermal diffusivity

$$a = \frac{\lambda}{\rho c} . \tag{130}$$

6.4 Formulation of the Discrete Element Method for a Coupled Thermo-Mechanical Problem

The mechanical and thermal phenomena can be analysed jointly as a coupled thermo-mechanical problem using the discrete element model (Rojek 2014). Thermo-mechanical problem defined in the framework of the discrete element method by the system of coupled equations formed by the equations of motion (1) and (2) and the heat balance equation (122)

$$m_i \ddot{\boldsymbol{u}}_i = \boldsymbol{F}_i , \tag{131}$$

$$J_i \dot{\boldsymbol{\omega}}_i = \boldsymbol{T}_i , \tag{132}$$

$$m_i c \dot{\theta}_i = Q_i \tag{133}$$

with appropriate initial conditions. Coupling of Eqs. (131) and (132) with Eq. (133) can be obtained considering such effects as:

- frictional heat generation,
- thermal expansion of the particles and its effect on particle interaction (thermal stresses),
- temperature dependence of mechanical contact parameters (due to temperature dependence of material macroscopic properties),
- modification of the geometrical configuration of thermal problem determined by the solution of mechanical problem.

6.5 Thermomechanical Contact

Friction considered in the mechanical problem is associated with heat generation. The heat is absorbed by the particles increasing their temperature and can be conducted to other particles by heat transfer at the contact. These coupled phenomena can be taken into account in the contact model considered in the formulation of the thermomechanical model. The schematic of the thermomechanical contact model is shown schematically in Fig. 33.

Heat generation through frictional dissipation is calculated using the following formula

$$Q^{\text{gen}} = \chi \, |F_{\text{t}} \, v_{\text{rt}}^{\text{ir}}| \, , \tag{134}$$

where F_{t} is the friction force, $v_{\text{rt}}^{\text{ir}}$ is the irreversible part of the relative tangential velocity, and $0 \leq \chi \leq 1$ is the part of the friction work converted to heat. Heat generated at the contact point is absorbed by the contacting particles

$$Q^{\text{gen}} = Q_{ij}^{\text{gen}} + Q_{ji}^{\text{gen}} \tag{135}$$

If the particles are of the same materials, it is assumed that the heat is absorbed equally by the particles:

$$Q_{ij}^{\text{gen}} = Q_{ji}^{\text{gen}} = 0.5 Q^{\text{gen}} \tag{136}$$

In a general case, when the particle thermal properties can be different, heat absorbed by each of the contacting particles is assumed to be proportional to the thermal effusivity of the particle materials

$$\frac{Q_{ij}^{\text{gen}}}{Q_{ji}^{\text{gen}}} = \frac{\alpha Q^{\text{gen}}}{(1-\alpha) Q^{\text{gen}}} = \frac{e_i}{e_j} \tag{137}$$

with the effusivity e being defined as

$$e = \sqrt{\lambda \rho c} \tag{138}$$

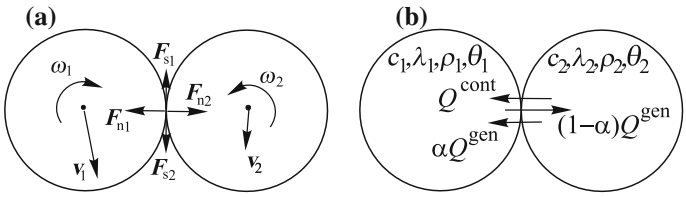

Fig. 33 Thermomechanical contact schematic for a pair of particles: **a** mechanical contact interaction, **b** thermal contact effects (reproduced from Rojek (2014))

where λ is the thermal conductivity, c is the specific heat capacity, and ρ is the density. The heat partition coefficient α is given by the following equation

$$\alpha = \frac{e_i}{e_i + e_j} \tag{139}$$

Solution of Thermomechanical Coupled Problem in the Discrete Element Method

The system of coupled equations is solved using the staggered solution scheme, in which the mechanical and thermal problems are analysed separately.

1. Solution of the mechanical problem.
   Equations (131) and (132) are integrated in time using the explicit central difference scheme. Employing the equations for the known configuration at the time t_n the solution for the time t_{n+1} is obtained in the following way:

$$\ddot{\boldsymbol{u}}_i^n = \frac{\boldsymbol{F}_i^n}{m_i}\,, \tag{140}$$

$$\dot{\boldsymbol{u}}_i^{n+1/2} = \dot{\boldsymbol{u}}_i^{n-1/2} + \ddot{\boldsymbol{u}}_i^n \Delta t\,, \tag{141}$$

$$\boldsymbol{u}_i^{n+1} = \boldsymbol{u}_i^n + \dot{\boldsymbol{u}}_i^{n+1/2} \Delta t\,, \tag{142}$$

$$\dot{\boldsymbol{\omega}}_i^n = \frac{\boldsymbol{T}_i^n}{J_i}\,, \tag{143}$$

$$\boldsymbol{\omega}_i^{n+1/2} = \boldsymbol{\omega}_i^{n-1/2} + \dot{\boldsymbol{\omega}}_i^n \Delta t\,, \tag{144}$$

$$\Delta\psi_i = \boldsymbol{\omega}_i^{n+1/2} \Delta t\,. \tag{145}$$

   Thermal expansion of the particles and resulting thermal components of the interaction forces are considered in the solution of the mechanical problem. In many problems, thermally induced stresses may material damage or failure (Wanne 2009; Leclerc et al. 2018). Heat generated by friction is evaluated in Eqs. (131) and (132) and passed to Eq. (133).
2. Solution of the thermal problem
   Heat conduction Eq. (133) is integrated in time using the explicit forward Euler scheme

$$\theta_i^{n+1} = \theta_i^n + \frac{Q_i^n \Delta t}{m_i c}\,. \tag{146}$$

   The thermal problem is solved on the modified particle configuration determined in the solution of the mechanical problem. Particle temperatures evaluated in the solution of the thermal problem are passed to the solution of the mechanical problem.

The explicit time integration scheme of the coupled thermomechanical problem is conditionally stable. The time integration step is limited by the critical step Δt_{cr}:

$$\Delta t_{\mathrm{cr}} = \min(\Delta t_{\mathrm{cr}}^{\mathrm{mech}}, \Delta t_{\mathrm{cr}}^{\mathrm{therm}}), \tag{147}$$

where $\Delta t_{\mathrm{cr}}^{\mathrm{mech}}$ is the critical time step for the solution of the mechanical problem depending on the highest eigenfrequency of the discrete system $\Omega_{\max}$

$$\Delta t_{\mathrm{cr}}^{\mathrm{mech}} = \frac{2}{\Omega_{\max}}, \tag{148}$$

and $\Delta t_{\mathrm{cr}}^{\mathrm{therm}}$ is the critical time step for the solution of the thermal problem. The critical time step for the solution of the thermal problem can be estimated according to Eq. (129).

Concluding remarks

A brief overview of the basic concepts of the discrete element method and most popular contact models used in this method has been made in this chapter. Contact models used in the DEM intend to reproduce complex phenomena associated with contact between particles (discrete elements) using models composed of relatively simple rheological elements reproducing elementary contact mechanisms.

Many applications of the DEM show that even with simple contact models, it is possible to reproduce the complex macroscopic behaviour of the bulk material modelled by a collection of discrete elements. The contact model in the DEM plays a role of a constitutive model at the micro- or mesoscopic level. The DEM can be used to model cohesionless granular materials as well as various cohesive materials. The DEM in a simple way takes into account discontinuities existing in the material or occurring under loading. The DEM is a suitable tool to model failure of materials and structures characterized by multiple fracturing.

It must be remarked that the DEM is not a simple method for a user. The choice of a suitable contact model and evaluation of appropriate model parameters requires a certain experience and knowledge. The author dares hope this work will be useful to understand the physical background and mathematical representation of the contact phenomena in the DEM.

Acknowledgements The author would like to thank Mr. Nikhil Madan for performing simulations and preparing the plots for the numerical examples included in this chapter.

References

J. Argyris. An excursion into large rotations. *Comput. Meth. Appl. Mech. Eng.*, 32: 85–155, 1982.

J.P. Bardet and J. Proubet. A numerical investigation of the structure of persistent shear bands in granular media. *Geotechnique*, 41: 599–613, 1991.

T. Belytschko, P. Smolinski, and W.K. Liu. Stability of multi-time step partitioned integrators for the first order finite element systems. *Comput. Meth. Appl. Mech. Eng.*, 49: 281–297, 1985.

Y.A. Cengel. *Heat and Mass Transfer: A Practical Approach*. McGraw-Hill, Third Edition edition, 2007.

M.G. Cooper, B.B. Mikic, and M.M. Yovanovich. Thermal contact conductance. *Int. J. Heat Mass Transfer*, 12: 279–300, 1969.

P.A. Cundall. A Computer Model for Simulating Progressive Large Scale Movements in Blocky Rock Systems. In *Proc. Int. Symp. Rock Fracture, ISRM*, pages 2–8, Nancy, France, 1971.

P.A. Cundall. Distinct element models of rock and soil structure. In *Analytical and computational models in engineering and rock mechanics*. Allen&Unwin, London, 1987.

P.A. Cundall. Formulation of a Three Dimensional Distinct Element Model — Part I. A Scheme to Detect and Represent Contacts in a System of Many Polyhedral Blocks. *Int. J. Rock Mech., Min. Sci. & Geomech. Abstr.*, 25 (3): 107–116, 1988.

P.A. Cundall and O.D.L. Strack. A discrete numerical method for granular assemblies. *Geotechnique*, 29: 47–65, 1979.

A. Curnier. Unilateral contact. mechanical modelling. In P. Wriggers and P. Panagiotopoulos, editors, *New Developments in Contact Problems*, pages 1–54. Springer, 1999.

F. Fleissner, T. Gaugele, and P. Eberhard. Applications of the discrete element method in mechanical engineering. *Multibody Syst. Dyn.*, 18: 81–94, 2007.

P.K. Haff and B.T. Werner. Collisional interaction of a small number of confined inelastic grains. In T. Ariman and T. N. Veziroglu, editors, *Colloidal and Interfacial Phenomena*, pages 483–501. Hemisphere Publishing, 1987.

S. Hentz, L. Daudeville, and F.V. Donzé. Identification and validation of a discrete element model for concrete. *ASCE J. Eng. Mech.*, 130: 709–719, 2004.

H. Hertz. Über die Berührung fester elastischer Körper (On the contact of elastic bodies). *J. Reine Angewandte Math.*, 94: 156–171, 1882.

D.C. Hong and J.A. McLennan. Molecular dynamics simulations of hard sphere granular particles. *Phys. A: Stat. Mech. Applicat.*, 187: 159–171, 1992.

T.J.R. Hughes. *The Finite Element Method. Linear Static and Dynamic Analysis*. Prentice-Hall, 1987.

K.H. Hunt and F.R.E. Crossley. Coefficient of restitution interpreted as damping in vibroimpact. *Journal of Applied Mechanics*, 42: 440, 1975.

R. Jankowski. Analytical expression between the impact damping ratio and the coefficient of restitution in the non-linear viscoelastic model of structural pounding. *Earthquake Engng Struct. Dyn.*, 35: 517–524, 2006.

K.J. Johnson. *Contact Mechanics*. Cambridge University Press, 1985.

G.R. Joldes, A. Wittek, and K. Miller. An adaptive Dynamic Relaxation method for solving nonlinear finite element problems. Application to brain shift estimation. *Int. J. Numer. Method Biomed. Eng.*, 27: 173–185, 2011.

A. Klarbring. Contact, friction, discrete mechanical structures and mathematical programming. In P. Wriggers and P. Panagiotopoulos, editors, *New Developments in Contact Problems*, pages 56–100. Springer, 1999.

H. Kruggel-Emden, S. Wirtz, and V. Scherer. A study on tangential force laws applicable to the discrete element method (DEM) for materials with viscoelastic or plastic behavior. *Chemical Engineering Science*, 63: 1523–1541, 2008.

P.-L. Larsson, S. Biwa, and B. Storåkers. Analysis of cold and hot isostatic compaction of spherical particles. *Acta mater.*, 44: 3655–3666, 1996.

W. Leclerc, H. Haddad, and M. Guessasma. On a discrete element method to simulate thermal-induced damage in 2d composite materials. *Computers and Structures*, 196: 277–291, 2018.

J. Lee. Density waves in the flows of granular media. *Phys. Rev. E*, 49 (1): 281, 1994.

Y.-Y. Lin and C.Y. Hui. Mechanics of contact and adhesion between viscoelastic spheres: An analysis of hysteresis during loading and unloading. *Journal of Polymer Science: Part B: Polymer Physics*, 40: 772–793, 2002.

F.P. Di Maio and A. Di Renzo. Modelling particle contacts in distinct element simulations. Linear and non-linear approach. *Chemical Engineering Research and Design*, 83(A11): 1287–1297, 2005.

I. Marczewska, J. Rojek, and R. Kačianauskas. Investigation of the effective elastic parameters in the discrete element model of granular material by the triaxial compression test. *Archives of Civil and Mechanical Engineering*, 16: 64–75, 2016.

C.L. Martin, D. Bouvard, and S. Shima. Study of particle rearrangement during powder compaction by the Discrete Element Method. *J. Mech. Phys. Solids*, 51: 667–693, 2003.

R.D. Mindlin and H. Deresiewicz. Elastic spheres in contact under varying oblique forces. *ASME Journal of Applied Mechanics*, 20: 327–344, 1953.

J.J. Moreau. Some numerical methods in multibody dynamics: application to granularmaterials. *European Journal of Mechanics A/Solids*, 13: 93–114, 1994.

M. Nagurka and S. Huang. A mass-spring-damper model of a bouncing ball. *Int. J. Engng Ed.*, 22: 393–401, 2006.

H.A. Navarro and M.P. de Souza Braun. Linear and nonlinear Hertizian contact models for materials in multibody dynamics. In *Proceedings of the 22nd Int. Congress of Mechanical Engineering (COBEM 2013) November 3–7, 2013, Ribeirao Preto, SP, Brazil*, pages 159–180, 2013.

E. Olsson and P.-L. Larsson. On the effect of particle size distribution in cold powder compaction. *Journal of Applied Mechanics*, 79: 1–8, 2012.

J.P. Plassiard, N. Belheine, and F.V. Donze. A spherical discrete element model: calibration procedure and incremental response. *Granular Matter*, 11: 293–306, 2009.

D.O. Potyondy and P.A. Cundall. A bonded-particle model for rock. *Int. J. Rock Mech. Min. Sci.*, 41: 1329–1364, 2004.

M. Raous. Quasistatic signorini problem with Coulomb friction and coupling to adhesion. In P. Wriggers and P. Panagiotopoulos, editors, *New Developments in Contact Problems*, pages 101–178. Springer, 1999.

A. Di Renzo and F.P. Di Maio. Comparison of contact–force models for the simulation of collisions in DEM-based granular flow codes. *Chemical Engineering Science*, 59: 525–541, 2004.

D.C. Richardson, K.J. Walsh, N. Murdoch, and P. Michel. Numerical simulations of granular dynamics: I. Hard-sphere discrete element method and tests. *Icarus*, 212: 427–437, 2011.

J. Rojek. Discrete element thermomechanical modelling of rock cutting with valuation of tool wear. *Computational Particle Mechanics*, 1: 71–84, 2014.

J. Rojek, E. Oñate, F. Zarate, and J. Miquel. Modelling of rock, soil and granular materials using spherical elements. In *2nd European Conference on Computational Mechanics ECCM-2001*, Cracow, 26-29 June, 2001.

J. Rojek, F. Zarate, C. Agelet de Saracibar, Ch. Gilbourne, and P. Verdot. Discrete element modelling and simulation of sand mould manufacture for the lost foam process. *Int. J. Num. Meth. Eng.*, 62: 1421–1441, 2005.

J. Rojek, E. Onate, C. Labra, and H. Kargl. Discrete element simulation of rock cutting. *International Journal of Rock Mechanics and Mining Sciences*, 48: 996–1010, 2011.

J. Rojek, C. Labra, O. Su, and E. Oñate. Comparative study of different discrete element models and evaluation of equivalent micromechanical parameters. *Int. J. Solids and Structures*, 49: 1497–1517, 2012.

J. Rojek, G.F. Karlis, L.J. Malinowski, and G. Beer. Setting up virgin stress conditions in discrete element models. *Computers and Geotechnics*, 48: 228–248, 2013.

L. Rothenburg and R. J. Bathurst. Micromechanical features of granular materials with planar elliptical particles. *Geotechnique*, 42 (1): 79–95, 1992.

R. Senapati and J. Zhang. Identifying fracture origin in ceramics by combination of nondestructive testing and discrete element analysis. In *AIP Conference Proceedings*, volume 1211, pages 1445–1451, 2010.

M. Shillor, M. Sofonea, and J.J. Telega. *Models and Analysis of Quasistatic Contact. Variational Methods*. Lect. Notes Phys. 655, Springer, 2004.

B. Storåkers, S. Biwa, and P.-L. Larsson. Similarity analysis of inelastic contact. *Int. J. Solids and Structures*, 34: 3061–3083, 1997.

B. Storåkers, N.A. Fleck, and R.M. McMeeking. The viscoplastic compaction of composite powders. *Journal of the Mechanics and Physics of Solids*, 47: 785–815, 1999.

H. Tao, W. Zhong, and B. Jin. Flow behavior of non-spherical particle flowing in hopper. *Frontiers in Energy*, 3: 315–321, 2014.

L.M. Taylor and D.S. Preece. Simulation of blasting induced rock motion. *Eng. Comput.*, 9 (2): 243–252, 1992.

C. Thornton. Coefficient of restitution for collinear collisions of elastic-perfectly plastic spheres. *J. Appl. Mech*, 64: 383–386, 1997.

Y. Tsuji, T. Tanaka, and T. Ishida. Lagrangian numerical simulation of plug flow of cohesionless particles in a horizontal pipe. *Powder Technology*, 71: 239–250, 1992.

L. Vu-Quoc, X. Zhang, and L. Lesburg. A normal force-displacement model for contacting spheres accounting for plastic deformation: Force-driven formulation. *Journal of Applied Mechanics*, 67: 363–371, 2000.

O.R. Walton. Explicit particle dynamics for granular materials. In *Proc. 4th Int. Conf. on Numerical Methods in Geomechanics*, pages 1261–1268, Eldmonton, Canada, 1982.

O.R. Walton. Particle dynamics calculations of shear flow. In J.T. Jenkins and M. Satake, editors, *Mechanics of Granular Materials: New Models and Constitutive Relations*, pages 327–338. Elsevier, 1983.

O.R. Walton and R.L. Braun. Stress calculations for assemblies of inelastic spheres in uniform shear. *Aeta Mechanica*, 63: 73–86, 1986.

Y. Wang, F. Alonso-Marroquin, and W.W. Guo. Rolling and sliding in 3-D discrete element models. *Particuology*, 23: 49–55, 2015.

T. Wanne. *Bonded-particle modeling of thermally induced damage in rock*. PhD thesis, University of Toronto, 2009.

L. Widuliński, J. Kozicki, and J. Tejchman. Numerical Simulations of Triaxial Test with Sand Using DEM. *Archives of Hydro-Engineering and Environmental Mechanics*, 56: 149–171, 2009.

J.R. Williams, G. Hocking, and G.G.W. Mustoe. The theoretical basis of the discrete element method. In *NUMETA 1985, Numerical Methods of Engineering, Theory and Applications*. A.A. Balkema, Rotterdam, 1985.

T. Wu, I. Temizer, and P. Wriggers. Computational thermal homogenization of concrete. *Cement and Concrete Composites*, 35: 59–70, 2013.

E. Zdancevičius, R. Kačianauskas, and D. Zabulionis. Improvement of viscoelastic damping for the Hertz contact of particles due to impact velocity. *Procedia Engineering*, 172: 1286–1290, 2017.

H.W. Zhang, Q. Zhou, H.L. Xing, and H. Muhlhaus. A DEM study on the effective thermal conductivity of granular assemblies. *Powder Technology*, 205: 172–183, 2011.

L.F.C. Zonetti, A.S.S. Camargo, J. Sartori, D.F. de Sousa, and L.A.O. Nunes. A demonstration of dry and viscous damping of an oscillating pendulum. *Eur. J. Phys.*, 20: 85–88, 1999.

A. Zubelewicz and Z. Mroz. Numerical simulation of rock burst processes treated as problems of dynamic instability. *Rock Mechanics and Rock Engineering*, 16: 253–274, 1983.