ISSN 2732–0189 (Online)
ISSN 1586–2070 (Print)

JOURNAL OF COMPUTATIONAL AND APPLIED MECHANICS

An Open Access International Journal

Published by the University of Miskolc

VOLUME 15, NUMBER 2 (2020)

MISKOLC UNIVERSITY PRESS

ISSN 2732–0189 (Online)
ISSN 1586–2070 (Print)

JOURNAL OF COMPUTATIONAL AND APPLIED MECHANICS

An Open Access International Journal

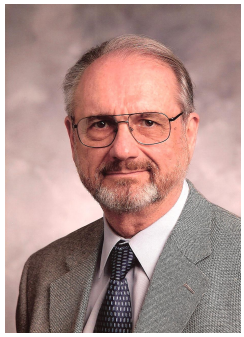
Published by the University of Miskolc

VOLUME 15 NUMBER 2 (2020)

MISKOLC UNIVERSITY PRESS

Journal of Computational and Applied Mechanics, Vol. 15, No. 2, (2020), pp. 79–80
DOI: 10.32973/jcam.2020.005

PREFACE



The papers in the present and next issues of JCAM are dedicated to an excellent scientist: Professor Barna Szabó, who turned 85 this fall. He was a student of the Faculty of Mining Engineering of the Technical University for Heavy Industry in Miskolc (now the University of Miskolc) between 1954 and 1956. For political reasons Professor Szabó had to leave Hungary following the failed Hungarian uprising in 1956. After emigrating from Hungary he completed his university studies in Canada, and then obtained a PhD degree and made a scientific career in the United States. It is worth mentioning his important contribution to the theory of the finite element method here. Readers who would like to have more details concerning Professor Szabó's scientific activity are referred to paper [1] in JCAM.

One of his fundamental results was the observation of the fact that increasing the polynomial degree p of elements on a fixed mesh results in a rate of convergence in energy norm that is faster than if fixed p and uniform or quasi-uniform mesh refinement, known as the h-version, are used. The term "p-version of the finite element method" first appeared in a 1981 publication [2] in which the theoretical foundations were established for a discretization strategy whereby the finite element mesh is fixed and the polynomial degree p of the elements is progressively increased. The results presented in this paper motivated research in the applied mathematics community on the properties of high order finite element methods, which still continues.

Szabó recognized that it was of fundamental importance from the engineering and scientific perspectives to formulate mathematical problems that simulate some specific aspects of a physical reality with sufficient reliability to justify basing engineering decisions on them.

In 1989 Szabó co-founded a company called Engineering Software Research and Development, Inc. (ESRD). The mission of this company is "to create and market software tools for the advancement of the quality, reliability and timeliness of information that serves the engineering decision-making process[1]". ESRD produces and markets the software StressCheck, which is the only finite element analysis software tool designed to meet the technical requirements of simulation governance [3]. It is used primarily in the aerospace sector.

Szabó has published over 150 papers and two textbooks [4],[5]. The second edition of book [4] is coming out soon. He is a founding member and Fellow of the US

[1] www.esrd.com

Association for Computational Mechanics. Among his honors are election to the Hungarian Academy of Sciences as External Member in 1995 and Doctor Honoris Causa, University of Miskolc in 1998.

Professor Szabó has never forgotten about his roots. He used to say that "Half of my heart beats for the country that helped me to become who I am now and the second part of my heart beats for my homeland Hungary." Since the early seventies – the communist authorities in Hungary had not allowed him to visit his homeland earlier – there has been a living cooperation between him and members of the Institute of Applied Mechanics at the Miskolc University. We are very grateful for his continuous support and the help he has been providing us even now.

Miskolc, November 15, 2020 László Baranyi, István Páczelt and György Szeidl

References

1. L. Baranyi, I. Páczelt, and G. Szeidl. "Preface." *Journal of Computational and Applied Mechanics,* **10**(2), (2015), pp. 105–112. DOI: 10.32973/jcam.2015.007.
2. I. Babuška, B. Szabó, and N. Katz. "The p-version of the finite element method." *SIAM Journal on Numerical Analysis,* **18**(3), (1981), pp. 515–545. DOI: 10.1137/0718033.
3. B. Szabó and R. Actis. "Simulation governance: Technical requirements for mechanical design." *Computer Methods in Applied Mechanics and Engineering,* **249-252**, (2012), pp. 158–168. DOI: 10.1016/j.cma.2012.02.008.
4. B. Szabó and I. Babuška. *Finite Element Analysis.* John Wiley & Sons Inc., New York, 1991.
5. B. Szabó and I. Babuška. *Introduction to Finite Element Analysis: Formulation, Verification and Validation.* John Wiley & Sons Inc., New York, 2011.

Journal of Computational and Applied Mechanics, Vol. 15, No. 2, (2020), pp. 81–100
DOI: 10.32973/jcam.2020.006

EQUILIBRATED STRESS SPACE FOR NONLINEAR DIMENSIONALLY REDUCED SHELL MODELS IN TERMS OF FIRST-ORDER STRESS FUNCTIONS

Edgár Bertóti
Institute of Applied Mechanics, University of Miskolc
H-3515 Miskolc-Egyetemváros, Hungary
edgar.bertoti@uni-miskolc.hu

[Received: August 18, 2020; Accepted: September 26, 2020]

Dedicated to Professor Barna Szabó on the occasion of his eighty-fifth birthday

Abstract. Considering the power series expansion of the three-dimensional variables with respect to the shell thickness coordinate, an equilibrated stress space for the first Piola-Kirchhoff stress vectors is derived in convective curvilinear coordinate system. The infinite series of the two-dimensional translational equilibrium equations are satisfied by introducing two first-order stress function vectors expanded into power series. For the important case of thin shells, the infinite number of two-dimensional equilibrium equations is truncated to obtain a 'first-order' model, where the equilibrated stress-space requires three vectorial stress function coefficients only. The formulation presented for thin shells is compared to the nonlinear equilibrium equations of the classical shell theories, written in terms of the first Piola-Kirchhoff stress resultants and stress couples and satisfied by the introduction of three first-order stress function vectors.

Mathematical Subject Classification: 05C38, 15A15
Keywords: shell theory, equilibrated stress space, first-order stress functions

1. Introduction

A self-equilibrated stress space in elasticity can be generated by stress functions through differentiation. A divergence-free second-order stress tensor requires first-order stress functions and the stress components are obtained by the combinations of their first-order derivatives. When the symmetry of the stress tensor is also *a priori* required, second-order stress functions are needed. In that case the stress components are generated by the combinations of the second-order derivatives of the stress functions.

The terminology 'first-order' and 'second-order' for describing stress functions has been introduced by Fraeijs de Veubeke [1–3]. A self-equilibrated divergence-free stress space requires six non-zero components of the tensor of first-order stress functions [2, 4, 5], whereas a symmetric and divergence-free stress space requires three non-zero components of a second-order stress-function tensor [2, 6]. The main advantage in the use of first-order stress functions, from the point of view of finite element analysis, is that they require C_0-continuous approximations, in contrast to the C_1-continuity

requirement for second-order stress functions. In addition, the use of first-order stress functions "not only simplifies the construction of stress elements by a reduction to C_0 continuity but opens the field to the three- or two-dimensional stress elements with curved boundaries" [2].

The stress function approach and the related complementary energy-based variational principles have been successfully applied in the finite element analysis of two-dimensional elasticity problems, see, e.g., [2, 7–9]. For three-dimensional problems, however, no efficient finite element model exists, which is due mainly to the C_1-continuity requirement for the second-order stress functions on the one hand, and to the structure of the zero-energy modes and the difficulties with their suppression for the case of first-order stress functions, on the other hand [10].

Complementary energy-based equilibrium models using stress functions have also been developed for structural members like plates and shells. Considering the analogy between the special forms of the linear kinematic equations and the equilibrium equations for shells, stress functions were introduced for cylindrical shells by Schaefer [11] and for general shells by Günther [12] and Goldenveizer [13]. The relationship between those shell stress functions and the Maxwell-Morera-type second-order stress functions has been investigated by Yamamoto [14]. Based on the Kirchhoff-Love kinematical assumption, a nonlinear shell theory with finite rotation and stress function vectors was proposed by Simmonds and Danielson [15] and later, using alternate stress and conjugate strain measures, by Atluri [16]. In those works the membrane equilibrium equation is satisfied by one first-order stress function vector and the bending strains are described by one rotation vector. The mixed variational principles proposed in those works rely on both the bending strain energy and the complementary membrane energy of the shell.

An equilibrium finite element model for plates has been developed by Punch and Atluri [17]. The intrinsic shell theory formulated by Valid [18] is also based on the Kirchhoff-Love kinematical assumption and assumes the existence of a two-dimensional surface complementary strain energy density. The linearized equilibrium equations, written in the terms of stress resultants and stress couples, are satisfied by the introduction of a first-order and a second-order stress function vector. For the nonlinear equilibrium equations of the shell theory considered in [18], however, "every attempt to find SA (statically admissible) stresses would be illusory", as stated on page 75 of [18]. Dual-mixed *hp* finite element models for cylindrical shells using first-order stress functions and rotations have been developed in [19] and [20].

The main goal of this paper is to derive an equilibrated stress space for nonlinear shell problems without any *a priori*-made kinematical assumptions for the deformation. Considering the shell as a three-dimensional body, the starting point of the derivation is the three-dimensional equilibrium equations for the first Piola-Kirchhoff stress vectors. The dimensional reduction procedure is based on the Taylor-series expansions of the variables, describing the geometry and the mechanical state of the shell, with respect to the thickness coordinate.

In Section 2 the basic concepts and the notation necessary for understanding the paper, are summarized. Section 3 presents a dimensional reduction procedure for the

derivation of an equilibrated stress space for the first Piola-Kirchhoff stress vectors in convective curvilinear frame, attached to the shell middle surface. The infinite series of the two-dimensional translational equilibrium equations, written in terms of the expanded stress coefficients, are satisfied by introducing two first-order stress function vectors expanded into power series. The equilibrated stress spaces are given for two cases: equilibrium without and with satisfied stress boundary conditions on the shell faces. Section 4 considers the important case of thin shells, for which a 'first-order model' is derived. The equilibrated stress space for the finite number of stress coefficients is compared to the equilibrated stress space derived for stress resultants and stress couples of the classical shell models, and the relationships between the first-order stress function vectors are derived.

2. Geometric preliminaries

This section summarizes some fundamental concepts used in shell theories and introduces the notation applied throughout the paper. Both invariant and index notations are applied. When the index notation is used, the summation convention is applied. Latin indices, as usual, take the range of values 1,2,3 and the Greek indices of the values 1 and 2. The scalar product between two tensors is indicated by one dot. The tensorial or dyadic product between two tensors of any order has no special sign.

2.1. **Reference domain.** To describe the deformation of the shell as a three-dimensional (3D) body, we define first a simply connected two-dimensional (2D) reference domain, a plain denoted by $\bar{S}$, which is bounded by a sufficiently smooth boundary curve $\bar{\ell}$ and parametrized by the coordinates ξ^α. Let a 3D reference domain be defined as

$$V = \{\boldsymbol{\xi} \,|\, \xi^\alpha \in \bar{S},\ |\xi^3| < d/2\}, \tag{2.1}$$

where the coordinate ξ^3 is measured along a straight line perpendicular to the coordinate lines ξ^α at each point of $\bar{S}$, and d is the thickness of the domain (Figure 1). It is assumed that d is independent of ξ^α, i.e., the thickness of V is constant. The surface $\bar{S}$ described by the equation $\xi^3 = 0$ is called the middle surface of V. The boundary surface of V, denoted by S, consists of the top and bottom surfaces

$$S^\pm = \{\boldsymbol{\xi} \,|\, \xi^\alpha \in \bar{S},\ \xi^3 = \pm d/2\} \tag{2.2}$$

and the lateral surface

$$S^\times = \{\boldsymbol{\xi} \,|\, \xi^\alpha \in \bar{\ell},\ |\xi^3| \leq d/2\} \tag{2.3}$$

for which the relations $S = S^\pm \cup S^\times$, $S^\pm \cap S^\times = \emptyset$ hold.

In the subsequent analysis it is assumed that an arbitrary tensor quantity $\boldsymbol{T}(\xi^i)$, defined at the point $\xi^i \in V$, can be differentiated sufficient number of times with respect to ξ^3 and can be expanded into Taylor-series around the point $\xi^3 = 0$:

$$\begin{aligned} \boldsymbol{T}(\xi^i) &= \boldsymbol{T}(\xi^\alpha, 0) + \left.\frac{\partial \boldsymbol{T}}{\partial \xi^3}\right|_{\xi^3=0} \xi^3 + \frac{1}{2}\left.\frac{\partial^2 \boldsymbol{T}}{\partial (\xi^3)^2}\right|_{\xi^3=0} (\xi^3)^2 + \cdots \\ &= {}_0\boldsymbol{T}(\xi^\alpha) + {}_1\boldsymbol{T}(\xi^\alpha)\,\xi^3 + {}_2\boldsymbol{T}(\xi^\alpha)\,(\xi^3)^2 + \cdots = \sum_{n=0}^{\infty} {}_n\boldsymbol{T}(\xi^\alpha)\,(\xi^3)^n, \end{aligned} \tag{2.4}$$

where the left subscript $n = 0, 1, 2, ...$ refers to the coefficient of the n-th power of ξ^3. Note that the coefficients ${}_n\boldsymbol{T}(\xi^\alpha)$, $n = 0, 1, 2, ...$ depend on the coordinates ξ^α only.

2.2. **Initial and current configuration of the shell.** The initial and the current configuration of the shell are denoted by ${}^0V$ and ${}^tV$, respectively, the material points of which are labeled by the Cartesian coordinates X^A and x^a (Figure 1). The 3D domains ${}^0V$ and ${}^tV$ are obtained as one-to-one nonlinear mappings from the reference configuration V, according to the equations

$$\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{\xi}), \qquad X^A = X^A(\xi^i) \tag{2.5}$$

and

$$\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{\xi}, t), \qquad x^a = x^a(\xi^i, t), \tag{2.6}$$

where ξ^i are the convective coordinates. The deformation of the shell is described by the nonlinear mapping

$$\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{X}, t), \qquad x^a = x^a(X^A, t). \tag{2.7}$$

The relationships between the mapping functions (2.5)-(2.7) are given by $\boldsymbol{x}(\boldsymbol{\xi}, t) = \boldsymbol{x}\,[\,\boldsymbol{X}(\boldsymbol{\xi}), t\,]$ and $\boldsymbol{x}(\boldsymbol{X}, t) = \boldsymbol{x}\,[\,\boldsymbol{\xi}^{-1}(\boldsymbol{X}), t\,]$. The faces of the shell in the initial and the

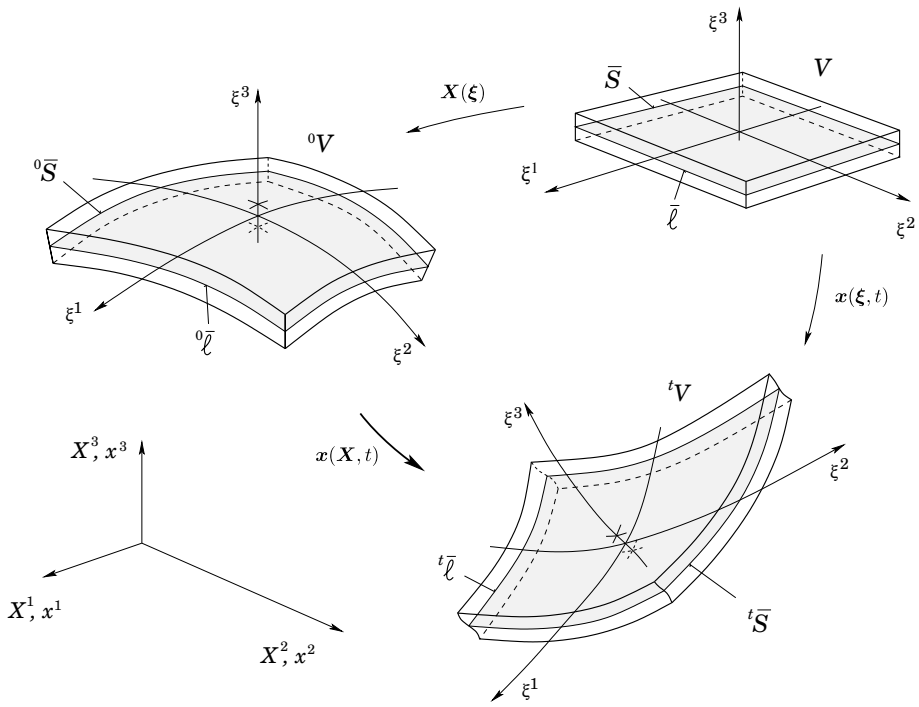


Figure 1. Initial and current configurations of the shell

current configuration are denoted, respectively, by ${}^0S^\pm$ and ${}^tS^\pm$, the lateral surfaces by ${}^0S^\times$ and ${}^tS^\times$, the middle surfaces by ${}^0\bar{S}$ and ${}^t\bar{S}$, and their boundary curves by ${}^0\bar{\ell}$

and ${}^t\bar{\ell}$ (Figure 1). The unit normals to the shell faces are denoted by ${}^0\boldsymbol{n}^{\pm}$ and ${}^t\boldsymbol{n}^{\pm}$, and those to the lateral surfaces by ${}^0\boldsymbol{n}^{\times}$ and ${}^t\boldsymbol{n}^{\times}$.

Applying the material description of the deformation, the mapping (2.5) that defines the initial configuration ${}^0V$ is assumed, without loss of generality, to be a linear function of ξ^3:

$$\boldsymbol{X}(\xi^i) = {}_0\boldsymbol{X}(\xi^\alpha) + {}_1\boldsymbol{X}(\xi^\alpha)\,\xi^3. \tag{2.8}$$

The coordinate lines ξ^3 in ${}^0V$ are straight and assumed to be perpendicular to the middle surface ${}^0\bar{S}$. The mapping (2.6) that gives the current configuration ${}^tV$ is, in general, a non-linear function of ξ^3 and can be given as

$$\boldsymbol{x}(\xi^i) = {}_0\boldsymbol{x}(\xi^\alpha) + {}_1\boldsymbol{x}(\xi^\alpha)\,\xi^3 + {}_2\boldsymbol{x}(\xi^\alpha)\,(\xi^3)^2 + \ldots \tag{2.9}$$

with an unknown number of coefficients ${}_n\boldsymbol{x}(\xi^\alpha)$, $n = 0, 1, 2, \ldots$. This means that the coordinate line ξ^3 in ${}^tV$ is usually neither straight nor perpendicular to the deformed middle surface ${}^t\bar{S}$.

Remark. Classical plate and shell models are usually based on kinematical assumptions that can be considered as restrictions to the mapping (2.6) and its Taylor series (2.9). The Reissner-Mindlin plate model [21, 22] and the Naghdi shell model [23, 24], for example, assume that the mapping (2.9) is linear in ξ^3. The Kirchhoff-plate model [25] and the Love-Novozhilov-Sanders-Koiter shell model [26–29] assume, in addition, that the coordinate line ξ^3 is perpendicular to the deformed middle surface, according to the Kirchhoff-Love hypothesis.

2.3. **Convective base vectors.** The convective covariant base vectors in the initial configuration are given by

$$\boldsymbol{A}_a(\xi^i) = \frac{\partial \boldsymbol{X}}{\partial \xi^a} \tag{2.10}$$

which, using (2.8), can be written as

$$\boldsymbol{A}_\alpha = \frac{\partial\, {}_0\boldsymbol{X}}{\partial \xi^\alpha} + \frac{\partial\, {}_1\boldsymbol{X}}{\partial \xi^\alpha}\,\xi^3 = {}_0\boldsymbol{A}_\alpha(\xi^\alpha) + {}_1\boldsymbol{A}_\alpha(\xi^\alpha)\,\xi^3, \tag{2.11}$$

$$\boldsymbol{A}_3 = {}_1\boldsymbol{X}(\xi^\alpha) = {}_0\boldsymbol{A}_3\,(\xi^\alpha), \tag{2.12}$$

where ${}_0\boldsymbol{A}_a(\xi^\alpha)$ denotes the (convective, covariant) tangent base vectors to the middle surface ${}^0\bar{S}$ (see Figure 2). According to (2.11)-(2.12), the variation of $\boldsymbol{A}_\alpha$ in ξ^3 is linear, whereas $\boldsymbol{A}_3 = {}_0\boldsymbol{A}_3$ is constant in ξ^3. Since the coordinate line ξ^3 is perpendicular to the middle surface ${}^0\bar{S}$, $\boldsymbol{A}_3 = {}_0\boldsymbol{A}_3$ is the unit normal to the surfaces $\xi^3 =$ constant in the initial configuration.

Taking into account (2.12), the coefficients ${}_1\boldsymbol{A}_\alpha$ in the power series (2.11) can be expressed by the base vectors ${}_0\boldsymbol{A}_\beta$ as

$${}_1\boldsymbol{A}_\alpha = \frac{\partial\, {}_1\boldsymbol{X}}{\partial \xi^\alpha} = \frac{\partial\, {}_0\boldsymbol{A}_3}{\partial \xi^\alpha} = -B_\alpha^\beta\, {}_0\boldsymbol{A}_\beta\,, \tag{2.13}$$

where

$$\boldsymbol{B} = B_\alpha^\beta\, {}_0\boldsymbol{A}_\beta\, {}_0\boldsymbol{A}^\alpha = -\,{}_1\boldsymbol{A}_\alpha\, {}_0\boldsymbol{A}^\alpha, \qquad B_\alpha^\beta = -\,{}_1\boldsymbol{A}_\alpha \cdot {}_0\boldsymbol{A}^\beta \tag{2.14}$$

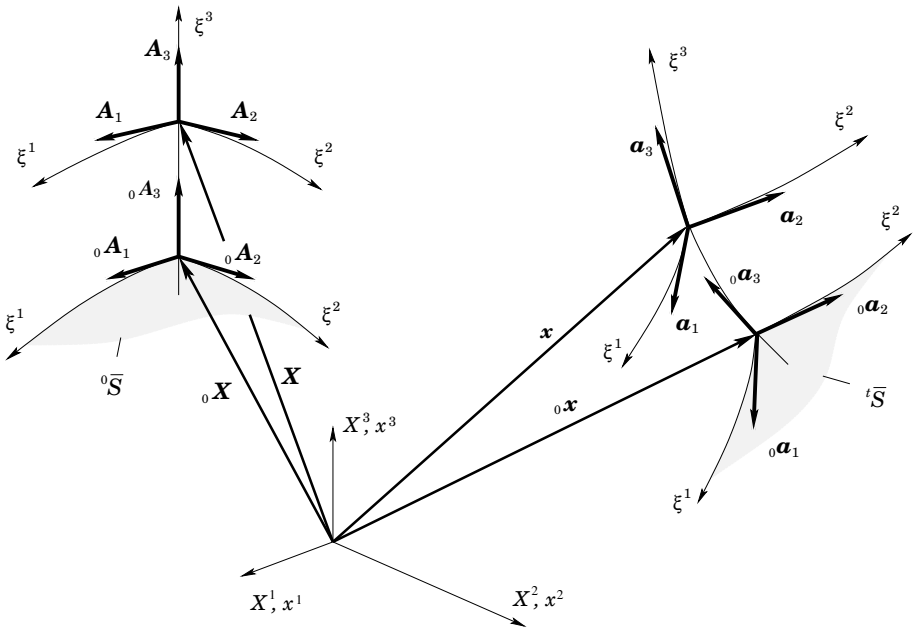


Figure 2. Base vectors in the reference and the current configuration

is the (symmetric) curvature tensor of the middle surface ${}^0\bar{S}$. Taking into account (2.13), equation (2.11) can be rewritten in the form

$$\boldsymbol{A}_\alpha = (\delta^\beta_\alpha - B^\beta_\alpha\,\xi^3)\,{}_0\boldsymbol{A}_\beta\,. \tag{2.15}$$

Introducing the shifter tensor as

$$\mu^b_a = \boldsymbol{A}_a \cdot {}_0\boldsymbol{A}^b, \qquad \mu^\beta_\alpha = \delta^\beta_\alpha - B^\beta_\alpha\,\xi^3, \qquad \mu^\beta_3 = \mu^3_\alpha = 0, \qquad \mu^3_3 = 1, \tag{2.16}$$

$$\boldsymbol{\mu} = \mu^b_a\ {}_0\boldsymbol{A}_b\ {}_0\boldsymbol{A}^a = \boldsymbol{A}_a\ {}_0\boldsymbol{A}^a \tag{2.17}$$

and its inverse

$$(\mu^{-1})^b_a = {}_0\boldsymbol{A}_a \cdot \boldsymbol{A}^b, \qquad \boldsymbol{\mu}^{-1} = (\mu^{-1})^b_a\ \boldsymbol{A}_b\ \boldsymbol{A}^a = {}_0\boldsymbol{A}_a\ \boldsymbol{A}^a \qquad \boldsymbol{\mu}\cdot\boldsymbol{\mu}^{-1} = \mathbf{1}, \tag{2.18}$$

the relationships between the covariant base vectors (2.12) and (2.15) can be given by

$$\boldsymbol{A}_a = \mu^b_a\ {}_0\boldsymbol{A}_b, \qquad {}_0\boldsymbol{A}_a = (\mu^{-1})^b_a\ \boldsymbol{A}_b\,, \tag{2.19}$$

and those between the contravariant base vectors by

$$\boldsymbol{A}^b = (\mu^{-1})^b_a\ {}_0\boldsymbol{A}^a, \qquad {}_0\boldsymbol{A}^b = \mu^b_a\ \boldsymbol{A}^a. \tag{2.20}$$

The determinant of the shifter $\boldsymbol{\mu}$ in terms of the curvature tensor is given by

$$\mu = \det\boldsymbol{\mu} = 1 - (\operatorname{tr}\boldsymbol{B})\,\xi^3 + (\det\boldsymbol{B})\,(\xi^3)^2, \tag{2.21}$$

where $\operatorname{tr}\boldsymbol{B} = B^\alpha_\alpha$ and $\det\boldsymbol{B}$ is the determinant of $\boldsymbol{B}$. For thin shells the approximations

$$\boldsymbol{\mu} = \mathbf{1}, \qquad \mu^b_a = \delta^b_a, \qquad \mu = 1 \tag{2.22}$$

can be applied for the shifter. The derivative of the shifter (2.16) and its determinant (2.21) with respect to ξ^3 are proportional to the curvature of the shell, i.e., they cannot be neglected either for thin shells.

The covariant components of the metric tensor at the arbitrary point of the shell in the initial configuration are given by $A_{kl} = \boldsymbol{A}_k \cdot \boldsymbol{A}_l$, and on the middle surface ${}^0\bar{S}$ by ${}_0A_{kl} = {}_0\boldsymbol{A}_k \cdot {}_0\boldsymbol{A}_l$. Their determinants, denoted by $A = \det A_{kl}$ and ${}_0A = \det {}_0A_{kl}$, respectively, are related to each other by

$$\sqrt{A} = \mu \sqrt{{}_0A}, \tag{2.23}$$

where $\sqrt{A}$ is also known as the Jacobian of the mapping $\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{\xi})$. Using the contravariant metric components A^{ij} and ${}_0A^{ij}$, the relationships between the co- and contravariant base vectors are given by

$$\boldsymbol{A}^i = A^{ij}\boldsymbol{A}_j, \qquad {}_0\boldsymbol{A}^i = {}_0A^{ij}\,{}_0\boldsymbol{A}_j\,. \tag{2.24}$$

The convective covariant base vectors in the current configuration are defined by

$$\boldsymbol{a}_p(\xi^i) = \frac{\partial \boldsymbol{x}}{\partial \xi^p}. \tag{2.25}$$

Taking into account (2.9), they can be written as

$$\boldsymbol{a}_p(\,\xi^i) = {}_0\boldsymbol{a}_p(\xi^\alpha) + {}_1\boldsymbol{a}_p(\xi^\alpha)\,\xi^3 + {}_2\boldsymbol{a}_p(\xi^\alpha)\,(\xi^3)^2 + \ldots\,, \tag{2.26}$$

where ${}_0\boldsymbol{a}_p(\xi^\alpha)$ are the base vectors on the deformed middle surface ${}^t\bar{S}$ (Figure 2). As can be seen from (2.26), $\boldsymbol{a}_p$ is, in general, a nonlinear function of ξ^3. When the deformation gradient

$$\boldsymbol{F} = \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{X}} = \boldsymbol{a}_p\,\boldsymbol{A}^p \tag{2.27}$$

is known, the convective base vectors in the current configuration can be obtained as

$$\boldsymbol{a}_p = \boldsymbol{F} \cdot \boldsymbol{A}_p, \qquad \boldsymbol{a}^q = \boldsymbol{F}^{-T} \cdot \boldsymbol{A}^q. \tag{2.28}$$

3. Equilibrated stress space of the shell

Nonlinear dual-mixed finite element models based on equilibrated stress spaces can be constructed by satisfying the translational equilibrium equation for the first Piola-Kirchhoff stress tensor through the introduction of first-order stress functions. The rotational equilibrium, i.e., the symmetry of the Cauchy stress tensor, is satisfied in a weak sense through the variational formulation and by the corresponding finite element model. For shells with arbitrary geometry, the construction of an equilibrated stress space in convective curvilinear coordinates requires special considerations. As will be shown in this section, the key issue is the use of the vectorial form of the translational equilibrium equation in the shell coordinate space. The Taylor-series expansion of the stress vectors leads to a sequence of two-dimensional translational equilibrium equations that can be satisfied by introducing first-order stress function vectors.

3.1. **Vectorial equilibrium equations.**

3.1.1. *Translational equilibrium.* The three-dimensional translational equilibrium equation for the first Piola-Kirchhoff stress tensor $\boldsymbol{P} = P^{k\ell}\boldsymbol{a}_k\,\boldsymbol{A}_\ell$ in the convective coordinate system is given by

$$\boldsymbol{P}\cdot{}^0\nabla + {}^0\rho\boldsymbol{b} = \boldsymbol{0} \qquad \boldsymbol{X}\in{}^0V, \tag{3.1}$$

where

$${}^0\nabla = \boldsymbol{A}^i\frac{\partial}{\partial\xi^i} \tag{3.2}$$

is the differential operator in the initial configuration, ${}^0\rho$ is the mass density in the undeformed state and $\boldsymbol{b}$ is the prescribed body force density per unit mass. Introducing the first Piola-Kirchhoff stress vectors as

$$\boldsymbol{p}^\ell = \boldsymbol{P}\cdot\boldsymbol{A}^\ell, \qquad\qquad \boldsymbol{P} = \boldsymbol{p}^\ell\boldsymbol{A}_\ell \tag{3.3}$$

and making use of the identity

$$\boldsymbol{P}\cdot{}^0\nabla = \frac{\partial\boldsymbol{p}^\ell}{\partial\xi^\ell} + \frac{1}{\sqrt{A}}\,\frac{\partial\sqrt{A}}{\partial\xi^\ell}\,\boldsymbol{p}^\ell \tag{3.4}$$

for the divergence of the second-order tensor $\boldsymbol{P}$ (see, e.g., [30]), the equilibrium equation (3.1) can be rewritten in vectorial form:

$$\frac{1}{\sqrt{A}}\,(\sqrt{A}\,\boldsymbol{p}^\ell)_{,\ell} + {}^0\rho\boldsymbol{b} = \boldsymbol{0} \qquad \boldsymbol{X}\in{}^0V. \tag{3.5}$$

Note that equation (3.1) contains the covariant derivative of the stress components $P^{k\ell}$, whereas the vectorial equation (3.5) contains the partial derivative of the stress vectors $\boldsymbol{p}^\ell$. This is an important fact when the construction of an equilibrated stress space is needed in a general curvilinear coordinate frame.

3.1.2. *Rotational equilibrium.* The symmetry of the Cauchy stress tensor $\boldsymbol{\sigma} = J^{-1}\boldsymbol{P}\cdot\boldsymbol{F}^T$ in terms of the first Piola-Kirchhoff stress vectors is expressed by the vectorial equation

$$\boldsymbol{a}_\ell\times\boldsymbol{p}^\ell = \boldsymbol{0}, \tag{3.6}$$

where $\boldsymbol{a}_\ell$ are the base vectors defined by (2.25). Satisfaction of the rotational equilibrium equation (3.6) is not required *a priori* in the stress-based dual-mixed shell model under consideration, it is given here only for the sake of completeness.

3.2. **Equilibrated stress vectors.** To derive an equilibrated space for the first Piola-Kirchhoff stress vectors $\boldsymbol{p}^\ell$ in the convective coordinate system, the three-dimensional translational equilibrium equation (3.5) is rewritten first by making use of (2.23):

$$\frac{1}{\sqrt{{}_0A}}\,(\sqrt{{}_0A}\,\mu\boldsymbol{p}^\ell)_{,\ell} + {}^0\rho\mu\boldsymbol{b} = \boldsymbol{0} \tag{3.7}$$

in which $\sqrt{{}_0A}$ depends on ξ^α only. The stress vectors $\mu\boldsymbol{p}^\ell(\xi^i)$ and the body force vector ${}^0\rho\mu\boldsymbol{b}$ are expanded next into power series with respect to the coordinate $\xi^3\equiv\zeta$:

$$\hat{\boldsymbol{p}}^\ell := \mu\boldsymbol{p}^\ell(\xi^\alpha,\zeta) \quad = \quad \sum_{n=0}^{\infty}{}_n\hat{\boldsymbol{p}}^\ell(\xi^\alpha)\,\zeta^n, \tag{3.8}$$

$$\hat{\boldsymbol{b}} := {}^{0}\rho\mu\boldsymbol{b}(\xi^{\alpha},\zeta) \quad = \quad \sum_{n=0}^{\infty} {}_{n}\hat{\boldsymbol{b}}(\xi^{\alpha})\,\zeta^{n}. \tag{3.9}$$

Inserting (3.8) and (3.9) in (3.7), the three-dimensional equilibrium equation is replaced by a series of two-dimensional equations, according to the powers of ζ:

$$\frac{1}{\sqrt{{}_{0}A}}\,(\sqrt{{}_{0}A}\;{}_{n}\hat{\boldsymbol{p}}^{\lambda})_{,\lambda} + (n+1)\;{}_{n+1}\hat{\boldsymbol{p}}^{3} + {}_{n}\hat{\boldsymbol{b}} = \boldsymbol{0}, \qquad n = 0,1,2,3,\dots\,. \tag{3.10}$$

Equilibrium equations in (3.10) can be satisfied by introducing two arbitrary but differentiable first-order stress function vectors, denoted by $\boldsymbol{\psi}$ and $\boldsymbol{\chi}$, and expanding them into power series with respect to ζ as

$$\boldsymbol{\psi}(\xi^{\alpha},\zeta) \quad = \quad \sum_{n=0}^{\infty} {}_{n}\boldsymbol{\psi}(\xi^{\alpha})\,\zeta^{n}, \tag{3.11}$$

$$\boldsymbol{\chi}(\xi^{\alpha},\zeta) \quad = \quad \sum_{n=0}^{\infty} {}_{n}\boldsymbol{\chi}(\xi^{\alpha})\,\zeta^{n}. \tag{3.12}$$

Then the equilibrium equations (3.10) are identically satisfied by the following stress vectors:

$${}_{n}\hat{\boldsymbol{p}}^{\lambda} \quad = \quad {}_{0}\epsilon^{\lambda\beta}\;{}_{n}\boldsymbol{\psi}_{,\beta} - (n+1)\;{}_{0}\epsilon^{3\lambda}\;{}_{n+1}\boldsymbol{\chi} - \frac{1}{2\sqrt{{}_{0}A}}\;{}_{n}\boldsymbol{f}^{\lambda}, \quad n = 0,1,2,3,\dots\,, \tag{3.13}$$

$${}_{n}\hat{\boldsymbol{p}}^{3} \quad = \quad {}_{0}\epsilon^{3\beta}\;{}_{n}\boldsymbol{\chi}_{,\beta}\,, \qquad n = 1,2,3,4,\dots\,, \tag{3.14}$$

where ${}_{0}\epsilon^{\lambda\beta}$ and ${}_{0}\epsilon^{3\beta}$ are two-dimensional permutation symbols defined by

$$\sqrt{{}_{0}A}\;{}_{0}\epsilon^{\lambda\beta} = \begin{cases} 1, \text{ if } \lambda = 1,\ \beta = 2 \\ -1, \text{ if } \lambda = 2,\ \beta = 1 \end{cases}, \qquad \sqrt{{}_{0}A}\;{}_{0}\epsilon^{3\beta} = \begin{cases} 1, \text{ if } \beta = 1 \\ -1, \text{ if } \beta = 2 \end{cases}, \tag{3.15}$$

and

$${}_{n}\boldsymbol{f}^{\lambda} = \int_{(\xi^{\lambda})} \sqrt{{}_{0}A}\;{}_{n}\hat{\boldsymbol{b}}\;\mathrm{d}\xi^{\lambda}, \quad n = 0,1,2,\dots \tag{3.16}$$

are known load vectors obtained from the prescribed body forces by integration along the coordinate lines ξ^{1} and ξ^{2}. Note that ${}_{0}\hat{\boldsymbol{p}}^{3}$ does not appear in (3.10) and, therefore, in (3.14), its value is arbitrary, at least from the point of view of equilibrium.

Dual-mixed finite element models based on the above stress space require the satisfaction of the stress boundary conditions on the top and bottom faces, as well as on the lateral surfaces of the shell. These boundary conditions can be satisfied by applying the Lagrange multiplier technique, i.e., using hybridized elements.

3.3. **Equilibrated stress vectors with satisfied stress boundary conditions on the faces.** In the classical, dimensionally reduced shell models it is generally assumed that on the top and bottom faces of the shells only surface loads are present and the prescribed tractions are naturally included in the equilibrium equations written in terms of stress resultants and stress couples. Considering the equilibrium equations in the form (3.10), the prescribed loads on the faces of the shell can be taken into account as follows.

3.3.1. *Stress boundary conditions on the faces of the shell.* The stress boundary conditions for the first Piola-Kirchhoff stress vectors (3.3) at the top and bottom faces ${}^0S^{\pm}$ of the shell with outward unit normals ${}^0\boldsymbol{n}^{\pm} = \pm\boldsymbol{A}^3 = \pm\boldsymbol{A}_3$ are given by

$$\pm\, \boldsymbol{p}^3 = {}^0\widetilde{\boldsymbol{p}}^{\pm} \qquad \boldsymbol{X} \in {}^0S^{\pm}, \tag{3.17}$$

and on the lateral surface ${}^0S_p^{\times}$ with outward unit normal ${}^0\boldsymbol{n}^{\times} = {}^0n_{\lambda}^{\times}\,\boldsymbol{A}^{\lambda}$ is given by

$${}^0n_{\lambda}^{\times}\,\boldsymbol{p}^{\lambda} = {}^0\widetilde{\boldsymbol{p}}^{\times} \qquad \boldsymbol{X} \in {}^0S_p^{\times}, \tag{3.18}$$

where ${}^0\widetilde{\boldsymbol{p}}^{\pm}$ and ${}^0\widetilde{\boldsymbol{p}}^{\times}$ are prescribed tractions.

3.3.2. *Vectorial equilibrium including the surface load components.* In view of the left side of (3.8), the stress boundary conditions (3.17), prescribed on the top and bottom faces of the shell for the first Piola-Kirchhoff stress vector $\boldsymbol{p}^3$, take the form

$$\pm\, \hat{\boldsymbol{p}}^3 = \mu^{\pm}\, {}^0\widetilde{\boldsymbol{p}}^{\pm} \qquad \boldsymbol{X} \in {}^0S^{\pm} \tag{3.19}$$

which, taking into account (3.8), is rewritten as

$${}_0\hat{\boldsymbol{p}}^3 + \frac{d}{2}\,{}_1\hat{\boldsymbol{p}}^3 + (\frac{d}{2})^2\,{}_2\hat{\boldsymbol{p}}^3 + ... + (\frac{d}{2})^n\,{}_n\hat{\boldsymbol{p}}^3 + ... = \pm\,\mu^{\pm}\,{}^0\widetilde{\boldsymbol{p}}^{\pm}. \tag{3.20}$$

Subtracting and adding the two equations corresponding to the signs $\pm$ in (3.20), the equations

$${}_0\hat{\boldsymbol{p}}^3 + (\frac{d}{2})^2\,{}_2\hat{\boldsymbol{p}}^3 + (\frac{d}{2})^4\,{}_4\hat{\boldsymbol{p}}^3 + ... \;=\; \frac{1}{2}\,(\mu^+\,{}^0\widetilde{\boldsymbol{p}}^+ - \mu^-\,{}^0\widetilde{\boldsymbol{p}}^-), \tag{3.21}$$

$${}_1\hat{\boldsymbol{p}}^3 + (\frac{d}{2})^2\,{}_3\hat{\boldsymbol{p}}^3 + (\frac{d}{2})^4\,{}_5\hat{\boldsymbol{p}}^3 + ... \;=\; \frac{1}{d}\,(\mu^+\,{}^0\widetilde{\boldsymbol{p}}^+ + \mu^-\,{}^0\widetilde{\boldsymbol{p}}^-) \tag{3.22}$$

are obtained. For the sake of compact notation, let the load vector components

$${}_0\widetilde{\boldsymbol{p}}(\xi^{\alpha}) \;=\; \frac{1}{2}\,(\mu^+\,{}^0\widetilde{\boldsymbol{p}}^+ - \mu^-\,{}^0\widetilde{\boldsymbol{p}}^-), \tag{3.23}$$

$${}_1\widetilde{\boldsymbol{p}}(\xi^{\alpha}) \;=\; \frac{1}{d}\,(\mu^+\,{}^0\widetilde{\boldsymbol{p}}^+ + \mu^-\,{}^0\widetilde{\boldsymbol{p}}^-) \tag{3.24}$$

be introduced. Then the load vector defined by

$$\widetilde{\boldsymbol{p}} \;=\; {}_0\widetilde{\boldsymbol{p}} \;+\; {}_1\widetilde{\boldsymbol{p}}\,\zeta \tag{3.25}$$

at $\zeta = \pm\, d/2$ takes the values of the prescribed loads on the top and bottom surfaces:

$$\widetilde{\boldsymbol{p}}(\xi^{\alpha},\, \pm\, d/2) = \pm\,\mu^{\pm}\,{}^0\widetilde{\boldsymbol{p}}^{\pm}. \tag{3.26}$$

Taking into account (3.23)-(3.24), the coefficients ${}_1\hat{\boldsymbol{p}}^3$ and ${}_2\hat{\boldsymbol{p}}^3$ can be expressed from the stress boundary conditions (3.21)-(3.22) as

$${}_1\hat{\boldsymbol{p}}^3 \;=\; {}_1\widetilde{\boldsymbol{p}} - \sum_{i=1}^{\infty}\left(\frac{d}{2}\right)^{2i}\,{}_{2i+1}\hat{\boldsymbol{p}}^3, \tag{3.27}$$

$${}_2\hat{\boldsymbol{p}}^3 \;=\; \frac{4}{d^2}(\,{}_0\widetilde{\boldsymbol{p}} \;-\; {}_0\hat{\boldsymbol{p}}^3) - \sum_{i=1}^{\infty}\left(\frac{d}{2}\right)^{2i}\,{}_{2i+2}\hat{\boldsymbol{p}}^3. \tag{3.28}$$

Inserting them in the first two equilibrium equations of (3.10) that correspond to $n = 0$ and $n = 1$, the translational equilibrium equations including the prescribed loads on the faces are given by

$$n = 0: \qquad \frac{1}{\sqrt{{}_0A}}\,(\sqrt{{}_0A}\;{}_0\hat{\boldsymbol{p}}^{\lambda})_{,\lambda} - \sum_{i=1}^{\infty}\left(\frac{d}{2}\right)^{2i}{}_{2i+1}\hat{\boldsymbol{p}}^{3} + {}_1\widetilde{\boldsymbol{p}} + {}_0\hat{\boldsymbol{b}} = \boldsymbol{0}, \tag{3.29}$$

$$n = 1: \qquad \frac{1}{\sqrt{{}_0A}}\,(\sqrt{{}_0A}\;{}_1\hat{\boldsymbol{p}}^{\lambda})_{,\lambda} - 2\sum_{i=1}^{\infty}\left(\frac{d}{2}\right)^{2i}{}_{2i+2}\hat{\boldsymbol{p}}^{3} + \frac{8}{d^2}({}_0\widetilde{\boldsymbol{p}} - {}_0\hat{\boldsymbol{p}}^{3}) + {}_1\hat{\boldsymbol{b}} = \boldsymbol{0}, \tag{3.30}$$

$$n \geq 2: \qquad \frac{1}{\sqrt{{}_0A}}\,(\sqrt{{}_0A}\;{}_n\hat{\boldsymbol{p}}^{\lambda})_{,\lambda} + (n+1)\;{}_{n+1}\hat{\boldsymbol{p}}^{3} + {}_n\hat{\boldsymbol{b}} = \boldsymbol{0}. \tag{3.31}$$

Equations (3.31) for the case $n \geq 2$ are identical with those in (3.10). Note also that the value of ${}_0\hat{\boldsymbol{p}}^{3}$ is not arbitrary in this case: due to (3.20) and (3.28), it appears in the equilibrium equations.

3.3.3. *Equilibrated stress vectors.* Equilibrium equations (3.29)-(3.31) can identically be satisfied by the stress function vectors $\boldsymbol{\psi}$ and $\boldsymbol{\chi}$, introduced in Section 3.2. Considering their power series expansion (3.11)-(3.12) with respect to $\xi^3 \equiv \zeta$, the equilibrated stress vectors are given by

$${}_0\hat{\boldsymbol{p}}^{\lambda} = {}_0\epsilon^{\lambda\beta}\;{}_0\boldsymbol{\psi}_{,\beta} - {}_0\epsilon^{3\lambda}\sum_{i=1}^{\infty}\left(\frac{d}{2}\right)^{2i}{}_{2i+1}\boldsymbol{\chi} - \frac{1}{2\sqrt{{}_0A}}\;{}_1\boldsymbol{f}^{\lambda}, \tag{3.32}$$

$${}_1\hat{\boldsymbol{p}}^{\lambda} = {}_0\epsilon^{\lambda\beta}\;{}_1\boldsymbol{\psi}_{,\beta} + \frac{8}{d^2}\;{}_0\epsilon^{3\lambda}\;{}_0\boldsymbol{\chi} - 2\;{}_0\epsilon^{3\lambda}\sum_{i=1}^{\infty}\left(\frac{d}{2}\right)^{2i}{}_{2i+2}\boldsymbol{\chi} - \frac{1}{2\sqrt{{}_0A}}\;{}_0\boldsymbol{f}^{\lambda}, \tag{3.33}$$

$${}_n\hat{\boldsymbol{p}}^{\lambda} = {}_0\epsilon^{\lambda\beta}\;{}_n\boldsymbol{\psi}_{,\beta} - (n+1)\;{}_0\epsilon^{3\lambda}\;{}_{n+1}\boldsymbol{\chi} - \frac{1}{2\sqrt{{}_0A}}\;{}_n\boldsymbol{f}^{\lambda}, \qquad n = 2,3,4,..., \tag{3.34}$$

$${}_n\hat{\boldsymbol{p}}^{3} = {}_0\epsilon^{3\lambda}\;{}_n\boldsymbol{\chi}_{,\lambda}, \qquad n = 0,3,4,5,\dots, \tag{3.35}$$

where

$${}_0\boldsymbol{f}^{\lambda} = \int_{(\xi^{\lambda})}\sqrt{{}_0A}\,(\,\frac{8}{d^2}\;{}_0\widetilde{\boldsymbol{p}} + {}_1\hat{\boldsymbol{b}})\,\mathrm{d}\xi^{\lambda}, \tag{3.36}$$

$${}_1\boldsymbol{f}^{\lambda} = \int_{(\xi^{\lambda})}\sqrt{{}_0A}\,(\,{}_1\widetilde{\boldsymbol{p}} + {}_0\hat{\boldsymbol{b}})\,\mathrm{d}\xi^{\lambda} \tag{3.37}$$

are known load vectors obtained from the prescribed body forces and surface loads by integration along the coordinate lines ξ^1 and ξ^2. The load vectors ${}_n\boldsymbol{f}^{\lambda}$, $n = 2,3,....$ in (3.34) are defined by (3.16) and the permutation symbols ${}_0\epsilon^{\lambda\beta}$ and ${}_0\epsilon^{3\beta}$ by (3.15).

Note that since the stress vectors ${}_1\hat{\boldsymbol{p}}^{3}$ and ${}_2\hat{\boldsymbol{p}}^{3}$ are computed in this case from the stress boundary conditions on the faces, according to (3.27) and (3.28), the equilibrated stress space (3.32)-(3.35) does not require the use of the stress function coefficients ${}_1\boldsymbol{\chi}$ and ${}_2\boldsymbol{\chi}$. This means that the finite element models based on the stress space (3.32)-(3.35), with satisfied stress boundary condition on the faces, are more efficient than those based on (3.13) and (3.14) of Section 3.2. The fact that the number of the unknown stress function components is less by 6 for the present case has

a special advantage in the modeling of thin shells, where a first-order approximation with two equilibrium equations from (3.10) is usually enough for an accurate solution (see Section 4).

3.4. **Equilibrated stress components.** The equilibrated components of the first Piola-Kirchhoff stress tensor can be obtained from the definitions of the stress vectors (3.3) and (3.8). Introducing the components of the stress vector $\hat{\boldsymbol{p}}^{\ell}$ with respect to the base vectors ${}_0\boldsymbol{A}_k$ as

$$\hat{\boldsymbol{p}}^{\ell} = \hat{P}^{k\ell}\,{}_0\boldsymbol{A}_k, \qquad \hat{P}^{k\ell} = \hat{\boldsymbol{p}}^{\ell}\cdot {}_0\boldsymbol{A}^k = \mu\,\boldsymbol{p}^{\ell}\cdot {}_0\boldsymbol{A}^k, \tag{3.38}$$

the power series of $\hat{\boldsymbol{p}}^{\ell}$ in (3.8) can be rewritten as

$$\hat{\boldsymbol{p}}^{\ell} = \sum_{n=0}^{\infty} {}_n\hat{\boldsymbol{p}}^{\ell}\,\zeta^n = \sum_{n=0}^{\infty} {}_n\hat{P}^{k\ell}\,{}_0\boldsymbol{A}_k\,\zeta^n, \tag{3.39}$$

from which the coefficients ${}_n\hat{\boldsymbol{p}}^{\ell}$ are obtained as

$${}_n\hat{\boldsymbol{p}}^{\ell} = {}_n\hat{P}^{k\ell}\,{}_0\boldsymbol{A}_k, \qquad n = 0,1,2,3,\dots . \tag{3.40}$$

Considering the stress function vectors (3.11) and (3.12) in the expanded forms

$$\boldsymbol{\psi} = \psi^k\,{}_0\boldsymbol{A}_k = \Big(\sum_{n=0}^{\infty} {}_n\psi^k\,\zeta^n\Big)\,{}_0\boldsymbol{A}_k, \tag{3.41}$$

$$\boldsymbol{\chi} = \chi^k\,{}_0\boldsymbol{A}_k = \Big(\sum_{n=0}^{\infty} {}_n\chi^k\,\zeta^n\Big)\,{}_0\boldsymbol{A}_k, \tag{3.42}$$

the equilibrated stress coefficients ${}_n\hat{P}^{k\ell}$ for the case of Section 3.2 are obtained by inserting (3.40)-(3.42) in (3.13) and (3.14):

$${}_n\hat{P}^{k\lambda} = {}_0\epsilon^{\lambda\beta}\,{}_n\psi^k_{\,|\beta} - (n+1)\,{}_0\epsilon^{3\lambda}\,{}_{n+1}\chi^k - \frac{1}{2\sqrt{{}_0A}}\,{}_n(f^k)^{\lambda}, \quad n = 0,1,2,3,\dots , \tag{3.43}$$

$${}_n\hat{P}^{k3} = {}_0\epsilon^{3\beta}\,{}_n\chi^k_{\,|\beta}, \qquad n = 1,2,3,4,\dots , \tag{3.44}$$

where ${}_n(f^k)^{\lambda}$, $n = 0,1,2,3,\dots$ are the components of the load vectors (3.16) with respect to the base vectors ${}_0\boldsymbol{A}_k$, and the sign $|$ in the right subscript followed by β refers to the three-dimensional covariant derivative on the middle surface ${}^0\bar{S}$ with respect to ξ^{β}:

$${}_n\psi^k_{\,|\beta} = {}_n\psi^k_{\,,\beta} + {}_n\psi^{\ell}\,\bar{\Gamma}^k_{\ell\beta} \tag{3.45}$$

with Christoffel-symbols $\bar{\Gamma}^k_{\ell m} = {}_0\boldsymbol{A}_{\ell,m}\cdot {}_0\boldsymbol{A}^k$, defined on the middle surface of the shell in the initial configuration.

The components of the equilibrated stress space for the case of Section 3.3 can be obtained from equations (3.32)-(3.35) by inserting (3.40)-(3.42) in them:

$${}_0\hat{P}^{k\lambda} = {}_0\epsilon^{\lambda\beta}\,{}_0\psi^k_{\,|\beta} - {}_0\epsilon^{3\lambda}\sum_{i=1}^{\infty}\Big(\frac{d}{2}\Big)^{2i}\,{}_{2i+1}\chi^k - \frac{1}{2\sqrt{{}_0A}}\,{}_1(f^k)^{\lambda}, \tag{3.46}$$

$${}_1\hat{P}^{k\lambda} = {}_0\epsilon^{\lambda\beta}\,{}_1\psi^k_{\,|\beta} + \frac{8}{d^2}\,{}_0\epsilon^{3\lambda}\,{}_0\chi^k - 2\,{}_0\epsilon^{3\lambda}\sum_{i=1}^{\infty}\Big(\frac{d}{2}\Big)^{2i}\,{}_{2i+2}\chi^k - \frac{1}{2\sqrt{{}_0A}}\,{}_0(f^k)^{\lambda}, \tag{3.47}$$

$$ {}_n\hat{P}^{k\lambda} = {}_0\epsilon^{\lambda\beta}\,{}_n\psi^{k}_{|\beta} - (n+1)\,{}_0\epsilon^{3\lambda}\,{}_{n+1}\chi^{k} - \frac{1}{2\sqrt{{}_0A}}\,{}_n(f^k)^{\lambda}, \qquad n=2,3,4,..., \tag{3.48} $$

$$ {}_n\hat{P}^{k3} = {}_0\epsilon^{3\lambda}\,{}_n\chi^{k}_{|\lambda}, \qquad n=0,3,4,5,..., \tag{3.49} $$

where ${}_n(f^k)^{\lambda}$, $n=0,1,2,3,...$ are the components of the load vectors (3.36)-(3.37) and (3.16) with respect to the base vectors ${}_0\boldsymbol{A}_k$. The stress components ${}_1\hat{P}^{k3}$ and ${}_2\hat{P}^{k3}$ are obtained from equations (3.27)-(3.28):

$$ {}_1\hat{P}^{k3} = {}_1\tilde{p}^{k} - \sum_{i=1}^{\infty}\left(\frac{d}{2}\right)^{2i}\,{}_{2i+1}\hat{P}^{k3}, \tag{3.50} $$

$$ {}_2\hat{P}^{k3} = \frac{4}{d^2}({}_0\tilde{p}^{k} - {}_0\hat{P}^{k3}) - \sum_{i=1}^{\infty}\left(\frac{d}{2}\right)^{2i}\,{}_{2i+2}\hat{P}^{k3}. \tag{3.51} $$

The shifted components $\bar{P}^{km}$ of the first Piola-Kirchhoff stress tensor with respect to the base vectors ${}_0\boldsymbol{A}_i$ can be obtained using the definition

$$ \boldsymbol{P} = \bar{P}^{km}\,{}_0\boldsymbol{A}_k\,{}_0\boldsymbol{A}_m \tag{3.52} $$

and the relations (3.3), (3.8), (3.38) and (2.19):

$$ \boldsymbol{P} = \boldsymbol{p}^{\ell}\boldsymbol{A}_{\ell} = \frac{1}{\mu}\,\hat{\boldsymbol{p}}^{\ell}\boldsymbol{A}_{\ell} = \frac{1}{\mu}\,\hat{P}^{k\ell}\,{}_0\boldsymbol{A}_k\boldsymbol{A}_{\ell} = \frac{1}{\mu}\,\mu^m_{\ell}\,\hat{P}^{k\ell}\,{}_0\boldsymbol{A}_k\,{}_0\boldsymbol{A}_m. \tag{3.53} $$

On comparing (3.52) and (3.53) the relation

$$ \bar{P}^{km} = \frac{1}{\mu}\,\mu^m_{\ell}\,\hat{P}^{k\ell} \tag{3.54} $$

is obtained. The expanded coefficients of the equilibrated stress components $\bar{P}^{k\ell}$ can be derived from (3.54) by inserting (3.43)-(3.44) or (3.46)-(3.51). The resulting, rather lengthy, expressions are not listed here.

4. The special case of thin shells

In the previous sections there was no limitation on the thickness of the shell. The number of the stress coefficients in the power series has been considered to be infinite, which resulted in an infinite number of two-dimensional equilibrium equations of the shell. When the numerical solution of a shell problem is needed, however, only a finite set of equations can be considered and solved efficiently. Depending on the number of the two-dimensional equilibrium equations chosen, a large variety of dimensionally reduced shell models can be derived, and hierarchic sets of shell models can also be constructed. In this section an equilibrated stress space for thin shells is derived. It is assumed that the stress boundary conditions on the shell faces are *a priori* satisfied, just like in the classical shell models, which means that the procedure described in Section 3.3 can be applied.

4.1. **Equilibrated stress space for thin shells.** In the case of thin shells, let the first two equilibrium equations of (3.10) be selected and let the other equations, corresponding to the higher powers of $\xi^3 \equiv \zeta$, be neglected with the assumption that they are identically satisfied. The two equations for $n=0$ and $n=1$ are

$$\frac{1}{\sqrt{{}_0A}}\,(\sqrt{{}_0A}\;{}_0\hat{\boldsymbol{p}}^{\lambda})_{,\lambda} + {}_1\hat{\boldsymbol{p}}^3 + {}_0\hat{\boldsymbol{b}} = \boldsymbol{0}, \tag{4.1}$$

$$\frac{1}{\sqrt{{}_0A}}\,(\sqrt{{}_0A}\;{}_1\hat{\boldsymbol{p}}^{\lambda})_{,\lambda} + 2\,{}_2\hat{\boldsymbol{p}}^3 + {}_1\hat{\boldsymbol{b}} = \boldsymbol{0}. \tag{4.2}$$

According to these equations, the first Piola-Kirchhoff stress vectors $\hat{\boldsymbol{p}}^{\ell}$, defined in (3.8), are approximated across the thickness of the shell by linear and quadratic polynomials in ζ:

$$\hat{\boldsymbol{p}}^{\lambda}(\xi^{\alpha},\zeta) = {}_0\hat{\boldsymbol{p}}^{\lambda}(\xi^{\alpha}) + {}_1\hat{\boldsymbol{p}}^{\lambda}(\xi^{\alpha})\,\zeta, \tag{4.3}$$

$$\hat{\boldsymbol{p}}^{3}(\xi^{\alpha},\zeta) = {}_0\hat{\boldsymbol{p}}^{3}(\xi^{\alpha}) + {}_1\hat{\boldsymbol{p}}^{3}(\xi^{\alpha})\,\zeta + {}_2\hat{\boldsymbol{p}}^{3}(\xi^{\alpha})\,\zeta^2. \tag{4.4}$$

Following now the method applied in Section 3.3, the coefficients ${}_1\hat{\boldsymbol{p}}^3$ and ${}_2\hat{\boldsymbol{p}}^3$ can be expressed from the stress boundary conditions (3.21)-(3.22) prescribed on the shell faces: in view of (4.4), equations (3.27)-(3.28) simplify to the form

$${}_1\hat{\boldsymbol{p}}^3 = {}_1\widetilde{\boldsymbol{p}}, \tag{4.5}$$

$${}_2\hat{\boldsymbol{p}}^3 = \frac{4}{d^2}\,{}_0\widetilde{\boldsymbol{p}} - \frac{4}{d^2}\,{}_0\hat{\boldsymbol{p}}^3. \tag{4.6}$$

Substituting (4.5)-(4.6) into (4.1)-(4.2), the equilibrium equations for the first Piola-Kirchhoff stress vectors ${}_0\hat{\boldsymbol{p}}^{\ell}$ and ${}_1\hat{\boldsymbol{p}}^{\lambda}$ take the form

$$\frac{1}{\sqrt{{}_0A}}\,(\sqrt{{}_0A}\;{}_0\hat{\boldsymbol{p}}^{\lambda})_{,\lambda} + {}_1\widetilde{\boldsymbol{p}} + {}_0\hat{\boldsymbol{b}} = \boldsymbol{0}, \tag{4.7}$$

$$\frac{1}{\sqrt{{}_0A}}\,(\sqrt{{}_0A}\;{}_1\hat{\boldsymbol{p}}^{\lambda})_{,\lambda} - \frac{8}{d^2}\,{}_0\hat{\boldsymbol{p}}^3 + \frac{8}{d^2}\,{}_0\widetilde{\boldsymbol{p}} + {}_1\hat{\boldsymbol{b}} = \boldsymbol{0}. \tag{4.8}$$

These equations can be deduced from (3.29)-(3.30), as well, by neglecting the higher-order terms. The equilibrated stress vectors, satisfying (4.7)-(4.8) are obtained from (3.32)-(3.35):

$${}_0\hat{\boldsymbol{p}}^{\lambda} = {}_0\epsilon^{\lambda\beta}\,{}_0\boldsymbol{\psi}_{,\beta} - \frac{1}{2\sqrt{{}_0A}}\,{}_1\boldsymbol{f}^{\lambda}, \tag{4.9}$$

$${}_0\hat{\boldsymbol{p}}^{3} = {}_0\epsilon^{3\lambda}\,{}_0\boldsymbol{\chi}_{,\lambda}\,, \tag{4.10}$$

$${}_1\hat{\boldsymbol{p}}^{\lambda} = {}_0\epsilon^{\lambda\beta}\,{}_1\boldsymbol{\psi}_{,\beta} + \frac{8}{d^2}\,{}_0\epsilon^{3\lambda}\,{}_0\boldsymbol{\chi} - \frac{1}{2\sqrt{{}_0A}}\,{}_0\boldsymbol{f}^{\lambda}\,, \tag{4.11}$$

where ${}_0\boldsymbol{f}^{\lambda}$ and ${}_1\boldsymbol{f}^{\lambda}$ are defined by (3.36)-(3.37). From (4.9)-(4.11) it follows that the power series of the stress function vectors (3.11)-(3.12) are restricted now to linear and constant functions, according to

$$\boldsymbol{\psi}(\xi^{\alpha},\zeta) = {}_0\boldsymbol{\psi}(\xi^{\alpha}) + {}_1\boldsymbol{\psi}(\xi^{\alpha})\,\zeta, \tag{4.12}$$

$$\boldsymbol{\chi}\,(\xi^{\alpha},\zeta) = {}_0\boldsymbol{\chi}\,(\xi^{\alpha}), \tag{4.13}$$

i.e., only three first-order stress function vectors, ${}_0\boldsymbol{\psi}$, ${}_1\boldsymbol{\psi}$ and ${}_0\boldsymbol{\chi}$, are required in this model for thin shells.

The equilibrated first Piola-Kirchhoff stress components can be obtained from equations (3.46)-(3.51) of Section 3.4 by taking into account the stress function space (4.12) and (4.13):

$$
\begin{aligned}
{}_0\hat{P}^{k\lambda} &= {}_0\epsilon^{\lambda\beta}\,{}_0\psi^{k}_{\ |\beta} - \frac{1}{2\sqrt{{}_0A}}\,{}_1(f^k)^{\lambda}, && (4.14)\\
{}_1\hat{P}^{k\lambda} &= {}_0\epsilon^{\lambda\beta}\,{}_1\psi^{k}_{\ |\beta} + \frac{8}{d^2}\,{}_0\epsilon^{3\lambda}\,{}_0\chi^{k} - \frac{1}{2\sqrt{{}_0A}}\,{}_0(f^k)^{\lambda}, && (4.15)\\
{}_0\hat{P}^{k3} &= {}_0\epsilon^{3\lambda}\,{}_0\chi^{k}_{\ |\lambda}, && (4.16)\\
{}_1\hat{P}^{k3} &= {}_1\tilde{p}^{\,k}, && (4.17)\\
{}_2\hat{P}^{k3} &= \frac{4}{d^2}\,{}_0\tilde{p}^{\,k} - \frac{4}{d^2}\,{}_0\hat{P}^{k3}, && (4.18)
\end{aligned}
$$

where ${}_i(f^k)^{\lambda}$, $i = 1, 2$ are the components of the load vectors (3.36)-(3.37) with respect to the base vectors ${}_0\boldsymbol{A}_k$. According to (4.14)-(4.18), the components $\hat{P}^{k\ell}$ of the first Piola-Kirchhoff stress tensor are approximated by linear and quadratic polynomials across the thickness:

$$
\begin{aligned}
\hat{P}^{k\lambda}(\xi^\alpha,\zeta) &= {}_0\hat{P}^{k\lambda}(\xi^\alpha) + {}_1\hat{P}^{k\lambda}(\xi^\alpha)\,\zeta, && (4.19)\\
\hat{P}^{k3}(\xi^\alpha,\zeta) &= {}_0\hat{P}^{k3}(\xi^\alpha) + {}_1\hat{P}^{k3}(\xi^\alpha)\,\zeta + {}_2\hat{P}^{k3}(\xi^\alpha)\,\zeta^2. && (4.20)
\end{aligned}
$$

This approximation is consistent with the stress state obtained from the classical shell models and equivalent with the stresses obtained from the three-dimensional equilibrium equations through *a posteriori* integration across the thickness of the shell.

The shifted components $\bar{P}^{k\ell}$ of the first Piola-Kirchhoff stress tensor can be derived using relation (3.54). When the approximations (2.22) for the shifter tensor and its determinant are applied, the equality $\bar{P}^{k\ell} = \hat{P}^{k\ell}$ is valid for the shifted stress components.

4.2. **Relationship with resultant-based shell equations.** In this section, first-order stress function vectors generating equilibrated stress resultants and stress couples used in classical shell theories are introduced and their relationships to the stress function vectors (4.12)-(4.13) that generate equilibrated stress vectors are investigated.

4.2.1. *Equilibrated stress resultants and stress couples.* The vectorial equilibrium equations for the first Piola-Kirchhoff stress resultants and stress couples can be derived from the three-dimensional equilibrium equations (3.7), rewritten in the form

$$
\frac{1}{\sqrt{{}_0A}}\,(\sqrt{{}_0A}\,\mu\boldsymbol{p}^{\lambda})_{,\lambda} + (\mu\boldsymbol{p}^{3})_{,3} + {}^0\rho\,\mu\boldsymbol{b} = \boldsymbol{0}. \qquad (4.21)
$$

Introducing the infinite number of first Piola-Kirchhoff stress resultant vectors as

$$ {}_i\boldsymbol{N}(\xi^\alpha) = \int_{(d)} \zeta^i \mu \boldsymbol{p}^\ell \mathrm{d}\zeta, \qquad i = 0, 1, 2, 3, \dots , \tag{4.22} $$

the 'classical' stress resultants and stress couples are obtained for $i = 0$ and $i = 1$:

$$ \boldsymbol{N}^\ell(\xi^\alpha) := {}_0\boldsymbol{N}^\ell = \int_{(d)} \mu \boldsymbol{p}^\ell \mathrm{d}\zeta, \tag{4.23} $$

$$ \boldsymbol{M}^\lambda(\xi^\alpha) := {}_1\boldsymbol{N}^\lambda = \int_{(d)} \zeta\, \mu \boldsymbol{p}^\lambda \, \mathrm{d}\zeta. \tag{4.24} $$

Note that ${}_1\boldsymbol{N}^3 = \boldsymbol{M}^3 = 0$.

Integrating equation (4.21) with respect to the thickness coordinate ζ and taking into account the stress boundary conditions on the faces, the vectorial equilibrium equation

$$ \frac{1}{\sqrt{{}_0A}} (\sqrt{{}_0A}\, \boldsymbol{N}^\lambda)_{,\lambda} + \mathfrak{p} = \mathbf{0} \tag{4.25} $$

is obtained for the first Piola-Kirchhoff stress resultants $\boldsymbol{N}^\lambda$, where the load vector $\mathfrak{p}$ is given by

$$ \mathfrak{p}(\xi^\alpha) = \mu^+ {}^0\widetilde{\boldsymbol{p}}^+ + \mu^- {}^0\widetilde{\boldsymbol{p}}^- + \int_{(d)} {}^0\rho\, \mu \boldsymbol{b} \, \mathrm{d}\zeta \, . \tag{4.26} $$

Multiplying equation (4.21) by ζ, integrating it then with respect to ζ and taking into account the stress boundary conditions on the faces again, the vectorial moment equilibrium equation is obtained:

$$ \frac{1}{\sqrt{{}_0A}} (\sqrt{{}_0A}\, \boldsymbol{M}^\lambda)_{,\lambda} - \boldsymbol{N}^3 + \mathfrak{m} = \mathbf{0}, \tag{4.27} $$

where the moment load vector $\mathfrak{m}$ is given by

$$ \mathfrak{m}(\xi^\alpha) = \frac{d}{2} (\mu^+ {}^0\widetilde{\boldsymbol{p}}^+ - \mu^- {}^0\widetilde{\boldsymbol{p}}^-) + \int_{(d)} \zeta\, {}^0\rho\, \mu \boldsymbol{b} \, \mathrm{d}\zeta \, . \tag{4.28} $$

Equilibrium equations (4.25) and (4.27) can identically be satisfied by introducing three arbitrary but differentiable first-order stress function vectors, denoted by $\boldsymbol{\mathcal{F}}(\xi^\alpha)$, $\boldsymbol{\mathcal{G}}(\xi^\alpha)$ and $\boldsymbol{\mathcal{H}}(\xi^\alpha)$, according to

$$ \boldsymbol{N}^\lambda = {}_0\epsilon^{\lambda\beta}\, \boldsymbol{\mathcal{F}}_{,\beta} - \frac{1}{2\sqrt{{}_0A}} \int_{(\xi^\lambda)} \sqrt{{}_0A}\, \mathfrak{p} \, \mathrm{d}\xi^\lambda \, , \tag{4.29} $$

$$ \boldsymbol{N}^3 = {}_0\epsilon^{3\lambda}\, \boldsymbol{\mathcal{G}}_{,\lambda} \, , \tag{4.30} $$

$$ \boldsymbol{M}^\lambda = {}_0\epsilon^{\lambda\beta}\, \boldsymbol{\mathcal{H}}_{,\beta} + {}_0\epsilon^{3\lambda}\, \boldsymbol{\mathcal{G}} - \frac{1}{2\sqrt{{}_0A}} \int_{(\xi^\lambda)} \sqrt{{}_0A}\, \mathfrak{m} \, \mathrm{d}\xi^\lambda. \tag{4.31} $$

Note that in the works [15] and [16], the stress function vector $\boldsymbol{\mathcal{F}}$ is already used to satisfy the membrane equilibrium equation (4.25), according to (4.29). However, the stress function vectors $\boldsymbol{\mathcal{G}}$ and $\boldsymbol{\mathcal{H}}$ are not introduced and used, for satisfying the bending equilibrium equation (4.27) in the works mentioned, as the bending moments were derived from the bending strains through constitutive equations.

4.2.2. *Relationships for the stress functions.* The relationships between the first-order stress function vectors ${}_0\boldsymbol{\psi}(\xi^\alpha)$, ${}_0\boldsymbol{\chi}(\xi^\alpha)$, ${}_1\boldsymbol{\psi}(\xi^\alpha)$, introduced in Section 4.1, and $\boldsymbol{\mathcal{F}}(\xi^\alpha)$, $\boldsymbol{\mathcal{G}}(\xi^\alpha)$, $\boldsymbol{\mathcal{H}}(\xi^\alpha)$ introduced above can be obtained as follows. Substituting (4.3)-(4.4) into (4.23)-(4.24), the relations between the stress vectors and the stress resultants are obtained:

$$ {}_0\hat{\boldsymbol{p}}^\lambda = \frac{\boldsymbol{N}^\lambda}{d}, \qquad {}_1\hat{\boldsymbol{p}}^\lambda = 12\,\frac{\boldsymbol{M}^\lambda}{d^3}, \tag{4.32} $$

$$ {}_0\hat{\boldsymbol{p}}^3 = \frac{3}{2}\frac{\boldsymbol{N}^3}{d} - \frac{1}{2}\,{}_0\widetilde{\boldsymbol{p}}, \qquad {}_1\hat{\boldsymbol{p}}^3 = {}_1\widetilde{\boldsymbol{p}}, \qquad {}_2\hat{\boldsymbol{p}}^3 = \frac{6}{d^2}\,({}_0\widetilde{\boldsymbol{p}} - \frac{\boldsymbol{N}^3}{d}). \tag{4.33} $$

Inserting (4.32) in (4.3) and (4.33) in (4.4), the relations

$$ \hat{\boldsymbol{p}}^\lambda = \frac{\boldsymbol{N}^\lambda}{d} + 12\,\frac{\boldsymbol{M}^\lambda}{d^3}\,\zeta\,, \tag{4.34} $$

$$ \hat{\boldsymbol{p}}^3 = \frac{3}{2}\frac{\boldsymbol{N}^3}{d}(1 - \frac{4}{d^2}\zeta^2) - {}_0\widetilde{\boldsymbol{p}}\,(\frac{1}{2} - \frac{6}{d^2}\zeta^2) + {}_1\widetilde{\boldsymbol{p}}\,\zeta \tag{4.35} $$

are obtained, which are partially known from the classical shell theories.

The relationships between the first-order stress function vectors $\boldsymbol{\mathcal{F}}$, $\boldsymbol{\mathcal{G}}$, $\boldsymbol{\mathcal{H}}$ and ${}_0\boldsymbol{\psi}$, ${}_0\boldsymbol{\chi}$, ${}_1\boldsymbol{\psi}$ are obtained by substituting (4.29)-(4.31) into (4.32)-(4.33) and comparing the result with (4.5)-(4.6) and (4.9)-(4.11):

$$ {}_0\boldsymbol{\psi} = \frac{\boldsymbol{\mathcal{F}}}{d}, \tag{4.36} $$

$$ {}_1\boldsymbol{\psi} = 12\,\frac{\boldsymbol{\mathcal{H}}}{d^3}, \tag{4.37} $$

$$ {}_0\epsilon^{3\lambda}({}_0\boldsymbol{\chi} - \frac{3}{2}\,\frac{\boldsymbol{\mathcal{G}}}{d})_{,\lambda} + \frac{1}{2}\,{}_0\widetilde{\boldsymbol{p}} = \boldsymbol{0}. \tag{4.38} $$

On inserting the relations (4.36) and (4.37) in (4.12), it follows that

$$ \boldsymbol{\psi}(\xi^\alpha, \zeta) = \frac{\boldsymbol{\mathcal{F}}}{d} + 12\,\frac{\boldsymbol{\mathcal{H}}}{d^3}\,\zeta, \tag{4.39} $$

and, thus,

$$ \boldsymbol{\mathcal{F}}(\xi^\alpha) = \int_{(d)} \boldsymbol{\psi}(\xi^\alpha, \zeta)\,\mathrm{d}\zeta, \tag{4.40} $$

$$ \boldsymbol{\mathcal{H}}(\xi^\alpha) = \int_{(d)} \zeta\,\boldsymbol{\psi}(\xi^\alpha, \zeta)\,\mathrm{d}\zeta, \tag{4.41} $$

which indicate that $\boldsymbol{\mathcal{F}}$ and $\boldsymbol{\mathcal{H}}$ can be considered as two-dimensional stress function resultants obtained from the three-dimensional stress function vector $\boldsymbol{\psi}$.

Equation (4.38) is a partial differential equation for ${}_0\boldsymbol{\chi}$ and $\boldsymbol{\mathcal{G}}$. For the special case when the loads on the shell faces are zero, the relation

$$ {}_0\boldsymbol{\chi} = \frac{3}{2}\,\frac{\boldsymbol{\mathcal{G}}}{d} \tag{4.42} $$

holds for them.

5. Concluding remarks

The equilibrated stress space given in this paper for nonlinear deformation of shells can serve as a starting point for the derivation of a stress-based nonlinear shell theory and a dual-mixed, complementary energy-based finite element model for shells. The variational background of this theory and the related finite element model is Fraeijs de Veubeke's variational principle for 3D elasticity problems [31]. The dimensional reduction procedure for the present shell model relies on *a priori* assumptions regarding the variation of the stress space across the thickness, instead of the standard Kirchhoff-Love or Reissner-Mindlin type kinematical assumptions on the deformed geometry of the shell. Since the transverse shear and the transverse normal stresses are also present in the formulation, three-dimensional constitutive equations can directly be applied. Beside this fact, the main advantage of the stress-based models and the related dual-mixed finite elements for shells is their expected locking-free behavior, i.e., their superior performance in those cases, when the classical, displacement-based models and elements exhibit different types of numerical over-stiffening phenomenon.

References

1. B. M. Fraeijs de Veubeke. "Diffusive equilibrium models." Lecture notes for the International Research Seminar on *'The theory and application of finite element methods',* University of Calgary, Alberta, Canada (1973).
2. B. M. Fraeijs de Veubeke. "Stress function approach." *Proceedings of the World Congress on Finite Element Methods.* Bournemouth, U.K., 1975, J.1–J.51.
3. B. M. Fraeijs de Veubeke. "Discretization of rotational equilibrium in the finite element method." *Lecture Notes in Mathematics. Proceedings of the Mathematical Aspects of Finite Element Methods held in Rome, December 10–12, 1975.* Ed. by I. Galligiani and E. Magenes. Vol. 606. Berlin: Springer-Verlag, 1977, pp. 87–112. DOI: 10.1007/BFb0064451.
4. I. Kozák and G. Szeidl. "The field equations and the boundary conditions with force stresses and couple stresses in the linearized theory of micropolar elastostatics." *Acta Technica Academiae Scientiarum Hungaricae,* **91**(1-2), (1980), pp. 57–80.
5. E. Bertóti. "Indeterminacy of first-order stress functions and the stress- and rotation-based formulation of linear elasticity." *Computational Mechanics,* **14**, (1994), pp. 249–265. DOI: 10.1007/BF00370076.
6. L. E. Malvern. *Introduction to the Mechanics of a Continuous Medium.* Englewood Cliffs, New Jersey: Prentice-Hall, 1969.
7. B. M. Fraeijs de Veubeke and A. Millard. "Discretization of stress fields in the finite element method." *Journal of the Franklin Institute,* **302**, (1976), pp. 389–412. DOI: 10.1016/0016-0032(76)90032-6.
8. H. Murakawa and S. N. Atluri. "Finite elasticity solutions using hybrid finite elements based on a complementary energy principle." *Journal of Applied Mechanics,* **45**(3), (1978), pp. 539–547. DOI: 10.1115/1.3424358.

9. H. Murakawa and S. N. Atluri. "Finite elasticity solutions using hybrid finite elements based on a complementary energy principle. Part 2: Incompressible materials." *Journal of Applied Mechanics,* **46**(1), (1979), pp. 71–77. DOI: 10.1115/1.3424531.
10. E. Bertóti. "On the stress function approach in three-dimensional elasticity." *Acta Mechanica,* **190**(1-4), (2007), pp. 197–204. DOI: 10.1007/s00707-006-0432-6.
11. H. Schaefer. "Die Analogie zwischen den Verschiebungen und den Spannungsfunktionen in der Biegetheorie der Kreiszylinderschale." *Ingenieur-Archiv,* **29**, (1960), pp. 125–133.
12. W. Günther. "Analoge Systeme von Schalengleichungen." *Ingenieur-Archiv,* **30**, (1961), pp. 160–186. DOI: 10.1007/BF00534754.
13. A. E. Goldenveizer. *Theory of Elastic Thin Shells.* Oxford: Pergamon Press, 1961.
14. Y. Yamamoto. "An intrinsic theory of shells." *International Journal of Solids and Structures,* **2**, (1966), pp. 235–248. DOI: 10.1016/0020-7683(66)90017-5.
15. J. G. Simmonds and D. A. Danielson. "Nonlinear shell theory with finite rotation and stress function vectors." *Journal of Applied Mechanics,* **39**(4), (1972), pp. 1085–1090. DOI: 10.1115/1.3422833.
16. S. N. Atluri. "Alternate stress and conjugate strain measures, and mixed variational formulations involving rigid rotations, for computational analyses of finitely deformed solids, with application to plates and shells – I. Theory." *Computers and Structures,* **18**, (1984), pp. 93–116. DOI: 10.1016/0045-7949(84)90085-3.
17. E. F. Punch and S. N. Atluri. "Large displacement analysis of plates by a stress-based finite element approach." *Computers and Structures,* **24**(1), (1986), pp. 107–117. DOI: 10.1016/0045-7949(86)90339-1.
18. R. Valid. *The Nonlinear Theory of Shells through Variational Principles: From Elementary Algebra to Differential Geometry.* Chichester: John Wiley & Sons, 1995.
19. L. G. Kocsán. "Derivation of a dual-mixed *hp*-finite element model for axisymmetrically loaded cylindrical shells." *Archive of Applied Mechanics,* **81**(12), (2011), pp. 1953–1971. DOI: 10.1007/s00419-011-0530-3.
20. B. Tóth and L. G. Kocsán. "Comparison of dual-mixed *h*- and *p*-version finite element models for axisymmetric problems of cylindrical shells." *Finite Element in Analysis and Design,* **65**, (2013), pp. 50–62. DOI: 10.1016/j.finel.2012.11.002.
21. E. Reissner. "The effect of transverse shear deformations on the bending of elastic plates." *Journal of Applied Mechanics,* **12**, (1945), A69–A77.
22. R. D. Mindlin. "Influence of rotatory inertia and shear on flexural motions of isotropic elastic plates." *Journal of Applied Mechanics,* **18**, (1951), pp. 31–38.
23. P. M. Naghdi. Foundations of Elastic Shell Theory. In *Progress in Solid Mechanics, Vol. IV* (I. N. Sneddon and R. Hill, Eds.) pp. 1–90. Amsterdam: North-Holland, 1963.

24. P. M. Naghdi. The Theory of Shells and Plates. In *Handbuch der Physik, Band VIa/2* (S. Flügge and C. Truesdell, Eds.) Berlin: Springer-Verlag, 1972, pp. 425–640. DOI: 10.1007/978-3-662-39776-3_5.
25. G. Kirchhoff. "Über das Gleichgewicht und die Bewegung einer elastischen Scheibe." *Journal für die reine und angewandte Mathematik,* **40**, (1850), pp. 51–58.
26. A. E. H. Love. "The small free vibrations and deformations of a thin elastic shell." *Philosophical Transactions of the Royal Society of London Series A,* **179**, (1888), pp. 491–546.
27. V. V. Novozhilov. *Thin Shell Theory.* (2nd ed., Wolters-Noordhoff, Groningen, 1970). Groningen: Noordhoff, 1959.
28. J. L. Sanders. "Non-linear theories for thin shells." *Quarterly of Applied Mathematics,* **21**, (1963), pp. 21–36.
29. W. T. Koiter. "On the foundations of the linear theory of thin elastic shells. I-II." *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen,* **B73**, (1970), pp. 169–195.
30. C. Truesdell and R. A. Toupin. The Classical Field Theories. In *Handbuch der Physik, Band III/1* (S. Flügge, Ed.) Berlin: Springer-Verlag, 1960. DOI: 10.1007/978-3-642-45943-6_2.
31. B. M. Fraeijs de Veubeke. "A new variational principle for finite elastic displacements." *International Journal for Engineering Sciences,* **10**, (1972), pp. 745–763. DOI: 10.1016/0020-7225(72)90079-1.

Journal of Computational and Applied Mechanics, Vol. 15, No. 2, (2020), pp. 101–114
DOI: 10.32973/jcam.2020.007

DAMAGE ASSESSMENT OF THE HISTORICAL SZÉCHENYI CHAIN BRIDGE

László Dunai[a] and Balázs Kövesdi[a]
[a]Department of Structural Engineering, Budapest University of Technology and Economics
H-1111 Budapest, Műegyetem rkp. 3, Hungary
dunai.laszlo@epito.bme.hu kovesdi.balazs@epito.bme.hu

[Received: June 18, 2020; Accepted: September 14, 2020]

Dedicated to Professor Barna Szabó on the occasion of his 85th birthday

Abstract. The Széchenyi Chain Bridge is a 170-year-old historical structure located in the downtown of Budapest. The superstructure of the bridge was reconstructed several times in its history and currently the renewal process of the bridge is under consideration. According to the current plans main girders, chain elements and cross-girders will remain the old structure and the deck system will be replaced by a new orthotropic steel deck. The Budapest University of Technology and Economics, Department of Structural Engineering was involved in the design process and in the assessment of the remaining elements' condition within the last 5 years. During the project authors were faced with numerous specific important and challenging structural problems, modelling specialties, advanced design methods and research interest. The main part of these unusual characteristics come from the layout of the historical structure, long time traffic and corrosion problems. One of the most important questions during the structural analysis is the condition and rotational capacity of the pins between the chain elements. The chain system is more than 100 years old and the rotational capacity of the pins is questionable due to corrosion and friction. This phenomenon significantly influences the static behaviour of the chain elements and the whole suspending system. The current paper presents the numerical and on-site experimental program on the investigation of the rotational capacity of the pins. A second important question was related to the condition of current deck system. Significant corrosion damage was observed on the steel stringers which might cause damage or local collapse of the bridge deck under public transportation loads. Advanced numerical model using probabilistic analysis (FORM) and measurement based corrosion models are applied to make a risk assessment of the deck system's capability to maintain and keep the current traffic on the bridge before the deck will be replaced. Via this bridge inspection and investigation project the authors would like to demonstrate the application of advanced numerical modelling based design techniques and the industrial application of research models for lifetime assessment and risk analysis of historical structures.

Keywords: Chain bridge, historical structure, probabilistic design, advanced numerical modeling

1. Introduction

The Széchenyi Chain Bridge (Figure 1) is an almost 170-year-old historical structure of Budapest, a major attraction of the downtown over the Danube River. Although the bridge is a historical monument of Hungary and popular place for tourists, significant daily roadway traffic crosses the bridge and it is also used by the Budapest public transportation. The superstructure of the bridge was reconstructed several times in the history, but the chain elements have reached their 100-year-lifetime and the deck system is more than 70-year-old. The renewal process of the bridge is currently under organization. Főmterv Co. and MSc Ltd. made the design of the reconstruction work. The independent static check and the expert reports concerning the renewal process and damage assessment of the remaining structural elements have been completed by the Budapest University of Technology and Economics, Department of Structural Engineering in cooperation with the designers.

Figure 1. The Széchenyi Chain Bridge [1]

The strategy of the currently planned reconstruction is that the old bridge deck system (concrete slab and longitudinal steel stringers) will be replaced by a new orthotropic steel deck system (Figure 2). In longitudinal direction the deck system replacement will be made step-by-step, minimizing the geometry change of the suspension system during construction. The chain elements, the steel stiffening girder and the cross-girder system will remain unchanged, only the corrosion protection is planned to be renewed. Therefore, the evaluation of the structural condition of the remaining elements is an important task of the renewal process to determine their load carrying capacity and remaining lifetime.

The current paper introduces how an advanced numerical model can be used for the evaluation of the load carrying capacity of a historical structure in combination with on-site measurements and probabilistic design approach. The assessment of two structural elements are presented: (i) corroded chain elements whose pins cannot rotate due to friction and/or corrosion, (ii) old deck system with significant corrosion

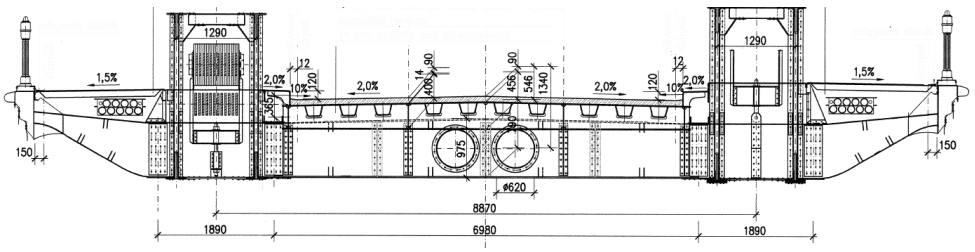


Figure 2. Planned cross-section after reconstruction [2]

damage whose risk of failure is evaluated for a short time period till reconstruction is started.

2. History of the Széchenyi Chain Bridge

The original Széchenyi Chain Bridge was erected between 1839-49 and it was considered as one of the greatest engineering works of its age (Figure 3). With a maximum middle span of 202.6 m this bridge was the largest chain bridge at that time; currently it is still the third one after the Herzilio Luz Bridge (339 m) in Brazil and Clifton Bridge (214 m) in England. The design was carried out by the English engineer William Tierney Clark while the construction was supervised by the Scottish engineer Adam Clark. The bridge was opened for traffic in 1849 and operated until 1914. Because the bridge had no stiffening girder and it had a light-weight timber deck system in this first time period, significant horizontal and vertical vibrations were observed on the bridge deck, which made the redesign of the construction necessary. The new bridge was built based on the static calculations of the Hungarian professor Antal Kherndl according to the plans of István Gállik and József Beke.

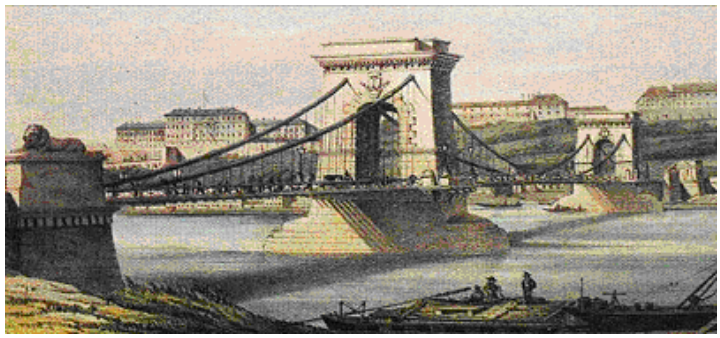


Figure 3. Planned cross-section after reconstruction [1]

The new supporting structure consisted of 25 carbon steel chain bars between each node whose length and bearing capacity was doubled compared to the previous structure (the distance between the suspension bars and the pins increased from 1.8

m to 3.6 m). With the new truss stiffening girder, the mass of the entire built-in ironwork grown to 5200 t using carbon steel with tensile strength of 48-56 kN/cm^2 (fits to the current S355 steel grade). During World War II.the bridge was demolished (in 1945) and the structure was rebuilt in its original form without any significant static changes. Two third of the chain elements were re-used, and the entire stiffening girder and deck system were replaced by new structural elements. The bridge was opened for traffic in its renewed form in 1949 [1]. In the meantime, no reconstruction work has been made on the system. Several investigations and measurements are made on the bridge. A loading test is performed to check the actual forces in the suspending bars; corrosion measurement were made on the chain elements in 2002; these results help the current renewal process and design.

3. Numerical modeling and design aspects

The structural behavior of the bridge is studied using a numerical model developed in the general finite element program Ansys [3]. The new orthotropic deck system and the old cross-girders are modeled using higher order (8-node) thin shell elements, while the stiffening truss girder and the suspension system (hangers and chain elements) are modeled by beam elements. The general layout of the global model is shown in Figure 4; the global model consists of 193 950 finite elements with a total number of 1 163 732 degrees of freedom. The erection phases of the bridge are simulated by birth and death process. The ultimate load calculation and risk analysis of the two analyzed structural members (chain elements and stringers of the deck system) are made using local numerical sub-models following the bridge characteristics (corrosion grade, lifetime, friction between pins and chain).

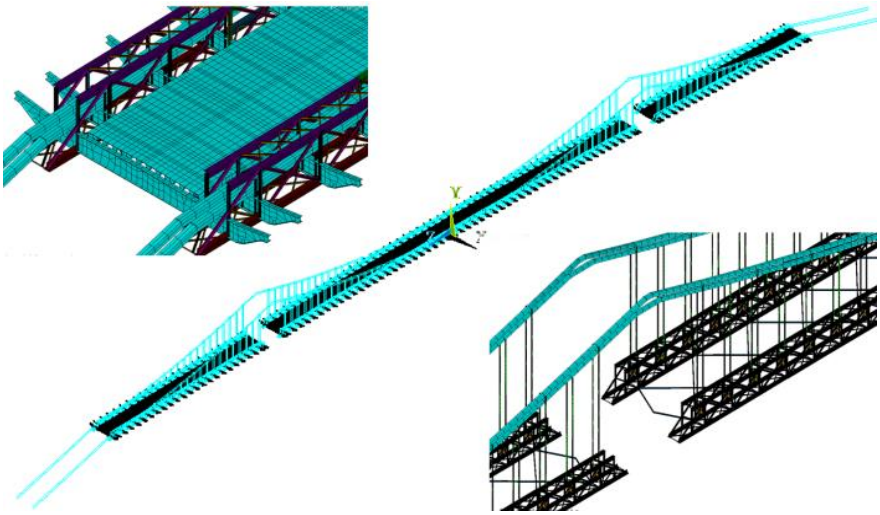


Figure 4. Global finite element model – combined shell and beam element model

In the numerical model the following structural specialties are considered:

- original erection phases having a significant impact on the internal force distribution between the different structural members (stiffening truss girders work only for traffic loads, self-weight is carried by the chain elements alone),
- second order effects (internal force distribution between the suspension system and the deck system depends on the pretension force acting in the chains),
- rotational capacity of the pins between the chain elements (chain is subjected to dominant tension force along the entire length even if pinned joints are fixed, but significant bending moments can act at the location of the pylon and at the abutment where chain system slope significantly changes), pins can get stuck due to corrosion and/or friction resulting in bending moment in the chain elements,
- corrosion of the chains, which reduces the net cross-section of the tension system, leading to reduced tensile strength and introducing stress concentration, resulting in reduced fatigue lifetime,
- corrosion of the longitudinal steel stringers and reinforcing bars of the concrete slab, which might influence the load carrying capacity of the deck system,
- uncertainties in the corrosion grade of the bridge and changes in time increasing the risk of damage under daily traffic.

On the global numerical model geometric nonlinear analysis (GNA) is carried out in each analyzed load case combination to determine the second order internal forces and stresses. These internal forces and direct loads are applied on the local submodels where the ultimate load of the analyzed structural details are determined by geometric and material nonlinear analysis using imperfections (GMNIA) according to EN 1993-1-5 [4]. The design resistance of the analyzed structural members is determined by first order reliability method (FORM) using Monte-Carlo simulation technique. This design approach can consider the uncertainties of the input data based on statistical analysis and able to perform risk assessment of the damaged structural items. Results of the FORM analysis are (i) the reliability index (β) and (ii) reliability of the structure calculated by $P_s = (1 - P_f)$, where P_f is the risk of failure against the investigated failure mode and design lifetime. If the calculated risk of failure is larger than the prescribed value (P_0), the structure does not fulfill the design criteria of the applied standard. The reliability index (β) is in correlation with the risk of the failure, the relationship is given in Table 1 according to the design philosophy of Eurocode EN 1990 [5].

Table 1. Correlation between reliability index (β) and risk of failure (P_f), [5]

P_f	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
β	1.28	2.32	3.09	3.72	4.27	4.75	5.20

One of the most important uncertainties in case of the Széchenyi Chain Bridge is the corrosion grade. In the international literature there are numerous deterministic and stochastic corrosion models available [6–8], which have been developed for

different steel grades and environmental conditions. Authors considered corrosion using a time-dependent statistical model taken from the international literature using real measurement data as input parameters. In the current numerical model, the corrosion model developed by Paik and Kim [7] in 2012 is applied, which was originally developed for a marine environment and extended for steel bridges by Tohodi and Sharifi [6] in 2016. In this corrosion model the damage grade is considered by the Weibull-type density function given by equation (3.1):

$$d_c = \frac{\alpha}{\beta} \cdot \left(\frac{Y_e}{\beta}\right)^{\alpha-1} \cdot \exp\left[-\left(\frac{Y_e}{\beta}\right)^{\alpha}\right], \tag{3.1}$$

where: d_c is the corrosion depth, α is the shape factor, β is size factor, $Y_e = Y - Y_c$; Y is the age of the bridge, Y_c is the age of the corrosion protection. Further details and the considered input parameters are given in [6]. This model can consider the change in the effectiveness of corrosion protection within the lifetime of the structure. On the other hand, it can also consider how the likelihood of the corrosion depth changes over time depending on the corrosion grade. In the applied numerical model of the chain elements, on-site measurement data from 2002 are implemented as initial input data of the corrosion model. Corroded surface is fitted on a large number of corrosion measurement data which is implemented in a solid finite element model.

4. On-site measurements on the chain elements

To check the rotational capacity of the pins an on-site loading test is executed on the bridge. The aim of the test was to determine the change in the normal force and bending moment within the chain elements. From the measured bending moment, conclusion could be made on the rotation ability of the pins (thus rotation reduces or totally eliminates bending moment). Results of the numerical model show the typical moment diagram of the suspension system (Figure 5b) if all pins become stuck (fixed connections are assumed in the numerical model). Results prove a significant bending moment can develop only within the chain elements at the two sides of the pylons and at the abutments. Therefore, these are the places where stress changes were measured during the loading test using strain gauges. All the other joints behave quasi-pinned, even if they are modeled by fixed joints. The bridge was loaded by 12 trucks

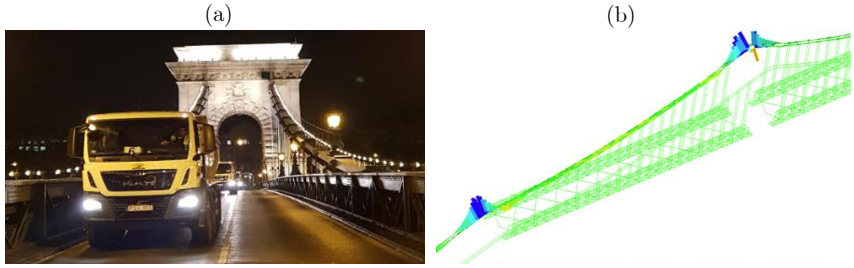


Figure 5. a) On-site loading test and b) calculated bending moment diagram with fixed pins

(Figure 5a) with an average weight of 200 kN, placed in 13 different loading arrangements simulating partial and total loading situations. The following measurements are executed within the loading test program:

- deflection measurements to check the global behavior of the system,
- strain measurements at 8 chain elements to determine the normal stress changes (normal force and bending moments separately),
- influence line measurements on the chain to check the global structural behavior due to moving loads,
- eigenfrequency measurements.

Deformation measurement results show the maximum measured deflection was slightly smaller than the numerically calculated maximum displacement, which predicted certain sticking effect and confirmed that the chain system has larger stiffness than predicted by the numerical model. Strain gauge measurements are also evaluated. Normal force and bending moment changes are determined in each investigated chain element, which are also compared to the numerical results. Results proved that from the 8 analyzed pins only one pin could rotate under the applied load. All other pins got stuck and could not rotate under the applied test load. One example for the measured normal stresses is presented in Figure 6. The diagram shows the strain measurement results while 3 trucks are moving on the bridge and take their planned position. Strain gauges H3/1 – H3/6 are placed on the outermost fiber of the first chain element next to the abutment and others are located on the second one, serving as reference measurements (dominant tension force was expected on the second chain elements). Results give evidence on the developed bending moment resulting from the stuck pins. Numbering of the strain gauges followed the same strategy at each measuring location. Strain gauges around the abutment are signed by H4 abutments are investigated during the on-site measurements marked by H1 – H4 and 10 strain gauges are placed on the chain elements at each measuring location, resulting in the code of the strain gauges (e.g. H3/1, etc.).

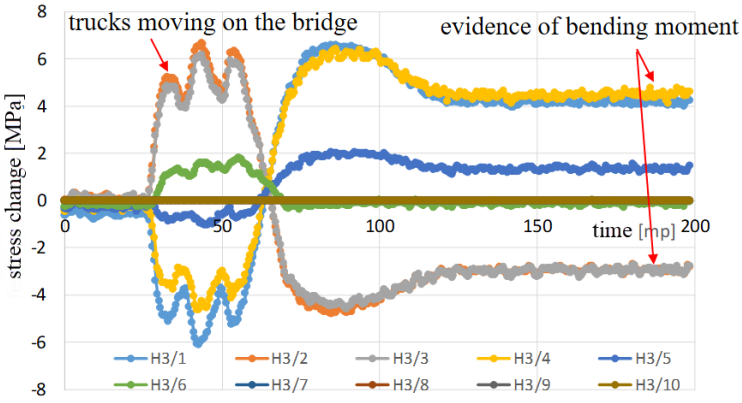


Figure 6. Stress changes in one chain element due to loading by 3 trucks

Strain measurement results show normal force level difference between the two sides of the pylon. This can only happen if roller supports cannot move on the top of the pylons, showing significant friction or sticking effect. This fact can also be seen on the measured influence lines presented in Figure 7.

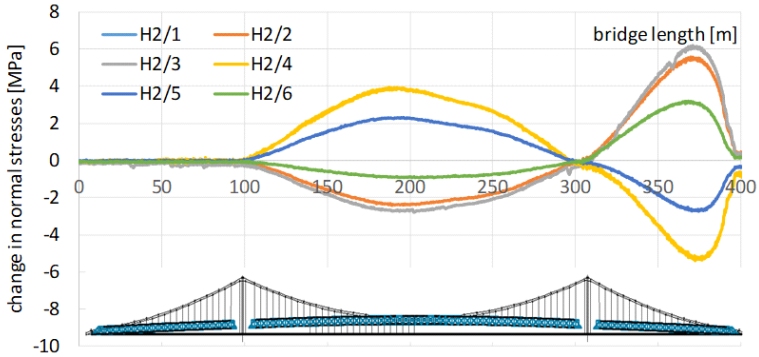


Figure 7. Influence line of one chain element

Measurement shows there are no measured strains for the left side span of the bridge, which can only happen if significant horizontal force is carried by the roller supports. On-site measurement results gave essential information to the designers and confirmed chain elements should be checked for bending and normal force interaction (N+M interaction).

5. Resistance calculation of corroded and stucked chain elements

It is well known that stuck pins change the structural behavior of the chain elements, which could lead to damage if they are not designed or statically checked. One of the most serious chain bridge failures occurred in the USA in 1967 and resulted in the death of 46 people [9]. The Silver Bridge collapsed due to a single chain element failure initiated from a fatigue crack, resulting in the total collapse of the chain system. The reason for the failure was the sticking of the pins due to corrosion. Therefore, it is a crucial point to check the resistance and risk of failure of the chain for combined

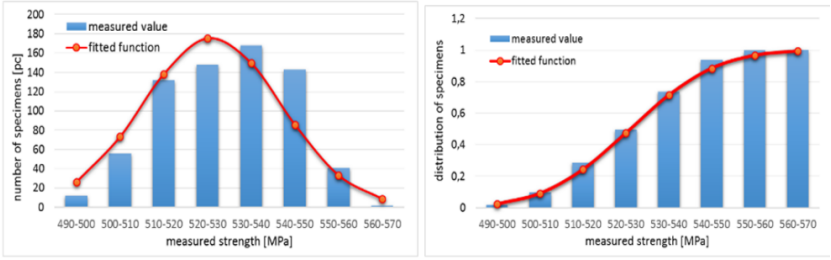


Figure 8. Material properties of the chain elements based on coupon tests

normal force and bending moment (N+M interaction) considering corrosion and stuck pins. The corrosion grade, old material properties and other design parameters have large scatter, which can only be considered in probability based design approaches. The main part (76%) of the chain elements was manufactured in 1915 [1] with the characteristic of the contemporary steel manufacturing techniques. Results of 702 coupon tests taken in 1912 were found by the authors in the literature. So, the mean and characteristic values of the steel material could be determined by statistical evaluation and probabilistic curve fitting, as shown in Figure 8.

Advanced solid element based numerical model is developed for the analysis of the chain elements. Contact elements are applied between the connected surfaces considering friction and sticking effect. The developed numerical models have the following capabilities:

- elasto-plastic material behaviour using damage criteria to model tensile fracture of the material,
- friction and sticking between the pin and chain elements, - considering the measured corrosion grade based on the measured surface properties and its probability on the surface of the chain elements,
- ultimate load is determined using geometrical and material nonlinear analysis using imperfections (GMNI analysis),
- coupling GMNI analysis (numerical simulations) with the probabilistic design approach.

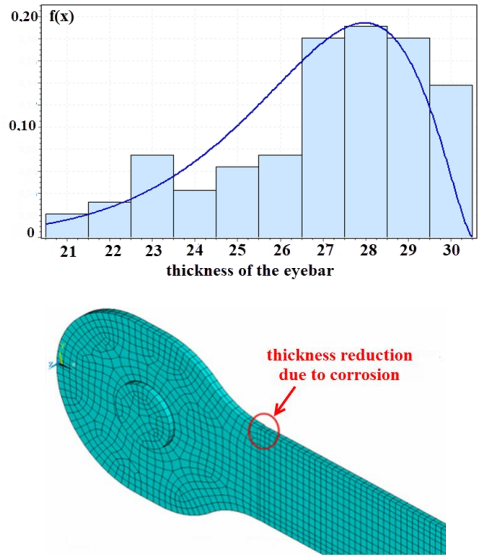


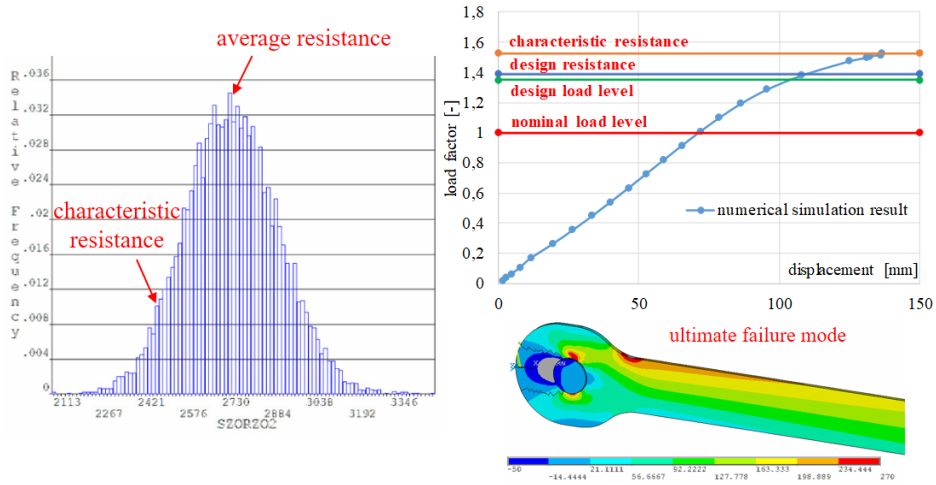
Figure 9. Model of the corroded surface and its probabilistic distribution

Considering corrosion of the chain elements, measured mean surface is implemented into the numerical model and distribution function is determined and graded by the

Kolmogorov-Szmirnov test. Beta distribution function was found to be accurate in most cases to model the corrosion uncertainties, therefore this probabilistic function was implemented in the numerical model. The mean corrosion surface and the applied beta distribution function are presented in Figure 9.

Probabilistic analysis is carried out on the numerical model using the Monte-Carlo simulation technique, generating pseudorandom values for the defined variable parameters using Latin hypercube sample selection. In each single case the ultimate resistance is determined using GMNI analysis by evaluation of the obtained load-displacement curve. During probabilistic analysis the following variables are considered:

- depth of the chain elements (using CoV =0.005),
- thickness of the chain elements (using CoV =0.05),
- yield strength of the steel (using CoV=0.03 - based on measurements),
- corrosion grade (using Beta distribution function - based on measurements).



a) distribution function of the characteristic resistances

b) typical load-displacement curve and obtained failure mode

Figure 10. Results of the combined probabilistic design method with the GMNI analyses

In the frame of the probabilistic analysis a total of 20 000 numerical analyses (using the surface fitting method) are executed within the Monte-Carlo simulation at a confidence level of 95%. One example for the numerical GMNI analysis result is presented in Figure 10, showing a load-displacement curve around the average resistance level from all the simulation results. In each calculation, the characteristic resistances are determined referring to different input values. Based on the obtained results the average, the characteristic and design values of the resistance are determined by statistical evaluation, as shown in Figure 10. Results of the probabilistic design showed that the average value of the resistance is 158% of the actual internal force. The 5% lower quantile value is equal to 148% representing the characteristic value. The 0.1%

quantile value is equal by 138% representing the design resistance level. This value is larger than the design value of the acting loads (135% of the nominal load level). So the corroded chain element satisfies the design criteria for N+M interaction with a maximum utilization ratio of 98%.

6. Risk assessment of the deck system

Another structural component where corrosion grade can lead to damage of the structure is the bridge deck system. The aim of the investigation was to determine the risk of failure of the deck considering on-site measurement data, real loading situation with public transportation using a probabilistic design approach. Locations and the extent of corrosion were taken into consideration in the numerical model based on on-site measurements carried out in 2019. Two photographs of the corroded longitudinal stringers are presented in Figure 11 showing the worst case locations.

expansion joint

lower flange of longitudinal stringer

Figure 11. Corroded structural details of the deck system

Based on the on-site corrosion measurements, the following average corrosion damage grades (or corrosion waste, i.e., reduction of thickness) are taken into account in the numerical model:

- upper and lower flanges of the outer longitudinal stringer: 50%,
- web of the outer longitudinal stringer: 30%,
- web, upper and lower flanges of the inner longitudinal stringer: 10%,
- upper and lower flanges of the outer longitudinal stringer near to the expansion joint: 50%,

– web of the outer longitudinal stringer near to the expansion joint: 80%,
– reinforcing bars at the bottom of the slab: 30%.

The given values are considered in the initial configuration for deterministic analysis, while in the stochastic analysis these are considered as mean values of the corresponding probability density function. Two standard cross-sections are analysed: i) longitudinal stringers near to the expansion joint and ii) internal longitudinal stringers in general location. Isometric and cross-sectional views of the sub-model are shown in Figure 12.

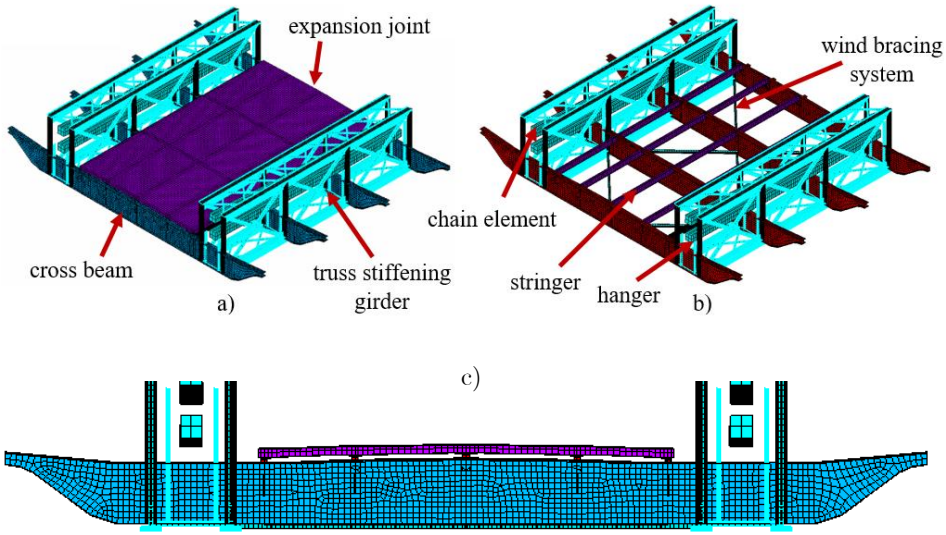


Figure 12. Isometric view of the model a) with and b) without reinforced concrete slab, c) cross-sectional view

Our check of the reinforced concrete slab in ultimate limit state is partially based on analytical calculations as height of the compression zone is pre-calculated in the analyzed cases, assuming that tensile strength of the concrete can be ignored. Thus, finite elements in the tensile zones having concrete material models are inactivated. An automatic algorithm is developed to determine the extension of cracking in all load cases and the numerical model is rebuilt considering crack locations and relevant material properties. On-site measurements show the largest damage risk is at the vicinity of expansion joints, where corrosion grade of the web reaches 80%, traffic load results in maximum shear force and dynamic effect is also at its maximum due to the expansion joint. At this location the failure mode obtained by the numerical calculations is shear buckling of the web, as shown in Figure 13. The numerical simulation based resistance is interpreted and calculated according to EN 1993-1-5 [4] recommendations.

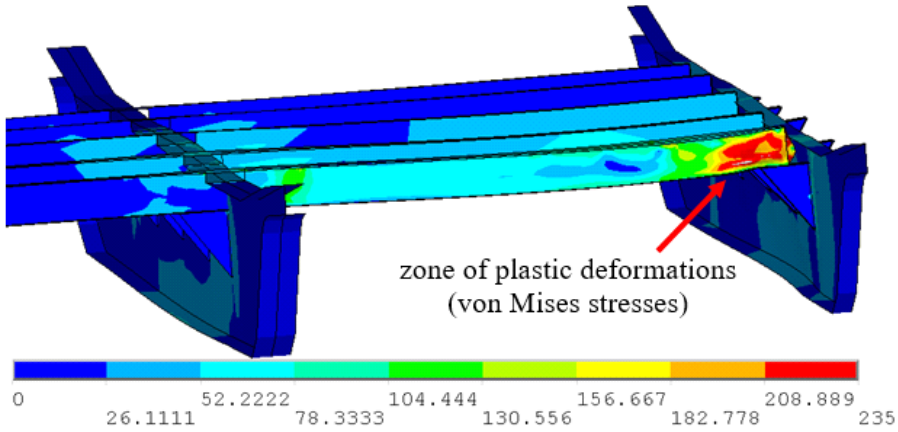


Figure 13. Failure mechanism and von Mises stresses [MPa]

Based on the deterministic calculation the characteristic value of the load factor is equal to $\alpha = 1.05$ (considering design load case combination). Deterministic calculations proved that the vicinity of the expansion joint has a high utilization ratio, therefore risk of failure is also determined. Stochastic analysis is carried out focusing on the worst case loading situation and corresponding failure mode (shear buckling of stringer web). The density function of the calculated ultimate load factors based on Monte-Carlo simulation is shown in Figure 14. The load factors are determined by GMNI analysis, which is combined with the probability approach.

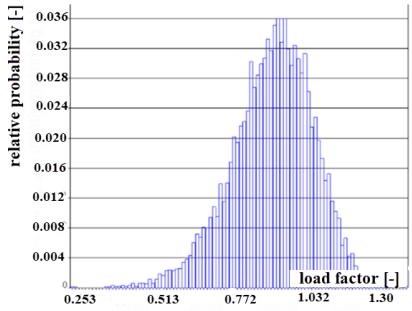


Figure 14. Density function of load factors based on Monte Carlo simulation

A relatively large coefficient of variation is obtained, which is a result of the corrosion model, with its high level of uncertainty associated with corrosion damage depth. The mean value of calculated ultimate load factors, denoting the failure of the structure, is 1.40, while the standard deviation is 0.206 (maximum axle load of buses in service on the bridge is 125 kN and considered dynamic amplification factor on the deck system is 1.40). From the statistical evaluation the calculated reliability index (β) is equal to 2.43, which corresponds to the failure probability of 0.007448 (1/134).

7. Summary

The present paper has the aim to demonstrate the application of advanced numerical modelling techniques and probabilistic design approaches in the civil engineering praxis. The Széchenyi Chain Bridge is one of the most important monuments and national symbols of Hungary, and its renewal process is currently under organization. The executed numerical calculations could help designers and decision makers in the evaluation of the actual condition, remaining lifetime, risk of damage of the corroded chain elements and deck system. In the applied numerical calculations a probabilistic design approach is combined with geometrical and material nonlinear analysis using imperfections. In the deterministic and stochastic analysis the results of the on-site measurements are considered regarding the corrosion grade, rotational capacity of the pins and material properties of the chain elements.

Acknowledgement. The authors would like to express special thanks to the Főmterv Zrt. design office for providing information related to the historical bridge, and for their cooperation concerning both the expert reports and the evaluation of the on-site measurements.

References

1. S. Domanovszky. *Száz éve, 1915. november 27-én helyezték forgalomba az átépített Lánchidat [The rebuilt Chane Bridge was put in circulation one hundred years ago on November 27, 1915].* TEXT Nyomdaipari Kft., 2015.
2. 10-I-002-01. *Static calculation of the superstructure of the Széchenyi chain bridge.* MSc Kft., Főmterv – MSc – Céh Konzorcium, tsz: 8300, 23.08.2016. Shop drawings.
3. ANSYS$^{\otimes}$ v15.0, Canonsburg, Pennsylvania, USA.
4. EN 1993-1-5. Eurocode 3: Design of steel structures – Part 1.5: Plated structural elements, 2006.
5. EN 1990. Eurocode – Bases of structural design, 2002.
6. S. Tohodi and Y. Sharifi. "Load-carrying capacity of locally corroded steel plate girder ends using artificial neural network." *Thin-Walled Structures,* **100**, (2016), pp. 48–61. DOI: 10.1016/j.tws.2015.12.007.
7. J. K. Paik and D. K. Kim. "Advanced method for the development of an empirical model to predict time-dependent corrosion wastage." *Corrosion Science,* **63**, (2012), pp. 51–58. DOI: 10.1016/j.corsci.2012.05.015.
8. S. Wang Y.and Xu, H. Wang, and A. Li. "Predicting the residual strength and deformability of corroded steel plate based on the corrosion morphology." *Construction and Building Materials,* **152**, (2017), p. 777.
9. B. Äkesson. *Understanding Bridge Collapses.* Taylor & Francis Group, 2008, pp. 139–148.

Journal of Computational and Applied Mechanics, Vol. 15, No. 2, (2020), pp. 115–133
DOI: 10.32973/jcam.2020.008

CONVERGENCE THEOREMS FOR THE NELDER-MEAD METHOD

Aurél Galántai
Óbuda University
1034 Budapest, Bécsi út 96/b, Hungary
galantai.aurel@nik.uni-obuda.hu

[Received: September 8, 2020, Accepted: October 14, 2020]

Dedicated to Professor Barna Szabó on the occasion of his 85th birthday

Abstract. We develop a matrix form of the Nelder-Mead method and after discussing the concept of convergence we prove a general convergence theorem. The new theorem is demonstrated in low dimensional spaces.

Mathematical Subject Classification: 65K10, 90C56
Keywords: Nelder-Mead simplex method, derivative-free algorithms, convergence

1. Introduction

We study the Nelder-Mead simplex method [1] for the solution of the unconstrained minimization problem

$$f(x) \to \min \quad (f : \mathbb{R}^n \to \mathbb{R}),$$

where f is continuous. Since 1965, the Nelder-Mead algorithm and its later variants have become highly popular in various application areas and derivative-free optimization [2–6]. Although the original paper [1] has a lot of citations (over 31 000 in Google Scholar on August 3, 2020), notably few theoretical results are known on the convergence (see Kelley [7] and Lagarias at al. [8, 9]).

In the paper we develop a matrix form of the Nelder-Mead method, discuss the concept of convergence and its consequences, prove a general convergence theorem under plausible assumptions and demonstrate the convergence of the algorithm for low dimensional spaces. This approach partially answers some of the questions raised by Wright [10] concerning the Nelder-Mead method.

Section 2 contains the description of the algorithm. The next section summarizes the most important earlier results on the convergence. The matrix reformulation of the Nelder-Mead method is given Section 4. The concept of convergence and some of its consequences are developed and discussed in Section 5. The spectra of the occurring matrices is investigated in Section 6. The general convergence theorem is developed in Section 7. Finally, the convergence of the Nelder-Mead method is demonstrated for $n = 1, 2, 3$ in the last section.

2. The Nelder-Mead simplex method

We use the following form of the original method [8]. The vertices of the initial simplex are denoted by $x_1, x_2, \ldots, x_{n+1} \in \mathbb{R}^n$. It is assumed that vertices $x_1, \ldots, x_{n+1}$ are ordered such that

$$f(x_1) \le f(x_2) \le \cdots \le f(x_{n+1}) \tag{2.1}$$

and this condition is maintained during the iterations of the Nelder-Mead algorithm. Define $x_c = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $x(\lambda) = (1+\lambda)\, x_c - \lambda x_{n+1}$. The related evaluation points are

$$x_r = x(1), \quad x_e = x(2), \quad x_{oc} = x\left(\frac{1}{2}\right), \quad x_{ic} = x\left(-\frac{1}{2}\right).$$

Then one iteration step of the method is the following.

Operation	Nelder-Mead simplex method
0. Ordering	$f(x_1) \le \cdots \le f(x_{n+1})$
1. Reflect	if $f(x_1) \le f(x_r) < f(x_n)$, then $x_{n+1} \leftarrow x_r$ and goto 0.
2. Expand	if $f(x_r) < f(x_1)$ and $f(x_e) < f(x_r)$, then $x_{n+1} \leftarrow x_e$ and goto 0. If $f(x_e) \ge f(x_r)$, then $x_{n+1} \leftarrow x_r$ and goto 0.
3. Contract outside	If $f(x_n) \le f(x_r) < f(x_{n+1})$ and $f(x_{oc}) \le f(x_r)$, then $x_{n+1} \leftarrow x_{oc}$ and goto 0.
4. Contract inside	If $f(x_r) \ge f(x_{n+1})$ and $f(x_{ic}) < f(x_{n+1})$ then $x_{n+1} \leftarrow x_{ic}$ and goto 0.
5. Shrink	$x_i \leftarrow (x_i + x_1)/2$, $f(x_i)$ (for all i) and goto 0.

There are two rules that apply to reindexing after each iteration. If a nonshrink step occurs, then x_{n+1} is discarded and a new point $v \in \{x_r, x_e, x_{oc}, x_{ic}\}$ is accepted. The following cases are possible:

$$f(v) < f(x_1), \quad f(x_1) \le f(v) \le f(x_n), \quad f(v) < f(x_{n+1}).$$

If

$$j = \begin{cases} 1, & \text{if } f(v) < f(x_1) \\ \max_{2 \le \ell \le n+1} \{ f(x_{\ell-1}) \le f(v) \le f(x_\ell) \}, & \text{otherwise} \end{cases}.$$

then the new simplex vertices are

$$x_i^{new} = x_i \ \ (1 \le i \le j-1), \ x_j^{new} = v, \ x_i^{new} = x_{i-1} \ \ (i = j+1, \ldots, n+1). \tag{2.2}$$

This rule inserts v into the ordering with the highest possible index. If shrinking occurs, then

$$x_1' = x_1, \quad x_i' = (x_i + x_1)/2 \quad (i = 2, \ldots, n+1)$$

plus a reordering takes place. By convention, if $f(x_1') \le f(x_i')$ $(i = 2, \ldots, n)$, then $x_1^{new} = x_1$.

Lagarias at al. [9] also investigated a restricted version, where expansion steps are not allowed.

We adopt the following notations. The simplex of iteration k is denoted by $S^{(k)} = \left[x_1^{(k)}, x_2^{(k)}, \ldots, x_{n+1}^{(k)}\right]$ with vertices that satisfy the condition

$$f\left(x_1^{(k)}\right) \leq f\left(x_2^{(k)}\right) \leq \cdots \leq f\left(x_{n+1}^{(k)}\right) \quad (k \geq 0)\,.$$

The initial simplex is $S^{(0)}$. The reflection, expansion and contraction points of simplex $S^{(k)}$ are denoted by $x_r^{(k)}$, $x_e^{(k)}$, $x_{oc}^{(k)}$ and $x_{ic}^{(k)}$, respectively. The function values at the vertices $x_j^{(k)}$ and the points $x_r^{(k)}$, $x_e^{(k)}$, $x_{oc}^{(k)}$ and $x_{ic}^{(k)}$ are denoted by $f\left(x_j^{(k)}\right) = f_j^{(k)}$ $(j = 1, \ldots, n+1)$, $f_r^{(k)} = f\left(x_r^{(k)}\right)$, $f_e^{(k)} = f\left(x_e^{(k)}\right)$, $f_{oc}^{(k)} = f\left(x_{oc}^{(k)}\right)$ and $f_{ic}^{(k)} = f\left(x_{ic}^{(k)}\right)$, respectively.

3. A review of the earlier convergence results

In 1998 McKinnon [11] constructed a function f, which is strictly convex and has continuous first derivatives for certain parameters. He showed that for this f, the Nelder-Mead simplex algorithm may fail to converge.

Lagarias et al. [8] proved several convergence properties of the simplex method for one and two-variable strictly convex functions by giving a deep insight look of the method. They summarize their main results as follows (see p. 114 of [8]):

1. *In dimension* 1*, the Nelder-Mead method converges to a minimizer, and convergence is eventually* M*-step linear.*
2. *In dimension* 2*, the function values at all simplex vertices in the standard Nelder-Mead algorithm converge to the same value.*
3. *In dimension* 2*, the simplices in the standard Nelder-Mead algorithm have diameters converging to zero.*

In 1999 Kelley [7, 12] developed a sufficient decrease condition for the average of the object function values (evaluated at the vertices) and proved that if this condition is satisfied during the process, then any accumulation point of the simplices is a critical point of f. For similar results on other variants of the Nelder-Mead algorithm, see Tseng [13], Nazareth and Tseng [14], Pryce at all. [15].

For the restricted Nelder-Mead method, Lagarias at al. [9] significantly improved the results of [8]. Let $\mathcal{F}$ denote the class of twice-continuously differentiable functions $\mathbb{R}^2 \to \mathbb{R}$ with bounded level sets and everywhere positive definite Hessian. Lagarias at al. proved that if the restricted Nelder-Mead algorithm is applied to a function $f \in \mathcal{F}$, starting from any nondegenerate simplex, then the algorithm converges to the unique minimizer of f.

Wright [10] raised several open questions concerning the Nelder-Mead method such as

- Why is it sometimes so effective (compared to other direct search methods) in obtaining a rapid improvement in f?

- One failure mode is known (McKinnon [11]) – but are there other failure modes?
- Why, despite its apparent simplicity, should the Nelder-Mead method be difficult to analyze mathematically?

Our purpose is to analyze and prove the convergence of the method using a matrix formalism. This technique will also shed light on these questions, even if only, partially. The failure modes are the subject of paper [16].

4. The Nelder-Mead simplex method in matrix form

Assume that simplex $S^{(k)}=\left[x_1^{(k)},x_2^{(k)},\dots,x_{n+1}^{(k)}\right]$ is such that

$$f\left(x_1^{(k)}\right)\le f\left(x_2^{(k)}\right)\le\cdots\le f\left(x_{n+1}^{(k)}\right).$$

If the incoming vertex v is of the form

$$x(\alpha)=\frac{1+\alpha}{n}\sum_{i=1}^{n}x_i^{(k)}-\alpha x_{n+1}^{(k)}$$

for some $\alpha\in\left\{1,2,\frac{1}{2},-\frac{1}{2}\right\}$, we can define the transformation matrix

$$T(\alpha)=\begin{bmatrix} I_n & \frac{1+\alpha}{n}e \\ 0 & -\alpha \end{bmatrix}\quad\left(e=[1,1,\dots,1]^T\right).$$

Since $S^{(k)}T(\alpha)=\left[x_1^{(k)},\dots,x_n^{(k)},x(\alpha)\right]$, we have to reorder the matrix columns according to the insertion rule (2.2). Define the permutation matrix

$$P_j=[e_1,\dots,e_{j-1},e_{n+1},e_j,\dots,e_n]\in\mathbb{R}^{(n+1)\times(n+1)}\quad(j=1,\dots,n+1).$$

Then $S^{(k)}T(\alpha)P_j$ is the new simplex $S^{(k+1)}$. Particularly, we have the following cases.

1. If the reflection point $x_r^{(k)}$ is the new incoming vertex, then

$$S^{(k+1)}=S^{(k)}T(1)P_j\quad(j=2,\dots,n).$$

2a) If the expansion point $x_e^{(k)}$ is the new incoming vertex, then

$$S^{(k+1)}=S^{(k)}T(2)P_1.$$

2b) If the expansion point is the reflection point $x_r^{(r)}$, then

$$S^{(k+1)}=S^{(k)}T(1)P_1.$$

3) If the outside contraction point $x_{oc}^{(k)}$ is the new incoming vertex, then

$$S^{(k+1)}=S^{(k)}T\left(\frac{1}{2}\right)P_j\quad(j=1,\dots,n+1).$$

4) If the inside contraction point $x_{ic}^{(k)}$ is the new incoming vertex, then

$$S^{(k+1)}=S^{(k)}T\left(-\frac{1}{2}\right)P_j\quad(j=1,\dots,n+1).$$

5) In the case of shrinking, the new vertices before reordering are

$$x_i' = \left(x_1^{(k)} + x_i^{(k)}\right)/2 \quad (i = 1, 2, \ldots, n+1)\,.$$

Hence $\left[x_1', x_2', \ldots, x_{n+1}'\right] = S^{(k)} T_{shr}$, where

$$T_{shr} = \begin{bmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{2} \\ 0 & \frac{1}{2} & \cdots & 0 \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{2} \end{bmatrix} = \frac{1}{2} I + \frac{1}{2} e_1 e^T.$$

Since the new vertices are subject to the ordering

$$f\left(x_1^{(k+1)}\right) \leq \cdots \leq f\left(x_k^{(k+1)}\right) \leq \cdots \leq f\left(x_{n+1}^{(k+1)}\right),$$

the new simplex is defined by

$$S^{(k+1)} = S^{(k)} T_{shr} P,$$

where $P \in \mathcal{P}_{n+1}$, which is the set of all possible permutation matrices of order $n+1$.

There are only $n+1$ permutation matrices of the type P_j, while there are $(n+1)!$ possible permutation matrices of the type $\mathcal{P}_{n+1}$

Hence in any of the above cases the new simplex $S^{(k+1)}$ is given by

$$S^{(k+1)} = S^{(k)} T_k P^{(k)},$$

where T_k is either $T(\alpha)$ $\left(\alpha \in \left\{-\frac{1}{2}, \frac{1}{2}, 1, 2\right\}\right)$ and $P^{(k)} \in \{P_1, \ldots, P_{n+1}\}$ or $T_k = T_{shr}$ and $P^{(k)} \in \mathcal{P}_{n+1}$. The number of different $T_i P^{(i)}$ matrices is at most $3n+3+(n+1)!$ indicating an increasing complexity if n increases.

Observe that matrices $T(\alpha)$, T_{shr}, $T(\alpha) P$ and $T_{shr} P$, for any $P \in \mathcal{P}_{n+1}$, have the property that their column sums are 1. We exploit the following simple results.

Claim 1. *(i) If $A \in \mathbb{R}^{n \times n}$ is a matrix whose column sums are 1, then A has an eigenvalue $\lambda = 1$ and a corresponding left eigenvector $x = e^T$. (ii) If $A, B \in \mathbb{R}^{n \times n}$ are two matrices whose column sums are 1, then $C = AB$ also has this property. (iii) If $A \in \mathbb{R}^{n \times n}$ is a matrix whose column sums are 1, then $\|A\| \geq 1$ in any induced matrix norm.*

Proof. By definition $e^T A = [\sum_{i=1}^n a_{i1}, \ldots, \sum_{i=1}^n a_{in}] = 1 \cdot e^T$. This implies $e^T B = e^T$, $e^T AB = e^T B = e^T$. Since $\rho(A) \geq 1$ and $\|A\| \geq \rho(A)$, (iii) also follows. □

Particularly, $\|T(\alpha)\|_1 = \|T(\alpha) P\|_1 = |1+\alpha| + |\alpha|$ and $\|T_{shr}\|_1 = \|T_{shr} P\|_1 = 1$.

A matrix A is called left stochastic if $a_{ij} \geq 0$ for all i,j and the column sums are 1. A matrix is called stochastic if $a_{ij} \geq 0$ for all i,j and both the column sums and the row sums are 1.

Matrix $T(\alpha)$ is left stochastic for $-1 \leq \alpha \leq 0$. The shrinking transformation matrix $T_{shr} = \frac{1}{2} I_{n+1} + \frac{1}{2} e_1 e^T$ is a left stochastic matrix, $T_{shr}^k = \frac{1}{2^k} I_{n+1} + \frac{2^k - 1}{2^k} e_1 e^T$ and $T_{shr}^k \to e_1 e^T$.

For any α, β, $T(\alpha)T(\beta) = T(-\alpha\beta)$. Hence $T(\alpha)^{-1} = T\left(\frac{1}{\alpha}\right)$ $(\alpha \neq 0)$ and $T(\alpha)^k = T\left((-1)^{k+1}\alpha^k\right)$. If $|\alpha| < 1$, then $\lim_{k\to\infty} T(\alpha)^k = T(0)$. Matrix $T(1)$ is an involution $(T(1)^2 = I_{n+1})$. If $T(1)$ is multiplied by a permutation matrix P, this property may change. For $n = 2$, $T(1)P_2$ (reflection) is a 6-involutory matrix (for k-involutory matrices, see Trench [17]). $T(\alpha)^k$ is unbounded if $|\alpha| > 1$ and uniformly bounded if $|\alpha| \leq 1$.

5. The concept of convergence and consequences

Simplex $S^{(k)}$ is given by

$$S^{(k)} = S^{(0)}B_k, \tag{5.1}$$

where

$$B_k = \prod_{i=1}^{k} T_i P^{(i)} \tag{5.2}$$

and

$$T_i P^{(i)} \in \left\{ T(\alpha)P_k : \alpha \in \left\{ -\frac{1}{2}, \frac{1}{2}, 1, 2 \right\},\ k = 1, \ldots, n+1 \right\} \tag{5.3}$$

or

$$T_i P^{(i)} \in \{ T_{shr} P : P \in \mathcal{P}_{n+1} \}. \tag{5.4}$$

Note that each $T_i P^{(i)}$ is nonsingular and $e^T T_i P^{(i)} = e^T$.

A simplex S is nondegenerate if the matrix

$$M = [x_1 - x_{n+1}, x_2 - x_{n+1}, \ldots, x_n - x_{n+1}]$$

is nonsingular. Then S must be affinely independent, which is equivalent to (see, e.g., [18], [19]) that vectors

$$\begin{bmatrix} 1 \\ x_1 \end{bmatrix}, \ldots, \begin{bmatrix} 1 \\ x_{n+1} \end{bmatrix}$$

are linearly independent. Hence $\operatorname{rank}\left(\begin{bmatrix} e^T \\ S \end{bmatrix}\right) = n + 1$.[1] Assume that the initial simplex $S^{(0)}$ is nondegenerate. Since $e^T B_k = e^T$ and

$$\begin{bmatrix} e^T \\ S^{(k)} \end{bmatrix} = \begin{bmatrix} e^T \\ S^{(0)} \end{bmatrix} B_k$$

is nonsingular, $S^{(k)}$ is also nondegenerate.

For the convergence of the Nelder-Mead algorithm, it is natural to require that the simplex vertices $x_j^{(k)}$ $(j = 1, 2, \ldots, n+1)$ should converge to the same vector $\widehat{x}$ as $k \to \infty$. In such a case

$$\lim_{k\to\infty} S^{(k)} = [\widehat{x}, \ldots, \widehat{x}] = \widehat{x}e^T. \tag{5.5}$$

Claim 2. *If $\{B_k\}$ is bounded, then $\left\{S^{(k)}\right\}$ converge to some S^∞ if and only if $\{B_k\}$ converge to some B.*

[1] It is assumed through the whole paper that the sizes of e and the unit vectors e_i are compatible with the operation and/or partition.

Proof. If $S^{(k)} \to S^{\infty}$ (whatever S^{∞} is) and $B_k \to B$, then $S^{(0)}B_k \to S^{(0)}B = S^{\infty}$. Assume that $S^{(k)} \to S^{\infty}$ and $\{B_k\}$ has no limit point. Since $\{B_k\}$ is bounded it must have at least one accumulation point, say B^* and there is a subsequence $\left\{B_{i_j}\right\} \subset \{B_k\}$ such that $B_{i_j} \to B^*$ and $S^{(i_j)} \to S^{(0)}B^* = S^{\infty}$. Assume that there exists a second accumulation point $B^{**} \neq B^*$ and a subsequence $\left\{B_{k_j}\right\} \subset \{B_k\}$ such that $B_{k_j} \to B^{**}$. It follows that

$$\begin{bmatrix} e^T \\ S^{(i_j)} \end{bmatrix} \to \begin{bmatrix} e^T \\ S^{(0)} \end{bmatrix} B^* = \begin{bmatrix} e^T \\ S^{(0)} \end{bmatrix} B^{**} \leftarrow \begin{bmatrix} e^T \\ S^{(k_j)} \end{bmatrix}.$$

Since $\begin{bmatrix} e^T \\ S^{(0)} \end{bmatrix}$ is nonsingular, we obtain that $B^* = B^{**}$, which is a contradiction. It follows that $\{B_k\}$ converges. □

Hence it is enough to study the convergence of $\{B_k\}$, or more precisely the convergence of the right infinite matrix product

$$B = \prod_{i=1}^{\infty} T_i P^{(i)}. \tag{5.6}$$

Let A be an $n \times n$ matrix. The 1-eigenspace of the matrix A is

$$E(A) = \{x : xA = x\}.$$

Lemma 3. *Assume that $B_k \to B$ and $T_s P^{(s)}$ occurs infinitely often in the product $\prod_{j=1}^{\infty} T_j P^{(j)}$, then every row of B is in $E\left(T_s P^{(s)}\right)$.*

Proof. Since $T_s P^{(s)}$ occurs infinitely often in the product $\prod_{i=1}^{\infty} T_i P^{(i)}$, there is a subsequence of $\left\{B_{i_j}\right\}$ with rightmost factor TP, say

$$B_{i_1} T_s P^{(s)}, B_{i_2} T_s P^{(s)}, \ldots,$$

where the B_{i_j}'s are products of $T_i P^{(i)}$'s. Since $B_{i_j} \to B$, so does $B_{i_j} T_s P^{(s)}$. Thus $B_{i_j} T_s P^{(s)} \to B T_s P^{(s)} = B$. □

In fact the rows of B, if not zero vectors, are the left eigenvectors of $T_s P^{(s)}$ belonging to $\lambda = 1$. If several $T_s P^{(s)}$ occur infinitely often in the product $\prod_{i=1}^{\infty} T_i P^{(i)}$, then the rows of B belong to $\cap E\left(T_s P^{(s)}\right)$, where the intersection is over all matrices $T_s P^{(s)}$ that occur infinitely often in $\prod_{k=1}^{\infty} T_i P^{(i)}$.

Each of the matrices $T_i P^{(i)}$ has at least one left eigenvector (e^T) belonging to $\lambda = 1$. Hence $e^T \in \cap E\left(T_s P^{(s)}\right)$. If $\cap E\left(T_s P^{(s)}\right) = \left\{\lambda e^T : \lambda \in \mathbb{R}\right\}$, then B has the form we^T for some $w \in \mathbb{R}^{n+1}$. However it is not always the case.

Example 4. *Let $n = 2$, $\{d_k\}_{k=0}^{\infty}$ be a strictly monotone decreasing sequence, $d_k > \delta_1 > \delta_2$ for all k, and define $f_1^{(k)} = \delta_2$, $f_2^{(k)} = \delta_1$, $f_3^{(k)} = d_{k+1}$, $f_r^{(k)} = d_{k+1}$ and $f_{ic}^{(k)} = d_{k+2}$. Then $x_1^{(k+1)} = x_1^{(k)}$, $x_2^{(k+1)} = x_2^{(k)}$, $x_3^{(k+1)} = x_{ic}^{(k)}$, $f\left(x_1^{(k+1)}\right) = \delta_2$,*

$f\left(x_2^{(k+1)}\right) = \delta_1$ *and* $f\left(x_3^{(k+1)}\right) = d_{k+2}$. *Hence a repeated inside contraction occurs,*

$$B_k = \left[T\left(-\frac{1}{2}\right)\right]^k \to B = \begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix} \neq we^T.$$

Here $x_1^{(k)} \to x_1^{(0)}$, $x_2^{(k)} \to x_2^{(0)}$, $x_3^{(k)} \to \frac{1}{2}\left(x_1^{(0)} + x_2^{(0)}\right)$, *and the simplex diameters do not converge to* 0. *Note that* B *is not a rank one matrix.*

Assume now that $B_k \to B = we^T$. Then $S^{(k)} = S^{(0)}B_k \to S^{(0)}we^T = \widehat{w}e^T$, $\operatorname{diam}\left(S^{(k)}\right) = \max_{i,j}\left\|S^{(k)}\left(e_i - e_j\right)\right\|$ and

$$\operatorname{diam}\left(S^{(k)}\right) \leq \max_{i,j}\left\|S^{(0)}\right\| \left\|B_k\left(e_i - e_j\right)\right\| \to 0. \tag{5.7}$$

Since $B\left(e_i - e_j\right) = 0$, $B_k\left(e_i - e_j\right) = \left(B_k - B\right)\left(e_i - e_j\right)$, we have the speed estimate

$$\operatorname{diam}\left(S^{(k)}\right) \leq \sqrt{2}\left\|S^{(0)}\right\| \left\|B_k - B\right\|. \tag{5.8}$$

Note again that in Example 4, where B is of rank 2, the simplex diameters do not converge to 0.

If B_k converges to a rank one matrix we^T, then all simplex vertices $x_i^{(k)}$ $(i = 1, \ldots, n+1)$ converge to the same limit $\widehat{x} = S^{(0)}w$ implying that $f_i^{(k)} \to f\left(S^{(0)}w\right)$ $(i = 1, \ldots, n+1)$ and $\operatorname{diam}\left(S^{(k)}\right) \to 0$. In such a case the results of Lagarias et al. [8] mentioned as 2. and 3. in Section 3, are direct consequences.

The next example indicates that the boundedness assumption on $\{B_k\}$ is also needed.

Example 5. *Let* $n = 2$, $\{d_k\}_{k=1}^{\infty}$ *be a strictly monotone decreasing sequence, and define*

$$f_1^{(k)} = d_{3+k}, \quad f_2^{(k)} = d_{2+k}, \quad f_3^{(k)} = d_{1+k}$$

and

$$f_1^{(k)} = d_{3+k} > f_r^{(k)} = \frac{1}{2}\left(d_{3+k} + d_{4+k}\right) > d_{4+k} = f_e^{(k)}.$$

This guarantees the selection of $x_e^{(k)}$ *as the incoming vertex for each iteration. The sequences* $\left\{f\left(x_1^{(k)}\right)\right\}$, $\left\{f\left(x_2^{(k)}\right)\right\}$ *and* $\left\{f\left(x_3^{(k)}\right)\right\}$ *are strictly monotone decreasing, while* $B_k = \left[T\left(2\right)P_1\right]^k$ *is unbounded. For* $d_k \to d$, *the function values are converging to* d, *while there is no convergence for the simplex vertices. A similar example can be given for* $x_r^{(k)}$ *if it is selected as an expansion point.*

For a given $S^{(0)}$ and f, the sequence $S^{(k)}$ (B_k) is uniquely defined. In fact, $S^{(k+1)}$ is determined by $S^{(k)}$ and the relative value distribution of f at the vertices of $S^{(k)}$ and trial points $x_r^{(k)}$, $x_e^{(k)}$, $x_{oc}^{(k)}$ and $x_{ic}^{(k)}$. Hence we study the convergence of the matrix product B_k and the convergence of the simplex vertices as a consequence. The selection of the initial simplex $S^{(0)}$ may also influence the convergence of the Nelder-Mead algorithm but it is difficult to consider it within this approach (for experimental observations on the initial simplex $S^{(0)}$, see [20], [21]).

6. Eigenvalues of the transformation matrices

The eigenvalues of the matrices $T_i P^{(i)}$ play a key role in the convergence of the infinite matrix product $\prod_{i=1}^{\infty} T_i P^{(i)}$. For $P^{(i)} = I_{n+1}$, we have the following simple result.

Lemma 6. *(i) Matrix $T(\alpha)$ has the eigenvalues $\lambda_i = 1$ $(i = 1, \ldots, n)$ and $\lambda_{n+1} = -\alpha$ and the diagonal Jordan form*

$$T(\alpha) = \begin{bmatrix} I_n & -\frac{1}{n}e \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I_n & 0 \\ 0 & -\alpha \end{bmatrix} \begin{bmatrix} I_n & \frac{1}{n}e \\ 0 & 1 \end{bmatrix}.$$

(ii) The eigenvalues of T_{shr} are $\lambda_1 = 1$ and $\lambda_i = \frac{1}{2}$ for $i = 2, \ldots, n+1$. The corresponding eigenvectors are $x_1 = e_1$, $x_i = -e_1 + e_i$. Furthermore T_{shr} has a diagonal Jordan normal form

$$T_{shr} = X \Lambda X^{-1}, \tag{6.1}$$

where $\Lambda = diag(\lambda_i) = \frac{1}{2}\left(I_{n+1} + e_1 e_1^T\right)$ and $X = I_{n+1} + e_1\left(e_1^T - e^T\right)$.

Lemma 7. *The matrix $T(\alpha) P_k$ $(1 \le k \le n)$ has $k-1$ eigenvalues $\lambda = 1$. The remaining $n-k+2$ eigenvalues are the zeros of the polynomial*

$$p_{n+2-k}(\lambda) = \lambda^{n+2-k} - c \sum_{i=1}^{n+1-k} \lambda^i + \alpha, \tag{6.2}$$

where $c = \frac{1+\alpha}{n}$. If $k = 1$, then $p_{n+1}(\lambda)$ has at least one eigenvalue $\lambda = 1$. If $\alpha = 1$, $p_{n+1}(\lambda)$ has at least two eigenvalues $\lambda = 1$. If $\alpha = 2$, $p_{n+1}(\lambda)$ has an eigenvalue in the interval $(1, 2)$. For $k \ge 2$, there are exactly $k-1$ eigenvalues $\lambda = 1$. If $2 \le k \le n$ and $\alpha < 0$, $p_{n-k+2}(\lambda)$ has all roots in the open unit disk.

Proof. For $1 \le k \le n$,

$$T(\alpha) P_k = \begin{bmatrix} I_{k-1} & cee_1^T \\ 0 & A_{n+2-k} \end{bmatrix},$$

where

$$A_{n+2-k} = \begin{bmatrix} c & 1 & 0 & \cdots & 0 \\ c & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ c & 0 & \cdots & 0 & 1 \\ -\alpha & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Since A_{n+2-k} is a companion matrix (for this form, see, e.g. [22]), its characteristic polynomial is

$$p_{n+2-k}(\lambda) = \lambda^{n+2-k} - c \sum_{i=1}^{n+1-k} \lambda^i + \alpha \quad (1 \le k \le n)$$

and the characteristic polynomial of $T(\alpha) P_k$ is

$$\det\left(T(\alpha) P_k - \lambda I_{n+1}\right) = (1-\lambda)^{k-1} p_{n+2-k}(\lambda).$$

Note that $p_{n+2-k}(1) = \frac{k-1}{n}(\alpha+1)$. If $k=1$, $p_{n+1}(1)=0$, that is $\lambda_1 = 1$ (the column sums of $T(\alpha)P_1$ are 1!). Since $p'_{n+1}(\lambda) = \frac{1-\alpha}{2}(n+1)$, there is a second zero $\lambda_2 = 1$ if $\alpha = 1$ (expansion by reflection). For $\alpha = 2$, $p'_{n+1}(1) < 0$ and $p_{n+1}(2) = 2^{n+1}\left(1-\frac{3}{n}\right)+2+\frac{6}{n} > 0$ $(n \geq 2)$. Hence $p_{n+1}(\lambda)$ has a zero in the interval $(1,2)$ for $\alpha = 2$ and $n \geq 2$. If $k \geq 2$, $p_{n+2-k}(1) \neq 0$. If $\alpha < 0$ and $2 \leq k \leq n$, the roots of $p_{n+2-k}(\lambda)$ are inside the unit disk since for $|\lambda| \geq 1$,

$$\begin{aligned}
|p_{n+2-k}(\lambda)| &= |\lambda|^{n+2-k}\left|1 - c\sum_{i=1}^{n+1-k}\frac{1}{\lambda^i} + \alpha\frac{1}{\lambda^{n+2-k}}\right| \\
&\geq |\lambda|^{n+2-k}\left(1 - c\sum_{i=1}^{n+1-k}\frac{1}{|\lambda|^i} - |\alpha|\frac{1}{|\lambda|^{n+2-k}}\right) \\
&\geq |\lambda|^{n+2-k}\left(1-(n+1-k)c - |nc-1|\right) \\
&\geq |\lambda|^{n+2-k}(k-1)c > 0.
\end{aligned}$$

□

For $\alpha > 0$, there is no estimate on the location of the roots of p_{n+2-k} relative to the open unit disk. However the Schur-Cohn test may help to decide if the roots of $p_{n+2-k}(\lambda)$ are in the open unit disk.

For the eigenvalues of $T_{shr}P$ $(P \in \mathcal{P}_{n+1})$, we cite the following result

Theorem 8. *(Langville and Meyer [23, 24]). If the spectrum of the stochastic matrix P is $\{1, \lambda_2, \ldots, \lambda_3\}$, then the spectrum of*

$$W = \alpha P + (1-\alpha)ev^T$$

is $\{1, \alpha\lambda_2, \alpha\lambda_3, \ldots, \alpha\lambda_n\}$, where v^T is a probability vector[2].

Corollary 9. *Since the eigenvalues of W and W^T coincide, we have the same result for the transposed matrix*

$$W^T = \alpha P^T + (1-\alpha)ve^T$$

as well.

Corollary 10. *The spectrum of $T_{shr}P = \frac{1}{2}P + \frac{1}{2}e_1e^T$ is $\left\{1, \frac{1}{2}\lambda_2, \frac{1}{2}\lambda_3, \ldots, \frac{1}{2}\lambda_{n+1}\right\}$. Since the eigenvalues of a permutation matrix are on the unit circle $|\lambda| = 1$, we have $\left|\frac{1}{2}\lambda_i\right| = \frac{1}{2}$ for $i = 2, \ldots, n+1$.*

7. General convergence results for the Nelder-Mead method

For every $T(\alpha)P_k$, the spectral radius is bigger than or equal to 1. Since $e^T \in E\left(T_iP^{(i)}\right)$ for all $T_iP^{(i)}$ occurring here, we first block triangularize them by a common similarity transformation (for left infinite matrix products, see Theorem 6.10 of Hartfiel [25]). We show that for

$$F = \begin{bmatrix} 1 & -e^T \\ 0 & I_n \end{bmatrix}, \quad F^{-1} = \begin{bmatrix} 1 & e^T \\ 0 & I_n \end{bmatrix} \tag{7.1}$$

[2]Vector v is such that $v_i \geq 1$ and $e^Tv = 1$.

and for all possible $T_i P^{(i)}$,

$$F^{-1} T_i P^{(i)} F = \begin{bmatrix} 1 & 0^T \\ b_i & C_i \end{bmatrix}, \tag{7.2}$$

where $b_i \in \mathbb{R}^n$ and $C_i \in \mathbb{R}^{n \times n}$ depends on $T_i P^{(i)}$.

Lemma 11. *For every $T(\alpha) P_k$ ($\alpha \in \left\{-\frac{1}{2}, \frac{1}{2}, 1, 2\right\}$, $P_k \in \{P_1, \ldots, P_{n+1}\}$), the matrix $F^{-1} T(\alpha) P_k F$ has the form (7.2).*

Proof. For $k > 1$ we can write

$$T(\alpha) P_k = \begin{bmatrix} 1 & c e_{k-1}^T \\ 0 & W \end{bmatrix} \quad \left(W \in \mathbb{R}^{n \times n}\right),$$

and so

$$F^{-1} T(\alpha) P_k F = \begin{bmatrix} 1 & -e^T + c e_{k-1}^T + e^T W \\ 0 & W \end{bmatrix}.$$

Since $e^T W e_{k-1} = (n-1)\, c - \alpha = 1 - c$, $e^T W = [1, \ldots, 1-c, 1, \ldots, 1]$, we obtain the form

$$F^{-1} T(\alpha) P_k F = \begin{bmatrix} 1 & 0 \\ 0 & W \end{bmatrix}.$$

For $k = 1$, we can write

$$T(\alpha) P_1 = \begin{bmatrix} c & e_1^T \\ z & W \end{bmatrix} \quad \left(W \in \mathbb{R}^{n \times n}\right)$$

with $z = [c, \ldots, c, -\alpha]^T$. Hence

$$F^{-1} T(\alpha) P_1 F = \begin{bmatrix} c + e^T z & -c e^T + e_1^T - e^T z e^T + e^T W \\ z & -z e^T + W \end{bmatrix}.$$

Since $e^T W = [0, 1, \ldots, 1]$, $e^T z = 1 - c$, $c + e^T z = 1$,

$$-c e^T + e_1^T - e^T z e^T + e^T W = -c e^T + e_1^T - (1-c)\, e^T + e^T W = 0.$$

The final result is

$$F^{-1} T(\alpha) P_1 F = \begin{bmatrix} 1 & 0 \\ z & -z e^T + W \end{bmatrix}.$$

□

Remark 12. *For $k > 1$, $b = 0$, and for $k = 1$, $\|b\|_2 = \left(\frac{(1+\alpha)^2}{n} + \alpha^2\right)^{\frac{1}{2}}$.*

Lemma 13. *For every $T_{shr} P$ ($P \in \mathcal{P}_{n+1}$), the matrix $F^{-1} T_{shr} P F$ has the form (7.2).*

Proof. Note that $T_{shr} P = \frac{1}{2} P + \frac{1}{2} e_1 e^T$ and $P = \left[e_{i_1}, \ldots, e_{i_{n+1}}\right]$. If $i_1 = 1$, then

$$T_{shr} P = \begin{bmatrix} 1 & \frac{1}{2} e^T \\ 0 & W_1 \end{bmatrix}$$

where W_1 is an $n \times n$ permutation matrix multiplied by $\frac{1}{2}$. Hence $e^T W_1 = \frac{1}{2} e^T$ and

$$F^{-1} T_{shr} P F = \begin{bmatrix} 1 & -\frac{1}{2} e^T + e^T W_1 \\ 0 & W_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & W_1 \end{bmatrix}.$$

If $i_1 > 1$ and $T_{shr} P e_j = e_1$, then

$$T_{shr} P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} e^T + \frac{1}{2} e_{j-1}^T \\ \frac{1}{2} e_{i_1 - 1} & W_2 \end{bmatrix},$$

where $W_2 e_{j-1} = 0$, $e^T W_2 e_i = \frac{1}{2}$ $(i \neq j-1)$ and $e^T W_2 = \frac{1}{2} \left(e^T - e_{j-1}^T \right)$. Since $e^T e_{i_1 - 1} = 1$,

$$F^{-1} T_{shr} P F = \begin{bmatrix} \frac{1}{2} + \frac{1}{2} e^T e_{i_1 - 1} & \frac{1}{2} e_{j-1}^T - \frac{1}{2} e^T e_{i_1 - 1} e^T + e^T W_2 \\ \frac{1}{2} e_{i_1 - 1} & -\frac{1}{2} e_{i_1 - 1} e^T + W_2 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 \\ \frac{1}{2} e_{i_1 - 1} & -\frac{1}{2} e_{i_1 - 1} e^T + W_2 \end{bmatrix}$$

□

Remark 14. *If $i_1 = 1$, then the first column entries are 0 except for entry $(1,1)$. If $i_1 \geq 2$, then entry $(i_1, 1)$ is $\frac{1}{2}$, while the remaining entries are 0 $(\ell \neq 1, i_1)$. Hence $\|b\|_2 \leq \frac{1}{2}$. The entries of submatrix C are only 0, $\frac{1}{2}$ and $-\frac{1}{2}$. In column j, there can be at most two nonzero elements. Theorem 8 and Corollaries 9 and 10 imply that $\rho(C) = \frac{1}{2}$. Note that $\|C\|_1 \leq 1$.*

We also need the following simple results.

Assume that for $i \geq 1$,

$$A_i = \begin{bmatrix} 1 & 0 \\ b_i & C_i \end{bmatrix} \in \mathbb{R}^{(n+1) \times (n+1)} \quad \left(C_i \in \mathbb{R}^{n \times n} \right). \tag{7.3}$$

It is easy to see that

$$L_k = \prod_{j=1}^{k} A_j = \begin{bmatrix} 1 & 0 \\ \sum_{i=1}^{k} \left(\prod_{j=1}^{i-1} C_j \right) b_i & \prod_{j=1}^{k} C_j \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ x_k & \prod_{j=1}^{k} C_j \end{bmatrix}. \tag{7.4}$$

Lemma 15. *Assume that $\left\| \prod_{j=1}^{k} C_j \right\| \leq c_k$, $\sum_{k=1}^{\infty} c_k$ is convergent $(< \infty)$ and $\|b_k\| \leq \gamma$ for all k. Then L_k converges and*

$$\lim_{k \to \infty} L_k = \begin{bmatrix} 1 & 0 \\ \widetilde{x} & 0 \end{bmatrix} \tag{7.5}$$

for some $\widetilde{x}$.

Proof. If $\sum_{k=1}^{\infty} c_k$ is convergent, then $c_k \to 0$. Hence $\prod_{j=1}^{k} C_j \to 0$ as $k \to \infty$. Since $s_k = \sum_{j=1}^{k} c_j$ is convergent, for any $\varepsilon > 0$ there is a number $k_0 = k_0(\varepsilon)$ such that for

$m > k \geq k_0$, $|s_m - s_k| < \varepsilon$. Thus for $m > k \geq k_0$, we obtain

$$\|x_m - x_k\| \leq \sum_{i=k+1}^{m} \left\| \prod_{j=1}^{i-1} C_j \right\| \|b_i\| \leq \gamma \sum_{i=k+1}^{m} c_{i-1} \leq \gamma\varepsilon.$$

Hence $x_k \to \widetilde{x}$ for some $\widetilde{x}$. □

Remark 16. *If $\|C_j\| \leq q < 1$ for $j \geq 1$, then $\left\| \prod_{j=1}^{k} C_j \right\| \leq q^k$ and the series $\sum_{i=1}^{\infty} q^i$ is convergent.*

Remark 17. *For the matrices C_i of (7.2), either $\rho(C_i) < 1$ or $\rho(C_i) \geq 1$. Since in any induced matrix norm $\|C_i\| \geq \rho(C_i)$, we cannot expect convergence improvement from the latter matrices, unless their effect is compensated.*

Assume now that for some indices j, $1 \leq \|C_j\| \leq Q$, while for other indices i, $\|C_i\| \leq q < 1$. Denote by $t_1(k)$ the number of those C_i's that satisfies $\|C_i\| \leq q < 1$ $(1 \leq i \leq k)$, and denote by $t_2(k)$ those C_i's that satisfiy $1 \leq \|C_i\| \leq Q$ $(1 \leq i \leq k)$. Clearly, $0 \leq t_i(k) \leq k$ and $t_1(k) + t_2(k) = k$. Then

$$\left\| \prod_{j=1}^{k} C_j \right\| \leq q^{t_1(k)} Q^{t_2(k)}.$$

Assume that $\kappa \in \mathbb{N}$ is such that $\frac{1}{q^{\kappa-1}} \leq Q \leq \frac{1}{q^{\kappa}}$. Then $\left\| \prod_{j=1}^{k} C_j \right\| \leq q^{t_1(k) - \kappa t_2(k)}$. If $\varphi(k) := t_1(k) - \kappa t_2(k)$ is a monotone increasing sequence converging to infinity, then $\prod_{j=1}^{k} C_j \to 0$ as $k \to \infty$. Note that $\varphi(k) \leq k$. The root test of infinite series guarantees that if for some $k_0 > 0$,

$$q^{\frac{\varphi(k)}{k}} \leq r < 1 \quad (k \geq k_0),$$

then $\sum_{i=1}^{\infty} q^{\varphi(i)}$ is convergent. This condition is certainly satisfied if $\frac{\varphi(k)}{k} \geq \mu$, where $0 < \mu < 1$ is a fixed number. Observe that in such a case $k \geq t_1(k) \geq \mu k + \kappa t_2(k)$ and $t_2(k) \leq \frac{1-\mu}{\kappa} k$. If $Q = 1$, then $\kappa = 1$.

We can also give an estimate for the speed of convergence. For $\prod_{i=1}^{k} C_j \to 0$, we have the estimate $\left\| \prod_{j=1}^{k} C_j \right\| \leq q^{\varphi(k)} \leq q^{\mu k}$. For the speed of the convergence $\sum_{i=1}^{k} \left(\prod_{j=1}^{i-1} C_j \right) b_i \to \widetilde{x}$, we have the estimate

$$\|\widetilde{x} - x_k\| = \left\| \sum_{i=k+1}^{\infty} \left(\prod_{j=1}^{i-1} C_j \right) b_i \right\| \leq \gamma \sum_{i=k}^{\infty} q^{\varphi(i)} \leq \gamma \sum_{i=k}^{\infty} q^{\mu i} \leq \frac{\gamma q^{\mu k}}{1 - q^{\mu}}.$$

We have just proved the following

Corollary 18. *Assume that $\|C_k\| \leq Q$ and $\|b_k\| \leq \gamma$ for all k. Denote by $t_1(k)$ the number of those C_i's that satisfies $\|C_i\| \leq q < 1$ $(1 \leq i \leq k)$, and denote by $t_2(k)$ those C_i's that satisfies $1 \leq \|C_i\| \leq Q$ $(1 \leq i \leq k)$. Define $\kappa \in \mathbb{N}$ by the inequality*

$\frac{1}{q^{\kappa-1}} \leq Q \leq \frac{1}{q^{\kappa}}$. *If for some* $k_0 \in \mathbb{N}$ *and* $\mu \in (0,1)$, $t_1(k) \geq \mu k + \kappa t_2(k)$ *holds* $(k \geq k_0)$, *then* L_k *converges and*

$$\lim_{k\to\infty} L_k = \begin{bmatrix} 1 & 0 \\ \widetilde{x} & 0 \end{bmatrix} = \widetilde{L} \tag{7.6}$$

for some $\widetilde{X}$. *For the speed of the convergence we have the estimate*

$$\left\| L_k - \widetilde{L} \right\| \leq \Gamma q^{\mu k}, \tag{7.7}$$

where $\Gamma > 0$ *is a suitable constant depending on* μ *and* γ.

Remark 19. *If* $t_2(k) = 0$ *for* $k \geq 1$, *then* $t_1(k) = k$ *and* $\mu = 1$. *If there are only a finite number of cases, when* $1 \leq \|C_i\| \leq Q$ *holds, we can omit it in the sense that if after iteration, say* k_0, *only* $\|C_i\| \leq q < 1$ *occurs, then* $t_2(k) \leq k_0$, $t_1(k) \geq k - k_0$ *and we can set* $\mu = 1$ *in the above Corollary 18.*

Given n, we have a finite set of matrices $T_i P^{(i)}$, say, $\mathcal{T}$ that may occur in the infinite product (5.6) if the Nelder-Mead method is applied to some function f. For each $T_i P^{(i)}$, we have the representation

$$T_i P^{(i)} = F \begin{bmatrix} 1 & 0^T \\ b_i & C_i \end{bmatrix} F^{-1}. \tag{7.8}$$

Hence $B_k = \prod_{i=1}^{k} T_i P^{(i)} = F L_k F^{-1}$ and B_k is convergent if and only if

$$L_k = \prod_{i=1}^{k} \begin{bmatrix} 1 & 0^T \\ b_i & C_i \end{bmatrix}$$

is convergent. For some of the C_i's, $\rho(C_i) \leq q \leq 1$, while for the others $\rho(C_i) \geq 1$. Define the sets

$$\mathcal{W}_1 = \left\{ T\left(\frac{1}{2}\right) P_j, T\left(-\frac{1}{2}\right) P_j : j = 1,2 \right\} \cup \{ T_{shr} P : P \in \mathcal{P}_{n+1} \}$$

and

$$\mathcal{W}_2 = \{ T(2) P_1, T(1) P_j : j = 1, \ldots, n \} \cup \left\{ T\left(\frac{1}{2}\right) P_j, T\left(-\frac{1}{2}\right) P_j : j = 3, \ldots, n+1 \right\}.$$

If $T_i P^{(i)} \in \mathcal{W}_2$, the matrix has a second eigenvalue $|\lambda_2| \geq 1$, and for the corresponding C_i, $\rho(C_i) \geq 1$ holds. Although Lemma 7 does not guarantee that for matrices $T_i P^{(i)} \in \left\{ T\left(\frac{1}{2}\right) P_j, T\left(-\frac{1}{2}\right) P_j : j = 1,2 \right\}$, the second largest eigenvalue $|\lambda_2| < 1$ or $\rho(C_i) < 1$, a thorough computer check shows that it is the case, at least up to $n = 20$. However $|\lambda_2|$'s are approaching 1 as n increases. Hence we assume the following:

(A) *If* $T_i P^{(i)} \in \mathcal{W}_1$, *then* $\rho(C_i) < 1$, *and if* $T_i P^{(i)} \in \mathcal{W}_2$, *then* $\rho(C_i) \geq 1$.
(B) *There is a matrix norm* $\|A\|_w$ *(induced by a vector norm* $\|x\|_w$*) such that if* $T_i P^{(i)} \in \mathcal{W}_1$, *then* $\|C_i\|_w < 1$.

Since the sets $\mathcal{W}_1$ and $\mathcal{W}_2$ are finite, there are numbers $0 < q < 1 \leq Q$ such that for $T_i P^{(i)} \in \mathcal{W}_1$, $\|C_i\|_w \leq q < 1$, and for $T_i P^{(i)} \in \mathcal{W}_2$, $1 \leq \|C_i\|_w \leq Q$. Also we can assume that for every $T_i P^{(i)} \in \mathcal{W}_1 \cup \mathcal{W}_2$, $\|b_i\|_w \leq \gamma$.

Theorem 20. *Assume (A) and (B) are satisfied and the initial simplex $S^{(0)}$ is non-degenerate. Let $t_1(k)$ be the number of Nelder-Mead steps (operations $T_i P^{(i)}$) that belong to $\mathcal{W}_1$, and $t_2(k)$ be the number of those steps (operations $T_i P^{(i)}$) that belong to $\mathcal{W}_2$ during the first k iterations of the Nelder-Mead method. Also assume that $\kappa \in \mathbb{N}$ is such that $\frac{1}{q^{\kappa-1}} \leq Q \leq \frac{1}{q^{\kappa}}$ If for some $k_0 \in \mathbb{N}$ and $\mu \in (0,1)$, $t_1(k) \geq \mu k + \kappa t_2(k)$ holds ($k \geq k_0$), then the Nelder-Mead algorithm converges in the sense that for all simplex vertices $x_j^{(k)} \to \widehat{x}$ ($j = 1, \ldots, n+1$) holds as $k \to \infty$ with a convergence speed proportional to $q^{\mu k}$. If f is continuous at $\widehat{x}$, then $f\left(x_j^{(k)}\right) \to f(\widehat{x})$ ($j = 1, \ldots, n+1$) holds as well.*

Proof. Under the assumptions

$$L_k = \prod_{i=1}^{k} \begin{bmatrix} 1 & 0^T \\ b_i & C_i \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ \widetilde{x} & 0 \end{bmatrix} = \widetilde{L}$$

with the speed

$$\left\| L_k - \widetilde{L} \right\|_w \leq \Gamma q^{\mu k}.$$

Hence

$$B_k \to F \begin{bmatrix} 1 & 0 \\ \widetilde{x} & 0 \end{bmatrix} F^{-1} = \begin{bmatrix} 1 - e^T \widetilde{x} \\ \widetilde{x} \end{bmatrix} e^T = \widehat{x} e^T = B \tag{7.9}$$

with the speed

$$\|B_k - B\|_w \leq \Gamma \mathrm{cond}(F) q^{\mu k}. \tag{7.10}$$

□

Corollary 21. *$diam\left(S^{(k)}\right) \to 0$ ($k \to \infty$) with a speed of $O\left(q^{\mu k}\right)$.*

For higher dimension, we can expect slower convergence, since q approaches 1.

8. The convergence of the Nelder-Mead method in low dimensions

Here we demonstrate the use of Theorem 20 for $n = 1, 2, 3$. The cases show an increasing technical complexity and also the growth of $\max\left\{\rho(C_i) : T_i P^{(i)} \in \mathcal{W}_1\right\}$.

8.1. The Nelder-Mead method in one dimension.

For $n = 1$,

$$\mathcal{W}_1 = \left\{ T\left(\frac{1}{2}\right) P_j, T\left(-\frac{1}{2}\right) P_j : j = 1, 2 \right\}$$

and

$$\mathcal{W}_2 = \{ T(1) P_2, T(2) P_2, T(1) P_1 \}.$$

Lemma 11 implies

$$F^{-1} B_k F = \prod_{i=1}^{k} \begin{bmatrix} 1 & 0 \\ b_i & c_i \end{bmatrix}.$$

Note that for $T_iP^{(i)} \in \mathcal{W}_1$, $|c_i| = \frac{1}{2} = q$, and for $T_iP^{(i)} \in \mathcal{W}_2$, $|c_i| \geq 1$ and $1 \leq |c_i| \leq 2 = Q$. Here the norm $\|\cdot\|_w = |\cdot|$. Hence Theorem 20 implies the convergence with $\kappa = 1$.

8.2. **The Nelder-Mead method in two dimensions.** For $n = 2$, the six permutation matrices of set $\mathcal{P}_3$ are

$$P_1 = [e_3, e_1, e_2], \quad P_2 = [e_1, e_3, e_2], \quad P_3 = [e_1, e_2, e_3],$$

$$P_4 = [e_2, e_1, e_3], \quad P_5 = [e_2, e_3, e_1], \quad P_6 = [e_3, e_2, e_1].$$

Define

$$\mathcal{W}_1 = \left\{ T\left(\frac{1}{2}\right) P_j, T\left(-\frac{1}{2}\right) P_j : j = 1, 2 \right\} \cup \{T_{shr} P_j : j = 1, \ldots, 6\}$$

and

$$\mathcal{W}_2 = \{T(1) P_2, T(2) P_1, T(1) P_1\} \cup \left\{ T\left(\frac{1}{2}\right) P_3, T\left(-\frac{1}{2}\right) P_3 \right\}.$$

The similarity transformation

$$F^{-1} T P F = \begin{bmatrix} 1 & 0^T \\ b & C \end{bmatrix} \tag{8.1}$$

on the elements of $\mathcal{W}_1 \cup \mathcal{W}_2$ yields matrices C_i for which $\rho(C_i) < 1$, while $\|C_i\|_2 > 1$.

Since all such C_i's have diagonal Jordan forms, it would be an ideal situation if for some norm $\|\cdot\|_w$, every $\|C_i\|_w$ would be close to $\rho(C_i)$. Such thing is possible, if the matrices C_i are simultaneously similar to diagonal matrices (see, e.g. [26–28]). However, this requires that all matrices C_i must be pairwise commuting, which is not the case here. Instead of this we look for a 2×2 matrix S such that $\|C_i\|_w = \left\|S^{-1} C_i S\right\|_2$ (induced by the vector norm $\|x\|_w = \left\|S^{-1} x\right\|_2$) is as close to $\rho(S_i)$ as possible. So we try to solve the optimization problem

$$\min_S \max \left\{ \left\|S^{-1} C_i S\right\|_2 : T_i P^{(i)} \in \left\{ T\left(\frac{1}{2}\right) P_j, T\left(-\frac{1}{2}\right) P_j : j = 1, 2 \right\} \right\}. \tag{8.2}$$

For the matrix

$$S = \begin{bmatrix} 2 & -\frac{4}{5} \\ 0 & \frac{8}{5} \end{bmatrix} \quad (\operatorname{cond}(S) \approx 1.640\,4),$$

we have the following numerical results.

Case	Operation	C_i	$\rho(C_i)$	$\|C_i\|_2$	$\left\|S^{-1}C_iS\right\|_2$
1	$F^{-1}T(1)P_2F$	C_1	1	1.6180	1.3765
2a	$F^{-1}T(2)P_1F$	C_2	1.6861	3.1787	2.7043
2b	$F^{-1}T(1)P_1F$	C_3	1	1.6180	1.3765
3a	$F^{-1}T\left(\frac{1}{2}\right)P_1F$	C_4	0.8431	0.9056	0.8438
3b	$F^{-1}T\left(\frac{1}{2}\right)P_2F$	C_5	0.7071	1.2892	0.8438
3c	$F^{-1}T\left(\frac{1}{2}\right)P_3F$	C_6	1	1.2892	1.0
4a	$F^{-1}T\left(-\frac{1}{2}\right)P_1F$	C_7	0.7071	0.9056	0.8438
4b	$F^{-1}T\left(-\frac{1}{2}\right)P_2F$	C_8	0.8431	1.0399	0.8438
4c	$F^{-1}T\left(-\frac{1}{2}\right)P_3F$	C_9	1	1.0399	1.0
5a	$F^{-1}T_{shr}P_1F$	C_{10}	$\frac{1}{2}$	0.8090	0.6882
5b	$F^{-1}T_{shr}P_2F$	C_{11}	$\frac{1}{2}$	0.5	0.6882
5c	$F^{-1}T_{shr}P_3F$	C_{12}	$\frac{1}{2}$	0.5	0.5
5d	$F^{-1}T_{shr}P_4F$	C_{13}	$\frac{1}{2}$	0.8090	0.5
5e	$F^{-1}T_{shr}P_5F$	C_{14}	$\frac{1}{2}$	0.8090	0.6882
5f	$F^{-1}T_{shr}P_6F$	C_{15}	$\frac{1}{2}$	0.8090	0.6882

Here we can select the values $q = 0.85$ ($q \approx \rho(C_4), \rho(C_8)$), $Q = 2.71$ and $\kappa = 7$. Note that for $T_iP^{(i)} \in \mathcal{W}_1$, $\|C_i\|_w \le q < 1$, and for $T_iP^{(i)} \in \mathcal{W}_2$, $1 \le \|C_i\|_w \le Q$.

Hence Theorem 20 implies convergence with $\kappa = 7$.

The applied optimization technique (8.2) is somewhat similar to the joint approximate diagonalization method (see, e.g., [29]).

8.3. **The Nelder-Mead method in three dimensions.** For $n = 3$, $\mathcal{W}_1$ has 28 elements, while $\mathcal{W}_2$ has 8. It can be checked that for $T_iP^{(i)} \in \mathcal{W}_1$, $\rho(C_i) \le 0.9275$ and $\|C_i\|_2 \le 1.2622$. A numerical solution of the optimization problem (8.2) gives the following matrix

$$S = \begin{bmatrix} -0.6012 & 1.5707 & 0.3968 \\ 1.4938 & -0.0616 & 0.4419 \\ -0.6500 & -0.7949 & 0.7620 \end{bmatrix},$$

for which $\|C_i\|_w = \left\|S^{-1}C_iS\right\|_2 \le 0.9293$ ($T_iP^{(i)} \in \mathcal{W}_1$) and $\|C_i\|_w \le 2.7729$ ($T_iP^{(i)} \in \mathcal{W}_2$). Selecting $q = 0.93$ and $Q = 2.78$, we have $\kappa = 15$, and Theorem 20 holds for $n = 3$, as well.

If we exclude the expansion steps, then we can set $Q = 1.3725$, resulting in $\kappa = 5$ and a faster convergence.

References

1. J. A. Nelder and R. Mead. "A simplex method for function minimization," *Computer Journal,* **7**, (1965), pp. 308–313. DOI: 10.1093/comjnl/7.4.308.
2. F. H. Walters, S. L. Morgan, L. R. Parker, and S.N. Deming. *Sequential Simplex Optimization.* CRC Press LLC, 1991.

3. R. M. Lewis, V. Torczon, and M. W. Trosset. "Direct search methods: then and now." *Journal of Computational and Applied Mathematics,* **124**, (2000), pp. 191–207. DOI: 10.1016/S0377-0427(00)00423-4.
4. T. G. Kolda, R. M. Lewis, and V. Torczon. "Optimization by Direct Search: New Perspectives on Some Classical and Modern Methods." *SIAM Review,* **45**(3), (2003), pp. 385–482. DOI: 10.1137/S003614450242889.
5. A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimizations.* MPS-SIAM Series on Optimization. SIAM, 2009. DOI: 10.1137/1.9780898718768.
6. C. Audet and W. Hare. *Derivative-free and Blackbox Optimization.* Springer, 2017. DOI: 10.1007/978-3-319-68913-5.
7. C. T. Kelley. "Detection and remediation of stagnation in the Nelder-Mead algorithm using an sufficient decrease condition." *SIAM Journal of Optimization,* **10**, (1999), pp. 33–45. DOI: 10.1137/S1052623497315203.
8. J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright. "Convergence properties of the Nelder-Mead simplex method in low dimensions." *SIAM Journal on Optimization,* **9**(1), (1998), pp. 112–147. DOI: 10.1137/S1052623496303470.
9. J. C. Lagarias, B. Poonen, and M. H. Wright. "Convergence of the restricted Nelder-Mead algorithm in two dimensions." *SIAM Journal on Optimization,* **22**(2), (2012), pp. 501–532. DOI: 10.1137/110830150.
10. M. H. Wright. "Nelder, Mead, and the other simplex method," *Documenta Mathematica, Extra Volume: Optimization Stories,* (2012), pp. 271–2766.
11. K. I. M. McKinnon. "Convergence of the Nelder-Mead simplex method to a nonstationary point," *SIAM Journal on Optimization,* **9**(1), (1998), pp. 148–158. DOI: 10.1137/S1052623496303482.
12. C. T. Kelley. *Iterative Methods for Optimization.* Frontiers in Applied Mathematics. SIAM, 1999. DOI: 10.1137/1.9781611970920.
13. P. Tseng. "Fortified-descent simplicial search method: A general approach," *SIAM Journal on Optimization,* **10**(1), (1999), pp. 269–288. DOI: 10.1137/S1052623495282857.
14. L. Nazareth and P. Tseng. "Gilding the Lily: A variant for the Nelder-Mead algorithm based on golden-section search," *Computational Optimization and Applications,* **22**, (2002), pp. 133–144. DOI: 10.1023/A:1014842520519.
15. C. J. Price, I. D. Coope, and D. Byatt. "A convergent variant of the Nelder-Mead algorithm," *Journal of Optimization Theory and Applications,* **113**(1), (2002), pp. 5–19. DOI: 10.1023/A:1014849028575.
16. A. Galántai. "A convergence analysis of the Nelder-Mead simplex method." *Acta Polytechnica Hungarica,* (2020). Submitted.
17. W. F. Trench. "Characterization and properties of matrices with k-involutory symmetries," *Linear Algebra and its Applications,* **429**, (2008), pp. 2278–2290. DOI: 10.1016/j.laa.2008.07.002.
18. J. G. Hocking and G. S. Young. *Topology.* Addison–Wesley, 1961.
19. J. Matousek. *Using the Borsuk-Ulam Theorem.* 2nd edition. Springer, 2008. DOI: 10.1007/978-3-540-76649-0.

20. J. M. Parkinson and D. Hutchinson. "An investigation into the efficiency of variants on the simplex method." *Numerical Method for Non-linear Optimization.* Ed. by F. A. Lootsma. Academic Press, 1972, pp. 115–135.
21. S. Wessing. "Proper initialization is crucial for the Nelder-Mead simplex search," *Optimization Letters,* **13**, (2019), pp. 847–856. DOI: 10.1007/s11590-018-1284-4.
22. R. A. Horn and C. R. Johnson. *Matrix Analysis.* 2nd edition. Cambridge University Press, 2013. DOI: 10.1017/9781139020411.
23. A. N. Langville and C. D. Meyer. "Deeper inside PageRank," *Internet Mathematics Journal,* **1**(3), (2005), pp. 335–380. DOI: 10.1080/15427951.2004.10129091.
24. A. N. Langville and C. D. Meyer. *Google's PageRank and Beyond: The Science of Search Engine Rankings.* Princeton University Press, 2006.
25. D. J. Hartfiel. *Nonhomogeneous Matrix Products.* World Scientific, 2001. DOI: 10.1142/4707.
26. L. Mirsky. *An Introduction to Linear Algebra.* Oxford University Press, 1955.
27. S. Friedland. *Matrices: Algebra, Analysis, and Applications.* World Scientific Publishing, 2015. DOI: 10.1142/9567.
28. K. C. O'Meara, J. Clark, and C. L. Vinsonhaler. *Advanced Topics in Linear Algebra: Weawing Matrix Problems through the Weyr Form.* Oxford University Press, 2011.
29. X. Shi. *Blind Signal Processing: Theory and Practice.* Springer, 2011. DOI: 10.1007/978-3-642-11347-5.

Journal of Computational and Applied Mechanics, Vol. 15, No. 2, (2020), pp. 135–172
DOI: 10.32973/jcam.2020.009

FORMULATION OF THE P-EXTENSION FINITE ELEMENTS FOR THE SOLUTION OF NORMAL CONTACT PROBLEMS

I. Páczelt[a], A. Baksa[a] and T. Szabó[b]
[a] Institute of Applied Mechanics, University of Miskolc, Hungary
[b]Institute of Machine Tools and Mechatronics, University of Miskolc,
istvan.paczelt@uni-miskolc.hu attila.baksa@uni-miskolc.hu
tamas.szabo@uni-miskolc.hu

[Received: August 10, 2020; Accepted: September 8, 2020]

Dedicated to Professor Barna Szabó on the occasion of his eighty-fifth birthday

Abstract. This work deals with normal contact problems. After a wide literature review, we look for the possibility of achieving a high-precision solution using the principle of minimum potential energy and the Hellinger-Reissner variational principle with penalty and augmented Lagrangian techniques. By positioning of the border of the contact elements, the whole surfaces of the eligible elements fall in contact or in gap regions. This reduces the error of the singularity in the border of the contact domain. Computations with *h*-, *p*- and *rp*-versions are performed. For the *rp*-version, the pre-fixed number of finite elements are moved so that small elements are placed in one or two element layers at the ends of the contact zone. A number of diagrams and tables showing the convergence of the solution (by increasing the number of polynomial degrees p) demonstrate the high efficiency of the proposed solution procedure.

Mathematical Subject Classification: 74S05, 74M15
Keywords: contact problems, p-version of finite elements

1. Introduction

The rapid progress in the computer sciences and information technology, the increase of computational speed, and improvements in graphic user-interface software provide new opportunities for modeling physical phenomena with a high degree of reliability and usefulness. It is now possible to perform very large number of arithmetic operations in a short time, making it possible for designers and analysts to employ advanced modeling concepts within the context of an everyday engineering decision-making process.

The solution of contact problems is difficult because the actual contact zone is unknown *a priori* and must be determined by an iterative procedure which seeks to satisfy not only equilibrium equations and boundary conditions but also one or more inequality contact conditions. Another source of difficulty is that the stress changes very substantially over short distances at the boundaries of the contact zone. An important design objective is to determine the shape of the contacting parts in such a way that the stress singularity will not occur, or the strain energy associated with

the singularity is within acceptable limits. Therefore the development of efficient and reliable procedures for the numerical treatment of contact is an important and challenging problem.

A number of methods have been published in the field of the mechanical contacts since the appearance of the Hertzian contact theory. Applying the electronic computers, better and better numerical processes were worked out from the late sixties. The paper [1] (Fridman and Chernina, 1967) was an important contribution to the numerical treatment of contact. Their iteration method was based on the search for the minimum of the complementary energy by ensuring positive contact pressure. Its further development could take into regard the relative rigid body displacements of the contacting bodies [2] (Páczelt, 1974). Mathematical programming methods were used to solve the minimum problem in [3](Conry and Seireg, 1971), [4] (Kalker and von Randen, 1972), [5](Fremond, 1973), [6](Frankavilla and Zienkiewicz, 1975), [7](Haug et al., 1977), [8](Páczelt and Herpai, 1977), [9](Páczelt, 1979), [10] (Hung and De Saxce, 1980). These methods were applied later as well, see [11](Klarbring, 1985) and [12](Björkman,1991) for instance.

The generality of the finite element method (FEM) has been utilized efficiently in the solution of contact problems [13](Chan and Tuba, 1971), [14] (Fredrikson, 1976), [15] (Páczelt, 1976), [16] (Huges et al., 1976), [17] (Bathe and Chaudhary, 1985), [18] (Nour-Omid and Wriggers, 1986), [19] (Chaudhary and Bathe, 1986), [20] (Égert and Altenbach, 1989), [21] (Mottershead et al., 1992), [22] (Papadopoulos and Taylor, 1992).

A Lagrangian multiplier technique was proposed in [15] (Páczelt, 1976) and [16] (Huges et al., 1976), and a penalty method was used first in [23] (Kikuchi et al., 1984). Contact conditions, such as contact and separation, are satisfied accurately for the contacting nodal point pairs in the Lagrangian multiplier method by implying the unknown contact pressure field as a multiplier. The fulfillment of the contact conditions in the solution of the penalty method strongly depends on the value of the penalty parameter. The combination of the two methods was applied in [24](Simo et al., 1985) for small displacement problems, and in [25] (Ju and Taylor, 1988), [26] (Simo and Laursen, 1992), [27] (Laursen and Simo, 1993) for large displacement cases. The mathematical backgrounds of these methods can be found in [28] (Berteskas, 1989), [29] (Fletcher, 1989).

A good review of static mechanical contact problems is given in [30] (Zhong and Mackerle, 1992).

The error of a finite element solution depends on the element size (h), and on polynomial degree (p) of the approximation.

The majority of the finite element applications use the h-version of the FEM. The solution accuracy is improved adaptively in [31] (Nackenhorst, 1995) by the help of the ZZ error estimator, its further development is given by [32] (Wriggers, 1997). Relatively few papers have dealt with the application of the p-version method [33] (Szabó and Babuska, 1991) for mechanical contact problems [34] (Lee and Oden, 1993), [35] (Gabbert and Graeff-Winberg, 1994), [36] (Buczkowski et al., 1994), [37] (Páczelt and Szabó, 1995), [38] (Páczelt and Baksa, 2009), [39] (Franke et al., 2010),

[40] (Konyukhov and Schweizerhof, 2009), [41] (Franke et al., 2011). A quasi-static problem was solved by finite elements with $p = 2$ applying mesh refinement in the contact region [34] (Lee and Oden, 1993).

A special pNh type element was used in [35] (Gabbert and Graeff-Winberg, 1994), [36] (Buczkowski et.al., 1994). There is piece-wise linear approximation on the contacting face of the element, and p-approximation is applied on the other sides. This sort of approximation provides non smooth stresses inside the element, which is a drawback concerning the efficiency of this element.

The shape optimization of the contacting surfaces was used in [37] (Páczelt and Szabó, 1995) by controlling the contact pressure. The high-quality approximation properties of the p-version make it possible to detect the presence of singularity, even weak singularity, by numerical means. This is possible theoretically but not feasible in practice with the h-version. The application of the augmented Lagrangian technique was investigated in [42] (Páczelt et al., 1999) concerning the solution accuracy. The contact conditions of the p-version elements are checked in the Gauss or Lobatto integration points. As a result of adaptive remeshing the whole sides of the elements are either in the contacting region or in the gap region. This preferable mesh performs no oscillation in the normal stresses along the possible contact region

There are a number of papers dealing with contact problems by the use of variational principles, e.g. [43] (Duvat and Lions, 1972), [44] (Hlavacek et.al., 1980), [45] (Panagiotopoulos, 1985), [46] (Oden and Kikuchi, 1988), [47] (Telega, 1987), [48] (Haslinger and Neittaanmaki, 1988).

Efficient algorithms have been proposed (Nour-Omid and Wriggers, 1986), [49] (Kalker, 1990), and [50] (Raos et al., 1988) in order to solve the system of inequalities established at frictional and frictionless contact problems. A broad range of application possibility of the mathematical programming techniques was analyzed by Klarbring [51]. A review for those papers devoted to the optimization of structures in unilateral mechanical contact is presented in [52] (Hilding et al., 1999).

The p-version was employed for locating the boundaries of the contact domain and the boundaries of the sliding-adhesive domain through the detection of the associated singularity. Two different procedures were used. In one case the standard polynomial basis functions were augmented by special basis functions which can approximate singularity [53] (Volpert, 1995), [54] (Volpert et.al., 1997). In the other case the boundaries of the elements were moved, using a special algorithm and a number of error indicators, in such a way that each element in the contact domain would either be in contact, separate, stick or slip along its full length [42] (Páczelt et.al., 1999), [55] (Páczelt and Szabó, 2002). The positioning technique was extended in Baksa et al. [56] for 3D contact problems including friction, where slip, stick zones have been separated. The borders of adhesion and slip zones and also of the outer border of the whole contact domain were approximated by NURBS. In this way the oscillation of the contact pressure and the oscillation of the shear stresses were practically eliminated, resulting in a very accurate solution of the problem.

In the past decades there have been some valuable books by Laursen 2002 [57], Wriggers 2006 [58], Konyukhov and Schweizerhof 2013 [59], Yastrebov 2013 [60], which describe the numerical treatment of contact problems. In these works not only small displacements and deformations but also large displacements and large deformations are taken into consideration for numerical analysis, both in normal and frictional contact examples.

High order finite elements are used for many different types of problems [61], [62]. The research in tribology, analytical solution of contact problems are examined by the book of Goryacheva [63], and we found a technique to analyze the wear process by p-extension finite elements [64].

In the book [65] many aspects of optimization of contact interaction of rigid, elastic, elastic-plastic bodies are investigated.

There have been some conferences e.g. CMM, Eccomas, HOFEIM, IUTAM, where not only the theoretical questions but also practical problems were analyzed [66–69]. Among the p-extension finite element computer systems, there is a simulation program called `StressCheck` developped by ESRD group leaded by Professor B. Szabó [70].

The aim of this investigation is to compare the results of non-frictional contact problems treated by traditional h-, p- and with adaptive $hp(rp)$-extension finite elements, and to show the computational algorithm for highly accurate $hp(rp)$-extension.

Using the notations introduced by papers [39], [41], we made some rp-extension computations as well. Here we use the principle of minimum of total potential energy and the mixed Hellinger-Reissner theory, which is based on approximating both the displacements and the stress fields [38]. In this paper a comparison is made to show the advantages of using the Hellinger-Reissner functional. The main question is whether we can get better results with the same number of unknowns of the displacement fields. However, the execution time is much longer because of the eliminating the stress field variables from the evolving stiffness matrix.

In case of the p- and hp- (rp)-versions, we use the augmented Lagrangian method for the solution of contact problems. It is assumed that displacements and deformations are small, and can be neglected as regard the adhesion, friction and dynamic effects between the contacting bodies.

This paper is organized as follows. In Section 2 the formulation of the contact problem is presented. In Section 3 two variational methods, minimum of potential energy and Hellinger-Reissner variational principle, are discussed with penalty and augmented Lagrangian form. In Section 4 finite element approximation and the system of algebraic inequalities are given, solving by Kalker-type iteration. Section 5 is dealing with the positioning technique of nodal contact points. Using this technique each eligible contact element along its whole side is either in contact or in the gap regions. Section 6 solves the same boundary value problem of contacting cylindrical bodies with different methods and techniques. These contact problems are solved by h-version, p-version and rp-version techniques. Stress states, distribution of the contact stresses, convergence of the contact radius, resultant of contact force, relative error in energy norm [33] and in contact force are demonstrated in figures. Results of

the principle of minimum potential energy and the Helliger-Reissner variational principle can be compared in order to choose the better one. In Section 7 Hertz probem with friction will be solved. Finally concluding remarks are provided in Section 8.

2. Formulation of the contact problem

Let us consider the contact problem of two elastic bodies ($e = 1, 2$). The surfaces of the bodies will be separated into three regions: $S_u^{(e)}$ denotes that part of the body where displacements $\boldsymbol{u}_o$ are given, in $S_t^{(e)}$ the traction $\boldsymbol{t}_o$ is applied, while $S_c^{(e)}$ represents that part of the bodies where contact is expected. The $S_c^{(e)}$ part of the body is called the proposed zone of the contact. The bodies are loaded with the body force $\boldsymbol{b}^{(e)}$, and initial stress $\boldsymbol{T}_o^{(e)}$ and initial strain $\boldsymbol{A}_o^{(e)}$ too. We are interested in finding the displacement vector field $\boldsymbol{u}$, strain $\boldsymbol{A}$ and stress $\boldsymbol{T}$ tensor fields. In the domain $V^{(e)}$ we have the equilibrium equation

$$\boldsymbol{T}^{(e)} \cdot \nabla + \boldsymbol{b}^{(e)} = \boldsymbol{0} \qquad \boldsymbol{r} \in V^{(e)}, \tag{2.1}$$

the strain displacement relationship

$$\boldsymbol{A}^{(e)} = \frac{1}{2}\left(\boldsymbol{u}^{(e)} \circ \nabla + \nabla \circ \boldsymbol{u}^{(e)}\right) \qquad \boldsymbol{r} \in V^{(e)}, \tag{2.2}$$

and the Hooke's constitutive law

$$\boldsymbol{T}^{(e)} = \boldsymbol{T}_o^{(e)} + \boldsymbol{D}^{(e)} \cdot\cdot \left(\boldsymbol{A}^{(e)} - \boldsymbol{A}_o^{(e)}\right) \qquad \boldsymbol{r} \in V^{(e)}, \tag{2.3}$$

where $\boldsymbol{D}^{(e)}$ is a fourth order tensor of the material parameters, "$\cdot$" , " $\cdot\cdot$ ", "$\circ$", are the symbols of a scalar, double scalar and tensor product, respectively, and ∇ is the Hamiltonian differential operator.

The boundary conditions are

$$\boldsymbol{u}^{(e)} = \boldsymbol{u}_o \qquad \boldsymbol{r} \in S_u^{(e)}, \tag{2.4}$$

and

$$\boldsymbol{T}^{(e)} \cdot \boldsymbol{n}^{(e)} = \boldsymbol{t}_o \qquad \boldsymbol{r} \in S_t^{(e)}. \tag{2.5}$$

For the examination of the contact/separation conditions in the proposed zone of contact we shall consider the projection of the displacement in a prescribed direction only (e.g., normal to the surface $\boldsymbol{n}_c$). The contact normal vector $\boldsymbol{n}_c$ determines the points Q_1, Q_2 on the corresponding surfaces $S_c^{(1)}$ and $S_c^{(2)}$, where the two surfaces may contact with each other (see Fig. 1). Therefore the contact surface will be denoted by S_c.

We denote the displacement projected in the direction $\boldsymbol{n}_c$ by $u_n^{(e)} = \boldsymbol{u}^{(e)} \cdot \boldsymbol{n}_c$, the normal stress by $\sigma_n^{(e)} = \boldsymbol{n}^{(e)} \cdot \boldsymbol{T}^{(e)} \cdot \boldsymbol{n}^{(e)}$ and the initial gap between bodies by h. We define the distance (gap) after deformation

$$d = d\,(\boldsymbol{u}) = u_n^{(2)} - u_n^{(1)} + h \geq 0, \tag{2.6}$$

and the contact pressure

$$p_n = -\boldsymbol{n}^{(1)} \cdot \boldsymbol{T}^{(1)} \cdot \boldsymbol{n}_c = \boldsymbol{n}^{(2)} \cdot \boldsymbol{T}^{(2)} \cdot \boldsymbol{n}_c \cong -\boldsymbol{n}^{(1)} \cdot \boldsymbol{T}^{(1)} \cdot \boldsymbol{n}^{(1)} = -\boldsymbol{n}^{(2)} \cdot \boldsymbol{T}^{(2)} \cdot \boldsymbol{n}^{(2)} \geq 0. \tag{2.7}$$

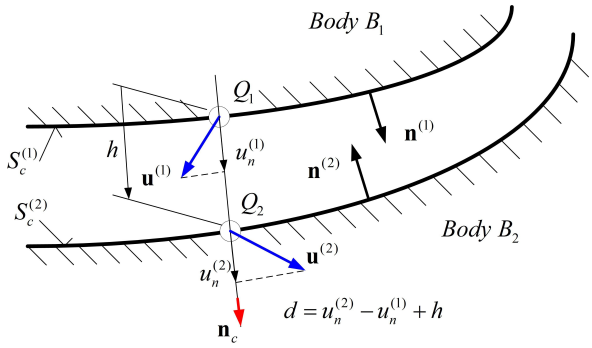


Figure 1. Normal displacements $u_n^{(e)}, e = 1, 2$

Denoting the contact zone by C and the separation (gap) zone by G $(S_c = C \cup G)$, we have

$$d = 0, \qquad p_n \geq 0 \qquad \boldsymbol{r} \in C, \tag{2.8}$$

$$d > 0, \qquad p_n = 0 \qquad \boldsymbol{r} \in G, \tag{2.9}$$

$$p_n d = 0 \qquad \boldsymbol{r} \in S_c. \tag{2.10}$$

From the condition of frictionless contact we have zero tangential stress

$$\tau^{(e)} = \boldsymbol{e}_\tau^{(e)} \cdot \boldsymbol{T}^{(e)} \cdot \boldsymbol{n}^{(e)} = 0 \qquad \boldsymbol{r} \in S_c^{(e)}, \tag{2.11}$$

where $\boldsymbol{e}_\tau^{(e)}$ is a tangential unit vector.

3. Variational formulations of the contact problem

In this section two variation principles will be discussed in parallel.

3.1. **Principles based on the total potential energy and Hellinger-Reissner functional.** For investigation of normal contact problem we can use the principle of minimum potential energy $\Pi(\boldsymbol{u})$ subjected to two types of kinematic conditions: $\boldsymbol{u} = \boldsymbol{u}_o$ on $\boldsymbol{r} \in S_u$ and $d \geq 0$ on $\boldsymbol{r} \in S_c$.

If we use the Hellinger-Reissner functional we have only kinematical constraints for contact conditions: $d \geq 0$ on $\boldsymbol{r} \in S_c$

Thus

$$\min \{\Pi(\boldsymbol{u}) \mid \quad \boldsymbol{u} = \boldsymbol{u}_o, \quad \boldsymbol{r} \in S_u, \quad d \geq 0, \quad \boldsymbol{r} \in S_c\} \tag{3.1}$$

and

$$\min \{\Pi_R(\boldsymbol{u}, \boldsymbol{T}) \mid \quad d \geq 0, \quad \boldsymbol{r} \in S_c\}\ , \tag{3.2}$$

which must be solved satisfying the variational inequality $\delta\Pi \geq 0$ or $\delta_u \Pi_R \geq 0$ and $\delta_T \Pi_R = 0$. The detailed mathematical discussion of this variational inequality and other variational principles can be found in Haslinger and Neittaanmaki [48], Oden and Kikuchi [46] and in Telega [47].

Here the total potential energy

$$\Pi(\boldsymbol{u}) = \sum_{e=1}^{2} \left\{ \frac{1}{2} \int\limits_{V^{(e)}} (\boldsymbol{A}(\boldsymbol{u}) - \boldsymbol{A}_o) \cdot\cdot \boldsymbol{D} \cdot\cdot (\boldsymbol{A}(\boldsymbol{u}) - \boldsymbol{A}_o) \, \mathrm{d}V + \int\limits_{V^{(e)}} \boldsymbol{A}(\boldsymbol{u}) \cdot\cdot \boldsymbol{T}_o \, \mathrm{d}V - \right.$$
$$\left. - \int\limits_{V^{(e)}} \boldsymbol{u} \cdot \boldsymbol{b} \, \mathrm{d}V - \int\limits_{S_t^{(e)}} \boldsymbol{u} \cdot \boldsymbol{t}_o \, \mathrm{d}S. \right\} \tag{3.3}$$

and the Hellinger-Reissner functional

$$\Pi_R(\boldsymbol{u}, \boldsymbol{T}) = \sum_{e=1}^{2} \left\{ \int\limits_{V^{(e)}} \left[\boldsymbol{T} \cdot\cdot (\boldsymbol{A}(\boldsymbol{u}) - \boldsymbol{A}_o) - \tfrac{1}{2} \boldsymbol{T} \cdot\cdot \boldsymbol{C} \cdot\cdot \boldsymbol{T} \right] \mathrm{d}V + \right.$$
$$\left. + \int\limits_{V^{(e)}} \boldsymbol{T} \cdot\cdot \boldsymbol{C} \cdot\cdot \boldsymbol{T}_o \, \mathrm{d}V - \int\limits_{V^{(e)}} \boldsymbol{u} \cdot \boldsymbol{b} \, \mathrm{d}V - \int\limits_{S_t^{(e)}} \boldsymbol{u} \cdot \boldsymbol{t}_o \, \mathrm{d}S. - \int\limits_{S_u^{(e)}} \boldsymbol{n} \cdot \boldsymbol{T} \cdot (\boldsymbol{u} - \boldsymbol{u}_o) \mathrm{d}S \right\}. \tag{3.4}$$

Here $\boldsymbol{C} = \boldsymbol{D}^{-1}$, and $()^{-1}$ denote the inverse of a tensor. In practice, instead of problem (3.1)and (3.2), functionals (3.3) and (3.4) are extended with penalty functions of the non-penetration contact constraint as

$$\mathcal{L}^{PE} = \mathcal{L}^{PE}(\boldsymbol{u}) = \Pi(\boldsymbol{u}) + \frac{1}{2} \int\limits_{S_c} c_n \left(d^{-}(\boldsymbol{u}) \right)^2 \mathrm{d}S \tag{3.5}$$

and

$$\mathcal{L}_R^{PE} = \mathcal{L}_R^{PE}(\boldsymbol{u}, \boldsymbol{T}) = \Pi_R(\boldsymbol{u}, \boldsymbol{T}) + \frac{1}{2} \int\limits_{S_c} c_n \left(d^{-}(\boldsymbol{u}) \right)^2 \mathrm{d}S, \tag{3.6}$$

where c_n is the penalty parameter, and d^{-} denotes the negative part of d. From the variational equation $\delta_u \mathcal{L}^{PE} = 0$ and $\delta_u \mathcal{L}_R^{PE} = 0$ we obtain a formula for the contact pressure p_n.

$$p_n = -c_n d^{-}(\boldsymbol{u}). \tag{3.7}$$

The higher the penalty parameter c_n, the smaller the violation of the non-penetration condition $\lim\limits_{c_n \to \infty} d^{-}(\boldsymbol{u}) = 0$, i.e., the non-penetration condition $d \geq 0$ is not strictly satisfied.

The correct choice of the penalty parameter is essential, because the condition number of the coefficient matrix increases as the penalty parameter increases. Using p-version finite elements [33], [42], $c_n \sim 100E - 1000E$ is recommended, where E is the Young's modulus.

Combining the Lagrangian and the penalty methods as it was proposed by Simo and Laursen in [26] (Simo and Laursen, 1992), the following augmented Lagrangian

functionals are obtained

$$\mathcal{L}^{AU} = \mathcal{L}^{AU}(\boldsymbol{u}) = \Pi(\boldsymbol{u}) - \int_C p_n d(\boldsymbol{u})\,\mathrm{d}S + \frac{1}{2}\int_C c_n \left(d(\boldsymbol{u})\right)^2 \mathrm{d}S, \tag{3.8}$$

and

$$\mathcal{L}_R^{AU} = \mathcal{L}_R^{AU}(\boldsymbol{u}, \boldsymbol{T}) = \Pi_R(\boldsymbol{u}, \boldsymbol{T}) - \int_C p_n d(\boldsymbol{u})\,\mathrm{d}S + \frac{1}{2}\int_C c_n \left(d(\boldsymbol{u})\right)^2 \mathrm{d}S, \tag{3.9}$$

where p_n is the Lagrangian multiplier, which is kept constant during an iteration loop. From the following two variational equations

$$\delta_u \mathcal{L}^{AU} = 0 \quad \text{or} \quad \delta_u \mathcal{L}_R^{AU} = 0 \tag{3.10}$$

the same formula is obtained for the normal contact stresses

$$\sigma_n^{(1)}\left(\boldsymbol{u}^1\right) = \sigma_n^{(2)}\left(\boldsymbol{u}^2\right) = -\left(p_n - c_n d(\boldsymbol{u})\right), \tag{3.11}$$

and

$$\sigma_n^{(1)}\left(\boldsymbol{T}^{(1)}\right) = \sigma_n^{(2)}\left(\boldsymbol{T}^{(2)}\right) = -\left(p_n - c_n d(\boldsymbol{u})\right). \tag{3.12}$$

During the iteration process, the contact pressure is updated using the formula

$$p_n^{(k)} = \left\langle p_n^{(k-1)} - c_n d\left(\boldsymbol{u}^{(k)}\right)\right\rangle, \tag{3.13}$$

where the operation $< \cdot >$ is defined by

$$\langle p_n \rangle = \frac{1}{2}\left(p_n + |p_n|\right). \tag{3.14}$$

In the $(k+1)th$ iteration loop the contact surface is subjected by $p_n^{(k)}$ as an external load in the variational formula:

$$\delta_u \mathcal{L}^{AU}\left(\boldsymbol{u}_{(k+1)}\right) = \delta\Pi\left(\boldsymbol{u}_{(k+1)}\right) - \int_{C^{(k)}} \delta d(\boldsymbol{u})\left(p_n^{(k)} - c_n d\left(\boldsymbol{u}_{(k+1)}\right)\right)\mathrm{d}S = 0 \tag{3.15}$$

and

$$\delta_u \mathcal{L}_R^{AU}\left(\boldsymbol{u}_{(k+1)}, \boldsymbol{T}_{(k+1)}\left(\boldsymbol{u}_{(k+1)}\right)\right) = \delta_u \Pi_R\left(\boldsymbol{u}_{(k+1)}, \boldsymbol{T}_{(k+1)}\left(\boldsymbol{u}_{(k+1)}\right)\right) - \int_{C^{(k)}} \delta d(\boldsymbol{u})\left(p_n^{(k)} - c_n d\left(\boldsymbol{u}_{(k+1)}\right)\right)\mathrm{d}S = 0$$

because of the next variational equation $\delta_T \mathcal{L}_R^{AU} = 0$ we are able to to gain the relation for $\boldsymbol{T}_{(k+1)} = \boldsymbol{T}_{(k+1)}\left(\boldsymbol{u}_{(k+1)}\right)$, i.e.

$$\boldsymbol{T}_{(k+1)}^{(e)}\left(\boldsymbol{u}_{(k+1)}^{(e)}\right) = \boldsymbol{T}_o^{(e)} + \boldsymbol{D}^{(e)} \cdot\cdot \left(\boldsymbol{A}^{(e)}\left(\boldsymbol{u}_{(k+1)}^{(e)}\right) - \boldsymbol{A}_o^{(e)}\right) \tag{3.16}$$

in the approximation process.

REMARK 1: If the second terms of the functionals (3.8 and 3.9) are vanished, we obtain two previous functionals of the penalty methods with $d(\boldsymbol{u}) = d^-(\boldsymbol{u}) \leq 0$ substitutions.

- REMARK 2: If the last terms of the functionals (3.8, 3.9) are omitted the funtionals of the Lagrangian multiplier techniques are received back. In these cases $\mathcal{L}^{LA} = \mathcal{L}^{LA}(\boldsymbol{u}, p_n \geq 0)$, $\mathcal{L}_R^{LA} = \mathcal{L}_R^{LA}(\boldsymbol{u}, \boldsymbol{T}, p_n \geq 0)$ for each functional variational equation or two equations and only one variational inequality are written:

$$\delta_u L^{LA} = 0, \qquad \delta_{p_n} L^{LA} \leq 0 \;\text{ or }\; \delta_u \mathcal{L}_R^{LA} = 0, \qquad \delta_T \mathcal{L}_R^{LA} = 0, \quad -\delta_{p_n} \mathcal{L}_R^{LA} \geq 0. \tag{3.17}$$

  In order to satisfy the contact conditions, positivity of the contact pressure must be ensured.

4. Finite element formulation

4.1. **Approximated fields.** The displacements of the contacting bodies are approximated in the usual form

$$\boldsymbol{u}^{(e)} = \boldsymbol{u}^{(e)}(\mathbf{x}) = \mathbf{N}^{(e)}(\mathbf{x})\,\mathbf{q}^{(e)}, \tag{4.1}$$

where the shape functions $\mathbf{N}^{(e)}(\mathbf{x})$ consist of nodal point modes, side modes and internal modes, and $\mathbf{q}^e$ is the vector of displacement parameters, $\mathbf{x}$ is the space coordinates [33]. The strain vector is given by the next formula:

$$\boldsymbol{A}^{(e)}(\boldsymbol{u}) \to \boldsymbol{\varepsilon}^{(e)} = \boldsymbol{\varepsilon}^{(e)}(\mathbf{x}) = \partial \mathbf{u}^{(e)} = \mathbf{B}^{(e)}(\mathbf{x})\,\mathbf{q}^{(e)}, \tag{4.2}$$

where $\mathbf{B}^e(\mathbf{x})$ is the strain-displacement matrix. The stress vector can be expressed as

$$\boldsymbol{T}^{(e)} \to \boldsymbol{\sigma}^{(e)} = \boldsymbol{\sigma}^{(e)}(\mathbf{x}) = \boldsymbol{\sigma}_o^{(e)}(\mathbf{x}) + \mathbf{D}^{(e)}(\mathbf{x})\left(\mathbf{B}^{(e)}(\mathbf{x})\,\mathbf{q}^{(e)} - \boldsymbol{\varepsilon}_o^{(e)}(\mathbf{x})\right), \tag{4.3}$$

for the minimum principle of total potential energy, where $\mathbf{D}^{(e)}(\mathbf{x})$ is the constitutive matrix, $\boldsymbol{\varepsilon}_o^{(e)}(\mathbf{x}), \boldsymbol{\sigma}_o^{(e)}(\mathbf{x})$ are the initial strain and stress vectors, respectively. Using (4.1)-(4.3) the total potential energy (3.3) can be written in discretized form:

$$\Pi^{(e)}(\boldsymbol{u}) \to \Pi^{(e)}\left(\mathbf{q}^{(e)}\right) = \frac{1}{2}\mathbf{q}^{(e),T}\left(\mathbf{K}^{(e)}\mathbf{q}^{(e)} - 2\mathbf{f}^{(e)}\right), \tag{4.4}$$

where

$$\mathbf{K}^{(e)} = \int\limits_{V^{(e)}} \mathbf{B}^{(e),T}\mathbf{D}^{(e)}\mathbf{B}^{(e)}\mathrm{d}V \tag{4.5}$$

is the element stiffness matrix, and

$$\mathbf{f}^{(e)} = \int\limits_{V^e} \mathbf{N}^{(e),T}\mathbf{b}^{(e)}\mathrm{d}V + \int\limits_{S_t^{(e)}} \mathbf{N}^{(e),T}\mathbf{t}_o\,\mathrm{d}S + \int\limits_{V^{(e)}} \mathbf{B}^{(e),T}\left(\mathbf{D}^{(e)}\boldsymbol{\varepsilon}_o^{(e)} - \boldsymbol{\sigma}_o^{(e)}\right)\mathrm{d}V \tag{4.6}$$

is the element load vector, T denotes the transpose of a matrix.

In the case of Hellinger-Reissner variational principle the stress is approximated by the next formula

$$\boldsymbol{T}^{(e)} \to \boldsymbol{\sigma}^{(e)} = \mathbf{F}^{(e)}(\mathbf{x})\;\boldsymbol{\beta}^{(e)}, \tag{4.7}$$

where $\mathbf{F}^{(e)}(\mathbf{x})$ is an approximation matrix which consists of the same functions that are used for $\mathbf{N}^{(e)}(\mathbf{x})$.

The vector $\boldsymbol{n}^{(e)} \cdot \boldsymbol{T}^{(e)}$ is approximated in the next form

$$\boldsymbol{n}^{(e)} \cdot \boldsymbol{T}^{(e)} \rightarrow \mathbf{R}^{(e)}(\mathbf{x})\ \boldsymbol{\beta}^{(e)}. \tag{4.8}$$

The discretized form of Hellinger-Reissner functional (3.4) has the next form:

$$\Pi_R^{(e)}(\boldsymbol{u},\boldsymbol{T}) \rightarrow \Pi_R^{(e)}\left(\mathbf{q}^{(e)},\boldsymbol{\beta}^{(e)}\right) = \\ = \boldsymbol{\beta}^{(e),T}\ \mathbf{S}^{(e)}\ \mathbf{q}^{(e)}\ - \frac{1}{2}\boldsymbol{\beta}^{(e),T}\ \widetilde{\mathbf{H}}^{(e)}\ \boldsymbol{\beta}^{(e)} - \mathbf{t}^{(e),T}\mathbf{q}^{(e)} + \ \boldsymbol{\beta}^{(e),T}\ \mathbf{g}^{(e)}, \tag{4.9}$$

where

$$\widetilde{\mathbf{H}}^{(e)} = \int\limits_{V^{(e)}} \mathbf{F}^T\ \mathbf{C}\ \mathbf{F}\ \mathrm{d}V, \quad \mathbf{S}^{(e)} = \int\limits_{V^{(e)}} \mathbf{F}^T\ \mathbf{B}\ \mathrm{d}V - \int\limits_{S_u^{(e)}} \mathbf{R}^T\ \mathbf{N}\ \mathrm{d}S, \tag{4.10a}$$

$$\mathbf{t}^{(e)} = \int\limits_{V^{(e)}} \mathbf{N}^T\ \mathbf{b}\ \mathrm{d}V\ + \int\limits_{S_t^{(e)}} \mathbf{N}^T\,\mathbf{t}_o\,\mathrm{d}S, \tag{4.10b}$$

$$\mathbf{g}^{(e)} = \int\limits_{S_u^{(e)}} \mathbf{R}^T\ \mathbf{u}_o\ \mathrm{d}S\ - \int\limits_{V^{(e)}} \mathbf{F}^T\ (\boldsymbol{\varepsilon}_o\ -\ \mathbf{C}\boldsymbol{\sigma}_o)\ \mathrm{d}V. \tag{4.10c}$$

From variational equation $\delta_T \Pi_R^{(e)} = 0 \rightarrow \delta_\beta \Pi_R^{(e)} = 0 = \delta\boldsymbol{\beta}^{(e),T}\ \frac{\partial \Pi_R^{(e)}}{\partial \boldsymbol{\beta}^{(e)}} \rightarrow \frac{\partial \Pi_R^{(e)}}{\partial \boldsymbol{\beta}^{(e)}} = \mathbf{0}$ the stress parameter's vector is

$$\boldsymbol{\beta}^{(e)} = \left(\widetilde{\mathbf{H}}^{(e)}\right)^{-1}\ \mathbf{S}^{(e)}\mathbf{q}^{(e)} + \left(\widetilde{\mathbf{H}}^{(e)}\right)^{-1}\ \mathbf{g}^{(e)} \tag{4.11}$$

and the reduced discretized form of $\Pi_R^{(e)}$ can be written as follows

$$\Pi_R^{(e)}(\boldsymbol{u}) \rightarrow \Pi_R^{(e)}\left(\mathbf{q}^{(e)}\right) = \frac{1}{2}\mathbf{q}^{(e),T}\left(\mathbf{K}^{(e)}\mathbf{q}^{(e)} - 2\mathbf{f}^{(e)}\right), \tag{4.12}$$

where the element stiffness matrix

$$\mathbf{K}^{(e)} = \mathbf{S}^{(e),T}\left(\widetilde{\mathbf{H}}^{(e)}\right)^{-1}\mathbf{S}^{(e)}, \tag{4.13}$$

and the load vector

$$\mathbf{f}^{(e),T} = \mathbf{t}^{(e),T} - \mathbf{g}^{(e),T}\left(\widetilde{\mathbf{H}}^{(e)}\right)^{-1}\mathbf{S}^{(e)}. \tag{4.14}$$

The gap after deformation is computed by the following projection

$$d = u_n^{(2)} - u_n^{(1)} + h = -\mathbf{L}^{(1)}(\mathbf{x})\mathbf{q}^{(1)} + \mathbf{L}^{(2)}(\mathbf{x})\mathbf{q}^{(2)} + h = \mathbf{L}(\mathbf{x})\mathbf{q} + h, \tag{4.15}$$

where the matrix of shape functions $\mathbf{L}^{(e)}(\mathbf{x})$ is constructed by the use of $\mathbf{N}^{(e)}(\mathbf{x})$ with $x \in S_c$ and the definition of the normal displacement $u_n^{(e)}$. The vector of displacement parameters for the whole system is given as $\mathbf{q}^T = \left[\mathbf{q}^{(1),T}\quad \mathbf{q}^{(2),T}\right]$.

The same penalty terms in (3.8, 3.9) can be written in discretized form

$$\begin{aligned}\tfrac{1}{2}\int\limits_C c_n \left(d\left(\boldsymbol{u}\right)\right)^2 \mathrm{d}S = \tfrac{1}{2}\mathbf{q}^T \left\{ \int\limits_C \begin{bmatrix} -\mathbf{L}^{(1),T} \\ \mathbf{L}^{(2),T} \end{bmatrix} c_n \begin{bmatrix} -\mathbf{L}^{(1)} & \mathbf{L}^{(2)} \end{bmatrix} \mathrm{d}S \; \mathbf{q} + \right. \\ \left. +2\int\limits_C c_n \begin{bmatrix} -\mathbf{L}^{(1),T} \\ \mathbf{L}^{(2),T} \end{bmatrix} h \, \mathrm{d}S \right\} + const = \tfrac{1}{2}\mathbf{q}^T \begin{bmatrix} \widetilde{\mathbf{C}}^{11} & -\widetilde{\mathbf{C}}^{12} \\ -\widetilde{\mathbf{C}}^{21} & \widetilde{\mathbf{C}}^{22} \end{bmatrix} \mathbf{q} + \\ +\mathbf{q}^T \begin{bmatrix} -\mathbf{f}_h^{(1)} \\ \mathbf{f}_h^{(2)} \end{bmatrix} + const = \tfrac{1}{2}\mathbf{q}^T \widetilde{\mathbf{C}}\mathbf{q} + \mathbf{q}^T \mathbf{f}_h + const,\end{aligned} \tag{4.16}$$

where $\widetilde{\mathbf{C}}$ is the contact stiffness matrix.

From the same Lagrangian terms in (3.8, 3.9) the load vector $\mathbf{f}_p^e$ associated to the contact pressure is

$$\int\limits_C d\left(\mathbf{u}\right) p_n \, \mathrm{d}S \to \mathbf{q}^T \int\limits_C \begin{bmatrix} -\mathbf{L}^{(1),T} \\ \mathbf{L}^{(2),T} \end{bmatrix} p_n \, \mathrm{d}S = \mathbf{q}^T \begin{bmatrix} -\mathbf{f}_p^{(1)} \\ \mathbf{f}_p^{(2)} \end{bmatrix} = \mathbf{q}^T \mathbf{f}_p \,. \tag{4.17}$$

4.2. **Discretized functionals.** Finally, the discretized funtionals of $\mathcal{L}^{AU}$ and $\mathcal{L}_R^{AU}$ expressed with nodal displacements are written in the same form as

$$\mathcal{L}^{AU} = \mathcal{L}^{AU}\left(\mathbf{q}\right) = \sum_e \left\{ \frac{1}{2}\mathbf{q}^{(e),T}\mathbf{K}^{(e)}\mathbf{q}^{(e)} - \mathbf{q}^{(e),T}\mathbf{f}^{(e)} \right\} + \frac{1}{2}\mathbf{q}^T \widetilde{\mathbf{C}}\mathbf{q} + \mathbf{q}^T \left(\mathbf{f}_h - \mathbf{f}_p\right). \tag{4.18}$$

The algebraic system of equation associated with (4.18) has the following form

$$\begin{bmatrix} \mathbf{K}^1 + \widetilde{\mathbf{C}}^{11} & -\widetilde{\mathbf{C}}^{12} \\ -\widetilde{\mathbf{C}}^{21} & \mathbf{K}^2 + \widetilde{\mathbf{C}}^{22} \end{bmatrix}^{(k)} \begin{bmatrix} \mathbf{q}^{(1)} \\ \mathbf{q}^{(2)} \end{bmatrix}_{(k+1)} = \begin{bmatrix} \mathbf{f}^{(1)} + \mathbf{f}_h^{(1),(k)} - \mathbf{f}_p^{(1),(k)} \\ \mathbf{f}^{(2)} - \mathbf{f}_h^{(2),(k)} + \mathbf{f}_p^{(2),(k)} \end{bmatrix}, \tag{4.19}$$

in which the matrix $\widetilde{\mathbf{C}}^{ij}$ is modified to fulfill the contact/separation conditions.

The iterational KALKER procedure [49], [37] with the control of the sign of p_n can be applied for solving (4.19). The contact conditions are checked in the Gauss or Lobatto integration points of the contact elements during the solution of (4.19). With the updated contact pressure $p_n^{(k+1)}$ the integrals (4.16) and (4.17) can be computed repeatedly, that is we have a new penalty (contact) matrix $\widetilde{\mathbf{C}}$ and new vectors $\mathbf{f}_h$ and $\mathbf{f}_p$. The $(k+1)th$ displacements are obtained from the solution of (4.19). The procedure is terminated when the following condition is fulfilled:

$$\frac{\int\limits_{S_c} \left| p_n^{(k+1)} - p_n^{(k)} \right| \mathrm{d}S}{\int\limits_{S_c} p_n^{(k+1)} \mathrm{d}S} \leq 10^{-4}. \tag{4.20}$$

Remarks:

1. Since $d(\boldsymbol{u})$ is computed in the local coordinate system, the elements which have boundaries on the contact surface, must be transformed from global coordinate system to the local one. The transformation is performed by least squares

fitting. A detailed explanation of this problem and the numerical reasoning of the transformation can be found in [42].

2. When the p-version is used then accuracy is typically high enough for singularities to induce oscillations in the numerical solutions. The oscillations are minimized when nodes are located at the boundary of the contact zone. The idea of the nodal positioning technique was first published in [42].
3. The system of inequalities according to Lagrangian technique is a Linear Complementary Problem, which can be solved by different algorithms as given in e.g. [11], [21] and [49].
4. In the work [18] a two-level algorithm is employed for solution of the contact problem using Lagrangian multipliers.

5. Positioning the nodal points

One of the advantages of the p-version is that for smooth problems only coarse meshes are needed, since the error in energy norm decreases exponentially when the polynomial degree of elements is increased [33] (Szabó and Babuska, 1991). At the border of the contact zone we have a singularity in the stress state. When the ends of contact zone C are not situated in nodal points (in 2D case) then the derivatives of the shape functions cannot have the appropriate jumps there, that is, the singularities induce oscillations in the pressure distribution. By positioning (moving) the nodal points to the ends of contact zone C, the jump in the derivatives can be represented in the discretized problem.

The positioning algorithm had been developed for contact problems of a two-dimensional problem [42] (Páczelt et.al., 1999). The positioning of the nodal points is performed in one or two phases, depending on the predefined tolerance. The first phase is a rough positioning of contact point to ensure that contact exists in each integration point of the contacting elements. In the second phase the border points are moved based on error indicators in order to increase the accuracy.

5.1. **The rough positioning technique.** We supposed that the contact zone of two axsymmetric bodies is situated on the left side of the possible contact region. Let r_i, r_j and r_k denote the radial coordinates of the nodal points of the contact element in the original mesh (see Figure 3). We search for the contact (that is $p_n > 0$) from the right-hand direction. The contact elements are integrated in N Gauss or Lobatto points. The star and triangular markers denote, respectively, the N^{th} integration point and the first contacting integration point from the right side, which is denoted by IC. The gap zone is denoted by a single line, the parallel double lines denote the compressed Winkler spring contact zone. The $(\widetilde{r})$ denotes the modified radial coordinate. The natural coordinate of the contact element is denoted by ξ.

Checking the contact conditions in integration points is started from the right-hand side. The first contact point r_{IC} is detected in the positive range of the standard coordinate. The location of the new nodal point is found by linear mapping (Step 3).

The algorithm has the following steps:

Step 1. The contact problem is solved with the given mesh.

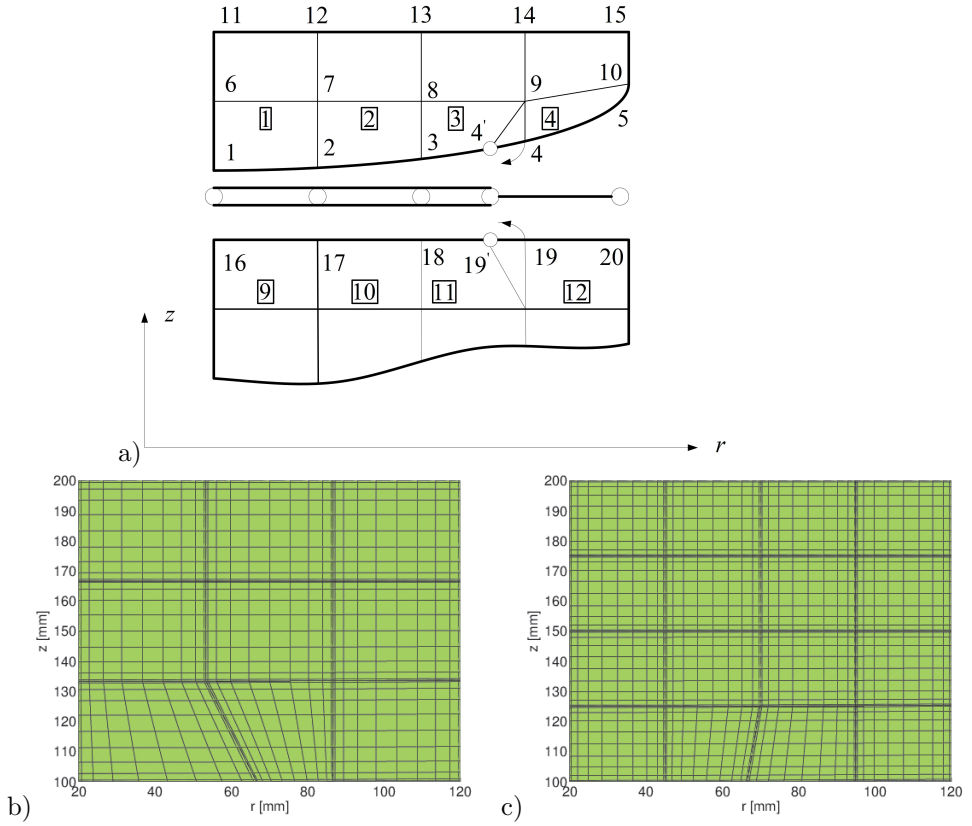


Figure 2. a) nodal points 4 and 19 are moved to $4'$ and $19'$, elements $1-3$ and $9-11$ are in contact, and between elements 4 and 12 there is no contact; b) modified (3×3) mesh c) modified (4×4) mesh.

Step 2. Search to find the end points of contact region. We seek the first contacting integration point located in r_{IC}, where the contact pressure p_n is positive.

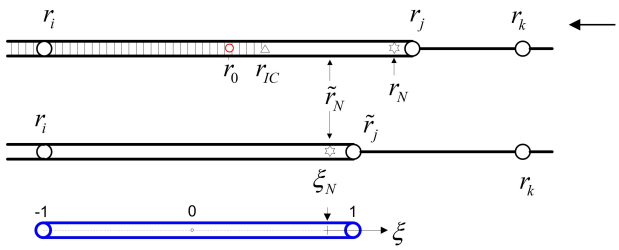
Step 3 The position of the new end point is calculated by linear mapping based on the position r_N of the new N^{th} integration point:

$$\widetilde{r}_j = \frac{[2\widetilde{r}_N - r_j\,(1-\xi_N)]}{(1+\xi_N)}, \tag{5.1}$$

where $\widetilde{r}_N$ is computed by the use of the bisection method:

$$\widetilde{r}_N = \begin{cases} \frac{1}{2}\,(r_{IC} + r_N) & if \quad \xi_{IC} \geq 0 \\ \frac{1}{2}\,(r_{IC} + r_\circ) & if \quad \xi_{IC} < 0 \end{cases} \tag{5.2}$$

The iteration Steps 1-3 are repeated unless each integration point of the contacting elements is in contact. As a result of this iteration process we have a new mesh, which will be called "the first correct mesh".



$$\tilde{r}_N = (r_{IC} + r_N)/2 \qquad \tilde{r}_j = (2\tilde{r}_N - r_i(1-\xi_N))/(1+\xi_N) \Rightarrow r_p$$

Figure 3. Schematic representation of the right hand side of the contact zone C. Calculation of the new position of the point j.

5.2. **The fine positioning technique.** After finding the first "correct mesh" we can further adjust the border of the contact zone C. The following indicators are introduced to monitor the fine adjustment of the nodal points:

- contact pressure at the contact border point (zero value is sought)
$$I_1 = \min |c_n \cdot d(r_j)| \tag{5.3}$$
where r_j is the coordinate of the computed contact border point,
- minimum of the potential energy
$$I_2 = \min \Pi_P \tag{5.4}$$
- minimum of the differences between contact pressure and normal stresses
$$I_3 = \min \left[\int \left\{ \left(\sigma_n^{(1)} + p_n\right)^2 + \left(\sigma_n^{(2)} + p_n\right)^2 \right\} \mathrm{d}S \right]^{0.5} \tag{5.5}$$
- minimum of shear stress on S_c
$$I_4 = \min \left[\int \left\{ \left(\tau_{rz}^{(1)}\right)^2 + \left(\tau_{rz}^{(2)}\right)^2 \right\} \mathrm{d}S \right]^{0.5}, \tag{5.6}$$
where the upper and lower bodies are denoted by superscripts 1 and 2, respectively.

In order to minimize the indicators the nodal point is positioned at the vicinity of the border of contact zone determined by the rough positioning technique. That is, the nodal point is moved either to the direction of the N^{th} integration point of the contacting element, or to the direction of the 1^{st} integration point of the gap element. The minimum of the indicator is searched within the interval determined by these two integration points.

5.3. **Position technique assuming friction.** The solution of the frictional contact problem is demonstrated in the middle of Figure 4. The end of the contact region is situated between elements 4_e and 8_e. The border of the stick-slip zones is between elements 2_e and 6_e. In the point P_p the normal stress σ_z is continuous but the derivatives on the right and left sides are not equal $\left(\frac{d\sigma_z}{dr}\right)^- \neq \left(\frac{d\sigma_z}{dr}\right)^+$.

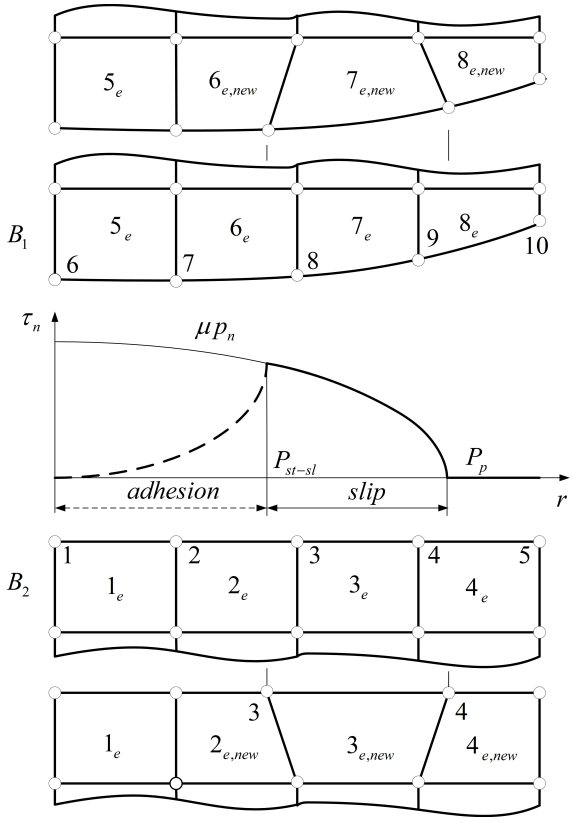


Figure 4. Modification of the mesh at friction contact problem

Since the approximated function is continuous along element edge and the derivatives are also continuous up to order $p-1$ the discrepancy between the derivatives is not represented, which induces oscillations in the results. Therefore new positions for the nodal points 4 and 9 shoul be found. A similar situation takes place for shear stresses in the point P_{st-sl}, which also requires mesh modifications. That is, nodal points 3 and 8 must be moved to left direction. After these positioning steps the elements $2_e, 3_e, 4_e$ and $6_e, 7_e, 8_e$ have new shapes $2_{e,new}, 3_{e,new}, 4_{e,new}$ and $6_{e,new}, 7_{e,new}, 8_{e,new}$, as shown in Figure 4.

Steps of the algorithm developed for friction contact problems are published in [56]. The shear stresses in the contact zone must satisfy the condition given for the slip surface $\Phi = |\tau_{rz}| - \mu p_n \leq 0$, which is performed with a predictor-corrector iterative algorithm. In the course of the time integration process the predicted value of the shear stress which is proportional to $c_\tau \Delta u_\tau$ must be calculated, where displacement in tangential direction is given as: $\Delta u_\tau = u_\tau^{t+\Delta t} - u_\tau^t$, at time t, with time step Δt. The number of integration points along an element side is denoted by *Nlobatto*. For the example let us suppose that on the left side of the element there is a stick zone, i.e.,

for integration Lobatto points ($ilobatto$=1,..., $istick$) the condition of slip surface is satisfied: $\Phi^{pred}(ilobatto) \leq 0$. However, for the rest of the integration points the condition of the slip surface is not satisfied: $\Phi^{pred}(ilobatto) > 0$, ($ilobatto$ = $istick$+1,..., $Nlobatto$). A return mapping algorithm is applied to satisfy the slip surface condition. This means that the right node point of the element must be moved to the left direction. At the end of iteration in each integration point of element $2_{e,new}$ there is stick: ($ilobatto$= 1,..., $Nlobatto$), i.e., the whole element side is in stick state.

6. Examples

Let us consider two contacting cylindrical bodies with equal dimensions (inner diameter $D_i = 2r_i = 40\,\text{mm}$ and outer diameter $D_u = 2r_u = 240\,\text{mm}$ and height $h_{cyl} = 100\,\text{mm}$, see Fig. 5. The meridian curve of the contacting surface of the upper body is given by the following function

$$h = 0.0004(r - 20)^2 \tag{6.1}$$

which is practically the function of the initial gap between the contacting bodies. There is no friction between the contacting bodies. The displacement is prescribed

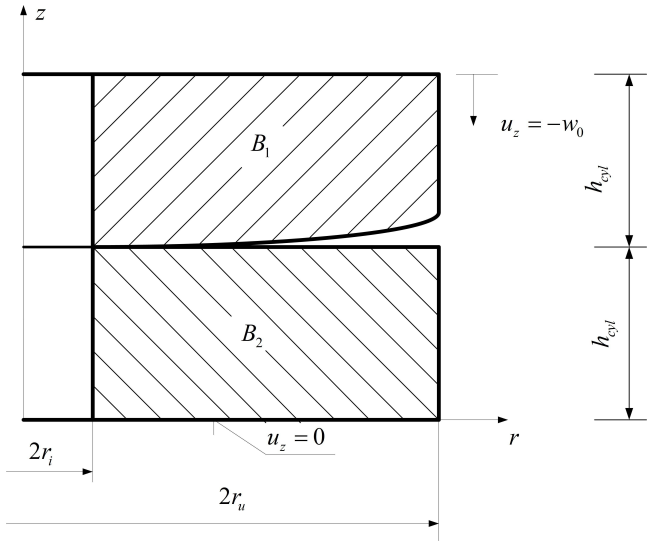


Figure 5. Two cylinders in contact. Initial gap $h = 0.0004(r - 20)^2$

on the top of the upper body in vertical direction $w_0 = 0.15\,\text{mm}$. Elastic material properties are given by Young's modulus $E = 200\,000\,\text{MPa}$ and Poisson's ratio $\nu = 0.3$. The boundary of the contact zone for body B_1 is defined by the blending function method [33].

The contact problem was solved by h-version, p-version and hp (rp)-version techniques using the error indicator I_1 for finding the radial coordinate of the contact border r_p.

The contact problems are solved as detailed below. The positioning technique introduced the previous section is used in all of the examples. The contact conditions are checked in the Gauss integration points. The penalty parameter is chosen as $c_n = 1\,000E$. The analyzed tasks are identified by the type of the applied mesh (number of elements in `columns` $\times$ `rows`).

Firstly, the problem is discretized with classical quadratic (low polynomial order) elements, i.e. h-version finite element method is used.

6.1. **Task 1.** For solving the contact problem the Hellinger-Reissner variational principle is applied using penalty technique. For the sake of convenience, instead of using the name Hellinger-Reissner variational hereafter. The polynomial order $p = 2$ is fixed, and the problem is solved by different meshes 4×4, 8×8, 16×16 and 32×32.

Figure 6a shows the final status of the mesh 8×8, the mesh also contains parametric lines crossing on the Gauss integration points. The distribution of the normal stress σ_z for body B_1 is given in Figure 3b. Figure 6c shows the results obtained for meshes 8×8, 16×16 and 32×32. Due to mesh refinement, the contact pressure p_n becomes smooth together with normal stress σ_z, but in the separation domain $r \geq r_p$ defined by the p_n, significant normal stresses are present, though in principle it should be zero. For mesh 8×8, the absolute value of the σ_z is higher than 10 MPa, while for the mesh 32×32 it does not fall below 7 MPa. These stress peaks occur at the end of the contact domain in oscillating form. Such oscillations can be experienced for each subsequent solution, only its extent will vary significantly. The smallest oscillation amplitude is obtained by the *rp*-version technique. It can also be observed that the differences in the contact stresses between the upper and lower bodies are very small.

Convergence diagrams of strain energy and contact border radius r_p calculated for polynomial degree $p = 2$ and $p = 3$ are shown in Figure 7. The change in strain energy is well converged (see Figure 7a) but the convergence of contact border radius is a bit slower. In these figures, we refer to converged results with Mesh 7×5 Type: `Imesh` $= 3$, which have been achieved with far fewer unknowns. The number of unknowns is 12 674 for mesh 32×32 with $p = 2$, and it is 4 474 for mesh 7×5 (Type: `Imesh` $= 3$ with $p = 8$). It is well known [33] that the speed of convergence for h-version is twice as slow as for p-version, however, the rate of convergence is exponential for $hp(rp)$-version computations.

6.2. **Task 2.** Here, we investigate the contact solution as a function of polynomial degree p. Our aim is to determine the position of the contact border as accurate as possible, i.e. the convergence of the solutions is examined.

The border points of the contact zone are determined by the nodal positioning technique detailed in Section 5 using error indicator I_1.

Augmented Lagrangian method is applied for solving these contact problems.

In this case computations are performed by p-extension finite elements. Here, the meshes of the contacting bodies are fixed, but the polynomial order p is increased from 2 to 8. The solution can be obtained with the positioning technique introduced in the previous section. We use rectangular elements with truncated space [33].

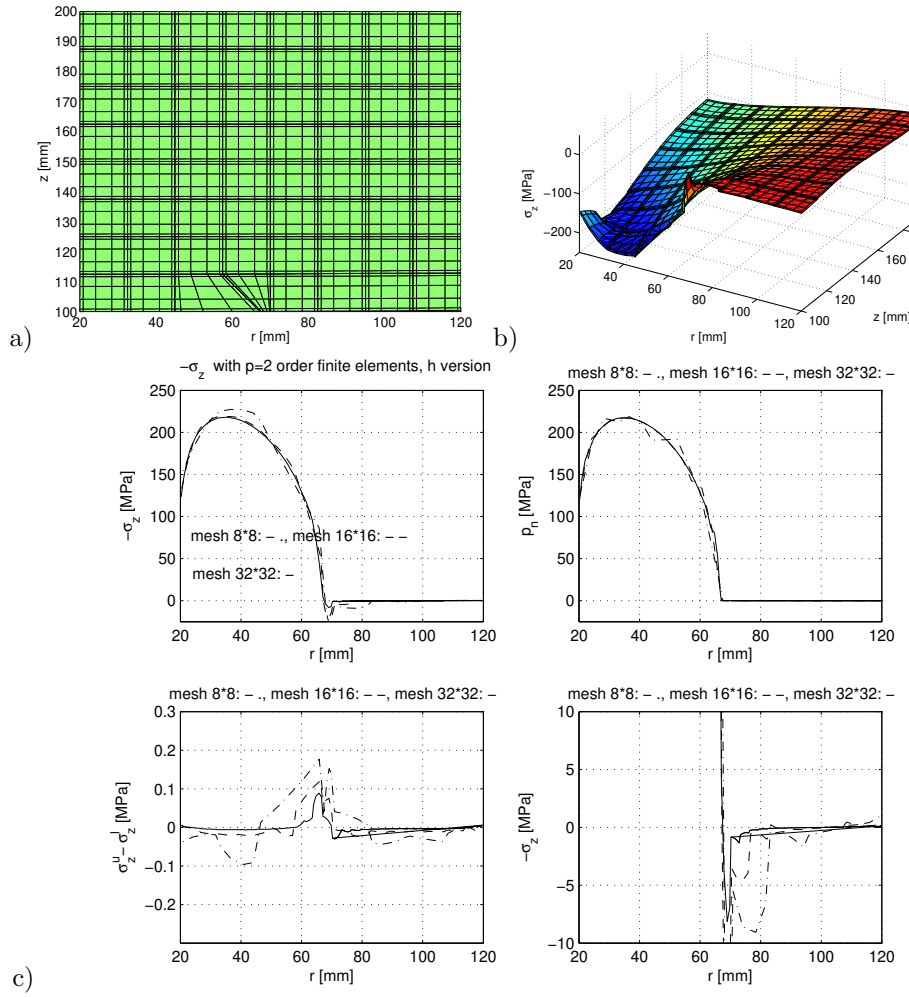


Figure 6. Task 1: solution with h-version, polynomial order for the displacements and stresses is $p = 2$; a) mesh 8×8; b) normal stress σ_z in the punch (upper body); c) distributions of normal stress σ_z, contact pressure p_n and differences between the normal stresses of the upper (σ_z^u) and lower (σ_z^l) bodies for different meshes

The Reissner functional is applied and the same degree of polynomials are used for the approxination of the displacements and stresses: i.e. `npu` = `nps` = p. The fixed mesh contains 7 columns and 5 rows (mesh 7 ×5). Figure 8a–g illustrate the results with polynomial order $p = 8$ for the different stress components.

Figure 9 shows different components of the results: the normal stress σ_z, the contact pressures p_n, the difference between the normal stresses σ_z^u and σ_z^l (computed on the

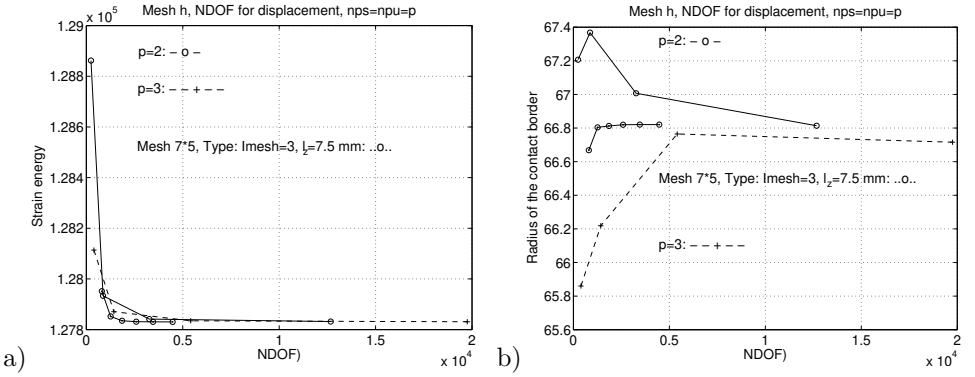


Figure 7. Convergence diagrams for h-version calculation with $p = 2$ and $p = 3$ (mesh 4×4, 8×8, 16×16, 32×32) and mesh 7×5 (identified by Type: `Imesh` $= 3$, $l_z = 7.5\,\mathrm{mm}$); a) for strain energy in Nmm; b) for contact border radius r_p in mm

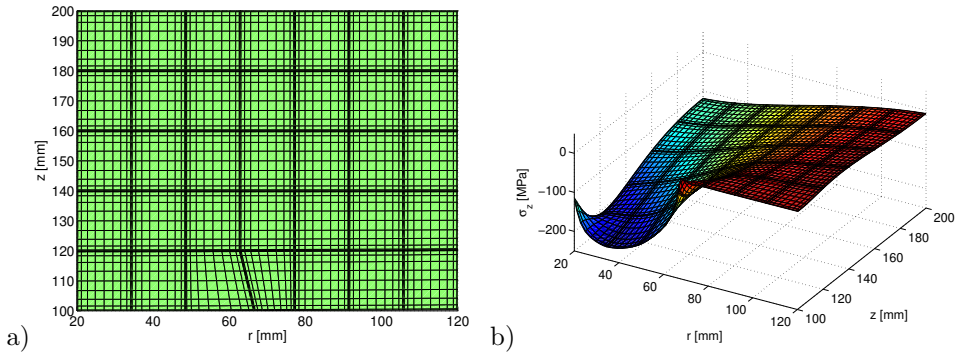


Figure 8. Solution of contact problem (Task 2 for body B_1) with uniform mesh 7×5, `Imesh` $= 0$: a) Final mesh after positioning, $r_p = 66.6168\,\mathrm{mm}$; b) vertical normal stresses

upper and the lower bodies) with polynomial order $p = 8$, and the distribution of σ_z^u as a function of $r \geq r_p$ with different polynomial orders ($p = 2$, $p = 6$ and $p = 8$).

Oscillations can also be observed here, but it much smoother than in Task 1, where only h-extension elements were used (compare to Figure. 6a). In the zone of gap (a certain distance from the border of the contact zone) the boundary condition $\sigma_z^u = 0$ is satisfied better with p-extension elements than with h-extension ones (compare Figure 6 to Figure 9).

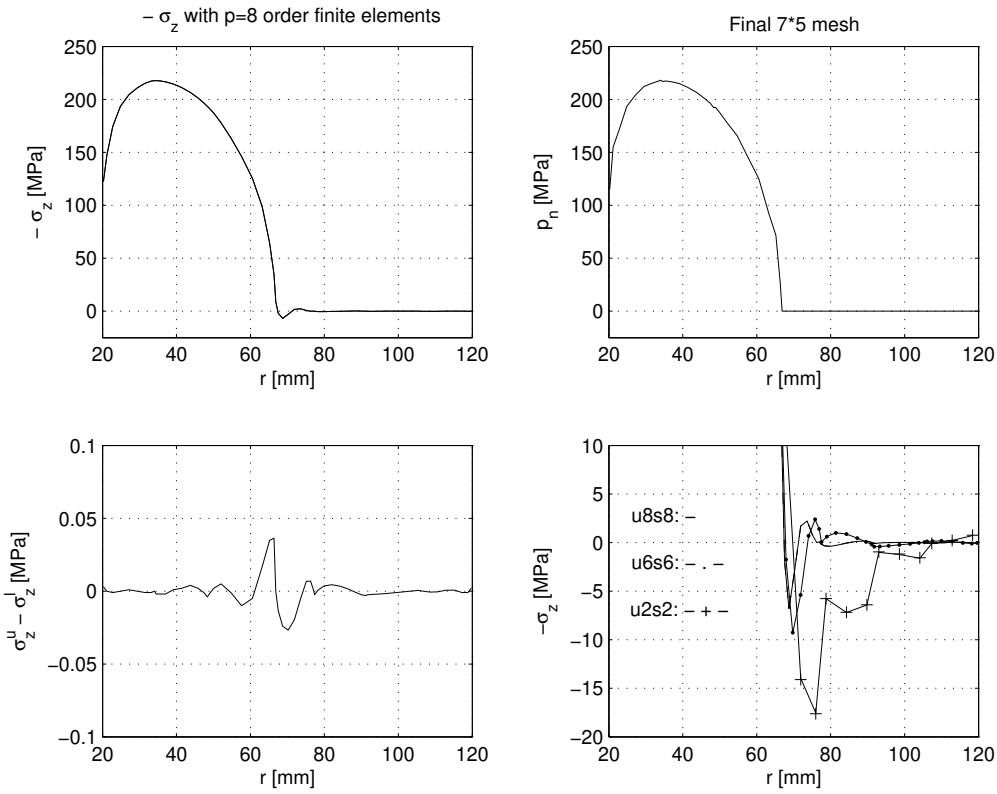


Figure 9. Solution of the contact problem with mesh type `Imesh` = 0, computed by using different polynomial orders: $p = 2(u2s2)$, $6(u6s6)$ and $8(u8s8)$

6.3. **Task 3.** Choosing the appropriate finite element mesh is an essential step to achieve high accuracy. In the mesh type `Imesh` = 0 there are equivalent elements. In the case of `Imesh` = 1 the initial `Imesh` = 0 is modified, i.e. near the border of the contact zone smaller elements are generated because the border point is a singular stress point.

The mesh (7×5) is applied with narrow elements on the right and left sides of the contact border point, i.e. single element layers are used on both sides of the border point in horizontal direction. The radial coordinate of the border point is r_p. The outer radius of the cylinder r_u and the length of the separation zone is $l_{pu} = r_u - r_p$. The sizes of the elements around of border point are $\Delta_{r1} = l_{pu} f^2$, $\Delta_{r2} = l_{pu} f$ where $f = 0.125$.

In vertical direction the size of the elements is increased gradually starting from the contacting elements and given by the following values for body B_1. In vertical direction there is a distance l_z and the sizes of the elements are $\Delta_{z1} = l_z f^2$, $\Delta_{z2} = l_z f$ and in this z direction the border of the next element is at l_z. The mesh for body B_2 is symmetrical to the x-axis.

Because the position r_p is determined with iteration, the length $l_{pu} = r_u - r_p$ can change. The iteration is finished when the contact pressure at the border point reaches its minimun value $I_1 = \min |c_n d(u)| \leq 0.0001$. In the computations we have three different values for $l_z = 7.5$, 15 and 20 mm.

`Imesh` = 1: Mesh (7×5) is used with single element layer in horizontal direction on the left and the right hand side of the border point ($\Delta_{rl} = l_{pu} f^2$).

`Imesh` = 2: Mesh (7×5) is applied with single element layer in horizontal direction on the left hand side of the border point ($\Delta_{rl} = l_{pu} f^2$), but on the right hand side of the border point the mesh is finer because of the double element layer ($\Delta_{r1} = l_{pu} f^2$, $\Delta_{r2} = l_{pu} f$).

`Imesh` = 3: Mesh (7×5) is used with double element layer on both the left and right hand side of the contact border point ($\Delta_{r1} = l_{pu} f^2$, $\Delta_{r2} = l_{pu} f$).

Firstly, let us compare the simulations of `Imesh` = 1 and `Imesh` = 2. The results are practically the same, see Figure 10. In Figure 10c the change of mesh 7×5 is shown, according to the *rp*-version of the finite element method. The initially uniform mesh 7×5 is modified to (Figure 10c) and then the final mesh (Figure 10c) was obtained after 10 iteration steps.

Secondly, the solution of the contact problem with `Imesh` = 3 is visualized by Figure 11. In Figure 12 one can see the results with original mesh (i.e. without positioning), and Figure 13 demonstrates the results using the node positioning (i.e. using the *rp*-extension method). It can also be realized that the oscillation of normal stress σ_z at the border point of contact zone is very small compared to simulation results received by *p*-extension finite elements (`Imesh` = 0), see Figure 14.

It is interesting that the distribution of the normal stress σ_z in the contact zone is smooth enough even in the case of using the original mesh. The discrepancy between the contact pressure and normal stress theoretically should be equal to zero, in this aspect the positioning technique gives better results, as is shown in Figure 12 and Figure 13.

Figure 15 shows different convergence diagrams: Figure 15a and 15b demonstrate the convergence of the position of contact border (r_p). It is clear that the *rp*-version of the computations is better than the *p*-version one.

We define the following error indicators using equations (5.5) and (5.6)

$$I_{3u} = I_3(\boldsymbol{u}), \quad I_{3s} = I_3(\boldsymbol{T}), \tag{6.2}$$

$$I_{4u} = I_4(\boldsymbol{u}), \quad I_{4s} = I_4(\boldsymbol{T}), \tag{6.3}$$

$$(\mathtt{Error}_F)_u = \frac{100 I_{3u}}{F_z}, \qquad (\mathtt{Error}_F)_s = \frac{100 I_{3s}}{F_z}, \tag{6.4}$$

where the resultant contact force is

$$F_z = -\int_{S_c} \sigma_n^{(1)} \, \mathrm{d}S, \qquad \sigma_n^{(i)} = \boldsymbol{n}^{(i)} \cdot \boldsymbol{T}^{(i)} \cdot \boldsymbol{n}^{(i)} \, . \tag{6.5}$$

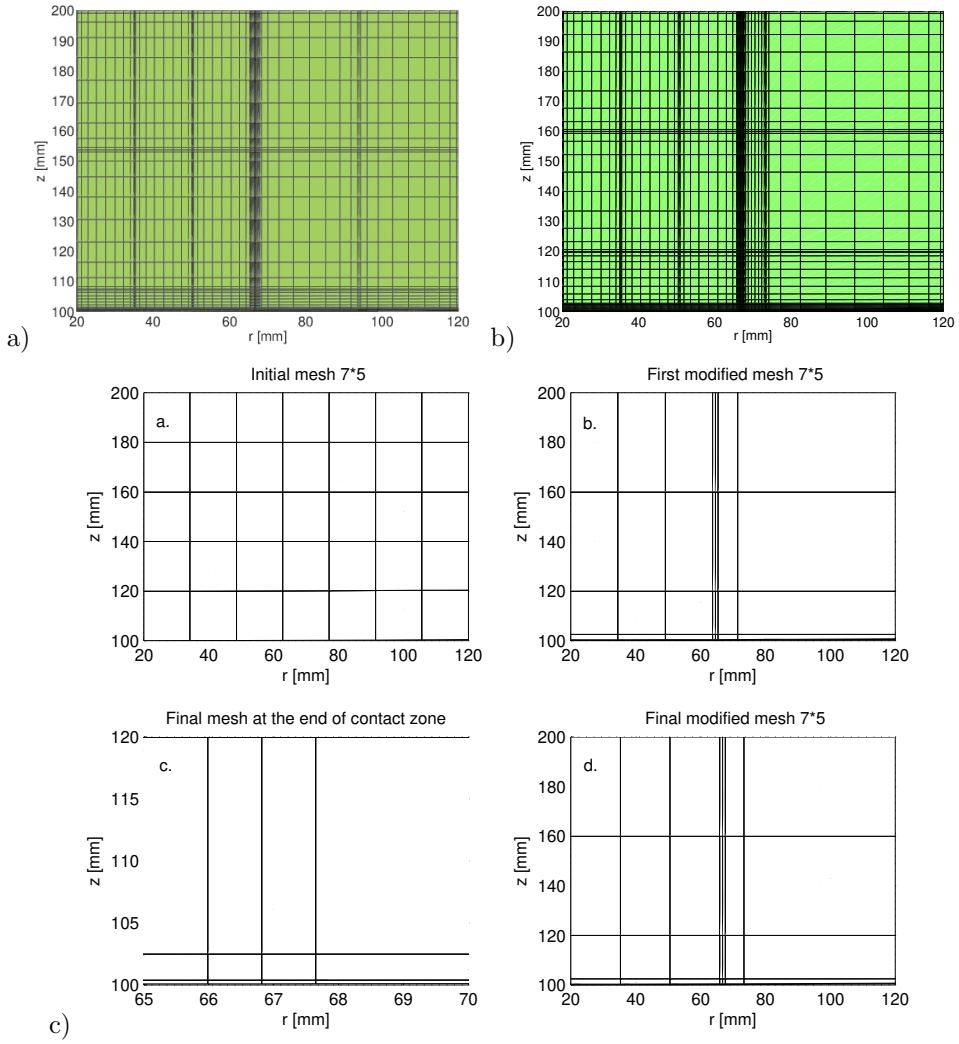


Figure 10. Solution with modified `Imesh` = 1, a) $l_z = 7.5\,\mathrm{mm}$, $r_p = 66.8220\,\mathrm{mm}$; b, c) `Imesh` = 2, $l_z = 20.0\,\mathrm{mm}$, $r_p = 66.8197\,\mathrm{mm}$

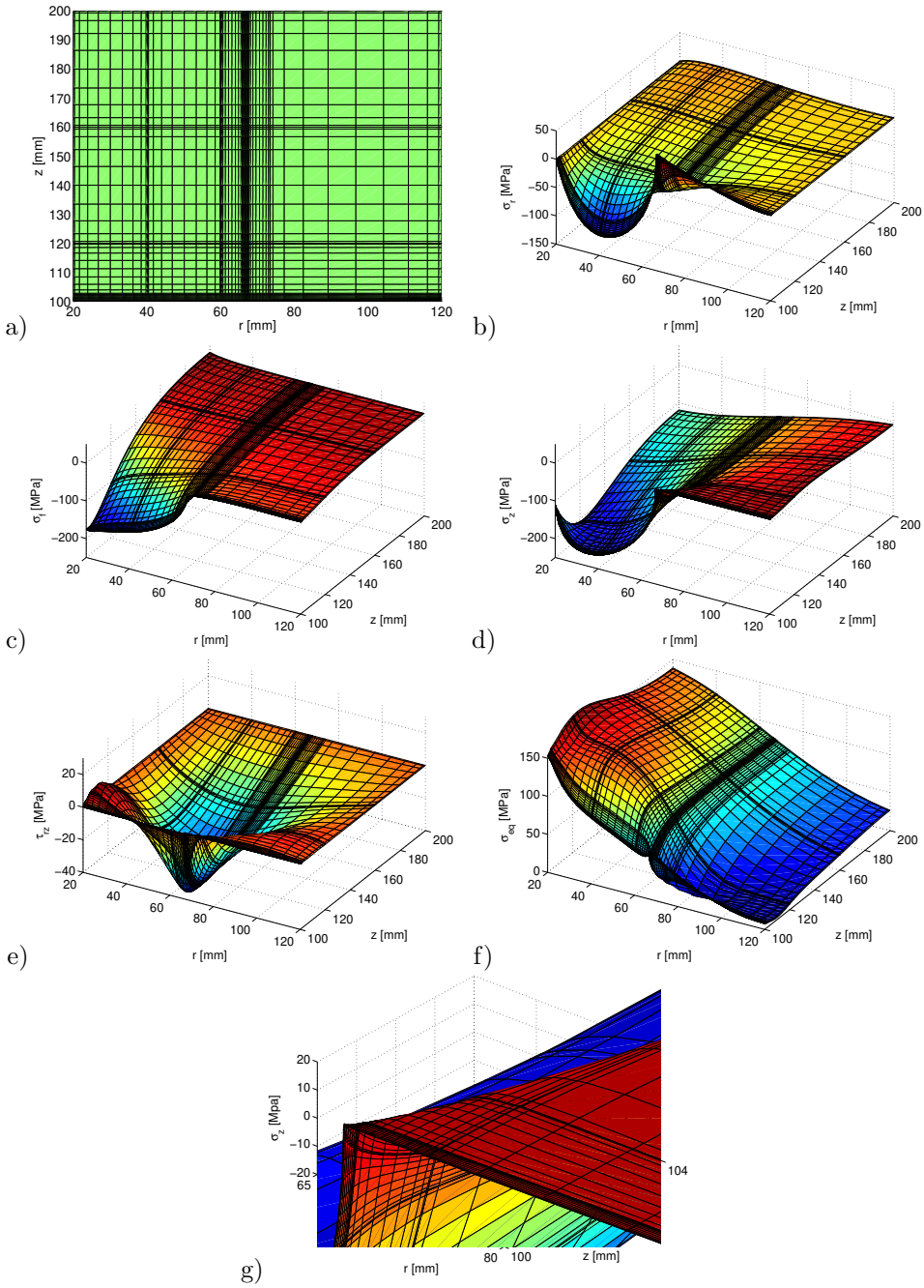


Figure 11. Solution of contact problem (Task 3 for body B_1) with small elements around contact border mesh (7×5) (`Imesh`$= 3$, $l_z = 20\,\text{mm}$, $r_p = 66.8262\,\text{mm}$): a) Final mesh after positioning; b)–d) radial, circumferential and vertical normal stresses; e) shear stress; f) Mieses equivalent stress; g) normal stress σ_z near the border of the contact zone. The polynomial order for displacements and stresses is $p = 8$

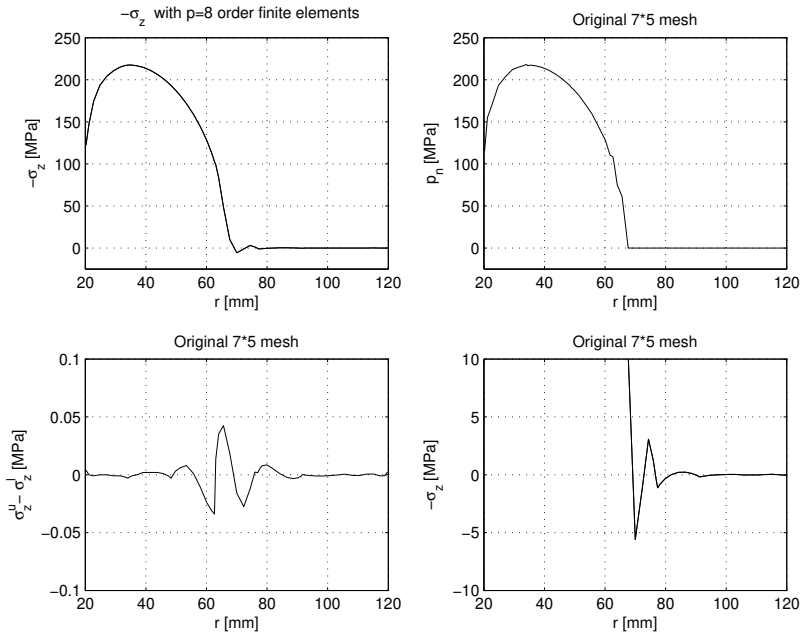


Figure 12. Contact stress distribution using initial mesh (contact pressure has large oscillation around the border of contact zone).

Figure 15c and 15d provide a good comparison of the results obtained for the resultant force F_z. The *rp*-version has very good convergence. In these diagrams the results obtained by the use of the total potential energy functional are also displayed. It is surprising that the Reissner variational principle was not proved to be superior to the principle of minimum total potential energy, though displacements and stresses as two independent unknown fields have been approximated. Due to higher computational effort of the Reissner variational principle one can conclude that it is not worth applying it to similar contact problems.

It is also investigated what the connection is between the polynomial order of the stress and the polynomial order of the displacement fields. Three variants: Variant 1 (V1), Variant 2 (V2) and Variant 3 (V3) are analyzed:

V1: $\mathtt{nps} = \mathtt{npu} - 1$;
V2: $\mathtt{nps} = \mathtt{npu}$;
V3: $\mathtt{nps} = \mathtt{npu} + 1$,

where `nps` means the polynomial order of stress approximation function, and the polynomial order of displacement approximation function is denoted by `npu`. Figure 16a makes the comparison easier.

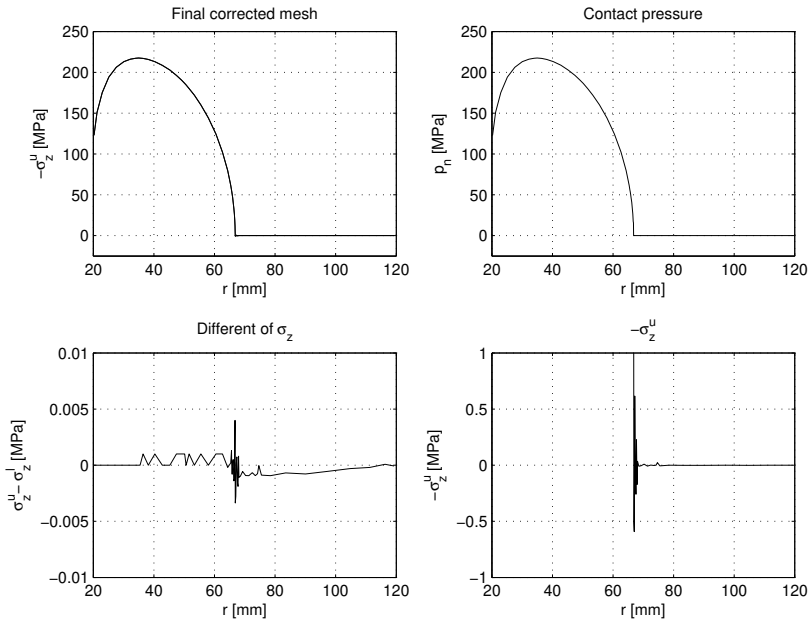


Figure 13. Contact stress distribution using the final mesh, `Imesh` = 3, $l_z = 20\,\mathrm{mm}$

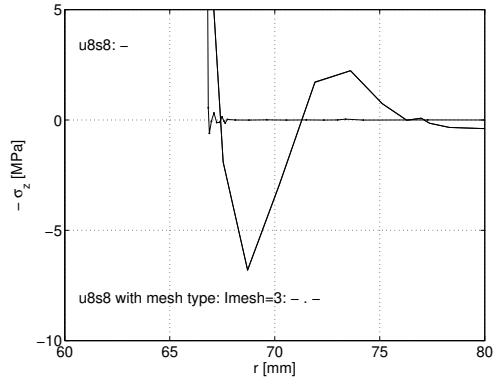


Figure 14. Oscillation of the normal stress around border point of contact zone (`Imesh` = 0 : −, `Imesh` = 3 : −.−, $l_z = 20\,\mathrm{mm}$)

Using the principle of minimum total potential energy with `Imesh` = 3, $l_z = 7.5\,\mathrm{mm}$ (see Table 1).

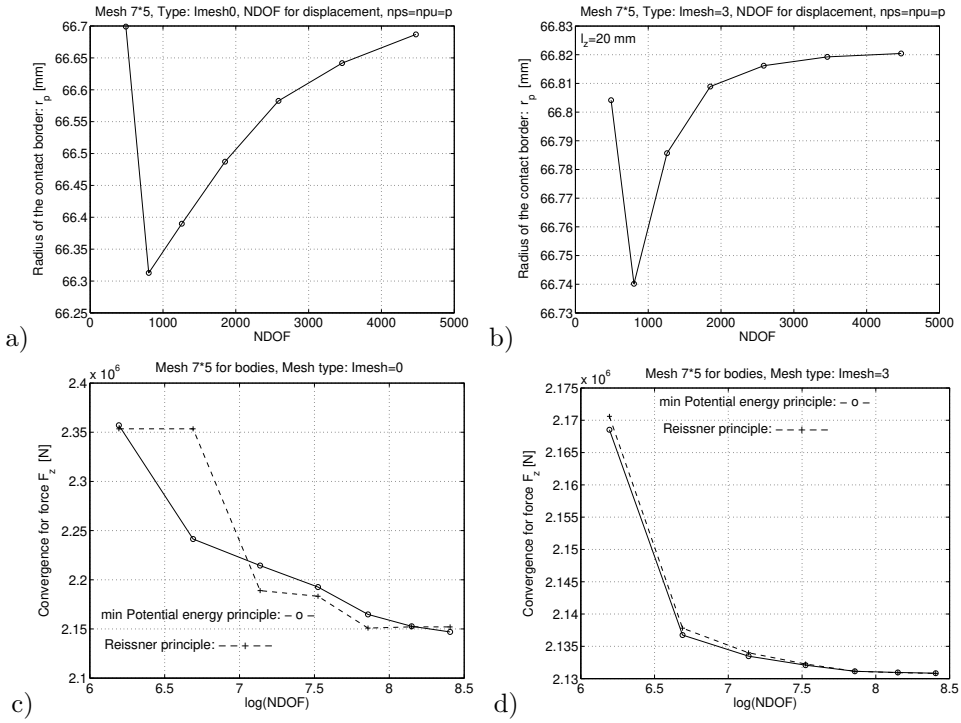


Figure 15. Convergence diagram for radius r_p a) uniform mesh (`lmesh` $= 0$); b) meshed with small elements around border point of contact zone (`lmesh` $= 3$). Resultant force F_z c) `lmesh` $= 0$; d) `lmesh` $= 2$, $l_z = 7.5\,\mathrm{mm}$

The Reissner principle is applied to solve the same contact problem and using the same *rp*-version finite element mesh with increasing polynomial order (`npu` $=$ `nps` $= p$) and the results are summarized in Table 2.

We can compare the results obtained by the total potential energy functional and Reissner principle. Practically the radius of the contact border is the same, the error indicators I_3 and I_4 are mainly smaller with the Reissner principle. The difference in resultant force F_z is very small. The strain energy of Reissner principle is computed from the approximated stress field. The convergence rate of the strain energy is a bit larger for the Reissner principle.

It is evident that using the (7×5) mesh the solution gives a small discrepancy when the polynomial order changinges from $p = 2$ to $p = 8$. However, the relative error for r_p is smaller than $0.2\,\%$ and for F_z this error shows $2\,\%$ ($[F_z(p = 2) - F_z(p = 8)]/F_z(p = 8)$).

Three mesh types (`lmesh` $= 1$, `lmesh` $= 2$ and `lmesh` $= 3$) are used in order to compare the results for these two minimization principles ($l_z = 7.5\,\mathrm{mm}$). Results are collected in Table 3 for `nps` $=$ `npu` $= p = 8$ and `NDOF` $= 4\,474$.

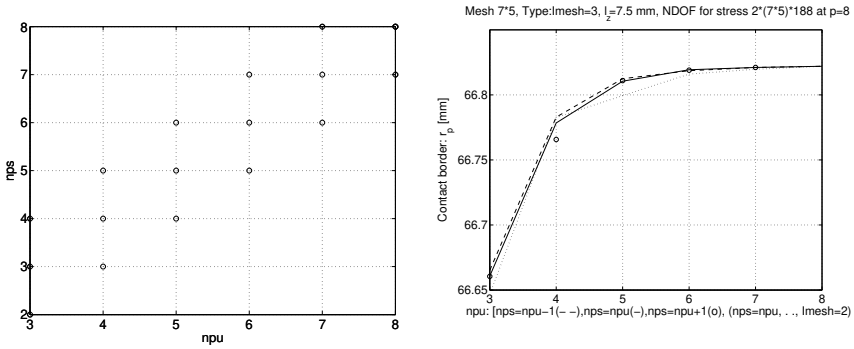


Figure 16. a) Change in polynomial order for displacements (`npu`) and for stresses (`nps`); b) Convergence diagram for border point of contact zone r_p when mesh is modified by small elements (`lmesh` = 3)

Table 1. Results are summarized as a function of polynomial order p, based on total potential energy.

p	NDOF	r_p [mm]	Pot.e [Nmm]	F_z [N]	I_3 [N/mm]	I_4 [N/mm]
2	490	66.6759	128 390.922	2 168 530.63	543.85	280.02
3	804	66.6806	127 940.062	2 136 748.84	119.48	110.73
4	1 258	66.7532	127 865.300	2 133 480.24	57.14	59.29
5	1 852	66.7987	127 840.504	2 132 085.58	35.23	43.61
6	2 586	66.8130	127 831.901	2 131 137.97	22.11	36.77
7	3 460	66.8228	127 830.607	2 130 957.02	21.53	34.86
8	4 474	66.8225	127 830.353	2 130 835.13	16.81	32.46

Table 2. Results are summarized as a function of polynomial order p, based on Reissner principle.

p	NDOF	r_p [mm]	Strain.e [Nmm]	F_z [N]	I_3 [N/mm]	I_4 [N/mm]
2	490	66.6632	128 443.511	2 170 611.24	614.04	274.43
3	804	66.6659	127 951.403	2 137 806.80	132.08	89.02
4	1 258	66.7649	127 871.373	2 133 994.69	70.21	53.16
5	1 852	66.8107	127 843.167	2 132 259.59	37.73	42.50
6	2 586	66.8194	127 832.525	2 131 129.16	18.51	34.05
7	3 460	66.8216	127 830.756	2 130 920.19	14.92	32.07
8	4 474	66.8222	127 830.400	2 130 804.34	11.77	30.88

Figure 18 shows the results for 7×5 mesh, at $l_z = 7.5$, 15 and 20 mm, and at `lmesh` = 1, 2, 3, by $p = 8$, when using the Reissner principle. It can be observed that the results are very similar. The smallest strain energy is supplied by `lmesh` = 3 and $l_z = 15$ mm. The minimum value for the contact border radius r_p is with `lmesh` = 1 and $l_z = 15$ mm. Indicator I_3 is the smallest with `lmesh` = 2 and $l_z = 7.5$ mm, and I_4 is the smallest with `lmesh` = 3 and $l_z = 7.5$ mm.

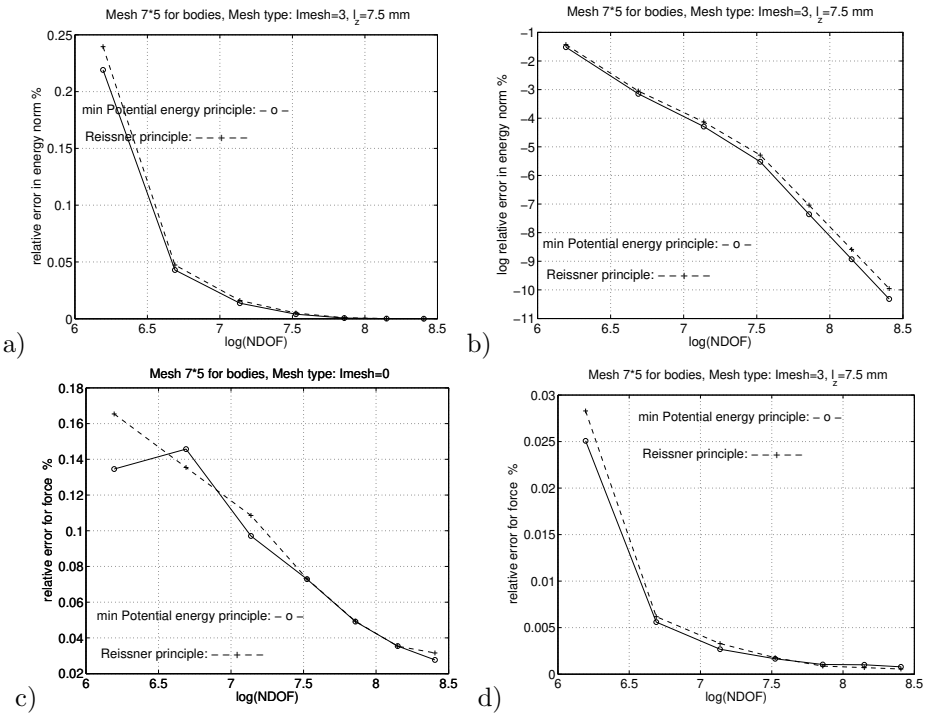


Figure 17. Comparing the convergence of the energy for finite element codes using minimum of the total potential energy and Reissner principle: a) relative error in the energy norm; b) logarithm of the relative error; c) $\texttt{Imesh} = 0$, $(\texttt{Error}_F)_u : - \circ -$, $(\texttt{Error}_F)_s : - + -$; d) $\texttt{Imesh} = 3$, $(\texttt{Error}_F)_u : - \circ -$, $(\texttt{Error}_F)_s : - + -$

The relative difference between the maximum force and the minimum force in F_z is less than 0.13 %.

Solving the above contact problem with the Reissner principle by mesh 8×5 ($\texttt{NDOF} = 5\,090$), the following results are obtained:

$\texttt{Imesh} = 2 \quad l_z = 7.5\,\text{mm} \quad r_p = 66.8221\,\text{mm} \quad l_z = 15\,\text{mm} \quad r_p = 66.8207\,\text{mm}$

$\texttt{Imesh} = 3 \quad l_z = 7.5\,\text{mm} \quad r_p = 66.8223\,\text{mm} \quad l_z = 15\,\text{mm} \quad r_p = 66.8208\,\text{mm}$

This indicates that the finer mesh 8×5 does not give a better or different solution, therefore applying mesh 7×5 gives sufficently accurate results using $\texttt{Imesh} = 2$ according to Table 3.

7. Example for Hertz problem with friction

The contact problem for a truncated sphere with radius $R = 800\,\text{mm}$ and a cylinder with outer radius $r_u = 10\,\text{mm}$ is investigated. The truncated sphere is the upper body denoted by B_1, the cylinder is the lower one denoted by B_2. The height of

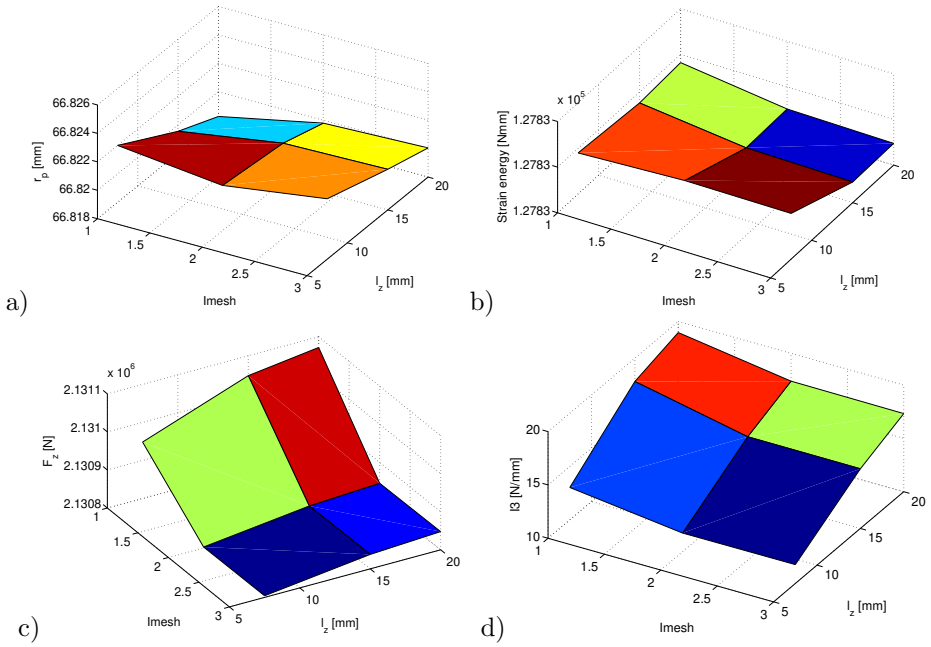


Figure 18. Results with mesh 7×5, `Imesh` $= 1, 2, 3$ and $l_z = 7.5, 15, 20$ mm by using the Reissner principle

the truncated sphere is 5 mm and its outer surface is a cylinder mantle with radius 10 mm. Height of the lower cylinder is equal to 5 mm.

Material parameters for the lower body are Young's modules $E_2 = 200$ GPa and Poisson ratio $\nu = 0.3$. Sphere is regarded to be rigid and its Young's modules is given as $E_1 = 1000E_2$. Vertical displacement $u_z = -w_0 = 0.005$ mm is prescribed on the top surface of the body B_1. The coefficient of friction is $\mu = 0.2$.

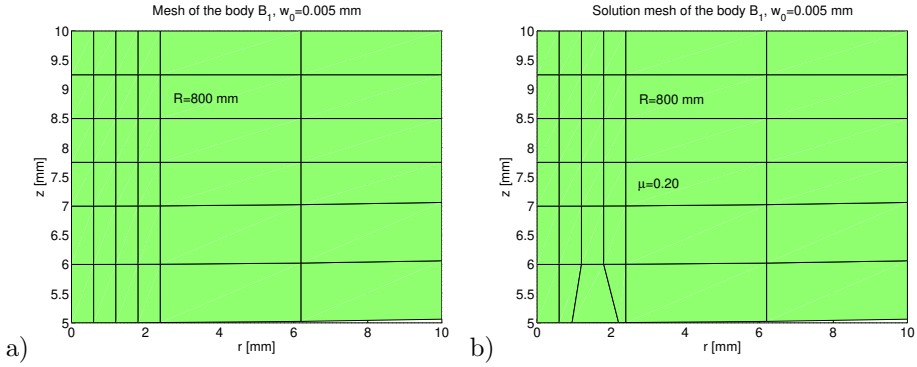
The initial mesh is shown in Figure 19a. The modified mesh after solution of the frictional contact problem is given in Figure 19b. The 3$^{\text{rd}}$ element has new nodal point positions, one of them is moved left to border of stick-slip region, and the other one is moved to right to the effective contact border.

The obtained distributions of the shear stresses along the contact domain are shown in Figure 20 for different load levels. In those points, where $\tau_n = \tau_{rz} \leq \mu p_n$ stick takes place, while in the rest of points, where $\tau_n = \tau_{rz} = \mu p_n$ here are slips. Distribution of the normal stress along the boundary $z = 5$ and $z = 10$ mm of the sphere is displayed in Figure 21. The resultant vertical forces of both surfaces theoretically should be equal to each other. This requirement is fulfilled with an acceptably small error, they are equal up to the first three digits.

Penalty parameter in the tangential direction is $c_\tau = 100E_2$, and in the vertical (normal) direction it is $c_n = 1000E_2$.

Table 3. Comparing results for different meshes ($l_z = 7.5\,\mathrm{mm}$).

Imesh	r_p [mm]	using Pot.e. min. princ.	using Reissner var. princ.
1		66.8228	66.8220
2		66.8174	66.8219
3		66.8225	66.8222
Imesh	**Strain e. [Nmm]**	**using Pot.e. min. princ.**	**using Reissner var. princ.**
1		127 830.352	127 830.38
2		127 830.353	127 830.40
3		127 830.353	127 830.40
Imesh	**F_z [N]**	**using Pot.e. min. princ.**	**using Reissner var. princ.**
1		2 130 904.48	2 130 948.39
2		2 130 860.09	2 130 802.48
3		2 130 835.13	2 130 804.34
Imesh	**I_3 [N/mm]**	**using Pot.e. min. princ.**	**using Reissner var. princ.**
1		17.49	13.04
2		12.26	11.75
3		16.81	11.77
Imesh	**I_4 [N/mm]**	**using Pot.e. min. princ.**	**using Reissner var. princ.**
1		33.09	31.67
2		31.92	30.89
3		32.46	30.88



a) b)

Figure 19. Meshes of body B_1: a) original b) at the end of solution of contact problem using position technique. The stick-slip border is at $r_{st-sl} = 0.947\,\mathrm{mm}$, the effective contact border is at $r_p = 2.203\,\mathrm{mm}$

The value of the friction coefficient μ essentially influences the length of stick zone. The results of the tangential stresses obtained for three different frictional coefficient are shown in Figure 22. However μ has only a little influence on contact border radius r_p as given here: $r_p(\mu = 0.1) = 2.213$, $r_p(\mu = 0.2) = 2.203$, $r_p(\mu = 0.3) = 2.193\,\mathrm{mm}$.

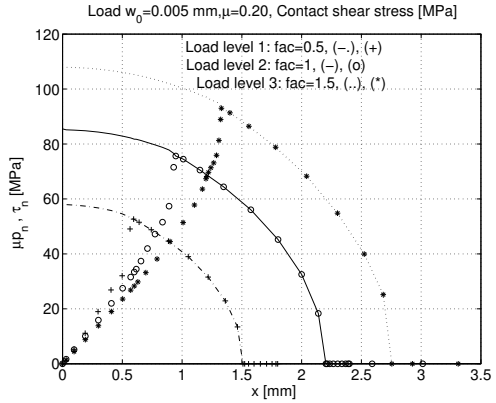


Figure 20. Shear stresses along the contact surface

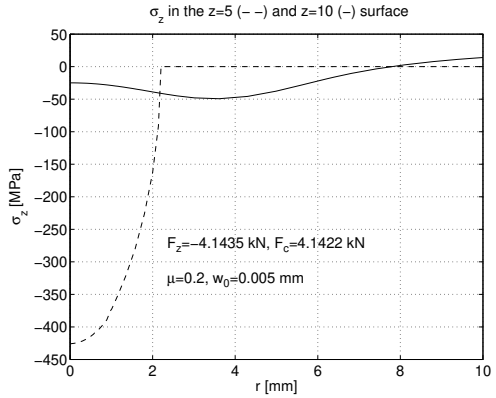


Figure 21. Distribution of σ_z along the contact surfaces at $z = 5$ and $z = 10$ mm where the vertical displacement is prescribed

The contact pressure maximums are 423.83, 427.49, 429.13 MPa at different friction coefficients, $\mu = 0.1$, 0.2, 0.3, respectively.

Finally problems are analyzed when only the rigidity of the sphere is changed.

The results are collected in Table 4. In order to compare the results of task 4 in Table 4 with Hertz theory, the maximum of contact pressure p_0 and resultant load F_z are calculated with $r_p(\mu = 0.2) = a = 2.232$ mm.

Using the Hertz formula the force $F_{z,Hertz} = 4E_*a^3/(3R) = 2.037$ kN and pressure $p_{0,Hertz} = \frac{3F_{z,Hertz}}{2\pi a^2} = 195.19$ MPa are calculated with $\frac{1}{E_*} = \frac{1-(\upsilon_1)^2}{E_1} + \frac{1-(\upsilon_2)^2}{E_2}$.

The relative errors are very small:

$$\text{error } F_z = 100\frac{F_{z,Hertz} - F_z}{F_{z,Hertz}} = -0.736\%, \quad \text{error } p_0 = 100\frac{p_{0,Hertz} - p_0}{p_{0,Hertz}} = -2.05\%.$$

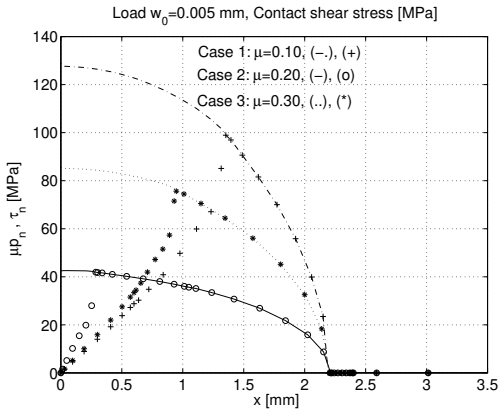


Figure 22. Shear stress along the contact surface at different coefficients of friction

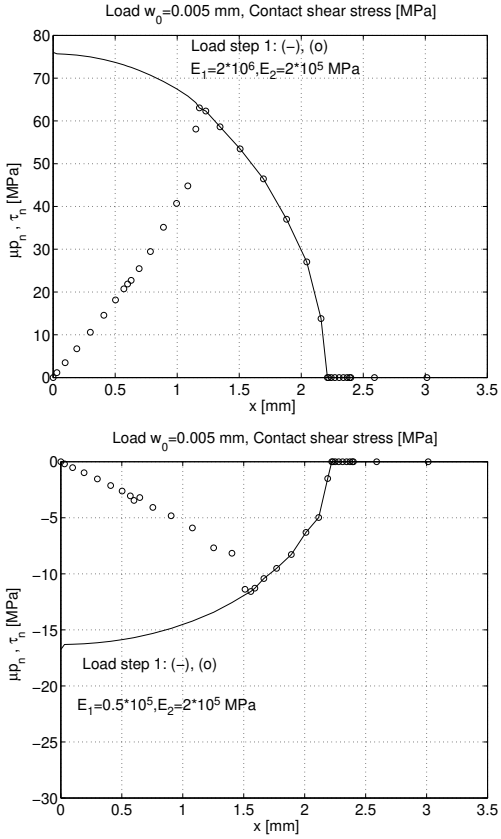


Figure 23. Shear stress distribution along the contact surface for different material parameters of the sphere, a) $E_1 = 2000\,\mathrm{GPa}$, b) $E_1 = 50\,\mathrm{GPa}$

Table 4. Results of the Hertz problem for different friction coefficient μ ($\nu = \nu_1 = \nu_2 = 0.3$)

$\mu = 0.2$	E_1	E_2	r_{st-sl}	r_p	p_0	F_z
Number of task	[GPa]	[GPa]	[mm]	[mm]	[MPa]	[kN]
1	∞	200	0.947	2.203	427.49	4.143
2	2000	200	1.179	2.211	380.18	3.761
3	1000	200	1.181	2.222	372.37	3.720
4 $\mu = 0$	200	200	-	2.232	199.19	2.052
5	50	200	1.558	2.222	83.779	0.825

In tasks 1-3 the tangential displacements of the sphere are smaller than of the cylinder, but the opposite was calculated for task 5. This is due to the fact that the tangential stress $\tau_n = \tau_{rz}$ of the sphere is positive for tasks 1-3, and shear stress is negative for task 5 (see Figure 23).

8. Conclusions

Axially symmetric normal contact problems have been investigated by the *h*-version, *p*-version and *hp(rp)*-version finite element methods. The computations have been performed by a special purpose finite element package. The code can be applied for contact problems which leads to only a single connected contact zone.

In this paper the contact conditions are checked at the Gauss integration points. In order to obtain highly accurate results we use a special nodal positioning technique which ensures the contact along the whole edge of the contacting finite elements and very small oscillation in the gap zone.

Comparing the numerical results we concluded the following.

- When using equal number of degrees of freedom for *h*- and *p*-versions the contact pressure distributions are similar. The maximum value of the contact pressure is bigger for *h*-version than for *p*-version. The boundary condition for the normal stress in the gap zone is satisfied better by *p*-version than by *h*-version.
- Accurate position of the contact border point can be determined by the use of the *hp(rp)*-extension finite elements with nodal positioning technique. The results are practically free of oscillations, and the dynamic boundary conditions are satisfied with high accuracy in the gap region as well.
- The convergence of the *hp*(*rp*)-version has been proved by numerical computations. Using finite element mesh with one or two layers of narrow elements at the contact border point and performing computations with a sequence of high values of p ($p = 5, 6, 7, 8$) give a convergent series for the position of the contact border point.
- The accuracy of the solution with the Reissner principle is not better (at the same `NDOF` for displacement field) than the using minimum principle of potential energy. For engineering contact problems it is recommended to use the principle of minimum potential energy because of faster computation.
- Application of the special positioning technique for the frictional contact problem gives good accuracy in solutions.

Acknowledgement. The present research was partially supported by the Hungarian Academy of Sciences, by grant NKFIH 115701.

References

1. V. M. Fridman and V. S. Chernina. "Iteration methods applied to the solution of contact problems of bodies." *Mehanika Tverdogo Tela AN, USSR,* **1**, (1967), pp. 116–120.
2. I. Páczelt. "Iterational method applied to the solution of contact problems of elastic systems having elements in unilateral relation (in Russian)." *Acta Technica Hungarorum,* **76**, (1974), pp. 217–241.
3. T. F. Conry and G. A. Seireg. "A mathematical programming method for design of elastic bodies in contact." *ASME, Journal of Applied Mechanics,* **38**, (1971), pp. 387–392. DOI: 10.1115/1.3408787.
4. J. J. Kalker and Van Y. Randen. "A minimum principle for frictionless elastic contact problems." *Journal of Engineering Mathematics,* **6**, (1972), pp. 193–206. DOI: 10.1007/BF01535102.
5. M. Fremond. "Dual formulation for potential and complementary energies. Unilateral boundary conditions application to the finite element method." *The Mathematics of Finite Elements and Applications.* Academic Press, 1973, pp. 175–188. ISBN: 978-1483-2688-42. URL: https://www.sciencedirect.com/book/9780127472508/the-mathematics-of-finite-elements-and-applications.
6. A. Frankavilla and O. C. Zienkiewicz. "A note on numerical computation of elastic contact problems." *International Journal for Numerical Methods in Engineering,* **9**, (1975), pp. 913–924. DOI: 10.1002/nme.1620090410.
7. E. Haug, R. Chand, and K. Pan. "Multibody elastic contact analysis by quadratic programming." *Journal of Optimization Theory and Applications,* **21**(2), (1977), pp. 189–198. DOI: 10.1007/BF00932519.
8. I. Páczelt and B. Herpai. "Some remarks on the solution of contact problems of shells." *Archiwum Budowy Maszyn XXIV,* **1**, (1977), pp. 117–202.
9. I. Páczelt. "Some remarks to the solution of quadratic programming problems." *Publicatins of the Technical University Heavy Industry, Series D. Natural Sciences,* **33**, (1979), pp. 137–156.
10. Nguyen Dan Hung and Géry De Saxce. "Frictionless contact of elastic bodies by finite element method and mathematical programming technique." *Computers and Structures,* **11**, (1980), pp. 55–67.
11. A. Klarbring. "A mathematical programming approach to three-dimensional contact problems." *Computer Methods in Applied Mechanics and Engineering,* 1985 (**58**,), pp. 175–200. DOI: 10.1016/0045-7825(86)90095-2.
12. G. Björkman. "The solution of large displacement frictionless contact problems using a sequence of linear complementary problems." *International Journal for Numerical Methods in Engineering,* **31** (1991), pp. 1533–1566. DOI: 10.1002/nme.1620310808.
13. S. Chan and I. Tuba. "A finite element method for contact problems of bodies." *International Journal of Mechanical Sciences,* **13**, (1971), pp. 615–639. DOI: 10.1016/0020-7403(71)90033-6.

14. B. Fredrikson. "Finite element solution of surface nonlinearities in structural mechanics problems." *Computers and Structures,* **6**, (1976), pp. 281–290.
15. I. Páczelt. "Solution of elastic contact problems by the finite element displacement method." *Acta Technica Hungarorum,* **82**, (1976), pp. 353–375. URL: 193.6.1.94:9080/?docId=28267.
16. T. R. J. Hughes and R. L. Taylor. "A finite element method for a class of contact impact problems." *Computer Methods in Applied Mechanics and Engineering,* **8**, (1976), pp. 249–276. DOI: 10.1016/0045-7825(76)90018-9.
17. K. J. Bathe and A. Chaudhary. "A solution method for planar and axisymmetric contact problems." *International Journal for Numerical Methods in Engineering,* **21** (1985), pp. 65–88. DOI: 10.1002/nme.1620210107.
18. B. Nour-Omid and P. Wriggers. "A two level iteration method for solution of contact problems." *Computer Methods in Applied Mechanics and Engineering,* **54**, (1986), pp. 131–144. DOI: 10.1016/0045-7825(86)90122-2.
19. A. B. Chaudhary and K. J. Bathe. "A solution method for static and dynamic analysis of three-dimensional contact problems with friction." *Computers and Structures,* **24**(6), (1986), pp. 855–873.
20. J. Égert and J. Altenbach. "Inkrementelle Iterationsmethoden zur Lösung elastischer Kontaktaufgaben mit Reibung mittels spezieller isoparametrischer 2D- und 3D-kontaktelemente." *Technische Mechanik,* **10**, (1989), pp. 120–137. URL: http://www15.ovgu.de/ifme/zeitschrift_tm/1989_Heft2/Egert.pdf.
21. J. E. Mottershead, S. K. Pascone, and R. G. English. "A general finite element approach for contact stress analysis." *International Journal for Numerical Methods in Engineering,* **33**, (1992), pp. 765–779. DOI: 10.1002/nme.1620330407.
22. P. Papadopoulos and R. L. Taylor. "A simple algorithm for three- dimensional finite element analysis of contact problems." *Computers and Structures,* **46**, (1992), pp. 1107–1118.
23. N. Kikuchi, J. T. Oden, and G. F. Carcey. *Finite Elements: Special Problems in Solid Mechanics*. Prentice Hall, Englewood, 1984.
24. 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**, (1985), pp. 163–180. DOI: 10.1016/0045-7825(85)90088-X.
25. J. W. Ju and R. L. Taylor. "A perturbed Lagrangian formulation for the finite element solution of nonlinear frictional contact problems." *Journal of Theoretical and Applied Mechanics, Special issue, Suplement No. 1.,* **7**, (1988), pp. 1–14.
26. J. C. Simo and T. A. Laursen. "An augmented lagrangian treatment of contact problems involving friction." *Computers and Structures,* **42**(7), (1992), pp. 97–116.
27. 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**, (1993), pp. 3451–3485. DOI: 10.1002/nme.1620362005.
28. D. P. Berteskas. *Constrained Optimization and Lagrange Multiplier Methods.* 2nd edition. Academic Press, NewYork, 2014. DOI: 10.1016/C2013-0-10366-2.

29. R. Fletcher. *Practical Methods of Optimization.* 2nd edition. John Wiley, NewYork, 1989.
30. Z. H. Zhong and J. Mackerle. "Static contact problems a review." *Engineering Computations,* **9** (1992), pp. 13–37. DOI: 10.1108/eb023846.
31. U. Nackenhorst. *An adaptive finite element method to analyse contact problems. In Contact Mechanics II, Computational Techniques.* Ed. by M. H. Aliabali and C. Alessandri. Southampton, Boston: Computational Mechanics Publications, 1991, pp. 241–248.
32. P. Wriggers. *Finite elements for thermomechanical contact and adaptive finite element analysis of contact problems. CISM Course: No. 388, Chapter 4 in New developments in contact problems.* Ed. by P. Wriggers and P. Panatiotopoulos. Wien: Springer Verlag, 1999, pp. 179–246. DOI: 10.1007/978-3-7091-2496-3.
33. B. A. Szabó and I. Babuska. *Finite Element Analysis.* John Wiley & Sons, New York, 1991.
34. C. Y. Lee and J. T. Oden. "Theory and approximation of quasistatic frictional contact problems." *Computer Methods in Applied Mechanics and Engineering,* **106**, (1993), pp. 407–429. DOI: 10.1016/0045-7825(93)90098-I.
35. U. Gabbert and K. Weinberg. "A pNh-element formulation for contact analysis." *Zeitschrift für Angewandte Mathematik und Mechanik (ZAMM),* **74**(1), (1994), pp. 195–197. DOI: 10.1002/zamm.19940740407.
36. R. Buczkowski, M. Kleiber, and U. Gabbert. "On linear and higher order standard finite elements for 3d-nonlinear contact problem." *Computers and Structures,* **53**, (1994), pp. 817–823.
37. I. Páczelt and T. Szabó. *Application of the augmented lagrangian technique for solution of contact optimisation problems. Contact Mechanics II, Computational Techniques.* Ed. by M. H. Aliabali and C. Alessandri. Southampton, Boston: WIT Press, 1995, pp. 249–256. DOI: 10.2495/CON950291.
38. I. Páczelt and A. Baksa. "Solution of contact problems using p-extension finite elements." *Proceedings of XXXVII Summer School: Conference Advanced Problems in Mechanics.* Ed. by D.A. Indeitsev and S. M. Kristov. 2009, pp. 507–519. URL: http://apm-conf.spg.ru.
39. D. Franke, A. Dünster, V. Nübel, and E. Rank. "A computation of the h-, p-, hp- and rp-version of the FEM for the solution of the 2D Hertzian contact problem." *Computational Mecheanics,* **45**, (2010), pp. 513–522. DOI: 10.1007/s00466-009-0464-6.
40. A. Konyukhov and K. Schweizerhof. "Incorporation of contact for high-order finite element in convariant form." *Computer Methods in Applied Mechanics and Engineering,* **198**, (2009), pp. 1213–1223. DOI: 10.1016/j.cma.2008.04.023.
41. D. Franke, E. Rank, and A. Dünster. "Computational contact mechanics based on the rp-version of the finite element method." *International Journal of Computational Methods,* **8**, (2011), pp. 493–512. DOI: 10.1142/S0219876211002630.
42. I. Páczelt, B. A. Szabó, and T. Szabó. "Solution of contact problem using the hp-version of finite elements, accepted for publication." *Computers & Mathematics with Applications,* **38** (1999), pp. 49–69. DOI: 10.1016/S0898-1221(99)00261-8.

43. G. Duvat and J. L. Lions. *Les Inequations en Mecanique et en Physique*. Dunod, Paris, 1972.
44. I. Hlavacek, J. Haslinger, J. Necas, and J. Lovicek. *Solution of Variational Inequalities in Mechanics, Applied Mathematical Sciences*. Volume 66. Series of Springer Verlag, Berlin, 1980.
45. P. D. Panagiotopoulos. *Inequality Problems in Mechanics and Applications: Convex and Nonconvex Energy Functions*. Birkhaüser, Basel, 1985. DOI: 10.1007/978-1-4612-5152-1.
46. J. T. Oden and N. Kikuchi. *Contact Problem in Elasticity: A Study of Variational Inequalities and Finite Element Methods*. SIAM, Philadelphia, 1988.
47. J. I. Telega. "Variational principles for mechanical contact problems (in Russian)." *Advances of Mechanics,* **10**, (1987), pp. 3–95.
48. J. Haslinger and P. Neittaanmaki. *Finite Element Approximation for Optimal Shape Design*. 2nd. John Wiley & Sons, 1996.
49. J. Kalker. *Three-dimensional Elastic Bodies in Rolling Contact*. Academic Publisher, Dordrecht, 1990. DOI: 10.1007/978-94-015-7889-9.
50. M. Raous, P. Charband, and F. Lebon. "Numerical methods for frictional contact problems and applications." *Journal of Theoretical and Applied Mechanics, Special Issue, Supplement No. 1,* **7**, (1988), pp. 111–128.
51. A. Klarbring. "Contact, friction, discrete mechanical structures and mathematical programming." *CISM course: Contact problems: No. 388 Chapter 2, New developments in contact problem*. Ed. by P. Wriggers and P. Panatiotopoulos. Wien: Springer Verlag, 1999, pp. 55–100. DOI: 10.1007/978-3-7091-2496-3.
52. S. Hilding, A. Klarbring, and J. Peterson. "Optimization of structures in unilateral contact." *Applied Mechanics Reviews,* **52**(4), (1999), pp. 139–160. DOI: 10.1115/1.3098931.
53. Y. Volpert. "Space Enrichment Methods for the Numerical Solution of Mechanical Contact Problems." PhD. Washington University, Department of Mechanical Engineering, St. Louis, 1995.
54. Y. Volpert, T. Szabó, I. Páczelt, and B. A. Szabó. "Application of the space enrichment method to problems of mechanical contact." *Finite elements in Analysis and Design,* **24**, (1997), pp. 157–170. DOI: 10.1016/S0168-874X(96)00047-9.
55. I. Páczelt and T. Szabó. "Solution of contaqct optimization problems of cylindrical bodies using hp-FEM." *International Journal for Numerical Methods in Engineering,* **53**, (2002), pp. 123–146. DOI: 10.1002/nme.395.
56. A. Baksa, I. Páczelt, and T. Szabó. "Solution of 3D contact problems using spline interpolation." *Journal of Computational and Applied Mechanics,* **9**(2), (2014), pp. 125–147. DOI: 10.32973/jcam.2014.007.
57. V. A. Yastrebov. *Numerical Mechanics in Contact Mechanics*. John Wiley & Sons Inc. Hoboken, 2013. DOI: 10.1002/9781118647974.
58. P. Wriggers. *Computational Contact Mechanics*. Springer, New York, 2006. DOI: 10.1007/978-94-015-9048-8.
59. A. Konyukhov and K. Schweizerhof. *Computational Contact Mechanics*. Springer-Verlag Berlin Heidelberg, 2013. DOI: 10.1007/978-3-642-31531-2.

60. T.A. Laursen. *Computational Contact and Impact Mechanics: Fundamentals of Modeling Interfacial Phenomena in Nonlinear Finite Element Analysis*. Springer-Verlag, New York, 2003. DOI: 10.1007/978-3-662-04864-1.
61. P. Solin, K. Segeth, and I. Dolezel. *Higher-Order Finite Element Methods*. Chapman & Hall/CRC, London, 2004.
62. L. Demkowicz, J. Kurtz, D. Pardo, M. Paszenski, W. Rachowicz, and A. Zdunek. *Computing with hp-adaptive finite elements, Frontiers: Three Dimensional Elliptic and Maxwell Problems with Applications*. Vol. 2nd. Chapman & Hall/CRC, 2008.
63. I. G. Goryacheva. *Contact Mechanics in Tribology*. Springer Netherlands, 1998. DOI: 10.1007/978-94-015-9048-8.
64. I. Páczelt and Z. Mróz. "Solution of wear problems for monoton and periodic sliding with p-version of finite element method." *Computer Methods in Applied Mechanics and Engineering,* **1**, (2012), pp. 249–252. DOI: 10.1016/j.cma.2012.02.012.
65. N. V. Banichuk and S. Yu. Ivanova. *Optimal structural design*. Walter de Gruyter GmbH, Berlin, 2017.
66. M. Raous, M. Jean, and J. J. Moreau, eds. *Contact Mechanics*. Springer US, NewYork, 1995. DOI: 10.1007/978-1-4615-1983-6.
67. J. A. C. Martins and M. O. P. Monteiro Marques, eds. *Contact Mechanics, Proceedings of the 3rd Contact Mechanics International Symposium*. Kluwer Academic Publishers, Dordrecht, 2002.
68. A. Popp and P. Wriggers. *Contact Modeling for Solids and Particals*. CISM International Centre for Mechanical Sciences 585. Springer International Publishing, 2018. DOI: 10.1007/978-3-319-90155-8.
69. P. Wriggers and U. Nackenhorst, eds. *IUTAM Symposium on Computational Methods in Contact Mechanics, Hannover, Germany*. Springer Netherlands, 2006. DOI: 10.1007/978-1-4020-6405-0.
70. ESRD Group. ***StressCheck*** *finite element simulation program*. 1989. URL: https://esrd.com.

Journal of Computational and Applied Mechanics, Vol. 15, No. 2, (2020), pp. 173–184
DOI: 10.32973/jcam.2020.010

GENERALIZED DISPLACEMENTS AND MOMENTA FORMULATIONS OF AN ELECTROMECHANICAL PLUNGER

Tamás Szabó[a] and László Rónai[a]
[a]Institute of Machine Tools and Mechatronics, University of Miskolc,
3515 Miskolc-Egyetemváros, Hungary
szabo.tamas@uni-miskolc.hu, ronai.laszlo@uni-miskolc.hu

[Received: June 22, 2020; Accepted: August 14, 2020]

Dedicated to Professor Barna Szabó on the occasion of his 85th birthday

Abstract. This paper deals with four different derivations of the governing equations of a solenoid plunger with lumped-parameter. Energy-based modeling is employed with extended Hamilton's principle with independent generalized coordinates and generalized momenta in order to be applicable to composite Lagrange's equations. In the electromechanical models, displacements and charges are regarded to be generalized coordinates, mechanical momenta and flux linkages are the generalized momenta. The derived systems of differential equations are solved numerically with the Runge-Kutta method.

Mathematical Subject Classification: 37N30, 70H03
Keywords: plunger, lumped-paramter, energy-based modeling, composite Lagrange's equation

1. Introduction

A mechatronic problem can include mechanical, electrical, hydraulic, and thermal phenomena. The application of an energy-based approach is suitable for the uniform management of all of them [1]. The following domain-independent generalized variables are advisable to use in complex problems: generalized coordinates, generalized momenta, generalized potentials and generalized velocities. The time derivative of the generalized potentials are often called efforts, and the generalized velocities are named flows. In this context generalized power, generalized energy, generalized potential energy, generalized potential co-energy, generalized kinetic energy, and generalized kinetic co-energy can be defined [1]–[2].

A plunger [3] is regarded as an electromechanical problem. A number of papers deal with modeling and simulation of hydraulic valves, which are also plungers [4]–[5]. In these papers, models of lumped-parameter are investigated using displacement, electrical current or flux linkage variables. Following that, in addition to the previously mentioned variables, mechanical momenta will also be used to derive the governing

equations. To the best knowledge of the authors of this article such equations for plungers written by momenta have not been presented yet.

Next an electromechanical plunger is modeled, in which variables displacement of the spool, electrical charge are generalized coordinates, the mechanical momentum of the spool and the flux linkage are generalized momenta. The model of the plunger to be investigated can be composed via purely generalized coordinates, or purely generalized momenta, and also their mixed combinations.

This paper is organized as follows: Section 2 defines the structure of the plunger. Section 3 details four different formulations of the governing equations of the model. Section 4 represents results of the numerical computations. Concluding remarks are given in Section 5.

2. A lumped-parameter electromechanical plunger

The structure of the plunger is shown in Figure 1, where m is the mass of the spool, k is the stiffness of the suspending spring, r is the damping coefficient of the dashpot, $L(y)$ is the inductance of the coil of the electromagnet, R is the resistor of the inductor, $U_0(t)$ is the voltage source of the circuit and k_c and r_c represent the resultant stiffness and the damping coefficients of the collision layer, respectively. The nominal gap of the plunger is denoted by y_0 in unexcited situation. When the switch is closed the coil produces a magnetic force, which attracts the spool downwards to the electromagnet. Collision takes place at $y = -y_0$.

At this point the viscoelastic layer is under compression, it means that two springs $k + k_c$ and two dashpots $r + r_c$ are connected parallel to the spool. The spring characteristic of the problem is shown in Figure 2, where $\tan\alpha = k$ and $\tan\alpha_c = k+k_c$. Due to this fact the investigated problem is nonlinear, also in its mechanical aspect. The penalty treatment of the mechanical contact problem is based on a similar elastic layer sometimes with nonlinear spring, as was recommended by Prof. B. Szabó [6].

It is assumed that the inductance depends on the displacement y of the spool [2]:

$$L(y) = \frac{L_0}{1 + (y_0 + y)/h}, \tag{2.1}$$

where L_0 is the inductance at zero gap, h is a geometric size, which is shown in Figure 1. The Coulomb charge and the flux linkage are denoted by q, and λ, respectively. It is self evident that the potential is zero at the vicinity of the ground.

In the sequel, the mathematical models of the plunger will be formulated on an energy basis using the extended Hamilton's principle with independent generalized coordinates, generalized momenta and their combinations.

3. Mathematical formulation of the problem

A mechatronic system usually contains energy conservative and non-conservative elements. To treat such system the extended Hamilton's principle is a suitable variational principle:

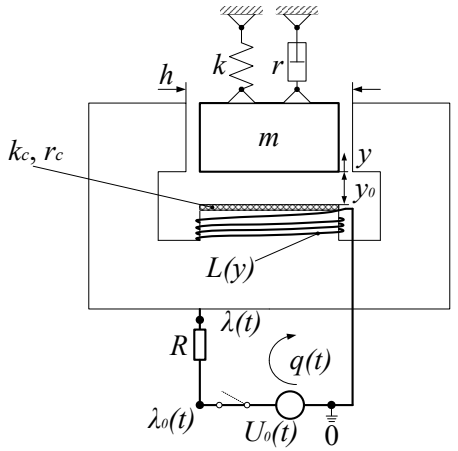


Figure 1. Schematic of the plunger

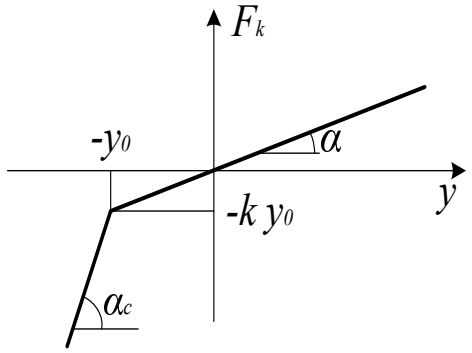


Figure 2. Spring characteristic of the electromechanical plunger

$$\int_{t_1}^{t_2} (\delta\mathcal{L} + \overline{\delta W}_{nc})dt = 0, \qquad \delta q_k = 0 \quad \text{or} \quad \delta p_k = 0, \quad k = 1, 2, ..., n; \quad t = t_1, t_2\,, \tag{3.1}$$

where $\delta\mathcal{L}$ is the variation of the Lagrangian function containing energies and co-energies of the conservative elements, $\overline{\delta W}_{nc}$ is the virtual work associated to the non-conservative elements, δq_k and δp_k are the variation of the general coordinates and momenta of a system, respectively.

An electromechanical problem can be treated by four combinations of displacement y and charge q as generalized coordinates, and also mechanical momentum p and flux linkage λ as generalized momenta:

$$\mathcal{L}(y,q) = T^*(\dot{y}) + W_m^*(\dot{q}) - V_p(y) - W_e(q), \quad \overline{\delta W}_{nc}(y,q), \tag{3.2}$$

$$\mathcal{L}(y,\lambda) = T^*(\dot{y}) + W_e^*(\dot{\lambda}) - V_p(y) - W_m(\lambda), \quad \overline{\delta W}_{nc}(y,\lambda), \tag{3.3}$$

$$\mathcal{L}(p,q) = V_p^*(\dot{p}) + W_m^*(\dot{q}) - T(p) - W_e(q), \quad \overline{\delta W}_{nc}(p,q), \tag{3.4}$$

$$\mathcal{L}(p,\lambda) = V_p^*(\dot{p}) + W_e^*(\dot{\lambda}) - T(p) - W_m(\lambda), \quad \overline{\delta W}_{nc}(p,\lambda), \tag{3.5}$$

where $T^*(\dot{y})$ is the kinetic co-energy, $W_m^*(\dot{q})$ is the magnetic co-energy, $V_p(y)$ is the potential energy, $W_m(\lambda)$ is the magnetic energy, $V_p^*(p)$ is the potential co-energy, and $T(p)$ is the kinetic energy. Since the investigated system has no capacitor the electric energy $W_e(q)$ and co-energy $W_e^*(\dot{\lambda})$ are zero.

Virtual works in equations (3.2)–(3.5) of the non-conservative elements are calculated by multiplication of the virtual change in generalized coordinates or generalized potentials and their energetically conjugated generalized forces or generalized velocities, respectively.

In order to apply the energy-based approach, it is noted that generally the admissibility requirements include kinematical constraints on the virtual displacements and the velocities, and Kirchhoff's current rule on the virtual charges and the currents. The admissibility requirements include the equation of Newton's second law on the virtual mechanical momenta and forces, and Kirchhoff's voltage rule on the virtual flux linkages and the voltages. Usually it is convenient to satisfy admissibility conditions on the virtual variations, i.e., to select a complete set of independent generalized coordinates [2].

Performing the variation of the Lagrangian function in equation (3.1), the Lagrange equation of the second kind can be obtained for independent generalized coordinates and momenta.

3.1. **Displacement and charge formulation (y, q).** The most natural formulation of the plunger is using displacement and charge variables [3]–[7], since it is easy to satisfy the kinematical constraints and the Kirchhoff's current rule.

If the spool is not in contact with the collision layer, i.e., $(y + y_0) > 0$:

$$\mathcal{L}(y,\dot{y},\dot{q}) = \frac{1}{2}m\dot{y}^2 + \frac{1}{2}L(y)\dot{q}^2 - \frac{1}{2}ky^2, \tag{3.6}$$

$$\overline{\delta W}_{nc} = -r\dot{y}\delta y + U_0(t)\delta q - R\dot{q}\delta q, \tag{3.7}$$

while in contact, i.e., $(y + y_0) \leq 0$:

$$\mathcal{L}(y,\dot{y},\dot{q}) = \frac{1}{2}m\dot{y}^2 + \frac{1}{2}L(y)\dot{q}^2 - \frac{1}{2}ky^2 - \frac{1}{2}k_c(y+y_0)^2, \tag{3.8}$$

$$\overline{\delta W}_{nc} = -(r+r_c)\dot{y}\delta y + U_0(t)\delta q - R\dot{q}\delta q, \tag{3.9}$$

where δy, δq are the variation of the displacement and the charge, respectively.

The governing equations are derived with the Lagrange equations:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}}\right) - \frac{\partial \mathcal{L}}{\partial q} = U_0(t) - R\dot{q}, \tag{3.10}$$

$$L(y)\ddot{q} + R\dot{q} + L'(y)\dot{y}\dot{q} = U_0(t), \tag{3.11}$$

where $L'(y) = \frac{dL(y)}{dy}$.

If the spool is not in contact with the collision layer, i.e., $(y + y_0) > 0$:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{y}}\right) - \frac{\partial \mathcal{L}}{\partial y} = -r\dot{y}, \tag{3.12}$$

$$m\ddot{y} + r\dot{y} + ky = L'(y)\frac{\dot{q}^2}{2}, \tag{3.13}$$

while in contact $(y + y_0) \leq 0$:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{y}}\right) - \frac{\partial \mathcal{L}}{\partial y} = -(r + r_c)\dot{y}, \tag{3.14}$$

$$m\ddot{y} + (r + r_c)\dot{y} + ky + k_c(y + y_0) = L'(y)\frac{\dot{q}^2}{2}. \tag{3.15}$$

The resulted ordinary differential equation equation (3.11) is the Kirchhoff's voltage rule, equations (3.13) and (3.15) are Newton's second law.

3.2. **Displacement and flux linkage formulation (y, λ).** The flux linkage is often used with the displacement to derive the equations [1, 2]. Instead of Kirchhoff's current rule, Kirchhoff's voltage rule should be satisfied for the flux linkages. The voltage source $U_0(t)$ can be expressed as the time derivative of a given function of the flux linkage $\dot{\lambda}_0(t)$. The selection of the flux linkage variable is based on Kirchhoff's voltage rule $U_0(t) = U_R + U_L = (\dot{\lambda}_0 - \dot{\lambda})_R + (\dot{\lambda} - 0)_L$.

If the spool is not in contact, i.e., $(y + y_0) > 0$:

$$\mathcal{L}(y, \dot{y}, \lambda) = \frac{1}{2}m\dot{y}^2 - \frac{1}{2}ky^2 - \frac{\lambda^2}{2L(y)}, \tag{3.16}$$

$$\overline{\delta W}_{nc} = -r\dot{y}\delta y - \frac{\dot{\lambda} - U_0(t)}{R}\delta\lambda, \tag{3.17}$$

while in contact, i.e., $(y + y_0) \leq 0$:

$$\mathcal{L}(y, \dot{y}, \lambda) = \frac{1}{2}m\dot{y}^2 - \frac{1}{2}ky^2 - \frac{1}{2}k_c(y + y_0)^2 - \frac{\lambda^2}{2L(y)}, \tag{3.18}$$

$$\overline{\delta W}_{nc} = -(r + r_c)\dot{y}\delta y - \frac{\dot{\lambda} - U_0(t)}{R}\delta\lambda, \tag{3.19}$$

where $\delta\lambda$ is the variation of the flux linkage.

The governing equations are also derived with the Lagrange equations:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{\lambda}}\right) - \frac{\partial \mathcal{L}}{\partial \lambda} = -\frac{\dot{\lambda} - U_0(t)}{R}\delta\lambda, \tag{3.20}$$

$$\frac{\dot{\lambda}}{R} + \frac{\lambda}{L(y)} = \frac{U_0(t)}{R}. \tag{3.21}$$

If the spool is not in contact $(y + y_0) > 0$:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{y}}\right) - \frac{\partial \mathcal{L}}{\partial y} = -r\dot{y}, \tag{3.22}$$

$$m\ddot{y} + r\dot{y} + ky = \frac{1}{2}\lambda^2 L(y)^{-2} L'(y), \tag{3.23}$$

while in contact $(y + y_0) \leq 0$:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{y}}\right) - \frac{\partial \mathcal{L}}{\partial y} = -(r + r_c)\dot{y}, \tag{3.24}$$

$$m\ddot{y} + (r + r_c)\dot{y} + ky + k_c(y + y_0) = \frac{1}{2}\lambda^2 L(y)^{-2} L'(y). \tag{3.25}$$

The resulting differential equation (3.21) is the Kirchhoff's current rule, equations (3.23) and (3.25) are Newton's second law.

3.3. **Mechanical momentum and charge formulation (p, q).** Mechanical momenta are applied less frequently compared to displacements in dynamical problems [1]. In order to satisfy the admissibility constraint on the mechanical momenta, Newton's second law should be considered:

$$\dot{p}_m = \dot{p}_k + \dot{p}_r - \dot{p}_M, \tag{3.26}$$

where $\dot{p}_m$ is the time derivative of the linear momentum of the mass m, $\dot{p}_k$ is the spring force, $\dot{p}_r$ is the damping force, and $\dot{p}_M$ is the electromagnetic force.

The constraint equation for adequate mechanical momenta is obtained by taking the time integration of (3.26):

$$p_m = p_k + p_r - p_M. \tag{3.27}$$

The Lagrangian function and the virtual work of the non-conservative elements are given as:
if the spool is not in contact, i.e., $(y + y_0) > 0$:

$$\mathcal{L}(p, \dot{p}, \dot{q}) = \frac{\dot{p}_k^2}{2k} + \frac{1}{2}L(y)\dot{q}^2 - \frac{(p_k + p_r - p_M)^2}{2m}, \tag{3.28}$$

$$\overline{\delta W}_{nc} = U_0(t)\delta q - R\dot{q}\delta q - \frac{\dot{p}_r}{r}\delta p_r, \tag{3.29}$$

while in contact, i.e., $(y + y_0) \leq 0$:

$$\mathcal{L}(p,\dot{p},\dot{q}) = \frac{(ky_0)^2}{2k} + y_0(\dot{p}_k - ky_0) + \frac{(\dot{p}_k - ky_0)^2}{2(k+k_c)} + \frac{1}{2}L(y)\dot{q}^2 - \frac{(p_k + p_r - p_M)^2}{2m}, \quad (3.30)$$

$$\overline{\delta W}_{nc} = U_0(t)\delta q - R\dot{q}\delta q - \frac{\dot{p}_r}{r+r_c}\delta p_r. \quad (3.31)$$

In equations (3.28)–(3.31) the displacement y and the momenta p_m, p_M are not independent variables; they can be given by two constitutive equations in which $\dot{q}$, p_k and p_r are the independent variables:

$$\dot{y} = \frac{(p_k + p_r - p_M)}{m}, \quad (3.32)$$

$$\dot{p}_M = -\frac{\partial W_m^*}{\partial y} = -\frac{1}{2}L'(y)\dot{q}^2. \quad (3.33)$$

The Lagrange equations are of the form

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}}\right) - \frac{\partial \mathcal{L}}{\partial q} = U_0(t) - R\dot{q}, \quad (3.34)$$

$$L(y)\ddot{q} + R\dot{q} + L'(y)\dot{y}\dot{q} = U_0(t. \quad (3.35)$$

Substituting equation (3.32) into equation (3.35) yields:

$$L(y)\ddot{q} + R\dot{q} + L'(y)\frac{(p_k + p_r - p_M)}{m}\dot{q} = U_0(t). \quad (3.36)$$

If the spool is not in contact $(y + y_0) > 0$:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{p}_k}\right) - \frac{\partial \mathcal{L}}{\partial p_k} = 0, \quad (3.37)$$

$$\frac{\ddot{p}_k}{k} + \frac{p_k + p_r - p_M}{m} = 0, \quad (3.38)$$

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{p}_r}\right) - \frac{\partial \mathcal{L}}{\partial p_r} = -\frac{\dot{p}_r}{r}, \quad (3.39)$$

$$\frac{\dot{p}_r}{r} + \frac{p_k + p_r - p_M}{m} = 0, \quad (3.40)$$

while in contact $(y + y_0) \leq 0$:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{p}_k}\right) - \frac{\partial \mathcal{L}}{\partial p_k} = 0, \quad (3.41)$$

$$\frac{\ddot{p}_k}{k+k_c} + \frac{p_k + p_r - p_M}{m} = 0, \quad (3.42)$$

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{p}_r}\right) - \frac{\partial \mathcal{L}}{\partial p_r} = -\frac{\dot{p}_r}{r + r_c}, \tag{3.43}$$

$$\frac{\dot{p}_r}{r + r_c} + \frac{p_k + p_r - p_M}{m} = 0. \tag{3.44}$$

The resulted ordinary differential equation (3.36) is Kirchhoff's voltage rule, equations (3.38), (3.40), (3.42), and (3.44) are the kinematical connection equations.

3.4. **Mechanical momentum and flux linkage formulation (p, λ).** This formulation is based on purely the generalized momenta, i.e., on the mechanical momenta and flux linkage. Here in addition to the admissibility constraints equations (3.26) and (3.27) valid for the mechanical momenta, Kirchhoff's voltage rule should also be satisfied for the flux linkage as in Subsection 3.2.

The Lagrangian function and the virtual work of the non-conservative elements are given as:
if the spool is not in contact, i.e., $(y + y_0) > 0$:

$$\mathcal{L}(p, \dot{p}, \lambda) = \frac{\dot{p}_k^2}{2k} - \frac{(p_k + p_r - p_M)^2}{2m} - \frac{\lambda^2}{2L(y)}, \tag{3.45}$$

$$\overline{\delta W}_{nc} = -\frac{\dot{\lambda} - U_0}{R}\delta\lambda - \frac{\dot{p}_r}{r}\delta p_r, \tag{3.46}$$

while in contact, i.e., $(y + y_0) \leq 0$:

$$\mathcal{L}(p, \dot{p}, \lambda) = \frac{(ky_0)^2}{2k} + y_0(\dot{p}_k - ky_0) + \frac{(\dot{p}_k - ky_0)^2}{2(k + k_c)} - \frac{(p_k + p_r - p_M)^2}{2m} - \frac{\lambda^2}{2L(y)}, \tag{3.47}$$

$$\overline{\delta W}_{nc} = -\frac{\dot{\lambda} - U_0}{R}\delta\lambda - \frac{\dot{p}_r}{r + r_c}\delta p_r. \tag{3.48}$$

In equations (3.45)–(3.48) y and p_M are also not independent; they should satisfy the following constitutive equations:

$$\dot{y} = \frac{(p_k + p_r - p_M)}{m}, \tag{3.49}$$

$$\dot{p}_M = \frac{\partial W_m}{\partial y} = -\frac{1}{2}L'(y)\lambda^2 L(y)^{-2}. \tag{3.50}$$

The Lagrange equations can be written as:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{\lambda}}\right) - \frac{\partial \mathcal{L}}{\partial \lambda} = -\frac{\dot{\lambda} - U_0}{R}, \tag{3.51}$$

$$\frac{\dot{\lambda} - U_0}{R} + \frac{\lambda}{L(y)} = 0. \tag{3.52}$$

If the spool is not in contact $(y + y_0) > 0$:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{p}_k}\right) - \frac{\partial \mathcal{L}}{\partial p_k} = 0, \tag{3.53}$$

$$\frac{\ddot{p}_k}{k} + \frac{p_k + p_r - p_M}{m} = 0, \tag{3.54}$$

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{p}_r}\right) - \frac{\partial \mathcal{L}}{\partial p_r} = -\frac{\dot{p}_r}{r}, \tag{3.55}$$

$$\frac{\dot{p}_r}{r} + \frac{p_k + p_r - p_M}{m} = 0, \tag{3.56}$$

while in contact $(y + y_0) \leq 0$:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{p}_k}\right) - \frac{\partial \mathcal{L}}{\partial p_k} = 0, \tag{3.57}$$

$$\frac{\ddot{p}_k}{k + k_c} + \frac{p_k + p_r - p_M}{m} = 0, \tag{3.58}$$

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{p}_r}\right) - \frac{\partial \mathcal{L}}{\partial p_r} = -\frac{\dot{p}_r}{r + r_c}, \tag{3.59}$$

$$\frac{\dot{p}_r}{r + r_c} + \frac{p_k + p_r - p_M}{m} = 0. \tag{3.60}$$

The resulting differential equation (3.52) is Kirchhoff's current rule, Equations (3.54), (3.56) (3.58), (3.60) are the kinematical connection equations.

4. Numerical analysis

A special purpose program has been developed using Scilab software for numerical solutions of the four set of systems of differential equations. The Runge-Kutta numerical method has been used in order to get the results. The problem shown in Figure 1 has been analyzed with the following parameters: $L_0 = 1.5\ mH$, $y_0 = 2.5\ mm$, $h = 10\ mm$, $R = 2.4\ \Omega$, $U_0 = 12\ V$, $m = 10\ g$, $k = 500\ N/m$, $r = 1.5\ Ns/m$, $k_c = 5 \cdot 10^6\ N/m$, and $r_c = 1500\ Ns/m$.

The displacement coordinate y versus time t has been computed with four different formulations and equal results have been obtained, which are shown in Figure 3. Equal diagrams of current $\dot{q}$ versus time t have been computed with (displacement – charge), and (momentum – charge) formulations (see Figure 4). Equal curves of flux linkage λ versus time t have been obtained by (displacement – flux linkage), and (momentum – flux linkage) formulations (see Figure 5). Equal curves have been determined for the momentum p_m of the spool versus time t by (momentum – charge) and (momentum – flux linkage) formulations (see Figure 6).

Multiple collisions of the spool with decreasing amplitude of rebounds can be recognized in Figure 3 and Figure 6, then the spool comes to a standstill. There are also

correlations between the curves of the current, the flux linkage and the motion of the spool (see Figure 5 and Figure 6).

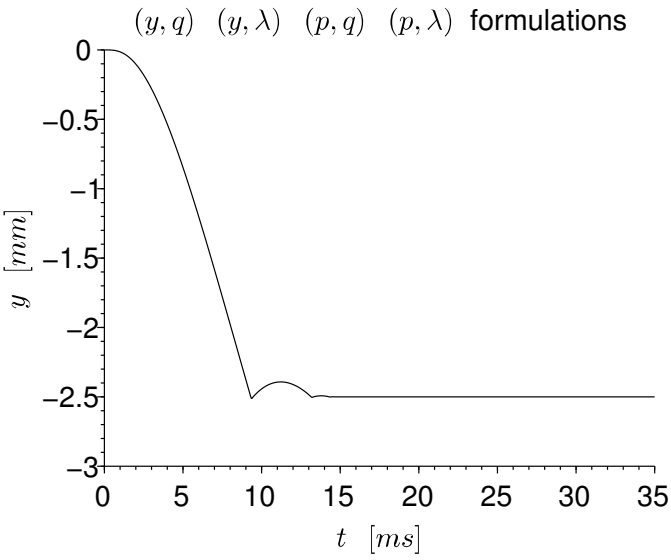


Figure 3. Displacement of the spool

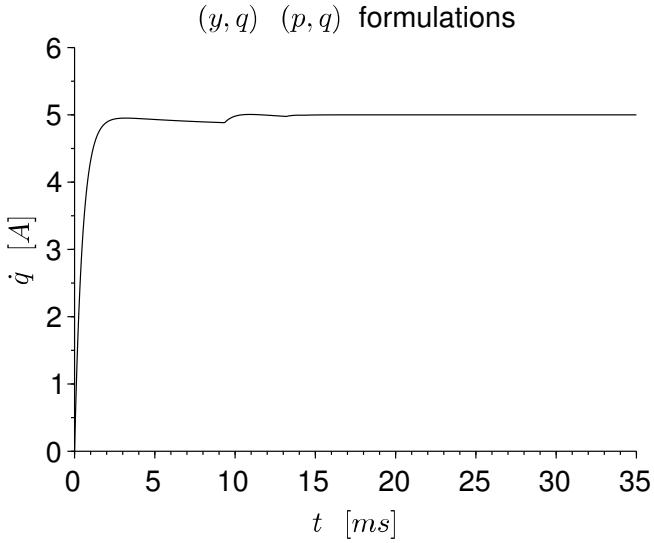


Figure 4. Current of the coil

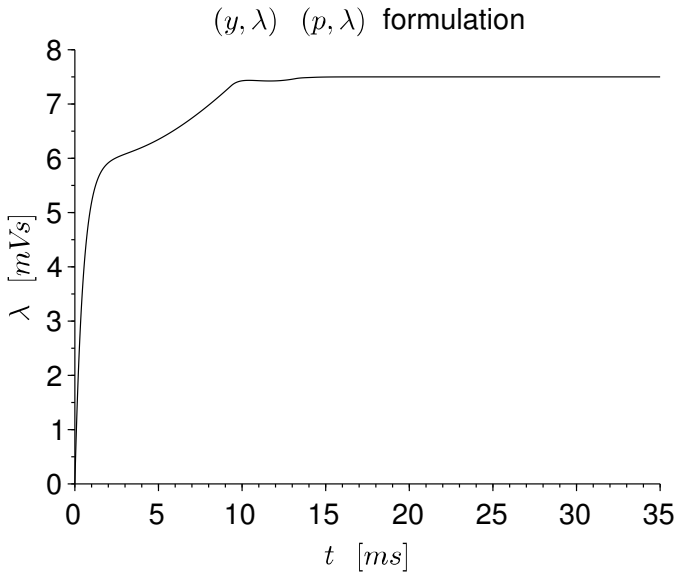


Figure 5. Flux linkage of the electromagnet

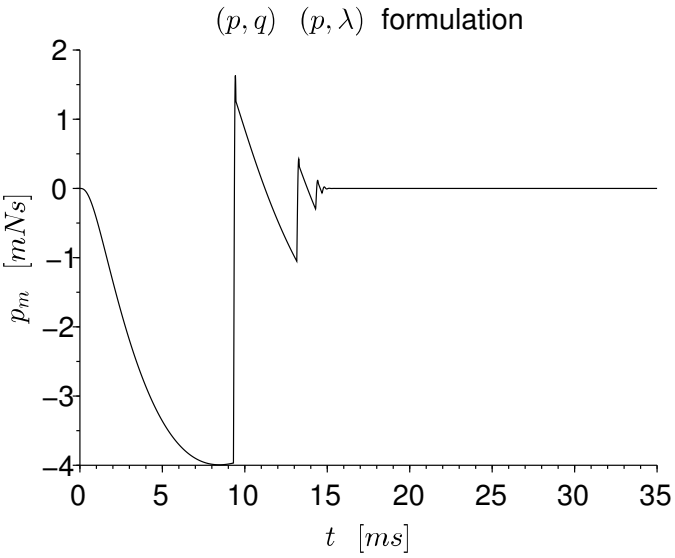


Figure 6. Momentum of the spool

5. Conclusions

Four equivalent formulations have been presented to model an electromechanical plunger using generalized coordinates and generalized momenta. The system of equations obtained by two formulations containing mechanical momenta can be regarded as novel results.

Numerical computations confirmed the equivalence of the four formulations. When a displacement coordinate is used in the formulations together with charge or flux linkage, three differential equations are obtained in both cases. However, seven differential equations are derived by mechanical momentum with charge or flux linkage formulation. Usually it is more convenient to satisfy the constraints for the generalized coordinates than their counterparts of momenta. Thus a formulation based on generalized coordinates is more popular compared to the generalized momenta formulations.

Acknowledgement. The described article was carried out as part of the EFOP-3.6.1-16-2016-00011 "Younger and Renewing University – Innovative Knowledge City – institutional development of the University of Miskolc aiming at intelligent specialisation" project implemented in the framework of the Szechenyi 2020 program. The realization of this project is supported by the European Union, co-financed by the European Social Fund.

References

1. P. E. Wellstead. *Introduction to Physical System Modelling.* Academic Press Ltd., London, 1979.
2. A. Preumont. *Mechatronics - Dynamics of Electromechanical and Piezoelectric Systems.* Springer-Verlag, Berlin, Heidelberg, 2006. DOI: 10.1007/1-4020-4696-0.
3. A. Demarchi, L. Farçoni, A. Pinto, R. Lang, R. Romero, and I. Silva. "Modelling a Solenoid's Valve Movement." *RoboCup 2017: Robot World Cup XXI.* Ed. by H. Akiyama, O. Obst, C. Sammut, and F. Tonidandel. Cham: Springer International Publishing, 2018, pp. 290–301. DOI: 10.1007/978-3-030-00308-1_24.
4. H. Tian and Y. Zhao. "Coil inductance model based solenoid on-off valve spool displacement sensing via laser calibration." *Sensors,* **18**(12), (2018), pp. 1–14. DOI: 10.3390/s18124492.
5. A. E. Balau and C. Lazar. "State-space model of an electro-hydraulic actuated wet clutch." *IFAC Proceedings Volumes, 6th IFAC Symposium Advances in Automotive Control, Munich, Germany.* IFAC, 2010, pp. 506–511. DOI: 10.3182/20100712-3-DE-2013.00039.
6. Y. Volpert, T. Szabó, I. Páczelt, and B. Szabó. "Application of the space enrichment method to problems of mechanical contact." *Finite Elements in Analysis and Design,* **24**(3), (1997), pp. 157–170. DOI: 10.1016/S0168-874X(96)00047-9.
7. A. C. Yudell and J. D. Van de Ven. "Predicting solenoid valve spool displacement through current analysis." *International Journal of Fluid Power,* **16**(3), (2015), pp. 133–140. DOI: 10.1080/14399776.2015.1068549.

Journal of Computational and Applied Mechanics, Vol. 15, No. 2, (2020), pp. 185–207
DOI: 10.32973/jcam.2020.011

HYPERSINGULAR BOUNDARY INTEGRAL EQUATIONS FOR PLANE ORTHOTROPIC ELASTICITY IN TERMS OF FIRST-ORDER STRESS FUNCTIONS

Sándor Szirbik
Institute of Applied Mechanics, University of Miskolc
H-3515 Miskolc, Miskolc–Egyetemváros, Hungary
sandor.szirbik@uni-miskolc.hu

[Received: August 18, 2020; Accepted: October 21, 2020]

Dedicated to Professor Barna Szabó on the occasion of his 85th birthday

Abstract. This paper is intended to present an implementation of the hypersingular boundary integral equations in terms of first-order stress functions for stress computations in plane orthotropic elasticity. In general, the traditional computational technique of the boundary element method used for computing the stress distribution on the boundary and close to it is not as accurate as it should be. In contrast, the accuracy of stress computations on the boundary is greatly increased by applying the hypersingular integral equations. Contrary to the method in which the solution is based on an approximation of displacement field, here the first-order stress functions and the rigid body rotation are the fundamental variables. An advantage of this approach is that the stress components can be obtained directly from the stress functions, there is, therefore, no need for Hooke's law, which should be used when they are computed from displacements. In addition, the computational work can be reduced when the stress distribution is computed at an arbitrary point on the boundary. The numerical examples presented prove the efficiency of this technique.

Mathematical Subject Classification: 74S15, 45F15
Keywords: First-order stress functions, plane orthotropic elasticity, hypersingular formulation

1. Introduction

The theory of elasticity for orthotropic bodies is well established in the boundary element method. Paper [1] studies orthotropic plane problems and presents the most important relations for inner regions by the boundary element method, in which the displacements and the stresses are the unknowns on the boundary and the numerical solution is based on a constant approximation of the displacements and the stresses over the boundary elements. One of the most important advantages of using first-order stress functions is the fact that calculating stresses requires determining first derivatives only, whereas with stress functions of order two, the stresses are obtained in terms of the second derivatives. First derivatives are generally more convenient in

boundary element applications, although an additional equation is needed for symmetrizing the stress tensor. If the body considered is isotropic the boundary integral equations of the direct method for plane problems in terms of first-order stress functions are established in paper [2]. For orthotropic bodies [3] presents the two fundamental solutions of the basic equation and the Somigliana relations for inner and exterior regions, provided that the first-order stress functions are the basic variables and constant strains and stresses are assumed at infinity. The authors of paper [4] modified the Somigliana formulas for plane orthotropic bodies and exterior regions by assuming that the strains are constant and accordingly the displacements are linear at infinity.

When using the direct boundary element method, weakly and strongly singular integrals have to be computed. This is the main reason for obtaining partly inaccurate results for the stresses at those points located to the region's boundary. However, the accuracy of the numerical solutions could be increased by using an appropriately chosen hypersingular formulation, because in this way the terms containing singularities can be computed in semi-analytical forms. [5] came to the conclusion that a deep analysis of the hypersingular formulation provides an opportunity to increase the accuracy of the stress components computed at interior points close to the contour and on the contour itself. The theory of the hypersingular boundary integral equations, in which the displacements are the fundamental variables, is developed further for 2D and 3D problems [6–14] and there are a number of issues (see for instance [15–18]) or [19] for plane orthotropic elasticity) to which they can be applied successfully.

The goal of this paper is to utilize the integral equations in terms of first-order stress functions with hypersingular kernels [20] for the problems of plane orthotropic elasticity in order to make the stress computations more accurate on the boundary. Section 2 outlines the most important preliminaries, i.e., it presents the governing equations of the plane strain problem in terms of first-order stress functions for the orthotropic case, details the corresponding fundamental solutions, and clarifies what the boundary conditions are for the first-order stress functions. The concept of the strain boundary conditions is also introduced. Section 3 is devoted to the problem of how to improve the hypersingular integral formulation for boundary value problems under plane strain conditions. The necessary equations are set up in a suitable form in Section 4, which is devoted to some further important computational issues so that an appropriate algorithm can be developed for the stress computations. The formulas for calculating the stresses are derived in semi-analytical forms. Section 5 presents the applicability of the algorithm for two simple problems and a more difficult one. The last section is the conclusions. Some longer formulas are presented in the Appendix.

2. Preliminaries

Throughout this paper a two-dimensional rectangular Cartesian coordinate system is used in which $x(x_1, x_2)$ and $y(y_1, y_2)$ are the source and field points. Greek subscripts are assumed to have the range (1, 2). The summation convention is applied to dummy indices. In accordance with the notations introduced: $\delta_{\rho\lambda}$ is the Kronecker symbol,

$\epsilon_{\rho\pi 3}$ is the permutation symbol and derivatives with respect to y_ρ are denoted by a Greek subscript preceded by a comma.

Consider a simply connected inner region $\mathcal{A}_I$ and its supplementary exterior region $\mathcal{A}_E$. We stipulate that the contour $\mathcal{L}_o$, which is divided into two parts denoted by $\mathcal{L}_t$ and $\mathcal{L}_u$, admits a nonsingular parametrization in terms of its arc length s. The positive direction on $\mathcal{L}_o$ is the one which leaves the region on the left side. The outward unit normal and the unit tangent on the contour are denoted by n_λ and τ_π.

The quantities that are prescribed on the boundary are denoted by hatted letters. [Stress functions $\mathfrak{u}_\lambda(s)$ (obtainable from the prescribed tractions)] {Strain boundary conditions, i.e., the derivatives $\mathfrak{t}_\lambda(s) = -\mathrm{d}\hat{u}_\lambda/\mathrm{d}s$} are prescribed on $[\mathcal{L}_t]$ $\{\mathcal{L}_u\}$. It is assumed that there are no body forces. The components of the elastic compliance tensor are denoted by s_{11}, $s_{12} = s_{21}$, s_{22} and s_{66}.

According to the Tonti scheme [21], stress functions are the fundamental variables, and the stresses and strains constitute the intermediate variables of the first and second kind in the dual system of elasticity. A prescribed incompatibility, which is in general zero, is the source variable. For the sake of a formal similarity to the boundary integral equations written in the usual forms, the opposite of the displacement derivative $\mathfrak{t}_\lambda = -\mathrm{d}\hat{u}_\lambda/\mathrm{d}s$ and the stress functions of order one $\mathfrak{u}_\lambda$ are further notations applied in our boundary element formalism. However, highlighting the difference between the method based on an approximation of displacements and the presented one, these variables and the related quantities are typeset in calligraphic fonts. In this way, the plane strain problem in terms of the first-order stress functions for the orthotropic case is governed by the dual kinematic equations

$$\begin{aligned} t_{11} &= \mathfrak{u}_{1,2}\,, & t_{12} &= \mathfrak{u}_{2,2}\,, \\ t_{21} &= -\mathfrak{u}_{1,1}\,, & t_{22} &= -\mathfrak{u}_{2,1}\,, \end{aligned} \tag{2.1}$$

the inverse Hooke's law

$$\begin{aligned} e_{11} &= s_{11}t_{11} + s_{12}t_{22}\,, \\ e_{12} = e_{21} &= \frac{s_{66}}{4}\left(t_{12} + t_{21}\right)\,, \\ e_{22} &= s_{21}t_{11} + s_{22}t_{22}\,, \end{aligned} \tag{2.2}$$

the compatibility conditions

$$e_{11,2} - e_{12,1} + \varphi_{3,1} = 0\,, \qquad e_{21,2} - e_{22,1} + \varphi_{3,2} = 0\,, \tag{2.3}$$

and the symmetry condition (equation of rotational equilibrium)

$$t_{12} = t_{21}\,. \tag{2.4}$$

With the prescribed traction $\hat{t}_\lambda(s)$, $(s \in \mathcal{L}_t)$ the first-order stress functions on the boundary are given by

$$\mathfrak{u}_\lambda(s) - \underbrace{\mathfrak{u}_\lambda(P)}_{C_\lambda} = \int_P^s \hat{t}_\lambda(\sigma)\mathrm{d}\sigma \qquad s \in \mathcal{L}_t\,, \tag{2.5}$$

where P is the starting point of the arc $\mathcal{L}_t$ and C_λ is an undetermined integration constant. Consequently, equation

$$\mathfrak{u}_\lambda(s) = \hat{\mathfrak{u}}_\lambda(s) + C_\lambda \qquad s \in \mathcal{L}_t \tag{2.6}$$

is the boundary condition imposed on the stress functions. Since no stresses belong to constant stress functions, C_λ can be set to zero.

The strain boundary conditions on $\mathcal{L}_u$ have the following form:

$$\mathfrak{t}_\lambda(s) = -\frac{\mathrm{d}\hat{u}_\lambda(s)}{\mathrm{d}s} = -n_\rho[\epsilon_{\rho\pi 3}e_{\pi\lambda} - \delta_{\rho\lambda}\varphi_3] \qquad s \in \mathcal{L}_u \,, \tag{2.7}$$

where $\hat{u}_\lambda(s)$ is the vector of the prescribed displacements on $\mathcal{L}_u$.

The outer region $\mathcal{A}_E$ is understood as the part of the coordinate plane outside the closed curve $\mathcal{L}_o$. The stresses at infinity denoted by $t_{11}(\infty)$, $t_{12}(\infty) = t_{21}(\infty)$, $t_{22}(\infty)$ are all constants. It is assumed that the rigid body rotation vanishes at infinity:

$$\varphi_3(\infty) = 0 \,. \tag{2.8}$$

Let $c_\kappa(\infty)$ be a constant. The stress functions

$$\tilde{\mathfrak{u}}_\kappa(y_\alpha) = \epsilon_{\alpha 3\rho}y_\alpha t_{\kappa\rho}(\infty) + c_\kappa(\infty) \tag{2.9}$$

result in a constant stress state $t_{11}(\infty)$, $t_{12}(\infty) = t_{21}(\infty)$, $t_{22}(\infty)$ at the point y. It is obvious that no stresses belong to the stress function $c_\kappa(\infty)$.

After eliminating the intermediate variables, equation system (2.1)-(2.4) yields the fundamental equations for the unknowns $\mathfrak{u}_\lambda$ and φ_3. The fundamental solutions of the dual system can be given in a relatively simple form if on the basis of [3] we introduce the following notations:

$$D_1 = \frac{2s_{21} + s_{66}}{2s_{11}} \,, \qquad D = \left[\frac{2s_{21} + s_{66}}{2s_{11}}\right]^2 - \frac{s_{22}}{s_{11}} \,, \tag{2.10a}$$

$$b_1 = \sqrt{D_1} - \sqrt{D} \,, \qquad b_2 = \sqrt{D_1} + \sqrt{D} \,, \tag{2.10b}$$

$$\beta_1^2 = -\frac{2s_{21} + s_{66}}{2s_{11}} + \sqrt{\left(\frac{2s_{21} + s_{66}}{2s_{11}}\right)^2 - \frac{s_{22}}{s_{11}}} \,, \tag{2.10c}$$

$$\beta_2^2 = -\frac{2s_{21} + s_{66}}{2s_{11}} - \sqrt{\left(\frac{2s_{21} + s_{66}}{2s_{11}}\right)^2 - \frac{s_{22}}{s_{11}}} \,, \tag{2.10d}$$

$$d_1 = -\begin{vmatrix} 1 & \tilde{\beta}_1 & \tilde{\beta}_1^2 \\ 1 & \beta_2 & \beta_2^2 \\ 1 & \tilde{\beta}_2 & \tilde{\beta}_2^2 \end{vmatrix} \,, \qquad d_2 = -\begin{vmatrix} 1 & \beta_1 & \beta_1^2 \\ 1 & \tilde{\beta}_1 & \tilde{\beta}_1^2 \\ 1 & \tilde{\beta}_2 & \tilde{\beta}_2^2 \end{vmatrix} \,, \tag{2.10e}$$

$$I_\alpha = 4\pi\frac{d_\alpha}{b_\alpha + 1}\left[b_\alpha^3 s_{11} + b_\alpha^2\left(s_{21} + s_{66}\right) - b_\alpha s_{12} - s_{22}\right] \,, \quad K = -\frac{1}{I_1 + I_2} \,, \tag{2.10f}$$

where $\tilde{\beta}_\alpha$ is the complex conjugate of β_α, and

$$\rho_\alpha = r_1 + \beta_\alpha r_2 \,, \tag{2.10g}$$

in which $r_\rho = x_\rho - y_\rho$.

The fundamental solution of order one $\mathfrak{U}_{\kappa\lambda}(x, y)$ in component form is given by the following equations:

$$\mathfrak{U}_{11}(x, y) = -K\,\mathrm{Im}\sum_{\alpha=1}^{2} d_\alpha\left(2\ln\rho_\alpha + 3\right)\beta_\alpha^2 \,, \tag{2.11a}$$

$$\mathfrak{U}_{12}(x,y) = K\mathrm{Im}\sum_{\alpha=1}^{2} d_\alpha \left(2\ln\rho_\alpha + 3\right)\beta_\alpha = \mathfrak{U}_{21}(x,y)\ , \tag{2.11b}$$

$$\mathfrak{U}_{22}(x,y) = -K\mathrm{Im}\sum_{\alpha=1}^{2} d_\alpha \left(2\ln\rho_\alpha + 3\right)\ . \tag{2.11c}$$

As regards the physical meaning of the fundamental solutions of order one the product $e_\kappa(x)\mathfrak{U}_{\kappa\lambda}(x,y)$ is the dual displacement $\mathfrak{u}_\lambda(y)$ due to a unit incompatibility $e_\kappa(x)$.

Let now y be a point on the contour. The fundamental solution of order two $\mathfrak{T}_{\kappa\lambda}(x,y)$ has the following components:

$$\mathfrak{T}_{11}(x,y) = K\mathrm{Im}\sum_{\alpha=1}^{2} \frac{2d_\alpha}{\rho_\alpha}\Big[n_2\left(-\beta_\alpha^3 s_{11} - s_{12}\beta_\alpha\right) - n_1\frac{s_{66}}{2}\beta_\alpha^2 - \\ -n_1(s_{22} + (s_{21} + \frac{s_{66}}{2})\beta_\alpha^2)\Big]\ , \tag{2.12a}$$

$$\mathfrak{T}_{12}(x,y) = K\mathrm{Im}\sum_{\alpha=1}^{2} \frac{2d_\alpha}{\rho_\alpha}\Big[\frac{n_2}{2}s_{66}\beta_\alpha^2 + n_1\left(s_{21}\beta_\alpha^3 + s_{22}\beta_\alpha\right) - \\ -n_2(s_{22} + (s_{21} + \frac{s_{66}}{2})\beta_\alpha^2)\Big]\ , \tag{2.12b}$$

$$\mathfrak{T}_{21}(x,y) = K\mathrm{Im}\sum_{\alpha=1}^{2} \frac{2d_\alpha}{\rho_\alpha}\Big[n_2\left(\beta_\alpha^2 s_{11} + s_{12}\right) + n_1\frac{s_{66}}{2}\beta_\alpha - \\ -n_1\beta_\alpha((s_{12} + \frac{s_{66}}{2}) + s_{11}\beta_\alpha^2)\Big]\ , \tag{2.12c}$$

$$\mathfrak{T}_{22}(x,y) = K\mathrm{Im}\sum_{\alpha=1}^{2} \frac{2d_\alpha}{\rho_\alpha}\Big[-\frac{n_2}{2}s_{66}\beta_\alpha - n_1\left(s_{21}\beta_\alpha^2 + s_{22}\right) - \\ -n_2\beta_\alpha((s_{21} + \frac{s_{66}}{2}) + s_{11}\beta_\alpha^2)\Big]\ , \tag{2.12d}$$

where the outward unit normal n_ρ is taken at y. The product $e_\kappa(x)\mathfrak{T}_{\kappa\lambda}(x,y)$ is the dual stress $\mathfrak{t}_\lambda(y)$ due to a unit incompatibility $e_\kappa(x)$.

It is important to note here that $\mathfrak{U}_{\kappa\lambda}(x,y)$ is weakly singular and $\mathfrak{T}_{\kappa\lambda}(x,y)$ is strongly singular. The elements of the elastic compliance tensor can be expressed with Young's moduli E_1, E_2, shear modulus μ_{12} and Poisson's ratios ν_{12}, ν_{21}

$$s_{11} = \frac{1}{E_1},\ s_{12} = -\frac{\nu_{21}}{E_2},\ s_{21} = -\frac{\nu_{12}}{E_1},\ s_{22} = \frac{1}{E_2},\ s_{66} = \frac{1}{\mu_{12}}\ . \tag{2.13}$$

For the inner region A_I the Somigliana formulas in the dual system of orthotropic plane elasticity are of the form

$$\mathfrak{u}_\kappa(y) = \oint_{\mathcal{L}_o} \mathfrak{U}_{\kappa\lambda}(x,y)\mathfrak{t}_\lambda(x)\,\mathrm{d}s_x - \oint_{\mathcal{L}_o} \mathfrak{T}_{\kappa\lambda}(x,y)\mathfrak{u}_\lambda(x)\,\mathrm{d}s_x \quad y \in \mathcal{A}_I, \tag{2.14a}$$

$$c_{\kappa\lambda}(y)\mathfrak{u}_\lambda(y) = \oint_{\mathcal{L}_o} \mathfrak{U}_{\kappa\lambda}(x,y)\mathfrak{t}_\lambda(x)\,\mathrm{d}s_x - \oint_{\mathcal{L}_o} \mathfrak{T}_{\kappa\lambda}(x,y)\mathfrak{u}_\lambda(x)\,\mathrm{d}s_x \quad y \in \mathcal{L}_o, \tag{2.14b}$$

$$0 = \oint_{\mathcal{L}_o} \mathfrak{U}_{\kappa\lambda}(x,y)\mathfrak{t}_\lambda(x)\,\mathrm{d}s_x - \oint_{\mathcal{L}_o} \mathfrak{T}_{\kappa\lambda}(x,y)\mathfrak{u}_\lambda(x)\,\mathrm{d}s_x \quad y \notin \mathcal{A}_I \cup \mathcal{L}_o, \tag{2.14c}$$

in which the matrix $c_{\kappa\lambda}(y)$ depends on the angle formed by the tangents to the contour at y ($c_{\kappa\lambda}(y) = \delta_{\kappa\lambda}/2$ if the contour is smooth at y).

For the outer region A_E equations (2.14) are slightly modified by taking the stress state at infinity into account:

$$\mathfrak{u}_\kappa(y) = \tilde{\mathfrak{u}}_\kappa(y) + \oint_{\mathcal{L}_o} \mathfrak{U}_{\kappa\lambda}(x,y)\mathfrak{t}_\lambda(x)\,\mathrm{d}s_x - \oint_{\mathcal{L}_o} \mathfrak{T}_{\kappa\lambda}(x,y)\mathfrak{u}_\lambda(x)\,\mathrm{d}s_x \quad y \in \mathcal{A}_E\ , \tag{2.15a}$$

$$c_{\kappa\lambda}(y)\mathfrak{u}_\lambda(y) = \tilde{\mathfrak{u}}_\kappa(y) + \oint_{\mathcal{L}_o} \mathfrak{U}_{\kappa\lambda}(x,y)\mathfrak{t}_\lambda(x)\,\mathrm{d}s_x - \oint_{\mathcal{L}_o} \mathfrak{T}_{\kappa\lambda}(x,y)\mathfrak{u}_\lambda(x)\,\mathrm{d}s_x \quad y \in \mathcal{L}_o\ , \tag{2.15b}$$

$$0 = \tilde{\mathfrak{u}}_\kappa(y) + \oint_{\mathcal{L}_o} \mathfrak{U}_{\kappa\lambda}(x,y)\mathfrak{t}_\lambda(x)\,\mathrm{d}s_x - \oint_{\mathcal{L}_o} \mathfrak{T}_{\kappa\lambda}(x,y)\mathfrak{u}_\lambda(x)\,\mathrm{d}s_x \quad y \notin \mathcal{A}_E \cup \mathcal{L}_o\,. \tag{2.15c}$$

The dual Somigliana formulas (2.14), (2.15) assume the same form as those established for the isotropic case [2] with the difference being in the fundamental solutions.

3. Hypersingular formulation for stress evaluations

In this section, the hypersingular formulation is applied to determine the stresses in the vicinity of the boundary with high accuracy. After solving integral equations (2.14b and (2.15b) of the direct method one obtains the values of $\mathfrak{u}_\lambda(x)$ and $\mathfrak{t}_\lambda(x)$ at every nodal point. The traditional approach to calculating the stress components at an arbitrary nodal point is the use of an equation system based partly on equations (2.1), (2.2) and on the polynomial approximation of $\mathfrak{u}_\lambda$ and $\mathfrak{t}_\lambda$ on the element that contains the nodal point selected [3]. In many cases the stresses obtained in this way are, however, not as accurate as they should be. For this reason we shall present a hypersingular integral equation formulation which significantly increases the accuracy of the stresses computed.

The line of thought is based on a technique developed by [5] in the primal system of plane elasticity. The partial derivatives of equation (2.15c) with respect to the coordinates y_ρ of the source point y should be determined taking equation (2.10g) into account, similarly to the case of isotropic materials [10], [20]. For this purpose, a small neighborhood of the source point y should be removed from the region of integration in order to obtain appropriate formulas. This way, singular integrals can be avoided – the kernels are singular at the point $x = y$. For convenience, the small circular domain $\mathcal{A}_\varepsilon$ with the source point as its center is removed from the original region $\mathcal{A}_E$. Consequently, the new contour consists of two arcs $\mathcal{L}_\varepsilon$ left from $\mathcal{L}_o$ after the removal of the small circle with radius ϵ and the circular arc s_ε.

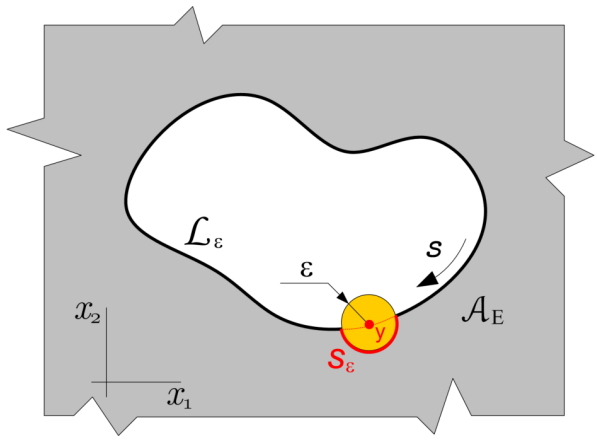


Figure 1. Outer region A_E bounded by the boundary curve $\mathcal{L}_\varepsilon \cup s_\varepsilon$.

Under this condition equation (2.15c) can be written in the form

$$0 = \lim_{\varepsilon\to 0}\left\{\tilde{\mathrm{u}}_\kappa(y) + \int_{\mathcal{L}_\varepsilon} [\mathfrak{U}_{\kappa\lambda}(x,y)\mathrm{t}_\lambda(x) - \mathfrak{T}_{\kappa\lambda}(x,y)\mathrm{u}_\lambda(x)]\,\mathrm{d}s_x + \right.$$
$$\left. + \int_{s_\varepsilon} [\mathfrak{U}_{\kappa\lambda}(x,y)\mathrm{t}_\lambda(x) - \mathfrak{T}_{\kappa\lambda}(x,y)\mathrm{u}_\lambda(x)]\,\mathrm{d}s_x\right\} . \quad (3.1)$$

After deriving equation (3.1) with respect to y_ρ, we arrive at

$$\lim_{\varepsilon\to 0}\left\{\tilde{\mathrm{u}}_{\kappa,\rho}(y) + \int_{\mathcal{L}_\varepsilon} [\mathfrak{U}_{\kappa\lambda,\rho}(x,y)\mathrm{t}_\lambda(x) - \mathfrak{T}_{\kappa\lambda,\rho}(x,y)\mathrm{u}_\lambda(x)]\,\mathrm{d}s_x + \right.$$
$$\left. + \int_{s_\varepsilon} \mathfrak{U}_{\kappa\lambda,\rho}(x,y)\mathrm{t}_\lambda(x) - \mathfrak{T}_{\kappa\lambda,\rho}(x,y)\mathrm{u}_\lambda(x)\,\mathrm{d}s_x\right\} = 0 , \quad (3.2)$$

where the derivatives $\mathfrak{U}_{\kappa\lambda,\rho}$ and $\mathfrak{T}_{\kappa\lambda,\rho}$ are presented in the Appendix – see equations (A.1) and (A.2). Due to the derivation, the order of singularity is increased, i.e., $\mathcal{D}_{\kappa\lambda\rho} = \mathfrak{U}_{\kappa\lambda,\rho}$ has a strong singularity of $O(\rho_\alpha^{-1})$, whereas $\mathcal{S}_{\kappa\lambda\rho} = \mathfrak{T}_{\kappa\lambda,\rho}$ is hypersingular of $O(\rho_\alpha^{-2})$.

Assume further that the stress functions of order one are continuous and differentiable in the neighborhood of the point y. Then it holds with good accuracy that

$$\mathrm{u}_\lambda(x) = \mathrm{u}_\lambda(y) + \mathrm{u}_{\lambda,\delta}(y)(x_\delta - y_\delta) + O(r^{1+\alpha}) . \quad (3.3)$$

It is also not too difficult to check that the opposite of the displacement derivative t_λ in the vicinity of the source point assumes the form

$$\mathrm{t}_\lambda(x) = -\frac{\partial u_\lambda}{\partial s} \cong -u_{\lambda,\pi}(y)\tau_\pi(x) = -u_{\lambda,\pi}(y)\epsilon_{3\rho\pi}n_\rho(x) . \quad (3.4)$$

Making use of these approximations, we can rewrite equation (3.2) into the following form:

$$\tilde{\mathfrak{u}}_{\kappa,\rho}(y)+$$
$$\lim_{\varepsilon\to 0}\left\{\int_{\mathcal{L}_\varepsilon}\left[\mathcal{D}_{\kappa\lambda\rho}\mathfrak{t}_\lambda(x)-\mathcal{S}_{\kappa\lambda\rho}\mathfrak{u}_\lambda(x)\right]\mathrm{d}s_x+\int_{s_\varepsilon}\mathcal{D}_{\kappa\lambda\rho}[\mathfrak{t}_\lambda(x)+u_{\lambda,\pi}(y)\tau_\pi(x)]\,\mathrm{d}s_x-\right.$$
$$-\int_{s_\varepsilon}\mathcal{S}_{\kappa\lambda\rho}[\mathfrak{u}_\lambda(x)-\mathfrak{u}_{\lambda,\delta}(y)(x_\delta-y_\delta)-\mathfrak{u}_\lambda(y)]\,\mathrm{d}s_x-\int_{s_\varepsilon}\mathcal{S}_{\kappa\lambda\rho}\mathrm{d}s_x\mathfrak{u}_\lambda(y)-$$
$$\left.-\int_{s_\varepsilon}\mathcal{D}_{\kappa\lambda\rho}u_{\lambda,\pi}(y)\tau_\pi(x)\mathrm{d}s_x-\int_{s_\varepsilon}\mathcal{S}_{\kappa\lambda\rho}\mathfrak{u}_{\lambda,\delta}(y)(x_\delta-y_\delta)\mathrm{d}s_x\right\}=0\,. \quad (3.5)$$

Observe that the second integral in the second line and the first integral in the third line are equal to zero if their limits are taken. The value of the last integral in the third line can be manipulated into the form

$$\lim_{\varepsilon\to 0}\left\{\int_{s_\varepsilon}\mathcal{S}_{\kappa\lambda\rho}\mathrm{d}s_x\mathfrak{u}_\lambda(y)\right\}=\lim_{\varepsilon\to 0}\left\{\frac{b_{\kappa\lambda\rho}(y)}{\varepsilon}\mathfrak{u}_\lambda(y)\right\}\,, \quad (3.6)$$

in which the coefficient $b_{\kappa\lambda\rho}(y)$ depends on the local geometry of S_ε at the source point y throughout the limiting process [10]. This formula clearly shows that the limit of the above integral – due to the fact that the kernel is singular – is unbounded and depends also on the value of $b_{\kappa\lambda\rho}(y)$. However, if the source point y is an internal point, then $b_{\kappa\lambda\rho}(y)=0$ – the proof is omitted here but can be done in the same way as in paper [5]. Hence taking the limit of the last two integrals in (3.5) yields

$$\lim_{\varepsilon\to 0}\left\{\int_{s_\varepsilon}\mathcal{D}_{\kappa\lambda\rho}u_{\lambda,\pi}(y)\tau_\pi(x)\mathrm{d}s_x+\int_{s_\varepsilon}\mathcal{S}_{\kappa\lambda\rho}\mathfrak{u}_{\lambda,\delta}(y)(x_\delta-y_\delta)\mathrm{d}s_x\right\}=$$
$$=d_{\kappa\alpha\beta\rho}(y)\mathfrak{u}_{\alpha,\beta}(y)\,, \quad (3.7)$$

where $d_{\kappa\alpha\beta\rho}(y)\mathfrak{u}_{\alpha,\beta}=0.5\mathfrak{u}_{\kappa,\rho}$ if the contour $\mathcal{L}_o$ is smooth at the source point y. Finally we have

$$\lim_{\varepsilon\to 0}\left\{\int_{\mathcal{L}_\varepsilon}\left[\mathcal{D}_{\kappa\lambda\rho}\mathfrak{t}_\lambda(x)-\mathcal{S}_{\kappa\lambda\rho}\mathfrak{u}_\lambda(x)\right]\mathrm{d}s_x\right\}=$$
$$=\lim_{\varepsilon\to 0}\left\{\frac{b_{\kappa\lambda\rho}}{\varepsilon}\mathfrak{u}_\lambda(y)\right\}+d_{\kappa\alpha\beta\rho}(y)\mathfrak{u}_{\alpha,\beta}(y)\,. \quad (3.8)$$

After performing the limiting process and assuming that the boundary curve is smooth, the Somigliana stress identity takes the following form at all boundary points y with the exception of 'non-smooth boundary points':

$$\tilde{\mathfrak{u}}_{\kappa,\rho}(y)+\frac{1}{2}\mathfrak{u}_{\kappa,\rho}(y)+\lim_{\varepsilon\to 0}\left\{\int_{\mathcal{L}_o}\left[\mathcal{S}_{\kappa\lambda\rho}\mathfrak{u}_\lambda(x)-\mathcal{D}_{\kappa\lambda\rho}\mathfrak{t}_\lambda(x)\right]\mathrm{d}s_x+\frac{b_{\kappa\lambda\rho}}{\varepsilon}\mathfrak{u}_\lambda(y)\right\}=0\,. \quad (3.9)$$

Assume further that contour $\mathcal{L}_o$ is divided into n_{BE} boundary elements $\mathcal{L}_e$ ($e=1,...,n_{BE}$), which have three nodes: the two end points and the middle point numbered locally as $a=1,...,3$ in the direction s. Thus, the usual isoparametric approximation is employed with quadratic boundary elements, which may be discontinuous if necessary. With the previous notations and the Somigliana stress identity (3.9), the elements of the stress tensor $t_{\psi\kappa}$ at the boundary point y can be obtained in the following form:

$$t_{\psi\kappa}(y) = \epsilon_{\psi 3\rho}\mathfrak{u}_{\kappa,\rho}(y) = 2\epsilon_{\psi 3\rho}\left\{-\tilde{\mathfrak{u}}_{\kappa,\rho}(y) + \right.$$

$$+ \sum_{e=1,\ e\neq m}^{n_{BE}} \int_{\mathcal{L}_e} \mathcal{D}_{\kappa\lambda\rho}(x,y)\,\mathfrak{t}_\lambda(x)\mathrm{d}s_x + \int_{\mathcal{L}_m} \mathcal{D}_{\kappa\lambda\rho}(x,y)\,\mathfrak{t}_\lambda(x)\mathrm{d}s_x -$$

$$\left. - \left[\sum_{e=1,\ e\neq m}^{n_{BE}} \int_{\mathcal{L}_e} \mathcal{S}_{\kappa\lambda\rho}(x,y)\,\mathfrak{u}^e_\lambda(x)\mathrm{d}s_x + \int_{\mathcal{L}_m} \mathcal{S}_{\kappa\lambda\rho}(x,y)\,\mathfrak{u}_\lambda(x)\mathrm{d}s_x\right]\right\} . \quad (3.10)$$

The superscript e identifies the e-th boundary elements, while the element containing the source point y is denoted by the index m. Therefore, the line integrals with singular kernels on the m-th boundary element can be separately treated in a semi-analytical form, while the other integrals can be determined numerically by using the standard Gauss integration technique.

4. Semi-analytical treatment of integrals containing singularities

Once the boundary fields $\mathfrak{u}_\lambda(x)$ and $\mathfrak{t}_\lambda(x)$ in terms of nodal values are known from the solution of the direct boundary integral equation method, the nodal values $\mathfrak{u}^a_\lambda$ and $\mathfrak{t}^a_\lambda$ $(a = 1,\ldots,3)$, which belong to boundary element e and local node number a, are set up in matrix form

$$[\mathfrak{u}^e_j]^T = \left[\begin{array}{cccccc} \mathfrak{u}^1_1 & \mathfrak{u}^1_2 & \mathfrak{u}^2_1 & \mathfrak{u}^2_2 & \mathfrak{u}^3_1 & \mathfrak{u}^3_2 \end{array}\right] \quad (4.1)$$

and

$$[\mathfrak{t}^e_j]^T = \left[\begin{array}{cccccc} \mathfrak{t}^1_1 & \mathfrak{t}^1_2 & \mathfrak{t}^2_1 & \mathfrak{t}^2_2 & \mathfrak{t}^3_1 & \mathfrak{t}^3_2 \end{array}\right] \quad (4.2)$$

on the e-th element $(j = 1,\ldots,6)$. It is obvious that the approximations of $\mathfrak{u}_\lambda$ and $\mathfrak{t}_\lambda$ over the e-th element are of the form:

$$\mathfrak{u}_\lambda = N_{\lambda j}(\xi)\mathfrak{u}^e_j\,, \qquad \mathfrak{t}_\lambda = N_{\lambda j}(\xi)\mathfrak{t}^e_j\,, \quad (4.3)$$

in which

$$[N_{\lambda j}(\xi)] = \left[\begin{array}{cccccc} N^1(\xi) & 0 & N^2(\xi) & 0 & N^3(\xi) & 0 \\ 0 & N^1(\xi) & 0 & N^2(\xi) & 0 & N^3(\xi) \end{array}\right] \quad (4.4)$$

is the shape function matrix constituted by the quadratic Lagrange polynomials

$$N^1(\xi) = \frac{1}{2}\xi(\xi-1)\,, \qquad N^2(\xi) = 1-\xi^2\,, \qquad N^3(\xi) = \frac{1}{2}\xi(\xi+1)\,, \quad (4.5)$$

which map the element onto the interval $\xi \in [-1,1]$ – isoparametric approximation. The line integral with the strong singularity $O(\rho_\alpha^{-1})$ taken over the m-th element in (3.10) is of the form

$$\tilde{I}^m_{\kappa j\rho} = \int_{\mathcal{R}^m} \mathcal{D}_{\kappa\lambda\rho}(\xi,\eta)\,N_{\lambda j}(\xi)J(\xi)\mathrm{d}\xi = \int_{\mathcal{R}^m} G^m_{\kappa j\rho}(\xi,\eta)\mathrm{d}\xi\,. \quad (4.6)$$

The line integral with the hypersingularity $O(\rho_\alpha^{-2})$ in (3.10) can be expressed in the following form:

$$I^m_{\kappa j\rho} = \int_{\mathcal{R}^m} \mathcal{S}_{\kappa\lambda\rho}\left(\xi,\eta\right) N_{\lambda j}(\xi) J(\xi) \mathrm{d}\xi = \int_{\mathcal{R}^m} F^m_{\kappa j\rho}(\xi,\eta) \mathrm{d}\xi \,. \tag{4.7}$$

Making use of the notations introduced the stress tensor $t_{\psi\kappa}$ at the boundary point y is given by

$$t_{\psi\kappa}(y) = 2\epsilon_{\psi 3\rho}\left\{-\tilde{\mathfrak{u}}_{\kappa,\rho}(y) + \sum_{e=1,e\neq m}^{n_{BE}} \int_{\mathcal{R}^e} \mathcal{D}_{\kappa\lambda\rho}\left(\xi,\eta\right) N_{\lambda j}(\xi) J(\xi) \mathrm{d}\xi \mathfrak{t}^e_j + \tilde{I}^m_{\kappa j\rho}\mathfrak{t}^m_j - \right.$$
$$\left. - \left[\sum_{e=1,e\neq m}^{n_{BE}} \int_{\mathcal{R}^e} \mathcal{S}_{\kappa\lambda\rho}\left(\xi,\eta\right) N_{\lambda j}(\xi) J(\xi) \mathrm{d}\xi \mathfrak{u}^e_j + I^m_{\kappa j\rho}\mathfrak{u}^m_j\right]\right\} . \tag{4.8}$$

Here $J(\xi)$ is the Jacobian, $\mathcal{R}^e$ is the mapping of $\mathcal{L}_e$ onto the interval $[-1,1]$. The meaning of $\mathcal{R}^m_\varepsilon$ can be read off Figure 2, which also shows the mapping of the boundary element $\mathcal{L}_m$ onto the axis ξ. The length of $\mathcal{R}^m_\varepsilon$ is denoted by $\alpha(\varepsilon)$, while η is the image of the source point y on this interval.

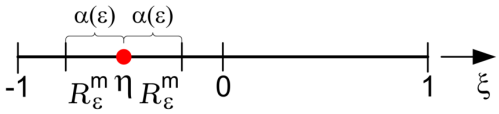


Figure 2. Illustration of the interval $\mathcal{R}^m_\varepsilon$ on the map of $\mathcal{L}_m$ on the axis ξ.

After determining the power series representation of $F^m_{\kappa j\rho}(\xi,\eta)$, the singular terms in the integral expression (4.7) can be isolated. If we want to give the power series of $\mathcal{S}_{\kappa\lambda\rho}\left(\xi,\eta\right)$ in the vicinity of the source point η, the power series expansion

$$\frac{1}{\rho_\alpha^2} = \frac{1}{(A_1+\beta_\alpha A_2)^2\delta^2} - 2\frac{B_1+\beta_\alpha B_2}{(A_1+\beta_\alpha A_2)^3\,\delta} + O(1) \tag{4.9}$$

is needed, in which $\delta = \xi - \eta$. The constants A_λ and B_λ can be obtained from the first two terms of the Taylor series of the position vector r_λ taken in the vicinity of the source point η:

$$r_\lambda = x_\lambda(\xi) - y_\lambda(\eta) = 0 + \left.\frac{\partial x_\lambda}{\partial\xi}\right|_{\xi=\eta}(\xi-\eta) + \frac{1}{2!}\left.\frac{\partial^2 x_\lambda}{\partial\xi^2}\right|_{\xi=\eta}(\xi-\eta)^2 + \ldots =$$
$$= A_\lambda\delta + B_\lambda\delta^2 + O(\delta^3)\,, \tag{4.10}$$

where

$$A_\lambda = (x^1_\lambda - 2x^2_\lambda + x^3_\lambda)\eta + \frac{1}{2}(x^3_\lambda - x^1_\lambda)\,, \quad B_\lambda = \frac{1}{2}(x^1_\lambda - 2x^2_\lambda + x^3_\lambda) \tag{4.11}$$

are constants at η; x_λ^1, x_λ^2 and x_λ^3 are the nodal coordinates in the usual representation. Moreover, let $A = J(\eta) = \sqrt{\sum_{\lambda=1}^{2} A_\lambda^2}$. Then the power series of the shape function matrix is of the form

$$N_{\lambda j}(\xi) = N_{\lambda j}(\xi)\Big|_{\xi=\eta} + \frac{\mathrm{d}}{\mathrm{d}\xi} N_{\lambda j}(\xi)\Big|_{\xi=\eta} (\xi-\eta) + O(\delta^2) = \\ = {}^0N_{\lambda j}(\eta) + {}^1N_{\lambda j}(\eta)\delta + O(\delta^2)\,, \quad (4.12)$$

where

$$[{}^0N_{\lambda j}(\eta)] = \begin{bmatrix} {}^0N^1 & 0 & {}^0N^2 & 0 & {}^0N^3 & 0 \\ 0 & {}^0N^1 & 0 & {}^0N^2 & 0 & {}^0N^3 \end{bmatrix}\,, \quad (4.13)$$

in which

$${}^0N^1(\eta) = \frac{1}{2}\eta(\eta-1)\,, \qquad {}^0N^2(\eta) = 1-\eta^2\,, \qquad {}^0N^3(\eta) = \frac{1}{2}\eta(\eta+1) \quad (4.14)$$

and

$$[{}^1N_{\lambda j}(\eta)] = \begin{bmatrix} {}^1N^1 & 0 & {}^1N^2 & 0 & {}^1N^3 & 0 \\ 0 & {}^1N^1 & 0 & {}^1N^2 & 0 & {}^1N^3 \end{bmatrix}\,, \quad (4.15)$$

in which

$${}^1N^1(\eta) = \frac{1}{2}(\eta-1)\,, \qquad {}^1N^2(\eta) = -2\eta\,, \qquad {}^1N^3(\eta) = \frac{1}{2}(\eta+1)\,. \quad (4.16)$$

For the product of the Jacobian and the vector n_ρ we obtain:

$$J_1(\xi) = n_1 J(\xi) = A_2 + 2B_2\delta + O(\delta^2)\,, \quad (4.17)$$

$$J_2(\xi) = n_2 J(\xi) = -A_1 - 2B_1\delta + O(\delta^2)\,. \quad (4.18)$$

After substituting power series (4.9)-(4.18) into integral (4.7), the integrand $F^m_{\kappa j\rho}(\xi,\eta)$ can be manipulated into the following Laurent series in the vicinity of the source point η:

$$F^m_{\kappa j\rho}(\xi,\eta) = \frac{{}^{-2}F^m_{\kappa j\rho}(\eta)}{(\xi-\eta)^2} + \frac{{}^{-1}F^m_{\kappa j\rho}(\eta)}{\xi-\eta} + O(1)\,, \quad (4.19)$$

where ${}^{-1}F^m_{\kappa j\rho}(\eta)$ and ${}^{-2}F^m_{\kappa j\rho}(\eta)$ are constant values at a fixed $\eta \in (-1,1)$. The details are presented in Appendix (A.3). Subtracting the first two terms of the power series of the integrand $F^m_{\kappa j\rho}(\xi,\eta)$ from equation (4.7) and investigating these terms separately, three line integrals are obtained:

$$I^m_{\kappa j\rho} = \lim_{\varepsilon\to 0}\left\{\int_{\mathcal{R}^m-2\mathcal{R}^m_\varepsilon}\left(F^m_{\kappa j\rho}(\xi,\eta) - \left[\frac{{}^{-2}F^m_{\kappa j\rho}(\eta)}{(\xi-\eta)^2} + \frac{{}^{-1}F^m_{\kappa j\rho}(\eta)}{\xi-\eta}\right]\right)\mathrm{d}\xi\right\} + \\ + \lim_{\varepsilon\to 0}\left\{\int_{\mathcal{R}^m-2\mathcal{R}^m_\varepsilon}\frac{{}^{-2}F^m_{\kappa j\rho}(\eta)}{(\xi-\eta)^2}\mathrm{d}\xi\right\} + \lim_{\varepsilon\to 0}\left\{\int_{\mathcal{R}^m-2\mathcal{R}^m_\varepsilon}\frac{{}^{-1}F^m_{\kappa j\rho}(\eta)}{\xi-\eta}\mathrm{d}\xi\right\} = \\ = {}^0I^m_{\kappa j\rho} + {}^{-2}I^m_{\kappa j\rho} + {}^{-1}I^m_{\kappa j\rho}\,. \quad (4.20)$$

Note that the integrand of ${}^{0}I^{m}_{\kappa j\rho}$ is a polynomial, i.e., it is not singular. Hence the limit

$$ {}^{0}I^{m}_{\kappa j\rho} = \int_{-1}^{1} \left\{ F^{m}_{\kappa j\rho}(\xi,\eta) - \left[\frac{{}^{-2}F^{m}_{\kappa j\rho}(\eta)}{(\xi-\eta)^2} + \frac{{}^{-1}F^{m}_{\kappa j\rho}(\eta)}{\xi-\eta} \right] \right\} \mathrm{d}\xi \tag{4.21} $$

is a regular integral, which can be computed by using the standard numerical integration techniques. We have chosen a 14-node Gauss-Legendre quadrature rule for the interval $[-1,1]$. The second and third integrals can be calculated analytically. After dividing $\mathcal{R}^m$ into two subintervals $[-1,-\alpha]$ and $[\alpha,1]$ separated from each other at the singular point η and making use of the series presented in the Appendix, the limit of the two last integrals in equation (4.20) can be obtained. Some paper and pencil calculations yield the following analytical result:

$$ {}^{-1}I^{m}_{\kappa j\rho} = \lim_{\varepsilon\to 0^-} \left\{ \int_{-1}^{\eta-\alpha} \frac{{}^{-1}F^{m}_{\kappa j\rho}(\eta)}{\xi-\eta}\mathrm{d}\xi \right\} + \lim_{\varepsilon\to 0^+} \left\{ \int_{\eta+\alpha}^{1} \frac{{}^{-1}F^{m}_{\kappa j\rho}(\eta)}{\xi-\eta}\mathrm{d}\xi \right\} = $$

$$ = \ln\left|\frac{1-\eta}{-1-\eta}\right| {}^{-1}F^{m}_{\kappa j\rho}(\eta) \ . \tag{4.22} $$

A similar procedure results in the last integral in equation (4.20) in a closed form:

$$ {}^{-2}I^{m}_{\kappa j\rho} = \lim_{\varepsilon\to 0^-} \left\{ \int_{-1}^{\eta-\alpha} \frac{{}^{-2}F^{m}_{\kappa j\rho}(\eta)}{(\xi-\eta)^2}\mathrm{d}\xi \right\} + \lim_{\varepsilon\to 0^+} \left\{ \int_{\eta+\alpha}^{1} \frac{{}^{-2}F^{m}_{\kappa j\rho}(\eta)}{(\xi-\eta)^2}\mathrm{d}\xi \right\} = $$

$$ = -\left[\frac{1}{1+\eta} + \frac{1}{1-\eta}\right] {}^{-2}F^{m}_{\kappa j\rho}(\eta) \ . \tag{4.23} $$

Consequently, formulas (4.21), (4.22) and (4.23) can be used to determine the integrals taken over the element that contains the singularity. It is obvious that the singular integrals can be given in closed form while the line integral without singularities is computed numerically. After collecting the results, the final formula for the evaluation of the hypersingular integral is presented in the following semi-analytical form:

$$ I^{m}_{\kappa j\rho} = \int_{-1}^{1} \left(F^{m}_{\kappa j\rho}(\xi,\eta) - \left[\frac{{}^{-2}F^{m}_{\kappa j\rho}(\eta)}{(\xi-\eta)^2} + \frac{{}^{-1}F^{m}_{\kappa j\rho}(\eta)}{\xi-\eta} \right] \right) \mathrm{d}\xi + $$

$$ + \ln\left|\frac{1-\eta}{-1-\eta}\right| {}^{-1}F^{m}_{\kappa j\rho}(\eta) - \left[\frac{1}{1+\eta} + \frac{1}{1-\eta}\right] {}^{-2}F^{m}_{\kappa j\rho}(\eta) \ , \tag{4.24} $$

where $\eta \in (-1,1)$. When the point y coincides with the endpoints ($\eta = -1$ or $\eta = 1$), one has to apply the formula below:

$$ I^{i}_{\kappa j\rho} = \int_{-1}^{1} \left(F^{i}_{\kappa j\rho}(\xi,\eta) - \left[\frac{{}^{-2}F^{i}_{\kappa j\rho}(\eta)}{(\xi-\eta)^2} + \frac{{}^{-1}F^{i}_{\kappa j\rho}(\eta)}{\xi-\eta} \right] \right) \mathrm{d}\xi + $$

$$ + \ln\left|\frac{2}{\vartheta_i}\right| sgn(\delta)\, {}^{-1}F^{i}_{\kappa j\rho}(\eta) - \left[\frac{\gamma_i}{\vartheta_i^2} sgn(\delta) + \frac{1}{2}\right] {}^{-2}F^{i}_{\kappa j\rho}(\eta) \ , \tag{4.25} $$

where $i = m$ if $\eta = 1$, $i = m+1$ if $\eta = -1$ and, in addition, $\vartheta_i(\eta) = A^{-1}$ and $\gamma_i(\eta) = -\dfrac{1}{A^4}(\sum\limits_{\lambda=1}^{2} A_\lambda B_\lambda)$ as proposed by [10].

The line integral with the strong singularity $O(\rho_\alpha^{-1})$ on the m-th element can be treated similarly to equation (4.7). The integrand $G^m_{\kappa j\rho}(\xi,\eta)$ in equation (4.6) is manipulated into the following series in the vicinity of the source point η:

$$G^m_{\kappa j\rho}(\xi,\eta) = \frac{{}^{-1}G^m_{\kappa j\rho}(\eta)}{\xi-\eta} + O(1)\ . \tag{4.26}$$

The details are presented in Appendix (A.4). The corresponding line integrals can be determined in the same manner as for equation (4.20):

$$\tilde{I}^m_{\kappa j\rho} = \lim_{\varepsilon\to 0}\left\{\int_{\mathcal{R}^m-2\mathcal{R}^m_\varepsilon}\left(G^m_{\kappa j\rho}(\xi,\eta) - \frac{{}^{-1}G^m_{\kappa j\rho}(\eta)}{\xi-\eta}\right)\mathrm{d}\xi\right\} + \\ + \lim_{\varepsilon\to 0}\left\{\int_{\mathcal{R}^m-2\mathcal{R}^m_\varepsilon}\frac{{}^{-1}G^m_{\kappa j\rho}(\eta)}{\xi-\eta}\mathrm{d}\xi\right\} = {}^{0}\tilde{I}^m_{\kappa j\rho} + {}^{-1}\tilde{I}^m_{\kappa j\rho}\ . \tag{4.27}$$

After finding the limit values in question, it turns out that the term ${}^{0}\tilde{I}^m_{\kappa j\rho}$ is not singular. It can, therefore, be computed numerically with sufficient accuracy. As regards the term ${}^{-1}\tilde{I}^m_{\kappa j\rho}$ we can express it analytically:

$$\tilde{I}^m_{\kappa j\rho} = \int_{-1}^{1}\left(G^m_{\kappa j\rho}(\xi,\eta) - \frac{{}^{-1}G^m_{\kappa j\rho}(\eta)}{\xi-\eta}\right)\mathrm{d}\xi + \ln\left|-\frac{1-\eta}{1+\eta}\right| {}^{-1}G^m_{\kappa j\rho}(\eta)\ . \tag{4.28}$$

If y coincides with an endpoint of two neighboring elements ($i = m$ for $\eta = 1$ and $i = m+1$ for $\eta = -1$) the following formula should be applied:

$$\tilde{I}^i_{\kappa j\rho} = \int_{-1}^{1}\left(G^i_{\kappa j\rho}(\xi,\eta) - \frac{{}^{-1}G^i_{\kappa j\rho}(\eta)}{\xi-\eta}\right)\mathrm{d}\xi + \ln\left|\frac{2}{\vartheta_i}\right| sgn(\delta)\,{}^{-1}G^i_{\kappa j\rho}(\eta)\ . \tag{4.29}$$

The above results show that the singularities can be integrated analytically over the m-th element. A program has been developed in Fortran 90 for finding numerical solutions.

5. Numerical examples

Problem 1. We shall consider a plane strain problem. It is assumed that there is a rigid circular inclusion in the plane of the strain with a radius $R = 10\,\mathrm{mm}$ and a constant stress state is prescribed at infinity: $t_{11} = 100$ [MPa], $t_{12}(\infty) = t_{21}(\infty) = t_{22}(\infty) = 0$. The stress functions that characterize a constant stress state are given by equation (2.9). The material of the body in plane strain is birch for which $s_{11} = 8.497\cdot 10^{-5}$ [mm^2/N], $s_{12} = s_{21} = -6.11\cdot 10^{-6}$ [mm^2/N], $s_{22} = 1.699\cdot 10^{-4}$ [mm^2/N] and $s_{66} = 1.456\cdot 10^{-3}$ [mm^2/N] are the elastic coefficients [3].

It is clear that the inner boundary is fixed. Then the strain boundary condition (2.7) that can be described on the circle is of Neumann type. Determination of the stresses on the inner boundary can serve as a benchmark for proving numerically that

the use of equations (4.24) and (4.25) results in increasing accuracy concerning the stress computations. The contour is divided into only 24 uniform quadratic elements. Though there are two undetermined integration constants in equation (2.6) which have, however, no effect on the stress computations, we must to set them to a prescribed value (to zero) at such nodal points on the boundary where this can be done with ease. For Problem 1 the stress function $\mathfrak{u}_1$ is set to zero at point A and the stress function $\mathfrak{u}_2$ is set to zero at point B – see Figure 3 for details.

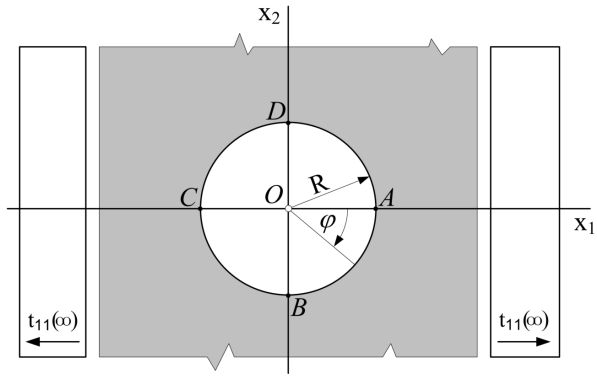


Figure 3. Outer region bounded by a circle for which a constant stress state is prescribed at infinity.

Table 1 contains the quotients $t_{rr}/t_{11}(\infty)$, $t_{r\varphi}/t_{11}(\infty)$ and $t_{\varphi\varphi}/t_{11}(\infty)$ obtained at various boundary points for the rigid circular inclusion using a polar coordinate system. Analytical results at given φ are obtained from Lekhnitskii's book [22].

Table 1. Results for the rigid inclusion.

	$\frac{t_{rr}}{t_{11}(\infty)}$			$\frac{t_{r\varphi}}{t_{11}(\infty)}$			$\frac{t_{\varphi\varphi}}{t_{11}(\infty)}$		
φ	HBIE	DBEM	anal.	HBIE	DBEM	anal.	HBIE	DBEM	anal.
$[rad]$		[3]	[22]		[3]	[22]		[3]	[22]
0	1.1364	1.1363	1.237	0.0000	0.0000	0.000	0.0449	0.0444	0.044
$\pi/12$	1.1562	1.1558	1.156	−2.9936	−2.9999	−0.299	0.0934	0.0936	0.093
$\pi/6$	0.9370	0.9364	0.937	−0.5185	−0.5188	−0.519	0.2696	0.2701	0.270
$\pi/4$	0.6377	0.6370	0.698	−0.5987	−0.5986	−0.599	0.5154	0.5158	0.516
$\pi/3$	0.3383	0.3377	0.338	−0.5185	−0.5181	−0.519	0.6989	0.6990	0.699
$5\pi/12$	0.1192	0.1188	0.119	−0.2994	−0.2987	−0.299	0.5637	0.5627	0.564
$\pi/2$	0.0390	0.0389	0.039	0.0000	0.0000	0.000	0.0044	0.0028	0.003

The present HBIE solutions show very good agreement with the literature data [22] and are also more accurate than the solutions from the direct boundary element method (DBEM) [3] with the same coarse mesh.

Figure 4 depicts the hypersingular solutions for the stresses on the quarter $\varphi \in [0,\ \pi/2]$ of the circle. The numerical results (lines) show good agreement with the analytical results (marks): t_{rr} (solid line, diamonds), $t_{\varphi\varphi}$ (dot dot dashed line, boxes)

and $t_{r\varphi}$ (dashed line, circles). The stress distributions plotted from the stress values computed at arbitrarily selected points per boundary element match very well with the marks. The further advantage of using HBIE, which is also demonstrated in the following figures (Figure 4, 5, 7 and 8), is its ability to directly calculate the stress components anywhere on the boundary, contrary to the direct method [3].

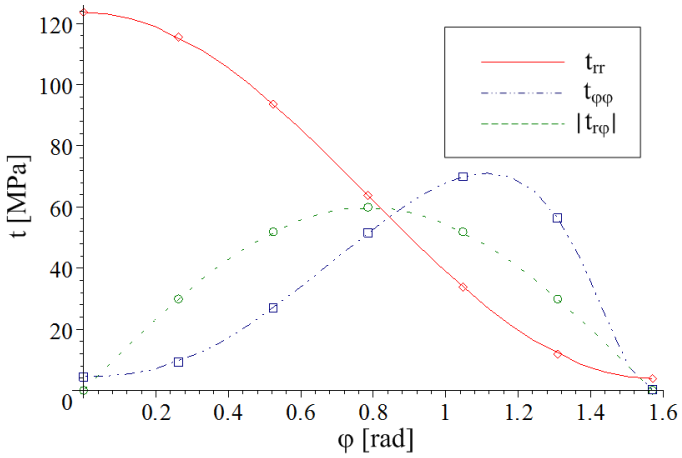


Figure 4. Exact and numerical solutions for the rigid circular inclusion as a function φ.

Problem 2. We consider the same outer region including a circular hole with radius $R = 10\,\text{mm}$ with the previous constant stress state prescribed at infinity. The inner boundary is, however, free; thus, the stress functions are zero on the whole boundary (Dirichlet boundary condition). Through this example, equations (4.28) and (4.29) can be separately validated for computational accuracy.

Using a polar coordinate system, Table 2 represents the numerical results for $t_{\varphi\varphi}/t_{11}(\infty)$ obtained at various boundary points.

Table 2. Results for the circular hole.

	$\frac{t_{\varphi\varphi}}{t_{11}(\infty)}$		
φ	HBIE	DBEM	anal.
$[rad]$		[3]	[22]
0	−0.7075	−0.7074	−0.707
$\pi/12$	−0.3399	−0.3393	−0.340
$\pi/6$	0.0691	0.0695	0.069
$\pi/4$	0.4039	0.4045	0.404
$\pi/3$	0.9664	0.9661	0.966
$5\pi/12$	2.5771	2.5774	2.577
$\pi/2$	5.4526	5.4541	5.453

Figure 5 depicts the hypersingular solutions for the stress $t_{\varphi\varphi}$ on the first quarter $\varphi \in [0, \pi/2]$ of the circle: $t_{rr} = 0$, $t_{\varphi\varphi}$ (dot dot dashed line, boxes) and $t_{r\varphi} = 0$.

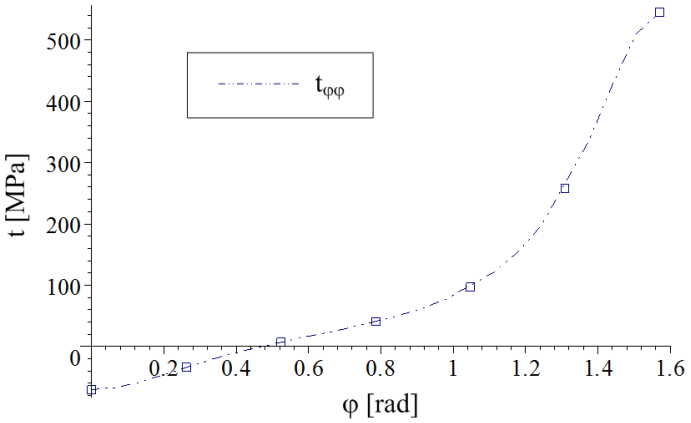


Figure 5. Exact and numerical solutions for the circular hole as function φ.

The results of the present HBIE solutions in comparison with the existing numerical [3] and analytical [22] solutions show a high accuracy and good fitting at various boundary point.

Problem 3. We consider an L-shape domain (see Figure 6) in order to investigate the effect of singularities arising at the reentrant corner, in the vicinity of point C. In the computations the following dimensions of the L-shape are chosen: $b_1 = h_1 = 100$ mm, $b_2 = h_2 = 50$ mm, and under the given loading ($p_1 = 5$ MPa and $p_2 = 10$ MPa) an equilibrium state is investigated. Our results are compared with those of the commercial software Abaqus, to assess the accuracy and efficiency of the method presented in this paper.

According to papers [23], [24], [25] the finite element method (FEM) mesh for Problem 3 has 5281 nodes and 16232 quadratic elements (CPE8R, 8-node biquadratic plane strain quadrilateral, reduced integration, approximate global size = 1.5 mm), and a bias ratio of 7 is used along the lines BC and CD to refine the mesh in the vicinity of point C. The components of the elastic compliance tensor are the same as in the previous examples, therefore, the elastic constants are $E_1 = 11768.86$ MPa, $E_2 = 5885.82$ MPa, $\mu_{12} = 686.81$ MPa and $\nu_{12} = 0.072$ (see equation (2.13)).

For our BEM solution, the traction boundary conditions with the zero values for C_λ at point O are, formally, given by equation (2.6), where $\hat{\mathrm{u}}_\lambda(s)$ is either constant or linear with gradients p_1 or p_2 on the boundary. The boundary of the L-shape domain has been discretized by 160 uniform quadratic elements with a length of 2.5 mm so that the lines BC and CD are divided into 20 elements. But even in that case, the

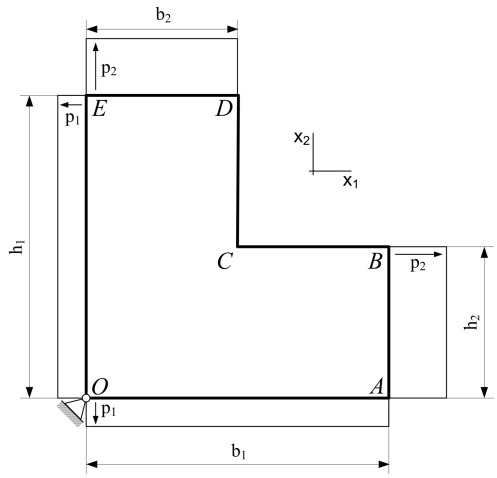


Figure 6. Geometry and boundary conditions for L-shape domain in an equilibrium state.

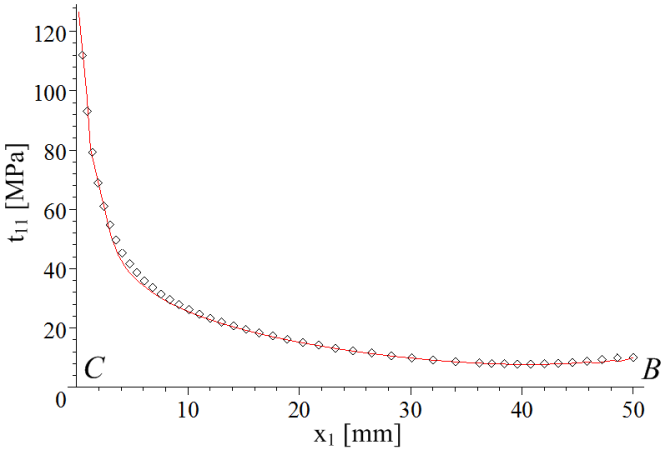


Figure 7. Comparison between FEM (◊) and HBIE solution (—) along the line BC.

stress solutions using HBIE show a good agreement with the Abaqus results denoted by diamond marks. Nevertheless, there is little difference between the FEM and BEM results in the vicinity of the corner point where the local mesh refinement is used in FEM analysis for stress singularity [24]. The stress t_{11} along line BC is plotted in Figure 7 to further demonstrate the efficiency of the method. $t_{12} = t_{22} = 0$. The stress t_{22} along line BC is plotted in Figure 8, while $t_{11} = t_{12} = 0$. Corner points B,

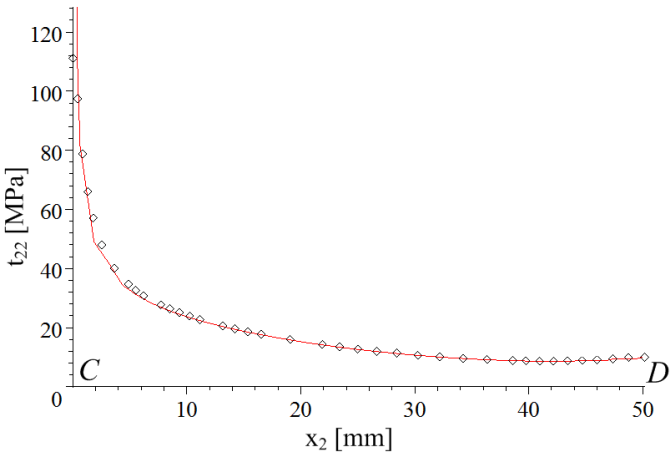


Figure 8. Comparison between FEM (◊) and HBIE solution (—) along the line CD.

C, D, E and F on the boundary of L-shaped domain do not lead to any difficulties since discontinuous boundary elements can be used at those corner points in BEM. This is a significant advantage over most collocation schemes.

6. Conclusions

This paper is devoted to an application of first-order stress functions in the hypersingular boundary integral formulation by which the stress components on the boundary can be computed with a high accuracy both in the plane region and on its boundary for plane orthotropic elasticity. For this case the hypersingular boundary integral equation method has been developed providing that (i) the region under consideration is a simply-connected region with a smooth boundary, (ii) there are no body forces, (iii) the material is orthotropic.

It has been proven that the formulas for calculating the stresses on the boundary can be derived in semi-analytical forms. The hypersingular boundary integrals are resolved into two parts. The first one, which is completely free from singularity, should be computed numerically – this computation results in, however, a very small numerical error. The second part (terms containing singularities) can be expressed analytically in a closed form. Consequently, in contrast to the conventional stress calculation, which uses polynomial approximations for the derivatives and thus might even lead to significant errors both in the formalism and in the numerical results in the vicinity of the boundary, the stress calculation method suggested in this paper is much more reliable and accurate due to the way it handles the hypersingular integrals.

A formulation of Somigliana stress identity for plane orthotropic elasticity has been developed and implemented in a Fortran 90 code for the numerical computations. The computational results we have obtained illustrate that the accuracy of the computations is increased significantly and confirm that a semi-analytical treatment of hypersingular integrals is very effective.

References

1. F. J. Rizzo and D. J. Shippy. "A method for stress determination in plane anisotropic elastic bodies." *Journal of Composite Materials,* **4**(1),1 (1970), pp. 36–61. DOI: 10.1177/002199837000400104.
2. G. Szeidl. "Boundary integral equations for plane problems in terms of stress functions of order one." *Journal of Computational and Applied Mechanics,* **2**(2), (2001), pp. 237–261.
3. G. Szeidl and J. Dudra. "On the direct BEM formulation in the dual system of plane elasticity for orthotropic bodies." *Journal of Computational and Applied Mechanics,* **10**(2),2 (2015), pp. 147–168. DOI: 10.32973/jcam.2015.010.
4. G. Szeidl and J. Dudra. "Boundary integral equations for plane orthotropic bodies and exterior regions." *Electronic Journal of Boundary Elements,* **2**, (2010), pp. 10–23. DOI: 10.14713/ejbe.v8i2.1023.
5. M. Guiggiani, G. Krishnasamy, F. J. Rizzo, and T. J. Randolphi. "Hypersingular boundary integral equations: a new approach to their numerical treatment." IABEM Symposium. 1990. DOI: 10.1007/978-3-642-85463-7_20.
6. G. Krishnasamy, T. J. Rudolphi, L. W. Schmerr, and F. J. Rizzo. "Hypersingular boundary integral equations: some applications in acoustic and elastic wave scattering," *ASME Journal of Applied Mechanics* **57**, (1990), pp. 404–414. DOI: 10.1115/1.2892004.
7. M. Guiggiani. "Direct evaluation of hypersingular integrals in 2D BEM, (7th GAMM Seminar on Numerical techniques for Boundary Element Method, Kiel, Germany, 1991)." *Notes in Numerical Fluid Mechanics.* Ed. by W. Hackbusch. 1992, pp. 23–24. DOI: 10.1007/978-3-663-14005-4.
8. Y. Liu and F. J. Rizzo. "A weakly singular form of the hypersingular boundary integral equation applied to 3-D acoustic wave problems." *Computer Methods in Applied Mechanics and Engineering,* **96**, (1992), pp. 271–287. DOI: 10.1016/0045-7825(92)90136-8.
9. J. Hildenbrand and G. Kuhn. "Numerical computation of hypersingular integrals and application to the boundary integral equation for the stress tensor." *Engineering Analysis with Boundary Elements,* **10**(3), (1992), pp. 209–217. DOI: 10.1016/0955-7997(92)90004-Q.
10. M. Guiggiani. "Hypersingular formulation for boundary stress evaluation." *Engineering Analysis with Boundary Elements,* **13**(2), (1994), pp. 169–179. DOI: 10.1016/0955-7997(94)90019-1.

11. S. Nintcheu Fata and L. J. Gray. "Semi-analytic integration of hypersingular Galerkin BIEs for three-dimensional potential problems." *Journal of Computational and Applied Mathematics,* **231**, (2009), pp. 561–576. DOI: 10.1016/j.cam.2009.04.003.
12. J. Rong, L. Wen, and J. Xiao. "Efficiency improvement of the polar coordinate transformation for evaluating BEM singular integrals on curved elements." *Engineering Analysis with Boundary Elements,* **38**, (2014), pp. 83–93. DOI: 10.1016/j.enganabound.2013.10.014.
13. J. Feng, Z. Yao, Y. Liu, and X. Zheng. "Evaluating hypersingular integrals of 3D acoustic problems on curved surfaces." *Engineering Analysis with Boundary Elements,* **60**, (2015), pp. 27–36. DOI: 10.1016/j.enganabound.2015.03.018.
14. Y. Liu. "On the displacement discontinuity method and the boundary element method for solving 3-D crack problems." *Engineering Fracture Mechanics,* **164** (2016), pp. 35–45. DOI: 10.1016/j.engfracmech.2016.07.009.
15. L. Huang, X. Sun, Y. Liu, and Z. Cen. "Parameter identification for two dimensional orthotropic material bodies by the boundary element method." *Engineering Analysis with Boundary Elements,* **28**(2), (2004), pp. 109–121. DOI: 10.1016/j.enganabound.2003.07.004.
16. Y. C. Shiah and C. L. Tan. "BEM treatment of two-dimensional anisotropic field problems by direct domain mapping." *Engineering Analysis with Boundary Elements,* **20**, (1997), pp. 347–351. DOI: 10.1016/S0955-7997(97)00103-3.
17. R. Avila, V. Mantic, and F. Paris. "Application of the boundary element method to elastic orthotropic materials in 2D: numerical aspects." *Transactions on Modelling and Simulation,* **18**, (1997), pp. 55–64. DOI: 10.2495/BE970061.
18. N. Kadioglu and S. Ataoglu. "A BEM implementation for 2D problems in plane orthotropic elasticity." *Structural Engineering & Mechanics,* **26**(5),5 (2007), pp. 591–615. DOI: 10.12989/sem.2007.26.5.591.
19. L. J. Gray and G. H. Paulino. "Symmetric galerkin boundary integral fracture analysis for plane orthotropic elasticity." *Computational Mechanics,* **20**, (1997), pp. 26–33. DOI: 10.1007/s004660050212.
20. S. Szirbik. "Hypersingular boundary integral formulations for plane elasticity in terms of first-order stress functions." *Journal of Computational and Applied Mechanics,* **11**(1), (2016), pp. 49–66. DOI: 10.32973/jcam.2016.004.
21. E. Tonti. "A mathematical model for physical theories I. II." *Rendiconti Accademia Nazionale dei Lincei,* VII, (1972), pp. 175–181; 351–356.
22. S. G. Lekhnitskii. *Theory of Elasticity of an Anisotropic Elastic Body.* Holden-Day, San Francisco, 1963.
23. M. L. Williams. "Stress singularities resulting from various boundary conditions in angular corners of plates in extension." *ASME Journal of Applied Mechanics,* **19**, (1952), pp. 526–528.
24. U. Brink, O. Klaas, R. Niekamp, and E. Stein. "Coupling of adaptively refined dual mixed finite elements and boundary elements in linear elasticity." *Advances in Engineering Software,* **24**, (1995), pp. 13–26. DOI: 10.1016/0965-9978(95)00055-0.

25. H. Wang and Q.-H. Qin. "Fundamental-solution-based finite element model for plane orthotropic elastic bodies." *European Journal of Mechanics A/Solids,* **29**, (2010), pp. 801–809. DOI: 10.1016/j.euromechsol.2010.05.003.

Appendix A.

A.1. **Derivatives of the fundamental solutions of order one.** Making use of equations (2.11), we can determine the derivatives $\mathcal{D}_{\kappa\lambda\rho} = \mathfrak{U}_{\kappa\lambda,\rho}$. The results are presented below:

$$\mathcal{D}_{111}(x,y) = 2K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{d_\alpha}{\rho_\alpha}\beta_\alpha^2\,, \qquad \mathcal{D}_{121}(x,y) = -2K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{d_\alpha}{\rho_\alpha}\beta_\alpha = \mathcal{D}_{211}(x,y)\,, \tag{A.1a}$$

$$\mathcal{D}_{211}(x,y) = -2K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{d_\alpha}{\rho_\alpha}\beta_\alpha\,, \qquad \mathcal{D}_{221}(x,y) = 2K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{d_\alpha}{\rho_\alpha}\,, \tag{A.1b}$$

$$\mathcal{D}_{112}(x,y) = 2K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{d_\alpha}{\rho_\alpha}\beta_\alpha^3\,, \qquad \mathcal{D}_{122}(x,y) = -2K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{d_\alpha}{\rho_\alpha}\beta_\alpha^2 = \mathcal{D}_{212}(x,y)\,, \tag{A.1c}$$

$$\mathcal{D}_{212}(x,y) = -2K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{d_\alpha}{\rho_\alpha}\beta_\alpha^2\,, \qquad \mathcal{D}_{222}(x,y) = 2K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{d_\alpha}{\rho_\alpha}\beta_\alpha\,. \tag{A.1d}$$

A.2. **Derivatives of the fundamental solutions of order two.** Making use of equations (2.12), we can determine the derivatives $\mathcal{S}_{\kappa\lambda\rho} = \mathfrak{T}_{\kappa\lambda,\rho}$. The results are presented below:

$$\mathcal{S}_{111}(x,y) = -K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{2d_\alpha}{\rho_\alpha^2}\Big[n_2\left(s_{11}\beta_\alpha^3 + s_{12}\beta_\alpha\right) + n_1\frac{s_{66}}{2}\beta_\alpha^2 + \\ + n_1(s_{22} + (s_{21} + \frac{s_{66}}{2})\beta_\alpha^2)\Big]\,, \tag{A.2a}$$

$$\mathcal{S}_{121}(x,y) = K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{2d_\alpha}{\rho_\alpha^2}\Big[\frac{n_2}{2}s_{66}\beta_\alpha^2 + n_1\left(s_{21}\beta_\alpha^3 + s_{22}\beta_\alpha\right) - \\ - n_2(s_{22} + (s_{21} + \frac{s_{66}}{2})\beta_\alpha^2)\Big]\,, \tag{A.2b}$$

$$\mathcal{S}_{211}(x,y) = K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{2d_\alpha}{\rho_\alpha^2}\Big[n_2\left(s_{11}\beta_\alpha^2 + s_{12}\right) + n_1\frac{s_{66}}{2}\beta_\alpha - \\ - n_1\beta_\alpha((s_{12} + \frac{s_{66}}{2}) + s_{11}\beta_\alpha^2)\Big]\,, \tag{A.2c}$$

$$\mathcal{S}_{221}(x,y) = K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{2d_\alpha}{\rho_\alpha^2}\Big[-\frac{n_2}{2}s_{66}\beta_\alpha - n_1\left(s_{21}\beta_\alpha^2 + s_{22}\right) - \\ - n_2\beta_\alpha((s_{21} + \frac{s_{66}}{2}) + s_{11}\beta_\alpha^2)\Big]\,, \tag{A.2d}$$

$$\mathcal{S}_{112}(x,y) = K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{2d_\alpha}{\rho_\alpha^2}\Big[-n_2\left(\beta_\alpha^4 s_{11} + s_{12}\beta_\alpha^2\right) - n_1\frac{s_{66}}{2}\beta_\alpha^3 - \\ - n_1(s_{22}\beta_\alpha + (s_{21} + \frac{s_{66}}{2})\beta_\alpha^3)\Big]\,, \tag{A.2e}$$

$$\mathcal{S}_{122}(x,y) = K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{2d_\alpha}{\rho_\alpha^2}\Big[n_1\left(s_{21}\beta_\alpha^4 + s_{22}\beta_\alpha^2\right) + \frac{n_2}{2}s_{66}\beta_\alpha^3 - \\ - n_2(s_{22}\beta_\alpha + (s_{21}+\frac{s_{66}}{2})\beta_\alpha^3)\Big]\ , \quad \text{(A.2f)}$$

$$\mathcal{S}_{212}(x,y) = K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{2d_\alpha}{\rho_\alpha^2}\Big[n_2\left(s_{11}\beta_\alpha^3 + s_{12}\beta_\alpha\right) + n_1\frac{s_{66}}{2}\beta_\alpha^2 - \\ - n_1\beta_\alpha^2((s_{12}+\frac{s_{66}}{2}) + s_{11}\beta_\alpha^2)\Big]\ , \quad \text{(A.2g)}$$

$$\mathcal{S}_{222}(x,y) = K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{2d_\alpha}{\rho_\alpha^2}\Big[-\frac{n_2}{2}s_{66}\beta_\alpha^2 - n_1\left(s_{21}\beta_\alpha^3 + s_{22}\beta_\alpha\right) - \\ - n_2\beta_\alpha^2((s_{21}+\frac{s_{66}}{2}) + s_{11}\beta_\alpha^2)\Big]\ , \quad \text{(A.2h)}$$

where the normal n_ρ is taken at the point x.

A.3. **Series expansion of the integrand $F^m_{\kappa j\rho}(\xi,\eta)$.** The Laurent series of $F^m_{\kappa\rho j}(\xi,\eta)$ in accordance with equation (4.19) is given in the series expansion at a given η as follows:

$$F^m_{\kappa j\rho}(\xi,\eta) = \mathcal{S}_{\kappa\lambda\rho}(\xi,\eta)J(\xi)N_{\lambda j}(\xi) = \underbrace{\left[K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{1}{(A_1+\beta_\alpha A_2)^2}{}^0a_{\kappa\lambda\rho}{}^0N_{\lambda j}(\eta)\right]}_{{}^{-2}F^m_{\kappa j\rho}(\eta)}\delta^{-2} + \\ + \underbrace{\left[\left(K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{1}{(A_1+\beta_\alpha A_2)^2}{}^1a_{\kappa\rho\lambda} - K\mathrm{Im}\sum_{\alpha=1}^{2}2\frac{B_1+\beta_\alpha B_2}{(A_1+\beta_\alpha A_2)^3}{}^0a_{\kappa\rho\lambda}\right){}^0N_{\lambda j}(\eta) + \right.}_{{}^{-1}F^m_{\kappa j\rho}(\eta)} \\ \underbrace{\left. + K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{1}{(A_1+\beta_\alpha A_2)^2}{}^0a_{\kappa\lambda\rho}{}^1N_{\lambda j}(\eta)\right]}_{{}^{-1}F^m_{\kappa j\rho}(\eta)}\delta^{-1} + \\ + O(1) \quad \text{(A.3a)}$$

and without entering into details, the functions are listed below:

$$
\begin{gathered}
{}^0a_{111} = 2d_\alpha\left(-A_2s_{22} + A_1s_{12}\beta_\alpha - A_2s_{21}\beta_\alpha^2 - A_2s_{66}\beta_\alpha^2 + A_1s_{11}\beta_\alpha^3\right) ,\\
{}^1a_{111} = -2d_\alpha\left(-2B_2s_{22} + 2B_1s_{12}\beta_\alpha - 2B_2s_{21}\beta_\alpha^2 - 2B_2s_{66}\beta_\alpha^2 + 2B_1s_{11}\beta_\alpha^3\right) ,\\
{}^0a_{121} = 2d_\alpha\left(s_{22}A_1 + s_{22}A_2\beta_\alpha + s_{21}A_1\beta_\alpha^2 + s_{21}A_2\beta_\alpha^3\right) ,\\
{}^1a_{121} = -2d_\alpha\left(2s_{22}B_1 + 2s_{22}B_2\beta_\alpha + 2s_{21}B_1\beta_\alpha^2 + 2s_{21}B_2\beta_\alpha^3\right) ,\\
{}^0a_{211} = 2d_\alpha\left(-s_{12}A_1 - s_{12}A_2\beta_\alpha - s_{11}A_1\beta_\alpha^2 - s_{11}A_2\beta_\alpha^3\right) ,\\
{}^1a_{211} = 2d_\alpha\left(-2s_{12}B_1 - 2s_{12}B_2\beta_\alpha - 2s_{11}B_1\beta_\alpha^2 - 2s_{11}B_2\beta_\alpha^3\right) ,\\
{}^0a_{221} = 2d_\alpha\left(-s_{22}A_2 + s_{21}A_1\beta_\alpha + s_{66}A_1\beta_\alpha - s_{21}A_2\beta_\alpha^2 + s_{11}A_1\beta_\alpha^3\right) ,\\
{}^1a_{221} = 2d_\alpha\left(-2s_{22}B_2 + 2s_{21}B_1\beta_\alpha + 2s_{66}B_1\beta_\alpha - 2s_{21}B_2\beta_\alpha^2 + 2s_{11}B_1\beta_\alpha^3\right) ,\\
{}^0a_{112} = 2d_\alpha\left(-s_{22}A_2\beta_\alpha + s_{12}A_1\beta_\alpha^2 - s_{21}A_2\beta_\alpha^3 - s_{66}A_2\beta_\alpha^3 + s_{11}A_1\beta_\alpha^4\right) ,\\
{}^1a_{112} = 2d_\alpha\left(-2s_{22}B_2\beta_\alpha + 2s_{12}B_1\beta_\alpha^2 - 2s_{21}B_2\beta_\alpha^3 - 2s_{66}B_2\beta_\alpha^3 + 2s_{11}B_1\beta_\alpha^4\right) ,\\
{}^0a_{122} = 2d_\alpha\left(s_{22}A_1\beta_\alpha + s_{22}A_2\beta_\alpha^2 + s_{21}A_1\beta_\alpha^3 + s_{21}A_2\beta_\alpha^4\right) ,\\
{}^1a_{122} = 2d_\alpha\left(2s_{22}B_1\beta_\alpha + 2s_{22}B_2\beta_\alpha^2 + 2s_{21}B_1\beta_\alpha^3 + 2s_{21}B_2\beta_\alpha^4\right) ,\\
{}^0a_{212} = 2d_\alpha\left(-s_{12}A_1\beta_\alpha - s_{12}A_2\beta_\alpha^2 - s_{11}A_1\beta_\alpha^3 - s_{11}A_2\beta_\alpha^4\right) ,\\
{}^1a_{212} = 2d_\alpha\left(-2s_{12}B_1\beta_\alpha - 2s_{12}B_2\beta_\alpha^2 - 2s_{11}B_1\beta_\alpha^3 - 2s_{11}B_2\beta_\alpha^4\right) ,\\
{}^0a_{222} = 2d_\alpha\left(-s_{22}A_2\beta_\alpha + s_{21}A_1\beta_\alpha^2 + s_{66}A_1\beta_\alpha^2 - s_{21}A_2\beta_\alpha^3 + s_{11}A_1\beta_\alpha^4\right) ,\\
{}^1a_{222} = 2d_\alpha\left(-2s_{22}B_2\beta_\alpha + 2s_{21}B_1\beta_\alpha^2 + 2s_{66}B_1\beta_\alpha^2 - 2s_{21}B_2\beta_\alpha^3 + 2s_{11}B_1\beta_\alpha^4\right) .
\end{gathered}
\tag{A.3b}
$$

A.4. **Series expansion of the integrand $G^m_{\kappa j\rho}(\xi,\eta)$.** The series of $G^m_{\kappa j\rho}(\xi,\eta)$ in accordance with equation (4.26) is given in the series expansion at a given η as follows:

$$
\begin{aligned}
G^m_{\kappa j\rho}(\xi,\eta) = \mathcal{D}_{\kappa\lambda\rho}(\xi,\eta)J(\xi)N_{\lambda j}(\xi) = \\
= \underbrace{K\mathrm{Im}\sum_{\alpha=1}^{2}\frac{1}{(A_1+\beta_\alpha A_2)}\tilde{a}_{\kappa\lambda\rho}\,{}^0N_{\lambda j}(\eta)\delta^{-1}}_{{}^{-1}G^m_{\kappa j\rho}(\eta)} + O(1) ,
\end{aligned}
\tag{A.4a}
$$

in which

$$
\begin{gathered}
\tilde{a}_{111} = 2d_\alpha\beta_\alpha^2A , \qquad \tilde{a}_{112} = 2d_\alpha\beta_\alpha^3A , \qquad \tilde{a}_{211} = -2d_\alpha\beta_\alpha A ,\\
\tilde{a}_{212} = -2d_\alpha\beta_\alpha^2A , \qquad \tilde{a}_{121} = -2d_\alpha\beta_\alpha A , \qquad \tilde{a}_{122} = -2d_\alpha\beta_\alpha^2A ,\\
\tilde{a}_{221} = 2d_\alpha A , \qquad \tilde{a}_{222} = 2d_\alpha\beta_\alpha A .
\end{gathered}
\tag{A.4b}
$$

Notes for Contributors

to the Journal of Computational and Applied Mechanics

Aims and scope. The aim of the journal is to publish research papers on theoretical and applied mechanics. Special emphasis is given to articles on computational mechanics, continuum mechanics (mechanics of solid bodies, fluid mechanics, heat and mass transfer) and dynamics. Review papers on a research field and materials effective for teaching can also be accepted and are published as review papers or classroom notes. Papers devoted to mathematical problems relevant to mechanics will also be considered.

Frequency of the journal. Two issues a year (approximately 80 pages per issue).

Submission of Manuscripts. Submission of a manuscript implies that the paper has not been published, nor is being considered for publication elsewhere. Papers should be written in standard grammatical English. The manuscript is to be submitted in electronic, preferably in pdf, format. The text is to be 130 mm wide and 190 mm long and the main text should be typeset in 10pt CMR fonts. Though the length of a paper is not prescribed, authors are encouraged to write concisely. However, short communications or discussions on papers published in the journal must not be longer than 2 pages. Each manuscript should be provided with an English Abstract of about 50–70 words, reporting concisely on the objective and results of the paper. The Abstract is followed by the Mathematical Subject Classification – in case the author (or authors) give the classification codes – then the keywords (no more than five). References should be grouped at the end of the paper in numerical order of appearance. Author's name(s) and initials, paper titles, journal name, volume, issue, year and page numbers should be given for all journals referenced.

The journal prefers the submission of manuscripts in LaTeX. Authors should select the $\mathcal{AMS}$-LaTeX article class and are not recommended to define their own LaTeX commands. Visit our home page for further details concerning how to edit your paper.

For the purpose of refereeing the manuscripts should be sent either to Balázs Tóth (Balazs.TOTH@uni-miskolc.hu) or György SZEIDL (Gyorgy.SZEIDL@uni-miskolc.hu).

The eventual supply of an accepted for publication paper in its final camera-ready form will ensure more rapid publication. Format requirements are provided by the home page of the journal from which sample LaTeX files can be downloaded:

http://www.mech.uni-miskolc.hu/jcam

These sample files can also be obtained directly (via e-mail) from Balázs TÓTH (Balazs.TOTH@uni-miskolc.hu), upon request.

One issue of the journal and ten offprints will be provided free of charge and mailed to the correspondent author. Since JCAM is an open access journal each paper can be downloaded freely from the homepage of the journal.

The Journal of Computational and Applied Mechanics is abstracted in Zentralblatt für Mathematik and in the Russian Referativnij Zhurnal.

Secretariat of the Vice-Rector for Research and International Relations, University of Miskolc
Responsible for publication: Prof. Dr. Zita Horváth, rector
Published by the Miskolc University Press under the leadership of Attila Szendi
Responsible for duplication: Works manager Erzsébet Pásztor
Number of copies printed: 75
Put to the Press on December 7, 2020
Number of permission: TNRT.2020-221.ME.
ISSN 2732–0189 (Online)
ISSN 1586–2070 (Print)

A Short History of the Publications of the University of Miskolc

The University of Miskolc (Hungary) is an important center of research in Central Europe. Its parent university was founded by the Empress Maria Teresia in Selmecbánya (today Banska Štiavnica, Slovakia) in 1735. After the first World War the legal predecessor of the University of Miskolc moved to Sopron (Hungary) where, in 1929, it started the series of university publications with the title *Publications of the Mining and Metallurgical Division of the Hungarian Academy of Mining and Forestry Engineering* (Volumes I.-VI.). From 1934 to 1947 the Institution had the name Faculty of Mining, Metallurgical and Forestry Engineering of the József Nádor University of Technology and Economic Sciences at Sopron. Accordingly, the publications were given the title *Publications of the Mining and Metallurgical Engineering Division* (Volumes VII.-XVI.). For the last volume before 1950 – due to a further change in the name of the Institution – *Technical University, Faculties of Mining, Metallurgical and Forestry Engineering, Publications of the Mining and Metallurgical Divisions* was the title.

For some years after 1950 the Publications were temporarily suspended.

After the foundation of the Mechanical Engineering Faculty in Miskolc in 1949 and the movement of the Sopron Mining and Metallurgical Faculties to Miskolc, the Publications restarted with the general title *Publications of the Technical University of Heavy Industry* in 1955. Four new series - Series A (Mining), Series B (Metallurgy), Series C (Machinery) and Series D (Natural Sciences) - were founded in 1976. These came out both in foreign languages (English, German and Russian) and in Hungarian.

In 1990, right after the foundation of some new faculties, the university was renamed to University of Miskolc. At the same time the structure of the Publications was reorganized so that it could follow the faculty structure. Accordingly three new series were established: Series E (Legal Sciences), Series F (Economic Sciences) and Series G (Humanities and Social Sciences). The latest series, i.e., the series H (European Integration Studies) was founded in 2001. The eight series are formed by some periodicals and such publications which come out with various frequencies.

Papers on computational and applied mechanics were published in the

Publications of the University of Miskolc, Series D, Natural Sciences.

This series was given the name Natural Sciences, Mathematics in 1995. The name change reflects the fact that most of the papers published in the journal are of mathematical nature though papers on mechanics also come out.

The series

Publications of the University of Miskolc, Series C, Fundamental Engineering Sciences

founded in 1995 also published papers on mechanical issues. The present journal, which is published with the support of the Faculty of Mechanical Engineering and Informatics as a member of the Series C (Machinery), is the legal successor of the above journal.

Journal of Computational and Applied Mechanics

Volume 15, Number 2 (2020)

Contents

Contributed Papers