UNIVERSITÄT
DUISBURG
ESSEN

Offen im Denken

Mixed Least Squares Finite Element Methods Based on Inverse Stress - Strain Relations in Hyperelasticity

Von der Fakultät für Mathematik der Universität Duisburg - Essen
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)

Genehmigte Dissertation von

Dipl. - Math. Benjamin Müller
geboren am 21.03.1985 in Alfeld (Leine)

Erstgutachter:	Prof. Dr. Gerhard Starke
Zweitgutachter:	Prof. Dr. - Ing. Jörg Schröder
Drittgutachter:	Prof. Dr. Christian Meyer

Ort und Datum der Einreichung:	Essen, den 11. Dezember 2014
Ort und Datum der mündlichen Prüfung:	Essen, den 17. April 2015

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeiten als wissenschaftlicher Mitarbeiter am Institut für Angewandte Mathematik der Leibniz Universität Hannover im Zeitraum März 2011 - März 2013 und im Anschluss an der Fakultät für Mathematik der Universität Duisburg - Essen im Rahmen des durch die Deutsche Forschungsgemeinschaft (DFG) geförderten Projektes „Gemischte Least - Squares Finite Elemente für geometrisch nichtlineare Probleme der Festkörpermechanik“ (Fördernummer STA402/11 - 1). Ich möchte mich für die finanzielle Unterstützung seitens der DFG herzlich bedanken.
Einen großen Dank möchte ich meinem Doktorvater Prof. Dr. Gerhard Starke aussprechen, der mir in all den Jahren mit seiner Expertise, aber auch in menschlicher Hinsicht, immer zur Seite stand. Ich danke ihm für seine geduldige, ruhige, ehrliche und humorvolle Art, welche den Arbeitsalltag allgemein sehr erleichtert hat. Ich danke ihm für den Freiraum, den er mir zur Anfertigung dieser Arbeit, besonders in den letzten Monaten, gegeben hat.
Ferner möchte ich mich bei Prof. Dr. - Ing. Jörg Schröder für die Übernahme des Zweitgutachtens und für die Kooperation in den letzten Jahren bedanken. Besonders zu erwähnen ist seine Expertise zur Modellierung anisotropen Materialverhaltens, die gerade zum Ende der vorliegenden Dissertation sehr hilfreich war. Zudem danke ich Prof. Dr. Christian Meyer für die Übernahme des Drittgutachtens.
Ich möchte mich bei meinem ehemaligen Arbeitskollegen Dr. Frank Samir Attia für die gemeinsame Zeit am Institut für Angewandte Mathematik in Hannover herzlich bedanken.
Meinen aktuellen Arbeitskollegen Dr. Fleurianne Bertrand und Dr. Steffen Münzenmaier danke ich für die konstruktive Zusammenarbeit in den letzten Jahren. Besonders danke ich Fleurianne, die die grundlegende Listenstruktur für 2d Triangulierungen auf 3d in MATLAB® erweitert hat und mir zur Verfügung gestellt hat. Ich danke Steffen für seine jederzeitige Bereitschaft mir bei auftretenden Problemen oder Fragen behilflich zu sein. Aber auch über die Arbeit hinaus ist Steffen zu einem Freund herangewachsen mit dem man viel Spaß haben kann.
Ich möchte mich bei meiner studentischen Hilfskraft Marcus O'Connor herzlich bedanken. Gemeinsam mit ihm wurde eine 3d - Verfeinerungsroutine in MATLAB® erfolgreich implementiert, die konsistente Tetraederzerlegungen bei lokaler und gleichmäßiger Verfeinerung erzeugt. Falls Probleme in der Routine auftraten, hat sich Marcus vorbildlich um die Behebung dieser gekümmert. Hierfür danke ich ihm sehr und wünsche ihm für seinen weiteren Werdegang alles Gute.
Ich danke meinem Cousin Dr. Sebastian Aeffner und meiner Verlobten Julia Riemer für die akribische Durchsicht meiner Arbeit. Bei meiner Verlobten möchte ich mich vor allem für die Unterstützung und Liebe, die sie mir in den letzten Jahren und besonders in den letzten Monaten immerwährend geschenkt hat, bedanken. Ohne sie hätte diese Arbeit nicht in der Form geschrieben werden können.
Zum Schluss möchte ich mich bei meiner gesamten Familie bedanken, ohne die ich an diesem Punkt meines Lebens nicht angekommen wäre. Vor allem möchte ich mich bei meinen Eltern und Großeltern für den Rückhalt bedanken, den sie mir immer gegeben haben und auch heute noch geben.

Abstract

Reliable simulation techniques for the description of elastic deformation processes in solid mechanics are nowadays of great importance. A reasonable model should take nonlinear kinematics and a nonlinear material law into account and should coincide with Hooke's law under small loads. In addition, a numerical method should be able to simulate compressible as well as (almost) incompressible material behavior. The calculation of good stress and displacement approximations is often of particular interest.

Therefore general mixed least squares finite element methods in the context of finite hyperelasticity are considered in this work. They are based on the conservation of linear momentum and inverse stress - strain relations and will be used for the simulation of homogeneous isotropic and homogeneous transverse isotropic material behavior. For the minimization of the nonlinear least squares functionals in finite dimensional spaces a Gauss - Newton framework is applied.

In the case of a specific homogeneous isotropic Neo - Hooke model an analysis is provided which proves reliability and efficiency of the nonlinear least squares functional as a - posteriori error estimator. The analysis remains valid in the incompressible limit and therefore the Poisson locking effect is excluded.

The analytical results for the Neo - Hooke model are used to propose an algorithm for model adaptivity which is based on the model of linear elasticity and the Neo - Hooke model. The algorithm automatically decides in which subdomain the linear model should be locally substituted by the Neo - Hooke model.

Two - and three - dimensional numerical examples for compressible and fully incompressible materials are given in order to illustrate the potential of our method. Here next - to - lowest - order Raviart - Thomas elements for the stress approximations are combined with conforming piecewise quadratic elements for the displacement approximations. A significant improvement of stress approximations in comparison to conventional discretization methods is demonstrated. In examples with corner or edge singularities almost optimal convergence rates for the nonlinear least squares functional using adaptive refinement strategies are achieved.

Key words:

first - order system least squares, mixed finite elements, Raviart - Thomas elements, Gauss - Newton algorithm, finite hyperelasticity, transverse isotropy, (model -) adaptivity

Kurzzusammenfassung

Zuverlässige Simulationstechniken zur Beschreibung von elastischen Verformungsprozessen in der Festkörpermechanik sind heutzutage von großer Bedeutung. Ein sinnvolles Modell sollte nichtlineare Kinematik und ein nichtlineares Materialgesetz berücksichtigen und mit dem Hookeschen Gesetz unter kleinen Belastungen übereinstimmen. Ferner sollte ein numerisches Verfahren sowohl kompressibles als auch (nahezu) inkompressibles Materialverhalten simulieren können. Die Berechnung von guten Spannungs - und Verschiebungsapproximationen ist oftmals von besonderem Interesse.
Aus diesen Gründen werden in dieser Arbeit allgemeine gemischte Least - Squares Finite - Element - Methoden im Rahmen der finiten Hyperelastizität betrachtet. Sie basieren auf der Impulserhaltung und inversen Spannungs - Verzerrungs - Relationen und werden zur Simulation homogen isotropen und homogen transversal - isotropen Materialverhaltens benutzt. Für die Minimierung der nichtlinearen Least - Squares Funktionale in endlichdimensionalen Räumen wird ein Gauß - Newton - Verfahren verwendet.
Im Falle eines speziellen homogen isotropen Neo - Hooke Modells wird eine Analysis bereitgestellt, welche die Zuverlässigkeit und Effizienz des nichtlinearen Least - Squares Funktionals als a - posteriori Fehlerschätzer beweist. Die Analysis bleibt gleichmäßig gültig im inkompressiblen Grenzfall womit der Poisson - Locking Effekt ausgeschlossen ist.
Die analytischen Resultate für das Neo - Hooke Modell werden benutzt um einen Algorithmus zur Modelladaptivität vorzuschlagen, welcher auf dem linearen Elastizitätsmodell und dem Neo - Hooke Modell basiert. Der Algorithmus entscheidet automatisch in welchem Teilgebiet das lineare Modell durch das Neo - Hooke Modell lokal ausgetauscht werden soll.
Zwei - und dreidimensionale numerische Beispiele für kompressible und inkompressible Materialien werden betrachtet, um das Potenzial unserer Methode zu verdeutlichen. Hierbei werden Raviart - Thomas Elemente zweitniedrigster Ordnung für die Spannungsapproximationen mit konformen, stückweise quadratischen, Elementen für die Verschiebungsapproximationen kombiniert. Eine signifikante Verbesserung von Spannungsapproximationen im Vergleich zu herkömmlichen Diskretisierungsmethoden wird nachgewiesen. In Beispielen mit Eck - oder Kantensingularitäten werden unter Verwendung adaptiver Verfeinerungsstrategien nahezu optimale Konvergenzraten für das nichtlineare Least - Squares Funktional erreicht.

Schlüsselwörter:
Least - Squares Finite - Element - Methoden basierend auf Systemen erster Ordnung, gemischte Finite Elemente, Raviart - Thomas Elemente, Gauß - Newton - Verfahren, finite Hyperelastizität, transversale Isotropie, (Modell -) Adaptivität

Table of contents

List of figures

List of tables

1 Introduction

1.1 Motivation

Numerical simulations and methods play a major role in many economical and industrial applications. For instance in insurance companies such methods are used for the simulation of natural catastrophes (e.g. earthquakes). Other applications for numerical methods can be found in mechanics, biomechanics, medicine and engineering. In all these fields the so - called Finite Element Method (abbrev. FEM) is an important tool.
In solid mechanics, if materials under load are considered, one generally distinguishes between plastic and elastic deformations. Plastic deformations are irreversible (e.g. crash tests in car manufacturing) whereas elastic deformation processes are reversible (e.g. small elongation of a spring). Reversible means that if one applies a load on a body, it will be firstly deformed and if the force does not act anymore, the body turns back into its original state. This work focuses on elastic deformations. Physical experiments show that the frequently used linear model (Hooke's law, cf. [Alt12]) is only valid up to a certain load. Therefore nonlinear models which describe the material behavior better for larger loads and correspond to the linear behavior of materials for small loads should be used.
Different discretization methods within FEM can be used to solve such problems. Generally one is interested in the primary variable, the deformation $\boldsymbol{\varphi}$ or equivalently the displacement $\mathbf{u}$. Additionally there is often a particular interest in secondary variables, for instance occurring strains and/or stresses. With this in mind one could distinguish discretization methods roughly into three categories:
The first and probably simplest one approximates only the primary variable $\mathbf{u}$ in a standard Galerkin framework (cf. [BS08], [Bra07] and [HR13]) and is therefore often called Galerkin or pure - displacement approach. In the context of this discretization method, using standard conforming piecewise polynomial elements, an undesirable effect has been observed in the past. It is called the Poisson locking effect and occurs if one combines almost incompressible materials, where the Lamé constant λ is very large, with a lower order polynomial in the FEM ansatz space. In this case the solution for the displacement within the Galerkin approach deteriorates. From a mathematical point of view Poisson locking occurs if the constant in the error estimate depends on λ and grows in the case $\lambda \to \infty$ (cf. [Bra07] and [BS08]). This problem cannot be solved by simple mesh refinement. However, there are some methods to overcome this problem. For instance one can use nonconforming finite elements (cf. Section 11.4 in [BS08]) or higher order conforming polynomial ansatz spaces. In [BS92] it is shown that one can eliminate the locking effect using at least piecewise polynomials of degree 4 in two dimensions and piecewise polynomials of degree 8 in three dimensions. Another problem that occurs by using the Galerkin approach is that the conservation of linear momentum is not satisfied very well in many cases.

The second category considers next to the displacements further variables, usually called mixed methods. In the field of linear elasticity common mixed methods are two - field methods for the approximation of displacements and stresses (Hellinger - Reissner principle) and three - field methods for the approximation of displacements, stresses and strains (Hu - Washizu principle), see [Bra07] for an overview. Other possible mixed methods are the so - called displacement - pressure approach (cf. [BBF13] for linear elasticity, [ABadVLR05] and [ABadVLR10] for nonlinear elasticity). In this approach the displacement $\mathbf{u}$ and additionally a scalar - valued pressure - like variable p are used as variables. A further mixed method for nonlinear hyperelasticity, based on the Hu - Washizu principle, was developed in [SWB11].
Mixed methods can overcome the Poisson locking effect and are well - suited for exact conservation of linear momentum. Nevertheless the main disadvantage is that one has to satisfy a discrete inf - sup - condition, also called Ladyženskaja - Babuška - Brezzi condition (abbrev. LBB condition), in order to obtain a stable formulation. This reduces the flexibility in choosing finite element spaces.
Another important question in nonlinear elasticity is the determination of bifurcation points, i.e. finding critical load values where a second solution of the problem occurs. Such situations are physically reasonable in nonlinear elasticity (cf. examples of non - uniqueness in [Wri08] and [Cia88]). In [ABadVLR10] it was shown for some concrete examples that some combinations of finite element spaces for the displacement - pressure approach fail in the approximation of critical load values. Moreover, it was shown that the exact satisfaction of the incompressibility constraint is very important in order to achieve good approximations.
Besides the displacement one is often interested in occurring stresses. Generally, using the Galerkin approach (respectively the displacement - pressure approach), one can calculate stress approximations as post - processing from displacement approximations (respectively from approximations of displacements and the pressure - like variable). This procedure leads to undesirable stress oscillations in many examples as we will see in this work.
The third category considers the so - called Least Squares Finite Element Method (abbrev. LSFEM). This method extends the common least squares method used in statistical regression analysis or data fitting to partial differential equations. An introduction into LSFEMs can be found for instance in [Jia98] and [BG09]. The method has in general some advantageous properties:
It has a unified formulation for all types of partial differential equations (elliptic, parabolic, hyperbolic) making it applicable to a wide array of problems. The stiffness matrices that occur in the linear systems of equations are in general symmetric and positive definite which is advantageous respectively necessary for iterative solvers, e.g. the conjugate gradient method. Another advantage is that the choice of finite element spaces is not restricted to the inf - sup - condition in contrast to mixed methods. Moreover, this method

automatically provides a candidate for an a - posteriori error estimator which one can use for adaptive mesh refinement. Local mesh refinement is generally desirable and often of great importance in numerical simulations.
In the context of fluid mechanics the least squares finite element method was applied succesfully in recent works, e.g. in [CW09] for viscoelastic fluids, in [ABL$^+$11] for two - phase flow using Navier - Stokes equations in both subdomains and for two - phase flow in [MS11] and [Mün12] using Stokes and Darcy flow in the different subdomains.
In the context of linear elasticity LSFEM was succesfully applied in [CS03], [CS04], [CKS05], [SSS10] and [SSS11]. The main differences in these works are the different weighting of the stress - strain relation and that the symmetry of the stress tensor is either taken into account or not as additional constraint in the least squares functional. For all of these approaches analytical results have been given.
To our best knowledge there is only one work in the context of LSFEM dealing with nonlinear elasticity and providing a detailed analysis, namely [MMSW06]. In this work a St. Venant - Kirchhoff material is considered and norm - equivalence of the linearized least squares functional to an appropriate norm in the case of pure displacement boundary conditions is proven. In [Sch09] some numerical results for finite elasticity, using a polyconvex stored energy function corresponding to a Neo - Hooke model and a LSFEM approach, are given without providing a mathematical analysis. Besides the conservation of momentum the usual stress - strain relation, without any weighting, is considered in the least squares functional in that work.
Typically one is interested in methods which can handle very general problems. In particular in solid mechanics the methods should cover compressible, almost incompressible (also called quasi - incompressible) and fully incompressible materials. In numerics the case of fully incompressible materials is the most difficult case, but of great importance (cf. [ABadVLR10] and [ADH$^+$14]).

1.2 Topics and outline of the work

This work extends the idea of [CS04] for linear elasticity to nonlinear problems in solid mechanics. The first Piola - Kirchhoff stress tensor $\mathbf{P}$ as secondary variable and the displacement $\mathbf{u}$ as primary variable are used in the proposed method. The considered models include nonlinear kinematics as well as nonlinear stress - strain relations for a given hyperelastic material. All of the used stored energy functions in this work satisfy the property of polyconvexity which is an important tool in the existence theory of Ball for the Galerkin approach (cf. [Bal77]).
The outline of this work is as follows:
In Section 2 we introduce essential features of functional analysis. Afterwards we give a brief introduction into the theory of mathematical elasticity including an introduction to

different stress and strain tensors, conditions for incompressibility, material - dependent and - independent properties, hyperelasticity and polyconvexity. General representation formulas for homogeneous isotropic materials will be derived. As specific models throughout the whole work we consider a Mooney - Rivlin and an associated Neo - Hooke model. Conditions to obtain consistency of the considered nonlinear model with the linear one are derived and are applied to the Mooney - Rivlin model. Polyconvexity is also proven for this model (cp. [Sch10]). At the end of the second section some suitable finite element spaces are introduced.

In Section 3 we explain the idea of the work [CS04] in more detail. The subsequent nonlinear extension is based on two partial differential equations, the conservation of linear momentum and an inverse stress - strain relation. The main motivation for inverting the stress - strain relation is the suitability to consider (quasi -) incompressible materials, similar to the linear case. We derive general least squares finite element methods for homogeneous isotropic materials.

We focus in particular on the Neo - Hooke model. We will see that the consideration of fully incompressible materials, i.e. $\lambda = \infty$, is possible in this model and the Poisson locking effect does not occur. Moreover, we show that the incompressibility constraint for the strain tensor according to the pair $(\mathbf{P}, \mathbf{u})$ is exactly satisfied. An analysis for the nonlinear problem will be provided. In particular we show that the nonlinear least squares functional is a reliable and efficient a - posteriori error estimator which one can use for adaptive refinement strategies. This result also implies an a - priori estimate for the error. We show that the approximation of the Kirchhoff stress tensor becomes symmetric in convergence as long as the nonlinear least squares functional converges to zero. We also obtain well - posedness of the corresponding linearized problems in $H_{\Gamma_N}(\mathrm{div};\, \Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$. The whole analysis is valid under some regularity assumptions and for solutions sufficiently close to the origin. For the minimization of the nonlinear least squares functional in finite dimensional spaces we use a Gauss - Newton scheme which solves the nonlinear minimization problem through a sequence of linearized problems.

At the end of the section we discuss several advantages and disadvantages of our approach in comparison to the Galerkin and the displacement - pressure approach.

Section 4 deals with the extension of the proposed least squares finite element approach for homogeneous isotropic materials to anisotropic materials. Suitable models are specified and the special case of transverse isotropic materials is studied in more detail. These materials are very important in engineering, e.g. in the context of fiber reinforced materials. In this application one inserts a strong material into a weaker one such that the material has a stronger load capacity in the direction of fibers. In this work a suitable nonlinear model, based on the explanations in [Sch10] and [BSN10], is derived in such a way that it is consistent with an appropriate linear one. Moreover, for a special choice of material parameters, we show that this model also covers the previous considered full isotropic case.

In Section 5 we use the analysis of Section 3 in the context of model adaptivity. We show under some suitable assumptions that it is possible to measure the quality of solutions of linear elasticity with respect to the Neo - Hooke model. We establish an algorithm which uses the model of linear elasticity as simple model at the beginning and adjusts the model appropriately to a more complex one, more precisely to the Neo - Hooke model in the subsequent steps, if necessary. This approach leads to locally nonlinear models which tend to describe the material behavior better.
In Section 6 we illustrate the performance of our method in two - and three - dimensional examples. We use next - to - lowest - order Raviart - Thomas elements for the approximation of the stresses and standard conforming piecewise quadratic polynomials for the approximation of the displacements. We show that our proposed method produces very good stress approximations without any unphysical oscillations. We obtain almost optimal convergence rates using adaptive refinement strategies. We will see that the term of linear momentum is also conserved quite well and we obtain an improved convergence rate for the balance of momentum in the $L^2(\Omega)$ - norm, similar to the results in [SSS10] and [SSS11] for linear elasticity. The displacement and stress approximations obtained with our proposed least squares formulation are compared with the results of the Galerkin method for compressible materials and the displacement - pressure approach for fully incompressible materials. The numerical experiments also include an example for the calculation of critical load values, some results for transverse isotropic materials and some results for model adaptivity.
In Section 7 a short conclusion, summarizing the main results, is given. Open questions which arose during this work are specified for further research in the outlook.

2 Preliminaries

In this chapter the essential tools for this work are provided, briefly speaking the fundament is presented. It contains basics in functional analysis and a general introduction into (nonlinear) elasticity theory. Moreover we consider a special class of models and introduce some finite element spaces at the end of this chapter.

2.1 Basics in functional analysis

2.1.1 Fréchet and Gâteaux derivative

Fréchet and Gâteaux derivatives are important tools in functional analysis. They generalize the total and directional differentiability of functions $g : \mathbb{R}^n \to \mathbb{R}^m$ to operators $f : V \to W$ between two arbitrary normed spaces V and W. In the following these derivatives will be introduced based on Section 5.3 in [AH09].
Let K be a subset of an arbitrary normed space V with norm $\|\cdot\|_V$ and $f : K \subset V \to W$ an operator, which maps an element of V into an element of a normed space W with norm $\|\cdot\|_W$. Further we assume that u_0 is an interior point of K, i.e. there exists $r > 0$ such that the ball $\mathcal{B}(u_0, r) := \{u \in V : \|u - u_0\|_V < r\}$ centered at u_0 with radius r is a subset of K. As a common abbreviation we use $\mathcal{L}(V, W)$ in the following as the set of all continuous linear operators from V to W.

Definition 2.1: (Fréchet derivative)
The operator f is Fréchet differentiable at $u_0 \in K$ if and only if there exists $A \in \mathcal{L}(V, W)$ such that

$$f(u_0 + v) = f(u_0) + A(v) + o(\|v\|_V)$$

for $v \to 0$, where $o(\cdot)$ describes the common little o-notation (see Appendix A).
The map A is called the Fréchet derivative of f at u_0 and we write $A = \partial f(u_0)$. If f is Fréchet differentiable at all points $K_0 \subset K$, f is called Fréchet differentiable on the set K_0 with derivative $\partial f : K_0 \subset V \to \mathcal{L}(V, W)$.

Remark 2.2:

(a) The Fréchet derivative $\partial f(u_0)$ is unique (cf. Section 5.3 in [AH09] below Definition 5.3.1).

(b) If $f : K \subset V \to W$ is m times Fréchet differentiable on K and each derivative is continuous, we say that f is m times continuously differentiable on K and denote it by $f \in C^m(K, W)$.

Definition 2.3: (Gâteaux derivative)
The Gâteaux derivative $f'(u_0)[v]$ of the operator f at u_0 in an arbitrary direction $v \in V$

is defined as

$$f'(u_0)[v] := \lim_{t \to 0} \frac{f(u_0 + tv) - f(u_0)}{t} = \frac{d}{dt} f(u_0 + tv)\big|_{t=0},$$

as long as the limit exists. If the limit exists for all elements in $K_0 \subset V$, we say that f is Gâteaux differentiable on K_0 with Gâteaux derivative $f' : K_0 \subset V \to \mathcal{L}(V, W)$.

Lemma 2.4: (Relation between Fréchet and Gâteaux derivative)

(a) If f is Fréchet differentiable at u_0, then f is also Gâteaux differentiable at u_0 and it holds

$$\partial f(u_0)(v) = f'(u_0)[v] \quad \forall\, v \in V \setminus \{0\}. \tag{2.1}$$

(b) If f is Gâteaux differentiable at u_0 and the Gâteaux derivative is continuous at u_0, then f is also Fréchet differentiable at u_0.

Proof:

(a) Due to the assumption that f is Fréchet differentiable at u_0 it follows by Definition 2.1 that $f(u_0 + v) = f(u_0) + \partial f(u_0)(v) + o(\|v\|_V)$ for $v \in V$ with $v \to 0$. We split an arbitrary $v \in V$ into $v = t\,\tilde{v}$ with $t := \|v\|_V$ and $\tilde{v} := \frac{v}{\|v\|_V}$. With this choice it holds $\|\tilde{v}\|_V = 1$ and we obtain $f(u_0 + t\tilde{v}) = f(u_0) + \partial f(u_0)(t\tilde{v}) + o(|t|)$ for $t \to 0$, which is equivalent to

$$\lim_{t \to 0} \frac{f(u_0 + t\tilde{v}) - f(u_0) - t\partial f(u_0)(\tilde{v})}{|t|} = 0 \Leftrightarrow \lim_{t \to 0} \frac{f(u_0 + t\tilde{v}) - f(u_0)}{t} = \partial f(u_0)(\tilde{v}),$$

i.e. $\partial f(u_0)(\tilde{v}) = f'(u_0)[\tilde{v}]\, \forall\, \tilde{v} \in V \setminus \{0\}$. The statement follows immediately.

(b) see proof of Theorem III. 5.4 (c) in [Wer05]

□

Remark 2.5:

If we know that a function f is Fréchet differentiable we know by this lemma that it is also Gâteaux differentiable. Furthermore we know that both derivatives are unique and coincide via (2.1). Under the assumption of Fréchet differentiability it is sufficient to determine the Gâteaux derivative, because then the Fréchet derivative must be given through equation (2.1). Due to this reason we are determining in this work often only the Gâteaux derivative of a function f, because the notation is more pleasant and the Gâteaux derivative is in general simpler to determine.

For the derivations in the following sections some calculation rules for the derivatives are needed. These rules are taken from [AH09] in a slightly different notation.

Definition 2.6: (partial Fréchet and Gâteaux derivative)
Let U, V, W be Banach spaces and $f : K_1 \times K_2 \subset U \times V \to W$ an operator on the product space $U \times V$. Assume that (u_0, v_0) is an interior point of $K_1 \times K_2$.
For fixed $v_0 \in K_2$, $f(u, v_0)$ is a function of $u \in K_1$. Then we call $\partial_u f(u_0, v_0) \in \mathcal{L}(U, W)$ the partial Fréchet derivative with respect to u of $f(u, v_0)$ at u_0, if it exists, and it holds by Definition 2.1

$$f(u_0 + u_1, v_0) = f(u_0, v_0) + \partial_u f(u_0, v_0)(u_1) + o(\|u_1\|_U), \quad u_1 \to 0.$$

The expression

$$f'_u(u_0, v_0)[u_1] = \lim_{t \to 0} \frac{f(u_0 + tu_1, v_0) - f(u_0, v_0)}{t} \quad \forall\, u_1 \in U,$$

if the limit exists, is called partial Gâteaux derivative of f with respect to u at u_0.
For fixed $u_0 \in K_1$, $f(u_0, v)$ is a function of $v \in K_2$. Then we call $\partial_v f(u_0, v_0) \in \mathcal{L}(V, W)$ the partial Fréchet derivative with respect to v of $f(u_0, v)$ at v_0, if it exists, and it holds by Definition 2.1

$$f(u_0, v_0 + v_1) = f(u_0, v_0) + \partial_v f(u_0, v_0)(v_1) + o(\|v_1\|_V), \quad v_1 \to 0.$$

The expression

$$f'_v(u_0, v_0)[v_1] = \lim_{t \to 0} \frac{f(u_0, v_0 + tv_1) - f(u_0, v_0)}{t} \quad \forall\, v_1 \in V,$$

if the limit exists, is called partial Gâteaux derivative of f with respect to v at v_0.

Remark 2.7:

a) If $f : K_1 \times K_2 \subset U \times V \to W$ is Fréchet differentiable at $(u_0, v_0) \in K_1 \times K_2$, then the partial Fréchet derivatives $\partial_u f(u_0, v_0)$ and $\partial_v f(u_0, v_0)$ exist and it holds

$$\partial f(u_0, v_0)(u_1, v_1) = \partial_u f(u_0, v_0)(u_1) + \partial_v f(u_0, v_0)(v_1) \tag{2.2}$$

for all $u_1 \in U$ and $v_1 \in V$ (see Proposition 5.3.15 in [AH09]).

b) If $f : K_1 \times K_2 \subset U \times V \to W$ is Gâteaux differentiable at $(u_0, v_0) \in K_1 \times K_2$, then the partial Gâteaux derivatives $f'_u(u_0, v_0)[u_1]$ and $f'_v(u_0, v_0)[v_1]$ exist and it holds

$$f'(u_0, v_0)[u_1, v_1] = f'_u(u_0, v_0)[u_1] + f'_v(u_0, v_0)[v_1]$$

for all $u_1 \in U$ and $v_1 \in V$.

c) If conversely $\partial_u f(u, v)$ and $\partial_v f(u, v)$ exist in a neighborhood of (u_0, v_0) and are continuous at (u_0, v_0), then f is Fréchet differentiable at (u_0, v_0) and the equation (2.2) holds (see Proposition 5.3.15 in [AH09]).

d) These results can be generalized in the same way for operators $f : K_1 \times \ldots \times K_n \subset V_1 \times \ldots \times V_n \to W$ with Banach spaces $V_1, \ldots, V_n, W$.

Proposition 2.8: (Linearity of the Fréchet and Gâteaux derivative)
Let V, W be two normed spaces and $f, g : K \subset V \to W$ Fréchet or Gâteaux differentiable at $u_0 \in K$. Then for any scalars α, β the operator $\alpha f + \beta g$ is Fréchet differentiable, respectively Gâteaux differentiable, at u_0.
For the Fréchet differentiable case it holds

$$\partial(\alpha f + \beta g)(u_0) = \alpha \partial f(u_0) + \beta \partial g(u_0).$$

For the Gâteaux differentiable case it holds $(\alpha f + \beta g)'(u_0)[v] = \alpha f'(u_0)[v] + \beta g'(u_0)[v]$ for all $v \in V$.

Proposition 2.9: (Product rule for the Fréchet and Gâteaux derivative)
Let V, V_1, V_2 and W be normed spaces. If $f : K \subset V \to V_1$ and $g : K \subset V \to V_2$ are Fréchet or Gâteaux differentiable at $u_0 \in K$, and $b : V_1 \times V_2 \to W$ is a bounded bilinear form, then the operator $h(u) := b\,(f(u), g(u))$ is Fréchet differentiable, respectively Gâteaux differentiable, at u_0.
For the Fréchet differentiable case it holds

$$\partial h(u_0)(v) = b\,(\partial f(u_0)(v), g(u_0)) + b\,(f(u_0), \partial g(u_0)(v)) \quad \forall\, v \in V.$$

For the Gâteaux differentiable case it holds

$$h'(u_0)[v] = b\left(f'(u_0)[v], g(u_0)\right) + b\left(f(u_0), g'(u_0)[v]\right) \quad \forall\, v \in V.$$

Proposition 2.10: (Chain rule for the Fréchet and Gâteaux derivative)
Let U, V, W be normed spaces and $f : K \subset U \to V$, $g : L \subset V \to W$ given operators with $f(K) \subset L$. Assume that u_0 is an interior point of K, $f(u_0)$ is an interior point of L.
If $\partial f(u_0)$ and $\partial g(f(u_0))$ exist as Fréchet derivatives, then the operator
$h(u) := g(f(u)) = (g \circ f)(u)$ is Fréchet differentiable at u_0 with

$$\partial h(u_0) = \partial g(f(u_0))(\partial f(u_0)).$$

If the Gâteaux derivative $f'(u_0)$ and the Fréchet derivative $\partial g(f(u_0))$ exist, then the operator h is Gâteaux differentiable at u_0 with

$$h'(u_0)[v] = g'(f(u_0))\left[f'(u_0)[v]\right] \quad \forall\, v \in V.$$

The following theorem, which guarantees local invertibility, can be found for instance in Theorem 1.2 - 4 in [Cia88] or in Section VII in [AE06b] (Theorem 7.3). This theorem will be of great importance for the derivation of our finite element approach in nonlinear elasticity.

Theorem 2.11: (Local inversion theorem)
Let V, W be two Banach spaces, K_2 an open subset of W with $a_2 \in K_2$, $g : K_2 \subset W \to V$ an operator satisfying

- $g \in C^1(K_2, V)$,
- $\partial g(a_2)$ is an isomorphism from W to V.

Then for $a_1 := g(a_2)$ there exist open subsets O_1, O_2 of the spaces V, W with $(a_1, a_2) \in O_1 \times O_2$, $O_2 \subset K_2$, and an implicit function $f : O_1 \subset V \to O_2 \subset W$ such that

(i) $\{(x_1, x_2) \in O_1 \times O_2 : x_1 = g(x_2)\} = \{(x_1, x_2) \in O_1 \times O_2 : x_2 = f(x_1)\}$

(ii) $O_2 = f(O_1)$ and $f : O_1 \subset V \to O_2 \subset W$ is a C^1 - diffeomorphism

(iii) The Fréchet derivative $\partial f : O_1 \subset V \to \mathcal{L}(V, W)$ can be determined via

$$\partial f(x_1) = \partial g(f(x_1))^{-1} \quad \forall\, x_1 \in O_1.$$

Proof:
see Theorem 7.3 in [AE06b] □

Remark 2.12:
Theorem 2.11 guarantees under some regularity assumptions the existence of a local inverse and provides a general formula to calculate the derivative of the inverse $g^{-1} := f$ with the help of the inverse of the derivative of g. For the Gâteaux derivative of the inverse g^{-1} we obtain the formula

$$\left(g^{-1}\right)'(x_1)[h] = g'(g^{-1}(x_1))^{-1}[h] \quad \forall\, x_1 \in O_1 \text{ and } h \in V.$$

2.1.2 The Hilbert space $V = \mathbb{R}^{n\times m}$

One of the basic vector spaces in this work is the vector space $V = \mathbb{R}^{n\times m}$ of all $n \times m$ matrices over $\mathbb{R}$ equipped with the standard addition of matrices and the scalar multiplication. On this vector space it is usual to define the inner product

$$\mathbf{A} : \mathbf{B} := \operatorname{tr}(\mathbf{A}^T\mathbf{B}) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}b_{ij}, \quad \mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times m}, \tag{2.3}$$

with induced norm $|\mathbf{A}| := (\mathbf{A} : \mathbf{A})^{1/2} = \left(\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2\right)^{\frac{1}{2}}$. This norm is often called **Frobenius norm** in literature, is submultiplicative and consistent with the Euclidean norm. It is also well known that this vector space with inner product defined by (2.3) is a Hilbert space.
Obviously for vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ the inner product $\mathbf{u} : \mathbf{v}$ is exactly the Euclidean inner

product, i.e. $\mathbf{u} \cdot \mathbf{v} := \mathbf{u}^T\mathbf{v} = \sum_{i=1}^{n} u_i v_i = \mathbf{u} : \mathbf{v}$, and $|\mathbf{v}| = \left(\sum_{i=1}^{n} v_i^2 \right)^{\frac{1}{2}}$ is the length of the vector $\mathbf{v}$ respectively the Euclidean norm.
It is simple to verify that the definition of $\mathbf{A} : \mathbf{B}$ above satisfies the three axioms (linearity, symmetry, positive definiteness) of an inner product. This Hilbert space, equipped with the inner product and the induced norm, is indispensable for elasticity theory.

2.1.3 Gradient of $f : K \subset \mathbb{R}^{n\times m} \to \mathbb{R}$

If an operator $f : K \subset V \to W$ is Fréchet differentiable on K we have $\partial f : K \to \mathcal{L}(V, W)$ by Definition 2.1. For $W = \mathbb{R}$ we get $\partial f(u_0) \in \mathcal{L}(V, \mathbb{R})$ for all $u_0 \in K$, i.e. $\partial f(u_0)$ is in the dual space of V.
If V is additionally a Hilbert space with inner product $(\cdot, \cdot)_V$, we know by the Riesz representation theorem (see Theorem 2.4.2 in [BS08]) that there exists a unique $u \in V$ with

$$\partial f(u_0)(v) = (u, v)_V \quad \forall\, v \in V.$$

For the Hilbert space $V = \mathbb{R}^{n\times m}$ with inner product defined by (2.3) and an operator $f : K \subset \mathbb{R}^{n\times m} \to \mathbb{R}, \mathbf{A} \mapsto f(\mathbf{A})$, that is Fréchet differentiable on K, we therefore get a unique matrix $\partial_{\mathbf{A}} f(\mathbf{A}) \in \mathbb{R}^{n\times m}$ such that

$$\partial f(\mathbf{A})(\mathbf{E}) = \partial_{\mathbf{A}} f(\mathbf{A}) : \mathbf{E} \quad \forall\, \mathbf{E} \in \mathbb{R}^{n\times m}. \tag{2.4}$$

In the next theorem we will show that the entries of the matrix $\partial_{\mathbf{A}} f(\mathbf{A})$ are exactly the partial derivatives $\frac{\partial f}{\partial a_{ij}}(\mathbf{A})$, where a_{ij} denotes the entries of the matrix $\mathbf{A}$. It becomes clear that the matrix $\partial_{\mathbf{A}} f(\mathbf{A})$ is the extension of the usual gradient ∇f in $\mathbb{R}^n$ to $\mathbb{R}^{n\times m}$. Another common notation for the gradient is $\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}}$. Both notations are used in this work. Altogether we have a relation between the Fréchet derivative and the gradient of such an operator.

Theorem 2.13: (Gradient in $\mathbb{R}^{n\times m}$)
Let $f : K \subset \mathbb{R}^{n\times m} \to \mathbb{R}$ be Fréchet differentiable in $\mathbf{A} \in K$ and let $\mathbf{E}^{(i,j)} \in \mathbb{R}^{n\times m}$ be the matrix with exactly one nonzero element in the i-th row and j-th column, more precisely $\left(\mathbf{E}^{(i,j)}\right)_{ij} = 1$ and everywhere else zero.
With the partial derivatives

$$\frac{\partial f(\mathbf{A})}{\partial a_{ij}} := \lim_{t \to 0} \frac{f\left(\mathbf{A} + t\,\mathbf{E}^{(i,j)}\right) - f(\mathbf{A})}{t} = f'(\mathbf{A}) \left[\mathbf{E}^{(i,j)}\right]$$

it holds

$$\left(\partial_{\mathbf{A}} f(\mathbf{A})\right)_{ij} = \frac{\partial f(\mathbf{A})}{\partial a_{ij}}, \quad 1 \le i \le n, 1 \le j \le m.$$

Proof:

With the definition of $\mathbf{E}^{(i,j)}$ we can decompose an arbitrary matrix $\mathbf{E} \in \mathbb{R}^{n\times m}$ into the sum $\mathbf{E} = \sum_{i=1}^{n}\sum_{j=1}^{m} e_{ij}\mathbf{E}^{(i,j)}$, where e_{ij} denotes the entry of $\mathbf{E}$ in the i-th row and j-th column. Due to Lemma 2.4 (a) it holds $\partial f(\mathbf{A})\left(\mathbf{E}^{(i,j)}\right) = f'(\mathbf{A})\left[\mathbf{E}^{(i,j)}\right] = \frac{\partial f}{\partial a_{ij}}(\mathbf{A})$. With the equation (2.4) and the definition of the inner product in $\mathbb{R}^{n\times m}$, we get

$$\begin{aligned}\partial_{\mathbf{A}} f(\mathbf{A}) : \mathbf{E} = \partial f(\mathbf{A})(\mathbf{E}) = \partial f(\mathbf{A})\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\mathbf{E}^{(i,j)}e_{ij}\right) &= \sum_{i=1}^{n}\sum_{j=1}^{m}\partial f(\mathbf{A})\left(\mathbf{E}^{(i,j)}\right)e_{ij} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{\partial f(\mathbf{A})}{\partial a_{ij}}e_{ij} = \mathbf{M} : \mathbf{E}\end{aligned}$$

for all $\mathbf{E} \in \mathbb{R}^{n\times m}$ and a matrix $\mathbf{M}$ with entries $m_{ij} = \frac{\partial f(\mathbf{A})}{\partial a_{ij}}$, i.e. the statement holds.

□

2.1.4 Function spaces

The basic function spaces dealing with partial differential equations are the so called Sobolev spaces $W^{k,p}(\Omega)$. A detailed introduction is for example given in [AF03]. The Sobolev spaces are based on the Lebesgue spaces $L^p(\Omega)$. In this work $\mathbb{N} := \{0, 1, 2, \ldots\}$ denotes the set of all nonnegative integers and Ω a nonempty, open, bounded and connected subset in $\mathbb{R}^d$.

For $0 < p < \infty$ we define the function space

$$L^p(\Omega) := \{v : \Omega \to \mathbb{R} \,|\, v \text{ measurable on } \Omega,\ \ \|v\|_{L^p(\Omega)} := \left(\int_\Omega |v(x)|^p\, dx\right)^{\frac{1}{p}} < \infty\}.$$

For $p = \infty$ we define the function space

$$L^\infty(\Omega) := \{v : \Omega \to \mathbb{R} \,|\, v \text{ measurable on } \Omega,\ \ \|v\|_{L^\infty(\Omega)} < \infty\}$$

with $\|v\|_{L^\infty(\Omega)} := \operatorname*{ess\,sup}_{x\in\Omega} |v(x)| := \inf_{\text{meas}(\Omega')=0} \sup_{x\in\Omega\setminus\Omega'} |v(x)|$.

More precisely we have to note that each of these function spaces consist actually of equivalence classes of functions, where a class is made of functions that differ from each other only on a subset of Ω with measure zero, i.e. these functions are equal almost everywhere on Ω. For $p \in [1, \infty]$, $L^p(\Omega)$ is a Banach space (cf. Theorem 2.16 in [AF03]). $L^2(\Omega)$ is even a Hilbert space (cf. Corollary 2.18 in [AF03]) with inner product

$$(u, v)_{L^2(\Omega)} := \int_\Omega u(x)v(x)\, dx \text{ for } u, v \in L^2(\Omega).$$

For $0 < p < 1$ the mapping $\|\cdot\|_{L^p(\Omega)} : L^p(\Omega) \to \mathbb{R}_{\geq 0}$ is no longer a norm, since the triangle inequality does not hold in this case. However it is still a quasi-norm.

For vector - valued functions $\mathbf{u} \in L^p(\Omega)^n$ or matrix - valued functions $\mathbf{A} \in L^p(\Omega)^{n\times m}$, i.e. each component of these functions are in $L^p(\Omega)$, we define the norms

$$\|\mathbf{u}\|_{L^p(\Omega)} := \left(\sum_{i=1}^{n} \|u_i\|_{L^p(\Omega)}^p\right)^{\frac{1}{p}}, \quad \|\mathbf{A}\|_{L^p(\Omega)} := \left(\sum_{i=1}^{n}\sum_{j=1}^{m} \|a_{ij}\|_{L^p(\Omega)}^p\right)^{\frac{1}{p}}$$

for $p \in [1, \infty)$.
For $\mathbf{u} \in L^\infty(\Omega)^n$ and $\mathbf{A} \in L^\infty(\Omega)^{n\times m}$, i.e. $p = \infty$, we define the norms

$$\|\mathbf{u}\|_{L^\infty(\Omega)} := \max_{1\le i\le n} \|u_i\|_{L^\infty(\Omega)}, \quad \|\mathbf{A}\|_{L^\infty(\Omega)} := \max_{1\le i\le n}\max_{1\le j\le m} \|a_{ij}\|_{L^\infty(\Omega)}.$$

For the case $p = 2$ we can also define an inner product for vector - valued functions $\mathbf{u}, \mathbf{v} \in L^2(\Omega)^n$ and matrix - valued functions $\mathbf{A}, \mathbf{B} \in L^2(\Omega)^{n\times m}$ as

$$(\mathbf{u}, \mathbf{v})_{L^2(\Omega)} := \sum_{i=1}^{n} (u_i, v_i)_{L^2(\Omega)}, \quad (\mathbf{A}, \mathbf{B})_{L^2(\Omega)} := \sum_{i=1}^{n}\sum_{j=1}^{m} (a_{ij}, b_{ij})_{L^2(\Omega)}.$$

Remark 2.14: (Alternative definition of $L^p(\Omega)$ - norms for matrix - valued functions)
For matrix - valued functions $\mathbf{A} \in L^p(\Omega)^{n\times m}$ one can easily prove that

$$\|\mathbf{A}\|_{L^p(\Omega)} \lesssim \left(\int_\Omega |\mathbf{A}(x)|^p \, dx\right)^{\frac{1}{p}} \lesssim \|\mathbf{A}\|_{L^p(\Omega)}, \quad p \in [1, \infty),$$
$$\|\mathbf{A}\|_{L^\infty(\Omega)} \lesssim \operatorname*{ess\,sup}_{x\in\Omega} |\mathbf{A}(x)| \lesssim \|\mathbf{A}\|_{L^\infty(\Omega)}, \quad p = \infty,$$

i.e. $\mathbf{A}$ is in $L^p(\Omega)^{n\times m}$ if and only if $|\mathbf{A}|$ is in $L^p(\Omega)$, $p \in [1, \infty]$. Here $\lesssim$ stands for inequalities up to positive constants. Hence an alternative definition of the $L^p(\Omega)$ - norms in form of

$$\|\mathbf{A}\|_{L^p(\Omega)} := \||\mathbf{A}|\|_{L^p(\Omega)}, \quad p \in [1, \infty],$$

for matrix - valued functions $\mathbf{A} \in L^p(\Omega)^{n\times m}$ is possible and both definitions are equivalent.

Remark 2.15: (Relation between the inner products $\mathbf{A} : \mathbf{B}$ and $(\mathbf{A}, \mathbf{B})_{L^2(\Omega)}$)
Let $\mathbf{A}, \mathbf{B} \in L^2(\Omega)^{n\times m}$ be two matrix - valued functions. Then it holds

$$(\mathbf{A}, \mathbf{B})_{L^2(\Omega)} = \sum_{i=1}^{n}\sum_{j=1}^{m} (a_{ij}, b_{ij})_{L^2(\Omega)} = \int_\Omega \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij} b_{ij} \, dx = \int_\Omega \mathbf{A} : \mathbf{B} \, dx$$

and therefore

$$\|\mathbf{A}\|_{L^2(\Omega)} = (\mathbf{A}, \mathbf{A})_{L^2(\Omega)}^{\frac{1}{2}} = \left(\int_\Omega \mathbf{A} : \mathbf{A} \, dx\right)^{\frac{1}{2}} = \left(\int_\Omega |\mathbf{A}|^2 \, dx\right)^{\frac{1}{2}}.$$

The relations $(\mathbf{u}, \mathbf{v})_{L^2(\Omega)} = \int_\Omega \mathbf{u} : \mathbf{v} \, dx = \int_\Omega \mathbf{u}^T \mathbf{v} \, dx$ and $\|\mathbf{u}\|_{L^2(\Omega)} = \left(\int_\Omega |\mathbf{u}|^2 \, dx\right)^{\frac{1}{2}}$ for vector - valued functions $\mathbf{u}, \mathbf{v} \in L^2(\Omega)^n$ are an immediate consequence.

With the help of the Lebesgue spaces we can now define the Sobolev spaces $W^{k,p}(\Omega)$ and additionally for vector - valued functions we can introduce the function space $W^p(\mathrm{div};\Omega)$ which generalizes the function space $H(\mathrm{div};\Omega)$ (cf. [BBF13]). These spaces are based on the terms „weak derivative“ and „weak divergence“. A more detailed introduction into weak derivatives, which extend the term of the (classical Fréchet) derivative and is crucial for the definition of the function spaces below, can be found for example in Chapter 7 of [AH09]. In the following definition we have used the fact that $L^p(\Omega)$ is a subspace of locally integrable functions (cf. Corollary 2.15 in [AF03]).

Definition 2.16: (Weak derivative and weak divergence in $L^p(\Omega)$)
Let $1 \leq p \leq \infty$, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d) \in \mathbb{N}^d$ be an arbitrary multi - index of order d with length $|\alpha| = \sum_{j=1}^{d} \alpha_j \in \mathbb{N}$.
A function $w \in L^p(\Omega)$ is called α^{th} **weak derivative** of $v \in L^p(\Omega)$ if

$$\int_\Omega v(x)\, \partial^\alpha \varphi(x)\, dx = (-1)^{|\alpha|} \int_\Omega w(x)\, \varphi(x)\, dx, \quad \varphi \in C_0^\infty(\Omega).$$

In this case we write $w = \partial^\alpha v$.
Consequently a function $\tilde{w} \in L^p(\Omega)$ is called **weak divergence** of $\mathbf{v} \in L^p(\Omega)^d$ if

$$\int_\Omega \mathbf{v}(x) \cdot \nabla \varphi(x)\, dx = -\int_\Omega \tilde{w}(x) \varphi(x)\, dx, \quad \varphi \in C_0^\infty(\Omega).$$

In this case we write $\tilde{w} = \operatorname{div} \mathbf{v}$.

Definition 2.17: (Sobolev spaces $W^{k,p}(\Omega)$)
Let $k \in \mathbb{N}$, $p \in [1, \infty]$ and $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d) \in \mathbb{N}^d$ be an arbitrary multi - index of order d with length $|\alpha| = \sum_{j=1}^{d} \alpha_j \in \mathbb{N}$. Then we define the Sobolev space $W^{k,p}(\Omega)$ as

$$W^{k,p}(\Omega) := \{v \in L^p(\Omega) : \partial^\alpha v \in L^p(\Omega) \text{ for all } \alpha \text{ with } |\alpha| \leq k\}$$

under the assumption that all possible weak derivatives $\partial^\alpha v$ of v exist.
We equip this space, as usual, with norm and semi - norm

$$\|v\|_{W^{k,p}(\Omega)} := \begin{cases} \left(\sum\limits_{|\alpha| \leq k} \|\partial^\alpha v\|_{L^p(\Omega)}^p \right)^{\frac{1}{p}} & , 1 \leq p < \infty \\ \max\limits_{|\alpha| \leq k} \|\partial^\alpha v\|_{L^\infty(\Omega)} & , p = \infty, \end{cases}$$

$$|v|_{W^{k,p}(\Omega)} := \begin{cases} \left(\sum\limits_{|\alpha| = k} \|\partial^\alpha v\|_{L^p(\Omega)}^p \right)^{\frac{1}{p}} & , 1 \leq p < \infty \\ \max\limits_{|\alpha| = k} \|\partial^\alpha v\|_{L^\infty(\Omega)} & , p = \infty. \end{cases}$$

For $k = 0$ we have $W^{0,p}(\Omega) = L^p(\Omega)$. For $p = 2$ we generally write $H^k(\Omega) := W^{k,2}(\Omega)$ with norm $\|v\|_{H^k(\Omega)}$ and semi - norm $|v|_{H^k(\Omega)}$. Obviously it holds $H^0(\Omega) = L^2(\Omega)$.

For the special case $k = 1$ in the general definition above we obtain the norm

$$\|v\|_{W^{1,p}(\Omega)} := \begin{cases} \left(\|v\|_{L^p(\Omega)}^p + \|\nabla v\|_{L^p(\Omega)}^p\right)^{\frac{1}{p}} & , 1 \leq p < \infty \\ \max\left\{\|v\|_{L^\infty(\Omega)}, \|\nabla v\|_{L^\infty(\Omega)}\right\} & , p = \infty. \end{cases}$$

The definitions can be generalized to vector - or matrix - valued functions similar as above for the Lebesgue spaces.
Due to Theorem 7.2.3 and Corollary 7.2.4 in [AH09] we know that the Sobolev spaces $W^{k,p}(\Omega)$ are Banach spaces and for $p = 2$ they are actually Hilbert spaces.

Definition 2.18: **($W^p(\mathbf{div}; \Omega)$)**
For $p \in [1, \infty]$ we define

$$W^p(\mathrm{div}; \Omega) := \left\{\mathbf{v} \in L^p(\Omega)^d : \mathrm{div}\, \mathbf{v} \in L^p(\Omega)\right\}$$

under the assumption that $\mathrm{div}\, \mathbf{v}$ exists in the weak sense and equip this space with the norm

$$\|\mathbf{v}\|_{W^p(\mathrm{div};\Omega)} := \begin{cases} \left(\|\mathbf{v}\|_{L^p(\Omega)}^p + \|\mathrm{div}\, \mathbf{v}\|_{L^p(\Omega)}^p\right)^{\frac{1}{p}} & , 1 \leq p < \infty \\ \max\left\{\|\mathbf{v}\|_{L^\infty(\Omega)}, \|\mathrm{div}\, \mathbf{v}\|_{L^\infty(\Omega)}\right\} & , p = \infty. \end{cases}$$

Remark 2.19:

(a) For $p = 2$ we get the space $H(\mathrm{div}; \Omega)$ (cf. Section 2.1.1 in [BBF13]).

(b) For $p \geq 2$ it holds $W^p(\mathrm{div}; \Omega) \subseteq H(\mathrm{div}; \Omega)$, since $L^p(\Omega) \subseteq L^2(\Omega)$ for $p \geq 2$ (cf. Theorem 2.14 in [AF03]).

2.2 Basics in elasticity theory

The aim of this work is to develop a new discretization method for nonlinear elastostatic deformation processes. Since the often used linear elasticity theory has its validity only up to a certain load and therefore does not cover real life problems in general we use nonlinear models. This generally leads to a physically more realistic consideration of such problems and therefore to more suitable results. A nice mathematical and detailed introduction into the elasticity theory can be found in [Cia88]. In this section the essential basics of nonlinear elasticity theory, based on this book, will be described briefly. These basics are crucial for the following chapters and the development of our new discretization scheme.

2.2.1 Description of a deformation problem

The initial situation is given by a nonempty, open, bounded and connected subset $\Omega \subset \mathbb{R}^3$ with Lipschitz - continuous boundary $\Gamma := \partial\Omega$. In practice the subset Ω is a given body which will be deformed through some applied forces. We split the boundary Γ into two non - overlapping open subsets Γ_D and Γ_N, i.e. $\bar{\Gamma}_D \cup \bar{\Gamma}_N = \Gamma$ and $\Gamma_D \cap \Gamma_N = \emptyset$. For $\Gamma_N = \emptyset$, we have a pure displacement problem and for $\Gamma_D = \emptyset$ we have a pure traction problem. For practical purposes one usually considers a mixed problem, i.e. neither Γ_D nor Γ_N is empty.
If the body $\bar{\Omega}$ is unloaded, respectively undeformed, it is called the **reference configuration**. Now we can apply some forces on the given body. On the one hand we have volume forces, for example gravity, which act on the whole body. Mathematically, the volume forces are described by a given density function $\mathbf{f} : \Omega \to \mathbb{R}^3$ representing the applied force per unit volume. On the other hand we can apply surface forces, which act only on the part Γ_N of the boundary. Examples for surface forces are traction and pressure forces. The surface forces are given mathematically through a density function $\mathbf{g} : \Gamma_N \to \mathbb{R}^3$ representing the applied force per unit area.
After applying these forces we get the so - called **deformed configuration**. A visualization of this deformation process under given forces can be seen in Figure 2.1.
The aim is now to determine the **deformation** $\boldsymbol{\varphi} : \bar{\Omega} \to \mathbb{R}^3$, i.e. the mapping from the reference to the deformed configuration. This mapping must be injective in Ω and orientation - preserving in $\bar{\Omega}$ to be physically acceptable (see Section 1.4 in [Cia88]).
Obviously one can split the deformation $\boldsymbol{\varphi}$ into

$$\boldsymbol{\varphi} = \mathbf{id} + \mathbf{u}$$

with the pointwise **displacement** $\mathbf{u} : \bar{\Omega} \to \mathbb{R}^3$.
If $\boldsymbol{\varphi}$ is Fréchet differentiable, we can define the deformation gradient $\mathbf{F}$ as

$$\mathbf{F} := \nabla\boldsymbol{\varphi} = \nabla(\mathbf{id} + \mathbf{u}) = \mathbf{I} + \nabla\mathbf{u} =: \mathbf{F}(\mathbf{u}).$$

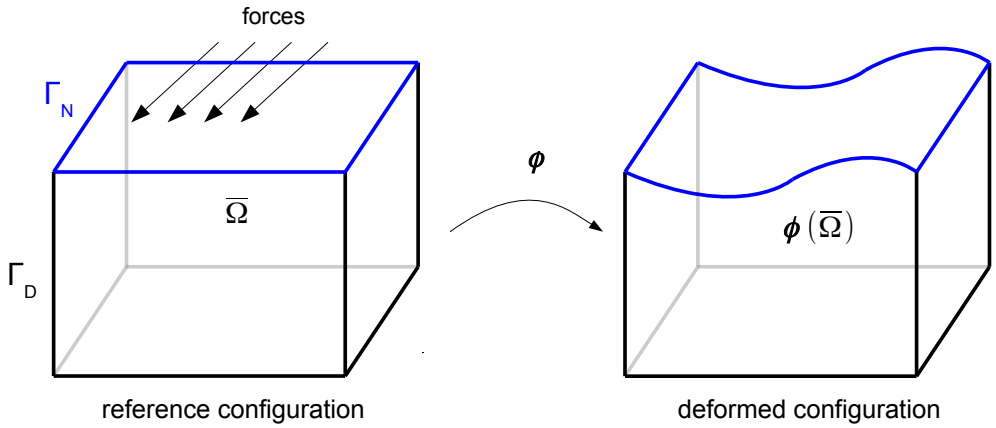


Figure 2.1: Schematical visualization of a deformation process

Here the common gradient operator ∇ is applied to each component of $\boldsymbol{\varphi}$ (respectively **id** and **u**) and forms the corresponding row of **F**. $\mathbf{id} : \bar{\Omega} \to \mathbb{R}^3$ is the identity mapping, i.e. $\mathbf{id}(\mathbf{x}) = \mathbf{x}$ for $\mathbf{x} \in \bar{\Omega}$, and **I** denotes the identity matrix. We use the notation $\mathbf{I} \in \mathbb{R}^{n \times n}$ for the identity matrix in the whole work, in each case with proper dimension $n \in \mathbb{N} \setminus \{0\}$, i.e. here $n = 3$.

Note that it holds $\det \mathbf{F} > 0$ in each point $x \in \bar{\Omega}$, since $\boldsymbol{\varphi}$ is orientation - preserving.

Besides the deformation $\boldsymbol{\varphi}$, engineers are also interested in the stresses that occur in the body. In this work the later determined stresses are mappings from $\bar{\Omega}$ to $\mathbb{R}^{3\times3}$, i.e. one obtains for each $x \in \bar{\Omega}$ a stress tensor which describes the mechanical stresses in this point. On the diagonal elements one has the normal stresses and on the nondiagonal elements one has the shear stresses (see Figure 2.2 for a visualization of a stress tensor **P** with its matrix entries).

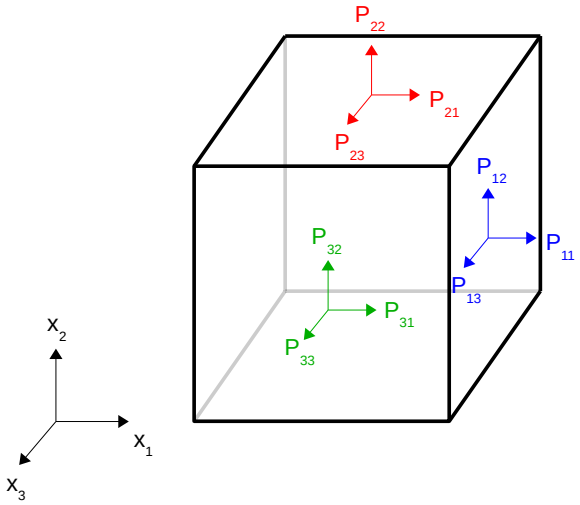


Figure 2.2: Schematical visualization of a stress tensor **P** in a small volume element

For the general elasticity theory two sets of equations are fundamental. The first set consists of the so - called equations of equilibrium that will be specified in Section 3. The second fundamental set of equations is the constitutive equation/material law. The material law describes a relation between stresses and strains. It is possible to describe the problem either with respect to the reference configuration (Lagrangian description) or the deformed configuration (Eulerian description). The Eulerian description has the disadvantage that it is expressed in the unknown $\boldsymbol{\varphi}(\mathbf{x})$. With the help of the so - called Piola transform the occurring equations can be transformed into the reference configuration, which is then independent of the deformation $\boldsymbol{\varphi}$. For a detailed introduction, distinction and derivation we refer to [Cia88] and [Sim98].

2.2.2 Stress and strain tensors, rigid body motions

For later purposes we define some stress and strain tensors which will appear in this work.
Stress tensors:
In this work we use the non - symmetric first Piola - Kirchhoff stress tensor $\mathbf{P} : \bar{\Omega} \to \mathbb{R}^{3\times 3}$ and the symmetric second Piola - Kirchhoff stress tensor $\boldsymbol{\Sigma} : \bar{\Omega} \to \mathbb{R}^{3\times 3}$ related by

$$\boldsymbol{\Sigma} = \mathbf{F}^{-1}\mathbf{P}. \tag{2.5}$$

Both Piola - Kirchhoff stress tensors are defined in the reference configuration. Another important stress tensor is the so - called Kirchhoff stress tensor $\boldsymbol{\tau}$. It is defined on the deformed configuration and is related to the Piola - Kirchhoff stress tensors by

$$\boldsymbol{\tau} = \mathbf{P}\mathbf{F}^T = \mathbf{F}\boldsymbol{\Sigma}\mathbf{F}^T.$$

A detailed introduction/derivation into the different stress tensors can be found again in [Cia88] and [Sim98].
Strain tensors:
For a given Fréchet differentiable deformation $\boldsymbol{\varphi} = \mathbf{id} + \mathbf{u}$ with deformation gradient $\mathbf{F} = \nabla\boldsymbol{\varphi} = \mathbf{I} + \nabla\mathbf{u} = \mathbf{F}(\mathbf{u})$ we define the following strain tensors

- $\mathbf{B} := \mathbf{F}\mathbf{F}^T$ (left Cauchy - Green strain tensor),
- $\mathbf{C} := \mathbf{F}^T\mathbf{F}$ (right Cauchy - Green strain tensor),
- $\mathbf{E} := \frac{1}{2}(\mathbf{C} - \mathbf{I})$ (Green - St. Venant strain tensor),

which are nonlinear in $\mathbf{u}$.
If we linearize $\mathbf{E}(\mathbf{u})$ about $\mathbf{u} = \mathbf{0}$ we obtain

$$\begin{aligned}\mathbf{E}(\mathbf{0}+\mathbf{v}) \approx \mathbf{E}(\mathbf{0}) + \mathbf{E}'(\mathbf{0})[\mathbf{v}] = \mathbf{0} + \frac{1}{2}\mathbf{C}'(\mathbf{0})[\mathbf{v}] &= \frac{1}{2}\left(\frac{d}{dt}\left[(\mathbf{F}(\mathbf{0}+t\mathbf{v}))^T\mathbf{F}(\mathbf{0}+t\mathbf{v})\right]\Big|_{t=0}\right) \\ &= \frac{1}{2}\left((\nabla\mathbf{v})^T\mathbf{F}(\mathbf{0}) + (\mathbf{F}(\mathbf{0}))^T\nabla\mathbf{v}\right) = \frac{1}{2}\left(\nabla\mathbf{v} + (\nabla\mathbf{v})^T\right) =: \boldsymbol{\varepsilon}(\mathbf{v}).\end{aligned}$$

Here $\boldsymbol{\varepsilon}(\mathbf{v})$ is the well - known strain tensor from linear elasticity theory.
The definition of $\mathbf{C}$ is motivated by the following considerations (see Section 1.8 in [Cia88]): If we consider an infinitesimal change of a point $\mathbf{x} \in \bar{\Omega}$ to $\mathbf{x} + \boldsymbol{\delta}\mathbf{x} \in \bar{\Omega}$ and have a Fréchet differentiable deformation $\boldsymbol{\varphi} : \bar{\Omega} \to \mathbb{R}^3$, we get by Definition 2.1

$$\boldsymbol{\varphi}(\mathbf{x} + \boldsymbol{\delta}\mathbf{x}) = \boldsymbol{\varphi}(\mathbf{x}) + \nabla\boldsymbol{\varphi}(\mathbf{x})\boldsymbol{\delta}\mathbf{x} + o(|\boldsymbol{\delta}\mathbf{x}|),$$

where $\nabla\boldsymbol{\varphi}(\mathbf{x}) \in \mathbb{R}^{3\times 3}$ here denotes the corresponding matrix representation of the Fréchet derivative $\partial\boldsymbol{\varphi}(\mathbf{x}) \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^3)$ of $\boldsymbol{\varphi}$, i.e. the row - wise applied gradient of $\boldsymbol{\varphi}$.
Therefore the points $\mathbf{x}$ and $\mathbf{x} + \boldsymbol{\delta}\mathbf{x}$ have the distance

$$\begin{aligned}|\boldsymbol{\varphi}(\mathbf{x} + \boldsymbol{\delta}\mathbf{x}) - \boldsymbol{\varphi}(\mathbf{x})|^2 &= |\nabla\boldsymbol{\varphi}(\mathbf{x})\boldsymbol{\delta}\mathbf{x} + o(|\boldsymbol{\delta}\mathbf{x}|)|^2 \\ &= |\nabla\boldsymbol{\varphi}(\mathbf{x})\boldsymbol{\delta}\mathbf{x}|^2 + 2\left(\nabla\boldsymbol{\varphi}(\mathbf{x})\boldsymbol{\delta}\mathbf{x}\right)^T o(|\boldsymbol{\delta}\mathbf{x}|) + |o(|\boldsymbol{\delta}\mathbf{x}|)|^2 \\ &= (\nabla\boldsymbol{\varphi}(\mathbf{x})\boldsymbol{\delta}\mathbf{x})^T\nabla\boldsymbol{\varphi}(\mathbf{x})\boldsymbol{\delta}\mathbf{x} + o(|\boldsymbol{\delta}\mathbf{x}|^2) = (\boldsymbol{\delta}\mathbf{x})^T\mathbf{C}\boldsymbol{\delta}\mathbf{x} + o(|\boldsymbol{\delta}\mathbf{x}|^2)\end{aligned}$$

after its deformation. In the next - to - last identity we have used the fact that the Frobenius norm is consistent with the Euclidean norm. Hence the tensor $\mathbf{C}$ is involved in transforming the distance of two points due to a deformation $\boldsymbol{\varphi}$ and measures therefore how the points are „strained“ after the deformation.

Rigid body motions:

Another important term in the context of strains are the so - called rigid body motions. A deformation $\boldsymbol{\varphi} \neq \mathbf{id}$ is called a **rigid body motion** (or rigid deformation) if it is of the form

$$\boldsymbol{\varphi}(\mathbf{x}) = \mathbf{a} + \mathbf{Q}\mathbf{x} \tag{2.6}$$

with a translation $\mathbf{a} \in \mathbb{R}^3$ and a rotation $\mathbf{Q} \in \mathbb{O} := \{\mathbf{R} \in \mathbb{R}^{3\times 3} : \mathbf{R}^T\mathbf{R} = \mathbf{I} = \mathbf{R}\mathbf{R}^T, \det\mathbf{R} = 1\}$ about the origin. If we provide $\det\nabla\boldsymbol{\varphi} > 0$, which is physically reasonable, and assume that $\boldsymbol{\varphi} \in C^1(\Omega, \mathbb{R}^3)$ then we get

$$\boldsymbol{\varphi} \text{ is a rigid body motion} \Leftrightarrow \boldsymbol{\varphi}(\mathbf{x}) = \mathbf{a} + \mathbf{Q}\mathbf{x} \ \ \forall\, \mathbf{x} \in \Omega \Leftrightarrow \mathbf{C} = (\nabla\boldsymbol{\varphi})^T\nabla\boldsymbol{\varphi} = \mathbf{I} \ \ \forall\, \mathbf{x} \in \Omega$$

with a translation $\mathbf{a} \in \mathbb{R}^3$ and a rotation $\mathbf{Q} \in \mathbb{O}$. This equivalence could be understood as the characterization of a rigid body motion. If the mapping $\boldsymbol{\varphi}$ is even continuously differentiable on $\bar{\Omega}$ then the statements mentioned here hold for all $\mathbf{x} \in \bar{\Omega}$. Note that due to

$$\mathbf{C} = (\nabla\boldsymbol{\varphi})^T\mathbf{B}(\nabla\boldsymbol{\varphi})^{-T} \Leftrightarrow \mathbf{B} = (\nabla\boldsymbol{\varphi})^{-T}\mathbf{C}(\nabla\boldsymbol{\varphi})^T$$

it holds $\mathbf{C} = \mathbf{I}$ if and only if $\mathbf{B} = \mathbf{I}$.
Another remarkable observation for two deformations $\boldsymbol{\varphi}, \boldsymbol{\psi} \in C^1(\Omega, \mathbb{R}^3)$, which are injective in Ω and orientation - preserving in $\bar{\Omega}$, is

$$\begin{aligned}&\boldsymbol{\varphi} \text{ and } \boldsymbol{\psi} \text{ have the same strain tensor } \mathbf{C} \text{ everywhere in the body} \\ &\Leftrightarrow (\nabla\boldsymbol{\varphi})^T\nabla\boldsymbol{\varphi} = (\nabla\boldsymbol{\psi})^T\nabla\boldsymbol{\psi} \quad \forall\, \mathbf{x} \in \Omega \\ &\Leftrightarrow \boldsymbol{\varphi}(\mathbf{x}) = \mathbf{a} + \mathbf{Q}\boldsymbol{\psi}(\mathbf{x}) \quad \forall\, \mathbf{x} \in \Omega \text{ with } \mathbf{a} \in \mathbb{R}^3 \text{ and rotation } \mathbf{Q} \in \mathbb{O}.\end{aligned} \tag{2.7}$$

Note that all these considerations are not restricted to the case $n = 3$ and thus are valid for an arbitrary dimension $n \in \mathbb{N} \setminus \{0\}$. For proofs and further details to rigid body motions we refer to Section 1.8 in [Cia88].
An immediate consequence of (2.7) is that if two deformations differ only in a rotation and a translation, then the corresponding strain tensors and also the stress tensors coincide (due to the given stress - strain relation). Logically the uniqueness of solutions is problematic if rigid body motions are not eliminated. For practical purposes the rigid body motions are eliminated by suitable boundary conditions. After eliminating all rigid body motions we know that $\mathbf{C} = \mathbf{I}$ if and only if $\boldsymbol{\varphi} = \mathbf{id}$ or equivalently $\mathbf{u} = \mathbf{0}$.
In three dimensions we have six rigid body motions, the three translations

$$\mathbf{a}_x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{a}_y = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{a}_z = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

in x -, y - and z - direction and the three rotations

$$\mathbf{Q}_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}, \quad \mathbf{Q}_y = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix}, \quad \mathbf{Q}_z = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

about the x -, y - and z - axis with arbitrary rotation angles $\alpha, \beta, \gamma \in [0, 2\pi)$. Each deformation $\boldsymbol{\varphi} = \mathbf{a} + \mathbf{Q}\mathbf{x}$, $\boldsymbol{\varphi} \neq \mathbf{id}$, with $\mathbf{a} \in \operatorname{span}\{\mathbf{a}_x, \mathbf{a}_y, \mathbf{a}_z\}$ and $\mathbf{Q}$ an arbitrary matrix product of the set $\{\mathbf{I}, \mathbf{Q}_x, \mathbf{Q}_y, \mathbf{Q}_z\}$ satisfies $\mathbf{C} = (\nabla\boldsymbol{\varphi})^T \nabla\boldsymbol{\varphi} = \mathbf{I}$ and is therefore a rigid body motion by the characterization above.
At the end of this section we remark that the strain tensor $\mathbf{E}$ is a measure of the deviation between a rigid body motion and a given deformation. By definition of $\mathbf{E}$ it holds

$$\boldsymbol{\varphi} \text{ is a rigid body motion} \Leftrightarrow \mathbf{C} = (\nabla\boldsymbol{\varphi})^T \nabla\boldsymbol{\varphi} = \mathbf{I} \text{ for all } \mathbf{x} \in \Omega \Leftrightarrow \mathbf{E} = \mathbf{0} \text{ for all } \mathbf{x} \in \Omega.$$

2.2.3 Lamé constants and incompressibility

In this section we will introduce the Lamé constants λ and μ based on the explanations in [Dem03] and [Cia88]. The Lamé constants will appear in our material law models later. One requirement of our discretization scheme is that it should be reliable for (quasi -) incompressible materials. In the following we will motivate the fact that $\lambda \to \infty$ is an indicator for (quasi -) incompressible materials in linear elasticity.
To introduce the Lamé constants and motivate that a large value of λ corresponds to an almost incompressible material we consider an uniaxial tension test on a rectangular cuboid with length l, thickness and height d and thus area cross section $q := d^2$.
A visualization of this tension test is depicted in Figure 2.3 where the reference configuration is drawn in blue and the deformed configuration is drawn in orange. The experiment

is configured in such a way that the displacement in x - direction on the right face, the displacement in y - direction on the back face and the displacement in z - direction on the bottom face is zero.

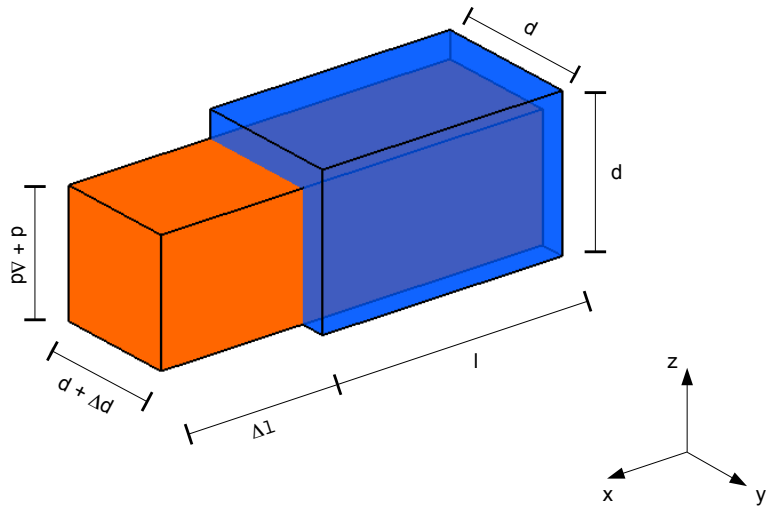


Figure 2.3: Visualization of an uniaxial tension test

For a traction force in x - direction the cuboid becomes longer and thinner. Conversely, for an applied force which compresses the body in x - direction, the cuboid becomes smaller and thicker. In the following Δd denotes the change of thickness and Δl denotes the change of length (see Figure 2.3).
The initial volume of the body is $V_1 = d^2 \cdot l$. The volume after the deformation is given by $V_2 = (d + \Delta d)^2 \cdot (l + \Delta l)$. The change of volume is therefore

$$\begin{aligned} \Delta V := V_2 - V_1 &= (d + \Delta d)^2 \cdot (l + \Delta l) - d^2 l \\ &= (d^2 + 2d\Delta d + (\Delta d)^2) \cdot (l + \Delta l) - d^2 l \\ &= d^2 \Delta l + 2dl\Delta d + 2d\Delta d\Delta l + (\Delta d)^2 \cdot (l + \Delta l). \end{aligned}$$

For small $\Delta l, \Delta d << 1$ we can neglect the higher order terms in the last two summands and obtain approximately

$$\Delta V \approx d^2 \Delta l + 2dl\Delta d.$$

Dividing this equation by $V = d^2 l$ leads to

$$\frac{\Delta V}{V} \approx \frac{d^2 \Delta l + 2dl\Delta d}{d^2 l} = \frac{\Delta l}{l} + 2\frac{\Delta d}{d}.$$

In mechanics, **Poisson's ratio** ν is defined as the negative quotient of the relative change of thickness and the relative change of length, i.e.

$$\nu := -\frac{\frac{\Delta d}{d}}{\frac{\Delta l}{l}}.$$

Obviously ν is a dimensionless physical quantity. With this definition and the relative change of length $\varepsilon_x := \frac{\Delta l}{l}$ in x - direction we get

$$\frac{\Delta V}{V} \approx \frac{\Delta l}{l}\left(1 + 2\frac{\frac{\Delta d}{d}}{\frac{\Delta l}{l}}\right) = \varepsilon_x(1-2\nu).$$

Applying a traction force on the body, the physical intuitive and normal behavior of a material is that the volume becomes larger, the length increases and the thickness decreases, i.e. $\Delta V > 0$, $\Delta l > 0$ and $\Delta d < 0$. Therefore we obtain $0 < \nu < \frac{1}{2}$. If we apply a force which compresses the body, the length shrinks, the thickness increases and the volume decreases, i.e. $\Delta l < 0$, $\Delta d > 0$ and $\Delta V < 0$. Also in this case we get $0 < \nu < \frac{1}{2}$. For $\nu = \frac{1}{2}$ the body does not change its volume, i.e. $\nu \to \frac{1}{2}$ is characteristic for an incompressible material in linear elasticity.
Another characteristic parameter for materials is the so - called **Young's modulus**. If we assume linear elastic behavior and apply an uniaxial force in x - direction then by Hooke's law the stress σ_x is proportional to the elongation ε_x, i.e.

$$\sigma_x = E\varepsilon_x \text{ (see Section 6.2 in [Dem03])}.$$

The constant E in this equation is Young's modulus and can be determined by such simple physical experiments. For given traction stress $\sigma_x := \frac{F}{q}$ and relative elongation $\varepsilon_x = \frac{\Delta l}{l}$ one can measure Δl and determine E through

$$E = \frac{F}{q} \cdot \frac{l}{\Delta l}.$$

Here F describes the value of a given force in x - direction. Obviously the physical SI unit of E is $\left[\frac{N}{m^2}\right]$.
For positive F we have $\Delta l > 0$ and for negative F we have $\Delta l < 0$. In both cases we obtain $E > 0$.
Altogether the characteristic parameters $E > 0$ and $0 < \nu < \frac{1}{2}$ of any material can be determined by simple uniaxial tension experiments. For given Young's modulus E and Poisson's ratio ν we define the **Lamé constants** λ, μ as

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

Obviously, both quantities have the physical SI unit $\left[\frac{N}{m^2}\right]$. Furthermore for $E > 0$ and $0 < \nu < \frac{1}{2}$ it holds $\lambda, \mu > 0$ and $\lambda \to \infty$ if and only if $\nu \to \frac{1}{2}$. Thus characteristical for an incompressible material is $\nu \to \frac{1}{2}$ or equivalently $\lambda \to \infty$.
For given Lamé constants we can determine E and ν by

$$\nu = \frac{\lambda}{2(\mu+\lambda)}, \quad E = \frac{\mu(2\mu+3\lambda)}{\mu+\lambda}.$$

Next to $\lambda \to \infty$ a further reasonable condition for an incompressible material is the constraint $\det \mathbf{F} = \det \nabla \boldsymbol{\varphi} = 1$ in the whole body, since then it holds for the deformation $\boldsymbol{\varphi} : \bar{\Omega} \to \bar{\Omega}^{\varphi} \subset \mathbb{R}^3$, $\boldsymbol{\varphi}(\bar{\Omega}) =: \bar{\Omega}^{\varphi}$ and $f : \bar{\Omega}^{\varphi} \to \mathbb{R}, \mathbf{x}^{\varphi} \mapsto 1$, with the help of integration by substitution

$$\mathrm{vol}(\bar{\Omega}^{\varphi}) = \int_{\bar{\Omega}^{\varphi}} dx^{\varphi} = \int_{\bar{\Omega}^{\varphi}} f(\mathbf{x}^{\varphi})\, dx^{\varphi} = \int_{\bar{\Omega}} \underbrace{f(\boldsymbol{\varphi}(\mathbf{x}))}_{=1} \underbrace{|\det \nabla \boldsymbol{\varphi}|}_{=1} dx = \int_{\bar{\Omega}} dx = \mathrm{vol}(\bar{\Omega}), \quad (2.8)$$

i.e. the volume of the reference configuration $\bar{\Omega}$ and the volume of the deformed configuration $\bar{\Omega}^{\varphi}$ are equal. Physically, this means that the material is incompressible.
If we assume vice versa that the volume of each subdomain $A \subseteq \bar{\Omega}$ with corresponding subdomain $A^{\varphi} = \boldsymbol{\varphi}(A) \subseteq \bar{\Omega}^{\varphi}$ in the deformed configuration is preserved, i.e. each subdomain is incompressible with $\mathrm{vol}(A^{\varphi}) = \mathrm{vol}(A)$, we get with $\mathrm{vol}(A^{\varphi}) = \int_A \det \mathbf{F}\, dx$ and $\mathrm{vol}(A) = \int_A 1\, dx$ the condition

$$\int_A (\det \mathbf{F} - 1) dx = 0 \text{ for an arbitrary } A \subseteq \bar{\Omega}.$$

Therefore it holds $\det \mathbf{F} = 1$ for all $\mathbf{x} \in \bar{\Omega}$.
The condition

$$\boxed{\det \mathbf{F} = \det \nabla \boldsymbol{\varphi} = 1 \text{ for all } \mathbf{x} \in \bar{\Omega}} \quad (2.9)$$

to the given deformation $\boldsymbol{\varphi}$ is called **incompressibility constraint**.
At the end of this excursion to the Lamé constants please note that the usage of these constants in nonlinear models are only reasonable if they are consistent with the linear elasticity model. For the consistency of nonlinear models with linear elasticity see Section 2.4.5.

2.2.4 Possible nonlinearities

In linear elasticity the strain tensor $\boldsymbol{\varepsilon} := \boldsymbol{\varepsilon}(\mathbf{u})$ is linear in $\mathbf{u}$ and the material law

$$\boldsymbol{\sigma}(\boldsymbol{\varepsilon}) = 2\mu\, \boldsymbol{\varepsilon} + \lambda\, \mathrm{tr}(\boldsymbol{\varepsilon})\mathbf{I} =: \mathcal{C}\boldsymbol{\varepsilon} \quad (2.10)$$

with constants $\lambda, \mu > 0$ is linear in $\boldsymbol{\varepsilon}$. Our aim in this work is to consider nonlinear problems. Therefore we introduce the following possible nonlinearities:

1. Nonlinear kinematics:
   The relation between the considered strain tensor and the displacement $\mathbf{u}$ is nonlinear, e.g. the strain tensors $\mathbf{B}(\mathbf{u}), \mathbf{C}(\mathbf{u}), \mathbf{E}(\mathbf{u})$ are nonlinear in $\mathbf{u}$.
   An example is the St. Venant - Kirchhoff material where indeed a linear stress - strain relation

   $$\boldsymbol{\Sigma}(\mathbf{E}) = 2\mu\, \mathbf{E} + \lambda\, \mathrm{tr}(\mathbf{E})\mathbf{I} \quad (2.11)$$

   is used, but nonlinear kinematics are taken into account.

2. Nonlinear material law:
   The relation between stress and strain is nonlinear, e.g. $\mathbf{\Sigma}(\mathbf{E})$ is nonlinear in $\mathbf{E}$. Practical nonlinear material laws are for instance Neo - Hooke - and Mooney - Rivlin models. These models will be introduced later and discussed in more detail.

Note that also other nonlinearities can additionally occur, see Section 5.9 in [Cia88]. However the examples in Section 6 are restricted to the two mentioned nonlinearities.

2.2.5 Hyperelasticity and important properties

In this section we are introducing the terms „elasticity“ and „hyperelasticity“ mathematically. We further introduce an important material independent property, the material frame - indifference, which must be satisfied due to physical reasons, and two material dependent properties (homogeneity and isotropy) that a given material can possess.
Let us start with the definition of an elastic material. In mathematical elasticity theory a material is **elastic** (cf. Section 3.1 in [Cia88]) if the two Piola - Kirchhoff stress tensors $\mathbf{P}, \mathbf{\Sigma} : \bar{\Omega} \to \mathbb{R}^{3\times 3}$ can be expressed in terms of $\mathbf{x}$ and $\mathbf{F} = \nabla\boldsymbol{\varphi}$ by

$$\mathbf{P}(\mathbf{x}) = \hat{\mathbf{P}}(\mathbf{x}, \mathbf{F}), \quad \mathbf{\Sigma}(\mathbf{x}) = \hat{\mathbf{\Sigma}}(\mathbf{x}, \mathbf{F}) \quad \forall\, \mathbf{x} \in \bar{\Omega}. \tag{2.12}$$

The functions $\hat{\mathbf{P}}, \hat{\mathbf{\Sigma}} : \bar{\Omega} \times \mathbb{M} \to \mathbb{R}^{3\times 3}$ are called **response functions** and characterize the material. Here $\mathbb{M} := \{\mathbf{F} \in \mathbb{R}^{3\times 3} : \det \mathbf{F} > 0\}$ is the set of all three - dimensional quadratic matrices with positive determinant. In this work we are dealing with so - called hyperelastic materials, following the definition of Section 4.1 in [Cia88]:

Definition 2.20: (Hyperelastic material)
Let an elastic material with response function $\hat{\mathbf{P}} : \bar{\Omega} \times \mathbb{M} \to \mathbb{R}^{3\times 3}$ be given, such that the first Piola - Kirchhoff stress tensor is expressed by $\mathbf{P}(\mathbf{x}) = \hat{\mathbf{P}}(\mathbf{x}, \mathbf{F})$, $\mathbf{x} \in \bar{\Omega}$. Then this material is called **hyperelastic** if there exists a function $\hat{\psi} : \bar{\Omega} \times \mathbb{M} \to \mathbb{R}$, Fréchet differentiable with respect to $\mathbf{F} \in \mathbb{M}$, such that

$$\hat{\mathbf{P}}(\mathbf{x}, \mathbf{F}) = \partial_{\mathbf{F}} \hat{\psi}(\mathbf{x}, \mathbf{F}), \quad \mathbf{x} \in \bar{\Omega}, \mathbf{F} \in \mathbb{M}. \tag{2.13}$$

The function $\hat{\psi}$ is called **stored energy function**.

Definition 2.21: (Cofactor of a matrix)
Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ be an arbitrary matrix with $n \geq 2$. Then the **cofactor** of $\mathbf{A}$ is defined as the matrix $\mathbf{Cof\, A} \in \mathbb{C}^{n\times n}$ with matrix entries

$$(\mathbf{Cof\, A})_{i,j} = (-1)^{i+j} \det \mathbf{A}'_{i,j}, \quad 1 \leq i, j \leq n,$$

where $\mathbf{A}'_{i,j} \in \mathbb{C}^{(n-1)\times(n-1)}$ denotes the matrix obtained by deleting the i - th row and j - th column of $\mathbf{A}$.
If $\mathbf{A}$ is additionally invertible we have the representation (see Section 1.1 in [Cia88])

$$\mathbf{Cof\, A} = (\det \mathbf{A})\, \mathbf{A}^{-T}. \tag{2.14}$$

Remark 2.22:
The cofactor of a matrix $\mathbf{A} \in \mathbb{C}^{3\times 3}$, whether invertible or not, has the representation

$$\mathbf{Cof\,A} = \begin{pmatrix} A_{22}A_{33} - A_{23}A_{32} & A_{23}A_{31} - A_{21}A_{33} & A_{21}A_{32} - A_{22}A_{31} \\ A_{13}A_{32} - A_{12}A_{33} & A_{11}A_{33} - A_{13}A_{31} & A_{12}A_{31} - A_{11}A_{32} \\ A_{12}A_{23} - A_{13}A_{22} & A_{13}A_{21} - A_{11}A_{23} & A_{11}A_{22} - A_{12}A_{21} \end{pmatrix}. \tag{2.15}$$

Remark 2.23: (Suitable properties of the stored energy function)
Physically reasonable requirements of the stored energy function $\hat{\psi} : \bar{\Omega} \times \mathbb{M} \to \mathbb{R}$ are (see Section 4.6 in [Cia88]):

1. $\det \mathbf{F} \to\ 0 \Rightarrow \hat{\psi}(\mathbf{x}, \mathbf{F}) \to \infty$, $\mathbf{F} \in \mathbb{M}, \mathbf{x} \in \bar{\Omega}$
   Roughly speaking this condition says that an infinitely large energy is necessary to compress the body to the volume 0.

2. $(|\mathbf{F}| + |\mathbf{Cof\,F}| + \det \mathbf{F}) \to \infty \Rightarrow \hat{\psi}(\mathbf{x}, \mathbf{F}) \to \infty$, $\mathbf{F} \in \mathbb{M}, \mathbf{x} \in \bar{\Omega}$
   Roughly speaking this condition says that an infinitely large energy is necessary to get extreme strains.

Material frame - indifference:

After these definitions and remarks we come to an important material - independent property, the **material frame - indifference**. This introduction is based on the explanations in [EGK11]. Roughly speaking material frame - indifference demands that the material behavior must not depend on the observer. Therefore scalar -, vector - and matrix - valued functions must be transformed in an appropriate way if one changes the coordinate system. Since every observer can be identified by an own coordinate system this property is also called **observer invariance** in literature. This property is generally physically necessary, since for instance the temperature of a body has to be clearly independent of the observer. For this purpose we consider two arbitrary orthonormal and positive oriented coordinate systems in the three - dimensional Euclidean space, spanned by $B^\tau := \{\mathbf{e}_1^\tau, \mathbf{e}_2^\tau, \mathbf{e}_3^\tau\}$ and $B^* := \{\mathbf{e}_1^*, \mathbf{e}_2^*, \mathbf{e}_3^*\}$ with origins $\mathbf{O}^\tau$ and $\mathbf{O}^*$.
For these given bases we can compute a vector $\mathbf{a} \in \mathbb{R}^3$ and a rotation $\mathbf{Q} \in \mathbb{O}$ such that $\mathbf{O}^\tau - \mathbf{O}^* = \sum_{j=1}^{3} a_j \mathbf{e}_j^*$ and $\mathbf{e}_j^\tau = \sum_{i=1}^{3} Q_{ij} \mathbf{e}_i^*$, $j = 1, 2, 3$. We can express an arbitrary point $\mathbf{x} \in \mathbb{R}^3$ through

$$\mathbf{x} = \mathbf{O}^\tau + \sum_{j=1}^{3} x_j^\tau \mathbf{e}_j^\tau \text{ and } \mathbf{x} = \mathbf{O}^* + \sum_{j=1}^{3} x_j^* \mathbf{e}_j^*$$

with coordinate/coefficient vectors $\mathbf{x}^\tau = (x_1^\tau, x_2^\tau, x_3^\tau)$ and $\mathbf{x}^* = (x_1^*, x_2^*, x_3^*)$ according to the two given coordinate systems. Both coordinate vectors are then related by $\mathbf{x}^* = \mathbf{a} + \mathbf{Q}\mathbf{x}^\tau$,

since

$$\begin{aligned}\mathbf{x}_k^* &= (\mathbf{e}_k^*)^T \sum_{j=1}^{3} x_j^* \mathbf{e}_j^* = (\mathbf{e}_k^*)^T (\mathbf{O}^\tau - \mathbf{O}^*) + (\mathbf{e}_k^*)^T \sum_{j=1}^{3} x_j^\tau \mathbf{e}_j^\tau \\ &= (\mathbf{e}_k^*)^T \sum_{j=1}^{3} a_j \mathbf{e}_j^* + \sum_{j=1}^{3} x_j^\tau \, (\mathbf{e}_k^*)^T \sum_{i=1}^{3} Q_{ij} \mathbf{e}_i^* \\ &= a_k + \sum_{j=1}^{3} Q_{kj} x_j^\tau = a_k + (\mathbf{Q}\mathbf{x}^\tau)_k = (\mathbf{a} + \mathbf{Q}\mathbf{x}^\tau)_k \,, \quad k = 1,2,3.\end{aligned}$$

Below we interpret the coordinate vectors $\mathbf{x}^\tau = \boldsymbol{\varphi}^\tau(\mathbf{x})$ and $\mathbf{x}^* = \boldsymbol{\varphi}^*(\mathbf{x})$ as Eulerian coordinates over the same reference configuration with coordinates $\mathbf{x}$ and invertible mappings $\boldsymbol{\varphi}^\tau, \boldsymbol{\varphi}^* : \mathbb{R}^3 \to \mathbb{R}^3$. For a scalar - valued function $T : \mathbb{R}^3 \to \mathbb{R}$, defined on the reference configuration, and a fixed point $\mathbf{x} \in \mathbb{R}^3$ with Eulerian coordinates $\mathbf{x}^\tau$ and $\mathbf{x}^*$ the axiom of material frame - indifference states

$$T^\tau(\mathbf{x}^\tau) = T^*(\mathbf{x}^*) \text{ with } \mathbf{x}^* = \mathbf{a} + \mathbf{Q}\mathbf{x}^\tau.$$

Here $T^\tau(\mathbf{x}^\tau) := T\left((\boldsymbol{\varphi}^\tau)^{-1}(\mathbf{x}^\tau)\right)$, $T^*(\mathbf{x}^*) := T\left((\boldsymbol{\varphi}^*)^{-1}(\mathbf{x}^*)\right)$ are the corresponding scalar - valued functions to T, expressed in the Eulerian coordinates.
For an arbitrary vector - valued function $\mathbf{v} : \mathbb{R}^3 \to \mathbb{R}^3$ and an arbitrary point $\mathbf{x} \in \mathbb{R}^3$, we consider the vector $\mathbf{q} := \mathbf{v}(\mathbf{x})$. With the Eulerian coordinates $\mathbf{x}^*$ and $\mathbf{x}^\tau$ to $\mathbf{x}$, we can find analogously as above corresponding vector - valued functions $\mathbf{v}^*$ and $\mathbf{v}^\tau$ with $\mathbf{v}^*(\mathbf{x}^*) = \mathbf{q} = \mathbf{v}^\tau(\mathbf{x}^\tau)$. We express the vector $\mathbf{q}$ in the two different coordinate systems as

$$\mathbf{q} = \sum_{j=1}^{3} q_j^*(\mathbf{x}^*)\mathbf{e}_j^* \quad \text{and} \quad \mathbf{q} = \sum_{j=1}^{3} q_j^\tau(\mathbf{x}^\tau)\mathbf{e}_j^\tau, \quad \mathbf{x}^* = \mathbf{a} + \mathbf{Q}\mathbf{x}^\tau,$$

with coefficient vectors $\mathbf{q}^*(\mathbf{x}^*) := (q_1^*(\mathbf{x}^*), q_2^*(\mathbf{x}^*), q_3^*(\mathbf{x}^*))$ and $\mathbf{q}^\tau(\mathbf{x}^\tau) := (q_1^\tau(\mathbf{x}^\tau), q_2^\tau(\mathbf{x}^\tau), q_3^\tau(\mathbf{x}^\tau))$, related by

$$\mathbf{q}^*(\mathbf{x}^*) = \mathbf{Q}\mathbf{q}^\tau(\mathbf{x}^\tau), \quad \mathbf{x}^* = \mathbf{a} + \mathbf{Q}\mathbf{x}^\tau. \tag{2.16}$$

The relation (2.16) between $\mathbf{q}^*(\mathbf{x}^*)$ and $\mathbf{q}^\tau(\mathbf{x}^\tau)$ can be proven similarly as above and is the condition of material frame - indifference for vector - valued functions.
For corresponding tensors $\boldsymbol{\sigma}^\tau(\mathbf{x}^\tau)$, $\boldsymbol{\sigma}^*(\mathbf{x}^*)$ to a given mapping $\boldsymbol{\sigma} : \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$ and vectors $\mathbf{n}^\tau(\mathbf{x}^\tau)$, $\mathbf{n}^*(\mathbf{x}^*)$ to a given mapping $\mathbf{n} : \mathbb{R}^3 \to \mathbb{R}^3$, related by

$$\mathbf{n}^*(\mathbf{x}^*) = \mathbf{Q}\mathbf{n}^\tau(\mathbf{x}^\tau) \Leftrightarrow \mathbf{n}^\tau(\mathbf{x}^\tau) = \mathbf{Q}^T\mathbf{n}^*(\mathbf{x}^*), \quad \mathbf{x}^* = \mathbf{a} + \mathbf{Q}\mathbf{x}^\tau, \tag{2.17}$$

we define the vector - valued functions $\mathbf{q}^\tau(\mathbf{x}^\tau) := (\boldsymbol{\sigma}^\tau(\mathbf{x}^\tau)) \cdot \mathbf{n}^\tau(\mathbf{x}^\tau)$ and
$\mathbf{q}^*(\mathbf{x}^*) := (\boldsymbol{\sigma}^*(\mathbf{x}^*)) \cdot \mathbf{n}^*(\mathbf{x}^*)$, insert it into condition (2.16) of material frame - indifference of vectors and obtain

$$(\boldsymbol{\sigma}^*(\mathbf{x}^*)) \cdot \mathbf{n}^*(\mathbf{x}^*) = (\mathbf{Q}\boldsymbol{\sigma}^\tau(\mathbf{x}^\tau)) \cdot \mathbf{n}^\tau(\mathbf{x}^\tau) = \left(\mathbf{Q}\boldsymbol{\sigma}^\tau(\mathbf{x}^\tau)\mathbf{Q}^T\right) \cdot \mathbf{n}^*(\mathbf{x}^*)$$

for arbitrary $\mathbf{n}^*(\mathbf{x}^*)$. This results in the condition of material frame - indifference of matrix - valued functions,

$$\boldsymbol{\sigma}^*(\mathbf{x}^*) = \mathbf{Q}\boldsymbol{\sigma}^\tau(\mathbf{x}^\tau)\mathbf{Q}^T, \quad \mathbf{x}^* = \mathbf{a} + \mathbf{Q}\mathbf{x}^\tau.$$

In elasticity theory one assumes now the material frame - indifference for the so - called Cauchy stress vector using (2.16) and (2.17) (cf. Axiom 3.3 - 1 in [Cia88]). This is equivalent to the requirement

$$\hat{\mathbf{P}}(\mathbf{x}, \mathbf{QF}) = \mathbf{Q}\hat{\mathbf{P}}(\mathbf{x}, \mathbf{F}) \,\forall\, \mathbf{F} \in \mathbb{M}, \mathbf{x} \in \bar{\Omega} \Leftrightarrow \hat{\boldsymbol{\Sigma}}(\mathbf{x}, \mathbf{QF}) = \hat{\boldsymbol{\Sigma}}(\mathbf{x}, \mathbf{F}) \,\forall\, \mathbf{F} \in \mathbb{M}, \mathbf{x} \in \bar{\Omega}$$

for the response functions of the Piola - Kirchhoff stress tensors with arbitrary rotations $\mathbf{Q} \in \mathbb{O}$ (cf. Section 3.3 in [Cia88]).
A hyperelastic material with stored energy function $\hat{\psi}$ satisfies the property of material frame - indifference if and only if

$$\hat{\psi}(\mathbf{x}, \mathbf{QF}) = \hat{\psi}(\mathbf{x}, \mathbf{F}), \quad \mathbf{F} \in \mathbb{M}, \mathbf{Q} \in \mathbb{O}, \mathbf{x} \in \bar{\Omega}. \tag{2.18}$$

If we assume a hyperelastic material with the material frame - indifferent property (2.18) for the corresponding stored energy function $\hat{\psi}$ there exists a function $\psi : \bar{\Omega} \times \mathbb{R}^{3\times3}_{\text{sym}} \to \mathbb{R}$ with

$$\psi(\mathbf{x}, \mathbf{C}) = \hat{\psi}(\mathbf{x}, \mathbf{F}), \quad \mathbf{C} = \mathbf{F}^T\mathbf{F}, \quad \mathbf{F} \in \mathbb{M}. \tag{2.19}$$

For a proof of (2.18) and (2.19) we refer to Theorem 4.2 - 1 in [Cia88]. A more detailed introduction into the Cauchy stress vector and the fact of material frame - indifference for elastic and hyperelastic materials can be found in Chapters 2, 3 and 4 of [Cia88].
The next purpose is to derive a relation between the second Piola - Kirchhoff stress tensor $\boldsymbol{\Sigma}$ and the gradient of ψ with respect to $\mathbf{C}$. Firstly, we state a simple lemma:

Lemma 2.24:
Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$ be matrices and $\mathbf{A}$ symmetric. Then it holds

$$\begin{aligned} \operatorname{tr}\left(\mathbf{A}\left(\mathbf{B} + \mathbf{B}^T\right)\mathbf{A}\right) &= 2\operatorname{tr}(\mathbf{ABA}) \\ \operatorname{tr}\left(\mathbf{A}\left(\mathbf{B} + \mathbf{B}^T\right)\right) &= 2\operatorname{tr}(\mathbf{AB}). \end{aligned}$$

Proof:
With the calculation rules of the trace operator and the assumption $\mathbf{A} = \mathbf{A}^T$ it holds

$$\begin{aligned} \operatorname{tr}\left(\mathbf{A}(\mathbf{B} + \mathbf{B}^T)\mathbf{A}\right) &= \operatorname{tr}(\mathbf{ABA}) + \operatorname{tr}\left(\mathbf{AB}^T\mathbf{A}\right) = \operatorname{tr}(\mathbf{ABA}) + \operatorname{tr}\left(\mathbf{A}^T\mathbf{BA}^T\right) \\ &= \operatorname{tr}(\mathbf{ABA}) + \operatorname{tr}(\mathbf{ABA}) = 2\operatorname{tr}(\mathbf{ABA}), \\ \operatorname{tr}\left(\mathbf{A}(\mathbf{B} + \mathbf{B}^T)\right) &= \operatorname{tr}(\mathbf{AB}) + \operatorname{tr}\left(\mathbf{AB}^T\right) = \operatorname{tr}(\mathbf{AB}) + \operatorname{tr}\left(\mathbf{BA}^T\right) \\ &= \operatorname{tr}(\mathbf{AB}) + \operatorname{tr}(\mathbf{BA}) = \operatorname{tr}(\mathbf{AB}) + \operatorname{tr}(\mathbf{AB}) = 2\operatorname{tr}(\mathbf{AB}). \end{aligned}$$

□

With the help of this lemma we obtain secondly the following lemma for a hyperelastic material with the property of frame - indifference:

Lemma 2.25:
Let $\hat{\psi} : \bar{\Omega} \times \mathbb{M} \to \mathbb{R}$ and $\psi : \bar{\Omega} \times \mathbb{R}^{3\times 3}_{\text{sym}} \to \mathbb{R}$ be Fréchet differentiable with respect to $\mathbf{F} \in \mathbb{M}$ (respectively $\mathbf{C} \in \mathbb{R}^{3\times 3}_{\text{sym}}$) and $\psi(\mathbf{x}, \mathbf{C}) = \hat{\psi}(\mathbf{x}, \mathbf{F})$ for $\mathbf{C} = \mathbf{F}^T\mathbf{F}$, $\mathbf{F} \in \mathbb{M}$. Then it holds for the Piola - Kirchhoff stress tensors $\mathbf{P}, \boldsymbol{\Sigma}$ in each $\mathbf{x} \in \bar{\Omega}$

$$\mathbf{P}(\mathbf{x}) = \partial_{\mathbf{F}}\hat{\psi}(\mathbf{x}, \mathbf{F}) = 2\mathbf{F}\partial_{\mathbf{C}}\psi(\mathbf{x}, \mathbf{C}) \Leftrightarrow \boldsymbol{\Sigma}(\mathbf{x}) = 2\partial_{\mathbf{C}}\psi(\mathbf{x}, \mathbf{C}).$$

Proof:
The mapping $\mathbf{F} \mapsto \mathbf{F}^T\mathbf{F}$ is Fréchet differentiable with derivative $\mathbf{H}^T\mathbf{F} + \mathbf{F}^T\mathbf{H}$, since

$$(\mathbf{F} + \mathbf{H})^T(\mathbf{F} + \mathbf{H}) = \mathbf{F}^T\mathbf{F} + \mathbf{H}^T\mathbf{F} + \mathbf{F}^T\mathbf{H} + \mathbf{H}^T\mathbf{H} = \mathbf{F}^T\mathbf{F} + \mathbf{H}^T\mathbf{F} + \mathbf{F}^T\mathbf{H} + o(|\mathbf{H}|).$$

With the function $h(\mathbf{F}) := \psi(\mathbf{x}, \mathbf{C}) = \psi(\mathbf{x}, \mathbf{F}^T\mathbf{F})$ for $\mathbf{F} \in \mathbb{M}$ it holds by the chain rule (Proposition 2.10), combined with Lemma 2.24 and equation (2.4)

$$\begin{aligned} h'(\mathbf{F})[\mathbf{H}] &= \psi'(\mathbf{x}, \mathbf{C})[\mathbf{H}^T\mathbf{F} + \mathbf{F}^T\mathbf{H}] = \partial_{\mathbf{C}}\psi(\mathbf{x}, \mathbf{C}) : \left[\mathbf{H}^T\mathbf{F} + \mathbf{F}^T\mathbf{H}\right] \\ &= 2\partial_{\mathbf{C}}\psi(\mathbf{x}, \mathbf{C}) : \mathbf{F}^T\mathbf{H} = 2\mathbf{F}\partial_{\mathbf{C}}\psi(\mathbf{x}, \mathbf{C}) : \mathbf{H} \end{aligned}$$

for an arbitrary $\mathbf{x} \in \bar{\Omega}$. Here we have used the fact that one can assume the symmetry of $\partial_{\mathbf{C}}\psi(\mathbf{x}, \mathbf{C})$ (see Section 4.2 in [Cia88] for a more detailed discussion).
Due to $h(\mathbf{F}) = \psi(\mathbf{x}, \mathbf{C}) = \hat{\psi}(\mathbf{x}, \mathbf{F})$ by assumption it holds $h'(\mathbf{F})[\mathbf{H}] = \hat{\psi}'(\mathbf{x}, \mathbf{F})[\mathbf{H}] = \partial_{\mathbf{F}}\hat{\psi}(\mathbf{x}, \mathbf{F}) : \mathbf{H}$. Altogether we get

$$\partial_{\mathbf{F}}\hat{\psi}(\mathbf{x}, \mathbf{F}) : \mathbf{H} = 2\mathbf{F}\partial_{\mathbf{C}}\psi(\mathbf{x}, \mathbf{C}) : \mathbf{H}, \quad \mathbf{H} \in \mathbb{M},$$

and therefore $\mathbf{P}(\mathbf{x}) = \hat{\mathbf{P}}(\mathbf{x}, \mathbf{F}) = \partial_{\mathbf{F}}\hat{\psi}(\mathbf{x}, \mathbf{F}) = 2\mathbf{F}\partial_{\mathbf{C}}\psi(\mathbf{x}, \mathbf{C})$. By relation (2.5) we also obtain $\boldsymbol{\Sigma}(\mathbf{x}) = \mathbf{F}^{-1}\mathbf{P}(\mathbf{x}) = 2\partial_{\mathbf{C}}\psi(\mathbf{x}, \mathbf{C})$.

□

Material - dependent properties are homogeneity and isotropy (respectively anisotropy). The following explanations are again based on [Cia88] and [EGK11]:

Homogeneity:

A material in the reference configuration $\bar{\Omega}$ is called **homogeneous**, if the response functions $\hat{\mathbf{P}}, \hat{\boldsymbol{\Sigma}}$ do not depend explicitly on $\mathbf{x} \in \bar{\Omega}$. In this case equation (2.12) reduces for both Piola - Kirchhoff stress tensors to

$$\mathbf{P} = \hat{\mathbf{P}}(\mathbf{F}), \quad \boldsymbol{\Sigma} = \hat{\boldsymbol{\Sigma}}(\mathbf{F}).$$

For hyperelastic materials equation (2.13) reduces to

$$\hat{\mathbf{P}}(\mathbf{F}) = \partial_{\mathbf{F}}\hat{\psi}(\mathbf{F}), \quad \mathbf{F} \in \mathbb{M}.$$

Keep in mind that the deformation gradient $\mathbf{F}$ still depends implicitly on $\mathbf{x} \in \bar{\Omega}$.

Isotropy:

The behavior of materials under loads often depends on the direction of the acting forces. For example if we consider wood we have a higher strength in direction of the wood fibers than in the direction orthogonal to these fibers. Therefore the behavior is for instance different if we apply forces in the fiber direction instead of its orthogonal direction. Such direction dependent materials are called **anisotropic**. If the behavior of the material in a point is the same no matter from which direction the forces act, we call them **isotropic**. Isotropy is therefore a property for materials which do not depend on the direction. The assumption of isotropy is an idealization, since the most materials in the real world are anisotropic. However the assumption of isotropy is often used in nonlinear material models and simplifies the model significantly.
Mathematically a material is called isotropic in $\mathbf{x} \in \bar{\Omega}$ if the response functions $\hat{\mathbf{P}}, \hat{\mathbf{\Sigma}}$ to the Piola - Kirchhoff stress tensors $\mathbf{P}, \mathbf{\Sigma}$ satisfy

$$\begin{aligned} \hat{\mathbf{P}}(\mathbf{x}, \mathbf{FQ}) &= \hat{\mathbf{P}}(\mathbf{x}, \mathbf{F})\mathbf{Q} \qquad \forall\, \mathbf{F} \in \mathbb{M}, \\ \hat{\mathbf{\Sigma}}(\mathbf{x}, \mathbf{FQ}) &= \mathbf{Q}^T \hat{\mathbf{\Sigma}}(\mathbf{x}, \mathbf{F})\mathbf{Q} \quad \forall\, \mathbf{F} \in \mathbb{M} \end{aligned} \tag{2.20}$$

with arbitrary rotations $\mathbf{Q} \in \mathbb{O}$. An elastic material is called isotropic if it is isotropic at all points in $\bar{\Omega}$ (cf. Section 3.4 in [Cia88]).
In hyperelasticity a material with stored energy function $\hat{\psi} : \bar{\Omega} \times \mathbb{M} \to \mathbb{R}$ is isotropic in $\mathbf{x} \in \bar{\Omega}$ if and only if

$$\hat{\psi}(\mathbf{x}, \mathbf{F}) = \hat{\psi}(\mathbf{x}, \mathbf{FQ}) \quad \forall\, \mathbf{F} \in \mathbb{M}$$

and rotations $\mathbf{Q} \in \mathbb{O}$ (cf. Theorem 4.3 - 1 in [Cia88]).

2.2.6 Polyconvexity

The stored energy function $\hat{\psi}(\mathbf{x}, \mathbf{F})$ for a matrix $\mathbf{F} \in \mathbb{R}^{3\times 3}$ with positive determinant, which was introduced in Section 2.2.5 for a hyperelastic material, must not be convex in general. For instance for strictly convex real - valued minimization problems it is known that if a solution exists it is unique (see Theorem 4.7 - 8 in [Cia88]). Therefore strictly convex minimization problems contradict the fact that in nonlinear elasticity the solutions are in general not unique. Also the assumption of a convex stored energy function is too strong, since it is incompatible to the first requirement of $\hat{\psi}$ in Remark 2.23 (see Theorem 4.8 - 1 in [Cia88] for a proof). Due to this fact the term „polyconvexity“ is introduced. Polyconvexity does not contradict any physical behavior and is weaker than convexity. Existence theory in nonlinear elasticity based on the minimization of

$$\tilde{\mathbf{I}}(\boldsymbol{\chi}) = \int_\Omega \hat{\psi}(\mathbf{x}, \nabla\boldsymbol{\chi})\, dx - \left(\int_\Omega \mathbf{f} \cdot \boldsymbol{\chi}\, dx + \int_{\Gamma_N} \mathbf{g} \cdot \boldsymbol{\chi}\, ds \right)$$

in a suitable set of admissible deformations $\boldsymbol{\chi}$ (see Section 3.6.1 for more details) is available if we assume among other things a polyconvex stored energy function $\hat{\psi}$. These existence theorems for different boundary conditions can be found in [Bal77].
Due to this fact a good material law should be based on a polyconvex stored energy function. In this section we define the term „polyconvexity“ (cf. Section 4.9 in [Cia88]) and state a proposition. With this proposition we can simply prove the polyconvexity of stored energy functions used in this work.

Definition 2.26: (Polyconvexity of a stored energy function)
Let $\hat{\psi} : \bar{\Omega} \times \mathbb{M} \to \mathbb{R}$ be a stored energy function defined for $\mathbb{M} = \{\mathbf{F} \in \mathbb{R}^{3\times 3} : \det \mathbf{F} > 0\}$. Then $\hat{\psi}$ is called **polyconvex** if for each $\mathbf{x} \in \bar{\Omega}$ there exists a convex function $g(\mathbf{x}, \cdot) : \mathbb{U} \to \mathbb{R}$ with

$$\mathbb{U} := \left\{(\mathbf{F}, \mathbf{Cof}\,\mathbf{F}, \det \mathbf{F}) \in \mathbb{R}^{3\times 3} \times \mathbb{R}^{3\times 3} \times (0, \infty)\right\}$$

such that

$$\hat{\psi}(\mathbf{x}, \mathbf{F}) = g(\mathbf{x}, \mathbf{F}, \mathbf{Cof}\,\mathbf{F}, \det \mathbf{F}) \quad \forall\, \mathbf{F} \in \mathbb{M}.$$

Thus if we can express $\hat{\psi}$ through a function g in $\mathbf{F}, \mathbf{Cof}\,\mathbf{F}$ and $\det \mathbf{F}$ and g is convex on $\mathbb{U}$, then $\hat{\psi}$ is polyconvex. Note that the set $\mathbb{U}$ is as the convex hull of $\{(\mathbf{F}, \mathbf{Cof}\,\mathbf{F}, \det \mathbf{F}) \in \mathbb{R}^{3\times 3} \times \mathbb{R}^{3\times 3} \times \mathbb{R} : \mathbf{F} \in \mathbb{M}\}$ naturally a convex set (see Theorem 4.7 - 4 in [Cia88]).
The following theorem gives necessary and sufficient conditions for the convexity of a sufficient smooth function on a convex set.

Proposition 2.27: (Conditions for convexity)
Let V be a normed space, $K \subset V$ a nonempty convex subset and $f : K \to \mathbb{R}$ a twice Gâteaux differentiable function. Then the following conditions are equivalent:

(a) f is convex on K, i.e. $f(\lambda v + (1-\lambda)u) \leq \lambda f(v) + (1-\lambda) f(u)$ for $u, v \in K, \lambda \in [0,1]$

(b) $f'(u)[v-u] + f(u) \leq f(v) \quad \forall\, u, v \in K$

(c) $f'(v)[v-u] - f'(u)[v-u] \geq 0 \quad \forall\, u, v \in K$

(d) $f''(u)[v-u, v-u] \geq 0 \quad \forall\, u, v \in K$

Proof:

A proof of the equivalences $(a) \Leftrightarrow (b) \Leftrightarrow (c)$ can be found in [AH09] (see proof of Theorem 5.3.17). It remains to show the equivalence of (c) and (d).
$(c) \Rightarrow (d)$:
Let $u, v \in K$. Then by assumption it holds $f'(v)[v-u] - f'(u)[v-u] \geq 0$. Since K is

convex we have $u+\lambda(v-u)=\lambda v+(1-\lambda)u \in K$ with $\lambda \in [0,1]$. Inserting this expression into the assumption (c) instead of v results in

$$f'(u+\lambda(v-u))[\lambda(v-u)] - f'(u)[\lambda(v-u)] \geq 0, \quad \lambda \in [0,1].$$

Dividing this term by λ^2 with $\lambda \in (0,1]$ leads to

$$\frac{f'(u+\lambda(v-u))[v-u] - f'(u)[v-u]}{\lambda} \geq 0. \qquad (*)$$

Since the second Gâteaux derivative of f in $u \in K$ is given by

$$f''(u)[w,v] = \lim_{t \to 0} \frac{f'(u+tw)[v] - f'(u)[v]}{t}, \quad v,w \in V,$$

we obtain from $(*)$ for $\lambda \to 0$ the result $f''(u)[v-u, v-u] \geq 0$.
$(d) \Rightarrow (c)$:
Let u, v be again in K. By assumption it holds $f''(u)[v-u, v-u] \geq 0$. We set
$h(\lambda) := f'(u+\lambda(v-u))[v-u]$ and get with the help of the chain rule the derivative

$$h'(\lambda) = f''(u+\lambda(v-u))[v-u, v-u].$$

By the mean value theorem there exists a $\bar{\lambda} \in (0,1)$ with

$$f'(v)[v-u] - f'(u)[v-u] = h(1) - h(0) = h'(\bar{\lambda}) = f''(u+\bar{\lambda}(v-u))[v-u, v-u].$$

If we insert $u+\bar{\lambda}(v-u) = \bar{\lambda}v + (1-\bar{\lambda})u \in K$ and $u \in K$ instead of u and v in assumption (d) it follows

$$\begin{aligned} f''(u+\bar{\lambda}(v-u))[-\bar{\lambda}(v-u), -\bar{\lambda}(v-u)] \geq 0 &\Leftrightarrow \bar{\lambda}^2 f''(u+\bar{\lambda}(v-u))[v-u, v-u] \geq 0 \\ &\Leftrightarrow f''(u+\bar{\lambda}(v-u))[v-u, v-u] \geq 0. \end{aligned}$$

With the result obtained above by the mean value theorem we get

$$f'(v)[v-u] - f'(u)[v-u] = f''(u+\bar{\lambda}(v-u))[v-u, v-u] \geq 0,$$

i.e. condition (c).

□

2.2.7 Plane strain model

For our two - dimensional examples in Section 6 a plane strain model is used. In a plane strain model we use the following assumptions:

- The displacements u_1 and u_2 of $\mathbf{u}$ depend only on x_1 and x_2, i.e. $u_1(x_1, x_2)$, $u_2(x_1, x_2)$.
- The displacement $\mathbf{u}$ is constant in x_3 - direction, i.e. $u_3 = const.$

The force densities $\mathbf{f}$ and $\mathbf{g}$ must be chosen such that they do not contradict these assumptions. We obtain the following consequences:
Due to $\partial_3 u_1 = 0$, $\partial_3 u_2 = 0$ and $\partial_j u_3 = 0$ for $j = 1, 2, 3$ we obtain the deformation gradient $\mathbf{F}(\mathbf{u})$ and the corresponding strain tensor $\mathbf{C}(\mathbf{u}) = (\mathbf{F}(\mathbf{u}))^T \mathbf{F}(\mathbf{u})$ as

$$\mathbf{F}(\mathbf{u}) = \mathbf{I} + \nabla \mathbf{u} = \begin{pmatrix} 1 + \partial_1 u_1 & \partial_2 u_1 & 0 \\ \partial_1 u_2 & 1 + \partial_2 u_2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \Rightarrow \mathbf{C}(\mathbf{u}) = \begin{pmatrix} C_{11}(\mathbf{u}) & C_{12}(\mathbf{u}) & 0 \\ C_{21}(\mathbf{u}) & C_{22}(\mathbf{u}) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Due to the simplified structure of the strain tensor $\mathbf{C}$ in a plane strain model we get in general a simplified corresponding stress tensor according to the given material law, see Section 2.4.2 in the case of a homogeneous isotropic frame - indifferent material.
An introduction and further details to the plane strain model and also to the plane stress model in elasticity theory can be found in [Bra07] and [EGK11].

2.3 Principal invariants of a matrix

In this section we define invariants and especially principal invariants of a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$. Then we consider the principal invariants in the case $n = 3$ which play a major role in three - dimensional elasticity theory. Further we prove some estimates and calculate the Fréchet derivatives and the corresponding gradients to these invariants.

2.3.1 Definition of the principal invariants

The following definitions and explanations are again based on [Cia88].

Definition 2.28: (Invariant of a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$)
An invariant of a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a real - valued function $\omega(\mathbf{A})$ with the property

$$\omega(\mathbf{A}) = w(\mathbf{B}^{-1}\mathbf{A}\mathbf{B})$$

for all invertible matrices $\mathbf{B} \in \mathbb{R}^{n \times n}$.

Definition 2.29: (Principal invariants of a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$)
We define the principal invariants of a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ as the n coefficients $\tau_1(\mathbf{A}), \ldots,$ $\tau_n(\mathbf{A})$ of the characteristic polynomial

$$\det(\mathbf{A} - \lambda \mathbf{I}) = (-1)^n \lambda^n + (-1)^{n-1} \tau_1(\mathbf{A}) \lambda^{n-1} + \ldots - \tau_{n-1}(\mathbf{A}) \lambda + \tau_n(\mathbf{A}).$$

Proposition 2.30: (Principal invariants for $n = 3$)
The principal invariants for an arbitrary matrix $\mathbf{A} \in \mathbb{R}^{3 \times 3}$ are

$$\begin{aligned} I_1(\mathbf{A}) &:= \tau_1(\mathbf{A}) = \operatorname{tr}(\mathbf{A}), \\ I_2(\mathbf{A}) &:= \tau_2(\mathbf{A}) = \operatorname{tr}(\mathbf{Cof}\, \mathbf{A}), \\ I_3(\mathbf{A}) &:= \tau_3(\mathbf{A}) = \det \mathbf{A}. \end{aligned} \tag{2.21}$$

Proof:
For an arbitrary matrix $\mathbf{A} \in \mathbb{R}^{3\times 3}$ with entries A_{ij} it holds

$$\det \mathbf{A} = A_{11}A_{22}A_{33} + A_{12}A_{23}A_{31} + A_{13}A_{21}A_{32} - A_{11}A_{23}A_{32} - A_{12}A_{21}A_{33} - A_{13}A_{22}A_{31}.$$

The representation (2.15) of $\mathbf{Cof}\,\mathbf{A} \in \mathbb{R}^{3\times 3}$ implies

$$\operatorname{tr}(\mathbf{Cof}\,\mathbf{A}) = A_{11}A_{22} + A_{11}A_{33} + A_{22}A_{33} - A_{12}A_{21} - A_{13}A_{31} - A_{23}A_{32}.$$

It follows

$$\begin{aligned}
\det(\mathbf{A} - \lambda\mathbf{I}) &= \begin{vmatrix} A_{11}-\lambda & A_{12} & A_{13} \\ A_{21} & A_{22}-\lambda & A_{23} \\ A_{31} & A_{32} & A_{33}-\lambda \end{vmatrix} \\
&= (A_{11}-\lambda)\begin{vmatrix} A_{22}-\lambda & A_{23} \\ A_{32} & A_{33}-\lambda \end{vmatrix} - A_{12}\begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33}-\lambda \end{vmatrix} + A_{13}\begin{vmatrix} A_{21} & A_{22}-\lambda \\ A_{31} & A_{32} \end{vmatrix} \\
&= (A_{11}-\lambda)(A_{22}-\lambda)(A_{33}-\lambda) - A_{23}A_{32}(A_{11}-\lambda) \\
&\quad - A_{12}A_{21}(A_{33}-\lambda) + A_{12}A_{23}A_{31} + A_{13}A_{21}A_{32} - A_{13}A_{31}(A_{22}-\lambda) \\
&= (A_{11}-\lambda)\left(A_{22}A_{33} - \lambda(A_{22}+A_{33}) + \lambda^2\right) \\
&\quad - \lambda\left[-A_{23}A_{32} - A_{12}A_{21} - A_{13}A_{31}\right] + \det\mathbf{A} - A_{11}A_{22}A_{33} \\
&= -\lambda^3 + \lambda^2(A_{11}+A_{22}+A_{33}) \\
&\quad - \lambda\left[A_{11}A_{22} + A_{11}A_{33} + A_{22}A_{33} - A_{23}A_{32} - A_{12}A_{21} - A_{13}A_{31}\right] + \det\mathbf{A} \\
&= -\lambda^3 + \lambda^2\operatorname{tr}(\mathbf{A}) - \lambda\operatorname{tr}(\mathbf{Cof}\,\mathbf{A}) + \det\mathbf{A}.
\end{aligned}$$

Therefore we obtain the coefficients $\tau_1(\mathbf{A}) = \operatorname{tr}(\mathbf{A})$, $\tau_2(\mathbf{A}) = \operatorname{tr}(\mathbf{Cof}\,\mathbf{A})$ and $\tau_3(\mathbf{A}) = \det\mathbf{A}$ of the characteristic polynomial and obtain the statement by Definition 2.29.

□

It is easy to check that the principal invariants in the case $n = 3$ satisfy the property of Definition 2.28.

2.3.2 Estimates for the principal invariants

In this part we prove important inequalities for the invariants $I_1(\mathbf{A})$, $I_2(\mathbf{A})$, $I_3(\mathbf{A})$. Although they were defined above only in the case $n = 3$ some of the following inequalities hold for arbitrary $n \in \mathbb{N} \setminus \{0\}$.

Lemma 2.31: (Estimate for the trace operator)
For $\mathbf{A} \in \mathbb{R}^{n\times n}$ it holds

$$|\operatorname{tr}(\mathbf{A})| \le \sqrt{n}|\mathbf{A}|.$$

Proof:

For the matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ with matrix entries A_{ij}, $1 \le i,j \le n$, it holds

$$(\mathrm{tr}(\mathbf{A}))^2 = \left(\sum_{i=1}^{n} A_{ii}\right)^2 \le \left(\sum_{i=1}^{n} 1^2\right) \cdot \left(\sum_{i=1}^{n} A_{ii}^2\right) = n\left(\sum_{i=1}^{n} A_{ii}^2\right) \le n\left(\sum_{i,j=1}^{n} A_{ij}^2\right) = n|\mathbf{A}|^2,$$

where we have used the definition of the trace operator, the Cauchy - Schwarz inequality for a sum and the definition of the Frobenius norm. The statement follows by extracting the square root.

□

Lemma 2.32: (Estimate for the cofactor in three dimensions)
For $\mathbf{A} \in \mathbb{R}^{3\times 3}$ it holds

$$|\mathbf{Cof\ A}| \le 6|\mathbf{A}|^2.$$

Proof:

We define the vectors $\mathbf{a}, \mathbf{a}^{\mathrm{cof}} \in \mathbb{R}^9$ which contain the entries of $\mathbf{A}$ and $\mathbf{Cof\ A}$. The Euclidean norm $|\cdot|$ is equivalent to the maximum norm $\|\mathbf{a}\|_\infty := \max\limits_{1\le i\le n} |a_i|$, $\mathbf{a} \in \mathbb{R}^n$, with

$$\|\mathbf{a}\|_\infty \le |\mathbf{a}| \le \sqrt{n}\|\mathbf{a}\|_\infty \quad \forall\, \mathbf{a} \in \mathbb{R}^n.$$

With this choice we get

$$|\mathbf{Cof\ A}|^2 = |\mathbf{a}^{\mathrm{cof}}|^2 \le 9\|\mathbf{a}^{\mathrm{cof}}\|_\infty^2 = 9\left(\max_{1\le i\le 9} |a_i^{\mathrm{cof}}|\right)^2 = 9\max_{1\le i\le 9} |a_i^{\mathrm{cof}}|^2.$$

By Remark 2.22 we know that each entry of $\mathbf{a}^{\mathrm{cof}}$ has the form $a_i^{\mathrm{cof}} = ab - cd$, $1 \le i \le 9$, where $a, b, c, d \in \mathbb{R}$ are matrix entries of $\mathbf{A}$. With the help of Young's inequality we get for each i the estimate

$$\begin{aligned}
|a_i^{\mathrm{cof}}|^2 = (ab-cd)^2 &= a^2b^2 - 2abcd + c^2d^2 \le a^2b^2 + 2|ab||cd| + c^2d^2 \\
&\le \frac{1}{2}(a^4+b^4) + |ab|^2 + |cd|^2 + \frac{1}{2}(c^4+d^4) \\
&\le \frac{1}{2}(a^4+b^4) + \frac{1}{2}(a^4+b^4+c^4+d^4) + \frac{1}{2}(c^4+d^4) = a^4+b^4+c^4+d^4 \\
&\le 4\max\{a^4,b^4,c^4,d^4\} \le 4\left(\max_{1\le i,j\le 3} A_{ij}^4\right) \\
&= 4\left(\max_{1\le i,j\le 3} |A_{ij}|\right)^4 = 4\|\mathbf{a}\|_\infty^4 \le 4|\mathbf{a}|^4 = 4|\mathbf{A}|^4.
\end{aligned}$$

Altogether we get $|\mathbf{Cof\ A}|^2 \le 9\max\limits_{1\le i\le 9} |a_i^{\mathrm{cof}}|^2 \le 9\cdot 4|\mathbf{A}|^4 = 36|\mathbf{A}|^4$. We get the statement by extracting the square root.

□

Corollary 2.33: (Estimate for tr(Cof A))
For $\mathbf{A} \in \mathbb{R}^{3\times 3}$ it holds

$$|\mathrm{tr}(\mathbf{Cof}\,\mathbf{A})| \leq 6\sqrt{3}|\mathbf{A}|^2.$$

Proof:

Combining Lemmata 2.31 and 2.32 with $n = 3$ leads to

$$|\mathrm{tr}(\mathbf{Cof}\,\mathbf{A})|^2 \leq 3|\mathbf{Cof}\,\mathbf{A}|^2 \leq 3 \cdot 36|\mathbf{A}|^4.$$

Extracting the square root results in the statement.

□

Lemma 2.34: (Classical Hadamard inequality)
Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ with $\mathbf{A} = (\mathbf{a}_1,\ \mathbf{a}_2,\ \ldots,\ \mathbf{a}_n)$, i.e. $\mathbf{a}_j \in \mathbb{R}^{n\times 1}$ for all $j = 1, \ldots, n$. Then it holds

$$|\det(\mathbf{A})| \leq \prod_{j=1}^{n} |\mathbf{a}_j|.$$

Proof:

Let $\mathbf{A} = \mathbf{QR}$ be the QR-decomposition to the matrix $\mathbf{A}$ with orthogonal $\mathbf{Q} \in \mathbb{R}^{n\times n}$ and upper triangular matrix $\mathbf{R} = (\mathbf{r}_1,\ \mathbf{r}_2,\ \ldots,\ \mathbf{r}_n)$, i.e. $\mathbf{r}_j \in \mathbb{R}^{n\times 1}$ for all $j = 1, \ldots, n$. Then it holds

$$\begin{aligned}|\det(\mathbf{A})| =& |\det(\mathbf{QR})| = \underbrace{|\det \mathbf{Q}|}_{=1} |\det \mathbf{R}| = |\det \mathbf{R}| = \left|\prod_{j=1}^{n} R_{jj}\right| = \prod_{j=1}^{n} |R_{jj}| \\ &= \prod_{j=1}^{n} \left(R_{jj}^2\right)^{\frac{1}{2}} \leq \prod_{j=1}^{n} \left(\sum_{i=1}^{n} R_{ij}^2\right)^{\frac{1}{2}} = \prod_{j=1}^{n} \left(\sum_{i=1}^{n} (\mathbf{r}_j)_i^2\right)^{\frac{1}{2}} = \prod_{j=1}^{n} |\mathbf{r}_j|,\end{aligned}$$

since $\mathbf{Q}$ is orthogonal, $\mathbf{R}$ is upper triangular and the matrix entries of $\mathbf{R}$ are $R_{ij} = (\mathbf{r}_j)_i$. Furthermore it holds $|\mathbf{r}_j| = |\mathbf{Q}\mathbf{r}_j| = |\mathbf{a}_j|$ for all $j = 1, \ldots, n$, since $\mathbf{Q}$ is orthogonal and $\mathbf{QR} = \mathbf{A}$. Altogether we obtain the statement.

□

Corollary 2.35: (Estimate for the determinant)
For $\mathbf{A} \in \mathbb{R}^{n\times n}$ it holds $|\det \mathbf{A}| \leq |\mathbf{A}|^n$.

Proof:

Let $\mathbf{A} = (\mathbf{a}_1,\ \mathbf{a}_2,\ \ldots,\ \mathbf{a}_n) \in \mathbb{R}^{n\times n}$ with $\mathbf{a}_j \in \mathbb{R}^{n\times 1}$ for all $j = 1, \ldots, n$. For each $\mathbf{a}_j$ with $(\mathbf{a}_j)_i = A_{ij}$, $i, j = 1, \ldots, n$, it holds

$$|\mathbf{a}_j| = \left(\sum_{i=1}^{n} (\mathbf{a}_j)_i^2\right)^{\frac{1}{2}} = \left(\sum_{i=1}^{n} A_{ij}^2\right)^{\frac{1}{2}} \leq \left(\sum_{i,k=1}^{n} A_{ik}^2\right)^{\frac{1}{2}} = |\mathbf{A}|.$$

With the classical Hadamard inequality (Lemma 2.34) it follows

$$|\det \mathbf{A}| \leq \prod_{j=1}^{n} |\mathbf{a}_j| \leq \prod_{j=1}^{n} |\mathbf{A}| = |\mathbf{A}|^n.$$

□

2.3.3 Fréchet derivatives and gradients for the principal invariants

For the computation of the Piola - Kirchhoff stress tensor $\mathbf{P}$ (respectively $\mathbf{\Sigma}$) for a homogeneous isotropic frame - indifferent hyperelastic material we need the gradients of the three principal invariants $I_1(\mathbf{C}), I_2(\mathbf{C}), I_3(\mathbf{C})$, $\mathbf{C} = \mathbf{F}^T\mathbf{F}$, with respect to $\mathbf{F}$ (respectively $\mathbf{C}$), see Section 2.4.1. For this purpose we derive the Fréchet derivatives and the corresponding gradients for these invariants. We are able to determine the derivatives of I_j, $j = 1, 2, 3$, for arbitrary $n \in \mathbb{N} \setminus \{0\}$, i.e. in this section $I_1(\mathbf{A}) := \operatorname{tr}(\mathbf{A})$, $I_2(\mathbf{A}) := \operatorname{tr}(\mathbf{Cof}\,\mathbf{A})$ and $I_3(\mathbf{A}) := \det \mathbf{A}$ are defined for matrices $\mathbf{A} \in \mathbb{R}^{n\times n}$.

Proposition 2.36: (Fréchet derivative of $I_1 : \mathbb{R}^{n\times n} \to \mathbb{R}, \mathbf{A} \mapsto \mathbf{tr}(\mathbf{A})$)
The Fréchet derivative of the mapping $I_1(\mathbf{A}) = \operatorname{tr}(\mathbf{A}), \mathbf{A} \in \mathbb{R}^{n\times n}$, is given by

$$\partial I_1(\mathbf{A})(\mathbf{H}) = \operatorname{tr}(\mathbf{H}) = \mathbf{I} : \mathbf{H}.$$

Proof:

It holds $I_1(\mathbf{A} + \mathbf{H}) = \operatorname{tr}(\mathbf{A} + \mathbf{H}) = \operatorname{tr}(\mathbf{A}) + \operatorname{tr}(\mathbf{H}) = I_1(\mathbf{A}) + \operatorname{tr}(\mathbf{H})$ and therefore the statement by Definition 2.1.

□

For the derivation of the Fréchet derivatives of $I_2(\mathbf{A})$ and $I_3(\mathbf{A})$ we need some crucial lemmata.

Lemma 2.37: (Linearization of the determinant about the identity)
Let $\mathbf{E} \in \mathbb{R}^{n\times n}$ with a positive integer n be given. Then it holds

$$\det(\mathbf{I} + \mathbf{E}) = 1 + \operatorname{tr}(\mathbf{E}) + o(|\mathbf{E}|), \quad \mathbf{E} \to \mathbf{0}.$$

Proof: (mathematical induction)
Base case: For $n = 1$ it holds $\det(\mathbf{I}+\mathbf{E}) = 1+e_{11} = 1+\operatorname{tr}(\mathbf{E})$ and therefore the statement.
Inductive hypothesis: The statement holds for $n \in \mathbb{N} \setminus \{0\}$.
Inductive step: $n \mapsto n + 1$
In the following we write $\mathbf{I}^{(n)}, \mathbf{E}^{(n)}$ for $\mathbf{I}$, $\mathbf{E} \in \mathbb{R}^{n\times n}$ to distinguish between $\mathbf{I}^{(n)}, \mathbf{E}^{(n)}$ and $\mathbf{I}^{(n+1)}, \mathbf{E}^{(n+1)}$. With this notation, using the formula of Laplace for the expansion of the

determinant with respect to the column $n+1$, it holds

$$\begin{aligned}\det\left(\mathbf{I}^{(n+1)}+\mathbf{E}^{(n+1)}\right) &= \sum_{i=1}^{n+1}(-1)^{n+1+i}\left(\mathbf{I}^{(n+1)}+\mathbf{E}^{(n+1)}\right)_{i,n+1}\det\left(\mathbf{I}^{(n+1)}+\mathbf{E}^{(n+1)}\right)'_{i,n+1}\\ &= \sum_{i=1}^{n}(-1)^{n+1+i}\left(\mathbf{E}^{(n+1)}\right)_{i,n+1}\det\left(\mathbf{I}^{(n+1)}+\mathbf{E}^{(n+1)}\right)'_{i,n+1}\\ &\quad+\left(1+\mathbf{E}^{(n+1)}_{n+1,n+1}\right)\det\left(\mathbf{I}^{(n)}+\mathbf{E}^{(n)}\right),\end{aligned} \tag{$*$}$$

where $\left(\mathbf{I}^{(n+1)}+\mathbf{E}^{(n+1)}\right)'_{i,n+1}\in\mathbb{R}^{n\times n}$ denotes the matrix obtained by deleting the i-th row and the $(n+1)$-st column in the matrix $\mathbf{I}^{(n+1)}+\mathbf{E}^{(n+1)}\in\mathbb{R}^{(n+1)\times(n+1)}$.
It can be seen in the following way that the first term is $o\left(|\mathbf{E}^{(n+1)}|\right)$:
For positive integers $n\geq 2$ and real numbers $a_1,\dots,a_n\geq 0$ holds generally

$$\begin{aligned}\prod_{k=1}^{n}a_k &\leq \sum_{k=1}^{n-1}\left(\frac{1}{2}\right)^k a_k^{2^k}+\left(\frac{1}{2}\right)^{n-1}a_n^{2^{n-1}}\leq\sum_{k=1}^{n-1}a_k^{2^k}+a_n^{2^{n-1}}\\ &\leq a_1^2+(\dots+(\dots+(a_{n-3}^2+(a_{n-2}^2+(a_{n-1}^2+a_n^2)^2)^2)^2)^2)^2.\end{aligned}$$

The validity of these inequalities can be proven separately with mathematical induction.
The expression $\left(\mathbf{E}^{(n+1)}\right)_{i,n+1}\det\left(\mathbf{I}^{(n+1)}+\mathbf{E}^{(n+1)}\right)'_{i,n+1}$ consists for each $i\in\{1,\dots,n\}$ of a sum where each summand is a product with at least 2 and at most $n+1$ matrix entries e_k of $\mathbf{E}^{(n+1)}$ as factors. One obtains for each summand

$$\begin{aligned}\left|\prod_{k=1}^{N}e_k\right| &\leq e_1^2+(\dots+(\dots+(e_{N-3}^2+(e_{N-2}^2+(e_{N-1}^2+e_N^2)^2)^2)^2)^2)^2\\ &\leq |\mathbf{E}^{(n+1)}|^2+(\dots+(\dots+(|\mathbf{E}^{(n+1)}|^2+(|\mathbf{E}^{(n+1)}|^2+(|\mathbf{E}^{(n+1)}|^2)^2)^2)^2)^2)^2\\ &\lesssim |\mathbf{E}^{(n+1)}|^2\end{aligned}$$

with $N\in\{2,\dots,n+1\}$ and $\mathbf{E}^{(n+1)}\to\mathbf{0}$.
Thus each summand of $\left(\mathbf{E}^{(n+1)}\right)_{i,n+1}\det\left(\mathbf{I}^{(n+1)}+\mathbf{E}^{(n+1)}\right)'_{i,n+1}$ is $o\left(|\mathbf{E}^{(n+1)}|\right)$ and thus the whole first term in $(*)$ is $o\left(|\mathbf{E}^{(n+1)}|\right)$.

To prove the statement finally we set

$$\mathbf{E}^{(n+1)}=\begin{pmatrix}\mathbf{E}^{(n)}_{1,1} & \cdots & \mathbf{E}^{(n)}_{1,n} & \mathbf{E}^{(n+1)}_{1,n+1}\\ \vdots & \ddots & \vdots & \vdots\\ \mathbf{E}^{(n)}_{n,1} & \cdots & \mathbf{E}^{(n)}_{n,n} & \mathbf{E}^{(n+1)}_{n,n+1}\\ \mathbf{E}^{(n+1)}_{n+1,1} & \cdots & \mathbf{E}^{(n+1)}_{n+1,n} & \mathbf{E}^{(n+1)}_{n+1,n+1}\end{pmatrix}$$

as extension of the matrix $\mathbf{E}^{(n)}\in\mathbb{R}^{n\times n}$ and get straight forward:

- $\left|\mathbf{E}^{(n+1)}_{n+1,n+1}o\left(|\mathbf{E}^{(n)}|\right)\right|\lesssim|\mathbf{E}^{(n+1)}|^2$ and in particular $\mathbf{E}^{(n+1)}_{n+1,n+1}o\left(|\mathbf{E}^{(n)}|\right)=o\left(|\mathbf{E}^{(n+1)}|\right)$
- $\left|\mathbf{E}^{(n+1)}_{n+1,n+1}\mathrm{tr}\left(\mathbf{E}^{(n)}\right)\right|\lesssim|\mathbf{E}^{(n+1)}|^2$ and in particular $\mathbf{E}^{(n+1)}_{n+1,n+1}\mathrm{tr}\left(\mathbf{E}^{(n)}\right)=o\left(|\mathbf{E}^{(n+1)}|\right)$

Additionally for functions $f : \mathbb{R}^{n\times n} \to \mathbb{R}$ and $g : \mathbb{R}^{(n+1)\times(n+1)} \to \mathbb{R}$ with $g\left(\mathbf{E}^{(n+1)}\right) := f\left(\mathbf{E}^{(n)}\right)$ and $f\left(\mathbf{E}^{(n)}\right) = o\left(\left|\mathbf{E}^{(n)}\right|\right)$ we get in a neighborhood of the zero matrix

$$\left|g\left(\mathbf{E}^{(n+1)}\right)\right| = \left|f\left(\mathbf{E}^{(n)}\right)\right| \leq \varepsilon \left|\mathbf{E}^{(n)}\right| \leq \varepsilon \left|\mathbf{E}^{(n+1)}\right| \quad \forall \varepsilon > 0.$$

Thus it holds $f\left(\mathbf{E}^{(n)}\right) = o\left(\left|\mathbf{E}^{(n)}\right|\right) \Rightarrow g\left(\mathbf{E}^{(n+1)}\right) = o\left(\left|\mathbf{E}^{(n+1)}\right|\right)$. With these estimates and the help of the inductive hypothesis it follows for the second term in $(*)$

$$\begin{aligned}\left(1 + \mathbf{E}^{(n+1)}_{n+1,n+1}\right) \det\left(\mathbf{I}^{(n)} + \mathbf{E}^{(n)}\right) &= \left(1 + \mathbf{E}^{(n+1)}_{n+1,n+1}\right)\left(1 + \operatorname{tr}\left(\mathbf{E}^{(n)}\right) + o\left(\left|\mathbf{E}^{(n)}\right|\right)\right) \\ &= 1 + \operatorname{tr}\left(\mathbf{E}^{(n)}\right) + \mathbf{E}^{(n+1)}_{n+1,n+1} + o\left(\left|\mathbf{E}^{(n+1)}\right|\right) \\ &= 1 + \operatorname{tr}\left(\mathbf{E}^{(n+1)}\right) + o\left(\left|\mathbf{E}^{(n+1)}\right|\right).\end{aligned}$$

Therefore we obtain by $(*)$ altogether $\det\left(\mathbf{I}^{(n+1)} + \mathbf{E}^{(n+1)}\right) = 1 + \operatorname{tr}\left(\mathbf{E}^{(n+1)}\right) + o\left(\left|\mathbf{E}^{(n+1)}\right|\right)$.

□

Lemma 2.38: (Fréchet derivative of the inverse of a given matrix)
The Fréchet derivative of the mapping $\mathbf{A} \mapsto \mathbf{A}^{-1}, \mathbf{A} \in \mathbb{R}^{n\times n}$, is given by

$$\partial \mathbf{A}^{-1}(\mathbf{H}) = -\mathbf{A}^{-1}\mathbf{H}\mathbf{A}^{-1}.$$

Proof:

For a matrix $\mathbf{E} \in \mathbb{R}^{n\times n}$ with $|\mathbf{E}| < \frac{1}{\sqrt{n}}$ it holds $|\operatorname{tr}(\mathbf{E})| \leq \sqrt{n}\,|\mathbf{E}| < 1$ (see Lemma 2.31). Therefore for sufficiently small $|\mathbf{E}| < 1$, using Lemma 2.37, one obtains

$$0 < 1 - |\operatorname{tr}(\mathbf{E})| = |1 - |\operatorname{tr}(\mathbf{E})|| \leq |1 + \operatorname{tr}(\mathbf{E})| \approx |\det(\mathbf{I} + \mathbf{E})|,$$

i.e. $\mathbf{I} + \mathbf{E}$ is invertible. For $\mathbf{E} = \mathbf{A}^{-1}\mathbf{H}$ the requirement $|\mathbf{E}| < \frac{1}{\sqrt{n}}$ is satisfied if $|\mathbf{H}| < \frac{1}{\sqrt{n}}|\mathbf{A}^{-1}|^{-1}$. Under this assumption it holds

$$\left(\mathbf{I} + \mathbf{A}^{-1}\mathbf{H}\right)\left(\mathbf{I} - \mathbf{A}^{-1}\mathbf{H}\right) = \mathbf{I} + o(|\mathbf{H}|),$$

since $|(\mathbf{A}^{-1}\mathbf{H})^2| \leq |\mathbf{A}^{-1}|^2|\mathbf{H}|^2 \to 0$ for $\mathbf{H} \to \mathbf{0}$ due to the submultiplicativity of the Frobenius norm. Multiplying this equation with $(\mathbf{I} + \mathbf{A}^{-1}\mathbf{H})^{-1}$ from left, we get

$$\left(\mathbf{I} + \mathbf{A}^{-1}\mathbf{H}\right)^{-1} = \mathbf{I} - \mathbf{A}^{-1}\mathbf{H} - \left(\mathbf{I} + \mathbf{A}^{-1}\mathbf{H}\right)^{-1} o(|\mathbf{H}|).$$

Inserting this equation recursively in itself we get

$$\left(\mathbf{I} + \mathbf{A}^{-1}\mathbf{H}\right)^{-1} = \mathbf{I} - \mathbf{A}^{-1}\mathbf{H} + o(|\mathbf{H}|),$$

since $\mathbf{A}^{-1}\mathbf{H}\, o(|\mathbf{H}|) = o(|\mathbf{H}|)$ and $o(|\mathbf{H}|) \cdot o(|\mathbf{H}|) = o(|\mathbf{H}|^2)$.
With the help of this relation it follows

$$\begin{aligned}(\mathbf{A} + \mathbf{H})^{-1} &= \left(\mathbf{A}(\mathbf{I} + \mathbf{A}^{-1}\mathbf{H})\right)^{-1} = (\mathbf{I} + \mathbf{A}^{-1}\mathbf{H})^{-1}\mathbf{A}^{-1} \\ &= \left(\mathbf{I} - \mathbf{A}^{-1}\mathbf{H} + o(|\mathbf{H}|)\right)\mathbf{A}^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{H}\mathbf{A}^{-1} + o(|\mathbf{H}|)\mathbf{A}^{-1} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{H}\mathbf{A}^{-1} + o(|\mathbf{H}|)\end{aligned}$$

for $\mathbf{H} \to \mathbf{0}$, i.e. the statement by Definition 2.1.

□

Proposition 2.39: (Fréchet derivative of $I_3 : \mathbb{R}^{n\times n} \to \mathbb{R}, \mathbf{A} \mapsto \det(\mathbf{A})$)
For an invertible arbitrary matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ the Fréchet derivative of $I_3(\mathbf{A}) = \det(\mathbf{A})$ is given by

$$\partial I_3(\mathbf{A})(\mathbf{H}) = \mathbf{Cof}\,\mathbf{A} : \mathbf{H}.$$

Proof:
With the help of Lemma 2.37 it holds

$$\begin{aligned}\det(\mathbf{A}+\mathbf{H}) &= \det\left(\mathbf{A}(\mathbf{I}+\mathbf{A}^{-1}\mathbf{H})\right) = \det(\mathbf{A})\det(\mathbf{I}+\mathbf{A}^{-1}\mathbf{H}) \\ &= \det(\mathbf{A})\left(1+\operatorname{tr}(\mathbf{A}^{-1}\mathbf{H}) + o(|\mathbf{A}^{-1}\mathbf{H}|)\right) \\ &= \det(\mathbf{A}) + \det(\mathbf{A})\mathbf{A}^{-T} : \mathbf{H} + \det(\mathbf{A})o(|\mathbf{A}^{-1}\mathbf{H}|) \\ &= \det(\mathbf{A}) + \mathbf{Cof}\,\mathbf{A} : \mathbf{H} + o(|\mathbf{H}|)\end{aligned}$$

for an arbitrary invertible matrix $\mathbf{A}$ and $\mathbf{H} \to \mathbf{0}$, i.e. the statement.

□

Lemma 2.40: (Fréchet derivative of the cofactor of an invertible matrix)
Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be an invertible arbitrary matrix. Then the Fréchet derivative of the mapping $\mathbf{A} \mapsto \mathbf{Cof}\,\mathbf{A} := \det(\mathbf{A})\mathbf{A}^{-T}$ is given by

$$\partial(\mathbf{Cof}\,\mathbf{A})(\mathbf{H}) = (\mathbf{Cof}\,\mathbf{A} : \mathbf{H})\mathbf{A}^{-T} - \det(\mathbf{A})\mathbf{A}^{-T}\mathbf{H}^T\mathbf{A}^{-T}.$$

Proof:

For the calculation of this derivative we use the product rule of Proposition 2.9 with $V_1 = \mathbb{R}$, $V = V_2 = \mathbb{R}^{n\times n}$, $f(\mathbf{A}) = \det(\mathbf{A})$, $g(\mathbf{A}) = \mathbf{A}^{-T}$ and the bounded bilinear form $b(a, \mathbf{A}) := a\mathbf{A}$, $a \in \mathbb{R}$, $\mathbf{A} \in \mathbb{R}^{n\times n}$. The derivative of f is $\partial f(\mathbf{A})(\mathbf{H}) = \mathbf{Cof}\,\mathbf{A} : \mathbf{H}$ (see Proposition 2.39). Since $(\mathbf{A}+\mathbf{H})^T = \mathbf{A}^T + \mathbf{H}^T$ the Fréchet derivative of the mapping $\mathbf{A} \mapsto \mathbf{A}^T$ is

$$\partial\mathbf{A}^T(\mathbf{H}) = \mathbf{H}^T. \tag{2.22}$$

With the help of the chain rule in Proposition 2.10 and Lemma 2.38 we get $\partial g(\mathbf{A})(\mathbf{H}) = -\mathbf{A}^{-T}\mathbf{H}^T\mathbf{A}^{-T}$. With the definitions above it holds $\mathbf{Cof}\,(\mathbf{A}) = b(f(\mathbf{A}), g(\mathbf{A}))$ and therefore

$$\begin{aligned}\partial(\mathbf{Cof}\,\mathbf{A})(\mathbf{H}) &= b(\partial f(\mathbf{A})(\mathbf{H}), g(\mathbf{A})) + b(f(\mathbf{A}), \partial g(\mathbf{A})(\mathbf{H})) \\ &= \partial f(\mathbf{A})(\mathbf{H})g(\mathbf{A}) + f(\mathbf{A})\partial g(\mathbf{A})(\mathbf{H}) \\ &= (\mathbf{Cof}\,\mathbf{A} : \mathbf{H})\mathbf{A}^{-T} - \det(\mathbf{A})\mathbf{A}^{-T}\mathbf{H}^T\mathbf{A}^{-T}.\end{aligned}$$

□

Proposition 2.41: (Fréchet derivative of $I_2 : \mathbb{R}^{n\times n} \to \mathbb{R}, \mathbf{A} \mapsto \mathbf{tr}(\mathbf{Cof}\,(\mathbf{A}))$)
Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be an invertible arbitrary matrix. Then the Fréchet derivative of the mapping $I_2(\mathbf{A}) = \mathrm{tr}(\mathbf{Cof}\,(\mathbf{A}))$ is given by

$$\partial I_2(\mathbf{A})(\mathbf{H}) = \left(\mathrm{tr}(\mathbf{A}^{-1})\mathbf{I} - \mathbf{A}^{-T}\right) \mathbf{Cof}\,\mathbf{A} : \mathbf{H}.$$

Proof:
Combining Proposition 2.36 and Lemma 2.40 and using the chain rule in Proposition 2.10 result in

$$\begin{aligned}
\partial I_2(\mathbf{A})(\mathbf{H}) &= \mathrm{tr}(\partial(\mathbf{Cof}\,\mathbf{A})(\mathbf{H})) = \mathrm{tr}\left((\mathbf{Cof}\,\mathbf{A} : \mathbf{H})\mathbf{A}^{-T} - \det(\mathbf{A})\mathbf{A}^{-T}\mathbf{H}^T\mathbf{A}^{-T}\right) \\
&= (\mathbf{Cof}\,\mathbf{A} : \mathbf{H})\mathrm{tr}(\mathbf{A}^{-T}) - \det(\mathbf{A})\mathrm{tr}\left(\mathbf{A}^{-T}\mathbf{H}^T\mathbf{A}^{-T}\right) \\
&= \left[\mathrm{tr}(\mathbf{A}^{-1})\mathbf{Cof}\,\mathbf{A}\right] : \mathbf{H} - \det(\mathbf{A})\mathrm{tr}\left(\mathbf{A}^{-1}\mathbf{H}\mathbf{A}^{-1}\right) \\
&= \left[\mathrm{tr}(\mathbf{A}^{-1})\mathbf{Cof}\,\mathbf{A}\right] : \mathbf{H} - \mathrm{tr}\left((\mathbf{Cof}\,\mathbf{A})^T\mathbf{A}^{-1}\mathbf{H}\right) \\
&= \left[\mathrm{tr}(\mathbf{A}^{-1})\mathbf{Cof}\,\mathbf{A}\right] : \mathbf{H} - \left[\mathbf{A}^{-T}\mathbf{Cof}\,\mathbf{A}\right] : \mathbf{H} \\
&= \left[\mathrm{tr}(\mathbf{A}^{-1})\mathbf{Cof}\,\mathbf{A} - \mathbf{A}^{-T}\mathbf{Cof}\,\mathbf{A}\right] : \mathbf{H} \\
&= \left[\left(\mathrm{tr}(\mathbf{A}^{-1})\mathbf{I} - \mathbf{A}^{-T}\right)\mathbf{Cof}\,\mathbf{A}\right] : \mathbf{H}.
\end{aligned}$$

□

With these considerations it is easy to obtain the derivatives and gradients of $I_j(\mathbf{C})$ for $j = 1, 2, 3$, $\mathbf{C} = \mathbf{F}^T\mathbf{F}$, with respect to the matrix $\mathbf{F}$. For this purpose we define $\hat{I}_j(\mathbf{F}) := I_j(\mathbf{C}) = I_j(\mathbf{F}^T\mathbf{F})$ and need again Lemma 2.24. We obtain with the help of the chain rule and the relation (2.4)

$$\partial_{\mathbf{F}}\hat{I}_j(\mathbf{F}) : \mathbf{H} = \partial\hat{I}_j(\mathbf{F})(\mathbf{H}) = \partial I_j(\mathbf{C})(\mathbf{H}^T\mathbf{F} + \mathbf{F}^T\mathbf{H}), \quad j = 1, 2, 3.$$

Here we recall that the mapping $\mathbf{F} \mapsto \mathbf{F}^T\mathbf{F}$ is Fréchet differentiable with derivative $\mathbf{H}^T\mathbf{F} + \mathbf{F}^T\mathbf{H}$. To achieve $\partial_{\mathbf{F}}\hat{I}_j(\mathbf{F})$ we use Propositions 2.36, 2.41, 2.39 and Lemma 2.24. It results

$$\begin{aligned}
\partial_{\mathbf{F}}\hat{I}_1(\mathbf{F}) : \mathbf{H} &= \partial\hat{I}_1(\mathbf{F})(\mathbf{H}) = \mathbf{I} : \left[\mathbf{H}^T\mathbf{F} + \mathbf{F}^T\mathbf{H}\right] = 2\mathbf{F} : \mathbf{H}, \\
\partial_{\mathbf{F}}\hat{I}_2(\mathbf{F}) : \mathbf{H} &= \partial\hat{I}_2(\mathbf{F})(\mathbf{H}) = \mathrm{tr}\left((\mathbf{Cof}\,\mathbf{C})^T\left(\mathrm{tr}(\mathbf{C}^{-1})\mathbf{I} - \mathbf{C}^{-T}\right)^T(\mathbf{H}^T\mathbf{F} + \mathbf{F}^T\mathbf{H})\right) \\
&= \mathrm{tr}\left(2(\mathbf{Cof}\,\mathbf{C})^T\left(\mathrm{tr}(\mathbf{C}^{-1})\mathbf{I} - \mathbf{C}^{-T}\right)^T\mathbf{F}^T\mathbf{H}\right) \\
&= \left[2\mathbf{F}\left(\mathrm{tr}(\mathbf{C}^{-1})\mathbf{I} - \mathbf{C}^{-T}\right)\mathbf{Cof}\,\mathbf{C}\right] : \mathbf{H}, \\
\partial_{\mathbf{F}}\hat{I}_3(\mathbf{F}) : \mathbf{H} &= \partial\hat{I}_3(\mathbf{F})(\mathbf{H}) = \mathbf{Cof}\,\mathbf{C} : \left[\mathbf{H}^T\mathbf{F} + \mathbf{F}^T\mathbf{H}\right] = \mathrm{tr}\left((\mathbf{Cof}\,\mathbf{C})^T(\mathbf{H}^T\mathbf{F} + \mathbf{F}^T\mathbf{H})\right) \\
&= \mathrm{tr}\left(2(\mathbf{Cof}\,\mathbf{C})^T\mathbf{F}^T\mathbf{H}\right) = [2\mathbf{F}\,\mathbf{Cof}\,\mathbf{C}] : \mathbf{H} = \left[2\mathbf{F}(\det\mathbf{F})^2(\mathbf{F}^T\mathbf{F})^{-T}\right] : \mathbf{H} \\
&= \left[2(\det\mathbf{F})^2\mathbf{F}^{-T}\right] : \mathbf{H}
\end{aligned}$$

under the assumption that $\mathbf{C}$ is invertible.
Therefore we obtain the gradients of the principal invariants with respect to $\mathbf{F}$ as

$$\boxed{\begin{aligned}&\partial_{\mathbf{F}}\hat{I}_1(\mathbf{F}) = \partial_{\mathbf{F}} I_1(\mathbf{C}) = 2\mathbf{F},\\ &\partial_{\mathbf{F}}\hat{I}_2(\mathbf{F}) = \partial_{\mathbf{F}} I_2(\mathbf{C}) = 2\mathbf{F}\left(\operatorname{tr}(\mathbf{C}^{-1})\mathbf{I} - \mathbf{C}^{-T}\right)\mathbf{Cof}\,\mathbf{C},\\ &\partial_{\mathbf{F}}\hat{I}_3(\mathbf{F}) = \partial_{\mathbf{F}} I_3(\mathbf{C}) = 2(\det\mathbf{F})^2\mathbf{F}^{-T}\end{aligned}} \tag{2.23}$$

with $\mathbf{C} = \mathbf{F}^T\mathbf{F}$.

2.4 Homogeneous isotropic materials

2.4.1 General representation formulas

By Theorem 31.1 in [Sim98] the stored energy function $\psi(\mathbf{C}) = \hat{\psi}(\mathbf{F})$, $\mathbf{C} = \mathbf{F}^T\mathbf{F}$, for a homogeneous isotropic frame - indifferent hyperelastic material can be expressed through a function $\tilde{\psi}$, depending on the three principal invariants I_1, I_2, I_3, i.e. it holds

$$\psi(\mathbf{C}) = \tilde{\psi}(I_1(\mathbf{C}), I_2(\mathbf{C}), I_3(\mathbf{C})). \tag{2.24}$$

Recall that the existence of such a function ψ to the given stored energy function $\hat{\psi}$ is guaranteed due to the frame - indifference property (see (2.19)).
We will show that for such a material it is possible to write the second Piola - Kirchhoff stress tensor $\boldsymbol{\Sigma}$ as an expression in $\mathbf{C}$ and the Kirchhoff stress tensor $\boldsymbol{\tau}$ as an expression in $\mathbf{B}$.
For this aim we firstly simplify the expression for the gradient of $I_2(\mathbf{C})$ in the case $n = 3$ and state some consequences.

Lemma 2.42:
For an arbitrary matrix $\mathbf{A} \in \mathbb{R}^{3\times3}$ it holds

$$\operatorname{tr}(\mathbf{A})\mathbf{A} - \mathbf{A}^2 = \operatorname{tr}(\mathbf{Cof}\,\mathbf{A})\mathbf{I} - (\mathbf{Cof}\,\mathbf{A})^T.$$

Proof:
For $\mathbf{A} \in \mathbb{R}^{3\times3}$ with matrix entries A_{ij} it obviously holds

$$\operatorname{tr}(\mathbf{A})\mathbf{I} - \mathbf{A} = \begin{pmatrix} A_{22}+A_{33} & -A_{12} & -A_{13}\\ -A_{21} & A_{11}+A_{33} & -A_{23}\\ -A_{31} & -A_{32} & A_{11}+A_{22}\end{pmatrix}. \tag{2.25}$$

Using the representation (2.15) for $\mathbf{Cof}\,\mathbf{A} \in \mathbb{R}^{3\times3}$ and the transpose of equation (2.25) for $\mathbf{Cof}\,\mathbf{A}$ instead of $\mathbf{A}$ leads to

$$\begin{aligned}\operatorname{tr}(\mathbf{Cof}\,\mathbf{A})\mathbf{I} - (\mathbf{Cof}\,\mathbf{A})^T &= \begin{pmatrix} A_{22}+A_{33} & -A_{12} & -A_{13}\\ -A_{21} & A_{11}+A_{33} & -A_{23}\\ -A_{31} & -A_{32} & A_{11}+A_{22}\end{pmatrix}\cdot\begin{pmatrix} A_{11} & A_{12} & A_{13}\\ A_{21} & A_{22} & A_{23}\\ A_{31} & A_{32} & A_{33}\end{pmatrix}\\ &= (\operatorname{tr}(\mathbf{A})\mathbf{I} - \mathbf{A})\,\mathbf{A} = \operatorname{tr}(\mathbf{A})\mathbf{A} - \mathbf{A}^2.\end{aligned}$$

□

Note that this result does not hold for arbitrary $n \in \mathbb{N} \setminus \{0\}$. An immediate consequence of this lemma is

$$\begin{aligned}\operatorname{tr}(\mathbf{A})\mathbf{I} - \mathbf{A} &= \operatorname{tr}(\mathbf{Cof}\,\mathbf{A})\mathbf{A}^{-1} - (\mathbf{Cof}\,\mathbf{A})^T\mathbf{A}^{-1} = \operatorname{tr}(\mathbf{A}^{-1})(\mathbf{Cof}\,\mathbf{A})^T - \mathbf{A}^{-1}(\mathbf{Cof}\,\mathbf{A})^T \\ &= \left(\operatorname{tr}(\mathbf{A}^{-1})\mathbf{I} - \mathbf{A}^{-1}\right)(\mathbf{Cof}\,\mathbf{A})^T\end{aligned}$$

for an invertible matrix $\mathbf{A} \in \mathbb{R}^{3\times 3}$.

If $\mathbf{A}$ is additionally symmetric it follows

$$\left(\operatorname{tr}(\mathbf{A}^{-1})\mathbf{I} - \mathbf{A}^{-1}\right)\mathbf{Cof}\,\mathbf{A} = \operatorname{tr}(\mathbf{A})\mathbf{I} - \mathbf{A}. \tag{2.26}$$

Using Propositions 2.36, 2.41 (combined with equation (2.26) for the symmetric strain tensor $\mathbf{C}$) and Proposition 2.39 in the three-dimensional case we obtain the gradients

$$\frac{\partial I_1(\mathbf{C})}{\partial \mathbf{C}} = \mathbf{I}, \quad \frac{\partial I_2(\mathbf{C})}{\partial \mathbf{C}} = \left(\operatorname{tr}(\mathbf{C}^{-1})\mathbf{I} - \mathbf{C}^{-T}\right)\mathbf{Cof}\,\mathbf{C} = \operatorname{tr}(\mathbf{C})\mathbf{I} - \mathbf{C}, \quad \frac{\partial I_3(\mathbf{C})}{\partial \mathbf{C}} = \mathbf{Cof}\,\mathbf{C}$$

of $I_1(\mathbf{C}) = \operatorname{tr}(\mathbf{C})$, $I_2(\mathbf{C}) = \operatorname{tr}(\mathbf{Cof}\,\mathbf{C})$ and $I_3(\mathbf{C}) = \det(\mathbf{C})$ with respect to $\mathbf{C}$.

If we apply the chain rule on (2.24) for a Fréchet differentiable function $\tilde{\psi}$, we get the gradient of ψ with respect to $\mathbf{C}$ as

$$\begin{aligned}\frac{\partial \psi(\mathbf{C})}{\partial \mathbf{C}} &= \frac{\partial \tilde{\psi}}{\partial I_1}\frac{\partial I_1(\mathbf{C})}{\partial \mathbf{C}} + \frac{\partial \tilde{\psi}}{\partial I_2}\frac{\partial I_2(\mathbf{C})}{\partial \mathbf{C}} + \frac{\partial \tilde{\psi}}{\partial I_3}\frac{\partial I_3(\mathbf{C})}{\partial \mathbf{C}} \\ &= \frac{\partial \tilde{\psi}}{\partial I_1}\mathbf{I} + \frac{\partial \tilde{\psi}}{\partial I_2}\left(\operatorname{tr}(\mathbf{C})\mathbf{I} - \mathbf{C}\right) + \frac{\partial \tilde{\psi}}{\partial I_3}\mathbf{Cof}\,\mathbf{C} \\ &= \left(\frac{\partial \tilde{\psi}}{\partial I_1} + \frac{\partial \tilde{\psi}}{\partial I_2}I_1(\mathbf{C})\right)\mathbf{I} - \frac{\partial \tilde{\psi}}{\partial I_2}\mathbf{C} + \frac{\partial \tilde{\psi}}{\partial I_3}I_3(\mathbf{C})\mathbf{C}^{-1}.\end{aligned}$$

Therefore we get by equation (2.5) and Lemma 2.25

$$\mathbf{F}^{-1}\mathbf{P} = \mathbf{\Sigma} = 2\frac{\partial \psi(\mathbf{C})}{\partial \mathbf{C}} = 2\left(\frac{\partial \tilde{\psi}}{\partial I_1} + \frac{\partial \tilde{\psi}}{\partial I_2}I_1(\mathbf{C})\right)\mathbf{I} - 2\frac{\partial \tilde{\psi}}{\partial I_2}\mathbf{C} + 2\frac{\partial \tilde{\psi}}{\partial I_3}I_3(\mathbf{C})\mathbf{C}^{-1}. \tag{2.27}$$

Multiplying this equation from left with $\mathbf{F}$ and then from right by $\mathbf{F}^T$ results in

$$\boldsymbol{\tau} = \mathbf{P}\mathbf{F}^T = 2\frac{\partial \tilde{\psi}}{\partial I_3}I_3(\mathbf{B})\mathbf{I} + 2\left(\frac{\partial \tilde{\psi}}{\partial I_1} + \frac{\partial \tilde{\psi}}{\partial I_2}I_1(\mathbf{B})\right)\mathbf{B} - 2\frac{\partial \tilde{\psi}}{\partial I_2}\mathbf{B}^2, \tag{2.28}$$

since $\mathbf{F}\mathbf{C}\mathbf{F}^T = \mathbf{F}\mathbf{F}^T\mathbf{F}\mathbf{F}^T = \mathbf{B}^2$, $\mathbf{F}\mathbf{C}^{-1}\mathbf{F}^T = \mathbf{F}\left(\mathbf{F}^T\mathbf{F}\right)^{-1}\mathbf{F}^T = \mathbf{I}$, $\mathbf{B} = \mathbf{F}\mathbf{F}^T$ and $I_j(\mathbf{C}) = I_j(\mathbf{B})$ for $j = 1, 2, 3$. Note that the derivatives $\frac{\partial \tilde{\psi}}{\partial I_j}$ generally still depend on all three principal invariants.

With the help of equation (2.26) we can further express $\mathbf{C}^{-1}$ through

$$\begin{aligned}\mathbf{C}^{-1} &= \operatorname{tr}(\mathbf{C}^{-1})\mathbf{I} - \left(\operatorname{tr}(\mathbf{C}^{-1})\mathbf{I} - \mathbf{C}^{-1}\right) \\ &= \frac{1}{\det \mathbf{C}}\left(\det \mathbf{C}\operatorname{tr}(\mathbf{C}^{-1})\mathbf{I} - \left(\operatorname{tr}(\mathbf{C}^{-1})\mathbf{I} - \mathbf{C}^{-1}\right)(\det \mathbf{C})\mathbf{C}^{-T}\mathbf{C}^T\right) \\ &= \frac{1}{\det \mathbf{C}}\left(\operatorname{tr}(\mathbf{Cof}\,\mathbf{C})\mathbf{I} - \left(\operatorname{tr}(\mathbf{C}^{-1})\mathbf{I} - \mathbf{C}^{-1}\right)(\mathbf{Cof}\,\mathbf{C})\mathbf{C}\right) \\ &= \frac{1}{\det \mathbf{C}}\left(\operatorname{tr}(\mathbf{Cof}\,\mathbf{C})\mathbf{I} - \left(\operatorname{tr}(\mathbf{C})\mathbf{I} - \mathbf{C}\right)\mathbf{C}\right) \\ &= (I_3(\mathbf{C}))^{-1}\left(I_2(\mathbf{C})\mathbf{I} - I_1(\mathbf{C})\mathbf{C} + \mathbf{C}^2\right).\end{aligned}$$

Inserting this expression into (2.27) leads to an alternative expression

$$\boldsymbol{\Sigma} = 2\left(\frac{\partial\tilde{\psi}}{\partial I_1} + \frac{\partial\tilde{\psi}}{\partial I_2}I_1(\mathbf{C}) + \frac{\partial\tilde{\psi}}{\partial I_3}I_2(\mathbf{C})\right)\mathbf{I} - 2\left(\frac{\partial\tilde{\psi}}{\partial I_2} + \frac{\partial\tilde{\psi}}{\partial I_3}I_1(\mathbf{C})\right)\mathbf{C} + 2\frac{\partial\tilde{\psi}}{\partial I_3}\mathbf{C}^2. \quad (2.29)$$

Thus if we consider a homogeneous isotropic frame - indifferent hyperelastic material, we can express the second Piola - Kirchhoff stress tensor in terms of $\mathbf{C}$ and the Kirchhoff stress tensor $\boldsymbol{\tau} = \mathbf{P}\mathbf{F}^T$ in terms of $\mathbf{B}$.

2.4.2 Stress tensors in a plane strain model

In Section 2.2.7 we have seen that a plane strain model leads to a simplified structure of the deformation gradient $\mathbf{F}$ and the corresponding strain tensor $\mathbf{C}$. With this structure and the equation (2.29) it becomes clear that also $\boldsymbol{\Sigma}$ has the structure

$$\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & 0 \\ \Sigma_{21} & \Sigma_{22} & 0 \\ 0 & 0 & \Sigma_{33} \end{pmatrix},$$

since the partial derivatives $\frac{\partial\tilde{\psi}}{\partial I_j}$ in equation (2.29) of $\tilde{\psi} : \mathbb{R}^3 \to \mathbb{R}$ are in $\mathcal{L}(\mathbb{R}, \mathbb{R})$ by Definition 2.6. Thus the terms in front of $\mathbf{I}, \mathbf{C}$ and $\mathbf{C}^2$ are real - valued. Since $\boldsymbol{\Sigma}$ has this simplified structure and it holds $\boldsymbol{\tau} = \mathbf{P}\mathbf{F}^T$, $\mathbf{P} = \mathbf{F}\boldsymbol{\Sigma}$ also the stress tensors $\boldsymbol{\tau}$ and $\mathbf{P}$ have this structure.

2.4.3 Representation formulas for Mooney - Rivlin

We consider a homogeneous hyperelastic material with stored energy function

$$\hat{\psi}_{MR}(\mathbf{F}) := \alpha\,|\mathbf{F}|^2 + \beta(\det\mathbf{F})^2 - \gamma\ln(\det\mathbf{F}) + \delta\,|\mathbf{Cof}\,\mathbf{F}|^2, \quad \mathbf{F} \in \mathbb{M}, \quad (2.30)$$

with the Frobenius norm $|\cdot|$ (see Section 2.1.2) and parameters $\alpha, \beta, \gamma > 0, \delta \geq 0$. This stored energy function is motivated by the fact that its structure is quite simple, it includes all three principal invariants, is polyconvex as we will see in Section 2.4.4 and obviously satisfies the requirements of Remark 2.23. This concrete stored energy function belongs to a Mooney - Rivlin material and is proposed in Section 4.10 in [Cia88].
For a rotation $\mathbf{Q} \in \mathbb{O}$ it holds by definition $\det\mathbf{Q} = 1$ and $\mathbf{Q}^T = \mathbf{Q}^{-1}$. Then it holds $\det(\mathbf{FQ}) = \det(\mathbf{QF}) = \det\mathbf{F}$. Additionally it holds with $\mathbf{C} = \mathbf{F}^T\mathbf{F}$ for $\mathbf{F} \in \mathbb{M}$

$$\begin{aligned} |\mathbf{FQ}|^2 &= \operatorname{tr}(\mathbf{Q}^T\mathbf{F}^T\mathbf{FQ}) = \operatorname{tr}(\mathbf{Q}^T\mathbf{CQ}) = \operatorname{tr}(\mathbf{C}) = |\mathbf{F}|^2, \\ |\mathbf{QF}|^2 &= \operatorname{tr}(\mathbf{F}^T\mathbf{Q}^T\mathbf{QF}) = \operatorname{tr}(\mathbf{C}) = |\mathbf{F}|^2, \end{aligned}$$

$$
\begin{aligned}
|\mathbf{Cof}\,(\mathbf{FQ})|^2 &= \operatorname{tr}\left((\mathbf{Cof}\,(\mathbf{FQ}))^T\mathbf{Cof}\,(\mathbf{FQ})\right) = \operatorname{tr}\left((\det(\mathbf{FQ}))^2(\mathbf{FQ})^{-1}(\mathbf{FQ})^{-T}\right)\\
&= \operatorname{tr}\left((\det\mathbf{F})^2\mathbf{Q}^{-1}\mathbf{F}^{-1}\mathbf{F}^{-T}\mathbf{Q}^{-T}\right) = \operatorname{tr}\left((\det\mathbf{F})^2\mathbf{F}^{-1}\mathbf{F}^{-T}\right)\\
&= \operatorname{tr}\left((\mathbf{Cof}\,\mathbf{F})^T\mathbf{Cof}\,\mathbf{F}\right) = |\mathbf{Cof}\,\mathbf{F}|^2,\\
|\mathbf{Cof}\,(\mathbf{QF})|^2 &= \operatorname{tr}\left((\mathbf{Cof}\,(\mathbf{QF}))^T\mathbf{Cof}\,(\mathbf{QF})\right) = \operatorname{tr}\left((\det(\mathbf{QF}))^2(\mathbf{QF})^{-1}(\mathbf{QF})^{-T}\right)\\
&= \operatorname{tr}\left((\det\mathbf{F})^2\mathbf{F}^{-1}\mathbf{F}^{-T}\right) = \operatorname{tr}\left((\mathbf{Cof}\,\mathbf{F})^T\mathbf{Cof}\,\mathbf{F}\right) = |\mathbf{Cof}\,\mathbf{F}|^2.
\end{aligned}
$$

Therefore by equation (2.30) it holds $\hat{\psi}_{MR}(\mathbf{FQ}) = \hat{\psi}_{MR}(\mathbf{F}) = \hat{\psi}_{MR}(\mathbf{QF})$ for all $\mathbf{F} \in \mathbb{M}$ and all rotations $\mathbf{Q} \in \mathbb{O}$, i.e. this material is frame - indifferent and isotropic.
For $\mathbf{C} = \mathbf{F}^T\mathbf{F}$ it holds $\det\mathbf{C} = (\det\mathbf{F})^2$, $\operatorname{tr}(\mathbf{C}) = \operatorname{tr}(\mathbf{F}^T\mathbf{F}) = |\mathbf{F}|^2$ and

$$
\begin{aligned}
\operatorname{tr}(\mathbf{Cof}\,\mathbf{C}) &= \operatorname{tr}(\mathbf{Cof}\,(\mathbf{F}^T\mathbf{F})) = \operatorname{tr}\left(\det(\mathbf{F}^T\mathbf{F})(\mathbf{F}^T\mathbf{F})^{-T}\right) = \operatorname{tr}\left((\det\mathbf{F})^2\mathbf{F}^{-1}\mathbf{F}^{-T}\right)\\
&= \operatorname{tr}\left((\mathbf{Cof}\,\mathbf{F})^T\mathbf{Cof}\,\mathbf{F}\right) = |\mathbf{Cof}\,\mathbf{F}|^2.
\end{aligned}
$$

The corresponding function $\psi_{MR} : \mathbb{R}^{3\times3}_{\text{sym}} \to \mathbb{R}$ to $\hat{\psi}_{MR}$ for this Mooney - Rivlin material according to (2.19) is therefore

$$
\psi_{MR}(\mathbf{C}) = \alpha\operatorname{tr}(\mathbf{C}) + \beta\det\mathbf{C} - \gamma\ln(\det\mathbf{C})^{\frac{1}{2}} + \delta\operatorname{tr}(\mathbf{Cof}\,\mathbf{C}). \tag{2.31}
$$

The function $\tilde{\psi}_{MR} : \mathbb{R}^3 \to \mathbb{R}$ in equation (2.24) to this material is obviously

$$
\tilde{\psi}_{MR}(I_1, I_2, I_3) = \alpha I_1 + \beta I_3 - \frac{\gamma}{2}\ln(I_3) + \delta I_2.
$$

The partial derivatives of $\tilde{\psi}_{MR}(I_1, I_2, I_3)$ are

$$
\frac{\partial\tilde{\psi}_{MR}}{\partial I_1} = \alpha, \quad \frac{\partial\tilde{\psi}_{MR}}{\partial I_2} = \delta, \quad \frac{\partial\tilde{\psi}_{MR}}{\partial I_3} = \beta - \frac{\gamma}{2I_3}.
$$

From equation (2.28) we achieve

$$
\begin{aligned}
\boldsymbol{\tau}_{MR} &= 2\frac{\partial\tilde{\psi}_{MR}}{\partial I_3}I_3(\mathbf{B})\mathbf{I} + 2\left(\frac{\partial\tilde{\psi}_{MR}}{\partial I_1} + \frac{\partial\tilde{\psi}_{MR}}{\partial I_2}I_1(\mathbf{B})\right)\mathbf{B} - 2\frac{\partial\tilde{\psi}_{MR}}{\partial I_2}\mathbf{B}^2\\
&= 2\left(\beta - \frac{\gamma}{2I_3(\mathbf{B})}\right)I_3(\mathbf{B})\mathbf{I} + 2\left(\alpha + \delta I_1(\mathbf{B})\right)\mathbf{B} - 2\delta\,\mathbf{B}^2\\
&= \left(2\beta\,I_3(\mathbf{B}) - \gamma\right)\mathbf{I} + 2\left(\alpha + \delta I_1(\mathbf{B})\right)\mathbf{B} - 2\delta\,\mathbf{B}^2\\
&= \left(2\beta\det\mathbf{B} - \gamma\right)\mathbf{I} + 2\left(\alpha + \delta\operatorname{tr}(\mathbf{B})\right)\mathbf{B} - 2\delta\,\mathbf{B}^2\\
&= 2\alpha\mathbf{B} + \left(2\beta\det\mathbf{B} - \gamma\right)\mathbf{I} + 2\delta\left(\operatorname{tr}(\mathbf{B})\mathbf{B} - \mathbf{B}^2\right)
\end{aligned} \tag{2.32}
$$

and from equation (2.27) we achieve

$$
\begin{aligned}
\boldsymbol{\Sigma}_{MR} &= 2\left(\frac{\partial\tilde{\psi}_{MR}}{\partial I_1} + \frac{\partial\tilde{\psi}_{MR}}{\partial I_2}I_1(\mathbf{C})\right)\mathbf{I} - 2\frac{\partial\tilde{\psi}_{MR}}{\partial I_2}\mathbf{C} + 2\frac{\partial\tilde{\psi}_{MR}}{\partial I_3}I_3(\mathbf{C})\mathbf{C}^{-1}\\
&= 2(\alpha + \delta I_1(\mathbf{C}))\mathbf{I} - 2\delta\mathbf{C} + 2\left(\beta - \frac{\gamma}{2I_3(\mathbf{C})}\right)I_3(\mathbf{C})\mathbf{C}^{-1}\\
&= 2\alpha\,\mathbf{I} + (2\beta\det\mathbf{C} - \gamma)\mathbf{C}^{-1} + 2\delta\left(\operatorname{tr}(\mathbf{C})\mathbf{I} - \mathbf{C}\right)
\end{aligned} \tag{2.33}
$$

as expressions for the Kirchhoff and the second Piola - Kirchhoff stress tensors in three - dimensional elasticity.

2.4.4 Polyconvexity of Mooney - Rivlin

In this part we show that the Mooney - Rivlin model with stored energy function $\hat{\psi}$, defined by equation (2.30), is polyconvex.
For this purpose we define the mappings

$$\begin{aligned} g_1 : \mathbb{R}^{3\times 3} \to \mathbb{R}, \quad & \mathbf{A} \mapsto |\mathbf{A}|^2, \\ g_2 : (0,\infty) \to \mathbb{R}, \quad & x \mapsto \beta x^2 - \gamma \ln(x) \end{aligned}$$

where β and γ are the positive constants in (2.30).
Since $g_1(\mathbf{A}) = |\mathbf{A}|^2 = \operatorname{tr}(\mathbf{A}^T\mathbf{A}) = I_1(\mathbf{A}^T\mathbf{A})$ we get the Gâteaux derivatives

$$\begin{aligned} g_1'(\mathbf{A})[\mathbf{H}] &= I_1'(\mathbf{A}^T\mathbf{A})[(\mathbf{A}^T\mathbf{A})'[\mathbf{H}]] = \operatorname{tr}((\mathbf{A}^T\mathbf{A})'[\mathbf{H}]) = \operatorname{tr}(\mathbf{H}^T\mathbf{A} + \mathbf{A}^T\mathbf{H}) = 2\mathbf{A} : \mathbf{H} \\ \Rightarrow g_1''(\mathbf{A})[\mathbf{E},\mathbf{H}] &= \frac{d}{dt} g'(\mathbf{A}+t\mathbf{E})[\mathbf{H}]\big|_{t=0} = \frac{d}{dt}\left([2(\mathbf{A}+t\mathbf{E})] : \mathbf{H}\right)\big|_{t=0} = 2\mathbf{E} : \mathbf{H} \end{aligned}$$

for all $\mathbf{E}, \mathbf{H} \in \mathbb{R}^{3\times 3}$ and

$$g_2'(x) = 2\beta\, x - \frac{\gamma}{x} \Rightarrow g_2''(x) = 2\beta + \frac{\gamma}{x^2} > 0 \quad \forall\, x \in (0,\infty).$$

With the help of Proposition 2.27 it follows that g_2 is convex on $(0,\infty)$ and g_1 is convex on $\mathbb{R}^{3\times 3}$, since

$$g_1''(\mathbf{H})[\mathbf{E}-\mathbf{H}, \mathbf{E}-\mathbf{H}] = 2(\mathbf{E}-\mathbf{H}) : (\mathbf{E}-\mathbf{H}) = 2|\mathbf{E}-\mathbf{H}|^2 \geq 0$$

for all $\mathbf{E}, \mathbf{H} \in \mathbb{R}^{3\times 3}$.
With the definitions of g_1 and g_2 and the mapping

$$\begin{aligned} g :\ & \mathbb{U} := \mathbb{R}^{3\times 3} \times \mathbb{R}^{3\times 3} \times (0,\infty) \to \mathbb{R} \\ & (\mathbf{A}, \mathbf{B}, x) \mapsto \alpha g_1(\mathbf{A}) + g_2(x) + \delta g_1(\mathbf{B}) = \alpha|\mathbf{A}|^2 + \beta x^2 - \gamma \ln(x) + \delta|\mathbf{B}|^2 \end{aligned}$$

it holds $\hat{\psi}_{MR}(\mathbf{F}) = g(\mathbf{F}, \mathbf{Cof}\,\mathbf{F}, \det\mathbf{F})$ for all $\mathbf{F} \in \mathbb{M}$.
Due to the convexity of g_1 on $\mathbb{R}^{3\times 3}$ and g_2 on $(0,\infty)$ it holds for $\mathbf{U}_1 := (\mathbf{A}_1, \mathbf{B}_1, x_1)$, $\mathbf{U}_2 := (\mathbf{A}_2, \mathbf{B}_2, x_2) \in \mathbb{U}$ and $\lambda \in [0,1]$ the inequality

$$\begin{aligned} g(\lambda\mathbf{U}_1 + (1-\lambda)\mathbf{U}_2) &= g(\lambda\mathbf{A}_1 + (1-\lambda)\mathbf{A}_2, \lambda\mathbf{B}_1 + (1-\lambda)\mathbf{B}_2, \lambda x_1 + (1-\lambda)x_2) \\ &= \alpha g_1(\lambda\mathbf{A}_1 + (1-\lambda)\mathbf{A}_2) + g_2(\lambda x_1 + (1-\lambda)x_2) + \delta g_1(\lambda\mathbf{B}_1 + (1-\lambda)\mathbf{B}_2) \\ &\leq \lambda\left(\alpha g_1(\mathbf{A}_1) + g_2(x_1) + \delta g_1(\mathbf{B}_1)\right) + (1-\lambda)\left(\alpha g_1(\mathbf{A}_2) + g_2(x_2) + \delta g_1(\mathbf{B}_2)\right) \\ &= \lambda g(\mathbf{A}_1, \mathbf{B}_1, x_1) + (1-\lambda) g(\mathbf{A}_2, \mathbf{B}_2, x_2) \\ &= \lambda g(\mathbf{U}_1) + (1-\lambda) g(\mathbf{U}_2), \end{aligned}$$

i.e. g is convex on $\mathbb{U}$ and by Definition 2.26 we obtain the polyconvexity of $\hat{\psi}_{MR}$.

2.4.5 Consistency with linear elasticity

In this work our aim is to deal with nonlinear hyperelastic material models, i.e. the nonlinearities that we have listed in Section 2.2.4 can and will occur. From physical experiments one knows that a material under sufficiently small loads has firstly a linear behavior, i.e. if one doubles the load one doubles also the displacement. However, one observes in physical experiments that there exists a point where the material behavior becomes nonlinear. A reasonable model should reflect both behaviors, the linear for „small“ loads and the nonlinear for „larger“ loads. A nonlinear material law must therefore turn into the linear model for small loads. If we apply no loads the displacement $\mathbf{u}$ is reasonably $\mathbf{0}$. Small loads mean that we get a displacement in the neighborhood of $\mathbf{u} = \mathbf{0}$. If a nonlinear model turns into the model of linear elasticity in a neighborhood of $\mathbf{u} = \mathbf{0}$ we say that the model is consistent with linear elasticity. In this case also the use of the Lamé constants, introduced in Section 2.2.3 for linear elastic behavior, is meaningful.
With these considerations it is reasonable to assume the following conditions for an (nonlinear) elasticity model:

1. If one has a zero displacement $\mathbf{u} = \mathbf{0}$, the given body is not strained and therefore the corresponding stresses, given by the stress - strain relation, should be zero. Therefore one assumes that one has no occurring stresses for $\mathbf{u} = \mathbf{0}$, i.e. mathematically for both Piola - Kirchhoff stress tensors one supposes

$$\boldsymbol{\Sigma}(\mathbf{u} = \mathbf{0}) = \mathbf{0} = \mathbf{P}(\mathbf{u} = \mathbf{0}). \tag{2.34}$$

   In this case one obtains a stress - free reference configuration.

2. The second condition assumes that the stress - strain relation in nonlinear elasticity turns into the linear stress - strain relation of linear elasticity in a neighborhood of $\mathbf{u} = \mathbf{0}$. Thus if we linearize a stress tensor about $\mathbf{u} = \mathbf{0}$, assuming that it is Fréchet differentiable at $\mathbf{u} = \mathbf{0}$, we should obtain the stress - strain relation (2.10) of linear elasticity. For both Piola - Kirchhoff stress tensors $\mathbf{P} = \mathbf{P}(\mathbf{u}), \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\mathbf{u})$, related by $\mathbf{P}(\mathbf{u}) = \mathbf{F}(\mathbf{u})\boldsymbol{\Sigma}(\mathbf{u})$ this results under the first assumption of a stress - free reference configuration into the conditions

$$\mathbf{P}(\mathbf{v}) = \mathbf{P}(\mathbf{0} + \mathbf{v}) \approx \mathbf{P}(\mathbf{0}) + \mathbf{P}'(\mathbf{0})[\mathbf{v}] = \mathbf{P}'(\mathbf{0})[\mathbf{v}] \overset{!}{=} 2\mu\,\boldsymbol{\varepsilon}(\mathbf{v}) + \lambda\,\mathrm{tr}(\boldsymbol{\varepsilon}(\mathbf{v}))\mathbf{I},$$
$$\boldsymbol{\Sigma}(\mathbf{v}) = \boldsymbol{\Sigma}(\mathbf{0} + \mathbf{v}) \approx \boldsymbol{\Sigma}(\mathbf{0}) + \boldsymbol{\Sigma}'(\mathbf{0})[\mathbf{v}] = \boldsymbol{\Sigma}'(\mathbf{0})[\mathbf{v}] \overset{!}{=} 2\mu\,\boldsymbol{\varepsilon}(\mathbf{v}) + \lambda\,\mathrm{tr}(\boldsymbol{\varepsilon}(\mathbf{v}))\mathbf{I},$$

   where $\boldsymbol{\varepsilon}(\mathbf{v}) = \frac{1}{2}\left(\nabla\mathbf{v} + (\nabla\mathbf{v})^T\right)$ denotes the linear strain tensor (cf. Section 2.2.2).
   Due to

$$\mathbf{P}'(\mathbf{0})[\mathbf{v}] = \nabla\mathbf{v}\boldsymbol{\Sigma}(\mathbf{0}) + \mathbf{F}(\mathbf{0})\boldsymbol{\Sigma}'(\mathbf{0})[\mathbf{v}] = \boldsymbol{\Sigma}'(\mathbf{0})[\mathbf{v}],$$

   both stress tensors are approximately the same in a neighborhood of $\mathbf{u} = \mathbf{0}$, i.e. $\mathbf{P}(\mathbf{v}) \approx \boldsymbol{\Sigma}(\mathbf{v})$. Therefore it is sufficient to assume the condition for one of the stress tensors. In

the following we use the condition for $\boldsymbol{\Sigma}(\mathbf{v})$. Taking the deviator and the trace of this condition we obtain

$$\begin{aligned} \mathbf{dev}\,\left(\boldsymbol{\Sigma}'(\mathbf{0})[\mathbf{v}]\right) &= 2\mu\,\mathbf{dev}\,\boldsymbol{\varepsilon}(\mathbf{v}), \\ \mathrm{tr}\,\left(\boldsymbol{\Sigma}'(\mathbf{0})[\mathbf{v}]\right) &= (2\mu+3\lambda)\,\mathrm{tr}(\boldsymbol{\varepsilon}(\mathbf{v})) \end{aligned} \tag{2.35}$$

with the deviator $\mathbf{dev}\,\mathbf{A} := \mathbf{A} - \frac{1}{3}\mathrm{tr}(\mathbf{A})\mathbf{I}$ for $\mathbf{A} \in \mathbb{R}^{3\times 3}$.

If one considers now a given stored energy function $\hat{\psi}$ to a hyperelastic material, one has firstly constraintless coefficients in front of the single terms. To satisfy polyconvexity it is reasonable to assume that the coefficients are nonnegative. To guarantee further consistency with linear elasticity, we have to satisfy the conditions (2.34) and (2.35) above, i.e. we have altogether three additional conditions for the calculation of these coefficients.

Application to Mooney - Rivlin:

For the stored energy function (2.31), i.e. a homogeneous isotropic frame - indifferent material of Mooney - Rivlin type, we have four unknowns $\alpha, \beta, \gamma, \delta$ which have to be determined such that the material is consistent with linear elasticity. By equation (2.33) we know

$$\boldsymbol{\Sigma}_{MR}(\mathbf{u}) = 2\alpha\,\mathbf{I} + (2\beta\det\mathbf{C}(\mathbf{u}) - \gamma)\mathbf{C}(\mathbf{u})^{-1} + 2\delta\left(\mathrm{tr}(\mathbf{C}(\mathbf{u}))\mathbf{I} - \mathbf{C}(\mathbf{u})\right). \tag{2.36}$$

The condition (2.34) results due to $\mathbf{C}(\mathbf{0}) = \mathbf{I}$ and therefore $\mathbf{C}^{-1}(\mathbf{0}) = \mathbf{I}$, $\det\mathbf{C}(\mathbf{0}) = 1$, $\mathrm{tr}(\mathbf{C}(\mathbf{0})) = 3$ into

$$2\alpha + 2\beta - \gamma + 4\delta = 0. \tag{2.37}$$

For the derivation of the two conditions in equation (2.35) we define the mappings $h_1(\mathbf{A}) := I_1(\mathbf{A}) = \mathrm{tr}(\mathbf{A})$, $h_2(\mathbf{A}) := I_3(\mathbf{A}) = \det\mathbf{A}$, $h_3(\mathbf{A}) := \mathbf{A}^{-1}$, $h_4(\mathbf{A}) := \mathbf{A}$. We recall the Fréchet/Gâteaux derivatives from Section 2.3.3 as

$$\begin{aligned} h_1'(\mathbf{A})[\mathbf{H}] &= \partial h_1(\mathbf{A})(\mathbf{H}) = \mathrm{tr}(\mathbf{H}), \\ h_2'(\mathbf{A})[\mathbf{H}] &= \partial h_2(\mathbf{A})(\mathbf{H}) = \mathbf{Cof}\,\mathbf{A} : \mathbf{H}, \\ h_3'(\mathbf{A})[\mathbf{H}] &= \partial h_3(\mathbf{A})(\mathbf{H}) = -\mathbf{A}^{-1}\mathbf{H}\mathbf{A}^{-1}, \\ h_4'(\mathbf{A})[\mathbf{H}] &= \partial h_4(\mathbf{A})(\mathbf{H}) = \mathbf{H} \end{aligned}$$

for arbitrary matrices $\mathbf{A}, \mathbf{H}$.
We know further that $\mathbf{F} \mapsto \mathbf{C} = \mathbf{F}^T\mathbf{F}$ is Fréchet differentiable with derivative $\mathbf{H}^T\mathbf{F} + \mathbf{F}^T\mathbf{H}$. The mapping $\mathbf{u} \mapsto \mathbf{F}(\mathbf{u}) = \mathbf{I} + \nabla\mathbf{u}$ is Fréchet differentiable with derivative $\nabla\mathbf{v}$, since $\mathbf{F}(\mathbf{u}+\mathbf{v}) = \mathbf{I} + \nabla(\mathbf{u}+\mathbf{v}) = \mathbf{F}(\mathbf{u}) + \nabla\mathbf{v}$, assuming that $\mathbf{u}, \mathbf{v} : \bar{\Omega} \to \mathbb{R}^3$ are themselves Fréchet differentiable. Thus altogether we know that the mapping $\mathbf{u} \mapsto \mathbf{C}(\mathbf{u}) = (\mathbf{F}(\mathbf{u}))^T\mathbf{F}(\mathbf{u})$ is Fréchet differentiable with derivative $(\nabla\mathbf{v})^T\mathbf{F}(\mathbf{u}) + (\mathbf{F}(\mathbf{u}))^T\nabla\mathbf{v}$.
If we consider (2.36) we have

$$\boldsymbol{\Sigma}_{MR}(\mathbf{u}) = 2\alpha\,\mathbf{I} + (2\beta g_2(\mathbf{u}) - \gamma)g_3(\mathbf{u}) + 2\delta\left(g_1(\mathbf{u})\mathbf{I} - g_4(\mathbf{u})\right)$$

with $g_i(\mathbf{u}) := h_i(\mathbf{C}(\mathbf{u}))$ for $i = 1, \dots, 4$. To compute $\boldsymbol{\Sigma}'_{MR}(\mathbf{u})[\mathbf{v}]$ in $\mathbf{u} = \mathbf{0}$ we need the Fréchet/Gâteaux derivatives of g_i. The derivatives are generally given by

$$g_i'(\mathbf{u})[\mathbf{v}] = \partial g_i(\mathbf{u})(\mathbf{v}) = h_i'(\mathbf{C}(\mathbf{u}))[(\nabla \mathbf{v})^T \mathbf{F}(\mathbf{u}) + (\mathbf{F}(\mathbf{u}))^T \nabla \mathbf{v}].$$

Individually we get

$$\begin{aligned}
g_1'(\mathbf{u})[\mathbf{v}] &= \operatorname{tr}\left((\nabla \mathbf{v})^T \mathbf{F}(\mathbf{u}) + (\mathbf{F}(\mathbf{u}))^T \nabla \mathbf{v}\right), \\
g_2'(\mathbf{u})[\mathbf{v}] &= \mathbf{Cof}\, \mathbf{C}(\mathbf{u}) : \left((\nabla \mathbf{v})^T \mathbf{F}(\mathbf{u}) + (\mathbf{F}(\mathbf{u}))^T \nabla \mathbf{v}\right), \\
g_3'(\mathbf{u})[\mathbf{v}] &= -(\mathbf{C}(\mathbf{u}))^{-1} \left((\nabla \mathbf{v})^T \mathbf{F}(\mathbf{u}) + (\mathbf{F}(\mathbf{u}))^T \nabla \mathbf{v}\right) (\mathbf{C}(\mathbf{u}))^{-1}, \\
g_4'(\mathbf{u})[\mathbf{v}] &= (\nabla \mathbf{v})^T \mathbf{F}(\mathbf{u}) + (\mathbf{F}(\mathbf{u}))^T \nabla \mathbf{v}.
\end{aligned}$$

For $\mathbf{u} = \mathbf{0}$ it follows due to $\mathbf{F}(\mathbf{0}) = \mathbf{C}(\mathbf{0}) = \mathbf{I}$ and $2\boldsymbol{\varepsilon}(\mathbf{v}) = \nabla \mathbf{v} + (\nabla \mathbf{v})^T$

$$g_1'(\mathbf{0})[\mathbf{v}] = g_2'(\mathbf{0})[\mathbf{v}] = 2\operatorname{tr}(\boldsymbol{\varepsilon}(\mathbf{v})), g_3'(\mathbf{0})[\mathbf{v}] = -2\boldsymbol{\varepsilon}(\mathbf{v}), g_4'(\mathbf{0})[\mathbf{v}] = 2\boldsymbol{\varepsilon}(\mathbf{v}).$$

Using these derivatives in $\mathbf{u} = \mathbf{0}$ and the chain rule we obtain

$$\begin{aligned}
\boldsymbol{\Sigma}'_{MR}(\mathbf{0})[\mathbf{v}] &= 2\beta(g_2'(\mathbf{0})[\mathbf{v}])g_3(\mathbf{0}) + (2\beta g_2(\mathbf{0}) - \gamma)g_3'(\mathbf{0})[\mathbf{v}] + 2\delta\left(g_1'(\mathbf{0})[\mathbf{v}]\mathbf{I} - g_4'(\mathbf{0})[\mathbf{v}]\right) \\
&= 4\beta \operatorname{tr}(\boldsymbol{\varepsilon}(\mathbf{v}))\mathbf{I} + (2\beta - \gamma)(-2\boldsymbol{\varepsilon}(\mathbf{v})) + 2\delta\left(2\operatorname{tr}(\boldsymbol{\varepsilon}(\mathbf{v}))\mathbf{I} - 2\boldsymbol{\varepsilon}(\mathbf{v})\right) \\
&= \left(-4\beta + 2\gamma - 4\delta\right)\boldsymbol{\varepsilon}(\mathbf{v}) + \left(4\beta + 4\delta\right)\operatorname{tr}(\boldsymbol{\varepsilon}(\mathbf{v}))\mathbf{I}
\end{aligned}$$

and with (2.35)

$$\begin{aligned}
\mathbf{dev}\left(\boldsymbol{\Sigma}'_{MR}(\mathbf{0})[\mathbf{v}]\right) &= (-4\beta + 2\gamma - 4\delta)\,\mathbf{dev}\,\boldsymbol{\varepsilon}(\mathbf{v}) \overset{!}{=} 2\mu\,\mathbf{dev}\,\boldsymbol{\varepsilon}(\mathbf{v}) \\
\operatorname{tr}\left(\boldsymbol{\Sigma}'_{MR}(\mathbf{0})[\mathbf{v}]\right) &= (-4\beta + 2\gamma - 4\delta + 12\beta + 12\delta)\operatorname{tr}(\boldsymbol{\varepsilon}(\mathbf{v})) \overset{!}{=} (2\mu + 3\lambda)\operatorname{tr}(\boldsymbol{\varepsilon}(\mathbf{v})).
\end{aligned} \tag{2.38}$$

Since $\mathbf{v}$ is arbitrary here it must hold $-2\beta + \gamma - 2\delta = \mu$ by the first condition in (2.38). Inserting this relation directly in the second equation of (2.38) results in $\lambda = 4(\beta + \delta)$. Combining these two conditions with (2.37) we can express α, β, γ through

$$\begin{aligned}
\alpha(\mu, \delta) &= \frac{\mu}{2} - \delta, \\
\beta(\lambda, \delta) &= \frac{\lambda}{4} - \delta, \\
\gamma(\mu, \lambda) &= \mu + \frac{\lambda}{2}
\end{aligned} \tag{2.39}$$

with a free parameter $\delta \geq 0$. α is even independent of λ, β is independent of μ and γ is independent of δ. Since we have assumed $\alpha, \beta > 0$ in (2.30), we get the constraint

$$0 \leq \delta < \min\left\{\frac{\lambda}{4}, \frac{\mu}{2}\right\}. \tag{2.40}$$

Thus if we choose the parameters $\alpha, \beta, \gamma, \delta$ in (2.30) according to (2.39) and (2.40), consistency of the nonlinear Mooney - Rivlin model to linear elasticity theory is guaranteed.

2.5 Finite element spaces

In this section, the finite elements used in this work will be explained. For this purpose let $\mathcal{T}_h$ be an admissible and shape-regular triangulation of a nonempty, open, bounded and connected polygonal subset $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N} \setminus \{0\}$, into elements $T \in \mathcal{T}_h$ (cf. Chapter II § 5 in [Bra07]). In the 2d plane strain case of elasticity theory we use triangles and in the full 3d case we use tetrahedra as elements. Due to the admissibility of $\mathcal{T}_h$ it holds in particular $\bar{\Omega} = \bigcup_{T \in \mathcal{T}_h} T$. Shape-regular means that in each element $T \in \mathcal{T}_h$ a n-dimensional sphere with radius ρ_T could be inscribed and there exists a constant $\kappa > 0$ such that $\kappa \geq \frac{h_T}{\rho_T}$ for all $T \in \mathcal{T}_h$. h_T denotes the diameter of an element and $h := \max\{h_T : T \in \mathcal{T}_h\}$ the mesh size.

Further n_p, n_t, n_e (and additionally n_f in 3d) denotes the number of points, the number of triangles/tetrahedra, the number of edges (and the number of faces) in the triangulation. $\mathcal{P}_k(T)$ is the set of polynomials of degree less than or equal to k, defined on T.

For the calculation of the dimension $\mathcal{P}_k(T)$ with variables $x_1, \ldots, x_n$ we split the space in $\mathcal{P}_k(T) = \bigoplus_{i=0}^{k} \tilde{\mathcal{P}}_i(T)$ where $\tilde{\mathcal{P}}_i(T)$ denotes the set of homogeneous polynomials of degree i, i.e. all monomials of a polynomial in $\tilde{\mathcal{P}}_i(T)$ are exactly of degree i. We consider all possible combinations of $\{x_1, \ldots, x_n\}$ (with repetition, order is not taken into account) to monomials of degree i. It holds $\dim \tilde{\mathcal{P}}_i(T) = \binom{n+i-1}{i} = \frac{(n+i-1)!}{i!\,(n-1)!}$ and therefore

$$\dim \mathcal{P}_k(T) = \sum_{i=0}^{k} \dim \tilde{\mathcal{P}}_i(T) = \sum_{i=0}^{k} \binom{n+i-1}{i} = 1 + \sum_{i=1}^{k} \binom{n+i-1}{i}.$$

It follows in two ($n = 2$) and three dimensions ($n = 3$)

$$\dim \mathcal{P}_k(T) = \begin{cases} 1 + \sum_{i=1}^{k} (i+1) & = \frac{1}{2}(k+1)(k+2) & , n = 2 \\ 1 + \sum_{i=1}^{k} \frac{1}{2}(i+1)(i+2) & = \frac{1}{6}(k+1)(k+2)(k+3) & , n = 3. \end{cases} \tag{2.41}$$

We also define the space $\mathcal{P}_k(\partial T)$ which consists of all polynomials of degree k defined on the boundary ∂T of an element $T \in \mathcal{T}_h$. Since for given dimension n the boundary of an element is $(n-1)$ dimensional, it holds

$$\dim \mathcal{P}_k(\partial T) = \#(\text{boundary segments of the element}) \cdot \left(1 + \sum_{i=1}^{k} \binom{(n-1)+i-1}{i}\right).$$

It follows in two ($n = 2$) and three dimensions ($n = 3$)

$$\dim \mathcal{P}_k(\partial T) = \begin{cases} 3 \cdot \left(1 + \sum_{i=1}^{k} \binom{i}{i}\right) & = 3(k+1) & , n = 2 \\ 4 \cdot \left(1 + \sum_{i=1}^{k} \binom{i+1}{i}\right) & = 2(k+1)(k+2) & , n = 3. \end{cases} \tag{2.42}$$

2.5.1 Piecewise polynomial elements

Continuous elements:

For the approximation of each component of the displacement $\mathbf{u}$ in elasticity theory we define a (scalar - valued) space for an integer $k \geq 0$ as

$$\mathcal{P}_k(\mathcal{T}_h) := \{v \in L^\infty(\Omega) : v|_T \in \mathcal{P}_k(T) \,\forall\, T \in \mathcal{T}_h\}.$$

In the following proposition we show that a function $v \in \mathcal{P}_k(\mathcal{T}_h)$ which is additionally continuous in the domain $\bar{\Omega}$ is also in the Sobolev space $W^{1,p}(\Omega)$. This means that continuous piecewise polynomial elements are suitable for $W^{1,p}(\Omega)$ - approximations.

Proposition 2.43: (Conformity in Sobolev spaces)
Let $p \in [1, \infty]$ be arbitrary and $k \geq 0$ an integer. A function $v \in \mathcal{P}_k(\mathcal{T}_h)$, which is additionally continuous in $\bar{\Omega}$, is also in $W^{1,p}(\Omega)$.

Proof:
Let $v \in \mathcal{P}_k(\mathcal{T}_h)$ be a function satisfying the additional assumption of continuity in $\bar{\Omega}$. By definition of $\mathcal{P}_k(\mathcal{T}_h)$ the function is in $L^\infty(\Omega) \subseteq L^p(\Omega)$ for $p \in [1, \infty]$. Furthermore we know that a function $v \in \mathcal{P}_k(\mathcal{T}_h)$ is piecewise in $W^{1,p}(T)$ for all $T \in \mathcal{T}_h$ and $p \in [1, \infty]$. With an arbitrary test function $\varphi \in C_0^\infty(\Omega)$, the decomposition $\bar{\Omega} = \bigcup_{T \in \mathcal{T}_h} T$ and element - wise partial integration (see Theorem 6.1 - 9 in [Cia88]) it holds for all multi - indices α with $|\alpha| = 1$

$$\begin{aligned}
\int_\Omega v(\mathbf{x})\,(\partial^\alpha \varphi(\mathbf{x}))\,dx &= \sum_{T \in \mathcal{T}_h} \int_T v(\mathbf{x})\,(\partial_i \varphi(\mathbf{x}))\,dx \\
&= \sum_{T \in \mathcal{T}_h} \left[-\int_T (\partial_i v(\mathbf{x}))\,\varphi(\mathbf{x})\,dx + \int_{\partial T} v(\mathbf{x})\,\varphi(\mathbf{x})\,n_i\,ds \right] \\
&= -\sum_{T \in \mathcal{T}_h} \int_T (\partial_i v(\mathbf{x}))\,\varphi(\mathbf{x})\,dx = (-1)^{|\alpha|} \int_\Omega (\partial_i v(\mathbf{x}))\,\varphi(\mathbf{x})\,dx.
\end{aligned}$$

Here we have set $\partial_i \varphi := \partial^\alpha \varphi$ for $\alpha = (0, \ldots, \alpha_i, \ldots, 0)$ with $\alpha_i = 1$ and have used the assumed continuity of the function v in $\bar{\Omega}$ and the fact that φ vanishes on $\partial\Omega$. n_i denotes the i - th component of the outer normal $\mathbf{n}$ on ∂T. Due to $v \in W^{1,p}(T)$ for all elements we know that $\partial^\alpha v$ is piecewise in $L^p(T)$ and therefore altogether $\partial^\alpha v \in L^p(\Omega)$ for $p \in [1, \infty]$. By Definition 2.16 we see that $\partial^\alpha v = \partial_i v$ is the weak derivative of v.

□

For practical purposes one usually defines nodal basis functions $v_i \in \mathcal{P}_k(\hat{T})$, $i = 1, \ldots, \dim \mathcal{P}_k(\hat{T})$ on a reference element $\hat{T}$ and uses an invertible Fréchet differentiable mapping $\mathbf{F}_T : \hat{T} \to T$ (with invertible Jacobi matrix $\mathbf{J}_{\mathbf{F}_T}$) from the reference element to an arbitrary element T of the triangulation to define basis functions

$$v_i(\mathbf{x}) := \hat{v}_i(\mathbf{F}_T^{-1}(\mathbf{x})), \quad \mathbf{x} \in T.$$

Note that we restrict ourselves to affine transformations $\mathbf{F}_T(\hat{\mathbf{x}}) = \mathbf{M}\hat{\mathbf{x}} + \mathbf{a}$ as mapping between the reference element $\hat{T}$ and an arbitrary element T in this work. This means that we use no isoparametric elements.
In the case $k = 2$, i.e. quadratic elements, we use the degrees of freedom depicted in Figure 2.4.

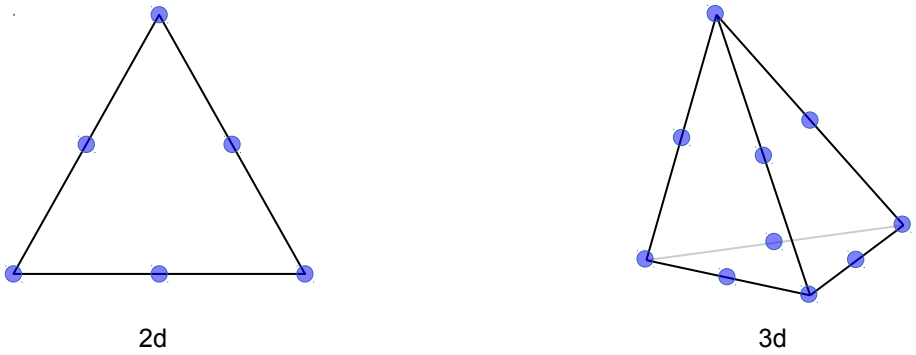


Figure 2.4: Piecewise quadratic elements $\mathcal{P}_2(T)$ in two and three dimensions

Thus we have locally 6 degrees of freedom in 2d and 10 degrees of freedom in 3d according to (2.41).
In finite element methods the common way is to build local matrices on each element and assembling them afterwards to a global matrix. In the case of continuous piecewise quadratic elements, i.e. a function in $\mathcal{P}_2(\mathcal{T}_h)$, one obtains altogether $n_p + n_e$ degrees of freedom in two and three dimensions. These degrees of freedom will be reduced afterwards due to prescribed boundary conditions in the problem. The following result can be found in Proposition 2.2.2 in [BBF13] and states an estimate for the approximation error using piecewise polynomial elements.

Proposition 2.44: (Approximation error in $H^s(\Omega)$)
Let the mapping $\mathbf{F}_T : \hat{T} \to T$ be affine and $I_h : H^s(\Omega) \to \mathcal{P}_k(\mathcal{T}_h)$ with $I_h p_k = p_k$ for all $p_k \in \mathcal{P}_k(T)$ and all $T \in \mathcal{T}_h$ the interpolation operator defined in [BBF13]. Let further $\Delta T := \{T' : \bar{T}' \cap \bar{T} \neq \emptyset\}$ be a patch around the element T , $\sigma_{\Delta T} := \max\limits_{T' \in \Delta T} \frac{h_{T'}}{\rho_{T'}}$ and $h_{\Delta T} := \max\limits_{T' \in \Delta T} h_{T'}$.
Then there exists a constant c, depending on k and $\sigma_{\Delta T}$, such that for $0 \leq m \leq s$, $1 \leq s \leq k+1$ it holds

$$|v - I_h v|_{H^m(T)} \leq c h_{\Delta T}^{s-m} |v|_{H^s(\Delta T)}, \quad v \in H^s(\Delta T). \tag{2.43}$$

Summing up this inequality over all $T \in \mathcal{T}_h$ and using $h \geq h_{\Delta T}$ for all possible patches ΔT leads to

$$|v - I_h v|_{H^m(\Omega)} \leq c h^{s-m} |v|_{H^s(\Omega)}, \quad v \in H^s(\Omega).$$

An immediate consequence for quadratic elements ($k = 2$), $v \in H^3(\Omega)$, i.e. $s = 3$, and $m = 1$ is $|v - I_h v|_{H^1(\Omega)} \leq ch^2$. For $v \in H^2(\Omega)$, i.e. $s = 2$, and $m = 0$ it follows $|v - I_h v|_{L^2(\Omega)} = |v - I_h v|_{H^0(\Omega)} \leq ch^2$. Combining these estimates we get under the assumption of $v \in H^3(\Omega)$ altogether

$$\|v - I_h v\|_{H^1(\Omega)} = \left(|v - I_h v|^2_{L^2(\Omega)} + |v - I_h v|^2_{H^1(\Omega)}\right)^{\frac{1}{2}} \lesssim h^2.$$

This means that if the solution is sufficiently regular we obtain a optimal convergence rate of two for piecewise quadratic elements.
An approximation result in Sobolev spaces $W^{m,p}(\Omega)$ for polyhedral domains $\Omega \subset \mathbb{R}^n$ and $1 \leq p \leq \infty$ is formulated in Corollary 4.4.24 in [BS08]. Again for quadratic elements and a function $v \in W^{3,p}(\Omega)$ a convergence rate of two is at most possible.

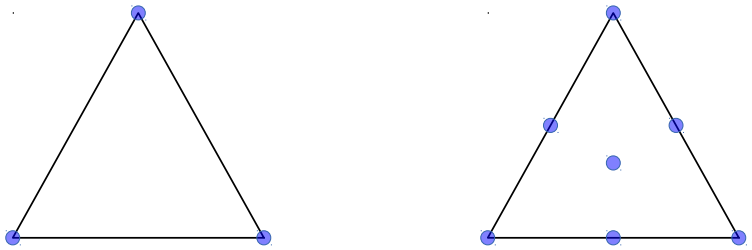
Further elements for the plane strain model:

In Section 2.4.2 we have seen that one in general gets a nonzero matrix entry P_{33} in the first Piola - Kirchhoff stress tensor $\mathbf{P}$ in the context of a plane strain model. In this case we approximate P_{33} by a discontinuous piecewise linear function. Per triangle one needs three degrees of freedom. We choose the vertices of the triangles as degrees of freedom. After assembling the local matrices one obtains a global matrix of dimension $3n_t$, due to the discontinuity. Discontinuous piecewise linear functions are suitable to approximate $L^2(\Omega)$ - functions.

In our numerical experiments in a plane strain model we will additionally compare the performance of continuous piecewise quadratic elements for approximating the displacement $\mathbf{u}$ with the so - called Fortin - Soulie elements introduced in [FS83]. This element is a piecewise quadratic element and uses besides the standard nodal basis for quadratic elements an additional basis function, a so - called bubble function. The additional basis function vanishes in the Gauss - Legendre points on the edges of $\hat{T}$ and has the value 1 in the barycenter. Altogether one obtains 7 degrees of freedom on an arbitrary element $T \in \mathcal{T}_h$. Globally, before including the boundary conditions, one has $n_p + n_e + n_t$ degrees of freedom. This element is no longer continuous on the boundary edges and therefore a non - conforming element.
The linear (discontinuous) element for the stress component P_{33} and the quadratic Fortin - Soulie element are depicted in Figure 2.5.

2.5.2 Raviart - Thomas elements

For the approximation of the single rows of the first Piola - Kirchhoff stress tensor $\mathbf{P}$ we use the well - studied Raviart - Thomas elements. A nice introduction into these elements can be found in [BBF13]. We discuss the essential facts about these elements briefly.



(a) Linear element $\mathcal{P}_1(T)$ (b) Quadratic Fortin - Soulie element

Figure 2.5: Elements for the plane strain model

For an arbitrary integer $k \geq 0$ we define on each $T \in \mathcal{T}_h$ the Raviart - Thomas space as

$$\mathcal{RT}_k(T) := \{\mathbf{v} : T \to \mathbb{R}^n | \, \mathbf{v} = (\mathcal{P}_k(T))^n + \mathbf{x}\,\mathcal{P}_k(T)\}\,, \quad \mathbf{x} := (x_1, \ldots, x_n).$$

By this definition it is clear, that $(\mathcal{P}_k(T))^n \subset \mathcal{RT}_k(T) \subset (\mathcal{P}_{k+1}(T))^n$. One can further write this space as the direct sum

$$\mathcal{RT}_k(T) = (\mathcal{P}_k(T))^n \oplus \mathbf{x}\,\tilde{\mathcal{P}}_k(T),$$

where $\tilde{\mathcal{P}}_k(T)$ denotes again the space of homogeneous polynomials of degree k. The dimension of the Raviart - Thomas space is given by

$$\dim \mathcal{RT}_k(T) = n \cdot (\dim \mathcal{P}_k(T)) + \dim \tilde{\mathcal{P}}_k(T)$$

with $\dim \mathcal{P}_k(T) = \sum_{i=0}^{k} \binom{n+i-1}{i}$ and $\dim \tilde{\mathcal{P}}_k(T) = \binom{n+k-1}{k}$, derived at the beginning of this section. For our cases of interest $n \in \{2, 3\}$ we get

$$\begin{aligned}
\dim \mathcal{RT}_k(T) &= n \cdot (\dim \mathcal{P}_k(T)) + \dim \tilde{\mathcal{P}}_k(T) \\
&= \begin{cases} 2 \cdot \left(\frac{1}{2}(k+1)(k+2)\right) + (k+1) & , \quad n = 2 \\ 3 \cdot \left(\frac{1}{6}(k+1)(k+2)(k+3)\right) + \frac{1}{2}(k+1)(k+2) & , \quad n = 3 \end{cases} \qquad (2.44) \\
&= \begin{cases} (k+1)(k+3), & , \quad n = 2 \\ \frac{1}{2}(k+1)(k+2)(k+4), & , \quad n = 3\,. \end{cases}
\end{aligned}$$

On the triangulation $\mathcal{T}_h$ we define the set of **Raviart - Thomas functions** as

$$\begin{aligned}
\mathcal{RT}_k(\mathcal{T}_h) := \{\mathbf{v} \in (L^\infty(\Omega))^n : \mathbf{v}|_T \in \mathcal{RT}_k(T)\, \forall\, T \in \mathcal{T}_h, \\
\mathbf{v} \cdot \mathbf{n} \text{ is continuous at the interfaces of elements}\}.
\end{aligned}$$

In the following proposition we show that a function $\mathbf{v} \in \mathcal{RT}_k(\mathcal{T}_h)$ is also in the function space $W^p(\mathrm{div}; \Omega)$ for all $p \in [1, \infty]$, i.e. conformity in $W^p(\mathrm{div}; \Omega)$ is ensured and therefore the Raviart - Thomas elements $\mathbf{v} \in \mathcal{RT}_k(\mathcal{T}_h)$ are suitable for $W^p(\mathrm{div}; \Omega)$ - approximations. For the proof of this conformity result the continuity of the normal component at the interfaces of elements is crucial.

Proposition 2.45: (Conformity of Raviart - Thomas elements in $W^p(\mathbf{div};\Omega)$)
Let $v \in \mathcal{RT}_k(\mathcal{T}_h)$ be a Raviart - Thomas function for given integer $k \geq 0$ and $p \in [1,\infty]$ arbitrary. Then it holds $v \in W^p(\mathrm{div};\Omega)$, i.e. $\mathcal{RT}_k(\mathcal{T}_h) \subset W^p(\mathrm{div};\Omega)$.

Proof:
The Raviart - Thomas functions $\mathbf{v} \in \mathcal{RT}_k(\mathcal{T}_h)$ are by definition in $(L^\infty(\Omega))^n \subseteq (L^p(\Omega))^n$. Furthermore they are as (vector - valued) polynomials on each element of course in $W^p(\mathrm{div};T)$ and additionally the partial weak derivatives in $L^p(T)$ exist. We set $\tilde{w}(\mathbf{x}) := \sum_{i=1}^{n} \partial_i v_i(\mathbf{x})$ which exists elementwise. It holds $\tilde{w} \in L^p(T)$ for all $T \in \mathcal{T}_h$ and therefore also $\tilde{w} \in L^p(\Omega)$. It remains to show that $\tilde{w}$ is the weak divergence of $\mathbf{v}$. For a test function $\varphi \in C_0^\infty(\Omega)$ it holds with the help of the decomposition $\bar{\Omega} = \bigcup_{T\in\mathcal{T}_h} T$ and partial integration on each of these elements

$$\begin{aligned}
\int_\Omega \tilde{w}(\mathbf{x})\varphi(\mathbf{x})\,dx &= \sum_{T\in\mathcal{T}_h} \int_T \tilde{w}(\mathbf{x})\varphi(\mathbf{x})\,dx = \sum_{T\in\mathcal{T}_h} \int_T \left(\sum_{i=1}^{n} \partial_i v_i(\mathbf{x})\right)\varphi(\mathbf{x})\,dx \\
&= \sum_{T\in\mathcal{T}_h} \sum_{i=1}^{n} \left[-\int_T v_i(\mathbf{x})\,(\partial_i\varphi(\mathbf{x}))\,dx + \int_{\partial T} v_i(\mathbf{x})\,n_i\,\varphi(\mathbf{x})\,ds \right] \\
&= \sum_{T\in\mathcal{T}_h} \left[-\int_T \mathbf{v}(\mathbf{x})\cdot\nabla\varphi(\mathbf{x})\,dx + \int_{\partial T} (\mathbf{v}(\mathbf{x})\cdot\mathbf{n})\varphi(\mathbf{x})\,ds \right] \\
&= -\int_\Omega \mathbf{v}(\mathbf{x})\cdot\nabla\varphi(\mathbf{x})\,dx.
\end{aligned}$$

The sum over the boundary integrals vanishes due to the assumed continuity of the normal component $\mathbf{v}\cdot\mathbf{n}$ with outer normals $\mathbf{n}$ and the fact that $\varphi = 0$ on $\partial\Omega$. By Definition 2.16 $\tilde{w} = \operatorname{div}\mathbf{v}$ is the weak divergence of $\mathbf{v}$.

□

For the construction of basis functions $\mathcal{RT}_k(T)$ in practice one starts again on a reference element $\hat{T}$ and takes again an invertible Fréchet differentiable mapping $\mathbf{F}_T : \hat{T} \to T$ (with invertible Jacobi matrix $\mathbf{J}_{\mathbf{F}_T}$) from the reference element to an arbitrary element T of the triangulation.
For any integer $k \geq 0$ one can define the vector - valued basis functions $\hat{\mathbf{v}}_i(\hat{\mathbf{x}})$, $i = 1,\ldots,$ $\dim\mathcal{RT}_k(\hat{T})$ on $\hat{T}$ with the help of the moments

- $$\int_{\partial\hat{T}} (\hat{\mathbf{v}}_i(\hat{\mathbf{x}})\cdot\hat{\mathbf{n}})\hat{p}_k(\hat{\mathbf{x}})\,d\hat{s}, \quad \hat{p}_k \in \mathcal{P}_k(\partial\hat{T})$$

These are $3(k+1)$ integrals in 2d and $2(k+1)(k+2)$ integrals in 3d due to the derived dimension of $\mathcal{P}_k(\partial T)$ for an arbitrary element T in equation (2.42).

- $$\int_{\hat{T}} \hat{\mathbf{v}}_i(\hat{\mathbf{x}})\cdot\hat{\mathbf{p}}_{k-1}(\hat{\mathbf{x}})\,d\hat{x}, \quad \hat{\mathbf{p}}_{k-1} \in (\mathcal{P}_{k-1}(\hat{T}))^n$$

These are $\dim\mathcal{RT}_k(\hat{T}) - \dim\mathcal{P}_k(\partial\hat{T})$ integrals, i.e. $k(k+1)$ integrals in 2d and $\frac{1}{2}k(k+1)(k+2)$ integrals in 3d.

For details concerning the linear independency of the resulting basis functions we refer to [BBF13].
Instead of using the moments to define basis functions for $\mathcal{RT}_k(\hat{T})$ one could also define basis functions by prescribing on the one hand their normal components in $k+1$ points on each of the edges in 2d (respectively $\frac{1}{2}(k+1)(k+2)$ points on each of the faces in 3d). On the other hand one uses additionally $\frac{k}{2}(k+1)$ different points in 2d (respectively $\frac{1}{6}k(k+1)(k+2)$ different points in 3d) and prescribes the x-, y-value (respectively the x-, y- and z-value) in these points. It is clear that 2 divides $k(k+1)$. It is also clear that 6 divides $k(k+1)(k+2)$, which can be proven simply with complete induction.

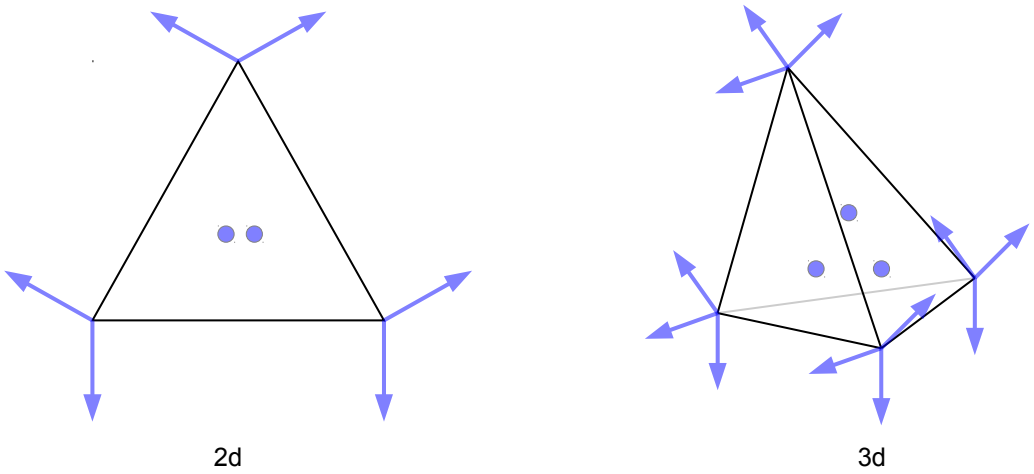


Figure 2.6: Raviart-Thomas elements $\mathcal{RT}_1(T)$ in two and three dimensions

This ansatz for defining basis functions is motivated by the fact that we have to satisfy continuity of the normal components at the element interfaces to obtain conformity in $W^p(\mathrm{div};\Omega)$. In the case $k=1$ one obtains for example the degrees of freedom drawn in Figure 2.6, i.e. we prescribe the normal components in all vertices of each edge/face and we use the barycenter inside the triangle/tetrahedron to define the basis functions. Thus in 2d we have locally 8 degrees of freedom and in 3d we have locally 15 degrees of freedom, according to the dimension of $\mathcal{RT}_1(T)$ in equation (2.44).
If one has determined the basis functions on the reference element the next step is again to transform them to an arbitrary element $T \in \mathcal{T}_h$. Since the standard transformation does not preserve normal components, we need here the so-called Piola transformation. We define the basis functions $\mathbf{v}_i$, $i=1,\dots,\dim\mathcal{RT}_k(T)$, on T in general as

$$\mathbf{v}_i(\mathbf{x}) := \frac{1}{|\det \mathbf{J}_{\mathbf{F}_T}(\mathbf{F}_T^{-1}(\mathbf{x}))|}\mathbf{J}_{\mathbf{F}_T}(\mathbf{F}_T^{-1}(\mathbf{x}))\hat{\mathbf{v}}_i(\mathbf{F}_T^{-1}(\mathbf{x})), \quad \mathbf{x}\in T.$$

For an affine orientation - preserving transformation $\mathbf{F}_T(\hat{\mathbf{x}}) = \mathbf{M}\hat{\mathbf{x}} + \mathbf{a}$ it follows

$$\mathbf{v}_i(\mathbf{x}) = \frac{1}{\det \mathbf{M}} \mathbf{M}\hat{\mathbf{v}}_i(\mathbf{F}_T^{-1}(\mathbf{x})), \quad \mathbf{x} \in T.$$

With this choice an important consequence is that the conditions above, defining the degrees of freedom, can be preserved (cf. Lemma 2.1.6 in [BBF13]).
Similar to the standard piecewise polynomial elements one assembles the local matrices of a function $\mathcal{RT}_1(\mathcal{T}_h)$ to a global matrix and gets $2(n_e + n_t)$ degrees of freedom in 2d and $3(n_f + n_t)$ degrees of freedom in 3d. Again these degrees of freedom are generally reduced through suitable boundary conditions.
The following result which can be found in [BBF13] is important to get a - priori error estimates in $H(\mathrm{div}; \Omega)$.

Proposition 2.46: (Approximation error in $H(\mathbf{div}; \Omega)$)
For the global interpolation operator $\Pi_h : H(\mathrm{div}; \Omega) \cap L^r(\Omega)^n \to \mathcal{RT}_k(\mathcal{T}_h)$ with fixed $r > 2$, defined in Section 2.5.1 and 2.5.2 in [BBF13], and $\mathbf{q} \in H^m(\Omega)^n$ it holds

$$\|\mathbf{q} - \Pi_h \mathbf{q}\|_{L^2(\Omega)} \leq c h^m |\mathbf{q}|_{H^m(\Omega)}$$

with constant c independent of h and $1 \leq m \leq k + 1$. Furthermore for $\mathrm{div}\, \mathbf{q} \in H^s(\Omega)$ it holds

$$\|\mathrm{div}(\mathbf{q} - \Pi_h \mathbf{q})\|_{L^2(\Omega)} \leq c h^s |\mathrm{div}\, \mathbf{q}|_{H^s(\Omega)}$$

with $s \leq k + 1$.

Proof:

See Proposition 2.5.4 and the statements before in [BBF13].

□

An immediate consequence for $k = 1$ and $s = m = k + 1 = 2$ and therefore a function $\mathbf{q} \in H(\mathrm{div}; \Omega) \cap L^r(\Omega)^n \cap H^2(\Omega)^n$ with $\mathrm{div}\, \mathbf{q} \in H^2(\Omega)$ is

$$\left(\|\mathbf{q} - \Pi_h \mathbf{q}\|^2_{L^2(\Omega)} + \|\mathrm{div}(\mathbf{q} - \Pi_h \mathbf{q})\|^2_{L^2(\Omega)} \right)^{\frac{1}{2}} \lesssim \left(h^4 \left(|\mathbf{q}|^2_{H^2(\Omega)} + |\mathrm{div}\, \mathbf{q}|^2_{H^2(\Omega)} \right) \right)^{\frac{1}{2}} \lesssim h^2,$$

i.e. for $k = 1$ and a $H(\mathrm{div}; \Omega)$ - conforming approximation one expects a optimal convergence rate of two.

3 Least Squares Finite Element Methods in elasticity

This section is the main chapter of this work. For the derivation of least squares finite element methods (abbr. LSFEMs) for nonlinear hyperelasticity, we use the idea of a LSFEM approach from linear elasticity. We are interested in developing a robust LSFEM method for nonlinear elasticity which approximates besides the displacement $\mathbf{u}$ also a full stress tensor. We will approximate the first Piola - Kirchhoff stress tensor $\mathbf{P}$. The simultaneous approximation of both quantities has the advantage that one needs no post - processing to determine $\mathbf{P}$. Furthermore one expects better stress approximations.
The outline of this chapter is as follows. At the beginning we state the partial differential equations that we have to solve in (nonlinear) elasticity, namely the equations of equilibrium and the stress - strain relation. Then we explain the least squares finite element method on the basis of the work [CS04] for linear elasticity. In LSFEM for linear problems one is usually interested in a „wanted property“, which leads to the well - posedness of the underlying problem. The mentioned work of Cai and Starke states such a „wanted property“ for linear elasticity.
This work is also the basis for the extension to the nonlinear case described afterwards. Here we explain the general idea of our approach for homogeneous isotropic frame - indifferent materials before we focus on the cases of a Mooney - Rivlin and a Neo - Hooke material. For the considered Neo - Hooke model we provide a detailed analysis for the nonlinear problem as well as for the corresponding linearized problem.
At the end of this chapter we explain two other possible standard discretization methods to compare our method with already existing ones in Section 6. The first method here is the simplest one in finite elements for elastic deformation problems, the so - called displacement approach or simply Galerkin method. Unfortunately, this method leads to the Poisson locking problem at least if one uses small polynomial degrees in an underlying conforming finite element space (cf. [BS92] for linear elasticity). Poisson locking means that the obtained approximations deteriorate if $\lambda \to \infty$ or equivalently if Poisson's ratio $\nu \to \frac{1}{2}$ (cf. Section 2.2.3). Therefore the displacement approach is either only suitable for compressible materials or with larger polynomial degrees. It is our aim that our approach works also in the (quasi -) incompressible case for quite small polynomial degrees. We will compare our LSFEM approach additionally with an existing **displacement - pressure** approach, which is proposed by Auricchio in [ABadVLR10] for incompressible materials.

3.1 First - order system in elasticity theory

We follow the notation of Section 2.2 and will describe the elastostatic problem in the reference configuration generally for frame - indifferent hyperelastic materials. We focus on the case of mixed boundary conditions which is more relevant for practical purposes, i.e. we have boundary conditions for $\mathbf{u}$ on Γ_D and for $\mathbf{P}$ on Γ_N. In the introduction of

elastic deformation problems we have already mentioned that two sets of equations are fundamental. The first set consists of the **equations of equilibrium**. In the reference configuration they state

$$\boxed{\begin{aligned} -\operatorname{div}\mathbf{P} &= \mathbf{f} && \text{in } \Omega, \\ \mathbf{P}\cdot\mathbf{n} &= \mathbf{g} && \text{on } \Gamma_N. \end{aligned}}$$

The first equation is an immediate consequence of the physically necessary conservation of linear momentum for a static problem (cf. Section 5.10 in [EGK11] respectively Section 2 in [Cia88]). The boundary conditions for $\mathbf{P}$ on Γ_N follow directly from the definition of the so - called Cauchy stress vector and its corresponding Cauchy stress tensor (cf. Section 2 in [Cia88]). Additionally it must hold $\mathbf{P}\mathbf{F}^T = \mathbf{F}\mathbf{P}^T$ in the domain Ω for the deformation gradient $\mathbf{F} = \nabla\boldsymbol{\varphi}$ of the deformation $\boldsymbol{\varphi}$ and the first Piola - Kirchhoff stress tensor $\mathbf{P}$. This follows directly from the physically necessary requirement of conservation of angular momentum for a static problem.

The second set of equations is given by a stress - strain relation. In linear elasticity theory we have the stress - strain relation (2.10). In nonlinear hyperelasticity the stress - strain relation can be obtained through Definition 2.20. Since we are dealing with mixed boundary conditions, we must prescribe additionally $\mathbf{u}$ on Γ_D to obtain a well - posed problem. We assume $\mathbf{u} = \mathbf{u}_D$ on Γ_D. Altogether this forms the first - order system/strong formulation for a frame - indifferent nonlinear hyperelastic material with given stored energy function $\psi : \bar{\Omega} \times \mathbb{R}^{3\times3}_{\text{sym}} \to \mathbb{R}$:

Seek the displacement $\mathbf{u} : \bar{\Omega} \to \mathbb{R}^3$ and the first Piola - Kirchhoff stress tensor $\mathbf{P} : \bar{\Omega} \to \mathbb{R}^{3\times3}$ with

$$\boxed{\begin{aligned} -\operatorname{div}\mathbf{P} &= \mathbf{f} && \text{in } \Omega, \\ \mathbf{P} &= \partial_{\mathbf{F}}\psi(\mathbf{x},\mathbf{C}) && \text{in } \Omega, \\ \mathbf{u} &= \mathbf{u}_D \text{ on } \Gamma_D, \quad \mathbf{P}\cdot\mathbf{n} = \mathbf{g} \text{ on } \Gamma_N \end{aligned}} \tag{3.1}$$

under given force densities $\mathbf{g} : \Gamma_N \to \mathbb{R}^3$ and $\mathbf{f} : \Omega \to \mathbb{R}^3$.

In linear elasticity it is not necessary to distinguish between the different stress tensors and one uses generally $\boldsymbol{\sigma}$ as notation for the stress tensor. The strong formulation with first - order system (3.1) reduces to:

Seek the displacement $\mathbf{u} : \bar{\Omega} \to \mathbb{R}^3$ and the stress tensor $\boldsymbol{\sigma} : \bar{\Omega} \to \mathbb{R}^{3\times3}$ with

$$\boxed{\begin{aligned} -\operatorname{div}\boldsymbol{\sigma} &= \mathbf{f} && \text{in } \Omega, \\ \boldsymbol{\sigma} &= 2\mu\,\boldsymbol{\varepsilon}(\mathbf{u}) + \lambda\operatorname{tr}(\boldsymbol{\varepsilon}(\mathbf{u}))\mathbf{I} =: \mathcal{C}\boldsymbol{\varepsilon}(\mathbf{u}) && \text{in } \Omega, \\ \mathbf{u} &= \mathbf{u}_D \text{ on } \Gamma_D, \quad \boldsymbol{\sigma}\cdot\mathbf{n} = \mathbf{g} \text{ on } \Gamma_N. \end{aligned}} \tag{3.2}$$

3.2 Inverse LSFEM approach for linear elasticity

The general idea of least squares finite element methods can be found in the book [BG09] of Bochev and Gunzburger. Generally one transforms a given system of partial differential

equations into a corresponding first - order system with vanishing right - hand side, i.e. in residual form. The next step in general is to put each of the single residuals into the L^2 - norm and square them. One defines the least squares functional as the sum of these squared L^2 - norms and seeks a minimizer of this functional in a suitable (problem dependent) function space. If the value of the functional is zero, one knows that one has found the exact solution. Roughly speaking, that is the idea of standard least squares finite element methods. We explain the method exemplarily and in more detail for the linear elastic problem (3.2) in the following, based on the work of [CS04].
In this work the authors start with the system described in (3.2). Obviously the material law $\boldsymbol{\sigma} = \mathcal{C}\boldsymbol{\varepsilon}(\mathbf{u})$ blows up in the limit $\lambda \to \infty$. Since the authors were interested also in the incompressible case $\lambda \to \infty$ their idea was to invert the stress - strain relation into

$$\boldsymbol{\varepsilon}(\mathbf{u}) = \frac{1}{2\mu}\left(\boldsymbol{\sigma} - \frac{\lambda}{3\lambda + 2\mu}\mathrm{tr}(\boldsymbol{\sigma})\mathbf{I}\right) =: \mathcal{C}^{-1}\boldsymbol{\sigma} =: \mathcal{A}_{lin}(\boldsymbol{\sigma}), \tag{3.3}$$

i.e. $\mathcal{A}_{lin}$ is now a mapping from stresses into strains. For $\lambda \to \infty$ one gets with the definition of the deviator

$$\lim_{\lambda\to\infty} \mathcal{A}_{lin}(\boldsymbol{\sigma}) = \lim_{\lambda\to\infty} \frac{1}{2\mu}\left(\boldsymbol{\sigma} - \frac{1}{3 + \frac{2\mu}{\lambda}}\mathrm{tr}(\boldsymbol{\sigma})\mathbf{I}\right) = \frac{1}{2\mu}\left(\boldsymbol{\sigma} - \frac{1}{3}\mathrm{tr}(\boldsymbol{\sigma})\mathbf{I}\right) = \frac{1}{2\mu}\mathbf{dev}\,\boldsymbol{\sigma}.$$

Please note that all stresses of $\boldsymbol{\sigma}$ of the form $\boldsymbol{\sigma} = c\,\mathbf{I}$, $c \in \mathbb{R}$, vanish in the incompressible limit. This means that the kernel of $\mathcal{A}_{lin}$ is non - trivial and therefore the mapping is no longer invertible in the incompressible case. We obtain the „inverse“ first - order system with vanishing right - hand side

$$\boxed{\begin{aligned} \mathrm{div}\,\boldsymbol{\sigma} + \mathbf{f} &= \mathbf{0} \quad \text{in } \Omega, \\ \mathcal{A}_{lin}(\boldsymbol{\sigma}) - \boldsymbol{\varepsilon}(\mathbf{u}) &= \mathbf{0} \quad \text{in } \Omega, \\ \mathbf{u} = \mathbf{u}_D \text{ on } \Gamma_D, &\quad \boldsymbol{\sigma}\cdot\mathbf{n} = \mathbf{g} \text{ on } \Gamma_N. \end{aligned}} \tag{3.4}$$

The boundary conditions can be imposed strongly or weakly. Both possibilities are discussed in [CS04]. For strongly imposed boundary conditions the least squares functional

$$\mathcal{F}_{lin}(\boldsymbol{\sigma}, \mathbf{u}; \mathbf{f}) := \|\mathrm{div}\,\boldsymbol{\sigma} + \mathbf{f}\|^2_{L^2(\Omega)} + \|\mathcal{A}_{lin}(\boldsymbol{\sigma}) - \boldsymbol{\varepsilon}(\mathbf{u})\|^2_{L^2(\Omega)} \tag{3.5}$$

is defined, following the general idea of LSFEM, and a minimizer $(\boldsymbol{\sigma}, \mathbf{u}) := (\boldsymbol{\sigma}^N + \hat{\boldsymbol{\sigma}}, \mathbf{u}_D + \hat{\mathbf{u}}) \in (H(\mathrm{div};\Omega)^3 + H_{\Gamma_N}(\mathrm{div};\Omega)^3) \times (H^1(\Omega)^3 + H^1_{\Gamma_D}(\Omega)^3)$ with $\boldsymbol{\sigma}^N \cdot \mathbf{n} = \mathbf{g}$ on Γ_N and $\mathbf{u} = \mathbf{u}_D$ on Γ_D of the functional is seeked. The subscripts Γ_N and Γ_D in the function spaces here denote functions in the same spaces, but with zero boundary conditions on Γ_N, respectively Γ_D. We will use this notation in the rest of this work, also for other function spaces.
Generally in finite element methods one is interested in estimating the error between the (in general unknown) exact and the approximated solution. In LSFEM the aim is to

estimate the error in a suitable norm from below and above by the defined least squares functional.
In the case of linear elasticity the main result in the work [CS04] is the following theorem.

Theorem 3.1: (Continuity and ellipticity of $\mathcal{F}_{lin}(\boldsymbol{\tau}, \mathbf{v}; \mathbf{0})$ in linear elasticity)
Let $\mathcal{V} := H_{\Gamma_N}(\mathrm{div}; \Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$. There exists a constant C, independent of λ, such that

$$\begin{aligned} \mathcal{F}_{lin}(\boldsymbol{\tau}, \mathbf{v}; \mathbf{0}) &\leq C \left(\|\boldsymbol{\varepsilon}(\mathbf{v})\|^2_{L^2(\Omega)} + \|\boldsymbol{\tau}\|^2_{L^2(\Omega)} + \|\mathrm{div}\, \boldsymbol{\tau}\|^2_{L^2(\Omega)} \right) \quad \text{(continuity)} \\ \mathcal{F}_{lin}(\boldsymbol{\tau}, \mathbf{v}; \mathbf{0}) &\geq \frac{1}{C} \left(\|\boldsymbol{\varepsilon}(\mathbf{v})\|^2_{L^2(\Omega)} + \|\boldsymbol{\tau}\|^2_{L^2(\Omega)} + \|\mathrm{div}\, \boldsymbol{\tau}\|^2_{L^2(\Omega)} \right) \quad \text{(ellipticity)} \end{aligned} \tag{3.6}$$

for all $(\boldsymbol{\tau}, \mathbf{v}) \in \mathcal{V}$.

Proof:

see Theorem 3.1 in [CS04]

□

With the norm $\|(\boldsymbol{\tau}, \mathbf{v})\|_{\mathcal{V}} := \left(\|\boldsymbol{\tau}\|^2_{H(\mathrm{div};\, \Omega)} + \|\mathbf{v}\|^2_{H^1(\Omega)} \right)^{\frac{1}{2}}$ on the space $\mathcal{V}$ an immediate consequence is

$$\boxed{\|(\boldsymbol{\tau}, \mathbf{v})\|^2_{\mathcal{V}} \lesssim \mathcal{F}_{lin}(\boldsymbol{\tau}, \mathbf{v}; \mathbf{0}) \lesssim \|(\boldsymbol{\tau}, \mathbf{v})\|^2_{\mathcal{V}}, \quad (\boldsymbol{\tau}, \mathbf{v}) \in \mathcal{V}.} \tag{3.7}$$

Here we have used on the one hand the simple estimate $\|\boldsymbol{\varepsilon}(\mathbf{v})\|^2_{L^2(\Omega)} \leq \|\mathbf{v}\|^2_{H^1(\Omega)}$ for $\mathbf{v} \in H^1(\Omega)^3$ and on the other hand Korn's inequality, i.e. $\|\boldsymbol{\varepsilon}(\mathbf{v})\|^2_{L^2(\Omega)} \gtrsim \|\mathbf{v}\|^2_{H^1(\Omega)}$, $\mathbf{v} \in H^1_{\Gamma_D}(\Omega)^3$, (see Corollary 11.2.22 in [BS08]). The abbreviations $\lesssim$ and $\gtrsim$ stands again for inequalities up to positive constants and are often used in the rest of this work.
Equation (3.7) is the „wanted property“ one is usually interested in least squares finite element methods for (linear) problems. With this property one obtains beneficial consequences. The following consequences are not restricted to the problem of linear elasticity, i.e. the explanations below work in the same way for general linear least squares problems of the form $\mathcal{F}(w; r) = \|\mathcal{L}(w) - r\|^2_{L^2(\Omega)}$, $w \in V, r \in L^2(\Omega)$, with a linear operator $\mathcal{L}$ which is defined on a suitable function space V and maps into a subspace of $L^2(\Omega)$.
In the context of linear elasticity we set for $(\boldsymbol{\sigma}, \mathbf{u}) := (\boldsymbol{\sigma}^N + \hat{\boldsymbol{\sigma}}, \mathbf{u}_D + \hat{\mathbf{u}}) \in \Big(H(\mathrm{div}; \Omega)^3 + H_{\Gamma_N}(\mathrm{div}; \Omega)^3 \Big) \times \Big(H^1(\Omega)^3 + H^1_{\Gamma_D}(\Omega)^3 \Big)$, $\boldsymbol{\sigma}^N \cdot \mathbf{n} = \mathbf{g}$ on Γ_N, $\mathbf{u} = \mathbf{u}_D$ on Γ_D

$$\mathcal{L}(\boldsymbol{\sigma}, \mathbf{u}) := \begin{pmatrix} \mathrm{div}\, \boldsymbol{\sigma} \\ \mathcal{A}_{lin}(\boldsymbol{\sigma}) - \boldsymbol{\varepsilon}(\mathbf{u}) \end{pmatrix}, \quad \mathbf{r} := \begin{pmatrix} -\mathbf{f} \\ \mathbf{0} \end{pmatrix}. \tag{3.8}$$

With this definition it holds (cf. (3.5))

$$\mathcal{F}_{lin}(\boldsymbol{\sigma}, \mathbf{u}; \mathbf{f}) = \|\mathcal{L}(\boldsymbol{\sigma}, \mathbf{u}) - \mathbf{r}\|^2_{L^2(\Omega)}. \tag{3.9}$$

Since we are seeking $(\hat{\boldsymbol{\sigma}}, \hat{\mathbf{u}}) \in \mathcal{V}$ such that $\mathcal{F}_{lin}(\boldsymbol{\sigma}, \mathbf{u}; \mathbf{f}) = \mathcal{F}_{lin}(\boldsymbol{\sigma}^N + \hat{\boldsymbol{\sigma}}, \mathbf{u}_D + \hat{\mathbf{u}}; \mathbf{f})$ is minimized in $\mathcal{V}$, the necessary condition is

$$\begin{aligned} &\frac{d}{dt}\mathcal{F}_{lin}(\boldsymbol{\sigma} + t\boldsymbol{\tau}, \mathbf{u} + t\mathbf{v}; \mathbf{f}))\big|_{t=0} = 0 \\ &\Leftrightarrow (\mathcal{L}(\boldsymbol{\sigma}, \mathbf{u}) - \mathbf{r}, \mathcal{L}(\boldsymbol{\tau}, \mathbf{v}))_{L^2(\Omega)} = 0 \\ &\Leftrightarrow \quad \underbrace{(\mathcal{L}(\hat{\boldsymbol{\sigma}}, \hat{\mathbf{u}}), \mathcal{L}(\boldsymbol{\tau}, \mathbf{v}))_{L^2(\Omega)}}_{=:\, a\left((\hat{\boldsymbol{\sigma}}, \hat{\mathbf{u}}), (\boldsymbol{\tau}, \mathbf{v})\right)} = \underbrace{\big(\mathbf{r} - \mathcal{L}(\boldsymbol{\sigma}^N, \mathbf{u}_D), \mathcal{L}(\boldsymbol{\tau}, \mathbf{v})\big)_{L^2(\Omega)}}_{=:\, F\left((\boldsymbol{\tau}, \mathbf{v})\right)} \end{aligned} \tag{3.10}$$

for all $(\boldsymbol{\tau}, \mathbf{v}) \in \mathcal{V}$ with corresponding bilinear form $a : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ and linear form $F : \mathcal{V} \to \mathbb{R}$. With these considerations we get the following proposition:

Proposition 3.2: (Existence and uniqueness in linear elasticity)
We consider the minimization problem of (3.9) and assume that the property (3.7) holds. Then the corresponding bilinear form, defined in (3.10), is symmetric, continuous and coercive on $\mathcal{V}$. Furthermore under the assumption of $\mathbf{f} \in L^2(\Omega)^3$, $\boldsymbol{\sigma}^N \in H(\mathrm{div};\, \Omega)^3$ and $\mathbf{u}_D \in H^1(\Omega)^3$ the linear form F in (3.10) is continuous.

Proof:

Symmetry of a:
Obviously it holds $a\left((\hat{\boldsymbol{\sigma}}, \hat{\mathbf{u}}), (\boldsymbol{\tau}, \mathbf{v})\right) = a\left((\boldsymbol{\tau}, \mathbf{v}), (\hat{\boldsymbol{\sigma}}, \hat{\mathbf{u}})\right)$ for all $(\hat{\boldsymbol{\sigma}}, \hat{\mathbf{u}}), (\boldsymbol{\tau}, \mathbf{v}) \in \mathcal{V}$.
Continuity of a:
With the help of the Cauchy - Schwarz inequality, (3.9) for $\mathbf{r} = \mathbf{0}\,(\Leftrightarrow \mathbf{f} = \mathbf{0})$ and (3.7) it holds for $(\hat{\boldsymbol{\sigma}}, \hat{\mathbf{u}}), (\boldsymbol{\tau}, \mathbf{v}) \in \mathcal{V}$

$$\begin{aligned} |a((\hat{\boldsymbol{\sigma}}, \hat{\mathbf{u}}), (\boldsymbol{\tau}, \mathbf{v}))| &= |\left(\mathcal{L}(\hat{\boldsymbol{\sigma}}, \hat{\mathbf{u}}), \mathcal{L}(\boldsymbol{\tau}, \mathbf{v})\right)_{L^2(\Omega)}| \\ &\leq \|\mathcal{L}(\hat{\boldsymbol{\sigma}}, \hat{\mathbf{u}})\|_{L^2(\Omega)} \|\mathcal{L}(\boldsymbol{\tau}, \mathbf{v})\|_{L^2(\Omega)} \\ &= \left(\mathcal{F}_{lin}(\hat{\boldsymbol{\sigma}}, \hat{\mathbf{u}}; \mathbf{0})\right)^{\frac{1}{2}} \left(\mathcal{F}_{lin}(\boldsymbol{\tau}, \mathbf{v}; \mathbf{0})\right)^{\frac{1}{2}} \lesssim \|(\hat{\boldsymbol{\sigma}}, \hat{\mathbf{u}})\|_{\mathcal{V}} \|(\boldsymbol{\tau}, \mathbf{v})\|_{\mathcal{V}}. \end{aligned}$$

Coercivity of a:
For $(\boldsymbol{\tau}, \mathbf{v}) \in \mathcal{V}$ it holds due to (3.7) and (3.9)

$$\begin{aligned} a((\boldsymbol{\tau}, \mathbf{v}), (\boldsymbol{\tau}, \mathbf{v})) &= \left(\mathcal{L}(\boldsymbol{\tau}, \mathbf{v}), \mathcal{L}(\boldsymbol{\tau}, \mathbf{v})\right)_{L^2(\Omega)} = \|\mathcal{L}(\boldsymbol{\tau}, \mathbf{v})\|^2_{L^2(\Omega)} \\ &= \mathcal{F}_{lin}(\boldsymbol{\tau}, \mathbf{v}; \mathbf{0}) \gtrsim \|(\boldsymbol{\tau}, \mathbf{v})\|^2_{\mathcal{V}}. \end{aligned}$$

Continuity of F:
By the assumption it is clear that $\mathbf{r} - \mathcal{L}\left(\boldsymbol{\sigma}^N, \mathbf{u}_D\right) \in L^2(\Omega)^3 \times L^2(\Omega)^{3\times 3}$. For $(\boldsymbol{\tau}, \mathbf{v}) \in \mathcal{V}$ it holds due to (3.7) and again the Cauchy - Schwarz inequality and (3.9)

$$|F((\boldsymbol{\tau}, \mathbf{v}))| \leq \|\mathbf{r} - \mathcal{L}\left(\boldsymbol{\sigma}^N, \mathbf{u}_D\right)\|_{L^2(\Omega)} \|\mathcal{L}(\boldsymbol{\tau}, \mathbf{v})\|_{L^2(\Omega)} \lesssim \left(\mathcal{F}_{lin}(\boldsymbol{\tau}, \mathbf{v}; \mathbf{0})\right)^{\frac{1}{2}} \lesssim \|(\boldsymbol{\tau}, \mathbf{v})\|_{\mathcal{V}}.$$

□

An immediate consequence of this proposition with the help of Lax - Milgram (see Theorem

2.7.7 in [BS08]) is that a unique solution of the variational problem in (3.10) exists. For the sake of completeness we mention that the variational problem in (3.10) is actually equivalent to the minimization problem of $\mathcal{F}_{lin}(\boldsymbol{\sigma}, \mathbf{u}; \mathbf{f})$, since for a solution $(\hat{\boldsymbol{\sigma}}, \hat{\mathbf{u}}) \in \mathcal{V}$ of $a\left((\hat{\boldsymbol{\sigma}}, \hat{\mathbf{u}}), (\boldsymbol{\tau}, \mathbf{v})\right) = F\left((\boldsymbol{\tau}, \mathbf{v})\right)$ for all $(\hat{\boldsymbol{\tau}}, \hat{\mathbf{v}}) \in \mathcal{V}$ it holds

$$\begin{aligned}\mathcal{F}_{lin}(\boldsymbol{\sigma}+\boldsymbol{\tau}, \mathbf{u}+\mathbf{v}; \mathbf{f}) &= \|\mathcal{L}(\boldsymbol{\sigma}+\boldsymbol{\tau}, \mathbf{u}+\mathbf{v}) - \mathbf{r}\|^2_{L^2\Omega)} = \|\mathcal{L}(\boldsymbol{\sigma}, \mathbf{u}) - \mathbf{r} + \mathcal{L}(\boldsymbol{\tau}, \mathbf{v})\|^2_{L^2\Omega)} \\ &= \|\mathcal{L}(\boldsymbol{\sigma}, \mathbf{u}) - \mathbf{r}\|^2_{L^2\Omega)} + 2\underbrace{(\mathcal{L}(\boldsymbol{\sigma}, \mathbf{u}) - \mathbf{r}, \mathcal{L}(\boldsymbol{\tau}, \mathbf{v}))_{L^2(\Omega)}}_{=0} + \underbrace{\|\mathcal{L}(\boldsymbol{\tau}, \mathbf{v})\|^2_{L^2\Omega)}}_{\geq 0} \\ &\geq \|\mathcal{L}(\boldsymbol{\sigma}, \mathbf{u}) - \mathbf{r}\|^2_{L^2\Omega)} = \mathcal{F}_{lin}(\boldsymbol{\sigma}, \mathbf{u}; \mathbf{f})\end{aligned}$$

with $\boldsymbol{\sigma} = \boldsymbol{\sigma}^N + \hat{\boldsymbol{\sigma}}$ and $\mathbf{u} = \mathbf{u}_D + \hat{\mathbf{u}}$ and arbitrary $(\boldsymbol{\tau}, \mathbf{v}) \in \mathcal{V}$.
With the help of the property (3.7) we can also show that the least squares functional is equivalent to the error, i.e. the least squares functional is a suitable a - posteriori error estimator and can be used for adaptive refinement.

Corollary 3.3: (Error estimator in linear elasticity)
Let $(\boldsymbol{\sigma}, \mathbf{u})$ be the exact solution of (3.4) and $(\boldsymbol{\tau}, \mathbf{v}) \in H(\mathrm{div}; \Omega)^3 \times H^1(\Omega)^3$, satisfying $\boldsymbol{\tau} \cdot \mathbf{n} = \mathbf{g}$ on Γ_N and $\mathbf{v} = \mathbf{u}_D$ on Γ_D. Then it holds

$$\|(\boldsymbol{\sigma} - \boldsymbol{\tau}, \mathbf{u} - \mathbf{v})\|^2_{\mathcal{V}} \lesssim \mathcal{F}_{lin}(\boldsymbol{\tau}, \mathbf{v}; \mathbf{f}) \lesssim \|(\boldsymbol{\sigma} - \boldsymbol{\tau}, \mathbf{u} - \mathbf{v})\|^2_{\mathcal{V}}. \tag{3.11}$$

Proof:

By assumption is $(\boldsymbol{\sigma}, \mathbf{u})$ the exact solution of (3.4) and therefore it holds $\mathcal{L}(\boldsymbol{\sigma}, \mathbf{u}) = \mathbf{r}$. This implies due to the linearity of $\mathcal{L}$ and the definition of $\mathcal{F}_{lin}$

$$\begin{aligned}\mathcal{F}_{lin}(\boldsymbol{\tau}, \mathbf{v}; \mathbf{f}) &= \|\mathcal{L}(\boldsymbol{\tau}, \mathbf{v}) - \mathbf{r}\|^2_{L^2(\Omega)} = \|\mathcal{L}(\boldsymbol{\tau}, \mathbf{v}) - \mathcal{L}(\boldsymbol{\sigma}, \mathbf{u})\|^2_{L^2(\Omega)} \\ &= \|\mathcal{L}(\boldsymbol{\sigma} - \boldsymbol{\tau}, \mathbf{u} - \mathbf{v})\|^2_{L^2(\Omega)} = \mathcal{F}_{lin}(\boldsymbol{\sigma} - \boldsymbol{\tau}, \mathbf{u} - \mathbf{v}; \mathbf{0})\end{aligned}$$

with $(\boldsymbol{\sigma} - \boldsymbol{\tau}) \cdot \mathbf{n} = \boldsymbol{\sigma} \cdot \mathbf{n} - \boldsymbol{\tau} \cdot \mathbf{n} = \mathbf{g} - \mathbf{g} = \mathbf{0}$ on Γ_N and $\mathbf{u} - \mathbf{v} = \mathbf{u}_D - \mathbf{u}_D = \mathbf{0}$ on Γ_D. Thus we can apply (3.7) and obtain the statement.

□

An immediate consequence for $\boldsymbol{\tau} = \boldsymbol{\sigma}_h$ and $\mathbf{v} = \mathbf{u}_h$ with a conforming approximation $(\boldsymbol{\sigma}_h, \mathbf{u}_h) \in H(\mathrm{div}; \Omega)^3 \times H^1(\Omega)^3$, satisfying $\boldsymbol{\sigma}_h \cdot \mathbf{n} = \mathbf{g}$ on Γ_N and $\mathbf{u}_h = \mathbf{u}_D$ on Γ_D, and the error $\mathbf{e} := (\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{u} - \mathbf{u}_h)$ is

$$\|\mathbf{e}\|^2_{\mathcal{V}} = \|(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{u} - \mathbf{u}_h)\|^2_{\mathcal{V}} \eqsim \mathcal{F}_{lin}(\boldsymbol{\sigma}_h, \mathbf{u}_h; \mathbf{f}). \tag{3.12}$$

The sign $\eqsim$ in (3.12) is an abbreviation for $\|\mathbf{e}\|^2_{\mathcal{V}} \lesssim \mathcal{F}_{lin}(\boldsymbol{\sigma}_h, \mathbf{u}_h; \mathbf{f}) \lesssim \|\mathbf{e}\|^2_{\mathcal{V}}$. The abbreviation $\eqsim$ will be used in the rest of this work in the same way.
(3.12) means that the least squares functional, evaluated in the approximation, is up to constants a reliable and efficient measure for the error $\mathbf{e}$ and can be used for adaptive refinement.

Furthermore if one uses for instance Raviart - Thomas elements $(\mathcal{RT}_{k-1}(\mathcal{T}_h))^3$ for $\boldsymbol{\sigma}_h$ and continuous elements $(\mathcal{P}_k(\mathcal{T}_h))^3$ for $\mathbf{u}_h$ with an integer $k \geq 1$, we get from the approximation error estimates in Propositions 2.44 and 2.46

$$\begin{aligned} \|\mathbf{e}\|_{\mathcal{V}} &= \left(\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|^2_{H(\mathrm{div};\,\Omega)} + \|\mathbf{u} - \mathbf{u}_h\|^2_{H^1(\Omega)} \right)^{\frac{1}{2}} \\ &\leq Ch^k \left(\|\boldsymbol{\sigma}\|^2_{H^k(\Omega)} + \|\mathrm{div}\,\boldsymbol{\sigma}\|^2_{H^k(\Omega)} + \|\mathbf{u}\|^2_{H^{k+1}(\Omega)} \right)^{\frac{1}{2}}, \end{aligned} \tag{3.13}$$

assuming that $\mathbf{u} \in H^{k+1}(\Omega)^3$, $\boldsymbol{\sigma} \in H(\mathrm{div};\Omega)^3 \cap H^k(\Omega)^{3\times 3} \cap L^r(\Omega)^{3\times 3}$ (with fixed $r > 2$ for the interpolation operator Π_h defined in Proposition 2.46) and $\mathrm{div}\,\boldsymbol{\sigma} \in H^k(\Omega)^3$. Equation (3.13) is an a - priori estimate for the error.

3.3 Extension to homogeneous isotropic hyperelastic models

Our aim in this section is to generalize the idea described in Section 3.2 to nonlinear homogeneous isotropic frame - indifferent hyperelastic materials. The point of departure is the first - order system (3.1) with stored energy function $\psi(\mathbf{C})$, now homogeneous and isotropic. We have seen in equation (2.28) and (2.27) (respectively (2.29)) that we can express the stress tensors $\boldsymbol{\tau} = \mathbf{P}\mathbf{F}^T$ and $\boldsymbol{\Sigma} = \mathbf{F}^{-1}\mathbf{P}$ in terms of $\mathbf{B} = \mathbf{F}\mathbf{F}^T$ and $\mathbf{C} = \mathbf{F}^T\mathbf{F}$, i.e. there exist mappings $\mathcal{G}, \tilde{\mathcal{G}} : \mathbb{R}^{3\times 3} \to \mathbb{R}^{3\times 3}$ with

$$\mathbf{P}\mathbf{F}^T = \mathcal{G}(\mathbf{B}) \text{ and } \mathbf{F}^{-1}\mathbf{P} = \tilde{\mathcal{G}}(\mathbf{C}). \tag{3.14}$$

These are mappings from strains into stresses similar to the mapping $\mathcal{C}$ in (3.2). Following the idea of Section 3.2, we want to invert these equations in order to obtain mappings from stresses to strains. However, since the mappings $\mathcal{G}, \tilde{\mathcal{G}}$ are nonlinear in the strains, this is in general impossible. The idea is now that the stress - strain relations are at least invertible in a neighborhood of $\mathbf{B} = \mathbf{I} = \mathbf{C}$. Assuming that we have eliminated all rigid body motions $\boldsymbol{\varphi} \neq \mathbf{id}$ (cf. Section 2.2.2), this condition is only possible if and only if $\boldsymbol{\varphi} = \mathbf{id} \Leftrightarrow \mathbf{u} = \mathbf{0}$. If we consider (3.14) in terms of the displacement $\mathbf{u}$, i.e.

$$\mathbf{P}(\mathbf{u})(\mathbf{F}(\mathbf{u}))^T = \mathcal{G}(\mathbf{B}(\mathbf{u})) \text{ and } \mathbf{F}(\mathbf{u})^{-1}\mathbf{P}(\mathbf{u}) = \tilde{\mathcal{G}}(\mathbf{C}(\mathbf{u})),$$

and assume that the material is consistent with linear elasticity (cf. Section 2.4.5) we get

$$\mathcal{G}(\mathbf{I}) = \mathcal{G}(\mathbf{B}(\mathbf{0})) = \mathbf{P}(\mathbf{0})(\mathbf{F}(\mathbf{0}))^T = \mathbf{0}, \quad \tilde{\mathcal{G}}(\mathbf{I}) = \tilde{\mathcal{G}}(\mathbf{C}(\mathbf{0})) = \mathbf{F}(\mathbf{0})^{-1}\mathbf{P}(\mathbf{0}) = \mathbf{0},$$

since $\mathbf{P}(\mathbf{0}) = \mathbf{0}$. We assume that $\mathcal{G}$ and $\tilde{\mathcal{G}}$ are continuously differentiable. Then we get as Fréchet/Gâteaux derivatives

$$\begin{aligned} &\mathcal{G}'(\mathbf{B}(\mathbf{u}))[\nabla\mathbf{v}(\mathbf{F}(\mathbf{u}))^T + \mathbf{F}(\mathbf{u})(\nabla\mathbf{v})^T] = \mathcal{G}'(\mathbf{B}(\mathbf{u}))[\mathbf{B}'(\mathbf{u})[\mathbf{v}]] = (\mathcal{G}(\mathbf{B}(\mathbf{u})))'[\mathbf{v}] \\ &\quad \overset{!}{=} \left(\mathbf{P}(\mathbf{u})(\mathbf{F}(\mathbf{u}))^T\right)'[\mathbf{v}] = \mathbf{P}'(\mathbf{u})[\mathbf{v}](\mathbf{F}(\mathbf{u}))^T + \mathbf{P}(\mathbf{u})(\nabla\mathbf{v})^T, \\ &\tilde{\mathcal{G}}'(\mathbf{C}(\mathbf{u}))[(\nabla\mathbf{v})^T\mathbf{F}(\mathbf{u}) + (\mathbf{F}(\mathbf{u}))^T\nabla\mathbf{v}] = \tilde{\mathcal{G}}'(\mathbf{C}(\mathbf{u}))[\mathbf{C}'(\mathbf{u})[\mathbf{v}]] = (\tilde{\mathcal{G}}(\mathbf{C}(\mathbf{u})))'[\mathbf{v}] \\ &\quad \overset{!}{=} \left(\mathbf{F}(\mathbf{u})^{-1}\mathbf{P}(\mathbf{u})\right)'[\mathbf{v}] = (-\mathbf{F}(\mathbf{u})^{-1}\nabla\mathbf{v}\mathbf{F}(\mathbf{u})^{-1})\mathbf{P}(\mathbf{u}) + \mathbf{F}(\mathbf{u})^{-1}\mathbf{P}'(\mathbf{u})[\mathbf{v}]. \end{aligned}$$

For $\mathbf{u} = \mathbf{0}$ it follows $\mathbf{B}(\mathbf{u}) = \mathbf{I} = \mathbf{C}(\mathbf{u})$, $\nabla\mathbf{v}(\mathbf{F}(\mathbf{u}))^T + \mathbf{F}(\mathbf{u})(\nabla\mathbf{v})^T = 2\boldsymbol{\varepsilon}(\mathbf{v}) = (\nabla\mathbf{v})^T\mathbf{F}(\mathbf{u}) + (\mathbf{F}(\mathbf{u}))^T\nabla\mathbf{v}$ and therefore by the assumed consistency with linear elasticity

$$\mathcal{G}'(\mathbf{I})[2\boldsymbol{\varepsilon}(\mathbf{v})] = \tilde{\mathcal{G}}'(\mathbf{I})[2\boldsymbol{\varepsilon}(\mathbf{v})] = \mathbf{P}'(\mathbf{0})[\mathbf{v}] = 2\mu\,\boldsymbol{\varepsilon}(\mathbf{v}) + \lambda\,\mathrm{tr}(\boldsymbol{\varepsilon}(\mathbf{v}))\mathbf{I}$$

for all $\mathbf{v}$ in a neighborhood of $\mathbf{u} = \mathbf{0}$. Since $\mathbf{C}(\mathbf{v}) - \mathbf{I} = \mathbf{C}(\mathbf{v}) - \mathbf{C}(\mathbf{0}) = 2\,\mathbf{E}(\mathbf{v}) \approx 2\,\boldsymbol{\varepsilon}(\mathbf{v})$ the equation

$$\mathcal{G}'(\mathbf{I})[\mathbf{E}] = \mu\,\mathbf{E} + \frac{\lambda}{2}\,\mathrm{tr}(\mathbf{E})\mathbf{I} = \tilde{\mathcal{G}}'(\mathbf{I})[\mathbf{E}]$$

is reasonable for the Green - St. Venant strain tensor $\mathbf{E}$ in a neighborhood of $\mathbf{E} = \mathbf{0}$ (respectively for the Cauchy - Green strain tensors $\mathbf{B}$ and $\mathbf{C}$ in a neighborhood of $\mathbf{I}$). Thus for small strains $\mathbf{E}$ and by definition of $\mathcal{C}$ in (3.2)

$$\mathcal{G}'(\mathbf{I})[\mathbf{E}] = \tilde{\mathcal{G}}'(\mathbf{I})[\mathbf{E}] = \frac{1}{2}\mathcal{C}\,\mathbf{E} = \mu\,\mathbf{E} + \frac{\lambda}{2}\,\mathrm{tr}(\mathbf{E})\mathbf{I} \tag{3.15}$$

is motivated.

If (3.15) holds and the mappings $\mathcal{G}$ and $\tilde{\mathcal{G}}$ are continuously differentiable it remains to show by Theorem 2.11 that $\mathcal{C}$ is an isomorphism. Then local invertibility of $\mathcal{G}$ and $\tilde{\mathcal{G}}$ is ensured. Obviously $\mathcal{C}$ and its inverse $\mathcal{C}^{-1}$, defined in (3.3), are linear. Furthermore for finite λ and μ the mappings $\mathcal{C}$ and $\mathcal{C}^{-1}$ are continuous since with $|\mathbf{I}| = \sqrt{3}$ and Lemma 2.31 it holds

$$\begin{aligned}
|\mathcal{C}\,\mathbf{E}| &= |2\mu\mathbf{E} + \lambda\mathrm{tr}(\mathbf{E})\mathbf{I}| \le 2\mu|\mathbf{E}| + \lambda|\mathrm{tr}(\mathbf{E})||\mathbf{I}| \le (2\mu + 3\lambda)|\mathbf{E}|, \quad \mathbf{E} \in \mathbb{R}^{3\times 3},\\
|\mathcal{C}^{-1}\boldsymbol{\sigma}| &= \frac{1}{2\mu}\left|\boldsymbol{\sigma} - \frac{\lambda}{3\lambda + 2\mu}\mathrm{tr}(\boldsymbol{\sigma})\mathbf{I}\right| \le \frac{1}{2\mu}\left(1 + \frac{3\lambda}{3\lambda + 2\mu}\right)|\boldsymbol{\sigma}|\\
&= \frac{1}{\mu}\left(\frac{3\lambda + \mu}{3\lambda + 2\mu}\right)|\boldsymbol{\sigma}| = \frac{1}{\mu}\left(\frac{3 + \frac{\mu}{\lambda}}{3 + \frac{2\mu}{\lambda}}\right)|\boldsymbol{\sigma}|, \quad \boldsymbol{\sigma} \in \mathbb{R}^{3\times 3}.
\end{aligned}$$

Thus $\mathcal{C}$ and therefore also $\partial\mathcal{G}(\mathbf{I}) = \frac{1}{2}\mathcal{C}$ and $\partial\tilde{\mathcal{G}}(\mathbf{I}) = \frac{1}{2}\mathcal{C}$ are isomorphisms. The consequence of the local inversion theorem is that the inverse mappings $\mathcal{G}^{-1}(\boldsymbol{\tau})$ and $\tilde{\mathcal{G}}^{-1}(\boldsymbol{\Sigma})$ are well - defined in a neighborhood of $\boldsymbol{\tau} = \mathbf{0} = \boldsymbol{\Sigma}$, i.e. at least for small stresses. We can therefore find, similar to the linear case, two first - order systems. On the one hand we get

$$\boxed{\begin{aligned}
\mathrm{div}\,\mathbf{P} + \mathbf{f} &= \mathbf{0} \quad \text{in } \Omega,\\
\mathcal{G}^{-1}(\mathbf{P}\mathbf{F}(\mathbf{u})^T) - \mathbf{B}(\mathbf{u}) &= \mathbf{0} \quad \text{in } \Omega,\\
\mathbf{u} = \mathbf{u}_D \text{ on } \Gamma_D, &\quad \mathbf{P}\cdot\mathbf{n} = \mathbf{g} \text{ on } \Gamma_N,
\end{aligned}} \tag{3.16}$$

using the representation in $\mathbf{B}$, and on the other hand we have

$$\boxed{\begin{aligned}
\mathrm{div}\,\mathbf{P} + \mathbf{f} &= \mathbf{0} \quad \text{in } \Omega,\\
\tilde{\mathcal{G}}^{-1}(\mathbf{F}(\mathbf{u})^{-1}\mathbf{P}) - \mathbf{C}(\mathbf{u}) &= \mathbf{0} \quad \text{in } \Omega,\\
\mathbf{u} = \mathbf{u}_D \text{ on } \Gamma_D, &\quad \mathbf{P}\cdot\mathbf{n} = \mathbf{g} \text{ on } \Gamma_N
\end{aligned}} \tag{3.17}$$

using the representation in $\mathbf{C}$.

3.3.1 General least squares formulations for hyperelastic materials

We have observed in Section 3.2, that the operator $\mathcal{A}_{lin} = \mathcal{C}^{-1}$ of linear elasticity is indeed well - defined in the incompressible limit but is no longer invertible in this case. The linearization of $\mathcal{G}$ and $\tilde{\mathcal{G}}$ about the matrix $\mathbf{I}$ is by construction up to a constant similar to $\mathcal{C}$. Thus also for the inverses $\mathcal{G}^{-1}$ and $\tilde{\mathcal{G}}^{-1}$ we expect a similar behavior as for $\mathcal{C}^{-1}$ in the incompressible limit. Due to this observation we use in (3.16) and (3.17) instead of $\mathcal{G}^{-1}$ and $\tilde{\mathcal{G}}^{-1}$ the notation $\mathcal{A}$ and $\tilde{\mathcal{A}}$ in the following. For finite λ we set $\mathcal{A} = \mathcal{G}^{-1}$ and $\tilde{\mathcal{A}} = \tilde{\mathcal{G}}^{-1}$. One question that arises is if the operators $\mathcal{A}$ and $\tilde{\mathcal{A}}$ are also well - defined in the incompressible limit $\lambda \to \infty$. For the case of a special Neo - Hooke material we will answer this question in Section 3.4. But let us first define general least squares functionals in hyperelasticity based on (3.16) and (3.17) using the notation $\mathcal{A}$, $\tilde{\mathcal{A}}$ instead of $\mathcal{G}^{-1}$, $\tilde{\mathcal{G}}^{-1}$. We follow the same idea as for linear elasticity.
For this purpose let $\mathbf{P} = \mathbf{P}^N + \hat{\mathbf{P}} \in W^q(\mathrm{div};\Omega)^3 + W^q_{\Gamma_N}(\mathrm{div};\Omega)^3$ (with $\mathbf{P}^N \cdot \mathbf{n} = \mathbf{g}$ on Γ_N), $\mathbf{u} = \mathbf{u}_D + \hat{\mathbf{u}} \in W^{1,p}(\Omega)^3 + W^{1,p}_{\Gamma_D}(\Omega)^3$ and $\mathbf{f} \in L^q(\Omega)^3$ for sufficiently large q and p. For such pairs $(\mathbf{P}, \mathbf{u})$ we define the nonlinear operators

$$\mathcal{R}(\mathbf{P},\mathbf{u}) := \begin{pmatrix} \omega_1 \left(\mathrm{div}\, \mathbf{P} + \mathbf{f}\right) \\ \omega_2 \left(\mathcal{A}(\mathbf{P}\mathbf{F}(\mathbf{u})^T) - \mathbf{B}(\mathbf{u})\right) \end{pmatrix}, \quad \mathcal{R}(\mathbf{P},\mathbf{u}) := \begin{pmatrix} \omega_1 \left(\mathrm{div}\, \mathbf{P} + \mathbf{f}\right) \\ \omega_2 \left(\tilde{\mathcal{A}}(\mathbf{F}(\mathbf{u})^{-1}\mathbf{P}) - \mathbf{C}(\mathbf{u})\right) \end{pmatrix} \tag{3.18}$$

for (3.16) (respectively for (3.17)) with scaling parameters $\omega_1, \omega_2 > 0$. We define general nonlinear least squares functionals as

$$\boxed{\begin{aligned} \mathcal{F}(\mathbf{P},\mathbf{u}) &:= \|\mathcal{R}(\mathbf{P},\mathbf{u})\|^2_{L^2(\Omega)} \\ &= \begin{cases} \omega_1^2 \|\mathrm{div}\, \mathbf{P} + \mathbf{f}\|^2_{L^2(\Omega)} + \omega_2^2 \|\mathcal{A}(\mathbf{P}\mathbf{F}(\mathbf{u})^T) - \mathbf{B}(\mathbf{u})\|^2_{L^2(\Omega)} \\ \omega_1^2 \|\mathrm{div}\, \mathbf{P} + \mathbf{f}\|^2_{L^2(\Omega)} + \omega_2^2 \|\tilde{\mathcal{A}}(\mathbf{F}(\mathbf{u})^{-1}\mathbf{P}) - \mathbf{C}(\mathbf{u})\|^2_{L^2(\Omega)}. \end{cases} \end{aligned}} \tag{3.19}$$

We call the first case inverse $\mathbf{B}$ - and the second case inverse $\mathbf{C}$ - formulation. The aim is again to find a minimizer of $\mathcal{F}(\mathbf{P},\mathbf{u})$, since the exact solution of the problem satisfies $\mathcal{F}(\mathbf{P},\mathbf{u}) = 0$.
The value of q and p has to be chosen sufficiently large such that $\mathcal{R}(\mathbf{P},\mathbf{u}) \in L^2(\Omega)^3 \times L^2(\Omega)^{3\times 3}$ is ensured. Since the strain tensors $\mathbf{B}(\mathbf{u}) = \mathbf{F}(\mathbf{u})(\mathbf{F}(\mathbf{u}))^T$ (respectively $\mathbf{C}(\mathbf{u}) = (\mathbf{F}(\mathbf{u}))^T\mathbf{F}(\mathbf{u})$) are involved it must at least hold $p \geq 4$. Since we are dealing with nonlinear problems $q > 2$ is additionally a reasonable assumption. We will specify the required values of q and p in the case of a Neo - Hooke material for the $\mathbf{B}$ - formulation in Section 3.5.1.
Furthermore, it would be desirable to prove

$$\mathcal{F}(\mathbf{P}_h,\mathbf{u}_h) = \|\mathcal{R}(\mathbf{P}_h,\mathbf{u}_h)\|^2_{L^2(\Omega)} \eqsim \|(\mathbf{P}-\mathbf{P}_h, \mathbf{u}-\mathbf{u}_h)\|^2_V, \tag{3.20}$$

similar to (3.12) in linear elasticity, for the exact solution $(\mathbf{P},\mathbf{u})$, a conforming finite element approximation $(\mathbf{P}_h,\mathbf{u}_h)$ and a suitable norm $\|\cdot\|_V$. This means that we can

estimate the error between the (unknown) exact solution and the calculated approximation by the nonlinear least squares functional. In the case of a special Neo - Hooke material we will prove this property in Section 3.5.1 with $V = \mathcal{V} = H_{\Gamma_N}(\mathrm{div};\,\Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$ and additional assumptions on $(\mathbf{P}, \mathbf{u})$ and $(\mathbf{P}_h, \mathbf{u}_h)$.

3.3.2 Linearized least squares formulation

For the minimization of (3.19) we consider a sequence of linearized problems. If we assume that the operator $\mathcal{R}(\mathbf{P}, \mathbf{u})$ is Fréchet differentiable with respect to $(\mathbf{P}, \mathbf{u})$ we can linearize this operator about a given $\left(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}\right) \in W^q(\mathrm{div};\Omega)^3 \times W^{1,p}(\Omega)^3$, satisfying $\mathbf{P}^{(k)} \cdot \mathbf{n} = \mathbf{g}$ on Γ_N and $\mathbf{u}^{(k)} = \mathbf{u}_D$ on Γ_D. We define the linearized least squares functional as

$$\mathcal{F}^{\mathrm{lin}}(\mathbf{Q}, \mathbf{v}) := \mathcal{F}^{\mathrm{lin}}(\mathbf{Q}, \mathbf{v}; \mathcal{R}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})) := \|\mathcal{R}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) + \mathcal{R}'(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{Q}, \mathbf{v}]\|^2_{L^2(\Omega)} \tag{3.21}$$

and seek the minimizer $(\mathbf{Q}, \mathbf{v})$ with zero boundary conditions in a suitable normed function space $\mathbf{\Pi}_{\Gamma_N} \times \mathbf{U}_{\Gamma_D}$, equipped with norm $\|\cdot\|_{\mathbf{\Pi} \times \mathbf{U}}$. Unfortunately, one needs in general more regularity for the linearized problem (3.21) as for the nonlinear problem (3.19) to ensure that also the derivative $\mathcal{R}'(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{Q}, \mathbf{v}]$ is in $L^2(\Omega)^3 \times L^2(\Omega)^{3\times 3}$ and therefore (3.21) exists. Similar to the derivation in (3.10) the necessary condition for a minimum of (3.21) is $\frac{d}{dt}\mathcal{F}^{\mathrm{lin}}(\mathbf{Q} + t\hat{\mathbf{Q}}, \mathbf{v} + t\hat{\mathbf{v}})\big|_{t=0} = 0$. We define a bilinear form and a linear form through

$$\begin{aligned} a((\mathbf{Q}, \mathbf{v}), (\hat{\mathbf{Q}}, \hat{\mathbf{v}})) &:= \left(\mathcal{R}'(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{Q}, \mathbf{v}], \mathcal{R}'(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\hat{\mathbf{Q}}, \hat{\mathbf{v}}]\right)_{L^2(\Omega)} \\ F((\hat{\mathbf{Q}}, \hat{\mathbf{v}})) &:= -\left(\mathcal{R}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}), \mathcal{R}'(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\hat{\mathbf{Q}}, \hat{\mathbf{v}}]\right)_{L^2(\Omega)} \end{aligned}$$

for all $(\hat{\mathbf{Q}}, \hat{\mathbf{v}}) \in \mathbf{\Pi}_{\Gamma_N} \times \mathbf{U}_{\Gamma_D}$. Then the corresponding variational problem to the minimization problem (3.21) is:

$$\boxed{\begin{aligned} &\text{Find } (\mathbf{Q}, \mathbf{v}) \in \mathbf{\Pi}_{\Gamma_N} \times \mathbf{U}_{\Gamma_D} \text{ with} \\ &\qquad a((\mathbf{Q}, \mathbf{v}), (\hat{\mathbf{Q}}, \hat{\mathbf{v}})) = F((\hat{\mathbf{Q}}, \hat{\mathbf{v}})) \\ &\text{for all } (\hat{\mathbf{Q}}, \hat{\mathbf{v}}) \in \mathbf{\Pi}_{\Gamma_N} \times \mathbf{U}_{\Gamma_D}. \end{aligned}} \tag{3.22}$$

The next lemma proves that the problems (3.21) and (3.22) are even equivalent.

Lemma 3.4:

Let $(\mathbf{Q}, \mathbf{v}) \in \mathbf{\Pi}_{\Gamma_N} \times \mathbf{U}_{\Gamma_D}$ the solution of (3.22). Then $(\mathbf{Q}, \mathbf{v})$ is also the minimizer of (3.21).

Proof:

By assumption it holds for the solution $(\mathbf{Q}, \mathbf{v}) \in \mathbf{\Pi}_{\Gamma_N} \times \mathbf{U}_{\Gamma_D}$ and arbitrary $(\hat{\mathbf{Q}}, \hat{\mathbf{v}}) \in \mathbf{\Pi}_{\Gamma_N} \times \mathbf{U}_{\Gamma_D}$

$$\begin{aligned} &a((\mathbf{Q}, \mathbf{v}), (\hat{\mathbf{Q}}, \hat{\mathbf{v}})) - F((\hat{\mathbf{Q}}, \hat{\mathbf{v}})) = 0 \\ \Leftrightarrow &\left(\mathcal{R}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) + \mathcal{R}'(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{Q}, \mathbf{v}], \mathcal{R}'(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\hat{\mathbf{Q}}, \hat{\mathbf{v}}]\right)_{L^2(\Omega)} = 0. \end{aligned}$$

With this property we get

$$\begin{aligned}
\mathcal{F}^{\text{lin}}(\mathbf{Q}+\hat{\mathbf{Q}},\mathbf{v}+\hat{\mathbf{v}}) &= \|\mathcal{R}(\mathbf{P}^{(k)},\mathbf{u}^{(k)})+\mathcal{R}'(\mathbf{P}^{(k)},\mathbf{u}^{(k)})[\mathbf{Q}+\hat{\mathbf{Q}},\mathbf{v}+\hat{\mathbf{v}}]\|^2_{L^2(\Omega)} \\
&= \|\mathcal{R}(\mathbf{P}^{(k)},\mathbf{u}^{(k)})+\mathcal{R}'(\mathbf{P}^{(k)},\mathbf{u}^{(k)})[\mathbf{Q},\mathbf{v}]+\mathcal{R}'(\mathbf{P}^{(k)},\mathbf{u}^{(k)})[\hat{\mathbf{Q}},\hat{\mathbf{v}}]\|^2_{L^2(\Omega)} \\
&= \|\mathcal{R}(\mathbf{P}^{(k)},\mathbf{u}^{(k)})+\mathcal{R}'(\mathbf{P}^{(k)},\mathbf{u}^{(k)})[\mathbf{Q},\mathbf{v}]\|^2_{L^2(\Omega)} + \underbrace{\|\mathcal{R}'(\mathbf{P}^{(k)},\mathbf{u}^{(k)})[\hat{\mathbf{Q}},\hat{\mathbf{v}}]\|^2_{L^2(\Omega)}}_{\geq 0} \\
&\quad + \underbrace{2\left(\mathcal{R}(\mathbf{P}^{(k)},\mathbf{u}^{(k)})+\mathcal{R}'(\mathbf{P}^{(k)},\mathbf{u}^{(k)})[\mathbf{Q},\mathbf{v}],\mathcal{R}'(\mathbf{P}^{(k)},\mathbf{u}^{(k)})[\hat{\mathbf{Q}},\hat{\mathbf{v}}]\right)_{L^2(\Omega)}}_{=0} \\
&\geq \|\mathcal{R}(\mathbf{P}^{(k)},\mathbf{u}^{(k)})+\mathcal{R}'(\mathbf{P}^{(k)},\mathbf{u}^{(k)})[\mathbf{Q},\mathbf{v}]\|^2_{L^2(\Omega)} = \mathcal{F}^{\text{lin}}(\mathbf{Q},\mathbf{v})
\end{aligned}$$

for all $(\hat{\mathbf{Q}},\hat{\mathbf{v}}) \in \mathbf{\Pi}_{\Gamma_N} \times \mathbf{U}_{\Gamma_D}$, i.e. $(\mathbf{Q},\mathbf{v})$ is a minimizer of (3.21).

□

Similar to the property (3.7) in linear elasticity one is generally interested in a property

$$\mathcal{F}^{\text{lin}}(\hat{\mathbf{Q}},\hat{\mathbf{v}};\mathbf{0}) = \|\mathcal{R}'(\mathbf{P}^{(k)},\mathbf{u}^{(k)})[\hat{\mathbf{Q}},\hat{\mathbf{v}}]\|^2_{L^2(\Omega)} \eqsim \|(\hat{\mathbf{Q}},\hat{\mathbf{v}})\|^2_{\mathbf{\Pi}\times\mathbf{U}}, \tag{3.23}$$

$(\hat{\mathbf{Q}},\hat{\mathbf{v}}) \in \mathbf{\Pi}_{\Gamma_N} \times \mathbf{U}_{\Gamma_D}$, for the linearized least squares problem, since from this property follows the well - posedness of the variational problem (3.22) and therefore a unique solution $(\mathbf{Q},\mathbf{v}) \in \mathbf{\Pi}_{\Gamma_N} \times \mathbf{U}_{\Gamma_D}$. The proof for the existence of a unique solution can be done in the same way as in Section 3.2 using the linear operator $\mathcal{R}'(\mathbf{P}^{(k)},\mathbf{u}^{(k)})$ instead of $\mathcal{L}$.
Furthermore if $(\mathbf{Q},\mathbf{v}) \in \mathbf{\Pi}_{\Gamma_N}\times\mathbf{U}_{\Gamma_D}$ is the exact solution of (3.21), i.e. it holds $\mathcal{F}^{\text{lin}}(\mathbf{Q},\mathbf{v}) = 0$, and the property (3.23) is satisfied, then it holds for arbitrary $(\hat{\mathbf{Q}},\hat{\mathbf{v}}) \in \mathbf{\Pi}_{\Gamma_N} \times \mathbf{U}_{\Gamma_D}$

$$\begin{aligned}
\mathcal{F}^{\text{lin}}(\hat{\mathbf{Q}},\hat{\mathbf{v}}) &= \mathcal{F}^{\text{lin}}(\hat{\mathbf{Q}},\hat{\mathbf{v}};\mathcal{R}(\mathbf{P}^{(k)},\mathbf{u}^{(k)})) = \|\mathcal{R}(\mathbf{P}^{(k)},\mathbf{u}^{(k)})+\mathcal{R}'(\mathbf{P}^{(k)},\mathbf{u}^{(k)})[\hat{\mathbf{Q}},\hat{\mathbf{v}}]\|^2_{L^2(\Omega)} \\
&= \|-\mathcal{R}'(\mathbf{P}^{(k)},\mathbf{u}^{(k)})[\mathbf{Q},\mathbf{v}]+\mathcal{R}'(\mathbf{P}^{(k)},\mathbf{u}^{(k)})[\hat{\mathbf{Q}},\hat{\mathbf{v}}]\|^2_{L^2(\Omega)} \\
&= \|\mathcal{R}'(\mathbf{P}^{(k)},\mathbf{u}^{(k)})[\mathbf{Q}-\hat{\mathbf{Q}},\mathbf{v}-\hat{\mathbf{v}}]\|^2_{L^2(\Omega)} \\
&= \mathcal{F}^{\text{lin}}(\mathbf{Q}-\hat{\mathbf{Q}},\mathbf{v}-\hat{\mathbf{v}};\mathbf{0}) \\
&\eqsim \|(\mathbf{Q}-\hat{\mathbf{Q}},\mathbf{v}-\hat{\mathbf{v}})\|^2_{\mathbf{\Pi}\times\mathbf{U}},
\end{aligned}$$

i.e. the linearized least squares functional is an efficient and reliable a - posteriori error estimator.
If we use for instance Raviart - Thomas elements $\mathbf{\Pi}^l_h := (\mathcal{RT}_{l-1}(\mathcal{T}_h))^3 \subset \mathbf{\Pi}_{\Gamma_N}$ for the approximation $\mathbf{Q}_h$ of $\mathbf{Q}$ and continuous elements $\mathbf{U}^l_h := (\mathcal{P}_l(\mathcal{T}_h))^3 \subset \mathbf{U}_{\Gamma_D}$ for the approximation $\mathbf{v}_h$ of $\mathbf{v}$ with an arbitrary integer $l \geq 1$ and $(\mathbf{Q}_h,\mathbf{v}_h)$ is the unique minimizer of $\mathcal{F}^{\text{lin}}(\mathbf{R}_h,\mathbf{w}_h)$ about all $(\mathbf{R}_h,\mathbf{w}_h) \in \mathbf{\Pi}^l_h \times \mathbf{U}^l_h \subset \mathbf{\Pi}_{\Gamma_N} \times \mathbf{U}_{\Gamma_D}$, we get the a - priori estimate

$$\begin{aligned}
\left(\mathcal{F}^{\text{lin}}(\mathbf{Q}_h,\mathbf{v}_h)\right)^{\frac{1}{2}} &= \inf_{(\mathbf{R}_h,\mathbf{w}_h)\in\mathbf{\Pi}^l_h\times\mathbf{U}^l_h}\left(\mathcal{F}^{\text{lin}}(\mathbf{R}_h,\mathbf{w}_h)\right)^{\frac{1}{2}} \\
&\leq \left(\mathcal{F}^{\text{lin}}(\Pi_h\mathbf{Q},I_h\mathbf{v})\right)^{\frac{1}{2}} \lesssim ||(\mathbf{Q}-\Pi_h\mathbf{Q},\mathbf{v}-I_h\mathbf{v})||_{\mathbf{\Pi}\times\mathbf{U}} \\
&\lesssim h^l\left(\|\mathbf{Q}\|^2_{H^l(\Omega)}+\|\operatorname{div}\mathbf{Q}\|^2_{H^l(\Omega)}+\|\mathbf{v}\|^2_{H^{l+1}(\Omega)}\right)^{\frac{1}{2}}
\end{aligned}$$

with the interpolation operators Π_h, I_h defined in Section 2.5, its componentwise application and the assumptions $\mathbf{Q} \in H(\mathrm{div};\Omega)^3 \cap H^l(\Omega)^{3\times 3} \cap L^r(\Omega)^{3\times 3}$ (with fixed $r > 2$), $\mathrm{div}\,\mathbf{Q} \in H^l(\Omega)^3$ and $\mathbf{v} \in H^{l+1}(\Omega)^3$.
We will prove the property (3.23) for a Neo - Hooke material in Section 3.5.2 with $\mathbf{\Pi}_{\Gamma_N} := H_{\Gamma_N}(\mathrm{div};\,\Omega)^3$, $\mathbf{U}_{\Gamma_D} := H^1_{\Gamma_D}(\Omega)^3$ and an additional assumption on $(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})$.

3.3.3 Discretization, Gauss - Newton method and implementation

In the following we describe the idea how to solve the nonlinear minimization problem (3.19) through a sequence of linearized problems (3.21) in a finite dimensional space $\mathbf{\Pi}_h \times \mathbf{U}_h$. In our numerical experiments later we use $\mathbf{\Pi}_h \times \mathbf{U}_h = (\mathcal{RT}_{l-1}(\mathcal{T}_h))^3 \times (\mathcal{P}_l(\mathcal{T}_h))^3$, $l \geq 1$, respectively another suitable combination of the introduced finite element spaces in Section 2.5 for the approximation of $(\mathbf{P}, \mathbf{u})$.
We start with an initial solution $\left(\mathbf{P}_h^{(0)}, \mathbf{u}_h^{(0)}\right) \in \mathbf{\Pi}_h \times \mathbf{U}_h$, satisfying $\mathbf{P}_h^{(0)} \cdot \mathbf{n} = \mathbf{g}$ on Γ_N and $\mathbf{u}_h^{(0)} = \mathbf{u}_D$ on Γ_D, set $k = 0$ and solve the discrete problem of (3.22) in the finite element space $\mathbf{\Pi}_h \times \mathbf{U}_h$ to obtain a correction term $(\mathbf{Q}_h^{(k)}, \mathbf{v}_h^{(k)}) \in \mathbf{\Pi}_h \times \mathbf{U}_h$, satisfying $\mathbf{Q}_h^{(k)} \cdot \mathbf{n} = \mathbf{0}$ on Γ_N and $\mathbf{v}_h^{(k)} = \mathbf{0}$ on Γ_D. If $\{\mathbf{\Phi}_i\}_{i=1}^N$ denotes a basis of $\mathbf{\Pi}_h \times \mathbf{U}_h$ with $N := \dim(\mathbf{\Pi}_h \times \mathbf{U}_h)$ we set

$$\left(\mathbf{Q}_h^{(k)}, \mathbf{v}_h^{(k)}\right) = \sum_{i=1}^{N} x_i^{(k)} \mathbf{\Phi}_i \tag{3.24}$$

and can build in each step the stiffness matrix $\mathbf{A}^{(k)} \in \mathbb{R}^{N\times N}$ and the right - hand side $\mathbf{r}^{(k)} \in \mathbb{R}^N$ with components

$$\begin{aligned} A_{ij}^{(k)} &= a(\mathbf{\Phi}_j, \mathbf{\Phi}_i) = \left(\mathcal{R}'(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)})[\mathbf{\Phi}_j], \mathcal{R}'(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)})[\mathbf{\Phi}_i]\right)_{L^2(\Omega)}, \quad i,j = 1,\dots,N, \\ r_i^{(k)} &= F(\mathbf{\Phi}_i) = -\left(\mathcal{R}(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)}), \mathcal{R}'(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)})[\mathbf{\Phi}_i]\right)_{L^2(\Omega)}, \quad i = 1,\dots,N. \end{aligned} \tag{3.25}$$

Hence we have to solve the linear system of equations

$$\mathbf{A}^{(k)}\mathbf{x}^{(k)} = \mathbf{r}^{(k)} \tag{3.26}$$

to get the correction term $(\mathbf{Q}_h^{(k)}, \mathbf{v}_h^{(k)})$. The stiffness matrices $\mathbf{A}^{(k)}$, $k \geq 0$, are obviously symmetric in each step. Additionally as long as the bilinear form $a(\cdot,\cdot)$ is coercive the matrices are even symmetric positive definite, since under this assumption it holds

$$\begin{aligned} \left(\mathbf{x}^{(k)}\right)^T \mathbf{A}^{(k)}\mathbf{x}^{(k)} &= \sum_{i=1}^{N} x_i^{(k)} \sum_{j=1}^{N} A_{ij}^{(k)} x_j^{(k)} = \sum_{i=1}^{N} x_i^{(k)} \sum_{j=1}^{N} a(\mathbf{\Phi}_j, \mathbf{\Phi}_i) x_j^{(k)} \\ &= a\left(\sum_{j=1}^{N} x_j^{(k)} \mathbf{\Phi}_j, \sum_{i=1}^{N} x_i^{(k)} \mathbf{\Phi}_i\right) = a\left(\left(\mathbf{Q}_h^{(k)}, \mathbf{v}_h^{(k)}\right), \left(\mathbf{Q}_h^{(k)}, \mathbf{v}_h^{(k)}\right)\right) > 0 \end{aligned}$$

for all $0 \neq \mathbf{x}^{(k)} \in \mathbb{R}^N$. After solving the problem (3.26) with a suitable solver, we set the new approximation as

$$\left(\mathbf{P}_h^{(k+1)}, \mathbf{u}_h^{(k+1)}\right) = \left(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)}\right) + \alpha^{(k)} \left(\mathbf{Q}_h^{(k)}, \mathbf{v}_h^{(k)}\right),$$

where we have additionally inserted a parameter $\alpha^{(k)}$ of a suitable damping strategy. For instance one can use any line search or trust region method. In our numerical experiments later we use a backtracking line search strategy (cf. Section 3.1 in [NW06]), often also called (classical) Armijo method in literature:
We start with $\alpha^{(k)} = 1$ and multiply $\alpha^{(k)}$ with given fixed $\rho \in (0,1)$ as long as

$$\begin{aligned}
&\mathcal{F}\left(\mathbf{P}_h^{(k+1)}, \mathbf{u}_h^{(k+1)}\right) = \mathcal{F}\left(\left(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)}\right) + \alpha^{(k)} \left(\mathbf{Q}_h^{(k)}, \mathbf{v}_h^{(k)}\right)\right) \\
&\overset{!}{\leq} \mathcal{F}\left(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)}\right) + \frac{tol_2}{2}\, \alpha^{(k)} \mathcal{F}'\left(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)}\right) [\mathbf{Q}_h^{(k)}, \mathbf{v}_h^{(k)}] \\
&= \mathcal{F}\left(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)}\right) + tol_2\, \alpha^{(k)} \left(\mathcal{R}(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)}), \mathcal{R}'(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)})[\mathbf{Q}_h^{(k)}, \mathbf{v}_h^{(k)}]\right)_{L^2(\Omega)} \\
&= \mathcal{F}\left(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)}\right) + tol_2\, \alpha^{(k)} \sum_{i=1}^{N} x_i^{(k)} \underbrace{\left(\mathcal{R}(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)}), \mathcal{R}'(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)})[\mathbf{\Phi}_i]\right)_{L^2(\Omega)}}_{= -r_i^{(k)}} \\
&= \mathcal{F}\left(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)}\right) - tol_2\, \alpha^{(k)} \left(\mathbf{x}^{(k)}\right)^T \mathbf{r}^{(k)},
\end{aligned} \tag{3.27}$$

for small given $tol_2 > 0$, is satisfied. If the right-hand side $\mathbf{r}^{(k)}$ in (3.26) equals $\mathbf{0}$ a minimum is found. As long as the matrix $\mathbf{A}^{(k)}$ is symmetric positive definite it holds further
$\|\mathbf{x}^{(k)}\|^2_{\mathbf{A}^{(k)}} := \left(\mathbf{x}^{(k)}\right)^T \mathbf{A}^{(k)} \mathbf{x}^{(k)} = \left(\mathbf{x}^{(k)}\right)^T \mathbf{r}^{(k)} > 0$ by (3.26) and therefore

$$\begin{aligned}
\|\mathbf{r}^{(k)}\|^2_{\left(\mathbf{A}^{(k)}\right)^{-1}} &= \left(\mathbf{r}^{(k)}\right)^T \left(\mathbf{A}^{(k)}\right)^{-1} \mathbf{r}^{(k)} = \left(\mathbf{x}^{(k)}\right)^T \left(\mathbf{A}^{(k)}\right)^T \left(\mathbf{A}^{(k)}\right)^{-1} \mathbf{A}^{(k)} \mathbf{x}^{(k)} \\
&= \left(\mathbf{x}^{(k)}\right)^T \mathbf{A}^{(k)} \mathbf{x}^{(k)} = \|\mathbf{x}^{(k)}\|^2_{\mathbf{A}^{(k)}} = \left(\mathbf{x}^{(k)}\right)^T \mathbf{r}^{(k)}.
\end{aligned}$$

In particular this means that $\|\mathbf{r}^{(k)}\|_{\left(\mathbf{A}^{(k)}\right)^{-1}} = \sqrt{\left(\mathbf{x}^{(k)}\right)^T \mathbf{r}^{(k)}}$ is a suitable measure within any stopping criterion for the sequence of linearized problems. Furthermore by (3.27) we ensure that the value of the nonlinear functional decreases in each step. The requirement of (3.27) is the first Wolfe condition, often also called Armijo condition (cf. Section 3.1 in [NW06]). To ensure that the parameter does not become too small one can prescribe a parameter $\alpha_{\min}$ and demand $\alpha^{(k)} \geq \alpha_{\min}$ in the algorithm.
After the determination of $\left(\mathbf{P}_h^{(k+1)}, \mathbf{u}_h^{(k+1)}\right)$, which satisfies by construction automatically the given boundary conditions, we increase k by one and use the new approximation in the following variational problem (3.22). We continue this procedure as long as the given stopping criterion is satisfied or a prescribed number of iterations is exceeded. As output one gets an approximation $(\mathbf{P}_h, \mathbf{u}_h) \in \mathbf{\Pi}_h \times \mathbf{U}_h$ of a minimizer of the nonlinear least

Algorithm 1 Damped Gauss - Newton for minimizing the nonlinear functional (3.19)

Require: $tol > 0$, $tol_2 > 0$; $i_{\max} \in \mathbb{N}$, $\alpha_{\min} > 0$, $\rho \in (0,1)$;
initial solution $\left(\mathbf{P}_h^{(0)}, \mathbf{u}_h^{(0)}\right) \in \mathbf{\Pi}_h \times \mathbf{U}_h$, satisfying $\mathbf{P}_h^{(0)} \cdot \mathbf{n} = \mathbf{g}$ on Γ_N and $\mathbf{u}_h^{(0)} = \mathbf{u}_D$ on Γ_D;

Set $k = 0$, determine $\mathbf{r}^{(k)}$ via (3.25) and set $r = |\mathbf{r}^{(k)}|$;
while $r > tol$ **and** $k < i_{\max}$ **do**
  Determine the stiffness matrix $\mathbf{A}^{(k)}$ via (3.25);
  Solve the linear system of equation $\mathbf{A}^{(k)}\mathbf{x}^{(k)} = \mathbf{r}^{(k)}$ to obtain the correction term $\left(\mathbf{Q}_h^{(k)}, \mathbf{v}_h^{(k)}\right) \in \mathbf{\Pi}_h \times \mathbf{U}_h$ via (3.24);
  Set $\alpha^{(k)} = 1$;
  while $\mathcal{F}\left(\left(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)}\right) + \alpha^{(k)}\left(\mathbf{Q}_h^{(k)}, \mathbf{v}_h^{(k)}\right)\right) > \mathcal{F}\left(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)}\right) - tol_2\,\alpha^{(k)}\left(\mathbf{x}^{(k)}\right)^T \mathbf{r}^{(k)}$ **and** $\alpha^{(k)} \geq \alpha_{\min}$ **do**
    $\alpha^{(k)} = \rho\alpha^{(k)}$;
  end while
  Set $\mathbf{P}_h^{(k+1)} = \mathbf{P}_h^{(k)} + \alpha^{(k)}\mathbf{Q}_h^{(k)}$, $\mathbf{u}_h^{(k+1)} = \mathbf{u}_h^{(k)} + \alpha^{(k)}\mathbf{v}_h^{(k)}$; {new approximation}
  Set $r = \sqrt{\left(\mathbf{x}^{(k)}\right)^T \mathbf{r}^{(k)}}$ and $k = k + 1$;
  Determine $\mathbf{r}^{(k)}$ via (3.25);
end while

squares functional (3.19). A pseudocode of the whole algorithm can be found above in Algorithm 1.

Please note that the Gauss - Newton method works on a fixed triangulation $\mathcal{T}_h$ of the given domain Ω. For a further improvement of the solution one can refine the mesh, either uniformly or adaptively. One solves the problem on the coarse mesh, interpolates the obtained solution to the finer mesh, ensures the satisfaction of the given boundary conditions on the fine mesh and uses this approximation as initial solution in Algorithm 1 on the finer mesh.

We have to consider another problem for the numerical implementation. It must be possible to evaluate $\mathcal{F}(\mathbf{P}_h, \mathbf{u}_h)$ for given $(\mathbf{P}_h, \mathbf{u}_h) \in \mathbf{\Pi}_h \times \mathbf{U}_h$ locally at each quadrature point. If we consider (3.18) the problematical part is the evaluation of $\mathcal{A}$, respectively $\tilde{\mathcal{A}}$. However we can solve the problem $\mathcal{G}(\mathbf{B}) = \boldsymbol{\tau}$, respectively $\tilde{\mathcal{G}}(\mathbf{C}) = \mathbf{\Sigma}$ for $\boldsymbol{\tau} := \mathbf{P}_h\mathbf{F}(\mathbf{u}_h)^T, \mathbf{\Sigma} := \mathbf{F}(\mathbf{u}_h)^{-1}\mathbf{P}_h$ and given $(\mathbf{P}_h, \mathbf{u}_h) \in \mathbf{\Pi}_h \times \mathbf{U}_h$ with the help of a Newton scheme. We assume here a finite λ and sufficiently small $\boldsymbol{\tau}, \mathbf{\Sigma}$. The sequence of Newton iterations is given by

$$\mathbf{B}^{(j+1)} = \mathbf{B}^{(j)} + \mathbf{\Delta}^{(j)}, \quad \mathbf{C}^{(j+1)} = \mathbf{C}^{(j)} + \tilde{\mathbf{\Delta}}^{(j)}$$

with $\mathbf{\Delta}^{(j)}, \tilde{\mathbf{\Delta}}^{(j)}$ determined through

$$\mathcal{G}'\left(\mathbf{B}^{(j)}\right)\left[\mathbf{\Delta}^{(j)}\right] = \boldsymbol{\tau} - \mathcal{G}\left(\mathbf{B}^{(j)}\right), \quad \tilde{\mathcal{G}}'\left(\mathbf{C}^{(j)}\right)\left[\tilde{\mathbf{\Delta}}^{(j)}\right] = \mathbf{\Sigma} - \tilde{\mathcal{G}}\left(\mathbf{C}^{(j)}\right), \tag{3.28}$$

provided that $\mathcal{G}$ and $\tilde{\mathcal{G}}$ are Gâteaux differentiable. The starting values $\mathbf{B}^{(0)} = \mathbf{C}^{(0)} = \mathbf{I}$ are at least for small $(\mathbf{P}_h, \mathbf{u}_h)$ reasonable, since for $(\mathbf{P}_h, \mathbf{u}_h) = (\mathbf{0}, \mathbf{0})$ the solution is $\mathbf{B} = \mathbf{C} = \mathbf{I}$. The equations in (3.28) are nothing else than linear systems of equations with nine unknowns, where the occurring matrices depend on the old approximations $\mathbf{B}^{(j)}, \mathbf{C}^{(j)}$ and the right - hand sides depend on $(\mathbf{P}_h, \mathbf{u}_h)$ and $\mathbf{B}^{(j)}, \mathbf{C}^{(j)}$. We have to use this Newton iteration for each quadrature point and on each given element $T \in \mathcal{T}_h$. This means that for a prescribed maximal number $i_{\max}$ of Newton steps, n_t elements and n_q quadrature points, we have to solve in the worst case $n_t \cdot n_q \cdot i_{\max}$ linear systems with nine equations and nine unknowns. In the case of a plane strain model the 9×9 systems reduces to 5×5 systems. Obviously this is numerically very expensive, but it is in general possible. For a special Neo - Hooke model which we consider in the following sections it is even possible to solve the problem without Newton's method. In fact it is possible to set $\lambda = \infty$.
For more complicated models, based on the special Neo - Hooke model, the solution of the Neo - Hooke model can be used as initial values $\mathbf{B}^{(0)}, \mathbf{C}^{(0)}$ for the more complicated models in the Newton scheme (3.28).

3.3.4 Mappings $\mathcal{G}$ and $\tilde{\mathcal{G}}$ and their derivatives for Mooney - Rivlin and Neo - Hooke

The first heuristic nonlinear candidate for an extension of linear elasticity is the St. Venant - Kirchhoff model with stored energy function

$$\hat{\psi}_{SV}(\mathbf{F}) = \psi_{SV}(\mathbf{C}) = \frac{\lambda}{8}\left(\operatorname{tr}(\mathbf{C}) - 3\right)^2 + \frac{\mu}{4}\operatorname{tr}\left((\mathbf{C} - \mathbf{I})^2\right),$$

since it leads to the stress - strain relation (2.11), i.e. it is the stress - strain relation from linear elasticity with nonlinear kinematics. However this stored energy function is not polyconvex (cf. [Rao10]) and is therefore in general not suitable, since it does not fit into the existence theory of Ball (cf. [Bal77]).
Further extension of the material model leads historically to the Neo - Hooke and the Mooney - Rivlin model, proposed in [Cia88] and defined already in (2.31). The considered Neo - Hooke model in this work is a special case of the Mooney - Rivlin model (2.31), more precisely (2.31) with $\delta = 0$. These models include nonlinear kinematics as well as nonlinearities in the material law and are additionally polyconvex (cf. Sections 2.2.4 and 2.4.4). In the following the mappings $\mathcal{G}_{MR}(\mathbf{B})$, $\tilde{\mathcal{G}}_{MR}(\mathbf{C})$ and $\mathcal{G}_{NH}(\mathbf{B})$, $\tilde{\mathcal{G}}_{NH}(\mathbf{C})$ will be specified. Furthermore we will confirm that condition (3.15) holds actually for these materials.

Derivation of $\mathcal{G}_{MR}$ and $\tilde{\mathcal{G}}_{MR}$

In the following we study the Mooney - Rivlin material with stored energy function (2.31). To ensure consistency with linear elasticity we have to satisfy the conditions in (2.39) and

(2.40). Inserting these conditions into the representations (2.32) and (2.33) we get

$$\begin{aligned}\boldsymbol{\tau}_{MR} &= 2\alpha\mathbf{B} + (2\beta \det\mathbf{B} - \gamma)\,\mathbf{I} + 2\delta\left(\mathrm{tr}(\mathbf{B})\mathbf{B} - \mathbf{B}^2\right)\\ &= 2\left(\frac{\mu}{2} - \delta\right)\mathbf{B} + \left(2\left(\frac{\lambda}{4} - \delta\right)\det\mathbf{B} - \left(\mu + \frac{\lambda}{2}\right)\right)\mathbf{I} + 2\delta\left(\mathrm{tr}(\mathbf{B})\mathbf{B} - \mathbf{B}^2\right)\\ &= \mu\mathbf{B} + \left(\frac{\lambda}{2}(\det\mathbf{B} - 1) - \mu\right)\mathbf{I} + 2\delta\left((\mathrm{tr}(\mathbf{B}) - 1)\mathbf{B} - (\det\mathbf{B})\mathbf{I} - \mathbf{B}^2\right)\\ &=: \mathcal{G}_{MR}(\mathbf{B})\end{aligned} \tag{3.29}$$

and

$$\begin{aligned}\boldsymbol{\Sigma}_{MR} &= 2\alpha\,\mathbf{I} + (2\beta \det\mathbf{C} - \gamma)\mathbf{C}^{-1} + 2\delta\,(\mathrm{tr}(\mathbf{C})\mathbf{I} - \mathbf{C})\\ &= 2\left(\frac{\mu}{2} - \delta\right)\mathbf{I} + \left(2\left(\frac{\lambda}{4} - \delta\right)\det\mathbf{C} - \left(\mu + \frac{\lambda}{2}\right)\right)\mathbf{C}^{-1} + 2\delta\,(\mathrm{tr}(\mathbf{C})\mathbf{I} - \mathbf{C})\\ &= \mu\mathbf{I} + \left(\frac{\lambda}{2}(\det\mathbf{C} - 1) - \mu\right)\mathbf{C}^{-1} + 2\delta\left((\mathrm{tr}(\mathbf{C}) - 1)\mathbf{I} - \mathbf{C} - (\mathbf{Cof}\,\mathbf{C})^T\right)\\ &=: \tilde{\mathcal{G}}_{MR}(\mathbf{C}).\end{aligned} \tag{3.30}$$

Obviously $\mathcal{G}_{MR}$ and $\tilde{\mathcal{G}}_{MR}$ map symmetric matrices to symmetric matrices. We have already derived the Fréchet derivatives of the components of these mappings in Section 2.3.3. Combining these derivatives leads to $\mathcal{G}'_{MR}(\mathbf{B})[\mathbf{E}] = \partial\mathcal{G}_{MR}(\mathbf{B})(\mathbf{E})$ and $\tilde{\mathcal{G}}'_{MR}(\mathbf{B})[\mathbf{E}] = \partial\tilde{\mathcal{G}}_{MR}(\mathbf{B})(\mathbf{E})$ for arbitrary matrices $\mathbf{E} \in \mathbb{R}^{3\times3}$ with

$$\begin{aligned}\mathcal{G}'_{MR}(\mathbf{B})[\mathbf{E}] &= \mu\mathbf{E} + \frac{\lambda}{2}\,(\mathbf{Cof}\,\mathbf{B} : \mathbf{E})\,\mathbf{I}\\ &\quad + 2\delta\left[\mathrm{tr}(\mathbf{E})\mathbf{B} + (\mathrm{tr}(\mathbf{B}) - 1)\mathbf{E} - (\mathbf{Cof}\,\mathbf{B} : \mathbf{E})\,\mathbf{I} - (\mathbf{EB} + \mathbf{BE})\right]\\ \tilde{\mathcal{G}}'_{MR}(\mathbf{C})[\mathbf{E}] &= \frac{\lambda}{2}\,(\mathbf{Cof}\,\mathbf{C} : \mathbf{E})\,\mathbf{C}^{-1} - \left(\frac{\lambda}{2}(\det\mathbf{C} - 1) - \mu\right)\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}\\ &\quad + 2\delta\left[\mathrm{tr}(\mathbf{E})\mathbf{I} - \mathbf{E} - (\mathbf{Cof}\,\mathbf{C} : \mathbf{E})\,\mathbf{C}^{-1} + (\mathbf{Cof}\,\mathbf{C})^T\mathbf{E}\mathbf{C}^{-1}\right].\end{aligned} \tag{3.31}$$

Thus we see that $\mathcal{G}_{MR} : \mathbb{R}^{3\times3} \to \mathbb{R}^{3\times3}$ and $\tilde{\mathcal{G}}_{MR} : \mathbb{R}^{3\times3} \to \mathbb{R}^{3\times3}$ are Fréchet differentiable with the derivatives above. Furthermore the derivatives $\partial\mathcal{G}_{MR}(\mathbf{B}) : \mathbb{R}^{3\times3} \to \mathbb{R}^{3\times3}$, $\partial\tilde{\mathcal{G}}_{MR}(\mathbf{C}) : \mathbb{R}^{3\times3} \to \mathbb{R}^{3\times3}$ are continuous in $\mathbf{B}$ (respectively in $\mathbf{C}$) since they are even again differentiable with respect to $\mathbf{B}$ (with respect to $\mathbf{C}$). Thus they are altogether at least continuously differentiable.
For $\mathbf{B} = \mathbf{I} = \mathbf{C}$ it follows due to $\det\mathbf{I} = 1$, $\mathbf{Cof}\,\mathbf{I} = \mathbf{I}$ and $\mathrm{tr}(\mathbf{I}) = 3$

$$\begin{aligned}\mathcal{G}'_{MR}(\mathbf{I})[\mathbf{E}] &= \mu\mathbf{E} + \frac{\lambda}{2}\,(\mathbf{I} : \mathbf{E})\,\mathbf{I} + 2\delta\left[\mathrm{tr}(\mathbf{E})\mathbf{I} + (\mathrm{tr}(\mathbf{I}) - 1)\mathbf{E} - (\mathbf{I} : \mathbf{E})\,\mathbf{I} - 2\mathbf{E}\right]\\ &= \mu\mathbf{E} + \frac{\lambda}{2}\mathrm{tr}(\mathbf{E})\mathbf{I} + 2\delta\left[\mathrm{tr}(\mathbf{E})\mathbf{I} + 2\mathbf{E} - \mathrm{tr}(\mathbf{E})\mathbf{I} - 2\mathbf{E}\right]\\ &= \mu\mathbf{E} + \frac{\lambda}{2}\mathrm{tr}(\mathbf{E})\mathbf{I},\\ \tilde{\mathcal{G}}'_{MR}(\mathbf{I})[\mathbf{E}] &= \frac{\lambda}{2}\,(\mathbf{I} : \mathbf{E})\,\mathbf{I} - \left(\frac{\lambda}{2}(1 - 1) - \mu\right)\mathbf{E} + 2\delta\left[\mathrm{tr}(\mathbf{E})\mathbf{I} - \mathbf{E} - (\mathbf{I} : \mathbf{E})\,\mathbf{I} + \mathbf{E}\right]\\ &= \mu\mathbf{E} + \frac{\lambda}{2}\mathrm{tr}(\mathbf{E})\mathbf{I}.\end{aligned} \tag{3.32}$$

Altogether the mappings $\mathcal{G}_{MR}$ and $\tilde{\mathcal{G}}_{MR}$ are continuously differentiable and the condition (3.15) is confirmed for this model. Therefore the mappings are at least invertible in a neighborhood of $\mathbf{I}$ (cf. introduction of Section 3.3) and thus suitable for our inverse first-order systems (3.16) and (3.17).

Mappings $\mathcal{G}_{NH}$ and $\tilde{\mathcal{G}}_{NH}$ for the Neo-Hooke model

For the Neo-Hooke model, i.e. $\delta = 0$, we conclude by (3.29), (3.30) and (3.31)

$$\begin{aligned}
&\mathcal{G}_{NH}(\mathbf{B}) := \mu\mathbf{B} + \left(\frac{\lambda}{2}(\det\mathbf{B} - 1) - \mu\right)\mathbf{I},\\
&\tilde{\mathcal{G}}_{NH}(\mathbf{C}) := \mu\mathbf{I} + \left(\frac{\lambda}{2}(\det\mathbf{C} - 1) - \mu\right)\mathbf{C}^{-1}\\
&\text{with Fréchet derivatives}\\
&\mathcal{G}'_{NH}(\mathbf{B})[\mathbf{E}] = \partial\mathcal{G}_{NH}(\mathbf{B})(\mathbf{E}) = \mu\mathbf{E} + \frac{\lambda}{2}\left(\mathbf{Cof}\,\mathbf{B} : \mathbf{E}\right)\mathbf{I},\\
&\tilde{\mathcal{G}}'_{NH}(\mathbf{C})[\mathbf{E}] = \partial\tilde{\mathcal{G}}_{NH}(\mathbf{C})(\mathbf{E})\\
&\qquad\qquad = \frac{\lambda}{2}\left(\mathbf{Cof}\,\mathbf{C} : \mathbf{E}\right)\mathbf{C}^{-1} - \left(\frac{\lambda}{2}(\det\mathbf{C} - 1) - \mu\right)\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}.
\end{aligned} \tag{3.33}$$

For our analysis and the implementation of our approach it would be advantageous if we could invert the derivatives of $\mathcal{G}$ and $\tilde{\mathcal{G}}$ directly. In the case of the Neo-Hooke model we can invert $\mathcal{G}'_{NH}$ simply with the help of the following lemma and get an exact representation for its inverse.

Lemma 3.5: (General inversion formula)
Let $\mathbf{A}, \mathbf{C} \in \mathbb{R}^{n\times n}$ be arbitrary matrices and $a, b \in \mathbb{R}$ which may depend on $\mathbf{A}, \mathbf{C}$. Then a mapping $\mathcal{H} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ with $\mathcal{H}(\mathbf{E}) := a\,\mathbf{E} + b\,(\mathbf{C} : \mathbf{E})\mathbf{A}$, $\mathbf{E} \in \mathbb{R}^{n\times n}$, is invertible with inverse

$$\mathcal{H}^{-1}(\mathbf{\Sigma}) := \frac{1}{a}\left(\mathbf{\Sigma} - \frac{b}{a + b(\mathbf{C} : \mathbf{A})}(\mathbf{C} : \mathbf{\Sigma})\mathbf{A}\right), \quad \mathbf{\Sigma} \in \mathbb{R}^{n\times n},$$

provided that $a + b(\mathbf{C} : \mathbf{A}) \neq 0$ and $a \neq 0$.

Proof:

We have to show that $\mathcal{H}(\mathcal{H}^{-1}(\mathbf{\Sigma})) = \mathbf{\Sigma}$ and $\mathcal{H}^{-1}(\mathcal{H}(\mathbf{E})) = \mathbf{E}$ hold for arbitrary matrices $\mathbf{E}, \mathbf{\Sigma} \in \mathbb{R}^{n\times n}$. Due to

$$\begin{aligned}
\mathbf{C} : \mathcal{H}^{-1}(\mathbf{\Sigma}) &= \mathbf{C} : \left[\frac{1}{a}\left(\mathbf{\Sigma} - \frac{b}{a + b(\mathbf{C} : \mathbf{A})}(\mathbf{C} : \mathbf{\Sigma})\mathbf{A}\right)\right]\\
&= \frac{1}{a}(\mathbf{C} : \mathbf{\Sigma}) - \frac{1}{a}\left(\frac{b}{a + b(\mathbf{C} : \mathbf{A})}\right)(\mathbf{C} : \mathbf{\Sigma})(\mathbf{C} : \mathbf{A})
\end{aligned}$$

it holds on the one hand

$$\begin{aligned}\mathcal{H}(\mathcal{H}^{-1}(\boldsymbol{\Sigma})) &= a\,\mathcal{H}^{-1}(\boldsymbol{\Sigma}) + b\,(\mathbf{C} : \mathcal{H}^{-1}(\boldsymbol{\Sigma}))\mathbf{A} \\ &= \boldsymbol{\Sigma} - \frac{b}{a + b(\mathbf{C} : \mathbf{A})}(\mathbf{C} : \boldsymbol{\Sigma})\mathbf{A} + b\,(\mathbf{C} : \mathcal{H}^{-1}(\boldsymbol{\Sigma}))\mathbf{A} \\ &= \boldsymbol{\Sigma} + \frac{b}{a}(\mathbf{C} : \boldsymbol{\Sigma})\mathbf{A} - \frac{b}{a + b(\mathbf{C} : \mathbf{A})}\left(1 + \frac{b}{a}(\mathbf{C} : \mathbf{A})\right)(\mathbf{C} : \boldsymbol{\Sigma})\mathbf{A} \\ &= \boldsymbol{\Sigma} + \frac{b}{a}(\mathbf{C} : \boldsymbol{\Sigma})\mathbf{A} - \frac{b}{a}(\mathbf{C} : \boldsymbol{\Sigma})\mathbf{A} = \boldsymbol{\Sigma}.\end{aligned}$$

Due to

$$\begin{aligned}\mathbf{C} : \mathcal{H}(\mathbf{E}) = \mathbf{C} : [a\,\mathbf{E} + b\,(\mathbf{C} : \mathbf{E})\mathbf{A}] &= a(\mathbf{C} : \mathbf{E}) + b(\mathbf{C} : \mathbf{E})(\mathbf{C} : \mathbf{A}) \\ &= (a + b(\mathbf{C} : \mathbf{A}))(\mathbf{C} : \mathbf{E})\end{aligned}$$

it holds one the other hand

$$\begin{aligned}\mathcal{H}^{-1}(\mathcal{H}(\mathbf{E})) &= \frac{1}{a}\left(\mathcal{H}(\mathbf{E}) - \frac{b}{a + b(\mathbf{C} : \mathbf{A})}(\mathbf{C} : \mathcal{H}(\mathbf{E}))\mathbf{A}\right) \\ &= \frac{1}{a}\left(\mathcal{H}(\mathbf{E}) - b(\mathbf{C} : \mathbf{E})\mathbf{A}\right) = \mathbf{E}.\end{aligned}$$

□

We can use Lemma 3.5 to obtain an expression for $\mathcal{G}'_{NH}(\mathbf{B})^{-1}[\boldsymbol{\Sigma}]$. If we set $a = \mu$, $b = \frac{\lambda}{2}$, $\mathbf{C} = \mathbf{Cof}\,\mathbf{B}$ and $\mathbf{A} = \mathbf{I}$ we obtain, after expanding the fraction by 2, directly

$$\mathcal{G}'_{NH}(\mathbf{B})^{-1}[\boldsymbol{\Sigma}] = \frac{1}{\mu}\left(\boldsymbol{\Sigma} - \frac{\lambda}{2\mu + \lambda\,\mathrm{tr}(\mathbf{Cof}\,\mathbf{B})}(\mathbf{Cof}\,\mathbf{B} : \boldsymbol{\Sigma})\mathbf{I}\right). \tag{3.34}$$

Unfortunately, we cannot directly find a formula for the inverse $\tilde{\mathcal{G}}'_{NH}(\mathbf{C})^{-1}[\boldsymbol{\Sigma}]$ with the help of Lemma 3.5. However we can find a remedy. We construct a mapping $\hat{\mathcal{G}}_{NH}$ whose derivative is directly invertible with the help of Lemma 3.5. The idea is to define

$$\hat{\mathcal{G}}_{NH}(\mathbf{A}) = \mu\mathbf{I} + \left(\frac{\lambda}{2}\left((\det\mathbf{A})^{-1} - 1\right) - \mu\right)\mathbf{A}$$

for invertible $\mathbf{A} \in \mathbb{R}^{3\times 3}$. Then it holds due to $\det(\mathbf{C}^{-1}) = (\det\mathbf{C})^{-1}$ the relation $\hat{\mathcal{G}}_{NH}(\mathbf{C}^{-1}) = \tilde{\mathcal{G}}_{NH}(\mathbf{C})$. The Fréchet derivative of $\hat{\mathcal{G}}_{NH}$ is

$$\hat{\mathcal{G}}'_{NH}(\mathbf{A})[\mathbf{E}] = -\frac{\lambda}{2(\det\mathbf{A})^2}(\mathbf{Cof}\,\mathbf{A} : \mathbf{E})\mathbf{A} + \left(\frac{\lambda}{2}\left((\det\mathbf{A})^{-1} - 1\right) - \mu\right)\mathbf{E}$$

and it holds

$$\tilde{\mathcal{G}}'_{NH}(\mathbf{C})[\mathbf{E}] = \hat{\mathcal{G}}'_{NH}(\mathbf{C}^{-1})[-\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}]. \tag{3.35}$$

Due to Lemma 3.5 the inverse of $\hat{\mathcal{G}}'_{NH}(\mathbf{A})[\mathbf{E}]$ is given by

$$\hat{\mathcal{G}}'_{NH}(\mathbf{A})^{-1}[\boldsymbol{\Sigma}] = \frac{1}{a}\left(\boldsymbol{\Sigma} - \frac{b}{a + 3b\,\det\mathbf{A}}(\mathbf{Cof}\,\mathbf{A} : \boldsymbol{\Sigma})\mathbf{A}\right)$$

with $a := \frac{\lambda}{2}\left((\det \mathbf{A})^{-1} - 1\right) - \mu$, $b := -\frac{\lambda}{2(\det \mathbf{A})^2}$, provided that $a + 3b \det \mathbf{A} \neq 0$ and $a \neq 0$. We can now define the inverse of $\tilde{\mathcal{G}}'_{NH}(\mathbf{C})[\mathbf{E}]$ as

$$\begin{aligned}\tilde{\mathcal{G}}'_{NH}(\mathbf{C})^{-1}[\boldsymbol{\Sigma}] :=& -\mathbf{C}\left(\hat{\mathcal{G}}'_{NH}(\mathbf{C}^{-1})^{-1}[\boldsymbol{\Sigma}]\right)\mathbf{C} \\ =& -\frac{1}{a}\mathbf{C}\left(\boldsymbol{\Sigma} - \frac{b}{a + 3b\,\det \mathbf{C}^{-1}}(\mathbf{Cof}\,\mathbf{C}^{-1} : \boldsymbol{\Sigma})\mathbf{C}^{-1}\right)\mathbf{C}\end{aligned}$$

with $a := \frac{\lambda}{2}(\det \mathbf{C} - 1) - \mu$ and $b := -\frac{\lambda}{2}(\det \mathbf{C})^2$, since with this choice and the help of relation (3.35) it holds

$$\begin{aligned}\tilde{\mathcal{G}}'_{NH}(\mathbf{C})[\tilde{\mathcal{G}}'_{NH}(\mathbf{C})^{-1}[\boldsymbol{\Sigma}]] &= \tilde{\mathcal{G}}'_{NH}(\mathbf{C})\left[-\mathbf{C}\left(\hat{\mathcal{G}}'_{NH}(\mathbf{C}^{-1})^{-1}[\boldsymbol{\Sigma}]\right)\mathbf{C}\right] \\ &= \hat{\mathcal{G}}'_{NH}(\mathbf{C}^{-1})\left[-\mathbf{C}^{-1}\left(-\mathbf{C}\left(\hat{\mathcal{G}}'_{NH}(\mathbf{C}^{-1})^{-1}[\boldsymbol{\Sigma}]\right)\mathbf{C}\right)\mathbf{C}^{-1}\right] \\ &= \hat{\mathcal{G}}'_{NH}(\mathbf{C}^{-1})\left[\hat{\mathcal{G}}'_{NH}(\mathbf{C}^{-1})^{-1}[\boldsymbol{\Sigma}]\right] = \boldsymbol{\Sigma}, \\ \tilde{\mathcal{G}}'_{NH}(\mathbf{C})^{-1}[\tilde{\mathcal{G}}'_{NH}(\mathbf{C})[\mathbf{E}]] &= -\mathbf{C}\left(\hat{\mathcal{G}}'_{NH}(\mathbf{C}^{-1})^{-1}\left[\tilde{\mathcal{G}}'_{NH}(\mathbf{C})[\mathbf{E}]\right]\right)\mathbf{C} \\ &= -\mathbf{C}\left(\hat{\mathcal{G}}'_{NH}(\mathbf{C}^{-1})^{-1}\left[\hat{\mathcal{G}}'_{NH}(\mathbf{C}^{-1})[-\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}]\right]\right)\mathbf{C} \\ &= -\mathbf{C}\left(-\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}\right)\mathbf{C} = \mathbf{E}.\end{aligned}$$

We have to remark that $a = 0$ if and only if $\det \mathbf{C} = 1 + \frac{2\mu}{\lambda}$. Furthermore it holds

$$a + 3b\det \mathbf{C}^{-1} = \frac{\lambda}{2}(\det \mathbf{C} - 1) - \mu - \frac{3\lambda}{2}\det \mathbf{C} = \frac{\lambda}{2}(-2\det \mathbf{C} - 1) - \mu = 0$$

if and only if $\det \mathbf{C} = -\frac{1}{2} - \frac{\mu}{\lambda}$.

Remark 3.6:

In general it is problematic, in most cases even impossible, to invert the Fréchet derivative $\partial\mathcal{G}$ (respectively $\partial\tilde{\mathcal{G}}$) of the general mappings $\mathcal{G}$ and $\tilde{\mathcal{G}}$ in (3.14) belonging to homogeneous isotropic frame - indifferent materials. Already for the Mooney - Rivlin case we could not find a representation for the inverse of the corresponding derivatives (cf. (3.31)). This leads to more computational time in numerical simulations (cf. Section 3.3.3).

3.4 Suitability of the LSFEM approach with Neo - Hooke in the incompressible limit

In this section we consider the Neo - Hooke case and show that the corresponding operators $\mathcal{A}_{NH}$ and $\tilde{\mathcal{A}}_{NH}$ (with $\mathcal{A}_{NH} = \mathcal{G}_{NH}^{-1}$ and $\tilde{\mathcal{A}}_{NH} = \tilde{\mathcal{G}}_{NH}^{-1}$ for finite λ) are also well - defined in the incompressible limit. Moreover we will show that the operators are in the incompressible case no longer invertible, similar to $\mathcal{A}_{lin}$ in linear elasticity. We will derive cubic equations to determine $\mathcal{A}_{NH}(\boldsymbol{\tau})$ and $\tilde{\mathcal{A}}_{NH}(\boldsymbol{\Sigma})$ for given stresses $\boldsymbol{\tau}, \boldsymbol{\Sigma}$, i.e. we calculate the corresponding strains for given stresses. The novelty is here that we can even set $\lambda = \infty$ in these equations and are therefore able to consider the fully incompressible case in our theory and our numerical simulations later. We distinguish between the inverse $\mathbf{B}$ - and the inverse $\mathbf{C}$ - formulation (cf. (3.19)).

Cubic equation and incompressibility for the inverse B - formulation

The following explanations for the inverse **B** - formulation are already partly published in [MSSS14]. Our aim is to determine $\mathcal{A}_{NH}(\boldsymbol{\tau}) =: \mathbf{B}$ for given $\boldsymbol{\tau} \in \mathbb{R}^{3\times 3}$ and to show that $\mathcal{A}_{NH}$ is even well - defined for $\lambda \to \infty$. Let us assume firstly that λ is finite and a matrix $\boldsymbol{\tau}$ is given. We seek the corresponding strain $\mathbf{B} \in \mathbb{R}^{3\times 3}$ to $\boldsymbol{\tau}$ with $\mathcal{G}_{NH}(\mathbf{B}) = \boldsymbol{\tau}$.
We split $\mathbf{B}$ and $\boldsymbol{\tau}$ into its trace and deviatoric part with the help of

$$\mathbf{A} = \mathbf{dev}\,\mathbf{A} + \frac{1}{3}\mathrm{tr}(\mathbf{A})\mathbf{I}$$

which obviously holds for arbitrary matrices $\mathbf{A} \in \mathbb{R}^{3\times 3}$, due to the definition of the deviator. Inserting this splitting into $\mathcal{G}_{NH}(\mathbf{B}) = \boldsymbol{\tau}$ and using the representation of $\mathcal{G}_{NH}$ in (3.33) leads to

$$\mu\,\mathbf{dev}\,\mathbf{B} + \frac{\mu}{3}\,\mathrm{tr}(\mathbf{B})\mathbf{I} + \left(\frac{\lambda}{2}(\det\mathbf{B} - 1) - \mu\right)\mathbf{I} = \mathbf{dev}\,\boldsymbol{\tau} + \frac{1}{3}\mathrm{tr}(\boldsymbol{\tau})\mathbf{I}.$$

Since the splitting into its trace and deviatoric part of a matrix is unique it must hold

$$\begin{aligned} \mu\,\mathbf{dev}\,\mathbf{B} &= \mathbf{dev}\;\boldsymbol{\tau}, \\ \mu\left(\frac{1}{3}\mathrm{tr}(\mathbf{B}) - 1\right) + \frac{\lambda}{2}(\det\mathbf{B} - 1) &= \frac{1}{3}\mathrm{tr}(\boldsymbol{\tau}). \end{aligned} \tag{3.36}$$

For the derivation of an expression for $\det\mathbf{B}$ we use the property $\mathrm{tr}(\mathbf{dev}\,\mathbf{B}) = 0$ of the deviator, the properties $\det(c\mathbf{B}) = c^3(\det\mathbf{B})$ and $\mathbf{Cof}\,(c\mathbf{B}) = c^2\mathbf{Cof}\,\mathbf{B}$ (cf. representation (2.15)) for $c \in \mathbb{R}$ and arbitrary matrices $\mathbf{B} \in \mathbb{R}^{3\times 3}$ and the identity

$$\det(\mathbf{B}_1 + \mathbf{B}_2) = \det\mathbf{B}_1 + \mathbf{Cof}\,\mathbf{B}_1 : \mathbf{B}_2 + \mathbf{B}_1 : \mathbf{Cof}\,\mathbf{B}_2 + \det\mathbf{B}_2$$

for arbitrary matrices $\mathbf{B}_1, \mathbf{B}_2 \in \mathbb{R}^{3\times 3}$ (cf. Exercise 1.3 in [Cia88]). Combining these properties with the first equation in (3.36) implies

$$\begin{aligned} \det\mathbf{B} &= \det\left(\mathbf{dev}\,\mathbf{B} + \frac{1}{3}\mathrm{tr}(\mathbf{B})\mathbf{I}\right) \\ &= \det(\mathbf{dev}\,\mathbf{B}) + \mathbf{Cof}\,(\mathbf{dev}\,\mathbf{B}) : \frac{1}{3}\mathrm{tr}(\mathbf{B})\mathbf{I} + \mathbf{dev}\,\mathbf{B} : \mathbf{Cof}\,\left(\frac{1}{3}\mathrm{tr}(\mathbf{B})\mathbf{I}\right) + \det\left(\frac{1}{3}\mathrm{tr}(\mathbf{B})\mathbf{I}\right) \\ &= \det\left(\frac{\mathbf{dev}\,\boldsymbol{\tau}}{\mu}\right) + \frac{1}{3}\mathrm{tr}(\mathbf{B})\,\mathrm{tr}\left(\mathbf{Cof}\,\left(\frac{\mathbf{dev}\,\boldsymbol{\tau}}{\mu}\right)\right) + \underbrace{\mathbf{dev}\,\mathbf{B} : \left(\frac{1}{9}(\mathrm{tr}(\mathbf{B}))^2\right)\mathbf{I}}_{=0} \\ &\quad + \frac{1}{27}(\mathrm{tr}(\mathbf{B}))^3\det(\mathbf{I}) \\ &= \frac{1}{27}(\mathrm{tr}(\mathbf{B}))^3 + \frac{1}{3\mu^2}\mathrm{tr}(\mathbf{B})\,\mathrm{tr}\,(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\tau})) + \frac{1}{\mu^3}\det\,(\mathbf{dev}\,\boldsymbol{\tau})\,. \end{aligned} \tag{3.37}$$

We plug this expression for $\det\mathbf{B}$ into the second equation of (3.36) and obtain

$$\begin{aligned} &\frac{\lambda}{2}\left(\frac{1}{27}(\mathrm{tr}(\mathbf{B}))^3 + \frac{1}{3\mu^2}\mathrm{tr}(\mathbf{B})\,\mathrm{tr}\,(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\tau})) + \frac{1}{\mu^3}\det\,(\mathbf{dev}\,\boldsymbol{\tau}) - 1\right) \\ &+ \mu\left(\frac{1}{3}\mathrm{tr}(\mathbf{B}) - 1\right) = \frac{1}{3}\mathrm{tr}(\boldsymbol{\tau}). \end{aligned}$$

Multiplying the whole equation with $\left(\frac{27\cdot 2}{\lambda}\right)$, subtracting the resulting $\mathrm{tr}(\boldsymbol{\tau})$-term and ordering the equation in powers of $\mathrm{tr}(\mathbf{B})$ leads to

$$\begin{aligned}(\mathrm{tr}(\mathbf{B}))^3 + \left(\frac{9}{\mu^2}\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\tau})) + \frac{18\mu}{\lambda}\right)\mathrm{tr}(\mathbf{B}) \\ + 27\left(\frac{1}{\mu^3}\det(\mathbf{dev}\,\boldsymbol{\tau}) - 1 - \frac{2\mu}{\lambda} - \frac{2}{3\lambda}\mathrm{tr}(\boldsymbol{\tau})\right) = 0.\end{aligned} \tag{3.38}$$

Thus with the coefficients

$$\begin{aligned}S &:= \frac{9}{\mu^2}\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\tau})) + \frac{18\mu}{\lambda}, \\ T &:= 27\left(\frac{1}{\mu^3}\det(\mathbf{dev}\,\boldsymbol{\tau}) - 1 - \frac{2\mu}{\lambda} - \frac{2}{3\lambda}\mathrm{tr}(\boldsymbol{\tau})\right)\end{aligned} \tag{3.39}$$

we have obtained a cubic equation

$$(\mathrm{tr}(\mathbf{B}))^3 + S\,\mathrm{tr}(\mathbf{B}) + T = 0 \tag{3.40}$$

with discriminant $D := \left(\frac{S}{3}\right)^3 + \left(\frac{T}{2}\right)^2$ for $\mathrm{tr}(\mathbf{B})$. This cubic equation can be solved with Cardano's formula (cf. Section 2.1.6.2 in [Zei13]). For $D < 0$ one obtains in general three different real solutions and for $D = 0$ one obtains also three real solutions and at least two of them are equal. Since we are interested in a unique solution for $\mathbf{B}$, only the case $D > 0$ makes sense. In this case one has one real solution and the other two solutions are complex conjugates. Since the strain $\mathbf{B}$ that corresponds to $\boldsymbol{\tau}$ should be real, only the real solution makes sense. This real unique solution is given by

$$\mathrm{tr}(\mathbf{B}) = \sqrt[3]{-\frac{T}{2} + \sqrt{D}} + \sqrt[3]{-\frac{T}{2} - \sqrt{D}}. \tag{3.41}$$

If we have determined $\mathrm{tr}(\mathbf{B})$ via (3.41), provided that $D > 0$, and $\mathbf{dev}\,\mathbf{B} = \frac{\mathbf{dev}\,\boldsymbol{\tau}}{\mu}$ via the first equation in (3.36) we obtain a unique strain

$$\mathbf{B} = \mathbf{dev}\,\mathbf{B} + \frac{1}{3}\mathrm{tr}(\mathbf{B})\mathbf{I}$$

which belongs to the given stress $\boldsymbol{\tau}$, i.e. $\mathbf{B} = \mathcal{A}_{NH}(\boldsymbol{\tau})$. For $\mathbf{u} = \mathbf{0}$ we have $\boldsymbol{\tau} = \mathbf{0}$ due to consistency with linear elasticity (cf. Section 2.4.5). In this case it holds $\mathbf{dev}\,\boldsymbol{\tau} = \mathbf{0}$, $\mathrm{tr}(\boldsymbol{\tau}) = 0$ and therefore $S = \frac{18\mu}{\lambda}$ and $T = -27\left(1 + \frac{2\mu}{\lambda}\right)$. We obtain the discriminant

$$D = \left(\frac{S}{3}\right)^3 + \left(\frac{T}{2}\right)^2 = \left(\frac{6\mu}{\lambda}\right)^3 + \left(-27\left(\frac{1}{2} + \frac{\mu}{\lambda}\right)\right)^2 = 216\left(\frac{\mu}{\lambda}\right)^3 + 729\left(\frac{1}{2} + \frac{\mu}{\lambda}\right)^2 \tag{3.42}$$

which is obviously positive for given Lamé constants $\lambda, \mu > 0$. This observation confirms that the mapping $\mathcal{G}_{NH}$ is invertible at least for small strains $\mathbf{B}$ in a neighborhood of $\mathbf{I}$ or equivalently for small enough stresses $\boldsymbol{\tau}$. We make this statement more precise in the following proposition:

Proposition 3.7:
Under the assumptions of

$$\left|\frac{\mathbf{dev}\,\boldsymbol{\tau}}{\mu}\right| \leq a, \quad \left|\frac{\mathrm{tr}(\boldsymbol{\tau})}{\lambda}\right| \leq b$$

with $d := a^3 + \frac{2}{3}b - 1 < 0$ and $d^2 > 96a^6\sqrt{3}$ it holds $T < 0$ and $D > 0$ in the cubic equation (3.40). In this case we have furthermore $\mathrm{tr}(\mathbf{B}) > 0$ for the strain $\mathbf{B} = \mathcal{A}_{NH}(\boldsymbol{\tau})$.

Proof:
The identity $\mathbf{Cof}\,(c\,\mathbf{A}) = c^2\mathbf{Cof}\,\mathbf{A}$ (cf. representation (2.15)) for arbitrary $c \in \mathbb{R}$, $\mathbf{A} \in \mathbb{R}^{3\times 3}$, implies $c^2\mathrm{tr}(\mathbf{Cof}\,\mathbf{A}) = \mathrm{tr}(\mathbf{Cof}\,(c\,\mathbf{A}))$. In combination with Corollary 2.33 it follows

$$\frac{1}{\mu^2}|\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\tau}))| = \left|\mathrm{tr}\left(\mathbf{Cof}\,\left(\frac{\mathbf{dev}\,\boldsymbol{\tau}}{\mu}\right)\right)\right| \leq 6\sqrt{3}\left|\frac{\mathbf{dev}\,\boldsymbol{\tau}}{\mu}\right|^2 \leq 6a^2\sqrt{3},$$

i.e. in particular $\frac{1}{\mu^2}\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\tau})) \geq -6a^2\sqrt{3}$. This implies

$$S = \frac{9}{\mu^2}\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\tau})) + \underbrace{\frac{18\mu}{\lambda}}_{\geq 0} \geq \frac{9}{\mu^2}\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\tau})) \geq -54a^2\sqrt{3}.$$

For the coefficient T it holds with the help of Corollary 2.35

$$\begin{aligned}
T &= 27\left(\frac{1}{\mu^3}\det(\mathbf{dev}\,\boldsymbol{\tau}) - 1 - \frac{2\mu}{\lambda} - \frac{2}{3\lambda}\mathrm{tr}(\boldsymbol{\tau})\right) \leq 27\left(\frac{1}{\mu^3}\det(\mathbf{dev}\,\boldsymbol{\tau}) - 1 - \frac{2}{3\lambda}\mathrm{tr}(\boldsymbol{\tau})\right) \\
&\leq 27\left(\frac{1}{\mu^3}|\det(\mathbf{dev}\,\boldsymbol{\tau})| - 1 + \left|-\frac{2}{3\lambda}\mathrm{tr}(\boldsymbol{\tau})\right|\right) = 27\left(\left|\det\left(\frac{\mathbf{dev}\,\boldsymbol{\tau}}{\mu}\right)\right| - 1 + \frac{2}{3\lambda}|\mathrm{tr}(\boldsymbol{\tau})|\right) \\
&\leq 27\left(\left|\frac{\mathbf{dev}\,\boldsymbol{\tau}}{\mu}\right|^3 - 1 + \frac{2}{3}\left|\frac{\mathrm{tr}(\boldsymbol{\tau})}{\lambda}\right|\right) \leq 27\left(a^3 + \frac{2}{3}b - 1\right) = 27d < 0
\end{aligned}$$

by assumption.
These considerations imply $\frac{S}{3} \geq -18a^2\sqrt{3}$ and $\frac{T}{2} \leq \frac{27}{2}\left(a^3 + \frac{2}{3}b - 1\right) < 0$. Since the mapping $x \mapsto x^3$ is monotonically increasing on $(-\infty, \infty)$ and $x \mapsto x^2$ is monotonically decreasing on $(-\infty, 0)$ we get

$$\left(\frac{S}{3}\right)^3 \geq -(18\sqrt{3})^3a^6, \quad \left(\frac{T}{2}\right)^2 \geq \left(\frac{27}{2}\right)^2\left(a^3 + \frac{2}{3}b - 1\right)^2$$

and therefore

$$\begin{aligned}
D &= \left(\frac{S}{3}\right)^3 + \left(\frac{T}{2}\right)^2 \geq -(18\sqrt{3})^3a^6 + \left(\frac{27}{2}\right)^2\left(a^3 + \frac{2}{3}b - 1\right)^2 \\
&= \left(\frac{27}{2}\right)^2\left(\left(a^3 + \frac{2}{3}b - 1\right)^2 - \left(\frac{2}{27}\right)^2(18\sqrt{3})^3a^6\right) \\
&= \left(\frac{27}{2}\right)^2\left(\underbrace{\left(a^3 + \frac{2}{3}b - 1\right)^2}_{=d^2} - 96a^6\sqrt{3}\right) > 0
\end{aligned}$$

by assumption. It remains to show that $\mathrm{tr}(\mathcal{A}_{NH}(\boldsymbol{\tau})) > 0$. With the considerations above it holds obviously $-T > 0 \Leftrightarrow -\frac{T}{2} > \frac{T}{2}$ and therefore $-\frac{T}{2}+\sqrt{D} > \frac{T}{2}+\sqrt{D} = -\left(-\frac{T}{2}-\sqrt{D}\right)$. The function $f(x) = \sqrt[3]{x}$ is due to $f'(x) = \frac{1}{3}x^{-\frac{2}{3}} = \frac{1}{3}\frac{1}{\sqrt[3]{x^2}} > 0$ for all $x \in \mathbb{R}$ strictly increasing in $\mathbb{R}$, i.e. it holds

$$\begin{aligned}\sqrt[3]{-\left(-\frac{T}{2}-\sqrt{D}\right)} < \sqrt[3]{-\frac{T}{2}+\sqrt{D}} &\Leftrightarrow \sqrt[3]{-\frac{T}{2}+\sqrt{D}} > -\sqrt[3]{-\frac{T}{2}-\sqrt{D}} \\ &\Leftrightarrow \underbrace{\sqrt[3]{-\frac{T}{2}+\sqrt{D}}+\sqrt[3]{-\frac{T}{2}-\sqrt{D}}}_{=\mathrm{tr}(\mathbf{B})} > 0\end{aligned}$$

for $\mathbf{B} := \mathcal{A}_{NH}(\boldsymbol{\tau})$ by equation (3.41).

□

Example 3.8:
For instance for $a = b = \frac{1}{3}$ it holds $a^3 + \frac{2}{3}b - 1 = -\frac{20}{27} < 0$ and $d^2 = \frac{400}{729} > \frac{96\sqrt{3}}{729} = 96a^6\sqrt{3}$. Thus the assumptions in Proposition 3.7 are satisfied and therefore for such stress tensors $\boldsymbol{\tau}$ the cubic equation (3.40) has a unique real solution.
We remark that this choice is not optimal. However in numerical experiments one can easily check for every approximation in the program if $D > 0$ and $T < 0$ is still satisfied or not.

With the derivation of the cubic equation above we can also state the following theorem concerning the well - posedness of $\mathcal{A}_{NH}$.

Theorem 3.9: (Well - posedness of $\mathcal{A}_{NH}$ for $\lambda \to \infty$)
Assume that the discriminant D of the cubic equation (3.40) is positive.
Then the mapping $\mathbf{B} = \mathcal{A}_{NH}(\boldsymbol{\tau})$, defined by the first equation in (3.36) and (3.41), is well - defined in the incompressible limit $\lambda \to \infty$. Its inverse does not exist in this case.

Proof:

We can take the limit $\lambda \to \infty$ in (3.39) and obtain the coefficients

$$\begin{aligned}\lim_{\lambda\to\infty} S &= \lim_{\lambda\to\infty}\left[\frac{9}{\mu^2}\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\tau})) + \frac{18\mu}{\lambda}\right] = \frac{9}{\mu^2}\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\tau})), \\ \lim_{\lambda\to\infty} T &= \lim_{\lambda\to\infty}\left[27\left(\frac{1}{\mu^3}\det(\mathbf{dev}\,\boldsymbol{\tau}) - 1 - \frac{2\mu}{\lambda} - \frac{2}{3\lambda}\mathrm{tr}(\boldsymbol{\tau})\right)\right] \\ &= 27\left(\frac{1}{\mu^3}\det(\mathbf{dev}\,\boldsymbol{\tau}) - 1\right).\end{aligned} \tag{3.43}$$

Thus also in this case, provided that $D > 0$, we get a unique solution for $\mathrm{tr}(\mathbf{B})$ and therefore for $\mathbf{B}$. It remains to show that in the incompressible case the inverse of $\mathcal{A}_{NH}$ does not exist anymore.

Obviously (3.43) does not depend on $\mathrm{tr}(\boldsymbol{\tau})$ anymore. For instance for $\boldsymbol{\tau}_2 := \boldsymbol{\tau}_1 + c\,\mathbf{I}$ with given arbitrary matrix $\boldsymbol{\tau}_1$ and $c \in \mathbb{R} \setminus \{0\}$ it holds $\boldsymbol{\tau}_1 \neq \boldsymbol{\tau}_2$ and $\mathbf{dev}\,\boldsymbol{\tau}_1 = \mathbf{dev}\,\boldsymbol{\tau}_2$. This implies due to (3.43) the same coefficients S and T in the cubic equation (3.40). With $\mathbf{dev}\,\mathbf{B} = \frac{\mathbf{dev}\,\boldsymbol{\tau}}{\mu}$, according to the first equation in (3.36), it follows

$$\mathbf{dev}\,\mathbf{B}_1 = \mathbf{dev}\,\mathbf{B}_2 \text{ and } \mathrm{tr}(\mathbf{B}_1) = \mathrm{tr}(\mathbf{B}_2) \Rightarrow \mathcal{A}_{NH}(\boldsymbol{\tau}_1) = \mathbf{B}_1 = \mathbf{B}_2 = \mathcal{A}_{NH}(\boldsymbol{\tau}_2).$$

This means that $\mathcal{A}_{NH}$ is not injective and therefore not invertible for $\lambda \to \infty$.

□

Remark 3.10: (Exact satisfaction of the incompressibility constraint)
Another remarkable fact is that it always holds $\det(\mathcal{A}_{NH}(\mathbf{PF}(\mathbf{u})^T)) = 1$ for any combination $(\mathbf{P}, \mathbf{u})$ in the incompressible limit. This can be seen in the following way. In the incompressible case we get the coefficients (3.43) and therefore the cubic equation

$$\begin{aligned}
&(\mathrm{tr}(\mathbf{B}))^3 + \left(\frac{9}{\mu^2}\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\tau}))\right)\mathrm{tr}(\mathbf{B}) + 27\left(\frac{1}{\mu^3}\det(\mathbf{dev}\,\boldsymbol{\tau}) - 1\right) = 0\\
\Leftrightarrow\ &\frac{1}{27}(\mathrm{tr}(\mathbf{B}))^3 + \left(\frac{1}{3\mu^2}\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\tau}))\right)\mathrm{tr}(\mathbf{B}) + \left(\frac{1}{\mu^3}\det(\mathbf{dev}\,\boldsymbol{\tau}) - 1\right) = 0\\
\Leftrightarrow\ &\frac{1}{27}(\mathrm{tr}(\mathbf{B}))^3 + \left(\frac{1}{3\mu^2}\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\tau}))\right)\mathrm{tr}(\mathbf{B}) + \frac{1}{\mu^3}\det(\mathbf{dev}\,\boldsymbol{\tau}) = 1.
\end{aligned}$$

Inserting this equation into (3.37) results in $\det\mathbf{B} = 1$ for $\mathbf{B} = \mathcal{A}_{NH}(\boldsymbol{\tau})$ with $\boldsymbol{\tau} = \mathbf{PF}(\mathbf{u})^T$. This means that our approach produces an exactly incompressible strain $\mathbf{B}$ (cf. Section 2.2.3).

Cubic equation and incompressibility for the inverse C - formulation

We can also show the well - posedness of the operator $\tilde{\mathcal{A}}_{NH}$ for $\lambda \to \infty$ as we will see in the following. The first step is again to determine $\tilde{\mathcal{A}}_{NH}(\boldsymbol{\Sigma}) =: \mathbf{C}$ for given $\boldsymbol{\Sigma} \in \mathbb{R}^{3\times3}$. Let us assume again firstly that λ is finite. We seek the corresponding strain $\mathbf{C} \in \mathbb{R}^{3\times3}$ to $\boldsymbol{\Sigma}$ with $\tilde{\mathcal{G}}_{NH}(\mathbf{C}) = \boldsymbol{\Sigma}$.
By the representation of $\tilde{\mathcal{G}}_{NH}$ in (3.33) and $\rho := \frac{\lambda}{2}(\det\mathbf{C} - 1) - \mu$ it follows

$$\tilde{\mathcal{G}}_{NH}(\mathbf{C}) = \boldsymbol{\Sigma} \Leftrightarrow \mu\,\mathbf{I} + \rho\mathbf{C}^{-1} = \boldsymbol{\Sigma} \Leftrightarrow \rho\mathbf{C}^{-1} = \boldsymbol{\Sigma} - \mu\mathbf{I}. \tag{3.44}$$

Obviously for $\mathbf{C}$ with $\det\mathbf{C} \neq 0$, which is usually valid for the seeked strain, it holds $\det(\boldsymbol{\Sigma} - \mu\mathbf{I}) = 0$ if and only if $\rho = 0$. We assume that $\det(\boldsymbol{\Sigma} - \mu\mathbf{I}) \neq 0$ and therefore $\rho \neq 0$. It follows $\mathbf{C}^{-1} = \frac{1}{\rho}(\boldsymbol{\Sigma} - \mu\mathbf{I})$ with determinant

$$\det(\mathbf{C}^{-1}) = \frac{1}{\rho^3}\det(\boldsymbol{\Sigma} - \mu\mathbf{I}) \Leftrightarrow \det\mathbf{C} = \frac{1}{\det(\mathbf{C}^{-1})} = \frac{\rho^3}{\det(\boldsymbol{\Sigma} - \mu\mathbf{I})}. \tag{3.45}$$

Inserting (3.45) and $\mathbf{C}^{-1} = \frac{1}{\rho}(\mathbf{\Sigma} - \mu\mathbf{I})$ into $\tilde{\mathcal{G}}_{NH}(\mathbf{C}) = \mathbf{\Sigma}$, with $\tilde{\mathcal{G}}_{NH}(\mathbf{C})$ given by (3.33), leads to

$$\begin{aligned}\tilde{\mathcal{G}}_{NH}(\mathbf{C}) &= \left(\frac{\lambda}{2}\left(\frac{\rho^3}{\det(\mathbf{\Sigma}-\mu\mathbf{I})} - 1\right) - \mu\right)\frac{\mathbf{\Sigma}-\mu\mathbf{I}}{\rho} + \mu\mathbf{I} \stackrel{!}{=} \mathbf{\Sigma} \quad |\cdot\rho \\ &\Leftrightarrow \left(\frac{\lambda}{2}\left(\frac{\rho^3}{\det(\mathbf{\Sigma}-\mu\mathbf{I})} - 1\right) - \mu\right)(\mathbf{\Sigma}-\mu\mathbf{I}) = \rho(\mathbf{\Sigma}-\mu\mathbf{I}) \\ &\Leftrightarrow \frac{\lambda}{2}\left(\frac{\rho^3}{\det(\mathbf{\Sigma}-\mu\mathbf{I})}\right)(\mathbf{\Sigma}-\mu\mathbf{I}) = \left(\frac{\lambda}{2}+\mu+\rho\right)(\mathbf{\Sigma}-\mu\mathbf{I}).\end{aligned}$$

Multiplying this equation with $\frac{2}{\lambda}\det(\mathbf{\Sigma}-\mu\mathbf{I}) \neq 0$ implies

$$\rho^3(\mathbf{\Sigma}-\mu\mathbf{I}) = \left(1 + \frac{2\mu}{\lambda} + \frac{2\rho}{\lambda}\right)\det(\mathbf{\Sigma}-\mu\mathbf{I})(\mathbf{\Sigma}-\mu\mathbf{I}).$$

Since we have assumed $\det(\mathbf{\Sigma}-\mu\mathbf{I}) \neq 0$, it must hold

$$\rho^3 - \left(\frac{2}{\lambda}\det(\mathbf{\Sigma}-\mu\mathbf{I})\right)\rho - \left(1+\frac{2\mu}{\lambda}\right)\det(\mathbf{\Sigma}-\mu\mathbf{I}) = 0. \tag{3.46}$$

Thus with

$$S := -\frac{2}{\lambda}\det(\mathbf{\Sigma}-\mu\mathbf{I}), \quad T := -\left(1+\frac{2\mu}{\lambda}\right)\det(\mathbf{\Sigma}-\mu\mathbf{I})$$

we have again a cubic equation of the form (3.40) to determine ρ. With the same arguments as for the inverse $\mathbf{B}$ - formulation we get a unique solution

$$\rho = \sqrt[3]{-\frac{T}{2} + \sqrt{D}} + \sqrt[3]{-\frac{T}{2} - \sqrt{D}} \tag{3.47}$$

provided that the corresponding discriminant

$$D := \left(\frac{S}{3}\right)^3 + \left(\frac{T}{2}\right)^2 = -\frac{8}{27\lambda^3}(\det(\mathbf{\Sigma}-\mu\mathbf{I}))^3 + \left(\frac{1}{2}+\frac{\mu}{\lambda}\right)^2(\det(\mathbf{\Sigma}-\mu\mathbf{I}))^2 \tag{3.48}$$

is positive. After determining ρ via (3.47) we obtain for a given stress $\mathbf{\Sigma}$ by construction (cf. (3.44))

$$\mathbf{C}^{-1} = \frac{\mathbf{\Sigma}-\mu\mathbf{I}}{\rho} \Leftrightarrow \mathbf{C} = \rho(\mathbf{\Sigma}-\mu\mathbf{I})^{-1}, \tag{3.49}$$

i.e. a corresponding strain $\mathbf{C}$ under the assumption that $\mathbf{\Sigma}-\mu\mathbf{I}$ is invertible (cp. Section 10.3 in [Wri08]).

For $\mathbf{\Sigma} = \mathbf{0} \in \mathbb{R}^{3\times 3}$ we get $\det(\mathbf{\Sigma}-\mu\mathbf{I}) = \det(-\mu\mathbf{I}) = -\mu^3$ and therefore

$$D = \frac{8\mu^9}{27\lambda^3} + \mu^6\left(\frac{1}{2}+\frac{\mu}{\lambda}\right)^2 > 0.$$

This observation confirms that the mapping $\tilde{\mathcal{G}}_{NH}$ is invertible at least for small strains $\mathbf{C}$ in a neighborhood of $\mathbf{I}$ or equivalently for small enough stresses $\mathbf{\Sigma}$.

With these considerations we can state the following theorem for the operator $\tilde{\mathcal{A}}_{NH}$ similar to Theorem 3.9.

Theorem 3.11: (Well - posedness of $\tilde{\mathcal{A}}_{NH}$ for $\lambda \to \infty$)
Assume that for a given stress tensor $\mathbf{\Sigma}$ with $\det(\mathbf{\Sigma} - \mu\mathbf{I}) \neq 0$ the discriminant D, defined by (3.48), is positive.
Then the mapping $\mathbf{C} = \tilde{\mathcal{A}}_{NH}(\mathbf{\Sigma})$, defined by (3.47) and (3.49), is well - defined in the incompressible limit $\lambda \to \infty$. Its inverse does not exist in this case.

Proof:

In the incompressible case $\lambda \to \infty$ the cubic equation (3.46) turns into

$$\rho^3 = \det(\mathbf{\Sigma} - \mu\mathbf{I}) \Leftrightarrow \rho = \sqrt[3]{\det(\mathbf{\Sigma} - \mu\mathbf{I})}, \tag{3.50}$$

i.e. we have a unique solution for ρ and therefore by (3.49) a unique solution for $\mathbf{C}$.
For an arbitrary matrix $\mathbf{\Sigma}_1 \in \mathbb{R}^{3\times 3}$ and $c \in \mathbb{R} \setminus \{1\}$ we set $\mathbf{\Sigma}_2 := c(\mathbf{\Sigma}_1 - \mu\mathbf{I}) + \mu\mathbf{I}$ such that $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$. For this choice and the fact that $\rho_i^3 = \det(\mathbf{\Sigma}_i - \mu\mathbf{I})$, $i = 1, 2$, by (3.50), it holds

$$\det(\mathbf{\Sigma}_2 - \mu\mathbf{I}) = \det(c(\mathbf{\Sigma}_1 - \mu\mathbf{I})) = c^3 \det(\mathbf{\Sigma}_1 - \mu\mathbf{I}) \Leftrightarrow \rho_2^3 = c^3\rho_1^3 \Leftrightarrow \rho_2 = c\rho_1.$$

Thus by (3.49) we conclude

$$\begin{aligned}\tilde{\mathcal{A}}_{NH}(\mathbf{\Sigma}_2) = \mathbf{C}_2 &= \rho_2(\mathbf{\Sigma}_2 - \mu\mathbf{I})^{-1} = c\rho_1 \left(c(\mathbf{\Sigma}_1 - \mu\mathbf{I})\right)^{-1} \\ &= \rho_1(\mathbf{\Sigma}_1 - \mu\mathbf{I})^{-1} = \mathbf{C}_1 = \tilde{\mathcal{A}}_{NH}(\mathbf{\Sigma}_1).\end{aligned}$$

This means that $\tilde{A}_{NH}$ is not injective and therefore not invertible for $\lambda \to \infty$.

□

Furthermore we get an analogous result as stated in Remark 3.10.

Remark 3.12: (Exact satisfaction of the incompressibility constraint)
Also in the case of the inverse $\mathbf{C}$ - formulation we satisfy the incompressibility constraint $\det\left(\tilde{A}_{NH}(\mathbf{\Sigma})\right) = 1$ for given stress tensor $\mathbf{\Sigma}$ with $\det(\mathbf{\Sigma} - \mu\mathbf{I}) \neq 0$ exactly, since it holds $\rho^3 = \det(\mathbf{\Sigma} - \mu\mathbf{I})$ in the incompressible case (cf. equation (3.50)) and therefore by (3.45) $\det \mathbf{C} = \rho^3 \left(\det(\mathbf{\Sigma} - \mu\mathbf{I})\right)^{-1} = 1$.

3.5 Analysis for the inverse B - formulation and Neo - Hooke material law

In this section of the work we will analyze the nonlinear least squares formulation (3.19) and its linearized problem (3.21) for the inverse **B** - formulation and a Neo - Hooke material.

3.5.1 The nonlinear problem

The general aim for the nonlinear problem (3.19) is to estimate the error from below and above by the nonlinear least squares functional, similar to (3.12), i.e. to obtain an estimate of the form (3.20). The analysis in [CS04] for linear elasticity is done without scaling the first - order system. Since we need this theory in some proofs below we set for simplicity $\omega_1 = \omega_2 = 1$ in the whole Section 3.5.
For the estimation of the error we need some preparations. The first preparation is a mapping property concerning the nonlinear operator $\mathcal{A}_{NH}$.

Lemma 3.13: (Mapping property of $\mathcal{A}_{NH}$)
The operator $\mathcal{A}_{NH}$, defined by the first equation in (3.36) and the cubic equation (3.38), provided that its discriminant is positive, maps functions in $L^2(\Omega)^{3\times 3}$ into $L^2(\Omega)^{3\times 3}$.

Proof:
We recall that the definition of the Lebesgue spaces $L^p(\Omega)$ in Section 2.1.4 is valid for $p \in (0,1)$. In this case $\|f\|^p_{L^p(\Omega)} := \int_\Omega |f|^p \, dx$, $f \in L^p(\Omega)$, is a quasi - norm. Additionally we need the generalized Hölder inequality. It states that for functions $f_j \in L^{p_j}(\Omega)$ with $p_j \in (0,\infty]$, $j = 1,\dots,m$, its product $\prod_{j=1}^{m} f_j$ is in $L^r(\Omega)$ with $\frac{1}{r} := \sum_{j=1}^{m} \frac{1}{p_j}$ (cf. Corollary 2.6 in [AF03]).
We have to show that for $\boldsymbol{\tau} \in L^2(\Omega)^{3\times 3}$ it holds $\mathbf{B} := \mathcal{A}_{NH}(\boldsymbol{\tau}) \in L^2(\Omega)^{3\times 3}$. For $\boldsymbol{\tau} \in L^2(\Omega)^{3\times 3}$ it follows immediately that $\mathrm{tr}(\boldsymbol{\tau}) = \tau_{11} + \tau_{22} + \tau_{33} \in L^2(\Omega)$. Then, by definition of the deviator and the first equation in (3.36), it follows

$$\mathbf{dev}\,\boldsymbol{\tau} = \boldsymbol{\tau} - \frac{1}{3}\mathrm{tr}(\boldsymbol{\tau})\mathbf{I} \in L^2(\Omega)^{3\times 3} \Rightarrow \mathbf{dev}\,\mathbf{B} = \frac{\mathbf{dev}\,\boldsymbol{\tau}}{\mu} \in L^2(\Omega)^{3\times 3}.$$

It remains to show that $\mathrm{tr}(\mathbf{B}) \in L^2(\Omega)$ to obtain

$$\mathbf{B} = \mathbf{dev}\,\mathbf{B} + \frac{1}{3}\mathrm{tr}(\mathbf{B})\mathbf{I} \in L^2(\Omega)^{3\times 3}.$$

The representation (2.15) of the cofactor and the generalized Hölder inequality for two functions in $L^2(\Omega)$ imply $\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\tau}) \in L^1(\Omega)^{3\times 3}$. Due to $\mathbf{dev}\,\boldsymbol{\tau} \in L^2(\Omega)^{3\times 3}$ and the fact that each term in $\det(\mathbf{dev}\,\boldsymbol{\tau})$ is a product of three matrix entries of $\mathbf{dev}\,\boldsymbol{\tau}$ we get by the generalized Hölder inequality $\det(\mathbf{dev}\,\boldsymbol{\tau}) \in L^{\frac{2}{3}}(\Omega)$. These considerations imply the

coefficients

$$S = \frac{9}{\mu^2} \underbrace{\operatorname{tr}(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\tau}))}_{\in L^1(\Omega)} + \frac{18\mu}{\lambda} \in L^1(\Omega),$$

$$T = 27 \left(\frac{1}{\mu^3} \underbrace{\det(\mathbf{dev}\,\boldsymbol{\tau})}_{\in L^{\frac{2}{3}}(\Omega)} - 1 - \frac{2\mu}{\lambda} - \frac{2}{3\lambda} \underbrace{\operatorname{tr}(\boldsymbol{\tau})}_{\in L^2(\Omega)} \right) \in L^{\frac{2}{3}}(\Omega)$$

in the cubic equation (3.40). Here we have to remark that $L^q(\Omega) \subseteq L^p(\Omega)$ holds, even in the case $0 < p \leq q \leq 1$, as long as Ω does not contain sets of arbitrarily large measure (cp. Theorem 2.14 in [AF03] and Theorem 2 in [Vil85]).
An arbitrary function f is due to

$$\|f^q\|_{L^p(\Omega)} = \left(\int_\Omega |f^q|^p \, dx \right)^{\frac{1}{p}} = \left(\int_\Omega |f|^{qp} \, dx \right)^{\frac{1}{p}} = \|f\|^q_{L^{qp}(\Omega)}$$

in $L^{qp}(\Omega)$ if and only if $f^q \in L^p(\Omega)$ with $0 < p, q < \infty$.
This means that $S \in L^1(\Omega)$ $(q = 3, p = \frac{1}{3})$ implies $S^3 \in L^{\frac{1}{3}}(\Omega)$ and $T \in L^{\frac{2}{3}}(\Omega)$ $(q = 2, p = \frac{1}{3})$ implies $T^2 \in L^{\frac{1}{3}}(\Omega)$. Altogether we get $D = \left(\frac{S}{3}\right)^3 + \left(\frac{T}{2}\right)^2 \in L^{\frac{1}{3}}(\Omega)$. This implies $\sqrt{D} \in L^{\frac{2}{3}}(\Omega)$ with $q = \frac{1}{2}$, $p = \frac{2}{3}$ and therefore $-\frac{T}{2} \pm \sqrt{D} \in L^{\frac{2}{3}}(\Omega)$. With $q = \frac{1}{3}$ and $p = 2$ we conclude $\sqrt[3]{-\frac{T}{2} \pm \sqrt{D}} \in L^2(\Omega)$ and by (3.41) $\operatorname{tr}(\mathbf{B}) \in L^2(\Omega)$.

□

At the end of Section 3.3.1 it was mentioned that we have to choose suitable values for q and p in the function spaces $W^q(\operatorname{div};\Omega)^3$ and $W^{1,p}(\Omega)^3$ such that $\mathcal{R}(\mathbf{P}, \mathbf{u})$ is in $L^2(\Omega)^3 \times L^2(\Omega)^{3\times 3}$. We specify now these values for the inverse $\mathbf{B}$ - formulation and the considered Neo - Hooke law:

Corollary 3.14:
For the inverse $\mathbf{B}$ - formulation (3.19) with Neo - Hooke law $\mathcal{A} = \mathcal{A}_{NH}$, defined by the first equation in (3.36) and the cubic equation (3.38), again provided that its discriminant is positive, it holds for $\mathbf{u} \in W^{1,4}(\Omega)^3$, $\mathbf{P} \in W^4(\operatorname{div};\Omega)^3$ and a volume force density $\mathbf{f} \in L^4(\Omega)^3$

$$\mathcal{R}_{NH}(\mathbf{P}, \mathbf{u}) \in L^4(\Omega)^3 \times L^2(\Omega)^{3\times 3}.$$

Proof:

By definition of the space $W^4(\operatorname{div};\Omega)^3$ (cf. Definition 2.18) it holds $\operatorname{div}\mathbf{P} \in L^4(\Omega)^3$ for given $\mathbf{P} \in W^4(\operatorname{div};\Omega)^3$, i.e. $\operatorname{div}\mathbf{P} + \mathbf{f} \in L^4(\Omega)^3$ is clear. By definition of $W^{1,4}(\Omega)^3$ (cf. Definition 2.17) it holds $\mathbf{F}(\mathbf{u}) = \mathbf{I} + \nabla\mathbf{u} \in L^4(\Omega)^{3\times 3}$ for $\mathbf{u} \in W^{1,4}(\Omega)^3$. The generalized Hölder inequality implies $\mathbf{P}\mathbf{F}(\mathbf{u})^T \in L^2(\Omega)^{3\times 3}$. By Lemma 3.13 we know that $\mathcal{A}_{NH}(\mathbf{P}\mathbf{F}(\mathbf{u})^T) \in L^2(\Omega)^{3\times 3}$. Since $\mathbf{F}(\mathbf{u}) \in L^4(\Omega)^{3\times 3}$ it follows $\mathbf{B}(\mathbf{u}) = \mathbf{F}(\mathbf{u})(\mathbf{F}(\mathbf{u}))^T \in L^2(\Omega)^{3\times 3}$ and therefore $\mathcal{A}_{NH}(\mathbf{P}\mathbf{F}(\mathbf{u})^T) - \mathbf{B}(\mathbf{u}) \in L^2(\Omega)^{3\times 3}$.

□

Due to this corollary the least squares functional (3.19) for the **B** - formulation and the Neo - Hooke law exists for $p = 4 = q$.
For our purposes we also need the derivative of $\mathcal{R}_{NH}(\mathbf{P}, \mathbf{u})$ with respect to $(\mathbf{P}, \mathbf{u})$. By the local inversion theorem (cf. Theorem 2.11) and the inverse (3.34) of $\mathcal{G}'_{NH}(\mathbf{B})[\mathbf{E}]$ we obtain the derivative

$$\begin{aligned}\mathcal{A}'_{NH}(\boldsymbol{\tau})[\boldsymbol{\Sigma}] &= \mathcal{G}'_{NH}(\mathcal{A}_{NH}(\boldsymbol{\tau}))^{-1}[\boldsymbol{\Sigma}] \\ &= \frac{1}{\mu}\left(\boldsymbol{\Sigma} - \frac{\lambda}{2\mu + \lambda \operatorname{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\tau})))}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\tau})) : \boldsymbol{\Sigma})\mathbf{I}\right).\end{aligned} \tag{3.51}$$

The cubic equation (3.38) is uniquely solvable for $\boldsymbol{\tau} = \mathbf{0}$ (cf. (3.42)) with solution $\operatorname{tr}(\mathcal{A}_{NH}(\mathbf{0})) = 3$. The corresponding strain is then given by $\mathcal{A}_{NH}(\mathbf{0}) = \mathbf{I}$. For $\boldsymbol{\tau} = \mathbf{0}$ we obtain therefore

$$\mathcal{A}'_{NH}(\mathbf{0})[\boldsymbol{\Sigma}] = \frac{1}{\mu}\left(\boldsymbol{\Sigma} - \frac{\lambda}{2\mu + 3\lambda}\operatorname{tr}(\boldsymbol{\Sigma})\mathbf{I}\right) = 2\mathcal{C}^{-1}\boldsymbol{\Sigma} = 2\mathcal{A}_{lin}(\boldsymbol{\Sigma}), \tag{3.52}$$

i.e. $\mathcal{A}'_{NH}(\mathbf{0})$ is up to a constant identical to the operator $\mathcal{A}_{lin} = \mathcal{C}^{-1}$ of linear elasticity. It follows by equation (3.18)

$$\mathcal{R}'_{NH}(\mathbf{P}, \mathbf{u})[\mathbf{Q}, \mathbf{v}] = \begin{pmatrix} \omega_1 \operatorname{div} \mathbf{Q} \\ \omega_2\Big(\mathcal{A}'_{NH}(\mathbf{P}\mathbf{F}(\mathbf{u})^T)[\mathbf{Q}\mathbf{F}(\mathbf{u})^T + \mathbf{P}(\nabla \mathbf{v})^T] - \nabla \mathbf{v}\mathbf{F}(\mathbf{u})^T - \mathbf{F}(\mathbf{u})(\nabla \mathbf{v})^T\Big) \end{pmatrix} \tag{3.53}$$

and with $(\mathbf{P}, \mathbf{u}) = (\mathbf{0}, \mathbf{0})$

$$\begin{aligned}\mathcal{R}'_{NH}(\mathbf{0}, \mathbf{0})[\mathbf{Q}, \mathbf{v}] &= \begin{pmatrix} \omega_1 \operatorname{div} \mathbf{Q} \\ \omega_2\Big(\mathcal{A}'_{NH}(\mathbf{0})[\mathbf{Q}] - \nabla \mathbf{v} - (\nabla \mathbf{v})^T\Big) \end{pmatrix} = \begin{pmatrix} \omega_1 \operatorname{div} \mathbf{Q} \\ 2\,\omega_2\,(\mathcal{A}_{lin}(\mathbf{Q}) - \boldsymbol{\varepsilon}(\mathbf{v})) \end{pmatrix} \\ &\left[= \begin{pmatrix} \operatorname{div} \mathbf{Q} \\ 2\,(\mathcal{A}_{lin}(\mathbf{Q}) - \boldsymbol{\varepsilon}(\mathbf{v})) \end{pmatrix} \quad (\text{for } \omega_1 = \omega_2 = 1)\right],\end{aligned} \tag{3.54}$$

i.e. $\mathcal{R}'_{NH}(\mathbf{0}, \mathbf{0})[\mathbf{Q}, \mathbf{v}]$ is up to a constant identical to the operator $\mathcal{L}(\mathbf{Q}, \mathbf{v})$ of linear elasticity (cf. equation (3.8)).
In what follows (cp. [MSSS14]), let for $\mathbf{P}^N \in W^\infty(\operatorname{div}; \Omega)^3$, satisfying $\mathbf{P}^N \cdot \mathbf{n} = \mathbf{g}$ on Γ_N, and for $\mathbf{u}_D \in W^{1,\infty}(\Omega)^3$, satisfying the boundary conditions on Γ_D,

$$\begin{aligned}\boldsymbol{\Pi}^\infty &:= \{\mathbf{Q} \in W^\infty(\operatorname{div}; \Omega)^3 : \|\mathbf{Q}\|_{L^\infty(\Omega)} \leq \theta\} \cap (\mathbf{P}^N + W^4_{\Gamma_N}(\operatorname{div}; \Omega)^3), \\ \mathbf{U}^\infty &:= \{\mathbf{u} \in W^{1,\infty}(\Omega)^3 : \|\nabla \mathbf{u}\|_{L^\infty(\Omega)} \leq \theta\} \cap (\mathbf{u}_D + W^{1,4}_{\Gamma_D}(\Omega)^3)\end{aligned} \tag{3.55}$$

be the restriction of the solution spaces to sufficiently small neighborhoods of the origin, i.e. for sufficiently small θ. We will assume in the following that $(\mathbf{P}, \mathbf{u}) \in \boldsymbol{\Pi}^\infty \times \mathbf{U}^\infty$ is an exact solution of $\mathcal{R}_{NH}(\mathbf{P}, \mathbf{u}) = \mathbf{0}$ for $\mathcal{R}_{NH}(\mathbf{P}, \mathbf{u})$ defined in (3.18). Since we have to satisfy $-\operatorname{div} \mathbf{P} = \mathbf{f}$ in Ω and $\mathbf{P} \cdot \mathbf{n} = \mathbf{P}^N \cdot \mathbf{n} = \mathbf{g}$ on Γ_N, we need $\mathbf{f} \in L^\infty(\Omega)^3$ and $\mathbf{g} \in L^\infty(\Gamma_N)^3$ for the densities of the given volume and surface forces. In the case of a pure displacement

boundary problem (i.e. $\Gamma_N = \emptyset$), $\mathbf{u}_D = \mathbf{0}$ on Γ_D, the existence of a unique solution is ensured for sufficiently small $\|\mathbf{f}\|_{L^\infty(\Omega)}$ and strong regularity assumptions (cf. Theorem 6.7.1 in [Cia88]).
We will prove (cp. (3.20))

$$||(\mathbf{P}-\mathbf{P}_h, \mathbf{u}-\mathbf{u}_h)||_{\mathcal{V}}^2 \lesssim \mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h) \lesssim ||(\mathbf{P}-\mathbf{P}_h, \mathbf{u}-\mathbf{u}_h)||_{\mathcal{V}}^2 \tag{3.56}$$

for $\mathcal{V} := H_{\Gamma_N}(\mathrm{div};\Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$ and an approximation $(\mathbf{P}_h, \mathbf{u}_h) \in \mathbf{\Pi}^\infty \times \mathbf{U}^\infty$. It would be great if the constants that appear in (3.56) are independent of λ such that the Poisson locking problem is eliminated in our approach. For the proof of (3.56) we need some further lemmata.

Lemma 3.15: (Estimate of the cofactor in three dimensions near the identity)
For an arbitrary matrix $\mathbf{A} \in \mathbb{R}^{3\times 3}$ it holds

$$|\mathbf{Cof}\,\mathbf{A} - \mathbf{I}|^2 \leq 6|\mathbf{A}-\mathbf{I}|^2 + 3|\mathbf{A}-\mathbf{I}|^4.$$

Proof:

For the proof of the statements we need the (complex) Schur decomposition of the matrix $\mathbf{A} \in \mathbb{R}^{3\times 3} \subset \mathbb{C}^{3\times 3}$ (cf. Theorem 2.3.1 in [HJ13]). For this purpose we must extend the definition of the Frobenius norm in Section 2.1.2 to matrices over $\mathbb{C}$. The Frobenius norm for a matrix $\mathbf{Q} \in \mathbb{C}^{n\times n}$ is defined by

$$|\mathbf{Q}| := (\mathrm{tr}(\mathbf{Q}^*\mathbf{Q}))^{\frac{1}{2}} = \left(\sum_{i,j=1}^{n} |q_{ij}|^2\right)^{\frac{1}{2}},$$

where $\mathbf{Q}^* := \overline{\mathbf{Q}}^T$ denotes the conjugate transpose of $\mathbf{Q}$.
With the help of the (complex) Schur decomposition we can find for $\mathbf{A} \in \mathbb{R}^{3\times 3}$ a unitary matrix $\mathbf{Q} \in \mathbb{C}^{3\times 3}$, i.e. $\mathbf{Q}^* = \mathbf{Q}^{-1}$, and an upper triangular matrix $\mathbf{R} \in \mathbb{C}^{3\times 3}$ with $\mathbf{Q}^*\mathbf{A}\mathbf{Q} = \mathbf{R}$. It follows

$$\begin{aligned}
\mathbf{Cof}\,\mathbf{R} = \mathbf{Cof}\,(\mathbf{Q}^*\mathbf{A}\mathbf{Q}) &= \det(\mathbf{Q}^*\mathbf{A}\mathbf{Q})(\mathbf{Q}^*\mathbf{A}\mathbf{Q})^{-T} = (\overline{\det\mathbf{Q}})(\det\mathbf{A})(\det\mathbf{Q})(\overline{\mathbf{Q}}^T\mathbf{A}\mathbf{Q})^{-T} \\
&= (\overline{\det\mathbf{Q}})(\det\mathbf{A})(\det\mathbf{Q})(\overline{\mathbf{Q}}^{-1}\mathbf{A}^{-T}\mathbf{Q}^{-T}) = (\overline{\det\mathbf{Q}})\overline{\mathbf{Q}}^{-1}(\det\mathbf{A})\mathbf{A}^{-T}(\det\mathbf{Q})\mathbf{Q}^{-T} \\
&= \left((\det\mathbf{Q})\mathbf{Q}^{-T}\right)^*(\mathbf{Cof}\,\mathbf{A})(\mathbf{Cof}\,\mathbf{Q}) = (\mathbf{Cof}\,\mathbf{Q})^* \cdot (\mathbf{Cof}\,\mathbf{A}) \cdot (\mathbf{Cof}\,\mathbf{Q}).
\end{aligned}$$

In this equation of the proof we have assumed for simplicity that $\mathbf{A} \in \mathbb{R}^{3\times 3}$ is invertible. However, this equation holds in general, i.e. is also valid for matrices which are non-invertible.
For unitary $\mathbf{Q}$ also $\mathbf{Cof}\,\mathbf{Q}$ is unitary, since

$$\begin{aligned}
(\mathbf{Cof}\,\mathbf{Q})^* &= \left((\det\mathbf{Q})\mathbf{Q}^{-T}\right)^* = (\overline{\det\mathbf{Q}})\left(\mathbf{Q}^{-T}\right)^* = (\det\mathbf{Q}^*)\left(\overline{\mathbf{Q}}^{-1}\right) = (\det\mathbf{Q}^{-1})\mathbf{Q}^T \\
&= (\det\mathbf{Q})^{-1}\mathbf{Q}^T = \left((\det\mathbf{Q})\mathbf{Q}^{-T}\right)^{-1} = (\mathbf{Cof}\,\mathbf{Q})^{-1}.
\end{aligned}$$

Until now we have proven that $\mathbf{Cof\,R}$ is similar to $\mathbf{Cof\,A}$, i.e. both matrices have the same eigenvalues, and that $\mathbf{Cof\,Q}$ is also unitary.
With the representation $\mathbf{R} = \begin{pmatrix} \lambda_1 & \varepsilon_1 & \varepsilon_2 \\ 0 & \lambda_2 & \varepsilon_3 \\ 0 & 0 & \lambda_3 \end{pmatrix}$, including the eigenvalues $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{C}$ of $\mathbf{A}$, it follows

$$\mathbf{Cof\,R} = \begin{pmatrix} \lambda_2\lambda_3 & 0 & 0 \\ -\lambda_3\varepsilon_1 & \lambda_1\lambda_3 & 0 \\ -\lambda_2\varepsilon_2 + \varepsilon_1\varepsilon_3 & -\lambda_1\varepsilon_3 & \lambda_1\lambda_2 \end{pmatrix}$$

by Definition 2.21 (respectively representation (2.15)).
We set $\delta_i := \lambda_i - 1$ for $i = 1, 2, 3$ and obtain with the representations above

$$|\mathbf{R} - \mathbf{I}|^2 = \sum_{i=1}^{3} \left(|\lambda_i - 1|^2 + |\varepsilon_i|^2\right) = \sum_{i=1}^{3} \left(|\delta_i|^2 + |\varepsilon_i|^2\right),$$
$$|\mathbf{Cof\,R} - \mathbf{I}|^2 = |\lambda_2\lambda_3 - 1|^2 + |\lambda_1\lambda_3 - 1|^2 + |\lambda_1\lambda_2 - 1|^2 + |\lambda_3\varepsilon_1|^2 + |\lambda_1\varepsilon_3|^2 + |\lambda_2\varepsilon_2 - \varepsilon_1\varepsilon_3|^2.$$

Now we estimate the single terms in $|\mathbf{Cof\,R} - \mathbf{I}|^2$, using only the triangle and Young's inequality. If we combine both inequalities we get the estimates

$$\begin{aligned} |a+b|^2 &\le (|a| + |b|)^2 = |a|^2 + |b|^2 + 2|a||b| \le 2\left(|a|^2 + |b|^2\right), \\ |a+b+c|^2 &\le (|a| + |b| + |c|)^2 = |a|^2 + |b|^2 + |c|^2 + 2\left(|a||b| + |a||c| + |b||c|\right) \\ &\le 3\left(|a|^2 + |b|^2 + |c|^2\right) \end{aligned} \tag{3.57}$$

for $a, b, c \in \mathbb{C}$.

(i) For the first three terms in $|\mathbf{Cof\,R} - \mathbf{I}|^2$ with $(i, j) \in \{(2,3), (1,3), (1,2)\}$ we conclude

$$\begin{aligned} |\lambda_i\lambda_j - 1|^2 = |(1+\delta_i)(1+\delta_j) - 1|^2 &= |\delta_i + \delta_j + \delta_i\delta_j|^2 \le 3\left(|\delta_i|^2 + |\delta_j|^2 + |\delta_i\delta_j|^2\right) \\ &\le 3\left(|\delta_i|^2 + |\delta_j|^2\right) + \frac{3}{2}\left(|\delta_i|^4 + |\delta_j|^4\right). \end{aligned}$$

(ii) For $(i, j) \in \{(3,1), (1,3)\}$ we conclude

$$|\lambda_i\varepsilon_j|^2 = |(\lambda_i - 1)\varepsilon_j + \varepsilon_j|^2 = |\delta_i\varepsilon_j + \varepsilon_j|^2 \le 2|\delta_i|^2|\varepsilon_j|^2 + 2|\varepsilon_j|^2.$$

(iii) For the last term in the representation of $|\mathbf{Cof\,R} - \mathbf{I}|^2$ we conclude

$$\begin{aligned} |\lambda_2\varepsilon_2 - \varepsilon_1\varepsilon_3|^2 = |(\lambda_2 - 1)\varepsilon_2 + \varepsilon_2 - \varepsilon_1\varepsilon_3|^2 &= |\delta_2\varepsilon_2 + \varepsilon_2 - \varepsilon_1\varepsilon_3|^2 \\ &\le 3\left(|\delta_2\varepsilon_2|^2 + |\varepsilon_2|^2 + |\varepsilon_1\varepsilon_3|^2\right). \end{aligned}$$

Inserting these inequalities in the expression of $|\mathbf{Cof}\,\mathbf{R}-\mathbf{I}|^2$ above results in

$$\begin{aligned}
|\mathbf{Cof}\,\mathbf{R}-\mathbf{I}|^2 &= |\lambda_2\lambda_3-1|^2+|\lambda_1\lambda_3-1|^2+|\lambda_1\lambda_2-1|^2+|\lambda_3\varepsilon_1|^2+|\lambda_1\varepsilon_3|^2+|\lambda_2\varepsilon_2-\varepsilon_1\varepsilon_3|^2\\
&\le 6\left(|\delta_1|^2+|\delta_2|^2+|\delta_3|^2\right)+3\left(|\delta_1|^4+|\delta_2|^4+|\delta_3|^4\right)\\
&\quad+2\left(|\delta_1|^2|\varepsilon_3|^2+|\delta_3|^2|\varepsilon_1|^2\right)+2\left(|\varepsilon_1|^2+|\varepsilon_3|^2\right)+3\left(|\delta_2\varepsilon_2|^2+|\varepsilon_2|^2+|\varepsilon_1\varepsilon_3|^2\right)\\
&\le 6\sum_{i=1}^{3}|\delta_i|^2+3\sum_{i=1}^{3}|\varepsilon_i|^2+\\
&\quad+3\underbrace{\left(|\delta_1|^4+|\delta_2|^4+|\delta_3|^4+|\delta_1|^2|\varepsilon_3|^2+|\delta_3|^2|\varepsilon_1|^2+|\delta_2|^2|\varepsilon_2|^2+|\varepsilon_1|^2|\varepsilon_3|^2\right)}_{\le|\mathbf{R}-\mathbf{I}|^4}.
\end{aligned}$$

Thus altogether we obtain

$$|\mathbf{Cof}\,\mathbf{R}-\mathbf{I}|^2\le 6|\mathbf{R}-\mathbf{I}|^2+3|\mathbf{R}-\mathbf{I}|^4.$$

Due to the similarity of $\mathbf{A}$ to $\mathbf{R}$ and the invariance of the (complex) Frobenius norm, i.e.

$$|\mathbf{Q}^*\mathbf{A}\mathbf{Q}|^2=\mathrm{tr}\left((\mathbf{Q}^*\mathbf{A}\mathbf{Q})^*\,\mathbf{Q}^*\mathbf{A}\mathbf{Q}\right)=\mathrm{tr}\left(\mathbf{Q}^*\mathbf{A}^*\mathbf{Q}\mathbf{Q}^*\mathbf{A}\mathbf{Q}\right)=\mathrm{tr}\left(\mathbf{A}^*\mathbf{A}\right)=|\mathbf{A}|^2$$

for unitary matrices $\mathbf{Q}$, it follows

$$|\mathbf{R}-\mathbf{I}|^2=|\mathbf{Q}^*\mathbf{A}\mathbf{Q}-\mathbf{I}|^2=|\mathbf{Q}^*(\mathbf{A}-\mathbf{I})\mathbf{Q}|^2=|\mathbf{A}-\mathbf{I}|^2. \tag{3.58}$$

We have shown above that also $\mathbf{Cof}\,\mathbf{Q}$ is unitary for unitary $\mathbf{Q}$. Using the same arguments again and the estimate for $|\mathbf{Cof}\,\mathbf{R}-\mathbf{I}|^2$ we obtain

$$\begin{aligned}
|\mathbf{Cof}\,\mathbf{A}-\mathbf{I}|^2&=|(\mathbf{Cof}\,\mathbf{Q})(\mathbf{Cof}\,\mathbf{R})(\mathbf{Cof}\,\mathbf{Q})^*-\mathbf{I}|^2=|(\mathbf{Cof}\,\mathbf{Q})(\mathbf{Cof}\,\mathbf{R}-\mathbf{I})(\mathbf{Cof}\,\mathbf{Q})^*|^2\\
&=|\mathbf{Cof}\,\mathbf{R}-\mathbf{I}|^2\le 6|\mathbf{R}-\mathbf{I}|^2+3|\mathbf{R}-\mathbf{I}|^4=6|\mathbf{A}-\mathbf{I}|^2+3|\mathbf{A}-\mathbf{I}|^4,
\end{aligned}$$

i.e. the statement.

□

Corollary 3.16:
For a matrix $\mathbf{A}\in\mathbb{R}^{3\times3}$ with $|\mathbf{A}-\mathbf{I}|\le 1$ it holds $|\mathbf{A}-\mathbf{I}|^4\le|\mathbf{A}-\mathbf{I}|^2$ and therefore by Lemma 3.15

$$|\mathbf{Cof}\,\mathbf{A}-\mathbf{I}|\le 3|\mathbf{A}-\mathbf{I}|.$$

Lemma 3.17:
Let $\mathbf{A}\in\mathbb{R}^{n\times n}$ be an arbitrary matrix and $\mathbf{dev}\,\mathbf{A}:=\mathbf{A}-\frac{1}{n}\mathrm{tr}(\mathbf{A})\mathbf{I}$ the deviator of $\mathbf{A}$. Then it holds for all $c\in\mathbb{R}$

$$|\mathbf{dev}\,\mathbf{A}+c\,\mathbf{I}|^2=|\mathbf{dev}\,\mathbf{A}|^2+n\,c^2.$$

Proof:

By definition of the inner product (2.3) and its induced (Frobenius) norm in Section 2.1.2 it holds

$$\begin{aligned}|\mathbf{dev}\,\mathbf{A} + c\,\mathbf{I}|^2 &= |\mathbf{dev}\,\mathbf{A}|^2 + 2\,(\mathbf{dev}\,\mathbf{A} : c\,\mathbf{I}) + |c\,\mathbf{I}|^2 \\ &= |\mathbf{dev}\,\mathbf{A}|^2 + 2c\underbrace{\mathrm{tr}(\mathbf{dev}\,\mathbf{A})}_{=0} + c^2|\mathbf{I}|^2 = |\mathbf{dev}\,\mathbf{A}|^2 + n\,c^2.\end{aligned}$$

□

Corollary 3.18:

For arbitrary $\mathbf{A} \in \mathbb{R}^{n\times n}$ it holds

$$|\mathbf{A} - \mathbf{I}|^2 = |\mathbf{dev}\,\mathbf{A}|^2 + n\left(\frac{1}{n}\mathrm{tr}(\mathbf{A}) - 1\right)^2.$$

Proof:

We split $\mathbf{A}$ into its trace and deviatoric part and obtain with the help of Lemma 3.17

$$\begin{aligned}|\mathbf{A} - \mathbf{I}|^2 = |\mathbf{dev}\,\mathbf{A} + \frac{1}{n}\mathrm{tr}(\mathbf{A})\mathbf{I} - \mathbf{I}|^2 &= |\mathbf{dev}\,\mathbf{A} + \underbrace{\left(\frac{1}{n}\mathrm{tr}(\mathbf{A}) - 1\right)}_{=:c}\mathbf{I}|^2 \\ &= |\mathbf{dev}\,\mathbf{A}|^2 + n\left(\frac{1}{n}\mathrm{tr}(\mathbf{A}) - 1\right)^2.\end{aligned}$$

□

Lemma 3.19:

For arbitrary matrix - valued functions $\mathbf{A} \in L^\infty(\Omega)^{n\times n}$ and $\mathbf{B} \in L^2(\Omega)^{n\times n}$ it holds

$$\|\mathbf{A} : \mathbf{B}\|_{L^2(\Omega)} \le \|\mathbf{A}\|_{L^\infty(\Omega)}\|\mathbf{B}\|_{L^2(\Omega)}.$$

Proof:

It holds with the help of the Cauchy - Schwarz inequality and Remark 2.15

$$\begin{aligned}\|\mathbf{A} : \mathbf{B}\|^2_{L^2(\Omega)} = \int_\Omega |\mathbf{A} : \mathbf{B}|^2\,dx &\le \int_\Omega |\mathbf{A}|^2|\mathbf{B}|^2\,dx \\ &\le \|\mathbf{A}\|^2_{L^\infty(\Omega)}\int_\Omega |\mathbf{B}|^2\,dx = \|\mathbf{A}\|^2_{L^\infty(\Omega)}\|\mathbf{B}\|^2_{L^2(\Omega)},\end{aligned}$$

i.e. the statement after extracting the square root.

□

The next lemma gives us an estimate for small perturbations of $\mathcal{A}'_{NH}(\boldsymbol{\Xi})[\boldsymbol{\Sigma}]$ about $\boldsymbol{\Xi} = \mathbf{0}$.

Lemma 3.20:

If $\boldsymbol{\Xi} \in L^\infty(\Omega)^{3\times 3}$ satisfies

$$\boxed{\begin{aligned}\left\|\frac{\mathbf{dev}\,\boldsymbol{\Xi}}{\mu}\right\|_{L^\infty(\Omega)} \le a, \quad \left\|\frac{\mathrm{tr}(\boldsymbol{\Xi})}{\lambda}\right\|_{L^\infty(\Omega)} \le b, \\ \mathrm{tr}(\mathcal{A}_{NH}(\boldsymbol{\Xi})) > 0, \quad \mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi}))) \ge 2\end{aligned}} \tag{3.59}$$

with $0 \leq a < \sqrt{\frac{1}{6\sqrt{3}}}$, $0 \leq b \leq \frac{\sqrt{3}}{2}$ and

$$\hat{C} := \frac{2b}{\sqrt{3} - 18a^2} + a^2 \left(\frac{18 + \sqrt{3}a}{1 - 6a^2\sqrt{3}} \right) + a \leq 1, \tag{3.60}$$

then it holds

$$\begin{aligned} \|\mathcal{A}'_{NH}(\boldsymbol{\Xi})[\boldsymbol{\Sigma}] - \mathcal{A}'_{NH}(\mathbf{0})[\boldsymbol{\Sigma}]\|_{L^2(\Omega)} \leq \frac{3\sqrt{3}}{\mu} \Bigg[\left(\frac{2}{\sqrt{3} - 18a^2} \right) \left\| \frac{\operatorname{tr}(\boldsymbol{\Xi})}{\lambda} \right\|_{L^\infty(\Omega)} \\ + \left(a \left(\frac{18 + \sqrt{3}a}{1 - 6a^2\sqrt{3}} \right) + 1 \right) \left\| \frac{\mathbf{dev}\, \boldsymbol{\Xi}}{\mu} \right\|_{L^\infty(\Omega)} \Bigg] \|\boldsymbol{\Sigma}\|_{L^2(\Omega)} \end{aligned} \tag{3.61}$$

for all $\boldsymbol{\Sigma} \in L^2(\Omega)^{3\times 3}$ and the operator $\mathcal{A}_{NH}$, defined by the first equation in (3.36) and the cubic equation (3.38), provided that its discriminant is positive. Moreover there exists a constant $C > 0$, depending on λ, μ and a, such that

$$\|\mathcal{A}'_{NH}(\boldsymbol{\Xi})[\boldsymbol{\Sigma}] - \mathcal{A}'_{NH}(\mathbf{0})[\boldsymbol{\Sigma}]\|_{L^2(\Omega)} \leq C \|\boldsymbol{\Xi}\|_{L^\infty(\Omega)} \|\boldsymbol{\Sigma}\|_{L^2(\Omega)}. \tag{3.62}$$

Proof:

By assumption it holds $\boldsymbol{\Xi} \in L^\infty(\Omega)^{3\times 3}$. Under this assumption, following the same steps as in the proof of Lemma 3.13, the corresponding strain $\mathcal{A}_{NH}(\boldsymbol{\Xi})$ is in $L^\infty(\Omega)^{3\times 3}$ and thus $\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})) \in L^\infty(\Omega)^{3\times 3}$. Obviously for arbitrary $f \in L^2(\Omega)$ it holds

$$\|f\,\mathbf{I}\|^2_{L^2(\Omega)} = \int_\Omega |f\,\mathbf{I}|^2\, dx = \int_\Omega |f|^2 |\mathbf{I}|^2\, dx = 3\|f\|^2_{L^2(\Omega)}. \tag{3.63}$$

With the help of (3.63), Lemma 3.19 in combination with the representations of $\mathcal{A}'_{NH}$ in (3.51) and (3.52) for general $\boldsymbol{\Xi}$ and $\boldsymbol{\Xi} = \mathbf{0}$ it follows for arbitrary $\boldsymbol{\Sigma} \in L^2(\Omega)^{3\times 3}$

$$\begin{aligned} \|\mathcal{A}'_{NH}(\boldsymbol{\Xi})[\boldsymbol{\Sigma}] - \mathcal{A}'_{NH}(\mathbf{0})[\boldsymbol{\Sigma}]\|_{L^2(\Omega)} &= \left\| \frac{1}{\mu} \left(\frac{\lambda \operatorname{tr}(\boldsymbol{\Sigma})}{2\mu + 3\lambda} \mathbf{I} - \frac{\lambda (\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})) : \boldsymbol{\Sigma})}{2\mu + \lambda \operatorname{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))} \mathbf{I} \right) \right\|_{L^2(\Omega)} \\ &= \frac{\sqrt{3}}{\mu} \left\| \frac{\lambda\, (\mathbf{I} : \boldsymbol{\Sigma})}{2\mu + 3\lambda} - \frac{\lambda (\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})) : \boldsymbol{\Sigma})}{2\mu + \lambda \operatorname{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))} \right\|_{L^2(\Omega)} \\ &= \frac{\sqrt{3}}{\mu} \left\| \left(\frac{\lambda\, \mathbf{I}}{2\mu + 3\lambda} - \frac{\lambda\, \mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi}))}{2\mu + \lambda \operatorname{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))} \right) : \boldsymbol{\Sigma} \right\|_{L^2(\Omega)} \\ &\leq \frac{\sqrt{3}}{\mu} \left\| \frac{\lambda\, \mathbf{I}}{2\mu + 3\lambda} - \frac{\lambda\, \mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi}))}{2\mu + \lambda \operatorname{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))} \right\|_{L^\infty(\Omega)} \|\boldsymbol{\Sigma}\|_{L^2(\Omega)}. \end{aligned} \tag{3.64}$$

It remains to show that

$$\left| \frac{\lambda\, \mathbf{I}}{2\mu + 3\lambda} - \frac{\lambda\, \mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi}))}{2\mu + \lambda \operatorname{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))} \right| \tag{3.65}$$

is bounded from above.

To this end we use Corollary 3.18 for $n = 3$ and obtain

$$|\mathcal{A}_{NH}(\boldsymbol{\Xi}) - \mathbf{I}|^2 = |\mathbf{dev}\, \mathcal{A}_{NH}(\boldsymbol{\Xi})|^2 + 3 \left(\frac{1}{3} \operatorname{tr}(\mathcal{A}_{NH}(\boldsymbol{\Xi})) - 1 \right)^2,$$

which leads after extracting the square root and the use of $(a_1^2 + a_2^2)^{\frac{1}{2}} \leq a_1 + a_2$ for $a_1, a_2 \geq 0$ to

$$|\mathcal{A}_{NH}(\boldsymbol{\Xi}) - \mathbf{I}| \leq |\mathbf{dev}\, \mathcal{A}_{NH}(\boldsymbol{\Xi})| + \sqrt{3} \left| \frac{1}{3}\mathrm{tr}(\mathcal{A}_{NH}(\boldsymbol{\Xi})) - 1 \right|. \tag{3.66}$$

Both terms are bounded individually as we will see in the following. For the deviator term we have by the first equation in (3.36)

$$|\mathbf{dev}\, \mathcal{A}_{NH}(\boldsymbol{\Xi})| = \left| \frac{\mathbf{dev}\, \boldsymbol{\Xi}}{\mu} \right|. \tag{3.67}$$

Equation (3.38) with $\mathbf{B} = \mathcal{A}_{NH}(\boldsymbol{\Xi})$ is after dividing it by 27 equivalent to

$$\begin{aligned} \left(\frac{1}{3}\mathrm{tr}(\mathcal{A}_{NH}(\boldsymbol{\Xi}))\right)^3 + \left(\frac{1}{\mu^2}\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\, \boldsymbol{\Xi})) + \frac{2\mu}{\lambda}\right)\left(\frac{1}{3}\mathrm{tr}(\mathcal{A}_{NH}(\boldsymbol{\Xi}))\right) - 1 - \frac{2\mu}{\lambda} \\ = \frac{2}{3\lambda}\mathrm{tr}(\boldsymbol{\Xi}) - \frac{1}{\mu^3}\det(\mathbf{dev}\, \boldsymbol{\Xi}). \end{aligned}$$

We subtract on both sides the term $-\frac{1}{\mu^2}\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\, \boldsymbol{\Xi}))$ and get

$$\begin{aligned} &\underbrace{\left(\frac{1}{3}\mathrm{tr}(\mathcal{A}_{NH}(\boldsymbol{\Xi})) - 1\right)}_{=:x} \underbrace{\left(\left(\frac{1}{3}\mathrm{tr}(\mathcal{A}_{NH}(\boldsymbol{\Xi}))\right)^2 + \frac{1}{3}\mathrm{tr}(\mathcal{A}_{NH}(\boldsymbol{\Xi})) + 1 + \frac{2\mu}{\lambda} + \frac{1}{\mu^2}\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\, \boldsymbol{\Xi}))\right)}_{=:y} \\ &\qquad = \underbrace{\frac{2}{3\lambda}\mathrm{tr}(\boldsymbol{\Xi}) - \frac{1}{\mu^3}\det(\mathbf{dev}\, \boldsymbol{\Xi}) - \frac{1}{\mu^2}\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\, \boldsymbol{\Xi}))}_{=:z}. \end{aligned} \tag{3.68}$$

Obviously Corollary 2.33 and assumption (3.59) lead to the estimate

$$\begin{aligned} &\frac{1}{\mu^2}\left|\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\, \boldsymbol{\Xi}))\right| \leq \frac{6\sqrt{3}}{\mu^2}|\mathbf{dev}\, \boldsymbol{\Xi}|^2 = 6\sqrt{3}\left|\frac{\mathbf{dev}\, \boldsymbol{\Xi}}{\mu}\right|^2 \leq 6a^2\sqrt{3} < 1 \\ \Leftrightarrow\ & 1 - \frac{1}{\mu^2}\left|\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\, \boldsymbol{\Xi}))\right| \geq 1 - 6a^2\sqrt{3} \quad (> 0) \\ \Leftrightarrow\ & \frac{1}{1 - \frac{1}{\mu^2}\left|\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\, \boldsymbol{\Xi}))\right|} \leq \frac{1}{1 - 6a^2\sqrt{3}}. \end{aligned} \tag{3.69}$$

With $y_1 := 1 + \frac{1}{\mu^2}\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\, \boldsymbol{\Xi}))$ and $y_2 := \left(\frac{1}{3}\mathrm{tr}(\mathcal{A}_{NH}(\boldsymbol{\Xi}))\right)^2 + \frac{1}{3}\mathrm{tr}(\mathcal{A}_{NH}(\boldsymbol{\Xi})) + \frac{2\mu}{\lambda}$ we have $y = y_1 + y_2$. By the identity (3.68) it holds $xy = z$ and therefore $|x| = \frac{|z|}{|y|}$, provided that $y \neq 0$. Due to the assumption $\mathrm{tr}(\mathcal{A}_{NH}(\boldsymbol{\Xi})) > 0$ and positive Lamé constants λ, μ it holds obviously $y_2 > 0$. y_1 is positive if and only if $\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\, \boldsymbol{\Xi})) > -\mu^2$. This is ensured, since we have either for $\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\, \boldsymbol{\Xi})) > 0$ the conclusion $\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\, \boldsymbol{\Xi})) > -\mu^2$ or for $\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\, \boldsymbol{\Xi})) \leq 0$ we have the implication $\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\, \boldsymbol{\Xi})) \geq (-6a^2\sqrt{3})\mu^2 > -\mu^2$ by (3.69). Altogether we have proven that it holds $y_1, y_2 > 0$. This implies $y = y_1 + y_2 > 0$ and the estimate

$$|y| = y = y_1 + y_2 \geq y_1 = |y_1| \Leftrightarrow \frac{1}{|y|} \leq \frac{1}{|y_1|} \Rightarrow |x| = \frac{|z|}{|y|} \leq \frac{|z|}{|y_1|}.$$

Additionally the reverse triangle inequality and (3.69) imply

$$y_1 = |y_1| = \left|1 + \frac{1}{\mu^2}\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\Xi}))\right| \geq \left|1 - \frac{1}{\mu^2}|\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\Xi}))|\right| \geq 1 - 6a^2\sqrt{3}.$$

Altogether we obtain

$$\begin{aligned}
\left|\frac{1}{3}\mathrm{tr}(\mathcal{A}_{NH}(\boldsymbol{\Xi})) - 1\right| = |x| \leq \frac{|z|}{|y_1|} &\leq \frac{\left|\frac{2}{3\lambda}\mathrm{tr}(\boldsymbol{\Xi}) - \frac{1}{\mu^3}\det(\mathbf{dev}\,\boldsymbol{\Xi}) - \frac{1}{\mu^2}\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\Xi}))\right|}{1 - \frac{1}{\mu^2}|\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\Xi}))|} \\
&\leq \left(\frac{1}{1 - 6a^2\sqrt{3}}\right)\left(\left|\frac{2}{3\lambda}\mathrm{tr}(\boldsymbol{\Xi})\right| + \left|\frac{1}{\mu^3}\det(\mathbf{dev}\,\boldsymbol{\Xi})\right| + \left|\frac{1}{\mu^2}\mathrm{tr}(\mathbf{Cof}\,(\mathbf{dev}\,\boldsymbol{\Xi}))\right|\right) \\
&\leq \left(\frac{1}{1 - 6a^2\sqrt{3}}\right)\left(\frac{2}{3}\left|\frac{\mathrm{tr}(\boldsymbol{\Xi})}{\lambda}\right| + 6\sqrt{3}\left|\frac{\mathbf{dev}\,\boldsymbol{\Xi}}{\mu}\right|^2 + \left|\frac{\mathbf{dev}\,\boldsymbol{\Xi}}{\mu}\right|^3\right) \\
&\leq \left(\frac{1}{1 - 6a^2\sqrt{3}}\right)\left(\frac{2}{3}\left|\frac{\mathrm{tr}(\boldsymbol{\Xi})}{\lambda}\right| + \left(6a\sqrt{3} + a^2\right)\left|\frac{\mathbf{dev}\,\boldsymbol{\Xi}}{\mu}\right|\right).
\end{aligned} \tag{3.70}$$

In the last steps here we have combined Corollary 2.33 and 2.35 and have used $\left|\frac{\mathbf{dev}\,\boldsymbol{\Xi}}{\mu}\right|^n \leq a^{n-1}\left|\frac{\mathbf{dev}\,\boldsymbol{\Xi}}{\mu}\right|$ for $n \in \mathbb{N} \setminus \{0\}$ which holds by assumption (3.59).
Plugging (3.70) into (3.66) and using (3.67) lead to

$$\begin{aligned}
|\mathcal{A}_{NH}(\boldsymbol{\Xi}) - \mathbf{I}| &\leq |\mathbf{dev}\,\mathcal{A}_{NH}(\boldsymbol{\Xi})| + \sqrt{3}\left|\frac{1}{3}\mathrm{tr}(\mathcal{A}_{NH}(\boldsymbol{\Xi})) - 1\right| \\
&\leq \left|\frac{\mathbf{dev}\,\boldsymbol{\Xi}}{\mu}\right| + \left(\frac{\sqrt{3}}{1 - 6a^2\sqrt{3}}\right)\left(\frac{2}{3}\left|\frac{\mathrm{tr}(\boldsymbol{\Xi})}{\lambda}\right| + \left(6a\sqrt{3} + a^2\right)\left|\frac{\mathbf{dev}\,\boldsymbol{\Xi}}{\mu}\right|\right) \\
&= \left(\frac{2}{\sqrt{3} - 18a^2}\right)\left|\frac{\mathrm{tr}(\boldsymbol{\Xi})}{\lambda}\right| + \left(a\left(\frac{18 + \sqrt{3}a}{1 - 6a^2\sqrt{3}}\right) + 1\right)\left|\frac{\mathbf{dev}\,\boldsymbol{\Xi}}{\mu}\right| (\leq \hat{C})
\end{aligned} \tag{3.71}$$

by assumption. Until now we have derived an estimate for $|\mathcal{A}_{NH}(\boldsymbol{\Xi}) - \mathbf{I}|$. To prove finally that (3.65) is bounded we observe

$$\begin{aligned}
&\left|\frac{\lambda\,\mathbf{I}}{2\mu + 3\lambda} - \frac{\lambda\,\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi}))}{2\mu + \lambda\,\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))}\right| = \left|\frac{\lambda\,\mathbf{I}}{2\mu + 3\lambda} - \frac{\lambda\,\mathbf{I}}{2\mu + \lambda\,\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))}\right. \\
&\qquad\qquad \left. + \frac{\lambda\,\mathbf{I}}{2\mu + \lambda\,\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))} - \frac{\lambda\,\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi}))}{2\mu + \lambda\,\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))}\right| \\
&\leq \left|\frac{\lambda\,\mathbf{I}}{2\mu + 3\lambda} - \frac{\lambda\,\mathbf{I}}{2\mu + \lambda\,\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))}\right| + \left|\frac{\lambda(\mathbf{I} - \mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))}{2\mu + \lambda\,\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))}\right| \\
&= \left|\frac{2\mu\lambda + \lambda^2\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi}))) - 2\mu\lambda - 3\lambda^2}{(2\mu + 3\lambda)(2\mu + \lambda\,\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi}))))}\mathbf{I}\right| + \left|\frac{\lambda(\mathbf{I} - \mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))}{2\mu + \lambda\,\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))}\right| \\
&= \left|\frac{\lambda^2\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})) - \mathbf{I})}{(2\mu + 3\lambda)(2\mu + \lambda\,\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi}))))}\mathbf{I}\right| + \left|\frac{\lambda(\mathbf{I} - \mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))}{2\mu + \lambda\,\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))}\right| \\
&\leq \sqrt{3}\left|\frac{\lambda^2\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})) - \mathbf{I})}{(2\mu + 3\lambda)(2\mu + 2\lambda)}\right| + \left|\frac{\lambda(\mathbf{I} - \mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))}{2\mu + 2\lambda}\right| \\
&\leq \frac{\sqrt{3}}{6}|\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})) - \mathbf{I})| + \frac{1}{2}|\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})) - \mathbf{I}| \\
&\leq \frac{1}{2}|\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})) - \mathbf{I}| + \frac{1}{2}|\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})) - \mathbf{I}| = |\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})) - \mathbf{I}|.
\end{aligned}$$

In the last steps we have used:

- $(2\mu+2\lambda)^{-1} \geq (2\mu+\lambda\,\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi}))))^{-1} \Leftrightarrow \mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi}))) \geq 2$
  (holds by assumption (3.59))
- $\frac{\lambda^2}{(2\mu+3\lambda)(2\mu+2\lambda)} \leq \frac{1}{6} \Leftrightarrow 6\lambda^2 \leq (2\mu+3\lambda)(2\mu+2\lambda) = 4\mu^2+10\mu\lambda+6\lambda^2$
  (holds obviously for $\lambda,\mu>0$)
- $\frac{\lambda}{2\mu+2\lambda} \leq \frac{1}{2} \Leftrightarrow \lambda \leq \frac{1}{2}(2\mu+2\lambda) = \mu+\lambda$ (holds obviously for $\mu>0$)

It follows immediately

$$\left\| \frac{\lambda\,\mathbf{I}}{2\mu+3\lambda} - \frac{\lambda\,\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi}))}{2\mu+\lambda\,\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))} \right\|_{L^\infty(\Omega)} \leq \|\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})) - \mathbf{I}\|_{L^\infty(\Omega)}. \tag{3.72}$$

As long as $|\mathcal{A}_{NH}(\boldsymbol{\Xi}) - \mathbf{I}| \leq 1$, which is satisfied by assumption, we obtain by Corollary 3.16 and (3.71)

$$\begin{aligned} |\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})) - \mathbf{I}| &\leq 3|\mathcal{A}_{NH}(\boldsymbol{\Xi}) - \mathbf{I}| \\ &\leq 3\left[\left(\frac{2}{\sqrt{3}-18a^2}\right)\left|\frac{\mathrm{tr}(\boldsymbol{\Xi})}{\lambda}\right| + \left(a\left(\frac{18+\sqrt{3}a}{1-6a^2\sqrt{3}}\right)+1\right)\left|\frac{\mathbf{dev}\,\boldsymbol{\Xi}}{\mu}\right|\right]. \end{aligned} \tag{3.73}$$

Combining (3.64), (3.72) and (3.73) ends up in

$$\begin{aligned} \|\mathcal{A}'_{NH}(\boldsymbol{\Xi})[\boldsymbol{\Sigma}] - \mathcal{A}'_{NH}(\mathbf{0})[\boldsymbol{\Sigma}]\|_{L^2(\Omega)} \leq \frac{3\sqrt{3}}{\mu}\Bigg[&\left(\frac{2}{\sqrt{3}-18a^2}\right)\left\|\frac{\mathrm{tr}(\boldsymbol{\Xi})}{\lambda}\right\|_{L^\infty(\Omega)} \\ &+\left(a\left(\frac{18+\sqrt{3}a}{1-6a^2\sqrt{3}}\right)+1\right)\left\|\frac{\mathbf{dev}\,\boldsymbol{\Xi}}{\mu}\right\|_{L^\infty(\Omega)}\Bigg]\,\|\boldsymbol{\Sigma}\|_{L^2(\Omega)}, \end{aligned}$$

i.e. the first statement (3.61).
We know due to Lemma 3.17

$$|\boldsymbol{\Xi}|^2 = \left|\mathbf{dev}\,\boldsymbol{\Xi} + \frac{1}{3}\mathrm{tr}(\boldsymbol{\Xi})\mathbf{I}\right|^2 = |\mathbf{dev}\,\boldsymbol{\Xi}|^2 + 3\left(\frac{1}{3}\mathrm{tr}(\boldsymbol{\Xi})\right)^2 \geq |\mathbf{dev}\,\boldsymbol{\Xi}|^2,$$

i.e. $|\mathbf{dev}\,\boldsymbol{\Xi}| \leq |\boldsymbol{\Xi}|$, and by Lemma 2.31 the inequality $|\mathrm{tr}(\boldsymbol{\Xi})| \leq \sqrt{3}|\boldsymbol{\Xi}|$. Consequently

$$\begin{aligned} \|\mathcal{A}'_{NH}(\boldsymbol{\Xi})[\boldsymbol{\Sigma}] - \mathcal{A}'_{NH}(\mathbf{0})[\boldsymbol{\Sigma}]\|_{L^2(\Omega)} &\leq \frac{3\sqrt{3}}{\mu}\Bigg[\left(\frac{2}{\sqrt{3}-18a^2}\right)\left\|\frac{\mathrm{tr}(\boldsymbol{\Xi})}{\lambda}\right\|_{L^\infty(\Omega)} \\ &\quad+\left(a\left(\frac{18+\sqrt{3}a}{1-6a^2\sqrt{3}}\right)+1\right)\left\|\frac{\mathbf{dev}\,\boldsymbol{\Xi}}{\mu}\right\|_{L^\infty(\Omega)}\Bigg]\,\|\boldsymbol{\Sigma}\|_{L^2(\Omega)} \\ &\leq \frac{3\sqrt{3}}{\mu}\left[\left(\frac{2}{\sqrt{3}-18a^2}\right)\frac{\sqrt{3}}{\lambda} + \frac{1}{\mu}\left(a\left(\frac{18+\sqrt{3}a}{1-6a^2\sqrt{3}}\right)+1\right)\right]\,\|\boldsymbol{\Xi}\|_{L^\infty(\Omega)}\|\boldsymbol{\Sigma}\|_{L^2(\Omega)} \\ &=: C\|\boldsymbol{\Xi}\|_{L^\infty(\Omega)}\|\boldsymbol{\Sigma}\|_{L^2(\Omega)}, \end{aligned}$$

i.e. the second statement (3.62).

□

Remark 3.21:

1. The parameter λ which is characteristic for an incompressible material appears in the constant C only in one denominator. Hence for $\lambda \geq 1$ the constant can be estimated by a constant that is independent of λ. In particular, the constant cannot blow up for $\lambda \to \infty$.

2. The assumption $\mathrm{tr}(\mathcal{A}_{NH}(\boldsymbol{\Xi})) > 0$ is automatically satisfied if $a^3 + \frac{2}{3}b - 1 < 0$ and $\left(a^3 + \frac{2}{3}b - 1\right)^2 > 96a^6\sqrt{3}$ (cf. Proposition 3.7).
   If we choose additionally the pair $(a, b) \in \mathbb{R}^2$ sufficiently small such that the condition $|\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi}))-\mathbf{I})| \leq 1$ is satisfied, then we automatically have $\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})))$ $\in [2, 4]$, i.e. in particular the condition $\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi}))) \geq 2$ of assumption (3.59) holds. Hence we seek pairs $(a, b) \in \mathbb{R}^2$ such that $|\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})) - \mathbf{I})| \leq 1$ is satisfied.
   Combining Lemma 2.31, Corollary 3.16 and equation (3.71) lead to

$$\begin{aligned} |\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})) - \mathbf{I})| &\leq \sqrt{3}|\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})) - \mathbf{I}| \\ &\leq 3\sqrt{3}|\mathcal{A}_{NH}(\boldsymbol{\Xi}) - \mathbf{I}| \leq 3\sqrt{3}\hat{C} =: \bar{C}. \end{aligned}$$

   Thus for $0 \leq \bar{C} \leq 1$ or equivalently $0 \leq \hat{C} \leq \frac{1}{3\sqrt{3}} \approx 0.1925$ the condition $\mathrm{tr}(\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi}))) \geq 2$ is ensured.

3. In the proof of Lemma 3.20 we have chosen a and b such that the constant $\hat{C}$ in (3.60) is less than or equal to one. With this choice and due to (3.71) it was possible to use Corollary 3.16 in (3.73). If we do not assume $\hat{C} \leq 1$, we can use Lemma 3.15 instead of Corollary 3.16 to obtain an estimate for $|\mathbf{Cof}\,(\mathcal{A}_{NH}(\boldsymbol{\Xi})) - \mathbf{I}|$. However, also in this case, an inequality of the form (3.62) can be achieved.

By Remark 3.21 we know that Lemma 3.20 holds at least for small stresses $\boldsymbol{\Xi}$. In numerical simulations one could easily prove the conditions in (3.59) and (3.60). In the following example we state exemplarily two values a and b for which the assumptions are satisfied. Note that this choice is not optimal.

Example 3.22:
Choosing for instance $a = \frac{1}{24} < \sqrt{\frac{1}{6\sqrt{3}}}$ and $b = \frac{1}{12} \leq \frac{\sqrt{3}}{2}$ the discriminant of the cubic equation (3.38) is positive and the condition $\mathrm{tr}(\mathcal{A}_{NH}(\boldsymbol{\Xi})) > 0$ is automatically satisfied (cf. Example 3.8). Furthermore with this choice we have $\hat{C} \approx 0.1716 < 0.1925 \leq 1$. Therefore with the help of the second part of Remark 3.21 all assumptions of Lemma 3.20 are satisfied.

Lemma 3.23: (Estimate for $\mathcal{R}'_{NH}(\mathbf{Q}, \mathbf{v})$ near the origin)
If $\theta > 0$ in the definition of $\boldsymbol{\Pi}^\infty$ and $\mathbf{U}^\infty$ (cf. (3.55)) is sufficiently small, then there is a

$\rho \in [0,1)$ such that

$$\|\mathcal{R}'_{NH}(\mathbf{Q},\mathbf{v})[\hat{\mathbf{Q}},\hat{\mathbf{v}}] - \mathcal{R}'_{NH}(\mathbf{0},\mathbf{0})[\hat{\mathbf{Q}},\hat{\mathbf{v}}]\|_{L^2(\Omega)} \leq \rho \|\mathcal{R}'_{NH}(\mathbf{0},\mathbf{0})[\hat{\mathbf{Q}},\hat{\mathbf{v}}]\|_{L^2(\Omega)}$$

holds for all $(\mathbf{Q},\mathbf{v}) \in \mathbf{\Pi}^\infty \times \mathbf{U}^\infty$ and $(\hat{\mathbf{Q}},\hat{\mathbf{v}}) \in H_{\Gamma_N}(\mathrm{div};\,\Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$.

Proof:

The proof follows the steps in the proof of Lemma 4.3 of [MSSS14], but is explained in more detail here.
Let $(\mathbf{Q},\mathbf{v}) \in \mathbf{\Pi}^\infty \times \mathbf{U}^\infty$ and $(\hat{\mathbf{Q}},\hat{\mathbf{v}}) \in H_{\Gamma_N}(\mathrm{div};\,\Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$ be arbitrary and therefore in particular $\hat{\mathbf{Q}} \in L^2(\Omega)^{3\times 3}$.
Inserting (3.53) and (3.54) (with $\omega_1 = \omega_2 = 1$) leads to

$$\begin{aligned}
&\|\mathcal{R}'_{NH}(\mathbf{Q},\mathbf{v})[\hat{\mathbf{Q}},\hat{\mathbf{v}}] - \mathcal{R}'_{NH}(\mathbf{0},\mathbf{0})[\hat{\mathbf{Q}},\hat{\mathbf{v}}]\|_{L^2(\Omega)} \\
&\qquad \leq \|\mathcal{A}'_{NH}(\mathbf{Q}\mathbf{F}(\mathbf{v})^T)[\hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T] - \mathcal{A}'_{NH}(\mathbf{0})[\hat{\mathbf{Q}}]\|_{L^2(\Omega)} \\
&\qquad\quad + \|\nabla\hat{\mathbf{v}}\mathbf{F}(\mathbf{v})^T + \mathbf{F}(\mathbf{v})(\nabla\hat{\mathbf{v}})^T - \nabla\hat{\mathbf{v}} - (\nabla\hat{\mathbf{v}})^T\|_{L^2(\Omega)}
\end{aligned} \tag{3.74}$$

by the triangle inequality. In the following we estimate both terms on the right - hand side of (3.74) individually. For the first term, by adding and subtracting an additional term at the same time and using the triangle inequality, we obtain

$$\begin{aligned}
&\|\mathcal{A}'_{NH}(\mathbf{Q}\mathbf{F}(\mathbf{v})^T)[\hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T] - \mathcal{A}'_{NH}(\mathbf{0})[\hat{\mathbf{Q}}]\|_{L^2(\Omega)} \\
&\leq \|\mathcal{A}'_{NH}(\mathbf{Q}\mathbf{F}(\mathbf{v})^T)[\hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T] - \mathcal{A}'_{NH}(\mathbf{0})[\hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T]\|_{L^2(\Omega)} \\
&\quad + \|\mathcal{A}'_{NH}(\mathbf{0})[\hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T] - \mathcal{A}'_{NH}(\mathbf{0})[\hat{\mathbf{Q}}]\|_{L^2(\Omega)}.
\end{aligned} \tag{3.75}$$

For the first term in (3.75) we use Lemma 3.20 with $\mathbf{\Xi} = \mathbf{Q}\mathbf{F}(\mathbf{v})^T \in L^\infty(\Omega)^{3\times 3}$, $\mathbf{\Sigma} = \hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T \in L^2(\Omega)^{3\times 3}$ and obtain

$$\begin{aligned}
&\|\mathcal{A}'_{NH}(\mathbf{Q}\mathbf{F}(\mathbf{v})^T)[\hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T] - \mathcal{A}'_{NH}(\mathbf{0})[\hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T]\|_{L^2(\Omega)} \\
&\qquad \leq C\|\mathbf{Q}\mathbf{F}(\mathbf{v})^T\|_{L^\infty(\Omega)}\|\hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T\|_{L^2(\Omega)},
\end{aligned} \tag{3.76}$$

where C is the constant in the proof of Lemma 3.20. Note that at least for sufficiently small θ the assumptions of Lemma 3.20 are satisfied.
Due to $\frac{2\mu}{2\mu+3\lambda} \leq 1 \Leftrightarrow 0 \leq \lambda$, which obviously holds, equation (3.52) and Lemma 3.17 $\Big($with $c = \frac{2\mu\,\mathrm{tr}(\mathbf{\Sigma})}{3(2\mu+3\lambda)}$ respectively $c = \frac{\mathrm{tr}(\mathbf{\Sigma})}{3}\Big)$ it follows for arbitrary $\mathbf{\Sigma} \in \mathbb{R}^{3\times 3}$

$$\begin{aligned}
|\mathcal{A}'_{NH}(\mathbf{0})[\mathbf{\Sigma}]|^2 &= \frac{1}{\mu^2}\left|\mathbf{\Sigma} - \frac{\lambda}{2\mu+3\lambda}\mathrm{tr}(\mathbf{\Sigma})\mathbf{I}\right|^2 = \frac{1}{\mu^2}\left|\mathbf{dev}\,\mathbf{\Sigma} + \left(\frac{1}{3} - \frac{\lambda}{2\mu+3\lambda}\right)\mathrm{tr}(\mathbf{\Sigma})\mathbf{I}\right|^2 \\
&= \frac{1}{\mu^2}\left|\mathbf{dev}\,\mathbf{\Sigma} + \frac{2\mu\,\mathrm{tr}(\mathbf{\Sigma})}{3(2\mu+3\lambda)}\mathbf{I}\right|^2 = \frac{1}{\mu^2}\left(|\mathbf{dev}\,\mathbf{\Sigma}|^2 + 3\left(\frac{2\mu\,\mathrm{tr}(\mathbf{\Sigma})}{3(2\mu+3\lambda)}\right)^2\right) \\
&\leq \frac{1}{\mu^2}\left(|\mathbf{dev}\,\mathbf{\Sigma}|^2 + 3\left(\frac{\mathrm{tr}(\mathbf{\Sigma})}{3}\right)^2\right) = \frac{1}{\mu^2}\left|\mathbf{dev}\,\mathbf{\Sigma} + \frac{1}{3}\mathrm{tr}(\mathbf{\Sigma})\mathbf{I}\right|^2 = \frac{1}{\mu^2}|\mathbf{\Sigma}|^2,
\end{aligned}$$

which implies $\|\mathcal{A}'_{NH}(\mathbf{0})[\boldsymbol{\Sigma}]\|^2_{L^2(\Omega)} \leq \frac{1}{\mu^2}\|\boldsymbol{\Sigma}\|^2_{L^2(\Omega)}$ (cf. Remark 2.15) for $\boldsymbol{\Sigma} \in L^2(\Omega)^{3\times 3}$. By assumption we know $\hat{\mathbf{Q}}(\nabla \mathbf{v})^T + \mathbf{Q}(\nabla \hat{\mathbf{v}})^T \in L^2(\Omega)^{3\times 3}$. Thus for the second term in (3.75) we obtain

$$\begin{aligned}
&\|\mathcal{A}'_{NH}(\mathbf{0})[\hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T] - \mathcal{A}'_{NH}(\mathbf{0})[\hat{\mathbf{Q}}]\|_{L^2(\Omega)} \\
&\quad = \|\mathcal{A}'_{NH}(\mathbf{0})[\hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T - \hat{\mathbf{Q}}]\|_{L^2(\Omega)} \\
&\quad = \|\mathcal{A}'_{NH}(\mathbf{0})[\hat{\mathbf{Q}}(\mathbf{I} + (\nabla\mathbf{v})^T) + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T - \hat{\mathbf{Q}}]\|_{L^2(\Omega)} \\
&\quad = \|\mathcal{A}'_{NH}(\mathbf{0})[\hat{\mathbf{Q}}(\nabla\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T]\|_{L^2(\Omega)} \\
&\quad \leq \frac{1}{\mu}\|\hat{\mathbf{Q}}(\nabla\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T\|_{L^2(\Omega)}.
\end{aligned} \tag{3.77}$$

Plugging (3.76) and (3.77) into (3.75) leads to

$$\begin{aligned}
&\|\mathcal{A}'_{NH}(\mathbf{Q}\mathbf{F}(\mathbf{v})^T)[\hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T] - \mathcal{A}'_{NH}(\mathbf{0})[\hat{\mathbf{Q}}]\|_{L^2(\Omega)} \\
&\leq C\|\mathbf{Q}\mathbf{F}(\mathbf{v})^T\|_{L^\infty(\Omega)}\|\hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T\|_{L^2(\Omega)} + \frac{1}{\mu}\|\hat{\mathbf{Q}}(\nabla\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T\|_{L^2(\Omega)}.
\end{aligned}$$

The norms $\|\mathbf{Q}\mathbf{F}(\mathbf{v})^T\|_{L^\infty(\Omega)}$, $\|\hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T\|_{L^2(\Omega)}$ and $\|\hat{\mathbf{Q}}(\nabla\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T\|_{L^2(\Omega)}$ can be further estimated by

$$\begin{aligned}
\|\mathbf{Q}\mathbf{F}(\mathbf{v})^T\|_{L^\infty(\Omega)} &= \|\mathbf{Q}\left(\mathbf{I} + (\nabla\mathbf{v})^T\right)\|_{L^\infty(\Omega)} \leq \|\mathbf{Q}\|_{L^\infty(\Omega)} + \|\mathbf{Q}(\nabla\mathbf{v})^T)\|_{L^\infty(\Omega)} \\
&\leq \|\mathbf{Q}\|_{L^\infty(\Omega)} + \|\mathbf{Q}\|_{L^\infty(\Omega)}\|\nabla\mathbf{v}\|_{L^\infty(\Omega)} \\
&\leq (\max\{1,\theta\})\left(\|\mathbf{Q}\|_{L^\infty(\Omega)} + \|\nabla\mathbf{v}\|_{L^\infty(\Omega)}\right), \\
\|\hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T\|_{L^2(\Omega)} &= \|\hat{\mathbf{Q}} + \hat{\mathbf{Q}}(\nabla\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T\|_{L^2(\Omega)} \\
&\leq \|\hat{\mathbf{Q}}\|_{L^2(\Omega)} + \|\hat{\mathbf{Q}}(\nabla\mathbf{v})^T\|_{L^2(\Omega)} + \|\mathbf{Q}(\nabla\hat{\mathbf{v}})^T\|_{L^2(\Omega)} \\
&\leq \|\hat{\mathbf{Q}}\|_{L^2(\Omega)} + \|\hat{\mathbf{Q}}\|_{L^2(\Omega)}\|\nabla\mathbf{v}\|_{L^\infty(\Omega)} + \|\mathbf{Q}\|_{L^\infty(\Omega)}\|\nabla\hat{\mathbf{v}}\|_{L^2(\Omega)} \\
&\leq \|\hat{\mathbf{Q}}\|_{L^2(\Omega)} + \theta\|\hat{\mathbf{Q}}\|_{L^2(\Omega)} + \theta\|\nabla\hat{\mathbf{v}}\|_{L^2(\Omega)} \\
&\leq (1+\theta)\left(\|\hat{\mathbf{Q}}\|_{L^2(\Omega)} + \|\nabla\hat{\mathbf{v}}\|_{L^2(\Omega)}\right), \\
\|\hat{\mathbf{Q}}(\nabla\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T\|_{L^2(\Omega)} &\leq \|\hat{\mathbf{Q}}\|_{L^2(\Omega)}\|\nabla\mathbf{v}\|_{L^\infty(\Omega)} + \|\nabla\hat{\mathbf{v}}\|_{L^2(\Omega)}\|\mathbf{Q}\|_{L^\infty(\Omega)} \\
&\leq \left(\|\mathbf{Q}\|^2_{L^\infty(\Omega)} + \|\nabla\mathbf{v}\|^2_{L^\infty(\Omega)}\right)^{\frac{1}{2}}\left(\|\hat{\mathbf{Q}}\|^2_{L^2(\Omega)} + \|\nabla\hat{\mathbf{v}}\|^2_{L^2(\Omega)}\right)^{\frac{1}{2}} \\
&\leq \left(\|\mathbf{Q}\|_{L^\infty(\Omega)} + \|\nabla\mathbf{v}\|_{L^\infty(\Omega)}\right)\left(\|\hat{\mathbf{Q}}\|_{L^2(\Omega)} + \|\nabla\hat{\mathbf{v}}\|_{L^2(\Omega)}\right)
\end{aligned}$$

using the triangle inequality, the generalized Hölder inequality (cf. Corollary 2.6 in [AF03]), the definition of $\boldsymbol{\Pi}^\infty$ and $\mathbf{U}^\infty$ (cf. (3.55)) and the inequalities

$$\begin{aligned}
ac + bd &\leq \left(a^2+b^2\right)^{\frac{1}{2}}\left(c^2+d^2\right)^{\frac{1}{2}}, \quad a,b,c,d \geq 0, \\
\left(a^2+b^2\right)^{\frac{1}{2}} &\leq a + b, \quad a,b \geq 0.
\end{aligned}$$

This ends up in

$$\begin{aligned}
&\|\mathcal{A}'_{NH}(\mathbf{Q}\mathbf{F}(\mathbf{v})^T)[\hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T] - \mathcal{A}'_{NH}(\mathbf{0})[\hat{\mathbf{Q}}]\|_{L^2(\Omega)} \\
&\leq C\|\mathbf{Q}\mathbf{F}(\mathbf{v})^T\|_{L^\infty(\Omega)}\|\hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T\|_{L^2(\Omega)} + \frac{1}{\mu}\|\hat{\mathbf{Q}}(\nabla\mathbf{v})^T + \mathbf{Q}(\nabla\hat{\mathbf{v}})^T\|_{L^2(\Omega)} \\
&\leq \underbrace{\left(C\ (\max\{1,\theta\})\,(1+\theta) + \frac{1}{\mu}\right)}_{=:\bar{C}(\theta,\mu)} \left(\|\mathbf{Q}\|_{L^\infty(\Omega)} + \|\nabla\mathbf{v}\|_{L^\infty(\Omega)}\right)\left(\|\hat{\mathbf{Q}}\|_{L^2(\Omega)} + \|\nabla\hat{\mathbf{v}}\|_{L^2(\Omega)}\right)
\end{aligned} \tag{3.78}$$

for the first term in (3.74).
For the second term in (3.74) we obtain

$$\begin{aligned}
&\|\nabla\hat{\mathbf{v}}\mathbf{F}(\mathbf{v})^T + \mathbf{F}(\mathbf{v})(\nabla\hat{\mathbf{v}})^T - \nabla\hat{\mathbf{v}} - (\nabla\hat{\mathbf{v}})^T\|_{L^2(\Omega)} \\
&\qquad = \|\nabla\hat{\mathbf{v}}(\mathbf{I} + (\nabla\mathbf{v})^T) + (\mathbf{I} + \nabla\mathbf{v})(\nabla\hat{\mathbf{v}})^T - \nabla\hat{\mathbf{v}} - (\nabla\hat{\mathbf{v}})^T\|_{L^2(\Omega)} \\
&\qquad = \|\nabla\hat{\mathbf{v}}(\nabla\mathbf{v})^T + \nabla\mathbf{v}(\nabla\hat{\mathbf{v}})^T\|_{L^2(\Omega)} \leq 2\|\nabla\mathbf{v}\|_{L^\infty(\Omega)}\|\nabla\hat{\mathbf{v}}\|_{L^2(\Omega)} \\
&\qquad \leq 2\left(\|\mathbf{Q}\|_{L^\infty(\Omega)} + \|\nabla\mathbf{v}\|_{L^\infty(\Omega)}\right)\left(\|\hat{\mathbf{Q}}\|_{L^2(\Omega)} + \|\nabla\hat{\mathbf{v}}\|_{L^2(\Omega)}\right).
\end{aligned} \tag{3.79}$$

Plugging (3.78) and (3.79) into (3.74) leads to

$$\begin{aligned}
&\|\mathcal{R}'_{NH}(\mathbf{Q},\mathbf{v})[\hat{\mathbf{Q}},\hat{\mathbf{v}}] - \mathcal{R}'_{NH}(\mathbf{0},\mathbf{0})[\hat{\mathbf{Q}},\hat{\mathbf{v}}]\|_{L^2(\Omega)} \\
&\qquad \leq \left(\bar{C}(\theta,\mu) + 2\right)\left(\|\mathbf{Q}\|_{L^\infty(\Omega)} + \|\nabla\mathbf{v}\|_{L^\infty(\Omega)}\right)\left(\|\hat{\mathbf{Q}}\|_{L^2(\Omega)} + \|\nabla\hat{\mathbf{v}}\|_{L^2(\Omega)}\right) \\
&\qquad \leq \left(\bar{C}(\theta,\mu) + 2\right)\left(\|\mathbf{Q}\|_{L^\infty(\Omega)} + \|\nabla\mathbf{v}\|_{L^\infty(\Omega)}\right)\left(\|\hat{\mathbf{Q}}\|_{L^2(\Omega)} + \|\hat{\mathbf{v}}\|_{H^1(\Omega)}\right).
\end{aligned}$$

Usage of Korn's inequality $\|\hat{\mathbf{v}}\|_{H^1(\Omega)} \leq \frac{1}{C_K}\|\boldsymbol{\varepsilon}(\hat{\mathbf{v}})\|_{L^2(\Omega)}$ for $\hat{\mathbf{v}} \in H^1_{\Gamma_D}(\Omega)^3$ (cf. Corollary 11.2.22 in [BS08]) with constant $C_K > 0$ leads to

$$\begin{aligned}
&\|\mathcal{R}'_{NH}(\mathbf{Q},\mathbf{v})[\hat{\mathbf{Q}},\hat{\mathbf{v}}] - \mathcal{R}'_{NH}(\mathbf{0},\mathbf{0})[\hat{\mathbf{Q}},\hat{\mathbf{v}}]\|^2_{L^2(\Omega)} \\
&\leq \left(\bar{C}(\theta,\mu) + 2\right)^2\left(\|\mathbf{Q}\|_{L^\infty(\Omega)} + \|\nabla\mathbf{v}\|_{L^\infty(\Omega)}\right)^2\left(\|\hat{\mathbf{Q}}\|_{L^2(\Omega)} + \frac{1}{C_K}\|\boldsymbol{\varepsilon}(\hat{\mathbf{v}})\|_{L^2(\Omega)}\right)^2 \\
&\lesssim \left(\|\mathbf{Q}\|_{L^\infty(\Omega)} + \|\nabla\mathbf{v}\|_{L^\infty(\Omega)}\right)^2\left(\|\hat{\mathbf{Q}}\|_{L^2(\Omega)} + \|\boldsymbol{\varepsilon}(\hat{\mathbf{v}})\|_{L^2(\Omega)}\right)^2 \\
&\lesssim \left(\|\mathbf{Q}\|_{L^\infty(\Omega)} + \|\nabla\mathbf{v}\|_{L^\infty(\Omega)}\right)^2\left(\|\hat{\mathbf{Q}}\|^2_{L^2(\Omega)} + \|\boldsymbol{\varepsilon}(\hat{\mathbf{v}})\|^2_{L^2(\Omega)}\right) \\
&\lesssim \left(\|\mathbf{Q}\|_{L^\infty(\Omega)} + \|\nabla\mathbf{v}\|_{L^\infty(\Omega)}\right)^2\left(\|\operatorname{div}\hat{\mathbf{Q}}\|^2_{L^2(\Omega)} + \|\hat{\mathbf{Q}}\|^2_{L^2(\Omega)} + \|\boldsymbol{\varepsilon}(\hat{\mathbf{v}})\|^2_{L^2(\Omega)}\right),
\end{aligned} \tag{3.80}$$

where we have used Young's inequality in the last but one estimate.
We have already observed in (3.54) and (3.8) that the operators

$$\mathcal{R}'_{NH}(\mathbf{0},\mathbf{0})[\hat{\mathbf{Q}},\hat{\mathbf{v}}] = \begin{pmatrix} \operatorname{div}\hat{\mathbf{Q}} \\ 2\left(\mathcal{A}_{lin}(\hat{\mathbf{Q}}) - \boldsymbol{\varepsilon}(\hat{\mathbf{v}})\right)\end{pmatrix}, \quad \mathcal{L}(\hat{\mathbf{Q}},\hat{\mathbf{v}}) = \begin{pmatrix} \operatorname{div}\hat{\mathbf{Q}} \\ \mathcal{A}_{lin}(\hat{\mathbf{Q}}) - \boldsymbol{\varepsilon}(\hat{\mathbf{v}})\end{pmatrix}$$

differ only up to a constant. Therefore it holds

$$\|\mathcal{L}(\hat{\mathbf{Q}},\hat{\mathbf{v}})\|_{L^2(\Omega)} \leq \|\mathcal{R}'_{NH}(\mathbf{0},\mathbf{0})[\hat{\mathbf{Q}},\hat{\mathbf{v}}]\|_{L^2(\Omega)} \leq 2\|\mathcal{L}(\hat{\mathbf{Q}},\hat{\mathbf{v}})\|_{L^2(\Omega)}. \tag{3.81}$$

We apply Theorem 3.1 of linear elasticity, use (3.9) and (3.81) to obtain

$$\begin{aligned}
&\|\mathcal{R}'_{NH}(\mathbf{Q},\mathbf{v})[\hat{\mathbf{Q}},\hat{\mathbf{v}}] - \mathcal{R}'_{NH}(\mathbf{0},\mathbf{0})[\hat{\mathbf{Q}},\hat{\mathbf{v}}]\|^2_{L^2(\Omega)} \\
&\qquad \lesssim \left(\|\mathbf{Q}\|_{L^\infty(\Omega)} + \|\nabla \mathbf{v}\|_{L^\infty(\Omega)}\right)^2 \left(\|\operatorname{div} \hat{\mathbf{Q}}\|^2_{L^2(\Omega)} + \|\hat{\mathbf{Q}}\|^2_{L^2(\Omega)} + \|\boldsymbol{\varepsilon}(\hat{\mathbf{v}})\|^2_{L^2(\Omega)}\right) \\
&\qquad \lesssim \left(\|\mathbf{Q}\|_{L^\infty(\Omega)} + \|\nabla \mathbf{v}\|_{L^\infty(\Omega)}\right)^2 \|\mathcal{L}(\hat{\mathbf{Q}},\hat{\mathbf{v}})\|^2_{L^2(\Omega)} \\
&\qquad \lesssim \left(\|\mathbf{Q}\|_{L^\infty(\Omega)} + \|\nabla \mathbf{v}\|_{L^\infty(\Omega)}\right)^2 \|\mathcal{R}'_{NH}(\mathbf{0},\mathbf{0})[\hat{\mathbf{Q}},\hat{\mathbf{v}}]\|^2_{L^2(\Omega)}
\end{aligned} \tag{3.82}$$

from inequality (3.80). We obtain the statement after extracting the square root and choosing θ in (3.55) sufficiently small such that the constant on the right - hand side in (3.82) becomes less than 1 and the assumptions of Lemma 3.20 for $\boldsymbol{\Xi} = \mathbf{Q}\mathbf{F}(\mathbf{v})^T$, $(\mathbf{Q},\mathbf{v}) \in \boldsymbol{\Pi}^\infty \times \mathbf{U}^\infty$, are additionally satisfied.

□

Remark 3.24:

The constant that appears in Lemma 3.23 depends on the constant of Lemma 3.20, the constant of Korn's inequality, the constant of Theorem 3.1 and θ and μ. Therefore it is cumbersome to specify. However, it is guaranteed that Lemma 3.23 holds for $(\mathbf{Q},\mathbf{v}) \in \boldsymbol{\Pi}^\infty \times \mathbf{U}^\infty$ with sufficiently small $\theta > 0$, i.e. for $(\mathbf{Q},\mathbf{v})$ sufficiently close to the origin. The constant does not depend on λ for $\lambda \geq 1$ (cf. Remark 3.21). This means in particular that the statement holds uniformly in the incompressible limit $\lambda \to \infty$.

Lemma 3.25:

Let V be a normed space with norm $\|\cdot\|_V$ and assume that

$$\|u - v\|_V \leq \rho \|v\|_V$$

for all $u, v \in V$ and $\rho \in [0,1)$. Then it holds

$$\|u\|_V \leq (1+\rho)\|v\|_V \text{ and } \|u\|_V \geq (1-\rho)\|v\|_V$$

for all $u, v \in V$.

Proof:

Let $u, v \in V$ be arbitrary. By assumption it holds $\|u - v\|_V \leq \rho\|v\|_V < \|v\|_V$, since $\rho < 1$. Therefore it holds on the one hand with the triangle inequality

$$\|u\|_V = \|u - v + v\|_V \leq \|u - v\|_V + \|v\|_V \leq \rho\|v\|_V + \|v\|_V = (1+\rho)\|v\|_V.$$

With the help of the reverse triangle inequality it holds on the other hand

$$\begin{aligned}
\|u\|_V = \|v - (v-u)\|_V &\geq \Big| \underbrace{\|v\|_V - \|v-u\|_V}_{>0} \Big| = \|v\|_V - \|u - v\|_V \\
&\geq \|v\|_V - \rho\|v\|_V = (1-\rho)\|v\|_V.
\end{aligned}$$

□

Lemma 3.26:

Let V be a normed function space, $R : V \to L^2(\Omega)$ continuously differentiable, $g : [0,1] \to V$, $s \mapsto v + s(u - v)$ for fixed but arbitrary $u, v \in V$ and $f : [0,1] \to L^2(\Omega)$, defined by $f(s) := R(g(s))$ for $s \in [0,1]$. Then it holds

$$\|R(u) - R(v)\|_{L^2(\Omega)} \leq \max_{s \in [0,1]} \|f'(s)\|_{L^2(\Omega)}.$$

Proof:

By assumption it holds $R'(v) \in \mathcal{L}(V, L^2(\Omega))$ for arbitrary $v \in V$, $g'(s) = u - v \in V$ for $s \in [0,1]$, $u, v \in V$, and by the chain rule the derivative

$$f'(s) = R'(g(s))[g'(s)] = R'(v + s(u - v))[u - v] \in C([0,1], L^2(\Omega)).$$

This implies

$$R(u) - R(v) = f(1) - f(0) = \int_0^1 f'(s)\, ds \in L^2(\Omega), \quad u, v \in V,$$

and results in

$$\|R(u) - R(v)\|_{L^2(\Omega)} = \left\| \int_0^1 f'(s)\, ds \right\|_{L^2(\Omega)} \leq \int_0^1 \|f'(s)\|_{L^2(\Omega)}\, ds \leq \max_{s \in [0,1]} \|f'(s)\|_{L^2(\Omega)}.$$

Here we have used well-known estimates for integrals over continuous functions mapping from compact intervals to Banach spaces (cf. Section VI.4 in [AE06b]).

□

Corollary 3.27:

Let V be a normed function space, $R : V \to L^2(\Omega)$ continuously differentiable. Then it holds

$$\|R(u) - R(v) - R'(0)[u - v]\|_{L^2(\Omega)} \leq \max_{s \in [0,1]} \|R'(v + s(u - v))[u - v] - R'(0)[u - v]\|_{L^2(\Omega)}.$$

Proof:

We set $\tilde{R}(u) := R(u) - R'(0)[u]$ for $u \in V$. Then $\tilde{R}$ is also a mapping from V into $L^2(\Omega)$. Since R is assumed to be continuously differentiable, $\tilde{R}$ is also continuously differentiable with respect to u. We obtain its derivative $\tilde{R}'(u)[v] = R'(u)[v] - R'(0)[v]$, $u, v \in V$, and use the mapping $\tilde{R}$ instead of R in Lemma 3.26. Using $\tilde{f}(s) := \tilde{R}(v + s(u - v))$ for $s \in [0,1]$, $u, v \in V$, we get

$$\tilde{f}'(s) = \tilde{R}'(v + s(u - v))[u - v] = R'(v + s(u - v))[u - v] - R'(0)[u - v].$$

Inserting this into the statement of Lemma 3.26 leads to

$$\begin{aligned} \|R(u) - R(v) - R'(0)[u - v]\|_{L^2(\Omega)} &= \|\tilde{R}(u) - \tilde{R}(v)\|_{L^2(\Omega)} \leq \max_{s \in [0,1]} \|\tilde{f}'(s)\|_{L^2(\Omega)} \\ &= \max_{s \in [0,1]} \|R'(v + s(u - v))[u - v] - R'(0)[u - v]\|_{L^2(\Omega)}, \end{aligned}$$

i.e. the statement.

□

Corollary 3.28:
Let N be a positive integer. For functions $R_i : V \to L^2(\Omega)$, $i = 1, \dots, N$, in a normed function space V we define

$$\mathbf{R}(v) := \begin{pmatrix} R_1(v) \\ \vdots \\ R_N(v) \end{pmatrix}, \quad \mathbf{f}(s) := \mathbf{R}(g(s)) = \begin{pmatrix} R_1(g(s)) \\ \vdots \\ R_N(g(s)) \end{pmatrix} =: \begin{pmatrix} f_1(s) \\ \vdots \\ f_N(s) \end{pmatrix}$$

with $g : [0,1] \to V, s \mapsto v + s(u - v)$ for $u, v \in V$. Consequently we obtain

$$\begin{aligned} \|\mathbf{R}(u) - \mathbf{R}(v)\|_{L^2(\Omega)} &= \left(\sum_{i=1}^{N} \|R_i(u) - R_i(v)\|^2_{L^2(\Omega)} \right)^{\frac{1}{2}} \leq \left(\sum_{i=1}^{N} \max_{s \in [0,1]} \|f_i'(s)\|^2_{L^2(\Omega)} \right)^{\frac{1}{2}} \\ &= \left(\max_{s \in [0,1]} \|\mathbf{f}'(s)\|^2_{L^2(\Omega)} \right)^{\frac{1}{2}} = \max_{s \in [0,1]} \|\mathbf{f}'(s)\|_{L^2(\Omega)}, \end{aligned}$$

where we have used Lemma 3.26 for each $i \in \{1, \dots, N\}$. Hence, Lemma 3.26 can be extended to a continuously differentiable function $\mathbf{R} : V \to L^2(\Omega)^N$. Analogously, following the steps in Corollary 3.27, we obtain moreover

$$\|\mathbf{R}(u) - \mathbf{R}(v) - \mathbf{R}'(0)[u - v]\|_{L^2(\Omega)} \leq \max_{s \in [0,1]} \|\mathbf{R}'(v + s(u - v))[u - v] - \mathbf{R}'(0)[u - v]\|_{L^2(\Omega)}. \tag{3.83}$$

Theorem 3.29: (Efficiency and reliability of the nonlinear least squares functional)
For the first-order system (3.18) to the inverse $\mathbf{B}$-formulation in the considered Neo-Hooke material, if $\theta > 0$ sufficiently small in $\mathbf{\Pi}^\infty, \mathbf{U}^\infty$ (cf. (3.55)), then

$$\begin{aligned} \|\mathcal{R}_{NH}(\hat{\mathbf{Q}}, \hat{\mathbf{v}}) - \mathcal{R}_{NH}(\mathbf{Q}, \mathbf{v})\|^2_{L^2(\Omega)} &\lesssim \|\hat{\mathbf{Q}} - \mathbf{Q}\|^2_{H(\mathrm{div};\Omega)} + \|\hat{\mathbf{v}} - \mathbf{v}\|^2_{H^1(\Omega)} \\ \|\mathcal{R}_{NH}(\hat{\mathbf{Q}}, \hat{\mathbf{v}}) - \mathcal{R}_{NH}(\mathbf{Q}, \mathbf{v})\|^2_{L^2(\Omega)} &\gtrsim \|\hat{\mathbf{Q}} - \mathbf{Q}\|^2_{H(\mathrm{div};\Omega)} + \|\hat{\mathbf{v}} - \mathbf{v}\|^2_{H^1(\Omega)} \end{aligned} \tag{3.84}$$

holds for all $(\hat{\mathbf{Q}}, \hat{\mathbf{v}}), (\mathbf{Q}, \mathbf{v}) \in \mathbf{\Pi}^\infty \times \mathbf{U}^\infty$.

Proof:

We recall that $\mathcal{R}_{NH} : \mathbf{\Pi}^\infty \times \mathbf{U}^\infty \subset H(\mathrm{div};\Omega)^3 \times H^1(\Omega)^3 \to L^2(\Omega)^3 \times L^2(\Omega)^{3\times 3}$. Since $\mathcal{R}_{NH}$ is continuously differentiable with derivative (3.53), we can use (3.83) in Corollary 3.28 for $\mathcal{R}_{NH}$ (instead of $\mathbf{R}$) and $V := H(\mathrm{div};\Omega)^3 \times H^1(\Omega)^3$. Using moreover Lemma 3.23 this leads immediately to

$$\begin{aligned} &\|\mathcal{R}_{NH}(\hat{\mathbf{Q}}, \hat{\mathbf{v}}) - \mathcal{R}_{NH}(\mathbf{Q}, \mathbf{v}) - \mathcal{R}'_{NH}(\mathbf{0}, \mathbf{0})[\hat{\mathbf{Q}} - \mathbf{Q}, \hat{\mathbf{v}} - \mathbf{v}]\|_{L^2(\Omega)} \\ &\quad \leq \max_{s \in [0,1]} \left\| \left(\mathcal{R}'_{NH}(\mathbf{Q} + s(\hat{\mathbf{Q}} - \mathbf{Q}), \mathbf{v} + s(\hat{\mathbf{v}} - \mathbf{v})) - \mathcal{R}'_{NH}(\mathbf{0}, \mathbf{0}) \right) [\hat{\mathbf{Q}} - \mathbf{Q}, \hat{\mathbf{v}} - \mathbf{v}] \right\|_{L^2(\Omega)} \\ &\quad \leq \max_{s \in [0,1]} \rho \|\mathcal{R}'_{NH}(\mathbf{0}, \mathbf{0})[\hat{\mathbf{Q}} - \mathbf{Q}, \hat{\mathbf{v}} - \mathbf{v}]\|_{L^2(\Omega)} = \rho \|\mathcal{R}'_{NH}(\mathbf{0}, \mathbf{0})[\hat{\mathbf{Q}} - \mathbf{Q}, \hat{\mathbf{v}} - \mathbf{v}]\|_{L^2(\Omega)}. \end{aligned}$$

We use Lemma 3.25 to obtain

$$\begin{aligned}\|\mathcal{R}_{NH}(\hat{\mathbf{Q}},\hat{\mathbf{v}})-\mathcal{R}_{NH}(\mathbf{Q},\mathbf{v})\|_{L^2(\Omega)} &\leq (1+\rho)\|\mathcal{R}'_{NH}(\mathbf{0},\mathbf{0})[\hat{\mathbf{Q}}-\mathbf{Q},\hat{\mathbf{v}}-\mathbf{v}]\|_{L^2(\Omega)},\\ \|\mathcal{R}_{NH}(\hat{\mathbf{Q}},\hat{\mathbf{v}})-\mathcal{R}_{NH}(\mathbf{Q},\mathbf{v})\|_{L^2(\Omega)} &\geq (1-\rho)\|\mathcal{R}'_{NH}(\mathbf{0},\mathbf{0})[\hat{\mathbf{Q}}-\mathbf{Q},\hat{\mathbf{v}}-\mathbf{v}]\|_{L^2(\Omega)}.\end{aligned} \tag{3.85}$$

The statement (3.84) follows immediately by combining (3.85) with (3.81) and (3.7).

□

An immediate consequence of Theorem 3.29 for the exact solution $(\mathbf{Q},\mathbf{v}) := (\mathbf{P},\mathbf{u}) \in \mathbf{\Pi}^\infty \times \mathbf{U}^\infty$ and an approximation $(\hat{\mathbf{Q}},\hat{\mathbf{v}}) := (\mathbf{P}_h,\mathbf{u}_h) \in \mathbf{\Pi}^\infty \times \mathbf{U}^\infty$ is

$$||(\mathbf{P}-\mathbf{P}_h,\mathbf{u}-\mathbf{u}_h)||^2_{\mathcal{V}} \lesssim \mathcal{F}_{NH}(\mathbf{P}_h,\mathbf{u}_h) \lesssim ||(\mathbf{P}-\mathbf{P}_h,\mathbf{u}-\mathbf{u}_h)||^2_{\mathcal{V}}$$

with $\mathcal{V} = H_{\Gamma_N}(\mathrm{div};\Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$. This is exactly the property (3.56) and is valid, since for the exact solution it holds $\mathcal{R}_{NH}(\mathbf{P},\mathbf{u}) = \mathbf{0}$ and by definition of the least squares functional in (3.19) for the Neo - Hooke case it holds $\|\mathcal{R}_{NH}(\mathbf{P}_h,\mathbf{u}_h)\|^2_{L^2(\Omega)} = \mathcal{F}_{NH}(\mathbf{P}_h,\mathbf{u}_h)$. Thus we have proven under quite strong regularity assumptions that we can estimate the error $\mathbf{e} := (\mathbf{P}-\mathbf{P}_h,\mathbf{u}-\mathbf{u}_h)$ from below and above by the nonlinear least squares functional, evaluated in the approximation. This holds at least for $(\mathbf{P},\mathbf{u})$ and $(\mathbf{P}_h,\mathbf{u}_h)$ sufficiently close to the origin. In this case it is proven that the nonlinear functional is a reasonable a - posteriori error estimator.

For instance if we combine Raviart - Thomas elements $\mathbf{\Pi}^l_h := (\mathcal{RT}_{l-1}(\mathcal{T}_h))^3 \subset \mathbf{\Pi}^\infty \subset H(\mathrm{div};\,\Omega)^3$ for the approximation of $\mathbf{P}_h$ with continuous elements $\mathbf{U}^l_h := (\mathcal{P}_l(\mathcal{T}_h))^3 \subset \mathbf{U}^\infty \subset H^1(\Omega)^3$ for the approximation of $\mathbf{u}_h$ with an arbitrary integer $l \geq 1$ and $(\mathbf{P}_h,\mathbf{u}_h)$ minimize $\mathcal{F}_{NH}(\mathbf{Q}_h,\mathbf{v}_h)$ about all $(\mathbf{Q}_h,\mathbf{v}_h) \in \mathbf{\Pi}^l_h \times \mathbf{U}^l_h \subset H(\mathrm{div};\,\Omega)^3 \times H^1(\Omega)^3$ we get the a - priori estimate

$$\begin{aligned}\|\mathbf{e}\|_{\mathcal{V}} = \|(\mathbf{P}-\mathbf{P}_h,\mathbf{u}-\mathbf{u}_h)\|_{\mathcal{V}} &\lesssim (\mathcal{F}_{NH}(\mathbf{P}_h,\mathbf{u}_h))^{\frac{1}{2}} = \inf_{(\mathbf{Q}_h,\mathbf{v}_h)\,\in\,\mathbf{\Pi}^l_h \times \mathbf{U}^l_h} (\mathcal{F}_{NH}(\mathbf{Q}_h,\mathbf{v}_h))^{\frac{1}{2}}\\ &\lesssim ||(\mathbf{P}-\Pi_h\mathbf{P},\mathbf{u}-I_h\mathbf{u})||_{\mathcal{V}} \lesssim h^l\left(\|\mathbf{P}\|^2_{H^l(\Omega)} + \|\mathrm{div}\,\mathbf{P}\|^2_{H^l(\Omega)} + \|\mathbf{u}\|^2_{H^{l+1}(\Omega)}\right)^{\frac{1}{2}}\end{aligned} \tag{3.86}$$

for the error respectively the nonlinear least squares functional (cf. (3.13) with the interpolation operators Π_h, I_h defined in Section 2.5). In particular we expect at most a behavior proportional to h^l of the square root of the nonlinear least squares functional, provided that the solution $(\mathbf{P},\mathbf{u})$ is sufficiently regular. Hence a convergence rate of order l is optimal for this choice of finite element spaces.

For the sake of completeness we want to prove at the end of this subsection that the approximation $\boldsymbol{\tau}_h = \mathbf{P}_h\mathbf{F}(\mathbf{u}_h)^T$ of the symmetric Kirchhoff stress tensor $\boldsymbol{\tau} = \mathbf{P}\mathbf{F}(\mathbf{u})^T$ is also symmetric in convergence, i.e. one obtains $\boldsymbol{\tau}_h = \boldsymbol{\tau}_h^T$ if $\mathcal{F}_{NH}(\mathbf{P}_h,\mathbf{u}_h) \to 0$ for $h \to 0$. For this purpose we start with the following lemma:

Lemma 3.30:

Let $(\mathbf{Q},\mathbf{v}),(\hat{\mathbf{Q}},\hat{\mathbf{v}})$ be in $\mathbf{\Pi}^\infty \times \mathbf{U}^\infty$, again provided that $\theta > 0$ is sufficiently small. Then

it holds

$$\|\mathbf{Q}\mathbf{F}(\mathbf{v})^T - \hat{\mathbf{Q}}\mathbf{F}(\hat{\mathbf{v}})^T\|_{L^2(\Omega)} \lesssim \|\mathcal{R}_{NH}(\mathbf{Q},\mathbf{v}) - \mathcal{R}_{NH}(\hat{\mathbf{Q}},\hat{\mathbf{v}})\|_{L^2(\Omega)}.$$

Proof:

By assumption we have $(\mathbf{Q},\mathbf{v}),(\hat{\mathbf{Q}},\hat{\mathbf{v}}) \in \mathbf{\Pi}^\infty \times \mathbf{U}^\infty$. With this choice we get on the one hand the estimate

$$\begin{aligned}
\|\mathbf{Q}\mathbf{F}(\mathbf{v})^T - \hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T\|_{L^2(\Omega)} &= \|(\mathbf{Q}-\hat{\mathbf{Q}})\mathbf{F}(\mathbf{v})^T\|_{L^2(\Omega)} \\
&= \|(\mathbf{Q}-\hat{\mathbf{Q}})(\mathbf{I}+(\nabla\mathbf{v})^T)\|_{L^2(\Omega)} \\
&\leq \|\mathbf{Q}-\hat{\mathbf{Q}}\|_{L^2(\Omega)} + \|(\mathbf{Q}-\hat{\mathbf{Q}})(\nabla\mathbf{v})^T\|_{L^2(\Omega)} \\
&\leq \|\mathbf{Q}-\hat{\mathbf{Q}}\|_{L^2(\Omega)} + \|\mathbf{Q}-\hat{\mathbf{Q}}\|_{L^2(\Omega)}\|\nabla\mathbf{v}\|_{L^\infty(\Omega)} \\
&\leq (1+\theta)\|\mathbf{Q}-\hat{\mathbf{Q}}\|_{L^2(\Omega)}.
\end{aligned} \tag{3.87}$$

On the other hand we obtain

$$\begin{aligned}
\|\hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T - \hat{\mathbf{Q}}\mathbf{F}(\hat{\mathbf{v}})^T\|_{L^2(\Omega)} &= \|\hat{\mathbf{Q}}(\mathbf{F}(\mathbf{v})-\mathbf{F}(\hat{\mathbf{v}}))^T\|_{L^2(\Omega)} = \|\hat{\mathbf{Q}}(\nabla(\mathbf{v}-\hat{\mathbf{v}}))^T\|_{L^2(\Omega)} \\
&\leq \|\hat{\mathbf{Q}}\|_{L^\infty(\Omega)}\|\nabla(\mathbf{v}-\hat{\mathbf{v}})\|_{L^2(\Omega)} \leq \theta\|\nabla(\mathbf{v}-\hat{\mathbf{v}})\|_{L^2(\Omega)}.
\end{aligned} \tag{3.88}$$

Combining (3.87) and (3.88) leads to

$$\begin{aligned}
\|\mathbf{Q}\mathbf{F}(\mathbf{v})^T - \hat{\mathbf{Q}}\mathbf{F}(\hat{\mathbf{v}})^T\|_{L^2(\Omega)} &= \|\mathbf{Q}\mathbf{F}(\mathbf{v})^T - \hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T + \hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T - \hat{\mathbf{Q}}\mathbf{F}(\hat{\mathbf{v}})^T\|_{L^2(\Omega)} \\
&\leq \|\mathbf{Q}\mathbf{F}(\mathbf{v})^T - \hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T\|_{L^2(\Omega)} + \|\hat{\mathbf{Q}}\mathbf{F}(\mathbf{v})^T - \hat{\mathbf{Q}}\mathbf{F}(\hat{\mathbf{v}})^T\|_{L^2(\Omega)} \\
&\leq (1+\theta)\|\mathbf{Q}-\hat{\mathbf{Q}}\|_{L^2(\Omega)} + \theta\|\nabla(\mathbf{v}-\hat{\mathbf{v}})\|_{L^2(\Omega)} \\
&\leq (1+\theta)\left(\|\mathbf{Q}-\hat{\mathbf{Q}}\|_{L^2(\Omega)} + \|\nabla(\mathbf{v}-\hat{\mathbf{v}})\|_{L^2(\Omega)}\right) \\
&\leq (1+\theta)\left(\|\mathbf{Q}-\hat{\mathbf{Q}}\|_{H(\mathrm{div};\,\Omega)} + \|\mathbf{v}-\hat{\mathbf{v}}\|_{H^1(\Omega)}\right).
\end{aligned}$$

Using Theorem 3.29 ends up in

$$\begin{aligned}
\|\mathbf{Q}\mathbf{F}(\mathbf{v})^T - \hat{\mathbf{Q}}\mathbf{F}(\hat{\mathbf{v}})^T\|^2_{L^2(\Omega)} &= (1+\theta)^2\left(\|\mathbf{Q}-\hat{\mathbf{Q}}\|_{H(\mathrm{div};\,\Omega)} + \|\mathbf{v}-\hat{\mathbf{v}}\|_{H^1(\Omega)}\right)^2 \\
&\leq 2(1+\theta)^2\left(\|\mathbf{Q}-\hat{\mathbf{Q}}\|^2_{H(\mathrm{div};\,\Omega)} + \|\mathbf{v}-\hat{\mathbf{v}}\|^2_{H^1(\Omega)}\right) \\
&\lesssim \|\mathcal{R}_{NH}(\mathbf{Q},\mathbf{v}) - \mathcal{R}_{NH}(\hat{\mathbf{Q}},\hat{\mathbf{v}})\|^2_{L^2(\Omega)},
\end{aligned}$$

i.e. the statement after extracting the square root.

□

Corollary 3.31: (Symmetry of $\boldsymbol{\tau}_h$)
Let $(\mathbf{P},\mathbf{u}) \in \mathbf{\Pi}^\infty \times \mathbf{U}^\infty$ be the exact solution, i.e. $\mathcal{R}_{NH}(\mathbf{P},\mathbf{u}) = \mathbf{0}$, and $(\mathbf{P}_h,\mathbf{u}_h) \in \mathbf{\Pi}^\infty \times \mathbf{U}^\infty$ be a conforming finite element approximation. Then we get for the approximation $\boldsymbol{\tau}_h := \mathbf{P}_h\mathbf{F}(\mathbf{u}_h)^T$ of the Kirchhoff stress tensor $\boldsymbol{\tau} = \mathbf{P}\mathbf{F}(\mathbf{u})^T$ the estimates

$$\begin{aligned}
\|\boldsymbol{\tau} - \boldsymbol{\tau}_h\|_{L^2(\Omega)} &\lesssim (\mathcal{F}_{NH}(\mathbf{P}_h,\mathbf{u}_h))^{\frac{1}{2}}, \\
\|\boldsymbol{\tau}_h - \boldsymbol{\tau}_h^T\|_{L^2(\Omega)} &\lesssim (\mathcal{F}_{NH}(\mathbf{P}_h,\mathbf{u}_h))^{\frac{1}{2}}.
\end{aligned}$$

Proof:

We use Lemma 3.30 with $(\mathbf{Q}, \mathbf{v}) = (\mathbf{P}, \mathbf{u})$, $(\hat{\mathbf{Q}}, \hat{\mathbf{v}}) = (\mathbf{P}_h, \mathbf{u}_h) \in \mathbf{\Pi}^\infty \times \mathbf{U}^\infty$ and obtain with the help of (3.19)

$$\begin{aligned}\|\boldsymbol{\tau} - \boldsymbol{\tau}_h\|_{L^2(\Omega)} &= \|\mathbf{P}\mathbf{F}(\mathbf{u})^T - \mathbf{P}_h\mathbf{F}(\mathbf{u}_h)^T\|_{L^2(\Omega)} \lesssim \|\mathcal{R}_{NH}(\mathbf{P}, \mathbf{u}) - \mathcal{R}_{NH}(\mathbf{P}_h, \mathbf{u}_h)\|_{L^2(\Omega)} \\ &= \|\mathcal{R}_{NH}(\mathbf{P}_h, \mathbf{u}_h)\|_{L^2(\Omega)} = (\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h))^{\frac{1}{2}},\end{aligned}$$

i.e. the first statement.

In Section 3.1 we have remarked that the conservation of angular momentum leads to a symmetric $\boldsymbol{\tau}$, i.e. it holds $\boldsymbol{\tau} = \boldsymbol{\tau}^T$ for the exact Kirchhoff stress tensor $\boldsymbol{\tau}$. With this property and the first statement we obtain

$$\begin{aligned}\|\boldsymbol{\tau}_h - \boldsymbol{\tau}_h^T\|_{L^2(\Omega)} &= \|\boldsymbol{\tau} - \boldsymbol{\tau}_h^T - (\boldsymbol{\tau} - \boldsymbol{\tau}_h)\|_{L^2(\Omega)} = \|(\boldsymbol{\tau} - \boldsymbol{\tau}_h)^T - (\boldsymbol{\tau} - \boldsymbol{\tau}_h)\|_{L^2(\Omega)} \\ &\leq 2\|\boldsymbol{\tau} - \boldsymbol{\tau}_h\|_{L^2(\Omega)} \lesssim (\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h))^{\frac{1}{2}},\end{aligned}$$

i.e. the second statement.

□

Corollary 3.31 tells us that as long as the value $\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$ converges to zero it is also ensured that the Kirchhoff stress approximation converges to the exact one and these approximations become symmetric. Note that the estimates in Corollary 3.31 can be also combined with (3.86) to obtain a - priori estimates for the Kirchhoff stress tensor.

3.5.2 The linearized problem

In the Neo - Hooke case we are also able to prove the property (3.23) in the space $\mathcal{V} := H_{\Gamma_N}(\mathrm{div}; \Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$ for the linearized problem. For this purpose we need the following lemma.

Lemma 3.32:

Let $(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) \in \mathbf{\Pi}^\infty \times \mathbf{U}^\infty$ and $(\hat{\mathbf{Q}}, \hat{\mathbf{v}}) \in \mathcal{V}$. Then it holds

$$\begin{aligned}\|\mathcal{R}'_{NH}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\hat{\mathbf{Q}}, \hat{\mathbf{v}}]\|_{L^2(\Omega)} &\leq (1 + \rho)\|\mathcal{R}'_{NH}(\mathbf{0}, \mathbf{0})[\hat{\mathbf{Q}}, \hat{\mathbf{v}}]\|_{L^2(\Omega)}, \\ \|\mathcal{R}'_{NH}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\hat{\mathbf{Q}}, \hat{\mathbf{v}}]\|_{L^2(\Omega)} &\geq (1 - \rho)\|\mathcal{R}'_{NH}(\mathbf{0}, \mathbf{0})[\hat{\mathbf{Q}}, \hat{\mathbf{v}}]\|_{L^2(\Omega)}\end{aligned}$$

with $\rho \in [0, 1)$.

Proof:

For sufficiently small θ in (3.55) we know by Lemma 3.23 that there exists $\rho \in [0, 1)$ such that

$$\|\mathcal{R}'_{NH}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\hat{\mathbf{Q}}, \hat{\mathbf{v}}] - \mathcal{R}'_{NH}(\mathbf{0}, \mathbf{0})[\hat{\mathbf{Q}}, \hat{\mathbf{v}}]\|_{L^2(\Omega)} \leq \rho\|\mathcal{R}'_{NH}(\mathbf{0}, \mathbf{0})[\hat{\mathbf{Q}}, \hat{\mathbf{v}}]\|_{L^2(\Omega)}$$

holds for $(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) \in \mathbf{\Pi}^\infty \times \mathbf{U}^\infty$ and arbitrary $(\hat{\mathbf{Q}}, \hat{\mathbf{v}}) \in \mathcal{V}$. Using this observation and Lemma 3.25 directly leads to the statement.

□

Corollary 3.33: („Wanted property“ for the linearized problem)
Let $(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) \in \mathbf{\Pi}^{\infty} \times \mathbf{U}^{\infty}$ be given. Then it holds

$$\mathcal{F}_{NH}^{\text{lin}}(\hat{\mathbf{Q}}, \hat{\mathbf{v}}; \mathbf{0}) \eqsim \|(\hat{\mathbf{Q}}, \hat{\mathbf{v}})\|_{\mathcal{V}}^2 \quad \text{for all } (\hat{\mathbf{Q}}, \hat{\mathbf{v}}) \in \mathcal{V}.$$

Proof:

By definition of $\mathcal{F}_{NH}^{\text{lin}}(\hat{\mathbf{Q}}, \hat{\mathbf{v}}; \mathbf{0})$ in (3.21) for the Neo - Hooke case, Lemma 3.32, the relation (3.81) between the operators $\mathcal{R}'_{NH}(\mathbf{0}, \mathbf{0})$ and $\mathcal{L}$ and the property (3.7) of linear elasticity we obtain

$$\begin{aligned}\mathcal{F}_{NH}^{\text{lin}}(\hat{\mathbf{Q}}, \hat{\mathbf{v}}; \mathbf{0}) &= \|\mathcal{R}'_{NH}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\hat{\mathbf{Q}}, \hat{\mathbf{v}}]\|_{L^2(\Omega)}^2 \eqsim \|\mathcal{R}'_{NH}(\mathbf{0}, \mathbf{0})[\hat{\mathbf{Q}}, \hat{\mathbf{v}}]\|_{L^2(\Omega)}^2 \\ &\eqsim \|\mathcal{L}(\hat{\mathbf{Q}}, \hat{\mathbf{v}})\|_{L^2(\Omega)}^2 \eqsim \|(\hat{\mathbf{Q}}, \hat{\mathbf{v}})\|_{\mathcal{V}}^2\end{aligned}$$

for all $(\hat{\mathbf{Q}}, \hat{\mathbf{v}}) \in \mathcal{V}$.

□

Therefore by the general considerations in Section 3.3.2 we get for each linearized problem in the algorithm a unique correction term $\left(\mathbf{Q}^{(k)}, \mathbf{v}^{(k)}\right) \in \mathcal{V}$. One open problem still remains, namely $\mathbf{\Pi}^{\infty} \times \mathbf{U}^{\infty}$ is only a subset of $H(\text{div}; \Omega)^3 \times H^1(\Omega)^3$ and therefore it is not guaranteed that the new solution

$$\left(\mathbf{P}^{(k+1)}, \mathbf{u}^{(k+1)}\right) := \left(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}\right) + \alpha^{(k)} \left(\mathbf{Q}^{(k)}, \mathbf{v}^{(k)}\right), \quad \alpha^{(k)} \in (0, 1],$$

is in $\mathbf{\Pi}^{\infty} \times \mathbf{U}^{\infty}$. However, the problem is not existent in the discrete problem, described in Section 3.3.3.

3.6 Comparison to other discretization methods

The least squares finite element methods for hyperelasticity proposed in Section 3.3, based on the inversion of given stress - strain relations, must be compared with already existing discretization schemes to show their suitability. For this purpose we introduce the standard Galerkin method (often called **pure displacement approach**) for compressible materials and a displacement - pressure approach for incompressible hyperelasticity, proposed by Auricchio (cf. [ABadVLR05] and [ABadVLR10]). In both discretization schemes we assume for simplicity that the applied forces are **dead loads**, which means that the given densities $\mathbf{f}$ and $\mathbf{g}$ of the volume and surface forces are independent of the deformation $\boldsymbol{\varphi}$ (or equivalently independent of the displacement $\mathbf{u}$). However, bear in mind that both schemes also work for more general conservative forces (cf. Section 5 in [Cia88]). In particular both discretization schemes will be formulated for the Mooney - Rivlin material (2.30).

3.6.1 Pure displacement approach

The point of departure is to find a minimizer of the **total energy**

$$\tilde{I}(\boldsymbol{\chi}) := \int_\Omega \hat{\psi}(\mathbf{x}, \nabla\boldsymbol{\chi})\, dx - \underbrace{\left(\int_\Omega \mathbf{f}\cdot\boldsymbol{\chi}\, dx + \int_{\Gamma_N} \mathbf{g}\cdot\boldsymbol{\chi}\, ds\right)}_{=:\, L(\boldsymbol{\chi})}$$

in an admissible set

$$\tilde{\boldsymbol{\Phi}} := \left\{\boldsymbol{\varphi} \in W^{1,p}(\Omega)^3 : \det\nabla\boldsymbol{\varphi} > 0 \text{ in } \bar{\Omega}, \boldsymbol{\varphi} = \boldsymbol{\varphi}_D \text{ on } \Gamma_D\right\}$$

of deformations (cf. Section 5 in [Cia88]), i.e. we seek $\boldsymbol{\varphi} \in \tilde{\boldsymbol{\Phi}}$ such that

$$\tilde{I}(\boldsymbol{\varphi}) = \inf\left\{\tilde{I}(\boldsymbol{\chi}) : \boldsymbol{\chi} \in \tilde{\boldsymbol{\Phi}}\right\}. \tag{3.89}$$

$\hat{\psi} : \bar{\Omega}\times\mathbb{M} \to \mathbb{R}$ with $\mathbb{M} := \{\mathbf{F} \in \mathbb{R}^{3\times3} : \det\mathbf{F} > 0\}$ is assumed to be a Fréchet differentiable stored energy function to a given hyperelastic material (cf. Definition 2.20) and $\boldsymbol{\varphi}_D \in W^{1,p}(\Omega)^3$ are prescribed boundary conditions on Γ_D.
An existence theory for minimizer(s) of (3.89) can be found in [Bal77] (respectively in Section 7.7 in [Cia88]). This theory is based on the polyconvexity (cf. Section 2.2.6) of the underlying stored energy function $\hat{\psi}$ and a coerciveness inequality of the form

$$\hat{\psi}(\mathbf{x}, \mathbf{F}) \geq c_1\left(|\mathbf{F}|^p + |\mathbf{Cof}\,\mathbf{F}|^q + (\det\mathbf{F})^r\right) + c_2, \tag{3.90}$$

for all $\mathbf{F} \in \mathbb{M}$, $\mathbf{x} \in \bar{\Omega}$, with $p \geq 2$, $q \geq \frac{p}{p-1}$, $r > 1$, $c_1 > 0$, $c_2 \in \mathbb{R}$.
In particular, the stored energy function (2.30) of the considered Mooney - Rivlin material satisfies by Theorem 4.10 - 2 in [Cia88] the coerciveness inequality (3.90) with $p = q = r = 2$ and is polyconvex (cf. Section 2.4.4). Thus by Theorem 7.7 - 1 in [Cia88], provided that $L(\boldsymbol{\chi})$ is continuous and $\inf\{\tilde{I}(\boldsymbol{\chi}) : \boldsymbol{\chi} \in \tilde{\boldsymbol{\Phi}}\} < +\infty$, a minimizer $\boldsymbol{\varphi}$ of (3.89) is in $W^{1,2}(\Omega)^3 = H^1(\Omega)^3$ with $\mathbf{Cof}\,\nabla\boldsymbol{\varphi} \in L^2(\Omega)^{3\times3}$ and $\det\nabla\boldsymbol{\varphi} \in L^2(\Omega)$.
In the following we rewrite the minimization problem (3.89) in terms of displacements.
For each deformation $\boldsymbol{\chi} \in \tilde{\boldsymbol{\Phi}}$ we write $\boldsymbol{\chi} = \mathbf{id} + \mathbf{v}$ and set $\mathbf{u}_D := \boldsymbol{\varphi}_D - \mathbf{id} \in W^{1,p}(\Omega)^3$ with displacements

$$\mathbf{v} \in \boldsymbol{\Phi} := \left\{\mathbf{u} \in W^{1,p}(\Omega)^3 : \det(\mathbf{I} + \nabla\mathbf{u}) > 0 \text{ in } \bar{\Omega}, \mathbf{u} = \mathbf{u}_D \text{ on } \Gamma_D\right\}$$

and seek $\mathbf{u} \in \boldsymbol{\Phi}$ such that

$$I(\mathbf{u}) = \inf\left\{I(\mathbf{v}) : \mathbf{v} \in \boldsymbol{\Phi}\right\}, \quad I(\mathbf{v}) := \tilde{I}(\mathbf{id} + \mathbf{v}). \tag{3.91}$$

We further decompose each $\mathbf{v} \in \boldsymbol{\Phi}$ into $\mathbf{v} = \hat{\mathbf{v}} + \mathbf{u}_D$ with

$$\hat{\mathbf{v}} \in \boldsymbol{\Phi}_{\Gamma_D} := \left\{\hat{\mathbf{u}} \in W^{1,p}_{\Gamma_D}(\Omega)^3 : \det(\mathbf{I} + \nabla(\hat{\mathbf{u}} + \mathbf{u}_D)) > 0 \text{ in } \bar{\Omega}\right\}.$$

Since the boundary conditions $\mathbf{u}_D$ are prescribed, the minimization problem (3.91) is equivalent to find the minimizer $\hat{\mathbf{u}} \in \mathbf{\Phi}_{\Gamma_D}$ with

$$I_0(\hat{\mathbf{u}}) = \inf\{I_0(\hat{\mathbf{v}}) : \hat{\mathbf{v}} \in \mathbf{\Phi}_{\Gamma_D}\}, \quad I_0(\hat{\mathbf{v}}) := I(\hat{\mathbf{v}} + \mathbf{u}_D) = I(\mathbf{v}).$$

Since $\hat{\psi}$ is Fréchet differentiable by assumption and it holds $\mathbf{P}(\mathbf{u}) = \partial_{\mathbf{F}}\hat{\psi}(\mathbf{x}, \mathbf{F})|_{\mathbf{F}=\mathbf{F}(\mathbf{u})} = \frac{\partial\hat{\psi}}{\partial\mathbf{F}}(\mathbf{x}, \mathbf{F}(\mathbf{u}))$ for $\mathbf{u} = \hat{\mathbf{u}} + \mathbf{u}_D \in \mathbf{\Phi}$, we get by the chain rule and (2.4)

$$\begin{aligned}\Big(\hat{\psi}(\mathbf{x}, \mathbf{F}(\mathbf{u}))\Big)'[\hat{\mathbf{v}}] &= \Big(\hat{\psi}(\mathbf{x}, \mathbf{I} + \nabla(\hat{\mathbf{u}} + \mathbf{u}_D))\Big)'[\hat{\mathbf{v}}] = \hat{\psi}'(\mathbf{x}, \mathbf{I} + \nabla(\hat{\mathbf{u}} + \mathbf{u}_D))[\nabla\hat{\mathbf{v}}] \\ &= \partial_{\mathbf{F}}\hat{\psi}(\mathbf{x}, \mathbf{F})|_{\mathbf{F}=\mathbf{F}(\mathbf{u})} : \nabla\hat{\mathbf{v}} = \mathbf{P}(\mathbf{u}) : \nabla\hat{\mathbf{v}}, \quad \hat{\mathbf{v}} \in \mathbf{\Phi}_{\Gamma_D}.\end{aligned}$$

The necessary condition of finding a minimizer is $0 \overset{!}{=} I_0'(\hat{\mathbf{u}})[\hat{\mathbf{v}}] = (I(\hat{\mathbf{u}} + \mathbf{u}_D))'[\hat{\mathbf{v}}] = I'(\hat{\mathbf{u}} + \mathbf{u}_D)[\hat{\mathbf{v}}] = I'(\mathbf{u})[\hat{\mathbf{v}}]$ for all $\hat{\mathbf{v}} \in \mathbf{\Phi}_{\Gamma_D}$ and equivalently

$$(\mathbf{P}(\mathbf{u}), \nabla\hat{\mathbf{v}})_{L^2(\Omega)} - (\mathbf{f}, \hat{\mathbf{v}})_{L^2(\Omega)} - \underbrace{\int_{\Gamma_N} \mathbf{g} \cdot \hat{\mathbf{v}}\, ds}_{=: \langle \mathbf{g}, \hat{\mathbf{v}} \rangle_{\Gamma_N}} \overset{!}{=} 0 \quad \forall\, \hat{\mathbf{v}} \in \mathbf{\Phi}_{\Gamma_D}, \tag{3.92}$$

i.e. this is the corresponding (nonlinear) variational problem of finding the minimizer $\mathbf{u} = \hat{\mathbf{u}} + \mathbf{u}_D \in \mathbf{\Phi}$ of (3.91). Note that (3.92) is the variational problem according to the strong formulation (3.1). Indeed both problems are equivalent if the solution $\mathbf{u}$ is regular enough.

We solve the nonlinear variational formulation (3.92) with the help of a Newton iteration. In the k-th step of the Newton iteration we set the new approximation as $\mathbf{u}^{(k+1)} := \mathbf{u}^{(k)} + \boldsymbol{\delta}\mathbf{u}$, use the Taylor approximation $\mathbf{P}(\mathbf{u}^{(k+1)}) \approx \mathbf{P}(\mathbf{u}^{(k)}) + \mathbf{P}'(\mathbf{u}^{(k)})[\boldsymbol{\delta}\mathbf{u}]$, provided that $\mathbf{P}(\mathbf{u})$ is Fréchet differentiable with respect to $\mathbf{u}$, to get the linearized variational formulation

$$\underbrace{\int_\Omega \mathbf{P}'(\mathbf{u}^{(k)})[\boldsymbol{\delta}\mathbf{u}] : \nabla\hat{\mathbf{v}}\, dx}_{=:\, a(\boldsymbol{\delta}\mathbf{u}, \hat{\mathbf{v}})} = \underbrace{-\int_\Omega \mathbf{P}(\mathbf{u}^{(k)}) : \nabla\hat{\mathbf{v}}\, dx + \int_\Omega \mathbf{f} \cdot \hat{\mathbf{v}}\, dx + \int_{\Gamma_N} \mathbf{g} \cdot \hat{\mathbf{v}}\, ds}_{=:\, F(\hat{\mathbf{v}})} \quad \forall\, \hat{\mathbf{v}} \in \mathbf{\Phi}_{\Gamma_D}, \tag{3.93}$$

depending on the old approximation $\mathbf{u}^{(k)} \in \mathbf{\Phi}$.

In short notation we have

$$\begin{aligned}a(\boldsymbol{\delta}\mathbf{u}, \hat{\mathbf{v}}) &= (\mathbf{P}'(\mathbf{u}^{(k)})[\boldsymbol{\delta}\mathbf{u}], \nabla\hat{\mathbf{v}})_{L^2(\Omega)}, \\ F(\hat{\mathbf{v}}) &= -(\mathbf{P}(\mathbf{u}^{(k)}), \nabla\hat{\mathbf{v}})_{L^2(\Omega)} + (\mathbf{f}, \hat{\mathbf{v}})_{L^2(\Omega)} + \langle \mathbf{g}, \hat{\mathbf{v}} \rangle_{\Gamma_N}\end{aligned} \tag{3.94}$$

and seek a correction $\boldsymbol{\delta}\mathbf{u} \in \mathbf{\Phi}_{\Gamma_D}$ such that

$$a(\boldsymbol{\delta}\mathbf{u}, \hat{\mathbf{v}}) = F(\hat{\mathbf{v}}) \quad \forall\, \hat{\mathbf{v}} \in \mathbf{\Phi}_{\Gamma_D}. \tag{3.95}$$

Of course one can additionally use a damping strategy, as usual, in this method.

Remark 3.34: (Pure displacement approach for $\mathbf{u}^{(k)} = \mathbf{0}$ and zero boundary conditions)
If we set $\mathbf{u}_D = \mathbf{0}$ and $\mathbf{u}^{(k)} = \mathbf{0}$ in (3.94) and assume consistency with linear elasticity (cf. Section 2.4.5), i.e. it holds $\mathbf{P}(\mathbf{u}^{(k)}) = \mathbf{P}(\mathbf{0}) = \mathbf{0}$ and $\mathbf{P}'(\mathbf{u}^{(k)})[\boldsymbol{\delta}\mathbf{u}] = \mathbf{P}'(\mathbf{0})[\boldsymbol{\delta}\mathbf{u}] = 2\mu\,\boldsymbol{\varepsilon}(\boldsymbol{\delta}\mathbf{u}) + \lambda\,\mathrm{tr}(\boldsymbol{\varepsilon}(\boldsymbol{\delta}\mathbf{u}))\mathbf{I}$, we obtain

$$\begin{aligned} a(\boldsymbol{\delta}\mathbf{u}, \hat{\mathbf{v}}) &= (2\mu\,\boldsymbol{\varepsilon}(\boldsymbol{\delta}\mathbf{u}) + \lambda\,\mathrm{tr}(\boldsymbol{\varepsilon}(\boldsymbol{\delta}\mathbf{u}))\mathbf{I}, \nabla\hat{\mathbf{v}})_{L^2(\Omega)} = (\mathcal{C}\boldsymbol{\varepsilon}(\boldsymbol{\delta}\mathbf{u}), \boldsymbol{\varepsilon}(\hat{\mathbf{v}}))_{L^2(\Omega)}, \\ F(\hat{\mathbf{v}}) &= (\mathbf{f}, \hat{\mathbf{v}})_{L^2(\Omega)} + \langle \mathbf{g}, \hat{\mathbf{v}} \rangle_{\Gamma_N} \end{aligned}$$

due to Lemma 2.24 and the symmetry of $\mathcal{C}\boldsymbol{\varepsilon}(\boldsymbol{\delta}\mathbf{u})$.
Thus for $\mathbf{u}^{(k)} = \mathbf{0}$ and $\mathbf{u}_D = \mathbf{0}$ the variational problem (3.95) reduces to the well-known variational problem of linear elasticity (cf. Section 11.2 in [BS08]). Therefore (3.95) can also be used to determine the solution of linear elasticity.

The derivation of the nonlinear variational problem (3.92) and the linear variational problem (3.95) holds in general, provided that $\mathbf{P}(\mathbf{u}) \in L^2(\Omega)^{3\times 3}$ and $\mathbf{P}'(\mathbf{u}^{(k)})[\boldsymbol{\delta}\mathbf{u}] \in L^2(\Omega)^{3\times 3}$ for given $\mathbf{u}, \mathbf{u}^{(k)} \in \boldsymbol{\Phi}$ and arbitrary $\boldsymbol{\delta}\mathbf{u} \in \boldsymbol{\Phi}_{\Gamma_D}$. In the following we will focus on a hyperelastic material law of Mooney-Rivlin type with stored energy function (2.31). Consistency with linear elasticity has led to $\alpha := \alpha(\mu, \delta) = \frac{\mu}{2} - \delta, \beta := \beta(\lambda, \delta) = \frac{\lambda}{4} - \delta, \gamma := \gamma(\mu, \lambda) = \mu + \frac{\lambda}{2}$ with $0 \leq \delta < \min\left\{\frac{\lambda}{4}, \frac{\mu}{2}\right\}$ in Section 2.4.5.
Due to the definition of a hyperelastic material in (2.13) and the gradients in (2.23) we get the first Piola-Kirchhoff stress tensor as

$$\begin{aligned} \mathbf{P}_{MR} = \partial_{\mathbf{F}}\psi_{MR}(\mathbf{C}) &= 2\alpha\mathbf{F} + (2\beta(\det\mathbf{F})^2 - \gamma)\mathbf{F}^{-T} + 2\delta\,\mathbf{F}\left(\mathrm{tr}(\mathbf{C}^{-1})\mathbf{I} - \mathbf{C}^{-T}\right)\mathbf{Cof}\,\mathbf{C} \\ &= 2\alpha\mathbf{F} + (2\beta(\det\mathbf{F})^2 - \gamma)\mathbf{F}^{-T} + 2\delta(\det\mathbf{F})^2\left(\mathrm{tr}(\mathbf{B}^{-1})\mathbf{I} - \mathbf{B}^{-1}\right)\mathbf{F}^{-T} \end{aligned}$$

with $\mathbf{C} = \mathbf{F}^T\mathbf{F}$ and $\mathbf{B} = \mathbf{F}\mathbf{F}^T$.
Inserting the parameters α, β, γ leads to

$$\begin{aligned} \mathbf{P}_{MR}(\mathbf{u}) &= 2\left(\frac{\mu}{2} - \delta\right)\mathbf{F}(\mathbf{u}) + \left(2\left(\frac{\lambda}{4} - \delta\right)(\det\mathbf{F}(\mathbf{u}))^2 - \left(\mu + \frac{\lambda}{2}\right)\right)\mathbf{F}(\mathbf{u})^{-T} \\ &\quad + 2\delta(\det\mathbf{F}(\mathbf{u}))^2\left(\mathrm{tr}(\mathbf{B}(\mathbf{u})^{-1})\mathbf{I} - \mathbf{B}(\mathbf{u})^{-1}\right)\mathbf{F}(\mathbf{u})^{-T} \\ &= \underbrace{\mu\,\mathbf{F}(\mathbf{u}) + \left[\frac{\lambda}{2}\left((\det\mathbf{F}(\mathbf{u}))^2 - 1\right) - \mu\right]\mathbf{F}(\mathbf{u})^{-T}}_{=:\,\mathbf{P}_{NH}(\mathbf{u})} \\ &\quad + \underbrace{2\delta\left[(\det\mathbf{F}(\mathbf{u}))^2\left\{(\mathrm{tr}(\mathbf{B}(\mathbf{u})^{-1}) - 1)\mathbf{I} - \mathbf{B}(\mathbf{u})^{-1}\right\}\mathbf{F}(\mathbf{u})^{-T} - \mathbf{F}(\mathbf{u})\right]}_{=:\,\mathbf{P}_{add}(\mathbf{u})}. \end{aligned} \tag{3.96}$$

$\mathbf{P}_{MR}(\mathbf{u})$ is therefore decomposed into the first Piola-Kirchhoff stress tensor for the Neo-Hooke case, i.e. $\mathbf{P}_{NH}(\mathbf{u})$, and an additional part $\mathbf{P}_{add}(\mathbf{u})$.
We set $g_1(\mathbf{u}) := (\det\mathbf{F}(\mathbf{u}))^2$, $g_2(\mathbf{u}) := \mathbf{B}(\mathbf{u})^{-1}$, $g_3(\mathbf{u}) := \mathrm{tr}\left(\mathbf{B}(\mathbf{u})^{-1}\right)$, $g_4(\mathbf{u}) := \mathbf{F}(\mathbf{u})^{-T}$

and $g_5(\mathbf{u}) := \mathbf{F}(\mathbf{u})$. As presented in Section 2.4.5 one analogously obtains the Fréchet derivatives with respect to $\mathbf{u}$ as

$$\begin{aligned}
g_1'(\mathbf{u})[\mathbf{v}] &= 2(\det \mathbf{F}(\mathbf{u}))^2 \mathrm{tr}\left(\mathbf{F}(\mathbf{u})^{-1}\nabla\mathbf{v}\right),\\
g_2'(\mathbf{u})[\mathbf{v}] &= -\mathbf{B}(\mathbf{u})^{-1}\left(\nabla\mathbf{v}(\mathbf{F}(\mathbf{u}))^T + \mathbf{F}(\mathbf{u})(\nabla\mathbf{v})^T\right)\mathbf{B}(\mathbf{u})^{-1},\\
g_3'(\mathbf{u})[\mathbf{v}] &= \mathrm{tr}\left(g_2'(\mathbf{u})[\mathbf{v}]\right) = -2\,\mathrm{tr}\left(\mathbf{B}(\mathbf{u})^{-1}\nabla\mathbf{v}(\mathbf{F}(\mathbf{u}))^T\mathbf{B}(\mathbf{u})^{-1}\right),\\
g_4'(\mathbf{u})[\mathbf{v}] &= -\mathbf{F}(\mathbf{u})^{-T}(\nabla\mathbf{v})^T\mathbf{F}(\mathbf{u})^{-T},\\
g_5'(\mathbf{u})[\mathbf{v}] &= \nabla\mathbf{v}.
\end{aligned} \tag{3.97}$$

For the calculation of $g_3'(\mathbf{u})[\mathbf{v}]$ we have used Lemma 2.24. The Fréchet derivative of $\mathbf{P}_{MR}(\mathbf{u})$ with respect to $\mathbf{u}$ is then

$$\mathbf{P}_{MR}'(\mathbf{u})[\mathbf{v}] = \mathbf{P}_{NH}'(\mathbf{u})[\mathbf{v}] + \mathbf{P}_{add}'(\mathbf{u})[\mathbf{v}]$$

with components

$$\begin{aligned}
\mathbf{P}_{NH}'(\mathbf{u})[\mathbf{v}] &= \mu\nabla\mathbf{v} - \left[\frac{\lambda}{2}\left((\det\mathbf{F}(\mathbf{u}))^2 - 1\right) - \mu\right]\mathbf{F}(\mathbf{u})^{-T}(\nabla\mathbf{v})^T\mathbf{F}(\mathbf{u})^{-T}\\
&\quad + \lambda\left(\det\mathbf{F}(\mathbf{u})\right)^2\mathrm{tr}\big(\mathbf{F}(\mathbf{u})^{-1}\nabla\mathbf{v}\big)\mathbf{F}(\mathbf{u})^{-T},\\
\mathbf{P}_{add}'(\mathbf{u})[\mathbf{v}] &= 2\delta\left[g_1'(\mathbf{u})[\mathbf{v}]\left\{(g_3(\mathbf{u})-1)\mathbf{I} - g_2(\mathbf{u})\right\}g_4(\mathbf{u})\right.\\
&\quad + g_1(\mathbf{u})\left(\left\{g_3'(\mathbf{u})[\mathbf{v}]\mathbf{I} - g_2'(\mathbf{u})[\mathbf{v}]\right\}g_4(\mathbf{u}) + \left\{(g_3(\mathbf{u})-1)\mathbf{I} - g_2(\mathbf{u})\right\}g_4'(\mathbf{u})[\mathbf{v}]\right)\\
&\quad \left. - g_5'(\mathbf{u})[\mathbf{v}]\right].
\end{aligned} \tag{3.98}$$

With these observations one can also show that the bilinear form, defined in (3.94), is symmetric for the considered Mooney - Rivlin material.
In the following we study the pure displacement approach in a plane strain model and the Mooney - Rivlin material law.

Restriction to a plane strain model:

In a plane strain model we recall that the deformation gradient reduces to

$$\mathbf{F}(\mathbf{u}) = \begin{pmatrix} 1+\partial_1 u_1 & \partial_2 u_1 & 0\\ \partial_1 u_2 & 1+\partial_2 u_2 & 0\\ 0 & 0 & 1\end{pmatrix} =: \begin{pmatrix}\hat{\mathbf{F}}(\mathbf{u}) & 0\\ 0 & 1\end{pmatrix}.$$

Consequently we recall that the stress tensor and the left Cauchy - Green strain tensor are given by

$$\begin{aligned}
\mathbf{P}_{MR} &= \begin{pmatrix} P_{11} & P_{12} & 0\\ P_{21} & P_{22} & 0\\ 0 & 0 & P_{33}\end{pmatrix} =: \begin{pmatrix}\hat{\mathbf{P}}_{MR} & 0\\ 0 & 1\end{pmatrix},\\
\mathbf{B}(\mathbf{u}) &= \mathbf{F}(\mathbf{u})(\mathbf{F}(\mathbf{u}))^T = \begin{pmatrix}\hat{\mathbf{F}}(\mathbf{u})(\hat{\mathbf{F}}(\mathbf{u}))^T & 0\\ 0 & 1\end{pmatrix} =: \begin{pmatrix}\hat{\mathbf{B}}(\mathbf{u}) & 0\\ 0 & 1\end{pmatrix}.
\end{aligned}$$

Obviously it follows $\mathrm{tr}(\mathbf{B}(\mathbf{u})^{-1}) = \mathrm{tr}(\hat{\mathbf{B}}(\mathbf{u})^{-1}) + 1$, $\det \mathbf{F}(\mathbf{u}) = \det \hat{\mathbf{F}}(\mathbf{u})$ and therefore with the help of equation (3.96)

$$\begin{aligned}
(\mathbf{P}_{MR}(\mathbf{u}))_{1:2,1:2} &= (\mathbf{P}_{NH}(\mathbf{u}))_{1:2,1:2} \\
&\quad + 2\delta \left(\left[(\det \mathbf{F}(\mathbf{u}))^2 \left\{ (\mathrm{tr}(\mathbf{B}(\mathbf{u})^{-1}) - 1)\mathbf{I} - \mathbf{B}(\mathbf{u})^{-1} \right\} \mathbf{F}(\mathbf{u})^{-T} - \mathbf{F}(\mathbf{u}) \right] \right)_{1:2,1:2} \\
&= (\mathbf{P}_{NH}(\mathbf{u}))_{1:2,1:2} + 2\delta \left[(\det \hat{\mathbf{F}}(\mathbf{u}))^2 \left\{ \left(\mathrm{tr}(\hat{\mathbf{B}}(\mathbf{u})^{-1}) + 1 - 1 \right) \mathbf{I} - \hat{\mathbf{B}}(\mathbf{u})^{-1} \right\} \hat{\mathbf{F}}(\mathbf{u})^{-T} - \hat{\mathbf{F}}(\mathbf{u}) \right] \\
&= (\mathbf{P}_{NH}(\mathbf{u}))_{1:2,1:2} + 2\delta \left[(\det \hat{\mathbf{F}}(\mathbf{u}))^2 \left(\mathrm{tr}(\hat{\mathbf{B}}(\mathbf{u})^{-1})\mathbf{I} - \hat{\mathbf{B}}(\mathbf{u})^{-1} \right) \hat{\mathbf{F}}(\mathbf{u})^{-T} - \hat{\mathbf{F}}(\mathbf{u}) \right]
\end{aligned}$$

for the components of $\mathbf{P}_{MR}(\mathbf{u})$ in the first two rows and columns.
In two dimensions it holds $\mathrm{tr}(\mathbf{Cof}\,\mathbf{A}) = \mathrm{tr}(\mathbf{A})$ and $\mathrm{tr}(\mathbf{A})\mathbf{I} - (\mathbf{Cof}\,\mathbf{A})^T = \mathbf{A}$ for an arbitrary matrix $\mathbf{A} \in \mathbb{R}^{2\times 2}$. With these ingredients and the relation $\hat{\mathbf{B}}(\mathbf{u}) = \hat{\mathbf{F}}(\mathbf{u})(\hat{\mathbf{F}}(\mathbf{u}))^T$ it actually holds

$$\begin{aligned}
&(\det \hat{\mathbf{F}}(\mathbf{u}))^2 \left(\mathrm{tr}(\hat{\mathbf{B}}(\mathbf{u})^{-1})\mathbf{I} - \hat{\mathbf{B}}(\mathbf{u})^{-1} \right) \hat{\mathbf{F}}(\mathbf{u})^{-T} - \hat{\mathbf{F}}(\mathbf{u}) \\
&\qquad = \left(\mathrm{tr}(\mathbf{Cof}\,\hat{\mathbf{B}}(\mathbf{u}))\mathbf{I} - (\mathbf{Cof}\,\hat{\mathbf{B}}(\mathbf{u}))^T \right) \hat{\mathbf{F}}(\mathbf{u})^{-T} - \hat{\mathbf{F}}(\mathbf{u}) \\
&\qquad = \left(\mathrm{tr}(\hat{\mathbf{B}}(\mathbf{u}))\mathbf{I} - (\mathbf{Cof}\,\hat{\mathbf{B}}(\mathbf{u}))^T \right) \hat{\mathbf{F}}(\mathbf{u})^{-T} - \hat{\mathbf{F}}(\mathbf{u}) \\
&\qquad = \hat{\mathbf{B}}(\mathbf{u})\hat{\mathbf{F}}(\mathbf{u})^{-T} - \hat{\mathbf{F}}(\mathbf{u}) = \hat{\mathbf{F}}(\mathbf{u})(\hat{\mathbf{F}}(\mathbf{u}))^T \hat{\mathbf{F}}(\mathbf{u})^{-T} - \hat{\mathbf{F}}(\mathbf{u}) = \mathbf{0}
\end{aligned}$$

and it follows $(\mathbf{P}_{MR}(\mathbf{u}))_{1:2,1:2} = (\mathbf{P}_{NH}(\mathbf{u}))_{1:2,1:2}$. The stress tensors $\mathbf{P}_{NH}(\mathbf{u})$ and $\mathbf{P}_{MR}(\mathbf{u})$ differ therefore only in one component, namely

$$\begin{aligned}
(\mathbf{P}_{MR}(\mathbf{u}))_{33} &= (\mathbf{P}_{NH}(\mathbf{u}))_{33} \\
&\quad + 2\delta \left(\left[(\det \mathbf{F}(\mathbf{u}))^2 \left\{ (\mathrm{tr}(\mathbf{B}(\mathbf{u})^{-1}) - 1)\mathbf{I} - \mathbf{B}(\mathbf{u})^{-1} \right\} \mathbf{F}(\mathbf{u})^{-T} - \mathbf{F}(\mathbf{u}) \right] \right)_{33} \\
&= (\mathbf{P}_{NH}(\mathbf{u}))_{33} + 2\delta \left[(\det \hat{\mathbf{F}}(\mathbf{u}))^2 \left\{ \left(\mathrm{tr}(\hat{\mathbf{B}}(\mathbf{u})^{-1}) + 1 - 1 \right) \cdot 1 - 1 \right\} \cdot 1 - 1 \right] \\
&= (\mathbf{P}_{NH}(\mathbf{u}))_{33} + 2\delta \left[(\det \hat{\mathbf{F}}(\mathbf{u}))^2 \left(\mathrm{tr}(\hat{\mathbf{B}}(\mathbf{u})^{-1}) - 1 \right) - 1 \right],
\end{aligned}$$

whose additional term $2\delta \left[\det(\hat{\mathbf{F}}(\mathbf{u}))^2 \left(\mathrm{tr}(\hat{\mathbf{B}}(\mathbf{u})^{-1}) - 1 \right) - 1 \right]$ is in general unequal to zero.

Proposition 3.35: (Pure displacement approach with Mooney - Rivlin and plane strain configuration)
Assume that we use a plane strain model. Then the pure displacement approach for the considered Mooney - Rivlin material with stored energy function (2.31) leads for all possible values $0 \leq \delta < \min\left\{\frac{\lambda}{4}, \frac{\mu}{2}\right\}$ to the same displacement approximation as using the Neo - Hooke material, i.e. (2.31) with $\delta = 0$.

Proof:

Since we are dealing with a plane strain configuration it holds $\nabla \mathbf{v} = \begin{pmatrix} \partial_1 v_1 & \partial_2 v_1 & 0 \\ \partial_1 v_2 & \partial_2 v_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ for $\mathbf{v} \in \mathbf{\Phi}_{\Gamma_D}$.
Inserting $\nabla \mathbf{v}$ into (3.98) leads to

$$\left(\mathbf{P}'_{add}(\mathbf{u})[\mathbf{v}]\right)_{33} = 2\delta \left(g_1'(\mathbf{u})[\mathbf{v}](g_3(\mathbf{u}) - 2) + g_1(\mathbf{u})g_3'(\mathbf{u})[\mathbf{v}]\right)$$

and $\left(\mathbf{P}'_{add}(\mathbf{u})[\mathbf{v}]\right)_{ij} = 0$ for $(i,j) \in \{1,2,3\} \times \{1,2,3\} \setminus \{(3,3)\}$. These identities could be proven with the help of a long calculation or simply with the help of the Symbolic Math Toolbox™ in MATLAB® as done here.
The variational formulations (3.92) and (3.93) do not consider the components in the last row and last column of $\mathbf{P}(\mathbf{u})$ and $\mathbf{P}'(\mathbf{u}^{(k)})[\boldsymbol{\delta}\mathbf{u}]$ due to test functions $\hat{\mathbf{v}}$ with vanishing matrix entries in $\nabla\hat{\mathbf{v}}$ in the last row and column. Only the first two rows and columns of these matrices are taken into account. But these submatrices are, by the considerations above, equal for both models. Therefore the usage of the Mooney - Rivlin model is no improvement compared to the Neo - Hooke model in a pure displacement approach with plane strain configuration.

□

Proposition 3.35 states that we can neglect the additional δ - term in the Mooney - Rivlin material compared to the Neo - Hooke model if we are using a pure displacement approach in combination with a plane strain model to approximate the displacement $\mathbf{u}$. Hence, it always holds $\mathbf{u}_{MR} = \mathbf{u}_{NH}$. If we calculate the corresponding stress tensors $\mathbf{P}_{MR}$ and $\mathbf{P}_{NH}$ in a post - processing, then these tensors can only differ in one component, i.e. $(\mathbf{P}_{MR})_{33} \neq (\mathbf{P}_{NH})_{33}$ and all other components are equal.
However, for a complete three dimensional problem without a plane strain model, the results for Mooney - Rivlin and Neo - Hooke will differ. Also for other constitutive laws that are based on the Neo - Hooke material with corresponding stress tensor $\mathbf{P} = \mathbf{P}_{NH} + \mathbf{P}_{add}$ with at least one additional non - vanishing entry of the components $(\mathbf{P}_{add})_{11}$, $(\mathbf{P}_{add})_{12}$, $(\mathbf{P}_{add})_{21}$, $(\mathbf{P}_{add})_{22}$ will lead generally to a different displacement approximation.

3.6.2 Displacement - pressure approach

Since the pure displacement approach leads to unwanted Poisson locking effects in the incompressible limit, at least if polynomials with small degree are used as conforming finite elements (cf. [BS92] for linear elasticity), we use a different discretization method for the incompressible limit $\lambda \to \infty$. The following discretization method is a mixed formulation and approximates, in addition to the displacement $\mathbf{u}$, a pressure - like variable p. This method is proposed by Auricchio et al. in [ABadVLR05] and [ABadVLR10] for a homogeneous Neo - Hooke material. The corresponding stored energy function that the authors used in their work is slightly different to the one proposed in (2.31) with $\delta = 0$. However, we use their idea and formulate a mixed method for the Mooney - Rivlin material with stored energy function (2.31) in the following:
By equation (3.96) it holds $\mathbf{P}_{MR}(\mathbf{u}) = \mathbf{P}_{NH}(\mathbf{u}) + \mathbf{P}_{add}(\mathbf{u})$ with

$$\begin{aligned}\mathbf{P}_{NH}(\mathbf{u}) &= \mu\,\mathbf{F}(\mathbf{u}) + \left[\frac{\lambda}{2}\left((\det\mathbf{F}(\mathbf{u}))^2 - 1\right) - \mu\right]\mathbf{F}(\mathbf{u})^{-T},\\ \mathbf{P}_{add}(\mathbf{u}) &= 2\delta\left[g_1(\mathbf{u})\left\{(g_3(\mathbf{u})-1)\mathbf{I} - g_2(\mathbf{u})\right\}g_4(\mathbf{u}) - g_5(\mathbf{u})\right]\end{aligned}$$

and the functions $g_1, \dots, g_5$, defined in Section 3.6.1. Here only the first part, namely $\mathbf{P}_{NH}(\mathbf{u})$, depends on λ. Introducing the pressure - like variable $p := \frac{\lambda}{2}\left((\det\mathbf{F}(\mathbf{u}))^2 - 1\right)$, or equivalently $(\det\mathbf{F}(\mathbf{u}))^2 - 1 = \frac{2p}{\lambda}$, we get

$$\mathbf{P}_{NH}(\mathbf{u}, p) := \mu\,\mathbf{F}(\mathbf{u}) + (p-\mu)\mathbf{F}(\mathbf{u})^{-T}$$

and $\mathbf{P}_{add}(\mathbf{u})$ remains unchanged and is independent of p. Hence we write $\mathbf{P}_{MR}(\mathbf{u}, p) := \mathbf{P}_{NH}(\mathbf{u}, p) + \mathbf{P}_{add}(\mathbf{u})$.
For the limit $\lambda \to \infty$ we get $(\det\mathbf{F}(\mathbf{u}))^2 - 1 = 0$. Since $\det\mathbf{F}(\mathbf{u}) > 0$ (cf. Section 2.2.1), this condition is equivalent to $\det\mathbf{F}(\mathbf{u}) - 1 = 0$ and confirms the incompressibility constraint (2.9). We get the nonlinear mixed formulation:
Find the pair $(\mathbf{u}, p) \in W^{1,s}(\Omega)^3 \times L^r(\Omega)$ with sufficiently large $s, r \geq 2$ such that

$$\begin{aligned}&(\mathbf{P}_{MR}(\mathbf{u}, p), \nabla\mathbf{v})_{L^2(\Omega)} = (\mathbf{f}, \mathbf{v})_{L^2(\Omega)} + \langle\mathbf{g}, \mathbf{v}\rangle_{\Gamma_N} \quad \forall\,\mathbf{v} \in W^{1,s}_{\Gamma_D}(\Omega)^3,\\ &(\det\mathbf{F}(\mathbf{u}) - 1, q)_{L^2(\Omega)} = 0 \quad \forall\,q \in L^r(\Omega).\end{aligned} \tag{3.99}$$

Here $\mathbf{u}$ has to satisfy again the prescribed boundary condition on Γ_D, i.e. $\mathbf{u} = \mathbf{u}_D$ on Γ_D. We obtain the Fréchet derivative of $\mathbf{P}_{MR}$ with respect to $(\mathbf{u}, p)$ (cf. Definition 2.6) as

$$\begin{aligned}\mathbf{P}'_{MR}(\mathbf{u}, p)[\boldsymbol{\delta}\mathbf{u}, \delta p] &= \mathbf{P}'_{NH}(\mathbf{u}, p)[\boldsymbol{\delta}\mathbf{u}, \delta p] + \mathbf{P}'_{add}(\mathbf{u})[\boldsymbol{\delta}\mathbf{u}]\\ &= \frac{d}{dt}\mathbf{P}_{NH}(\mathbf{u} + t\boldsymbol{\delta}\mathbf{u}, p)\big|_{t=0} + \mathbf{P}'_{add}(\mathbf{u})[\boldsymbol{\delta}\mathbf{u}] + \frac{d}{dt}\mathbf{P}_{NH}(\mathbf{u}, p + t\delta p)\big|_{t=0}\\ &= \mu\nabla\boldsymbol{\delta}\mathbf{u} + (\mu - p)\,\mathbf{F}(\mathbf{u})^{-T}\,(\nabla\boldsymbol{\delta}\mathbf{u})^T\,\mathbf{F}(\mathbf{u})^{-T} + \mathbf{P}'_{add}(\mathbf{u})[\boldsymbol{\delta}\mathbf{u}] + \delta p\,\mathbf{F}(\mathbf{u})^{-T}\\ &= \mu\nabla\boldsymbol{\delta}\mathbf{u} + (\mu - p)\,\mathbf{F}(\mathbf{u})^{-T}\,(\nabla\boldsymbol{\delta}\mathbf{u})^T\,\mathbf{F}(\mathbf{u})^{-T} + \mathbf{P}'_{add}(\mathbf{u})[\boldsymbol{\delta}\mathbf{u}] + \delta p\,\mathbf{Cof}\,\mathbf{F}(\mathbf{u}),\end{aligned}$$

where we have inserted the incompressibility constraint $\det \mathbf{F}(\mathbf{u}) = 1$ in the last step. The derivative $\mathbf{P}'_{add}(\mathbf{u})[\boldsymbol{\delta u}]$ was already determined in (3.98).
We solve the nonlinear variational formulation (3.99) again by a Newton iteration. With the help of the derivative $\mathbf{P}'_{MR}(\mathbf{u}, p)$ and Taylor's formula of order one we linearize the system (3.99) about the pair $(\mathbf{u}^{(k)}, p^{(k)}) \in W^{1,s}(\Omega)^3 \times L^r(\Omega)$, where $\mathbf{u}^{(k)} = \mathbf{u}_D$ satisfies the boundary condition on Γ_D.
Since $\left(\det \mathbf{F}(\mathbf{u}^{(k)})\right)' [\mathbf{v}] = \mathbf{Cof}\, \mathbf{F}(\mathbf{u}^{(k)}) : \nabla \mathbf{v}$ we obtain the linearized system

$$\begin{aligned} \Big(\mathbf{P}'_{MR}(\mathbf{u}^{(k)}, p^{(k)})[\boldsymbol{\delta u}, \delta p], \nabla \mathbf{v}\Big)_{L^2(\Omega)} &= -\Big(\mathbf{P}_{MR}(\mathbf{u}^{(k)}, p^{(k)}), \nabla \mathbf{v}\Big)_{L^2(\Omega)} \\ &\quad + (\mathbf{f}, \mathbf{v})_{L^2(\Omega)} + \langle \mathbf{g}, \mathbf{v} \rangle_{\Gamma_N}, \\ \Big(\mathbf{Cof}\, \mathbf{F}(\mathbf{u}^{(k)}) : \nabla \boldsymbol{\delta u}, q\Big)_{L^2(\Omega)} &= -\Big(\det \mathbf{F}(\mathbf{u}^{(k)}) - 1, q\Big)_{L^2(\Omega)} \end{aligned} \tag{3.100}$$

of (3.99) for all $(\mathbf{v}, q) \in W^{1,s}_{\Gamma_D}(\Omega)^3 \times L^r(\Omega)$.
Defining bilinear forms $a : W^{1,s}_{\Gamma_D}(\Omega)^3 \times W^{1,s}_{\Gamma_D}(\Omega)^3 \to \mathbb{R}$, $b : W^{1,s}_{\Gamma_D}(\Omega)^3 \times L^r(\Omega) \to \mathbb{R}$ and linear forms $F : W^{1,s}_{\Gamma_D}(\Omega)^3 \to \mathbb{R}$ respectively $G : L^r(\Omega) \to \mathbb{R}$ as

$$\begin{aligned} a(\boldsymbol{\delta u}, \mathbf{v}) &:= \Big(\mu \nabla \boldsymbol{\delta u} + \Big(\mu - p^{(k)}\Big)\, \mathbf{F}(\mathbf{u}^{(k)})^{-T} (\nabla \boldsymbol{\delta u})^T\, \mathbf{F}(\mathbf{u}^{(k)})^{-T}, \nabla \mathbf{v}\Big)_{L^2(\Omega)} \\ &\quad + \Big(\mathbf{P}'_{add}(\mathbf{u}^k)[\boldsymbol{\delta u}], \nabla \mathbf{v}\Big)_{L^2(\Omega)}, \\ b(\mathbf{v}, \delta p) &:= \Big(\delta p\, \mathbf{Cof}\, \mathbf{F}(\mathbf{u}^{(k)}), \nabla \mathbf{v}\Big)_{L^2(\Omega)}, \\ F(\mathbf{v}) &:= -\Big(\mathbf{P}_{MR}(\mathbf{u}^{(k)}, p^{(k)}), \nabla \mathbf{v}\Big)_{L^2(\Omega)} + (\mathbf{f}, \mathbf{v})_{L^2(\Omega)} + \langle \mathbf{g}, \mathbf{v} \rangle_{\Gamma_N}, \\ G(q) &:= -\Big(\det \mathbf{F}(\mathbf{u}^{(k)}) - 1, q\Big)_{L^2(\Omega)} \end{aligned} \tag{3.101}$$

and using $\big(\mathbf{Cof}\, \mathbf{F}(\mathbf{u}^{(k)}) : \nabla \boldsymbol{\delta u}, q\big)_{L^2(\Omega)} = \big(q\, \mathbf{Cof}\, \mathbf{F}(\mathbf{u}^{(k)}), \nabla \boldsymbol{\delta u}\big)_{L^2(\Omega)} = b(\boldsymbol{\delta u}, q)$ leads to the following linearized problem:
Find for given $(\mathbf{u}^{(k)}, p^{(k)}) \in W^{1,s}(\Omega)^3 \times L^r(\Omega)$, satisfying $\mathbf{u}^{(k)} = \mathbf{u}_D$ on Γ_D, the correction term $(\boldsymbol{\delta u}, \delta p) \in W^{1,s}_{\Gamma_D}(\Omega)^3 \times L^r(\Omega)$ such that

$$\begin{aligned} a(\boldsymbol{\delta u}, \mathbf{v}) + b(\mathbf{v}, \delta p) &= F(\mathbf{v}) \qquad \forall\, \mathbf{v} \in W^{1,s}_{\Gamma_D}(\Omega)^3, \\ b(\boldsymbol{\delta u}, q) &= G(q) \qquad \forall\, q \in L^r(\Omega). \end{aligned} \tag{3.102}$$

After solving this typical saddle point problem we set the new displacement and pressure approximations as

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \alpha^{(k)}\, \boldsymbol{\delta u} \quad \text{and} \quad p^{(k+1)} = p^{(k)} + \alpha^{(k)}\, \delta p,$$

where $0 < \alpha^{(k)} \leq 1$ is again a parameter of a globalization strategy. By construction the new approximation satisfies $\mathbf{u}^{(k+1)} = \mathbf{u}_D$ on Γ_D and will be used in the following linearized problem of the used Newton scheme. We continue this until any given stopping criterion is satisfied.

In the following of this section we show that the saddle point problem (3.102) generalizes a well - known mixed finite element method for incompressible materials in linear elasticity: If we choose $(\mathbf{u}^{(k)}, p^{(k)}) = (\mathbf{0}, 0)$ under the assumption $\mathbf{u}_D = \mathbf{0}$, it holds by definition of $g_1, \dots, g_5$ in Section 3.6.1 and their derivatives in (3.97)

$$\begin{aligned}
g_1(\mathbf{0}) &= 1, \quad g_1'(\mathbf{0})[\mathbf{v}] = 2\,\mathrm{tr}(\nabla\mathbf{v}) = 2\,\mathrm{tr}(\boldsymbol{\varepsilon}(\mathbf{v})),\\
g_2(\mathbf{0}) &= \mathbf{I}, \quad g_2'(\mathbf{0})[\mathbf{v}] = -\left(\nabla\mathbf{v} + (\nabla\mathbf{v})^T\right) = -2\boldsymbol{\varepsilon}(\mathbf{v}),\\
g_3(\mathbf{0}) &= 3, \quad g_3'(\mathbf{0})[\mathbf{v}] = -2\,\mathrm{tr}(\boldsymbol{\varepsilon}(\mathbf{v})),\\
g_4(\mathbf{0}) &= \mathbf{I}, \quad g_4'(\mathbf{0})[\mathbf{v}] = -(\nabla\mathbf{v})^T,\\
g_5(\mathbf{0}) &= \mathbf{I}, \quad g_5'(\mathbf{0})[\mathbf{v}] = \nabla\mathbf{v}.
\end{aligned}$$

With the help of (3.98) we obtain

$$\begin{aligned}
\mathbf{P}_{add}'(\mathbf{0})[\mathbf{v}] &= 2\delta\left[g_1'(\mathbf{0})[\mathbf{v}]\left\{(g_3(\mathbf{0})-1)\mathbf{I} - g_2(\mathbf{0})\right\} g_4(\mathbf{0})\right.\\
&\quad + g_1(\mathbf{0})\left(\left\{g_3'(\mathbf{0})[\mathbf{v}]\mathbf{I} - g_2'(\mathbf{0})[\mathbf{v}]\right\} g_4(\mathbf{0}) + \left\{(g_3(\mathbf{0})-1)\mathbf{I} - g_2(\mathbf{0})\right\} g_4'(\mathbf{0})[\mathbf{v}]\right)\\
&\quad \left. - g_5'(\mathbf{0})[\mathbf{v}]\right]\\
&= 2\delta\left[2\,\mathrm{tr}(\boldsymbol{\varepsilon}(\mathbf{v}))\mathbf{I} - 2\,\mathrm{tr}(\boldsymbol{\varepsilon}(\mathbf{v}))\mathbf{I} + 2\boldsymbol{\varepsilon}(\mathbf{v}) - (\nabla\mathbf{v})^T - \nabla\mathbf{v}\right] = \mathbf{0}
\end{aligned}$$

and due to consistency with linear elasticity of course $\mathbf{P}_{MR}(\mathbf{0}, 0) = \mathbf{0}$.
By the definition of the bilinear forms and linear forms in (3.101) it follows

$$\begin{aligned}
a(\boldsymbol{\delta}\mathbf{u}, \mathbf{v}) &= \left(\mu\nabla\boldsymbol{\delta}\mathbf{u} + \mu(\nabla\boldsymbol{\delta}\mathbf{u})^T, \nabla\mathbf{v}\right)_{L^2(\Omega)} = (2\mu\,\boldsymbol{\varepsilon}(\boldsymbol{\delta}\mathbf{u}), \nabla\mathbf{v})_{L^2(\Omega)}\\
&= (2\mu\,\boldsymbol{\varepsilon}(\boldsymbol{\delta}\mathbf{u}), \boldsymbol{\varepsilon}(\mathbf{v}))_{L^2(\Omega)}\\
b(\mathbf{v}, \delta p) &= (\delta p\,\mathbf{I}, \nabla\mathbf{v})_{L^2(\Omega)} = \int_\Omega \delta p\,\mathrm{tr}(\nabla\mathbf{v})\,dx = \int_\Omega \delta p\,\mathrm{div}\,\mathbf{v}\,dx = (\delta p, \mathrm{div}\,\mathbf{v})_{L^2(\Omega)}.
\end{aligned}$$

Thus by (3.102) we obtain the saddle point formulation

$$\begin{aligned}
(2\mu\,\boldsymbol{\varepsilon}(\boldsymbol{\delta}\mathbf{u}), \boldsymbol{\varepsilon}(\mathbf{v}))_{L^2(\Omega)} + (\delta p, \mathrm{div}\,\mathbf{v})_{L^2(\Omega)} &= (\mathbf{f}, \mathbf{v})_{L^2(\Omega)} + \langle\mathbf{g}, \mathbf{v}\rangle_{\Gamma_N} \quad \forall\,\mathbf{v} \in W^{1,s}_{\Gamma_D}(\Omega)^3,\\
(q, \mathrm{div}\,\boldsymbol{\delta}\mathbf{u})_{L^2(\Omega)} &= 0 \qquad \forall\,q \in L^r(\Omega),
\end{aligned} \tag{3.103}$$

for which the choice $s = r = 2$ is sufficient in linear elasticity. The saddle point formulation (3.103) is exactly the well - known mixed formulation of linear elasticity for $\lambda \to \infty$ (cf. Section 8 in [BBF13]) and is close to the Stokes problem for an incompressible fluid.
Finally, we remark that also the displacement - pressure approach leads to the same displacement results for the Mooney - Rivlin and the Neo - Hooke model in a plane strain configuration (cp. Proposition 3.35), since also in this case it holds $(\mathbf{P}_{MR}(\mathbf{u}, p))_{1:2,1:2} = (\mathbf{P}_{NH}(\mathbf{u}, p))_{1:2,1:2}$, $(\mathbf{P}_{MR}'(\mathbf{u}, p)[\mathbf{v}, q])_{1:2,1:2} = (\mathbf{P}_{NH}'(\mathbf{u}, p)[\mathbf{v}, q])_{1:2,1:2}$ and exactly these components are taken into account in (3.99) and (3.102) (cf. also (3.101)).

3.7 Advantages and disadvantages of the LSFEM approach

At the end of this chapter and the description of our least squares finite element method we would like to discuss the advantages and drawbacks of this approach.
The first drawback is that we need much regularity in our theory. We had to assume in Section 3.5 that the stress tensor is in $W^{\infty}(\mathrm{div};\,\Omega)^3$ and the displacement is in $W^{1,\infty}(\Omega)^3$. This is a quite strong regularity assumption that is not satisfied in all applications (cf. [HMW11]). However, we will see in our numerical experiments in Section 6 that our method provides even good results if these assumptions are not satisfied.
Secondly, since we are approximating the whole stress tensor in our approach, we have much more degrees of freedom in contrast to the other two approaches. Consequently the resulting linear systems of equations are larger and the computational costs for their solution increase. A challenge is to provide reasonable (preconditioned) iterative solvers for the linear systems (3.26).
Thirdly, a further issue is unfortunately existent in our approach: In numerical experiments we have observed that it is necessary to scale the first term in the least squares functional sufficiently large in order to obtain good solutions. That is the reason why we have introduced a scaling parameter ω_1 in (3.19). ω_1 has to be chosen in such a way that the size of the domain is taken into account. An unscaled functional could be used if we would rescale the domain (cf. Section 6.1.3). For practical purposes one could for instance start with $\omega_1 = 1$ and increase the number by the factor 10 as long as the displacement in a particular point remains unchanged. The scaling issue is not existent in the Galerkin and the displacement - pressure approach, introduced in Section 3.6.
Besides these disadvantages our least squares finite element approach provides also many advantages. First of all, despite the fact that the appearing linear systems of equations are large, they have a beneficial structure. The stiffness matrix is always symmetric and, as long as the corresponding bilinear form to the linearized problem is coercive, it is moreover positive definite (cf. Section 3.3.3). This is a pleasant property for the development of suitable solvers and preconditioners.
Secondly, besides the displacement we automatically obtain an approximation of the occurring stresses with our method. The stresses of a deformed body are very important for engineers, since high stresses could practically lead for instance to cracks. In contrast to the other both discretization schemes we need no post - processing to obtain the stress tensor $\mathbf{P}$. Moreover, we do not lose any approximation quality in our approach. We will see exemplarily in Section 6 that the stress approximations in our LSFEM approach are better than in the other two approaches.
Thirdly, as a general advantage of least squares finite element methods, our approach is not restricted to any discrete inf - sup condition (cf. Chapter III § 4 in [Bra07] and [BG09]). Thus we can combine arbitrary finite element spaces for the stresses and displacements in

contrast to mixed methods with saddle point structure.
Fourthly, we will see in Section 6 that we can determine critical loads correctly with our approach. Auricchio et al. tried in [ABadVLR10] to determine critical load values in several benchmark tests with the help of the displacement - pressure approach. The authors observed that their results are for many combinations of finite elements unsatisfactory. We will see exemplarily in Section 6.1.4 that our approach can determine the right values for the same problems without any difficulty and with quite simple elements.
Fifthly, our method for the Neo - Hooke material is robust in the incompressible limit $\lambda \to \infty$. As we have observed it is actually possible to set $\lambda = \infty$ in this case. In particular we have no unwanted Poisson locking effect as often observed for the Galerkin method using small polynomial degrees for the approximation. Also for more complicated nonlinear models our approach is promising for quasi - incompressible materials, since by the inversion of the stress - strain relation the inverse material law should not blow up for $\lambda \to \infty$.
Sixthly and lastly, the least squares finite element method generally provides a candidate for an a - posteriori error estimator as by - product, namely the least squares functional itself. In the case of the Neo - Hooke material and the $\mathbf{B}$ - formulation we have proven that the corresponding least squares functional is a reliable and efficient error estimator, at least for small stresses and displacements close to the origin (cf. Section 3.5). Hence adaptive mesh refinement is possible without any difficulty. Practically we have observed that this estimator also identifies „problematical“ regions for larger displacements and stresses, i.e. beyond the theoretically guaranteed range. Furthermore we will introduce the so - called model adaptivity in Section 5 which is again based on the nonlinear least squares functional as error estimator. We will see that our approach is also promising concerning this direction.

4 Inverse LSFEM approach for transverse isotropy

In Section 3.3 we have derived least squares finite element methods for homogeneous isotropic frame - indifferent hyperelastic materials. The main idea was to invert the given stress - strain relation and express the material law in terms of strains instead of stresses. In this part of the work we will show that this approach, at least for the $\mathbf{C}$ - formulation, can be easily extended to anisotropic hyperelastic materials. Only the modeling, more precisely the stored energy function, has to be adjusted.
We follow the explanations in [Sch10] and [BSN10] for the modeling. After defining suitable additional invariants for general anisotropic materials, we consider in particular the case of transverse isotropy. In this case one has a so - called **preferred direction**, denoted by a vector $\mathbf{a} \in \mathbb{R}^3$. In the planes perpendicular to $\mathbf{a}$, the elasticity properties of a material remain independent of the direction. An example for transverse isotropy is wood with preferred direction in the wood fibers. In this chapter we denote the set of all rotations in $\mathbb{R}^3$ as $\mathbb{O} := \{\mathbf{Q} \in \mathbb{R}^{3\times 3} : \mathbf{Q}^T\mathbf{Q} = \mathbf{I} = \mathbf{Q}\mathbf{Q}^T, \det \mathbf{Q} = 1\}$ and the set of all matrices with positive determinant as $\mathbb{M} := \{\mathbf{F} \in \mathbb{R}^{3\times 3} : \det \mathbf{F} > 0\}$, as before.

4.1 Modeling of anisotropic materials

In Section 2.2.5 we have defined the isotropy of a stored energy function $\hat{\psi} : \bar{\Omega} \times \mathbb{M} \to \mathbb{R}$ as the condition

$$\hat{\psi}(\mathbf{x}, \mathbf{F}) = \hat{\psi}(\mathbf{x}, \mathbf{F}\mathbf{Q}), \quad \mathbf{x} \in \bar{\Omega}, \mathbf{F} \in \mathbb{M}, \mathbf{Q} \in \mathbb{O}. \tag{4.1}$$

The physically necessary property of material frame - indifference in hyperelasticity was also introduced in Section 2.2.5 as

$$\hat{\psi}(\mathbf{x}, \mathbf{F}) = \hat{\psi}(\mathbf{x}, \mathbf{Q}\mathbf{F}), \quad \mathbf{x} \in \bar{\Omega}, \mathbf{F} \in \mathbb{M}, \mathbf{Q} \in \mathbb{O}. \tag{4.2}$$

If we combine both properties we obtain

$$\hat{\psi}(\mathbf{x}, \mathbf{Q}_1\mathbf{F}\mathbf{Q}_2) = \hat{\psi}(\mathbf{x}, \mathbf{Q}_1\mathbf{F}) = \hat{\psi}(\mathbf{x}, \mathbf{F}), \quad \mathbf{x} \in \bar{\Omega}, \mathbf{F} \in \mathbb{M}, \mathbf{Q}_1, \mathbf{Q}_2 \in \mathbb{O}, \tag{4.3}$$

often called **orthogonal invariance**.
For anisotropic materials the property (4.1) holds only in a subset $\mathbb{G} \subset \mathbb{O}$. (4.2) is for anisotropic materials still a necessary requirement for all $\mathbf{Q} \in \mathbb{O}$. Altogether (4.3) holds for an anisotropic material only in the subset $\mathbb{G}$. The set $\mathbb{G}$ is called **material symmetry group**.
Following the explanations in [Sch10] and [BSN10], we split the total stored energy function of an anisotropic hyperelastic material into an isotropic and an anisotropic part

$$\hat{\psi}(\mathbf{x}, \mathbf{F}, \boldsymbol{\Xi}) = \hat{\boldsymbol{\psi}}_{iso}(\mathbf{x}, \mathbf{F}) + \hat{\boldsymbol{\psi}}_{aniso}(\mathbf{x}, \mathbf{F}, \boldsymbol{\Xi}), \quad \mathbf{x} \in \bar{\Omega}, \mathbf{F} \in \mathbb{M}, \tag{4.4}$$

where $\boldsymbol{\Xi} := \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{M}_1, \mathbf{M}_2, \ldots\}$ denotes a set of **structural tensors**. The stored energy function $\hat{\psi}(\mathbf{x}, \mathbf{F}, \boldsymbol{\Xi})$ then shall satisfy

$$\hat{\psi}(\mathbf{x}, \mathbf{F}, \boldsymbol{\Xi}) = \hat{\psi}(\mathbf{x}, \mathbf{Q}\mathbf{F}\mathbf{Q}^T, \mathbf{Q}\mathbf{a}_1, \mathbf{Q}\mathbf{a}_2, \ldots, \mathbf{Q}\mathbf{M}_1\mathbf{Q}^T, \mathbf{Q}\mathbf{M}_2\mathbf{Q}^T, \ldots) = \hat{\psi}(\mathbf{x}, \mathbf{Q}\mathbf{F}\mathbf{Q}^T, \mathbf{Q} * \boldsymbol{\Xi}) \tag{4.5}$$

for all $\mathbf{Q} \in \mathbb{O}$ with the abbreviation $\mathbf{Q} * \boldsymbol{\Xi} := (\mathbf{Q}\mathbf{a}_1, \mathbf{Q}\mathbf{a}_2, \ldots, \mathbf{Q}\mathbf{M}_1\mathbf{Q}^T, \mathbf{Q}\mathbf{M}_2\mathbf{Q}^T, \ldots)$. In the following we define, based on [Sch10], suitable ingredients of the stored energy function of an anisotropic material.

Definition 4.1: (Mixed invariants)
Let $\mathbf{a} \in \mathbb{R}^3$ be given. Then we define for arbitrary $\mathbf{F} \in \mathbb{R}^{3\times3}$ the **mixed invariants**

$$\begin{aligned}
\hat{J}_4(\mathbf{F}, \mathbf{a}) &:= |\mathbf{F}\mathbf{a}|^2, \\
\hat{J}_5(\mathbf{F}, \mathbf{a}) &:= |\mathbf{F}^T\mathbf{F}\mathbf{a}|^2, \\
\hat{K}_1(\mathbf{F}, \mathbf{a}) &:= \hat{J}_5(\mathbf{F}, \mathbf{a}) - \hat{I}_1(\mathbf{F})\hat{J}_4(\mathbf{F}, \mathbf{a}) + \hat{I}_2(\mathbf{F}), \\
\hat{K}_2(\mathbf{F}, \mathbf{a}) &:= \hat{I}_1(\mathbf{F}) - \hat{J}_4(\mathbf{F}, \mathbf{a}), \\
\hat{K}_3(\mathbf{F}, \mathbf{a}) &:= \hat{I}_1(\mathbf{F})\hat{J}_4(\mathbf{F}, \mathbf{a}) - \hat{J}_5(\mathbf{F}, \mathbf{a}),
\end{aligned}$$

where $\hat{I}_1(\mathbf{F}) = |\mathbf{F}|^2$ and $\hat{I}_2(\mathbf{F}) = |\mathbf{Cof}\,\mathbf{F}|^2$, $\hat{I}_3(\mathbf{F}) = (\det \mathbf{F})^2$ were already introduced in Section 2.3.3.

Note that $\hat{K}_1(\mathbf{F}, \mathbf{a})$ and $\hat{K}_2(\mathbf{F}, \mathbf{a})$ can be expressed by linear combinations of the set $\mathcal{I} := \{\hat{I}_1(\mathbf{F}), \hat{I}_2(\mathbf{F}), \hat{I}_3(\mathbf{F}), \hat{J}_4(\mathbf{F}, \mathbf{a}), \hat{K}_3(\mathbf{F}, \mathbf{a})\}$ and therefore $\mathcal{I}$ is for instance one possibility for an independent set of invariants. The stored energy function (4.4) will be defined later in terms of the principal and mixed invariants. In the following we often use the identity $\mathrm{tr}(\mathbf{y} \cdot \mathbf{y}^T) = \sum_{i=1}^{n} y_i^2 = |\mathbf{y}|^2$ which is valid for arbitrary column vectors $\mathbf{y} \in \mathbb{R}^n$.
Moreover, we define for arbitrary $\mathbf{A}, \mathbf{M} \in \mathbb{R}^{3\times3}$ the mappings $J_4(\mathbf{A}, \mathbf{M}) := \mathrm{tr}(\mathbf{A}\mathbf{M})$, $J_5(\mathbf{A}, \mathbf{M}) := \mathrm{tr}(\mathbf{A}^2\mathbf{M})$, $K_1(\mathbf{A}, \mathbf{M}) := J_5(\mathbf{A}, \mathbf{M}) - I_1(\mathbf{A})J_4(\mathbf{A}, \mathbf{M}) + I_2(\mathbf{A})$, $K_2(\mathbf{A}, \mathbf{M}) := I_1(\mathbf{A}) - J_4(\mathbf{A}, \mathbf{M})$ and $K_3(\mathbf{A}, \mathbf{M}) := I_1(\mathbf{A})J_4(\mathbf{A}, \mathbf{M}) - J_5(\mathbf{A}, \mathbf{M})$. For the special choice $\mathbf{A} = \mathbf{C} = \mathbf{F}^T\mathbf{F}$, $\mathbf{M} = \mathbf{a} \cdot \mathbf{a}^T$ we obtain similar to $I_1(\mathbf{C}) = \mathrm{tr}(\mathbf{C}) = \hat{I}_1(\mathbf{F})$, $I_2(\mathbf{C}) = \mathrm{tr}(\mathbf{Cof}\,\mathbf{C}) = \hat{I}_2(\mathbf{F})$, $I_3(\mathbf{C}) = \det \mathbf{C} = \hat{I}_3(\mathbf{F})$ (cf. Section 2.3.3) the relations

$$\begin{aligned}
J_4(\mathbf{C}, \mathbf{M}) &= \mathrm{tr}(\mathbf{C}\mathbf{M}) = \mathrm{tr}(\mathbf{F}^T\mathbf{F}\mathbf{a}\mathbf{a}^T) = \mathrm{tr}(\mathbf{F}\mathbf{a}(\mathbf{F}\mathbf{a})^T) = |\mathbf{F}\mathbf{a}|^2 = \hat{J}_4(\mathbf{F}, \mathbf{a}), \\
J_5(\mathbf{C}, \mathbf{M}) &= \mathrm{tr}(\mathbf{C}^2\mathbf{M}) = \mathrm{tr}(\mathbf{F}^T\mathbf{F}\mathbf{F}^T\mathbf{F}\mathbf{a}\mathbf{a}^T) = \mathrm{tr}(\mathbf{F}^T\mathbf{F}\mathbf{a}\mathbf{a}^T\mathbf{F}^T\mathbf{F}) \\
&= \mathrm{tr}(\mathbf{F}^T\mathbf{F}\mathbf{a}(\mathbf{F}^T\mathbf{F}\mathbf{a})^T) = |\mathbf{F}^T\mathbf{F}\mathbf{a}|^2 = \hat{J}_5(\mathbf{F}, \mathbf{a}), \\
K_1(\mathbf{C}, \mathbf{M}) &= J_5(\mathbf{C}, \mathbf{M}) - I_1(\mathbf{C})J_4(\mathbf{C}, \mathbf{M}) + I_2(\mathbf{C}) \\
&= \hat{J}_5(\mathbf{F}, \mathbf{a}) - \hat{I}_1(\mathbf{F})\hat{J}_4(\mathbf{F}, \mathbf{a}) + \hat{I}_2(\mathbf{F}) = \hat{K}_1(\mathbf{F}, \mathbf{a}), \\
K_2(\mathbf{C}, \mathbf{M}) &= I_1(\mathbf{C}) - J_4(\mathbf{C}, \mathbf{M}) = \hat{I}_1(\mathbf{F}) - \hat{J}_4(\mathbf{F}, \mathbf{a}) = \hat{K}_2(\mathbf{F}, \mathbf{a}), \\
K_3(\mathbf{C}, \mathbf{M}) &= I_1(\mathbf{C})J_4(\mathbf{C}, \mathbf{M}) - J_5(\mathbf{C}, \mathbf{M}) = \hat{I}_1(\mathbf{F})\hat{J}_4(\mathbf{F}, \mathbf{a}) - \hat{J}_5(\mathbf{F}, \mathbf{a}) = \hat{K}_3(\mathbf{F}, \mathbf{a}).
\end{aligned}$$

Thus, for instance if we have the set $\boldsymbol{\Xi} = (\mathbf{a}, \mathbf{M})$ of structural tensors with $\mathbf{M} := \mathbf{a} \cdot \mathbf{a}^T$, $\mathbf{a} \in \mathbb{R}^3$ given, and choose a stored energy function $\hat{\psi}(\mathbf{x}, \mathbf{F}, \mathbf{a}) := \hat{\psi}(\mathbf{x}, \mathbf{F}, \boldsymbol{\Xi})$ in terms of the principal and mixed invariants, we can express the stored energy function, similar to (2.19), in terms of $\mathbf{C}$, i.e.

$$\psi(\mathbf{x}, \mathbf{C}, \mathbf{M}) = \hat{\psi}(\mathbf{x}, \mathbf{F}, \mathbf{a}), \quad \mathbf{x} \in \bar{\Omega}, \mathbf{C} = \mathbf{F}^T\mathbf{F}, \mathbf{F} \in \mathbb{M}.$$

For our later purpose, the computation of the mapping $\tilde{\mathcal{G}}$ and its Gâteaux derivative for an anisotropic material, which will be done similar to the derivations in Section 3.3, it is necessary to compute the derivatives of the principal and mixed invariants and their gradients (cf. Section 2.1.3). For the principal invariants this was already done in Section 2.3.3. In three dimensions we recall the Gâteaux derivatives with respect to an invertible matrix $\mathbf{A} \in \mathbb{R}^{3\times3}$ and its evaluation in $\mathbf{A} = \mathbf{I}$ as

$$\begin{aligned} I_1'(\mathbf{A})[\mathbf{H}] &= \mathbf{I} : \mathbf{H}, & I_1'(\mathbf{I})[\mathbf{H}] &= \operatorname{tr}(\mathbf{H}), \\ I_2'(\mathbf{A})[\mathbf{H}] &= \left(\operatorname{tr}(\mathbf{A}^{-1})\mathbf{I} - \mathbf{A}^{-T}\right) \mathbf{Cof}\, \mathbf{A} : \mathbf{H}, & I_2'(\mathbf{I})[\mathbf{H}] &= 2\operatorname{tr}(\mathbf{H}), \\ I_3'(\mathbf{A})[\mathbf{H}] &= \mathbf{Cof}\, \mathbf{A} : \mathbf{H}, & I_3'(\mathbf{I})[\mathbf{H}] &= \operatorname{tr}(\mathbf{H}) \end{aligned} \tag{4.6}$$

with arbitrary $\mathbf{H} \in \mathbb{R}^{3\times3}$. For $\mathbf{C} = \mathbf{F}^T\mathbf{F}$ the calculation of the gradients with respect to $\mathbf{F} \in \mathbb{R}^{3\times3}$ were an immediate consequence in Section 2.3.3. We recall the result as

$$\begin{aligned} \partial_{\mathbf{F}} I_1(\mathbf{C}) &= 2\mathbf{F}, \\ \partial_{\mathbf{F}} I_2(\mathbf{C}) &= 2\mathbf{F}\left(\operatorname{tr}(\mathbf{C})\mathbf{I} - \mathbf{C}\right), \\ \partial_{\mathbf{F}} I_3(\mathbf{C}) &= 2(\det \mathbf{F})^2\mathbf{F}^{-T} = 2(\det \mathbf{C})\mathbf{F}^{-T}. \end{aligned} \tag{4.7}$$

In the following lemma we itemize the Gâteaux derivatives of the mixed invariants $J_4(\mathbf{A}, \mathbf{M})$, $J_5(\mathbf{A}, \mathbf{M})$, $K_1(\mathbf{A}, \mathbf{M})$, $K_2(\mathbf{A}, \mathbf{M})$ and $K_3(\mathbf{A}, \mathbf{M})$ with respect to a not necessarily symmetric but invertible matrix $\mathbf{A} \in \mathbb{R}^{3\times3}$ and its evaluation in $\mathbf{A} = \mathbf{I}$. Moreover, for the choice $\mathbf{A} = \mathbf{C} = \mathbf{F}^T\mathbf{F}$ we state their gradients with respect to $\mathbf{F}$.

Lemma 4.2: (Gâteaux derivatives and gradients of mixed invariants)
For an invertible arbitrary matrix $\mathbf{A} \in \mathbb{R}^{3\times3}$, $\mathbf{M} = \mathbf{a} \cdot \mathbf{a}^T$ with given normed $\mathbf{a} \in \mathbb{R}^3$, i.e. $|\mathbf{a}| = 1$, and $\mathbf{H} \in \mathbb{R}^{3\times3}$ it holds

$$\begin{aligned} J_4'(\mathbf{A}, \mathbf{M})[\mathbf{H}] &= \operatorname{tr}(\mathbf{MH}) = \mathbf{M} : \mathbf{H}, \\ J_5'(\mathbf{A}, \mathbf{M})[\mathbf{H}] &= \operatorname{tr}((\mathbf{AM} + \mathbf{MA})\mathbf{H}) = \left[\mathbf{MA}^T + \mathbf{A}^T\mathbf{M}\right] : \mathbf{H}, \\ K_1'(\mathbf{A}, \mathbf{M})[\mathbf{H}] &= \left[\mathbf{MA}^T + \mathbf{A}^T\mathbf{M} - \operatorname{tr}(\mathbf{AM})\mathbf{I} - \operatorname{tr}(\mathbf{A})\mathbf{M} + \left(\operatorname{tr}(\mathbf{A}^{-1})\mathbf{I} - \mathbf{A}^{-T}\right) \mathbf{Cof}\, \mathbf{A}\right] : \mathbf{H}, \\ K_2'(\mathbf{A}, \mathbf{M})[\mathbf{H}] &= \operatorname{tr}(\mathbf{H}) - \operatorname{tr}(\mathbf{MH}) = [\mathbf{I} - \mathbf{M}] : \mathbf{H}, \\ K_3'(\mathbf{A}, \mathbf{M})[\mathbf{H}] &= \left[\operatorname{tr}(\mathbf{AM})\mathbf{I} + \operatorname{tr}(\mathbf{A})\mathbf{M} - \left(\mathbf{MA}^T + \mathbf{A}^T\mathbf{M}\right)\right] : \mathbf{H}. \end{aligned} \tag{4.8}$$

For $\mathbf{A} = \mathbf{I}$ we obtain in particular

$$\begin{aligned} J_4'(\mathbf{I},\mathbf{M})[\mathbf{H}] &= \operatorname{tr}(\mathbf{MH}),\\ J_5'(\mathbf{I},\mathbf{M})[\mathbf{H}] &= 2\operatorname{tr}(\mathbf{MH}),\\ K_1'(\mathbf{I},\mathbf{M})[\mathbf{H}] &= \operatorname{tr}(\mathbf{H}) - \operatorname{tr}(\mathbf{MH}),\\ K_2'(\mathbf{I},\mathbf{M})[\mathbf{H}] &= \operatorname{tr}(\mathbf{H}) - \operatorname{tr}(\mathbf{MH}),\\ K_3'(\mathbf{I},\mathbf{M})[\mathbf{H}] &= \operatorname{tr}(\mathbf{H}) + \operatorname{tr}(\mathbf{MH}). \end{aligned} \tag{4.9}$$

For the special choice $\mathbf{A} = \mathbf{C} = \mathbf{F}^T\mathbf{F}$ the gradients of the mixed invariants with respect to $\mathbf{F}$ are then given by

$$\begin{aligned} \partial_{\mathbf{F}} J_4(\mathbf{C},\mathbf{M}) &= 2\mathbf{FM},\\ \partial_{\mathbf{F}} J_5(\mathbf{C},\mathbf{M}) &= 2\mathbf{F}(\mathbf{CM}+\mathbf{MC}),\\ \partial_{\mathbf{F}} K_1(\mathbf{C},\mathbf{M}) &= 2\mathbf{F}(\mathbf{CM}+\mathbf{MC} - \operatorname{tr}(\mathbf{CM})\mathbf{I} - \operatorname{tr}(\mathbf{C})\mathbf{M} + \operatorname{tr}(\mathbf{C})\mathbf{I} - \mathbf{C}),\\ \partial_{\mathbf{F}} K_2(\mathbf{C},\mathbf{M}) &= 2\mathbf{F}(\mathbf{I}-\mathbf{M}),\\ \partial_{\mathbf{F}} K_3(\mathbf{C},\mathbf{M}) &= 2\mathbf{F}(\operatorname{tr}(\mathbf{CM})\mathbf{I} + \operatorname{tr}(\mathbf{C})\mathbf{M} - (\mathbf{CM}+\mathbf{MC})). \end{aligned} \tag{4.10}$$

Proof:

A straightforward calculation leads to

$$J_4'(\mathbf{A},\mathbf{M})[\mathbf{H}] = \frac{d}{dt} J_4(\mathbf{A}+t\mathbf{H},\mathbf{M})\big|_{t=0} = \frac{d}{dt}\operatorname{tr}\left((\mathbf{A}+t\mathbf{H})\mathbf{M}\right)\big|_{t=0} = \operatorname{tr}(\mathbf{HM}) = \mathbf{M}:\mathbf{H},$$

due to the symmetry of $\mathbf{M}$ and the calculation rules of the trace operator.
Moreover, again with the help of the calculation rules for the trace operator, we obtain

$$\begin{aligned} J_5'(\mathbf{A},\mathbf{M})[\mathbf{H}] &= \frac{d}{dt} J_5(\mathbf{A}+t\mathbf{H},\mathbf{M})\big|_{t=0} = \frac{d}{dt}\operatorname{tr}\left((\mathbf{A}+t\mathbf{H})^2\mathbf{M}\right)\big|_{t=0} = \operatorname{tr}\left((\mathbf{HA}+\mathbf{AH})\mathbf{M}\right)\\ &= \operatorname{tr}(\mathbf{AMH}) + \operatorname{tr}(\mathbf{MAH}) = \operatorname{tr}\left((\mathbf{AM}+\mathbf{MA})\mathbf{H}\right) = \left[\mathbf{MA}^T + \mathbf{A}^T\mathbf{M}\right]:\mathbf{H}. \end{aligned}$$

With the help of the definitions above of $K_j(\mathbf{A},\mathbf{M})$ for $j = 1,2,3$, the derivatives in (4.6) and the previous calculations in this proof, we obtain

$$\begin{aligned} K_1'(\mathbf{A},\mathbf{M})[\mathbf{H}] &= J_5'(\mathbf{A},\mathbf{M})[\mathbf{H}] - I_1'(\mathbf{A})[\mathbf{H}]J_4(\mathbf{A},\mathbf{M}) - I_1(\mathbf{A})J_4'(\mathbf{A},\mathbf{M})[\mathbf{H}] + I_2'(\mathbf{A})[\mathbf{H}]\\ &= \left[\mathbf{MA}^T + \mathbf{A}^T\mathbf{M} - \operatorname{tr}(\mathbf{AM})\mathbf{I} - \operatorname{tr}(\mathbf{A})\mathbf{M} + \left(\operatorname{tr}(\mathbf{A}^{-1})\mathbf{I} - \mathbf{A}^{-T}\right)\mathbf{Cof}\,\mathbf{A}\right]:\mathbf{H},\\ K_2'(\mathbf{A},\mathbf{M})[\mathbf{H}] &= I_1'(\mathbf{A})[\mathbf{H}] - J_4'(\mathbf{A},\mathbf{M})[\mathbf{H}] = \mathbf{I}:\mathbf{H} - \mathbf{M}:\mathbf{H} = [\mathbf{I}-\mathbf{M}]:\mathbf{H},\\ K_3'(\mathbf{A},\mathbf{M})[\mathbf{H}] &= \left(-K_1(\mathbf{A},\mathbf{M}) + I_2(\mathbf{A})\right)'[\mathbf{H}]\\ &= I_1'(\mathbf{A})[\mathbf{H}]J_4(\mathbf{A},\mathbf{M}) + I_1(\mathbf{A})J_4'(\mathbf{A},\mathbf{M})[\mathbf{H}] - J_5'(\mathbf{A},\mathbf{M})[\mathbf{H}]\\ &= \left[\operatorname{tr}(\mathbf{AM})\mathbf{I} + \operatorname{tr}(\mathbf{A})\mathbf{M} - \left(\mathbf{MA}^T + \mathbf{A}^T\mathbf{M}\right)\right]:\mathbf{H}. \end{aligned}$$

Until now we have proven that (4.8) holds. Inserting $\mathbf{A} = \mathbf{I}$ in (4.8) and using the identity $\operatorname{tr}(\mathbf{M}) = 1$ directly result into (4.9).

For the proof of (4.10) we recall that by construction it holds $\hat{J}_j(\mathbf{F},\mathbf{a}) = J_j(\mathbf{C},\mathbf{M}) = J_j(\mathbf{F}^T\mathbf{F},\mathbf{M})$ for $j = 4,5$ and $\hat{K}_j(\mathbf{F},\mathbf{a}) = K_j(\mathbf{C},\mathbf{M}) = K_j(\mathbf{F}^T\mathbf{F},\mathbf{M})$ for $j = 1,2,3$ with $\mathbf{C} = \mathbf{F}^T\mathbf{F}$. Moreover, if we insert $\mathbf{A} = \mathbf{C} = \mathbf{F}^T\mathbf{F}$ in (4.8), we observe that all the terms in the first argument of the inner products in (4.8) are symmetric. By Riesz representation theorem these arguments are exactly the gradients of J_j $(j = 4,5)$ and K_j $(j = 1,2,3)$ with respect to $\mathbf{A}$, evaluated in $\mathbf{A} = \mathbf{C}$ (cf. Section 2.1.3). Thus we can use Lemma 2.24 and obtain similarly to the considerations below Proposition 2.41 the equations

$$\begin{aligned}\partial_{\mathbf{F}} J_j(\mathbf{C},\mathbf{M}) : \mathbf{H} &= \left(\partial_{\mathbf{A}} J_j(\mathbf{A},\mathbf{M})\big|_{\mathbf{A}=\mathbf{C}}\right) : \left[\mathbf{H}^T\mathbf{F} + \mathbf{F}^T\mathbf{H}\right] \\ &= 2\left(\partial_{\mathbf{A}} J_j(\mathbf{A},\mathbf{M})\big|_{\mathbf{A}=\mathbf{C}}\right) : \mathbf{F}^T\mathbf{H} = 2\mathbf{F}\left(\partial_{\mathbf{A}} J_j(\mathbf{A},\mathbf{M})\big|_{\mathbf{A}=\mathbf{C}}\right) : \mathbf{H}\end{aligned}$$

for $j = 4,5$ and analogously $\partial_{\mathbf{F}} K_j(\mathbf{C},\mathbf{M}) : \mathbf{H} = 2\mathbf{F}\left(\partial_{\mathbf{A}} K_j(\mathbf{A},\mathbf{M})\big|_{\mathbf{A}=\mathbf{C}}\right) : \mathbf{H}$ for $j = 1,2,3$, i.e. altogether the gradients in (4.10). Note that for the proof of the gradient of $K_1(\mathbf{C},\mathbf{M})$ the identity $\left(\mathrm{tr}(\mathbf{C}^{-1})\mathbf{I} - \mathbf{C}^{-T}\right)\mathbf{Cof}\,\mathbf{C} = \mathrm{tr}(\mathbf{C})\mathbf{I} - \mathbf{C}$ (cf. (2.26)), which is valid for symmetric invertible matrices $\mathbf{C} \in \mathbb{R}^{3\times3}$, must additionally be taken into account.

□

Remark 4.3:
If we consider the gradients of the mixed invariants in (4.10), it becomes clear that we cannot express the right - hand sides in terms of the left Cauchy - Green strain tensor $\mathbf{B} = \mathbf{F}\mathbf{F}^T$ after multiplying them with $\mathbf{F}^T$ from the right. However, one can multiply the gradients from left with $\mathbf{F}^{-1}$. Then the resulting right - hand sides can be expressed in terms of $\mathbf{C} = \mathbf{F}^T\mathbf{F}$. This motivates us to use the proposed $\mathbf{C}$ - formulation of Section 3.3 instead of the $\mathbf{B}$ - formulation for the extension to anisotropic materials.

In the following lemma we show that the mixed invariants with the exception of $\hat{J}_5(\mathbf{F},\mathbf{a})$ are polyconvex. For the non - polyconvexity of $\hat{J}_5$ we refer to [Sch10].

Lemma 4.4: (Polyconvexity of $\hat{J}_4(\mathbf{F},\mathbf{a})$ and $\hat{K}_j(\mathbf{F},\mathbf{a}), j = 1,2,3$)
Let $\mathbf{M} = \mathbf{a}\cdot\mathbf{a}^T$ for normed $\mathbf{a} \in \mathbb{R}^3$ (i.e. $|\mathbf{a}| = 1$) and $\mathbf{F} \in \mathbb{M}$. Then the mixed invariants $\hat{J}_4(\mathbf{F},\mathbf{a})$ and $\hat{K}_j(\mathbf{F},\mathbf{a})$, $j = 1,2,3$, are polyconvex.

Proof:
The following proof is based on the explanations in [SN03] and [Sch10]. For the proof we use the Definition 2.26 of polyconvexity and Proposition 2.27.

(a) For the polyconvexity of $\hat{J}_4(\mathbf{F},\mathbf{a}) = |\mathbf{F}\mathbf{a}|^2 = \mathrm{tr}(\mathbf{F}^T\mathbf{F}\mathbf{M})$, we have to show that $g(\mathbf{A}) := |\mathbf{A}\mathbf{a}|^2 = \mathrm{tr}(\mathbf{A}^T\mathbf{A}\mathbf{M})$, $\mathbf{A} \in \mathbb{R}^{3\times3}$, is convex on $\mathbb{R}^{3\times3}$. We obtain the Gâteaux derivatives with respect to $\mathbf{A}$ as

$$\begin{aligned}g'(\mathbf{A})[\mathbf{H}_1] &= \mathrm{tr}\left(\left(\mathbf{H}_1^T\mathbf{A} + \mathbf{A}^T\mathbf{H}_1\right)\mathbf{M}\right) = 2\,\mathrm{tr}\left(\mathbf{A}^T\mathbf{H}_1\mathbf{M}\right) \\ &= 2\,\mathrm{tr}\left(\mathbf{M}\mathbf{A}^T\mathbf{H}_1\right) = [2\mathbf{A}\mathbf{M}] : \mathbf{H}_1, \\ g''(\mathbf{A})[\mathbf{H}_1,\mathbf{H}_2] &= [2\,\mathbf{H}_2\mathbf{M}] : \mathbf{H}_1,\end{aligned}$$

since $\mathbf{M}^T = \mathbf{M}$. Hence it holds

$$\begin{aligned} g''(\mathbf{A})[\mathbf{H},\mathbf{H}] = [2\,\mathbf{HM}] : \mathbf{H} &= 2\,\mathrm{tr}\left(\mathbf{M}\,\mathbf{H}^T\mathbf{H}\right) = 2\,\mathrm{tr}\left(\mathbf{H}\mathbf{a}\mathbf{a}^T\mathbf{H}^T\right), \\ &= 2\,\mathrm{tr}\left(\mathbf{H}\mathbf{a}(\mathbf{H}\mathbf{a})^T\right) = 2|\mathbf{H}\mathbf{a}|^2 \geq 0 \end{aligned}$$

for arbitrary $\mathbf{A}, \mathbf{H} \in \mathbb{R}^{3\times 3}$. By Proposition 2.27 follows the convexity of g on $\{\mathbf{A} \in \mathbb{R}^{3\times 3}\}$ and therefore $\hat{J}_4(\mathbf{F},\mathbf{a}) = g(\mathbf{F})$ is polyconvex.

(b) For the polyconvexity of $\hat{K}_1(\mathbf{F},\mathbf{a})$ we need at first some simple considerations:
From linear algebra it is well - known that a real square matrix is a root of its own characteristic polynomial (Cayley - Hamilton theorem), i.e. it holds

$$\mathbf{C}^3 - I_1(\mathbf{C})\mathbf{C}^2 + I_2(\mathbf{C})\mathbf{C} - I_3(\mathbf{C})\mathbf{I} = \mathbf{0}$$

for the right Cauchy - Green strain tensor $\mathbf{C} = \mathbf{F}^T\mathbf{F}$. This identity is equivalent to

$$\mathbf{Cof}\,\mathbf{C} = I_3(\mathbf{C})\mathbf{C}^{-T} = I_3(\mathbf{C})\mathbf{C}^{-1} = \mathbf{C}^2 - I_1(\mathbf{C})\mathbf{C} + I_2(\mathbf{C})\mathbf{I}$$

using the symmetry of $\mathbf{C}$ and the definition of the cofactor. If we multiply this equation with $\mathbf{M}$ from the right and take the trace of it, we obtain

$$\begin{aligned} \mathrm{tr}((\mathbf{Cof}\,\mathbf{C})\mathbf{M}) &= \mathrm{tr}(\mathbf{C}^2\mathbf{M}) - I_1(\mathbf{C})\mathrm{tr}(\mathbf{C}\mathbf{M}) + I_2(\mathbf{C})\mathrm{tr}(\mathbf{M}) \\ &= J_5(\mathbf{C},\mathbf{M}) - I_1(\mathbf{C})J_4(\mathbf{C},\mathbf{M}) + I_2(\mathbf{C}) = K_1(\mathbf{C},\mathbf{M}) = \hat{K}_1(\mathbf{F},\mathbf{a}), \end{aligned}$$

due to $\mathrm{tr}(\mathbf{M}) = 1$ and the definition of $K_1(\mathbf{C},\mathbf{M})$ respectively $\hat{K}_1(\mathbf{F},\mathbf{a})$.
With the help of

$$\begin{aligned} \mathrm{tr}((\mathbf{Cof}\,\mathbf{C})\mathbf{M}) &= \mathrm{tr}((\det\mathbf{F})\mathbf{F}^{-1}(\det\mathbf{F})\mathbf{F}^{-T}\mathbf{a}\mathbf{a}^T) \\ &= \mathrm{tr}((\mathbf{Cof}\,\mathbf{F})^T(\mathbf{Cof}\,\mathbf{F})\mathbf{a}\mathbf{a}^T) = |(\mathbf{Cof}\,\mathbf{F})\mathbf{a}|^2 \end{aligned}$$

we end up with the identity

$$\hat{K}_1(\mathbf{F},\mathbf{a}) = |(\mathbf{Cof}\,\mathbf{F})\mathbf{a}|^2.$$

Using the same mapping g as in the first part (a) of this proof, it holds $\hat{K}_1(\mathbf{F},\mathbf{a}) = g(\mathbf{Cof}\,\mathbf{F})$ and therefore the polyconvexity of $\hat{K}_1(\mathbf{F},\mathbf{a})$.

(c) For the polyconvexity of $\hat{K}_2(\mathbf{F},\mathbf{a}) = \hat{I}_1(\mathbf{F}) - \hat{J}_4(\mathbf{F},\mathbf{a}) = |\mathbf{F}|^2 - |\mathbf{F}\mathbf{a}|^2$ we follow the same idea as before and consider the mapping $g(\mathbf{A}) := |\mathbf{A}|^2 - |\mathbf{A}\mathbf{a}|^2 = \mathrm{tr}(\mathbf{A}^T\mathbf{A}) - \mathrm{tr}(\mathbf{A}^T\mathbf{A}\mathbf{M})$, $\mathbf{A} \in \mathbb{R}^{3\times 3}$. We have to show that this mapping is convex on $\mathbb{R}^{3\times 3}$.
Analogously as above it hold

$$\begin{aligned} g'(\mathbf{A})[\mathbf{H}_1] &= 2\mathbf{A} : \mathbf{H}_1 - [2\mathbf{A}\mathbf{M}] : \mathbf{H}_1, \\ g''(\mathbf{A})[\mathbf{H}_1,\mathbf{H}_2] &= 2\mathbf{H}_2 : \mathbf{H}_1 - [2\mathbf{H}_2\mathbf{M}] : \mathbf{H}_1. \end{aligned}$$

Thus we obtain for arbitrary $\mathbf{A}, \mathbf{H} \in \mathbb{R}^{3\times 3}$

$$g''(\mathbf{A})[\mathbf{H},\mathbf{H}] = 2|\mathbf{H}|^2 - 2|\mathbf{Ha}|^2 = 2\left(|\mathbf{H}|^2 - |\mathbf{Ha}|^2\right) \geq 0,$$

since the Frobenius norm is consistent with the Euclidean norm $|\cdot|$ and therefore $|\mathbf{Ha}| \leq |\mathbf{H}||\mathbf{a}| = |\mathbf{H}| \Leftrightarrow |\mathbf{H}|^2 - |\mathbf{Ha}|^2 \geq 0$. Here we have used the assumption $|\mathbf{a}| = 1$. By Proposition 2.27 the convexity of g on $\{\mathbf{A} \in \mathbb{R}^{3\times 3}\}$ follows and therefore $\hat{K}_2(\mathbf{F},\mathbf{a}) = g(\mathbf{F})$ is polyconvex.

(d) For the polyconvexity of $\hat{K}_3(\mathbf{F},\mathbf{a})$ it holds with the help of the derivations in (b) and Definition 4.1 for arbitrary $\mathbf{F} \in \mathbb{M}$ and normed $\mathbf{a} \in \mathbb{R}^3$ the equation

$$\hat{K}_3(\mathbf{F},\mathbf{a}) = \hat{I}_2(\mathbf{F}) - \hat{K}_1(\mathbf{F},\mathbf{a}) = |\mathbf{Cof}\,\mathbf{F}|^2 - |(\mathbf{Cof}\,\mathbf{F})\mathbf{a}|^2.$$

Thus we can use the same mapping g as in the proof of the polyconvexity of $\hat{K}_2(\mathbf{F},\mathbf{a})$ in part (c) of this proof. Due to the convexity of g on $\mathbb{R}^{3\times 3}$ and the relation $\hat{K}_3(\mathbf{F},\mathbf{a}) = g(\mathbf{Cof}\,\mathbf{F})$, $\hat{K}_3(\mathbf{F},\mathbf{a})$ is polyconvex.

□

4.2 Application to transverse isotropy

In the case of a transversely isotropic hyperelastic material we consider one preferred direction, described by a column vector $\mathbf{a} \in \mathbb{R}^3$. Without loss of generality we assume that $\mathbf{a}$ is normed. By [Sch10] the material symmetry group in this case is defined by $\mathbb{G} := \{\pm\mathbf{I}, \mathbf{Q}(\phi,\mathbf{a})\}$, where $\mathbf{Q}(\phi,\mathbf{a})$ denotes a rotation about the $\mathbf{a}$-axis with arbitrary angle $0 < \phi < 2\pi$. Thus it holds $\mathbf{Qa} = \mathbf{a}$ and thus for $\mathbf{M} := \mathbf{a}\cdot\mathbf{a}^T$ we obtain $\mathbf{QMQ}^T = \mathbf{Qaa}^T\mathbf{Q}^T = \mathbf{Qa}(\mathbf{Qa})^T = \mathbf{aa}^T = \mathbf{M}$. The principal and mixed invariants, which will be ingredients for the stored energy function (4.4), do not depend explicitly on $\mathbf{x} \in \bar{\Omega}$. Thus, for simplicity, we assume in the following that (4.4) is completely homogeneous.
Before we define a concrete suitable stored energy function we introduce the so-called Macaulay brackets:

Definition 4.5: (Macaulay brackets)
Let f be a real-valued function. Then we define the **Macaulay brackets** pointwise as

$$\langle f\rangle := \frac{1}{2}\left(f + |f|\right) = \max\{f, 0\} = \begin{cases} f & , f \geq 0 \\ 0 & , f < 0. \end{cases}$$

Thus Definition 4.5 is nothing else than the positive part of a real-valued function f. However, this notation is often used in engineering and mechanics and we will stick to this notation in the following.

An immediate consequence for the function $f(x) = x$, $x \in \mathbb{R}$, $m \in \mathbb{R}_{\geq 0}$, is

$$\langle x \rangle^m = \begin{cases} x^m & , x \geq 0 \\ 0 & , x < 0, \end{cases}$$

which is continuous for positive values m and, due to $\lim_{x \to 0^+} \langle x \rangle^0 = \lim_{x \to 0^+} x^0 = 1$, discontinuous for $m = 0$.
Obviously for given $n \in \mathbb{N}$, $m \geq n$ and $i \in \{1, \dots, n\}$ it holds

$$\frac{d^i}{dx^i} \langle x \rangle^m = \begin{cases} \prod_{k=0}^{i-1} (m-k) x^{m-i} & , x \geq 0 \\ 0 & , x < 0 \end{cases} = \prod_{k=0}^{i-1} (m-k) \langle x \rangle^{m-i}. \tag{4.11}$$

Consequently for $m > n$ it holds $i \leq n < m$ or equivalently $m - i > 0$ for all $i \in \{1, \dots, n\}$ and in particular this leads to $\langle x \rangle^m \in C^n(\mathbb{R}, \mathbb{R}_{\geq 0})$ for $i \in \{1, \dots, n\}$.
For the isotropic part $\hat{\psi}_{iso}$ in the general stored energy function (4.4) we use the Mooney-Rivlin model, described by (2.30), i.e.

$$\hat{\psi}_{iso}(\mathbf{F}) := \hat{\psi}_{MR}(\mathbf{F}) = \alpha \left|\mathbf{F}\right|^2 + \beta (\det \mathbf{F})^2 - (2\alpha + 2\beta + 4\delta) \ln(\det \mathbf{F}) + \delta \left|\mathbf{Cof}\, \mathbf{F}\right|^2, \quad \mathbf{F} \in \mathbb{M}, \tag{4.12}$$

with $\alpha, \beta > 0$ and $\delta \geq 0$. Recall that for this choice consistency with linear elasticity is ensured if $\alpha = \frac{\mu}{2} - \delta$, $\beta = \frac{\lambda}{4} - \delta$ with $0 \leq \delta < \min\left\{\frac{\lambda}{4}, \frac{\mu}{2}\right\}$ (cf. Section 2.4.5).
For the anisotropic part $\hat{\psi}_{aniso}$ in (4.4) we use

$$\hat{\psi}_{aniso}(\mathbf{F}, \mathbf{a}) := \varepsilon_1 \hat{\psi}^{(1)}_{aniso}(\mathbf{F}, \mathbf{a}) + \varepsilon_2 \hat{\psi}^{(2)}_{aniso}(\mathbf{F}, \mathbf{a}) + \varepsilon_3 \hat{\psi}^{(3)}_{aniso}(\mathbf{F}, \mathbf{a}) = \sum_{i=1}^{3} \varepsilon_i \hat{\psi}^{(i)}_{aniso}(\mathbf{F}, \mathbf{a}) \tag{4.13}$$

with parameters $\varepsilon_1, \varepsilon_2, \varepsilon_3 \geq 0$,

$$\begin{aligned}
\hat{\psi}^{(1)}_{aniso}(\mathbf{F}, \mathbf{a}) &:= \langle \hat{J}_4(\mathbf{F}, \mathbf{a}) - 1 \rangle^2, \\
\hat{\psi}^{(2)}_{aniso}(\mathbf{F}, \mathbf{a}) &:= \frac{1}{a_1} \hat{J}_4(\mathbf{F}, \mathbf{a})^{a_1} + \frac{1}{a_2} \hat{K}_1(\mathbf{F}, \mathbf{a})^{a_2} + \frac{1}{a_3} \hat{I}_3(\mathbf{F})^{-a_3}, \\
\hat{\psi}^{(3)}_{aniso}(\mathbf{F}, \mathbf{a}) &:= \frac{1}{b_1} \left(\frac{1}{2} \hat{K}_2(\mathbf{F}, \mathbf{a}) \right)^{b_1} + \frac{1}{b_2} \left(\frac{1}{2} \hat{K}_3(\mathbf{F}, \mathbf{a}) \right)^{b_2} + \frac{1}{b_3} \hat{I}_3(\mathbf{F})^{-b_3},
\end{aligned}$$

containing real parameters $a_1, a_2, b_1, b_2 \geq 1$ and nonzero $a_3, b_3 \geq -\frac{1}{2}$ (cf. [Sch10] and [BSN10]). We denote the whole stored energy function for this choice in the following as

$$\hat{\psi}_{ti}(\mathbf{F}, \mathbf{a}) := \hat{\psi}_{iso}(\mathbf{F}) + \hat{\psi}_{aniso}(\mathbf{F}, \mathbf{a}). \tag{4.14}$$

Since $\hat{I}_j(\mathbf{Q}\mathbf{F}\mathbf{Q}^T) = \hat{I}_j(\mathbf{F})$ $(j = 1, 2, 3)$, $\hat{J}_j(\mathbf{Q}\mathbf{F}\mathbf{Q}^T, \mathbf{Q}\mathbf{a}) = \hat{J}_j(\mathbf{F}, \mathbf{a})$ for $j = 4, 5$ and all $\mathbf{Q} \in \mathbb{O}$, it holds by Definition 4.1 automatically $\hat{K}_j(\mathbf{Q}\mathbf{F}\mathbf{Q}^T, \mathbf{Q}\mathbf{a}) = \hat{K}_j(\mathbf{F}, \mathbf{a})$ for $j = 1, 2, 3$ and all $\mathbf{Q} \in \mathbb{O}$. Therefore condition (4.5) is automatically satisfied for (4.14).

Polyconvexity of the stored energy function

We have seen in Section 3.6.1 that polyconvexity is an important tool for existence theory of minimizers. Since we will not exclude the possibility a - priori to use this theory for anisotropic materials, we demand that our used stored energy function (4.4) is polyconvex. For our special choice (4.14) we have already proven polyconvexity of $\hat{\psi}_{iso}(\mathbf{F})$ in Section 2.4.4. It remains to show that $\hat{\psi}_{aniso}(\mathbf{F}, \mathbf{a})$ is polyconvex, since then the total stored energy function (4.14) is obviously polyconvex. Furthermore, since $\hat{\psi}_{aniso}(\mathbf{F}, \mathbf{a})$ is a linear combination of $\hat{\psi}^{(i)}_{aniso}(\mathbf{F}, \mathbf{a})$, $i = 1, 2, 3$, and ε_i, $i = 1, 2, 3$, are nonnegative, it is sufficient to show that each $\hat{\psi}^{(i)}_{aniso}(\mathbf{F}, \mathbf{a})$ is polyconvex. For the proof of polyconvexity of these terms we start with the following considerations:
For $f(x) := \frac{1}{a}\langle x\rangle^m$ with real parameters $a > 0, m \geq 2$ and arbitrary twice Gâteaux - differentiable function $g : \mathbb{R}^{3\times 3} \to \mathbb{R}$ we set $h(\mathbf{A}) := f(g(\mathbf{A})) = \frac{1}{a}\langle g(\mathbf{A})\rangle^m$, $\mathbf{A} \in \mathbb{R}^{3\times 3}$. Obviously it holds by (4.11)

$$f'(x)[y] = \frac{d}{dt}\frac{1}{a}\langle x + ty\rangle^m\big|_{t=0} = \frac{m}{a}\langle x\rangle^{m-1}\, y.$$

Therefore we obtain the Gâteaux derivatives

$$h'(\mathbf{A})[\mathbf{H}_1] = f'(g(\mathbf{A}))[g'(\mathbf{A})[\mathbf{H}_1]] = \frac{m}{a}\langle g(\mathbf{A})\rangle^{m-1} g'(\mathbf{A})[\mathbf{H}_1],$$
$$h''(\mathbf{A})[\mathbf{H}_1, \mathbf{H}_2] = \frac{m}{a}(m-1)\langle g(\mathbf{A})\rangle^{m-2} g'(\mathbf{A})[\mathbf{H}_2] g'(\mathbf{A})[\mathbf{H}_1] + \frac{m}{a}\langle g(\mathbf{A})\rangle^{m-1} g''(\mathbf{A})[\mathbf{H}_1, \mathbf{H}_2]$$

and in particular

$$h''(\mathbf{A})[\mathbf{H}, \mathbf{H}] = \frac{m}{a}(m-1)\langle g(\mathbf{A})\rangle^{m-2}\left(g'(\mathbf{A})[\mathbf{H}]\right)^2 + \frac{m}{a}\langle g(\mathbf{A})\rangle^{m-1} g''(\mathbf{A})[\mathbf{H}, \mathbf{H}]. \quad (4.15)$$

The Macaulay bracket in the first term of (4.15) is defined for $m \geq 2$ and the whole term is nonnegative in this case. Hence by Proposition 2.27 h is convex on $\mathbb{R}^{3\times 3}$ if and only if $g''(\mathbf{H}_1)[\mathbf{H}_2 - \mathbf{H}_1, \mathbf{H}_2 - \mathbf{H}_1] \geq 0$ for arbitrary $\mathbf{H}_1, \mathbf{H}_2 \in \mathbb{R}^{3\times 3}$. For the function $g_1(\mathbf{A}) := g(\mathbf{A}) = |\mathbf{A}\mathbf{a}|^2 - 1$ this is obviously true (cf. proof of Lemma 4.4). With this choice we have

$$\hat{\psi}^{(1)}_{aniso}(\mathbf{F}, \mathbf{a}) := \langle \hat{J}_4(\mathbf{F}, \mathbf{a}) - 1\rangle^2 = \langle g_1(\mathbf{F})\rangle^2 =: h_1(\mathbf{F})$$

for $m = 2$, $a = 1$ and a convex function h_1, i.e. by Definition 2.26 $\hat{\psi}^{(1)}_{aniso}(\mathbf{F}, \mathbf{a})$ is polyconvex. If we change the function f to $f(x) := \frac{1}{a}x^m$ with real parameters $a \neq 0$ and m, provided that $x \geq 0$, we can follow the same steps and obtain for mappings $g : \mathbb{R}^{3\times 3} \to \mathbb{R}_{\geq 0}$ and $h(\mathbf{A}) := f(g(\mathbf{A})) = \frac{1}{a}g(\mathbf{A})^m$ the condition

$$h''(\mathbf{A})[\mathbf{H}, \mathbf{H}] = \frac{m}{a}(m-1)g(\mathbf{A})^{m-2}\left(g'(\mathbf{A})[\mathbf{H}]\right)^2 + \frac{m}{a}g(\mathbf{A})^{m-1} g''(\mathbf{A})[\mathbf{H}, \mathbf{H}]. \quad (4.16)$$

(4.16) is at least nonnegative if $g''(\mathbf{A})[\mathbf{H}, \mathbf{H}] \geq 0$ for arbitrary $\mathbf{A}, \mathbf{H} \in \mathbb{R}^{3\times 3}$, $\frac{m}{a} \geq 0$ and $\frac{m}{a}(m-1) \geq 0$. Obviously $\frac{m}{a}$ is nonnegative if and only if m and a have the same algebraic

sign or $m = 0$. $\frac{m}{a}(m-1)$ is nonnegative for the sets $\{(m,a) \in \mathbb{R}^2 : m \geq 1 \wedge a > 0\}$, $\{(m,a) \in \mathbb{R}^2 : m \leq 0 \wedge a > 0\}$ and $\{(m,a) \in \mathbb{R}^2 : m \in [0,1] \wedge a < 0\}$. This means that (4.16) is definitely nonnegative for $\{(m,a) \in \mathbb{R}^2 : m \geq 1 \wedge a > 0\}$ under the assumption that $g''(\mathbf{A})[\mathbf{H},\mathbf{H}] \geq 0$ for all $\mathbf{A},\mathbf{H} \in \mathbb{R}^{3\times 3}$. In particular this is valid for $m = a \geq 1$.
Moreover, the function $f(x) = \frac{1}{a}x^m$, $x > 0$, $m \in \mathbb{R}$, $0 \neq a \in \mathbb{R}$, is convex on $\mathbb{R}_{>0}$ if and only $f''(x) = \frac{m}{a}(m-1)x^{m-2} \geq 0$ for all $x > 0$. This is again ensured for pairs $(m,a) \in \mathbb{R}^2$, $a \neq 0$, in the union of the three mentioned sets above. In particular, if we choose $m = -2n$ and $a = n$, f is convex on $\mathbb{R}_{>0}$ if and only if $n \geq -\frac{1}{2}$.
We define $g_2(\mathbf{A}) := |\mathbf{A}\mathbf{a}|^2$, $g_3(\mathbf{A}) := |\mathbf{A}|^2 - |\mathbf{A}\mathbf{a}|^2$ for arbitrary $\mathbf{A} \in \mathbb{R}^{3\times 3}$ and $g_4(x,n) := \frac{1}{n}x^{-2n}$ on $(0,\infty)$ with fixed $n \geq -\frac{1}{2}$, $n \neq 0$. Obviously it holds $g_i(\mathbf{A}) \geq 0$ and $g_i''(\mathbf{A})[\mathbf{H},\mathbf{H}] \geq 0$, $i = 2,3$, for all $\mathbf{A},\mathbf{H} \in \mathbb{R}^{3\times 3}$ (cf. proof of Lemma 4.4). With this choice and the considerations above we obtain for $\mathbf{F} \in \mathbb{M}$

$$\begin{aligned}
\hat{\psi}^{(2)}_{aniso}(\mathbf{F},\mathbf{a}) &= \frac{1}{a_1}\hat{J}_4(\mathbf{F},\mathbf{a})^{a_1} + \frac{1}{a_2}\hat{K}_1(\mathbf{F},\mathbf{a})^{a_2} + \frac{1}{a_3}\hat{I}_3(\mathbf{F})^{-a_3} \\
&= \frac{1}{a_1}g_2(\mathbf{F})^{a_1} + \frac{1}{a_2}g_2(\mathbf{Cof}\,\mathbf{F})^{a_2} + g_4(\det\mathbf{F}, a_3), \\
\hat{\psi}^{(3)}_{aniso}(\mathbf{F},\mathbf{a}) &= \frac{1}{b_1}\left(\frac{1}{2}\hat{K}_2(\mathbf{F},\mathbf{a})\right)^{b_1} + \frac{1}{b_2}\left(\frac{1}{2}\hat{K}_3(\mathbf{F},\mathbf{a})\right)^{b_2} + \frac{1}{b_3}\hat{I}_3(\mathbf{F})^{-b_3} \\
&= \frac{1}{b_1}\left(\frac{1}{2}g_3(\mathbf{F})\right)^{b_1} + \frac{1}{b_2}\left(\frac{1}{2}g_3(\mathbf{Cof}\,\mathbf{F})\right)^{b_2} + g_4(\det\mathbf{F}, b_3)
\end{aligned}$$

with convex right-hand sides on the set $\{(\mathbf{F},\mathbf{Cof}\,\mathbf{F},\det\mathbf{F}) \in \mathbb{R}^{3\times 3}\times\mathbb{R}^{3\times 3}\times(0,\infty)\}$. Therefore $\hat{\psi}^{(2)}_{aniso}(\mathbf{F},\mathbf{a})$ and $\hat{\psi}^{(3)}_{aniso}(\mathbf{F},\mathbf{a})$ are by Definition 2.26 polyconvex. Altogether polyconvexity of (4.14) is proven.

4.3 Consistency with the linear model of transverse isotropic materials

Remark 4.3 has motivated us to use the proposed $\mathbf{C}$-formulation of Section 3.3 for our least squares finite element approach in the case of anisotropic materials. In this subsection we formulate the corresponding mapping $\tilde{\mathcal{G}}_{ti}(\mathbf{C}) := \mathbf{F}^{-1}\partial_{\mathbf{F}}\hat{\psi}_{ti}(\mathbf{F},\mathbf{a})$ to the stored energy function (4.14). For our least squares formulation we also need the Gâteaux derivative of $\tilde{\mathcal{G}}_{ti}(\mathbf{C})$. $\tilde{\mathcal{G}}_{ti}(\mathbf{C})$ and $\tilde{\mathcal{G}}'_{ti}(\mathbf{C})[\mathbf{E}]$ will be derived firstly. With these expressions we can then ensure consistency of the nonlinear model with a linear model for transverse isotropy following the same steps as in Section 2.4.5. For consistency we have to ensure $\tilde{\mathcal{G}}_{ti}(\mathbf{I}) = \mathbf{0}$ to get a stress-free reference configuration. Moreover, we have to guarantee $2\tilde{\mathcal{G}}'_{ti}(\mathbf{I})[\mathbf{E}] = \mathcal{C}_{ti}\mathbf{E}$ for all $\mathbf{E} \in \mathbb{R}^{3\times 3}$ (cp. the introduction of Section 3.3), where $\mathcal{C}_{ti}$ is now a symmetric fourth-order tensor describing the stress-strain relation of form $\boldsymbol{\sigma} = \mathbf{C}_{ti}\boldsymbol{\varepsilon}$ in a linear model for transverse isotropy. Thus $\mathcal{C}_{ti}$ replaces the elasticity tensor $\mathcal{C}$ in the stress-strain relation of linear elasticity (cf. (2.10)). With these requirements the linearized nonlinear model, more precisely the linearization of $\tilde{\mathcal{G}}_{ti}(\mathbf{C})$ about $\mathbf{C} = \mathbf{I}$, then coincides again with the linear model up to a constant $\frac{1}{2}$. The required consistency with the linear model again

influences the choice of material parameters. Our aim is to determine the set of material parameters $(\alpha, \beta, \varepsilon_1, \varepsilon_2, \varepsilon_3)$ for arbitrary but fixed parameters $(\delta, a_1, a_2, a_3, b_1, b_2, b_3)$ such that consistency with a linear behavior is guaranteed. Recall that the choice $\alpha, \beta > 0$, $\varepsilon_1, \varepsilon_2, \varepsilon_3 \geq 0$, $\delta \geq 0$, $a_1, a_2, b_1, b_2 \geq 1$ and nonzero $a_3, b_3 \geq -\frac{1}{2}$ is necessary by the considerations in Sections 4.1 and 4.2.
We start with the calculation of $\tilde{\mathcal{G}}_{ti}(\mathbf{C}) = \mathbf{F}^{-1}\partial_{\mathbf{F}}\hat{\psi}_{ti}(\mathbf{F}, \mathbf{a})$ and its Gâteaux derivative with respect to $\mathbf{C}$ using Lemma 4.2 and (4.7):
We decompose $\tilde{\mathcal{G}}_{ti}(\mathbf{C})$ into

$$\tilde{\mathcal{G}}_{ti}(\mathbf{C}) = \tilde{\mathcal{G}}_{iso}(\mathbf{C}) + \tilde{\mathcal{G}}_{aniso}(\mathbf{C}) = \tilde{\mathcal{G}}_{iso}(\mathbf{C}) + \varepsilon_1\tilde{\mathcal{G}}^{(1)}_{aniso}(\mathbf{C}) + \varepsilon_2\tilde{\mathcal{G}}^{(2)}_{aniso}(\mathbf{C}) + \varepsilon_3\tilde{\mathcal{G}}^{(3)}_{aniso}(\mathbf{C}) \quad (4.17)$$

with

$$\begin{aligned}
\tilde{\mathcal{G}}_{iso}(\mathbf{C}) &:= \mathbf{F}^{-1}\partial_{\mathbf{F}}\hat{\psi}_{iso}(\mathbf{F}) = 2\alpha\mathbf{I} + 2\left(\beta(\det\mathbf{C}) - (\alpha+\beta+2\delta)\right)\mathbf{C}^{-1} + 2\delta(\mathrm{tr}(\mathbf{C})\mathbf{I} - \mathbf{C}),\\
\tilde{\mathcal{G}}^{(1)}_{aniso}(\mathbf{C}) &:= \mathbf{F}^{-1}\partial_{\mathbf{F}}\hat{\psi}^{(1)}_{aniso}(\mathbf{F}, \mathbf{a}) = 4\langle J_4(\mathbf{C}, \mathbf{M}) - 1\rangle\mathbf{M},\\
\tilde{\mathcal{G}}^{(2)}_{aniso}(\mathbf{C}) &:= \mathbf{F}^{-1}\partial_{\mathbf{F}}\hat{\psi}^{(2)}_{aniso}(\mathbf{F}, \mathbf{a}) = \mathbf{F}^{-1}\partial_{\mathbf{F}}\left(\frac{1}{a_1}\hat{J}_4(\mathbf{F}, \mathbf{a})^{a_1} + \frac{1}{a_2}\hat{K}_1(\mathbf{F}, \mathbf{a})^{a_2} + \frac{1}{a_3}\hat{I}_3(\mathbf{F})^{-a_3}\right)\\
&= \mathbf{F}^{-1}\big(\hat{J}_4(\mathbf{F}, \mathbf{a})^{a_1-1}\partial_{\mathbf{F}}\hat{J}_4(\mathbf{F}, \mathbf{a}) + \hat{K}_1(\mathbf{F}, \mathbf{a})^{a_2-1}\partial_{\mathbf{F}}\hat{K}_1(\mathbf{F}, \mathbf{a})\\
&\qquad - \hat{I}_3(\mathbf{F})^{-a_3-1}\partial_{\mathbf{F}}\hat{I}_3(\mathbf{F})\big)\\
&= 2\mathbf{F}^{-1}\left(J_4(\mathbf{C}, \mathbf{M})^{a_1-1}\mathbf{F}\mathbf{M} + K_1(\mathbf{C}, \mathbf{M})^{a_2-1}\mathbf{F}\left[\mathbf{C}\mathbf{M} + \mathbf{M}\mathbf{C} - \mathrm{tr}(\mathbf{C}\mathbf{M})\mathbf{I}\right.\right.\\
&\qquad \left.\left. -\mathrm{tr}(\mathbf{C})\mathbf{M} + \mathrm{tr}(\mathbf{C})\mathbf{I} - \mathbf{C}\right] - I_3(\mathbf{C})^{-a_3}\mathbf{F}^{-T}\right)\\
&= 2\left(J_4(\mathbf{C}, \mathbf{M})^{a_1-1}\mathbf{M} + K_1(\mathbf{C}, \mathbf{M})^{a_2-1}\left[\mathbf{C}\mathbf{M} + \mathbf{M}\mathbf{C} - \mathrm{tr}(\mathbf{C}\mathbf{M})\mathbf{I}\right.\right.\\
&\qquad \left.\left. -\mathrm{tr}(\mathbf{C})\mathbf{M} + \mathrm{tr}(\mathbf{C})\mathbf{I} - \mathbf{C}\right] - I_3(\mathbf{C})^{-a_3}\mathbf{C}^{-1}\right),\\
\tilde{\mathcal{G}}^{(3)}_{aniso}(\mathbf{C}) &:= \mathbf{F}^{-1}\partial_{\mathbf{F}}\left(\frac{1}{b_1}\left(\frac{1}{2}\hat{K}_2(\mathbf{F}, \mathbf{a})\right)^{b_1} + \frac{1}{b_2}\left(\frac{1}{2}\hat{K}_3(\mathbf{F}, \mathbf{a})\right)^{b_2} + \frac{1}{b_3}\hat{I}_3(\mathbf{F})^{-b_3}\right)\\
&= \mathbf{F}^{-1}\left(\frac{1}{2}\left(\frac{1}{2}\hat{K}_2(\mathbf{F}, \mathbf{a})\right)^{b_1-1}\partial_{\mathbf{F}}\hat{K}_2(\mathbf{F}, \mathbf{a}) + \frac{1}{2}\left(\frac{1}{2}\hat{K}_3(\mathbf{F}, \mathbf{a})\right)^{b_2-1}\partial_{\mathbf{F}}\hat{K}_3(\mathbf{F}, \mathbf{a})\right.\\
&\qquad \left. - \hat{I}_3(\mathbf{F})^{-b_3-1}\partial_{\mathbf{F}}\hat{I}_3(\mathbf{F})\right)\\
&= \mathbf{F}^{-1}\left(\left(\frac{1}{2}K_2(\mathbf{C}, \mathbf{M})\right)^{b_1-1}\mathbf{F}(\mathbf{I} - \mathbf{M}) + \left(\frac{1}{2}K_3(\mathbf{C}, \mathbf{M})\right)^{b_2-1}\mathbf{F}\right.\\
&\qquad \left. \cdot\left[\mathrm{tr}(\mathbf{C}\mathbf{M})\mathbf{I} + \mathrm{tr}(\mathbf{C})\mathbf{M} - (\mathbf{C}\mathbf{M} + \mathbf{M}\mathbf{C})\right] - 2I_3(\mathbf{C})^{-b_3}\mathbf{F}^{-T}\right)\\
&= \left(\frac{1}{2}K_2(\mathbf{C}, \mathbf{M})\right)^{b_1-1}(\mathbf{I} - \mathbf{M}) + \left(\frac{1}{2}K_3(\mathbf{C}, \mathbf{M})\right)^{b_2-1}\\
&\qquad \cdot\left[\mathrm{tr}(\mathbf{C}\mathbf{M})\mathbf{I} + \mathrm{tr}(\mathbf{C})\mathbf{M} - (\mathbf{C}\mathbf{M} + \mathbf{M}\mathbf{C})\right] - 2I_3(\mathbf{C})^{-b_3}\mathbf{C}^{-1}.
\end{aligned}$$

Due to $I_1(\mathbf{I}) = 3$, $I_2(\mathbf{I}) = 3$, $I_3(\mathbf{I}) = 1$, $J_4(\mathbf{I},\mathbf{M}) = 1$, $J_5(\mathbf{I},\mathbf{M}) = 1$, $K_1(\mathbf{I},\mathbf{M}) = 1$, $K_2(\mathbf{I},\mathbf{M}) = 2$ and $K_3(\mathbf{I},\mathbf{M}) = 2$ we obtain for $\mathbf{C} = \mathbf{I}$

$$\begin{aligned}
\tilde{\mathcal{G}}_{iso}(\mathbf{I}) &= 2\alpha\mathbf{I} + 2\left(\beta - (\alpha+\beta+2\delta)\right)\mathbf{I} + 2\delta(2\mathbf{I}) = \mathbf{0},\\
\tilde{\mathcal{G}}^{(1)}_{aniso}(\mathbf{I}) &= 4\langle 1-1\rangle\mathbf{M} = \mathbf{0},\\
\tilde{\mathcal{G}}^{(2)}_{aniso}(\mathbf{I}) &= 2\left(1^{a_1-1}\mathbf{M} + 1^{a_2-1}\left[\mathbf{M}+\mathbf{M}-\mathrm{tr}(\mathbf{M})\mathbf{I}-3\mathbf{M}+3\mathbf{I}-\mathbf{I}\right] - 1^{-a_3}\mathbf{I}\right)\\
&= 2\left(\mathbf{M} + [\mathbf{I}-\mathbf{M}] - \mathbf{I}\right) = \mathbf{0},\\
\tilde{\mathcal{G}}^{(3)}_{aniso}(\mathbf{I}) &= \left(\frac{1}{2}\cdot 2\right)^{b_1-1}(\mathbf{I}-\mathbf{M}) + \left(\frac{1}{2}\cdot 2\right)^{b_2-1}\left[\mathrm{tr}(\mathbf{M})\mathbf{I} + 3\mathbf{M} - (\mathbf{M}+\mathbf{M})\right] - 2\cdot 1^{-b_3}\mathbf{I}\\
&= (\mathbf{I}-\mathbf{M}) + [\mathbf{I}+\mathbf{M}] - 2\mathbf{I} = \mathbf{0},
\end{aligned}$$

i.e. with the help of (4.17) it holds $\tilde{\mathcal{G}}_{ti}(\mathbf{I}) = \mathbf{0}$. Thus for this choice of stored energy function the reference configuration is automatically stress - free.
Due to the linearity of the Gâteaux derivative and (4.17) we get

$$\tilde{\mathcal{G}}'_{ti}(\mathbf{C})[\mathbf{E}] = \tilde{\mathcal{G}}'_{iso}(\mathbf{C})[\mathbf{E}] + \sum_{i=1}^{3}\varepsilon_i\left(\tilde{\mathcal{G}}^{(i)}_{aniso}(\mathbf{C})\right)'[\mathbf{E}] \tag{4.18}$$

for $\mathbf{C},\mathbf{E} \in \mathbb{R}^{3\times 3}$. For the calculation of the Gâteaux derivatives with respect to $\mathbf{C}$ in the single terms we use Lemma 4.2 and (4.6) to obtain

$$\begin{aligned}
\tilde{\mathcal{G}}'_{iso}(\mathbf{C})[\mathbf{E}] &= 2\big[\beta(\mathbf{Cof}\,\mathbf{C} : \mathbf{E})\mathbf{C}^{-1} - (\beta(\det\mathbf{C}) - (\alpha+\beta+2\delta))\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}\\
&\qquad + \delta(\mathrm{tr}(\mathbf{E})\mathbf{I} - \mathbf{E})\big],\\
\left(\tilde{\mathcal{G}}^{(1)}_{aniso}(\mathbf{C})\right)'[\mathbf{E}] &= 4\langle J_4(\mathbf{C},\mathbf{M}) - 1\rangle^0 J'_4(\mathbf{C},\mathbf{M})[\mathbf{E}]\mathbf{M} = \begin{cases} 4\mathrm{tr}(\mathbf{M}\mathbf{E})\mathbf{M}, & J_4(\mathbf{C},\mathbf{M}) \geq 1\\ \mathbf{0}, & J_4(\mathbf{C},\mathbf{M}) < 1\end{cases}\\
\left(\tilde{\mathcal{G}}^{(2)}_{aniso}(\mathbf{C})\right)'[\mathbf{E}] &= 2\Big((a_1-1)J_4(\mathbf{C},\mathbf{M})^{a_1-2}J'_4(\mathbf{C},\mathbf{M})[\mathbf{E}]\mathbf{M} + (a_2-1)K_1(\mathbf{C},\mathbf{M})^{a_2-2}\\
&\qquad K'_1(\mathbf{C},\mathbf{M})[\mathbf{E}]\big\{\mathbf{C}\mathbf{M}+\mathbf{M}\mathbf{C} - \mathrm{tr}(\mathbf{C}\mathbf{M})\mathbf{I} - \mathrm{tr}(\mathbf{C})\mathbf{M} + \mathrm{tr}(\mathbf{C})\mathbf{I} - \mathbf{C}\big\}\\
&\qquad + K_1(\mathbf{C},\mathbf{M})^{a_2-1}\big\{\mathbf{E}\mathbf{M}+\mathbf{M}\mathbf{E} - \mathrm{tr}(\mathbf{E}\mathbf{M})\mathbf{I} - \mathrm{tr}(\mathbf{E})\mathbf{M} + \mathrm{tr}(\mathbf{E})\mathbf{I} - \mathbf{E}\big\}\\
&\qquad + a_3 I_3(\mathbf{C})^{-a_3-1}(\mathbf{Cof}\,\mathbf{C} : \mathbf{E})\mathbf{C}^{-1} + I_3(\mathbf{C})^{-a_3}\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}\Big),\\
\left(\tilde{\mathcal{G}}^{(3)}_{aniso}(\mathbf{C})\right)'[\mathbf{E}] &= \frac{1}{2}(b_1-1)\left(\frac{1}{2}K_2(\mathbf{C},\mathbf{M})\right)^{b_1-2}K'_2(\mathbf{C},\mathbf{M})[\mathbf{E}](\mathbf{I}-\mathbf{M})\\
&\qquad + \frac{1}{2}(b_2-1)\left(\frac{1}{2}K_3(\mathbf{C},\mathbf{M})\right)^{b_2-2}K'_3(\mathbf{C},\mathbf{M})[\mathbf{E}]\big\{\mathrm{tr}(\mathbf{C}\mathbf{M})\mathbf{I} + \mathrm{tr}(\mathbf{C})\mathbf{M}\\
&\qquad - (\mathbf{C}\mathbf{M}+\mathbf{M}\mathbf{C})\big\} + \left(\frac{1}{2}K_3(\mathbf{C},\mathbf{M})\right)^{b_2-1}\big\{\mathrm{tr}(\mathbf{E}\mathbf{M})\mathbf{I} + \mathrm{tr}(\mathbf{E})\mathbf{M}\\
&\qquad - (\mathbf{E}\mathbf{M}+\mathbf{M}\mathbf{E})\big\} + 2b_3 I_3(\mathbf{C})^{-b_3-1}(\mathbf{Cof}\,\mathbf{C} : \mathbf{E})\mathbf{C}^{-1}\\
&\qquad + 2I_3(\mathbf{C})^{-b_3}\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}.
\end{aligned}$$

If we insert $\mathbf{C} = \mathbf{I}$ we obtain with the help of (4.6), (4.9) and some elementary simplifications the Gâteaux derivatives

$$\begin{aligned}
\tilde{\mathcal{G}}'_{iso}(\mathbf{I})[\mathbf{E}] &= 2(\alpha+\delta)\mathbf{E} + 2(\beta+\delta)\mathrm{tr}(\mathbf{E})\mathbf{I},\\
\Big(\tilde{\mathcal{G}}^{(1)}_{aniso}(\mathbf{I})\Big)'[\mathbf{E}] &= 4\mathrm{tr}(\mathbf{ME})\mathbf{M},\\
\Big(\tilde{\mathcal{G}}^{(2)}_{aniso}(\mathbf{I})\Big)'[\mathbf{E}] &= 2(a_2+a_3)\mathrm{tr}(\mathbf{E})\mathbf{I} + 2(a_1+a_2-2)\mathrm{tr}(\mathbf{ME})\mathbf{M}\\
&\quad - 2a_2\big(\mathrm{tr}(\mathbf{ME})\mathbf{I} + \mathrm{tr}(\mathbf{E})\mathbf{M}\big) + 2(\mathbf{EM}+\mathbf{ME}),\\
\Big(\tilde{\mathcal{G}}^{(3)}_{aniso}(\mathbf{I})\Big)'[\mathbf{E}] &= 2\mathbf{E} + \left(\frac{1}{2}(b_1+b_2)+2b_3-1\right)\mathrm{tr}(\mathbf{E})\mathbf{I} + \left(\frac{1}{2}(b_1+b_2)-1\right)\mathrm{tr}(\mathbf{ME})\mathbf{M}\\
&\quad + \left(\frac{1}{2}(b_2-b_1)+1\right)\big(\mathrm{tr}(\mathbf{ME})\mathbf{I} + \mathrm{tr}(\mathbf{E})\mathbf{M}\big) - (\mathbf{EM}+\mathbf{ME}).
\end{aligned} \tag{4.19}$$

One observes that all right - hand sides in (4.19) can be expressed by linear combinations of the set of matrices $\{\mathbf{E}, \mathrm{tr}(\mathbf{E})\mathbf{I}, \mathrm{tr}(\mathbf{ME})\mathbf{M}, \mathrm{tr}(\mathbf{ME})\mathbf{I} + \mathrm{tr}(\mathbf{E})\mathbf{M}, \mathbf{EM}+\mathbf{ME}\}$ with fixed $\mathbf{M} = \mathbf{a}\cdot\mathbf{a}^T$ for given normed $\mathbf{a} \in \mathbb{R}^3$ and arbitrary matrix $\mathbf{E} \in \mathbb{R}^{3\times 3}$.
The expression

$$\begin{aligned}
2\tilde{\mathcal{G}}'_{ti}(\mathbf{I})[\mathbf{E}] &= 2\tilde{\mathcal{G}}'_{iso}(\mathbf{I})[\mathbf{E}] + 2\sum_{i=1}^{3}\varepsilon_i\Big(\tilde{\mathcal{G}}^{(i)}_{aniso}(\mathbf{I})\Big)'[\mathbf{E}]\\
&= \Big(4(\alpha+\delta)+4\varepsilon_3\Big)\mathbf{E} + \Big(4(\beta+\delta)+4(a_2+a_3)\varepsilon_2 + (b_1+b_2+4b_3-2)\varepsilon_3\Big)\mathrm{tr}(\mathbf{E})\mathbf{I}\\
&\quad + \Big(8\varepsilon_1 + 4(a_1+a_2-2)\varepsilon_2 + (b_1+b_2-2)\,\varepsilon_3\Big)\mathrm{tr}(\mathbf{ME})\mathbf{M}\\
&\quad + \Big(-4a_2\varepsilon_2 + (b_2-b_1+2)\,\varepsilon_3\Big)\big(\mathrm{tr}(\mathbf{ME})\mathbf{I} + \mathrm{tr}(\mathbf{E})\mathbf{M}\big)\\
&\quad + \Big(4\varepsilon_2 - 2\varepsilon_3\Big)(\mathbf{EM}+\mathbf{ME})
\end{aligned} \tag{4.20}$$

follows directly from (4.18) and (4.19).

Calculation of material parameters:

As already mentioned at the beginning of Section 4.3 the aim is to determine $(\alpha, \beta, \varepsilon_1, \varepsilon_2, \varepsilon_3)$ for given $\delta \geq 0$, $a_1, a_2, b_1, b_2 \geq 1$, nonzero $a_3, b_3 \geq -\frac{1}{2}$ such that our nonlinear model is consistent with a linear model. Recall that the restrictions for $\delta, a_1, a_2, a_3, b_1, b_2, b_3$ are necessary in order to guarantee polyconvexity of (4.14) (cp. the derivations in Section 4.2). For this purpose we have to introduce a linear model for transversely isotropic materials. The following introduction is based on [Alt12]. The stress - strain relation in the small - strain regime is given by $\boldsymbol{\sigma} = \mathcal{C}_{ti}\boldsymbol{\varepsilon}$ and has the same structure as in linear (isotropic) elasticity. $\mathcal{C}_{ti}$ is again a symmetric positive definite fourth - order tensor which maps symmetric strains $\boldsymbol{\varepsilon}$ into symmetric stresses $\boldsymbol{\sigma}$. The difference between the operators $\mathcal{C}$ for linear elasticity and $\mathcal{C}_{ti}$ for linear transverse isotropy is that $\mathcal{C}_{ti}$ contains now five independent physical material constants which describe the given material with transverse

isotropic behavior. In contrast, the linear elastic isotropic behavior was described by only two material parameters, either Young's modulus E and Poisson's ratio ν or equivalently the Lamé constants λ and μ.

If we choose the preferred direction as x_3 - direction, i.e. $\mathbf{a} = (0,0,1)^T$, the behavior can be described by material constants $E_1, E_3, \nu_{12}, \nu_{31}$ and G_{31} (cf. Section 12.4 in [Alt12]). E_3 is the elastic modulus in the preferred direction, E_1 is the elastic modulus in the isotropic x_1 - x_2 - plane (plane perpendicular to the preferred direction), ν_{12}, ν_{31} are two Poisson's ratios and G_{31} is the shear modulus in the x_3 - x_1 - plane. In general Poisson's ratios ν_{ij} characterize the transverse contraction between the directions i (direction of load) and j (direction of transverse strain). The shear moduli G_{ij} are necessary for the description of shearing strains in the x_i - x_j - plane. For $\mathbf{a} = (0,0,1)^T$ the stress - strain relation $\boldsymbol{\sigma} = \mathcal{C}_{ti}\boldsymbol{\varepsilon}$ is equivalent to

$$\tilde{\boldsymbol{\sigma}} := \begin{pmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{13} \\ \sigma_{12} \end{pmatrix} = \begin{pmatrix} N_{11} & N_{12} & N_{13} & 0 & 0 & 0 \\ N_{12} & N_{11} & N_{13} & 0 & 0 & 0 \\ N_{13} & N_{13} & N_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & N_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & N_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & N_{11} - N_{12} \end{pmatrix} \cdot \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{33} \\ \varepsilon_{23} \\ \varepsilon_{13} \\ \varepsilon_{12} \end{pmatrix} =: \tilde{\mathcal{C}}_{ti}\tilde{\boldsymbol{\varepsilon}}$$

in vector - matrix representation with constants

$$\begin{aligned} N_{11} &:= \frac{1 - \nu_{31}^2 \frac{E_1}{E_3}}{D}, & N_{12} &:= \frac{\nu_{12} + \nu_{31}^2 \frac{E_1}{E_3}}{D}, \\ N_{13} &:= \frac{(1+\nu_{12})\nu_{31}}{D}, & N_{33} &:= \frac{(1-\nu_{12})E_3}{1 - \nu_{12} - 2\nu_{31}^2 \frac{E_1}{E_3}}, \\ N_{44} &:= 2G_{31}, & D &:= \frac{1+\nu_{12}}{E_1}\left(1 - \nu_{12} - 2\nu_{31}^2 \frac{E_1}{E_3}\right), \end{aligned} \tag{4.21}$$

depending on the material constants $E_1, E_3, \nu_{12}, \nu_{31}$ and G_{31}. For the given material constants one has additional requirements

- $E_1 > 0$, $E_3 > 0$, $G_{31} > 0$, $-1 < \nu_{12} < 1$,
- $\nu_{31}^2 < \frac{E_3}{E_1} \Leftrightarrow 1 - \nu_{31}^2 \frac{E_1}{E_3} > 0$
- $1 - 2\nu_{31}^2 \frac{E_1}{E_3} > \nu_{12}$

such that $D > 0$ holds and the matrix entries of $\tilde{\mathcal{C}}_{ti}$ on the diagonal are positive (cf. [Alt12]).

The choice of $\nu := \nu_{12} = \nu_{31}$, $E := E_1 = E_3$ and $G := G_{31} = \frac{E}{2(1+\nu)} = \mu$, corresponds to an isotropic material. Thus for this choice we expect an isotropic behavior and the material parameters ε_i, $i = 1,2,3$, in the stored energy function must vanish in this case.

For a more detailed introduction into the material constants of anisotropic materials and

in particular of transverse isotropic materials we refer to [Alt12].
One can recompute that for $\mathbf{a} = (0,0,1)^T$, the corresponding matrix

$$\mathbf{M} = \mathbf{a} \cdot \mathbf{a}^T = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and a symmetric matrix $\mathbf{E} \in \mathbb{R}^{3\times 3}$, it holds

$$\begin{aligned} \mathcal{C}_{ti}\mathbf{E} = (N_{11} - N_{12})\mathbf{E} + N_{12}\mathrm{tr}(\mathbf{E})\mathbf{I} + (N_{11} - 2N_{13} + N_{33} - 2N_{44})\mathrm{tr}(\mathbf{ME})\mathbf{M} \\ + (N_{13} - N_{12})(\mathrm{tr}(\mathbf{ME})\mathbf{I} + \mathrm{tr}(\mathbf{E})\mathbf{M}) + (-N_{11} + N_{12} + N_{44})(\mathbf{EM} + \mathbf{ME}), \end{aligned} \tag{4.22}$$

i.e. the right - hand side is also expressed in terms of the set $\{\mathbf{E}, \mathrm{tr}(\mathbf{E})\mathbf{I}, \mathrm{tr}(\mathbf{ME})\mathbf{M}, \mathrm{tr}(\mathbf{ME})\mathbf{I} + \mathrm{tr}(\mathbf{E})\mathbf{M}, \mathbf{EM} + \mathbf{ME}\}$. Since we have to satisfy the condition $2\tilde{\mathcal{G}}'_{ti}(\mathbf{I})[\mathbf{E}] = \mathcal{C}_{ti}\mathbf{E}$ for all symmetric matrices $\mathbf{E} \in \mathbb{R}^{3\times 3}$, one can compare the coefficients of (4.20) and (4.22) in the case $\mathbf{a} = (0,0,1)^T$. This results in the linear system of equations

$$\begin{aligned} 4(\alpha + \delta) + 4\varepsilon_3 &= N_{11} - N_{12}, \\ 4(\beta + \delta) + 4(a_2 + a_3)\varepsilon_2 + (b_1 + b_2 + 4b_3 - 2)\varepsilon_3 &= N_{12}, \\ 8\varepsilon_1 + 4(a_1 + a_2 - 2)\varepsilon_2 + (b_1 + b_2 - 2)\,\varepsilon_3 &= N_{11} - 2N_{13} + N_{33} - 2N_{44}, \\ -4a_2\varepsilon_2 + (b_2 - b_1 + 2)\,\varepsilon_3 &= N_{13} - N_{12}, \\ 4\varepsilon_2 - 2\varepsilon_3 &= -N_{11} + N_{12} + N_{44}. \end{aligned}$$

In matrix notation the system is given by $\mathbf{Ax} = \mathbf{b}$ with

$$\mathbf{A} := \begin{pmatrix} 4 & 0 & 0 & 0 & 4 \\ 0 & 4 & 0 & 4(a_2 + a_3) & b_1 + b_2 + 4b_3 - 2 \\ 0 & 0 & 8 & 4(a_1 + a_2 - 2) & b_1 + b_2 - 2 \\ 0 & 0 & 0 & -4a_2 & b_2 - b_1 + 2 \\ 0 & 0 & 0 & 4 & -2 \end{pmatrix}, \ \mathbf{b} := \begin{pmatrix} N_{11} - N_{12} - 4\delta \\ N_{12} - 4\delta \\ N_{11} - 2N_{13} + N_{33} - 2N_{44} \\ N_{13} - N_{12} \\ -N_{11} + N_{12} + N_{44} \end{pmatrix}$$

for the unknown vector $\mathbf{x} = (\alpha, \beta, \varepsilon_1, \varepsilon_2, \varepsilon_3)^T$. The matrix $\mathbf{A}$ and the right - hand side $\mathbf{b}$ depend only on the given values $\delta \geq 0, a_1, a_2, b_1, b_2 \geq 1$, nonzero $a_3, b_3 \geq -\frac{1}{2}$ and the physical material constants $E_1, E_3, \nu_{12}, \nu_{31}, G_{31}$. Note that for a unique solution of the linear system of equations $\mathbf{Ax} = \mathbf{b}$ the free parameters $a_1, a_2, a_3, b_1, b_2, b_3$ have to be chosen in such a way that the rank of the matrix is full. Obviously one obtains a unique solution if and only if the subsystem

$$\begin{pmatrix} -4a_2 & b_2 - b_1 + 2 \\ 4 & -2 \end{pmatrix} \begin{pmatrix} \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} = \begin{pmatrix} N_{13} - N_{12} \\ -N_{11} + N_{12} + N_{44} \end{pmatrix}$$

is uniquely solvable. This subsystem is uniquely solvable if and only if the free parameters are chosen such that $b_2 - b_1 - 2a_2 + 2 \neq 0$.

Moreover one has to be careful in the choice of the free parameters δ and $a_1, a_2, a_3, b_1, b_2, b_3$ in order to guarantee polyconvexity of (4.14). For polyconvexity the entries in the solution $\mathbf{x}$ should be nonnegative. However, one can choose free parameters $\delta \geq 0$, $a_1, a_2, b_1, b_2 \geq 1$, nonzero $a_3, b_3 \geq -\frac{1}{2}$ and material parameters $(E_1, E_3, \nu_{12}, \nu_{31}, G_{31})$ such that some entries of $\mathbf{x}$ are negative. This is indeed physically meaningful, but does not satisfy the polyconvexity requirement. The free and material parameters should be chosen such that polyconvexity of (4.14) is also satisfied. This will be done in the numerical simulation in Section 6.3.

Remark 4.6: (Transition to the isotropic case)
For the transition to the isotropic case we set $E := E_1 = E_3$, $\nu := \nu_{12} = \nu_{31}$ and $G_{31} = \frac{E}{2(1+\nu)} = \mu$ which corresponds to material parameters for an isotropic material. Inserting these values into (4.21) leads to

$$N_{11} = N_{33} = \frac{E(1-\nu)}{(1+\nu)(1-2\nu)}, \quad N_{12} = N_{13} = \frac{E\nu}{(1+\nu)(1-2\nu)} = \lambda, \quad N_{44} = \frac{E}{1+\nu} = 2\mu$$

and therefore we get the right-hand side $\mathbf{b} = (2\mu - 4\delta, \lambda - 4\delta, 0, 0, 0)^T$. Under the assumption that the rank of $\mathbf{A}$ is full, we get the unique solution $\mathbf{x} = \left(\frac{\mu}{2} - \delta, \frac{\lambda}{4} - \delta, 0, 0, 0\right)^T$. This means that the anisotropic part in the stored energy function (4.14) plays no role in this case and the constants α and β correspond to the values determined in Section 2.4.5. Thus our model automatically tends reasonably to an isotropic model in this special case and therefore it can be used for the simulation of transversely isotropic and full isotropic materials.

4.4 Least squares formulation for transverse isotropic hyperelastic materials

In Section 4.3 we have determined the coefficients $(\alpha, \beta, \varepsilon_1, \varepsilon_2, \varepsilon_3)$ of the stored energy function (4.14) in such a way that our model is consistent with a linear model for transverse isotropic materials. Due to $\tilde{\mathcal{G}}'_{ti}(\mathbf{I}) = \frac{1}{2}\mathcal{C}_{ti}$ this means in particular, using Theorem 2.11, that the mapping $\tilde{\mathcal{G}}_{ti}(\mathbf{C})$ itself is locally invertible, at least in a neighborhood of $\mathbf{C} = \mathbf{I}$.
Following the steps similar to Section 3.3, we define for $(\mathbf{P}, \mathbf{u})$ (itself lying in a suitable function space)

$$\mathcal{R}_{ti}(\mathbf{P}, \mathbf{u}) := \begin{pmatrix} \omega_1 \left(\operatorname{div} \mathbf{P} + \mathbf{f} \right) \\ \omega_2 \left(\tilde{\mathcal{A}}_{ti}(\mathbf{F}(\mathbf{u})^{-1}\mathbf{P}) - \mathbf{C}(\mathbf{u}) \right) \end{pmatrix} \tag{4.23}$$

with $\tilde{\mathcal{A}}_{ti} := \tilde{\mathcal{G}}_{ti}^{-1}$ for compressible materials and a nonlinear least squares functional

$$\boxed{\begin{aligned} \mathcal{F}_{ti}(\mathbf{P}, \mathbf{u}) &:= \|\mathcal{R}_{ti}(\mathbf{P}, \mathbf{u})\|^2_{L^2(\Omega)} \\ &= \omega_1^2 \|\operatorname{div} \mathbf{P} + \mathbf{f}\|^2_{L^2(\Omega)} + \omega_2^2 \|\tilde{\mathcal{A}}_{ti}(\mathbf{F}(\mathbf{u})^{-1}\mathbf{P}) - \mathbf{C}(\mathbf{u})\|^2_{L^2(\Omega)}. \end{aligned}} \tag{4.24}$$

ω_1 and ω_2 are again scaling parameters. To find a minimizer we follow the same steps as in Section 3.3.2 for the linearization of (4.24) and Section 3.3.3 for its discretization. For the

computation of $\tilde{\mathcal{A}}_{ti}(\mathbf{F}(\mathbf{u})^{-1}\mathbf{P})$, $(\mathbf{P}, \mathbf{u})$ given, Newton's method is always applicable and is used for the numerical implementation.

At the end of this section we would like to point out that this method can be also extended to even more complicated materials (cp. for instance [Alt12] and [ESN10]). The most complicated case is a fully anisotropic solid with 21 material parameters instead of 5. In order to satisfy consistency of nonlinear models with appropriate linear ones the stored energy function will be much more complicated.

5 Model error and model adaptivity

In numerical mathematics we have to pay attention to different errors that can occur in an algorithm. On the one hand a certain quadrature formula for the integrations in (3.25) has to be chosen by the programmer. Quadrature formulas integrate by construction polynomials up to a certain degree exactly. If the occurring polynomials in the discretized method exceed this degree of exactness one gets a quadrature error, which tends normally to zero as the mesh size h tends to zero. On the other hand, especially in three dimensional problems or equivalently for huge linear systems of equations, one has to use iterative methods for solving the occurring linear systems of equations (3.26), since direct solvers generally have high memory requirements and quickly exceed the available resources. Using an iterative method then leads to an algebraic error.
Two further important errors in finite element methods are the discretization error, which occurs in a fixed model and normally vanishes for $h \to 0$, and additionally the **modeling error**. The modeling error is always present, since a mathematical model reflects the reality only up to a certain quality. For example the linear model of elasticity theory has its validity up to a certain load. Beyond this point a nonlinear model has to be used. But also here different models come into consideration, for instance a Neo - Hooke model versus a Mooney - Rivlin model, or even a still more complex model.
In this part of the work we would like to present a possibility to decide whether we only need a simple model on a particular element of our given triangulation $\mathcal{T}_h$ or a more complex one. Since an analysis for the Neo - Hooke case in the $\mathbf{B}$ - formulation is provided in Section 3.5, the explanations below are focused on the linear model as simple model and the Neo - Hooke model (cf. (2.30) with $\delta = 0$) as complex model. In general the explanations below can be extended to other choices of simple and complex models, for instance Neo - Hooke as simple and Mooney - Rivlin as complex model, and so on.

5.1 Preparations

The point of departure is on the one hand the first - order system operator

$$\begin{aligned} \mathcal{R}_{lin}(\mathbf{P},\mathbf{u}) &:= \begin{pmatrix} \omega_1^{\text{lin}}\,(\operatorname{div}\mathbf{P}+\mathbf{f}) \\ \omega_2^{\text{lin}}\,(\mathcal{A}_{lin}(\mathbf{P})-\boldsymbol{\varepsilon}(\mathbf{u})) \end{pmatrix} \\ &= \begin{pmatrix} \omega_1^{\text{lin}}\operatorname{div}\mathbf{P} \\ \omega_2^{\text{lin}}\,(\mathcal{A}_{lin}(\mathbf{P})-\boldsymbol{\varepsilon}(\mathbf{u})) \end{pmatrix} - \begin{pmatrix} -\omega_1^{\text{lin}}\,\mathbf{f} \\ \mathbf{0} \end{pmatrix} =: \mathcal{L}(\mathbf{P},\mathbf{u}) - \mathbf{r} \end{aligned} \tag{5.1}$$

of linear elasticity with the operator $\mathcal{A}_{lin}$ defined in (3.3) and $(\mathbf{P},\mathbf{u}) \in H(\operatorname{div};\Omega)^3 \times H^1(\Omega)^3$, satisfying the boundary conditions $\mathbf{P}\cdot\mathbf{n} = \mathbf{g}$ on Γ_N and $\mathbf{u} = \mathbf{u}_D$ on Γ_D. We use this model as simple model in our modeling error discussion. In comparison to Section 3.2 we have introduced scaling parameters $\omega_1^{\text{lin}}, \omega_2^{\text{lin}}$ here, similar to the nonlinear considerations in Section 3.3.1. The corresponding least squares functional, which has to be

minimized, is given by (cf. (3.9))

$$\mathcal{F}_{lin}(\mathbf{P},\mathbf{u}) := \|\mathcal{L}(\mathbf{P},\mathbf{u}) - \mathbf{r}\|^2_{L^2(\Omega)} = \|\mathcal{R}_{lin}(\mathbf{P},\mathbf{u})\|^2_{L^2(\Omega)}, \tag{5.2}$$

again defined for $(\mathbf{P},\mathbf{u}) \in H(\mathrm{div};\,\Omega)^3 \times H^1(\Omega)^3$, satisfying the prescribed boundary conditions. To obtain a structure that coincides with that in (3.21), we introduce in addition to (5.2) a second linear least squares functional as

$$\mathcal{F}^{\mathrm{lin}}_{lin}(\mathbf{R},\mathbf{w}) := \|\mathcal{R}_{lin}(\mathbf{P}^{(k)},\mathbf{u}^{(k)}) + \mathcal{R}'_{lin}(\mathbf{P}^{(k)},\mathbf{u}^{(k)})[\mathbf{R},\mathbf{w}]\|^2_{L^2(\Omega)} \tag{5.3}$$

for $(\mathbf{R},\mathbf{w}) \in H_{\Gamma_N}(\mathrm{div};\Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$ and fixed $(\mathbf{P}^{(k)},\mathbf{u}^{(k)}) \in H(\mathrm{div};\,\Omega)^3 \times H^1(\Omega)^3$, satisfying the prescribed boundary conditions.
The next Lemma shows that there is a close relation between the minimizer $(\mathbf{P}_{lin},\mathbf{u}_{lin})$ of (5.2) and the minimizer $(\mathbf{Q},\mathbf{v})$ of (5.3).

Lemma 5.1: (Relation between the two linear minimization problems)
$(\mathbf{P}_{lin},\mathbf{u}_{lin}) \in H(\mathrm{div};\,\Omega)^3 \times H^1(\Omega)^3$, satisfying the boundary conditions $\mathbf{P}_{lin}\cdot\mathbf{n} = \mathbf{g}$ on Γ_N and $\mathbf{u}_{lin} = \mathbf{u}_D$ on Γ_D, is the minimizer of (5.2) if and only if $(\mathbf{Q},\mathbf{v}) := (\mathbf{P}_{lin},\mathbf{u}_{lin}) - (\mathbf{P}^{(k)},\mathbf{u}^{(k)}) \in H_{\Gamma_N}(\mathrm{div};\Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$ is the minimizer of (5.3).

Proof:

We know by the considerations in Section 3.2 that the minimization problem (5.2) is equivalent to solve the variational problem:
Seek $(\mathbf{P}_{lin},\mathbf{u}_{lin}) \in H(\mathrm{div};\,\Omega)^3 \times H^1(\Omega)^3$ with

$$(\mathcal{L}(\mathbf{P}_{lin},\mathbf{u}_{lin}) - \mathbf{r}, \mathcal{L}(\mathbf{R},\mathbf{w}))_{L^2(\Omega)} = 0 \quad \forall\,(\mathbf{R},\mathbf{w}) \in H_{\Gamma_N}(\mathrm{div};\Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3.$$

Furthermore for arbitrary $(\mathbf{R},\mathbf{w}) \in H_{\Gamma_N}(\mathrm{div};\Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$ it holds
$\mathcal{R}'_{lin}(\mathbf{P}^{(k)},\mathbf{u}^{(k)})[\mathbf{R},\mathbf{w}] = \mathcal{L}(\mathbf{R},\mathbf{w})$ and thus

$$\begin{aligned}\mathcal{F}^{\mathrm{lin}}_{lin}(\mathbf{R},\mathbf{w}) &= \|\mathcal{R}_{lin}(\mathbf{P}^{(k)},\mathbf{u}^{(k)}) + \mathcal{R}'_{lin}(\mathbf{P}^{(k)},\mathbf{u}^{(k)})[\mathbf{R},\mathbf{w}]\|^2_{L^2(\Omega)}\\ &= \|\mathcal{L}(\mathbf{P}^{(k)},\mathbf{u}^{(k)}) - \mathbf{r} + \mathcal{L}(\mathbf{R},\mathbf{w})\|^2_{L^2(\Omega)}\\ &= \|\mathcal{L}(\mathbf{R},\mathbf{w}) - (\mathbf{r} - \mathcal{L}(\mathbf{P}^{(k)},\mathbf{u}^{(k)}))\|^2_{L^2(\Omega)}.\end{aligned} \tag{5.4}$$

Using (5.4), the minimization problem (5.3) has the same structure as (5.2) with $\hat{\mathbf{r}} := \mathbf{r} - \mathcal{L}(\mathbf{P}^{(k)},\mathbf{u}^{(k)})$ instead of $\mathbf{r}$. Then, again by the considerations in Section 3.2, this minimization problem is equivalent to find $(\mathbf{Q},\mathbf{v}) \in H_{\Gamma_N}(\mathrm{div};\Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$ with

$$(\mathcal{L}(\mathbf{Q},\mathbf{v}) - \hat{\mathbf{r}}, \mathcal{L}(\mathbf{R},\mathbf{w}))_{L^2(\Omega)} = 0 \quad \forall\,(\mathbf{R},\mathbf{w}) \in H_{\Gamma_N}(\mathrm{div};\Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3.$$

With this observation and due to $\mathcal{L}(\mathbf{Q},\mathbf{v}) = \mathcal{L}(\mathbf{P}_{lin},\mathbf{u}_{lin}) - \mathcal{L}(\mathbf{P}^{(k)},\mathbf{u}^{(k)})$ by definition of $(\mathbf{Q},\mathbf{v})$ and the linearity of the operator $\mathcal{L}$, the statement follows immediately.
□

Thus if we solve the minimization problem (5.3) and obtain its unique minimizer $(\mathbf{Q},\mathbf{v})$

we can set $(\mathbf{P}_{lin}, \mathbf{u}_{lin}) = (\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) + (\mathbf{Q}, \mathbf{v})$ and get the unique minimizer of (5.2) for arbitrary $(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})$, satisfying the boundary conditions.
On the other hand we use the first - order system operator (cf. (3.18))

$$\mathcal{R}_{NH}(\mathbf{P}, \mathbf{u}) := \begin{pmatrix} \omega_1 \left(\operatorname{div} \mathbf{P} + \mathbf{f}\right) \\ \omega_2 \left(\mathcal{A}_{NH}(\mathbf{P}\mathbf{F}(\mathbf{u})^T) - \mathbf{B}(\mathbf{u})\right) \end{pmatrix} \tag{5.5}$$

for the Neo - Hooke case in the $\mathbf{B}$ - formulation with operator $\mathcal{A}_{NH}$, defined by the first equation in (3.36) and (3.38), and the corresponding nonlinear least squares functional (cf. (3.19))

$$\mathcal{F}_{NH}(\mathbf{P}, \mathbf{u}) := \|\mathcal{R}_{NH}(\mathbf{P}, \mathbf{u})\|^2_{L^2(\Omega)} \tag{5.6}$$

as complex model. The linearized problem is then given by (cf. Section 3.3.2)

$$\mathcal{F}^{\text{lin}}_{NH}(\mathbf{R}, \mathbf{w}) := \|\mathcal{R}_{NH}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) + \mathcal{R}'_{NH}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{R}, \mathbf{w}]\|^2_{L^2(\Omega)} \tag{5.7}$$

with $\mathcal{R}'_{NH}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{R}, \mathbf{w}]$ defined by (3.53). These definitions are by the considerations in Section 3.5.2 reasonable for $(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) \in \mathbf{\Pi}^\infty \times \mathbf{U}^\infty$ and $(\mathbf{R}, \mathbf{w}) \in H_{\Gamma_N}(\operatorname{div}; \Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$.

Remark 5.2:
For $(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) = (\mathbf{0}, \mathbf{0})$, assuming zero boundary conditions, we have $\mathcal{R}_{NH}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) = \mathcal{R}_{NH}(\mathbf{0}, \mathbf{0}) = \begin{pmatrix} \omega_1 \mathbf{f} \\ \mathbf{0} \end{pmatrix}$, since $\mathcal{A}_{NH}(\mathbf{0}) = \mathbf{I} = \mathbf{B}(\mathbf{0})$ (cf. explanations below Corollary 3.14), and $\mathcal{R}_{lin}(\mathbf{0}, \mathbf{0}) = \begin{pmatrix} \omega_1^{\text{lin}} \mathbf{f} \\ \mathbf{0} \end{pmatrix}$. Moreover, due to (3.54) and the considerations above, we have

$$\mathcal{R}'_{NH}(\mathbf{0}, \mathbf{0})[\mathbf{R}, \mathbf{w}] = \begin{pmatrix} \omega_1 \operatorname{div} \mathbf{R} \\ 2\,\omega_2 \left(\mathcal{A}_{lin}(\mathbf{R}) - \boldsymbol{\varepsilon}(\mathbf{w})\right) \end{pmatrix}, \; \mathcal{R}'_{lin}(\mathbf{0}, \mathbf{0})[\mathbf{R}, \mathbf{w}] = \begin{pmatrix} \omega_1^{\text{lin}} \operatorname{div} \mathbf{R} \\ \omega_2^{\text{lin}} \left(\mathcal{A}_{lin}(\mathbf{R}) - \boldsymbol{\varepsilon}(\mathbf{w})\right) \end{pmatrix}.$$

Thus with $(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) = (\mathbf{0}, \mathbf{0})$, $\omega_1^{\text{lin}} = \omega_1$, $\omega_2^{\text{lin}} = 1$ and $\omega_2 = \frac{1}{2}$ we obtain

$$\begin{aligned}
\mathcal{F}^{\text{lin}}_{NH}(\mathbf{R}, \mathbf{w}) &= \|\mathcal{R}_{NH}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) + \mathcal{R}'_{NH}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{R}, \mathbf{w}]\|^2_{L^2(\Omega)} \\
&= \left\| \begin{pmatrix} \omega_1 \mathbf{f} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \omega_1 \operatorname{div} \mathbf{R} \\ 2\,\omega_2 \left(\mathcal{A}_{lin}(\mathbf{R}) - \boldsymbol{\varepsilon}(\mathbf{w})\right) \end{pmatrix} \right\|^2_{L^2(\Omega)} \\
&= w_1^2 \|\operatorname{div} \mathbf{R} + \mathbf{f}\|^2_{L^2(\Omega)} + \|\mathcal{A}_{lin}(\mathbf{R}) - \boldsymbol{\varepsilon}(\mathbf{w})\|^2_{L^2(\Omega)} \\
&= \left(\omega_1^{\text{lin}}\right)^2 \|\operatorname{div} \mathbf{R} + \mathbf{f}\|^2_{L^2(\Omega)} + \left(\omega_2^{\text{lin}}\right)^2 \|\mathcal{A}_{lin}(\mathbf{R}) - \boldsymbol{\varepsilon}(\mathbf{w})\|^2_{L^2(\Omega)} \\
&= \|\mathcal{R}_{lin}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) + \mathcal{R}'_{lin}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{R}, \mathbf{w}]\|^2_{L^2(\Omega)} = \mathcal{F}^{\text{lin}}_{lin}(\mathbf{R}, \mathbf{w}).
\end{aligned}$$

Thus for this choice the linear least squares functionals (5.3) and (5.7) coincide. This is expected from the considered consistency with linear elasticity in Section 2.4.5 and the observations at the beginning of Section 3.3. In this simple case the stiffness matrices and right - hand sides to the discrete problems coincide.

The considerations above motivate us to use the following formulations for our model error discussion with fixed $(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) \in \mathbf{\Pi}^{\infty} \times \mathbf{U}^{\infty}$:

Initial data:

Let $(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) \in \mathbf{\Pi}^{\infty} \times \mathbf{U}^{\infty}$ be given (with $\mathbf{\Pi}^{\infty}, \mathbf{U}^{\infty}$ defined by (3.55)) and choose the scaling parameters $\omega_1^{\text{lin}} = \omega_1$, $\omega_2^{\text{lin}} = 1$ and $\omega_2 = \frac{1}{2}$ in (5.1) and (5.5).

Simple model:

Minimize $\mathcal{F}_{lin}^{\text{lin}}(\mathbf{R}, \mathbf{w}) = \|\mathcal{R}_{lin}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) + \mathcal{R}'_{lin}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{R}, \mathbf{w}]\|^2_{L^2(\Omega)}$ about all $(\mathbf{R}, \mathbf{w}) \in H_{\Gamma_N}(\text{div}; \Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$ or equivalently find $(\mathbf{Q}, \mathbf{v}) \in H_{\Gamma_N}(\text{div};\, \Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$ with

$$\begin{aligned} &\Big(\mathcal{R}'_{lin}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{Q}, \mathbf{v}], \mathcal{R}'_{lin}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{R}, \mathbf{w}]\Big)_{L^2(\Omega)} \\ &\qquad = -\Big(\mathcal{R}_{lin}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}), \mathcal{R}'_{lin}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{R}, \mathbf{w}]\Big)_{L^2(\Omega)} \end{aligned} \tag{5.8}$$

for all $(\mathbf{R}, \mathbf{w}) \in H_{\Gamma_N}(\text{div};\, \Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$.

Linearized complex model:

Minimize $\mathcal{F}_{NH}^{\text{lin}}(\mathbf{R}, \mathbf{w}) = \|\mathcal{R}_{NH}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) + \mathcal{R}'_{NH}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{R}, \mathbf{w}]\|^2_{L^2(\Omega)}$ about all $(\mathbf{R}, \mathbf{w}) \in H_{\Gamma_N}(\text{div}; \Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$ or equivalently find $(\mathbf{Q}, \mathbf{v}) \in H_{\Gamma_N}(\text{div};\, \Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$ with

$$\begin{aligned} &\Big(\mathcal{R}'_{NH}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{Q}, \mathbf{v}], \mathcal{R}'_{NH}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{R}, \mathbf{w}]\Big)_{L^2(\Omega)} \\ &\qquad = -\Big(\mathcal{R}_{NH}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}), \mathcal{R}'_{NH}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{R}, \mathbf{w}]\Big)_{L^2(\Omega)} \end{aligned} \tag{5.9}$$

for all $(\mathbf{R}, \mathbf{w}) \in H_{\Gamma_N}(\text{div};\, \Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$.

Obviously the variational problems (5.8) and (5.9) have the same structure. Thus we can implement both problems in the same way.

5.2 Idea and algorithm for model adaptivity

An a - posteriori error estimator permits to decide in which part of the triangulated domain one should refine. Usually for the modeling of the underlying problem one fixed model is used in the whole domain Ω and in each of possible refinement levels. We are interested now in considering two different models simultaneously on one fixed mesh and in particular we want to decide on which part of the domain we can use the simpler model (i.e. in our

case the model of linear elasticity) and on which part we have to use the more complex model (i.e. in our case the nonlinear Neo - Hooke model).
In the rest of this chapter we assume that the Neo - Hooke model is an exact model and we have no quadrature and no algebraic error. Our aim is to measure the quality of the solution of linear elasticity totally and on single elements. We provide an algorithm which automatically decides on which part of the domain the simple model is insufficient and the complex model should be used. The main idea of model adaptivity is therefore that we only use the nonlinear model in a subdomain $\Omega_1 \subseteq \Omega$, where it is necessary and reasonable. We define the remaining domain as $\Omega_2 := \Omega \setminus \Omega_1$ and use the simple model on this subdomain. Thus instead of minimizing (5.6) we want to minimize the least squares functional

$$\mathcal{F}_{red}(\mathbf{P}, \mathbf{u}) := \|\mathcal{R}_{NH}(\mathbf{P}, \mathbf{u})\|^2_{L^2(\Omega_1)} + \|\mathcal{R}_{lin}(\mathbf{P}, \mathbf{u})\|^2_{L^2(\Omega_2)}. \tag{5.10}$$

$\mathcal{F}_{red}(\mathbf{P}, \mathbf{u})$ is still nonlinear on Ω_1 and for its minimization we use the damped Gauss - Newton method (described in Algorithm 1 for the discretized problem), i.e. we minimize again a sequence of linearized problems

$$\begin{aligned} \mathcal{F}^{\text{lin}}_{red}(\mathbf{R}, \mathbf{w}) := &\|\mathcal{R}_{NH}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) + \mathcal{R}'_{NH}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{R}, \mathbf{w}]\|^2_{L^2(\Omega_1)} \\ &+ \|\mathcal{R}_{lin}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) + \mathcal{R}'_{lin}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{R}, \mathbf{w}]\|^2_{L^2(\Omega_2)}, \end{aligned} \tag{5.11}$$

where $(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})$ is an old solution, satisfying the boundary conditions on Γ_N and Γ_D. We set the new solution again as $(\mathbf{P}^{(k+1)}, \mathbf{u}^{(k+1)}) = (\mathbf{P}^{(k)}, \mathbf{u}^{(k)}) + \alpha^{(k)}(\mathbf{Q}^{(k)}, \mathbf{v}^{(k)})$, where $(\mathbf{Q}^{(k)}, \mathbf{v}^{(k)})$ denotes the minimizer of (5.11) (cf. Section 3.3.3) with zero boundary conditions.
For the minimization of (5.11) in a finite element space we can use the discrete formulation of (5.9) on each element $T \in \Omega_1$ of the given triangulation $\mathcal{T}_h$. We call all elements $T \in \Omega_1$ **complex elements**. Analogously we use the discrete formulation of (5.8) for all elements $T \in \Omega_2$. We call these elements **simple elements**.
Since both variational problems (5.8) and (5.9) have the same structure, we can determine the local stiffness matrices and right - hand sides in the same way. Afterwards we assemble the global stiffness matrix and global right - hand side with the help of the local ones as usual in finite element methods. We call this mix of both models in the following **reduced model** and denote its solution, i.e. the minimizer of (5.10) as $(\mathbf{P}_{red}, \mathbf{u}_{red})$, and assume that it is still in the set $\mathbf{\Pi}^\infty \times \mathbf{U}^\infty$, defined by (3.55). At the beginning we always set $\Omega_1 = \emptyset$, i.e. we use the simple (respectively linear) model on all elements and the solution $(\mathbf{P}_{red}, \mathbf{u}_{red})$ equals $(\mathbf{P}_{lin}, \mathbf{u}_{lin})$. The following corollary provides a „measure of quality“ in order to decide on which part of the domain we have to switch to the complex (respectively Neo - Hooke) model.

Corollary 5.3: (Measure of quality)
Let $\theta > 0$ be sufficiently small in $\mathbf{\Pi}^\infty, \mathbf{U}^\infty$ (cf. (3.55)), $(\mathbf{P}_{NH}, \mathbf{u}_{NH}) \in \mathbf{\Pi}^\infty \times \mathbf{U}^\infty$ the

exact solution of (5.6), i.e. $\mathcal{R}_{NH}(\mathbf{P}_{NH}, \mathbf{u}_{NH}) = \mathbf{0}$, and assume that the Neo - Hooke model is an exact model. Moreover we assume that the minimizer $(\mathbf{P}_{red}, \mathbf{u}_{red})$ of (5.10) is also in $\mathbf{\Pi}^{\infty} \times \mathbf{U}^{\infty}$. Then the nonlinear functional $\mathcal{F}_{NH}$, evaluated in the reduced solution $(\mathbf{P}_{red}, \mathbf{u}_{red})$, measures the quality of $(\mathbf{P}_{red}, \mathbf{u}_{red})$ with respect to the exact model and the suitability of the reduced model.

Proof:

An immediate consequence of Theorem 3.29 for $(\mathbf{Q}, \mathbf{v}) = (\mathbf{P}_{NH}, \mathbf{u}_{NH})$ and $(\hat{\mathbf{Q}}, \hat{\mathbf{v}}) = (\mathbf{P}_{red}, \mathbf{u}_{red})$ is

$$\mathcal{F}_{NH}(\mathbf{P}_{red}, \mathbf{u}_{red}) \eqsim \|(\mathbf{P}_{NH} - \mathbf{P}_{red}, \mathbf{u}_{NH} - \mathbf{u}_{red})\|_{\mathcal{V}}^2$$

for $\mathcal{V} = H_{\Gamma_N}(\mathrm{div};\, \Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$, i.e. $\mathcal{F}_{NH}(\mathbf{P}_{red}, \mathbf{u}_{red})$ is up to constants equivalent to the error between the solution of the reduced model and the solution of the Neo - Hooke model. Since $(\mathbf{P}_{NH}, \mathbf{u}_{NH})$ is assumed to be the correct solution, $\mathcal{F}_{NH}(\mathbf{P}_{red}, \mathbf{u}_{red})$ measures the quality of the solution $(\mathbf{P}_{red}, \mathbf{u}_{red})$. Furthermore, since the Neo - Hooke model is assumed to be the exact model, $\mathcal{F}_{NH}(\mathbf{P}_{red}, \mathbf{u}_{red})$ also measures the suitability of the reduced model, i.e. its quality.

□

Corollary 5.3 provides a possibility to measure the quality of the reduced solution/model. In particular it provides a possibility to measure the quality of the solution $(\mathbf{P}_{lin}, \mathbf{u}_{lin})$ of linear elasticity. With the help of the least squares functional $\mathcal{F}_{NH}$, evaluated in the reduced solution, we can decide locally where we have to modify our model. Thus we are able to establish a method which automatically switches from the simple linear model to the complex Neo - Hooke model, if necessary. Thus the algorithm can adapt the model itself and we speak about **model adaptivity**. This is at least for small stresses and displacements near the origin theoretically ensured. Before we state the algorithm, we have to remark that $\mathcal{F}_{NH}(\mathbf{P}_{red}, \mathbf{u}_{red})$ is a value which reflects the total error as sum of discretization and model error. It would be more advantageous if one could split the total error into its two parts and could measure both errors individually. Then one could decide independently in which part of the domain one should refine and in which part one should use the complex model. However, $\mathcal{F}_{NH}(\mathbf{P}_{red}, \mathbf{u}_{red})$ measures the quality of the reduced solution with respect to the exact model and can be used a few times to adapt the model. After adapting the model on a fixed mesh several times one could do a step of (adaptive) refinement and use the model adaptivity on this new finer mesh, and so on.
Algorithm 2 needs besides a measure of quality a marking strategy. A marking strategy marks some elements of the given triangulation with the help of the given measure of quality. On the marked elements the simple model is substituted by the complex model. A logical assumption for Algorithm 2 is that an element $T \in \mathcal{T}_h$ which once becomes complex remains complex in the subsequent steps.

Algorithm 2 Model adaptivity on a fixed mesh

Require: Fixed triangulation $\mathcal{T}_h$ of the domain Ω, $tol_{\text{mod}} > 0$, $i_{\max}^{\text{mod}} \in \mathbb{N}$;
Marking strategy;

Set $i = 0$; $\Omega_1^{(i)} = \emptyset$; $\Omega_2^{(i)} = \Omega$ and calculate $\left(\mathbf{P}_{red}^{(i)}, \mathbf{u}_{red}^{(i)}\right)$ via discrete formulation of (5.8);
Choose $\Omega_1^{(i+1)}$ with the help of $\mathcal{F}_{NH}\left(\mathbf{P}_{red}^{(i)}, \mathbf{u}_{red}^{(i)}\right)$ and the given marking strategy;
Set $\Omega_2^{(i+1)} = \Omega \setminus \Omega_1^{(i+1)}$;
while $\mathcal{F}_{NH}\left(\mathbf{P}_{red}^{(i)}, \mathbf{u}_{red}^{(i)}\right) > tol_{\text{mod}}$ **and** $i < i_{\max}^{\text{mod}}$ **do**
  Determine $\left(\mathbf{P}_{red}^{(i+1)}, \mathbf{u}_{red}^{(i+1)}\right)$ with the help of Algorithm 1 $\Big($using $\left(\mathbf{P}_h^{(0)}, \mathbf{u}_h^{(0)}\right) = \left(\mathbf{P}_{red}^{(i)}, \mathbf{u}_{red}^{(i)}\right)$ as initial guess$\Big)$ and the reduced model $\Big($use discrete formulation of (5.9) on $\Omega_1^{(i+1)}$ and discrete formulation of (5.8) on $\Omega_2^{(i+1)}\Big)$;
  Set $i = i + 1$;
  Choose $\Omega_1^{(i+1)}$ as above;
  Set $\Omega_1^{(i+1)} = \Omega_1^{(i+1)} \cup \Omega_1^{(i)}$; {complex elements remain complex}
  Set $\Omega_2^{(i+1)} = \Omega \setminus \Omega_1^{(i+1)}$;
end while

In the i-th step of model adaptivity and therefore fixed $\Omega_1^{(i)}$ and $\Omega_2^{(i)}$, the reduced solution is determined by the damped Gauss-Newton method. Thus we use Algorithm 1, where now in each step of the Gauss-Newton iteration (5.8) is used on $\Omega_2^{(i)}$ and (5.9) is used on $\Omega_1^{(i)}$. We continue this until $\mathcal{F}_{NH}\left(\mathbf{P}_{red}^{(i)}, \mathbf{u}_{red}^{(i)}\right)$ goes below a given tolerance tol_{mod} or we exceed a prescribed number of model adaptivity steps $i_{\max}^{\text{mod}}$. As the output one obtains a sequence of „nonlinear“ domains $\Omega_1^{(i)}$ and a sequence of „linear“ domains $\Omega_2^{(i)}$. Furthermore we have determined an approximated minimizer of (5.10) for each of these i.
At the end of this chapter we would like to mention some benefits of this algorithm and in general of model adaptivity: For fixed $\Omega_1^{(i)}$ and $\Omega_2^{(i)}$ it is not necessary to recalculate the local stiffness matrices on $\Omega_2^{(i)}$ in the process of Gauss-Newton iterations, since $\mathcal{R}'_{lin}(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})[\mathbf{Q}, \mathbf{v}] = \mathcal{L}(\mathbf{Q}, \mathbf{v})$ is independent of $(\mathbf{P}^{(k)}, \mathbf{u}^{(k)})$ and thus by (5.8) the stiffness matrix on this part of the domain remains unchanged. Consequently one can save computational time. The possibility of reusing matrix entries of the linear model can be considered as general advantage of model adaptivity.
A second general advantage of model adaptivity can be found in the context of quadrature formulas: The usage of a nonlinear model needs in general a higher quadrature formula than a linear model for exact integration. Using a fixed quadrature formula could lead to situations where one integrates exactly in the parts $\Omega_2^{(i)}$ but not in the parts $\Omega_1^{(i)}$. In such situations the usage of the reduced model then leads to a smaller quadrature error than the usage of the full nonlinear model. Alternatively, one can also use higher quadrature formulas which integrate all polynomials exactly, i.e. also in the nonlinear part. But this leads in general to more effort and computational time.

6 Numerical examples

In this part of the work we present several numerical results using the developed least squares finite element methods for isotropic and transversely isotropic hyperelastic materials. Within the usage of the isotropic Neo - Hooke model the compressible as well as the fully incompressible case is considered.
The outline of this section is as follows: We start with some two - dimensional examples using the plane strain model (cf. Section 2.2.7 and 2.4.2). We continue with some three - dimensional examples using in addition to the Neo - Hooke model also the more complex Mooney - Rivlin model. For all these (isotropic) examples we restrict ourselves to the $\mathbf{B}$ - formulation where an analysis was provided in Section 3.5. In some of these examples we compare the results obtained with our LSFEM approach with the results of the pure displacement approach for compressible materials respectively with the results of the displacement - pressure approach for fully incompressible materials (cf. Section 3.6). Moreover, at the end of this chapter we consider one example for transversely isotropic materials in three dimensions (cf. Section 4) and one two - dimensional example for model adaptivity (cf. Section 5).
All examples are implemented in MATLAB®. For the occurring integrals in the discretized problems we use a quadrature rule which integrates polynomials up to degree 5 exactly (cf. Appendix B). Furthermore, as long as the memory resources are sufficient, we use the „backslash“/„divided into“ operator of MATLAB® (cf. [Att12]) for solving the occurring linear systems of equations. If we are close to exceed the available memory resources we use instead of the backslash solver the (iterative) preconditioned conjugate gradient method combined with an incomplete Cholesky factorization as preconditioner. In particular for the considered three - dimensional problems on finer meshes this is indispensable .
For adaptive refinement we use the nonlinear least squares functional, evaluated in the approximations, as a - posteriori error estimator to decide in which elements the error is locally large. Moreover we use the marking strategies described in Appendix C and combine them with standard refinement strategies (cf. [Riv84] and [Car04] for two dimensions and [Bey95] for three dimensions).
The physical units in the examples are neglected. But note that the deformations and the displacements have the physical unit of a length and the Lamé constants as well as the stress components have the physical unit of force per length squared.
In addition to the approximation of the displacements and stresses we are also interested in the numerically obtained convergence order of our algorithm as the mesh size h decreases. We have shown in Section 3.5.1 for the $\mathbf{B}$ - formulation that the nonlinear functional, evaluated in the approximation, is equivalent to the error at least if the loads are sufficiently small. Due to (3.86) we expect an optimal convergence rate of 2 for the error and $\sqrt{\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)}$ as long as the regularity assumptions in Propositions 2.44 and 2.46 and

Theorem 3.29 are satisfied and a combination of Raviart - Thomas elements $\mathcal{RT}_1(\mathcal{T}_h)$ for the stress approximations and piecewise quadratic elements $\mathcal{P}_2(\mathcal{T}_h)$ for the displacement approximations are used as finite element space. Note that also the usage of discontinuous elements for the stress approximation of P_{33} in a plane strain model leads to an optimal convergence rate of 2 (cf. Proposition 2.44 for a $L^2(\Omega)$ estimate) if the solution is sufficiently regular, i.e. $P_{33} \in H^2(\Omega)$.
It is well - known that the number of elements n_t is proportional to h^{-2} in two dimensions and proportional to h^{-3} in three dimensions. Thus with this choice we expect an optimal convergence order of

$$\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h) \lesssim h^4 \sim n_t^{-s} \tag{6.1}$$

with $s = 2$ in two dimensions and $s = \frac{4}{3}$ in three dimensions, provided that the regularity assumptions are satisfied.
In the following we consider two successive triangulations $\mathcal{T}_{h_l}$ and $\mathcal{T}_{h_{l+1}}$ with mesh sizes $h_{l+1} \leq h_l$ and number of elements $n_t^{(l+1)} \geq n_t^{(l)}$, $l \in \mathbb{N}$. On these meshes we can calculate the approximations $\left(\mathbf{P}_h^{(l+1)}, \mathbf{u}_h^{(l+1)}\right)$ and $\left(\mathbf{P}_h^{(l)}, \mathbf{u}_h^{(l)}\right)$.
We use the abbreviation $\mathcal{F}_l := \mathcal{F}_l\left(n_t^{(l)}\right) := \mathcal{F}_{NH}\left(\mathbf{P}_h^{(l)}, \mathbf{u}_h^{(l)}\right)$ and make the ansatz

$$\mathcal{F}_l = C \cdot \left(n_t^{(l)}\right)^{-r}$$

with unknowns $C > 0$ and $r > 0$. Due to $\log(\mathcal{F}_l) = \log(C) - r \log\left(n_t^{(l)}\right)$ we get a straight line with gradient $-r$ and intercept $\log(C)$ on the ordinate if we use a double logarithmic scaled diagram.
For the pairs $\left(n_t^{(l)}, \mathcal{F}_l\right)$ and $\left(n_t^{(l+1)}, \mathcal{F}_{l+1}\right)$, corresponding to the approximations $\left(\mathbf{P}_h^{(l)}, \mathbf{u}_h^{(l)}\right)$, $\left(\mathbf{P}_h^{(l+1)}, \mathbf{u}_h^{(l+1)}\right)$ on the triangulations $\mathcal{T}_{h_l}$, $\mathcal{T}_{h_{l+1}}$, we obtain the equations

$$\log(\mathcal{F}_{l+1}) = -r \log\left(n_t^{(l+1)}\right) + \log(C) \quad \text{and} \quad \log(\mathcal{F}_l) = -r \log\left(n_t^{(l)}\right) + \log(C). \tag{6.2}$$

Subtracting the second equation of (6.2) from the first one leads to

$$\log(\mathcal{F}_{l+1}) - \log(\mathcal{F}_l) = r\left(\log\left(n_t^{(l)}\right) - \log\left(n_t^{(l+1)}\right)\right) \Leftrightarrow r = \frac{\log\left(\frac{\mathcal{F}_{l+1}}{\mathcal{F}_l}\right)}{\log\left(\frac{n_t^{(l)}}{n_t^{(l+1)}}\right)}.$$

Note that r is the numerically obtained **convergence order (convergence rate)** and s is the theoretical convergence order which can be obtained if the solution is sufficiently regular. For such „regular“ problems one usually gets $r \approx s$, also with uniform refinement. But for „irregular“ problems one usually gets worse convergence rates r, i.e. $r < s$, using uniform refinement. In these cases adaptive refinement strategies play an important role. With these strategies one usually obtains convergence rates r close to s although the regularity assumptions are not satisfied in the considered problem. We will see this fact in

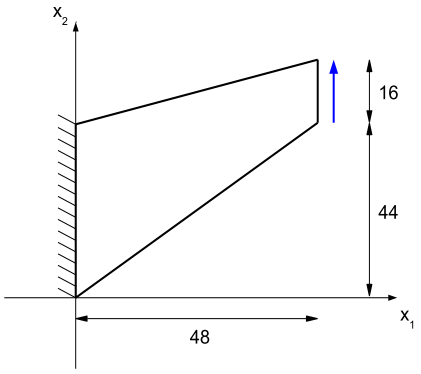
the concrete examples below.
Besides the evolution of the nonlinear functional we are also interested in the single term $\|\operatorname{div}\mathbf{P}_h + \mathbf{f}\|_{L^2(\Omega)}$ which describes the (linear) momentum in the $L^2(\Omega)$ - norm. Two questions arise: The first one is how good the conservation of momentum for the obtained approximations $\mathbf{P}_h$ is satisfied. Recall that the exact solution $\mathbf{P}$ satisfies $\operatorname{div}\mathbf{P} + \mathbf{f} = \mathbf{0}$. The second question considers the convergence rate of

$$\|\operatorname{div}(\mathbf{P} - \mathbf{P}_h)\|_{L^2(\Omega)} = \| - \mathbf{f} - \operatorname{div}\mathbf{P}_h\|_{L^2(\Omega)} = \|\operatorname{div}\mathbf{P}_h + \mathbf{f}\|_{L^2(\Omega)}.$$

By the interpolation estimate for the divergence in Proposition 2.46 one expects, similar as above, $\|\operatorname{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)} \lesssim h^4$, provided that the regularity assumptions in the proposition are satisfied and using Raviart - Thomas elements $(\mathcal{RT}_1(\mathcal{T}_h))^3$ for the approximation $\mathbf{P}_h$ of $\mathbf{P}$. By this interpolation estimate we expect a convergence order of 2 for $\|\operatorname{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)}$ in two dimensions and $\frac{4}{3}$ in three dimensions with respect to the number of elements n_t (cp. (6.1)). We will see in the results below that we actually get better convergence rates for $\|\operatorname{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)}$ using adaptive refinement. Moreover one gets better convergence orders for $\|\operatorname{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)}$ than for $\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$, regardless whether using uniform or adaptive refinement. Such an improvement was already observed in numerical experiments in [SSS10] for different least squares formulations. In that work one has approximately obtained twice as large convergence rates for the balance of momentum as for the least squares functional using uniform refinement. An corresponding analysis and further examples for the improvement of momentum balance can be found in [SSS11]. We will observe this interesting fact also numerically for the nonlinear case and the proposed least squares formulation in this work.

6.1 Two - dimensional problems for isotropic materials and a plane strain configuration

For our LSFEM approach in a plane strain model we use the space $\mathbf{\Pi}_h := (\mathcal{RT}_1(\mathcal{T}_h))^2 \times \mathcal{P}_{1,\text{disc}}(\mathcal{T}_h)$ for the stress approximations and $\mathbf{U}_h := (\mathcal{P}_2(\mathcal{T}_h))^2$ for the displacement approximations. $\mathcal{P}_{1,\text{disc}}(\mathcal{T}_h)$ denotes piecewise linear discontinuous elements for the approximation of the stress component P_{33}. For the pure displacement approach we use piecewise quadratic elements $\mathcal{P}_2(\mathcal{T}_h)$ respectively the non - conforming piecewise quadratic Fortin - Soulie elements for each component of $\mathbf{u}$ (cf. Section 2.5). For the displacement - pressure mixed finite element method we combine the Fortin - Soulie elements for the displacements with discontinuous piecewise linear pressure approximations. This pair of finite elements is inf - sup - stable for the mixed problem (3.103) in linear elasticity (cp. [FS83], Sections 4 and 8 in [BBF13] and Section 12 in [BS08]).
In Algortihm 1 which is essential for our least squares finite element method we choose in particular as input values the tolerance $tol = 10^{-6}$ in the stopping criterion and $i_{\max} = 50$ as maximal number of Gauss - Newton steps. Moreover, we use on the coarsest mesh (level



constraints on Γ_D:

- $\mathbf{u} = \mathbf{0}$ (left)

constraints on Γ_N:

- $\mathbf{P} \cdot \mathbf{n} = \mathbf{0}$ (top, bottom)
- $\mathbf{P} \cdot \mathbf{n} = \mathbf{g}$ (right)

Figure 6.1: Problem description of Cook's membrane in two dimensions

$l = 0$) the initial solution $\mathbf{P}_h^{(0)} = \mathbf{P}^N$, $\mathbf{u}_h^{(0)} = \mathbf{u}_D$ such that the boundary conditions are satisfied. On finer meshes ($l > 0$) we use the already computed solution from the previous mesh, interpolate this to the finer mesh and use the resulting interpolated solution as initial solution $\left(\mathbf{P}_h^{(0)}, \mathbf{u}_h^{(0)}\right)$ for the Gauss - Newton scheme on the finer mesh.
For the marking of elements in adaptive refinement strategies we use the percent marking strategy (cf. Appendix C.1) with $p = 10$, i.e. one - tenth of the elements are marked for regular refinement in each refinement step.
Note that, also in a plane strain model, the densities $\mathbf{f}$ and $\mathbf{g}$ are vector - valued with three components (cf. Sections 2.2.7 and 2.4.2), although the given domain Ω in the following examples is two - dimensional. Thus we assume in the following two - dimensional examples that the third component of $\mathbf{f}$ and $\mathbf{g}$ is always zero and therefore specify the densities only by two - dimensional vector - valued functions.

6.1.1 Cook's membrane with compressible Neo - Hooke

As a first example we study the so - called Cook membrane problem firstly considered in [CA 69] and [Coo74] by Robert D. Cook. The reference configuration and the prescribed boundary conditions are depicted in Figure 6.1. A surface force, more precisely an upward orientated traction force, is applied to the body on the right boundary. We do not apply any volume forces, i.e. we set $\mathbf{f} = \mathbf{0}$. For this example we use Poisson's ratio $\nu = 0.35$, Young's modulus $E = 200$ and $\mathbf{g} = (0, \gamma^{\text{load}})^T$ with load parameter $\gamma^{\text{load}} = 4$. Note that the domain of Cook's membrane contains a so - called corner singularity at $(0, 44)$ where the boundary conditions change from hard clamped ($\mathbf{u} = \mathbf{0}$) to a stress - free normal component ($\mathbf{P} \cdot \mathbf{n} = \mathbf{0}$) and the interior angle is larger than the critical one (cf. [Rös00]). Thus we expect a strong local refinement near this vertex using adaptive refinement strategies. As scaling parameters in the least squares functional (cf. (3.19)) we use $\omega_1 = 10^2$ and $\omega_2 = 1$. In Tables 6.1 and 6.2 the results for the considered problem, obtained with our LSFEM

Level l	(# Triangles)	$\dim \mathbf{\Pi}_h$	$\dim \mathbf{U}_h$	$\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$	(order)	# GN steps	$u_2(48, 60)$
0	186	2378	784	$2.3034 \cdot 10^{-2}$		6	5.9945
1	264	3380	1108	$9.5899 \cdot 10^{-3}$	(2.502)	5	6.0844
2	378	4854	1572	$4.1877 \cdot 10^{-3}$	(2.308)	5	6.1193
3	546	7018	2264	$1.8679 \cdot 10^{-3}$	(2.196)	5	6.1344
4	825	10635	3390	$8.1150 \cdot 10^{-4}$	(2.020)	5	6.1418
5	1243	16041	5090	$3.4713 \cdot 10^{-4}$	(2.072)	5	6.1449
6	1852	23936	7548	$1.4823 \cdot 10^{-4}$	(2.134)	6	6.1464
7	2738	35438	11108	$6.3157 \cdot 10^{-5}$	(2.182)	6	6.1469

Table 6.1: Results with adaptive refinement (compressible Neo - Hooke, 2d)

Level l	(# Triangles)	$\dim \mathbf{\Pi}_h$	$\dim \mathbf{U}_h$	$\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$	(order)	# GN steps	$u_2(48, 60)$
0	186	2378	784	$2.3034 \cdot 10^{-2}$		6	5.9945
1	744	9592	3056	$9.3014 \cdot 10^{-3}$	(0.654)	5	6.0869
2	2976	38528	12064	$3.8853 \cdot 10^{-3}$	(0.630)	5	6.1225
3	11904	154432	47936	$1.6405 \cdot 10^{-3}$	(0.622)	5	6.1372
4	47616	618368	191104	$6.9173 \cdot 10^{-4}$	(0.623)	5	6.1433

Table 6.2: Results with uniform refinement (compressible Neo - Hooke, 2d)

adaptive refinement		
Level l	$\|\mathrm{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)}$	(order)
0	$7.3330 \cdot 10^{-13}$	
1	$1.3494 \cdot 10^{-13}$	(4.833)
2	$2.4982 \cdot 10^{-14}$	(4.699)
3	$4.6497 \cdot 10^{-15}$	(4.572)
4	$8.4998 \cdot 10^{-16}$	(4.117)
5	$1.4788 \cdot 10^{-16}$	(4.266)
6	$2.6127 \cdot 10^{-17}$	(4.347)
7	$4.7685 \cdot 10^{-18}$	(4.351)

uniform refinement		
Level l	$\|\mathrm{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)}$	(order)
0	$7.3330 \cdot 10^{-13}$	
1	$1.3101 \cdot 10^{-13}$	(1.242)
2	$2.3676 \cdot 10^{-14}$	(1.234)
3	$4.2340 \cdot 10^{-15}$	(1.242)
4	$7.3776 \cdot 10^{-16}$	(1.260)

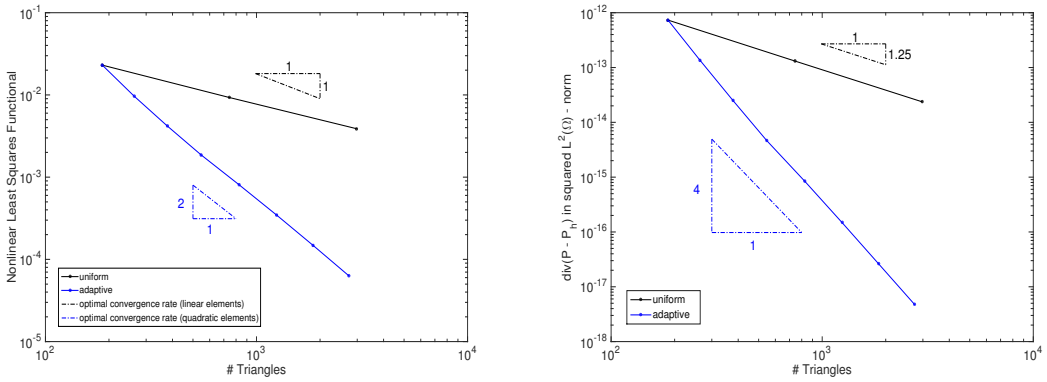
Table 6.3: Improved convergence rates for balance of momentum (compressible Neo - Hooke, 2d)

approach (**B** - formulation) and using adaptive respectively uniform refinement, are demonstrated.

In the third column of each table the values of the nonlinear functional $\mathcal{F}_{NH}$, evaluated in the computed approximations, and the corresponding convergence orders can be observed. One directly observes that the method using adaptive refinement is superior. One can achieve the theoretical optimal convergence rate of 2 and the method using uniform refinement is essential worse where one gets a convergence rate of merely approximately 0.63. This is as expected, since the problem is not sufficiently regular to obtain an optimal convergence order with uniform refinement. In Figure 6.2 (left) both behaviors are

graphically depicted. In the fourth column of Tables 6.1 and 6.2 the number of necessary Gauss - Newton steps, until the prescribed stopping criterion is achieved, are illustrated. We see that the number of steps is more or less constant and similar using adaptive or uniform refinement.

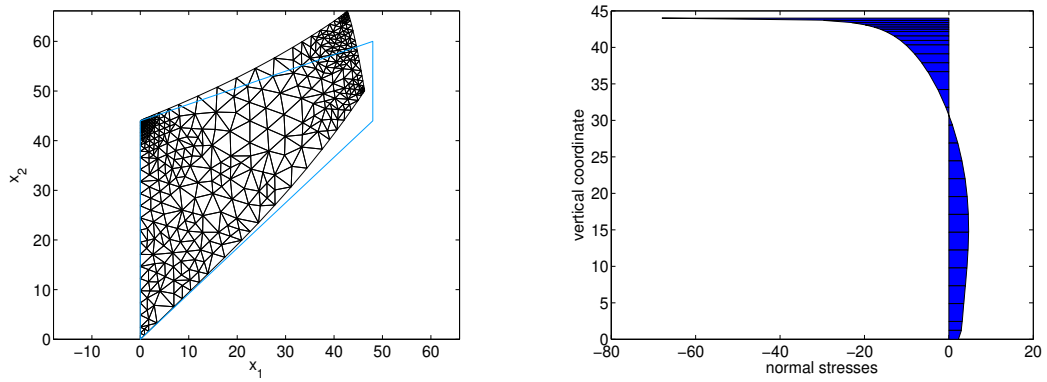
In Table 6.3 we observe that the conservation of linear momentum is satisfied very well. Moreover, we can observe an improved convergence rate of the term $\|\mathrm{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)}$ compared to the convergence rate of $\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$. The convergence rate for the balance of momentum is approximately doubled. The corresponding graphical impression can be found in Figure 6.2 (right).



(a) $\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$ (b) $\|\mathrm{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)}$

Figure 6.2: Comp. of adaptive and uniform refinement (compressible Neo - Hooke, 2d)

In Figure 6.3 the deformed mesh with its triangulation (left picture) and the normal stresses $\mathbf{n} \cdot \mathbf{P} \cdot \mathbf{n} = P_{11}$ on Γ_D (right picture) are drawn in level 4. Although our method produces a piecewise linear discontinuous stress along the left boundary, we see that the



(a) Deformed configuration (b) Normal stresses on Γ_D

Figure 6.3: Results in level 4 with adaptive LSFEM (compressible Neo - Hooke, 2d)

result is quite smooth. The expected singular behavior at $(0, 44)$ and therefore strong local refinement in this region can be observed in these plots.
In Figure 6.4 the first two rows of the Kirchhoff stress tensor approximation $\boldsymbol{\tau}_h = \mathbf{P}_h \mathbf{F}(\mathbf{u}_h)^T$ are plotted in level 4. The nondiagonal components seem equal, i.e. the approximation reflects the theoretical necessary symmetry of the exact Kirchhoff stress tensor (cf. Section 3.1 and Corollary 3.31). If we plot in comparison the nondiagonal elements of $\mathbf{P}_h$ in level 4, we see in Figure 6.5 that the approximation of the first Piola - Kirchhoff stress tensor $\mathbf{P}$ is not symmetric. Note that the singular behavior at the left upper vertex is also visible in all these stress plots.
At the end of this example we are interested in comparing our LSFEM approach with the displacement approach. Firstly we consider the vertical displacements of the vertex $(48, 60)$ for both approaches if we increase the number of elements. It is clear that both

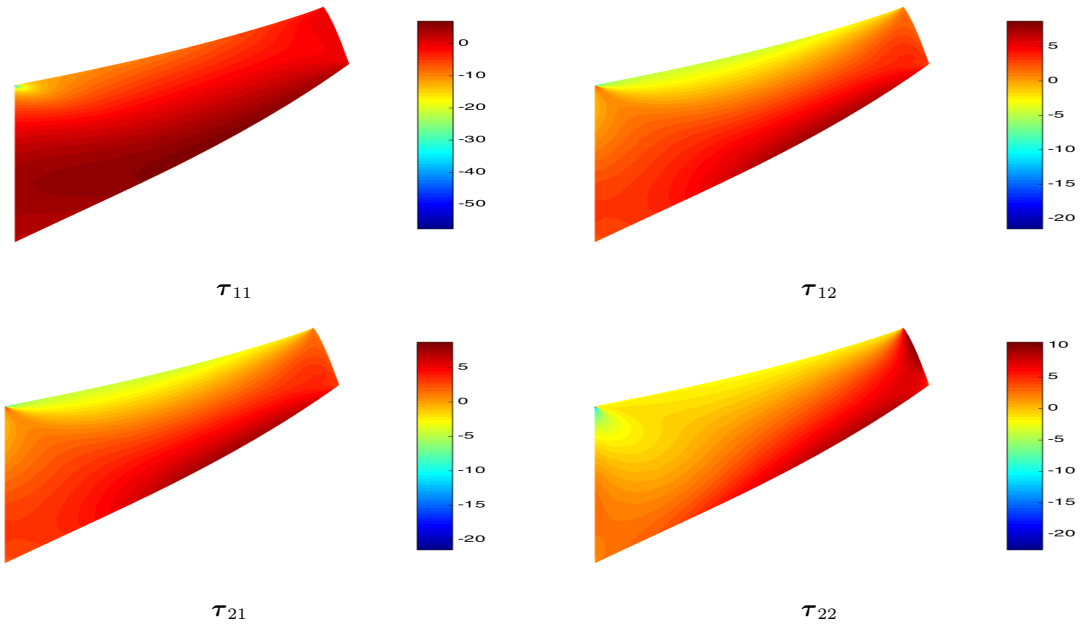


Figure 6.4: Components of the Kirchhoff stress (compressible Neo - Hooke, 2d)

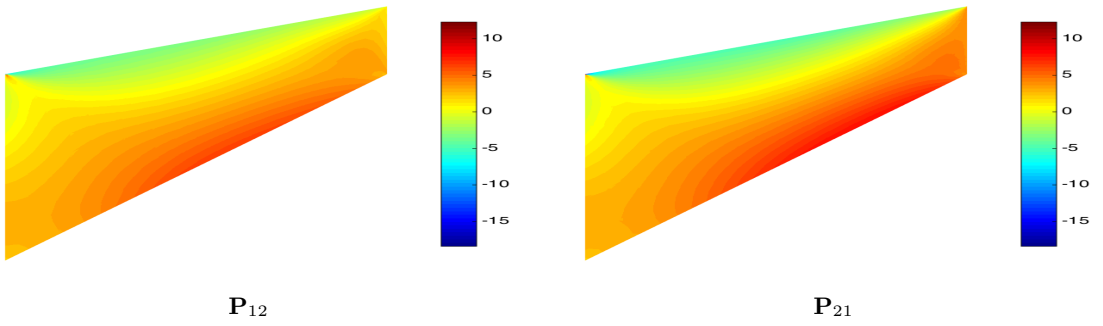
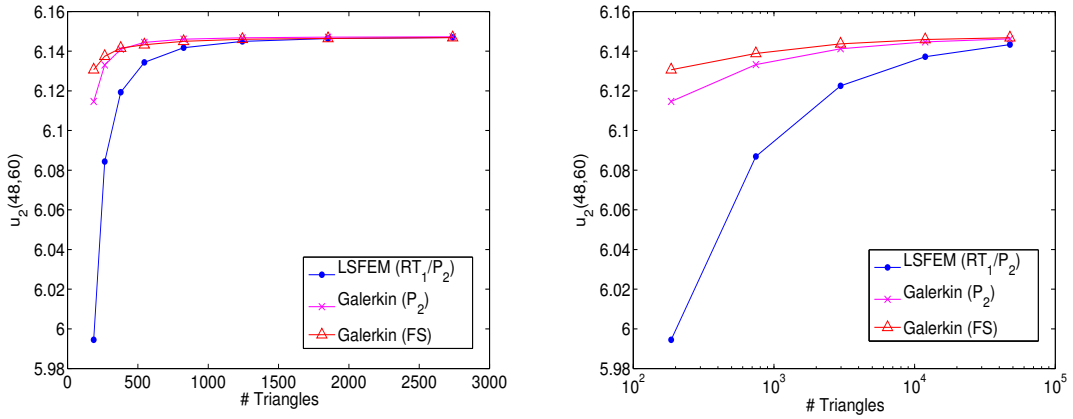


Figure 6.5: Nondiagonal components of $\mathbf{P}$ (compressible Neo - Hooke, 2d)

approaches should converge to the same displacement in this particular node. The values for the LSFEM approach can be found in the last column of Tables 6.1 (adaptive refinement) and 6.2 (uniform refinement). A graphical comparison of both approaches can be found in Figure 6.6 (left: adaptive refinement, right: uniform refinement). Here we have used the same meshes for the displacement approach that we have generated with our LSFEM approach.



(a) adaptive refinement

(b) uniform refinement

Figure 6.6: Vertical displacement in right upper node (compressible Neo-Hooke, 2d)

One observes that the displacement approximations of the Galerkin method are superior in comparison to the approximations of the LSFEM approach, at least on coarse meshes. Furthermore the results using Fortin-Soulie elements (abbrev. FS) are better than using standard piecewise quadratic elements (abbrev. $\mathcal{P}_2$). Additionally, it can be seen that both approaches converge to the same displacement value (approx. 6.1470).

Secondly we go back to the stress boundary approximations on Γ_D, where we will compare the values of the boundary integrals $\mathrm{Val}_1 := \int_{\Gamma_D} P_{11}\, ds$ and $\mathrm{Val}_2 := \int_{\Gamma_D} P_{21}\, ds$ for both approaches. For the displacement approach we distinguish moreover between Fortin-Soulie elements and standard continuous piecewise quadratic elements.

Before we show the results we state some preliminary considerations:

On the one hand, for an arbitrary vector $\mathbf{v} \in \mathbb{R}^2$, arbitrary load value $\gamma^{\mathrm{load}} \in \mathbb{R}$ and the prescribed boundary conditions on Γ_N, we obtain the equation

$$\int_{\Gamma_N} \mathbf{v} \cdot \mathbf{P} \cdot \mathbf{n}\, ds = \int_{\Gamma_R} \mathbf{v} \cdot \begin{pmatrix} 0 \\ \gamma^{\mathrm{load}} \end{pmatrix} ds = \int_{\Gamma_R} v_2 \gamma^{\mathrm{load}}\, ds = v_2 \gamma^{\mathrm{load}} |\Gamma_R| = 16 v_2 \gamma^{\mathrm{load}}, \quad (6.3)$$

where $\Gamma_R := \{(48, x_2) : 44 < x_2 < 60\}$ denotes the right part of the boundary Γ with length $|\Gamma_R| = 16$.

On the other hand for the choice $\mathbf{f} = \mathbf{0}$ we obtain $\operatorname{div} \mathbf{P} = \mathbf{0}$ (cf. (3.1)) and therefore with

the help of the divergence theorem the equation

$$\begin{aligned} \mathbf{0} = \int_\Omega \operatorname{div} \mathbf{P}\, dx = \int_\Gamma \mathbf{P}\cdot\mathbf{n}\, ds = \int_{\Gamma_N} \mathbf{P}\cdot\mathbf{n}\, ds + \int_{\Gamma_D} \mathbf{P}\cdot\mathbf{n}\, ds \\ \Leftrightarrow \int_{\Gamma_N} \mathbf{P}\cdot\mathbf{n}\, ds = -\int_{\Gamma_D} \mathbf{P}\cdot\mathbf{n}\, ds. \end{aligned} \tag{6.4}$$

Using (6.3) for the outer normal $\mathbf{v} = \mathbf{n} := (-1,0)^T$ on Γ_D, respectively the tangential vector $\mathbf{v} = \mathbf{t} := (0,1)^T$ orthogonal to $\mathbf{n}$, and combining this with (6.4) leads to

$$\begin{aligned} \int_{\Gamma_D} P_{11}\, ds &= \int_{\Gamma_D} \mathbf{n}\cdot\mathbf{P}\cdot\mathbf{n}\, ds = -\int_{\Gamma_N} \mathbf{n}\cdot\mathbf{P}\cdot\mathbf{n}\, ds = -16\cdot 0\cdot\gamma^{\text{load}} = 0, \\ \int_{\Gamma_D} P_{21}\, ds &= -\int_{\Gamma_D} \mathbf{t}\cdot\mathbf{P}\cdot\mathbf{n}\, ds = \int_{\Gamma_N} \mathbf{t}\cdot\mathbf{P}\cdot\mathbf{n}\, ds = 16\gamma^{\text{load}}, \end{aligned} \tag{6.5}$$

i.e. $\text{Val}_1 = 0$ and $\text{Val}_2 = 16\gamma^{\text{load}}$ ($= 64$ for $\gamma^{\text{load}} = 4$) are the exact values if one inserts the correct stress components P_{11}, P_{21} of $\mathbf{P}$.
In Table 6.4 the boundary integral values, obtained with our LSFEM approach and adaptive refinement, can be observed. One can see that these approximations are very close to the exact values, already on a very coarse mesh.

Level l	1	2	3	4	5	6	7
Val_1	$6.5207\cdot 10^{-6}$	$2.8131\cdot 10^{-6}$	$1.2118\cdot 10^{-6}$	$4.5390\cdot 10^{-7}$	$2.0002\cdot 10^{-7}$	$5.6404\cdot 10^{-7}$	$6.2720\cdot 10^{-7}$
Val_2	$6.4000\cdot 10^{1}$	$6.4000\cdot 10^{1}$	$6.4000\cdot 10^{1}$	$6.4000\cdot 10^{1}$	$6.4000\cdot 10^{1}$	$6.4000\cdot 10^{1}$	$6.4000\cdot 10^{1}$

Table 6.4: Values of boundary integrals on Γ_D (compressible Cook, adaptive LSFEM)

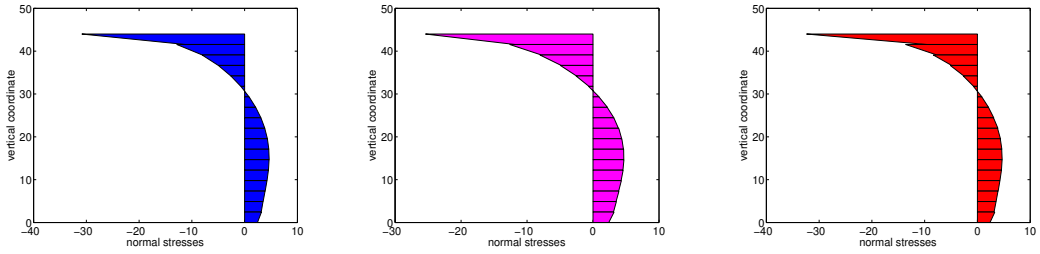
In Table 6.5 the values of the boundary integrals of the LSFEM approach can be compared with the values obtained with the pure displacement approach, using either $\mathcal{P}_2$ or Fortin - Soulie elements. Here a sequence of uniform refined meshes was used. The stress tensor in the pure displacement approach was computed in a post - processing with the help of the calculated approximation $\mathbf{u}$ (cf. $\mathbf{P}_{NH}(\mathbf{u})$ in (3.96)).

Level l	LSFEM ($\mathcal{RT}_1/\mathcal{P}_2$)		Galerkin ($\mathcal{P}_2$)		Galerkin (FS)	
	Val_1	Val_2	Val_1	Val_2	Val_1	Val_2
0	$1.5051\cdot 10^{-5}$	$6.4000\cdot 10^{1}$	$1.2720\cdot 10^{1}$	$6.3365\cdot 10^{1}$	$-2.6416\cdot 10^{-1}$	$6.3475\cdot 10^{1}$
1	$6.4612\cdot 10^{-6}$	$6.4000\cdot 10^{1}$	$8.2786\cdot 10^{0}$	$6.3764\cdot 10^{1}$	$-6.2475\cdot 10^{-2}$	$6.3653\cdot 10^{1}$
2	$2.7995\cdot 10^{-6}$	$6.4000\cdot 10^{1}$	$5.6891\cdot 10^{0}$	$6.4075\cdot 10^{1}$	$-8.6648\cdot 10^{-3}$	$6.3716\cdot 10^{1}$
3	$1.3555\cdot 10^{-6}$	$6.4000\cdot 10^{1}$	$4.0135\cdot 10^{0}$	$6.4264\cdot 10^{1}$	$-3.5480\cdot 10^{-3}$	$6.3733\cdot 10^{1}$
4	$-1.9792\cdot 10^{-6}$	$6.4000\cdot 10^{1}$	$2.8820\cdot 10^{0}$	$6.4368\cdot 10^{1}$	$-1.6818\cdot 10^{-2}$	$6.3730\cdot 10^{1}$

Table 6.5: Comparison of boundary stress approximations (compressible Cook)

One observes again that the LSFEM approach produces very good results, also for uniform refinement. The results for the displacement approach are overall poor and in particular do not converge to the correct values. The results with Fortin - Soulie elements are essentially better than with $\mathcal{P}_2$ elements, but still bad in comparison with the results of the LSFEM

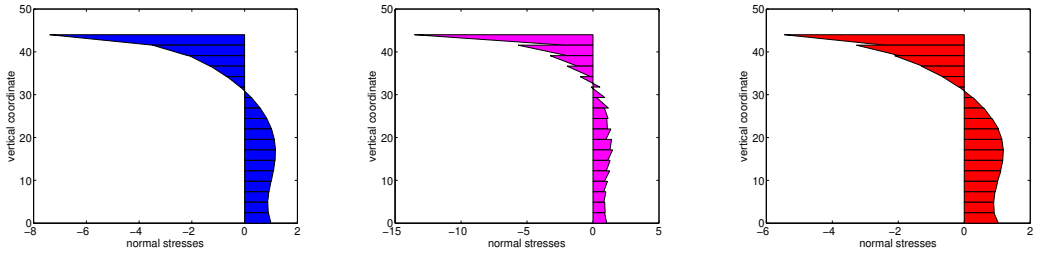
approach. Note that the results of the boundary integrals Val_1 and Val_2 of the Galerkin method were also checked in more detail in the case of very small loads and compressible materials. In this case, being in a regime of linear elasticity, the results are significantly better and converge to the correct values.
If one compares the normal stresses P_{11} on Γ_D in level 1 using uniform refinement (see Figure 6.7), one observes at first glance that they look fine for the Galerkin method.



LSFEM Galerkin ($\mathcal{P}_2$) Galerkin (FS)

Figure 6.7: Normal stresses on Γ_D (compressible Neo - Hooke, 2d, $\gamma^{\mathrm{load}} = 4$)

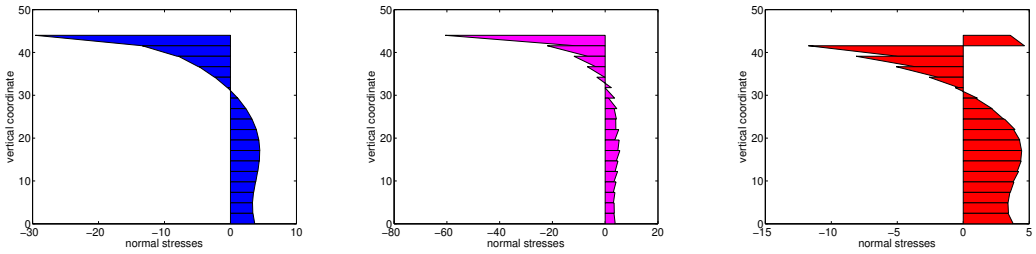
In order to convince the reader that the stress results on the boundary Γ_D using the Galerkin method are in general worse compared to the results of the LSFEM approach, we consider the same problem but with a less compressible material. More precisely we use $\nu = 0.499$ instead of $\nu = 0.35$. Figure 6.8 shows the results for $\gamma^{\mathrm{load}} = 1$ and Figure 6.9 displays the results for $\gamma^{\mathrm{load}} = 4$. Both figures corresponds again to the results on level 1 using uniform refinement. One observes that the LSFEM approach always yields excellent



LSFEM Galerkin ($\mathcal{P}_2$) Galerkin (FS)

Figure 6.8: Normal stresses on Γ_D (quasi - incompressible Neo - Hooke, 2d, $\gamma^{\mathrm{load}} = 1$)

results. The Galerkin method with $\mathcal{P}_2$ has strong discontinuities at the edge interfaces, also in the case of smaller loads. In the case $\gamma^{\mathrm{load}} = 1$ the Galerkin method with Fortin - Soulie elements seems okay, but if one increase the load value the results obviously fail. Altogether one can say that the Galerkin method cannot produce good stress approximations on the boundary if one tends to incompressible materials ($\nu \to \frac{1}{2}$) and/or increases the load value.



LSFEM Galerkin ($\mathcal{P}_2$) Galerkin (FS)

Figure 6.9: Normal stresses on Γ_D (quasi - incompressible Neo - Hooke, 2d, $\gamma^{\text{load}} = 4$)

We can conclude that our developed LSFEM approach produces essentially better stress approximations than the Galerkin method. We will examine in the following example if this is also true in the fully incompressible case. Here we will compare the LSFEM approach with the displacement - pressure formulation.

6.1.2 Cook's membrane with incompressible Neo - Hooke

We consider again the Cook membrane problem depicted in Figure 6.1. The main difference to the example in Section 6.1.1 is that we use now a fully incompressible material, i.e. we set actually $\lambda = \infty$. Furthermore we use $\mu = 1$ as second Lamé constant, the force densities $\mathbf{f} = \mathbf{0}$, $\mathbf{g} = (0, \gamma^{\text{load}})^T$ with $\gamma^{\text{load}} = 0.05$ and again the scaling parameters $\omega_1 = 10^2$, $\omega_2 = 1$.
The aim of this example is to confirm the results of compressible materials also for incompressible materials.
Table 6.6, using our LSFEM method with adaptive refinement, and Table 6.7, using our LSFEM approach with uniform refinement, show the results we have obtained for this problem. In comparison to the example with a compressible material in Section 6.1.1, we see in these tables that we need more steps in the Gauss - Newton scheme until the given stopping criterion is achieved and that the number of necessary steps vary stronger from mesh to mesh.
In Figure 6.10 (left), using the values of Tables 6.6 and 6.7, a graphical comparison between adaptive and uniform refinement can be found. Here again, the values of the nonlinear functional $\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$ are plotted against the number of triangles. One observes that we can achieve almost optimal convergence rates using adaptive refinement. The convergence rates for uniform refinement are worse, as expected, since the considered problem is still a singular problem. If we compare the convergence rates of Tables 6.6 and 6.7 with these of Tables 6.1 and 6.2 for compressible materials, we see that the achieved numerical con-

Level l	(# Triangles)	$\dim \mathbf{\Pi}_h$	$\dim \mathbf{U}_h$	$\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$	(order)	# GN steps	$u_2(48, 60)$
0	186	2378	784	$2.9972 \cdot 10^{-2}$		11	4.5092
1	275	3525	1150	$1.4042 \cdot 10^{-2}$	(1.939)	9	4.6120
2	390	5010	1620	$6.7178 \cdot 10^{-3}$	(2.110)	11	4.6586
3	559	7189	2314	$3.2427 \cdot 10^{-3}$	(2.023)	14	4.6810
4	821	10583	3374	$1.5525 \cdot 10^{-3}$	(1.916)	10	4.6921
5	1211	15633	4954	$7.3322 \cdot 10^{-4}$	(1.930)	12	4.6974
6	1796	23208	7324	$3.3695 \cdot 10^{-4}$	(1.973)	13	4.6999
7	2622	33918	10656	$1.4855 \cdot 10^{-4}$	(2.165)	14	4.7011

Table 6.6: Results with adaptive refinement (incompressible Neo - Hooke, 2d)

Level l	(# Triangles)	$\dim \mathbf{\Pi}_h$	$\dim \mathbf{U}_h$	$\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$	(order)	# GN steps	$u_2(48, 60)$
0	186	2378	784	$2.9972 \cdot 10^{-2}$		11	4.5092
1	744	9592	3056	$1.3800 \cdot 10^{-2}$	(0.559)	9	4.6141
2	2976	38528	12064	$6.4895 \cdot 10^{-3}$	(0.544)	10	4.6611
3	11904	154432	47936	$3.0743 \cdot 10^{-3}$	(0.539)	11	4.6830
4	47616	618368	191104	$1.4538 \cdot 10^{-3}$	(0.540)	13	4.6934

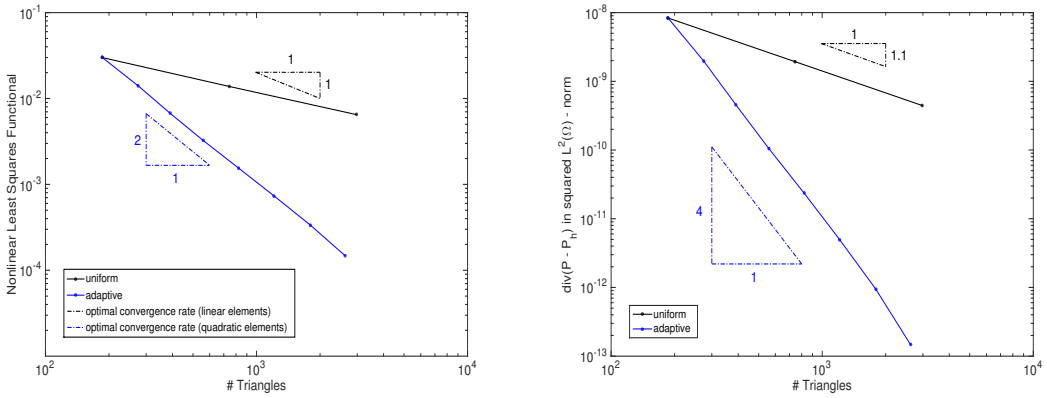
Table 6.7: Results with uniform refinement (incompressible Neo - Hooke, 2d)

adaptive refinement		
Level l	$\|\mathrm{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)}$	(order)
0	$8.3534 \cdot 10^{-9}$	
1	$1.9602 \cdot 10^{-9}$	(3.707)
2	$4.5544 \cdot 10^{-10}$	(4.177)
3	$1.0488 \cdot 10^{-10}$	(4.079)
4	$2.3596 \cdot 10^{-11}$	(3.881)
5	$4.9448 \cdot 10^{-12}$	(4.021)
6	$9.4024 \cdot 10^{-13}$	(4.212)
7	$1.4687 \cdot 10^{-13}$	(4.907)

uniform refinement		
Level l	$\|\mathrm{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)}$	(order)
0	$8.3534 \cdot 10^{-9}$	
1	$1.9315 \cdot 10^{-9}$	(1.056)
2	$4.4487 \cdot 10^{-10}$	(1.059)
3	$1.0116 \cdot 10^{-10}$	(1.068)
4	$2.2323 \cdot 10^{-11}$	(1.090)

Table 6.8: Improved convergence rates for balance of momentum (incompressible Neo - Hooke, 2d)

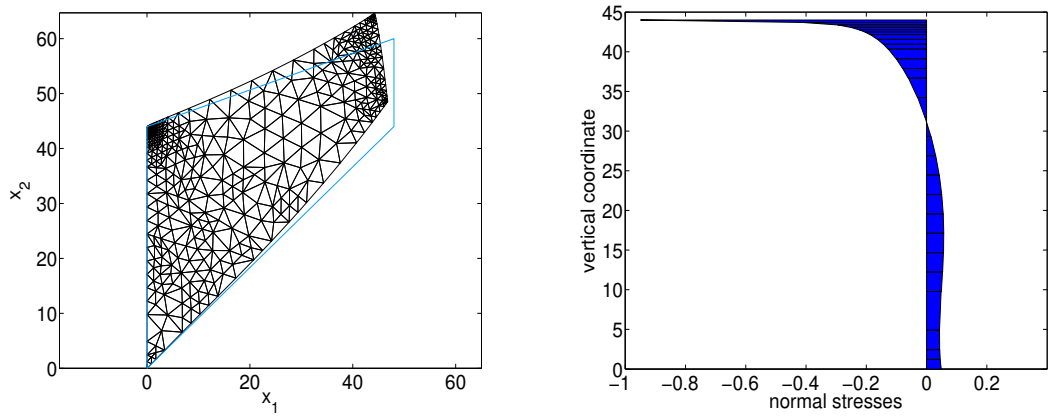
vergence orders are slightly worse in the incompressible case. In our opinion this is self - evident, since the problem is numerically much harder due to the incompressibility.
In Table 6.8 conservation of linear momentum is almost satisfied and an improved convergence rate can be again observed. Their convergence rates are approximately twice the convergence rates of $\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$, similar to the compressible case. A graphical impression can be found in Figure 6.10 (right).



(a) $\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$ (b) $\|\mathrm{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)}$

Figure 6.10: Comp. of adaptive and uniform refinement (incompressible Neo - Hooke, 2d)

In Figure 6.11 the deformed mesh (left picture) and the normal stresses on Γ_D (right picture) on level 4 using adaptive refinement are plotted. In the left picture also the reference configuration is drafted in cyan blue. We see in these pictures, analogously to the



Deformed configuration Normal stresses on Γ_D

Figure 6.11: Results in level 4 with adaptive LSFEM (incompressible Neo - Hooke, 2d)

compressible example, that we have a strong local refinement near the corner singularity at $(0, 44)$ and near the right boundary where the traction force is applied. Moreover, although the LSFEM method produces piecewise linear and discontinuous normal stress approximations on Γ_D, the results look quite smooth.
In Figure 6.12 the first two rows of the Kirchhoff stress tensor approximation on the same triangulation, are plotted. The nondiagonal components seem identical. In comparison to the Kirchhoff stress tensor the nondiagonal components of the first Piola - Kirchhoff stress

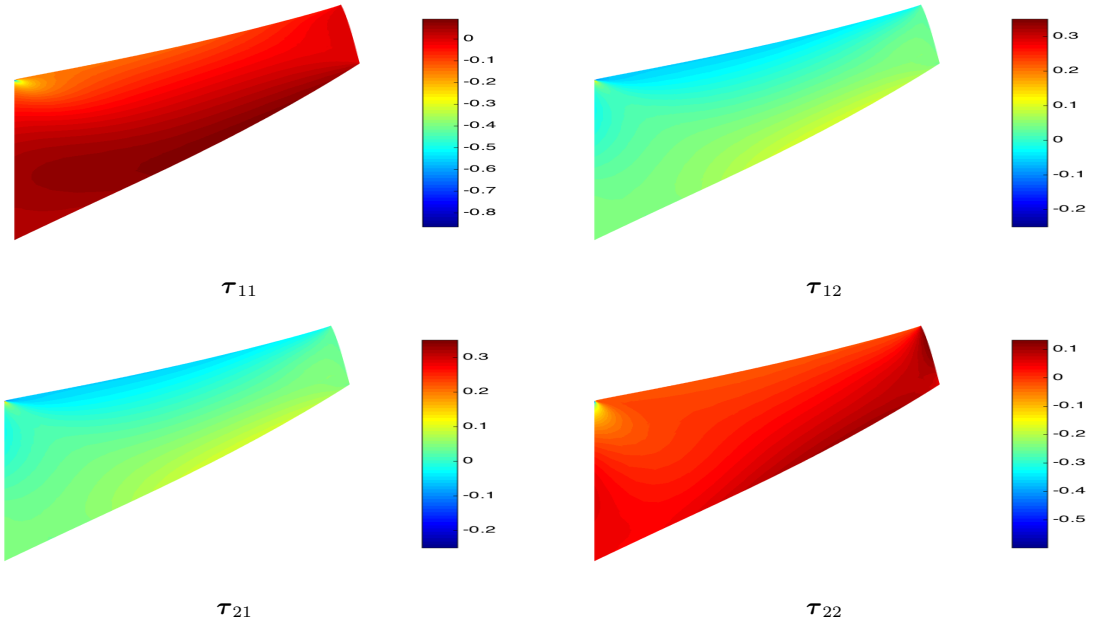


Figure 6.12: Components of the Kirchhoff stress (incompressible Neo - Hooke, 2d)

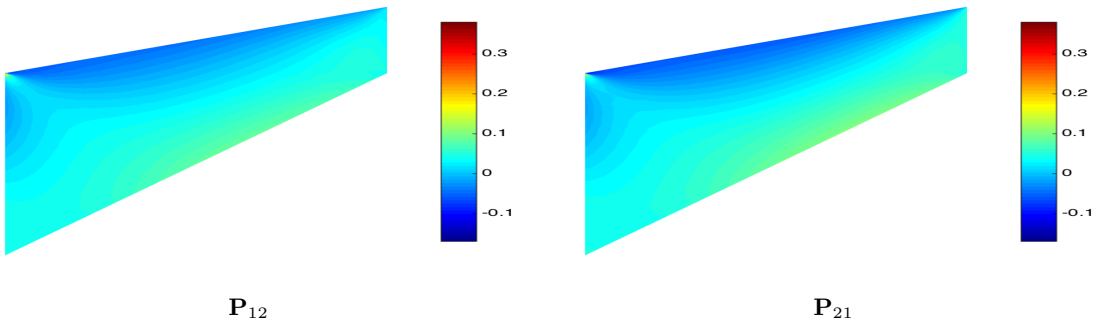
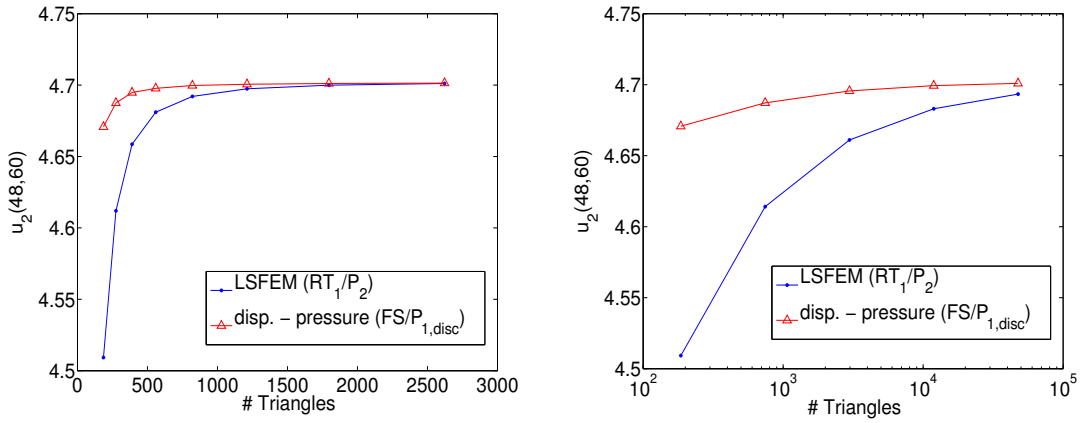


Figure 6.13: Nondiagonal components of $\mathbf{P}$ (incompressible Neo - Hooke, 2d)

tensor are drawn in Figure 6.13. Although it is difficult to see a difference at first glance, one can observe minor differences if one compare the components carefully.

In order to convince the reader that the results obtained with our LSFEM approach are reasonable we compare the vertical displacements in the node $(48, 60)$ with the displacement - pressure approach, similar as in Section 6.1.1 (see Figure 6.14). One obtains similar results as in the compressible case: The displacement approximations of the LSFEM approach are quite bad on a coarse mesh and the displacement - pressure approach is obviously superior. However, it is also obvious that both approaches converge to the same displacement value if one increases the number of elements. The results in the left picture with adaptive refinement are advantageous, since we need much less elements in order to be close to the correct displacement value (≈ 4.7013). Here we have used again the same meshes for the displacement - pressure approach that we have generated with our

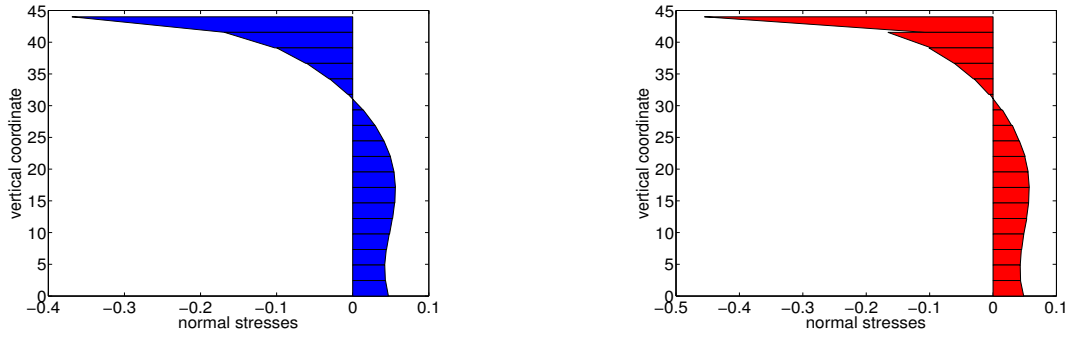


(a) adaptive refinement (b) uniform refinement

Figure 6.14: Vertical displacement in right upper node (incompressible Neo - Hooke, 2d)

adaptive LSFEM approach. Thus we can confirm the results for compressible materials also for incompressible materials, and record that the LSFEM approach is evidently worse in displacement approximations on coarse meshes.

In contrast, we have observed in the example of compressible materials (cf. Section 6.1.1) that the stress approximations are superior in the LSFEM approach. In order to confirm also this consideration in the incompressible case we firstly compare again the normal stress approximation P_{11} on Γ_D. We see in Figure 6.15 that the results for the displacement - pressure approach are more discontinuous close to the singularity at $(0, 44)$. This effect is enforced if we increase the forces to $\gamma^{\text{load}} = 0.25$ (cf. Figure 6.16).

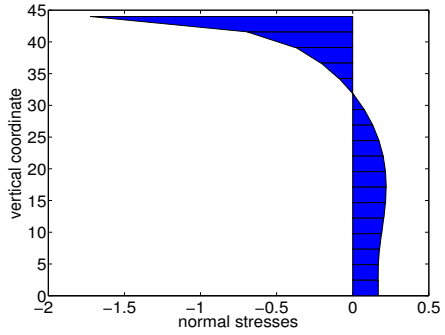


LSFEM ($\mathcal{RT}_1/\mathcal{P}_2$) displacement - pressure (FS/$\mathcal{P}_{1,\text{disc}}$)

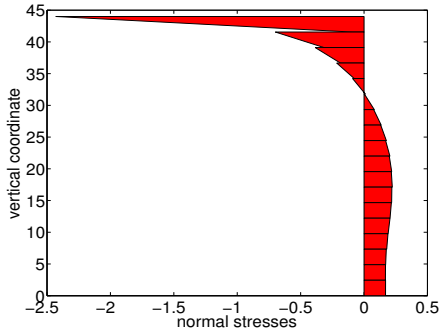
Figure 6.15: Normal stresses on Γ_D (incompressible Neo - Hooke, 2d, $\gamma^{\text{load}} = 0.05$)

Thus, the LSFEM approach seems superior with respect to stress approximations. We can confirm this statement considering the boundary integrals in equation (6.5):

For this example with $\gamma^{\text{load}} = 0.05$ the correct values are $\text{Val}_1 = 0$ and $\text{Val}_2 = 16 \cdot 0.05 = 8 \cdot 10^{-1}$. A comparison of the obtained values for both approaches with uniform refinement



LSFEM ($\mathcal{RT}_1/\mathcal{P}_2$)



displacement - pressure ($\mathrm{FS}/\mathcal{P}_{1,\mathrm{disc}}$)

Figure 6.16: Normal stresses on Γ_D (incompressible Neo - Hooke, 2d, $\gamma^{\mathrm{load}} = 0.25$)

can be found in Table 6.9.

We observe that the boundary stress approximations are essentially better using the LSFEM approach. In the displacement - pressure approach we cannot observe any convergence to the exact values. Note again, that these boundary integral approximations in the displacement - pressure method become better if one decreases the load value.

For the sake of completeness we look at Table 6.10 where good approximations of Val_1 and Val_2, using our LSFEM approach with adaptive refinement, are illustrated. Furthermore,

Level l	LSFEM ($\mathcal{RT}_1/\mathcal{P}_2$)		displacement - pressure ($\mathrm{FS}/\mathcal{P}_{1,\mathrm{disc}}$)	
	Val_1	Val_2	Val_1	Val_2
0	$1.6806 \cdot 10^{-3}$	$7.9764 \cdot 10^{-1}$	$-7.4612 \cdot 10^{-4}$	$7.9133 \cdot 10^{-1}$
1	$8.1169 \cdot 10^{-4}$	$7.9886 \cdot 10^{-1}$	$1.1027 \cdot 10^{-3}$	$7.9393 \cdot 10^{-1}$
2	$3.8982 \cdot 10^{-4}$	$7.9945 \cdot 10^{-1}$	$1.5488 \cdot 10^{-3}$	$7.9447 \cdot 10^{-1}$
3	$1.8587 \cdot 10^{-4}$	$7.9974 \cdot 10^{-1}$	$1.6826 \cdot 10^{-3}$	$7.9403 \cdot 10^{-1}$
4	$8.7255 \cdot 10^{-5}$	$7.9988 \cdot 10^{-1}$	$1.9403 \cdot 10^{-3}$	$7.9299 \cdot 10^{-1}$

Table 6.9: Comparison of boundary stress approximations (incompressible Cook)

Level l	1	2	3	4	5	6	7
Val_1	$8.1463 \cdot 10^{-4}$	$3.9221 \cdot 10^{-4}$	$1.8788 \cdot 10^{-4}$	$8.9001 \cdot 10^{-5}$	$4.0611 \cdot 10^{-5}$	$1.7616 \cdot 10^{-5}$	$6.8998 \cdot 10^{-6}$
Val_2	$7.9885 \cdot 10^{-1}$	$7.9945 \cdot 10^{-1}$	$7.9974 \cdot 10^{-1}$	$7.9987 \cdot 10^{-1}$	$7.9994 \cdot 10^{-1}$	$7.9998 \cdot 10^{-1}$	$7.9999 \cdot 10^{-1}$

Table 6.10: Values of boundary integrals on Γ_D (incompressible Cook, adaptive LSFEM)

convergence to the exact boundary integral values can be observed.

Altogether we can conclude that the results for the incompressible and compressible case are very similar. Moreover, we note that the LSFEM approach is inferior with respect to displacement approximations, at least on coarse meshes, and superior with respect to stress approximations. We will emphasize that the stress approximations of $\boldsymbol{\tau}$ in the compressible as well as in the incompressible case are almost symmetric (cf. Figures 6.4 and

6.12). This confirms numerically Corollary 3.31 where it was proven that $\boldsymbol{\tau}_h = \mathbf{P}_h\mathbf{F}(\mathbf{u}_h)^T$ converge to the symmetric Kirchhoff stress tensor as long as $\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$ tends to zero.

6.1.3 Cook's membrane with triple length and incompressible Neo - Hooke

In this problem we show the importance to scale the term of balance of (linear) momentum sufficiently large. We will see exemplarily that one can obtain poor results choosing wrong scaling parameters. Moreover we show in this example, provided that we have scaled the balance of momentum sufficiently large, that we get also excellent results for more bending dominated problems, even in the fully incompressible case.
For this purpose we consider a Cook membrane with triple length. The reference configuration of the Cook membrane with triple length is defined by the vertices $(0, 0)$, $(144, 44)$, $(144, 60)$ and $(0, 44)$. Also for this problem we use $\mathbf{f} = \mathbf{0}$ and the same boundary conditions as in Figure 6.1. More precisely we set $\mathbf{g} = (0, \gamma^{\text{load}})^T$ with $\gamma^{\text{load}} = 0.05$. Moreover, we set the Lamé constants again to $\lambda = \infty$, $\mu = 1$ and the scaling parameter corresponding to the inverse stress - strain relation (cf. (3.18)) to $\omega_2 = 1$.
In Tables 6.11 and 6.12 the dependence on the scaling parameter ω_1 of the horizontal and vertical displacement in the particular node $(144, 60)$ can be observed. Here the different scaling parameters $\omega_1 \in \left\{10^0, 10^1, \ldots, 10^4\right\}$ are taken into account in the adaptive LSFEM

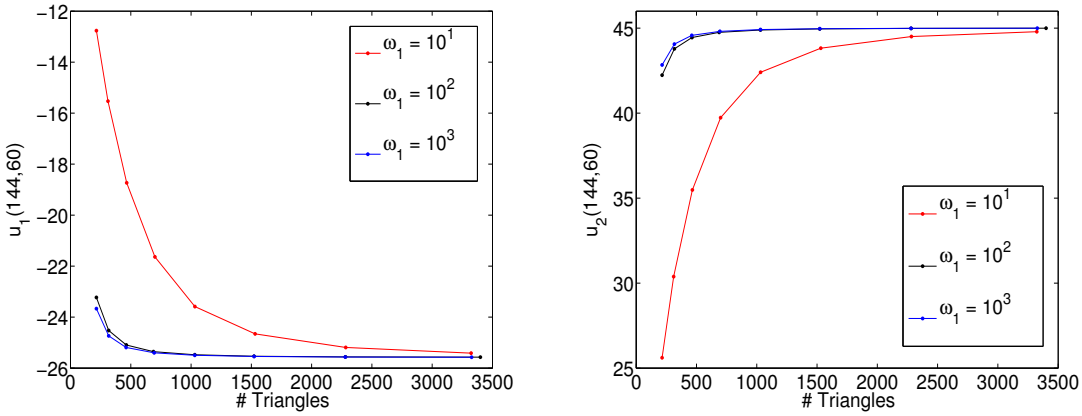
Level l	$\omega_1 = 10^0$	$\omega_1 = 10^1$	$\omega_1 = 10^2$	$\omega_1 = 10^3$	$\omega_1 = 10^4$
0	−8.2096	−12.7656	−23.2275	−23.6676	−23.6721
1	−8.3313	−15.5299	−24.5254	−24.7353	−24.7375
2	−8.3210	−18.7317	−25.0930	−25.1892	−25.1902
3	−8.5862	−21.6370	−25.3539	−25.3995	−25.3999
4	−9.4137	−23.5856	−25.4764	−25.4973	−25.4975
5	−10.1742	−24.6572	−25.5342	−25.5430	−25.5430
6	−10.7988	−25.1888	−25.5607	−25.5643	−25.5644
7	−11.8603	−25.4122	−25.5708	−25.5723	−25.5723

Table 6.11: Comparison of displacements $u_1(144, 60)$ using different ω_1

Level l	$\omega_1 = 10^0$	$\omega_1 = 10^1$	$\omega_1 = 10^2$	$\omega_1 = 10^3$	$\omega_1 = 10^4$
0	17.2227	25.6091	42.2354	42.8338	42.8399
1	17.2742	30.3833	43.7839	44.0616	44.0644
2	17.2024	35.4820	44.4502	44.5761	44.5774
3	17.6105	39.7345	44.7518	44.8106	44.8112
4	19.1545	42.4061	44.8910	44.9177	44.9180
5	20.5333	43.8196	44.9563	44.9676	44.9677
6	21.7195	44.5070	44.9846	44.9893	44.9894
7	23.7646	44.7934	44.9962	44.9981	44.9981

Table 6.12: Comparison of displacements $u_2(144, 60)$ using different ω_1

approach. A graphical impression of this scaling issue is depicted in Figure 6.17. In this figure the poor results for $\omega_1 = 10^0 = 1$ are neglected. Furthermore, since the displacement results for $\omega_1 = 10^3$ and $\omega_1 = 10^4$ are almost equal, the curve for $\omega_1 = 10^4$ is also omitted.



(a) horizontal displacement in $(144, 60)$ (b) vertical displacement in $(144, 60)$

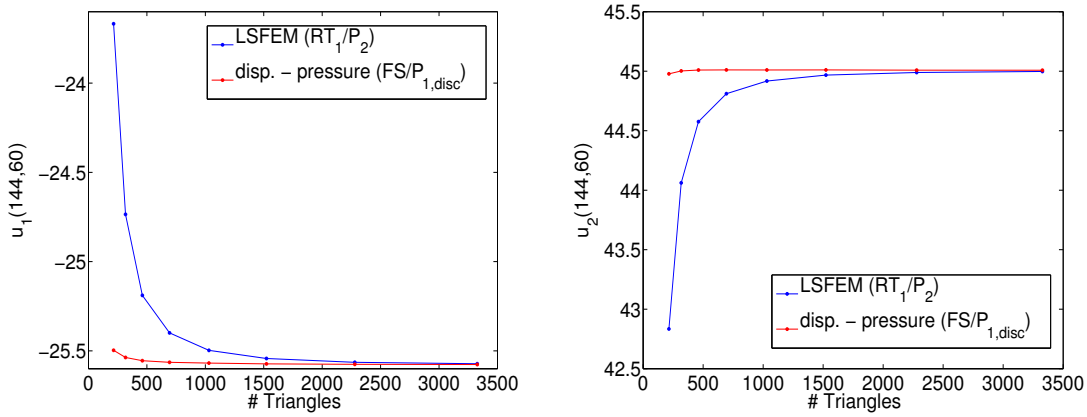
Figure 6.17: Comparison of displacements using different scaling parameters ω_1

As a result we get that $\omega_1 = 10^3$ is the best choice, since a further increase of this parameter does not lead to essentially better approximations and a very large scaling parameter ω_1 would lead to singular matrices in the algorithm. Note that for a rescaled Cook membrane problem with vertices $(0, 0)$, $(0.144, 0.044)$, $(0.144, 0.06)$ and $(0, 0.044)$, i.e. scaling factor $\frac{1}{1000}$ in the domain Ω, a scaling parameter $\omega_1 = 10^0 = 1$ is sufficient (cp. Table 6.13).

Level l	n_t	$u_1(0.144, 0.06)$		$u_2(0.144, 0.06)$	
		$\omega_1 = 10^0$	$\omega_1 = 10^1$	$\omega_1 = 10^0$	$\omega_1 = 10^1$
0	72	$-2.1814 \cdot 10^{-2}$	$-2.1822 \cdot 10^{-2}$	$4.0659 \cdot 10^{-2}$	$4.0670 \cdot 10^{-2}$
1	108	$-2.3781 \cdot 10^{-2}$	$-2.3785 \cdot 10^{-2}$	$4.3082 \cdot 10^{-2}$	$4.3087 \cdot 10^{-2}$
2	157	$-2.4696 \cdot 10^{-2}$	$-2.4698 \cdot 10^{-2}$	$4.4073 \cdot 10^{-2}$	$4.4076 \cdot 10^{-2}$
3	257	$-2.5146 \cdot 10^{-2}$	$-2.5147 \cdot 10^{-2}$	$4.4559 \cdot 10^{-2}$	$4.4560 \cdot 10^{-2}$
4	398	$-2.5372 \cdot 10^{-2}$	$-2.5372 \cdot 10^{-2}$	$4.4794 \cdot 10^{-2}$	$4.4794 \cdot 10^{-2}$
5	610	$-2.5475 \cdot 10^{-2}$	$-2.5475 \cdot 10^{-2}$	$4.4907 \cdot 10^{-2}$	$4.4908 \cdot 10^{-2}$
6	937	$-2.5533 \cdot 10^{-2}$	$-2.5533 \cdot 10^{-2}$	$4.4964 \cdot 10^{-2}$	$4.4964 \cdot 10^{-2}$
7	1416	$-2.5558 \cdot 10^{-2}$	$-2.5558 \cdot 10^{-2}$	$4.4987 \cdot 10^{-2}$	$4.4987 \cdot 10^{-2}$

Table 6.13: Comp. of displacements for a rescaled Cook's membrane using different ω_1

This means that the necessary value of the scaling parameter ω_1 depends on the size of the domain respectively the used physical unit of the problem.
Let us go back to the problem with triple length. In Figure 6.18 a comparison of the displacements in the node $(144, 60)$ between the LSFEM approach, choosing $\omega_1 = 10^3$ and using



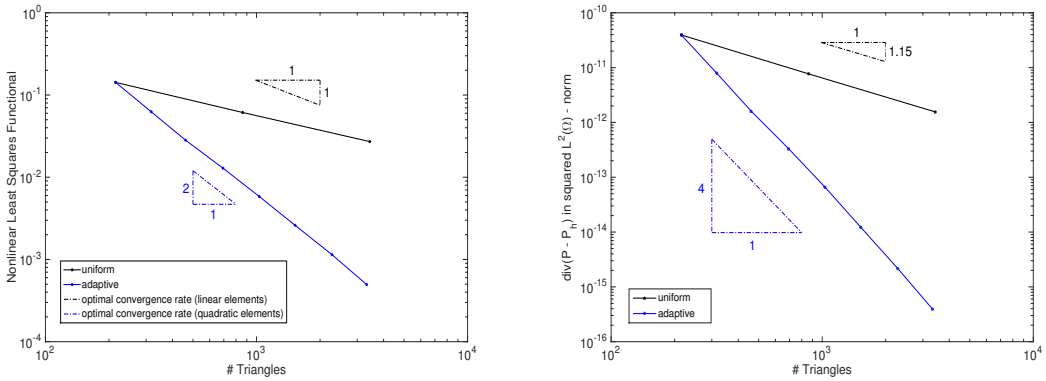
(a) horizontal displacement in $(144, 60)$ (b) vertical displacement in $(144, 60)$

Figure 6.18: Comparison of LSFEM and displacement - pressure approach (triple Cook)

adaptive refinement, and the displacement - pressure approach can be observed. Both approaches converge to the same displacements ($u_1(144, 60) \approx -25.58$, $u_2(144, 60) \approx 45.01$) and, similar to the previous examples, the LSFEM approach is inferior concerning displacement approximations.

In what follows also the other obtained results of the examples in Sections 6.1.1 and 6.1.2 can be confirmed for this bending dominated problem:

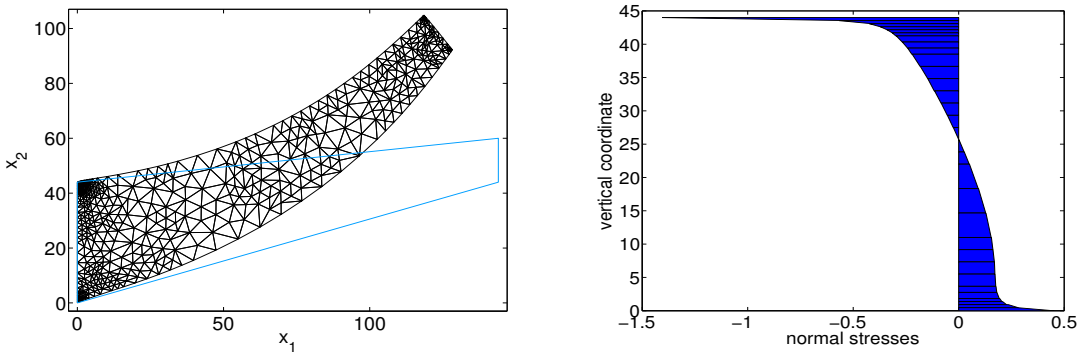
In Figure 6.19 a graphical comparison of adaptive and uniform refinement can be regarded. More precisely the values $\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$ (left) and $\|\mathrm{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)}$ (right) are plotted against the number of elements for different levels. Also for this problem one observes a numerically obtained convergence order of $\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$ close to the optimal one using adaptive refinement. Uniform refinement is worse. Moreover, one observes again an im-



(a) $\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$ (b) $\|\mathrm{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)}$

Figure 6.19: Comp. of adaptive and uniform refinement (triple Cook, incompressible Neo-Hooke)

proved convergence rate for the momentum term in the $L^2(\Omega)$-norm. One remarkable observation can be made in Figure 6.20. One evidently obtains a second singularity at the origin, since also in this vertex strong local refinement is performed. These results correspond again to the fourth level using adaptive refinement. The corresponding triangulation has 1031 triangles.
In Figure 6.21 the first two rows of the Kirchhoff stress tensor approximation can be found. Also in this example the nondiagonal components look equal in contrast to the nondia-



(a) Deformed configuration

(b) Normal stresses on Γ_D

Figure 6.20: Results in level 4 with adaptive LSFEM (triple Cook, incompressible Neo-Hooke)

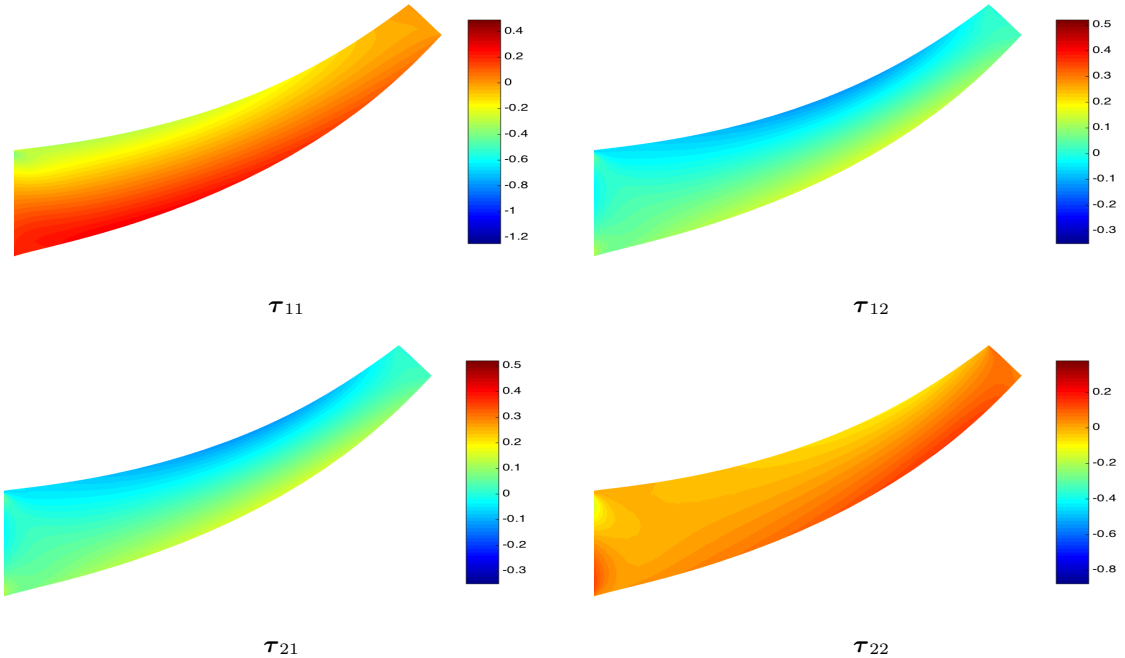
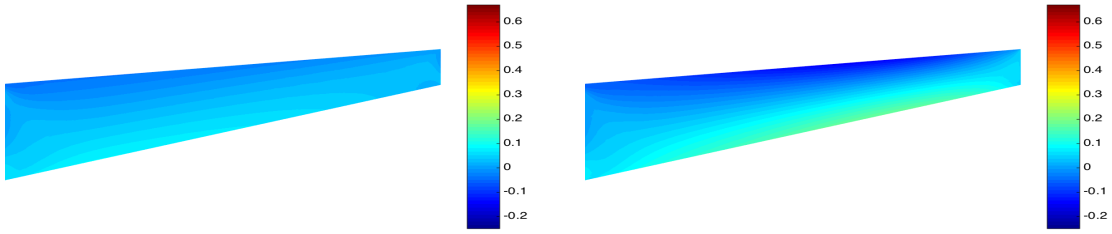


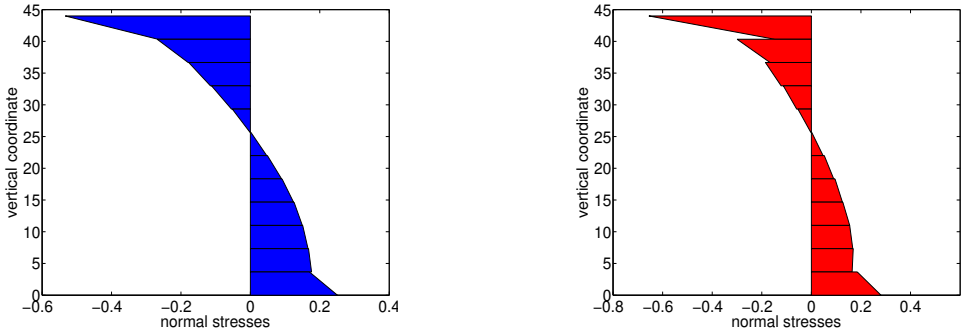
Figure 6.21: Components of the Kirchhoff stress (triple Cook, incompressible Neo-Hooke)



$\mathbf{P}_{12}$ $\mathbf{P}_{21}$

Figure 6.22: Nondiagonal components of $\mathbf{P}$ (triple Cook, incompressible Neo - Hooke)

gonal components of the first Piola - Kirchhoff stress tensor (cf. Figure 6.22).
At the end of this example we briefly compare the normal stress approximations of our LSFEM approach and the displacement - pressure approach in the first level using uniform refinement (cf. Figure 6.23). The discontinuities at the edge interfaces are more pronounced in this example in comparison to the other two examples in Sections 6.1.1 and 6.1.2.



LSFEM ($\mathcal{RT}_1/\mathcal{P}_2$) displacement - pressure (FS/$\mathcal{P}_{1,\text{disc}}$)

Figure 6.23: Normal stresses on Γ_D (level 1, triple Cook, incompressible Neo - Hooke)

Level l	n_t	LSFEM ($\mathcal{RT}_1/\mathcal{P}_2$)		displacement - pressure (FS/$\mathcal{P}_{1,\text{disc}}$)	
		Val_1	Val_2	Val_1	Val_2
0	215	$1.4163 \cdot 10^{-4}$	$7.9967 \cdot 10^{-1}$	$1.7869 \cdot 10^{-2}$	$7.5990 \cdot 10^{-1}$
1	860	$6.5348 \cdot 10^{-5}$	$7.9985 \cdot 10^{-1}$	$1.3127 \cdot 10^{-2}$	$7.7253 \cdot 10^{-1}$
2	3440	$3.0344 \cdot 10^{-5}$	$7.9993 \cdot 10^{-1}$	$9.1529 \cdot 10^{-3}$	$7.7478 \cdot 10^{-1}$
3	13760	$1.3936 \cdot 10^{-5}$	$7.9997 \cdot 10^{-1}$	$7.0386 \cdot 10^{-3}$	$7.7194 \cdot 10^{-1}$
4	55040	$6.2226 \cdot 10^{-6}$	$7.9999 \cdot 10^{-1}$	$6.8604 \cdot 10^{-3}$	$7.6614 \cdot 10^{-1}$

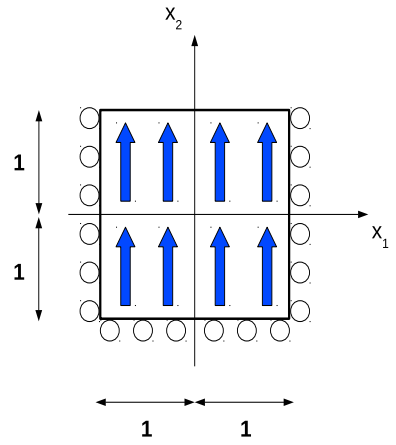
Table 6.14: Comparison of boundary stress approximations (triple Cook, incompressible)

In Table 6.14 the boundary stress integrals Val_1 and Val_2 can be again compared for the LSFEM and the displacement - pressure method using uniform refined meshes. The exact

values for the triple Cook membrane problem are $\text{Val}_1 = 0$ and $\text{Val}_2 = 0.05 \cdot 16 = 8 \cdot 10^{-1}$. Altogether we obtain similar results as in the previous considered examples concerning the boundary integral stress approximations.

6.1.4 Calculation of critical loads

In this example we consider the problem illustrated in Figure 6.24. The reference configuration is given by the unit square $\Omega := (-1, 1)^2$ where an uniform body load with density $\mathbf{f} = (0, \gamma^{\text{load}}, 0)^T$, $\gamma^{\text{load}} \in \mathbb{R}$, is applied. The boundary Γ_D contains the boundary segments on the left, right and bottom part where the boundary conditions $\mathbf{u} \cdot \mathbf{n} = 0$ and $(\mathbf{P} \cdot \mathbf{n}) \cdot \mathbf{t} = 0$ are prescribed. Here $\mathbf{n}$ denotes again the outer normal and $\mathbf{t}$ a tangential



constraints on Γ_D (left, right, bottom):

- $\mathbf{u} \cdot \mathbf{n} = 0$
- $(\mathbf{P} \cdot \mathbf{n}) \cdot \mathbf{t} = 0$

constraints on Γ_N (top):

- $\mathbf{P} \cdot \mathbf{n} = \mathbf{0}$

exact solution (for $\lambda \to \infty$):

- $\mathbf{P}(x_1, x_2) = \gamma^{\text{load}}(1 - x_2)\mathbf{I}$, $\mathbf{u} \equiv \mathbf{0}$

Figure 6.24: Problem description for the calculation of critical loads

vector. On the remaining boundary part Γ_N (top) no traction forces are applied, i.e. we prescribe $\mathbf{P} \cdot \mathbf{n} = \mathbf{0}$ on Γ_N. As Lamé constants we set $\lambda = \infty$ and $\mu = 1$, i.e. we consider again a fully incompressible material. As scaling parameters we have chosen $\omega_1 = 1 = \omega_2$. Note that this problem was already considered in [ABadVLR10]. The aim of this problem is to detect so - called critical load values, i.e. load values where the uniqueness of the solution is lost. In this manner one also speaks about bifurcation points.
Before we present the numerical results we state some preliminary considerations:
Firstly the solution of the problem is given by $\mathbf{u} \equiv \mathbf{0}$ and $\mathbf{P}(x_1, x_2) = \gamma^{\text{load}}(1 - x_2)\mathbf{I}$. This pair $(\mathbf{P}, \mathbf{u})$ obviously satisfies the prescribed boundary conditions. To verify this solution we have to show additionally that it holds $\mathcal{R}_{NH}(\mathbf{P}, \mathbf{u}) = \mathbf{0}$ with $\mathcal{R}_{NH}$ defined by the left equation in (3.18) and using the Neo - Hooke model. Obviously it holds on the one hand

$$\operatorname{div} \mathbf{P} = \gamma^{\text{load}} \cdot \nabla(1 - x_2) = \begin{pmatrix} 0 \\ -\gamma^{\text{load}} \\ 0 \end{pmatrix},$$

i.e. $\operatorname{div} \mathbf{P} + \mathbf{f} = \mathbf{0}$. On the other hand we obtain $\mathbf{F}(\mathbf{u}) = \mathbf{I}$ and therefore

$$\mathbf{dev}\, \boldsymbol{\tau} = \mathbf{dev}\, \left(\mathbf{P}\mathbf{F}(\mathbf{u})^T\right) = \mathbf{dev}\, \mathbf{P} = \mathbf{0}$$

for the Kirchhoff stress tensor $\boldsymbol{\tau} = \mathbf{P}\mathbf{F}(\mathbf{u})^T$ in Ω.
Inserting $\mathbf{dev}\, \boldsymbol{\tau} = \mathbf{0}$ into (3.39) with $\lambda = \infty$ results in the coefficients $S = 0$ and $T = -27$ of the cubic equation (3.40). One obtains its discriminant as $D = \left(\frac{S}{3}\right)^3 + \left(\frac{T}{2}\right)^2 = \left(\frac{27}{2}\right)^2$ and therefore (cf. (3.41))

$$\begin{aligned}\operatorname{tr}\left(\mathcal{A}_{NH}(\boldsymbol{\tau})\right) &= \sqrt[3]{-\frac{T}{2} + \sqrt{D}} + \sqrt[3]{-\frac{T}{2} - \sqrt{D}} = \sqrt[3]{\frac{27}{2} + \frac{27}{2}} + \sqrt[3]{\frac{27}{2} - \frac{27}{2}} \\ &= \sqrt[3]{\frac{54}{2}} = \sqrt[3]{27} = 3.\end{aligned}$$

Thus we obtain the strain tensor $\mathcal{A}_{NH}(\boldsymbol{\tau}) = \frac{\mathbf{dev}\, \boldsymbol{\tau}}{\mu} + \frac{1}{3} \operatorname{tr}\left(\mathcal{A}_{NH}(\boldsymbol{\tau})\right) \mathbf{I} = \mathbf{0} + \frac{1}{3} \cdot 3 \cdot \mathbf{I} = \mathbf{I}$. Due to $\mathbf{F}(\mathbf{u}) = \mathbf{I}$ it holds $\mathbf{B}(\mathbf{u}) = \mathbf{I}$ and hence the equation $\mathcal{A}_{NH}\left(\mathbf{P}\mathbf{F}(\mathbf{u})^T\right) - \mathbf{B}(\mathbf{u}) = \mathbf{0}$ is confirmed. Altogether we have shown that the pair $(\mathbf{P}, \mathbf{u})$ is indeed a solution of $\mathcal{R}_{NH}(\mathbf{P}, \mathbf{u}) = \mathbf{0}$.
Secondly we observe that the right hand side of the occurring linear systems in (3.26) in the Gauss - Newton iteration is zero inserting the exact solution $(\mathbf{P}, \mathbf{u})$ as $\left(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)}\right)$ (cf. (3.25)). This means that for this choice we get linear systems of equations of the form $\mathbf{A}^{(k)}\mathbf{x}^{(k)} = \mathbf{0}$. As long as the stiffness matrices $\mathbf{A}^{(k)}$ are positive definite the solution is $\mathbf{x}^{(k)} = \mathbf{0}$ and the new iteration would be $\left(\mathbf{P}_h^{(k+1)}, \mathbf{u}_h^{(k+1)}\right) = \left(\mathbf{P}_h^{(k)}, \mathbf{u}_h^{(k)}\right) = (\mathbf{P}, \mathbf{u})$, i.e. if one would use the Gauss - Newton scheme with exact solution as initial guess, the new solution remains the exact solution.
A difficulty occurs if the linear systems of equations $\mathbf{A}^{(k)}\mathbf{x}^{(k)} = \mathbf{0}$ has a solution $\mathbf{x}^{(k)} \neq \mathbf{0}$. This is possible if and only if the stiffness matrix has at least one zero eigenvalue or equivalently the matrix is singular. Note, that in this case the coerciveness property of the bilinear form is no longer satisfied (cf. Section 3.3.3), the problem loses its stability and a second solution unequal to $(\mathbf{P}, \mathbf{u})$ occurs.
In Figure 6.25 the smallest eigenvalue of the stiffness matrix $\mathbf{A}$ with components $A_{ij} = (\mathcal{R}'_{NH}(\mathbf{P}, \mathbf{u})[\boldsymbol{\Phi}_j], \mathcal{R}'_{NH}(\mathbf{P}, \mathbf{u})[\boldsymbol{\Phi}_i])_{L^2(\Omega)}$, where $\boldsymbol{\Phi}_j$, $j = 1, \ldots, N$, denote basis functions to the space $\boldsymbol{\Pi}_h \times \mathbf{U}_h$ with dimension $N := \dim\left(\boldsymbol{\Pi}_h \times \mathbf{U}_h\right)$ (cf. Section 3.3.3), is plotted against the load value γ^{load}. In this example γ^{load} varies between 0 and 8 choosing a load step size of 0.1. Furthermore three different triangulations were used.
One observes that the first zero eigenvalue which tends to zero as the mesh size decreases occurs between a load value of 3.1 and 3.3. This is the first critical load value. The second critical load value occurs between 6.2 and 6.4 (cp. also the values in Tables 6.15 and 6.16). If we zoom into the intervals $[3.1, 3.3]$ and $[6.2, 6.4]$ (cf. Figure 6.26) we can specify the critical load values as approximately 3.23 and 6.28. The first critical load value approximation 3.23 is identical to the theoretically obtained value in [ABadVLR10].

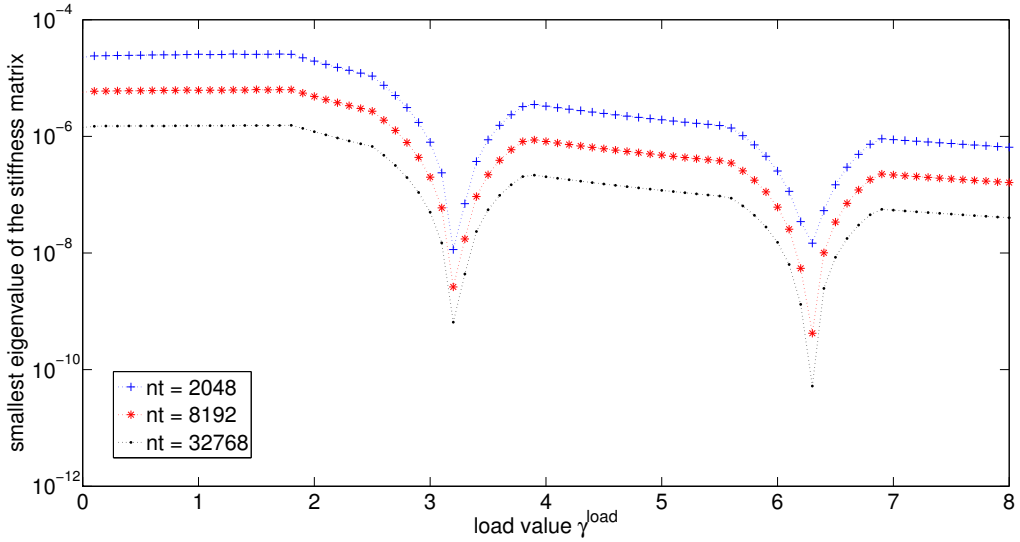


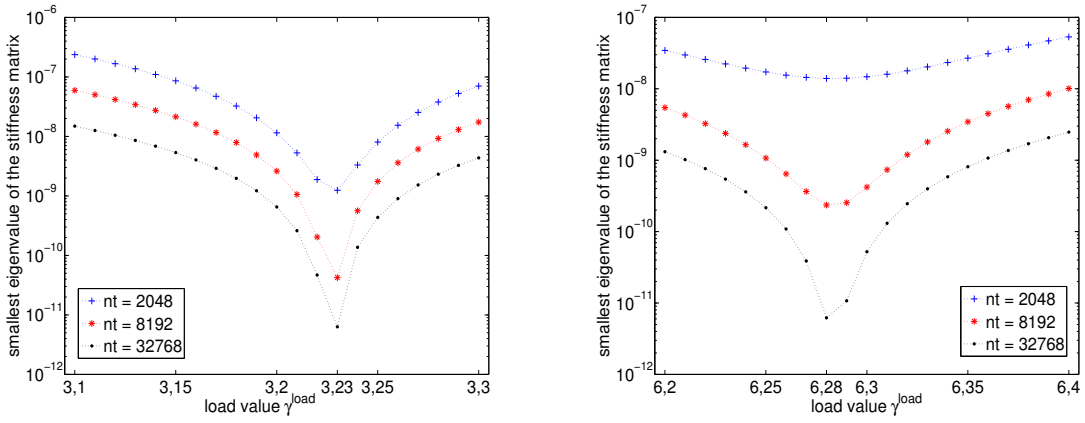
Figure 6.25: Identification of critical load values

γ^{load}	3.0	3.1	3.2	3.3	3.4	3.5
λ_1	$5.0013 \cdot 10^{-8}$	$1.4883 \cdot 10^{-8}$	$6.5141 \cdot 10^{-10}$	$4.3592 \cdot 10^{-9}$	$2.3340 \cdot 10^{-8}$	$5.5134 \cdot 10^{-8}$

Table 6.15: Smallest eigenvalue λ_1 of stiffness matrix ($n_t = 32768$, $\gamma^{\text{load}} \in [3, 3.5]$)

γ^{load}	6.0	6.1	6.2	6.3	6.4	6.5
λ_1	$1.5235 \cdot 10^{-8}$	$6.3616 \cdot 10^{-9}$	$1.3152 \cdot 10^{-9}$	$5.2258 \cdot 10^{-11}$	$2.4738 \cdot 10^{-9}$	$8.4330 \cdot 10^{-9}$

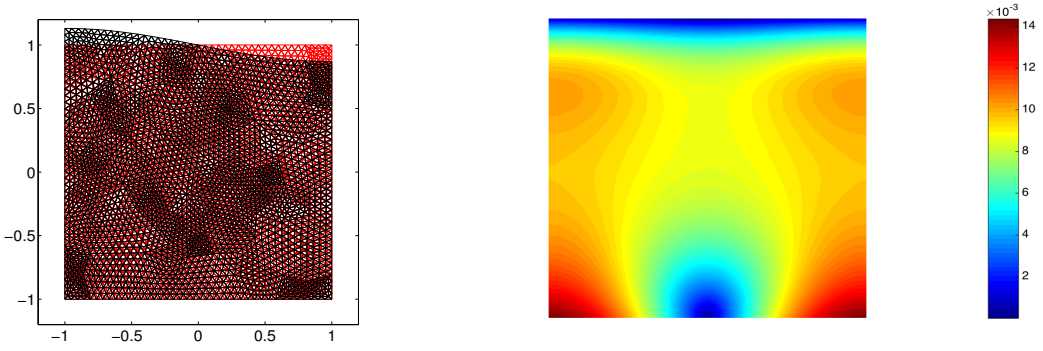
Table 6.16: Smallest eigenvalue λ_1 of stiffness matrix ($n_t = 32768$, $\gamma^{\text{load}} \in [6, 6.5]$)



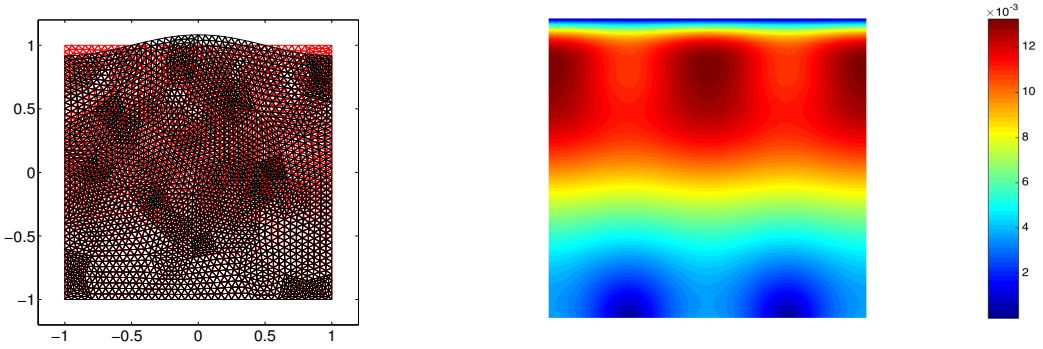
$\gamma^{\text{load}} \in [3.1, 3.3]$ $\gamma^{\text{load}} \in [6.2, 6.4]$

Figure 6.26: Zoom into critical intervals

In Figure 6.27 the displacement and stress eigenfunctions corresponding to the smallest eigenvalue of the critical loads are plotted (first row: $\gamma^{\text{load}} = 3.23$, second row: $\gamma^{\text{load}} = 6.28$). On the left side of this figure the displacement eigenfunctions can be regarded. One can observe that a second displacement solution (black mesh) occurs which is obviously unequal to the zero displacement solution (red mesh). On the right side of this figure the stress eigenfunctions are plotted for both critical load values.
We can conclude that our least squares approach provides very good approximations of the exact critical load values in this example.



displacement eigenfunction ($\gamma^{\text{load}} = 3.23$) stress eigenfunction ($\gamma^{\text{load}} = 3.23$)



displacement eigenfunction ($\gamma^{\text{load}} = 6.28$) stress eigenfunction ($\gamma^{\text{load}} = 6.28$)

Figure 6.27: Eigenfunctions to $\gamma^{\text{load}} = 3.23$ (1st row) and $\gamma^{\text{load}} = 6.28$ (2nd row)

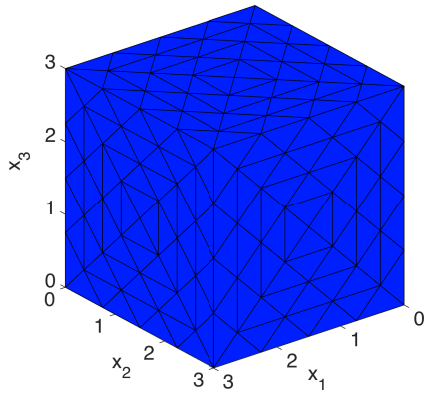
6.2 Three - dimensional problems for isotropic materials

The aim of this part of the work is to show that our proposed least squares finite element method works also for three - dimensional problems. In this subsection we will consider two examples for isotropic materials.
In general we use the space $\mathbf{\Pi}_h := (\mathcal{RT}_1(\mathcal{T}_h))^3$ for the stress approximations and $\mathbf{U}_h = (\mathcal{P}_2(\mathcal{T}_h))^3$ for the displacement approximations in our three - dimensional LSFEM simulations.
In Algorithm 1 we use the tolerance $tol = 10^{-6}$ inside the proposed stopping criterion

and $i_{\max} = 20$ as maximal number of Gauss - Newton steps. As initial solution for the Gauss - Newton scheme we use $\left(\mathbf{P}_h^{(0)}, \mathbf{u}_h^{(0)}\right) = \left(\mathbf{P}^N, \mathbf{u}_D\right)$.

6.2.1 Uniaxial tension test with compressible Mooney - Rivlin

The first example is a quite simple problem. We consider an uniaxial tension test on the cube $\Omega = (0,3)^3$ with fixed triangulation into $n_t = 2816$ tetrahedra and compressible isotropic material behavior. As material parameters we choose $E = 200$, $\nu = 0.35$ and as scaling parameters in (3.18) we use $\omega_1 = 1 = \omega_2$. The aim of this example is to compare different models using our proposed least squares finite element method.
Three different models are taken into account: The first one is the model of linear elasticity. The second one is the nonlinear Neo - Hooke model and the third one is the nonlinear Mooney - Rivlin model. Recall that the Neo - Hooke model is exactly the Mooney - Rivlin model for the choice $\delta = 0$ in (2.30).
Furthermore we compare the obtained results with the pure displacement approach that we have introduced in Section 3.6.1 as reference method for compressible materials. In this approach we use continuous piecewise quadratic elements $(\mathcal{P}_2(\mathcal{T}_h))^3$ for the approximation of $\mathbf{u}$.
The whole description of the problem is depicted in Figure 6.28. As force densities we



constraints on Γ_D (right, back, bottom):

- $\mathbf{u} \cdot \mathbf{n} = 0, (\mathbf{P} \cdot \mathbf{n}) \cdot \mathbf{t} = 0$

constraints on Γ_N:

- $\mathbf{P} \cdot \mathbf{n} = \mathbf{0}$ (front, top)
- $\mathbf{P} \cdot \mathbf{n} = \mathbf{g}$ (left)

Figure 6.28: Problem description of an uniaxial tension test in 3d

use $\mathbf{f} \equiv \mathbf{0}$, $\mathbf{g} = (\gamma^{\text{load}}, 0, 0)^T$ with $\gamma^{\text{load}} \in \mathbb{R}$. The boundary $\Gamma = \partial\Omega$ is divided into Γ_D, consisting of the right ($x_1 = 0$), back ($x_2 = 0$) and bottom ($x_3 = 0$) lateral face, and Γ_N, consisting of the left ($x_1 = 3$), front ($x_2 = 3$) and top ($x_3 = 3$) lateral face. The constraint $\mathbf{u} \cdot \mathbf{n} = 0$ and $(\mathbf{P} \cdot \mathbf{n}) \cdot \mathbf{t} = 0$ on Γ_D is equivalent to the boundary conditions $u_1 = 0$, $P_{21} = 0 = P_{31}$ on the right, $u_2 = 0$, $P_{12} = 0 = P_{32}$ on the back and $u_3 = 0$, $P_{13} = 0 = P_{23}$ on the bottom. On the part Γ_N we prescribe traction forces: On the front and top part we specify them to $\mathbf{P} \cdot \mathbf{n} = \mathbf{0}$, i.e. traction - free boundary conditions. On the left part we

apply a traction force in x_1 - direction with load parameter $\gamma^{\text{load}} \in \mathbb{R}$, according to the given force density $\mathbf{g}$. In Figure 6.29 a comparison using different models is illustrated: The displacement in x_1 - direction of the point $(3, 3, 3)$ is plotted against different load values $\gamma^{\text{load}} \in \{-100, -75, -50, -25, 25, 50, 75, 100\}$. Different values $\delta \in \{0, 10, 15, 20, 25\}$ in the stored energy function (2.30) of the Mooney - Rivlin model are taken into account. For each model one observes consistency with the model of linear elasticity, as expected by Section 2.4.5. Furthermore one notes that the displacements concerning the different nonlinear models for negative loads differ only slightly. In contrast the difference of displacements in the nonlinear models for larger positive load values is much more pronounced. For instance for $\gamma^{\text{load}} = 100$ the displacement for the Mooney - Rivlin model with $\delta = 25$ in the considered point is more than twice as large as the corresponding displacement for the Neo - Hooke model. With this in mind it is for example possible to fit hyperelastic models to given experimental data such that the theoretical model matches better with the physical experiment.

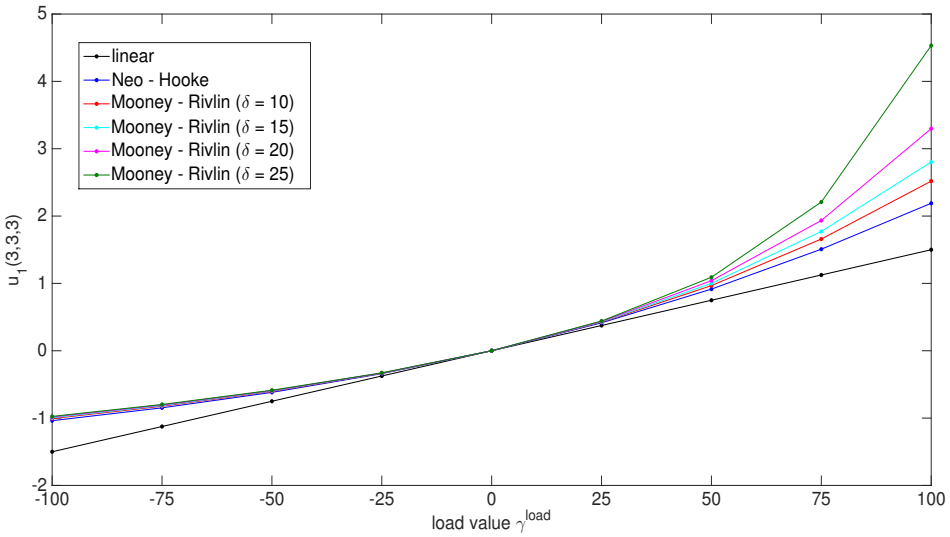


Figure 6.29: Comparison of different models (3d uniaxial tension test)

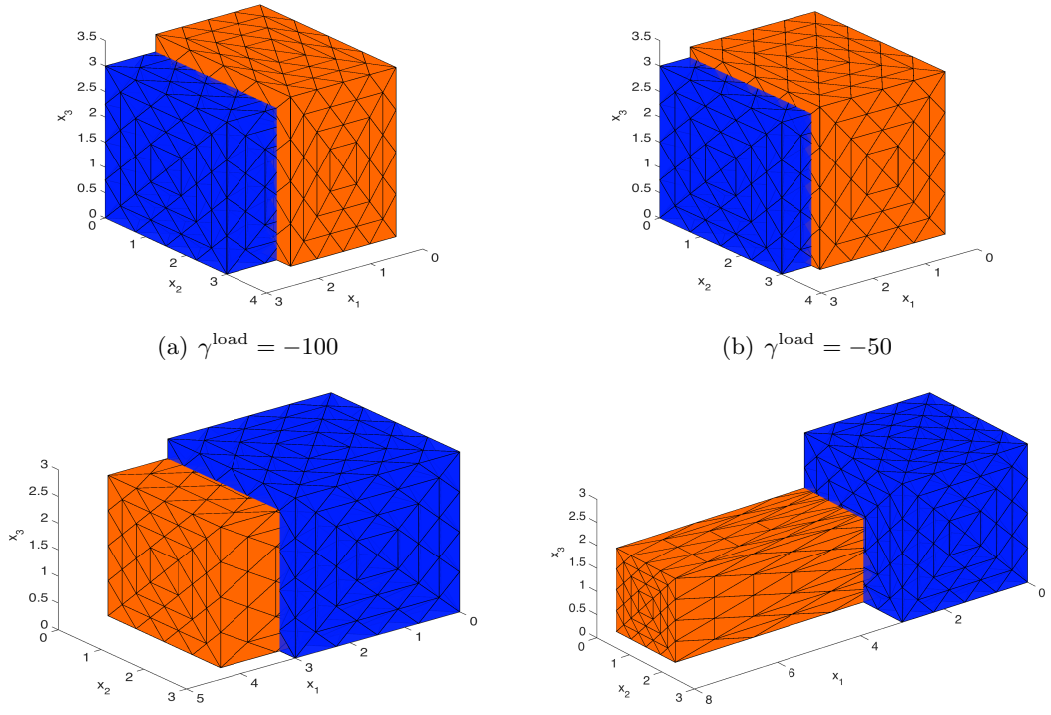
The displacement approximations plotted in Figure 6.29 can be found in more detail in Table 6.17. In this table the abbreviations „MR“ for the Mooney - Rivlin model and „linear“ for the model of linear elasticity are used. In the table one can observe quantitatively that the displacements increase in each row from left to right, i.e. for a fixed load value γ^{load} the displacements increase if one increment the value of δ in the model. Moreover, for a fixed model the displacement values increase columnwise.
For this problem we have also checked if the displacement values in Table 6.17 are reasonable. For this purpose we have compared the displacement approximations of $(3, 3, 3)$ in x_1 - direction obtained with the pure displacement approach with the values in Table

γ^{load}	linear	MR ($\delta = 0$)	MR ($\delta = 10$)	MR ($\delta = 15$)	MR ($\delta = 20$)	MR ($\delta = 25$)
-100	-1.5000	-1.0384	-1.0073	-0.9951	-0.9843	-0.9749
-75	-1.1250	-0.8475	-0.8239	-0.8142	-0.8056	-0.7978
-50	-0.7500	-0.6181	-0.6038	-0.5976	-0.5919	-0.5867
-25	-0.3750	-0.3398	-0.3348	-0.3325	-0.3304	-0.3283
25	0.3750	0.4146	0.4241	0.4294	0.4352	0.4414
50	0.7500	0.9154	0.9670	1.0000	1.0403	1.0915
75	1.1250	1.5082	1.6590	1.7716	1.9346	2.2089
100	1.5000	2.1901	2.5185	2.8023	3.2974	4.5319

Table 6.17: Displacements $u_1(3,3,3)$ for different models and load values γ^{load}

6.17. One obtains the same approximations up to a tolerance 10^{-7} for all the considered models (linear, Mooney-Rivlin with $\delta \in \{0, 10, 15, 20, 25\}$) and all considered load values $\gamma^{\text{load}} \in \{-100, -75, -50, -25, 25, 50, 75, 100\}$. Thus we can state that both discretization schemes for this simple problem lead to the same displacement approximations of $u_1(3,3,3)$. Hence, our proposed least squares finite element method yields reasonable results.

For the special choice of $\delta = 25$ four different deformed configurations, corresponding to the load values $\gamma^{\text{load}} \in \{-100, -50, 50, 100\}$, are depicted in Figure 6.30.



(a) $\gamma^{\text{load}} = -100$ (b) $\gamma^{\text{load}} = -50$

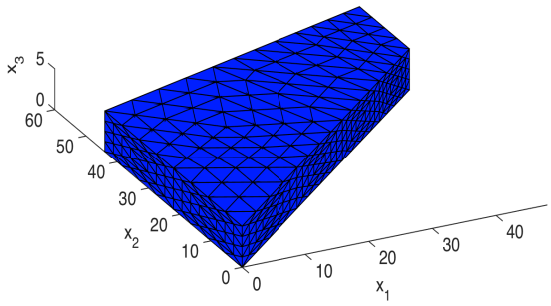
(c) $\gamma^{\text{load}} = 50$ (d) $\gamma^{\text{load}} = 100$

Figure 6.30: Deformed configuration for different loads (orange) and reference configuration (blue) (Mooney-Rivlin model with $\delta = 25$, $n_t = 2816$)

In this figure the reference configuration is plotted in blue and the deformed configuration for each load is plotted in orange. One observes a behavior which one expects due to the problem description in Figure 6.28. The satisfaction of the boundary condition $\mathbf{u} \cdot \mathbf{n} = 0$ on Γ_D is visible in all these plots.
Altogether we can confirm that our proposed least squares finite element method (for the $\mathbf{B}$ - formulation) works also in three dimensions. Moreover, it also works for more complex material models than Neo - Hooke. Recall that for the Mooney - Rivlin model with $\delta \neq 0$ a Newton scheme for the evaluation of $\mathcal{A}_{MR}\left(\mathbf{PF}(\mathbf{u})^T\right)$ is necessary. For such a model the simulation of quasi - incompressible materials $\left(\nu \text{ close to } \frac{1}{2}\right)$ is also possible whereas the fully incompressible case $\nu = \frac{1}{2}$ is not possible without further efforts.

6.2.2 Cook's membrane with incompressible Neo - Hooke and adaptive refinement

In the second three - dimensional example we extend the Cook membrane problem from two dimensions. For this purpose we use the domain of Cook's membrane (cf. Figure 6.1) in two dimensions as base area and expand it in x_3 - direction with thickness $d := 5$. The corresponding reference configuration and prescribed boundary conditions for this problem are summarized in Figure 6.31.



constraints on Γ_D:

- $\mathbf{u} = \mathbf{0}$ (left)

constraints on Γ_N:

- $\mathbf{P} \cdot \mathbf{n} = \mathbf{0}$ (top, bottom, front, back)
- $\mathbf{P} \cdot \mathbf{n} = \mathbf{g}$ (right)

Figure 6.31: Problem description of Cook's membrane in three dimensions

The boundary $\Gamma = \partial\Omega$ is splitted into the left lateral face $\Gamma_D := \{(0, x_2, x_3) : 0 < x_2 < 44, 0 < x_3 < d\}$ and Γ_N consisting of the remaining five lateral faces. We clamp the body on Γ_D and apply a surface force $\mathbf{g} = (0, \gamma^{\text{load}}, 0)^T$ with load value $\gamma^{\text{load}} \in \mathbb{R}$ on the right part of the boundary $\Gamma_R := \{(48, x_2, x_3) : 44 < x_2 < 60, 0 < x_3 < d\}$. On the other parts of Γ_N no surface forces act. As body force density we use again $\mathbf{f} = \mathbf{0}$. For the concrete example below we have chosen $\gamma^{\text{load}} = 0.05$, Lamé constants $\mu = 1$, $\lambda = \infty$, i.e. we consider a fully incompressible material, and scaling parameters $\omega_1 = 10^2$, $\omega_2 = 1$.
For the marking of elements in adaptive simulations we use now the Dörfler marking strategy (cp. Appendix C.2) which is in general superior in comparison to the percent marking

strategy, since not only a fixed number of elements is marked but also a particular rate of the total error is taken into account.
In Tables 6.18 and 6.19 numerical obtained convergence rates corresponding to $\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$ and $\|\mathrm{div}(\mathbf{P}-\mathbf{P}_h)\|^2_{L^2(\Omega)}$, using adaptive respectively uniform refinement, can be compared. Note that the initial triangulation is quite coarse with $n_t = 880$ tetrahedra and recall that l denotes the refinement level.

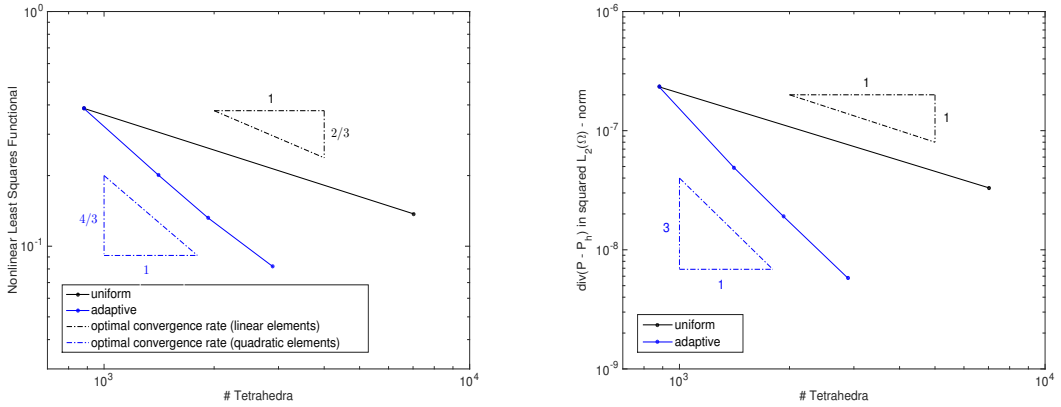
l	n_t	$\dim \mathbf{\Pi}_h$	$\dim \mathbf{U}_h$	$\sigma_{\text{dörf}}$	$\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$	(order)	$\|\mathrm{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)}$	(order)
0	880	22968	4104		$3.8682 \cdot 10^{-1}$		$2.3313 \cdot 10^{-7}$	
1	1410	37161	6321	0.800	$2.0062 \cdot 10^{-1}$	(1.393)	$4.8949 \cdot 10^{-8}$	(3.311)
2	1928	50859	8607	0.650	$1.3179 \cdot 10^{-1}$	(1.343)	$1.8969 \cdot 10^{-8}$	(3.030)
3	2892	76734	12576	0.450	$8.1998 \cdot 10^{-2}$	(1.170)	$5.7679 \cdot 10^{-9}$	(2.936)

Table 6.18: Convergence rates with adaptive refinement I (incompressible Neo - Hooke, 3d)

l	n_t	$\dim \mathbf{\Pi}_h$	$\dim \mathbf{U}_h$	$\sigma_{\text{dörf}}$	$\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$	(order)	$\|\mathrm{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)}$	(order)
0	880	22968	4104		$3.8682 \cdot 10^{-1}$		$2.3313 \cdot 10^{-7}$	
1	7040	186912	30384	1.000	$1.3719 \cdot 10^{-1}$	(0.498)	$3.3031 \cdot 10^{-8}$	(0.940)

Table 6.19: Convergence rates with uniform refinement (incompressible Neo - Hooke, 3d)

One observes in Table 6.18 that we obtain good convergence rates, close to the optimal value $\frac{4}{3}$, for the nonlinear functional using adaptive refinement. Moreover we see, similar as in the two - dimensional examples, that the convergence rates to the balance of momentum is greater than for the nonlinear functional. Here they are even more than doubled. Moreover, the value $\|\mathrm{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)}$ in each considered level is again close to zero, i.e. linear momentum is conserved quite well.
The convergence rate for $\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$ in the case of uniform refinement (see Table 6.19) is worse than the optimal convergence rate $\frac{2}{3}$ for linear elements. The convergence rate corresponding to the conservation of momentum is approximately doubled.
Altogether we can confirm the observations made in two dimensions. For the sake of completeness a visualization of the convergence rates, corresponding to the values in Tables 6.18 and 6.19, can be found in Figure 6.32.
If we have a closer look at the parameter $\sigma_{\text{dörf}}$ in the fourth column of Table 6.18 within the marking strategy of Dörfler, we see that we have reduced the parameter in each step in order to obtain good convergence rates. The question arises if such a reduction of $\sigma_{\text{dörf}}$ is always necessary or if the necessity in this example is based on the used coarse meshes. For this purpose we consider another simulation with adaptive refinement, using this time a finer initial mesh with $n_t = 7040$ tetrahedra. We observe in Table 6.20 that such a drastic reduction of $\sigma_{\text{dörf}}$ is not necessary in this example. However, a slight modification of $\sigma_{\text{dörf}}$ from level to level is also here needed in order to get convergence rates close to the optimal



(a) $\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$ (b) $\|\mathrm{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)}$

Figure 6.32: Comp. of uniform and adaptive refinement I (incompressible Neo - Hooke, 3d)

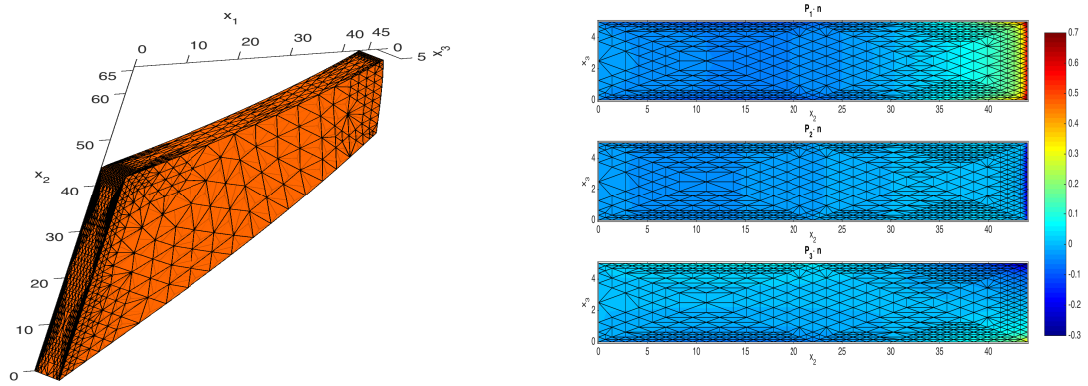
l	n_t	$\dim \mathbf{\Pi}_h$	$\dim \mathbf{U}_h$	$\sigma_{\text{dörf}}$	$\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$	(order)	$\|\mathrm{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)}$	(order)
0	7040	186912	30384		$1.5280 \cdot 10^{-1}$		$4.0382 \cdot 10^{-8}$	
1	14284	381132	60456	0.950	$5.7710 \cdot 10^{-2}$	(1.376)	$5.6971 \cdot 10^{-9}$	(2.768)
2	18628	498042	78216	0.900	$2.7646 \cdot 10^{-2}$	(2.772)	$1.1405 \cdot 10^{-9}$	(6.058)
2	23640	633231	98469	0.925	$2.5911 \cdot 10^{-2}$	(1.589)	$1.0739 \cdot 10^{-9}$	(3.312)
2	39442	1056294	164328	0.950	$2.3432 \cdot 10^{-2}$	(0.887)	$1.0000 \cdot 10^{-9}$	(1.713)

Table 6.20: Convergence rates with adaptive refinement II (incompressible Neo - Hooke, 3d)

one. In more detail: Starting from an initial triangulation (level $l = 0$) with 7040 tetrahedra we use $\sigma_{\text{dörf}} = 0.95$ in the first refinement step which results in a triangulation with 14284 tetrahedra (level $l = 1$). In this step the convergence rate 1.376 corresponding to the nonlinear functional is close to the optimal one $\frac{4}{3}$. From level $l = 1$ to $l = 2$ we have compared the results for three different parameters $\sigma_{\text{dörf}} \in \{0.9, 0.925, 0.95\}$. If one chooses $\sigma_{\text{dörf}} = 0.9$, we see that the convergence rate 2.772 is too good. Choosing $\sigma_{\text{dörf}} = 0.925$ leads to a convergence rate of 1.589 which is closer to the optimal one. If one chooses the same parameter $\sigma_{\text{dörf}}$ as in the first refinement step, i.e. $\sigma_{\text{dörf}} = 0.95$, we observe that the obtained convergence rate 0.887 becomes too bad. This means that one value $\sigma_{\text{dörf}} \in (0.925, 0.95)$ should lead to an optimal convergence rate.

The convergence rates belonging to the conservation of momentum are improved independently of the choice of $\sigma_{\text{dörf}}$ in all these case. They are approximately twice as large as the convergence rates to the nonlinear functional, similar as observed in the two - dimensional examples.

We can conclude that the numerical convergence rates are quite sensitive with respect to $\sigma_{\text{dörf}}$ and it is difficult to choose an „optimal“ parameter $\sigma_{\text{dörf}}$. In addition we have seen that if we start with a coarse mesh we must reduce the parameter $\sigma_{\text{dörf}}$ in the refinement



(a) Deformed configuration

(b) Normal component of $\mathbf{P}$ on Γ_D

Figure 6.33: Results in level 2 with adaptive refinement II (incompressible Neo - Hooke, 3d)

strategy stronger in order to get convergence rates close to the optimal one. In general, note that an optimal convergence rate is only asymptotically expectable, i.e. in fact we have to consider much more refinement steps to get a more precise statement.

On the left side of Figure 6.33 the deformed configuration with $n_t = 23640$ tetrahedra after two refinement steps, belonging to $\sigma_{\text{dörf}} = 0.925$ and the finer initial triangulation, is plotted. We see that we have a strong local refinement near the edge $\{(0, 44, x_3) : 0 < x_3 < 5\}$. This means that the point singularity in $(0, 44)$ of two dimensions becomes an edge singularity in three dimensions. Moreover, one also observes stronger local refinement at the transition of boundary conditions from hard - clamped ($\mathbf{u} = \mathbf{0}$) to stress - free ($\mathbf{P} \cdot \mathbf{n} = \mathbf{0}$) and at the right part where the surface force acts.

On the right side of Figure 6.33 the approximated normal component of $\mathbf{P}$ with outer normal $\mathbf{n} = (-1, 0, 0)^T$ is depicted on Γ_D. One can observe that the absolute values of the components P_{11} and P_{21} of $\mathbf{P}$ increase near the singularity edge. In Figure 6.34 stress approximations for each component of the Kirchhoff stress tensor $\boldsymbol{\tau}$ are illustrated on the same mesh. One observes that the approximated stress tensor is quite symmetric, according to our theory (cf. Corollary 3.31). Differences between the nondiagonal elements of $\boldsymbol{\tau}$ occur only near the singularity edge where the discretization error is large.

At the end of this example we are also interested in the resultant normal and traction forces on Γ_D, similar to Tables 6.4 and 6.10 in the two - dimensional case.

Analogously to (6.3) we obtain for an arbitrary vector $\mathbf{v} \in \mathbb{R}^3$, arbitrary load value $\gamma^{\text{load}} \in \mathbb{R}$ and the prescribed boundary conditions on Γ_N the equation

$$\int_{\Gamma_N} \mathbf{v} \cdot \mathbf{P} \cdot \mathbf{n}\, ds = \int_{\Gamma_R} \mathbf{v} \cdot \begin{pmatrix} 0 \\ \gamma^{\text{load}} \\ 0 \end{pmatrix} ds = \int_{\Gamma_R} v_2 \gamma^{\text{load}}\, ds = v_2 \gamma^{\text{load}} |\Gamma_R| = 80 v_2 \gamma^{\text{load}}, \quad (6.6)$$

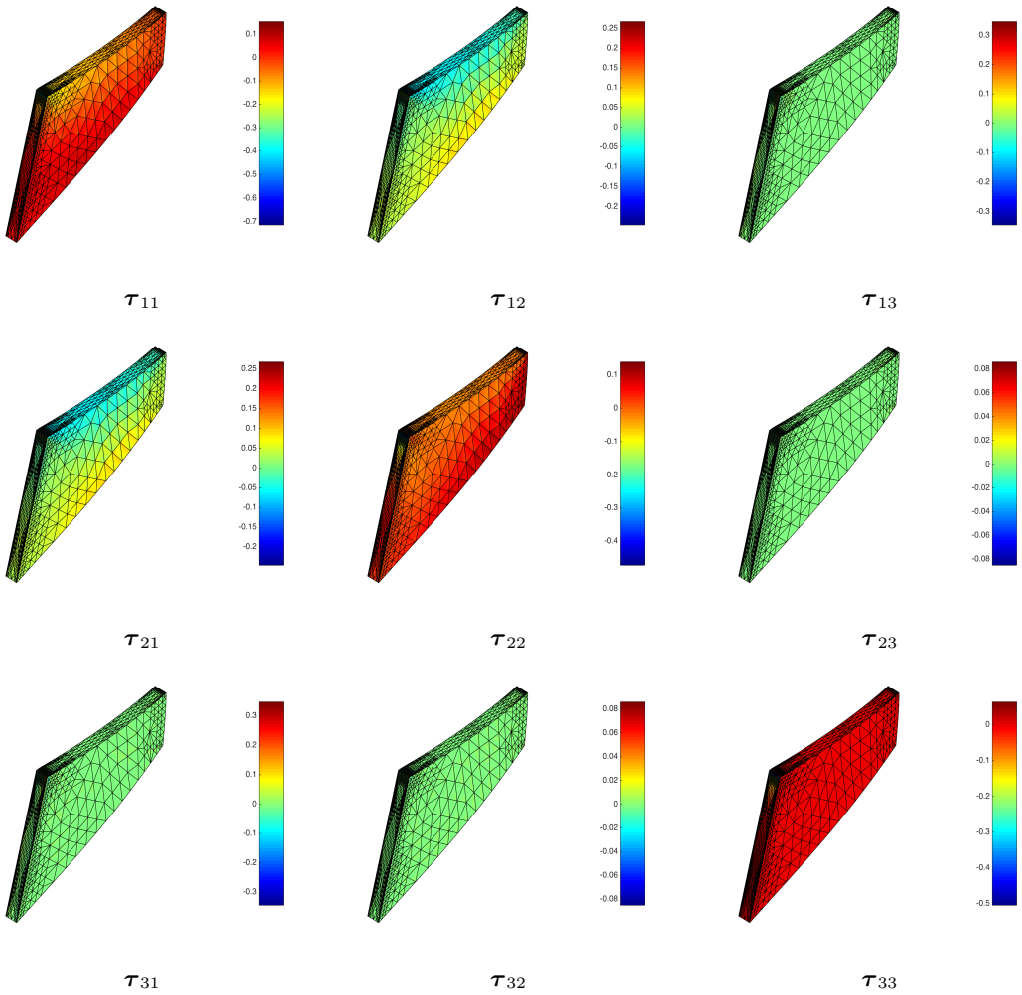


Figure 6.34: Components of the Kirchhoff stress (incompressible Neo - Hooke, 3d)

where the right part Γ_R of the boundary Γ has the area $|\Gamma_R| = 16d = 80$. Using (6.6) for the outer normal $\mathbf{v} = \mathbf{n} := (-1, 0, 0)^T$ on Γ_D, respectively the tangential vectors $\mathbf{v} = \mathbf{t}_1 := (0, 1, 0)^T$ and $\mathbf{v} = \mathbf{t}_2 := (0, 0, 1)^T$ orthogonal to $\mathbf{n}$, and combining this with (6.4) leads to

$$\begin{aligned}
\int_{\Gamma_D} P_{11}\, ds &= \int_{\Gamma_D} \mathbf{n}\cdot\mathbf{P}\cdot\mathbf{n}\, ds = -\int_{\Gamma_N} \mathbf{n}\cdot\mathbf{P}\cdot\mathbf{n}\, ds = -80\cdot 0\cdot\gamma^{\text{load}} = 0,\\
\int_{\Gamma_D} P_{21}\, ds &= -\int_{\Gamma_D} \mathbf{t}_1\cdot\mathbf{P}\cdot\mathbf{n}\, ds = \int_{\Gamma_N} \mathbf{t}_1\cdot\mathbf{P}\cdot\mathbf{n}\, ds = 80\gamma^{\text{load}},\\
\int_{\Gamma_D} P_{31}\, ds &= -\int_{\Gamma_D} \mathbf{t}_2\cdot\mathbf{P}\cdot\mathbf{n}\, ds = \int_{\Gamma_N} \mathbf{t}_2\cdot\mathbf{P}\cdot\mathbf{n}\, ds = 80\cdot 0\cdot\gamma^{\text{load}} = 0,
\end{aligned} \tag{6.7}$$

similar to (6.5). In particular for $\gamma^{\text{load}} = 0.05$ this means that $\text{Val}_1 := \int_{\Gamma_D} P_{11}\, ds = 0$, $\text{Val}_2 := \int_{\Gamma_D} P_{21}\, ds = 80\gamma^{\text{load}} = 80\cdot 0.05 = 4$ and $\text{Val}_3 := \int_{\Gamma_D} P_{31}\, ds = 0$ are the exact boundary integral values on Γ_D if one insert the correct stress components P_{11}, P_{21}, P_{31} of $\mathbf{P}$.

Some results can be found in Table 6.21 using adaptive and uniform refinement. In the adaptive case we consider the results on two different sequences of meshes, similar as above. „Adaptive refinement I“ corresponds to the coarse initial mesh with 880 tetrahedra and „adaptive refinement II“ to the finer initial mesh with 7040 tetrahedra. For the finer initial mesh we distinguish again between $\sigma_{\text{dörf}} \in \{0.9, 0.925, 0.95\}$ in the second refinement step. In the case of uniform refinement we start with the coarse initial mesh. We observe convergence of Val_1 and Val_2 to the correct values in all cases. Adaptive refinement leads to better results than uniform refinement, as expected. Concerning Val_3 we have to say that the values sometimes oscillate, e.g. observable in „adaptive refinement I“ from level $l = 2$ to $l = 3$. This behavior seems to occur if the previous solution is close to the used tolerance $tol = 10^{-6}$ in the Gauss - Newton framework.

	adaptive refinement I				
Level l	n_t	$\sigma_{\text{dörf}}$	Val_1	Val_2	Val_3
0	880		$1.7462 \cdot 10^{-2}$	$3.9723 \cdot 10^{0}$	$-1.1473 \cdot 10^{-4}$
1	1410	0.800	$6.6751 \cdot 10^{-3}$	$3.9872 \cdot 10^{0}$	$8.6541 \cdot 10^{-6}$
2	1928	0.650	$3.0716 \cdot 10^{-3}$	$3.9921 \cdot 10^{0}$	$5.3635 \cdot 10^{-6}$
3	2892	0.450	$1.8159 \cdot 10^{-3}$	$3.9959 \cdot 10^{0}$	$-6.7773 \cdot 10^{-5}$

	adaptive refinement II				
Level l	n_t	$\sigma_{\text{dörf}}$	Val_1	Val_2	Val_3
0	7040		$7.4262 \cdot 10^{-3}$	$3.9884 \cdot 10^{0}$	$-8.7433 \cdot 10^{-6}$
1	14284	0.950	$2.8350 \cdot 10^{-3}$	$3.9956 \cdot 10^{0}$	$1.9505 \cdot 10^{-5}$
2	18628	0.900	$1.2603 \cdot 10^{-3}$	$3.9981 \cdot 10^{0}$	$-1.3150 \cdot 10^{-5}$
2	23640	0.925	$1.2337 \cdot 10^{-3}$	$3.9981 \cdot 10^{0}$	$-1.3283 \cdot 10^{-5}$
2	39442	0.950	$1.2068 \cdot 10^{-3}$	$3.9982 \cdot 10^{0}$	$-5.7961 \cdot 10^{-6}$

	uniform refinement				
Level l	n_t	$\sigma_{\text{dörf}}$	Val_1	Val_2	Val_3
0	880		$1.7462 \cdot 10^{-2}$	$3.9723 \cdot 10^{0}$	$-1.1473 \cdot 10^{-4}$
1	7040	1.000	$6.7975 \cdot 10^{-3}$	$3.9895 \cdot 10^{0}$	$-3.8141 \cdot 10^{-6}$

Table 6.21: Values of boundary integrals on Γ_D (incompressible 3d Cook, LSFEM)

We can conclude that our least squares method, using suitable parameters $\sigma_{\text{dörf}}$, lead to optimal convergence rates. We have seen that the obtained convergence rates are very sensitive with respect to the choice of this parameter. We have again observed an improved convergence rate for the conservation of linear momentum, similar as in two dimensions. Good stress approximations can be achieved with our method. This includes that the axiom of force and momentum balance is satisfied quite well (cp. Table 6.21 and Figure 6.34).

6.3 Transverse isotropy in three dimensions

In this Section our aim is to test the proposed least squares finite element method of Section 4 for transverse isotropic materials. In the numerical simulations we use the same finite element spaces as in Section 6.2. Before we start we mention that all Young's moduli and shear moduli have the physical unit of force per length squared. In common tables, e.g. in literature or in the internet, these moduli are usually given in the unit of megapascal which is identical to Newton per square millimeter. The units of physical constants and forces are again neglected in the following.

In real applications so - called fiber reinforced materials are of great importance. Such materials are composites consisting of a basic material and some fibers of a second material. The fibers strengthen the material in a particular direction. In practical applications the basic material is weaker than the fiber material. This implies that Young's modulus of the basic material is usually less than Young's modulus of the fiber material. Vice versa, Poisson's ratio of the basic material is usually greater than Poisson's ratio of the fiber material.

With this in mind we consider a composite of a weak basic material (e.g. an elastomer) and a stronger material (e.g. steel fibers). If we use the weak material in the isotropic planes and perpendicular to them the strong material we are in the situation of a transverse isotropic material studied in Section 4. We need the material parameters $E_1, E_3, \nu_{12}, \nu_{31}$ and G_{31} as input for the calculation of the coefficients $(\alpha, \beta, \varepsilon_1, \varepsilon_2, \varepsilon_3)$ within the used transverse isotropic model (cf. Section 4.3). For an elastomer the values $\nu_{12} = 0.4$ as Poisson's ratio and $E_1 = 10^3$ as Young's modulus are quite reasonable in real applications and are used in our numerical simulation below. For the strong material we use a fixed Poisson's ratio $\nu_{31} = 0.3 < \nu_{12}$ and vary Young's modulus $E_3 = 10^{3+j} > E_1$ for $j = 1, 2, 3$ in order to show the robustness of our method for increasing E_3. We set the remaining necessary material parameter G_{31}, the shear modulus in the x_3 - x_1 - plane, as $G_{31} = 400$. The chosen set of material parameters satisfies the conditions below equation (4.21).

As free material parameters we choose $\delta = 0$, $a_1 = 9$, $a_2 = 1$, $a_3 = -\frac{1}{2}$, $b_1 = 1$, $b_2 = 3$ and $b_3 = -\frac{1}{2}$. With this choice and the chosen material parameters above we obtain nonnegative coefficients in (4.14) (respectively in (4.12) and (4.13)) after solving the linear system of equation derived in Section 4.3. Consequently, polyconvexity of the underlying stored energy function (4.14) is ensured (cf. Section 4.2).

As geometry and boundary conditions for our numerical example below we use again the Cook membrane problem in three dimensions described in Figure 6.31. Furthermore we choose $\mathbf{g} = (0, 40, 0)^T$ and $\mathbf{f} = \mathbf{0}$ as surface and volume force densities and $\omega_1 = 10^2$, $\omega_2 = 1$ as scaling parameters in (4.24).

In the first test we would like to illustrate numerically the dependence of displacements relative to the preferred direction $\mathbf{a}$. One expects that the displacement approximations

vary if one changes $\mathbf{a}$. For a confirmation of this behavior we start with preferred direction $\mathbf{a} = \frac{1}{\sqrt{3}}(\sqrt{2}, 0, 1)^T$ and rotate it about the x_3 - axis with the help of the orthogonal matrix

$$\mathbf{Q}_z = \begin{pmatrix} \cos(\phi) & -\sin(\phi) & 0 \\ \sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

depending on the rotation angle $\phi \in [0, \pi)$.
We start with $\phi = 0$ and choose an angle step size of $\frac{\pi}{8}$ ($\hat{=} 22.5°$). The corresponding results can be found in Table 6.22 for three different Young's moduli $E_3 = 10^{3+j}$, $j = 1, 2, 3$.

angle ϕ	$u_1(48, 60, 5)$			$u_2(48, 60, 5)$			$u_3(48, 60, 5)$		
	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
0°	−9.4052	−9.3698	−9.3662	10.4766	10.4609	10.4593	0.2044	0.0629	0.0468
22.5°	−8.3774	−8.2460	−8.2327	9.3860	9.2612	9.2486	−0.2799	−0.5373	−0.5649
45°	−8.2653	−8.0892	−8.0710	8.8394	8.6366	8.6162	−0.6477	−0.9286	−0.9607
67.5°	−9.3688	−9.2917	−9.2831	9.7851	9.6754	9.6640	−0.5118	−0.6002	−0.6143
90°	−10.3970	−10.4251	−10.4278	10.9611	10.9725	10.9737	−0.4106	−0.4225	−0.4234
112.5°	−10.6179	−10.6710	−10.6759	11.1998	11.2386	11.2423	−0.4461	−0.4579	−0.4567
135°	−10.5331	−10.5832	−10.5882	11.0457	11.0795	11.0830	−0.5923	−0.6050	−0.6069
157.5°	−10.3475	−10.3901	−10.3945	10.9905	11.0164	11.0190	−0.5439	−0.4898	−0.4833

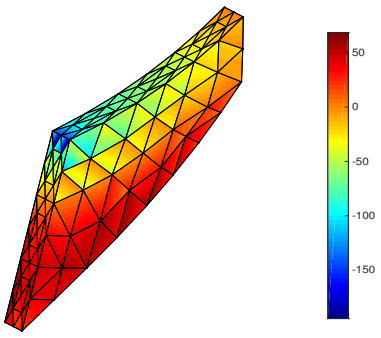
Table 6.22: Transverse isotropy: Dependence of displacements relative to preferred direction $\mathbf{a}$ for $E_3 = 10^{3+j}$, $j = 1, 2, 3$

One observes the displacement dependence in the particular point $(48, 60, 5)$ with respect to the preferred direction $\mathbf{a}$ in each case. The results belong to a fixed mesh with $n_t = 1144$ tetrahedra. One can also observe that if the displacements in x_1 - direction increase the displacements in x_2 - direction decrease and vice versa. This is reasonable for the considered problem. For each considered rotation angle ϕ one can additionally observe a kind of convergence of u_i, $i = 1, 2, 3$. Thus our proposed least squares method seems robust in E_3. One can also observe in this table that the absolute values of the displacement approximations in x_3 - direction reach its maximum in the case of a preferred direction $\mathbf{a} = \frac{1}{\sqrt{3}}(1, 1, 1)^T$, corresponding to the rotation angle $\phi = \frac{\pi}{4}$ ($\hat{=} 45°$).
After these observations we consider the convergence of the least squares functional $\mathcal{F}_{ti}$, evaluated in the obtained approximations $(\mathbf{P}_h, \mathbf{u}_h) \in \mathbf{\Pi}_h \times \mathbf{U}_h$, and the convergence of $\operatorname{div} \mathbf{P}_h + \mathbf{f}$ in the squared $L^2(\Omega)$ - norm for the choice of $\mathbf{a} = \frac{1}{\sqrt{3}}(1, 1, 1)^T$ and $E_3 = 10^6$. We see in Table 6.23 that one can achieve optimal convergence rates for the nonlinear functio-

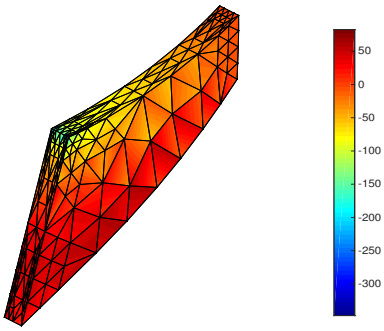
l	n_t	$\sigma_{\text{dörf}}$	$\mathcal{F}_{ti}(\mathbf{P}_h, \mathbf{u}_h)$	(order)	$\|\operatorname{div}(\mathbf{P} - \mathbf{P}_h)\|^2_{L^2(\Omega)}$	(order)	$u_1(48, 60, 5)$	$u_2(48, 60, 5)$	$u_3(48, 60, 5)$
0	880		$7.0826 \cdot 10^{-1}$		$8.9815 \cdot 10^{-13}$		−7.9973	8.5615	−0.9645
1	1499	0.75	$3.5686 \cdot 10^{-1}$	(1.287)	$2.0582 \cdot 10^{-13}$	(2.766)	−8.2149	8.7174	−1.1015
2	1617	0.30	$3.2136 \cdot 10^{-1}$	(1.383)	$1.6467 \cdot 10^{-13}$	(2.944)	−8.2288	8.7307	−1.1155

Table 6.23: Transverse isotropy: Results for $\mathbf{a} = \frac{1}{\sqrt{3}}(1, 1, 1)^T$ and $E_3 = 10^6$ with adaptive refinement

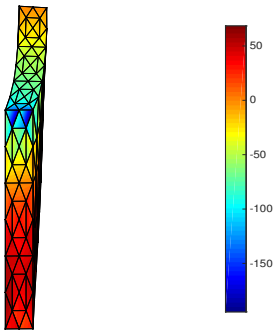
nal if one uses suitable parameter $\sigma_{\text{dörf}}$ in the marking strategy of Dörfler (cp. Appendix C.2). However the optimal choice of $\sigma_{\text{dörf}}$ is also here a difficult task, similar as observed in Section 6.2.2. The convergence rates to the momentum term is again improved and the conservation of momentum is also satisfied quite well in this example. In the last three columns of Table 6.23 the displacements in the particular point $(48, 60, 5)$ can be found. They seems to converge.



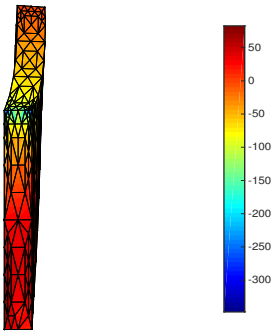
(a) uniform, total view



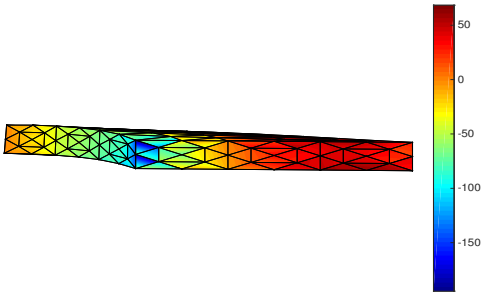
(b) adaptive, total view



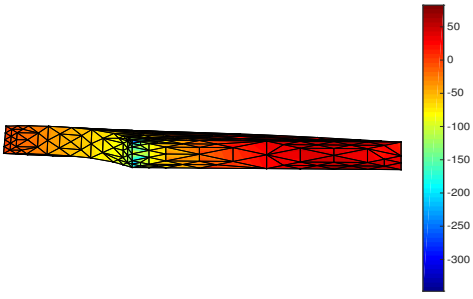
(c) uniform, view along x_2 - axis



(d) adaptive, view along x_2 - axis



(e) uniform, view along x_3 - axis



(f) adaptive, view along x_3 - axis

Figure 6.35: Transverse isotropy: Visualization of approximated $\boldsymbol{\tau}_{11}$ (left: uniform mesh with $n_t = 1144$, right: adaptive mesh with $n_t = 1617$)

In the next step our aim is to study the stress approximations. For this purpose we compare exemplarily the approximation of the Kirchhoff stress tensor component $\boldsymbol{\tau}_{11}$ on a uniform ($n_t = 1144$) as well as on a locally refined mesh ($n_t = 1617$). The results correspond again to $\mathbf{a} = \frac{1}{\sqrt{3}}(1,1,1)^T$ and $E_3 = 10^6$ and are depicted in Figure 6.35, on the left for the uniform mesh and on the right for the adaptive refined mesh. The stress approximations are plotted with respect to different views and look reasonable. One observes a bending behavior in x_3 - direction similar to the results in [SWB11]. Moreover, the occurring singularity at the edge $\{(0, 44, x_3) : 0 < x_3 < 5\}$ on Γ_D can be observed again.
Finally, we consider the boundary integral values Val_1, Val_2 and Val_3 for this example (cf. equation (6.7)). The results are given in Table 6.24.

Level l	0	1	2	exact values
n_t	880	1499	1617	
Val_1	$2.8993 \cdot 10^{-5}$	$1.1112 \cdot 10^{-5}$	$7.9571 \cdot 10^{-6}$	0
Val_2	$3.2000 \cdot 10^{3}$	$3.2000 \cdot 10^{3}$	$3.2000 \cdot 10^{3}$	3200
Val_3	$-1.2398 \cdot 10^{-5}$	$-1.3816 \cdot 10^{-5}$	$-1.3774 \cdot 10^{-5}$	0

Table 6.24: Values of boundary integrals on Γ_D (3d Cook, transverse isotropy, adaptive LSFEM)

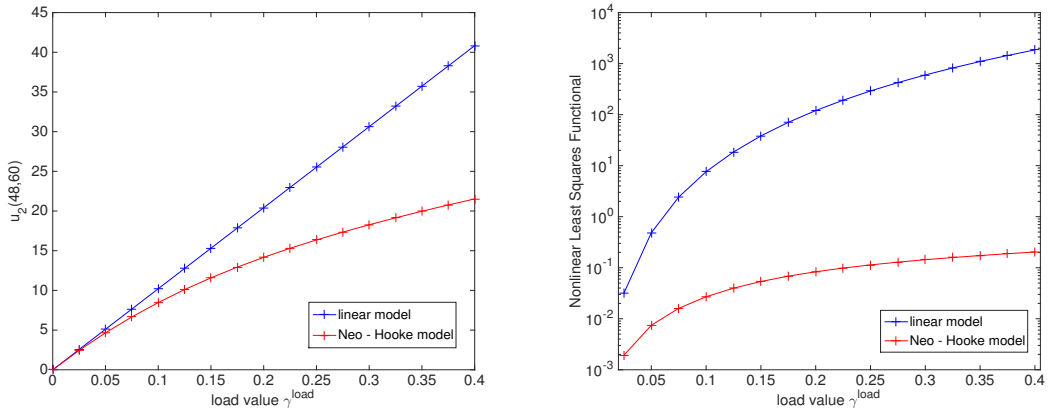
One can also observe for this example, dealing with transverse isotropic materials, that the approximations of the exact boundary integral values are very good.
Altogether we can conclude that our proposed least squares finite element method is also promising for the simulation of anisotropic materials. First results for materials with transverse isotropic behavior are presented in this section. They look quite reasonable although no analysis is provided in this work. In particular we have shown numerically that the results depend reasonably on the preferred direction and that we can get optimal convergence rates and good stress approximations.

6.4 Model adaptivity in two dimensions

In this example our aim is to apply Algorithm 2 and show that the considerations in Section 5 are reasonable. We use the same finite element spaces $\mathbf{\Pi}_h$ and $\mathbf{U}_h$ as in Section 6.1 within the numerical simulations below. For this purpose we consider again Cook's membrane in two dimensions (cp. Figure 6.1) with body force density $\mathbf{f} = \mathbf{0}$ and surface force density $\mathbf{g} = (0, \gamma^{\text{load}})^T$, $\gamma^{\text{load}} \in \mathbb{R}$. As scaling parameters in (5.1) and (5.5) we use $\omega_1 = 10^2 = \omega_1^{\text{lin}}$, $\omega_2 = \frac{1}{2}$ and $\omega_2^{\text{lin}} = 1$. The example below corresponds to a fixed mesh with $n_t = 2096$ triangles. In this section $(\mathbf{P}_{lin}, \mathbf{u}_{lin})$ denotes the solution of the minimization problem (5.2) of linear elasticity and $(\mathbf{P}_{red}, \mathbf{u}_{red})$ the solution of the minimization problem (5.10) in the finite dimensional space $\mathbf{\Pi}_h \times \mathbf{U}_h$.

Before we present some results for this problem, we firstly mention that the assumption $(\mathbf{P}_{NH}, \mathbf{u}_{NH}) \in \mathbf{\Pi}^\infty \times \mathbf{U}^\infty$ of Corollary 5.3 is not satisfied for the Cook membrane problem. We have observed this lack of regularity numerically in finite dimensional spaces near the point $(0, 44)$ in two dimensions (cf. Sections 6.1.1, 6.1.2 and 6.1.3) and near the edge $\{(0, 44, x_3) : 0 < x_3 < 5\}$ in three dimensions (cf. Section 6.2.2) in the previous examples. Despite this regularity problem, the nonlinear least squares functional $\mathcal{F}_{NH}(\mathbf{P}_h, \mathbf{u}_h)$, $(\mathbf{P}_h, \mathbf{u}_h) \in \mathbf{\Pi}_h \times \mathbf{U}_h$, has worked reasonable as a-posteriori error estimator in these numerical simulations. For this reason we assume that the nonlinear functional $\mathcal{F}_{NH}(\mathbf{P}_{lin}, \mathbf{u}_{lin})$ (respectively $\mathcal{F}_{NH}(\mathbf{P}_{red}, \mathbf{u}_{red})$) is also a measure of quality for the solution of linear elasticity (respectively the reduced solution) with respect to the Neo-Hooke model for this problem (cf. Corollary 5.3).

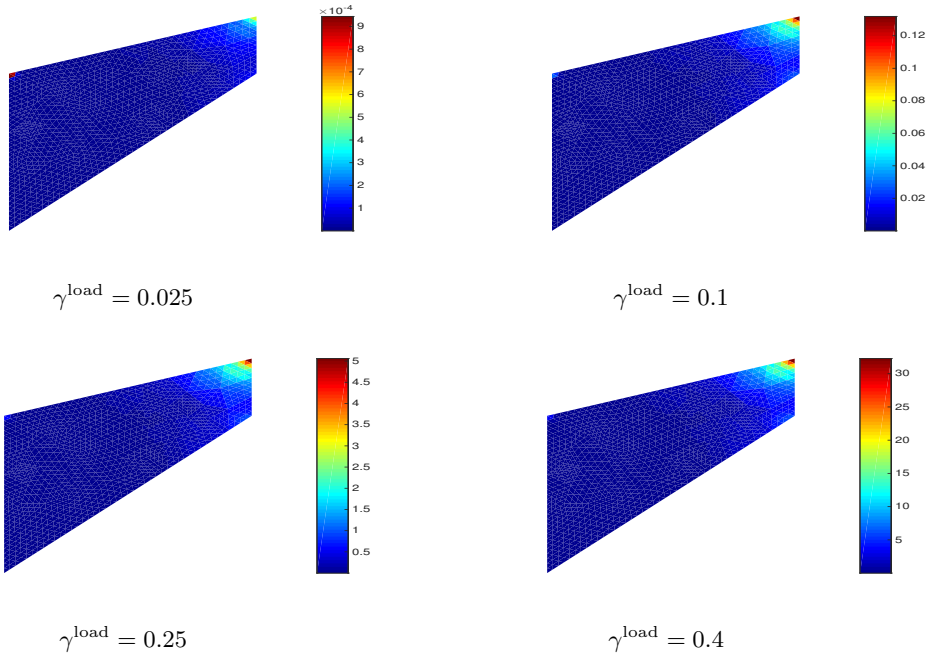
In Figure 6.36 a comparison between the linear model and the Neo-Hooke model is illustrated. On the left part of this figure the vertical displacement $u_2(48, 60)$ is plotted for load values $\gamma^{\text{load}} \in [0, 0.4]$ (linear model in blue, nonlinear model in red). At first glance



(a) Displacements $u_2(48, 60)$ (b) Quality of solutions

Figure 6.36: Model adaptivity: Comparison of linear and Neo-Hooke model

one observes consistency of the nonlinear model with the linear model, as expected by Section 2.4.5, and an appropriate linear load - displacement relation of the linear model. Moreover, one can observe that the linear solution becomes worse if one increases the load value. The right part of Figure 6.36 confirms this observation in a more general way. Here, instead of taking the displacement in only one particular point into account, the whole nonlinear least squares functional is considered in the approximations. More precisely, the values $\mathcal{F}_{NH}\big((\mathbf{P}_{NH})_h, (\mathbf{u}_{NH})_h\big)$ (red curve), $\big((\mathbf{P}_{NH})_h, (\mathbf{u}_{NH})_h\big) \in \mathbf{\Pi}_h \times \mathbf{U}_h$, and $\mathcal{F}_{NH}(\mathbf{P}_{lin}, \mathbf{u}_{lin})$ (blue curve) are plotted for different load values. Similar to the left plot both curves drift apart if one increases the load. The results are plausible and reflects the observations in physical experiments.
In Figure 6.37 the distribution of $\mathcal{F}_{NH}(\mathbf{P}_{lin}, \mathbf{u}_{lin})$ is plotted on the domain Ω for four different load values.



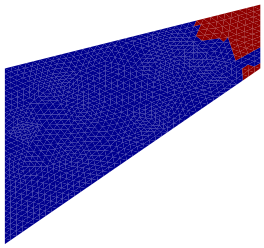
$\gamma^{\text{load}} = 0.025$ $\gamma^{\text{load}} = 0.1$

$\gamma^{\text{load}} = 0.25$ $\gamma^{\text{load}} = 0.4$

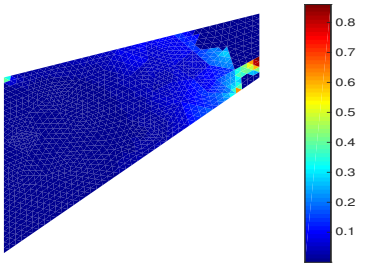
Figure 6.37: Error distribution of linear solution $(\mathbf{P}_{lin}, \mathbf{u}_{lin})$ for different load values

Recall that we have neglected the quadrature error in the theoretical part of Section 5 and therefore the total error contains only the model and the discretization error. Due to the previously considered examples we know that we have a singularity near the node $(0, 44)$ in this problem. Thus we expect a large discretization error in this part of the domain, independently of the considered model. Looking at the plots in Figure 6.37 we see another part where the error is locally quite large. This part seems to be close to the right boundary and in particular near the point $(48, 60)$. This area is exactly the region where the surface force is applied. Intuitively this is the part where the difference between the nonlinear and the linear model should be large. This observation is valid for all considered

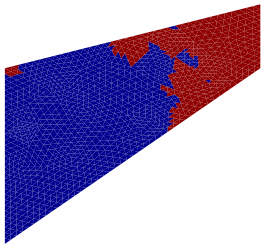
load values in Figure 6.37.
With this in mind we apply Algortihm 2 for the load value $\gamma^{\text{load}} = 0.25$. By Figure 6.36 it is clear that we are far away from the regime of linear elasticity and the displacements are quite large for this choice. In the algorithm we use the marking strategy of Dörfler with $\sigma_{\text{dörf}} = 0.7$ (cp. Appendix C.2), the tolerance $tol_{\text{mod}} = 10^{-1}$ and a maximal number $i_{\max}^{\text{mod}} = 15$ of model adaptivity steps. For the damped Gauss - Newton algorithm (cp. Algorithm 1) inside Algorithm 2 we use a tolerance $tol = 10^{-6}$ in the stopping criterion and $i_{\max} = 20$ as maximal number of Gauss - Newton steps per model adaptive step. In the following we denote $i \in \mathbb{N}$ as level in the model adaptivity scheme, according to the considerations in Section 5.2. Recall that $i = 0$ corresponds to a fully linear model.
Starting with the linear model on the whole domain, i.e. $\Omega_2^{(0)} = \Omega$ and $\Omega_1^{(0)} = \emptyset$, we expect by the considerations above that in the areas near the left upper node and near the right boundary the linear model will be substituted by the nonlinear model. This can be clearly seen by using Algortihm 2. The propagation of the nonlinear regime is illustrated in Figure 6.38 in the left column after one, three, six and nine steps of model adaptivity. One observes that the simple elements near the right boundary are exchanged in the first step by complex elements. The exchange of simple elements to complex elements near the singularity at the left upper node is visible after the third step. In general the complex elements propagate in each step more and more from the right into the left part of Ω. After the ninth step also the change of boundary conditions from hard - clamped to stress - free in the origin is taken into account.
In the right column of Figure 6.38 the distribution of $\mathcal{F}_{NH}\left(\mathbf{P}_{red}^{(i)}, \mathbf{u}_{red}^{(i)}\right)$, $\left(\mathbf{P}_{red}^{(i)}, \mathbf{u}_{red}^{(i)}\right) \in \mathbf{\Pi}_h \times \mathbf{U}_h$, is plotted over the entire domain. One observes that the error on the complex elements is quite small and the error dominates at the transition of the nonlinear and linear part and at the singularity. After nine steps of model adaptivity one has a quite smooth and small error distribution on the whole domain with exception near the singularity. In this part local mesh refinement is necessary to improve the results. For a closer quantitative consideration of the propagation of the nonlinear and linear regime we refer to Table 6.25. In the first steps the number of new complex elements from $\Omega_1^{(i)}$ to $\Omega_1^{(i+1)}$ increases and later the number of them decreases. For instance from $\Omega_1^{(2)}$ to $\Omega_1^{(3)}$ we get 270 new complex elements and from $\Omega_1^{(14)}$ to $\Omega_1^{(15)}$ we get only 2 new complex elements. This means that the model has to be adjusted quite strongly at the beginning and just a bit in the later steps of Algorithm 2.
In the second to last column the value of the nonlinear functional $\mathcal{F}_{NH}$, evaluated in the reduced solution $\left(\mathbf{P}_{red}^{(i)}, \mathbf{u}_{red}^{(i)}\right)$, and in the last column the values of the vertical displacement in $(48, 60)$ are listed. Convergence of these values can be observed in this example although the displacement values are oscillating in the first steps. Since the change of the nonlinear functional values $\mathcal{F}_{NH}\left(\mathbf{P}_{red}^{(i)}, \mathbf{u}_{red}^{(i)}\right)$ between two successive meshes is quite small after a certain number of steps, it is reasonable to stop the algorithm at this point.



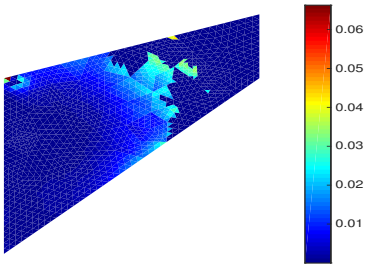
$|\Omega_1^{(1)}| = 107, |\Omega_2^{(1)}| = 1989$



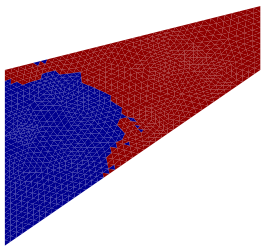
error distribution of $\left(\mathbf{P}_{red}^{(1)}, \mathbf{u}_{red}^{(1)}\right)$



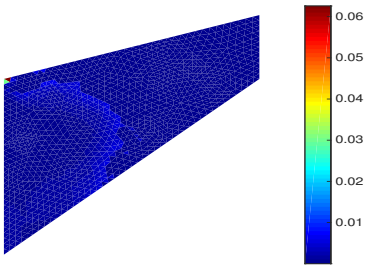
$|\Omega_1^{(3)}| = 545, |\Omega_2^{(3)}| = 1551$



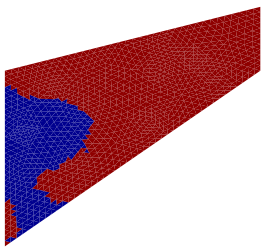
error distribution of $\left(\mathbf{P}_{red}^{(3)}, \mathbf{u}_{red}^{(3)}\right)$



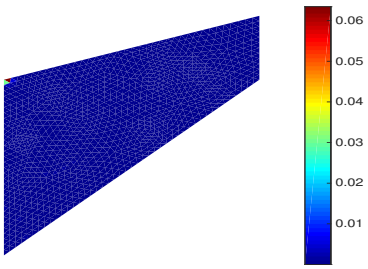
$|\Omega_1^{(6)}| = 1106, |\Omega_2^{(6)}| = 990$



error distribution of $\left(\mathbf{P}_{red}^{(6)}, \mathbf{u}_{red}^{(6)}\right)$



$|\Omega_1^{(9)}| = 1551, |\Omega_2^{(9)}| = 545$



error distribution of $\left(\mathbf{P}_{red}^{(9)}, \mathbf{u}_{red}^{(9)}\right)$

Figure 6.38: Visualization of model adaptivity on a fixed mesh ($n_t = 2096$) (left: decomposition into linear domain $\Omega_2^{(i)}$ and nonlinear domain $\Omega_1^{(i)}$; right: error distribution for reduced solutions)

Level i	$\lvert\Omega_1^{(i)}\rvert$	$\lvert\Omega_2^{(i)}\rvert$	$\mathcal{F}_{NH}(\mathbf{P}_{red}^{(i)}, \mathbf{u}_{red}^{(i)})$	$u_2(48, 60)$
0	0	2096	$2.9120 \cdot 10^{2}$	25.5216
1	107	1989	$7.0948 \cdot 10^{1}$	18.4849
2	275	1821	$1.9726 \cdot 10^{1}$	16.3284
3	545	1551	$9.3872 \cdot 10^{0}$	16.0075
4	745	1351	$4.9352 \cdot 10^{0}$	16.0405
5	942	1154	$2.5737 \cdot 10^{0}$	16.0640
6	1106	990	$1.4340 \cdot 10^{0}$	16.1579
7	1259	837	$8.2623 \cdot 10^{-1}$	16.2261
8	1419	677	$5.2130 \cdot 10^{-1}$	16.2876
9	1551	545	$3.6855 \cdot 10^{-1}$	16.3177
10	1634	462	$2.9418 \cdot 10^{-1}$	16.3289
11	1680	416	$2.5333 \cdot 10^{-1}$	16.3342
12	1703	393	$2.3458 \cdot 10^{-1}$	16.3373
13	1715	381	$2.2526 \cdot 10^{-1}$	16.3384
14	1721	375	$2.2031 \cdot 10^{-1}$	16.3390
15	1723	373	$2.1867 \cdot 10^{-1}$	16.3391

Table 6.25: Results for model adaptivity

Otherwise the computational costs are higher compared to the benefit. In this example a reasonable point to terminate the algorithm is probably between the eighth and tenth step.

At the end of this example we would like to confirm numerically one main advantage of model adaptivity. We have already realized at the end of Section 5.2 that one can reuse the local stiffness matrices on the domain Ω_2, provided that we have a fixed disjunct decomposition of Ω into a linear part Ω_2 and a nonlinear part Ω_1. This affects the number of entries in the global stiffness matrix that must be recomputed in the single steps of the Gauss - Newton scheme. Moreover, also a part of the stiffness matrix in the transition $\left(\Omega_1^{(i)}, \Omega_2^{(i)}\right) \rightarrow \left(\Omega_1^{(i+1)}, \Omega_2^{(i+1)}\right)$ can be reused. This part corresponds to the linear unchanged part, i.e. the elements which are in the intersection of $\Omega_2^{(i)}$ and $\Omega_2^{(i+1)}$. Since $\Omega_2^{(i+1)} \subseteq \Omega_2^{(i)}$ by construction, the linear unchanged part is exactly the set $\Omega_2^{(i+1)}$.

Some numerical results concerning the entries of the occurring stiffness matrices for this example can be found in Table 6.26.

Here $\mathbf{A}_{red}^{(i)}$ denote the stiffness matrices that occur in the Gauss - Newton scheme of the reduced model in level i of model adaptivity. In general they consist of a nonlinear part $\mathbf{A}_{nonlin}^{(i)}$, corresponding to $\Omega_1^{(i)}$, and a linear part, corresponding to $\Omega_2^{(i)}$. For $i = 0$ the stiffness matrix $\mathbf{A}_{red}^{(i)}$ coincides by construction with the stiffness matrix $\mathbf{A}_{lin}$ of linear elasticity. We split $\mathbf{A}_{red}^{(i)}$ additively into a linear and a nonlinear part. We determine the nonlinear part $\mathbf{A}_{nonlin}^{(i)}$ as all entries of $\mathbf{A}_{red}^{(i)} - \mathbf{A}_{lin}$ which are greater than a given tolerance tol. In particular this means that $n\left(\mathbf{A}_{nonlin}^{(i)}\right) := \#\left\{(j,k) : \left|\left(\mathbf{A}_{red}^{(i)}\right)_{j,k} - \left(\mathbf{A}_{lin}\right)_{j,k}\right| > tol\right\}$ denotes the number of nonlinear entries in the matrix $\mathbf{A}_{red}^{(i)}$ up to a given tolerance. The

Level i	$nnz\left(\mathbf{A}_{red}^{(i)}\right)$	$n\left(\mathbf{A}_{nonlin}^{(i)}\right)$	$1-\frac{n(\mathbf{A}_{nonlin}^{(i)})}{nnz(\mathbf{A}_{red}^{(i)})}$	$\frac{\lvert\Omega_2^{(i)}\rvert}{n_t}$
0	1674148	0	1.0000	1.0000
1	1674664	84223	0.9497	0.9490
2	1674912	218879	0.8693	0.8688
3	1674934	437317	0.7389	0.7400
4	1674966	597853	0.6431	0.6446
5	1674998	755954	0.5487	0.5506
6	1675024	887998	0.4699	0.4723
7	1675058	1010499	0.3967	0.3993
8	1675086	1140367	0.3192	0.3230
9	1675284	1245715	0.2564	0.2600
10	1675428	1313254	0.2162	0.2204
11	1675538	1350576	0.1939	0.1985
12	1675554	1369495	0.1827	0.1875
13	1675560	1379015	0.1770	0.1818
14	1675566	1384033	0.1740	0.1789
15	1675586	1385751	0.1730	0.1780

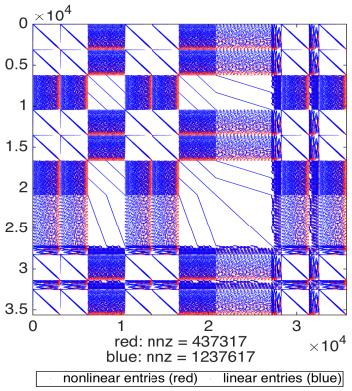
Table 6.26: Development of nonlinear entries in stiffness matrices

other entries of $\mathbf{A}_{red}^{(i)}$ are similar to the corresponding entries in $\mathbf{A}_{lin}$. For the results in Table 6.26 the tolerance was chosen as $tol = 10^{-9}$.

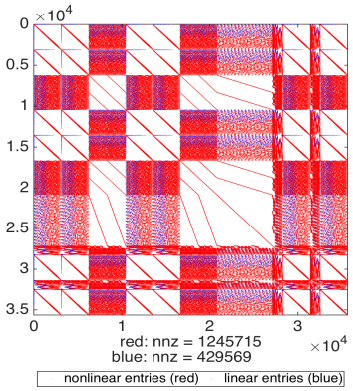
In the second column of Table 6.26 the number of nonzero entries of $\mathbf{A}_{red}^{(i)}$, abbreviated as $nnz\left(\mathbf{A}_{red}^{(i)}\right)$, can be found for each level of our model adaptivity scheme. In the third column the number of nonlinear entries in $\mathbf{A}_{red}^{(i)}$ are listed. They increase during the process of model adaptivity steps. This is reasonable, since more and more elements become complex elements in the algorithm. $\frac{n\left(\mathbf{A}_{nonlin}^{(i)}\right)}{nnz\left(\mathbf{A}_{red}^{(i)}\right)}$ is the ratio of nonlinear entries in the stiffness matrix $\mathbf{A}_{red}^{(i)}$. Thus $1-\frac{n\left(\mathbf{A}_{nonlin}^{(i)}\right)}{nnz\left(\mathbf{A}_{red}^{(i)}\right)}$ indicates the ratio of entries in this matrix that are similar to linear elasticity. The number $1-\frac{n\left(\mathbf{A}_{nonlin}^{(i)}\right)}{nnz\left(\mathbf{A}_{red}^{(i)}\right)}$ is therefore a ratio of matrix entries which need not be recomputed, i.e. a measure of saving computational time. This ratio coincides approximately with the ratio of linear elements to all elements (cp. the last two columns in Table 6.26). Furthermore, it is reasonable that these values decrease within increasing the level i, since more and more elements become complex.

All occurring stiffness matrices in this example have the dimension $\dim(\mathbf{\Pi}_h \times \mathbf{U}_h) = 35632$ and are sparse. For instance in Figure 6.39 the structure of two occurring stiffness matrices after three and nine steps of model adaptivity can be observed. The entries are divided into a blue part, corresponding to the linear model, and a red part corresponding to the nonlinear model.

At the end of this example we would like to point out that this algorithm for model ad-



Level $i = 3$



Level $i = 9$

Figure 6.39: Decomposition of stiffness matrices for two different levels

aptivity is far from optimal. A more improved algorithm needs a decomposition of the total error into model and discretization error. Such a splitting is desirable and highly recommended similar to [SRO07]. One should combine model adaptivity with usual mesh refinement in an appropriate way. However, from our point of view the observations here and in Section 5 provide some first helpful considerations dealing with model adaptivity in the context of least squares finite element methods, even without such a decomposition of the total error. The general aim of model adaptivity should be to speed up an existing algorithm combining model adaptivity and usual mesh refinement. Provided that the approximation quality of the solutions are the same, an algorithm using model adaptivity should be faster than another algorithm without using model adaptivity.

Generally such algorithms using model adaptivity are very interesting, since one can start with the simplest possible model and the algorithm automatically decides in which elements one must use a more complex one. Also an extension using a hierarchy of different models is possible, e.g. using the model of linear elasticity, Neo - Hooke, Mooney - Rivlin and maybe even more complex models.

7 Conclusion and outlook

7.1 Conclusion

In this work polyconvex stored energy functions in the context of hyperelasticity for the description of nonlinear material behavior have been considered. The coefficients in these functions have been determined such that the nonlinear model is consistent with appropriate linear ones (cf. Section 2.4.5 for the homogeneous isotropic Mooney - Rivlin model and Section 4.3 for a special model within transverse isotropy).
Based on the physical necessary conservation of linear momentum and the usual stress - strain relation, derived from the given stored energy function, the idea in our approach is to invert the nonlinear stress - strain relation, similar as done in [CS04] for linear elasticity. We have shown with the local inversion theorem that this is at least possible for small strains although an exact representation for the inverse is not available in general. With this in hand we have formulated general nonlinear least squares functionals for homogeneous isotropic models in terms of $\mathbf{B}$ and $\mathbf{C}$ (cf. (3.19)). These functionals depend on the first Piola - Kirchhoff stress tensor $\mathbf{P}$ and the displacement $\mathbf{u}$ and lead therefore to mixed finite element methods. This approach can be used in general for all kinds of given stored energy functions, provided that consistency of the considered model with linear elasticity is satisfied.
For the minimization of the nonlinear functionals in finite dimensional spaces we have used the Gauss - Newton scheme, i.e. we have replaced the nonlinear problem by a sequence of linearized problems. The practical result is Algorithm 1.
Focusing on a special model of Neo - Hooke type, we have shown that it is possible to derive cubic equations for the $\mathbf{B}$ - and the $\mathbf{C}$ - formulation (cf. (3.38) and (3.46)). With the help of them one is able to determine $\mathcal{A}_{NH}(\mathbf{P}\mathbf{F}(\mathbf{u})^T)$ (respectively $\tilde{\mathcal{A}}_{NH}(\mathbf{F}(\mathbf{u})^{-1}\mathbf{P})$) for given $(\mathbf{P}, \mathbf{u})$ exactly, i.e. in particular without using Newton's method. This is an essential advantage in the resulting numerical scheme. Another remarkable fact is that one can set $\lambda = \infty$ in these equations and in the resulting method. In particular we have shown the well - posedness of $\mathcal{A}_{NH}$ and $\tilde{\mathcal{A}}_{NH}$ (cf. Theorem 3.9 and Theorem 3.11) for $\lambda \to \infty$ and that in this case these mappings are no longer invertible, similar to $\mathcal{C}^{-1}$ in linear elasticity. Moreover, we have proven that the incompressibility constraint to the strain $\mathcal{A}_{NH}\left(\mathbf{P}\mathbf{F}(\mathbf{u})^T\right)$ $\left(\text{respectively } \tilde{\mathcal{A}}_{NH}\left(\mathbf{F}(\mathbf{u})^{-1}\mathbf{P}\right)\right)$, corresponding to any combination $(\mathbf{P}, \mathbf{u})$, is satisfied in the incompressible limit (see Remarks 3.10 and 3.12). For the $\mathbf{B}$ - formulation we have set conditions where the cubic equation has definitely only one real solution (Proposition 3.7). We have established an analysis for the nonlinear functional in the case of the Neo - Hooke material and the $\mathbf{B}$ - formulation starting with some necessary regularity assumptions for $(\mathbf{P}, \mathbf{u})$ and volume force density $\mathbf{f}$ (Corollary 3.14) such that the nonlinear functional (3.19) exists. Our main theoretical result is Theorem 3.29 which proves efficiency and reliability of the nonlinear functional $\mathcal{F}_{NH}$ and is uniformly valid in the incompressible limit $\lambda \to \infty$

such that Poisson locking is excluded within this model. Due to this theorem it is also theoretically proven that $\mathcal{F}_{NH}$ is a suitable a - posteriori error estimator which can be used in adaptive refinement strategies. An a - priori error estimate is an immediate consequence (cf. (3.86)). Moreover we have shown that the approximation $\boldsymbol{\tau}_h$ of the Kirchhoff stress tensor $\boldsymbol{\tau}$ becomes symmetric for $h \to 0$ (see Corollary 3.31). For the linearized problems we have proven that they are well - posed in $H_{\Gamma_N}(\mathrm{div};\,\Omega)^3 \times H^1_{\Gamma_D}(\Omega)^3$ (cf. Corollary 3.33). In Section 3.6 we have formulated some reference methods in order to compare our proposed least squares method with them in numerical experiments. In particular we have seen in this section that the usage of the Mooney - Rivlin model in a two - dimensional plane - strain leads to the same displacement approximations as the Neo - Hooke model (cf. Proposition 3.35). Moreover these formulations cover also the material behavior of linear elasticity if one uses the zero solution as initial guess (cf. Remark 3.34 and equation (3.103)).

In Section 4 we have extended our idea to materials with transversely isotropic behavior which is of great importance for concrete applications in engineering, e.g. fiber reinforced materials. We have formulated a suitable stored energy function based on [Sch10] and [BSN10] for such problems. Again consistency with an appropriate linear model and the transition to the isotropic case (cf. Remark 4.6) is ensured.

In Section 5 we have proposed an algorithm dealing with model adaptivity in the context of least squares finite element methods (cf. Algorithm 2). In particular the potential of measuring the quality of solutions of the reduced/simple model with respect to another model is of crucial importance in this context. The idea is realized in more detail for the model of linear elasticity as simple model and the Neo - Hooke model as more complex model (cf. Corollary 5.3). An extension to other models or a hierarchy of models is conceivable.

In Section 6 we have tested our proposed least squares finite element method succesfully in two and three dimensions. We have seen in several examples that we can set $\lambda = \infty$ in simulations for the Neo - Hooke model and that we can achieve almost optimal convergence rates, regardless of considering compressible or fully incompressible materials. The nonlinear least squares functional works well and reliable as a - posteriori error estimator, even for examples where the regularity assumptions of our theory are not satisfied. We have seen that the method also leads to good results for bending dominated problems (cf. Section 6.1.3). In this context we have pointed out the occurring scaling issue within our method. Considering the size of the domain and the physical parameters in more detail one can handle this problem quite well. One remarkable effect is the improved convergence rate for the conservation of momentum that we have observed in several examples (cp. [SSS11]).

After studying these examples our proposed least squares method seems well - suited to obtain good stress approximations. Stress oscillations as observed in the cases of the Ga-

lerkin or displacement - pressure approach are not present in the considered examples. The resulting stress approximations are one major advantage of our method compared to others. However, if one is only interested in good displacement approximations our least squares approach would not be the method of choice.
The extension of our method to more complicated homogeneous isotropic materials as Neo - Hooke is shown in Section 6.2.1 where the displacement approximations using a Mooney - Rivlin model for different parameters δ are compared within a three - dimensional uniaxial tension test.
At the end we would like to emphasize that our method has no difficulties in approximating the correct critical load values in the examples of [ABadVLR10]. Furthermore we have applied our method succesfully to an example with transversely isotropic material behavior (cf. Section 6.3). Here we have observed numerically the dependence of the displacement with respect to the preferred direction, as expected. And lastly we have shown exemplarily in Section 6.4 that our proposed algorithm for model adaptivity works quite well. We have seen a reasonable expansion of the nonlinear region and that one can reuse entries of older stiffness matrices for newer stiffness matrices in the Gauss - Newton framework. This idea might aid in saving computational time.

7.2 Outlook

In the following we would like to point out some open questions which arose during this work. Moreover we discuss some opportunities for further research.
Our theory in Section 3.5 is based on the convex sets $\mathbf{\Pi}^\infty$ and $\mathbf{U}^\infty$. This choice includes quite strong regularity assumptions which are not satisfied in general (cp. [HMW11]). It would be advantageous if one could weaken the assumptions to $W^q(\mathrm{div};\,\Omega)^3$ for the stresses and $W^{1,p}(\Omega)^3$ for the displacements with finite $q, p \geq 2$. Under these assumptions an analysis which includes efficiency and reliability of the nonlinear functional $\mathcal{F}(\mathbf{P}, \mathbf{u})$ with respect to appropriate norms is desirable.
Another problem occurred in Section 3.5.2. We have pointed out that we cannot guarantee that the new solution $\left(\mathbf{P}^{(k+1)}, \mathbf{u}^{(k+1)}\right)$ lies in $\mathbf{\Pi}^\infty \times \mathbf{U}^\infty$. But this is essential for Theorem 3.29. The aim should be to prove a regularity theorem such that the sequence $\left(\mathbf{P}^{(k)}, \mathbf{u}^{(k)}\right)$, $k \in \mathbb{N}$, stays in $\mathbf{\Pi}^\infty \times \mathbf{U}^\infty$.
Moreover it would be advantageous if one could extend the derivation of cubic equations, similar to (3.38) and (3.46), to more complicated models. But already for the Mooney - Rivlin model (2.30), $\delta > 0$, the coupling between $\mathbf{dev}\,\mathbf{B}$ and $\mathbf{dev}\,\boldsymbol{\tau}$ (cp. (3.36)) becomes much more complicated and prevents a simple analogous derivation. Well - posedness of $\mathcal{A}$ and $\tilde{\mathcal{A}}$ in the incompressible limit and an analysis for more complicated models would be desirable, but hard to achieve.
Besides these concrete issues, one could improve the computational time of Algorithm 1 in the following way: The linear systems of equations (cf. (3.26)) which occur in the Gauss -

Newton framework are extremely costly in three dimensions. They should be solved in an efficient way. One needs a suitable preconditioner in order to reduce the condition number of the occurring stiffness matrices and must combine them with a suitable solver. Domain decomposition methods and/or algebraic multigrid techniques could be helpful.
A further concrete improvement in the context of model adaptivity (cf. Section 5) would be an additive splitting of the total error into a part describing only the discretization error and a part describing only the model error. The considerations in [SRO07] could be helpful for further investigations.
The examples in Section 6 are all based on a polygonal bounded domain Ω. Bodys in real applications, e.g. cars or aircraft, obviously have a curved boundary. In order to improve results one can use so - called isoparametric elements. Here, roughly speaking, one increases the polynomial degree in the usual mapping from the reference element to an arbitrary element. Further explanations in the context of isoparametric elements can be found in [Bra07] and [BBF13]. Some investigations in the context of least squares finite element methods using isoparametric (Raviart - Thomas) elements have been provided in the recent works [BMS14b] and [BMS14a].
In the introduction at the beginning of this work we have distinguished between plastic and elastic deformations. The whole work has only considered elastic deformation processes. In order to simulate for instance crash tests in engineering one must extend the proposed method to plasticity. A mathematical introduction into plasticity can be found in [HR13]. In [Sta07] a least squares finite element method in the context of small - strain elasto - plasticity is realized. Some further modeling aspects and numerical examples in the context of finite multiplicative plasticity can be found for instance in [NW03].

Appendix

A Little o - and big $\mathcal{O}$ - notation

The following definitions can be found in [AE06a].

Little o - notation:

For normed spaces X and E, $D \neq \emptyset$, $f : D \subset X \to E$, $\alpha \geq 0$ and $a \in \bar{D}$ it holds

$$f(x) = o(\|x-a\|_X^\alpha) \ (x \to a) :\Leftrightarrow \lim_{x \to a} \frac{f(x)}{\|x-a\|_X^\alpha} = 0.$$

Or equivalently $\forall\, \varepsilon > 0$ there exists a neighborhood U of a in D with

$$\|f(x)\|_E \leq \varepsilon \|x-a\|_X^\alpha, \quad x \in U.$$

Big $\mathcal{O}$ - notation:

For normed spaces X and E, $D \neq \emptyset$, $f : D \subset X \to E$, $\alpha \geq 0$ and $a \in \bar{D}$ it holds $f(x) = \mathcal{O}(\|x-a\|_X^\alpha)$ $(x \to a)$ if and only if there exists $r > 0$ and $K > 0$ with $\|f(x)\|_E \leq K\|x-a\|_X^\alpha$ for all $x \in \mathcal{B}(a,r) \cap D$. $\mathcal{B}(a,r)$ denotes the open ball in X centered at a with radius r.

B Quadrature rules

The following quadrature rules can be found in [Cow73] for a triangulation into triangles (2d) and in [GH91] for a triangulation into tetrahedra (3d). Both quadrature rules integrate polynomials up to degree 5 exactly.

B.1 7 - point quadrature formula for triangles (2d)

In two dimensions we consider the reference triangle $\hat{T}$ with vertices $(0,0)$, $(1,0)$, $(0,1)$, the nodes

$$\hat{\mathbf{x}}_1 := \frac{1}{3}(1,1)^T, \quad \hat{\mathbf{x}}_2 := \frac{6-\sqrt{15}}{21}(1,1)^T, \quad \hat{\mathbf{x}}_3 := \left(\frac{6-\sqrt{15}}{21}, \frac{9+2\sqrt{15}}{21}\right)^T,$$
$$\hat{\mathbf{x}}_4 := \left(\frac{9+2\sqrt{15}}{21}, \frac{6-\sqrt{15}}{21}\right)^T, \quad \hat{\mathbf{x}}_5 := \left(\frac{6+\sqrt{15}}{21}, \frac{6+\sqrt{15}}{21}\right)^T,$$
$$\hat{\mathbf{x}}_6 := \left(\frac{6+\sqrt{15}}{21}, \frac{9-2\sqrt{15}}{21}\right)^T, \quad \hat{\mathbf{x}}_7 := \left(\frac{9-2\sqrt{15}}{21}, \frac{6+\sqrt{15}}{21}\right)^T$$

and the weights $\omega_1 = \frac{9}{40}$, $\omega_2 = \omega_3 = \omega_4 = \frac{155-\sqrt{15}}{1200}$, $\omega_5 = \omega_6 = \omega_7 = \frac{155+\sqrt{15}}{1200}$.
Then we define the quadrature rule

$$I(\hat{u}) := \int_{\hat{T}} \hat{u}(\hat{\mathbf{x}})\, d\hat{x} \approx \frac{1}{2} \sum_{i=1}^{7} \omega_i\, \hat{u}(\hat{\mathbf{x}}_i) =: \hat{I}(\hat{u})$$

on the reference element for functions $\hat{u} : \hat{T} \subset \mathbb{R}^2 \to \mathbb{R}$ and it holds $I(\hat{u}) = \hat{I}(\hat{u})$ for $\hat{u} \in \mathcal{P}_5(\hat{T})$.
Thus, with the standard affine transformation $\mathbf{F}_T : \hat{T} \to T$, $\hat{\mathbf{x}} \mapsto \mathbf{F}_T(\hat{\mathbf{x}}) = \mathbf{M}\hat{\mathbf{x}} + \mathbf{a}$, $\mathbf{M} \in \mathbb{R}^{2\times 2}$, $\mathbf{a} \in \mathbb{R}^2$, from $\hat{T}$ to an arbitrary element $T \in \mathcal{T}_h$, we obtain the quadrature rule

$$\begin{aligned} I(u) := \int_T u(\mathbf{x})\, dx = \int_{\hat{T}} u(\mathbf{F}_T(\hat{\mathbf{x}}))|\det \mathbf{M}|\, d\hat{x} = 2\,\mathrm{vol}(T) \int_{\hat{T}} u(\mathbf{F}_T(\hat{\mathbf{x}}))\, d\hat{x} \\ \approx \mathrm{vol}(T) \sum_{i=1}^{7} \omega_i\, u(\mathbf{F}_T(\hat{\mathbf{x}}_i)) =: \hat{I}(u) \end{aligned}$$

for functions $u : T \subset \mathbb{R}^2 \to \mathbb{R}$, since $|\det \mathbf{M}| = 2\,\mathrm{vol}(T)$, and it holds $I(u) = \hat{I}(u)$ for $u \in \mathcal{P}_5(T)$.

B.2 14 - point quadrature formula for tetrahedra (3d)

In three dimensions we start with the reference tetrahedron $\hat{T}$ defined by the vertices $(0,0,0)$, $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$. Moreover let

$$\begin{aligned} &g := \frac{1}{46\sqrt{46}}, \quad h := \arccos(g) + \frac{2}{3}\arcsin(g), \quad k := \frac{104 + 8\sqrt{46}\cos(h)}{3}, \\ &s := \sqrt{49 - k}, \quad b := \frac{7+s}{k}, \quad a := 1 - 3b, \quad d := \frac{7-s}{k}, \quad c := 1 - 3d, \\ &p := \frac{98 - k - 14s}{1680 s (b-a)^3}, \quad q := \frac{98 - k + 14s}{1680 s (c-d)^3}, \quad r := \frac{1 - 4(p+q)}{6}, \\ &e := \frac{1 + \left(\frac{2}{105r}\right)^{\frac{1}{4}}}{4}, \quad f := \frac{1-2e}{2} \end{aligned}$$

be some successively defined constants. With these constants we define the nodes

$$\begin{aligned} &\hat{\mathbf{x}}_1 := (a,b,b)^T, \quad \hat{\mathbf{x}}_2 := (b,a,b)^T, \quad \hat{\mathbf{x}}_3 := (b,b,a)^T, \quad \hat{\mathbf{x}}_4 := (b,b,b)^T, \\ &\hat{\mathbf{x}}_5 := (c,d,d)^T, \quad \hat{\mathbf{x}}_6 := (d,c,d)^T, \quad \hat{\mathbf{x}}_7 := (d,d,c)^T, \quad \hat{\mathbf{x}}_8 := (d,d,d)^T, \\ &\hat{\mathbf{x}}_9 := (e,e,f)^T, \quad \hat{\mathbf{x}}_{10} := (e,f,e)^T, \quad \hat{\mathbf{x}}_{11} := (e,f,f)^T, \quad \hat{\mathbf{x}}_{12} := (f,e,e)^T, \\ &\hat{\mathbf{x}}_{13} := (f,e,f)^T, \quad \hat{\mathbf{x}}_{14} := (f,f,e)^T, \end{aligned}$$

the weights $\omega_1 = \ldots = \omega_4 = p$, $\omega_5 = \ldots = \omega_8 = q$, $\omega_9 = \ldots = \omega_{14} = r$ and finally the quadrature rule

$$I(\hat{u}) := \int_{\hat{T}} \hat{u}(\hat{\mathbf{x}})\, d\hat{x} \approx \frac{1}{6} \sum_{i=1}^{14} \omega_i\, \hat{u}(\hat{\mathbf{x}}_i) =: \hat{I}(\hat{u})$$

on the reference element for functions $\hat{u} : \hat{T} \subset \mathbb{R}^3 \to \mathbb{R}$ and it holds $I(\hat{u}) = \hat{I}(\hat{u})$ for $\hat{u} \in \mathcal{P}_5(\hat{T})$.
Analogously as above for the two - dimensional case, using the standard affine transformation $\mathbf{F}_T : \hat{T} \to T$, $\hat{\mathbf{x}} \mapsto \mathbf{F}_T(\hat{\mathbf{x}}) = \mathbf{M}\hat{\mathbf{x}} + \mathbf{a}$, $\mathbf{M} \in \mathbb{R}^{3\times 3}$, $\mathbf{a} \in \mathbb{R}^3$, from $\hat{T}$ to an arbitrary

element $T \in \mathcal{T}_h$, we obtain the quadrature rule

$$I(u) := \int_T u(\mathbf{x})\,dx = \int_{\hat{T}} u(\mathbf{F}_T(\hat{\mathbf{x}}))|\det \mathbf{M}|\,d\hat{x} = 6\,\mathrm{vol}(T)\int_{\hat{T}} u(\mathbf{F}_T(\hat{\mathbf{x}}))\,d\hat{x}$$
$$\approx \mathrm{vol}(T)\sum_{i=1}^{14}\omega_i\,u(\mathbf{F}_T(\hat{\mathbf{x}}_i)) =: \hat{I}(u)$$

for functions $u : T \subset \mathbb{R}^3 \to \mathbb{R}$, since $|\det \mathbf{M}| = 6\,\mathrm{vol}(T)$, and it holds $I(u) = \hat{I}(u)$ for $u \in \mathcal{P}_5(T)$.

C Marking strategies

Assuming that $\mathcal{T}_h$ is an admissible triangulation of the given body Ω with decomposition $\bar{\Omega} = \bigcup_{T\in\mathcal{T}_h} T$, the least squares functionals in this work, evaluated in an approximation $(\mathbf{P}_h, \mathbf{u}_h)$, have the structure

$$\mathcal{F}(\mathbf{P}_h, \mathbf{u}_h) := \|\mathcal{R}(\mathbf{P}_h, \mathbf{u}_h)\|^2_{L^2(\Omega)} = \sum_{T\in\mathcal{T}_h}\|\mathcal{R}(\mathbf{P}_h, \mathbf{u}_h)\|^2_{L^2(T)} =: \sum_{T\in\mathcal{T}_h}\eta_T^2,$$

where η_T, $T \in \mathcal{T}_h$, are called **local error indicators**.
Without loss of generality we assume that the local error indicators are sorted in a descent order, i.e. $\eta_{T_1} \geq \eta_{T_2} \geq \ldots \geq \eta_{T_{n_t-1}} \geq \eta_{T_{n_t}}$, where n_t denotes again the number of elements in the triangulation.

C.1 Percent marking strategy

Let $p \in [0, 100]$ be arbitrary. Then we define $n \in \mathbb{N}$ as

$$n := \left\lceil \frac{p}{100}\cdot n_t \right\rceil,$$

where $\lceil x\rceil := \min\{k \in \mathbb{N} : k \geq x\}$ denotes the ceiling of a given $x \in \mathbb{R}_{\geq 0}$.
With this defined $0 \leq n \leq n_t$ the elements $T_1, \ldots, T_n$, will be marked for refinement. By construction these n elements are exactly the elements of the triangulation with largest local error indicators. The case $p = 100$ ($\Leftrightarrow n = n_t$) corresponds to an uniform refinement and the case $p = 0$ ($\Leftrightarrow n = 0$) corresponds to no refinement. Altogether one obtains a subset $\mathcal{S} := \{T_1, \ldots, T_n\} \subseteq \mathcal{T}_h$ which consists of elements that are marked for refinement.

C.2 Marking strategy of Dörfler

Let $\sigma_{\mathrm{dörf}} \in [0, 1]$ be arbitrary. The aim of Dörfler's marking strategy (cp. Section 4.2 in [Dör96]) is to seek the smallest subset $\mathcal{S} \subseteq \mathcal{T}_h$ such that

$$\sum_{T\in\mathcal{S}}\eta_T^2 \geq \sigma^2_{\mathrm{dörf}}\sum_{T\in\mathcal{T}_h}\eta_T^2. \tag{C1}$$

One starts with $\mathcal{S} = \emptyset$ and increase the set in each step by one element, starting with T_1 (the element with the largest local error indicator) and stopping at the latest on T_{n_T} (the element with the smallest local error indicator), as long as the inequality (C1) is not satisfied. The elements $T \in \mathcal{S}$ will be marked for refinement. For $\sigma_{\text{dörf}} = 0$ one obtains $\mathcal{S} = \emptyset$, i.e. no refinement will be performed, and for $\sigma_{\text{dörf}} = 1$ one gets $\mathcal{S} = \mathcal{T}_h$, i.e. all elements will be refined.

Bibliography

[ABadVLR05] F. Auricchio, L. Beirão da Veiga, C. Lovadina, and A. Reali. A stability study of some mixed finite elements for large deformation elasticity problems. *Comput. Methods Appl. Mech. Engrg.*, 194:1075 – 1092, 2005.

[ABadVLR10] F. Auricchio, L. Beirão da Veiga, C. Lovadina, and A. Reali. The importance of the exact satisfaction of the incompressibility constraint in nonlinear elasticity: mixed FEMs versus NURBS - based approximations. *Comput. Methods Appl. Mech. Engrg.*, 199:314 – 323, 2010.

[ABL+11] J. H. Adler, J. Brannick, C. Liu, T. Manteuffel, and L. Zikatanov. First - order system least squares and the energetic variational approach for two - phase flow. *J. Comput. Phys.*, 230:6647 – 6663, 2011.

[ADH+14] J. H. Adler, L. Dorfmann, D. Han, S. MacLachlan, and C. Paetsch. Mathematical and computational models of incompressible materials subject to shear. *IMA J. Appl. Math.*, 2014. To Appear.

[AE06a] H. Amann and J. Escher. *Analysis I.* Birkhäuser, Basel, 3rd edition, 2006.

[AE06b] H. Amann and J. Escher. *Analysis II.* Birkhäuser, Basel, 2nd corrected edition, 2006.

[AF03] R. A. Adams and J. J. F. Fournier. *Sobolev Spaces.* Pure and Applied Mathematics. Elsevier, Amsterdam, 2nd edition, 2003.

[AH09] K. Atkinson and W. Han. *Theoretical Numerical Analysis: A Functional Analysis Framework.* Number 39 in Texts in Applied Mathematics. Springer, Dordrecht, 3rd edition, 2009.

[Alt12] H. Altenbach. *Kontinuumsmechanik - Einführung in die materialunabhängigen und materialabhängigen Gleichungen.* Springer, Berlin, 2nd edition, 2012.

[Att12] S. Attaway. *MATLAB®: A Practical Introduction to Programming and Problem Solving.* Elsevier Butterworth - Heinemann, Oxford, 2nd edition, 2012.

[Bal77] J. M. Ball. Convexity conditions and existence theorems in nonlinear elasticity. *Arch. Rational Mech. Anal.*, 63(4):337 – 403, 1977.

[BBF13] D. Boffi, F. Brezzi, and M. Fortin. *Mixed Finite Element Methods and Applications.* Number 44 in Springer Series in Computational Mathematics. Springer, Heidelberg, 2013.

[Bey95] J. Bey. Tetrahedral grid refinement. *Computing*, 55:355 – 378, 1995.

[BG09] P. B. Bochev and M. D. Gunzburger. *Least - Squares Finite Element Methods*. Number 166 in Applied Mathematical Sciences. Springer, New York, 2009.

[BMS14a] F. Bertrand, S. Münzenmaier, and G. Starke. First - order system least squares on curved boundaries: Higher - order Raviart - Thomas elements. *SIAM J. Numer. Anal.*, 2014. To Appear.

[BMS14b] F. Bertrand, S. Münzenmaier, and G. Starke. First - order system least squares on curved boundaries: Lowest - order Raviart - Thomas elements. *SIAM J. Numer. Anal.*, 52(2):880 – 894, 2014.

[Bra07] D. Braess. *Finite elements: Theory, fast solvers, and applications in solid mechanics*. Cambridge University Press, Cambridge, 3rd edition, 2007.

[BS92] I. Babuška and M. Suri. Locking effects in the finite element approximation of elasticity problems. *Numer. Math.*, 62:439 – 463, 1992.

[BS08] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. Number 15 in Texts in Applied Mathematics. Springer, New York, 3rd edition, 2008.

[BSN10] D. Balzani, J. Schröder, and P. Neff. Applications of anisotropic polyconvex energies: thin shells and biomechanics of arterial walls. In J. Schröder and P. Neff, editors, *Poly -, Quasi - and Rank - One Convexity in Applied Mechanics*, volume 516 of *CISM Courses and Lectures*, pages 131 – 175. Springer, Wien, 2010.

[CA 69] R. D. Cook and J. K. Al - Abdulla. Some plane quadrilateral "hybrid" finite elements. *AIAA J.*, 7(11):2184 – 2185, 1969.

[Car04] C. Carstensen. An adaptive mesh - refining algorithm allowing for an H^1 stable L^2 projection onto Courant finite element spaces. *Constr. Approx.*, 20:549 – 564, 2004.

[Cia88] P. G. Ciarlet. *Mathematical Elasticity Volume I: Three - dimensional elasticity*, volume 20 of *Studies in Mathematics and its Applications*. Elsevier North - Holland, Amsterdam, 1988.

[CKS05] Z. Cai, J. Korsawe, and G. Starke. An adaptive least squares mixed finite element method for the stress - displacement formulation of linear elasticity. *Numer. Methods Partial Differential Equations*, 21(1):132 – 148, 2005.

[Coo74] R. D. Cook. Improved two - dimensional finite element. *J. Struct. Div. ASCE*, 100(9):1851 – 1863, 1974.

[Cow73] G. R. Cowper. Gaussian quadrature formulas for triangles. *Int. J. Numer. Meth. Engng.*, 7(3):405 – 408, 1973.

[CS03] Z. Cai and G. Starke. First - order system least squares for the stress - displacement formulation: Linear elasticity. *SIAM J. Numer. Anal.*, 41(2):715 – 730, 2003.

[CS04] Z. Cai and G. Starke. Least - squares methods for linear elasticity. *SIAM J. Numer. Anal.*, 42(2):826 – 842, 2004.

[CW09] Z. Cai and C. R. Westphal. An adaptive mixed least - squares finite element method for viscoelastic fluids of Oldroyd type. *J. Non - Newt. Fluid Mech.*, 159:72 – 80, 2009.

[Dem03] W. Demtröder. *Experimentalphysik 1: Mechanik und Wärme*. Springer, Berlin, 3rd edition, 2003.

[Dör96] W. Dörfler. A convergent adaptive algorithm for Poisson's equation. *SIAM J. Numer. Anal.*, 33(3):1106 – 1124, 1996.

[EGK11] C. Eck, H. Garcke, and P. Knabner. *Mathematische Modellierung*. Springer, Heidelberg, 2nd revised edition, 2011.

[ESN10] V. Ebbing, J. Schröder, and P. Neff. Construction of polyconvex energies for non - trivial anisotropy classes. In J. Schröder and P. Neff, editors, *Poly -, Quasi - and Rank - One Convexity in Applied Mechanics*, volume 516 of *CISM Courses and Lectures*, pages 107 – 130. Springer, Wien, 2010.

[FS83] M. Fortin and M. Soulie. A non - conforming piecewise quadratic finite element on triangles. *Int. J. Numer. Meth. Engng.*, 19:505 – 520, 1983.

[GH91] M. Gellert and R. Harbord. Moderate degree cubature formulas for 3 - d tetrahedral finite - element approximations. *Commun. Appl. Numer. Methods*, 7:487 – 495, 1991.

[HJ13] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 2nd edition, 2013.

[HMW11] R. Herzog, C. Meyer, and G. Wachsmuth. Integrability of displacement and stresses in linear and nonlinear elasticity with mixed boundary conditions. *J. Math. Anal. Appl.*, 382:802 – 813, 2011.

[HR13] W. Han and B. D. Reddy. *Plasticity: Mathematical Theory and Numerical Analysis*. Number 9 in Interdisciplinary Applied Mathematics. Springer, New York, 2nd edition, 2013.

[Jia98] B.-n. Jiang. *The Least-Squares Finite Element Method-Theory and Applications in Computational Fluid Dynamics and Electromagnetics*. Springer, Berlin, 1998.

[MMSW06] T. A. Manteuffel, S. F. McCormick, J. G. Schmidt, and C. R. Westphal. First-order system least squares for geometrically nonlinear elasticity. *SIAM J. Numer. Anal.*, 44(5):2057–2081, 2006.

[MS11] S. Münzenmaier and G. Starke. First-order system least squares for coupled Stokes-Darcy flow. *SIAM J. Numer. Anal.*, 49(1):387–404, 2011.

[MSSS14] B. Müller, G. Starke, A. Schwarz, and J. Schröder. A first-order system least squares method for hyperelasticity. *SIAM J. Sci. Comput.*, 36(5):B795–B816, 2014.

[Mün12] S. Münzenmaier. *Least-Squares Finite Element Methods for Coupled Generalized Newtonian Stokes-Darcy Flow*. Ph.d. thesis, Leibniz Universität Hannover, 2012.

[NW03] P. Neff and C. Wieners. Comparison of models for finite plasticity: A numerical study. *Comput Vis. Sci.*, 6:23–35, 2003.

[NW06] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2nd edition, 2006.

[Rao10] A. Raoult. Quasiconvex envelopes in nonlinear elasticity. In J. Schröder and P. Neff, editors, *Poly-, Quasi- and Rank-One Convexity in Applied Mechanics*, volume 516 of *CISM Courses and Lectures*, pages 17–51. Springer, Wien, 2010.

[Riv84] M. C. Rivara. Algorithms for refining triangular grids suitable for adaptive and multigrid techniques. *Int. J. Numer. Meth. Engng.*, 20:745–756, 1984.

[Rös00] A. Rössle. Corner singularities and regularity of weak solutions for the two-dimensional Lamé equations on domains with angular corners. *J. Elasticity*, 60:57–75, 2000.

[Sch09] A. Schwarz. *Least-Squares Mixed Finite Elements for Solid Mechanics*. Bericht Nr. 7/ Institut für Mechanik, Universität Duisburg-Essen, Abt. Bauwissenschaften, Essen, 2009. Ph. D. thesis.

[Sch10] J. Schröder. Anisotropic polyconvex energies. In J. Schröder and P. Neff, editors, *Poly-, Quasi- and Rank-One Convexity in Applied Mechanics*, volume 516 of *CISM Courses and Lectures*, pages 53–105. Springer, Wien, 2010.

[Sim98] J. C. Simo. Numerical analysis and simulation of plasticity. In P. G. Ciarlet and J. L. Lions, editors, *Handbook of Numerical Analysis*, volume VI, pages 183 – 499. Elsevier North - Holland, Amsterdam, 1998.

[SN03] J. Schröder and P. Neff. Invariant formulation of hyperelastic transverse isotropy based on polyconvex free energy functions. *Int. J. Solids Struct.*, 40:401 – 445, 2003.

[SRO07] E. Stein, M. Rüter, and S. Ohnimus. Error - controlled adaptive goal - oriented modeling and finite element approximations in elasticity. *Comput. Methods Appl. Mech. Engrg.*, 196:3598 – 3613, 2007.

[SSS10] A. Schwarz, J. Schröder, and G. Starke. A modified least - squares mixed finite element with improved momentum balance. *Int. J. Numer. Meth. Engng.*, 81:286 – 306, 2010.

[SSS11] G. Starke, A. Schwarz, and J. Schröder. Analysis of a modified first - order system least squares method for linear elasticity with improved momentum balance. *SIAM J. Numer. Anal.*, 49(3):1006 – 1022, 2011.

[Sta07] G. Starke. An adaptive least - squares mixed finite element method for elasto - plasticity. *SIAM J. Numer. Anal.*, 45(1):371 – 388, 2007.

[SWB11] J. Schröder, P. Wriggers, and D. Balzani. A new mixed finite element based on different approximations of the minors of deformation tensors. *Comput. Methods Appl. Mech. Engrg.*, 200:3583 – 3600, 2011.

[Vil85] A. Villani. Another note on the inclusion of $L^p(\mu) \subset L^q(\mu)$. *Am. Math. Mon.*, 92(7):485 – 487, 1985.

[Wer05] D. Werner. *Funktionalanalysis*. Springer, Berlin, 5th extended edition, 2005.

[Wri08] P. Wriggers. *Nonlinear Finite Element Methods*. Springer, Berlin, 2008.

[Zei13] E. Zeidler, editor. *Springer - Taschenbuch der Mathematik*. Springer, Wiesbaden, 3rd edition, 2013.

Benjamin Müller

Curriculum vitae

Personal data

Date of birth	March 21, 1985
Place of birth	31061 Alfeld (Leine), Germany
Nationality	German

Professional career

since April 2013 **Research associate**, *Faculty of Mathematics of the University Duisburg - Essen*, Essen, Workgroup Prof. Dr. Starke: Numerical Mathematics.

Activities: research (working on the DFG project: „Gemischte Least - Squares Finite Elemente für geometrisch nichtlineare Probleme der Festkörpermechanik“, plasticity), educational assistance (lecture, exercise and programming courses)

2011 - 2013 **Research associate**, *Institute of Applied Mathematics of the Leibniz University Hanover*, Workgroup Prof. Dr. Starke: Scientific Computing.

Activities: research (working on the DFG project: „Gemischte Least - Squares Finite Elemente für geometrisch nichtlineare Probleme der Festkörpermechanik“), educational assistance (exercise courses)

Education

Academic studies

since April 2013 **Ph.D. student in Mathematics**, *University Duisburg - Essen*, Essen.

2011 - 2013 **Ph.D. student in Mathematics**, *Leibniz University Hanover*, Hanover.

2007 - 2010 **Diploma in Mathematics**, *Leibniz University Hanover*, 2nd degree program.

Final grade: *with distinction*, minor subject: physics

Specialization: Numerical methods for partial differential equations

Diploma thesis: „A posteriori Fehlerkontrolle für elliptische Eigenwertprobleme“

2004 - 2010 **Teaching degree for secondary schools (1st state examination)**, *Leibniz University Hanover*, 1st degree program.

Final grade: 1.3, major subjects: mathematics, physics

Schooling

1997 - 2004 Secondary school in Alfeld

Graduation: Abitur (Final grade: 2.8)

1995 - 1997 Secondary school in Alfeld (orientation stage)

Thea - Leymann - Straße 9 – 45127 Essen, Germany
☎ +49 201 183 4199 • FAX +49 201 183 2601
✉ benjamin.mueller@uni-due.de
https://www.uni-due.de/mathematik/agstarke/bmueller.php

1994 - 1995	Bürgerschule Alfeld, primary school
1992 - 1994	Dohnser Schule in Alfeld, primary school
1991 - 1992	Primary school in Föhrste

Programming - and computer skills

Programming	MATLAB, C++, Scheme, LaTeX	Office	Microsoft Office (Word, Excel, PowerPoint)

Language skills

German	**native**
English	**very good**
French	**basic**

Publications

[1] B. Müller, G. Starke, A. Schwarz, and J. Schröder. *A First - Order System Least Squares Method for Hyperelasticity.* SIAM Journal on Scientific Computing, Vol. 36, No. 5 (2014), pp. B795 - B816

Personal interests

Sport	fitness, cycling, swimming, having a sauna, go - kart, soccer
Movies	cinema, Blu - ray at home

Thea - Leymann - Straße 9 – 45127 Essen, Germany
☎ +49 201 183 4199 • FAX +49 201 183 2601
✉ benjamin.mueller@uni-due.de
https://www.uni-due.de/mathematik/agstarke/bmueller.php