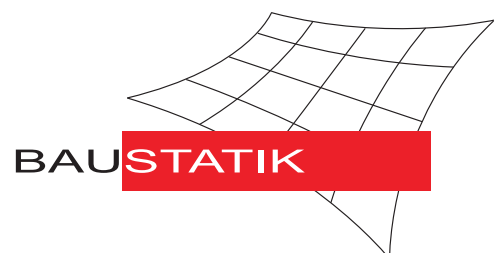


Physikalisch konsistente Modellierung von elastischem und plastischem Materialverhalten mit künstlichen neuronalen Netzen

Patrick Weber



**Physikalisch konsistente Modellierung von
elastischem und plastischem Materialverhalten
mit künstlichen neuronalen Netzen**

Zur Erlangung des akademischen Grades eines

DOKTOR-INGENIEURS

von der KIT-Fakultät für

Bauingenieur-, Geo- und Umweltwissenschaften

des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Patrick Thomas Weber, M.Sc.

aus Saarlouis

Tag der mündlichen Prüfung: 20. Oktober 2023

Referent: Prof. Dr.-Ing. habil. Werner Wagner

Korreferenten: Prof. Dr.-Ing. habil. Sven Klinkel

Prof. Dr.-Ing. Steffen Freitag

Karlsruhe 2023

Herausgeber:

Prof. Dr.-Ing. Steffen Freitag
Prof. Dr.-Ing. habil. Werner Wagner

Organisation und Verwaltung:

Institut für Baustatik
Karlsruher Institut für Technologie
Kaiserstr. 12
76131 Karlsruhe

Telefon: (0721) 608-42280
Telefax: (0721) 608-46015
E-Mail: info@ibs.kit.edu

© 2023 Patrick Thomas Weber
Institut für Baustatik
Karlsruher Institut für Technologie
Kaiserstraße 12
76131 Karlsruhe

Das vorliegende Werk darf von der KIT-Bibliothek frei im Internet angeboten werden. Die Nutzung erfolgt ausschließlich zu wissenschaftlichen Zwecken und zum Eigengebrauch. Die Urheberrechte liegen bei den Autorinnen bzw. Autoren. Für den Inhalt sind alleine die Autorinnen und Autoren verantwortlich. Jegliche Formen der kommerziellen Nutzung und Abänderung der Publikation sind ohne vorherige Zustimmung und Absprache mit den Autorinnen bzw. Autoren ausdrücklich verboten. Die Namen der Autorinnen und Autoren müssen stets genannt werden. Die Nutzerinnen und Nutzer sind für die Einhaltung der Rechtsvorschriften selbst verantwortlich und können bei Missbrauch haftbar gemacht werden.

ISBN 978-3-935322-28-7

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als Akademischer Mitarbeiter am Institut für Baustatik des Karlsruher Instituts für Technologie.

Mein besonderer Dank gilt meinem Hauptreferenten und ehemaligen Institutsleiter Prof. Dr.-Ing. habil. Werner Wagner, der mir den Weg zur Promotion geebnet hat. Die aus meiner Sicht ideale Mischung aus wissenschaftlichem Freiraum und flexibler fachlicher Unterstützung hat maßgeblich zum Gelingen dieser Arbeit beigetragen. Ferner führte das von Prof. Wagner vorgelebte konstruktive Arbeitsklima zu exzellenten Rahmenbedingungen für Forschung und Lehre.

Dem neuen Institutsleiter, Prof. Dr.-Ing. Steffen Freitag, danke ich ebenfalls vielmals für das mir entgegengebrachte Vertrauen, die eingebrachte fachliche Expertise und die Übernahme eines der beiden Korreferate. Weiter danke ich Prof. Dr.-Ing. habil. Sven Klinkel für sein Interesse an meiner wissenschaftlichen Arbeit und der Übernahme eines Korreferats.

Alle Mitarbeiterinnen und Mitarbeiter des Instituts für Baustatik trugen in erheblichem Maße zu einer angenehmen und familiären Atmosphäre bei. Mein Dank gilt insbesondere den mit mir promovierenden Kolleginnen und Kollegen, die mich durch fachliche Diskussionen und Ratschläge stets unterstützt haben. Im Rahmen gemeinsamer Konferenzbesuche sowie zahlreicher Freizeitunternehmungen wurden diese Kollegen schließlich zu Freunden, sodass ich in Zukunft immer mit einem Lächeln an die gemeinsame Zeit am Institut zurückblicken werde.

Die Lehre hat mir unentwegt tiefe Freude bereitet. Ich habe in hohem Maße fachlich sowie menschlich von ihr profitiert. Dabei hatte ich das Glück, mit motivierten und begeisterten studentischen Hilfskräften zu arbeiten. Außerdem durfte ich beim Betreuen spannender Bachelor- und Masterarbeiten meinen Horizont weit über die eigenen Forschungsthemen hinaus erweitern. Das dabei entgegengebrachte Engagement und die Wissbegierde haben mich stets dazu motiviert, viel Zeit und Leidenschaft in die Lehre zu investieren.

Von ganzem Herzen danke ich meinen Eltern Petra und Ralf, meiner Schwester Lisa-Marie und den jeweiligen Partnern für den Rückhalt und das aufgebrachte Verständnis während dieses zeitintensiven Lebensabschnitts. Ganz besonders gilt mein Dank meiner Freundin Malwine, die mir mit ihrer uneingeschränkten Unterstützung, Liebe und Motivation stets zur Seite gestanden hat.

Karlsruhe, im Oktober 2023

Patrick Weber

Kurzfassung

Die vorliegende Arbeit befasst sich mit der physikalisch konsistenten Modellierung von elastischem und plastischem Materialverhalten mit künstlichen neuronalen Netzen (KNN).

Die besagte Konsistenz wird hauptsächlich durch Einforderung physikalisch motivierter Nebenbedingungen innerhalb des KNN-Trainingsprozesses erzielt. Dazu wird das Training der KNN im Kontext restringierter Optimierungsprobleme formuliert. Die Lösung der resultierenden Integrale erfolgt mithilfe der Monte-Carlo-Integration. Im Fall plastischer KNN-Materialmodelle wird eine effiziente Methode zur Generierung der zugehörigen Integrationspunkte präsentiert. Für die beschriebenen KNN-Architekturen werden darüber hinaus effiziente Algorithmen zur Implementierung bereitgestellt.

Die Nebenbedingungen fordern physikalische Eigenschaften schwach ein. Bei der Modellierung elastischen Materialverhaltens werden diese Nebenbedingungen mit KNN-Architekturen verglichen, die gewisse Eigenschaften a priori und exakt erfüllen. Dabei zeigt sich, dass die positiven Effekte der zusätzlichen Informationsquelle den Trainingsprozess des KNN und die nachträgliche Nutzung des KNN-Materialmodells in ähnlichem Maße stabilisieren. Dies motiviert die Nutzung der Nebenbedingungen auch für Materialverhalten, bei dem noch keine vergleichbaren KNN-Architekturen vorliegen.

Neben dem Einfluss der Nebenbedingungen werden außerdem verschiedene, bei der praktischen Anwendung relevante Sachverhalte untersucht. Dies schließt das Verhalten elastischer KNN-Materialmodelle bei Quasi-Inkompressibilität und die Schrittweitenabhängigkeit im Fall plastischer KNN-Materialmodelle ein.

Die elastischen und plastischen KNN-Materialmodelle werden jeweils innerhalb der Methode der finiten Elemente (FEM) zur Struktursimulation genutzt. Die am Materialpunkt gewonnenen Erkenntnisse können dabei auf die Anwendung am Gesamttragwerk übertragen werden. Für die plastischen KNN-Materialmodelle wird zudem die Implementierung in ein inkrementell-iteratives Lösungsverfahren der FEM detailliert beschrieben.

Das Prinzip der KNN-Materialmodellierung mit physikalisch motivierten Nebenbedingungen wird auf reale Versuchsdaten dreier unterschiedlicher Materialien angewendet. Diese Daten umfassen das elastische Verhalten eines Elastomers unter großen Verzerrungen, das anisotrope und nichtlineare Verhalten von biaxial und quer zur Faser belastetem Fichtenholz sowie das gekoppelte Verhalten einer uniaxial mechanisch und elektrisch beanspruchten Probe einer Piezokeramik. In allen drei Fällen erlauben die Nebenbedingungen das Erstellen stabiler und zuverlässiger KNN-Materialmodelle.

Abstract

The present work deals with physically consistent modeling of elastic and plastic material behavior with artificial neural networks (ANN).

This consistency is mainly achieved by enforcing physically motivated restrictions within the ANN training process. For this purpose, the training process is formulated in the context of constrained optimization problems. The resulting integrals are solved by Monte Carlo integration. For plastic ANN material models, an efficient method for generating the necessary integration points is presented. For all described ANN architectures efficient algorithms for implementation are provided.

The constraints enforce physical properties in a weak sense only. Therefore, in the case of elastic material behavior, these constraints are compared to ANN architectures which satisfy certain physical properties a priori and exactly. It turns out, that the positive effects of the additional information source, i.e. the constraints, stabilize the training process of the ANN and the subsequent use as material model equally well. This motivates the use of physically motivated constraints for material behaviors, where comparable architectures are not yet available.

In addition to the influence of the constraints, various aspects relevant to the application of ANN material models are investigated. These include the case of near incompressibility for elasticity and the step-size dependency of plastic ANN material models.

The elastic and plastic ANN material models are used for structural simulation within the finite element method (FEM). The insights gained at the material point level can be transferred to the structural level. In the case of plasticity, the implementation into the FEM's incremental iterative solution scheme is described in detail.

The principle of ANN material modeling with physically motivated constraints is applied to real world experimental data with different properties. These databases include the elastic behavior of a rubber material under large deformations, the anisotropic and nonlinear behavior of biaxially and transversely loaded spruce wood and the coupled behavior of a uniaxially mechanically and electrically loaded sample of a piezoceramic. In all three cases, the constraints allow the construction of stable and reliable ANN material models.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Stand der Forschung	3
1.3	Ziele und Gliederung der Arbeit	5
2	Mechanische und numerische Grundlagen	8
2.1	Grundgleichungen der Kontinuumsmechanik	8
2.1.1	Kinematik: Geometrie und Deformation	8
2.1.2	Das Konzept der Spannungen	12
2.1.3	Die starke Form des Gleichgewichts	14
2.1.4	Die schwache Form des Gleichgewichts und das PVV . . .	16
2.1.5	Linearisierung des PVV	18
2.1.6	Reduzierte Notation des Materialgesetzes	19
2.2	Die Methode der finiten Elemente	20
2.2.1	Diskretisierung eines Körpers in finite Elemente	20
2.2.2	Die diskretisierte Linearisierung der schwachen Form . . .	23
2.2.3	Ein inkrementell-iteratives Lösungsverfahren	26
2.2.4	Spezielle Strukturelemente	27
3	Phänomenologische Materialmodelle	33
3.1	Passive Transformation von Tensoren	34
3.2	Physikalische Prinzipien der Materialmodellierung	37
3.2.1	Die Dissipationsungleichung der Thermodynamik	38
3.2.2	Das Prinzip der materiellen Objektivität	39
3.2.3	Materielle Symmetrien	42
3.2.4	Inkompressibilität als kinematische Zwangsbedingung . . .	43
3.3	Klassische elastische Materialmodelle	44
3.3.1	Isotrope hyperelastische Materialmodelle	44
3.3.2	Lineares Green'sches Materialgesetz	47
3.3.3	OGDEN Materialgesetz	49
3.3.4	Grenzfall der Inkompressibilität	51

3.3.5	Polykonvexität	52
3.4	Klassische elastoplastische Materialmodelle	54
3.4.1	Einordnung in den thermodynamischen Rahmen	57
3.4.2	Das Prinzip der maximalen plastischen Dissipationsleistung	58
3.4.3	Projektionsverfahren	59
3.4.4	Von Mises'sche Fließbedingung ohne Verfestigung	60
3.4.5	Von Mises'sche Fließbedingung mit gemischter Verfestigung	64
4	Künstliche neuronale Netze	67
4.1	Normalisierung und Daten-Fehlerfunktion	71
4.2	Das klassische vollverknüpfte Feedforward-Netz	72
4.2.1	Die FNN-Vorwärtsrechnung	73
4.2.2	Die FNN-JACOBI-Matrix	76
4.2.3	Erweiterte FNN-Vorwärtsberechnungen für alle Trainings- daten	79
4.2.4	Der Fehlergradient bezüglich FNN-Ausgangsvariablen . . .	80
4.2.5	Der Fehlergradient bezüglich FNN-Ableitungen	83
4.3	Weitere KNN-Architekturen	87
4.3.1	Das Input Convex Neural Network (ICNN)	87
4.3.2	Das mehrschichtige rekurrente Elman-Netz	93
4.4	Trainingsalgorithmen für KNN	100
4.4.1	Das Adam-Optimierungsverfahren	101
4.4.2	Das Quasi-Newton Optimierungsverfahren	102
4.4.3	Abbruchkriterien für das Training	105
4.5	KNN-Training als restringiertes Optimierungsproblem	107
4.5.1	Die klassische L2-Penalty-Methode	109
4.5.2	Die exakte L1-Penalty-Methode	110
4.5.3	Die Augemented-Lagrange-Methode	110
4.5.4	Nichtlineare Zwangsbedingungen über ZB-Daten	111
4.6	Die 2D-Sinc-Funktion als Prinzipbeispiel	114
4.6.1	Generierung der Trainings-, Validierungs- und NB-Punkte	115
4.6.2	Parameterstudien und Untersuchungen zum Training . . .	116

4.6.3	Training mit der Nebenbedingung Rotationssymmetrie . . .	118
4.6.4	Schlussfolgerungen über das Training mit Nebenbedingungen	121
5	Modellierung von elastischem	
	Materialverhalten mit KNN	122
5.1	Datengenerierung mit Datenpunkt-Materialversuchen	122
5.1.1	Ein gemischt gesteuerter Materialpunkt-Versuch	123
5.1.2	Anreicherung der Trainingsdaten	125
5.2	KNN-Materialmodelle für elastisches	
	Materialverhalten (KNN-E)	126
5.2.1	Anisotropes KNN-Materialmodell FNN-E-Ani	127
5.2.2	Anisotropes KNN-Materialmodell NCNN-E-Ani	127
5.2.3	Isotropes KNN-Materialmodell FNN-E-Iso	128
5.2.4	Polykonvexes isotropes KNN-Materialmodell ICNN-E-Poly	133
5.2.5	Exakte Einforderung der Inkompressibilitätsbedingung . .	135
5.2.6	Erfüllung der Wachstumsbedingungen	136
5.3	Physikalische Nebenbedingungen für E-KNN	137
5.3.1	Spannungsfreie Referenzkonfiguration	137
5.3.2	Energieerhaltung bzw. Symmetrie der Materialtangente . .	138
5.3.3	Isotropie	140
5.4	Untersuchungen am Materialpunkt für E-KNN	141
5.4.1	Funktionsfähigkeit aller KNN-Materialmodelle	142
5.4.2	Untersuchungen zur Energieerhaltung	144
5.4.3	KNN-Materialmodelle bei Quasi-Inkompressibilität	147
5.4.4	Verhalten bei großen Verzerrungen	151
5.5	Anwendung der KNN-E-Materialmodelle in der FEM	152
5.5.1	Scheibe mit Loch: Systembeschreibung und Referenzlösung	153
5.5.2	Vergleich der vier KNN-Materialmodelle an der Lochscheibe	155
5.5.3	Abschließende Bemerkungen	157

6	Modellierung von plastischem Materialverhalten mit KNN	160
6.1	KNN-Materialmodelle für plastisches Materialverhalten (KNN-P)	161
6.1.1	Geschichtsvariablen zur Modellierung von Verfestigung	162
6.1.2	Inkrementelles KNN-Materialmodell FNN-P-Ink	164
6.1.3	Sequenzielles KNN-Materialmodell FNN-P-Seq	166
6.1.4	Inkrementelles KNN-Materialmodell NCNN-P-Ink	167
6.1.5	RNN-Materialmodell RNN-P	168
6.2	Physikalische Nebenbedingungen für KNN-P	169
6.2.1	Datengenerierung innerhalb einer konvexen Hülle	170
6.2.2	Inkrementelle Normalisierung	171
6.2.3	Stationarität der Normalisierung	172
6.2.4	Symmetrie der Materialtangente	173
6.2.5	Inkrementelle Stabilität	173
6.2.6	Energiedissipation in direktem geschlossenen Verzerrungszyklus	174
6.3	Untersuchungen am Materialpunkt für P-KNN	176
6.3.1	Zufällige Generierung von Verzerrungspfaden	177
6.3.2	Funktionsfähigkeit aller KNN-Materialmodelle	178
6.3.3	Diskussion zum RNN-Materialmodell	180
6.3.4	Diskussion zur KNN-Materialtangente	181
6.3.5	Einfluss der physikalischen Nebenbedingungen	183
6.3.6	Diskussion zur Schrittweitenabhängigkeit	185
6.4	Approximation einer Fließfläche	187
6.5	Anwendung der KNN-P-Materialmodelle in der FEM	192
6.5.1	Aluminiumscheibe mit zentralem Loch	192
6.5.2	Eingespanntes U-Profil	198
6.5.3	Abschließende Bemerkungen	201

7	Anwendung auf experimentelle Daten	204
7.1	Treloars Daten von vulkanisiertem Gummi	205
7.1.1	Aufbereitung der Versuchsdaten	206
7.1.2	Training des KNN-Materialmodells	208
7.1.3	Anwendung des KNN-Materialmodells auf die Daten	210
7.1.4	Abschließende Bemerkungen	212
7.2	Biaxial beanspruchtes Fichtenholz	213
7.2.1	Versuchsdaten: Beschreibung und Aufbereitung	214
7.2.2	Eine Grundsatzdiskussion anhand der linearen Approxima- tion	219
7.2.3	Anwendung des nichtlinearen KNN-Materialmodells	222
7.2.4	Abschließende Bemerkungen	228
7.3	Ferroelektrische Funktionskeramik	230
7.3.1	Aufbereitung der Versuchsdaten	233
7.3.2	Anwendung eines KNN-Materialmodells	236
7.3.3	Abschließende Bemerkungen	239
8	Zusammenfassung und Ausblick	240
A	Elemente der Differenzialrechnung	244
B	Reduzierte Notation von Tensoren	245
C	Strategien der inexakten Liniensuche	247
C.1	Armijo-Liniensuche	247
C.2	Wolfe-Powell-Liniensuche	248
D	Ein inkompressibler Materialpunkt-Versuch	250
	Literaturverzeichnis	253

1 Einleitung

1.1 Motivation

Der Mensch zeichnet sich aus Sicht des Autors insbesondere durch seinen Drang aus, die Welt um ihn herum zu begreifen, zu interpretieren, durch das gewonnene Wissen in die Zukunft zu extrapolieren und somit letztlich die Umwelt zu seinen Gunsten zu verändern. Im Rahmen dieses philosophischen Kontextes schlagen Ingenieurwissenschaften wie das Bauingenieurwesen die Brücke zwischen den in den Naturwissenschaften gewonnenen, grundlegenden Erkenntnissen und deren technischer Umsetzung. Die **Baustatik** ist im Bauingenieurwesen ein zentrales Forschungsgebiet, das sich sowohl im Bereich der Grundlagenforschung als auch anwendungsbezogen mit der Modellierung von Tragwerken sowie der Beurteilung ihrer Standsicherheit und Zuverlässigkeit beschäftigt.

KURRER [109] lädt mit seinem Buch auf eine spannende Reise durch die Geschichte der Baustatik ein. Die Historie mechanischer und dadurch u. a. auch baustatischer Prinzipien ist zudem in SZABO [167] zu finden. Wie und mit welchen Methoden geforscht wird und welche Verfahren in der Praxis angewendet werden, hat sich im Laufe der letzten Jahrhunderte stetig geändert. Die Baustatik durchlief dabei – wie viele Wissenschaftszweige – einige **Paradigmenwechsel**. Bauen war für sehr lange Zeit vor allem empirischer Natur: Erfahrungen wurden gesammelt und im Laufe der Jahrhunderte konsolidiert. Dieses erste Paradigma wurde mit dem Aufkommen mathematischer und mechanischer Methoden und den daraus entstehenden Modellen weitestgehend abgelöst, was dem zweiten Paradigma entspricht. Zunächst meist noch geometrisch, dann auch theoretisch bspw. auf Basis der Festigkeitslehre wurden Gleichungen für Strukturen wachsender Komplexität definiert sowie deren Lösungen zur Dimensionierung von Bauteilen genutzt. Bis heute werden diese Modelle weiter verfeinert. Sie gingen schließlich fließend in die Zeit des dritten Paradigmas über, welche die numerische Lösung der Gleichungen mithilfe immer leistungsstärkerer Computer umfasst. Dabei werden die theoretischen Modelle nicht ersetzt, sondern vielmehr ergänzt. So ist die Methode der finiten Elemente (FEM) zur Simulation von Tragwerken aus der Praxis nicht mehr wegzudenken. Heute rückt ein viertes Paradigma der Wissenschaft in den Vordergrund: **datenbasierte Methoden**, vgl. HEY ET AL. [77] sowie AGRAWAL und CHOUDHARY [3]. Modelle werden nicht vorgegeben und an wenigen Daten kalibriert, sondern direkt mithilfe einer großen Menge Daten erzeugt – oder die Modelle werden gar nicht erst benötigt. Überall dort, wo stets viele Daten zur Verfügung standen, nahm die Entwicklung hin zu datenbasierten Methoden bereits Fahrt auf. Aus Sicht des Autors wird dieses vierte Paradigma die vorherrschenden Methoden der Baustatik nicht ersetzen, wohl aber punktuell ergänzen.

Durch die Vielfalt einsetzbarer Konstruktionsbaustoffe war das Bauingenieurwesen stets eng verknüpft mit der **qualitativen und quantitativen Beschreibung des zugehörigen Materialverhaltens**. Auch dieser Teilbereich wird sich in Zukunft mit datenbasierten Methoden auseinandersetzen. Die klassische Materialmodellierung basiert auf der Definition geeigneter Funktionsgleichungen, deren Parameter mit experimentellen Daten kalibriert werden. Je komplexer das Materialverhalten und je mehr physikalische Einschränkungen dabei berücksichtigt werden sollen, desto schwieriger ist das Finden solcher Gleichungen, vgl. TRUESDELL und NOLL [175]. Da durch die Verbesserung von Mess- und Auswertungsmethoden die Menge an verfügbaren Daten stetig wächst, ist die Ergänzung von klassischen physikalisch motivierten, starren Modellen durch vorwiegend mathematische Methoden, die diese Daten direkt verarbeiten können, nur natürlich, vgl. ALTENBACH [4].

So basiert die u. a. von ORTIZ ins Leben gerufene Richtung *data driven mechanics* auf einer Neuausrichtung des numerischen Lösers. Klassischerweise wird das Gleichgewicht als zu lösende Gleichung direkt eingefordert. Im Fall dieses neuen Lösungsschemas wird jedoch die Distanz eines jeden Materialpunkts zu gegebenen Datenpunkten minimiert, während Gleichgewicht und Kinematik als Nebenbedingungen eingehen. Seit dem Pionierartikel von KIRCHDOERFER und ORTIZ [101] wurde die Methode bereits u. a. auf die Mehrskalenmodellierung (vgl. KARAPIPERIS ET AL. [97]), die Bruchmechanik (vgl. CARRARA ET AL. [25]) und die Dynamik (vgl. KIRCHDOERFER und ORTIZ [102]) übertragen. Ein anderer Ansatz wird bspw. von der Gruppe um DE LORENZIS verfolgt, die das Erzeugen von Materialmodellen auf Grundlage vordefinierter Basisfunktionen anstrebt. Die Methode funktioniert mit Strukturdaten, also Verschiebungsfeldern und Kräften, die in sehr großer Zahl generiert werden, und benötigt daher keine Information über die Verzerrungen und Spannungen. Deren Methode wurde bspw. auf Hyperelastizität und Elastoplastizität angewendet, vgl. FLASCHEL ET AL. [43,44].

Während der *data driven*-Ansatz gar kein Modell mehr für das Materialverhalten definiert, bedient sich der Ansatz auf Basis der Strukturdaten einer Reihenentwicklung von zulässigen und schon bekannten Materialfunktionen, die indirekt über Strukturantworten kalibriert werden. Ein alternativer Weg ist die **Materialmodellierung mit künstlichen neuronalen Netzen** (KNN). Das KNN entspricht dem Modell und besteht aus einer beliebig komplexen Verkettung einfacher Basisfunktionen. Durch diese Struktur sind KNN *universelle Funktionsapproximatoren*: Sie können bei endlichem Verkettungsgrad Funktionen mit beliebiger Genauigkeit approximieren, vgl. CYBENKO [31] sowie KIDGER und LYONS [99]. Das macht sie zum idealen Kandidaten für einen Mittelweg zwischen den beiden genannten Extremen der klassischen und der vollständig modellfreien Methoden.

Ganz im Zeitgeist aufkommender Methoden der künstlichen Intelligenz gewinnen KNN in der Mechanik bzw. Baustatik im Rahmen der Materialmodellierung stetig mehr Raum – zumindest in der Forschung. Zum Beispiel wurden sog. *Physics informed neural networks* (PINN) von RAISSI ET AL. [145] eingeführt zur Lösung partieller Differenzialgleichungen, welche u. a. auf mikromechanische Feldprobleme angewendet wurden von HENKES ET AL. [75]. FREITAG ET AL. [49] reformulieren ein rekurrentes neuronales Netz (RNN) mithilfe von Fuzzy-Logik zur Berücksichtigung von Unschärfen in zeitabhängigen Prozessen. Darüber hinaus wurde Unschärfe im Rahmen der KNN-Materialmodellierung bspw. von FREITAG ET AL. [50] und BESSA ET AL. [14] berücksichtigt. Eine kompakte Übersicht über die Anwendung allgemeiner datenbasierter Methoden im Ingenieurwesen bietet MONTANS ET AL. [132]. Eine Übersicht Daten-basierter Methoden mit dem Fokus auf die kontinuumsmechanische Materialmodellierung geben BOCK ET AL. [19]. Zur allgemeinen Nutzung von KNN im Bauingenieurwesen haben FLOOD und KARTAM [46,47] sowie ADELI [2] frühe Arbeiten veröffentlicht, ohne sich auf Materialmodellierung zu beschränken.

1.2 Stand der Forschung

Die folgende Übersicht über KNN-Materialmodellierung ist umgekehrt ebenfalls nicht auf das Bauingenieurwesen beschränkt und stellt durch die Dynamik und Breite dieses interdisziplinären Forschungsgebiets nur eine persönliche Auswahl von Autoren und Literaturen dar, ohne Anspruch auf Vollständigkeit. Der Fokus liegt dabei auf Veröffentlichungen, die mit den KNN-Algorithmen gearbeitet und die zugrundeliegenden Methoden weiterentwickelt haben.

Die **ersten Schritte der KNN-Materialmodellierung** wurden im Jahr 1991 getan, als GHABOUSSI ET AL. [58] ein KNN zur Modellierung von zyklischem Verhalten von Betonscheiben nutzten. Auf diese Pionierarbeit folgten ab Ende der 1990er Jahre weitere Arbeiten u. a. von GHABOUSSI und SIDARTA [59] über modulare Netzarchitekturen zur Modellierung von drainiertem und undrainiertem Sand, von FURUKAWA und YAGAWA [52] oder JUNG und GHABOUSSI [94] zur Modellierung viskoplastischen Materialverhaltens sowie die Berücksichtigung von Schädigung durch HAJ-ALI ET AL. [70]. Eine Implementierung in die FEM wurde von LEFIK und SCHREFLER [112], HASHASH und GHABOUSSI [72] sowie LIANG und CHANDRASHEKHARA [116] durchgeführt. OESER und FREITAG [138] nutzten RNN zur Modellierung viskoelastischen Materialverhaltens. UNGER und KÖNKE [177,178] beschäftigten sich schon früh mit der Anwendung in der Mehrskalenmodellierung.

Alle diese bis etwa 2010 erschienenen Artikel haben gemeinsam, dass sie sich meist auf monotone Belastung, vorgegebene Lastpfade oder auf ein- bzw. zweiachsiges

Materialverhalten begrenzen. Außerdem sind die Modelle frei von physikalischer Information. Die KNN werden ausschließlich mit den Daten trainiert, ohne zusätzliches physikalisches Wissen zu berücksichtigen. In den auf 2010 folgenden Jahren wurde im Bereich der KNN-Materialmodellierung relativ wenig publiziert. In der Zwischenzeit nahm die Verfügbarkeit von Rechenleistung weiter zu, was vor allem im Bereich der Mehrskalenmodellierung eine Rolle spielt. Außerdem wurden parallel weitere KNN-Architekturen sowie zugehörige Trainingsalgorithmen entwickelt. Möglicherweise hat sich dadurch in den vergangenen fünf Jahren ein neuer Aufschwung im Bereich der KNN-Materialmodellierung entwickelt, der bis heute anhält. Dabei liegt der Fokus u. a. auf der Berücksichtigung physikalischer Eigenschaften sowie auf der Erstellung von robusten und für beliebige Belastungsfälle einsetzbaren Modellen.

Im Bereich **elastischen Materialverhaltens** haben FERNANDEZ ET AL. [42] KNN mit einem analytischen Ansatz zur Modellierung anisotroper hyperelastischer Metamaterialien kombiniert. LINKA ET AL. [118] führen eine neue KNN-Architektur aus mehreren Einzel-KNN ein, mit der sie die Menge der nötigen Daten reduzieren können. Durch nachträgliche Einführung eines Pseudopotentials gewährleisten KALINA ET AL. [96] die thermodynamische Konsistenz des zuvor trainierten KNN a posteriori. WEBER ET AL. [189] sowie AS'AD ET AL. [9] fordern physikalische Nebenbedingungen während des Trainingsprozesses schwach ein und verringern somit die Anzahl nötiger Trainingsdaten erheblich. Durch Nutzung spezieller KNN-Architekturen entwickeln KLEIN ET AL. [103] ein KNN-Materialmodell, welches per Konstruktion die Stabilitätsbedingungen der Polykonvexität erfüllt. Darauf aufbauend diskutieren LINDEN ET AL. [117] die zusätzliche Erfüllung weiterer physikalischer und mathematischer Eigenschaften per Konstruktion. Im Rahmen der Mehrskalenmodellierung nutzen VLASSIS ET AL. [181] topologische Eigenschaften des repräsentativen Volumenelements (RVE) zur Modellierung anisotroper Hyperelastizität, während KALINA ET AL. [95] eine effiziente Methode zur Generierung der zum KNN-Training nötigen RVEs beschreiben.

Im Rahmen **plastischen Materialverhaltens** haben HUANG ET AL. [86] mithilfe der *proper orthogonal decomposition* (POD) die Spannungen als Ausgangsvariablen dekorreliert, um so die Berechnung effizienter zu machen. Durch eine Kombination von inneren Variablen und Potentialen definieren MASI ET AL. [125, 126] sog. *thermodynamics-based artificial neural networks* (TANN) und fordern über eine Nebenbedingung eben jene thermodynamische Konsistenz ein. XU ET AL. [200] definieren die CHOLESKY-Faktoren der Materialtangente als Ausgangsvariablen, um deren positive Definitheit zu gewährleisten. Gerade im Bereich der Mehrskalenmodellierung wurde in den letzten Jahren viel veröffentlicht. So nutzen SETTGAST ET AL. [154] sowie WANG ET AL. [188] ein KNN zur Modellierung poröser Medien auf Basis numerischer Homogenisierung. Sog.

Deep Material Networks wurden von LIU ET AL. [120, 121] als effiziente Ersatzmodelle zur numerischen Homogenisierung entwickelt und bspw. von GAJEK ET AL. [53, 54] mechanisch analysiert und auf thermoplastisch gekoppelte Materialien angewendet. Durch die Pfadabhängigkeit plastischen Materialverhaltens bieten sich RNN als sequenzbasierte neuronale Netze an, was GORJI ET AL. [65] grundlegend anhand VON MISES'scher Plastizität untersucht haben. Aus der gleichen Arbeitsgruppe folgen von BONATTI und MOHR [21] und BONATTI ET AL. [20] Untersuchungen zur Schrittweite und die Anwendung auf die Kristallplastizität. Ein Vergleich mit Feedforward-KNN wurde von LI [115] durchgeführt. MOZAFAR ET AL. [135] approximieren mithilfe von RNN das elastoplastische Verhalten heterogener RVE im ebenen Spannungszustand. HEIDER ET AL. [74] untersuchen verschiedene Invarianten als Eingangsvariablen zur Modellierung von Anisotropie und in Hinblick auf die materielle Objektivität des KNN-Materialmodells. BORKOWSKI ET AL. [22] fordern die Nichtnegativität plastischer Dissipationsleistung während des Trainings eines RNN ein.

1.3 Ziele und Gliederung der Arbeit

Ein Großteil der zitierten Arbeiten ist im Laufe der letzten zwei bis drei Jahre veröffentlicht worden. Die vorliegende Dissertationsschrift steht mit zwei bereits veröffentlichten Artikeln (WEBER ET AL. [189, 190]) im Kontext dieses dynamischen und interdisziplinären Umfelds und der damit verbundenen rasanten Entwicklung neuer Algorithmen, die vor allem den Einbezug physikalischer Informationen in die ursprünglich rein mathematischen, datenbasierten KNN anstrebt. Im Umfeld dieses forschungsgeschichtlichen Kontextes werden für die vorliegende Arbeit die folgenden Ziele definiert:

- Einführung des Trainings künstlicher neuronaler Netze im Kontext restringierter Optimierungsstrategien. Während in der Fachliteratur das Erweitern der Fehlerfunktion zu finden ist, wurde dieses Vorgehen nach besten Wissen des Autors noch nicht im Rahmen mathematischer Optimierungsstrategien grundlegend hergeleitet. Zentraler Aspekte sind dabei die Approximation der zugehörigen Penalty-Terme über eine Monte-Carlo-Integration und die dazu nötigen Strategien zur Generierung der Integrationspunkte.
- Formulierung diverser physikalischer Nebenbedingungen für die elastische sowie die ratenunabhängige plastische KNN-Materialmodellierung. Im Gegensatz zu spezifischen Architekturen, welche physikalische Eigenschaften per Konstruktion erfüllen, werden die Nebenbedingungen nicht exakt erfüllt, sind dafür jedoch flexibel definierbar und gehen über die Gleichungen der klassischen Materialmodellierung hinaus. Sie stabilisieren nicht nur den

KNN-Trainingsprozess, sondern verbessern ebenfalls die physikalischen und numerischen Eigenschaften der KNN in der späteren Anwendung. Für eine ausgewogene Definition des Optimierungsprozesses werden darüber hinaus geeignete Normalisierungen für alle Nebenbedingungen vorgeschlagen.

- Ausführliche Untersuchungen verschiedener KNN-Materialmodelle mit und ohne Anwendung der genannten Nebenbedingungen. Dabei liegt der Fokus auf Eigenschaften der Materialmodelle, die in einer späteren Anwendung relevant sind, wie bspw. eine Schrittweitenabhängigkeit, das Verhalten bei Quasi-Inkompressibilität sowie Energieerhaltung im Rahmen elastischer Materialien. Darüber hinaus wird das Verhalten der KNN-Materialmodelle innerhalb der FEM an ausgewählten Beispielen untersucht und diskutiert.
- Übertragung der gewonnenen Erkenntnisse auf die KNN-Materialmodellierung mit echten Versuchsdaten. Reale Daten bieten weit über die Anwendung auf synthetisch erzeugte Daten hinaus andere Herausforderungen, wie bspw. begrenzte Datenmengen und große Streuungen. Dabei werden Daten von vulkanisiertem Gummi als Beispiel finiter elastischer Deformationen, Daten von biaxial beanspruchtem Fichtenholz als Beispiel eines nichtlinearen und anisotropen Materials sowie Daten einer ferroelektrischen Keramik als Beispiel für eine multiphysikalische Kopplung verwendet. In der Fachliteratur ist die Übertragung der KNN-Materialmodellierung auf Daten realer Experimente selten zu finden. Dies hebt die vorliegende Arbeit noch einmal stärker von den meisten genannten Artikeln ab.

Zur Operationalisierung der gesetzten Fragestellung ist die Arbeit wie nachfolgend beschrieben gegliedert. Die Zusammenhänge und Abhängigkeiten der Kapitel 2 bis 7 sind ebenfalls in Abbildung 1.1 dargestellt. Zu beachten ist dabei die illustrierte wechselseitige Abhängigkeit von Mechanik und KNN, ohne die das Gebäude der KNN-Materialmodellierung nicht mehr standsicher ist.

In **Kapitel 2** werden die materialunabhängigen analytischen und numerischen Grundlagen zur Modellierung von Festkörpern erläutert. Dabei werden wichtige Notationen und Annahmen festgelegt, die im Laufe der Arbeit gültig sind.

Die materialabhängigen Gleichungen sowie die zugehörigen physikalischen Grundlagen folgen in **Kapitel 3**. Auf diese physikalischen Grundlagen wird im Rahmen der KNN-Materialmodellierung wiederholt Bezug genommen. Dabei wird in diesem Kapitel sowohl in die elastischen als auch in die ratenunabhängigen elastoplastischen Materialtheorien eingeführt. Abschließend werden jeweils Beispiele für klassische Materialmodelle gegeben.

Kapitel 4 stellt eine ausführliche Einführung in die Theorie und Numerik von KNN dar. Drei KNN-Architekturen werden beschrieben und die zugehörigen Al-

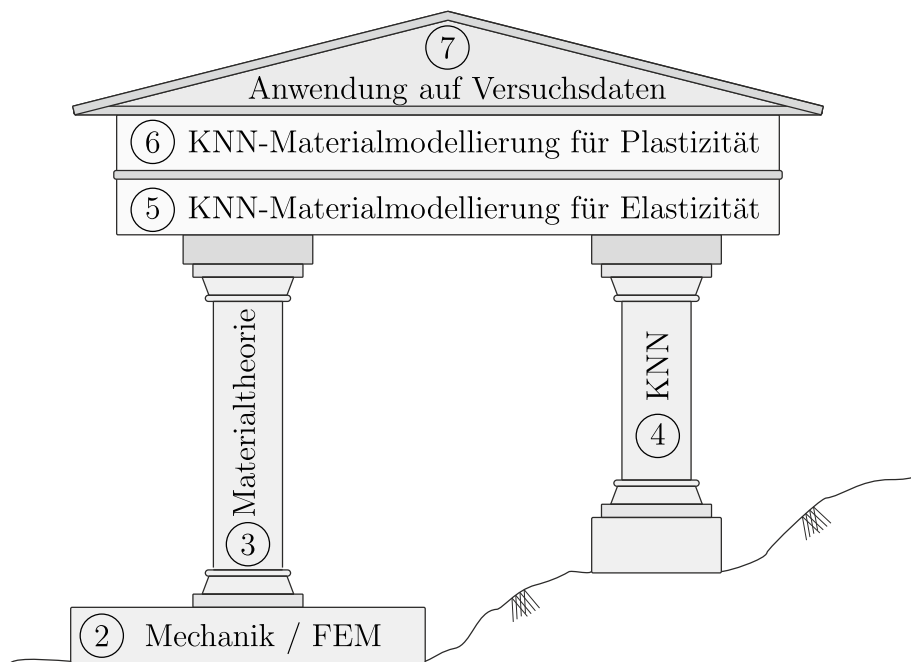


Abbildung 1.1: Visualisierung der Kapitelzusammenhänge

gorithmen für eine effiziente Implementierung zusammengefasst. Darauf aufbauend, führt die zweite Hälfte des Kapitels grundlegend in das KNN-Training unter Nebenbedingungen im Rahmen restringierter Optimierung ein. Dabei werden drei Verfahren zur Einforderung von Nebenbedingungen beschrieben und anhand eines Einführungsbeispiels miteinander verglichen.

In **Kapitel 5** wird die elastische KNN-Materialmodellierung untersucht. Dazu werden zunächst verschiedene elastische KNN-Materialmodelle beschrieben. Nach der Definition diverser physikalischer Nebenbedingungen werden die KNN-Materialmodelle an Materialpunkt-Versuchen sowie bei der Anwendung innerhalb der Methode der finiten Elemente (FEM) miteinander verglichen.

Analog zu Kapitel 5 werden in **Kapitel 6** KNN-Materialmodelle für ratenunabhängiges plastisches Materialverhalten beschrieben. Neben den zugehörigen physikalischen Nebenbedingungen wird außerdem eine effiziente Strategie für die dazu nötige Datengenerierung vorgestellt. Im Anschluss folgen Auswertungen am Materialpunkt und innerhalb der FEM.

Die in den vorherigen Kapiteln und an synthetischen Daten gewonnenen Erkenntnisse werden in **Kapitel 7** in die Anwendung gebracht. Dabei werden Daten zu vulkanisiertem Gummi, Daten zu biaxial beanspruchten Fichtenholzscheiben und Daten zu ferroelektrischen Funktionskeramiken verwendet.

Eine kurze Zusammenfassung, eine kritische Einordnung der Arbeit sowie Ausblicke für weitere mögliche Untersuchungen folgen in **Kapitel 8**.

2 Mechanische und numerische Grundlagen

In diesem Kapitel werden die materialunabhängigen Grundlagen zur analytischen Beschreibung und zur numerischen Berechnung von Festkörpern zusammengefasst. Dabei werden nur rein mechanische und quasi-statische Prozesse betrachtet. Sowohl dynamische als auch thermische, elektrische und chemische Effekte bleiben unberücksichtigt. Darüber hinaus wird Massenerhaltung vorausgesetzt.

2.1 Grundgleichungen der Kontinuumsmechanik

Um das Verhalten von Festkörpern analytisch zu beschreiben, sind drei verschiedene Gruppen von Gleichungen nötig. Die *kinematischen Gleichungen* sind eine rein mathematische Beschreibung der Bewegung des Körpers. Dabei sind die lokalen, relativen Deformationen in der Festkörpermechanik besonders relevant. Diese werden mit Verzerrungsgrößen quantifiziert. *Gleichgewichtsbedingungen* bilanzieren Kräfte innerhalb des Körpers und solche, die auf dessen Rand einwirken. Dazu wird das Konzept der Spannungen eingeführt. Darüber hinaus liefern Materialmodelle die *konstitutiven Gleichungen*, die den Zusammenhang zwischen Deformation und innerem Kraftzustand beschreiben. In diesem Kapitel werden zunächst nur die beiden ersten materialunabhängigen Zusammenhänge erläutert. Für eine ausführlichere Darstellung der Kontinuumsmechanik sei bspw. auf die Literatur von ALTENBACH [4], HOLZAPFEL [81], LIU [119] sowie TRUESDELL und NOLL [175] verwiesen.

2.1.1 Kinematik: Geometrie und Deformation

Im Rahmen der Kontinuumsmechanik wird ein Körper \mathcal{B} als eine zusammenhängende Menge materieller Punkte $\mathbf{X} \in \mathcal{B}$ definiert. Die räumliche Lage aller materiellen Punkte wird als dessen zeitabhängige Konfiguration bezeichnet. Die Zeit t ist dabei oft nur eine Prozesszeitkoordinate. Vom Punkt O aus werden die Konfigurationen mithilfe der Basisvektoren $\{\mathbf{e}_i\}_{i=1}^3$ im dreidimensionalen euklidischen Raum beschrieben. Dies ist in Abbildung 2.1 dargestellt. Per Definition befindet sich der Körper zum Zeitpunkt $t_0 = 0$ in der Referenzkonfiguration und nimmt den Raum V_0 ein. Der zugehörige Rand wird mit ∂V_0 bezeichnet. Jeder materielle Punkt \mathbf{X} befindet sich am zugehörigen initialen räumlichen Punkt $\mathbf{x}(\mathbf{X}, 0)$. In Kapitel 2 wird zunächst die Annahme getroffen, dass die materiellen Punkte \mathbf{X} mit dem gleichen kartesischen Koordinatensystem definiert werden, welches auch

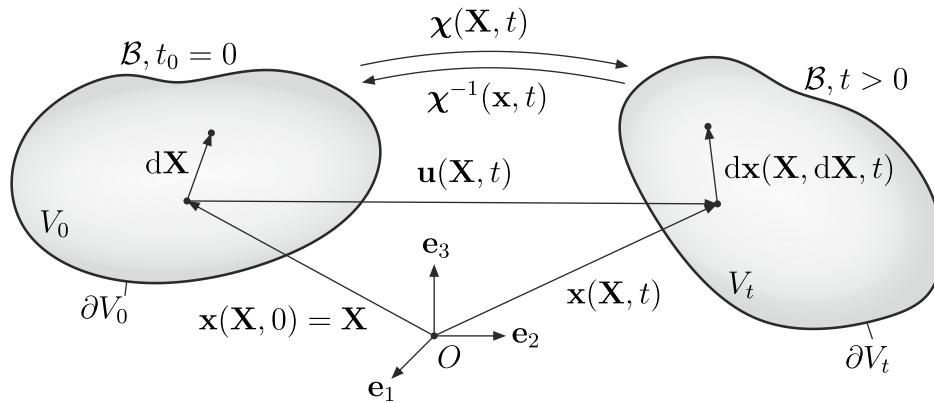


Abbildung 2.1: Bewegung eines Körpers \mathcal{B} aus der Referenzkonfiguration zum Zeitpunkt $t_0 = 0$ in die Momentankonfiguration zum Zeitpunkt t

die räumliche Lage \mathbf{x} beschreibt¹. Damit werden im Folgenden die materiellen Punkte über ihre räumliche Position in der Referenzkonfiguration definiert:

$$\mathbf{X} = \mathbf{x}(\mathbf{X}, 0) . \quad (2.1)$$

Zum Zeitpunkt $t > 0$ befindet sich der Körper in der Momentankonfiguration und nimmt den Raum V_t ein, mit zugehörigem Rand ∂V_t . Die materiellen Punkte \mathbf{X} haben sich jeweils von $\mathbf{x}(\mathbf{X}, 0) = \mathbf{X}$ nach $\mathbf{x}(\mathbf{X}, t)$ bewegt. Die bijektive und stetig differenzierbare Funktion $\chi(\mathbf{X}, t)$ mit

$$\mathbf{x} = \chi(\mathbf{X}, t) \quad \text{und} \quad \mathbf{X} = \chi^{-1}(\mathbf{x}, t) \quad (2.2)$$

bildet die materiellen Punkte \mathbf{X} auf die räumlichen Koordinaten \mathbf{x} ab. Im Folgenden wird \mathbf{X} auch als materielle Koordinate bezeichnet, mit der in dieser Arbeit hauptsächlich die Bewegung des Körpers beschrieben wird. Dies wird auch als LAGRANGE'sche Formulierung bezeichnet. So lässt sich

$$\mathbf{u}(\mathbf{X}, t) = \mathbf{x}(\mathbf{X}, t) - \mathbf{X} \quad (2.3)$$

als das materielle Verschiebungsfeld des Körpers durch die Koordinate \mathbf{X} und der Pseudozeit t definieren.

Die Funktionen χ und \mathbf{u} beschreiben die absolute und vollständige Bewegung des Körpers \mathcal{B} . In der Festkörpermechanik sind jedoch vor allem lokale, relative Verschiebungen und Formänderungen wichtig. Dazu wird zunächst die TAYLOR-Reihe der Lage \mathbf{x} in der Umgebung um \mathbf{X} aufgestellt

$$\mathbf{x}(\mathbf{X} + d\mathbf{X}, t) = \mathbf{x}(\mathbf{X}, t) + \frac{\partial \chi(\mathbf{X}, t)}{\partial \mathbf{X}} d\mathbf{X} + \mathcal{O}(\|d\mathbf{X}\|^2) . \quad (2.4)$$

¹ Diese Annahme erleichtert im Folgenden die Notation, da bei den Gleichungen kein Bezug auf verschiedene Koordinatensysteme genommen werden muss. Im Allgemeinen können die materiellen Punkte mit einem beliebigen Koordinatensystem beschrieben werden. Dies muss bei der Diskussion zur materiellen Objektivität in Abschnitt 3.2.2 ebenfalls berücksichtigt werden.

Der materielle Deformationsgradient² wird definiert als

$$\mathbf{F}(\mathbf{X}, t) := \frac{\partial \mathbf{x}(\mathbf{X}, t)}{\partial \mathbf{X}} = \frac{\partial \mathbf{x}(\mathbf{X}, t)}{\partial \mathbf{X}} . \quad (2.5)$$

Unter Vernachlässigung der Terme höherer Ordnung in Gleichung (2.4) kann \mathbf{F} als Abbildung zwischen den materiellen und räumlichen Linienelementen $d\mathbf{X}$ und $d\mathbf{x} = \mathbf{x}(\mathbf{X}, t) - \mathbf{x}(\mathbf{X} + d\mathbf{X}, t)$ identifiziert werden, mit

$$d\mathbf{x} = \mathbf{F} d\mathbf{X} \quad \text{und} \quad d\mathbf{X} = \mathbf{F}^{-1} d\mathbf{x} . \quad (2.6)$$

Mithilfe des materiellen Verschiebungsgradienten

$$\mathbf{H}(\mathbf{X}, t) := \frac{\partial \mathbf{u}}{\partial \mathbf{X}} \quad (2.7)$$

und dem Einheitstensor \mathbf{I} lässt sich der Deformationsgradient umformen zu

$$\mathbf{F} = \mathbf{I} + \mathbf{H} . \quad (2.8)$$

Der Körper \mathcal{B} darf sich bei Deformation nicht selbst durchdringen. Dadurch folgt die Nichtnegativität der Determinante des Deformationsgradienten

$$J(\mathbf{X}, t) := \det \mathbf{F} > 0 . \quad (2.9)$$

Mit ihr können materielle Volumenelemente dV und materielle Flächenelemente $\mathbf{N} dA$, bestehend aus Flächenelement dA und Normalenvektor \mathbf{N} , in die Momentankonfiguration transformiert werden:

$$dv = J dV , \quad (2.10)$$

$$\mathbf{n} da = J \mathbf{F}^{-T} \mathbf{N} dA . \quad (2.11)$$

Dabei ist $\mathbf{n} da$ das räumliche Flächenelement. Gleichung (2.11) ist auch als NANSON'sche Formel bekannt. Beide Transformationen können mit dem Volumen

$$dv = \left| \det \left([d\mathbf{x}_a, d\mathbf{x}_b, d\mathbf{x}_c] \right) \right| = \left| (d\mathbf{x}_a \times d\mathbf{x}_b) \cdot d\mathbf{x}_c \right| \quad (2.12)$$

eines infinitesimalen räumlichen Spats hergeleitet werden, welcher aus drei Linienelementen $d\mathbf{x}_i = \mathbf{F} d\mathbf{X}_i$ konstruiert wird.

Die Tensoren \mathbf{F} und \mathbf{H} beschreiben lokale und relative Verschiebungen. Sie enthalten allerdings auch Rotationen und Translationen, die nur eine Lageänderung

² An dieser Stelle sei angemerkt, dass es sich bei \mathbf{F} um einen Zwei-Feld-Tensor handelt, da \mathbf{x} und \mathbf{X} grundsätzlich mit verschiedenen Koordinatensystemen definiert werden können. Darüber hinaus sind alle eingeführten Tensoren eigentlich Tensorfelder, definiert auf den zugehörigen Konfigurationen. Aus Gründen der Lesbarkeit wird die Abhängigkeit von \mathbf{X} und t nicht durchgehend gekennzeichnet.

des Körpers im Raum beschreiben. Um nur die gestaltändernden Deformationen zu berücksichtigen, werden Verzerrungstensoren definiert. Durch die rechte polare Zerlegung des Deformationsgradienten

$$\mathbf{F} = \mathbf{R}\mathbf{U} \ , \quad \text{mit} \quad \mathbf{R}^T\mathbf{R} = \mathbf{I}, \det \mathbf{R} = 1 \quad \text{und} \quad \mathbf{U}^T = \mathbf{U} \text{ pos. def.} \ , \quad (2.13)$$

resultiert der Rotationstensor \mathbf{R} und der Rechtsstrecktensor \mathbf{U} . Dabei wurden mit der Bedingung $J > 0$ Spiegelungen mit $\det \mathbf{R} \neq -1$ ausgeschlossen. Weiter wird der Rechts-CAUCHY-GREEN-Tensor

$$\mathbf{C}(\mathbf{X}, t) := \mathbf{F}^T\mathbf{F} = \mathbf{U}^T\mathbf{U} = \mathbf{U}^2 \quad (2.14)$$

definiert. Mit ihm können über

$$d\mathbf{x}_a \cdot d\mathbf{x}_b = d\mathbf{X}_a \cdot (\mathbf{F}^T\mathbf{F}) d\mathbf{X}_b = d\mathbf{X}_a \cdot \mathbf{C} d\mathbf{X}_b \quad (2.15)$$

Streckungen von materiellen Linienelementen ($a = b$) und Winkeländerungen zwischen materiellen Linienelementen ($a \neq b$) am festen Punkt \mathbf{X} berechnet werden. Der Deformationstensor \mathbf{C} entspricht bei reinen Translationen und Rotationen genau dem Einheitstensor \mathbf{I} . Dieser wird daher noch subtrahiert, um als brauchbares Verzerrungsmaß den GREEN-LAGRANGE'schen Verzerrungstensor

$$\mathbf{E}(\mathbf{X}, t) := \frac{1}{2}(\mathbf{C} - \mathbf{I}) \quad (2.16)$$

zu definieren. Er lässt sich mit Gleichung (2.8) ebenfalls über

$$\mathbf{E} = \frac{1}{2}(\mathbf{H} + \mathbf{H}^T + \mathbf{H}^T\mathbf{H}) \quad (2.17)$$

in Abhängigkeit des Verschiebungsgradienten \mathbf{H} ausdrücken. Bezüglich der kartesischen Basen \mathbf{e}_i lassen sich die Komponenten von \mathbf{E} in Abhängigkeit der Verschiebungsfunktion \mathbf{u} wie folgt berechnen.

$$E_{ii} = u_{i,i} + \frac{1}{2}(u_{1,i}^2 + u_{2,i}^2 + u_{3,i}^2) \ , \quad i = 1, 2, 3 \quad (2.18)$$

$$2E_{ij} = u_{i,j} + u_{j,i} + u_{1,i}u_{1,j} + u_{2,i}u_{2,j} + u_{3,i}u_{3,j} \ , \quad i, j = 1, 2, 3, \ i \neq j \quad (2.19)$$

Neben \mathbf{E} können weitere materielle und auch räumliche Verzerrungstensoren definiert werden. Allen ist gemeinsam, dass sie für infinitesimale Deformationen zum linearen Verzerrungstensor $\boldsymbol{\varepsilon}$ konvergieren müssen. Wird in Gleichung (2.17) der nichtlineare Term $\mathbf{H}^T\mathbf{H}$ vernachlässigt, folgt

$$\boldsymbol{\varepsilon} := \frac{1}{2}(\mathbf{H} + \mathbf{H}^T) \ . \quad (2.20)$$

Mit diesem Grenzwert ist auch der Faktor (1/2) in der Definition von \mathbf{E} zu begründen. Im Rahmen dieser Arbeit finden ausschließlich der Deformationstensor \mathbf{C} und die Verzerrungstensoren \mathbf{E} und $\boldsymbol{\varepsilon}$ Verwendung.

2.1.2 Das Konzept der Spannungen

Der Körper \mathcal{B} wird durch äußere Lasten, die entweder innerhalb des Körpers oder auf dessen Rand wirken, belastet. Durch gedankliches Frei- oder Durchschneiden des Körpers entstehen Teilkörper und Schnittflächen, auf denen innere Kraftgrößen definiert werden. Dies ist in Abbildung 2.2 am Beispiel eines ebenen Schnitts dargestellt. Per Annahme stehen dabei die Kraftgrößen auf gegenüberliegenden Schnittufern im Kräftegleichgewicht und jeder resultierende Teilkörper steht für sich im Kräfte- und Momentengleichgewicht. Dieses Gedankenexperiment ist auf LEONHARD EULER zurückzuführen, was bspw. in SZABO [167] zu lesen ist.

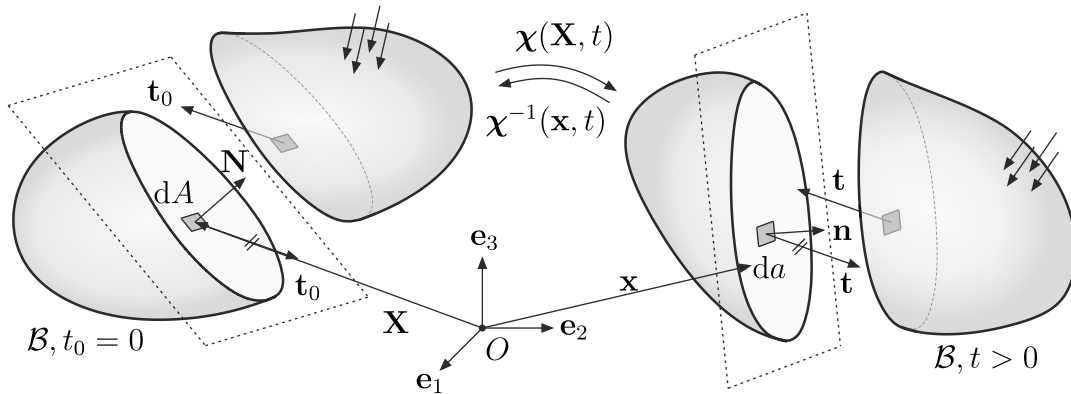


Abbildung 2.2: Gedanklicher Schnitt durch den Körper \mathcal{B}

In der Momentanconfiguration wirkt auf einem finiten Flächenelement Δa ein resultierender Kraftvektor $\Delta \mathbf{f}$. Durch Bildung des Grenzwertes

$$\mathbf{t}(\mathbf{x}, \mathbf{n}, t) := \lim_{\Delta a \rightarrow 0} \frac{\Delta \mathbf{f}}{\Delta a} = \frac{d\mathbf{f}}{da} \quad (2.21)$$

resultiert per Definition der CAUCHY-Spannungsvektor \mathbf{t} . Er hängt neben der räumlichen Koordinate \mathbf{x} auch von \mathbf{n} , der Orientierung des Schnitts, ab. Das Fundamentaltheorem nach CAUCHY

$$\mathbf{t}(\mathbf{x}, \mathbf{n}, t) = \boldsymbol{\sigma}(\mathbf{x}, t)^T \mathbf{n} \quad (2.22)$$

besagt, dass der Spannungsvektor im Punkt \mathbf{x} linear abhängig von \mathbf{n} ist. Dabei ist $\boldsymbol{\sigma}$ der symmetrische³ CAUCHY'sche Spannungstensor, der den vollständigen, von der Schnitorientierung unabhängigen Spannungszustand im räumlichen Punkt \mathbf{x} charakterisiert. Er wird in der Gleichung transponiert, sodass die Komponenten

³Die Symmetrie kann z. B. über das Momentengleichgewicht an einem herausgeschnittenen infinitesimalen Quader gezeigt werden.

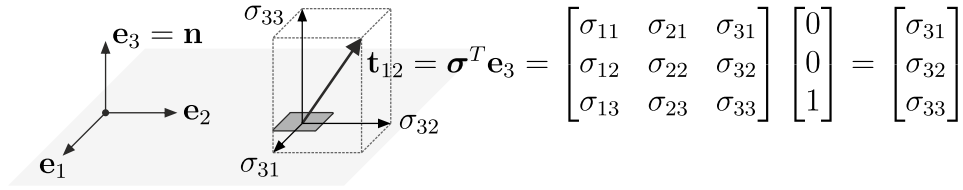


Abbildung 2.3: Komponenten des Spannungsvektors am Beispiel eines Schnitts entlang der $\mathbf{e}_1 - \mathbf{e}_2$ -Ebene

der Spannungsvektoren \mathbf{t} bei zu den Koordinatenachsen orthogonalen Schnittflächen mit $\mathbf{n} = \mathbf{e}_i$ der folgenden Index-Definition folgen: Der erste Index definiert die Schnittfläche, der zweite Index die Richtung der Spannungskomponente. Dies ist in Abbildung 2.3 dargestellt. Die CAUCHY-Spannungen werden auch als *wahre Spannungen* bezeichnet, da sie den aktuellen Spannungszustand in der Momentankonfiguration charakterisieren.

Wird der gleiche finite Kraftvektor $\Delta \mathbf{f}$ auf eine Fläche ΔA der Referenzkonfiguration bezogen und anschließend der Grenzwert $\Delta A \rightarrow 0$ analog zu Gleichung (2.21) gebildet, kann der 1. PIOLA-KIRCHHOFF'sche Spannungsvektor \mathbf{t}_0 definiert werden. Für den differentiellen Kraftvektor $d\mathbf{f}$ gilt somit die Gleichung

$$d\mathbf{f} = \mathbf{t} da = \mathbf{t}_0 dA . \quad (2.23)$$

Analog zu Gleichung (2.22) wird mit

$$\mathbf{t}_0(\mathbf{X}, \mathbf{N}, t) = \mathbf{P}(\mathbf{X}, t)^T \mathbf{N} \quad (2.24)$$

der unsymmetrische 1. PIOLA-KIRCHHOFF'sche Spannungstensor \mathbf{P} definiert. Mit Gleichung (2.11) folgen die Transformationsbeziehungen

$$\mathbf{P} = J \mathbf{F}^{-1} \boldsymbol{\sigma} \neq \mathbf{P}^T \quad \text{und} \quad \boldsymbol{\sigma} = \frac{1}{J} \mathbf{F} \mathbf{P} . \quad (2.25)$$

Die 1. PIOLA-KIRCHHOFF'schen Spannungen werden als *Nennspannungen* oder *technische Spannungen* bezeichnet. Sie beziehen die tatsächlich wirkenden Kräfte auf die Geometrie der Referenzkonfiguration.

Wird der infinitesimale Spannungsvektor $d\mathbf{f}$ aus Gleichung (2.23) mit Gleichung (2.6) auf die Referenzkonfiguration transformiert, kann der symmetrische 2. PIOLA-KIRCHHOFF'sche Spannungstensor \mathbf{S} mit

$$\mathbf{F}^{-1} d\mathbf{f} = \underbrace{\mathbf{F}^{-1} \mathbf{P}^T}_{=: \mathbf{S}^T} \mathbf{N} dA \quad (2.26)$$

definiert werden. Die Transformationsbeziehungen lassen sich angeben als

$$\mathbf{S} = J \mathbf{F}^{-1} \boldsymbol{\sigma} \mathbf{F}^{-T} = \mathbf{S}^T \quad \text{und} \quad \boldsymbol{\sigma} = \frac{1}{J} \mathbf{F} \mathbf{S} \mathbf{F}^T . \quad (2.27)$$

Die 2. PIOLA-KIRCHHOFF'schen Spannungen werden als *Pseudo-Spannungen* bezeichnet. Sie sind nicht physikalisch und dienen nur als Rechengröße.

Neben den hier eingeführten Größen existieren noch andere Spannungstensoren. Im Rahmen dieser Arbeit werden ausschließlich $\boldsymbol{\sigma}$, \mathbf{P} und \mathbf{S} verwendet. Im Falle infinitesimaler Verzerrungen sind alle Spannungsmaße äquivalent und das Symbol $\boldsymbol{\sigma}$ wird verwendet.

2.1.3 Die starke Form des Gleichgewichts

Mit den Verzerrungs- und Spannungstensoren aus den Abschnitten 2.1.1 und 2.1.2 lassen sich die Deformation und der Kraftzustand eines Körpers beschreiben. Um sie zu berechnen, müssen Gleichungen eingefordert und Randbedingungen definiert werden. Die Zielgröße ist dabei im Rahmen dieser Arbeit immer das materielle Verschiebungsfeld $\mathbf{u}(\mathbf{X}, t)$. Das zugehörige Randwertproblem des Körpers \mathcal{B} ist in Abbildung 2.4 dargestellt.

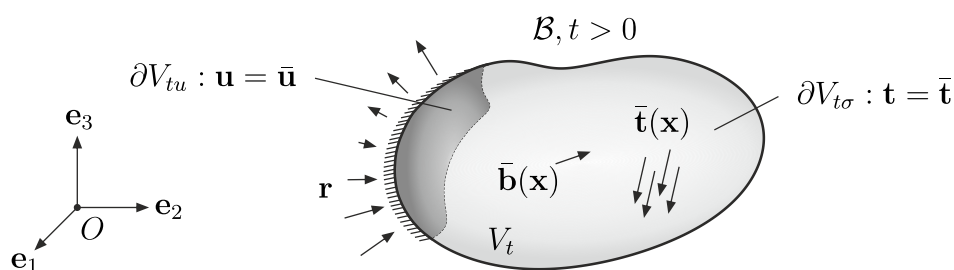


Abbildung 2.4: Randwertproblem des Körpers \mathcal{B}

Der Rand ∂V_t wird in einen Verschiebungsrand ∂V_{tu} und einen Krastrand $\partial V_{t\sigma}$ aufgeteilt, auf denen jeweils Verschiebungen $\bar{\mathbf{u}}$ und Spannungsvektoren $\bar{\mathbf{f}}$ vorgegeben sind. Es gilt $\partial V_{tu} \cup \partial V_{t\sigma} = \partial V_t$ und $\partial V_{tu} \cap \partial V_{t\sigma} = \emptyset$. Im Inneren des Körpers wirkt die Massenkraftdichte $\bar{\mathbf{b}}$, die mit der Dichte ρ_t in eine Volumenkraftdichte umgerechnet werden kann. Am Verschiebungsrand entstehen durch die aufgezungenen Verschiebungen Spannungsvektoren \mathbf{r} als Reaktionskraftgrößen.

Zu jedem Zeitpunkt t muss der Körper im globalen Kräftegleichgewicht

$$\int_{V_t} \rho_t \bar{\mathbf{b}} \, dv + \int_{\partial V_{tu}} \mathbf{r} \, da + \int_{\partial V_{t\sigma}} \bar{\mathbf{t}} \, da = \mathbf{0} \quad (2.28)$$

stehen. Unter Berücksichtigung des lokalen Gleichgewichts an den Randflächen und des CAUCHY-Theorems (2.22) lassen sich die Randterme mit

$$\int_{\partial V_{tu}} \mathbf{r} \, da + \int_{\partial V_{t\sigma}} \bar{\mathbf{t}} \, da = \int_{\partial V_{tu} \cup \partial V_{t\sigma}} \mathbf{t} \, da = \int_{\partial V_t} \boldsymbol{\sigma}^T \mathbf{n} \, da \quad (2.29)$$

umformen. Der resultierende Term wird mit dem GAUSS'schen Integralsatz zu

$$\int_{\partial V_t} \boldsymbol{\sigma}^T \mathbf{n} \, da = \int_{V_t} \operatorname{div}(\boldsymbol{\sigma}^T) \, dv \quad (2.30)$$

transformiert, wodurch Gleichung (2.28) die Form

$$\int_{V_t} (\operatorname{div}(\boldsymbol{\sigma}^T) + \rho_t \bar{\mathbf{b}}) \, dv = \mathbf{0} \quad (2.31)$$

annimmt. Der Integrand muss dabei für jeden Punkt \mathbf{x} gleich null sein. Er liefert somit die lokale Form des Kräftegleichgewichts in der Momentankonfiguration

$$\operatorname{div}(\boldsymbol{\sigma}^T) + \rho_t \bar{\mathbf{b}} = \mathbf{0} . \quad (2.32)$$

Darüber hinaus liefert das globale Momentengleichgewicht

$$\int_{V_t} \mathbf{x} \times \rho_t \bar{\mathbf{b}} \, dv + \int_{\partial V_{tu}} \mathbf{x} \times \mathbf{r} \, da + \int_{\partial V_{t\sigma}} \mathbf{x} \times \bar{\mathbf{t}} \, da = \mathbf{0} \quad (2.33)$$

nach einigen Umformungen die lokale Drehimpulsbilanz, vgl. HOLZAPFEL [81]. Daraus folgt die Symmetriebedingung des CAUCHY-Spannungstensors

$$\boldsymbol{\sigma}^T = \boldsymbol{\sigma} . \quad (2.34)$$

Werden die Spannungsvektoren auf die Referenzkonfiguration bezogen, nimmt das globale Kräftegleichgewicht nach Zusammenfassung der Last- und Reaktionsvektoren die Form

$$\int_{V_0} \rho_0 \bar{\mathbf{b}} \, dV + \int_{\partial V_0} \mathbf{t}_0 \, dA = \mathbf{0} \quad (2.35)$$

an. Dabei ist $\rho_0 = J\rho_t$ die Dichte des Körpers in der Referenzkonfiguration. Diese Umformung gilt nur, da Massenerhaltung mit $\dot{m} = 0$ angenommen wurde. Durch äquivalente Umformungen wie die bzgl. der Momentankonfiguration resultiert schließlich die lokale Gleichgewichtsbedingung in der Referenzkonfiguration

$$\operatorname{div}(\mathbf{P}^T) + \rho_0 \bar{\mathbf{b}} = \mathbf{0} . \quad (2.36)$$

Diese partielle Differentialgleichung in den Spannungen \mathbf{P} muss für jeden materiellen Punkt \mathbf{X} des Kontinuums erfüllt sein. Sie wird daher auch *starke Form* des Gleichgewichts genannt. Würden an dieser Stelle ein Stoffgesetz für $\mathbf{P}(\mathbf{F})$ und die kinematischen Beziehungen $\mathbf{F}(\mathbf{u})$ eingesetzt werden, könnte eine Differentialgleichung (DGL) für die unbekannte Verschiebungsfunktion \mathbf{u} aufgestellt werden. Diese DGL ist allerdings nur in Sonderfällen analytisch lösbar. Im nächsten Abschnitt wird daher zunächst die zugehörige *schwache Form* hergeleitet, die als Basis für numerische Lösungsverfahren dient. Ab hier erfolgen die Herleitungen ausschließlich in der Referenzkonfiguration.

2.1.4 Die schwache Form des Gleichgewichts und das PVV

Zur Umformung der starken Form in die schwache Form wird Gleichung (2.36) mit einer Testfunktion $\delta \mathbf{u}$ skalar multipliziert und über das Volumen der Referenzkonfiguration V_0 integriert. Das resultierende Integral muss null sein, es gilt

$$\int_{V_0} \left(\operatorname{div}(\mathbf{P}^T) + \rho_o \bar{\mathbf{b}} \right) \cdot \delta \mathbf{u} \, dV = 0 . \quad (2.37)$$

Die Testfunktionen $\delta \mathbf{u}$ sind dabei beliebige, aber feste Funktionen, die auf dem Verschiebungsrand ∂V_{0u} verschwinden:

$$\delta \mathbf{u}(\mathbf{X}, t) = \mathbf{0} , \quad \forall \mathbf{X} \in \partial V_{0u} , \quad \forall t \geq 0 . \quad (2.38)$$

Gleichung (2.37) lässt sich folgendermaßen interpretieren: Durch die Beliebigkeit der Testfunktionen $\delta \mathbf{u}$ kann das Integral nur dann null sein, wenn die starke Form in der Klammer in allen Punkten \mathbf{X} null und damit erfüllt ist. Beide Gleichungen sind somit äquivalent und besitzen die gleiche Lösung \mathbf{u} . Sie lassen sich stets ineinander überführen. Mathematisch ist die schwache Form Gegenstand der Variationsrechnung. Die Größe

$$\delta \mathbf{a} = \frac{\partial \mathbf{a}}{\partial \mathbf{u}} \delta \mathbf{u} = \lim_{h \rightarrow 0} \frac{\mathbf{a}(\mathbf{u} + h \delta \mathbf{u}) - \mathbf{a}(\mathbf{u})}{h} \quad (2.39)$$

wird dabei auch als *erste Variation* von \mathbf{a} bezüglich \mathbf{u} bezeichnet und gleicht einer Richtungsableitung der Größe \mathbf{a} in Richtung der Testfunktion $\delta \mathbf{u}$.

Zunächst wird Gleichung (2.37) weiter umgeformt. Der erste Summand kann zu

$$\int_{V_0} \operatorname{div}(\mathbf{P}^T) \cdot \delta \mathbf{u} \, dV = \int_{V_0} \left(\operatorname{div}(\mathbf{P} \delta \mathbf{u}) - \mathbf{P}^T : \delta \mathbf{H} \right) \, dV \quad (2.40)$$

umgeschrieben werden. Dabei wurden das Skalarprodukt für Tensoren 2. Stufe $\mathbf{A} : \mathbf{B}$ und die 1. Variation des materiellen Verschiebungsgradienten

$$\delta \mathbf{H} = \frac{\partial}{\partial \mathbf{u}} \left(\frac{\partial \mathbf{u}}{\partial \mathbf{X}} \right) \delta \mathbf{u} = \frac{\partial(\delta \mathbf{u})}{\partial \mathbf{X}} \quad (2.41)$$

verwendet. Mit dem GAUSS'schen Integralsatz lässt sich der linke Summand der rechten Seite aus Gleichung (2.40) weiter umformen zu

$$\int_{V_0} \operatorname{div}(\mathbf{P} \delta \mathbf{u}) \, dV = \int_{\partial V_0} (\mathbf{P} \delta \mathbf{u}) \cdot \mathbf{N} \, dA = \int_{\partial V_0} (\mathbf{P}^T \mathbf{N}) \cdot \delta \mathbf{u} \, dA , \quad (2.42)$$

wobei im zweiten Schritt die Identität $(\mathbf{A} \mathbf{b}) \cdot \mathbf{c} = (\mathbf{A}^T \mathbf{c}) \cdot \mathbf{b}$ verwendet wurde, mit den Tensoren 1. Stufe \mathbf{b} und \mathbf{c} sowie dem Tensor 2. Stufe \mathbf{A} . Unter Berücksichtigung des Verschwindens der Testfunktionen $\delta \mathbf{u}$ auf dem Verschiebungsrand ∂V_{0u}

sowie des CAUCHY-Theorems aus Gleichung (2.24) auf dem Spannungsrand $\partial V_{0\sigma}$ kann der Randterm wie folgt umgeformt werden.

$$\int_{\partial V_0} (\mathbf{P}^T \mathbf{N}) \cdot \delta \mathbf{u} \, dA = \int_{\partial V_{0\sigma}} \mathbf{t}_0 \cdot \delta \mathbf{u} \, dA = \int_{\partial V_{0\sigma}} \bar{\mathbf{t}}_0 \cdot \delta \mathbf{u} \, dA \quad (2.43)$$

Dabei ist $\bar{\mathbf{t}}_0 = d\bar{\mathbf{f}}/dA = \bar{\mathbf{t}} da/dA$ der Spannungsvektor, der auf die Oberflächen der Referenzkonfiguration bezogen ist⁴. Schließlich resultiert die schwache Form in den Größen \mathbf{P} und $\delta \mathbf{H}$ zu

$$\int_{V_0} \mathbf{P}^T : \delta \mathbf{H} \, dV - \int_{V_0} \rho_o \bar{\mathbf{b}} \cdot \delta \mathbf{u} \, dV - \int_{\partial V_{0\sigma}} \bar{\mathbf{t}}_0 \cdot \delta \mathbf{u} \, dA = 0 . \quad (2.44)$$

Der Term $\mathbf{P}^T : \delta \mathbf{H}$ lässt sich mit Definition (2.26) und der Identität $(\mathbf{A}\mathbf{B}) : \mathbf{C} = \mathbf{B} : (\mathbf{A}^T \mathbf{C})$ für die Tensoren 2. Stufe \mathbf{A} , \mathbf{B} und \mathbf{C} umschreiben zu

$$\mathbf{P}^T : \delta \mathbf{H} = \mathbf{F} \mathbf{S}^T : \delta \mathbf{H} = \mathbf{S}^T : \mathbf{F}^T \delta \mathbf{H} . \quad (2.45)$$

Da das Skalarprodukt des symmetrischen Tensors $\mathbf{S} = \mathbf{S}^T$ mit einem schiefsymmetrischen Tensor verschwindet, kann Gleichung (2.45) nur mit dem symmetrischen Anteil von $\mathbf{F}^T \delta \mathbf{H}$ geschrieben werden als

$$\mathbf{S}^T : \mathbf{F}^T \delta \mathbf{H} = \mathbf{S}^T : \frac{1}{2} (\mathbf{F}^T \delta \mathbf{H} + \delta \mathbf{H}^T \mathbf{F}) . \quad (2.46)$$

Dieser symmetrische Anteil wird mit

$$\delta \mathbf{E} = \frac{1}{2} (\delta \mathbf{H} + \delta \mathbf{H}^T + \delta \mathbf{H}^T \mathbf{H} + \mathbf{H}^T \delta \mathbf{H}) = \frac{1}{2} (\delta \mathbf{H}^T \mathbf{F} + \mathbf{F}^T \delta \mathbf{H}) \quad (2.47)$$

als 1. Variation des GREEN-LAGRANGE'schen Verzerrungstensors identifiziert. Dadurch resultiert die schwache Form in den Größen $\mathbf{S} = \mathbf{S}^T$ und $\delta \mathbf{E}$ mit

$$\int_{V_0} \mathbf{S} : \delta \mathbf{E} \, dV - \int_{V_0} \rho_o \bar{\mathbf{b}} \cdot \delta \mathbf{u} \, dV - \int_{\partial V_{0\sigma}} \bar{\mathbf{t}}_0 \cdot \delta \mathbf{u} \, dA = 0 . \quad (2.48)$$

Sie ist im Folgenden die Basis für weitere Herleitungen.

Die schwache Form des Gleichgewichts ist äquivalent zum *Prinzip der virtuellen Verschiebungen* (PVV). Dieses physikalische Prinzip besagt sinngemäß das Folgende: Steht der Körper \mathcal{B} im Gleichgewicht, dann ist bei einer beliebigen, aber festen virtuellen Deformation $\delta \mathbf{u}$ die Bilanz der virtuellen Arbeiten $\delta \pi$ null. In

⁴ In vielen Fällen wird die Änderung der Bezugsfläche während der Deformation vernachlässigt, wodurch $\bar{\mathbf{t}}_0 \approx \bar{\mathbf{t}}$ gilt. Ein Gegenbeispiel dazu wäre das Aufblasen eines Luftballons. Dort muss die Änderung der Oberfläche zur korrekten Erfassung des Innendrucks berücksichtigt werden.

Gleichung (2.48) lassen sich die Terme der virtuellen inneren Arbeit $\delta\pi_i$ und die der virtuellen äußeren Arbeit $\delta\pi_a$ identifizieren als

$$\delta\pi_i = \int_{V_0} \mathbf{S} : \delta\mathbf{E} \, dV \quad \text{und} \quad (2.49)$$

$$\delta\pi_a = \int_{V_0} \rho_o \bar{\mathbf{b}} \cdot \delta\mathbf{u} \, dV + \int_{\partial V_{0\sigma}} \bar{\mathbf{t}}_0 \cdot \delta\mathbf{u} \, dA . \quad (2.50)$$

Somit lässt sich die Aussage des PVV kompakt schreiben als

$$\delta\pi = \delta\pi_i - \delta\pi_a = 0 . \quad (2.51)$$

Mathematisch führen die Aussagen der schwachen Form und die des PVV zur gleichen zu lösenden Gleichung. Jedoch ist nur das PVV physikalisch begründet. Darüber hinaus sei an dieser Stelle angemerkt, dass diese Gleichung zunächst unabhängig vom betrachteten Materialverhalten ist.

2.1.5 Linearisierung des PVV

Ist das PVV (2.51) eine nichtlineare Gleichung, muss diese im Allgemeinen mithilfe eines Iterationsverfahrens numerisch gelöst werden. Im Rahmen dieser Arbeit wird dazu das NEWTON-Verfahren verwendet. Nicht die Nullstelle der Funktion $\delta\pi$, sondern die ihrer Tangente wird dabei in jedem Iterationsschritt gesucht. Dazu wird die Linearisierung

$$L[\delta\pi(\mathbf{u})] = \delta\pi(\mathbf{u}) + \Delta\delta\pi(\mathbf{u}, \Delta\mathbf{u}) = \delta\pi(\mathbf{u}) + \frac{\partial\delta\pi(\mathbf{u})}{\partial\mathbf{u}} \cdot \Delta\mathbf{u} \quad (2.52)$$

der Funktion $\delta\pi$ an der festen Entwicklungsstelle \mathbf{u} benötigt. Das Inkrement $\Delta(\bullet)$ lässt sich dabei, äquivalent zu Gleichung (2.39), als Richtungsableitung in Richtung des Verschiebungsincrements $\Delta\mathbf{u}$ berechnen. Im Folgenden wird die Annahme getroffen, dass die äußeren Lasten $\bar{\mathbf{b}}$ und $\bar{\mathbf{t}}_0$ unabhängig von den Verschiebungen \mathbf{u} sind. Darüber hinaus gilt $\Delta\delta\mathbf{u} = \mathbf{0}$, da die virtuellen Verschiebungen fest sind. Beides führt zu $\Delta\delta\pi_a = 0$. Das Inkrement

$$\Delta\delta\pi = \Delta\delta\pi_i = \int_{V_0} (\Delta\mathbf{S} : \delta\mathbf{E} + \mathbf{S} : \Delta\delta\mathbf{E}) \, dV \quad (2.53)$$

resultiert aus der Produktregel $\Delta(\mathbf{A} : \mathbf{B}) = \Delta\mathbf{A} : \mathbf{B} + \mathbf{A} : \Delta\mathbf{B}$ und der virtuellen inneren Arbeit aus Gleichung (2.49). Das Spannungsincrement $\Delta\mathbf{S}$ hängt dabei vom verwendeten Materialmodell ab, welches in diesem Kapitel noch nicht näher spezifiziert wird. Ganz allgemein schreibt sich das Spannungsincrement $\Delta\mathbf{S}$ als

$$\Delta\mathbf{S} = \frac{\partial\mathbf{S}}{\partial\mathbf{E}} \Delta\mathbf{E} = \mathbb{C}_T \Delta\mathbf{E} , \quad \text{bzw.} \quad \Delta S_{ij} = C_{ijkl} \Delta E_{kl} . \quad (2.54)$$

Dabei ist $\mathbb{C}_T(\mathbf{E})$ die nichtlineare Materialtangente und ein Tensor 4. Stufe.

2.1.6 Reduzierte Notation des Materialgesetzes

Im Rahmen der numerischen Umsetzung ist es effizient, die Komponenten der Tensoren 2. und 4. Stufe in eine reduzierte Vektor- bzw. Matrix-Notation umzuschreiben, welche auch deren Symmetrien berücksichtigt. Die Komponenten des 2. PIOLA-KIRCHHOFF'schen Spannungstensors⁵ $\mathbf{S} = S_{ij} \mathbf{e}_i \otimes \mathbf{e}_j$ und des GREEN-LAGRANGE'schen Verzerrungstensors $\mathbf{E} = E_{ij} \mathbf{e}_i \otimes \mathbf{e}_j$ werden in die Vektoren

$$\mathbf{S} = [S_{11}, S_{22}, S_{33}, S_{12}, S_{13}, S_{23}]^T \quad \text{und} \quad (2.55)$$

$$\mathbf{E} = [E_{11}, E_{22}, E_{33}, 2E_{12}, 2E_{13}, 2E_{23}]^T \quad (2.56)$$

einsortiert. Auf eine symbolische Unterscheidung zwischen den beiden Notationen wird verzichtet, da der Fall immer aus dem Kontext ableitbar oder explizit erwähnt sein wird. Die Komponenten der Materialtangente $\mathbb{C}_T = C_{ijkl} \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k \otimes \mathbf{e}_l$ werden in die Matrix

$$\mathbf{C}_T := \begin{bmatrix} C_{1111} & C_{1122} & C_{1133} & C_{1112} & C_{1113} & C_{1123} \\ C_{2211} & C_{2222} & C_{2233} & C_{2212} & C_{2213} & C_{2223} \\ C_{3311} & C_{3322} & C_{3333} & C_{3312} & C_{3313} & C_{3323} \\ C_{1211} & C_{1222} & C_{1233} & C_{1212} & C_{1213} & C_{1223} \\ C_{1311} & C_{1322} & C_{1333} & C_{1312} & C_{1313} & C_{1323} \\ C_{2311} & C_{2322} & C_{2333} & C_{2312} & C_{2313} & C_{2323} \end{bmatrix} \quad (2.57)$$

passend einsortiert⁶. Alle anderen Komponenten C_{ijkl} sind null⁷. Die Linearisierung des Stoffgesetzes (2.54) ist damit $\Delta \mathbf{S} = \mathbf{C}_T \Delta \mathbf{E}$.

In dieser Notation sind wichtige Tensor-Eigenschaften nicht mehr gültig, wie bspw. das Transformationsverhalten bei Wechsel des Koordinatensystems. Dadurch ist immer eine Rücktransformation in die Tensornotation notwendig. An dieser Stelle sei angemerkt, dass es mehrere Möglichkeiten gibt, eine reduzierte Notation zu definieren. Sie unterscheiden sich hauptsächlich durch die Reihenfolge und Definition der Schubterme. Bspw. sind bei der nach VOIGT [183] benannten Notation die 12- und 23-Komponenten in den Vektoren vertauscht. Dementsprechend ändert sich auch die \mathbf{C}_T -Matrix.

⁵ Dabei wird für die Tensornotation die EINSTEIN'sche Summenkonvention verwendet, bei der über doppelt auftretende Indizes von 1 bis 3 aufsummiert wird.

⁶ Diese Darstellung gilt nur durch die Symmetrieeigenschaft $C_{ijkl} = C_{ijlk}$

⁷ Durch die Symmetrien $E_{ij} = E_{ji}$ und $S_{ij} = S_{ji}$ reduzieren sich die 81 Komponenten von \mathbb{C} auf 36. Weitere Reduzierungen werden durch zusätzliche Anforderungen an das Materialgesetz erzielt, vgl. Kapitel 3.

2.2 Die Methode der finiten Elemente

Die Methode der finiten Elemente (FEM) ist ein numerisches Verfahren zur Lösung der schwachen Form und damit der zugrunde liegenden Differentialgleichungen. Die Anwendungsgebiete reichen vom Bauingenieurwesen über die Luft- und Raumfahrttechnik bis hin zur Medizintechnik. Sie umfassen u. a. die Simulation von Festkörpern, Wärmeströmen und gekoppelten physikalischen Problemen. Ihre vollständige Entwicklung muss über viele Jahrzehnte betrachtet werden und ist bis heute ein Zusammenspiel aus Mathematik und Ingenieurwissenschaften. Die folgenden Punkte stellen eine Auswahl zentraler historischer Meilensteine dar.

18. Jh.	Die Entwicklung der Variationsrechnung insbesondere von EULER und LAGRANGE, siehe z. B. GRAY [66]
1915	Das Verfahren der gewichteten Residuen von GALERKIN [55]
1943	Die Einführung lokaler Ansätze von COURANT [30]
1954	Die Matrizenformulierung der Verschiebungsmethode für Stabtragwerke von ARGYRIS [8]
1956	Die erste Anwendung der FEM mit einem Computerprogramm zur Berechnung eines Flugzeugflügels von TURNER ET AL. [176]
1967	Das erste Lehrbuch von ZIENKIEWICZ und CHEUNG [204]

Die parallel wachsende Verfügbarkeit von Rechenleistung treibt die Entwicklung der FEM bis heute an. So ist der Titel von ARGYRIS berühmtem Vortrag *The computer shapes the theory* [7] aus dem Jahr 1965 bis heute in der FEM-Gemeinschaft ein geflügeltes Wort. Zu den modernsten Entwicklungen der FEM zählen die numerische Mehrskalenmodellierung, siehe bspw. ZOHDI und WRIGGERS [207], sowie die isogeometrische Analyse von HUGHES ET AL. [89], die eine Brücke zwischen der CAD-Modellierung und der FE-Simulation schlägt.

In diesem Kapitel werden die wesentlichen Bestandteile der FEM für die nichtlineare Simulation dreidimensionaler Festkörper vorgestellt. Im Anschluss folgt eine Übersicht über weitere in den numerischen Beispielen verwendete Elementformulierungen. Für eine ausführlichere Darstellung sei auf die Literatur von WRIGGERS [198], IBRAHIMBEGOVIC [90], ZIENKIEWICZ und TAYLOR [205, 206] sowie HUGHES [88] verwiesen. Alle Elemente sind in einer erweiterten Version des FE-Programms FEAP (*Finite Element Analysis Programm*) [168] implementiert.

2.2.1 Diskretisierung eines Körpers in finite Elemente

Im Rahmen der FEM wird ein Körper \mathcal{B} , hier in seiner Referenzkonfiguration, in endlich große Teilkörper unterteilt, die als Elemente bezeichnet werden. Das

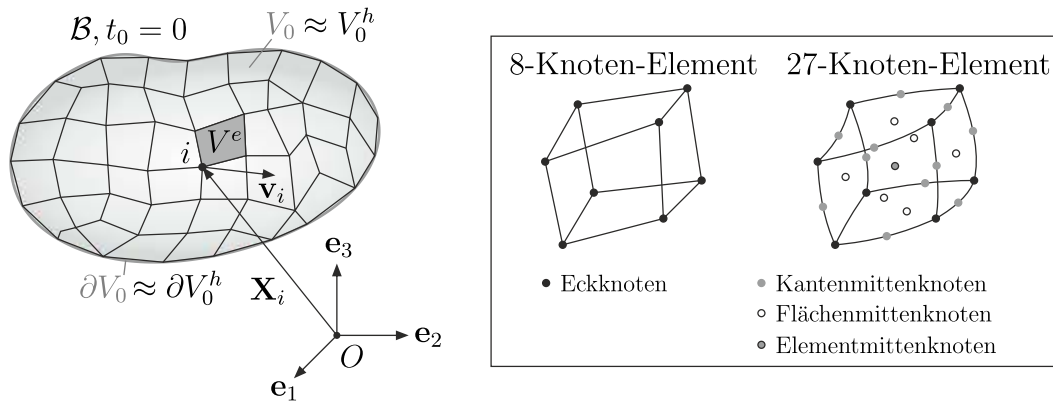


Abbildung 2.5: Diskretisierung eines Körpers \mathcal{B} in finite Elemente

ist in Abbildung 2.5 dargestellt. Das Volumen V_0 und die Oberfläche ∂V_0 werden dabei approximiert. Für das Volumen gilt bspw.

$$V_0 \approx V_0^h = \bigcup_{e=1}^{nelem} V^e, \quad (2.58)$$

wobei V^e jeweils das Volumen eines der insgesamt $nelem$ Elemente ist. Der Index h steht dabei für eine charakteristische Länge des Elements und kennzeichnet so die FE-Diskretisierung. Die Elemente sind über ihre Außenknoten miteinander verbunden. Wie in Abbildung 2.5 dargestellt, können dabei bspw. 8- oder 27-Knoten-Elemente verwendet werden. Innerhalb eines finiten Elements e wird die gesuchte Funktion, hier die Verschiebungsfunktion

$$\mathbf{u}^e \approx \mathbf{u}^h = \sum_{I=1}^{nel} N_I(\boldsymbol{\xi}) \mathbf{v}_I^e, \quad (2.59)$$

mithilfe der Knotenverschiebungsvektoren $\mathbf{v}_I^e = [u_{I1}^e, u_{I2}^e, u_{I3}^e]^T$ und Interpolationsfunktionen, sogenannten Ansatzfunktionen $N_I^e(\boldsymbol{\xi})$, approximiert. Die Anzahl der Knoten im Element wird mit nel bezeichnet. Die Ansatzfunktionen werden in einem natürlichen Koordinatenraum $\boldsymbol{\xi} = [\xi, \eta, \zeta]^T$ definiert. In diesem natürlichen Raum ist das Element ein Würfel mit Kantenlänge 2 und nimmt den Raum V^* ein. Die Gegenüberstellung beider Koordinatenräume ist in Abbildung 2.6 dargestellt. Die Ansatzfunktionen werden den jeweiligen Knoten innerhalb des Elements zugeordnet. Unter Anwendung des *isoparametrischen Konzepts* wird die Position

$$\mathbf{X}^e \approx \mathbf{X}^h = \sum_{I=1}^{nel} N_I(\boldsymbol{\xi}) \mathbf{X}_I^e \quad (2.60)$$

innerhalb des Elements mit den gleichen Ansatzfunktionen $N_I^e(\boldsymbol{\xi})$ und den zugehörigen Knotenkoordinaten $\mathbf{X}_I^e = [X_{I1}^e, X_{I2}^e, X_{I3}^e]^T$ approximiert.

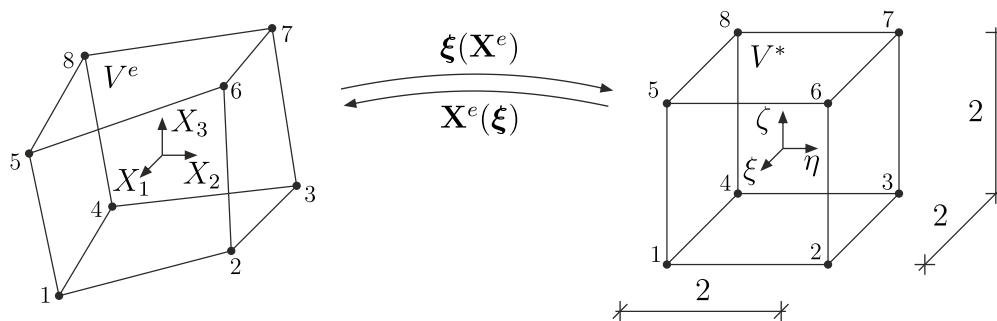


Abbildung 2.6: Ein 8-Knoten-Element e im physikalischen Raum $\mathbf{X} = [X_1, X_2, X_3]^T$ und im natürlichen Raum $\boldsymbol{\xi} = [\xi, \eta, \zeta]^T$

Mithilfe der JACOBI-Matrix

$$\mathbf{J}^e = \frac{d\mathbf{X}^e}{d\boldsymbol{\xi}} = \begin{bmatrix} X_{1,\xi}^e & X_{1,\eta}^e & X_{1,\zeta}^e \\ X_{2,\xi}^e & X_{2,\eta}^e & X_{2,\zeta}^e \\ X_{3,\xi}^e & X_{3,\eta}^e & X_{3,\zeta}^e \end{bmatrix} \approx \sum_{I=1}^{nel} \begin{bmatrix} N_{I,\xi} X_{I1}^e & N_{I,\eta} X_{I1}^e & N_{I,\zeta} X_{I1}^e \\ N_{I,\xi} X_{I2}^e & N_{I,\eta} X_{I2}^e & N_{I,\zeta} X_{I2}^e \\ N_{I,\xi} X_{I3}^e & N_{I,\eta} X_{I3}^e & N_{I,\zeta} X_{I3}^e \end{bmatrix} \quad (2.61)$$

werden Differentiale zwischen beiden Räumen transformiert. Für das Volumendifferential gilt

$$dV = \det \mathbf{J}^e d\xi d\eta d\zeta = \det \mathbf{J}^e dV^* . \quad (2.62)$$

Die Ansatzfunktionen und die Anzahl der Knoten hängen miteinander zusammen. In dieser Arbeit werden LAGRANGE'sche Interpolationsfunktionen verwendet. Diese können in die drei Raumrichtungen getrennt konstruiert werden,

$$l_I(\xi) = \prod_{\substack{k=1 \\ k \neq I}}^{p-1} \frac{\xi - \xi_k}{\xi_I - \xi_k}, \quad l_I(\eta) = \prod_{\substack{k=1 \\ k \neq I}}^{p-1} \frac{\eta - \eta_k}{\eta_I - \eta_k}, \quad l_I(\zeta) = \prod_{\substack{k=1 \\ k \neq I}}^{p-1} \frac{\zeta - \zeta_k}{\zeta_I - \zeta_k}, \quad (2.63)$$

um im Anschluss die Ansatzfunktion N_I^e des Knotens I über das Produkt

$$N_I^e(\boldsymbol{\xi}) = l_I(\xi) l_I(\eta) l_I(\zeta) \quad (2.64)$$

zu erhalten. Dabei ist p der gewünschte Polynomgrad. Für ein 8-Knoten-Element gilt bspw.

$$N_I^e(\boldsymbol{\xi}) = \frac{1}{8}(1 + \xi\xi_I)(1 + \eta\eta_I)(1 + \zeta\zeta_I), \quad I = 1, \dots, 8, \quad (2.65)$$

mit $\xi_I, \eta_I, \zeta_I \in \{-1, 1\}$. Für ein 27-Knoten-Element gilt $\xi_I, \eta_I, \zeta_I \in \{-1, 0, 1\}$. Allerdings lässt sich keine allgemeine Formel für N_I angeben, da eine Unterscheidung in die Knotentypen gemacht werden muss, siehe Abbildung 2.5.

Für eine detailliertere Darstellung sowie weitere mögliche Ansatzfunktionen wird auf die Literatur, bspw. WRIGGERS [198], verwiesen. Im Allgemeinen hat die Wahl der Ansatzfunktionen einen großen Einfluss auf die Approximationsgüte der FE-Näherungslösung.

2.2.2 Die diskretisierte Linearisierung der schwachen Form

Der FE-Ansatz (2.59) wird in die zu lösende Gleichung, hier die Linearisierung $L[\delta\pi] = 0$ der schwachen Form aus Gleichung (2.52), eingesetzt. Es gilt

$$L[\delta\pi] \approx \delta\pi^h + \Delta\delta\pi^h = \bigcup_{e=1}^{nelem} (\delta\pi^e + \Delta\delta\pi^e) = 0. \quad (2.66)$$

Die elementbezogenen Terme $\delta\pi^e$ und $\Delta\delta\pi^e$ werden mithilfe des FE-Ansatzes aus Kapitel 2.2.1 approximiert. An der Entwicklungsstelle sind \mathbf{u} und $\mathbf{S}(\mathbf{u})$ konstant. Die virtuelle Arbeit

$$\delta\pi^e = \int_{V^e} \delta\mathbf{E}^T \mathbf{S} \, dV - \int_{V^e} \rho_o \delta\mathbf{u}^T \bar{\mathbf{b}} \, dV - \int_{\partial V^e} \delta\mathbf{u}^T \bar{\mathbf{t}}_0 \, dA \quad (2.67)$$

wird als Elementresiduum bezeichnet und ist hier in reduzierter Notation geschrieben. Dabei wird in diesem Kapitel im Integranden auf die Kennzeichnung des Elementbezugs mit $(\bullet)^e$ und auf die der FE-Diskretisierung mit $(\bullet)^h$ verzichtet. Darüber hinaus wird der Term der Oberflächenlasten $\bar{\mathbf{t}}_0$ im Folgenden nicht dargestellt. Dieser kann mit Ansätzen niedrigerer Dimension diskretisiert werden, vgl. WRIGGERS [198]. Der Ansatz für die virtuellen Verschiebungen

$$\delta\mathbf{u} \approx \sum_{I=1}^{nel} N_I(\boldsymbol{\xi}) \delta\mathbf{v}_I \quad (2.68)$$

ist äquivalent zu dem der Verschiebungen \mathbf{u} aus Gleichung (2.59) und wird mit dem Vektor der virtuellen Knotenverschiebungen $\delta\mathbf{v}_I = [\delta u_{I1}, \delta u_{I2}, \delta u_{I3}]^T$ aufgestellt. Mit diesem Ansatz resultiert der virtuelle GREEN-LAGRANGE'sche Verzerrentensor in reduzierter Notation

$$\delta\mathbf{E} = \begin{bmatrix} \delta E_{11} \\ \delta E_{22} \\ \delta E_{33} \\ 2\delta E_{12} \\ 2\delta E_{13} \\ 2\delta E_{23} \end{bmatrix} = \sum_{I=1}^{nel} \mathbf{B}_I(\mathbf{u}) \delta\mathbf{v}_I. \quad (2.69)$$

Die von der Entwicklungsstelle (\mathbf{u} ist fest) abhängigen B-Matrizen

$$\mathbf{B}_I(\mathbf{u}) = \begin{bmatrix} N_{I,1}(1+u_{1,1}) & N_{I,1}u_{2,1} & N_{I,1}u_{3,1} \\ N_{I,2}u_{1,2} & N_{I,2}(1+u_{2,2}) & N_{I,2}u_{3,2} \\ N_{I,3}u_{1,3} & N_{I,3}u_{2,3} & N_{I,3}(1+u_{3,3}) \\ N_{I,2}(1+u_{1,1}) & N_{I,1}(1+u_{2,2}) & N_{I,1}u_{3,2} \\ +N_{I,1}u_{1,2} & +N_{I,2}u_{2,1} & +N_{I,2}u_{3,1} \\ N_{I,3}(1+u_{1,1}) & N_{I,1}u_{2,3} & N_{I,1}(1+u_{3,3}) \\ +N_{I,1}u_{1,3} & +N_{I,3}u_{2,1} & +N_{I,3}u_{3,1} \\ N_{I,2}u_{1,3} & N_{I,3}(1+u_{2,2}) & N_{I,2}(1+u_{3,3}) \\ +N_{I,3}u_{1,2} & +N_{I,2}u_{2,3} & +N_{I,3}u_{3,2} \end{bmatrix} \quad (2.70)$$

enthalten die Ableitungen der Ansatzfunktionen nach den physikalischen Koordinaten. Diese lassen sich aus den Ableitungen nach den natürlichen Koordinaten und der JACOBI-Matrix aus Definition (2.61) mit

$$\begin{bmatrix} N_{I,1} \\ N_{I,2} \\ N_{I,3} \end{bmatrix} = (\mathbf{J}^e)^{-T} \begin{bmatrix} N_{I,\xi} \\ N_{I,\eta} \\ N_{I,\zeta} \end{bmatrix} \quad (2.71)$$

berechnen. Sie werden auch für die Verschiebungsgradienten der aktuellen Entwicklungsstelle

$$u_{i,j} = \sum_{J=1}^{nel} N_{J,j} u_{Ji}^e \quad (2.72)$$

benötigt. Werden darüber hinaus die Integrationsgrenzen mit Gleichung (2.62) transformiert, resultiert die diskretisierte Form des Elementresiduums in den natürlichen Koordinaten

$$\begin{aligned} \delta\pi^e &= \sum_{I=1}^{nel} \delta\mathbf{v}_I^T \left\{ \underbrace{\left[\int_{V^*} \mathbf{B}_I^T \mathbf{S} \det \mathbf{J} dV^* \right]}_{\mathbf{F}_I^e} - \underbrace{\left[\int_{V^*} N_I \rho_0 \bar{\mathbf{b}} \det \mathbf{J} dV^* \right]}_{\mathbf{P}_I^e} \right\} \\ &= \sum_{I=1}^{nel} \delta\mathbf{v}_I^T \left\{ \mathbf{F}_I^e - \mathbf{P}_I^e \right\} = \sum_{I=1}^{nel} \delta\mathbf{v}_I^T \left\{ \mathbf{G}_I^e \right\}. \end{aligned} \quad (2.73)$$

Dabei sind \mathbf{F}_I^e und \mathbf{P}_I^e die auf den Knoten I bezogenen Anteile des Vektors der inneren Kräfte \mathbf{F}^e und des Elementlastvektors \mathbf{P}^e . Ihre Differenz wird als Residual- oder Fehlkraftvektor \mathbf{G}_I^e definiert. Die zugehörige Integration im Raum V^* der natürlichen Koordinaten $(\xi, \eta, \zeta) \in [-1, 1]^3$ erfolgt in der Regel numerisch mit der GAUSS-Quadratur. So gilt für den Vektor der inneren Kräfte

$$\mathbf{F}_I^e = \int_{V^*} \mathbf{B}_I^T \mathbf{S} \det \mathbf{J} d\xi d\eta d\zeta \approx \sum_{l=1}^{n_g} \mathbf{B}_I(\boldsymbol{\xi}_l)^T \mathbf{S}(\boldsymbol{\xi}_l) \det \mathbf{J}(\boldsymbol{\xi}_l) W_l, \quad (2.74)$$

mit den insgesamt n_g Stützstellen ξ_i und den zugehörigen Gewichten W_i . Stützstellen und Gewichte können bspw. aus WRIGGERS [198] entnommen werden. Im Fall eines trilinearen Ansatzes werden bspw. acht Stützstellen mit $\xi_i, \eta_i, \zeta_i \in \{-1/\sqrt{3}, 1/\sqrt{3}\}$ und $W_i \equiv 1$ gewählt. Bei einem quaderförmigen 8-Knoten-Element, welches hier dargestellt wurde, führt dies zu einer exakten Integration. In allen anderen Fällen ist sie nur eine Näherung, die mit dem Verzerrungsgrad der Elementgeometrie schlechter wird. Die Stützstellen werden auch Integrations- oder GAUSS-Punkte genannt. An dieser Stelle sei angemerkt, dass nur an diesen GAUSS-Punkten das Stoffgesetz $\mathbf{S}(\mathbf{E})$ ausgewertet wird.

Das lineare Inkrement des Elementresiduums

$$\Delta\delta\pi^e = \int_{V^e} \left(\delta\mathbf{E}^T \mathbf{C}_T \Delta\mathbf{E} + \Delta\delta\mathbf{E}^T \mathbf{S} \right) dV \quad (2.75)$$

wird auf ähnliche Weise diskretisiert. Für die linearisierten Verschiebungen

$$\Delta\mathbf{u} \approx \sum_{I=1}^{nel} N_I(\boldsymbol{\xi}) \Delta\mathbf{v}_I \quad (2.76)$$

gilt der gleiche Ansatz wie für die virtuellen Verschiebungen, mit den linearisierten Knotenverschiebungen $\Delta\mathbf{v}_I = [\Delta u_{I1}, \Delta u_{I2}, \Delta u_{I3}]^T$. Aus der Äquivalenz der Berechnungsvorschriften für die erste Variation und die Linearisierung folgen die diskretisierten linearisierten Verzerrungen

$$\Delta\mathbf{E} = \sum_{K=1}^{nel} \mathbf{B}_K(\mathbf{u}) \Delta\mathbf{v}_K \quad (2.77)$$

mit den B-Matrizen \mathbf{B} aus Definition (2.70). Die linearisierten virtuellen Verzerrungen $\Delta\delta\mathbf{E}$ lassen sich kompakt im Skalarprodukt mit den konstanten Spannungen des aktuellen Entwicklungspunkts \mathbf{S} schreiben als

$$\Delta\delta\mathbf{E}^T \mathbf{S} = \sum_{I=1}^{nel} \sum_{K=1}^{nel} \delta\mathbf{v}_I^T [G_{IK} \mathbf{I}] \Delta\mathbf{v}_K, \quad (2.78)$$

mit der 3×3 -Einheitsmatrix \mathbf{I} und den Koeffizienten

$$\begin{aligned} G_{IK} = & N_{I,1} N_{K,1} S_{11} + N_{I,2} N_{K,2} S_{22} + N_{I,3} N_{K,3} S_{33} \\ & + (N_{I,1} N_{K,2} + N_{I,2} N_{K,1}) S_{12} \\ & + (N_{I,1} N_{K,3} + N_{I,3} N_{K,1}) S_{13} \\ & + (N_{I,2} N_{K,3} + N_{I,3} N_{K,2}) S_{23}. \end{aligned} \quad (2.79)$$

Nach der Transformation der Integration in den natürlichen Raum V^* folgt die diskretisierte Form des linearen Inkrements

$$\Delta\delta\pi^e = \sum_{I=1}^{nel} \delta\mathbf{v}_I^T \left\{ \underbrace{\sum_{K=1}^{nel} \left[\int_{V^*} (\mathbf{B}_I^T \mathbf{C}_T \mathbf{B}_K + G_{IK} \mathbf{I}) \det \mathbf{J} dV^* \right]}_{\mathbf{K}_{TIK}^e} \Delta\mathbf{v}_K \right\}. \quad (2.80)$$

Dabei ist \mathbf{K}_{TK}^e der auf das Knotenpaar IK bezogene Anteil der Elementtangentialmatrix \mathbf{K}_T^e . Das Integral wird ebenfalls mit der GAUSS-Quadratur berechnet.

Die Vektoren der virtuellen und der linearisierten Verschiebungen lassen sich im Element e zusammenfassen, bspw. $\delta\mathbf{v}^e = [\delta\mathbf{v}_1^T, \dots, \delta\mathbf{v}_{nel}^T]^T \in \mathbb{R}^{nen}$. Dabei ist $nen = 3nel$ die Anzahl der unbekannt Elementverschiebungen. Dementsprechend können der Element-Residualvektor $\mathbf{G}^e \in \mathbb{R}^{nen}$ und die Elementtangentialmatrix $\mathbf{K}_T^e \in \mathbb{R}^{nen \times nen}$ blockweise zusammengebaut werden. Auf Systemebene wiederum werden diese Elementgrößen knotenbezogen zusammengebaut:

$$\delta\mathbf{v} = \bigcup_{e=1}^{nelem} \delta\mathbf{v}^e, \quad \Delta\mathbf{v} = \bigcup_{e=1}^{nelem} \Delta\mathbf{v}^e, \quad \mathbf{K}_T = \bigcup_{e=1}^{nelem} \mathbf{K}_T^e, \quad \mathbf{G} = \bigcup_{e=1}^{nelem} \mathbf{G}^e. \quad (2.81)$$

Dieser Zusammenbau wird auch als Assemblierung bezeichnet. Letztlich resultiert die diskretisierte Linearisierung des PVV, basierend auf Gleichung (2.66), mit

$$\delta\mathbf{v}^T \{ \mathbf{G} + \mathbf{K}_T \Delta\mathbf{v} \} = 0. \quad (2.82)$$

Aus der Beliebigkeit der Testfunktionen $\delta\mathbf{u}$ und der daraus folgenden Beliebigkeit der virtuellen Knotenverschiebungen $\delta\mathbf{v}$ resultiert das lineare Gleichungssystem

$$\mathbf{K}_T \Delta\mathbf{v} = -\mathbf{G}, \quad (2.83)$$

welches pro Iterationsschritt zu lösen ist. Dies ist erst möglich, wenn die Verschiebungsrandbedingungen $\mathbf{u} = \bar{\mathbf{u}}$ durch Elimination der zugehörigen Freiheitsgrade berücksichtigt werden. Die Einbettung in ein inkrementell-iterative Lösungsverfahren wird im folgenden Kapitel beschrieben.

2.2.3 Ein inkrementell-iteratives Lösungsverfahren

Das nichtlineare PVV wird mithilfe des NEWTON-Verfahrens iterativ gelöst. Durch die FE-Diskretisierung aus Kapitel 2.2.2 wird dabei nicht mehr die Verschiebungsfunktion \mathbf{u} , sondern der Vektor der Knotenverschiebungen \mathbf{v} gesucht. Dazu ist es zunächst zweckmäßig, einen Lastfaktor λ einzuführen, mit dessen Hilfe der Residualvektor des Systems mit

$$\mathbf{G}(\mathbf{v}, \lambda) = \mathbf{F}(\mathbf{v}) - \lambda\mathbf{P}_0 \quad (2.84)$$

geschrieben werden kann. Ausgehend von einem Lastschritt λ_n mit bekannten Verschiebungen \mathbf{v}_n werden nun die Verschiebungen \mathbf{v}_{n+1} für das neue und vorgegebene Lastniveau λ_{n+1} gesucht. Pro Iterationsschritt i wird dabei die diskretisierte Linearisierung

$$\mathbf{K}_T^{(i)} \Delta\mathbf{v}^{(i)} = -\mathbf{G}^{(i)} \quad (2.85)$$

aus Gleichung (2.83) gelöst. Im Anschluss wird der Vektor der Knotenverschiebungen aktualisiert mit

$$\mathbf{v}_{n+1}^{(i+1)} = \mathbf{v}_{n+1}^{(i)} + \Delta \mathbf{v}^{(i)} . \quad (2.86)$$

Zu Beginn der Iteration gilt $i = 0$ und $\mathbf{v}_{n+1}^{(0)} = \mathbf{v}_n$. Die Iteration wird abgebrochen, sobald eine gewünschte Toleranz unterschritten ist, bspw. bezüglich der Norm des Residualvektors $\|\mathbf{G}\|$. Eine grafische Interpretation dieses inkrementell-iterativen Lösungsalgorithmus ist in Abbildung 2.7 für eine einzelne unbekannte Verschiebung v dargestellt.

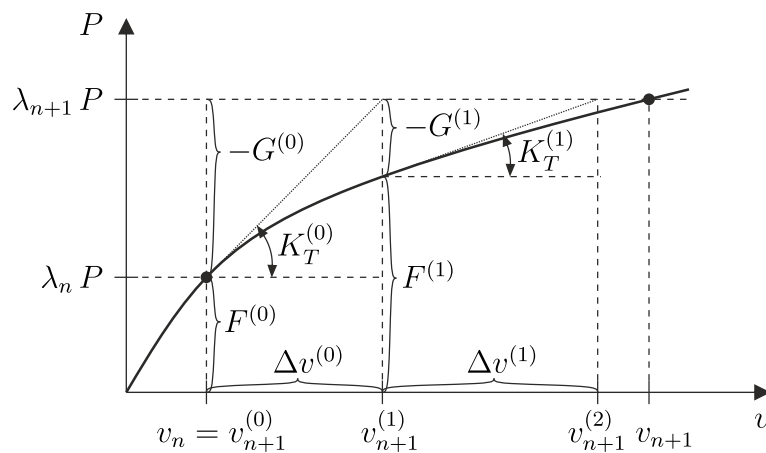


Abbildung 2.7: Grafische Interpretation des Newton-Verfahrens

Der Vorteil des NEWTON-Verfahrens ist die lokale quadratische Konvergenz. Durch Aufteilen des gewünschten Lastniveaus in kleinere Lastschritte kann dieses Verhalten besser genutzt werden. Oft ist so eine Konvergenz überhaupt erst möglich. Allerdings ist bei einer Laststeuerung $\mathbf{P} = \lambda \mathbf{P}_0$ eine Konvergenz unter Umständen gar nicht möglich, wenn $\det \mathbf{K}_T \approx 0$ gilt, wie z. B. bei Stabilitätsproblemen. In diesen Fällen wird entweder eine Verschiebungssteuerung mit $\bar{\mathbf{v}} = \lambda \bar{\mathbf{v}}_0$ genutzt oder das Bogenlängenverfahren verwendet. Letzteres erlaubt im Rahmen der Gleichgewichtsiteration simultan unbekannte Lastfaktoren und Verschiebungsgrößen, indem über die Vorgabe einer Bogenlänge passende Nebenbedingungen definiert werden. Dieses Verfahren wird hier nicht weiter vertieft. Es kann detailliert z. B. in WAGNER und WRIGGERS [187] nachgelesen werden.

2.2.4 Spezielle Strukturelemente

In einigen Problemstellungen ist es zweckmäßig anstatt des allgemeinen dreidimensionalen Körpers alternative mechanische Modelle zu wählen. Dazu gehören bspw. Balken-, Scheiben-, Platten- oder Schalenmodelle. Diese basieren auf

passenden statischen und kinematischen Annahmen. Die daraus resultierenden Elementformulierungen können sich maßgeblich von denen des dreidimensionalen Körpers unterscheiden. Sind das mechanische Modell und das zugehörige finite Element richtig gewählt, kann der notwendige Rechenaufwand erheblich reduziert werden. In diesem Kapitel werden weitere mechanische Modelle und die zugehörigen Elemente vorgestellt, die in den numerischen Beispielen Verwendung finden.

Ein Verschiebungselement für Scheibentragwerke

Ebene Tragwerke, deren Dicke d relativ klein ist gegenüber den restlichen Abmessungen und die nur in ihrer Fläche belastet sind, werden als Scheibentragwerke bezeichnet. Wenn die Verschiebungen u_3 in Dickenrichtung nicht behindert sind, liegt ein ebener Spannungszustand vor. Es gilt $S_{i3} = S_{3i} = 0$ und $E_{13} = E_{23} = E_{31} = E_{32} = 0$. Die restlichen Spannungen und Verzerrungen verlaufen in Dickenrichtung konstant. Die Scheibe wird in ihrer Mittelfläche A_0 modelliert. Dies ist in Abbildung 2.8 dargestellt.

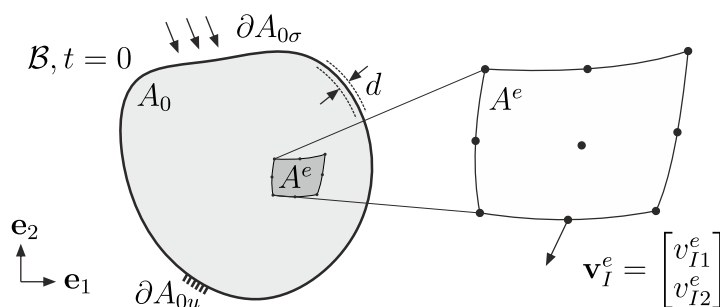


Abbildung 2.8: Diskretisierung einer Scheibe mit finiten Elementen

Der Vektor $\mathbf{u} = [u_1, u_2]^T$ enthält zwei unbekannte Verschiebungen. Die reduzierte Vektornotation der Verzerrungen und Spannungen enthält jeweils nur drei Komponenten:

$$\mathbf{E} = [E_{11}, E_{22}, 2E_{12}]^T \quad (2.87)$$

$$\mathbf{S} = [S_{11}, S_{22}, S_{12}]^T . \quad (2.88)$$

Die Verzerrung in Dickenrichtung E_{33} ist nicht null, sie hat jedoch keinen Anteil im PVV

$$\delta\pi(\mathbf{u}, \delta\mathbf{u}) = \int_{A_0} \delta\mathbf{E}^T \mathbf{S} dA - \delta\pi_a(\delta\mathbf{u}) = 0 . \quad (2.89)$$

Die Linearisierung und Herleitung der zugehörigen FE-Formulierung verlaufen äquivalent zur denjenigen eines dreidimensionalen Körpers. Dabei muss mit $dV =$

$d dA$ nur eine Flächenintegration durchgeführt werden. Die zugehörigen Elementvektoren und -matrizen haben die Größe $nen = 2 nel$. Das Beispiel eines 9-Knoten-Scheibenelements ist in Abbildung 2.8 dargestellt. Eine detailliertere Beschreibung kann bspw. in WRIGGERS [198] gefunden werden.

Ein gemischtes Schalenelement mit finiten Rotationen

Gekrümmte Strukturen, deren Dicke d relativ klein gegenüber den restlichen Abmessungen ist, werden als Schalenträgerwerke bezeichnet. Sie können in ihrer Fläche oder orthogonal dazu belastet werden. Ihre Geometrie wird in der Regel über konvektive, nicht unbedingt kartesische, Koordinaten $\{\xi^i\}_{i=1}^3$ beschrieben, wobei ξ^3 in Dickenrichtung und somit in Richtung des Direktorvektors \mathbf{D} zeigt, der normal auf der Referenzfläche Ω_0 steht. Diese ist mit $\xi^3 = 0$ definiert. In Abbildung 2.9 sind diese Sachverhalte dargestellt.

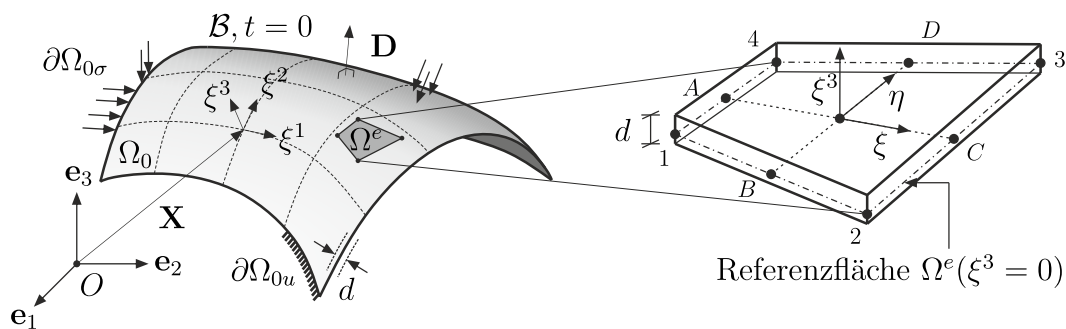


Abbildung 2.9: Diskretisierung einer Schale mit finiten Elementen

Schalen werden in der Regel mit Verzerrungs- und Schnittgrößen beschrieben, die auf die Referenzfläche Ω_0 bezogen sind. Die geometrischen Schalenverzerrungen

$$\boldsymbol{\varepsilon}_g(\mathbf{v}) = [\varepsilon_{11}, \varepsilon_{22}, 2\varepsilon_{12}, \kappa_{11}, \kappa_{22}, 2\kappa_{12}, \gamma_1, \gamma_2]^T \quad (2.90)$$

enthalten Membranverzerrungen $\varepsilon_{\alpha\beta}$, Biegeverzerrungen $\kappa_{\alpha\beta}$ und Schubverzerrungen γ_α . Sie werden nicht direkt über die Verschiebungsgrößen $\mathbf{v} = [\mathbf{u}, \boldsymbol{\omega}]^T$ berechnet, sondern mit den konvektiven Ableitungen der Ortsvektoren \mathbf{X} und $\mathbf{x}(\mathbf{X}, \mathbf{v})$ sowie der Direktorvektoren $\mathbf{D}(\mathbf{X})$ und $\mathbf{d}(\mathbf{X}, \mathbf{v})$. Im Vektor \mathbf{v} sind Verschiebungen \mathbf{u} und Verdrehungen $\boldsymbol{\omega}$ enthalten. Details können in WAGNER ET AL. [185] nachgelesen werden. Die Schalenschnittgrößen

$$\boldsymbol{\sigma}_s = [n^{11}, n^{22}, n^{12}, m^{11}, m^{22}, m^{12}, q^1, q^2]^T \quad (2.91)$$

sind die arbeitskonformen Kraftgrößen zu den Schalenverzerrungen $\boldsymbol{\varepsilon}_g$. Sie werden durch Integration der Spannungen über die Schalendicke definiert und enthalten Membrankräfte $n^{\alpha\beta}$, Biegemomente $m^{\alpha\beta}$ und Querkräfte q^α . Im Rahmen der

numerischen Beispiele werden Schalentragwerke unter Berücksichtigung großer Deformationen berechnet. Das hier dazu verwendete 4-Knoten-Schalenelement basiert auf der REISSNER-MINDLIN-Theorie mit $\gamma_\alpha \neq 0$. Es wurde 2005 von WAGNER ET AL. [185] veröffentlicht und 2008 von KLINKEL ET AL. [106] erweitert, um beliebige dreidimensionale Materialmodelle verwenden zu können. Dieses Schalenelement wird im Folgenden kurz beschrieben.

Die Elementformulierung basiert auf dem HU-WASHIZU-Variationsprinzip

$$\delta\pi(\boldsymbol{\theta}, \delta\boldsymbol{\theta}) = \delta\pi_i(\boldsymbol{\theta}, \delta\boldsymbol{\theta}) - \delta\pi_a(\delta\mathbf{u}) = 0, \quad (2.92)$$

mit

$$\delta\pi_i(\boldsymbol{\theta}, \delta\boldsymbol{\theta}) = \int_{\Omega_0} \left(\underbrace{\delta\boldsymbol{\varepsilon}_g^T \boldsymbol{\sigma}_s}_I + \underbrace{\delta\boldsymbol{\varepsilon}^T (\partial_\varepsilon \Psi - \tilde{\boldsymbol{\sigma}})}_{II} + \underbrace{\delta\boldsymbol{\sigma}_s^T (\boldsymbol{\varepsilon}_g - \boldsymbol{\varepsilon}_p)}_{III} \right) dA. \quad (2.93)$$

Im Vektor der Unbekannten $\boldsymbol{\theta} = [\mathbf{v}, \boldsymbol{\sigma}_s, \boldsymbol{\varepsilon}]^T$ sind neben den Verschiebungsgrößen \mathbf{v} auch die Verzerrungsgrößen $\boldsymbol{\varepsilon}$ und die Schnittgrößen $\boldsymbol{\sigma}_s$ in $\tilde{\boldsymbol{\sigma}} = [\boldsymbol{\sigma}_s, 0, 0]^T$ als unabhängige Felder enthalten. Die von \mathbf{v} abhängigen Verzerrungen $\boldsymbol{\varepsilon}_g$ sind nicht Teil des Vektors $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_p, \boldsymbol{\varepsilon}_z]^T$ der unabhängigen Verzerrungen. Dieser teilt sich auf in den Anteil der klassischen Schalenverzerrungen $\boldsymbol{\varepsilon}_p$ nach Definition (2.90) und einen zusätzlichen Anteil $\boldsymbol{\varepsilon}_z$, der lineare Verzerrungen in Dickenrichtung berücksichtigt. Somit fordert die Stationaritätsbedingung (2.92) neben den Gleichgewichtsbedingungen in Term *I* die folgenden Gleichungen schwach ein.

Das Stoffgesetz	$\partial_{\boldsymbol{\varepsilon}_p} \Psi = \boldsymbol{\sigma}_s$	in Term <i>II</i>
Der ebene Spannungszustand	$\partial_{\boldsymbol{\varepsilon}_z} \Psi = [0, 0]^T$	in Term <i>II</i>
Die Geometrie	$\boldsymbol{\varepsilon}_g = \boldsymbol{\varepsilon}_p$	in Term <i>III</i>

Der Term

$$\partial_\varepsilon \Psi = \begin{bmatrix} \partial_{\boldsymbol{\varepsilon}_p} \Psi \\ \partial_{\boldsymbol{\varepsilon}_z} \Psi \end{bmatrix} := \int_d \mathbf{A}^T \mathbf{S} \bar{\mu} d\xi^3 \quad (2.94)$$

wird durch die Dickenintegration des 2. PIOLA-KIRCHHOFF'schen Spannungsvektors \mathbf{S} berechnet, wobei mit \mathbf{A} über $\mathbf{E} = \mathbf{A}\boldsymbol{\varepsilon}$ die Schalenverzerrungen $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_p, \boldsymbol{\varepsilon}_z]^T$ auf die GREEN-LAGRANGE'schen Verzerrungen \mathbf{E} abgebildet werden. Die Determinante $\bar{\mu}$ des Shiftertensors berücksichtigt die Rotation der ξ^3 -Achse über die Höhe und kann für dünne Schalen zu $\bar{\mu} = 1$ angenommen werden. Die Linearisierung

$$L[\delta\pi(\delta\boldsymbol{\theta}, \Delta\boldsymbol{\theta})] = \delta\pi(\delta\boldsymbol{\theta}) + \Delta\delta\pi(\delta\boldsymbol{\theta}, \Delta\boldsymbol{\theta}) \quad (2.95)$$

von Gleichung (2.92) enthält das Inkrement

$$\begin{aligned} \Delta\delta\pi(\delta\boldsymbol{\theta}, \Delta\boldsymbol{\theta}) = \int_{\Omega_0} & \left(\delta\boldsymbol{\varepsilon}^T (\mathbf{D}\Delta\boldsymbol{\varepsilon} - \Delta\tilde{\boldsymbol{\sigma}}) + \delta\boldsymbol{\sigma}_s^T (\Delta\boldsymbol{\varepsilon}_g - \Delta\boldsymbol{\varepsilon}_p) \right. \\ & \left. + \delta\boldsymbol{\varepsilon}_g^T \Delta\boldsymbol{\sigma}_s + \Delta\delta\boldsymbol{\varepsilon}_g^T \boldsymbol{\sigma}_s \right) dA. \end{aligned} \quad (2.96)$$

Die Materialtangente \mathbf{C}_T aus Gleichung (2.54) ist im Integral

$$\mathbf{D} = \int_d \mathbf{A}^T \mathbf{C}_T \mathbf{A} \bar{\mu} d\xi^3 \quad (2.97)$$

über eine Dickenintegration enthalten. Mit den Gleichungen (2.97) und (2.94) ist ersichtlich, wie ein dreidimensionales Stoffgesetz, mit $\mathbf{S}(\mathbf{E})$ und $\Delta \mathbf{S} = \mathbf{C}_T \Delta \mathbf{E}$, im Rahmen dieser Schalenformulierung berücksichtigt werden kann.

Die Linearisierung (2.96) wird auf Elementebene nach dem isoparametrischen Konzept diskretisiert. Das finite Element ist in Abbildung 2.9 dargestellt. Die Ansätze für die geometrischen Größen sind stetig und bilinear:

$$\mathbf{X}^h = \sum_{I=1}^4 N_I(\xi, \eta) \mathbf{X}_I, \quad \mathbf{D}^h = \sum_{I=1}^4 N_I(\xi, \eta) \mathbf{D}_I, \quad (2.98)$$

$$\mathbf{x}^h = \sum_{I=1}^4 N_I(\xi, \eta) \mathbf{x}_I, \quad \mathbf{d}^h = \sum_{I=1}^4 N_I(\xi, \eta) \mathbf{d}_I. \quad (2.99)$$

Mit den Ableitungen in Richtung der konvektiven Koordinaten $\xi_\alpha, \alpha = 1, 2$ können so die zugehörigen geometrischen Membran- und Biegeverzerrungen $\varepsilon_{\alpha\beta}$ und $\kappa_{\alpha\beta}$ berechnet werden, siehe KLINKEL ET AL. [106]. Würden die Schubverzerrungen $\gamma_\alpha \neq 0$ dagegen über die gleichen Ansätze approximiert werden, würden sich die Elemente zu steif verhalten. Dieser Schub-Versteifungseffekt (engl. *locking*) resultiert aus einer unzureichenden Approximation des Verzerrungsfeldes. Dabei spielen sowohl die Ansätze als auch die numerische Integration eine Rolle. Die oft als Lösung genutzte reduzierte Integration der zugehörigen Integrale führt zu zusätzlichen Kinematiken, sogenannten *Hourglass-Modes*. Daher werden die Schubverzerrungen mithilfe eines unabhängigen, linearen Ansatzes und den Werten γ_α an den Knoten A,B,C und D nach DVORKIN und BATHE [39] definiert, vgl. Abbildung 2.9. Dadurch tritt das Locking-Phänomen nicht auf, trotz gleicher Integrationsordnung aller Integrale. Die unabhängigen Spannungen

$$\boldsymbol{\sigma}_s^h = \mathbf{N}_\sigma \hat{\boldsymbol{\sigma}} \quad (2.100)$$

werden linear und diskontinuierlich über die Elementgrenzen diskretisiert. Der Vektor $\hat{\boldsymbol{\sigma}}$ enthält 14 Variablen. Die Komponenten der Ansatzmatrix \mathbf{N}_σ können aus KLINKEL ET AL. [106] entnommen werden. Der Ansatz für das Verzerrungsfeld

$$\boldsymbol{\varepsilon}^h = \mathbf{N}_\varepsilon \hat{\boldsymbol{\varepsilon}} = \begin{bmatrix} \mathbf{N}_\varepsilon^1 & \mathbf{N}_\varepsilon^2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\varepsilon}}_1 \\ \hat{\boldsymbol{\varepsilon}}_2 \end{bmatrix} \quad (2.101)$$

enthält zwei Anteile. Der erste Anteil $\mathbf{N}_\varepsilon^1 \hat{\boldsymbol{\varepsilon}}_1$ entspricht dem Spannungsansatz aus Gleichung (2.100), ist linear und enthält ebenfalls 14 Variablen in $\hat{\boldsymbol{\varepsilon}}_1$. Der zweite Anteil $\mathbf{N}_\varepsilon^2 \hat{\boldsymbol{\varepsilon}}_2$ enthält zu (2.100) orthogonale Ansätze mit insgesamt $\beta = \beta_m + \beta_b + \beta_z$

zusätzlichen Ansatzparametern, bestehend aus Membran-, Biege- und Dickenverzerrungsanteilen. Die Wahl dieser EAS-Parameter (engl.: *Enhanced Assumed Strain*) nach SIMO und RIFAI [159]) kann wie folgt festgelegt werden.

$$\text{Membranverzerrung: } \beta_m \in \{0, 2, 4, 7, 9\}$$

$$\text{Biegeverzerrung: } \beta_b \in \{0, 2, 4, 6\}$$

$$\text{Dickenverzerrung: } \beta_z \in \{2, 8\}$$

Die Wahl der EAS-Parameter kann einen Einfluss auf die Lösung haben und sollte daher stets in den numerischen Beispielen angegeben werden. Die Ansätze für die drei unabhängigen Felder, ihre 1. Variationen und Linearisierungen werden in Gleichung (2.96) auf Elementebene eingesetzt. Durch die diskontinuierlichen Ansätze für die Spannungen und die Verzerrungen können die Unbekannten $\Delta\hat{\boldsymbol{\sigma}}$ und $\Delta\hat{\boldsymbol{\varepsilon}}$ auf Elementebene eliminiert werden. Dadurch müssen die Elementtangentialmatrix \mathbf{K}_T^e und das Elementresiduum \mathbf{G}^e auf Systemebene nur bezüglich der Verschiebungsfreiheitsgrade \mathbf{v} assembliert werden. Die zugehörigen Flächenintegrationen werden mit vier GAUSS-Punkten durchgeführt. Innerhalb der Schalfläche hat das Element fünf Freiheitsgrade, an Schnittkanten sechs.

Um bspw. im Rahmen plastischen Materialverhaltens die Genauigkeit der Dickenintegration in den Gleichungen (2.94) und (2.97) zu verbessern, können numerische Schichten definiert werden. Dabei wird jede Schicht mit jeweils mindestens zwei GAUSS-Punkten integriert. Dies ist in Abbildung 2.10 dargestellt.

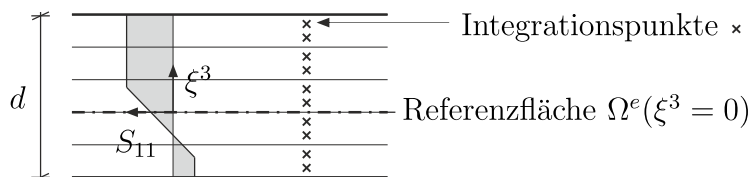


Abbildung 2.10: Dickenintegration mithilfe numerischer Schichten

Diese Schalenelementklasse wird stetig weiterentwickelt. In GRUTTMANN ET AL. [67] wurde das Element auf physikalische Schichtungen erweitert, um den komplexen Spannungszustand in geschichteten Bauteilen zu berücksichtigen. In WAGNER ET AL. [186] wurden intensive Untersuchungen zum Einfluss von quadratischen Ansätzen mit $\beta_m = 11$ durchgeführt.

Nachdem innerhalb dieses Kapitels die Form des Materialgesetzes $\mathbf{S}(\mathbf{E})$ zunächst ausgeklammert wurde, folgt in Kapitel 3 eine detailliertere Übersicht über die Methoden und physikalischen Hintergründe der Modellierung von elastischem und plastischem Materialverhalten. Auf diese Hintergründe wird im Rahmen der Materialmodellierung mit künstlichen neuronalen Netzen im späteren Verlauf der Arbeit häufig Bezug genommen.

3 Phänomenologische Materialmodelle

Zur Anwendung von künstlichen neuronalen Netzen in der Materialmodellierung ist das Wissen über klassische, phänomenologische Materialmodelle und die ihnen zugrunde liegenden physikalischen Prinzipien unerlässlich. Ziel dieses Kapitels ist es daher, diese Grundlagen und die für spätere Untersuchungen nötigen Materialgleichungen bereitzustellen. Materialmodelle sind *phänomenologisch*, wenn sie auf experimentellen Beobachtungen des makroskopischen Materialverhaltens basieren. So geht beispielsweise der mikrostrukturelle Aufbau nicht direkt in die Formulierung der Materialgleichungen ein.

Die zugehörigen Gleichungen verknüpfen makroskopische Konstitutivgrößen am materiellen Punkt \mathbf{X} miteinander, unabhängig von der Geometrie des Gesamtkörpers. Im rein mechanischen Fall sind das die Spannungen und die Dehnungen. Schon früh hat dies LEONHARD EULER erkannt, als er im Jahr 1727 eine Materialkonstante im Sinn des heutigen Elastizitätsmoduls verwendete, vgl. SZABO [167]. Heute ist die Materialmodellierung ein eigenständiges Forschungsgebiet der Kontinuumsmechanik. Dabei wird das zu beschreibende Materialverhalten durch wachsende Anforderungen und technische Fortschritte immer komplexer. Unter anderem sind in ALTENBACH [4], HOLZAPFEL [81] sowie LEMAITRE und CHABOCHE [114] ausführliche Kapitel zur Materialmodellierung enthalten.

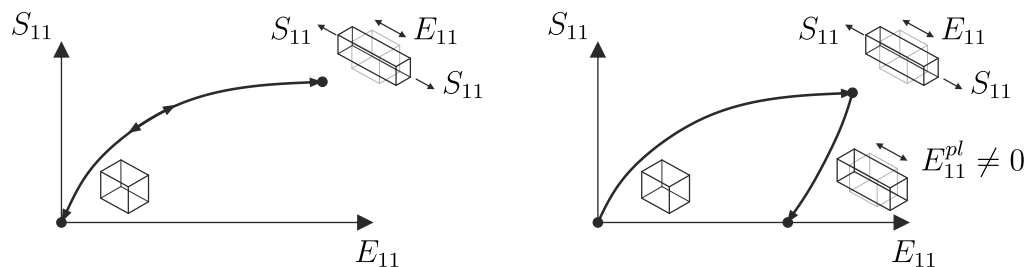


Abbildung 3.1: Rein elastisches (links) und plastisches Materialverhalten (rechts) am Beispiel einaxialer Spannungszustände. Plastisches Verhalten ist charakterisiert durch verbleibende Deformationen nach Entlastung.

In dieser Arbeit wird nur elastisches und ratenunabhängiges, plastisches Materialverhalten betrachtet. Diese sind in Abbildung 3.1 am Beispiel einaxialer Spannungszustände dargestellt. Weitere Effekte wie bspw. Ratenabhängigkeiten und Schädigung werden nicht berücksichtigt⁸.

⁸ Echte Materialien zeigen oft mehrere dieser Eigenschaften. Je nach Anwendungsfall sind diese Eigenschaften jedoch verschieden stark ausgeprägt und können ggf. vernachlässigt werden. So verhält sich bspw. Stahl bei hohen Temperaturen viskos, was bei Raumtemperatur im Normalfall nicht berücksichtigt werden muss.

Im Folgenden werden hauptsächlich die GREEN-LAGRANGE'schen Verzerrungen \mathbf{E} und die 2. PIOLA-KIRCHHOFF'schen Spannungen \mathbf{S} als konstitutive Größen verwendet. Für elastisches Materialverhalten lassen sich die Spannungen

$$\mathbf{S}(\mathbf{X}, t) = \mathcal{S}(\mathbf{X}, \mathbf{E}(\mathbf{X}, t)) \stackrel{\text{hom}}{=} \mathcal{S}(\mathbf{E}(\mathbf{X}, t)) \quad (3.1)$$

allgemein als Funktion \mathcal{S} der aktuellen Verzerrungen $\mathbf{E}(\mathbf{X}, t)$ angeben. Ist der Werkstoff homogen (hom), entfällt die direkte Abhängigkeit von der materiellen Koordinate \mathbf{X} . Elastoplastisches Materialverhalten ist abhängig von der Belastungsgeschichte. Die aktuellen Spannungen müssen daher als Funktional

$$\mathbf{S}(\mathbf{X}, t) = \mathcal{S}(\mathbf{X}, \mathbf{E}(\mathbf{X}, \tilde{t})) \stackrel{\text{hom}}{=} \mathcal{S}(\mathbf{E}(\mathbf{X}, \tilde{t})) \quad \tilde{t} \in [0, t] , \quad (3.2)$$

beschrieben werden. Die allgemeinen Gleichungen (3.1) und (3.2) sind die Grundlage für die nachfolgenden Ausführungen. Zunächst wird im Abschnitt 3.1 das Transformationsverhalten von Tensoren bei Wechsel des Koordinatensystems beschrieben. Dies ist eine wichtige Basis für die in Abschnitt 3.2 folgenden physikalische Prinzipien der Materialmodellierung. In den Abschnitten 3.3 und 3.4 werden klassische elastische und elastoplastische Materialmodelle behandelt, auf die im Rahmen dieser Arbeit Bezug genommen wird.

3.1 Passive Transformation von Tensoren

Die Transformation eines im dreidimensionalen kartesischen Koordinatensystem $\{\mathbf{e}_i\}_{i=1}^3$ definierten Tensors kann aktiv oder passiv sein. Im aktiven Fall wird der Tensor selbst transformiert, im passiven Fall das Koordinatensystem. Für eine ausführliche Beschreibung dieser Sachverhalte siehe BRANNON [23]. Zur Materialmodellierung ist vor allem der passive Fall aufschlussreich, der in Abbildung 3.2 dargestellt ist. Unter Vernachlässigung reiner Translationen werden die Basisvektoren $\{\mathbf{e}_i\}_{i=1}^3$ über

$$\mathbf{e}_i^* = \mathbf{Q} \mathbf{e}_i , \quad \mathbf{Q} \in \text{O}(3) \quad (3.3)$$

zu den Basisvektoren $\{\mathbf{e}_i^*\}_{i=1}^3$ transformiert. Die sog. orthogonale Gruppe $\text{O}(3)$ enthält alle Tensoren \mathbf{Q} mit der Eigenschaft $\mathbf{Q}^T = \mathbf{Q}^{-1}$, woraus $|\det \mathbf{Q}| = 1$ folgt⁹. Die Komponenten von \mathbf{Q} können bzgl. des Koordinatensystems $\{\mathbf{e}_i\}_{i=1}^3$ mit dem Skalarprodukt $Q_{ij} = \mathbf{e}_i \cdot \mathbf{e}_j^*$ berechnet werden.

Durch die Transformation der Basen ändern sich ebenfalls die Komponenten eines Tensors, nicht jedoch der Tensor selbst als physikalisches Objekt. Für einen Tensor 2. Stufe gilt bspw.

$$\mathbf{A} = A_{ij} \mathbf{e}_i \otimes \mathbf{e}_j = A_{ij}^* \mathbf{e}_i^* \otimes \mathbf{e}_j^* = \mathbf{A}^* . \quad (3.4)$$

⁹ Bei Spiegelungen ($\det \mathbf{Q} = -1$) ist das Koordinatensystem $\{\mathbf{e}_i^*\}_{i=1}^3$ nicht mehr rechtshändig.

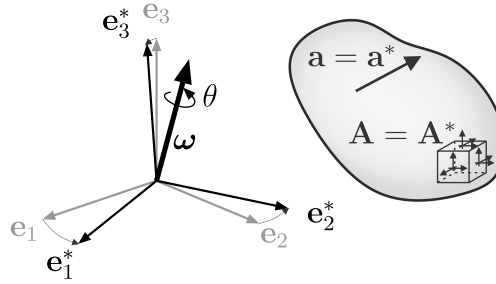


Abbildung 3.2: Transformation des Koordinatensystems $\{\mathbf{e}_i\}_{i=1}^3$ zu $\{\mathbf{e}_i^*\}_{i=1}^3$ durch eine Rotation um den Einheitsvektor $\boldsymbol{\omega}$ mit dem Winkel θ

Aus dieser Äquivalenz resultieren, durch beidseitiges skalares Multiplizieren der Einheitsvektoren, die folgenden Transformationsregeln für ein-, zwei- und vierstufige Tensoren.

$$a_k^* = a_i Q_{ik} \quad \text{oder} \quad [\mathbf{a}]^* = [\mathbf{Q}]^T [\mathbf{a}] \quad (3.5)$$

$$A_{kl}^* = A_{ij} Q_{ik} Q_{jl} \quad \text{oder} \quad [\mathbf{A}]^* = [\mathbf{Q}]^T [\mathbf{A}] [\mathbf{Q}] \quad (3.6)$$

$$A_{mnop}^* = A_{ijkl} Q_{im} Q_{jn} Q_{ko} Q_{lp} \quad (3.7)$$

Eine präzise Notation sollte grundsätzlich zwischen den Tensoren \mathbf{a} und \mathbf{A} und den Matrizen ihrer Komponenten $[\mathbf{a}]$ und $[\mathbf{A}]$ unterscheiden. Aus ästhetischen Gründen wird im Folgenden allerdings darauf verzichtet, es sei denn der Unterschied ist essenziell oder aus dem Kontext nicht direkt ableitbar.

Beschreibt \mathbf{Q} eine Rotation, gilt $\det(\mathbf{Q}) = 1$ und $\mathbf{Q} \in \text{SO}(3) \subset \text{O}(3)$, der *speziellen* orthogonalen Gruppe. Eine solche Rotation lässt sich bspw. um einen Einheitsvektor $\boldsymbol{\omega}$ mit dem Winkel θ definieren, wie in Abbildung 3.2 dargestellt ist. Die zugehörige Komponentenmatrix \mathbf{Q}^R kann in diesem Fall über eine Form der EULER-RODRIGUES-Formel

$$\mathbf{Q}^R = \mathbf{I} + \sin \theta \boldsymbol{\Omega} + (1 - \cos \theta) \boldsymbol{\Omega}^2 \quad (3.8)$$

berechnet werden. Dabei ist die schiefsymmetrische Kreuzproduktmatrix

$$\boldsymbol{\Omega} := \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} \quad (3.9)$$

über die Äquivalenz zum Kreuzprodukt $\boldsymbol{\omega} \times \mathbf{a} = \boldsymbol{\Omega} \mathbf{a}$ definiert. Eine Spiegelung ($\det \mathbf{Q} = -1$) an einer Ebene kann mit der HOUSEHOLDER-Matrix

$$\mathbf{Q}^S = \mathbf{I} - 2 \boldsymbol{\nu} \boldsymbol{\nu}^T \quad (3.10)$$

berechnet werden. Dabei ist $\boldsymbol{\nu}$ der normierte Normalenvektor der Ebene und hier als Spaltenvektor definiert. Eine Kombination aus mehreren Transformationen $\mathbf{Q}_1, \dots, \mathbf{Q}_n$ hintereinander kann mit $\mathbf{Q} = \mathbf{Q}_n \dots \mathbf{Q}_1$ zusammengefasst werden.

Komponenten-assozierte Größen von Tensoren, die sich bei Koordinatentransformation nicht ändern, werden als *Invarianten* bezeichnet. So sind die Eigenwerte $\{\lambda_a\}_{a=1}^3$ und Eigenvektoren $\{\mathbf{n}_a\}_{a=1}^3$ des speziellen Eigenwertproblems

$$\mathbf{A}\mathbf{n} = \lambda\mathbf{n} \quad (3.11)$$

unabhängig vom Koordinatensystem. Die Eigenvektoren bilden das orthogonale¹⁰ *Hauptachsensystem*, in welchem der zweistufige Tensor \mathbf{A} diagonalisiert ist. Die zugehörige *spektrale* Darstellung lautet

$$\mathbf{A} = \sum_{a=1}^3 \lambda_a \mathbf{n}_a \otimes \mathbf{n}_a . \quad (3.12)$$

Die Ableitung eines Eigenwerts nach seinem Tensor lautet

$$\frac{\partial \lambda_a}{\partial \mathbf{A}} = \mathbf{n}_a \otimes \mathbf{n}_a \quad (3.13)$$

und wird im späteren Verlauf der Arbeit noch benötigt. Ferner sind für einen Tensor 2. Stufe \mathbf{A} die drei Hauptinvarianten

$$I_{\mathbf{A}} = \text{Sp}(\mathbf{A}) = A_{ii} = A_{11} + A_{22} + A_{33} = \lambda_1 + \lambda_2 + \lambda_3 \quad (3.14)$$

$$\begin{aligned} II_{\mathbf{A}} &= \frac{1}{2} (\text{Sp}(\mathbf{A})^2 - \text{Sp}(\mathbf{A}^2)) \\ &= \frac{1}{2} (A_{ii}A_{jj} - A_{ij}A_{ji}) = \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3 \end{aligned} \quad (3.15)$$

$$III_{\mathbf{A}} = \det(\mathbf{A}) = \epsilon_{ijk} A_{1i} A_{2j} A_{3k} = \lambda_1 \lambda_2 \lambda_3 \quad (3.16)$$

unabhängig vom Koordinatensystem. Dabei sind ϵ_{ijk} die Komponenten des dreistufigen Permutationstensors¹¹. Die Ableitungen der Invarianten

$$\frac{\partial I_{\mathbf{A}}}{\partial \mathbf{A}} = \mathbf{I} , \quad \frac{\partial II_{\mathbf{A}}}{\partial \mathbf{A}} = \text{Sp}(\mathbf{A})\mathbf{I} - \mathbf{A}^T \quad \text{und} \quad \frac{\partial III_{\mathbf{A}}}{\partial \mathbf{A}} = \det(\mathbf{A})\mathbf{A}^{-T} \quad (3.17)$$

sind bspw. in HOLZAPFEL [81] zu finden.

¹⁰ Die Länge der Eigenvektoren ist unbestimmt und muss festgelegt werden. Für $\|\mathbf{n}_a\| = 1$ ist das Hauptachsensystem orthonormal.

¹¹ Die Komponenten des auch als LEVI-LEVITA-Symbol bezeichneten dreistufigen Tensors $\epsilon = \epsilon_{ijk} \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k$ haben die folgenden Eigenschaften: $\epsilon_{ijk} = 1$ bei gerader Permutation von $\{1,2,3\}$, $\epsilon_{ijk} = -1$ bei ungerader Permutation, $\epsilon_{ijk} = 0$ sonst.

3.2 Physikalische Prinzipien der Materialmodellierung

Die Physik stellt einen axiomatischen Rahmen bereit, in dem sich Materialgleichungen bewegen dürfen. Eine Übersicht darüber, wie diese Restriktion im klassischen Sinne a priori oder a posteriori für Materialmodelle sichergestellt werden, ist bspw. in ALTENBACH [4] zu finden. Bei der Materialmodellierung mit künstlichen neuronalen Netzen ist das Wissen über diese Prinzipien und ihre Anwendungsgrenzen von großer Bedeutung. Einige davon sind im Folgenden aufgelistet und werden im Verlauf dieses Abschnitts erläutert. Für weitere Informationen siehe bspw. BERTRAM [13].

- **Determinismus:** Der aktuelle Spannungszustand im Punkt \mathbf{X} wird durch die gesamte Vorgeschichte der Bewegung des Körpers \mathcal{B} bestimmt.
- **Lokale Wirkung:** Der Spannungszustand im Punkt \mathbf{X} wird nur durch die Deformationsgeschichte in einer kleinen Umgebung dieses Punktes beeinflusst. Durch den Abbruch der Taylorreihe (2.4) nach dem ersten Glied wurde diese Umgebung bereits als infinitesimal klein definiert. In Kombination mit dem Prinzip des Determinismus resultieren die Gleichungen (3.1) und (3.2) für sogenannte *Körper 1. Grades*.
- **Konsistenz mit physikalischen Bilanzgleichungen:** Die Zustandsänderung eines Körpers wird als thermodynamisch konsistent bezeichnet, wenn sie die nachfolgenden Erhaltungssätze nicht verletzen.
 1. Massenerhaltung: Durch $\rho_t = J\rho_0$ erfüllt, vgl. Kapitel 2.
 2. Impulserhaltung: Als zu lösende Gleichung (2.36) im Rahmen des Randwertproblems eingefordert, vgl. Kapitel 2.
 3. Drehimpulserhaltung: Das Materialmodell darf die Symmetriebedingung ($\boldsymbol{\sigma} = \boldsymbol{\sigma}^T$ bzw. $\mathbf{S} = \mathbf{S}^T$) nicht verletzen, vgl. Kapitel 2.
 - 4./5. Energieerhaltung und Entropieungleichung: Das Materialmodell muss die Dissipationsungleichung erfüllen, vgl. Abschnitt 3.2.1.
- **Materielle Objektivität:** Das Materialverhalten muss unabhängig von der Wahl des Bezugssystems sein. Dieses wichtige Konzept wird in Abschnitt 3.2.2 näher betrachtet.

Neben diesen axiomatischen Prinzipien lassen sich die Materialmodelle durch Einforderung materieller Symmetrien oder kinematischer Zwangsbedingungen weiter spezifizieren. Darauf wird in den Abschnitten 3.2.3 und 3.2.4 näher eingegangen.

3.2.1 Die Dissipationsungleichung der Thermodynamik

Die Thermodynamik beschäftigt sich mit der Wärme- und Energieübertragung in makroskopischen Systemen. Allgemeine Grundlagen sind beispielsweise in STEPHAN ET AL. [164] oder STIERSTADT [165] zu finden. Die Übertragung auf das Gebiet der Kontinuumsmechanik geht auf COLEMAN und GURTIN [27] sowie auf COLEMAN und NOLL [29] zurück. Ziel dieses Abschnitts ist die Herleitung der für die Materialmodellierung wichtigen Dissipationsungleichung, die sich aus dem ersten und zweiten Hauptsatz der Thermodynamik ergibt. Dazu müssen zunächst die folgenden kalorischen Feldgrößen eingeführt werden: die Temperatur $T(\mathbf{X}, t) > 0$ (in Kelvin), der materielle Wärmeflussvektor $\mathbf{q}(\mathbf{X}, t)$, die spezifische innere Energiedichte¹² $e(\mathbf{X}, t)$, die lokale Wärmezufuhrdichte $r(\mathbf{X}, t)$ und die spezifische Entropiedichte¹³ $s(\mathbf{X}, t)$. Dabei sind die letzten drei Größen massebezogen.

Der erste Hauptsatz der Thermodynamik postuliert u. a. in geschlossenen Systemen Energieerhaltung. Die zugehörige materielle lokale Form

$$\dot{e} = \frac{1}{\rho_0} \mathbf{S} : \dot{\mathbf{E}} + r - \frac{1}{\rho_0} \operatorname{div}(\mathbf{q}) \quad (3.18)$$

muss für jeden Punkt \mathbf{X} erfüllt sein. Sie sagt aus, dass sich die Änderung der spezifischen inneren Energiedichte aus der lokalen Spannungsleistung, der lokalen Wärmeproduktion und der äußeren Wärmezufuhr zusammensetzen muss. Der zweite Hauptsatz legt die zeitliche Richtung thermodynamischer Prozesse fest¹⁴. Die Ungleichung

$$\dot{s} + \frac{1}{\rho_0} \operatorname{div}\left(\frac{1}{T}\mathbf{q}\right) - \frac{r}{T} \geq 0 \quad (3.19)$$

ist die zugehörige materielle lokale Form. Sie folgt aus der Bedingung, dass die Entropieproduktionsdichte nichtnegativ sein darf; sie wird auch als CLAUSIUS-DUHEM-Ungleichung bezeichnet. Wird r in beiden Gleichungen eliminiert und das Resultat auf beiden Seiten mit $\rho_0 T > 0$ multipliziert, gilt

$$\rho_0 T \dot{s} + T \operatorname{div}\left(\frac{1}{T}\mathbf{q}\right) - \rho_0 \dot{e} - \operatorname{div}(\mathbf{q}) + \mathbf{S} : \dot{\mathbf{E}} \geq 0. \quad (3.20)$$

¹² Die innere Energie $\int_{V_0} e \rho_0 dV$ ist ein Maß für die gesamte zur Verfügung stehende Energie des sich im Gleichgewicht befindlichen Körpers.

¹³ Die Entropie ist ein Maß für die Menge an Möglichkeiten, wie Energie in einem Mehrteilchensystem verteilt werden kann. Durch Einführung dieser extensiven (sich mit Größe des Systems ändernden) Größe konnte CLAUSIUS um das Jahr 1850 erklären, wieso sich Wärme nicht immer vollständig in Arbeit umwandeln lässt, siehe STIERSTADT [165].

¹⁴ Tatsächlich ist dies eine der wenigen physikalischen Gleichungen, die einen Zeitpfeil definiert. Die Lösung anderer Gleichungen, wie bspw. die der Bewegungsgleichung, können stets vorwärts und rückwärts ablaufen.

Durch die analytische Umformung

$$T \operatorname{div}\left(\frac{1}{T}\mathbf{q}\right) = \operatorname{div}(\mathbf{q}) - \frac{1}{T}\mathbf{q} \cdot \operatorname{grad}(T) , \quad (3.21)$$

siehe bspw. DE BOER [33], folgt zunächst

$$\rho_0(T\dot{s} - \dot{e}) - \frac{1}{T}\mathbf{q} \cdot \operatorname{grad}(T) + \mathbf{S} : \dot{\mathbf{E}} \geq 0 . \quad (3.22)$$

Über die LEGENDRE-Transformation der inneren Energiedichte e bezüglich der Entropiedichte s wird die spezifische freie Energiedichte

$$\psi = e - T s , \quad \text{mit} \quad \dot{\psi} = \dot{e} - \dot{T} s - T \dot{s} \quad (3.23)$$

hergeleitet. Sie wird auch freie HELMHOLTZ-Energiedichte genannt. Nach Einsetzen in die Ungleichung (3.22) resultiert die Dissipationsungleichung

$$\mathbf{S} : \dot{\mathbf{E}} - \rho_0(\dot{T} s + \dot{\psi}) - \frac{1}{T}\mathbf{q} \cdot \operatorname{grad}(T) \geq 0 . \quad (3.24)$$

Die freie Energie ψ wird zu $\Psi := \rho_0\psi$ umgeformt, um sie auf das Volumen der Referenzkonfiguration zu beziehen. Für isotherme Prozesse ($\dot{T} = 0$) und unter Vernachlässigung eines Wärmestroms ($\mathbf{q} = \mathbf{0}$) folgt letztlich die innerhalb der vorliegenden Arbeit wichtige, rein mechanische Dissipationsungleichung

$$\dot{\mathcal{D}} := \mathbf{S} : \dot{\mathbf{E}} - \dot{\Psi} \geq 0 . \quad (3.25)$$

Sie wird auch CLAUSIUS-PLANCK-Ungleichung genannt. Dabei ist $\dot{\mathcal{D}}$ die spezifische Dissipationsleistung und Ψ im Folgenden die *Verzerrungsenergiedichtefunktion*. Im elastischen Fall gilt $\dot{\mathcal{D}} = 0$. Umgestellt zu $\dot{\Psi} \leq \mathbf{S} : \dot{\mathbf{E}}$ lässt sie sich so deuten: die Änderungsrate der gespeicherten Energie darf nicht größer sein als die dazu zur Verfügung stehende Leistung. Erfüllt ein rein mechanisches Materialmodell die Ungleichung (3.25), so ist es automatisch konsistent mit dem ersten und zweiten Hauptsatz der Thermodynamik. Dieser Umstand wird von vielen Materialmodellen genutzt, auch von einigen, die mit künstlichen neuronalen Netzen erstellt werden.

3.2.2 Das Prinzip der materiellen Objektivität

Physikalische Prozesse müssen unabhängig von der Wahl des Bezugssystems sein, welches Positionen \mathbf{x} und Zeitpunkte t festlegt. Zur Veranschaulichung kann auch von einem *Beobachter* gesprochen werden. In diesem Abschnitt werden gewisse Annahmen aus Kapitel 2 aufgehoben. Die materiellen Punkte werden über ein

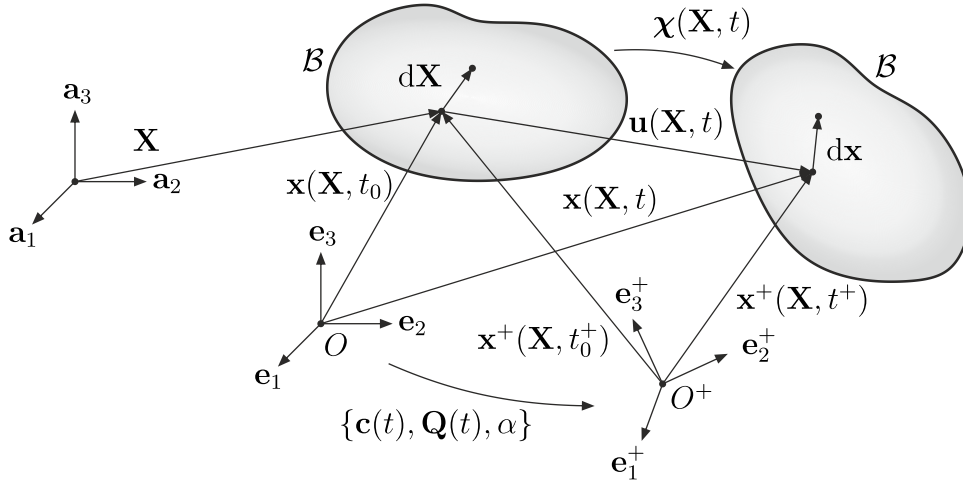


Abbildung 3.3: Beschreibung der Bewegung eines Körpers \mathcal{B} von zwei Beobachtern O und O^+ , die über eine euklidische Transformation in Zusammenhang stehen

unabhängiges kartesisches Koordinatensystem $\{\mathbf{a}_i\}_{i=1}^3$ beschrieben¹⁵, siehe Abbildung 3.3. Ohne Beschränkung der Allgemeinheit befindet sich der Beobachter O in Ruhe und beschreibt die Bewegung des Körpers \mathcal{B} mit dem kartesischen Koordinatensystem $\{\mathbf{e}_i\}_{i=1}^3$. Für den Beobachter O befindet sich der Körper zum Zeitpunkt t_0 in der Referenzkonfiguration. Ein zweiter Beobachter O^+ bewegt sich mit konstanter Geschwindigkeit (Inertialsystem) und benutzt das Koordinatensystem $\{\mathbf{e}_i^+\}_{i=1}^3$ sowie den Startzeitpunkt t_0^+ . Beide Beobachter stehen über die euklidische Transformation

$$\begin{aligned} \mathbf{e}_i^+(t) &= \mathbf{c}(t) + \mathbf{Q}(t) \mathbf{e}_i, & i = 1, 2, 3 \\ t^+ &= t + \beta \end{aligned} \quad (3.26)$$

miteinander in Zusammenhang. Dabei sind $\mathbf{c}(t)$ und β konstante Translationen in Raum und Zeit und $\mathbf{Q}(t) \in \text{SO}(3)$ beschreibt eine Rotation¹⁶. Ein räumlicher Tensor wird als materiell objektiv bezeichnet, wenn bei Wechsel des Bezugssystems sowohl Zeit- als auch Längendifferenzen erhalten bleiben. So folgt bspw. für die räumlich beschriebenen Verschiebungskomponenten

$$\begin{aligned} \mathbf{u}^+ &= \mathbf{x}^+(t^+) - \mathbf{x}^+(t_0^+) \\ &= \mathbf{c}(t) + \mathbf{Q}(t)\mathbf{x}(t) - (\mathbf{c}(t) + \mathbf{Q}(t)\mathbf{x}(t_0)) \\ &= \mathbf{Q}(t)(\mathbf{x}(t) - \mathbf{x}(t_0)) = \mathbf{Q}(t)\mathbf{u}. \end{aligned} \quad (3.27)$$

Neben $t - t_0 = t^+ - t_0^+ = \beta$ gilt mit $\mathbf{Q}(t) \in \text{SO}(3)$ auch $\|\mathbf{u}^+\| = \|\mathbf{u}\|$, womit die materielle Objektivität des räumlichen Verschiebungsvektors bestätigt ist.

¹⁵ In bspw. HOLZAPFEL [81] bspw. wird sich dazu einer dritten Konfiguration bedient, in der die materiellen Punkte definiert sind. In dieser Arbeit ist die Verwendung eines unabhängigen Koordinatensystems ausreichend und für die Fragestellung zudem übersichtlicher.

¹⁶ Durch die Bedingung $J > 0$ werden Spiegelungen hier ausgeschlossen.

Für allgemeine rein räumliche Tensoren verschiedener Stufen lassen sich analog zu den Gleichungen (3.5) und (3.6) die folgenden Bedingungen zur materiellen Objektivität ableiten.

$$\text{Tensor 0. Stufe} \quad a \quad \text{objektiv, wenn} \quad \alpha^+(t) = \alpha(t) \quad (3.28)$$

$$\text{Tensor 1. Stufe} \quad \mathbf{a} \quad \text{objektiv, wenn} \quad a_k^+(t) = Q_{ik}(t)a_i \quad (3.29)$$

$$\text{Tensor 2. Stufe} \quad \mathbf{A} \quad \text{objektiv, wenn} \quad A_{kl}^+(t) = Q_{ik}(t)Q_{jl}(t)A_{ij} \quad (3.30)$$

Der CAUCHY-Spannungstensor ist objektiv und wird mit der Vorschrift (3.30) transformiert. Dies kann über die Transformation des CAUCHY-Theorems (2.22) gezeigt werden. Der Deformationsgradient \mathbf{F} und der 1. PIOLA-KIRCHHOFF'sche Spannungstensor \mathbf{P} sind 2-Feld-Tensoren mit

$$\mathbf{F} = F_{ij} \mathbf{e}_i \otimes \mathbf{a}_j \quad \text{und} \quad \mathbf{P} = P_{ij} \mathbf{e}_i \otimes \mathbf{a}_j . \quad (3.31)$$

Für sie gelten somit die Objektivitätsbedingungen (3.29), welche auch erfüllt sind¹⁷. Der Rechts-CAUCHY-GREEN-Tensor \mathbf{C} , der GREEN-LAGRANGE'sche Verzerrungstensor \mathbf{E} und der 2. PIOLA-KIRCHHOFF'sche Spannungstensor \mathbf{S} sind unabhängig vom räumlichen Koordinatensystem $\{\mathbf{e}_i\}_{i=1}^3$. Sie sind daher wie alle rein materiellen Tensoren objektiv, und es gilt

$$\mathbf{C}^+ = \mathbf{C} , \quad \mathbf{E}^+ = \mathbf{E} \quad \text{und} \quad \mathbf{S}^+ = \mathbf{S} . \quad (3.32)$$

für die Matrizen ihrer Komponenten. Alle genannten räumlichen und materiellen Tensoren eignen sich somit zur Beschreibung objektiven Materialverhaltens.

Inelastische Materialmodelle können in Ratenform definiert werden. In diesem Fall müssen neben den Tensoren selbst auch deren Raten objektive Größen sein. So gilt bspw. für die Rate der CAUCHY-Spannungen

$$\dot{\boldsymbol{\sigma}}^+ = \frac{d}{dt}(\mathbf{Q}\boldsymbol{\sigma}\mathbf{Q}^T) = \dot{\mathbf{Q}}\boldsymbol{\sigma}\mathbf{Q}^T + \mathbf{Q}\dot{\boldsymbol{\sigma}}\mathbf{Q}^T + \mathbf{Q}\boldsymbol{\sigma}\dot{\mathbf{Q}}^T \neq \mathbf{Q}\dot{\boldsymbol{\sigma}}\mathbf{Q}^T . \quad (3.33)$$

Wie auch Geschwindigkeiten und Beschleunigungen sind die Spannungsraten im Allgemeinen nicht objektiv. Daher werden neue, objektive Spannungsraten eingeführt, in denen Materialmodelle modelliert werden müssen. So kann z. B. mithilfe des Spin-Tensors

$$\mathbf{W} = \frac{1}{2}(\dot{\mathbf{F}}\mathbf{F}^{-1} - \mathbf{F}^{-T}\dot{\mathbf{F}}^T) \quad (3.34)$$

die objektive JAUMANN'sche Spannungsrate

$$\overset{\circ}{\boldsymbol{\sigma}} := \dot{\boldsymbol{\sigma}} + \boldsymbol{\sigma}\mathbf{W} - \mathbf{W}\boldsymbol{\sigma} \quad (3.35)$$

¹⁷ Hier unterscheidet sich die allgemeine Lehrmeinung. In bspw. ALTENBACH [4] und BERTRAM [13] wird \mathbf{F} als nicht objektiv bezeichnet, in HOLZAPFEL [81] dagegen schon. Dies kann mit der im Allgemeinen nicht einheitlichen Definition materieller Objektivität zusammenhängen. Im Rahmen dieser Arbeit ist dieser Sachverhalt nur von theoretischem Interesse.

definiert werden, die sich nach Bedingung (3.30) transformiert. Neben ihr gibt es viele weitere objektive Spannungsraten, siehe bspw. HOLZAPFEL [81]. Aus der Unabhängigkeit vom räumlichen Koordinatensystem folgt für die Raten materieller Tensoren ebenfalls, dass sie objektiv sind. Für ihre Komponenten gilt

$$\dot{\mathbf{C}}^+ = \dot{\mathbf{C}} \quad , \quad \dot{\mathbf{E}}^+ = \dot{\mathbf{E}} \quad \text{und} \quad \dot{\mathbf{S}}^+ = \dot{\mathbf{S}} \quad . \quad (3.36)$$

Nicht nur die Variablen, auch die Materialfunktionen \mathcal{S} , bspw. aus Gleichung (3.1), müssen objektiv sein. Dies ist im Allgemeinen nicht dadurch gewährleistet, dass objektive Tensoren verwendet werden. Zum Beispiel wäre ein elastisches Materialmodell $\boldsymbol{\sigma} = \mathcal{S}(\mathbf{F})$ nur dann objektiv, wenn für die Komponenten $\mathbf{Q}\boldsymbol{\sigma}\mathbf{Q}^T = \mathcal{S}(\mathbf{Q}^T\mathbf{F})$ gilt. Werden \mathbf{S} und \mathbf{E} oder \mathbf{C} als Variablen verwendet, sind die Materialfunktionen automatisch objektiv, da sie sich über die Identität (3.32) transformieren. Dieser Sachverhalt wird sich im Laufe der vorliegenden Arbeit zunutze gemacht, da alle KNN-Materialmodelle mit nicht infinitesimalen Verzerrungen in \mathbf{E} und \mathbf{S} definiert werden.

3.2.3 Materielle Symmetrien

Ist das Materialverhalten richtungsunabhängig, wird es als *isotrop* bezeichnet, andernfalls als *anisotrop*. Spezielle Anisotropieklassen können durch materielle Symmetrien definiert werden. Diese müssen sich auf die undeformierte Referenzkonfiguration beziehen. Ändern sich die Materialeigenschaften nicht bei einer Transformation des materiellen Koordinatensystems $\{\mathbf{a}_i\}_{i=1}^3$ mit \mathbf{Q} , so ist das Materialverhalten bzgl. \mathbf{Q} symmetrisch. Dabei können Rotationen \mathbf{Q}^R aus Gleichung (3.8), Spiegelungen \mathbf{Q}^S aus Gleichung (3.10) oder im Allgemeinen Kombinationen \mathbf{Q} daraus verwendet werden. Spezifische Anisotropieklassen resultieren aus der Mikrostruktur des Materials. Ihr Studium steht vor allem im Zusammenhang mit den Materialeigenschaften von Kristallen, siehe bspw. NYE [137] oder VOIGT [182].

Für die allgemeinen Materialgleichungen (3.1) und (3.2) gilt folglich stets

$$\mathbf{S}^* = \mathbf{Q}^T \mathbf{S} \mathbf{Q} = \mathcal{S}(\mathbf{E}^*) = \mathcal{S}(\mathbf{Q} \mathbf{E} \mathbf{Q}^T) = \mathcal{S}(\mathbf{E}) = \mathbf{S} \quad \text{bzw.} \quad (3.37)$$

$$\mathbf{S}^* = \mathbf{Q}^T \mathbf{S} \mathbf{Q} = \mathcal{S}(\mathbf{E}^*(\tilde{t})) = \mathcal{S}(\mathbf{Q} \mathbf{E}(\tilde{t}) \mathbf{Q}^T) = \mathcal{S}(\mathbf{E}(\tilde{t})) = \mathbf{S} \quad , \quad \tilde{t} \in [0, t] \quad . \quad (3.38)$$

Zu jedem Zeitpunkt \tilde{t} resultiert dadurch für die Matrizen der Komponenten die Bedingung

$$[\mathbf{Q}]^T \mathbf{S} [\mathbf{Q}] = \mathcal{S}([\mathbf{Q}]^T \mathbf{E} [\mathbf{Q}]) \quad , \quad (3.39)$$

wobei zur Verdeutlichung die eckigen Klammern aus Abschnitt 3.1 wieder hinzugefügt wurden. Diese Eigenschaft ist äußerst praktisch, um zum Training künstlicher neuronaler Netze zusätzliche Datenpunkte zu erzeugen, auch im Fall elastoplastischen Materialverhaltens. Für Materialien mit kristalliner Struktur können

32 verschiedene Symmetrieklassen definiert werden, vgl. ALTENBACH [4]. Im Folgenden werden drei Sonderfälle genannt, auf die in dieser Arbeit Bezug genommen wird. Zwei von ihnen sind in Abbildung 3.4 skizziert.

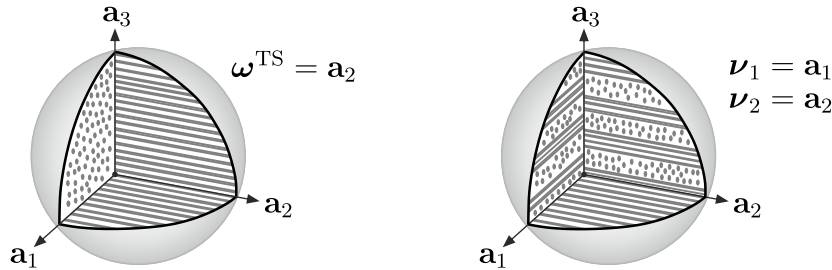


Abbildung 3.4: Illustration einer transversal isotropen Materialstruktur (links) und einer orthotropen Materialstruktur (rechts), inklusive zugehöriger Strukturvektoren

Orthotropie (OS) liegt vor, wenn sich bei Spiegelungen bezüglich zweier orthogonaler Ebenen das Materialverhalten nicht ändert. Die zugehörigen Transformationen sind

$$\begin{aligned} \mathbf{Q}^{\text{OS}} &= \mathbf{Q}_2^S \mathbf{Q}_1^S = (\mathbf{I} - 2\nu_2 \nu_2^T)(\mathbf{I} - 2\nu_1 \nu_1^T) \\ &= \mathbf{I} - 2\nu_2 \nu_2^T - 2\nu_1 \nu_1^T, \end{aligned} \quad (3.40)$$

mit den normierten Normalenvektoren zu den beiden Ebenen ν_1 und ν_2 unter Berücksichtigung von $\nu_2^T \nu_1 = 0$. Die dritte zu den beiden Ebenen orthogonale Ebene ist automatisch auch eine Symmetrieebene.

Transversale Isotropie (TS) liegt vor, wenn sich bei einer beliebigen Rotation um eine materielle Achse ω^{TS} das Materialverhalten nicht ändert. Die zugehörige Menge aller Transformationen ist

$$\mathbf{Q}^{\text{TS}} = \mathbf{Q}^R(\omega, \theta) \in \{\text{SO}(3) | \omega = \omega^{\text{TS}}, \theta \in (0, 2\pi]\}, \quad (3.41)$$

mit den Rotationsmatrizen \mathbf{Q}^R aus Definition (3.8).

Isotropie (IS) liegt vor, wenn sich bei einer beliebigen Transformation mit $\mathbf{Q}^{\text{IS}} \in \text{O}(3)$ das Materialverhalten nicht ändert. Dabei wird sich aus physikalischer Sicht meist auf die spezielle orthogonale Gruppe mit $\mathbf{Q}^{\text{IS}} \in \text{SO}(3)$ beschränkt.

3.2.4 Inkompressibilität als kinematische Zwangsbedingung

Einige Materialien verformen sich näherungsweise volumenerhaltend. Im Rahmen der Kontinuumsmechanik kann die kinematische Zwangsbedingung $J - 1 = 0$ eingefordert werden. Sie lässt sich mit $\det \mathbf{C} = J^2$ auch über

$$\det \mathbf{C} - 1 = 0 \quad (3.42)$$

darstellen. Mit dem Zusammenhang zum GREEN-LAGRANGE'schen Verzerrungstensor $\mathbf{C} = 2\mathbf{E} + \mathbf{I}$ und der Identität

$$\det(\mathbf{A} - \mathbf{I}) = \det \mathbf{A} + \text{Sp} \mathbf{A} - \frac{1}{2}(\text{Sp}(\mathbf{A})^2 - \text{Sp}(\mathbf{A}^2)) - 1 \quad (3.43)$$

kann die folgende Form der Inkompressibilitätsbedingung hergeleitet werden.

$$4 \det \mathbf{E} + \text{Sp} \mathbf{E} + \text{Sp}(\mathbf{E})^2 - \text{Sp}(\mathbf{E}^2) = 0 \quad (3.44)$$

Im Fall infinitesimaler Verzerrungen $\mathbf{E} \approx \boldsymbol{\varepsilon}$ gilt die Bedingung

$$\text{Sp} \mathbf{E} \approx \text{Sp} \boldsymbol{\varepsilon} = \varepsilon_{11} + \varepsilon_{22} + \varepsilon_{33} = 0, \quad (3.45)$$

wobei die nichtlinearen Terme in Gleichung (3.44) vernachlässigt werden.

3.3 Klassische elastische Materialmodelle

Elastische Materialmodelle zeichnen sich unter anderem dadurch aus, dass bei vollständiger Entlastung und fehlendem äußeren Zwang keine Deformation zurück bleibt, vgl. Abbildung 3.1. Ferner gibt es spezifischere Definitionen von Elastizität: CAUCHY-elastisches Material ist durch eine Spannungsfunktion $\mathcal{S}(\mathbf{E})$ nach Definition (3.1) charakterisiert, während GREEN-elastische bzw. *hyperelastische* Materialien darüber hinaus die Existenz einer Verzerrungsenergiedichtefunktion Ψ verlangen. In diesem Abschnitt werden letztgenannte hyperelastische Materialmodelle näher beschrieben. Eine Übersicht über solche Modelle ist darüber hinaus z. B. in BERTRAM [13], HOLZAPFEL [81] oder STEINMANN ET AL. [163] zu finden.

3.3.1 Isotrope hyperelastische Materialmodelle

Hyperelastische Materialmodelle sind durch die Existenz einer Verzerrungsenergiedichtefunktion Ψ definiert. Diese wird in der vorliegenden Arbeit in Abhängigkeit des Rechts-CAUCHY-GREEN-Tensors \mathbf{C} formuliert. Mit Gleichung (3.25) für elastische Materialien ($\dot{\mathcal{D}} = 0$) und der Kettenregel folgt die Dissipationsgleichung

$$\dot{\mathcal{D}} = \left(\mathbf{S} - \frac{\partial \Psi}{\partial \mathbf{C}} \frac{\partial \mathbf{C}}{\partial \mathbf{E}} \right) \dot{\mathbf{E}} = 0. \quad (3.46)$$

Aus der Umformung $\mathbf{C} = 2\mathbf{E} + \mathbf{I}$ resultieren schließlich die Bedingungen

$$\mathbf{S} = \mathcal{S}(\mathbf{E}) = \frac{\partial \Psi(\mathbf{C})}{\partial \mathbf{E}} = 2 \frac{\partial \Psi(\mathbf{C})}{\partial \mathbf{C}} \quad \text{und} \quad (3.47)$$

$$\mathbf{C}_T = \frac{\partial^2 \Psi(\mathbf{C})}{\partial \mathbf{E}^2} = 4 \frac{\partial^2 \Psi(\mathbf{C})}{\partial \mathbf{C}^2} \quad (3.48)$$

für die Spannungs- und Elastizitätstensoren. In der Regel werden an die Verzerrungsenergiedichtefunktion weitere Anforderungen gestellt:

$$\text{Normalisierung} \quad \Psi(\mathbf{I}) = 0 \quad (3.49)$$

$$\text{Positivität} \quad \Psi(\mathbf{C}) > 0, \quad \forall \mathbf{C} \neq \mathbf{I} \quad (3.50)$$

$$\text{Grenzzustände} \quad \lim_{J \rightarrow \infty} \Psi \rightarrow +\infty \quad (3.51)$$

$$\lim_{J \rightarrow 0} \Psi \rightarrow +\infty. \quad (3.52)$$

Die Bedingungen (3.49) und (3.50) garantieren durch das globale Minimum bei $\mathbf{C} = \mathbf{I}$ eine spannungsfreie Referenzkonfiguration¹⁸. Die Wachstumsbedingungen (3.51) und (3.52) garantieren realistisches Grenzverhalten bei unendlicher Expansion und Kompression.

Im Fall isotroper hyperelastischer Materialien kann die Verzerrungsenergiedichtefunktion Ψ in den Invarianten $I_{\mathbf{C}}$, $II_{\mathbf{C}}$ und $III_{\mathbf{C}}$ oder in Eigenwerten von \mathbf{C} definiert werden. Aus $\mathbf{C} = \mathbf{F}^T \mathbf{F}$ folgt, dass diese die Quadrate der Eigenwerte von \mathbf{F} sind, welche als Hauptstreckungen $\{\lambda_a\}_{a=1}^3$ bezeichnet werden. Die Eigenwerte von \mathbf{C} werden somit über das spezielle Eigenwertproblem

$$(\mathbf{C} - \lambda^2 \mathbf{I}) \mathbf{n} = \mathbf{0} \quad (3.53)$$

berechnet. Im Fall isotroper Hyperelastizität sind \mathbf{C} und \mathbf{S} koaxial, ihre Eigenvektoren $\{\mathbf{n}_a\}_{a=1}^3$ sind gleich¹⁹. Demzufolge werden über das Eigenwertproblem (3.53) auch die Eigenvektoren von \mathbf{S} berechnet. Auf Basis einer Verzerrungsenergiedichtefunktion $\Psi(\lambda_1, \lambda_2, \lambda_3)$ werden zunächst die Spannungen

$$\mathbf{S} = 2 \frac{\partial \Psi}{\partial \mathbf{C}} = 2 \sum_{a=1}^3 \frac{\partial \Psi}{\partial \lambda_a} \frac{\partial \lambda_a}{\partial \mathbf{C}} = \sum_{a=1}^3 \underbrace{\frac{1}{\lambda_a} \frac{\partial \Psi}{\partial \lambda_a}}_{S_a} \mathbf{n}_a \otimes \mathbf{n}_a \quad (3.54)$$

in spektraler Form berechnet, wobei die Ableitungsvorschrift²⁰

$$\frac{\partial \lambda_a}{\partial \mathbf{C}} = \frac{1}{2\lambda_a} \mathbf{n}_a \otimes \mathbf{n}_a. \quad (3.55)$$

genutzt wurde, vgl. SIMO und TAYLOR [161] oder Gleichung (3.13) mit $\partial \lambda_a^2 / \partial \lambda_a = 2\lambda_a$. Die zugehörige spektrale Form des Elastizitätstensors lautet

$$\begin{aligned} \mathbb{C} = & \sum_{a=1}^3 \sum_{b=1}^3 \hat{C}_{aabb} \mathbf{n}_a \otimes \mathbf{n}_a \otimes \mathbf{n}_b \otimes \mathbf{n}_b \\ & + \sum_{a=1}^3 \sum_{\substack{b=1 \\ b \neq a}}^3 \hat{C}_{abab} \left(\mathbf{n}_a \otimes \mathbf{n}_b \otimes \mathbf{n}_a \otimes \mathbf{n}_b + \mathbf{n}_a \otimes \mathbf{n}_b \otimes \mathbf{n}_b \otimes \mathbf{n}_a \right), \end{aligned} \quad (3.56)$$

¹⁸ Bei der Modellierung von granularen Materialien wäre das nicht korrekt, da deren Steifigkeit erst durch Reibung infolge Druck aktiviert wird.

¹⁹ Das kann über das RIVLIN-ERICKSEN-Theorem gezeigt werden, siehe HOLZAPFEL [81].

²⁰ Das gilt nur für \mathbf{C} , da λ_a die Wurzel der Eigenwerte von \mathbf{C} ist.

mit den Komponenten

$$\hat{C}_{aabb} = \frac{1}{\lambda_b} \frac{\partial S_a}{\partial \lambda_b} = 2 \frac{\partial S_a}{\partial \lambda_b^2} \quad (3.57)$$

$$\hat{C}_{abab} = \frac{S_b - S_a}{\lambda_b^2 - \lambda_a^2} \xrightarrow{\lambda_a = \lambda_b} \frac{1}{2\lambda_b} \left(\frac{\partial S_b}{\partial \lambda_b} - \frac{\partial S_a}{\partial \lambda_b} \right) = \frac{\partial S_b}{\partial \lambda_b^2} - \frac{\partial S_a}{\partial \lambda_b^2}. \quad (3.58)$$

Der Beweis dieser allgemeinen Form ist in HOLZAPFEL [81] zu finden. Der Grenzfall $\lambda_a \rightarrow \lambda_b$ muss für die Komponente (3.58) mit dem Satz von DE L'HOSPITAL abgefangen werden. Nach Berechnung der spektralen Komponenten S_a , \hat{C}_{aabb} und \hat{C}_{abab} werden diese in die kartesischen Basen $\{\mathbf{e}_i\}_{i=1}^3$ mit $\mathbf{S} = S_{ij} \mathbf{e}_i \otimes \mathbf{e}_j$ und $\mathbb{C} = C_{ijkl} \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k \otimes \mathbf{e}_l$ zurück transformiert. Die Komponenten des Transformationstensors \mathbf{Q} sind $Q_{ai} = \mathbf{n}_a \cdot \mathbf{e}_i$, mit denen die Transformationsgleichungen

$$S_{ij} = \sum_{a=1}^3 S_a Q_{ai} Q_{aj} \quad (3.59)$$

sowie

$$\begin{aligned} C_{ijkl} = & \sum_{a=1}^3 \sum_{b=1}^3 \hat{C}_{aabb} Q_{ai} Q_{aj} Q_{bk} Q_{bl} \\ & + \sum_{a=1}^3 \sum_{\substack{b=1 \\ b \neq a}}^3 \hat{C}_{abab} (Q_{ai} Q_{bj} Q_{ak} Q_{bl} + Q_{ai} Q_{bj} Q_{bk} Q_{al}) \end{aligned} \quad (3.60)$$

folgen. Mit den Basisvektoren $\mathbf{e}_1 = [1, 0, 0]^T$, $\mathbf{e}_2 = [0, 1, 0]^T$ und $\mathbf{e}_3 = [0, 0, 1]^T$ lassen sich die folgenden kompakten Transformationsbeziehungen in reduzierter Notation angeben, vgl. KLINKEL [104] und WRIGGERS [198]. Die spektralen Komponenten werden in den Vektor

$$\hat{\mathbf{S}} := \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix} \quad (3.61)$$

und die Matrizen

$$\hat{\mathbf{C}}_1 := \begin{bmatrix} \hat{C}_{1111} & \hat{C}_{1122} & \hat{C}_{1133} \\ \hat{C}_{2211} & \hat{C}_{2222} & \hat{C}_{2233} \\ \hat{C}_{3311} & \hat{C}_{3322} & \hat{C}_{3333} \end{bmatrix} \quad \hat{\mathbf{C}}_2 := \begin{bmatrix} \hat{C}_{1212} & 0 & 0 \\ 0 & \hat{C}_{1313} & 0 \\ 0 & 0 & \hat{C}_{2323} \end{bmatrix} \quad (3.62)$$

zusammengefasst. Mit den Transformationsmatrizen

$$\hat{\mathbf{Q}}_1 := \begin{bmatrix} n_{11}^2 & n_{21}^2 & n_{31}^2 \\ n_{12}^2 & n_{22}^2 & n_{32}^2 \\ n_{13}^2 & n_{23}^2 & n_{33}^2 \\ n_{11}n_{12} & n_{21}n_{22} & n_{31}n_{32} \\ n_{11}n_{13} & n_{21}n_{23} & n_{31}n_{33} \\ n_{12}n_{13} & n_{22}n_{23} & n_{32}n_{33} \end{bmatrix} \quad (3.63)$$

und

$$\hat{\mathbf{Q}}_2 := \begin{bmatrix} 2n_{11}n_{21} & 2n_{11}n_{31} & 2n_{21}n_{31} \\ 2n_{12}n_{22} & 2n_{12}n_{32} & 2n_{22}n_{32} \\ 2n_{13}n_{23} & 2n_{13}n_{33} & 2n_{23}n_{33} \\ n_{11}n_{22} + n_{21}n_{12} & n_{11}n_{32} + n_{31}n_{12} & n_{21}n_{32} + n_{31}n_{22} \\ n_{11}n_{23} + n_{21}n_{13} & n_{11}n_{33} + n_{31}n_{13} & n_{21}n_{33} + n_{31}n_{23} \\ n_{12}n_{23} + n_{22}n_{13} & n_{12}n_{33} + n_{32}n_{13} & n_{22}n_{33} + n_{32}n_{23} \end{bmatrix} \quad (3.64)$$

werden die spektralen Matrizen in die kartesischen über

$$\mathbf{S} = \hat{\mathbf{Q}}_1 \hat{\mathbf{S}} \quad \text{und} \quad \mathbf{C}_T = \hat{\mathbf{Q}}_1 \hat{\mathbf{C}}_1 \hat{\mathbf{Q}}_1^T + \hat{\mathbf{Q}}_2 \hat{\mathbf{C}}_2 \hat{\mathbf{Q}}_2^T \quad (3.65)$$

umgeformt.

3.3.2 Lineares Green'sches Materialgesetz

Das verallgemeinerte lineare HOOKE'sche Gesetz $\boldsymbol{\sigma} = \mathbb{C}\boldsymbol{\varepsilon}$ kann für den geometrisch nichtlinearen Fall erweitert werden zu

$$\mathbf{S} = \mathbb{C}\mathbf{E}, \quad \text{bzw.} \quad S_{ij} = C_{ijkl}E_{kl}, \quad (3.66)$$

mit dem konstanten Elastizitätstensor 4. Stufe \mathbb{C} . Per Annahme existiert die zugehörige Verzerrungsenergiedichtefunktion

$$\Psi = \frac{1}{8}(\mathbf{C} - \mathbf{I}) : (\mathbb{C}(\mathbf{C} - \mathbf{I})) = \frac{1}{2}\mathbf{E} : \mathbb{C}\mathbf{E}. \quad (3.67)$$

Da die Kompressionsbedingung (3.52) mit $\Psi(\mathbf{C} = \mathbf{0}) < \infty$ verletzt wird, ist ersichtlich, wieso dieses Materialmodell für große Verzerrungen nicht geeignet ist.

Die im Allgemeinen 36 unabhängigen Komponenten $\{C_{ijkl}\}_{i,j,k,l=1}^3$ reduzieren sich durch die Existenz der Verzerrungsenergiedichtefunktion $\Psi(\mathbf{C})$ auf 21. Für die Komponenten des Elastizitätstensors gilt $C_{ijkl} = C_{klij}$. Daraus folgt ebenfalls die Symmetrie der reduzierten Matrix $\mathbf{C}_T = \mathbf{C}_T^T$ aus Definition (2.57). Durch die Berücksichtigung von materiellen Symmetrien nach Abschnitt 3.2.3 wird im Folgenden die Anzahl der unabhängigen Komponenten weiter reduziert. Mit der Transformationsbeziehung (3.37) folgt die Bedingung

$$C_{mnop}^* = Q_{im}Q_{jn}Q_{ko}Q_{lp}C_{ijkl} \stackrel{!}{=} C_{ijkl} \quad (3.68)$$

für die Komponenten des Elastizitätstensors. Durch Vorgabe spezifischer Transformationstensoren \mathbf{Q} nach Abschnitt 3.2.3 lassen sich somit zusätzliche Abhängigkeiten herleiten. Im Fall von Orthotropie aus Gleichung (3.40) gilt mit $\boldsymbol{\nu}_1 = \mathbf{e}_1$ und $\boldsymbol{\nu}_2 = \mathbf{e}_2$ für $Q_{11} = Q_{22} = -1$, $Q_{33} = 1$ und für alle weiteren Komponenten $Q_{ij} = 0$. Damit folgt unter anderem die Bedingung $-C_{1113} = C_{1113}$, was

nur für $C_{1113} = 0$ im Allgemeinen erfüllt ist. Die vollständige Inverse²¹ \mathbb{C}^{-1} für Orthotropie in reduzierter Notation nach Definition (2.57) ist

$$\mathbf{C}_{OS}^{-1} = \begin{bmatrix} 1/E_1 & -\nu_{12}/E_1 & -\nu_{13}/E_1 & 0 & 0 & 0 \\ -\nu_{12}/E_1 & 1/E_2 & -\nu_{23}/E_2 & 0 & 0 & 0 \\ -\nu_{13}/E_1 & -\nu_{23}/E_2 & 1/E_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/\mu_{12} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/\mu_{13} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/\mu_{23} \end{bmatrix}, \quad (3.69)$$

siehe auch ALTENBACH ET AL. [5]. Sie gilt nur, wenn die Materialvorzugsrichtungen mit den Koordinatenachsen \mathbf{e}_i zusammenfallen und Energieerhaltung vorausgesetzt wird, sonst erhöht sich die Anzahl der unbekannt Parameter. Ein Beispiel dazu wird in Abschnitt 7.2 gezeigt. Hier, im Fall mit Energieerhaltung, werden die unabhängigen Koeffizienten mit drei Elastizitätsmoduln E_1, E_2, E_3 , drei Schubmoduln $\mu_{12}, \mu_{13}, \mu_{23}$ und drei Querdehnzahlen $\nu_{12}, \nu_{13}, \nu_{23}$ parametrisiert²². Unter anderem folgen aus der Nicht-Singularität von \mathbf{C} und \mathbf{C}^{-1} die Stabilitätsbedingungen in den angegebenen Ingenieurkonstanten:

$$\begin{aligned} E_i > 0, \quad \mu_{ij} > 0, \quad |\nu_{ij}| < \sqrt{E_i/E_j}, \\ 1 - \nu_{12}^2 \frac{E_2}{E_1} - \nu_{23}^2 \frac{E_3}{E_2} - \nu_{13}^2 \frac{E_3}{E_1} - 2\nu_{12}\nu_{13}\nu_{23} \frac{E_3}{E_1} > 0. \end{aligned} \quad (3.70)$$

Für den Fall der Transversalisotropie, um bspw. die \mathbf{e}_3 -Achse, reduzieren sich die obigen Angaben durch die Bedingungen

$$E_1 = E_2, \quad \nu_{13} = \nu_{23}, \quad \mu_{13} = \mu_{23} \quad \text{und} \quad \mu_{12} = \frac{E_1}{2(1 + \nu_{12})}. \quad (3.71)$$

Somit verbleiben fünf unabhängige Ingenieurkonstanten $E_1, E_3, \nu_{12}, \nu_{13}$ und μ_{13} . Bei vollständiger Isotropie reduzieren sich diese weiter mit

$$\begin{aligned} E_1 = E_2 = E_3 = E, \quad \nu_{12} = \nu_{13} = \nu_{23} = \nu \quad \text{und} \\ \mu_{12} = \mu_{13} = \mu_{23} = \mu = \frac{E}{2(1 + \nu)} \end{aligned} \quad (3.72)$$

auf zwei unabhängige Konstanten $E > 0$ und $-1 < \nu < 0,5$. Dieser Fall wird auch als ST. VENANT-KIRCHHOFF-Materialmodell bezeichnet. Die zugehörige Verzerrungsenergiefunktion lässt sich in diesem Fall mit

$$\Psi(\mathbf{E}) = \frac{\Lambda}{2} (\text{Sp}(\mathbf{E}))^2 + \mu \text{Sp}(\mathbf{E}^2) \quad (3.73)$$

²¹ Die Darstellung der Matrix \mathbf{C}_{OS} von \mathbb{C} ist unübersichtlicher.

²² Diese Ingenieurkonstanten lassen sich in Grundversuchen direkt ermitteln.

schreiben, wobei der erste Summand die volumetrischen und der zweite die deviatorischen Anteile enthält. Die LAMÉ'schen Konstanten Λ und μ (Schubmodul) können über

$$\Lambda = \frac{E\nu}{(1+\nu)(1-2\nu)} \quad \text{und} \quad \mu = \frac{E}{2(1+\nu)} \quad (3.74)$$

aus den isotropen Ingenieurkonstanten E und ν berechnet werden. Diese konstitutive Gleichung lässt sich ebenfalls dadurch motivieren, dass für sehr kleine Verzerrungen jede isotrope Verzerrungsenergiedichtefunktion durch

$$\Psi(\mathbf{E}) = \frac{\Lambda}{2}(\text{Sp}(\mathbf{E}))^2 + \mu \text{Sp}(\mathbf{E}^2) + \mathcal{O}(\|\mathbf{E}\|^2) \quad (3.75)$$

dargestellt werden kann, siehe CIARLET [26]. Somit stellt das ST. VENANT-KIRCHHOFF-Materialmodell eine konsistente Linearisierung für kleine Verzerrungen ($\|\mathbf{E}\|^2 \approx 0$) dar.

Aus diesem Grund, und unter Berücksichtigung der Anwendungsgrenzen bei großen Druckspannungen, wird das lineare GREEN'sche Material in der Praxis vor allem im Bereich großer Verschiebungen und kleiner Verzerrungen angewendet. Dies ist bspw. bei Schalenstrukturen der Fall, die durch ihre geringe Dicke trotz großer Verdrehungen kleine Verzerrungen tangential zur Fläche aufweisen. So wird dieses Materialmodell in WAGNER und GRUTTMANN [184] zur Berechnung geschichteter Schalenstrukturen genutzt, unter anderem zur dynamischen Simulation einer Propfan-Schaukel. Bei der Verwendung dieses Materialmodells ist stets Vorsicht bei größeren Druckspannungen geboten, da dort ggf. materielle Instabilitäten auftreten können.

3.3.3 Ogden Materialgesetz

Die um einem kompressiblen Anteil erweiterte Verzerrungsenergiedichtefunktion

$$\Psi(\mathbf{C}) = \sum_{r=1}^m \left[\frac{\mu_r}{\alpha_r} (\lambda_1^{\alpha_r} + \lambda_2^{\alpha_r} + \lambda_3^{\alpha_r} - 3) \right] + \Psi_{\text{vol}}(J) \quad (3.76)$$

geht auf zwei Arbeiten von OGDEN [139, 140] zurück, mit $J = \lambda_1 \lambda_2 \lambda_3$. Sie ist für große Verzerrungen und isotropes Materialverhalten gültig. Für den kompressiblen Anteil wird hier die Funktion

$$\Psi_{\text{vol}}(J) = \frac{\Lambda}{4} (J^2 - 1 - \ln(J^2)) - \sum_{r=1}^m \mu_r \ln(J) \quad (3.77)$$

gewählt, vgl. SIMO [158] oder auch KLINKEL [104] und REESE [147]. Eine Übersicht über verschiedene volumetrische Ergänzungen ist in DOLL und SCHWEIZERHOF [36] zu finden.

Für die Materialparameter $\{\mu\}_{r=1}^m$ und $\{\alpha\}_{r=1}^m$ gelten die Stabilitätsbedingungen

$$\mu_r \alpha_r > 0, \quad \forall r = 1, \dots, m. \quad (3.78)$$

Das Materialmodell muss für infinitesimale Verzerrungen mit der linearen Theorie konsistent sein. Aus einer Grenzbetrachtung für $\lambda_a \rightarrow 1$ resultiert der Zusammenhang

$$\sum_{r=1}^m \mu_r \alpha_r = 2\mu, \quad (3.79)$$

mit dem isotropen Schubmodul der linearen Theorie μ .

Mit $\partial J / \partial \lambda_a = J / \lambda_a$ und Gleichung (3.54) lassen sich die Hauptspannungen über

$$S_a = \frac{1}{\lambda_a^2} \left[\sum_{r=1}^m \mu_r (\lambda_a^{\alpha_r} - 1) + \frac{\Lambda}{2} (J^2 - 1) \right] \quad (3.80)$$

berechnen. Für die spektralen Komponenten \hat{C}_{aabb} des Elastizitätstensors nach Gleichung (3.57) müssen zwei Fälle $a = b$ und $a \neq b$ unterschieden werden:

$$\hat{C}_{aabb} = \frac{1}{\lambda_b} \frac{\partial S_a}{\partial \lambda_b} = \begin{cases} \frac{1}{\lambda_a^4} \left[\sum_{r=1}^m (\mu_r \lambda_a^{\alpha_r} (\alpha_r - 2) + 2\mu_r) + \Lambda \right], & a = b \\ \Lambda \frac{J^2}{\lambda_a^2 \lambda_b^2}, & a \neq b \end{cases}. \quad (3.81)$$

Für die Komponenten \hat{C}_{abab} gilt wiederum

$$\hat{C}_{abab} = \frac{S_b - S_a}{\lambda_b^2 - \lambda_a^2} \xrightarrow{\lambda_a = \lambda_b} \frac{1}{2\lambda_a^4} \left[\sum_{r=1}^m (\mu_r \lambda_a^{\alpha_r} (\alpha_r - 2) + 2\mu_r) + \Lambda(1 - J^2) \right], \quad (3.82)$$

wobei der Grenzfall $\lambda_a = \lambda_b$ mit dem Satz von DE L'HOSPITAL abgefangen wurde. Mit diesen Ergebnissen und den Transformationsbeziehungen (3.65) lassen sich die Spannungen \mathbf{S} und die Steifigkeiten \mathbf{C}_T in reduzierter Notation aufstellen. An dieser Stelle sei erwähnt, dass die materiellen Symmetriebedingungen (3.68) des ST.VENANT-KIRCHHOFF-Materials hier nicht gültig sind. Die Matrix \mathbf{C}_T kann auch bei Isotropie voll besetzt sein.

Dieses Materialmodell kann unter anderem durch die Erfüllung der Wachstumsbedingungen (3.51) und (3.52) zur Berücksichtigung großer Verzerrungen genutzt werden. Somit eignet es sich zur Modellierung von Elastomeren oder auch biologischem Gewebe, siehe HOLZAPFEL ET AL. [82] oder WEX ET AL. [193]. Das OGDEN-Materialmodell beinhaltet durch seine Parametrisierung auch andere, einfachere Modelle. Das Neo-HOOKE-Materialgesetz nach TRELOAR [171] folgt mit $m = 1$ und $\alpha_1 = 2$. Das Materialmodell nach MOONEY [133] und RIVLIN [148] folgt mit $m = 2$, $\alpha_1 = 2$ und $\alpha_2 = -2$. Generell gilt: Je mehr Reihenglieder die Verzerrungsenergiedichtefunktion (3.76) enthält, desto komplexeres Materialverhalten kann modelliert werden.

3.3.4 Grenzfall der Inkompressibilität

Zur Modellierung inkompressiblen Materialverhaltens kann die in Abschnitt 3.2.4 definierte Zwangsbedingung $J - 1 = 0$ durch die Zerlegung der Verzerrungsenergiegedichtefunktion

$$\Psi(p, \mathbf{C}) = \Psi_{\text{dev}}(\mathbf{C}) - p(J - 1) \quad (3.83)$$

exakt eingefordert werden. Einzig der gestaltändernde Anteil Ψ_{dev} hängt von den Verzerrungsgrößen \mathbf{C} ab und entspricht somit dem Materialgesetz. Der hydrostatische Druck p ist eine zusätzliche – vom Material unabhängige – Unbekannte, welche durch die äußeren Randbedingungen²³ bestimmt werden muss. Somit fungiert der Druck p im Rahmen der globalen Optimierung als LAGRANGE-Multiplikator für die Inkompressibilitäts-Nebenbedingung. Durch Differenzierung der Verzerrungsenergiegedichtefunktion (3.83) und mit dem Volumenverhältnis $J = 1$ resultiert die Aufteilung der 2. PIOLA-KIRCHHOFF'schen Spannungen zu

$$\mathbf{S} = -p\mathbf{C}^{-1} + \mathbf{S}_{\text{dev}}(\mathbf{E}) . \quad (3.84)$$

Dabei ist zu beachten, dass $p \approx -\text{Sp}(\mathbf{S})/3$ nur für kleine Verzerrungen als Näherung gültig ist. Im Fall des linearen und isotropen ST.VENANT-KIRCHHOFF-Materialmodells aus Abschnitt 3.3.2 gilt $\mathbf{S}_{\text{dev}} = 2\mu\mathbf{E}$. Für das OGDEN-Materialmodell aus Abschnitt 3.3.3 muss in Gleichung (3.76) der volumetrische Anteil Ψ_{vol} zu null gesetzt werden, wodurch $S_a = \sum_{r=1}^m \mu_r \lambda_a^{\alpha_r - 2}$ für die deviatorischen Anteile der Hauptspannungen folgt.

Im Rahmen einer FE-Berechnung ist $p(\mathbf{X})$ neben den Verschiebungen $\mathbf{u}(\mathbf{X})$ eine zusätzliche unabhängige Feldgröße, welche mit geeigneten Ansätzen diskretisiert werden muss, vgl. SIMO ET AL. [162].

Die Inkompressibilitätsbedingung kann auch näherungsweise erfasst werden, was auch als *Quasi-Inkompressibilität* bezeichnet wird. Für beide erwähnten Materialgesetze fungiert der 1. LAMÉ-Parameter Λ dabei als Penalty-Parameter. Im linearen Fall wird die Nebenbedingung $\text{Sp}\mathbf{E} = 0$ schwach eingefordert. Im Fall des hier beschriebenen OGDEN-Materialgesetzes ist die schwach eingeforderte Nebenbedingung $J^2 - 1 - \ln(J^2) = 0$, vgl. Gleichung (3.77). Je größer Λ gewählt wird, desto schärfer wird die Inkompressibilität erzwungen. Im linearen Fall konvergiert die Querdehnzahl entsprechend $\nu \rightarrow 0,5$, vgl. Gleichung (3.74)₁. Für $\Lambda \rightarrow \infty$ werden die Materialtangente \mathbf{C}_T jeweils singulär. Dementsprechend ist eine numerisch zweckmäßige Wahl von Λ notwendig.

²³ Beispiel: biaxialer Zugversuch in 1- und 2-Richtung. Durch die Bedingung $S_3 = 0$ lässt sich der Druck p berechnen und in die Gleichungen für S_1 und S_2 einsetzen.

3.3.5 Polykonvexität

Bisher wurden ausschließlich physikalisch motivierte Bedingungen an die Verzerrungsenergie-dichtefunktion $\Psi(\mathbf{C})$ gestellt. Im Hinblick auf die materielle Stabilität und die eindeutige Lösbarkeit des zugehörigen Randwertproblems sind weitere, rein mathematische Forderungen zu stellen. HILL [78] hat im Jahr 1957 gezeigt, dass die Konvexität von Ψ bzgl. des Verschiebungsgradienten \mathbf{H} , bzw. bzgl. des Deformationsgradienten $\mathbf{F} = \mathbf{H} + \mathbf{I}$, zu eben diesen gewünschten Eigenschaften führt. Diese Konvexitäts-Bedingung schränkt allerdings die Menge der möglichen Funktionen Ψ so stark ein, dass einige der zuvor eingeführten physikalischen Eigenschaften nicht mehr erfüllt werden können. Bspw. steht die Konvexität von $\Psi(\mathbf{F})$ im Widerspruch mit der materiellen Objektivität, vgl. COLEMAN und NOLL [28]. Im Jahr 1976 hat BALL [10] die schwächere Forderung nach *Polykonvexität* eingeführt, welche ebenfalls die Lösbarkeit des Randwertproblems impliziert ohne dabei im Widerspruch zu physikalischen Eigenschaften zu stehen²⁴.

Eine Funktion $\Psi(\mathbf{F})$ ist *polykonvex*, wenn eine Funktion

$$\hat{\Psi}(\mathbf{F}, \text{cof } \mathbf{F}, J) = \Psi(\mathbf{F}) \quad (3.85)$$

existiert, die jeweils konvex in \mathbf{F} , $\text{cof } \mathbf{F}$ und $\det \mathbf{F}$ ist²⁵. Der Kofaktor-Tensor $\text{cof } \mathbf{F}$ ist für $J > 0$ gleich $J\mathbf{F}^{-T}$. Kann diese Bedingung für eine Verzerrungsenergie-dichtefunktion $\Psi(\mathbf{F})$ gezeigt werden, ist die Lösbarkeit des zugehörigen Randwertproblems der nichtlinearen Elastizitätstheorie garantiert und das Material ist stabil²⁶.

Um vorhandene analytische Funktionen $\Psi(\mathbf{F})$ – und im späteren Verlauf der Arbeit auch KNN-Materialmodelle – auf Polykonvexität zu untersuchen, werden hier zunächst allgemeine Eigenschaften konvexer Funktionen beschrieben. Eine Funktion $f(\mathbf{x})$ ist konvex, wenn die Bedingung

$$f(\alpha \mathbf{x}_1 + (1 - \alpha) \mathbf{x}_2) \leq \alpha f(\mathbf{x}_1) + (1 - \alpha) f(\mathbf{x}_2) \quad (3.86)$$

für alle Eingangsvariablen $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^{n_i}$ und $\alpha \in [0, 1]$ gilt. Dies ist in Abbildung 3.5 am Beispiel einer eindimensionalen Funktion grafisch dargestellt. Für konvexe Funktionen lassen sich u. a. die zwei nachfolgenden Zusammenhänge angeben.

²⁴ Eine Funktion $f(g(x))$, die nicht konvex bzgl. x ist, kann trotzdem konvex bzgl. g sein. Somit wäre die Forderung der Konvexität bzgl. g schwächer als die bzgl. x .

²⁵ Diese drei Argumente stehen für die Transformation von Linien-, Flächen- und Volumenelementen von der Referenz- in die Momentankonfiguration, vgl. Abschnitt 2.1.1

²⁶ Zur Definition materieller Stabilität kann bspw. das Kriterium von HADAMARD [68] herangezogen werden, siehe REESE [147]. Dessen mathematische Folge ist die Elliptizität der partiellen Differentialgleichung des Randwertproblems. Die physikalische Folge ist, dass sich Beschleunigungswellen innerhalb des hyperelastischen Mediums mit reeller Geschwindigkeit ausbreiten.

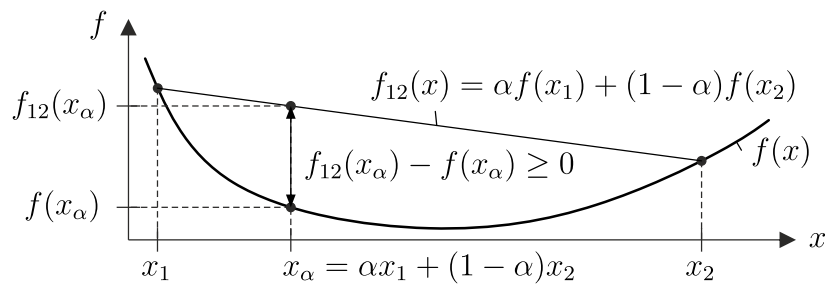


Abbildung 3.5: Darstellung der Konvexitätsbedingung am Beispiel einer eindimensionalen Funktion $f(x)$

- Sind $f_1(\mathbf{x})$ und $f_2(\mathbf{x})$ konvexe Funktionen, dann ist auch die Linearkombination $a f_1(\mathbf{x}) + b f_2(\mathbf{x})$ mit positiven Faktoren $a, b \geq 0$ konvex.
- Ist $f_1(f_2)$ konvex und monoton wachsend bzgl. f_2 sowie $f_2(\mathbf{x})$ konvex bzgl. \mathbf{x} , so ist die Verkettung $f_1(f_2(\mathbf{x}))$ ebenfalls konvex bzgl. \mathbf{x} .

Darüber hinaus ist die HESSE-Matrix $\partial^2 f / \partial \mathbf{x}^2$ einer konvexen Funktion $f(\mathbf{x})$ positiv definit.

Die Invarianten des Rechts-CAUCHY-GREEN-Tensors \mathbf{C} sind jeweils konvex²⁷ in \mathbf{F} , $\text{cof } \mathbf{F}$ und $\det \mathbf{F} = J$. Daraus folgt mit Eigenschaft b), dass eine in den Invarianten von \mathbf{C} konvexe und monoton wachsende Funktion $\hat{\Psi}(I_{\mathbf{C}}, II_{\mathbf{C}}, III_{\mathbf{C}})$ nach Gleichung (3.85) die Polykonvexität von $\Psi(\mathbf{F})$ impliziert. Diese Bedingungen sind strenger als die ursprünglichen Polykonvexitätsanforderungen, da die in der Funktion (3.85) geforderte Konvexität bzgl. J schwächer ist als die Konvexität und Nichtnegativität bzgl. J^2 .

Eine in den Hauptstreckungen $\{\lambda_a\}_{a=1}^3$ definierte Funktion Ψ der Form

$$\Psi(\lambda_1, \lambda_2, \lambda_3) = a(\lambda_1^\alpha + \lambda_2^\alpha + \lambda_3^\alpha) + bJ^\beta(\lambda_1^{-\beta} + \lambda_2^{-\beta} + \lambda_3^{-\beta}) + \Psi_{\text{vol}}(J) \quad (3.87)$$

ist polykonvex, wenn $a, b > 0$ und $\alpha, \beta > 1$ gilt sowie $\Psi_{\text{vol}}(J)$ konvex in $J \in (0, \infty)$ ist. Dies ist aus CIARLET [26] entnommen. Der in Gleichung (3.77) gewählte Ansatz für $\Psi_{\text{vol}}(J)$ erfüllt die letztgenannte Bedingung. Mit Eigenschaft a) konvexer Funktionen sowie der Form des OGDEN-Materialmodells (3.76) folgen aus den Bedingungen für $a, (bJ^\beta) \hat{=} \mu_r / \alpha_r$ und $\alpha, -\beta \hat{=} \alpha_r$ die Ungleichungen

$$\begin{aligned} \mu_r > 0, \quad \alpha_r > 1 \quad \text{oder} \\ \mu_r < 0, \quad \alpha_r < 1, \end{aligned} \quad (3.88)$$

welche strenger sind als die Stabilitätsanforderungen (3.78) von OGDEN.

Das lineare ST. VENANT-KIRCHHOFF-Materialgesetz nach Abschnitt 3.3.2 ist nicht polykonvex, vgl. RAOULT [146].

²⁷ Es gilt jeweils $I_{\mathbf{C}} = \|\mathbf{F}\|^2$, $II_{\mathbf{C}} = \|\text{cof}(\mathbf{F})\|^2$ und $III_{\mathbf{C}} = J^2$.

3.4 Klassische elastoplastische Materialmodelle

Das *Fließen* von Festkörpern wurde erstmalig von HENRI-ÉDOUARD TRESCA [174] in seinen ab 1864 erschienenen sechs Memoiren beschrieben. Mit seinen semi-empirischen Arbeiten legte er den Grundstein für die klassische Elastoplastizitätstheorie (EP). Grundlegende Nachschlagewerke zum Thema EP sind z. B. LEMAITRE und CHABOCHE [114], KHAN und HUANG [98] sowie HOULSBY und PUZRIN [85]. Im Bauingenieurwesen findet die EP vor allem in der Anwendung der Traglasttheoreme Gebrauch, die unter anderen auf DRUCKER ET AL. [38] im Jahr 1952 zurückgehen. Beispiele sind die plastische Bemessung von Stahl- und Stahlbetonquerschnitten, die Bemessung von Holzbauverbindungen, Traglastberechnungen von statischen Systemen und die Anwendung der Kollapstheoreme in der Geotechnik.

Im Folgenden werden große Verschiebungen und kleine Verzerrungen²⁸ vorausgesetzt. Daher werden der GREEN-LAGRANGE'sche Verzerrungstensor \mathbf{E} und der 2. PIOLA-KIRCHHOFF'sche Spannungstensor \mathbf{S} verwendet, obwohl die EP klassischerweise in den linearisierten Verzerrungs- und Spannungsmaßen $\boldsymbol{\varepsilon}$ und $\boldsymbol{\sigma}$ formuliert wird. Dadurch sind die zugehörigen Raten objektiv, siehe auch Abschnitt 3.2.2. Die maßgeblichen Modellgrundlagen der EP sind:

1. **Zerlegung der Verzerrungen** in einen elastischen und plastischen Anteil. Für kleine Verzerrungen wird diese Zerlegung als additiv²⁹ postuliert, mit

$$\mathbf{E} = \mathbf{E}^{el} + \mathbf{E}^{pl} . \quad (3.89)$$

2. Die **Fließbedingung** im Spannungsraum

$$F(\mathbf{S}, \mathbf{q}_1, \dots, \mathbf{q}_{n_\alpha}) \begin{cases} < 0 & \text{elastisch} \\ = 0 & \text{plastisch} \\ > 0 & \text{unzulässig} \end{cases} \quad (3.90)$$

definiert elastische und plastische Bereiche. Die Hyperfläche $F = 0$ wird als *Fließfläche* bezeichnet. Im elastischen Bereich $F < 0$ gilt das elastische Teilstoffgesetz $\mathbf{S} = \mathcal{S}(\mathbf{E}^{el})$ unabhängig von plastischen Effekten. Die Variablen $\{\mathbf{q}_i\}_{i=1}^{n_\alpha}$ sind zusätzliche spannungsäquivalente Variablen. Im Folgenden werden sie zunächst allgemein wie Tensoren 2. Stufe behandelt.

²⁸ Die Grenze ist im unteren einstelligen Prozentbereich.

²⁹ Dies kann bspw. bei Metallen durch verschiedene Effekte physikalisch begründet werden, die auf der Gitterebene stattfinden, vgl. LEMAITRE und CHABOCHE [114]. Für große Verzerrungen kann ein multiplikativer Split des Deformationsgradienten $\mathbf{F} = \mathbf{F}^{el} \mathbf{F}^{pl}$ angenommen werden, vgl. HOLZAPFEL [81].

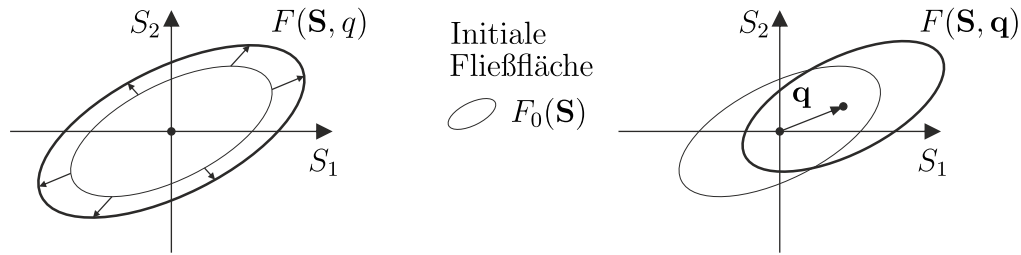


Abbildung 3.6: Darstellung isotroper (links) und kinematischer Verfestigung (rechts) in einem fiktiven Spannungsraum mit zwei Komponenten S_1 und S_2

3. **Verfestigungsregeln** beschreiben die Änderung der Fließfläche während des Fließens über die Abhängigkeiten $\mathbf{q}_i(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_{n_\alpha})$, mit verzerrungsäquivalenten inneren Variablen $\{\boldsymbol{\alpha}_i\}_{i=1}^{n_\alpha}$. Im Folgenden wird die Vereinfachung getroffen, dass es jeweils nur eine eins-zu-eins-Abhängigkeit $\mathbf{q}_i(\boldsymbol{\alpha}_i)$ gibt³⁰. Zwei übliche Spezialfälle sind in Abbildung 3.6 dargestellt. Sie lauten:

- Isotrope Verfestigung: selbstähnliches Aufblähen der Fließfläche. Dazu ist eine skalare Variable $q(\alpha)$ nötig.
- Kinematische Verfestigung: Verschiebung der Fließfläche im Spannungsraum bei konstanter Gestalt. Dazu ist eine tensorielle Variable $\mathbf{q}(\boldsymbol{\alpha})$ (engl.: *back stress*) nötig.

4. **Fließregeln** legen die Entwicklung der plastischen Verzerrungen \mathbf{E}^{pl} und der zusätzlichen inneren Variablen $\{\boldsymbol{\alpha}_i\}_{i=1}^{n_\alpha}$ fest, mit

$$\dot{\mathbf{E}}^{pl} = \lambda \mathbf{m}_{\mathbf{E}} \quad (3.91)$$

$$\dot{\boldsymbol{\alpha}}_i = \lambda \mathbf{m}_{\boldsymbol{\alpha}_i}, \quad i = 1, \dots, n_\alpha. \quad (3.92)$$

Dabei ist λ der sogenannte plastische Multiplikator, $\mathbf{m}_{\mathbf{E}}$ bzw. $\mathbf{m}_{\boldsymbol{\alpha}_i}$ sind die Fließrichtungen.

5. Die **Be- und Entlastungsbedingungen**

$$\lambda \geq 0, \quad F \leq 0, \quad \lambda F = 0 \quad (3.93)$$

garantieren eine eindeutige Unterscheidung in plastische Belastung und elastische Entlastung.

6. **Konsistenzbedingung:** Die Fließbedingung muss während des Fließens zu jeder Zeit erfüllt sein:

$$\dot{F} = 0 \Big|_{F=0}. \quad (3.94)$$

³⁰ Das trifft bei den hier betrachteten Materialmodellen zu und hat vor allem ästhetische Gründe bei der Herleitung der Materialtangente.

Mit diesem Satz an Gleichungen lässt sich darüber hinaus die elastoplastische Materialtangente $\mathbf{C}_T^{\text{kont}}$ mit $\dot{\mathbf{S}} = \mathbf{C}_T^{\text{kont}} \dot{\mathbf{E}}$ herleiten. Dies wird im Folgenden in reduzierter Notation für $\mathbf{C}_T^{\text{kont}}$ gezeigt. Aus dem elastischen Teilstoffgesetz, der additiven Zerlegung der Verzerrungen und der Fließregel folgt zunächst die Spannungsrate

$$\dot{\mathbf{S}} = \mathbf{C}_T^{\text{el}} \dot{\mathbf{E}}^{\text{el}} = \mathbf{C}_T^{\text{el}} (\dot{\mathbf{E}} - \dot{\mathbf{E}}^{\text{pl}}) = \mathbf{C}_T^{\text{el}} (\dot{\mathbf{E}} - \lambda \mathbf{m}_{\mathbf{E}}) , \quad (3.95)$$

mit der elastischen Materialtangente \mathbf{C}_T^{el} . Die Entwicklung der spannungsäquivalenten inneren Variablen kann mithilfe der zugehörigen Fließregeln über

$$\dot{\mathbf{q}}_i = \frac{\partial \mathbf{q}_i}{\partial \boldsymbol{\alpha}_i} \dot{\boldsymbol{\alpha}}_i = -\mathbf{H}_{T_i} \dot{\boldsymbol{\alpha}}_i = -\mathbf{H}_{T_i} \lambda \mathbf{m}_{\boldsymbol{\alpha}_i} \quad (3.96)$$

beschrieben werden, wobei die sogenannten *Verfestigungsmoduln* \mathbf{H}_{T_i} hier ebenfalls in reduzierter Notation angegeben sind. Das Minus ist üblich, muss prinzipiell jedoch nicht definiert werden. Darüber hinaus wurde eine eins-zu-eins-Abhängigkeit $\mathbf{q}_i(\boldsymbol{\alpha}_i)$ vorausgesetzt. Die Konsistenzbedingung

$$\begin{aligned} \dot{F}(\mathbf{S}, \mathbf{q}_1, \dots, \mathbf{q}_{n_\alpha}) = 0 &= \left(\frac{\partial F}{\partial \mathbf{S}} \right) \dot{\mathbf{S}} + \sum_{i=1}^{n_\alpha} \left(\frac{\partial F}{\partial \mathbf{q}_i} \right) \dot{\mathbf{q}}_i \\ &= \left(\frac{\partial F}{\partial \mathbf{S}} \right) \mathbf{C}_T^{\text{el}} (\dot{\mathbf{E}} - \lambda \mathbf{m}_{\mathbf{E}}) - \sum_{i=1}^{n_\alpha} \left(\frac{\partial F}{\partial \mathbf{q}_i} \right) \mathbf{H}_{F_i} \lambda \mathbf{m}_{\boldsymbol{\alpha}_i} \end{aligned} \quad (3.97)$$

kann folglich nach dem plastischen Multiplikator

$$\lambda = \frac{\left(\frac{\partial F}{\partial \mathbf{S}} \right) \mathbf{C}_T^{\text{el}} \dot{\mathbf{E}}}{\left(\frac{\partial F}{\partial \mathbf{S}} \right) \mathbf{C}_T^{\text{el}} \mathbf{m}_{\mathbf{E}} + \sum_{i=1}^{n_\alpha} \left(\frac{\partial F}{\partial \mathbf{q}_i} \right) \mathbf{H}_{T_i} \mathbf{m}_{\boldsymbol{\alpha}_i}} \quad (3.98)$$

aufgelöst werden. Wird dieser wiederum in die Spannungsrate (3.95) eingesetzt und $\dot{\mathbf{E}}$ ausgeklammert, folgt die elastoplastische Materialtangente

$$\mathbf{C}_T^{\text{kont}} = \mathbf{C}_T^{\text{el}} - \frac{\left(\mathbf{C}_T^{\text{el}} \mathbf{m}_{\mathbf{E}} \right) \left(\left(\frac{\partial F}{\partial \mathbf{S}} \right) \mathbf{C}_T^{\text{el}} \right)}{\left(\frac{\partial F}{\partial \mathbf{S}} \right) \mathbf{C}_T^{\text{el}} \mathbf{m}_{\mathbf{E}} + \sum_{i=1}^{n_\alpha} \left(\frac{\partial F}{\partial \mathbf{q}_i} \right) \mathbf{H}_{T_i} \mathbf{m}_{\boldsymbol{\alpha}_i}} . \quad (3.99)$$

Im elastischen Fall ($\lambda = 0$) gilt selbstverständlich $\mathbf{C}_T^{\text{kont}} = \mathbf{C}_T^{\text{el}}$.

In den Abschnitten 3.4.1 und 3.4.2 wird die EP-Theorie zunächst thermodynamisch eingeordnet. Im Anschluss wird die numerische Umsetzung skizziert. Die Anwendung auf das in dieser Arbeit verwendete VON MISES'sche Materialgesetz folgt schließlich ab Abschnitt 3.4.4.

3.4.1 Einordnung in den thermodynamischen Rahmen

Die Basis für die thermodynamische Untersuchung ist die Festlegung der Argumente der Verzerrungsenergiegedichtefunktion Ψ und damit der thermodynamischen Zustandsvariablen. Zunächst wird die Trennung von Ψ in einen elastischen und einen dissipativen Anteil postuliert, mit

$$\Psi(\mathbf{E}^{el}, \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_{n_\alpha}) = \Psi^s(\mathbf{E}^{el}) + \Psi^q(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_{n_\alpha}), \quad (3.100)$$

wobei die elastischen Verzerrungen \mathbf{E}^{el} und die inneren Variablen $\{\boldsymbol{\alpha}_i\}_{i=1}^{n_\alpha}$ den thermodynamischen Zustand beschreiben. Die daraus resultierende Rate

$$\dot{\Psi} = \frac{\partial \Psi^s}{\partial \mathbf{E}^{el}} : \dot{\mathbf{E}}^{el} + \sum_{i=1}^{n_\alpha} \frac{\partial \Psi^q}{\partial \boldsymbol{\alpha}_i} : \dot{\boldsymbol{\alpha}}_i \quad (3.101)$$

wird in die Dissipationsungleichung (3.25) eingesetzt, wodurch

$$\dot{D} = \left(\mathbf{S} - \frac{\partial \Psi^s}{\partial \mathbf{E}^{el}} \right) : \dot{\mathbf{E}}^{el} + \mathbf{S} : \dot{\mathbf{E}}^{pl} - \sum_{i=1}^{n_\alpha} \frac{\partial \Psi^q}{\partial \boldsymbol{\alpha}_i} : \dot{\boldsymbol{\alpha}}_i \geq 0 \quad (3.102)$$

folgt. Sie muss im elastischen Fall für beliebige $\dot{\mathbf{E}}^{el}$ erfüllt sein, was zur Bedingung

$$\mathbf{S} = \frac{\partial \Psi^s}{\partial \mathbf{E}^{el}} = \mathcal{S}(\mathbf{E}^{el}) \quad (3.103)$$

für das elastische Teilstoffgesetz führt. Die spannungsäquivalenten Variablen

$$\mathbf{q}_i := - \frac{\partial \Psi^q}{\partial \boldsymbol{\alpha}_i} \quad (3.104)$$

werden über den Gradienten von Ψ^q definiert³¹. Durch Einsetzen der Bedingung (3.103) und der Definition (3.104) in Gleichung (3.102), folgt mit

$$\dot{D} = \mathbf{S} : \dot{\mathbf{E}}^{pl} + \sum_{i=1}^{n_\alpha} \mathbf{q}_i : \dot{\boldsymbol{\alpha}}_i \geq 0 \quad (3.105)$$

die Rest-Dissipationsungleichung, welche von den Entwicklungsgleichungen $\dot{\mathbf{E}}^{pl}$ und $\dot{\boldsymbol{\alpha}}_i$ erfüllt werden muss. Darüber hinaus kann ein sogenanntes *plastisches Potential* $\phi(\mathbf{S}, \mathbf{q}_1, \dots, \mathbf{q}_{n_\alpha})$ eingeführt werden, über dessen Gradienten

$$\dot{\mathbf{E}}^{pl} = \lambda \frac{\partial \phi}{\partial \mathbf{S}} \quad (3.106)$$

$$\dot{\boldsymbol{\alpha}}_i = \lambda \frac{\partial \phi}{\partial \mathbf{q}_i}, \quad i = 1, \dots, n_\alpha, \quad (3.107)$$

die Entwicklungsgleichungen ableitbar sind. Ein Vergleich mit den Fließregeln (3.91) und (3.92) zeigt, dass die Gradienten dieser Potentialfunktion in diesem

³¹ Auch hier ist das Minus mathematisch nicht notwendig, vgl. Gleichung (3.96).

Fall gerade den Fließrichtungen $\mathbf{m}_{\mathbf{E}}$ und \mathbf{m}_{α_i} entsprechen. Für den Fall *assoziierter* Fließregeln wird die Fließfläche F als Potential verwendet mit $F = \phi$. Im allgemeinen *nicht assoziierten* Fall müssen drei skalare Funktionen definiert werden: die freie Energie $\Psi = \Psi^s + \Psi^q$, die Potentialfunktion ϕ für die Fließregeln und die Funktion F für die Fließbedingung.

Kommentar zu den getroffenen Annahmen

Ein elastoplastisches Materialmodell ist automatisch konsistent zum ersten und zweiten Hauptsatz der Thermodynamik, wenn die zuvor eingeführten skalaren Funktionen $\Psi = \Psi^s + \Psi^q$ und ϕ existieren und sie die Ungleichung (3.105) nicht verletzen. In diesem Fall wird auch von *generalisierten Standardmaterialien* gesprochen, vgl. HALPHEN und SON NGUYEN [71]. Dies ist ein praktischer Rahmen, wenn es um die Formulierung von elastoplastischen Materialgesetzen geht. Gleiches gilt für die additive Zerlegung der Verzerrungen. Alle diese Annahmen sind *hinreichend*, jedoch nicht *notwendig* zur Erfüllung der thermodynamischen Hauptsätze. Im Gegenteil: Gewisse Materialien zeigen in Experimenten, dass die Annahme eines plastischen Potentials ϕ nicht zutrifft. So reicht die klassische EP nicht aus, um bspw. bei der Modellierung von Böden das Kontraktanzverhalten nach einer Pfadumkehr zu erfassen. Bei der Beurteilung der physikalischen Konsistenz von KNN-Materialmodellen ist somit stets zu beachten, dass das klassische thermodynamische Korsett nicht in dem hier vorliegenden strikten Maß erfüllt sein muss.

3.4.2 Das Prinzip der maximalen plastischen Dissipationsleistung

Für gegebene plastische Verzerrungen \mathbf{E}^{pl} und innere Variablen $\{\alpha_i\}_{i=1}^{n_\alpha}$ maximiert unter allen *möglichen* Spannungszuständen $(\tilde{\mathbf{S}}, \{\tilde{\mathbf{q}}_i\}_{i=1}^{n_\alpha})$ der *echte* Spannungszustand $(\mathbf{S}, \{\mathbf{q}_i\}_{i=1}^{n_\alpha})$ die plastische Dissipationsleistung. Dies wird durch das *Prinzip der maximalen plastischen Dissipationsleistung* postuliert. Es lässt sich mit der Rest-Dissipationsgleichung (3.105) als Optimierungsproblem

$$\dot{\mathcal{D}} = \max_{\tilde{\mathbf{S}}, \tilde{\mathbf{q}}_i} \left\{ \tilde{\mathbf{S}} : \dot{\mathbf{E}}^{pl} + \sum_{i=1}^{n_\alpha} \tilde{\mathbf{q}}_i : \dot{\alpha}_i \right\} \quad \text{u. d. N.} \quad F(\tilde{\mathbf{S}}, \tilde{\mathbf{q}}_1, \dots) \leq 0 \quad (3.108)$$

aufstellen, mit u. d. N. als Abkürzung für *unter der Nebenbedingung*. Erfüllt ein gegebenes EP-Materialmodell dieses Prinzip, impliziert dies die nachfolgenden Eigenschaften.

- Alle Fließbedingungen sind assoziiert, es gilt $\phi = F$. Daraus ergibt sich mit Bedingung (3.103) die Symmetrie der Kontinuumstangente $\mathbf{C}_T^{\text{kont}}$.

- Die Be- und Entlastungsbedingungen sind als notwendige KARUSH-KUHN-TUCKER-Bedingungen des Optimierungsproblems automatisch erfüllt.
- Die Fließfläche F ist konvex.

Die zugehörigen Beweise sind bspw. in SIMO und HUGHES [87] zu finden. Durch das Prinzip der maximalen plastischen Dissipationsleistung können Existenz- und Eindeutigkeitsbedingungen für das zugehörige Randwertproblem formuliert werden, siehe ebenfalls SIMO und HUGHES [87]. Es ist allerdings nicht notwendig zur Erfüllung der thermodynamischen Konsistenz eines Materialmodells.

3.4.3 Projektionsverfahren

In der numerischen Umsetzung werden die Ratengleichungen (3.91) und (3.92) in der (Pseudo-)Zeit diskretisiert. Zwischen zwei Zeitpunkten t_n und t_{n+1} wird dazu das implizite EULER-Verfahren verwendet, da es für beliebige Zeitschritte stabil ist. Es gilt

$$\mathbf{E}_{n+1}^{pl} = \mathbf{E}_n^{pl} + \gamma \mathbf{m}_{\mathbf{E},n+1} \quad (3.109)$$

$$\boldsymbol{\alpha}_{i,n+1} = \boldsymbol{\alpha}_{i,n} + \gamma \mathbf{m}_{\boldsymbol{\alpha}_i,n+1}, \quad i = 1, \dots, n_\alpha. \quad (3.110)$$

Die Verzerrungen und Spannungen sind im Folgenden in der reduzierten Vektornotation gegeben. Der Faktor $\gamma := \lambda \Delta t$ fasst den plastischen Multiplikator λ und das (Pseudo-)Zeitinkrement Δt zusammen. Er muss mit den Spannungen \mathbf{S}_{n+1} , den plastischen Verzerrungen \mathbf{E}_{n+1}^{pl} und den inneren Variablen $\{\boldsymbol{\alpha}_i\}_{i=1}^{n_\alpha}$ ermittelt werden. Dazu wird das Projektionsverfahren verwendet, welches im Detail in SIMO und HUGHES [87] oder auch DE SOUZA ET AL. [34] zu finden ist. Eine Übersicht des Verfahrens ist in Tafel 3.1 dargestellt.

Die Verzerrungen \mathbf{E}_{n+1} werden als bekannt vorausgesetzt³². In einem elastischen Versuchsschritt (engl.: *trial*) werden die plastischen Verzerrungen und die inneren Variablen zunächst eingefroren. Liefert die Auswertung der Fließfläche F^{tr} zulässige Werte, ist der Schritt elastisch mit $\gamma = 0$. Ist $F^{tr} > 0$, werden die Spannungen im Allgemeinen iterativ auf die Fließfläche $F = 0$ projiziert. Details zu dieser Iteration sind jeweils bei den Materialmodellen ab Abschnitt 3.4.4 gegeben. Nach der Berechnung der Ausgangsdaten zum Zeitpunkt t_{n+1} muss für die übergeordnete Gleichgewichtssiteration die *algorithmisch konsistente Tangente* $\mathbf{C}_T = \mathbf{C}_T^{\text{alg}}$ berechnet werden. Durch die spezielle Form des Projektionsalgorithmus muss diese in Abhängigkeit der beteiligten numerischen Integrationsgleichungen hergeleitet

³² Bspw. im Rahmen der FEM gehen die Verzerrungen innerhalb der globalen Gleichgewichtssiteration als Eingangsgrößen im Integrationspunkt ein.

Eingangsdaten	$\mathbf{E}_n, \mathbf{E}_n^{pl}, \mathbf{E}_{n+1}, \mathbf{S}_n, \{\boldsymbol{\alpha}_{i,n}\}_{i=1}^{n_\alpha}$
Elastischer Prädiktor	$\gamma^{tr} = 0, \mathbf{E}^{pl,tr} = \mathbf{E}_n^{pl}, \boldsymbol{\alpha}_i^{tr} = \boldsymbol{\alpha}_{i,n}, i = 1, \dots, n_\alpha$
Auswertung der Fließbedingung	$F^{tr} = F(\mathbf{S}^{tr}, \boldsymbol{\alpha}_1^{tr}, \dots, \boldsymbol{\alpha}_{n_\alpha}^{tr})$
$F^{tr} \leq 0$	$\mathbf{S}_{n+1} = \mathbf{S}^{tr}, \mathbf{C}_T = \mathbf{C}^{el}, \mathbf{E}_{n+1}^{pl} = \mathbf{E}^{pl,tr},$ $\boldsymbol{\alpha}_{i,n+1} = \boldsymbol{\alpha}_i^{tr}, i = 1, \dots, n_\alpha$
$F^{tr} > 0$	Plastischer Korrektorschritt nötig 1) Projektion der Spannungen \mathbf{S}_{n+1} auf die Fließfläche 2) Berechnung der algorithmisch konsistenten Tangente $\mathbf{C}_T = \mathbf{C}_T^{alg}$
Ausgangsdaten	$\mathbf{S}_{n+1}, \mathbf{C}_T, \mathbf{E}_{n+1}^{pl}, \{\boldsymbol{\alpha}_{i,n+1}\}_{i=1}^{n_\alpha}$

Tafel 3.1: Übersicht des EP-Projektionsverfahrens

werden, siehe SIMO und TAYLOR [160]. \mathbf{C}_T^{alg} unterscheidet sich von der *Kontinuumstangente* \mathbf{C}_T^{kont} aus Gleichung (3.99) und ist Abhängig von der 'Zeit'-Diskretisierung.

3.4.4 Von Mises'sche Fließbedingung ohne Verfestigung

Nach der VON MISES'schen Gestaltänderungshypothese liefern hydrostatische Spannungen keinen Beitrag zum plastischen Fließen. Die VON MISES'sche Vergleichsspannung

$$\begin{aligned}
 S_V &= \sqrt{\frac{1}{2} \left((S_{11} - S_{22})^2 + (S_{11} - S_{33})^2 + (S_{22} - S_{33})^2 \right) + 3(S_{12}^2 + S_{13}^2 + S_{23}^2)} \\
 &= \sqrt{\frac{1}{2} \left((S_1 - S_2)^2 + (S_1 - S_3)^2 + (S_2 - S_3)^2 \right)} \\
 &= \sqrt{\frac{3}{2} \mathbf{S}^D : \mathbf{S}^D} = \sqrt{3 J_2}
 \end{aligned} \tag{3.111}$$

ist im Rahmen dieser Hypothese ein brauchbares skalares Maß zur Charakterisierung mehraxiger Spannungszustände. Sie kann mit der 2. Invarianten $J_2 := II_{\mathbf{S}^D}$ des Spannungsdeviators in Tensornotation

$$\mathbf{S}^D = \mathbf{S} - \frac{1}{3} \text{Sp}(\mathbf{S}) \mathbf{I} \tag{3.112}$$

berechnet werden. Für die reduzierte Notation ist es zunächst zweckmäßig, die Projektionsmatrix \mathbf{P} einzuführen, die über

$$\mathbf{S}^D : \mathbf{S}^D = \begin{bmatrix} S_{11} \\ S_{22} \\ S_{33} \\ S_{12} \\ S_{13} \\ S_{23} \end{bmatrix}^T \begin{bmatrix} 2/3 & -1/3 & -1/3 & 0 & 0 & 0 \\ -1/3 & 2/3 & -1/3 & 0 & 0 & 0 \\ -1/3 & -1/3 & 2/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} S_{11} \\ S_{22} \\ S_{33} \\ S_{12} \\ S_{13} \\ S_{23} \end{bmatrix} = \mathbf{S}^T \mathbf{P} \mathbf{S} \quad (3.113)$$

das Skalarprodukt des Spannungsdeviators mit sich selbst berechnet wird. Die VON MISES'sche Fließbedingung lässt sich damit in quadratischer Form³³ zu

$$F(\mathbf{S}) = \underbrace{\frac{1}{2} \mathbf{S}^T \mathbf{P} \mathbf{S}}_{S_V^2/3} - \frac{1}{3} Y_0^2 \leq 0 \quad (3.114)$$

schreiben. Die skalare Fließgrenze Y_0 ist ein Materialparameter und wird zunächst als konstant angenommen. Die isotrope Fließfläche lässt sich im Hauptachsensystem als Zylinder darstellen, was in Abbildung 3.7 dargestellt ist. Im Fall des ebenen Spannungszustands ($S_{33} = 0$ bzw. $S_3 = 0$) reduziert sich die Fließfläche auf eine Ellipse.

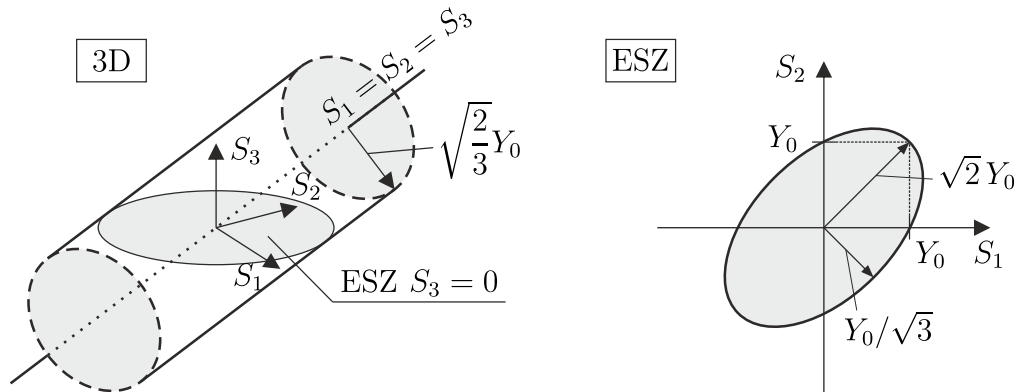


Abbildung 3.7: VON MISES'sche Fließfläche im dreidimensionalen Raum der Hauptspannungen (links) und im zweidimensionalen Raum der Hauptspannungen im ebenen Spannungszustand (rechts)

Die Fließregel ist assoziativ. Die für den plastischen Korrektorschritt ($F^{tr} > 0$) aus Abschnitt 3.4.3 nötigen Gleichungen werden im Folgenden hergeleitet. Die

³³Diese quadratische Form ist gleichwertig mit der in der Theorie üblicheren Wurzelform $F = S_V - Y_0$, liefert aber anschaulichere Gleichungen.

diskretisierte, assoziative Fließregel lautet

$$\mathbf{E}_{n+1}^{pl} = \mathbf{E}_n^{pl} + \gamma \left(\frac{\partial F(\mathbf{S}_{n+1})}{\partial \mathbf{S}} \right)^T = \mathbf{E}_n^{pl} + \gamma \mathbf{P} \mathbf{S}_{n+1} , \quad (3.115)$$

mit der Ableitung der Fließfläche nach dem Spannungsvektor $\partial F / \partial \mathbf{S} = (\mathbf{P} \mathbf{S})^T$ und der Symmetrie von $\mathbf{P} = \mathbf{P}^T$. Der Parameter γ wird berechnet, indem die Erfüllung der Fließbedingung (3.114) am noch unbekanntem Zeitpunkt t_{n+1} mit

$$F(\mathbf{S}_{n+1}(\gamma)) \stackrel{!}{=} 0 \quad (3.116)$$

eingefordert wird. In Sonderfällen ist diese Gleichung analytisch lösbar³⁴. Im Allgemeinen kann z. B. das NEWTON-Verfahren angewendet werden. Dazu wird die Tangente

$$\frac{\partial F}{\partial \gamma} = \left(\frac{\partial F}{\partial \mathbf{S}} \right) \frac{\partial \mathbf{S}}{\partial \gamma} = (\mathbf{P} \mathbf{S})^T \frac{\partial \mathbf{S}}{\partial \gamma} \quad (3.117)$$

in jedem Iterationsschritt benötigt. Die Ableitung $\partial \mathbf{S} / \partial \gamma$ wird am noch unbekanntem Spannungszustand ermittelt. Gilt bspw. ein lineares elastisches Teilstoffgesetz $\mathbf{S} = \mathcal{S}(\mathbf{E}^{el}) = \mathbf{C} \mathbf{E}^{el}$, folgt mit $\mathbf{E}_{n+1}^{el} = \mathbf{E}_{n+1} - \mathbf{E}_{n+1}^{pl}$ und der diskretisierten Fließregel (3.115) die implizite Gleichung

$$\mathbf{S}_{n+1} = \mathbf{C}(\mathbf{E}_{n+1} - \mathbf{E}_n^{pl} - \gamma \mathbf{P} \mathbf{S}_{n+1}) , \quad (3.118)$$

die nach den Spannungen

$$\mathbf{S}_{n+1} = (\mathbf{C}^{-1} + \gamma \mathbf{P})^{-1} (\mathbf{E}_{n+1} - \mathbf{E}_n^{pl}) = \mathbf{C}_{ep}(\gamma) \mathbf{E}^{tr} \quad (3.119)$$

umgeformt werden kann. Mithilfe der Ableitungsvorschrift

$$\frac{\partial \mathbf{A}^{-1}(\gamma)}{\partial \gamma} = -\mathbf{A}^{-1} \left(\frac{\partial \mathbf{A}}{\partial \gamma} \right) \mathbf{A}^{-1} , \quad (3.120)$$

siehe bspw. SELBY [153], folgt zunächst

$$\frac{\partial \mathbf{S}}{\partial \gamma} = -\mathbf{C}_{ep} \mathbf{P} \mathbf{C}_{ep} \mathbf{E}^{tr} = -\mathbf{C}_{ep} \mathbf{P} \mathbf{S} \quad (3.121)$$

und schließlich für das lokale NEWTON-Verfahren die Ableitung

$$\frac{\partial F}{\partial \gamma} = -(\mathbf{P} \mathbf{S})^T \mathbf{C}_{ep} \mathbf{P} \mathbf{S} . \quad (3.122)$$

Das lokale NEWTON-Verfahren zur Spannungsprojektion ist in Tafel 3.2 zusammengefasst. Im Anschluss an die lokale Iteration können die plastischen Verzerrungen \mathbf{E}_{n+1}^{pl} aktualisiert werden.

³⁴ Dazu kann bspw. die isotrope linear elastische Steifigkeit $\mathbf{C}_T^{el} \equiv \mathbf{C}$ in einen volumetrischen und einen deviatorischen Anteil aufgeteilt werden, siehe SIMO und HUGHES [87]. Zur Darlegung des allgemeinen Vorgehens bei der nichtlinearen Spannungsprojektion, wird an dieser Stelle darauf verzichtet, trotz dieser effizienteren Implementierungsmöglichkeit.

Initialisierung	$\gamma^{(0)} = 0, F^{(0)} = F^{tr}, \mathbf{S}_{n+1}^{(0)} = \mathbf{S}_{n+1}^{tr}, \mathbf{C}_{ep}^{(0)} = \mathbf{C}, i = 0$
Fließbedingung	$F(\mathbf{S}_{n+1}^{(i)}) \leq 0?$
ja: Abbruch	$\gamma = \gamma^{(i)}, \mathbf{S}_{n+1} = \mathbf{S}_{n+1}^{(i)}, \mathbf{C}_{ep,n+1} = \mathbf{C}_{ep}^{(i)}$
nein: Update	$\gamma^{(i+1)} = \gamma^{(i)} + \frac{F(\mathbf{S}_{n+1}^{(i)})}{(\mathbf{PS}_{n+1}^{(i)})^T \mathbf{C}_{ep}^{(i)} (\mathbf{PS}_{n+1}^{(i)})}$
	$\mathbf{C}_{ep}^{(i+1)} = [\mathbf{C}^{-1} + \gamma^{(i+1)} \mathbf{P}]^{-1}$
	$\mathbf{S}_{n+1}^{(i+1)} = \mathbf{C}_{ep}^{(i+1)} \mathbf{E}^{tr}$
	$i = i + 1$

Tafel 3.2: Lokale NEWTON-Iteration des Projektionsalgorithmus für die quadratische VON MISES'sche Fließbedingung ohne Verfestigung

Die algorithmisch konsistente Tangente $\mathbf{C}_T = \mathbf{C}_T^{\text{alg}}$ wird unter Berücksichtigung der beteiligten numerischen Integrationsgleichungen zur Ermittlung von γ hergeleitet über

$$\mathbf{C}_T = \frac{d\mathbf{S}_{n+1}}{d\mathbf{E}} = \frac{\partial \mathbf{S}}{\partial \mathbf{E}^{tr}} + \frac{\partial \mathbf{S}}{\partial \gamma} \frac{\partial \gamma}{\partial \mathbf{E}^{tr}}. \quad (3.123)$$

Die Konsistenzbedingung $\dot{F} = 0$ liefert in der ratenunabhängigen Form

$$dF = \left(\frac{\partial F}{\partial \mathbf{S}} + \frac{\partial F}{\partial \gamma} \frac{\partial \gamma}{\partial \mathbf{E}^{tr}} \frac{\partial \mathbf{E}^{tr}}{\partial \mathbf{S}} \right) d\mathbf{S} = 0 \quad (3.124)$$

für beliebige $d\mathbf{S}$ mit $\partial \mathbf{E}^{tr} / \partial \mathbf{S} = \mathbf{C}_{ep}^{-1}$ die partielle Ableitung

$$\frac{\partial \gamma}{\partial \mathbf{E}^{tr}} = - \left(\frac{\partial F}{\partial \gamma} \right)^{-1} \frac{\partial F}{\partial \mathbf{S}} \left(\frac{\partial \mathbf{E}^{tr}}{\partial \mathbf{S}} \right)^{-1} = \frac{(\mathbf{PS})^T \mathbf{C}_{ep}}{(\mathbf{PS})^T \mathbf{C}_{ep} (\mathbf{PS})}. \quad (3.125)$$

Die darin enthaltenen Anteile wurden schon für die lokale NEWTON-Iteration benötigt und weiter oben berechnet. Die symmetrische, algorithmisch konsistente Tangente folgt letztlich in reduzierter Notation zu

$$\mathbf{C}_T^{\text{alg}} = \mathbf{C}_{ep} - \frac{\mathbf{C}_{ep} (\mathbf{PS}) (\mathbf{PS})^T \mathbf{C}_{ep}}{(\mathbf{PS})^T \mathbf{C}_{ep} (\mathbf{PS})} \neq \mathbf{C} - \frac{\mathbf{C} (\mathbf{PS}) (\mathbf{PS})^T \mathbf{C}}{(\mathbf{PS})^T \mathbf{C} (\mathbf{PS})} = \mathbf{C}_T^{\text{kont}}. \quad (3.126)$$

Im Vergleich mit der analytischen Kontinuumstangente $\mathbf{C}_T^{\text{kont}}$ aus Gleichung (3.99) ist der algorithmische Charakter gut zu erkennen. Für unendlich kleine ‚Zeit‘-Schritte $\gamma \rightarrow 0$ gälte $\mathbf{C}_{ep} \rightarrow \mathbf{C}$ und somit $\mathbf{C}_T^{\text{alg}} \rightarrow \mathbf{C}_T^{\text{kont}}$.

Elastoplastische Materialgesetze mit VON MISES'scher Fließbedingung werden vor allem bei der Modellierung von Metallen verwendet. Ideale Plastizität – also ohne Verfestigung – ist z. B. bei S235-Baustahl als Annahme gerechtfertigt. Für diesen folgt nach dem Eurocode 3 [35] mit der Streckgrenze $Y_0 = 23,5 \text{ kN/cm}^2$, der Zugfestigkeit $Y_u = 36 \text{ kN/cm}^2$ und dem E-Modul $E = 21\,000 \text{ kN/cm}^2$ sowie der Annahme einer Dehngrenze von $\varepsilon_y = 0,002$ und einer Mindestbruchdehnung von $\varepsilon_y = 0,15$ ein Tangentenmodul während des Plastifizierens von $C_T < 0,005 E$.

3.4.5 Von Mises'sche Fließbedingung mit gemischter Verfestigung

Die ideal-plastische VON MISES'sche Fließbedingung aus Abschnitt 3.4.4 wird um eine lineare isotrope und eine lineare kinematische Verfestigung erweitert. Die Fließbedingung

$$F(\mathbf{S}, \mathbf{q}, q) = \frac{1}{2}(\mathbf{S} - \mathbf{q})^T \mathbf{P}(\mathbf{S} - \mathbf{q}) - \frac{1}{3}Y(q)^2 \leq 0 \quad (3.127)$$

ist zusätzlich vom deviatorischen Spannungstensor \mathbf{q} und der skalaren Spannung q abhängig, vgl. Abbildung 3.6. Die Spannungsgrößen

$$\mathbf{S} = \mathbf{C}(\mathbf{E} - \mathbf{E}^{pl}) = \mathbf{C}\mathbf{E}^{el}, \quad (3.128)$$

$$\mathbf{q} = -C_P(1 - \theta)\boldsymbol{\alpha} \quad \text{und} \quad (3.129)$$

$$q = -C_P\theta\alpha \quad (3.130)$$

sind hier linear abhängig von den zugehörigen assoziierten inneren Variablen \mathbf{E}^{el} , $\boldsymbol{\alpha}$ und α ³⁵. Der Materialparameter $\theta \in [0, 1]$ steuert die Aufteilung des plastischen Verfestigungsmoduls C_P zwischen isotroper und kinematischer Verfestigung. Die aktuelle Fließgrenze

$$Y(q) = Y_0 - q(\alpha) = Y_0 + C_P\theta\alpha \quad (3.131)$$

ist linear in q definiert. Die diskretisierten assoziativen Fließregeln sind für den plastischen Korrektorschritt ($F = 0$) mit

$$\mathbf{E}_{n+1}^{pl} = \mathbf{E}_n^{pl} + \gamma\mathbf{P}\hat{\mathbf{S}}_{n+1} \quad (3.132)$$

$$\boldsymbol{\alpha}_{n+1} = \boldsymbol{\alpha}_n - \gamma\mathbf{P}\hat{\mathbf{S}}_{n+1} \quad (3.133)$$

$$\alpha_{n+1} = \alpha_n + \gamma\frac{2}{3}Y_{n+1} = \alpha_n + \gamma\sqrt{\frac{2}{3}\hat{\mathbf{S}}_{n+1}^T \mathbf{P}\hat{\mathbf{S}}_{n+1}} \quad (3.134)$$

gegeben. Dabei sind $\hat{\mathbf{S}}_{n+1} := \mathbf{S}_{n+1} - \mathbf{q}_{n+1}$ die relativen, deviatorischen Spannungen. Die Umformung in Gleichung (3.134) kann durchgeführt werden, da im noch unbekanntem Zustand die Fließbedingung (3.127) erfüllt sein muss. In der lokalen NEWTON-Iteration werden die drei Variablen \mathbf{S}_{n+1} , \mathbf{q}_{n+1} und γ mit $\mathbf{V} := [\mathbf{S}_{n+1}, \mathbf{q}_{n+1}, \gamma]^T$ gleichzeitig berechnet. Der Residualvektor

$$\mathbf{R}(\mathbf{V}) = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \\ R_3 \end{bmatrix} = \begin{bmatrix} \mathbf{C}^{-1}\mathbf{S}_{n+1} - \mathbf{E}^{tr} + \gamma\mathbf{P}\hat{\mathbf{S}}_{n+1} \\ \mathbf{q}_{n+1} + C_P(1 - \theta)\boldsymbol{\alpha}_n - C_P(1 - \theta)\gamma\mathbf{P}\hat{\mathbf{S}}_{n+1} \\ \frac{1}{2}\hat{\mathbf{S}}_{n+1}^T \mathbf{P}\hat{\mathbf{S}}_{n+1} - \frac{1}{3}Y_{n+1}^2 \end{bmatrix} \quad (3.135)$$

³⁵ Die skalare Verfestigungsvariable α zur Berücksichtigung der isotropen Verfestigung steht in Zusammenhang mit der plastischen Vergleichsdehnung, ist aber nicht exakt gleich. Im Allgemeinen gibt es verschiedene Möglichkeiten, Verfestigungsvariablen zu definieren.

fordert das elastische Teilstoffgesetz (\mathbf{R}_1), die Evolutionsgleichung der kinematischen Verfestigung (\mathbf{R}_2) und die Fließbedingung (R_3) ein. Die zugehörige lokale Iterationstangente

$$\mathbf{T} = \frac{d\mathbf{R}}{d\mathbf{V}} = \begin{bmatrix} \mathbf{C}^{-1} + \gamma\mathbf{P} & -\gamma\mathbf{P} & \mathbf{P}\hat{\mathbf{S}}_{n+1} \\ -C_P(1-\theta)\gamma\mathbf{P} & \mathbf{1}^6 + C_P(1-\theta)\gamma\mathbf{P} & -C_P(1-\theta)\mathbf{P}\hat{\mathbf{S}}_{n+1} \\ (\mathbf{P}\hat{\mathbf{S}}_{n+1})^T & -(\mathbf{P}\hat{\mathbf{S}}_{n+1})^T & -\frac{4}{9}Y_{n+1}^2 \left(\frac{C_P\theta}{1-(2/3)C_P\theta\gamma} \right) \end{bmatrix} \quad (3.136)$$

ist nicht symmetrisch, dafür allerdings nicht-singulär für alle $\theta \in [0, 1]$. $\mathbf{1}^6$ ist eine 6×6 -Einheitsmatrix. Das lokale NEWTON-Verfahren ist in Tafel 3.3 zusammengefasst. Im Anschluss an die lokale Iteration können die plastischen Verzerrungen \mathbf{E}_{n+1}^{pl} und die inneren Variablen $\boldsymbol{\alpha}_{n+1}$ und α_{n+1} aktualisiert werden.

Initialisierung	$\gamma^{(0)} = 0, F^{(0)} = F^{tr}, \mathbf{S}_{n+1}^{(0)} = \mathbf{S}_{n+1}^{tr}, \mathbf{q}_{n+1}^{(0)} = \mathbf{q}_{n+1}^{tr}, i = 0$
Fließbedingung	$\ \mathbf{R}^{(i)}\ \leq TOL?$
ja: Abbruch	$\gamma = \gamma^{(i)}, \mathbf{S}_{n+1} = \mathbf{S}_{n+1}^{(i)}, \mathbf{q}_{n+1} = \mathbf{q}_{n+1}^{(i)}$
nein: Update	$\Delta\mathbf{V}^{(i)} = -\mathbf{T}^{(i)-1}\mathbf{R}^{(i)}$ mit $\Delta\mathbf{V}^{(i)} = [\Delta\mathbf{S}_{n+1}^{(i)}, \Delta\mathbf{q}_{n+1}^{(i)}, \Delta\gamma^{(i)}]^T$ und $\mathbf{T}^{(i)}$ nach Gleichung (3.136)
	$\mathbf{S}_{n+1}^{(i+1)} = \mathbf{S}_{n+1}^{(i)} + \Delta\mathbf{S}_{n+1}^{(i)}$
	$\mathbf{q}_{n+1}^{(i+1)} = \mathbf{q}_{n+1}^{(i)} + \Delta\mathbf{q}_{n+1}^{(i)}$
	$\gamma^{(i+1)} = \gamma^{(i)} + \Delta\gamma^{(i)}$
	$i = i + 1$

Tafel 3.3: Lokale NEWTON-Iteration des Projektionsalgorithmus für die quadratische VON MISES'sche Fließbedingung mit gemischter linearer Verfestigung

Die algorithmisch konsistente Materialtangente kann, dank geschickter Wahl von $\mathbf{R}_1 = \mathbf{E}_{n+1}^{el}$ im Residualvektor (3.135), durch Invertierung der letzten lokalen Tangente \mathbf{T} bestimmt werden:

$$(\mathbf{T}^{(i)})^{-1} = \begin{bmatrix} [\mathbf{C}_T]_{6 \times 6} & [\dots] & [\dots] \\ [\dots] & [\dots] & [\dots] \\ [\dots] & [\dots] & [\dots] \end{bmatrix}. \quad (3.137)$$

Alternativ lässt sie sich direkt über die Formel

$$\mathbf{C}_T = \hat{\mathbf{C}}_{ep} - \frac{\hat{\mathbf{C}}_{ep}(\mathbf{H}\mathbf{P}\hat{\mathbf{S}}_{n+1})(\mathbf{H}\mathbf{P}\hat{\mathbf{S}}_{n+1})^T\hat{\mathbf{C}}_{ep}}{(1+\beta)(\mathbf{H}\mathbf{P}\hat{\mathbf{S}}_{n+1})^T\hat{\mathbf{C}}_{ep}(\mathbf{H}\mathbf{P}\hat{\mathbf{S}}_{n+1})} \quad (3.138)$$

berechnen, mit den Einzelkomponenten

$$\hat{\mathbf{H}} = \left(\frac{1}{(1-\theta)C_P} \mathbf{1} + \gamma \mathbf{P} \right)^{-1} \xrightarrow{\theta=1} \mathbf{0} \quad (3.139)$$

$$\mathbf{H} = \mathbf{1} - \gamma \mathbf{P} \hat{\mathbf{H}} \quad (3.140)$$

$$\hat{\mathbf{C}}_{ep} = (\mathbf{C}^{-1} + \mathbf{H} \gamma \mathbf{P})^{-1} \quad (3.141)$$

$$\beta = \frac{\frac{2\theta C_P}{3-2\theta C_P \gamma} \hat{\mathbf{S}}_{n+1}^T \mathbf{P} \hat{\mathbf{S}}_{n+1} + \hat{\mathbf{S}}_{n+1}^T \mathbf{P} \hat{\mathbf{H}} \mathbf{P} \hat{\mathbf{S}}_{n+1}}{(\mathbf{H} \mathbf{P} \hat{\mathbf{S}}_{n+1})^T \hat{\mathbf{C}}_{ep} (\mathbf{H} \mathbf{P} \hat{\mathbf{S}}_{n+1})} . \quad (3.142)$$

Auf die Herleitung wird hier verzichtet. Eine Darstellung für die zugehörige Wurzelform ist in SIMO und HUGHES [87] zu finden.

4 Künstliche neuronale Netze

Das menschliche Gehirn ist ein *biologisches neuronales Netz* und besteht aus ungefähr 86 Milliarden Neuronen, vgl. HERCULANO-HOUZEL [76]. Jedes Neuron erhält elektrische Signale von tausenden benachbarten Neuronen, verarbeitet sie und gibt bei Erreichen einer Stimulationsgrenze elektrische Impulse an benachbarte Neuronen weiter. Obwohl die Informationsverarbeitung eines einzelnen Neurons vergleichsweise einfach ist, kann das Gehirn durch die Vernetzung vieler Neuronen die Lösung hochkomplexer Aufgaben erlernen. *Künstliche neuronale Netze* (KNN) sind Abstraktionen dieser biologischen Informationsverarbeitungsmaschinen. Sie bestehen aus *künstlichen Neuronen*, die über gewichtete Verbindungen miteinander Informationen austauschen können. Eine populärwissenschaftliche Gegenüberstellung beider Netze ist in Abbildung 4.1 dargestellt.

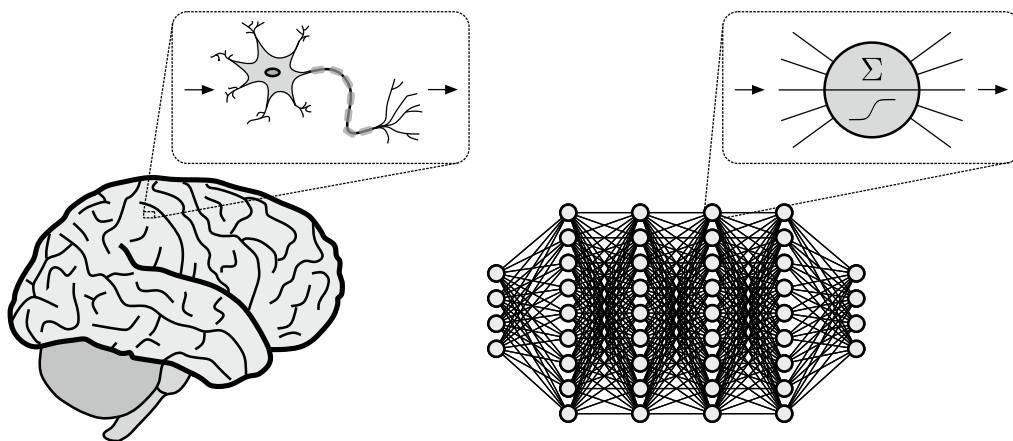


Abbildung 4.1: Skizze eines biologischen neuronalen Netzes mit zugehörigem Neuron (links) und eines künstlichen neuronalen Netzes mit künstlichem Neuron (rechts)

Einem KNN wird die zu erledigende Aufgabe nicht explizit einprogrammiert. Stattdessen soll das zugehörige Wissen mit Methoden des *maschinellen Lernens* implizit generiert werden. Im Zusammenhang mit KNN bedeutet *Lernen*, die gewichteten Verbindungen zielgerichtet anzupassen. Dies wird in der vorliegenden Arbeit ausschließlich mit *überwachtem Lernen* getan: mit der Diskrepanz zu bereitgestellten Beispielen können die gewichteten Verbindungen angepasst werden. Daneben existieren andere Lernformen, wie unüberwachtes, teilüberwachtes oder bestärkendes Lernen. Maschinelles Lernen sollte nicht mit *künstlicher Intelligenz* (KI) verwechselt werden. KI ist ein Oberbegriff für die Übertragung aller kognitiven menschlichen Fähigkeiten auf die Maschine³⁶.

³⁶ Die Vielzahl biologisch motivierter Begriffe sollte nicht von den mathematischen und numerischen Hintergründen ablenken, gerade im Bereich der KI.

Im Folgenden wird eine Auswahl historischer Meilensteine im Rahmen künstlicher neuronaler Netze gegeben.

- | | |
|-------------|---|
| 1943 | Erste Definition eines künstlichen Neurons und Vernetzung zu einem KNN von MCCULLOCH und PITTS [128], um logische und arithmetische Funktionen berechnen zu können. |
| 1949 | Formulierung der HEBB'schen Lernregel [73], als erste ihrer Art, zum Anpassen der Gewichte in KNN. |
| 1957 | Entwicklung des einlagigen Perzeptrons von ROSENBLATT [151], welches schon einfache Ziffern erkennen konnte. |
| 1960 | Erste kommerzielle Verbreitung des mline-KNN von WIDROW und HOFF [194] zur Echtzeit-Echofilterung in Analogtelefonen. |
| 1969 | Mathematischer Beweis von MINSKY und PAPERT [131], dass das einlagige Perzeptron bestimmte Probleme nicht lösen kann. Damit fiel das Forschungsgebiet in eine rund 15 Jahre andauernde Krise. |
| 1974 | Die Entwicklung des Backpropagation-Algorithmus in der Dissertation von WERBOS [191]. Die Arbeit fand jedoch wenig Aufmerksamkeit, da auf dem Gebiet zu jener Zeit wenig geforscht wurde. |
| 1985 | Lösung des Travelling-Salesman-Problems durch ein drei Jahre zuvor von HOPFIELD [83] veröffentlichtes rekurrentes neuronales Netz. |
| 1985 | Unabhängige Entwicklung des Backpropagation-Verfahrens von RUMELHART ET AL. [152] für ein mehrschichtiges Perzeptron. |

Ab Mitte der 1980er Jahre erlebte das Forschungsgebiet schließlich eine Renaissance, die bis heute anhält. Aktuelle Durchbrüche werden meist im Bereich des *Deep Learning* verzeichnet. So wird das Training von KNN mit vielen Neuronschichten genannt. Die Anwendungsgebiete sind unter anderem Muster- und Gesichtserkennung, Sprachsynthese, Bildbearbeitung und medizinische Diagnostik. Die Heterogenität der Anwendungsgebiete spiegelt sich auch in den Lehr- und Fachbüchern wider. Ein Fundamentalwerk ist das Buch von GOODFELLOW ET AL. [64], während KRIESEL [108] ein gutes Skript zum Einstieg in die Thematik zur Verfügung stellt. Eine Übersicht über die Anwendung in der numerischen Mechanik ist in KOLLMANNBERGER [107] zu finden. Zur Nutzung künstlicher neuronaler Netze im Bauingenieurwesen haben FLOOD und KARTAM [46, 47] sowie ADELI [2] frühe Arbeiten veröffentlicht.

In dieser Arbeit werden hauptsächlich Feedforward-Netze zur Funktionsapproximation verwendet. Ein Feedforward-KNN lässt sich als Vektorfunktion

$$\mathbf{z} = \mathbf{f}^{\text{KNN}}(\mathbf{x}, \mathbf{w}) \quad (4.1)$$

angeben, die den n_i -dimensionalen Eingangsvektor \mathbf{x} (input) auf den n_o -dimensionalen Ausgangsvektor \mathbf{z} (output) abbildet. Im Rahmen der KNN-Materialmo-

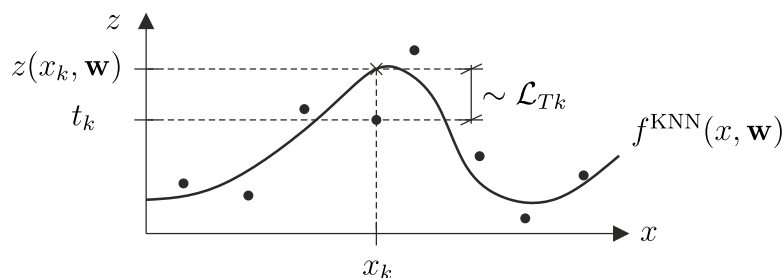


Abbildung 4.2: Das KNN als Approximationsmodell gegebener Datenpunkte

dellierung sind im einfachsten Fall die Eingangsvariablen die Verzerrungen und die Ausgangsvariablen die Spannungen. Wie genau die Vektoren \mathbf{x} und \mathbf{z} zur Modellierung spezifischen Materialverhaltens definiert werden müssen, folgt in den Kapiteln 5 und 6. Die freien Parameter – die *Gewichte* – werden im n_w -dimensionalen Vektor \mathbf{w} zusammengefasst (weights) und müssen im *Trainingsprozess* bestimmt werden.

Grundlage des Trainingsprozesses ist eine Menge

$$T := \{(\mathbf{x}_k, \mathbf{t}_k)\}_{k=1}^{P_T} \quad (4.2)$$

von Trainingsdatenpaaren (auch kurz Trainingsdaten), bestehend aus Eingangsvektoren \mathbf{x}_k und bekannten Ausgangsvektoren \mathbf{t}_k . Die Eingangsvektoren \mathbf{x}_k werden im Folgenden auch als Datenpunkte bezeichnet. Innerhalb des Trainingsprozesses wird eine Konfiguration von Gewichten $\hat{\mathbf{w}}$ gesucht, sodass die Abbildungsfunktion \mathbf{f}^{KNN} die gegebenen Daten ausreichend gut approximiert:

$$\mathbf{z}_k = \mathbf{f}^{\text{KNN}}(\mathbf{x}_k, \hat{\mathbf{w}}) \approx \mathbf{t}_k, \quad \forall k = 1, \dots, P_T. \quad (4.3)$$

Dies ist in Abbildung 4.2 dargestellt. Darüber hinaus sollte das KNN gut generalisieren, was einer *Interpolation* zwischen den gegebenen Datenpunkten entspricht. Eine *Extrapolation* außerhalb der gegebenen Datenpunkte ist im Allgemeinen nicht möglich. Der Trainingsprozess wird durch ein Optimierungsproblem

$$\min_{\mathbf{w} \in \mathbb{R}^{n_w}} \mathcal{L}(\mathbf{w}) \quad (4.4)$$

mathematisch formuliert, welches unter Verwendung des Gradienten

$$\nabla \mathcal{L}(\mathbf{w}) := \frac{d\mathcal{L}}{d\mathbf{w}} \quad (4.5)$$

mit iterativen Optimierungsalgorithmen gelöst wird. Die zu minimierende Fehlerfunktion $\mathcal{L}(\mathbf{w})$ kann verschiedene Terme enthalten, die im Laufe des Kapitels eingeführt werden.

Ziel dieses Kapitels ist die Bereitstellung aller nötigen Informationen für ein fundiertes Verständnis über die hier verwendeten KNN-Architekturen sowie die zugehörigen Algorithmen für eine effiziente Implementierung³⁷. Zunächst wird in Abschnitt 4.1 das Thema der Datennormierung behandelt und die klassische Fehlerfunktion bzgl. der Trainingsdaten beschrieben. In den Abschnitten 4.2 und 4.3 folgt die Beschreibung der KNN-Architekturen, die in dieser Arbeit verwendet werden. Dies schließt die Berechnung der folgenden Größen ein:

- Die Abbildungsfunktion $\mathbf{z} = \mathbf{f}^{\text{KNN}}(\mathbf{x}, \mathbf{w})$ und deren JACOBI-Matrix

$$\mathbf{j}(\mathbf{x}, \mathbf{w}) := \frac{\partial \mathbf{z}(\mathbf{x}, \mathbf{w})}{\partial \mathbf{x}} \in \mathbb{R}^{n_o \times n_i} . \quad (4.6)$$

Diese Größen werden zur Berechnung der Fehlerfunktion \mathcal{L} und ihres Gradienten $\nabla \mathcal{L}$ benötigt. Außerdem werden mit ihnen innerhalb der KNN-Materialmodelle die Spannungen \mathbf{S} und die Materialtangente \mathbf{C}_T berechnet. Somit sind die Algorithmen zur Berechnung von \mathbf{z} und \mathbf{j} diejenigen, die in einem FE-Programm zur Nutzung von KNN-Materialmodellen implementiert werden müssen.

- Der Gradient $\nabla \mathcal{L}(\mathbf{z}(\mathbf{w}))$ einer Fehlerfunktion, welche KNN-Ausgangsvariablen \mathbf{z} enthält. Ein Beispiel ist der Fehler bzgl. der Trainingsdaten.
- Der Gradient $\nabla \mathcal{L}(\mathbf{j}(\mathbf{w}))$ einer Fehlerfunktion, welche die KNN-Ableitungen aus \mathbf{j} enthält. Diese Gradienten werden für die physikalisch motivierten Zusatzterme im Optimierungsproblem (4.4) benötigt.

Weiter werden in Abschnitt 4.4 diverse Trainingsverfahren zur Lösung des Optimierungsproblems aus Gleichung (4.4) vorgestellt. Zur Einforderung von physikalischen Nebenbedingungen und zur Regularisierung des Trainings werden in Abschnitt 4.5 die dazu nötigen Methoden eingeführt. Abschließend folgt ein numerisches Beispiel in Abschnitt 4.6. Alle nachfolgend beschriebenen Verfahren und die zugehörigen Algorithmen sind in MATLAB [170] implementiert und werden im restlichen Verlauf der Arbeit zum Training von KNN verwendet.

³⁷ Durch mächtige KNN-Bibliotheken, wie bspw. TensorFlow [169], in Verbindung mit automatischer Differentiation, ist zur reinen Anwendung von KNN eine ausführliche Einarbeitung in deren Theorie und Implementierung heutzutage nicht mehr notwendig. Geht man diesen Schritt dennoch, gewinnt man aus Sicht des Autors tiefe Einblicke in ein spannendes und vielschichtiges Teilgebiet der Numerik. Darüber hinaus ist man in der Anwendung im Rahmen ingenieurtechnischer Fragestellungen sattelfester im Umgang mit diesen oft fälschlicherweise als *black box* bezeichneten Algorithmen.

4.1 Normalisierung und Daten-Fehlerfunktion

Die Datenpaare (\mathbf{x}, \mathbf{t}) werden zunächst jeweils linear und unabhängig voneinander transformiert bzw. normalisiert.

$$\hat{x}_i = (x_i - m_{xi})/s_{xi} \quad , \quad i = 1, \dots, n_i \quad (4.7)$$

$$\hat{t}_j = (t_j - m_{zj})/s_{zj} \quad , \quad j = 1, \dots, n_o \quad (4.8)$$

Die Transformationsparameter werden in dieser Arbeit als die Mittelwerte

$$m_{xi} = \frac{1}{P_T} \sum_{k=1}^{P_T} x_{ik} \quad \text{und} \quad m_{zj} := \frac{1}{P_T} \sum_{k=1}^{P_T} t_{jk} \quad (4.9)$$

sowie die (korrigierten) Standardabweichungen

$$s_{xi} := \sqrt{\frac{1}{P_T - 1} \sum_{k=1}^{P_T} (x_{ik} - m_{xi})^2} \quad \text{und} \quad s_{zj} := \sqrt{\frac{1}{P_T - 1} \sum_{k=1}^{P_T} (t_{jk} - m_{zj})^2} \quad (4.10)$$

der jeweiligen Eingangs- und Ausgangsvariablen berechnet. Diese Parameter werden jeweils in die Spaltenvektoren und Diagonalmatrizen

$$\mathbf{m}_x := [m_{x1}, \dots, m_{xn_i}]^T \in \mathbb{R}^{n_i} \quad \mathbf{S}_x := \text{diag}(s_{x1}, \dots, s_{xn_i}) \in \mathbb{R}^{n_i \times n_i} \quad (4.11)$$

$$\mathbf{m}_z := [m_{z1}, \dots, m_{zn_o}]^T \in \mathbb{R}^{n_o} \quad \mathbf{S}_z := \text{diag}(s_{z1}, \dots, s_{zn_o}) \in \mathbb{R}^{n_o \times n_o} \quad (4.12)$$

einsortiert. Nicht nur die Trainingsdaten, sondern auch die KNN Eingangs- und Ausgangsvektoren \mathbf{x} und \mathbf{z} werden nach den Gleichungen (4.7) und (4.8) transformiert. Daher schließt die KNN-Abbildungsvorschrift (4.1) stets die Transformation der Eingangsvariablen

$$\hat{\mathbf{x}} = \mathbf{S}_x^{-1}(\mathbf{x} - \mathbf{m}_x) \quad (4.13)$$

und die Rücktransformation der Ausgangsvariablen

$$\mathbf{z} = \mathbf{m}_z + \mathbf{S}_z \hat{\mathbf{z}} \quad (4.14)$$

mit ein.

Um die gewünschte Approximation (4.3) der gegebenen Trainingsdaten zu erzielen, entspricht die zu minimierende Fehlerfunktion $\mathcal{L}(\mathbf{w})$ klassischerweise einem über alle Trainingspaare der Menge T gemittelten *Trainingsdatenfehler*

$$\begin{aligned} \mathcal{L}_T(\mathbf{w}) &:= \frac{1}{P_T} \sum_{k=1}^{P_T} \mathcal{L}_{Tk}(\mathbf{w}) = \frac{1}{P_T} \sum_{k=1}^{P_T} \frac{1}{2} \sum_{o=1}^{n_o} \frac{1}{s_{zo}^2} (z_o(\mathbf{x}_k, \mathbf{w}) - t_{ok})^2 \\ &= \frac{1}{P_T} \sum_{k=1}^{P_T} \frac{1}{2} \sum_{o=1}^{n_o} (\hat{z}_o(\mathbf{x}_k, \mathbf{w}) - \hat{t}_{ok})^2 . \end{aligned} \quad (4.15)$$

Der auf ein einzelnes Trainingspaar $(\mathbf{x}_k, \mathbf{t}_k)$ bezogene Fehlerterm \mathcal{L}_{T_k} ist ein Maß für die Diskrepanz zwischen Zielwert \mathbf{t}_k und KNN-Vorhersage $\mathbf{z}(\mathbf{x}_k, \mathbf{w})$, vgl. Abbildung 4.2. In dieser Arbeit wird dieser Fehlerterm über den normalisierten MSE³⁸ (engl.: *Mean Squared Error*) berechnet. Die Normalisierung der Eingangs- und Ausgangsvariablen des KNN sowie der Fehlerterme \mathcal{L}_{T_k} im Minimierungsproblem haben die folgenden wichtigen Vorteile (vgl. LE CUN ET AL. [110]):

- Alle Ausgangsvariablen z_j gehen gleichwertig in die Fehlerfunktion \mathcal{L} ein. Unterschiedliche physikalische Variablen können sich allein schon durch ihre Einheiten um Magnituden unterscheiden. Sind sie normalisiert, werden sie innerhalb des Trainingsprozesses jedoch gleichwertig minimiert.
- Die Wertebereiche der Gewichte und der Fehlerfunktion \mathcal{L} variieren nicht erheblich zwischen verschiedenen Problemstellungen. Daher müssen die Hyperparameter der Trainingsverfahren nicht individuell eingestellt werden. So werden bspw. bei der plastischen KNN-Materialmodellierung Verzerrungen und Spannungen Teil des Eingangsvektors sein. Diese würden, abhängig von der Wahl der Einheiten, immer zu anderen Wertebereichen für die Gewichte führen.

4.2 Das klassische vollverknüpfte Feedforward-Netz

Das klassische vollverknüpfte Feedforward-Netz (FNN), auch *Multilayer Perceptron* (MLP) genannt, besteht aus sortierten Neuronenschichten. Es ist in Abbildung 4.3 dargestellt. Ausgehend von der Eingabeschicht $L = 0$ werden Informationen über n_h Zwischenschichten $L = 1, \dots, n_h$ bis zur Ausgabeschicht $L = n_h + 1$ weitergegeben. In jeder Schicht L befinden sich n_L Neuronen. Das FNN ist *vollverknüpft*, weil jedes Neuron einer Schicht mit jedem Neuron der nächsten Schicht verbunden ist. Eine bestimmte FNN-Topologie wird mit $[n_i - n_1 - \dots - n_{n_h} - n_o]$ bezeichnet.

Zentraler Bestandteil des FNN ist das Neuron, welches Informationen aus der vorherigen Schicht verarbeitet und an die Neuronen der folgenden Schicht weitergibt. Jedes Neuron wird dabei nach seinem skalaren Ausgabewert benannt. Die Eingangsneuronen in der Schicht $L = 0$ geben jeweils die transformierten Eingangsvariablen $y_i^{[0]} = \hat{x}_i$ weiter. Die Ausgangsneuronen geben die transformierten Ausgangsvariablen $y_j^{[n_h+1]} = \hat{z}_j$ aus. Innerhalb des FNN ist das Neuron $y_m^{[L]}$ mit allen Neuronen $y_l^{[L-1]}$ der vorherigen Schicht $L - 1$ über die mit $w_{ml}^{[L]}$ gewichteten

³⁸ Alternativ kann bspw. auch der MAE (engl.: *Mean Absolute Error*) verwendet werden, der, im Vergleich zum MSE, Ausreißer durch das fehlende Quadrieren nicht zu stark gewichtet.

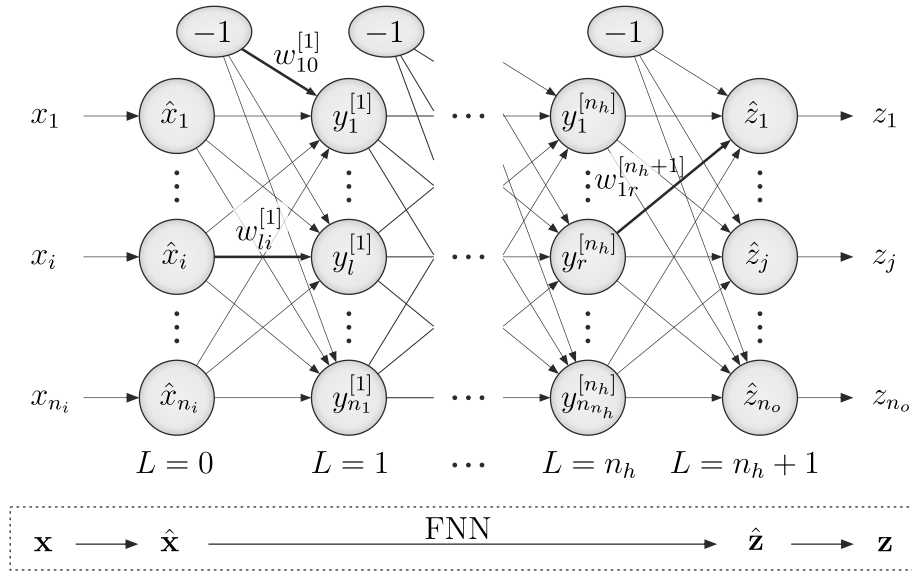


Abbildung 4.3: Informationsverarbeitung und Topologie eines FNN

Verbindungen verknüpft. Der erste Index bezieht sich dabei auf das Zielneuron, der zweite auf das Startneuron. Alle Gewichte des FNN werden im Spaltenvektor

$$\mathbf{w} = [w_{10}^{[1]}, \dots, w_{1n_i}^{[1]}, w_{20}^{[1]}, \dots, w_{10}^{[2]}, \dots, w_{n_o n_{n_h}}^{[n_h+1]}]^T \in \mathbb{R}^{n_w} \quad (4.16)$$

gruppiert. Die Reihenfolge orientiert sich dabei zunächst an den Schichten und anschließend nach den Ziel- und Startneuronen. Die Sortierung innerhalb des Fehlergradienten

$$\nabla \mathcal{L} = \left[\frac{\partial \mathcal{L}}{\partial w_{10}^{[1]}}, \dots, \frac{\partial \mathcal{L}}{\partial w_{1n_i}^{[1]}}, \frac{\partial \mathcal{L}}{\partial w_{20}^{[1]}}, \dots, \frac{\partial \mathcal{L}}{\partial w_{10}^{[2]}}, \dots, \frac{\partial \mathcal{L}}{\partial w_{n_o n_{n_h}}^{[n_h+1]}} \right]^T \in \mathbb{R}^{n_w} \quad (4.17)$$

von Gleichung (4.5) muss äquivalent sein.

4.2.1 Die FNN-Vorwärtsrechnung

In diesem Abschnitt wird die Berechnungsvorschrift $\mathbf{z} = \mathbf{f}^{\text{FNN}}(\mathbf{x}, \mathbf{w})$ der FNN-Ausgangsvariablen beschrieben. Dazu wird zunächst ein beliebiges Neuron $y_m^{[L]}$ innerhalb des FNN betrachtet. Die gewichtete Summe

$$s_m^{[L]} = \sum_{l=1}^{n_{L-1}} w_{ml}^{[L]} y_l^{[L-1]} - w_{m0}^{[L]} = \sum_{l=0}^{n_{L-1}} w_{ml}^{[L]} y_l^{[L-1]} \quad (4.18)$$

ist der skalare Eingangswert in das Neuron $y_m^{[L]}$. Der Schwellenwert $w_{m0}^{[L]}$ wird in Form eines zusätzlichen Gewichts zwischen dem Neuron $y_m^{[L]}$ und einem *Bias*-

Neuron berücksichtigt, welches stets den Wert $y_0^{[L-1]} \equiv -1$ ausgibt. Dabei ist das Bias-Neuron nicht mit der vorherigen Schicht verbunden. Die Neuronenausgabe

$$y_m^{[L]} = g(s_m^{[L]}) \quad (4.19)$$

wird mit der Aktivierungsfunktion $g(s)$ berechnet. Angelehnt an die Funktionsweise in biologischen neuronalen Netzen wurden in frühen Anwendungen Sprungfunktionen gewählt, die unter dem Schwellenwert $s_m^{[L]} < w_{m0}^{[L]}$ den Wert $g = 0$ (Neuron ist inaktiv) und darüber den Wert $g = 1$ (Neuron ist aktiv) zurückgeben. Ein differenzierbares Äquivalent ist der Tangens hyperbolicus³⁹, der mit den ersten beiden Ableitungen in Abbildung 4.4 gegeben ist. Er hat den Nachteil, dass

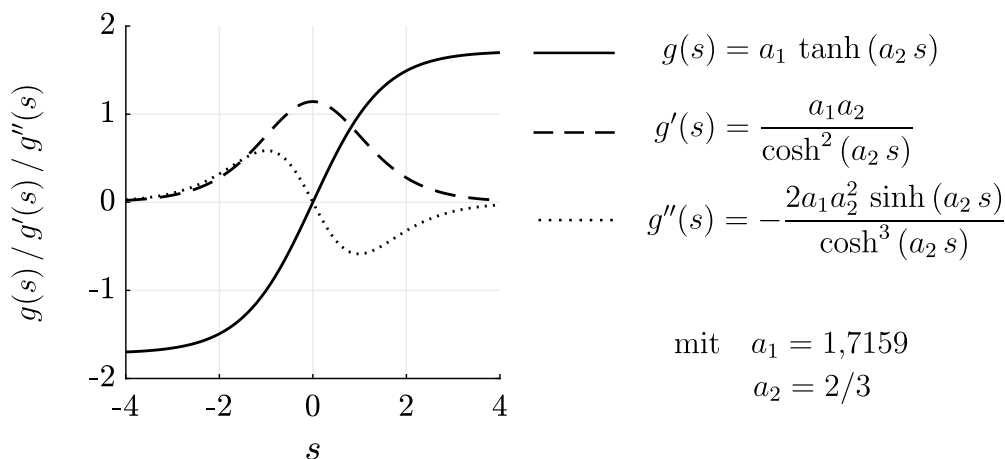


Abbildung 4.4: Die Funktionsgleichungen und Graphen der Aktivierungsfunktion Tangens hyperbolicus sowie ihrer ersten und zweiten Ableitung

auf beiden Seiten in den Plateaubereichen $s \gg 0$ und $s \ll 0$ die Ableitungen g' verschwinden, was letztlich zu sehr kleinen Werten im Gradienten $\nabla \mathcal{L}$ führt. Dies bremst das Training einzelner Gewichte. Jenes Verhalten ist auch als das Problem des verschwindenden Gradienten (engl.: *vanishing gradient problem*) bekannt. Der Effekt nimmt mit der Netztiefe zu. Die SoftPlus-Funktion in Abbildung 4.5 löst dieses Problem für $s > 0$. Die Funktionen lassen sich auch parametrisieren, was in Kombination mit zugeschnittenen Initialisierungen des Trainings zu einer besseren Konvergenz führen kann. Daher werden hier für den Tangens hyperbolicus die in Abbildung 4.4 dargestellten Zahlen von LE CUN ET AL. [110] verwendet. Die Aktivierungsfunktion der Ausgabeschicht wird mit $g_o(s)$ bezeichnet. Für das klassische FNN wird dazu die Identität gewählt mit

$$g_o(s) = s, \quad g'_o(s) = 1 \quad \text{und} \quad g''_o(s) = 0. \quad (4.20)$$

³⁹Die Symmetrie des Tangens hyperbolicus ist darüber hinaus beim Training vorteilhaft gegenüber der Sigmoid-Funktion, vgl. LECUN ET AL. [110].

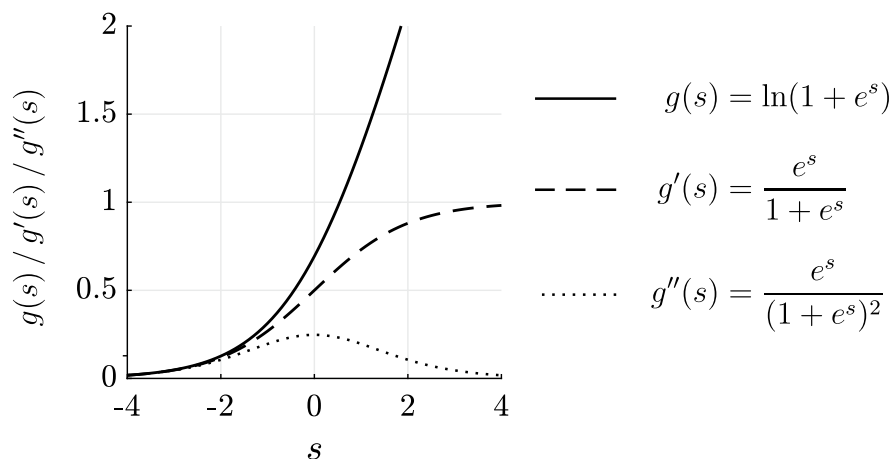


Abbildung 4.5: Die Funktionsgleichungen und Graphen der Aktivierungsfunktion SoftPlus sowie ihrer ersten und zweiten Ableitung

Weitere Aktivierungsfunktionen sind nicht Bestandteil der vorliegenden Arbeit. Mithilfe der bisher genannten Bausteine wird im Folgenden die vollständige Berechnung von $\mathbf{z} = \mathbf{f}^{\text{FNN}}(\mathbf{x}, \mathbf{w})$ für einen einzelnen Eingangsvektor \mathbf{x} in Vektor-Matrix-Notation gezeigt. Die Neuronenausgaben $y_m^{[L]}$, inklusive Bias-Wert $y_0 = -1$, und die gewichteten Summen $s_m^{[L]}$ werden jeweils in den Spaltenvektoren

$$\mathbf{y}^{[L]} := [-1, y_1^{[L]}, \dots, y_m^{[L]}, \dots, y_{n_L}^{[L]}]^T \in \mathbb{R}^{n_L+1} \quad \text{und} \quad (4.21)$$

$$\mathbf{s}^{[L]} := [s_1^{[L]}, \dots, s_m^{[L]}, \dots, s_{n_L}^{[L]}]^T \in \mathbb{R}^{n_L} \quad (4.22)$$

für die Schicht L zusammengefasst. Ohne Biaswert $y_0 = -1$ an erster Stelle wird der Vektor $\mathbf{y}^{[L]}$ mit einem Querstrich gekennzeichnet: $\bar{\mathbf{y}}^{[L]} \in \mathbb{R}^{n_L}$. Der transformierte Eingangsvektor $\hat{\mathbf{x}} = \bar{\mathbf{y}}^{[0]}$ ist somit Teil von $\mathbf{y}^{[0]} = [-1, \hat{\mathbf{x}}^T]^T$. Der transformierte Ausgangsvektor $\hat{\mathbf{z}}$ ist gleichbedeutend mit der Ausgabe der letzten Neuronschicht $\hat{\mathbf{z}} = \bar{\mathbf{y}}^{[n_h+1]}$. Die Gewichte werden für jede Schicht in der Matrix

$$\mathbf{W}^{[L]} := \begin{bmatrix} w_{10}^{[L]} & w_{11}^{[L]} & \dots & w_{1n_{L-1}}^{[L]} \\ w_{20}^{[L]} & w_{21}^{[L]} & \dots & w_{2n_{L-1}}^{[L]} \\ \vdots & & & \vdots \\ w_{n_L 0}^{[L]} & w_{n_L 1}^{[L]} & \dots & w_{n_L n_{L-1}}^{[L]} \end{bmatrix} \in \mathbb{R}^{n_L \times (n_{L-1}+1)} \quad (4.23)$$

gruppiert. Ohne die Biasgewichte w_{m0} in der ersten Spalte wird die Matrix $\mathbf{W}^{[L]}$ mit einem Querstrich gekennzeichnet: $\bar{\mathbf{W}}^{[L]} \in \mathbb{R}^{n_L \times n_{L-1}}$. Mithilfe dieser Definitionen lässt sich die Informationsübertragung von Schicht $L-1$ zu Schicht L kompakt mit

$$\mathbf{s}^{[L]} = \mathbf{W}^{[L]} \mathbf{y}^{[L-1]} \quad \text{und} \quad (4.24)$$

$$\bar{\mathbf{y}}^{[L]} = g(\mathbf{s}^{[L]}) \quad (4.25)$$

schreiben. Auch wenn sie für die Berechnung von $\mathbf{z} = \mathbf{f}^{\text{FNN}}(\mathbf{x}, \mathbf{w})$ nicht direkt benötigt werden, kann es für nachfolgende Algorithmen zweckmäßig sein, die Ableitungen

$$\mathbf{g}'^{[L]} := g'(\mathbf{s}^{[L]}) \quad \text{und} \quad (4.26)$$

$$\mathbf{g}''^{[L]} := g''(\mathbf{s}^{[L]}) \quad (4.27)$$

schichtweise abzuspeichern. Der vollständige Algorithmus zur Berechnung von $\mathbf{z} = \mathbf{f}^{\text{FNN}}(\mathbf{x}, \mathbf{w})$ für feste \mathbf{x} und \mathbf{w} ist in Tafel 4.1 beschrieben. Zur Vermeidung von numerischen Problemen empfiehlt es sich für die Aktivierungsfunktionen ab bspw. $|s| > 50$ die Grenzwerte direkt vorzugeben, bspw. $\tanh(s > 50) \approx a_1 = 1,7159$ oder $\ln(1 + e^s, s > 50) \approx s$.

Eingangstransformation	$\hat{\mathbf{x}} = \mathbf{S}_x^{-1}(\mathbf{x} - \mathbf{m}_x)$
Eingangsschicht	$\mathbf{y}^{[0]} = [-1, \hat{\mathbf{x}}^T]^T$
Schleife über Zwischenschichten	for $L = 1, \dots, n_h$ do $\mathbf{s}^{[L]} = \mathbf{W}^{[L]} \mathbf{y}^{[L-1]}, (\mathbf{g}'^{[L]}, \mathbf{g}''^{[L]})$ $\bar{\mathbf{y}}^{[L]} = g(\mathbf{s}^{[L]})$ $\mathbf{y}^{[L]} = [-1, \bar{\mathbf{y}}^{[L]T}]^T$ end for
Ausgangsschicht	$\mathbf{s}^{[n_h+1]} = \mathbf{W}^{[n_h+1]} \mathbf{y}^{[n_h]}$ $\hat{\mathbf{z}} = g_o(\mathbf{s}^{[n_h+1]})$
Ausgangstransformation	$\mathbf{z} = \mathbf{m}_z + \mathbf{S}_z \hat{\mathbf{z}}$

Tafel 4.1: FNN Vorwärtsberechnung für $\mathbf{z} = \mathbf{f}^{\text{FNN}}(\mathbf{x}, \mathbf{w})$

4.2.2 Die FNN-Jacobi-Matrix

Dieser Abschnitt beschreibt die Berechnung der FNN JACOBI-Matrix $\mathbf{j}(\mathbf{x}, \mathbf{w})$ nach Definition (4.6). Sie wird im Rahmen der KNN-Materialmodellierung hauptsächlich zur Berechnung der Materialtangente \mathbf{C}_T benötigt. Mit den Matrizen \mathbf{S}_x und \mathbf{S}_z aus (4.11) und (4.12) gilt zunächst die Transformationsbeziehung

$$\mathbf{j} = \frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \mathbf{S}_z \left(\frac{\partial \hat{\mathbf{z}}}{\partial \hat{\mathbf{x}}} \right) \mathbf{S}_x^{-1} = \mathbf{S}_z \hat{\mathbf{j}} \mathbf{S}_x^{-1}. \quad (4.28)$$

Die Ableitungen der transformierten Ausgangsvariablen nach den transformierten Eingangsvariablen folgen unter Beachtung der Kettenregel mit

$$\hat{\mathbf{j}} = \frac{\partial \hat{\mathbf{z}}}{\partial \hat{\mathbf{x}}} = \frac{\partial \bar{\mathbf{y}}^{[n_h+1]}}{\partial \bar{\mathbf{y}}^{[0]}} = \prod_{L=n_h+1}^1 \frac{\partial \bar{\mathbf{y}}^{[L]}}{\partial \bar{\mathbf{y}}^{[L-1]}}. \quad (4.29)$$

Die Matrix-Multiplikation ist nicht kommutativ und muss daher von der letzten Schicht zur ersten erfolgen. Innerhalb einer Schicht gilt durch die Definition der gewichteten Summe (4.18) für eine einzelne partielle Ableitung

$$\frac{\partial y_m^{[L]}}{\partial y_l^{[L-1]}} = g'(s_m^{[L]}) w_{ml}^{[L]} . \quad (4.30)$$

Mit einem Spaltenvektor bestehend aus Einsen

$$\mathbf{1}^m := [1, 1, \dots, 1]^T \in \mathbb{R}^m \quad (4.31)$$

sowie dem Vektor der Ableitungen der Aktivierungsfunktion $\mathbf{g}'^{[L]}$ nach Definition (4.26) lassen sich die Matrizen aller partiellen Ableitungen zweier aufeinanderfolgender Schichten kompakt schreiben mit

$$\frac{\partial \bar{\mathbf{y}}^{[L]}}{\partial \bar{\mathbf{y}}^{[L-1]}} = \begin{bmatrix} g'(s_1^{[L]}) & \dots & g'(s_1^{[L]}) \\ \vdots & \dots & \vdots \\ g'(s_{n_L}^{[L]}) & \dots & g'(s_{n_L}^{[L]}) \end{bmatrix} \circ \bar{\mathbf{W}}^{[L]} = \left(\mathbf{g}'^{[L]} \otimes \mathbf{1}^{n_{L-1}} \right) \circ \bar{\mathbf{W}}^{[L]} . \quad (4.32)$$

Die mit \circ gekennzeichnete Multiplikation wird komponentenweise ausgeführt und wird auch HADAMARD-Produkt genannt. Die Vektoren $\mathbf{g}'^{[L]}$ müssen dazu für den Eingangsvektor \mathbf{x} in der Vorwärtsberechnung nach Tafel 4.1 im Vorfeld berechnet und abgespeichert worden sein. Daraus, und mit den Gleichungen (4.28) und (4.29), resultiert mit

$$\mathbf{j}(\mathbf{x}, \mathbf{w}) = \mathbf{S}_z \left[\prod_{L=n_h+1}^1 \left(\mathbf{g}'^{[L]} \otimes \mathbf{1}^{n_{L-1}} \right) \circ \bar{\mathbf{W}}^{[L]} \right] \mathbf{S}_x^{-1} \quad (4.33)$$

eine kompakte Berechnungsvorschrift für die vollständige JACOBI-Matrix \mathbf{j} . Diese kann mithilfe des Algorithmus in Tafel 4.1 zur Berechnung aller partiellen Ableitungen für einen bestimmten Eingangsvektor \mathbf{x} verwendet werden. Eine Erweiterung auf mehrere Eingangsvektoren gleichzeitig ist allerdings nicht zweckmäßig.

Insbesondere für das Training mit Nebenbedingungen werden aus der JACOBI-Matrix \mathbf{j} nur einzelne Spalten $\mathbf{z}_{,i} = \partial \mathbf{z} / \partial x_i$ mit $i \in \mathbb{I}$ benötigt. Die Menge \mathbb{I} enthält dabei alle Indizes der Eingangsvariablen, nach denen mindestens einmal abgeleitet werden muss. Außerdem müssen im Rahmen der ICNN-Architektur aus dem folgenden Abschnitt 4.3.1 die Ableitungen für viele Datenpunkte \mathbf{x} gleichzeitig berechnet werden.

Um diesen Herausforderungen gewachsen zu sein, werden im Folgenden zugehörige Berechnungsvorschriften vorgestellt, die im Gegensatz zu Gleichung (4.33) in den Schichten *vorwärts* erfolgen und sich folglich mit dem Algorithmus aus

Tafel 4.1 verknüpfen lassen. Mit den Transformationen (4.7) und (4.8) wird eine einzelne partielle Ableitung umgeformt zu

$$z_{j,i} = \frac{\partial z_j}{\partial x_i} = \left(\frac{s_{zj}}{s_{xi}} \right) \frac{\partial \hat{z}_j}{\partial \hat{x}_i} = \left(\frac{s_{zj}}{s_{xi}} \right) \hat{z}_{j,i} . \quad (4.34)$$

Die Ableitungen der Neuronenausgaben nach den transformierten Eingangsvariablen $y_{m,i}^{[L]}$ lassen sich schichtweise berechnen mit

$$y_{m,i}^{[L]} = \frac{\partial y_m^{[L]}}{\partial \hat{x}_i} = \frac{\partial y_m^{[L]}}{\partial s_m^{[L]}} \frac{\partial s_m^{[L]}}{\partial \hat{x}_i} = g'(s_m^{[L]}) \sum_{l=1}^{n_{L-1}} w_{ml}^{[L]} y_{l,i}^{[L-1]} . \quad (4.35)$$

Dabei wurde die Kettenregel in Verbindung mit der Definition der gewichteten Summe (4.18) genutzt. Die Ableitungen der einzelnen Größen können abermals in den Spaltenvektoren

$$\bar{\mathbf{y}}_{,i}^{[L]} := [y_{1,i}^{[L]}, \dots, y_{m,i}^{[L]}, \dots, y_{n_L,i}^{[L]}]^T \in \mathbb{R}^{n_L} \quad \text{und} \quad (4.36)$$

$$\mathbf{s}_{,i}^{[L]} := [s_{1,i}^{[L]}, \dots, s_{m,i}^{[L]}, \dots, s_{n_L,i}^{[L]}]^T \in \mathbb{R}^{n_L} \quad (4.37)$$

zusammengefasst werden. Für die Ableitungen der gewichteten Summen

$$\mathbf{s}_{,i}^{[L]} = \frac{\partial \mathbf{s}^{[L]}}{\partial \hat{x}_i} = \begin{cases} \bar{\mathbf{W}}_{(:,i)}^{[1]} & (i\text{-te Spalte von } \bar{\mathbf{W}}^{[1]}) \quad L = 1 \\ \bar{\mathbf{W}}^{[L]} \bar{\mathbf{y}}_{,i}^{[L-1]} & L = 2, \dots, n_h + 1 \end{cases} \quad (4.38)$$

muss zwischen der ersten Zwischenschicht und allen weiteren Zwischenschichten unterschieden werden. Daraus folgt für die Ableitungen der Neuronenausgaben

$$\bar{\mathbf{y}}_{,i}^{[L]} = \mathbf{g}'^{[L]} \circ \mathbf{s}_{,i}^{[L]} . \quad (4.39)$$

In der Ausgangsschicht muss \mathbf{g}'_o genutzt werden, und es gilt $\bar{\mathbf{y}}_{,i}^{[n_h+1]} = \hat{\mathbf{z}}_{,i}$. Der Algorithmus zur Berechnung der Ableitungen $\mathbf{z}_{,i} = \partial \mathbf{z} / \partial x_i$ ist in Tafel 4.2 gegeben.

Initialisierung	$\mathbf{s}_{,i}^{[1]} = \bar{\mathbf{W}}_{(:,i)}^{[1]}$ (i -te Spalte von $\bar{\mathbf{W}}^{[1]}$)
Schleife über Zwischenschichten	for $L = 1, \dots, n_h$ do $\bar{\mathbf{y}}_{,i}^{[L]} = \mathbf{g}'^{[L]} \circ \mathbf{s}_{,i}^{[L]}$ $\mathbf{s}_{,i}^{[L+1]} = \bar{\mathbf{W}}^{[L+1]} \bar{\mathbf{y}}_{,i}^{[L]}$ end for
Ausgangsschicht	$\hat{\mathbf{z}}_{,i} = \partial \hat{\mathbf{z}} / \partial \hat{x}_i = \mathbf{g}'_o \circ \mathbf{s}_{,i}^{[n_h+1]}$
Ausgangstransformation	$\mathbf{z}_{,i} = (1/s_{xi}) \mathbf{S}_z \hat{\mathbf{z}}_{,i}$

Tafel 4.2: FNN Vorwärtsberechnung für die Ableitungen $\mathbf{z}_{,i} = \partial \mathbf{z} / \partial x_i$

4.2.3 Erweiterte FNN-Vorwärtsberechnungen für alle Trainingsdaten

Für das Training des FNN müssen i. d. R. sowohl die Fehlerfunktion \mathcal{L} als auch ihr Gradient $\nabla\mathcal{L}$ für eine größere Anzahl Daten ($P \gg 1$) aufgestellt werden. Dazu folgt zunächst eine erweiterte FNN-Vorwärtsrechnung in Form eines Algorithmus in Matrixnotation, der die Schleife über eine Menge Datenpunkte $\{\mathbf{x}_k\}_{k=1}^P$ in die Matrixoperationen eingliedert. Die Eingangs- und Ausgangsvariablen \mathbf{x} und \mathbf{z} , ihre transformierten Äquivalente $\hat{\mathbf{x}}$ und $\hat{\mathbf{z}}$ sowie die Transformationsvektoren \mathbf{m}_x und \mathbf{m}_z des FNN werden dazu in den Matrizen

$$\mathbf{X} := [\mathbf{x}_1, \dots, \mathbf{x}_k, \dots, \mathbf{x}_P] \in \mathbb{R}^{n_i \times P} \quad \mathbf{Z} := [\mathbf{z}_1, \dots, \mathbf{z}_k, \dots, \mathbf{z}_P] \in \mathbb{R}^{n_o \times P} \quad (4.40)$$

$$\hat{\mathbf{X}} := [\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_k, \dots, \hat{\mathbf{x}}_P] \in \mathbb{R}^{n_i \times P} \quad \hat{\mathbf{Z}} := [\hat{\mathbf{z}}_1, \dots, \hat{\mathbf{z}}_k, \dots, \hat{\mathbf{z}}_P] \in \mathbb{R}^{n_o \times P} \quad (4.41)$$

$$\mathbf{M}_x := [\mathbf{m}_x, \dots, \mathbf{m}_x] \in \mathbb{R}^{n_i \times P} \quad \mathbf{M}_z := [\mathbf{m}_z, \dots, \mathbf{m}_z] \in \mathbb{R}^{n_o \times P} \quad (4.42)$$

zusammengefasst. Die mit den Eingangsvektoren \mathbf{x}_k berechneten Vektoren $\mathbf{y}_k^{[L]}$, $\mathbf{s}_k^{[L]}$, $\mathbf{g}_k^{[L]}$, $\mathbf{g}_k^{\prime[L]}$, $\mathbf{g}_k^{\prime\prime[L]}$, $\bar{\mathbf{y}}_{k,i}^{[L]}$ und $\mathbf{s}_{k,i}^{[L]}$ aus den Definitionen (4.21), (4.22), (4.26), (4.27), (4.36) und (4.37) werden für alle P Datenpunkte schichtweise in die Matrizen

$$\mathbf{Y}^{[L]} := [\mathbf{y}_1^{[L]}, \dots, \mathbf{y}_k^{[L]}, \dots, \mathbf{y}_P^{[L]}] \in \mathbb{R}^{(n_L+1) \times P} \quad (4.43)$$

$$\mathbf{S}^{[L]} := [\mathbf{s}_1^{[L]}, \dots, \mathbf{s}_k^{[L]}, \dots, \mathbf{s}_P^{[L]}] \in \mathbb{R}^{n_L \times P} \quad (4.44)$$

$$\mathbf{G}^{\prime[L]} := [\mathbf{g}_1^{\prime[L]}, \dots, \mathbf{g}_k^{\prime[L]}, \dots, \mathbf{g}_P^{\prime[L]}] \in \mathbb{R}^{n_L \times P} \quad (4.45)$$

$$\mathbf{G}^{\prime\prime[L]} := [\mathbf{g}_1^{\prime\prime[L]}, \dots, \mathbf{g}_k^{\prime\prime[L]}, \dots, \mathbf{g}_P^{\prime\prime[L]}] \in \mathbb{R}^{n_L \times P} \quad (4.46)$$

$$\bar{\mathbf{Y}}_{,i}^{[L]} := [\bar{\mathbf{y}}_{1,i}^{[L]}, \dots, \bar{\mathbf{y}}_{k,i}^{[L]}, \dots, \bar{\mathbf{y}}_{P,i}^{[L]}] \in \mathbb{R}^{n_L \times P} \quad (4.47)$$

$$\mathbf{S}_{,i}^{[L]} := [\mathbf{s}_{1,i}^{[L]}, \dots, \mathbf{s}_{k,i}^{[L]}, \dots, \mathbf{s}_{P,i}^{[L]}] \in \mathbb{R}^{n_L \times P} \quad (4.48)$$

einsortiert. Die Matrix $\mathbf{Y}^{[L]}$ setzt sich dabei aus der ersten Zeile $(-\mathbf{1}^P)^T$, vgl. Definition (4.31), und $\bar{\mathbf{Y}}^{[L]}$ zusammen. Die Vorwärtsberechnung der FNN-Aufgabevektoren \mathbf{z}_k für alle Datenpunkte kann folglich mit den Matrixoperationen

$$\mathbf{S}^{[L]} = \mathbf{W}^{[L]} \mathbf{Y}^{[L-1]} \quad (4.49)$$

und

$$\mathbf{Y}^{[L]} = [-\mathbf{1}^P, g(\mathbf{S}^{[L]})^T]^T \quad (4.50)$$

schichtweise durchgeführt werden. Die Berechnungsvorschrift für die partiellen Ableitungen $\partial\mathbf{z}/\partial x_i$ aus Gleichung (4.39) kann mit den gegebenen Matrixdefinitionen ebenfalls für alle Daten folgendermaßen zusammengefasst werden:

$$\bar{\mathbf{Y}}_{,i}^{[L]} = \mathbf{G}^{\prime[L]} \circ \mathbf{S}_{,i}^{[L]} \quad (4.51)$$

$$\mathbf{S}_{,i}^{[L+1]} = \bar{\mathbf{W}}^{[L+1]} \bar{\mathbf{Y}}_{,i}^{[L]} \quad (4.52)$$

Auf dieser Herleitung aufbauend, lässt sich der erweiterte Algorithmus in Tafel 4.3 angeben, der für alle $k = 1, \dots, P$ Datenpunkte \mathbf{x}_k die FNN-Ausgabevektoren \mathbf{z}_k und die nötigen Ableitungen $\partial\mathbf{z}_k/\partial x_i$ mit $i \in \mathbb{I}$ berechnet.

Eingangstransformation	$\hat{\mathbf{X}} = \mathbf{S}_x^{-1}(\mathbf{X} - \mathbf{M}_x)$
Initialisierung	$\mathbf{Y}^{[0]} = [-\mathbf{1}^{[L]}, \hat{\mathbf{X}}]^T$ $\mathbf{s}_{:,i}^{[1]} \stackrel{*}{=} \bar{\mathbf{W}}_{(:,i)}^{[1]}, \forall i \in \mathbb{I}$
Schleife über Zwischenschichten	for $L = 1, \dots, n_h$ do $\mathbf{S}^{[L]} = \mathbf{W}^{[L]} \mathbf{Y}^{[L-1]}$ $\mathbf{Y}^{[L]} = [-\mathbf{1}^P, g(\mathbf{S}^{[L]})^T]^T$ $\mathbf{G}^{[L]} = g'(\mathbf{S}^{[L]})$ und $\mathbf{G}''^{[L]} \stackrel{*}{=} g''(\mathbf{S}^{[L]})$
Schleife über Eingangsvariablen	for $i \in \mathbb{I}$ do $\bar{\mathbf{Y}}_{:,i}^{[L]} \stackrel{*}{=} \mathbf{G}'^{[L]} \circ \mathbf{S}_{:,i}^{[L]}$ $\mathbf{S}_{:,i}^{[L+1]} \stackrel{*}{=} \bar{\mathbf{W}}^{[L+1]} \bar{\mathbf{Y}}_{:,i}^{[L]}$ end for
	end for
Ausgangsschicht	$\mathbf{S}^{[n_h+1]} = \mathbf{W}^{[n_h+1]} \mathbf{Y}^{[n_h]}$ $\hat{\mathbf{Z}} = g_o(\mathbf{S}^{[n_h+1]})$ $\hat{\mathbf{Z}}_{:,i} \stackrel{*}{=} \mathbf{G}'_o \circ \mathbf{S}_{:,i}^{[n_h+1]}, \forall i \in \mathbb{I}$
Ausgangstransformation	$\mathbf{Z}_k = \mathbf{S}_z \hat{\mathbf{Z}} + \mathbf{M}_z$ $\mathbf{Z}_{:,i} \stackrel{*}{=} (1/s_{xi}) \mathbf{S}_z \hat{\mathbf{Z}}_{:,i}, \forall i \in \mathbb{I}$

Tafel 4.3: Erweiterte FNN-Vorwärtsberechnung für \mathbf{z}_k und $\mathbf{z}_{k,i}$ mit $i \in \mathbb{I}$ für alle Datenpunkte \mathbf{x}_k , $k = 1, \dots, P$. Die mit * markierten Operationen sind nur für den Fehlergradient mit Ableitungen aus Abschnitt 4.2.5 notwendig.

4.2.4 Der Fehlergradient bezüglich FNN-Ausgangsvariablen

Im Folgenden wird die Berechnung eines Fehlergradienten $\nabla \mathcal{L}(\mathbf{z}(\mathbf{w}))$ beschrieben, der von KNN-Ausgängen \mathbf{z} abhängig ist. Klassischerweise ist das der Gradient des Trainingsdatenfehlers \mathcal{L}_T aus Gleichung (4.15). In den folgenden Kapiteln werden in Zusammenhang mit physikalischen Nebenbedingungen jedoch auch andere Fehlerterme definiert, die eine ähnliche Struktur haben. Daher wird hier zunächst eine allgemein von \mathbf{z} abhängige Fehlerfunktion

$$\mathcal{L}(\mathbf{w}) = \frac{1}{P} \sum_{k=1}^P \mathcal{L}_k = \frac{1}{P} \sum_{k=1}^P \tilde{\mathcal{L}}(\mathbf{z}(\mathbf{x}_k, \mathbf{w})) \quad (4.53)$$

betrachtet, die auf den Datenpunkten $\{\mathbf{x}_k\}_{k=1}^P$ definiert ist. Der auf einen einzelnen Punkt definierte Fehlerterm $\tilde{\mathcal{L}}$ wird an gegebener Stelle spezifiziert. Im allgemeinen Fall kann der zugehörige Gradient in der Form

$$\nabla \mathcal{L}(\mathbf{w}) = \frac{1}{P} \sum_{k=1}^P \sum_{j=1}^{n_o} d_{jk} \frac{\partial \hat{z}_{jk}}{\partial \mathbf{w}} \quad (4.54)$$

geschrieben werden. Die Komponenten d_{jk} sind die partiellen Ableitungen des Fehlerterms $\tilde{\mathcal{L}}$ nach den transformierten Ausgangsvariablen, ausgewertet am Punkt \mathbf{x}_k . Auch sie werden im folgenden Verlauf der Arbeit spezifiziert und zunächst in der Matrix

$$\mathbf{D} := \begin{bmatrix} d_{11} & \dots & d_{1P} \\ \vdots & d_{jk} & \vdots \\ d_{n_o1} & \dots & d_{n_oP} \end{bmatrix} \in \mathbb{R}^{n_o \times P} \quad (4.55)$$

zusammengefasst. Im Fall des Trainingsdatenfehlers \mathcal{L}_T aus Gleichung (4.15) entsprechen sie gerade der Ableitung des normalisierten MSE

$$d_{Tjk} = \frac{\partial \mathcal{L}_{Tk}}{\partial \hat{z}_j} = \hat{z}_j(\mathbf{x}_k, \mathbf{w}) - \hat{t}_{jk} . \quad (4.56)$$

Für die Ableitungen $\partial \hat{z} / \partial \mathbf{w}$ wird der klassische Backpropagation-Algorithmus (BP) genutzt, der nachfolgend beschrieben wird. Wie eingangs erwähnt, geht dieser für KNN auf die Dissertation von WERBOS [191] zurück, wobei die Veröffentlichung von RUMELHART ET AL. [152] diesem Algorithmus zur heutigen Popularität verholfen hat.

Zu Beginn ist eine Vorwärtsberechnung nach Tafel 4.3 für alle Datenpunkte $\{\mathbf{x}_k\}_{k=1}^P$ notwendig. Dabei werden die Größen \mathbf{Z} , \mathbf{G}'_o , sowie $\{\mathbf{Y}^{[L]}\}_{L=0}^{n_h}$, $\{\mathbf{G}'^{[L]}\}_{L=1}^{n_h}$ abgespeichert. Mit den FNN-Ergebnissen \mathbf{Z} lassen sich, abhängig von der Definition der $\tilde{\mathcal{L}}$ -Terme, die Fehlerfunktion (4.53) und die Matrix \mathbf{D} berechnen. Zur Berechnung des Gradienten wird zunächst die einzelne Komponente

$$\nabla \mathcal{L}_{ml}^{[L]} = \frac{\partial \mathcal{L}}{\partial w_{ml}^{[L]}} = \frac{1}{P} \sum_{k=1}^P \frac{\partial \mathcal{L}_k}{\partial s_m^{[L]}} \frac{\partial s_{mk}^{[L]}}{\partial w_{ml}^{[L]}} = \frac{1}{P} \sum_{k=1}^P \frac{\partial \mathcal{L}_k}{\partial s_m^{[L]}} y_l^{[L-1]} \quad (4.57)$$

betrachtet, die mithilfe der Kettenregel und der Definition der gewichteten Summe (4.18) umgeformt wurde. Der Summenindex startet bei $l = 1$, da das Bias-Neuron keine Verbindung zur vorherigen Schicht hat. Die Gradientenkomponenten der Schicht L werden äquivalent zu den Gewichten in Definition (4.23) in der Matrix

$$\nabla \mathcal{L}^{[L]} := \begin{bmatrix} \nabla \mathcal{L}_{10}^{[L]} & \nabla \mathcal{L}_{11}^{[L]} & \dots & \nabla \mathcal{L}_{1n_{L-1}}^{[L]} \\ \nabla \mathcal{L}_{20}^{[L]} & \nabla \mathcal{L}_{21}^{[L]} & \dots & \nabla \mathcal{L}_{2n_{L-1}}^{[L]} \\ \vdots & & \nabla \mathcal{L}_{ml}^{[L]} & \vdots \\ \nabla \mathcal{L}_{n_L0}^{[L]} & \nabla \mathcal{L}_{n_L1}^{[L]} & \dots & \nabla \mathcal{L}_{n_L n_{L-1}}^{[L]} \end{bmatrix} \in \mathbb{R}^{n_L \times (n_{L-1} + 1)} \quad (4.58)$$

zusammengefasst. Die klassischen Delta-Werte, die *Fehlersignale* der Neuronen, werden über

$$\delta_{mk}^{[L]} := \frac{\partial \mathcal{L}_k}{\partial s_m^{[L]}} \quad (4.59)$$

definiert und für alle Datenpunkte und Neuronen in Schicht L in der Matrix

$$\boldsymbol{\delta}^{[L]} := \begin{bmatrix} \delta_{11}^{[L]} & \cdots & \delta_{1P}^{[L]} \\ \vdots & \delta_{mk}^{[L]} & \vdots \\ \delta_{n_L 1}^{[L]} & \cdots & \delta_{n_L P}^{[L]} \end{bmatrix} \in \mathbb{R}^{n_L \times P} \quad (4.60)$$

gruppiert. Mit diesen Definitionen lässt sich die Matrix der Gradiententerme (4.58) für alle Datenpunkte kompakt mit

$$\nabla \mathcal{L}^{[L]} = \frac{1}{P} \boldsymbol{\delta}^{[L]} \mathbf{Y}^{[L-1]T} \quad (4.61)$$

berechnen. Die übrigen Delta-Werte in $\boldsymbol{\delta}^{[L]}$ werden *rückwärts* berechnet, von der Ausgangsschicht ausgehend bis hin zur Eingangsschicht. Die *Startwerte* in der Ausgangsschicht

$$\delta_{mk}^{[n_h+1]} = \frac{\partial \mathcal{L}_k}{\partial s_m^{[n_h+1]}} = \frac{\partial \mathcal{L}_k}{\partial \hat{z}_m} \frac{\partial \hat{z}_{mk}}{\partial s_m^{[n_h+1]}} = d_{mk} g'_o(s_{mk}^{[n_h+1]}) \quad (4.62)$$

können direkt mithilfe der Matrix \mathbf{D} aus Definition (4.55) berechnet werden über

$$\boldsymbol{\delta}^{[n_h+1]} = \mathbf{G}'_o \circ \mathbf{D} . \quad (4.63)$$

Zur Herleitung der BP-Aktualisierungsvorschrift für die Delta-Werte innerhalb des KNN wird die neuronale Nachbarschaft eines beliebigen Neurons $y_m^{[L]}$ in Abbildung 4.6 betrachtet. Die Bezeichnungen der Indizes folgen dabei dem Sinn *links*, *mittig* und *rechts*.

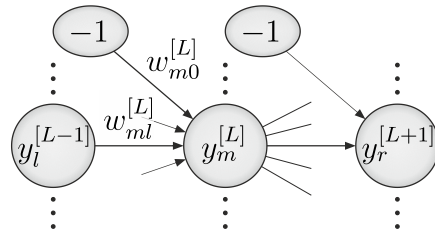


Abbildung 4.6: Neuronale Nachbarschaft des Neurons $y_m^{[L]}$ in Schicht L

Der Fehlerterm \mathcal{L}_k ist über die gewichteten Summen $s_{rk}^{[L+1]}$ mit $r = 1, \dots, n_{L+1}$ indirekt von der gewichteten Summe $s_{mk}^{[L]}$ abhängig:

$$\mathcal{L}_k(s_{mk}^{[L]}, \dots) = \mathcal{L}_k(s_{1k}^{[L+1]}(s_{mk}^{[L]}, \dots), \dots, s_{n_{L+1}k}^{[L+1]}(s_{mk}^{[L]}, \dots)) . \quad (4.64)$$

Unter Anwendung der Kettenregel

$$\delta_{mk}^{[L]} = \frac{\partial \mathcal{L}_k}{\partial s_m^{[L]}} = \sum_{r=1}^{n_{L+1}} \frac{\partial \mathcal{L}_k}{\partial s_r^{[L+1]}} \frac{\partial s_{rk}^{[L+1]}}{\partial s_m^{[L]}} \quad (4.65)$$

lässt sich mithilfe der Definition der gewichteten Summe (4.18) die Aktualisierungsvorschrift

$$\delta_{mk}^{[L]} = g'(s_{mk}^{[L]}) \sum_{r=1}^{n_{L+1}} \delta_{rk}^{[L+1]} w_{rm}^{[L+1]} \quad (4.66)$$

für die Delta-Werte herleiten. Mit den Definitionen der Matrizen (4.45), (4.60) und (4.23) ergibt sich diese Aktualisierungsvorschrift in Matrixnotation zu

$$\boldsymbol{\delta}^{[L]} = \mathbf{G}'^{[L]} \circ (\bar{\mathbf{W}}^{[L+1]T} \boldsymbol{\delta}^{[L+1]}) . \quad (4.67)$$

Die Berechnung des Gradienten $\nabla \mathcal{L}$ für alle Datenpunkte $\{\mathbf{x}_k\}_{k=1}^P$ ist in Tafel 4.4 zusammengefasst. Dabei werden die Terme des Gradienten $\nabla \mathcal{L}$ schichtweise über die Matrizen $\nabla \mathcal{L}^{[L]} \in \mathbb{R}^{n_L \times (n_{L-1}+1)}$ aus Definition (4.58) berechnet. Viele Optimierungsalgorithmen benötigen den Gradienten allerdings in der Vektorform $\nabla \mathcal{L} \in \mathbb{R}^{n_w}$ aus Definition (4.17). Daher ist im Anschluss an den Algorithmus eine manuelle Umsortierung notwendig.

Vorwärtsrechnung nach Tafel 4.3	für \mathbf{Z} , \mathbf{G}'_o und $\mathbf{Y}^{[L]}$, $\mathbf{G}'^{[L]}$, $\forall L$
Startwerte abh. von Fehlerterm	\mathbf{D}
Ausgangsschicht	$\boldsymbol{\delta}^{[n_h+1]} = \mathbf{G}'_o \circ \mathbf{D}$ $\nabla \mathcal{L}^{[n_h+1]} = (1/P) \boldsymbol{\delta}^{[n_h+1]} \mathbf{Y}^{[n_h]T}$
Schleife über restliche Schichten	for $L = n_h, \dots, 1$ do $\boldsymbol{\delta}^{[L]} = \mathbf{G}'^{[L]} \circ (\bar{\mathbf{W}}^{[L+1]T} \boldsymbol{\delta}^{[L+1]})$ $\nabla \mathcal{L}^{[L]} = (1/P) \boldsymbol{\delta}^{[L]} \mathbf{Y}^{[L-1]T}$ end for
Sortierung von Matrizen in Vektor	$\nabla \mathcal{L}^{[L]} \rightarrow \nabla \mathcal{L}, \forall L$

Tafel 4.4: Rückwärtsrechnung für Gradienten $\nabla \mathcal{L}$ von FNN-Ausgängen \mathbf{z}

4.2.5 Der Fehlergradient bezüglich FNN-Ableitungen

Im Folgenden wird der Backpropagation-Algorithmus zur Berechnung eines Fehlergradienten $\nabla \mathcal{L}(\mathbf{j}(\mathbf{w}))$ beschrieben, der von den Ableitungen der KNN-Ausgänge \mathbf{z} nach den KNN-Eingängen \mathbf{x} abhängig ist. Das kann bspw. dann der Fall sein, wenn im Rahmen der KNN-Materialmodellierung physikalische Nebenbedingungen innerhalb des Trainingsprozesses eingefordert werden. Der Algorithmus geht auf die Arbeiten von BISHOP [16] zurück, wobei er hier in Teilen abgewandelt wurde. Nicht immer sind alle möglichen Ableitungen $z_{j,i}$ aus \mathbf{j} im Fehlerterm enthalten. Daher ist es zweckmäßig die Indizes der Ausgangsvariablen nach denen

der Eingangsvariablen zu sortieren. In der Menge \mathbb{I} sind alle Indizes i der Eingangsvariablen x_i enthalten, nach denen innerhalb des Fehlerterms mindestens einmal abgeleitet wird. In den Mengen $\{\mathbb{J}(i)\}_{i \in \mathbb{I}}$ sind alle Indizes j enthalten, deren Ausgangsvariablen nach x_i abgeleitet werden. Der Fehlerterm $\mathcal{L}(\mathbf{w})$ selbst ist stark abhängig von der jeweiligen Anwendung. Der Gradient des Fehlerterms lässt sich allerdings stets in der Form

$$\nabla \mathcal{L} = \frac{1}{P} \sum_{i \in \mathbb{I}} \sum_{k=1}^P \sum_{j \in \mathbb{J}(i)} f_{ijk} \frac{\partial}{\partial \mathbf{w}} \left(\frac{\partial \hat{z}_{jk}}{\partial \hat{x}_i} \right) \quad (4.68)$$

angeben. Die Komponenten f_{ijk} hängen vom spezifischen Fehlerterm und der Variablentransformation ab. Sie werden an gegebener Stelle spezifiziert und zunächst in den Matrizen

$$\mathbf{F}_i := \begin{bmatrix} f_{i11} & \cdots & f_{i1P} \\ \vdots & f_{ijk} & \vdots \\ f_{in_01} & \cdots & f_{in_0P} \end{bmatrix} \in \mathbb{R}^{n_o \times P}, \quad \forall i \in \mathbb{I} \quad (4.69)$$

zusammengefasst. Zu Beginn ist eine Vorwärtsberechnung nach Tafel 4.3 für alle Datenpunkte \mathbf{x}_k , mit $k = 1, \dots, P$, notwendig. Dabei werden die Größen \mathbf{Z} , \mathbf{G}'_o , \mathbf{G}''_o und $\{\mathbf{Y}^{[L]}\}_{L=0}^{n_h}$, $\{\mathbf{G}'^{[L]}\}_{L=1}^{n_h}$, $\{\mathbf{G}''^{[L]}\}_{L=1}^{n_h}$ sowie $\{\mathbf{Y}_{,i}^{[L]}\}_{L=0}^{n_h}$ und $\{\mathbf{S}_{,i}^{[L]}\}_{L=0}^{n_h+1}$ für alle Indizes $i \in \mathbb{I}$ abgespeichert. Im Anschluss können die Matrizen \mathbf{F}_i aus Gleichung (4.69) abhängig vom definierten Fehlerterm berechnet werden. Zur Herleitung des Algorithmus wird ausgehend von Gleichung (4.68) zunächst die Komponente

$$\nabla \mathcal{L}_{ml}^{[L]} = \frac{1}{P} \sum_{i \in \mathbb{I}} \sum_{k=1}^P \sum_{j \in \mathbb{J}(i)} f_{ijk} \frac{\partial}{\partial w_{ml}^{[L]}} \left(\frac{\partial \hat{z}_{jk}}{\partial \hat{x}_i} \right) \quad (4.70)$$

betrachtet. Die Reihenfolge der partiellen Ableitungen $\partial/\partial w_{ml}^{[L]}$ und $\partial/\partial \hat{x}_i$ darf getauscht werden, da die Eingangsvariablen \mathbf{x} und \mathbf{w} unabhängig voneinander sind. Unter Beachtung der gewichteten Summe (4.18) folgt die Umformung

$$\nabla \mathcal{L}_{ml}^{[L]} = \frac{1}{P} \sum_{i \in \mathbb{I}} \sum_{k=1}^P \sum_{j \in \mathbb{J}(i)} f_{ijk} \frac{\partial}{\partial \hat{x}_i} \left(\frac{\partial \hat{z}_{jk}}{\partial s_m^{[L]}} y_{lk}^{[L]} \right). \quad (4.71)$$

Wie die Delta-Werte in Kapitel 4.2.4 werden die folgenden Variablen definiert:

$$\delta_{imk}^{[L]} := \sum_{j \in \mathbb{J}(i)} f_{ijk} \frac{\partial \hat{z}_{jk}}{\partial s_m^{[L]}}, \quad (4.72)$$

$$\gamma_{imk}^{[L]} := \sum_{j \in \mathbb{J}(i)} f_{ijk} \frac{\partial}{\partial \hat{x}_i} \left(\frac{\partial \hat{z}_{jk}}{\partial s_m^{[L]}} \right). \quad (4.73)$$

Sie können für alle $j \in \mathbb{J}(i)$ zusammengefasst werden, was die numerische Umsetzung wesentlich effizienter macht. Daher wurden die Summen in Gleichung (4.68) in dieser Art definiert. Mit diesen Variablen lässt sich Gleichung (4.71) zu

$$\nabla \mathcal{L}_{ml}^{[L]} = \frac{1}{P} \sum_{i \in \mathbb{I}} \sum_{k=1}^P \left(\delta_{imk}^{[L]} y_{ik,i}^{[L-1]} + \gamma_{imk}^{[L]} y_{ik}^{[L-1]} \right) \quad (4.74)$$

umformen. Die Variablen $\delta_{imk}^{[L]}$ und $\gamma_{imk}^{[L]}$ können wie die klassischen Delta-Werte schichtweise in den Matrizen

$$\boldsymbol{\delta}_i^{[L]} := \begin{bmatrix} \delta_{i11}^{[L]} & \cdots & \delta_{i1P}^{[L]} \\ \vdots & \delta_{imk}^{[L]} & \vdots \\ \delta_{in_L1}^{[L]} & \cdots & \delta_{in_LP}^{[L]} \end{bmatrix} \in \mathbb{R}^{n_L \times P} \quad (4.75)$$

und

$$\boldsymbol{\gamma}_i^{[L]} := \begin{bmatrix} \gamma_{i11}^{[L]} & \cdots & \gamma_{i1P}^{[L]} \\ \vdots & \gamma_{imk}^{[L]} & \vdots \\ \gamma_{in_L1}^{[L]} & \cdots & \gamma_{in_LP}^{[L]} \end{bmatrix} \in \mathbb{R}^{n_L \times P} \quad (4.76)$$

zusammengefasst werden. Mit diesen Definitionen lassen sich alle Gradiententerme aus Gleichung (4.74) für Schicht L mit der Matrixoperation

$$\nabla \mathcal{L}^{[L]} = \frac{1}{P} \sum_{i \in \mathbb{I}} \left(\boldsymbol{\delta}_i^{[L]} \mathbf{Y}_{,i}^{[L-1]T} + \boldsymbol{\gamma}_i^{[L]} \mathbf{Y}^{[L-1]T} \right) \quad (4.77)$$

ermitteln.

Wie im Fall der klassischen Delta-Werte lassen sich die Variablen $\delta_{imk}^{[L]}$ und $\gamma_{imk}^{[L]}$ über BP-Algorithmen berechnen. Für die Ausgangsschicht gilt für die $\delta_{imk}^{[L]}$ -Werte nach Definition (4.72)

$$\boldsymbol{\delta}_i^{[n_h+1]} = \mathbf{G}'_o \circ \mathbf{F}_i . \quad (4.78)$$

Für sie lässt sich die Aktualisierungsvorschrift von Schicht $L+1$ zu Schicht L äquivalent wie die der klassischen Delta-Werte in Kapitel 4.2.4 herleiten. In Matrixnotation lautet sie

$$\boldsymbol{\delta}_i^{[L]} = \mathbf{G}'^{[L]} \circ \left(\bar{\mathbf{W}}^{[L+1]T} \boldsymbol{\delta}_i^{[L+1]} \right) . \quad (4.79)$$

Die Variablen $\gamma_{imk}^{[L]}$ lassen sich auf ähnliche Weise berechnen. In der Ausgangsschicht gilt nach Definition (4.73)

$$\boldsymbol{\gamma}_i^{[n_h+1]} = \mathbf{G}''_o \circ \mathbf{F}_i \circ \mathbf{S}_{,i}^{[n_h+1]} . \quad (4.80)$$

Vorwärtsrechnung nach Tafel 4.3

für $\hat{\mathbf{Z}}, \mathbf{G}'_o, \mathbf{G}''_o$ und $\{\mathbf{Y}^{[L]}\}_{L=0}^{n_h}, \{\mathbf{G}'^{[L]}\}_{L=1}^{n_h}, \{\mathbf{G}''^{[L]}\}_{L=1}^{n_h}$
sowie $\{\mathbf{Y}_{,i}^{[L]}\}_{L=0}^{n_h}$ und $\{\mathbf{S}_{,i}^{[L]}\}_{L=0}^{n_h+1}$ für alle Indizes $i \in \mathbb{I}$

Startwerte $\mathbf{F}_i, \forall i \in \mathbb{I}$

Initialisierung $\nabla \mathcal{L}^{[L]} = \mathbf{0}, \forall L = 1, \dots, n_h + 1$

Schleife über Eingangsvariablen $i \in \mathbb{I}$

for $i \in \mathbb{I}$

$$\delta_i^{[n_h+1]} = \mathbf{G}'_o \circ \mathbf{F}_i$$

$$\gamma_i^{[n_h+1]} = \mathbf{G}''_o \circ \mathbf{F}_i \circ \mathbf{S}_{,i}^{[n_h+1]}$$

$$\nabla \mathcal{L}^{[n_h+1]} = \nabla \mathcal{L}^{[n_h+1]} + (1/P) \left(\delta_i^{[n_h+1]} \mathbf{Y}_i^{[n_h]T} + \gamma_i^{[n_h+1]} \mathbf{Y}^{[n_h]T} \right)$$

for $L = n_h, \dots, 1$

$$\delta_i^{[L]} = \mathbf{G}'^{[L]} \circ \left(\bar{\mathbf{W}}^{[L+1]T} \delta_i^{[L+1]} \right)$$

$$\gamma_i^{[L]} = \mathbf{G}'^{[L]} \circ \left(\bar{\mathbf{W}}^{[L+1]T} \gamma_i^{[L+1]} \right) + \mathbf{G}''^{[L]} \circ \mathbf{S}_{,i}^{[L]} \circ \left(\bar{\mathbf{W}}^{[L+1]T} \delta_i^{[L+1]} \right)$$

$$\nabla \mathcal{L}^{[L]} = \nabla \mathcal{L}^{[L]} + (1/P) \left(\delta_i^{[L]} \mathbf{Y}_i^{[L-1]T} + \gamma_i^{[L]} \mathbf{Y}^{[L-1]T} \right)$$

end for

end for

Umsortierung in den Gradientenvektor $\nabla \mathcal{L}^{[L]} \rightarrow \nabla \mathcal{L}, \forall L$

Tafel 4.5: Rückwärtsrechnung für Gradienten $\nabla \mathcal{L}$ von FNN-Ableitungen $z_{j,i}$

Die Aktualisierungsvorschrift für die $\gamma_{imk}^{[L]}$ -Werte von Schicht $L + 1$ zu Schicht L wird über die Abhängigkeit der gewichteten Summen aus der Idee von (4.64) berechnet. Dadurch formt sich Definition (4.73) zunächst zu

$$\gamma_{imk}^{[L]} = \sum_{j \in \mathbb{J}(i)} f_{ijk} \sum_{r=1}^{n_{L+1}} \left[\frac{\partial}{\partial \hat{x}_i} \left(\frac{\partial \hat{z}_{jk}}{\partial s_r^{[L+1]}} \right) g'(s_{mk}^{[L]}) w_{rm}^{[L+1]} + \frac{\partial \hat{z}_{jk}}{\partial s_r^{[L+1]}} g''(s_{mk}^{[L]}) \frac{\partial s_{mk}^{[L]}}{\partial \hat{x}_i} w_{rm}^{[L+1]} \right] \quad (4.81)$$

um, wobei die Definition der gewichteten Summe (4.18) und die Produktregel verwendet wurden. Die Summen über j und r können vertauscht und die Ableitungen der Aktivierungsfunktionen aus beiden Summen ausgeklammert werden. Nach Identifikation der $\delta_{irk}^{[L+1]}$ - und $\gamma_{irk}^{[L+1]}$ -Werte lässt sich die Aktualisierungsvorschrift letztlich mit

$$\gamma_{imk}^{[L]} = g'(s_{mk}^{[L]}) \sum_{r=1}^{n_{L+1}} w_{rm}^{[L+1]} \gamma_{irk}^{[L+1]} + g''(s_{mk}^{[L]}) \frac{\partial s_{mk}^{[L]}}{\partial \hat{x}_i} \sum_{r=1}^{n_{L+1}} w_{rm}^{[L+1]} \delta_{irk}^{[L+1]}, \quad (4.82)$$

bzw. in Matrix-Notation mit

$$\gamma_i^{[L]} = \mathbf{G}'^{[L]} \circ \left(\bar{\mathbf{W}}^{[L+1]T} \gamma_i^{[L+1]} \right) + \mathbf{G}''^{[L]} \circ \mathbf{S}_{,i}^{[L]} \circ \left(\bar{\mathbf{W}}^{[L+1]T} \delta_i^{[L+1]} \right) \quad (4.83)$$

angeben. Damit sind alle notwendigen Gleichungen hergeleitet. Der vollständige Algorithmus zur Berechnung des Gradienten $\nabla \mathcal{L}$ einer Fehlerfunktion, die partielle Ableitungen enthält, ist in Tafel 4.5 angegeben.

4.3 Weitere KNN-Architekturen

Neben dem klassischen Feedforward-Netz aus Kapitel 4.2 existieren viele weitere KNN-Architekturen. Diese haben sich über die Jahre in verschiedenen Anwendungsgebieten entwickelt und sind folglich oft spezialisiert zur Lösung der zugehörigen Problemstellungen. Auch im Rahmen der Materialmodellierung haben sich einige dieser vom klassischen FNN abweichenden Architekturen als praktisch herausgestellt. Im Folgenden werden von diesen Architekturen zwei näher beschrieben, mit denen dann die FNN-Ergebnisse in den numerischen Beispielen verglichen werden.

4.3.1 Das Input Convex Neural Network (ICNN)

Das *Input Convex Neural Network* (ICNN) wurde im Jahr 2017 von AMOS ET AL. [6] veröffentlicht. Von KLEIN ET AL. [103] wurde es erstmals zur Modellierung polykonvexer anisotroper elastischer Materialien verwendet. Im Rahmen der Materialmodellierung wird das ICNN üblicherweise als klassisches FNN bezeichnet und in Verbindung mit Sobolev-Training (vgl. CZARNECKI ET AL. [32]) beschrieben, welches das Training auf die Ableitungen der FNN-Ausgangsvariablen beschreibt. Für eine pragmatische Unterscheidbarkeit innerhalb der Untersuchungen der nachfolgenden Kapitel ist es an dieser Stelle zweckmäßig, das ICNN als eigenständige Architektur zu bezeichnen.

Das ICNN macht es möglich die Existenz einer Verzerrungsenergie-dichtefunktion zu gewährleisten, ohne dass diese explizit vorgegeben wird. Das macht das ICNN gerade im Fall elastischer KNN-Materialmodellierung sehr interessant. Die folgenden Herleitungen und Bezeichnungen sind an die Definitionen des FNN in den vorherigen Abschnitten angepasst. Die Eingangs- und Ausgangsvariablen (\mathbf{x}, \mathbf{z}) des ICNN werden wieder transformiert zu $(\hat{\mathbf{x}}, \hat{\mathbf{z}})$. Das ICNN besitzt eine beliebige Anzahl an Zwischenschichten, allerdings nur ein Neuron in der Ausgangsschicht, welches im Folgenden als *ICNN-Neuron* mit Ausgabewert $\hat{\Psi}$ bezeichnet wird. Die gesamte Informationsverarbeitung des ICNN ist in Abbildung 4.7 dargestellt.

Die transformierten Ausgangsvariablen $\hat{\mathbf{z}}$ des ICNN werden als Gradient

$$\hat{\mathbf{z}} = \left(\frac{\partial \hat{\Psi}(\mathbf{x}, \mathbf{w})}{\partial \hat{\mathbf{x}}} \right)^T \quad (4.84)$$

dieser Funktion $\hat{\Psi}$ definiert. Die HESSE-Matrix von $\hat{\Psi}$ ist dadurch die transformierte JACOBI-Matrix

$$\hat{\mathbf{j}}(\mathbf{x}, \mathbf{w}) = \frac{\partial \hat{\mathbf{z}}}{\partial \hat{\mathbf{x}}} = \frac{\partial^2 \hat{\Psi}}{\partial \hat{\mathbf{x}}^2}. \quad (4.85)$$

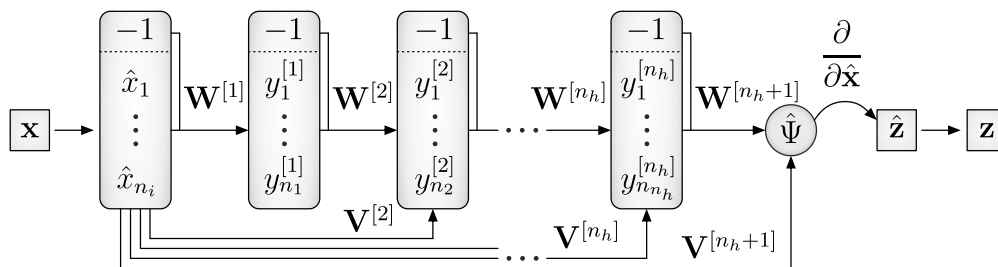


Abbildung 4.7: Informationsverarbeitung und Topologie eines ICNN mit Direktverbindungen vom Eingang zu allen Zwischenschichten

Durch die Existenz des Potentials $\hat{\Psi}$ ist $\hat{\mathbf{j}}$ symmetrisch – und zwar per Konstruktion des KNN. Die echte JACOBI-Matrix $\mathbf{j} = \mathbf{S}_z \hat{\mathbf{j}} \mathbf{S}_x^{-1}$ ist aufgrund der Transformationen (4.34) allerdings nicht automatisch auch symmetrisch, ein Potential Ψ mit $(\partial\Psi(\mathbf{x})/\partial\mathbf{x})^T = \mathbf{z}$ existiert also nicht. Dies wird im Folgenden dadurch erzwungen, dass die Streckungsparameter s_{xi} und s_{zj} gemittelt werden mit

$$\bar{s}_x = \frac{1}{n_i} \sum_{i=1}^{n_i} s_{xi} \quad \text{und} \quad \bar{s}_z = \frac{1}{n_o} \sum_{j=1}^{n_o} s_{zj} . \quad (4.86)$$

Die Eingangs- und Ausgangsvariablen werden somit jeweils mit den gleichen Streckungsparametern \bar{s}_x und \bar{s}_z transformiert. Die Mittelungsparameter m_{xi} und m_{zj} bleiben unabhängig. Dieses Vorgehen hat in der praktischen Anwendung dann keinen Nachteil, wenn die Wertebereiche der Eingangsvariablen alle in der gleichen Größenordnung liegen. Dies ist bspw. bei der Modellierung elastischen Materialverhaltens meist der Fall. Somit existiert ein Potential

$$\Psi(\mathbf{x}) := \bar{s}_x \bar{s}_z \hat{\Psi}(\mathbf{x}, \mathbf{w}) + \bar{s}_x \mathbf{m}_z^T \hat{\mathbf{x}} , \quad (4.87)$$

dessen HESSE-Matrix

$$\frac{\partial^2 \Psi}{\partial \mathbf{x}^2} = \frac{\bar{s}_z}{\bar{s}_x} \frac{\partial^2 \hat{\Psi}}{\partial \hat{\mathbf{x}}^2} = \mathbf{j}(\mathbf{x}, \mathbf{w}) \quad (4.88)$$

symmetrisch ist. Unter Vernachlässigung einer Daten-Normalisierung wären diese Überlegungen nicht vonnöten. Alternativ könnte man die Normalisierung auch innerhalb des Fehlerterms \mathcal{L}_T durchführen. Dann geht allerdings der Vorteil ähnlicher Gewichts-Wertebereiche verloren, vgl. Abschnitt 4.1.

Neben den klassischen gewichteten Verbindungen des FNN führen AMOS ET AL. [6] auch Direktverbindungen (engl. original: *passthroughs*) ein. Diese verbessern die Flexibilität des ICNN, die es durch Einfordern der Konvexität zunächst einbüßt, was weiter unten beschrieben wird. Ab der zweiten Zwischenschicht lässt sich die gewichtete Summe mit

$$s_m^{[L]} = \sum_{l=0}^{n_{L-1}} w_{ml}^{[L]} y_l^{[L-1]} + \sum_{i=1}^{n_i} v_{mi}^{[L]} \hat{x}_i \quad (4.89)$$

berechnen⁴⁰. Die zusätzlichen Gewichte $v_{mi}^{[L]}$ werden in der Implementierung sinngemäß hinter den Vektor der klassischen FNN-Gewichte \mathbf{w}^{FNN} aus Definition (4.16) in den Vektor \mathbf{w} aller Gewichte einsortiert

$$\mathbf{w} := [\mathbf{w}^{\text{FNN}^T}, v_{11}^{[2]}, v_{12}^{[2]}, \dots, v_{1n_i}^{[2]}, \dots, v_{n_2 n_i}^{[2]}, \dots, v_{11}^{[3]}, \dots, v_{1n_i}^{[n_h+1]}]^T \in \mathbb{R}^{n_w}. \quad (4.90)$$

Innerhalb einer Schicht L können die zusätzlichen Gewichte in der Matrix

$$\mathbf{V}^{[L]} := \begin{bmatrix} v_{11}^{[L]} & \dots & v_{1n_i}^{[L]} \\ \vdots & \ddots & \vdots \\ v_{n_L 1}^{[L]} & \dots & v_{n_L n_i}^{[L]} \end{bmatrix} \in \mathbb{R}^{n_L \times n_i} \quad (4.91)$$

zusammengefasst werden. Die Vektoren der gewichteten Summen $\mathbf{s}^{[L]}$ aus Definition (4.22) lassen sich folglich angeben mit

$$\mathbf{s}^{[L]} = \begin{cases} \mathbf{W}^{[1]} \mathbf{y}^{[0]} & L = 1 \\ \mathbf{W}^{[L]} \mathbf{y}^{[L-1]} + \mathbf{V}^{[L]} \hat{\mathbf{x}} & L = 2, \dots, n_h + 1 \end{cases}. \quad (4.92)$$

Die zugehörigen Ableitungen nach den transformierten Eingangsvariablen sind

$$\mathbf{s}_{,i}^{[L]} = \frac{\partial \mathbf{s}^{[L]}}{\partial \hat{x}_i} = \begin{cases} \bar{\mathbf{W}}_{(:,i)}^{[1]} \quad (i\text{-te Spalte von } \bar{\mathbf{W}}^{[1]}) & L = 1 \\ \bar{\mathbf{W}}^{[L]} \bar{\mathbf{y}}_{,i}^{[L-1]} + \mathbf{V}^{[L]} & L = 2, \dots, n_h + 1 \end{cases}. \quad (4.93)$$

Die Neuronenausgaben $\mathbf{y}^{[L]}$, ihre Ableitungen $\mathbf{y}_{,i}^{[L]}$ sowie die Ableitungen der Aktivierungsfunktionen $\mathbf{g}'^{[L]}$ und $\mathbf{g}''^{[L]}$ werden mithilfe der angepassten gewichteten Summen (4.92) und deren Ableitungen (4.93) sinngemäß mit den Gleichungen (4.25), (4.39), (4.26) und (4.27) berechnet.

Die ICNN-Vorwärtsrechnung

Der ICNN-Ausgang $\mathbf{z} = \mathbf{f}^{\text{ICNN}}(\mathbf{x}, \mathbf{w})$ lässt sich mit dem FNN-Algorithmus in Tafel 4.2 berechnen, wenn in den einzelnen Gleichungen die modifizierten gewichteten Summen (4.92) und deren Ableitungen (4.93) verwendet werden. Dabei gilt für die Indizes $i \in \mathbb{I} = \{1, \dots, n_i\}$. Für die Ausgangstransformation ist zu beachten, dass es sich hier um den direkten Ausgang des ICNN handelt und nicht um die Ableitung des FNN. Daher gilt, anders als in Tafel 4.2 geschrieben,

$$z_i = m_{zi} + \bar{s}_z \hat{z}_i = m_{zi} + \bar{s}_z \frac{\partial \hat{\Psi}}{\partial \hat{x}_i}. \quad (4.94)$$

Alternativ kann auch die direkte Formel (4.33) verwendet werden, wobei hier ebenfalls auf die korrekte Transformation zu achten ist.

⁴⁰In AMOS ET AL. [6] werden die Gewichte $\mathbf{W}^{[1]}$ als Passthrough-Gewichte $\mathbf{V}^{[1]}$ definiert.

Die ICNN-Jacobi-Matrix

Zur Berechnung der transformierten ICNN-JACOBI-Matrix $\hat{\mathbf{j}}(\mathbf{x}, \mathbf{w})$ werden die zweifachen Ableitungen von $\hat{\Psi}$ nach den Eingangsvariablen $\hat{\mathbf{x}}$ benötigt. Die zweifache Ableitung eines beliebigen Neuronen-Ausgangs $y_m^{[L]}$ innerhalb des ICNN lässt sich mit den Ableitungen der gewichteten Summen $s_{m,i}^{[L]}$ aus Gleichung (4.93) mit

$$y_{m,ij}^{[L]} = \frac{\partial^2 y_m^{[L]}}{\partial \hat{x}_i \partial \hat{x}_j} = g''(s_m^{[L]}) s_{m,i}^{[L]} s_{m,j}^{[L]} + g'(s_m^{[L]}) \left(\sum_{l=1}^{n_L-1} w_{ml}^{[L]} y_{i,lj}^{[L-1]} \right) \quad (4.95)$$

berechnen. Für die erste Zwischenschicht gilt $y_{m,ij}^{[1]} = g''(s_m^{[1]}) w_{mi}^{[1]} w_{mj}^{[1]}$, und für die Ausgangsschicht muss Gleichung (4.95) mit $y_m = \hat{\Psi}$ und g_o angewendet werden. Eine kompakte Vektor-Matrix-Notation kann hergeleitet werden, wenn zunächst i konstant angenommen wird. Dadurch wird eine Zeile $\hat{\mathbf{j}}_{(i)}$ der transformierten JACOBI-Matrix $\hat{\mathbf{j}}$ berechnet. Die zweifachen Ableitungen $y_{m,ij}^{[L]}$ werden dann in der Matrix

$$\hat{\mathbf{Y}}_{,i}^{[L]} := \begin{bmatrix} y_{1,i1}^{[L]} & \cdots & y_{1,in_i}^{[L]} \\ \vdots & y_{m,ij}^{[L]} & \vdots \\ y_{n_L,i1}^{[L]} & \cdots & y_{n_L,in_i}^{[L]} \end{bmatrix} \in \mathbb{R}^{n_L \times n_i} \quad (4.96)$$

zusammengefasst. Sie ist nicht zu verwechseln mit der Matrix $\bar{\mathbf{Y}}_{,i}^{[L]}$ aus Definition (4.47), welche die einfachen Ableitung nach \hat{x}_i für alle Datenpunkte beinhaltet. Hier wird die JACOBI-Matrix nur für einen Datenpunkt berechnet. In der ersten Zwischenschicht gilt

$$\hat{\mathbf{Y}}_{,i}^{[1]} = (\mathbf{g}''^{[1]} \otimes \mathbf{1}^{n_i}) \circ (\mathbf{1}^{n_1} \otimes \bar{\mathbf{W}}_{(m,:)}^{[1]}) \circ \bar{\mathbf{W}}^{[1]}, \quad (4.97)$$

mit dem Vektor $\mathbf{1}^m$ bestehend aus m Einsen, vgl. Definition (4.31), und der Bezeichnung $\bar{\mathbf{W}}_{(m,:)}^{[1]}$ für die m -te Zeile von $\bar{\mathbf{W}}^{[1]}$. In den Zwischenschichten gilt mit der Definition (4.93) und der Matrix

$$\bar{\mathbf{Y}}'^{[L]} := [\bar{\mathbf{y}}_{,1}^{[L]}, \dots, \bar{\mathbf{y}}_{,n_i}^{[L]}] \in \mathbb{R}^{n_L \times n_i} \quad (4.98)$$

die Gleichung

$$\begin{aligned} \hat{\mathbf{Y}}_{,i}^{[L]} &= (\mathbf{g}''^{[L]} \otimes \mathbf{1}^{n_i}) \circ (\bar{\mathbf{W}}^{[L]} \bar{\mathbf{Y}}'^{[L-1]} + \mathbf{V}^{[L]}) \circ ((\bar{\mathbf{W}}^{[L]} \bar{\mathbf{y}}_{,i}^{[L-1]} + \mathbf{V}_{(:,i)}^{[L]}) \otimes \mathbf{1}^{n_i}) \\ &\quad + (\mathbf{g}'^{[L]} \otimes \mathbf{1}^{n_i}) \circ (\bar{\mathbf{W}}^{[L]} \hat{\mathbf{Y}}_{,i}^{[L-1]}). \end{aligned} \quad (4.99)$$

Dabei kann die Matrix $\bar{\mathbf{Y}}'^{[L]}$ mit Gleichung

$$\bar{\mathbf{Y}}'^{[L]} = (\mathbf{g}'^{[L]} \otimes \mathbf{1}^{n_i}) \circ (\bar{\mathbf{W}}^{[L]} \bar{\mathbf{Y}}'^{[L-1]} + \mathbf{V}^{[L]}) \quad (4.100)$$

in den Schichten vorwärts aktualisiert werden. In der Ausgangsschicht $L = n_h + 1$ muss dementsprechend die Aktivierungsfunktion g_o verwendet werden. Es bietet sich an, die Vorwärtsrechnungen für \mathbf{z} und \mathbf{j} zu verknüpfen, da für \mathbf{j} auch die ersten Ableitungen benötigt werden. Dies ist in Tafel 4.6 dargestellt.

Eingangstransformation $\hat{\mathbf{x}} = \mathbf{S}_x^{-1}(\mathbf{x} - \mathbf{m}_x)$

Initialisierung (Eingangsschicht und erste Zwischenschicht)

$$\begin{aligned}\mathbf{y}^{[0]} &= [-1, \hat{\mathbf{x}}^T]^T \\ \mathbf{s}^{[1]} &= \mathbf{W}^{[1]}\mathbf{y}^{[0]} \\ \mathbf{y}^{[1]} &= [-1, g(\mathbf{s}^{[1]})^T]^T \\ \mathbf{g}'^{[1]} &= g'(\mathbf{s}^{[1]}) \text{ und } \mathbf{g}''^{[1]} = g''(\mathbf{s}^{[1]}) \\ \bar{\mathbf{Y}}^{[1]} &= (\mathbf{g}'^{[1]} \otimes \mathbf{1}^{n_i}) \circ \bar{\mathbf{W}}^{[1]}\end{aligned}$$

Schleife über die Eingangsvariablen $i = 1, \dots, n_i$

$$\begin{aligned}\text{for } i = 1, \dots, n_i \text{ do} \\ \hat{\mathbf{Y}}_{,i}^{[1]} &= ((\mathbf{g}''^{[1]} \circ \bar{\mathbf{W}}_{(:,i)}^{[1]}) \otimes \mathbf{1}^{n_i}) \circ \bar{\mathbf{W}}^{[1]} \\ \text{for } L = 2, \dots, n_h \text{ do} \\ \mathbf{s}^{[L]} &\stackrel{*}{=} \mathbf{W}^{[L]}\mathbf{y}^{[L-1]} + \mathbf{V}^{[L]}\hat{\mathbf{x}} \\ \mathbf{y}^{[L]} &\stackrel{*}{=} [-1, g(\mathbf{s}^{[L]})^T]^T \\ \mathbf{g}'^{[L]} &\stackrel{*}{=} g'(\mathbf{s}^{[L]}) \text{ und } \mathbf{g}''^{[L]} \stackrel{*}{=} g''(\mathbf{s}^{[L]}) \\ \bar{\mathbf{Y}}'^{[L]} &\stackrel{*}{=} (\mathbf{g}'^{[L]} \otimes \mathbf{1}^{n_i}) \circ (\bar{\mathbf{W}}^{[L]}\bar{\mathbf{Y}}'^{[L-1]} + \mathbf{V}^{[L]}) \\ \hat{\mathbf{Y}}_{,i}^{[L]} &= (\mathbf{g}''^{[L]} \otimes \mathbf{1}^{n_i}) \circ (\bar{\mathbf{W}}^{[L]}\bar{\mathbf{Y}}'^{[L-1]} + \mathbf{V}^{[L]}) \\ &\quad \circ ((\bar{\mathbf{W}}^{[L]}\bar{\mathbf{y}}_{,i}^{[L-1]} + \mathbf{V}_{(:,i)}^{[L]}) \otimes \mathbf{1}^{n_i}) + (\mathbf{g}'^{[L]} \otimes \mathbf{1}^{n_i}) \circ (\bar{\mathbf{W}}^{[L]}\hat{\mathbf{Y}}_{,i}^{[L-1]}) \\ \text{end for} \\ \mathbf{s}^{[n_h+1]} &\stackrel{*}{=} \mathbf{W}^{[n_h+1]}\mathbf{y}^{[n_h]} + \mathbf{V}^{[n_h+1]}\hat{\mathbf{x}} \\ g'_o &\stackrel{*}{=} g'_o(\mathbf{s}^{[n_h+1]}) \text{ und } g''_o \stackrel{*}{=} g''_o(\mathbf{s}^{[n_h+1]}) \\ \hat{\mathbf{z}} &\stackrel{*}{=} (g'_o \mathbf{1}^{n_i}) \circ (\bar{\mathbf{W}}^{[n_h+1]}\bar{\mathbf{Y}}'^{[n_h]} + \mathbf{V}^{[n_h+1]})^T \\ \hat{\mathbf{j}}_{(i)} &= (g''_o \mathbf{1}^{n_i})^T \circ (\bar{\mathbf{W}}^{[n_h+1]}\bar{\mathbf{Y}}'^{[n_h]} + \mathbf{V}^{[n_h+1]}) \\ &\quad \circ ((\bar{\mathbf{W}}^{[n_h+1]}\bar{\mathbf{y}}_{,i}^{[n_h]} + v_{1i}^{[n_h+1]}) \mathbf{1}^{n_i})^T + (g'_o \mathbf{1}^{n_i})^T \circ (\bar{\mathbf{W}}^{[n_h+1]}\hat{\mathbf{Y}}_{,i}^{[n_h]}) \\ \text{end for}\end{aligned}$$

Ausgangstransformation

$$\begin{aligned}\mathbf{z} &= \mathbf{m}_z + \mathbf{S}_z \hat{\mathbf{z}} \\ \mathbf{j} &= \mathbf{S}_z \hat{\mathbf{j}} \mathbf{S}_x^{-1}\end{aligned}$$

Tafel 4.6: Vorwärtsrechnung für ICNN-Ausgang $\mathbf{z} = \mathbf{f}^{\text{ICNN}}(\mathbf{x}, \mathbf{w})$ und die JACOBI-Matrix $\mathbf{j}(\mathbf{x}, \mathbf{w})$. Die mit * markierten Operationen sind von der äußeren Schleife unabhängig und können prinzipiell ausgelagert werden.

Der ICNN-Fehlergradient

Da der transformierte ICNN-Ausgang $\hat{\mathbf{z}} = \hat{\Psi}_{,\hat{\mathbf{x}}}$ ein Gradient ist, muss für alle transformierten Trainingsdaten $(\hat{\mathbf{x}}_k, \hat{\mathbf{t}}_k)$, $k = 1, \dots, P_T$ der Trainingsdatenfehler

$$\mathcal{L}_T(\mathbf{w}) = \frac{1}{P_T} \sum_{k=1}^{P_T} \mathcal{L}_{Tk} = \frac{1}{P_T} \sum_{k=1}^{P_T} \frac{1}{2} \sum_{o=1}^{n_o} \left(\frac{\partial \hat{\Psi}(\mathbf{x}_k, \mathbf{w})}{\partial \hat{x}_o} - \hat{t}_{ok} \right)^2 \quad (4.101)$$

minimiert werden, und nicht jener aus Gleichung (4.53). Der zugehörige Gradient

$$\nabla \mathcal{L}_T(\mathbf{w}) = \frac{1}{P_T} \sum_{o=1}^{n_o} \sum_{k=1}^{P_T} \frac{\partial \mathcal{L}_{Tk}}{\partial \hat{\Psi}_{,o}} \frac{\partial}{\partial \mathbf{w}} \left(\frac{\partial \hat{\Psi}}{\partial \hat{x}_o} \right) \quad (4.102)$$

hat die gleiche Form wie der von Abschnitt 4.2.5, mit $i \in \mathbb{I} = \{1, \dots, n_o\}$, $\mathbb{J} \equiv \{1\}$ und den Vorfaktoren

$$f_{i1k}^{\text{ICNN}} = \frac{\partial \hat{\Psi}(\mathbf{x}_k, \mathbf{w})}{\partial \hat{x}_i} - \hat{t}_{ik}, \quad (4.103)$$

die in die Matrix (4.69) eingesetzt werden können. Der in Abschnitt 4.2.5 beschriebene Algorithmus ist somit uneingeschränkt für den Fehlergradient des ICNN anwendbar.

Konvexität der Potentialfunktion

Je nach Anwendungsfall kann es zweckmäßig sein, die Konvexität der Funktion $\hat{\Psi}$ bezüglich der transformierten Eingangsvariablen $\hat{\mathbf{x}}$ zu fordern. Dadurch ist die Funktion $\Psi(\mathbf{x})$ auch konvex (s. u.). Dies motiviert schließlich den Namen ICNN. Dazu werden die Eigenschaften a) und b) konvexer Funktionen aus Abschnitt 3.3.5 im Folgenden auf das ICNN übertragen. Der Ausgangswert des letzten Neurons ist eine verkettete Funktion $\hat{\Psi}(g_o(\mathbf{x}))$ mit

$$\hat{\Psi}(\hat{\mathbf{x}}) = g_0 \left((-1)w_{10}^{[n_h]} + \sum_{l=1}^{n_{n_h-1}} w_{1l}^{[n_h]} y_l^{[n_h-1]}(\hat{\mathbf{x}}) + \sum_{i=1}^{n_i} v_{1i}^{[n_h]} \hat{x}_i \right). \quad (4.104)$$

$\hat{\Psi}$ ist genau dann konvex, wenn die Aktivierungsfunktion g_o monoton wachsend ist, die Funktionen $y_l^{[n_h-1]}(\hat{\mathbf{x}})$ selbst konvex sind und für die Gewichte $w_{1l}^{[n_h]} \geq 0$ gilt. Eine lineare Transformation mittels Bias-Wert ändert die Eigenschaften der Konvexität nicht. Diese Argumentationskette lässt sich bis zur Eingangsschicht weiterführen, sodass insgesamt die folgenden Bedingungen für die Konvexität von $\hat{\Psi}(\hat{\mathbf{x}})$ erfüllt sein müssen:

1. Alle Aktivierungsfunktionen $g(s)$ und $g_o(s)$ müssen monoton wachsend und konvex in s sein.
2. Die Gewichte ab der zweiten Zwischenschicht ($\bar{\mathbf{W}}^{[2]}, \dots, \bar{\mathbf{W}}^{[n_h+1]}$) müssen nichtnegativ sein.

Für die Bias-Gewichte $w_{m0}^{[L]}$ und die Gewichte der Direktverbindungen $v_{mi}^{[L]}$ gelten keine Einschränkungen⁴¹, da lineare und konstante Funktionen immer konvex

⁴¹Für das polykonvexe Materialmodell aus Abschnitt 5.2.4 müssen allerdings weitere Restriktionen berücksichtigt werden.

sind. Das ist auch der Grund, wieso die linearen Transformationen (4.7) und (4.8) nichts an der Konvexität ändern: ist $\hat{\Psi}(\hat{\mathbf{x}})$ konvex, dann auch $\Psi(\mathbf{x})$ aus Definition (4.87) – wenn die Transformationen wie beschrieben gemittelt werden. AMOS ET AL. [6] erklären ausdrücklich, dass diese Einschränkungen ohne die zusätzlichen Direktverbindungen zu stark wären und die Generalisierbarkeit des ICNN reduzieren würden. Während sich die Bedingung an die Aktivierungsfunktionen leicht durch die Wahl der SoftPlus-Funktion aus Abbildung 4.5 erfüllen lässt, ist die Bedingung positiver Gewichte dagegen nicht ohne Weiteres zu lösen. Dazu werden in dieser Arbeit die betroffenen Gewichte modifiziert. Mit

$$w_{ml}^{[L]} = \frac{1}{2}(\tilde{w}_{ml}^{[L]})^2 \geq 0 \quad \text{und} \quad \frac{\partial w_{ml}^{[L]}}{\partial \tilde{w}_{ml}^{[L]}} = \tilde{w}_{ml}^{[L]} \quad (4.105)$$

werden nicht die Gewichte $w_{ml}^{[L]}$ der gewichteten Summe direkt, sondern ihre *Erzeugergewichte* $\tilde{w}_{ml}^{[L]} \in \mathbb{R}$ trainiert und in den Vektor der Freiheitsgrade \mathbf{w} eingepflegt. Mit der Definition (4.105) ist die Nichtnegativität der $w_{ml}^{[L]}$ garantiert. Im Gradienten $\nabla \hat{\mathcal{L}}$ muss unter Beachtung der Kettenregel demzufolge nachdifferenziert werden.

$$\nabla \mathcal{L}_{ml}^{[L]} = \frac{\partial \mathcal{L}}{\partial \tilde{w}_{ml}^{[L]}} = \frac{\partial \mathcal{L}}{\partial w_{ml}^{[L]}} \tilde{w}_{ml}^{[L]} \quad (4.106)$$

Das Einfordern der Konvexität motiviert die Bezeichnung des ICNN. Dies ist allerdings nicht in allen Anwendungsfällen erstrebenswert. Im Rahmen dieser Arbeit wird die Konvexität des Ausgangs des ICNN-Neurons nur zur Einforderung der Polykonvexitätsbedingung nach Abschnitt 3.3.5 genutzt. Die Definition der Ausgangsgrößen als Ableitung einer Potentialfunktion wiederum ist öfter praktisch im Rahmen der KNN-Materialmodellierung. Wird die Konvexität im Folgenden nicht eingefordert, so wird es in dieser Arbeit als *NCNN* bezeichnet, um die Nähe zur Ursprungsarchitektur zu wahren.

4.3.2 Das mehrschichtige rekurrente Elman-Netz

Im Gegensatz zum FNN und ICNN können *rekurrente neuronale Netze* (RNN) auch Verbindungen zwischen Neuronen der gleichen Schicht oder Verbindungen zu Neuronen einer vorherigen Schicht berücksichtigen. Im Folgenden wird nur die erstgenannte Art betrachtet. Um solche Verbindungen zur gleichen Schicht zu ermöglichen, muss den Neuronenausgaben $y_{m(t)}^{[L]}$ eine Zeitvariable t hinzugefügt werden. Dadurch ist das RNN in der Lage, zeitliche Abhängigkeiten in Datensätzen zu identifizieren. Daher kann diese KNN-Architektur im Zusammenhang mit der Modellierung geschichtsabhängigen Materialverhaltens genutzt werden. Dabei gibt es eine Vielzahl verschiedener RNN-Architekturen. Im Rahmen dieser Arbeit wird ausschließlich das mehrschichtige ELMAN-RNN verwendet, das auf

ELMAN [41] im Jahr 1990 zurückgeht. Sie werden im Folgenden mit RNN bezeichnet. Eine Beispiel-Topologie ist in Abbildung 4.8 dargestellt. Dabei ist jedes

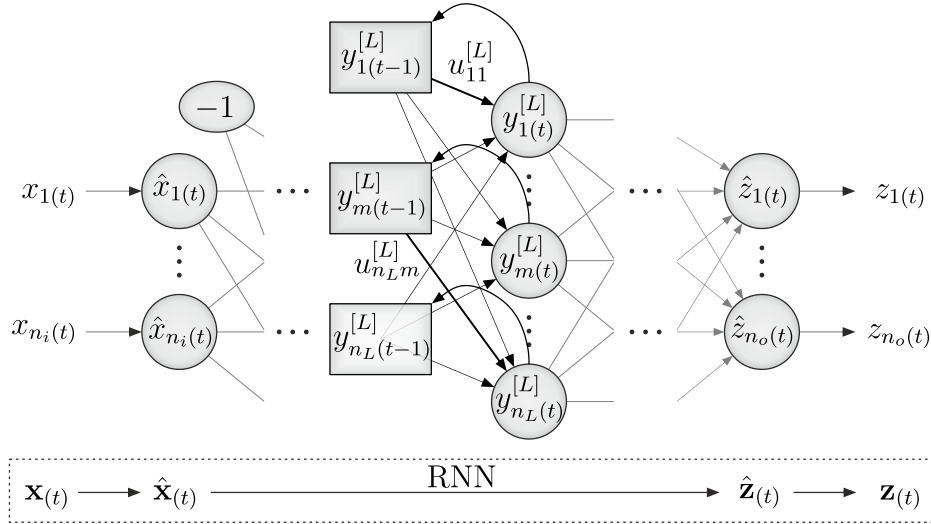


Abbildung 4.8: Rückkopplungen in einer Zwischenschicht eines RNN. Die FNN-Verbindungen sind nicht gezeichnet.

Neuron $y_m^{[L]}$ einer Zwischenschicht zum Zeitpunkt t zusätzlich zu allen Neuronen der gleichen Schicht zum Zeitpunkt $t - 1$ verbunden⁴². Die zugehörigen Gewichte werden im Folgenden mit $u_{mh}^{[L]}$ bezeichnet. Die gewichtete Summe

$$s_m^{[L]} = \sum_{l=0}^{n_L-1} w_{ml}^{[L]} y_l^{[L-1]} + \sum_{h=1}^{n_L} u_{mh}^{[L]} y_h^{[L]}(t-1) \quad (4.107)$$

enthält neben dem FNN-Term auch die Anteile aus der Rückkopplung. Die zusätzlichen Gewichte $u_{mh}^{[L]}$ werden sinngemäß hinter den Vektor der FNN-Gewichte \mathbf{w}^{FNN} aus Definition (4.16) in den Vektor \mathbf{w} aller Gewichte einsortiert

$$\mathbf{w} := [\mathbf{w}^{\text{FNN}T}, u_{11}^{[1]}, u_{12}^{[1]}, \dots, u_{1n_1}^{[1]}, \dots, u_{n_1n_1}^{[1]}, \dots, u_{11}^{[2]}, \dots, u_{n_h n_h}^{[n_h]}]^T \in \mathbb{R}^{n_w} . \quad (4.108)$$

Innerhalb einer Schicht L können die zusätzlichen Gewichte in der Matrix

$$\mathbf{U}^{[L]} := \begin{bmatrix} u_{11}^{[L]} & \dots & u_{1n_L}^{[L]} \\ \vdots & \ddots & \vdots \\ u_{n_L 1}^{[L]} & \dots & u_{n_L n_L}^{[L]} \end{bmatrix} \in \mathbb{R}^{n_L \times n_L} \quad (4.109)$$

zusammengefasst werden. Die nun zeitabhängigen Vektoren der gewichteten Summen $\mathbf{s}_{(t)}^{[L]}$ aus Definition (4.22) lassen sich folglich mit

$$\mathbf{s}_{(t)}^{[L]} = \mathbf{W}^{[L]} \mathbf{y}_{(t)}^{[L-1]} + \mathbf{U}^{[L]} \bar{\mathbf{y}}_{(t-1)}^{[L]} \quad (4.110)$$

⁴² In diesem Zusammenhang wird auch von *Kontextneuronen* gesprochen, in denen der Zustand des letzten Zeitschritts abgespeichert ist.

angeben. Die Neuronenausgaben ohne Bias $\bar{\mathbf{y}}_{(t)}^{[L]} = g(\mathbf{s}_{(t)}^{[L]})$ werden für jeden Zeitpunkt t mit der Aktivierungsfunktion $g(s)$ berechnet, vgl. Definition (4.25). Die Ableitungen der gewichteten Summen $\mathbf{s}_{i(t)}^{[L]}$ werden für jeden Zeitpunkt t weiter mit Gleichung (4.38) ermittelt, da beim Ableiten von Gleichung (4.110) der rekurrente Anteil wegfällt, mit

$$\frac{\partial \bar{\mathbf{y}}_{(t-1)}^{[L]}}{\partial \mathbf{x}_{(t)}} = \mathbf{0}. \quad (4.111)$$

Die Ableitungen der Neuronenausgaben $y_{i(t)}^{[L]}$ können analog zum FNN berechnet und im Vektor $\bar{\mathbf{y}}_{i(t)}^{[L]}$ zusammengefasst werden, vgl. Definition (4.39).

Die RNN-Vorwärtsrechnung

Im Kontext des RNN sind die Eingangsvektoren $\mathbf{x}_{(t)}$ Teil einer Zeitreihe

$$\mathcal{X} = \{\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(t)}, \dots, \mathbf{x}_{(\tau)}\} \quad (4.112)$$

und müssen in der richtigen Reihenfolge an das RNN übergeben werden. Am Anfang der Berechnung werden die Zustände aller Zwischenneuronen zu null initialisiert. Diese gehen über die gewichtete Summe (4.110) in die erste Vorwärtsrechnung $\mathbf{z}_{(1)} = \mathbf{f}^{\text{RNN}}(\mathbf{x}_{(1)}, \mathbf{w})$ ein. Die in dieser Berechnung ermittelten Zustände $y_{m(1)}^{[L]}$ gehen wiederum in die nächste Vorwärtsrechnung $\mathbf{z}_{(2)} = \mathbf{f}^{\text{RNN}}(\mathbf{x}_{(2)}, \mathbf{w})$ ein, usw. Dieser Prozess wird so lange durchgeführt, bis der letzte Vektor $\mathbf{x}_{(\tau)}$ verwendet wurde. Die somit entstandene Zeitreihe

$$\mathcal{Z} = \{\mathbf{z}_{(1)}, \dots, \mathbf{z}_{(t)}, \dots, \mathbf{z}_{(\tau)}\} \quad (4.113)$$

ist die Antwort des RNN auf die gegebene Zeitreihe \mathcal{X} . Aus diesen Gründen ist es sinnvoll, anstatt der FNN-Definition (4.1) die Abbildungsvorschrift

$$\mathcal{Z} = \mathcal{F}^{\text{RNN}}(\mathcal{X}, \mathbf{w}) \quad (4.114)$$

für das RNN zu verwenden⁴³. Eine einzelne Vorwärtsberechnung zum Zeitpunkt t lässt sich mit dem Algorithmus aus Tafel 4.1 berechnen, wenn für die gewichtete Summe nicht die FNN-Gleichung (4.24) verwendet wird, sondern die RNN-Gleichung (4.110). Die Zustände der Zwischenneuronen $\bar{\mathbf{y}}_{(t-1)}^{[L]}$ müssen dabei geeignet zwischengespeichert und berücksichtigt werden.

Die RNN-Jacobi-Matrix

Zu jedem Zeitpunkt t kann die RNN-JACOBI-Matrix

$$\mathbf{j}_{(t)}(\mathbf{x}_{(t)}, \mathbf{w}) := \frac{\partial \mathbf{z}_{(t)}(\mathbf{x}_{(t)}, \mathbf{w})}{\partial \mathbf{x}_{(t)}} = \mathbf{S}_z \hat{\mathbf{j}}_{(t)} \mathbf{S}_x^{-1} \in \mathbb{R}^{n_o \times n_i} \quad (4.115)$$

⁴³ Es sei noch einmal hervorgehoben, dass die Gewichte \mathbf{w} für alle Zeitschritte konstant sind.

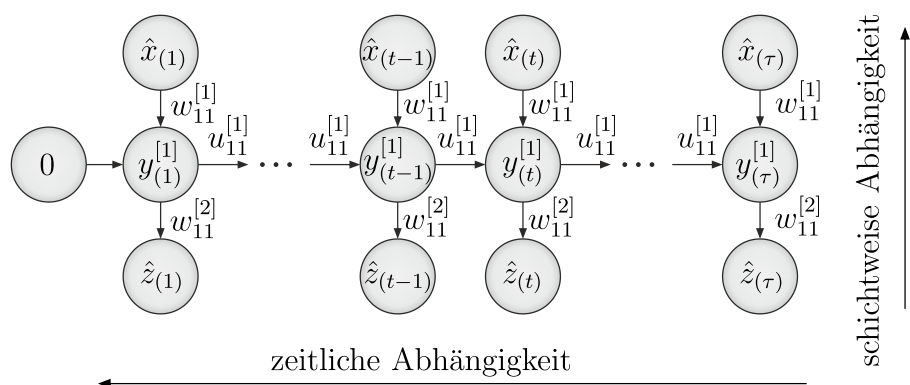


Abbildung 4.9: Ein in der Zeit aufgeklapptes RNN mit einem Zwischenneuron und jeweils einer Eingangs- und Ausgangsvariablen zur Verdeutlichung der Informationsverarbeitung

berechnet werden, die ebenfalls von den Zwischenzuständen $\bar{\mathbf{y}}_{(t-1)}^{[L]}$ abhängig ist. Durch die Unabhängigkeit der aufeinanderfolgenden Eingangsvariablen aus Gleichung (4.111) kann zur Berechnung der RNN-JACOBI-Matrix $\mathbf{j}_{(t)}$ der FNN-Algorithmus aus Tafel 4.2 oder die kompakte Gleichung (4.33) verwendet werden. Dabei muss die korrekte gewichtete Summe (4.110) benutzt werden.

Der RNN-Fehlergradient: der BP-Trough-Time-Algorithmus

Durch die temporale Abhängigkeit wird im Folgenden der Trainingsdatenfehler

$$\mathcal{L}_T(\mathbf{w}) = \frac{1}{P_T} \sum_{k=1}^{P_T} \frac{1}{\tau_k} \sum_{t=1}^{\tau_k} \mathcal{L}_{k(t)} = \frac{1}{P_T} \sum_{k=1}^{P_T} \frac{1}{\tau_k} \sum_{t=1}^{\tau_k} \frac{1}{2} \sum_{o=1}^{n_o} \left(\hat{z}_o(\mathbf{x}_{(t)k}, \mathbf{w}) - \hat{t}_{o(t)k} \right)^2 \quad (4.116)$$

definiert, wobei die gegebenen Trainingsdaten in P_T geordneten Zeitreihen

$$T = \{T_k\} = \{\{\mathcal{X}_k, \mathcal{T}_k\}\} \quad k = 1, \dots, P_T \quad (4.117)$$

nach den Definitionen (4.112) und (4.113) gegeben sind. Bei der Berechnung der Gradiententerme muss neben den Abhängigkeiten zwischen aufeinanderfolgenden Schichten auch die Abhängigkeit aufeinanderfolgender Zeitschritte beachtet werden. Dazu wird im Folgenden der BPTT-Algorithmus (engl.: *Backpropagation through time*) beschrieben, welcher von WERBOS [192] im Jahr 1988 vorgestellt wurde. Zum besseren Verständnis ist es zweckmäßig, die Informationsverarbeitung eines einfachen RNN in einer *aufgeklappten* Darstellung zu betrachten, siehe Abbildung 4.9. In die Definition der zeitabhängigen Delta-Werte

$$\delta_{mk(t)}^{[L]} := \frac{1}{\tau_k} \frac{\partial \mathcal{L}_{k(t)}}{\partial s_{mk(t)}^{[L]}} \quad (4.118)$$

wird hier die Anzahl der Zeitschritte der Zeitreihe \mathcal{X}_k eingepflegt. Daraus folgt die Aktualisierungsvorschrift für beide Gewichtskategorien zu

$$\frac{\partial \mathcal{L}}{\partial w_{ml}^{[L]}} = \frac{1}{P_T} \sum_{k=1}^{P_T} \sum_{t=1}^{\tau_k} \delta_{mk(t)}^{[L]} y_{lk(t)}^{[L-1]} \quad \text{und} \quad (4.119)$$

$$\frac{\partial \mathcal{L}}{\partial u_{mh}^{[L]}} = \frac{1}{P_T} \sum_{k=1}^{P_T} \sum_{t=1}^{\tau_k} \delta_{mk(t)}^{[L]} y_{hk(t)}^{[L]}. \quad (4.120)$$

Die Startwerte für den BPTT-Algorithmus

$$\delta_{jk(t)}^{[n_h+1]} = \frac{1}{\tau_k} g'_o(s_{jk(t)}^{[n_h+1]}) \frac{\partial \mathcal{L}_{k(t)}}{\partial \hat{z}_j} = \frac{1}{\tau_k} g'_o(s_{jk(t)}^{[n_h+1]}) (\hat{z}_j(\mathbf{x}_{(t)k}, \mathbf{w}) - \hat{t}_{j(t)k}) \quad (4.121)$$

können für alle Zeitschritte parallel berechnet werden, da in den Ausgangsschichten keine rekurrenten Verbindungen definiert sind. Die Aktualisierungsvorschrift

$$\delta_{mk(t)}^{[L]} = g'(s_{mk(t)}^{[L]}) \left[\sum_{r=1}^{n_{L+1}} \delta_{rk(t)}^{[L+1]} w_{rm}^{[L+1]} + \sum_{h=1}^{n_L} \delta_{hl(t+1)}^{[L]} u_{hm}^{[L]} \right] \quad (4.122)$$

wird durch Berücksichtigung der Abhängigkeiten in Zeit- und Schichtrichtung hergeleitet, vgl. Abbildung 4.9. Die Ausführung geschieht rückwärts, sowohl für die Schichten, als auch für die Zeitpunkte. Dadurch ist ersichtlich, dass die Fehlerrückführung mindestens zwei (nicht parallelisierbare) Schleifen benötigt, eine über die Schichten (vgl. Tafel 4.4) und eine über die Zeitschritte. Die auf die einzelnen Datenpunkte \mathbf{x}_k bezogenen RNN-Größen lassen sich folglich nicht pro Zeitreihe in einer Vektor-Matrix-Notation zusammenfassen. Daher wird im Folgenden der Algorithmus so abgewandelt, dass die voneinander unabhängigen Zeitreihen in die Felder einsortiert werden.

Die Ableitungen der Aktivierungsfunktionen $g'(s_{mk(t)}^{[L]})$, die Delta-Werte $\delta_{mk(t)}^{[L]}$ und die Neuronenausgaben $y_{mk(t)}^{[L]}$ werden für einen Zeitpunkt t in

$$\mathbf{G}'_{(t)}^{[L]} := [\mathbf{g}'(\mathbf{s}_{1(t)}^{[L]}), \dots, \mathbf{g}'(\mathbf{s}_{k(t)}^{[L]}), \dots, \mathbf{g}'(\mathbf{s}_{P(t)}^{[L]})] \in \mathbb{R}^{n_L \times P_T}, \quad (4.123)$$

$$\boldsymbol{\delta}_{(t)}^{[L]} := \begin{bmatrix} \delta_{11(t)}^{[L]} & \cdots & \delta_{1P(t)}^{[L]} \\ \vdots & \delta_{mk(t)}^{[L]} & \vdots \\ \delta_{n_L 1(t)}^{[L]} & \cdots & \delta_{n_L P(t)}^{[L]} \end{bmatrix} \in \mathbb{R}^{n_L \times P_T} \quad \text{und} \quad (4.124)$$

$$\mathbf{Y}_{(t)}^{[L]} := [\mathbf{y}_{1(t)}^{[L]}, \dots, \mathbf{y}_{k(t)}^{[L]}, \dots, \mathbf{y}_{P(t)}^{[L]}] \in \mathbb{R}^{(n_L+1) \times P_T} \quad (4.125)$$

zusammengefasst, wobei $\mathbf{G}'_{o(t)}$ sinngemäß für die letzte Schicht $L = n_h + 1$ und alle Zeitpunkte gilt. Die Vorwärtsberechnung kann für alle Zeitreihen parallel durchgeführt werden. Dabei ist es für die folgenden Ausführungen unerheblich, ob die Zeitreihen verschiedene Längen τ_k haben. Die für alle Zeitreihen maximale Länge

wird mit τ_m bezeichnet. Für einen beliebigen Zeitpunkt $t < \tau_m$ und für Schicht $L < n_h + 1$ lässt sich die Aktualisierungsvorschrift (4.122) kompakt schreiben über

$$\boldsymbol{\delta}_{(t)}^{[L]} = \mathbf{G}'_{(t)}^{[L]} \circ \left(\bar{\mathbf{W}}^{[L+1]T} \boldsymbol{\delta}_{(t)}^{[L+1]} + \mathbf{U}^{[L]T} \boldsymbol{\delta}_{(t+1)}^{[L]} \right). \quad (4.126)$$

Durch die Definition der zeitabhängigen Fehlermatrix

$$\mathbf{D}_{(t)} := \begin{bmatrix} \frac{1}{\tau_1} \frac{\partial \mathcal{L}_{1(t)}}{\partial \hat{z}_1} & \cdots & \frac{1}{\tau_P} \frac{\partial \mathcal{L}_{P_T(t)}}{\partial \hat{z}_1} \\ \vdots & \frac{1}{\tau_k} \frac{\partial \mathcal{L}_{k(t)}}{\partial \hat{z}_j} & \vdots \\ \frac{1}{\tau_1} \frac{\partial \mathcal{L}_{1(t)}}{\partial \hat{z}_{n_o}} & \cdots & \frac{1}{\tau_P} \frac{\partial \mathcal{L}_{P_T(t)}}{\partial \hat{z}_{n_o}} \end{bmatrix} \in \mathbb{R}^{n_o \times P_T} \quad (4.127)$$

können die Initialwerte für alle Zeitpunkte parallel mit

$$\boldsymbol{\delta}_{(t)}^{[n_h+1]} = \mathbf{D}_{(t)} \circ \mathbf{G}'_{o(t)} \quad (4.128)$$

berechnet werden. Wenn zum Zeitpunkt t eine Zeitreihe \mathcal{X}_k keine Daten mehr vorweisen kann, sind die zugehörigen Spalteneinträge null. Die Gradientenkomponenten werden für die Gewichte $\mathbf{W}^{[L]}$ in $\nabla \mathcal{L}_W^{[L]}$ aus Definition (4.58) und sinngemäß für $\mathbf{U}^{[L]}$ in $\nabla \mathcal{L}_U^{[L]}$ zusammengefasst und zu Beginn des Algorithmus genullt. Über die Gleichungen

$$\nabla \mathcal{L}_W^{[L]} = \nabla \mathcal{L}_W^{[L]} + \frac{1}{P_T} \boldsymbol{\delta}_{(t)}^{[L]} \mathbf{Y}_{(t)}^{[L-1]T} \quad \text{und} \quad (4.129)$$

$$\nabla \mathcal{L}_U^{[L]} = \nabla \mathcal{L}_U^{[L]} + \frac{1}{P_T} \boldsymbol{\delta}_{(t)}^{[L]} \mathbf{Y}_{(t-1)}^{[L]T} \quad (4.130)$$

werden sie aktualisiert. Der zugehörige Algorithmus ist in Tafel 4.7 skizziert. Die Vorwärtsrechnung kann dabei auch für alle Zeitreihen parallel erfolgen.

Eine Alternative zum BPTT-Algorithmus ist der RTRL-Algorithmus (engl.: *Real Time Recurrent Learning*), vgl. ROBINSON und FALLSIDE [150] oder WILLIAMS und ZIPSER [195]. Er ist darauf ausgelegt, den Gradienten simultan zur Vorwärtsberechnung zu berechnen. Der Speicherbedarf ist geringer als beim BPTT-Algorithmus, allerdings sind mehr Rechenoperationen notwendig. Der RTRL-Algorithmus wird im Rahmen der Arbeit nicht verwendet.

Weitere RNN-Architekturen und das Problem der Langzeiteffekte

Das hier ausführlich beschriebene ELMAN-RNN kann als flexiblere Version der JORDAN- und HOPFIELD-RNNs gesehen werden. Ersteres wurde im Jahr 1986 von JORDAN [93] vorgestellt und besitzt ausschließlich rekurrente Verbindungen von der Ausgangsschicht zur ersten Zwischenschicht. Das im Jahr 1982 von HOPFIELD [83] vorgestellte RNN besitzt nur eine Schicht, welche gleichzeitig als Eingangs- und Ausgangsschicht fungiert. Alle diese RNN-Architekturen leiden am

Vorwärtsrechnung	$\mathcal{Z}_k = \mathcal{F}^{\text{RNN}}(\mathcal{X}_k, \mathbf{w}), k = 1, \dots, P_T$ sowie $\mathbf{G}'_{o(t)}, \mathbf{G}'_{(t)}^{[L]}$ und $\mathbf{Y}_{(t)}^{[L]} \forall L, t$
Fehlerterme	$\mathbf{D}_{(t)}, \forall t$
Initialisierung	$\nabla \mathcal{L}_W^{[L]} = \mathbf{0}$ und $\nabla \mathcal{L}_U^{[L]} = \mathbf{0} \forall L$
Ausgangsschicht	for $t = 1, \dots, \tau_m$ do $\delta_{(t)}^{[n_h+1]} = \hat{\mathbf{D}}_{(t)} \circ \mathbf{G}'_{o(t)}$ $\nabla \mathcal{L}_W^{[n_h+1]} = \nabla \mathcal{L}_W^{[n_h+1]} + (1/P_T) \delta_{(t)}^{[n_h+1]} \mathbf{Y}_{(t)}^{[n_h]T}$ end for
Schichtschleife	for $L = n_h, \dots, 1$ do $\delta_{(\tau_m)}^{[L]} = \mathbf{G}'_{(\tau_m)}^{[L]} \circ (\bar{\mathbf{W}}^{[L+1]T} \delta_{(\tau_m)}^{[L+1]})$ $\nabla \mathcal{L}_W^{[L]} = \nabla \mathcal{L}_W^{[L]} + (1/P_T) \delta_{(\tau_m)}^{[L]} \mathbf{Y}_{(\tau_m)}^{[L-1]T}$ $\nabla \mathcal{L}_U^{[L]} = \nabla \mathcal{L}_U^{[L]} + (1/P_T) \delta_{(\tau_m)}^{[L]} \mathbf{Y}_{(\tau_m-1)}^{[L]T}$ for $t = \tau_m - 1, \dots, 1$ do $\delta_{(t)}^{[L]} = \mathbf{G}'_{(t)}^{[L]} \circ (\bar{\mathbf{W}}^{[L+1]T} \delta_{(t)}^{[L+1]} + \mathbf{U}^{[L]T} \delta_{(t+1)}^{[L]})$ $\nabla \mathcal{L}_W^{[L]} = \nabla \mathcal{L}_W^{[L]} + (1/P_T) \delta_{(t)}^{[L]} \mathbf{Y}_{(t)}^{[L-1]T}$ $\nabla \mathcal{L}_U^{[L]} = \nabla \mathcal{L}_U^{[L]} + (1/P_T) \delta_{(t)}^{[L]} \mathbf{Y}_{(t-1)}^{[L]T}$ end for end for
Umsortierung	$\nabla \mathcal{L}_W^{[L]}, \nabla \mathcal{L}_U^{[L]} \rightarrow \nabla \mathcal{L}, \forall L$

Tafel 4.7: Rückwärtsrechnung für den RNN-Fehlergradienten $\nabla \mathcal{L}$

Phänomen des *explodierenden* oder *verschwindenden Gradienten*. Durch die multiplikative Operation in Gleichung (4.126) werden die Gewichte so oft miteinander multipliziert, wie es Zeitschritte in der Zeitreihe gibt. Dies kann zu einem der genannten Probleme führen, abhängig vom Wertebereich der Gewichte. Bei Nutzung von Trainingsverfahren höherer Ordnung stellt dies durch die Liniensuchmethoden zumindest aus numerischer Sicht zunächst kein maßgebliches Problem dar. Eine alternative Lösungsstrategie sieht vor die Aktualisierungsschleifen (4.129) und (4.130) frühzeitig abubrechen. Dies ist als *Gradient Clipping* bekannt. Dadurch sind jedoch Verfahren höherer Ordnung nicht mehr brauchbar. Die Gradienten $\nabla \mathcal{L}$ sind dadurch nicht mehr exakt. Im Rahmen stochastischer Optimierungsverfahren stellt dies allerdings nur ein untergeordnetes Problem dar. Moderne RNN-Architekturen wie das LSTM-Netzwerk (engl.: *Long Short Term Memory*) wurden entwickelt⁴⁴, um dieses Problem zu umgehen, vgl. HOCHREI-

⁴⁴ Im Jahr 1991 hat SEPP HOCHREITER in seiner Diplomarbeit an der TU München das Problem der verschwindenden Gradienten untersucht. Vier Jahre später veröffentlichte er mit seinem Betreuer JÜRGEN SCHMIDHUBER die erste Arbeit zum LSTM.

TER und SCHMIDHUBER [80]. Auf sie wird im Rahmen dieser Arbeit nicht weiter eingegangen. Allerdings sind sie für die Materialmodellierung außerordentlich vielversprechend, da sie Kurz- und Langzeiteffekte erlernen können, ohne an dem genannten Problem zu leiden.

4.4 Trainingsalgorithmen für KNN

In diesem Abschnitt wird die iterative Lösung des nichtlinearen Minimierungsproblems (4.4) beschrieben. Dabei ist es im Folgenden unerheblich, mit welcher KNN-Architektur und mit welchen Termen die zu minimierende Fehlerfunktion $\mathcal{L}(\mathbf{w})$ und ihr Gradient $\nabla\mathcal{L}(\mathbf{w})$ berechnet werden.

Zu Beginn des Trainings sind die Gewichte \mathbf{w}_0 zu initialisieren. Hierzu werden normalverteilte, von der Anzahl der Neuronen abhängige Zufallszahlen

$$\mathbf{w}_{ml0}^{[L]} \sim \mathcal{N}\left(0, \frac{1}{n_{L-1}}\right) \quad (4.131)$$

genutzt ($n_0 = n_i$), auf welche die Parameter des Tangens hyperbolicus (vgl. Abbildung 4.4) angepasst sind⁴⁵. Dadurch liegen die gewichteten Summen zu Beginn möglichst im dynamischen Bereich und nicht in den Plateaubereichen, vgl. auch LE CUN ET AL. [110]. Im Laufe des Trainings werden die Gewichte über

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \alpha_n \mathbf{d}_n \quad (4.132)$$

aktualisiert, mit der *Schrittweite* α_n und der *Schrittrichtung* \mathbf{d}_n . Deren Spezifizierung wird vom jeweiligen Trainingsverfahren festgelegt. Dazu folgen zunächst mehrere Definitionen: Ein Intervall innerhalb des Trainingsprozesses, in dem die Menge aller Trainingsdaten $T = \{(\mathbf{x}_k, \mathbf{t}_k)\}_{k=1}^{P_T}$ genutzt wird, heißt *Epoche*. Ein *Offline*-Training liegt vor, wenn die Gewichtsaktualisierung (4.132) innerhalb der Epoche ein einziges Mal durchgeführt wird – unter Verwendung aller Trainingsdaten. Der zugehörige Gradient ist in diesem Fall der exakte Gradient des Optimierungsproblems (4.4); die Verfahren werden als Gradientenverfahren bezeichnet. Erfolgt die Berechnung des Gradienten und folglich die Gewichtsaktualisierung für jedes Datenpaar einzeln, wird von *Online*-Training gesprochen. Ein *Batch*-Training liegt vor, wenn die Trainingsdaten T zufällig in n_B disjunkte Untermengen $T = T_1 \cup T_2 \cup \dots \cup T_{n_B}$ aufgeteilt werden, mit bspw. $P_B = \lceil P_T/n_B \rceil$ Datenpaaren $\{(\mathbf{x}_k, \mathbf{t}_k)\}_{k=1}^{P_B}$. Dabei rundet die Operation $\lceil \bullet \rceil$ auf die nächste ganze Zahl auf. Die Differenz zur Gesamtanzahl P_T wird durch das letzte Batch kom-

⁴⁵ Der dynamische Bereich der SoftPlus-Funktion ist größer, wodurch auch für diese Aktivierungsfunktion die Initialisierung (4.131) angemessen ist.

pensiert⁴⁶. Die Aktualisierung wird für jedes Batch durchgeführt. Bei Batch- und Online-Training wird von *stochastischen* Gradientenverfahren gesprochen.

4.4.1 Das Adam-Optimierungsverfahren

Der Adam-Optimierer (engl.: *Adaptive moment estimation*) wurde im Jahr 2015 von KINGMA und BA [100] vorgestellt. Dieser Algorithmus ist ein Vertreter der stochastischen Gradientenverfahren. Die Grundidee liegt in der Betrachtung des Fehlers \mathcal{L} und dessen Gradienten $\nabla\mathcal{L}$ als stochastische Funktionen: Nicht \mathcal{L} , sondern der Erwartungswert $\mathbb{E}(\mathcal{L})$ wird minimiert. Die Schritttrichtung

$$\mathbf{d}_n = -\mathbb{E}(\nabla\mathcal{L})_n \oslash \sqrt{\mathbb{E}(\nabla\mathcal{L} \circ \nabla\mathcal{L})_n} \quad (4.133)$$

ist definiert als der negative Erwartungswert des Gradienten, normiert mit der Wurzel des Erwartungswerts des quadrierten Gradienten⁴⁷. Die mit \oslash gekennzeichnete Division wird komponentenweise ausgeführt. Beide (unbekannten) Erwartungswerte werden im Aktualisierungsschritt n über ihre exponentiell geglätteten Mittelwerte

$$\mathbb{E}(\nabla\mathcal{L})_n \approx \mathbf{m}_n = \beta_1 \mathbf{m}_{n-1} + (1 - \beta_1) \nabla\mathcal{L}_n \quad (4.134)$$

$$\mathbb{E}(\nabla\mathcal{L} \circ \nabla\mathcal{L})_n \approx \mathbf{v}_n = \beta_2 \mathbf{v}_{n-1} + (1 - \beta_2) (\nabla\mathcal{L}_n \circ \nabla\mathcal{L}_n) \quad (4.135)$$

abgeschätzt. Dies motiviert letztlich den Namen des Algorithmus. Mit $\nabla\mathcal{L}_n$ wird $\nabla\mathcal{L}(\mathbf{w}_n)$ abgekürzt. Die Parameter $\beta_1 \in [0, 1)$ und $\beta_2 \in [0, 1)$ müssen neben der festen Schrittweite α vom Benutzer gewählt werden und steuern die Stärke der exponentiellen Glättung. Die Vektoren \mathbf{m}_{-1} und \mathbf{v}_{-1} werden jeweils zu null initialisiert. Gerade zu Beginn des Trainings erhalten die Vektoren \mathbf{m}_n und \mathbf{v}_n dadurch eine starke Tendenz zu $\mathbf{0}$. Daher werden die Mittel (4.134) und (4.135) korrigiert, was letztlich zu einer Anpassung der Schrittweite

$$\alpha_n = \alpha \frac{\sqrt{(1 - \beta_2^n)}}{(1 - \beta_1^n)} \quad (4.136)$$

führt. Damit folgt die Aktualisierungsvorschrift

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \alpha_n \mathbf{m}_n \oslash (\sqrt{\mathbf{v}_n} + \hat{\epsilon} \mathbf{1}^{nw}), \quad (4.137)$$

wobei auf jede Komponente von $\sqrt{\mathbf{v}_n}$ eine kleine Zahl $0 < \hat{\epsilon} \ll 1$ addiert wird, um nicht ggf. durch null zu teilen. Der gesamte Algorithmus inklusive Batch-Management ist in Tafel 4.8 gegeben. In den Beispielen werden die Parameter für

⁴⁶ Durch die Aufrundung können insgesamt weniger Batches berechnet werden, als zuvor definiert. Dies kann jedoch algorithmisch leicht abgefangen werden – falls gewünscht.

⁴⁷ Diese Normierung wird auch in anderen Trainingsverfahren (Rprop, RMSProp) genutzt. Grundsätzlich basiert die Idee darauf, nicht die absoluten Werte, sondern nur die Vorzeichen der Gradientenkomponenten zur Gewichtsaktualisierung zu nutzen ($x/\sqrt{x^2} = \text{sign}(x)$).

Wahl Parameter	$\alpha > 0, \beta_1 \in [0, 1), \beta_2 \in [0, 1), 0 < \hat{\epsilon} \ll 1, n_B \geq 1$
Initialisierung	$\mathbf{w}_0, \mathbf{m}_{-1} = \mathbf{0}, \mathbf{v}_{-1} = \mathbf{0}, \text{epo} = 0, n = 0$
Training	while (kein Abbruch) do
	$\bar{T} = T$
	for $b = 1, \dots, n_B$ do
Batch-Management	$P_B = \min(\bar{T} , \lceil P/n_B \rceil)$
	$T_b = \{(\mathbf{x}_k, \mathbf{t}_k)\}_{k=1}^{P_B} \subset \bar{T}$ zufällig, $\bar{T} \leftarrow \bar{T} \setminus T_b$
Adam lokal	$\mathcal{L}_n, \nabla \mathcal{L}_n$ bzgl. Batch T_b (KNN)
	$\mathbf{m}_n = \beta_1 \mathbf{m}_{n-1} + (1 - \beta_1) \nabla \mathcal{L}_n$
	$\mathbf{v}_n = \beta_2 \mathbf{v}_{n-1} + (1 - \beta_2) (\nabla \mathcal{L}_n \circ \nabla \mathcal{L}_n)$
	$\alpha_n = \alpha \sqrt{(1 - \beta_2^n) / (1 - \beta_1^n)}$
	$\mathbf{w}_{n+1} = \mathbf{w}_n - \alpha_n \mathbf{m}_n \odot (\sqrt{\mathbf{v}_n} + \hat{\epsilon} \mathbf{1}^{n_w})$
Update-Zähler	$n \leftarrow n + 1$
	end for
	$\text{epo} \leftarrow \text{epo} + 1$
	end while

Tafel 4.8: Adam-Optimierer mit Batch-Management innerhalb der Epochen

das Adam-Verfahren stets nach Tabelle 4.1 gewählt, vgl. KINGMA und BA [100].

Schrittweite	Mittelung \mathbf{m}	Mittelung \mathbf{v}	Numerik	Anzahl Batches
α	β_1	β_2	$\hat{\epsilon}$	n_B
0,001	0,9	0,999	10^{-8}	100

Tabelle 4.1: Parameterwahl für den Adam-Optimierer

4.4.2 Das Quasi-Newton Optimierungsverfahren

Verfahren wie der Adam-Optimierer werden heutzutage häufig genutzt, da sie wenig Speicherplatz benötigen und durch die stochastische Komponente weniger stark durch lokale Minima gefährdet sind. In Anwendungsgebieten wie Bild- oder Spracherkennung, bei denen Millionen von Gewichten optimiert werden müssen, wiegen diese Vorteile schwer. Im Rahmen der Materialmodellierung sind die Netze allerdings kleiner und die zu optimierenden Fehlerfunktionen glatter. Außerdem sind die Anforderungen an die lokale Genauigkeit der Approximation wesentlich höher, da durch die zu berechnende Materialtangente auch die Ableitungen des KNN gut approximiert sein müssen. Daher ist die Anwendung sogenannter

Verfahren höherer Ordnung hier wesentlich interessanter, während sie in anderen Anwendungsgebieten schon längst keine Berücksichtigung mehr finden. Das Quasi-NEWTON-Verfahren (QN) hat sich dazu in vielen numerischen Untersuchungen des Autors als besonders geeignet herausgestellt. Es stellt einen guten Kompromiss zwischen Genauigkeit und Effizienz dar. Ausführliche Grundlagen sind bspw. in GEIGER und KANZOW [56] zu finden. Weitere Verfahren höherer Ordnung zum Training von KNN sind bspw. das CG-Verfahren (vgl. JOHANSSON ET AL. [91]) sowie das LEVENBERG-MARQUARDT-Verfahren (vgl. HAGAN und MENHAJ [69]). Vgl. auch SHEPHERD [157] für eine allgemeine Übersicht zu Verfahren höherer Ordnung im Rahmen des KNN-Trainings.

Das QN-Verfahren ist ein Offline-Verfahren mit $n_B = 1$ und $P_B = P$. Der Fehler \mathcal{L}_n und sein Gradient $\nabla \mathcal{L}_n$ werden innerhalb der Epoche $n = \text{epo}$ stets mit allen Trainingsdaten berechnet. Grundsätzlich basiert das Verfahren auf der Idee, in der Aktualisierungsvorschrift des klassischen NEWTON-Verfahrens

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \alpha_n \mathbf{d}_n = \mathbf{w}_n - \alpha_n \left(\nabla^2 \mathcal{L}_n \right)^{-1} \nabla \mathcal{L}_n \quad (4.138)$$

die exakte HESSE-Matrix $\nabla^2 \mathcal{L}_n = \nabla^2 \mathcal{L}(\mathbf{w}_n)$ durch eine Approximation \mathbf{H}_n zu ersetzen. Eine effizientere Abwandlung approximiert direkt die Inverse der HESSE-Matrix, mit $\mathbf{B}_n := \mathbf{H}_n^{-1}$, sodass lokal weder eine Invertierung durchgeführt noch ein lineares Gleichungssystem gelöst werden muss. Somit gilt für die Schrittichtung

$$\mathbf{d}_n = -\mathbf{B}_n \nabla \mathcal{L}_n . \quad (4.139)$$

Wird $\mathbf{B}_0 = \mathbf{I}$ initialisiert, entspricht die erste Suchrichtung dem negativen Gradienten und somit der Richtung des lokal steilsten Abstiegs. In jeder Epoche wird nach Aktualisierung der Gewichte die neue Approximation $\mathbf{B}_{n+1}(\mathbf{B}_n, \dots)$ berechnet. Die zugehörige Formel muss die sogenannte (inverse) QN-Gleichung

$$\mathbf{B}_{n+1} (\nabla \mathcal{L}_{n+1} - \nabla \mathcal{L}_n) = \mathbf{w}_{n+1} - \mathbf{w}_n \quad (4.140)$$

erfüllen. Sie garantiert u. a. lokale super-lineare Konvergenz, siehe GEIGER und KANZOW [56]. Dies ist gerade bei der Materialmodellierung ein großer Vorteil gegenüber dem Adam-Verfahren, und zwar aufgrund der oben genannten hohen Genauigkeitsanforderungen. Aus der Fülle der in der Literatur zur Verfügung stehenden Möglichkeiten wird in dieser Arbeit auf die inverse BFGS-Formel⁴⁸

$$\mathbf{B}_{n+1}^{\text{BFGS}} = \mathbf{B}_n + \frac{(\mathbf{s} - \mathbf{B}_n \mathbf{y}) \mathbf{s}^T + \mathbf{s} (\mathbf{s} - \mathbf{B}_n \mathbf{y})^T}{\mathbf{y}^T \mathbf{s}} - \frac{(\mathbf{s} - \mathbf{B}_n \mathbf{y})^T \mathbf{y} \mathbf{s}^T}{(\mathbf{y}^T \mathbf{s})^2} \quad (4.141)$$

⁴⁸ BFGS steht für die Mathematiker BROYDEN [24], FLETCHER [45], GOLDFARB [63] und SHANNO [155], die diese Methode unabhängig voneinander publiziert haben. Für weitere Details, vgl. GEIGER und KANZOW [56].

zurückgegriffen, mit den Abkürzungen $\mathbf{s} := \mathbf{w}_{n+1} - \mathbf{w}_n$ und $\mathbf{y} := \nabla \mathcal{L}_{n+1} - \nabla \mathcal{L}_n$. Zur Korrektur etwaiger numerischer Fehlerakkumulationen wird empfohlen, das QN-Training bspw. alle n_w Epochen *neu zu starten*, indem \mathbf{B} wieder zur Einheitsmatrix \mathbf{I} initialisiert wird.

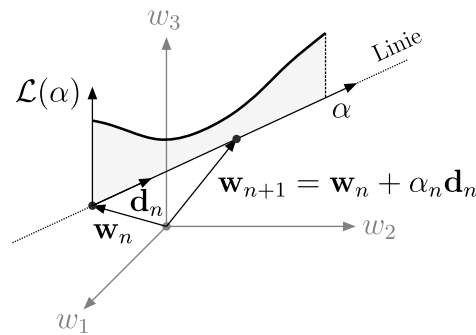


Abbildung 4.10: Prinzip der Liniensuche in einem dreidimensionalen Raum $\mathbf{w} \in \mathbb{R}^3$

Die Schrittweite α_n wird nicht fest vorgegeben, sondern durch eine inexakte Liniensuche (LS) bestimmt, deren Prinzip in Abbildung 4.10 dargestellt ist. Sie ist *inexakt*, weil nicht das exakte Minimum

$$\alpha_{min} = \arg \min_{\alpha > 0} \mathcal{L}(\mathbf{w}_n + \alpha \mathbf{d}_n) \quad (4.142)$$

auf der durch \mathbf{w}_n und \mathbf{d}_n aufgespannten Linie gesucht wird, sondern eine Lösung α_n , welche den Funktionswert mindestens reduziert und dabei gewisse Eigenschaften erfüllt. Dazu muss die Fehlerfunktion $\mathcal{L}(\mathbf{w})$ und ihr Gradient $\nabla \mathcal{L}(\mathbf{w})$ ggf. mehrmals in einer Epoche berechnet werden. Eine *exakte* Liniensuche, welche das echte Minimum sucht, wäre überdies numerisch zu aufwendig. In Verbindung mit dem QN-Verfahren wird eine Liniensuche entweder mit der ARMIJO-Regel oder mit der WOLFE-POWELL-Regel durchgeführt. Letztere hat in Verbindung mit dem QN-Verfahren den Vorteil, dass sie die positive Definitheit der einzelnen \mathbf{B}_n und damit die Lösbarkeit des linearen Gleichungssystems (4.138) garantiert. Die ARMIJO-Regel ist allerdings stabiler, erlaubt kleinere Schrittweiten und eignet sich daher bspw. gut für die erste Epoche, in der es je nach Zusammensetzung der $\mathcal{L}(\mathbf{w})$ -Terme zu numerischen Problemen kommen kann. Daher wird in dieser Arbeit der in Tafel 4.9 beschriebene QN-Algorithmus verwendet, welcher die genannten Liniensuchen gestaffelt einsetzt. Auch die WP-LS wird in eine ‚scharfe‘ und eine ‚schnelle‘ Einstellung unterteilt. Die Parameter der jeweiligen Liniensuchen sind in Tabelle 4.2 gegeben. Beide Liniensuch-Strategien sind im Anhang C zusammengefasst.

Initialisierung	$epo = n = 0, \mathbf{w}_0, \mathbf{B}_0 = \mathbf{I}, \mathcal{L}_0 = (\mathbf{w}_0), \nabla \mathcal{L}_0 = \nabla(\mathbf{w}_0)$
Training	ARMIJO-LS nach Tafel C.1 $\rightarrow \alpha_0$ $\mathbf{w}_1 = \mathbf{w}_0 - \alpha_0 \mathbf{B}_0 \nabla \mathcal{L}_0$ $\mathbf{B}_1(\mathbf{B}_0, \dots)$ mit BFGS-Formel (4.141) while (kein Abbruch) do
Gestaffelte LS	<i>schnelle</i> WOLFE-POWELL-LS-1 if (LS-1 erfolgreich) then $\rightarrow \alpha_n$ else <i>scharfe</i> WL-LS-2 if (LS-2 erfolgreich) then $\rightarrow \alpha_n$ else $\mathbf{B}_n = \mathbf{I}$ (interner Neustart) <i>scharfe</i> WL-LS-3 if (LS-3 erfolgreich) then $\rightarrow \alpha_n$ else STOP Training end if end if
Gewichts-Update	$\mathbf{w}_{n+1} = \mathbf{w}_n - \alpha_n \mathbf{B}_n \nabla \mathcal{L}_n$
Neustart	if $(\text{mod}(n+1, n_w) = 0)$ then $\mathbf{B}_{n+1} = \mathbf{I}$ else $\mathbf{B}_{n+1}(\mathbf{B}_n, \dots)$ mit BFGS-Formel (4.141) end if $n \leftarrow n + 1$ end while

Tafel 4.9: QN-Optimierer mit gestaffelter Liniensuche (LS)

Liniensuche	σ	ρ	β	i_{max}	γ	τ_1	τ_1
LS-WP schnell	10^{-4}	0,2	–	10	5	0,1	0,2
LS-WP scharf	10^{-4}	0,2	–	100	1.5	0,01	0,25
Armijo scharf	10^{-6}	–	0,5	30*	–	–	–

Tabelle 4.2: Parameterwahl für die Liniensuchen innerhalb des QN-Optimierers. *Bei mehr als 30 Iterationen ist $\alpha_n < \alpha_{n-1} \cdot 10^{-10}$ und das Training wird damit abgebrochen.

4.4.3 Abbruchkriterien für das Training

Die zu optimierende Fehlerfunktion \mathcal{L} ist nichtlinear, nichtkonvex und besitzt viele lokale Minima, wodurch die Berechnung des exakten globalen Minimums \mathbf{w}_{\min} i. d. R. nicht möglich ist. Demzufolge werden Kriterien definiert, zu wel-

chem Zeitpunkt der Trainingsprozess abgebrochen wird und welche Gewichte $\hat{\mathbf{w}}$ schließlich als dessen Ergebnis hervorgehen. Neben der Menge der Trainingsdaten $T = \{(\mathbf{x}_k, \mathbf{t}_k)\}_{k=1}^{P_T}$ werden zusätzliche Datenpaare $V := \{(\mathbf{x}_k, \mathbf{t}_k)\}_{k=1}^{P_V}$ vorgehalten, mit denen in jeder Epoche epo der *Validierungsfehler*

$$\mathcal{L}_V^{epo} := \sqrt{\frac{1}{P_V} \sum_{k=1}^{P_V} \sum_{o=1}^{n_o} (z_o(\mathbf{x}_k, \mathbf{w}_{epo}) - t_{ok})^2} \quad (4.143)$$

ermittelt wird. Als Fehlerfunktion wird dazu der RMSE (engl.: *Root Mean Squared Error*) genutzt, um korrekte Einheiten zu gewährleisten. Damit ist ein direkter Vergleich zum Trainingsdatenfehler über $\sqrt{2\mathcal{L}_T}$ möglich, vgl. dessen Definition (4.15). Da die Datenpunkte aus V nicht zum Training verwendet werden, ist \mathcal{L}_V ein brauchbares Maß zur Beurteilung der Generalisierungsfähigkeit des KNN – geeigneter als der Trainingsdatenfehler \mathcal{L}_T selbst. Darauf aufbauend werden die folgenden drei Kriterien definiert, die mit den Parametern δ^{abs} , δ^{rel} und epo_{max} eingestellt werden können:

1. Ein *absolutes* Kriterium für die aktuelle Epoche epo : Abbruch, wenn im Vergleich zum Wertebereich der Trainingsdaten ein kleiner Fehler vorliegt.

$$\mathcal{L}_V^{epo} \leq \mathcal{L}_V^{abs} = \frac{\delta^{abs}}{P_T n_o} \sum_{k=1}^{P_T} \sum_{j=1}^{n_o} |t_{jk}| \quad (4.144)$$

2. Ein *relatives* Kriterium für die aktuelle Epoche $epo \geq 500$: Abbruch, wenn sich im Intervall $I_c := [\lceil \sqrt{epo} \rceil, epo]$ keine signifikanten Änderungen mehr ergeben haben.

$$\frac{6 \sum_{m=1}^{|I_c|} \mathcal{L}_V^m (1 + |I_c| - 2m)}{|I_c|(|I_c| + 1)} = \bar{\mathcal{L}}_V^{I_c} \leq \max \left\{ \frac{\delta^{rel}}{|I_c|} \sum_{m \in I_c} \mathcal{L}_V^m, \mathcal{L}_V^{abs} \right\} \quad (4.145)$$

Die Formel für $\bar{\mathcal{L}}_V^{I_c}$ stammt aus einer linearen Regression. Dabei ist $|I_c|$ die Anzahl der Elemente in der Menge der Epochen I_c .

3. Die maximale Anzahl an Epochen ist erreicht: $epo = epo_{max}$.

Die Parameter δ^{abs} , δ^{rel} und epo_{max} werden bei den Untersuchungen jeweils angegeben. Ist eine dieser Bedingungen erfüllt, wird das Training abgebrochen. Diejenigen Gewichte $\hat{\mathbf{w}}$, die während des gesamten Trainingsprozesses zum kleinsten Validierungsfehler \mathcal{L}_V^{min} führen, werden abgespeichert.

Häufig wird in der Literatur neben der Validierung auch ein abschließender Test empfohlen, um das KNN mit solchen Daten zu bestätigen, die weder zum Training noch zu dessen Abbruch genutzt wurden. Diese Tests können im Rahmen der KNN-Materialmodellierung mit Materialpunkt-Versuchen durchgeführt werden. Sie sind in Bezug auf den Nutzen als Materialmodell aussagekräftiger als der Fehler bzgl. einzelner Testdaten.

4.5 KNN-Training als restringiertes Optimierungsproblem

Im Rahmen des KNN-Trainings wird klassischerweise der Trainingsdatenfehler \mathcal{L}_T nach Definition (4.15) als zu minimierende Zielfunktion definiert. Dabei erhält das KNN alle Informationen einzig über die Menge der Trainingsdatenpaare $\{(\mathbf{x}_k, \mathbf{t}_k)\}_{k=1}^{P_T}$. Unterliegen diesen Datenpaaren physikalische Eigenschaften, müssen sie implizit erlernt werden⁴⁹. Falls die Trainingsdaten jedoch – im Vergleich zur Komplexität der zu erlernenden Aufgabe – nur in geringer Zahl vorhanden sind oder starken Streuungen unterliegen, so kann dies nicht funktionieren; der Trainingsprozess ist möglicherweise nicht wohl definiert und kann scheitern. In vielen Anwendungsgebieten ist über das Verhalten der gegebenen Datenpaare mehr bekannt als nur die Werte ihrer Eingangs- und Ausgangsvariablen. Im Rahmen der Materialmodellierung sind dies bspw. physikalische Eigenschaften, die dem Materialverhalten a priori zugesprochen werden können. Daher kann es zweckmäßig sein, die Minimierung des Trainingsdatenfehlers als ein *restringiertes* Optimierungsproblem

$$\begin{aligned} \min_{\mathbf{w} \in \mathbb{R}^{n_w}} \mathcal{L}_T(\mathbf{w}) \quad \text{u. d. N.} \quad & h_i(\mathbf{w}) = 0, \quad i = 1, \dots, n_{eq} \\ & g_j(\mathbf{w}) \leq 0, \quad j = 1, \dots, n_{ie} \end{aligned} \quad (4.146)$$

zu definieren, mit u. d. N. als Abkürzung für *unter der Nebenbedingung*. Die exakte Lösung, welche die Nebenbedingungen erfüllt, wird im Folgenden mit \mathbf{w}_{min} bezeichnet. Die im Allgemeinen nichtlinearen Gleichheits- und Ungleichheits-Nebenbedingungen beschränken den Lösungsraum für die Gewichte \mathbf{w} . Sie werden im Folgenden allgemein mit NB abgekürzt. Grundlagen zur restringierten Optimierung können bspw. in GEIGER und KANZOW [57] nachgelesen werden. Sie werden im Folgenden kurz skizziert. Sind alle beteiligten Funktionen stetig differenzierbar, kann die LAGRANGE-Funktion

$$\begin{aligned} \mathcal{L}_L(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu}) & := \mathcal{L}_T(\mathbf{w}) + \sum_{i=1}^{n_{eq}} \mu_i h_i(\mathbf{w}) + \sum_{j=1}^{n_{ie}} \lambda_j g_j(\mathbf{w}) \\ & = \mathcal{L}_T(\mathbf{w}) + \boldsymbol{\mu}^T \mathbf{h}(\mathbf{w}) + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{w}) \end{aligned} \quad (4.147)$$

definiert werden, mit den unbekanntem Vektoren $\boldsymbol{\mu} = [\mu_1, \dots, \mu_{n_{ie}}]^T$ und $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_{n_{eq}}]^T$ der sog. LAGRANGE-Multiplikatoren sowie der Vektoren der Nebenbedingungen $\mathbf{h}(\mathbf{w}) = [h_1(\mathbf{w}), \dots, h_{n_{eq}}(\mathbf{w})]^T$ und $\mathbf{g}(\mathbf{w}) = [g_1(\mathbf{w}), \dots, g_{n_{ie}}(\mathbf{w})]^T$. Ein Punkt $(\mathbf{w}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, der die KARUSH-KUHN-TUCKER-Bedingungen

$$\begin{aligned} \nabla_{\mathbf{w}} \mathcal{L}_L &= \mathbf{0} \\ \nabla_{\boldsymbol{\mu}} \mathcal{L}_L &= \mathbf{h} = \mathbf{0} \\ \boldsymbol{\lambda}^* &\geq \mathbf{0} \quad \nabla_{\boldsymbol{\lambda}} \mathcal{L}_L = \mathbf{g} \leq \mathbf{0} \quad \boldsymbol{\lambda}^{*T} \mathbf{g} = 0 \end{aligned} \quad (4.148)$$

⁴⁹ In Abschnitt 7.2 wird gezeigt werden, dass diese Annahme nicht immer der Fall ist.

erfüllt, wird KKT-Punkt genannt⁵⁰. Ein solcher KKT-Punkt existiert, wenn gewisse Regularitätsvoraussetzungen erfüllt sind. Eine davon ist bspw. die lineare Unabhängigkeit aller Gradienten $\nabla h_i(\mathbf{w}^*)$ und $\nabla g_j(\mathbf{w}^*)$ (für $g_j(\mathbf{w}^*) = 0$). Dies wird LICQ (engl.: *Linear Independence Constraint Qualification*) genannt. Diese Voraussetzungen sind i. d. R. für allgemein nichtlineare Optimierungsprobleme schwer zu beweisen. Ferner stellen die KKT-Bedingungen (4.148) nur eine *notwendige* Bedingungen für das Minimum \mathbf{w}_{min} dar. Zusätzlich muss auch ein *hinreichendes* Kriterium erfüllt sein. Das ist in diesem Fall die positive Definitheit der Hessematrix $\nabla^2 \mathcal{L}_L(\mathbf{w}^*)$ innerhalb eines von den Nebenbedingungen abhängigen Tangentialkegels an \mathbf{w}^* . Für nichtkonvexe Problemstellungen⁵¹ sind diese Bedingungen zur Bestimmung eines *exakten* Minimums \mathbf{w}_{min} , welches die Nebenbedingungen $\mathbf{h} = \mathbf{0}$ und $\mathbf{g} \leq \mathbf{0}$ exakt erfüllt, ebenfalls schwer zu beweisen.

Die Nebenbedingungen im Rahmen des KNN-Trainings *exakt* einzufordern, ist von vornherein nicht sinnvoll. Das iterative Trainingsverfahren wird in der Nähe eines lokalen Minimums des Optimierungsproblems (4.149) abgebrochen, auf Basis der Kriterien aus Abschnitt 4.4.3. Daher können die Nebenbedingungen von vorneherein nicht exakt erfüllt werden. Darüber hinaus spielt bei einem KNN die Topologie eine entscheidende Rolle, ob und vor allem wie gut das KNN einen funktionalen Zusammenhang approximieren *könnte*, lägen unendlich viele Daten vor. Eine ausführliche Diskussion darüber ist in VAN DER SMAGT [180] zu finden. Es wäre folglich selbst mit einer *exakten Einforderung* der Nebenbedingungen nicht möglich zu gewährleisten, dass eine Nebenbedingung exakt erfüllt werden *kann*. Im schlimmsten Fall führt dies dazu, dass der Trainingsdatenfehler \mathcal{L}_T nicht mehr minimiert wird, durch Konkurrenz zu stark gewichteten Nebenbedingungen. So wurden bspw. in MÁRQUEZ-NEILA ET AL. [124] Nebenbedingungen mit LAGRANGE-Multiplikatoren exakt eingefordert, was gegenüber einer inexakten Einforderung zu einer Verschlechterung der Ergebnisse geführt hat.

Im Folgenden werden daher drei Methoden vorgestellt, mit denen die Nebenbedingungen im Rahmen einer restringierten Optimierung im KNN-Trainingsprozess *näherungsweise* berücksichtigt werden können. Die vermeintliche mathematische Exaktheit dieser Methoden in der Erfüllung der Nebenbedingungen spielt dabei im Vergleich der Verfahren untereinander eine untergeordnete Rolle.

Diese Methoden formen – ähnlich zur LAGRANGE-Funktion (4.147) – das restringierte Optimierungsproblem (4.146) zunächst in ein äquivalentes unrestringiertes

⁵⁰ Die dritte Zeile entspricht im Kontext des Prinzips der maximalen plastischen Dissipationsleistung gerade den Be- und Entlastungsbedingungen (3.93) aus der Elastoplastizitätstheorie, mit der Nebenbedingung $F \leq 0$. Auf diese Beobachtung wird hier nur am Rande verwiesen.

⁵¹ Die Methode der LAGRANGE-Multiplikatoren erfreut sich im Falle konvexer Optimierung großer Beliebtheit, weil dort ein KKT-Punkt automatisch ein notwendiges und ein hinreichendes Kriterium für das Minimum \mathbf{w}_{min} ist.

Optimierungsproblem

$$\min_{\mathbf{w} \in \mathbb{R}^{n_w}} \mathcal{L}(\mathbf{w}) = \min_{\mathbf{w} \in \mathbb{R}^{n_w}} (\mathcal{L}_T(\mathbf{w}) + \mathcal{L}_C(\mathbf{w})) \quad (4.149)$$

um. Im Zuge dessen muss die gesamte Fehlerfunktion $\mathcal{L} = \mathcal{L}_T + \mathcal{L}_C$ minimiert werden, bestehend aus dem Trainingsdatenfehler \mathcal{L}_T und dem Fehlerterm bzgl. der Nebenbedingungen \mathcal{L}_C . Ein wichtiger Unterschied ist jedoch, dass keine zusätzlichen Unbekannten $\boldsymbol{\lambda}$ und $\boldsymbol{\mu}$ in der Fehlerfunktion $\mathcal{L}(\mathbf{w})$ auftreten. Daher können die Trainingsalgorithmen aus Abschnitt 4.4 uneingeschränkt auch für das Training unter Nebenbedingungen genutzt werden. Im Folgenden wird stets nur von einer einzelnen Gleichheits-Nebenbedingung $h(\mathbf{w}) = 0$ und von einer einzelnen Ungleichheits-Nebenbedingung $g(\mathbf{w}) \leq 0$ ausgegangen. Die Erweiterung auf beliebig viele Nebenbedingungen ist durch Summation der jeweiligen Funktionen uneingeschränkt möglich.

4.5.1 Die klassische L2-Penalty-Methode

Die klassische L2-Penalty-Methode (oder nur Penalty-Methode) definiert den zusätzlichen Term \mathcal{L}_C über

$$\mathcal{L}_C^{PM}(\mathbf{w}) := \frac{\epsilon}{2} \left[h(\mathbf{w})^2 + \left(\max(0, g(\mathbf{w})) \right)^2 \right]. \quad (4.150)$$

Die Funktion $\mathcal{L}_C^{PM}(\mathbf{w})$ ist durch die Maximumsfunktion nur einmal stetig differenzierbar. Dies stellt mit den vorgestellten Optimierungsalgorithmen jedoch nur ein untergeordnetes Problem dar. Der *Penalty-Parameter* $\epsilon > 0$ steuert die Härte der Bestrafung. Je größer dieser gewählt wird, desto stärker werden die Nebenbedingungen h_i und g_j gegenüber der Fehlerfunktion \mathcal{L}_T gewichtet.

Nur für $\epsilon \rightarrow \infty$ sind die Optimierungsprobleme (4.146) und (4.149) äquivalent, die Nebenbedingungen *exakt* eingefordert und die zugehörigen Lösungsvektoren \mathbf{w}_{min} identisch. Ein Penalty-Parameter $\epsilon < \infty$ stellt somit immer eine Approximation dar. Auf der anderen Seite wachsen für $\epsilon \rightarrow \infty$ die Beträge der Eigenwerte von $\nabla^2 \mathcal{L}$ über alle Schranken. Innerhalb des Optimierungsverfahrens kann dies zu numerischen Problemen führen. Daher ist im Allgemeinen die Wahl des Penalty-Parameters stets eine Abwägung aus Exaktheit der Nebenbedingungen und der numerischen Umsetzbarkeit. Im Rahmen des KNN-Trainings ist darüber hinaus die gewählte Topologie zu berücksichtigen, was zu Beginn des Abschnitts ausführlich diskutiert wurde.

L2-Regularisierung

In der KNN-Anwendung wird häufig die L2-Regularisierung

$$\mathcal{L}_C^{L2}(\mathbf{w}) = \frac{\epsilon}{2} \|\mathbf{w}\|^2 \quad (4.151)$$

verwendet, welche nach dieser Interpretation ein Penalty-Term ist. Sie bestraft betragsmäßig große Gewichte und führt somit letztlich zu glatteren Funktionen \mathbf{f}^{KNN} . Für diese Nebenbedingung wäre die Wahl $\epsilon \rightarrow \infty$ nicht zielführend, da man $\|\mathbf{w}\| = 0$ nicht exakt einfordern möchte. Im Rahmen einer solchen *Regularisierung* ist der Penalty-Parameter demzufolge nur ein Wichtungsfaktor, der problemabhängig vom Anwender gewählt werden muss. Im Gegensatz dazu wäre bei physikalisch motivierten Nebenbedingungen der Grenzwert $\epsilon \rightarrow \infty$ grundsätzlich erstrebenswert. Allerdings führt dies bei KNN mit zu wenigen Gewichten zu einer Vernachlässigung des Trainingsdatenfehlers, sodass auch hier stets ein Kompromiss gefunden werden muss.

4.5.2 Die exakte L1-Penalty-Methode

Es existieren Penalty-Funktionen, die für finite $\epsilon < \infty$ eine Äquivalenz der Optimierungsprobleme (4.146) und (4.149) gewährleisten und damit theoretisch eine exakte Erfüllung der Nebenbedingungen. Eine dieser Funktionen ist die L1-Penalty-Funktion

$$\mathcal{L}_C^{EP}(\mathbf{w}) := \epsilon \left[|h(\mathbf{w})| + \max(0, g(\mathbf{w})) \right]. \quad (4.152)$$

Die Funktion (4.152) ist im Allgemeinen jedoch nicht mehr überall differenzierbar⁵², was im Rahmen des Trainingsverfahrens zu Problemen führen kann. Die gestaffelte Liniensuche im QN-Verfahren aus Abschnitt 4.4.2 löst dieses Problem innerhalb der vorliegenden Arbeit zufriedenstellend. Das Adam-Verfahren aus Abschnitt 4.4.1 hat damit grundsätzlich keine Probleme gezeigt.

4.5.3 Die Augmented-Lagrange-Methode

Die Augmented-LAGRANGE-Methode (ALM) zählt ebenfalls zu den Penalty-Methoden. Sie erfüllt die Nebenbedingungen mathematisch exakt für $\epsilon < \infty$ und ist trotzdem lokal differenzierbar. Die zugehörige Funktion ist

$$\mathcal{L}_{Cm}^{AL}(\mathbf{w}) = \mu_m h(\mathbf{w}) + \frac{\epsilon}{2} h(\mathbf{w})^2 + \frac{1}{2\epsilon} \left[\max(0, \lambda_m + \epsilon g(\mathbf{w}))^2 - \lambda_m^2 \right]. \quad (4.153)$$

Sie basiert auf der Idee, die LAGRANGE-Funktion (4.147) nicht in Bezug auf die Originalfunktion $\mathcal{L}_T(\mathbf{w})$ zu bilden, sondern bzgl. der zugehörigen Penalty-Funktion (4.150). Die LAGRANGE-Multiplikatoren μ_m und λ_m sind dabei keine unabhängigen Variablen. Sie werden über die HESTENS-POWELL-Vorschriften

$$\begin{aligned} \mu_{m+1} &= \mu_m + \epsilon h(\mathbf{w}_m) \\ \lambda_{m+1} &= \max(0, \lambda_m + \epsilon g(\mathbf{w}_m)) \end{aligned} \quad (4.154)$$

⁵² Die gleiche Problematik tritt auch bei der Verwendung des MAE als Datenfehlerterm auf.

mithilfe des Penalty-Parameters ϵ aktualisiert. Dieser fungiert somit nicht nur als Wichtungsfaktor, sondern auch als Änderungsrate der LAGRANGE-Multiplikatoren. Letztere werden jeweils zu null initialisiert. In der Regel wird das Verfahren in Form zweier verschachtelter Schleifen durchgeführt: Die innere Iterationsschleife optimiert die Funktion (4.153) mit konstanten μ_m und λ_m . In der äußeren Schleife werden die LAGRANGE-Multiplikatoren jeweils mit den gefundenen Minima \mathbf{w}_m aktualisiert. Man kann zeigen, dass die Werte $(\mathbf{w}_m, \boldsymbol{\lambda}_m, \boldsymbol{\mu}_m)$ gegen einen KKT-Punkt konvergieren.

Für das KNN-Training sind die verschachtelten Schleifen nicht sinnvoll, weil die exakte Lösung innerhalb eines (inneren) Trainingsverfahrens ohnehin nicht gefunden werden kann. Daher wird die ALM hier in einer einzigen Iterationsschleife durchgeführt. Die LAGRANGE-Multiplikatoren werden dabei alle n_w Epochen aktualisiert. Für das QN-Verfahren aus Abschnitt 4.4.2 ist es dabei wichtig, dass die Aktualisierungen (4.154) während eines Neustarts ($\mathbf{B}_n = \mathbf{I}$) durchgeführt werden, da sonst die BFGS-Formel (4.141) nicht mehr anwendbar ist. Dies ist bei allen Verfahren zweiter Ordnung zu beachten.

4.5.4 Nichtlineare Zwangsbedingungen über ZB-Daten

Nebenbedingungen sind im Allgemeinen nicht direkt in den Gewichten \mathbf{w} definiert, wie es bspw. bei der L2-Regularisierung (4.151) der Fall ist. Besonders physikalische Nebenbedingungen im Rahmen der Materialmodellierung sind abhängig von Spannungen, Dehnungen oder Tangentenkomponenten, die mit den Eingangs- und Ausgangsvariablen \mathbf{x} und \mathbf{z} sowie deren Ableitungskomponenten $\partial z_j / \partial x_i$ in Zusammenhang stehen können. So lässt sich bspw. die Gleichheits-Nebenbedingung $h(\mathbf{z}(\mathbf{x}, \mathbf{w}))$ nicht immer in eine rein von den Gewichten abhängige Repräsentation umformen, da die KNN-Abbildungsvorschrift $\mathbf{z} = \mathbf{f}^{\text{KNN}}(\mathbf{x}, \mathbf{w})$ auch von \mathbf{x} abhängig ist. Wollte man diese Nebenbedingungen mit der klassischen Penalty-Methode einfordern, müsste man dies über das Integral

$$\mathcal{L}_C^{PM} = \frac{\epsilon}{2} \int_{\mathbb{R}^{n_i}} h(\mathbf{z}(\mathbf{x}, \mathbf{w}))^2 \mathrm{d}\mathbf{x} \quad (4.155)$$

tun. Die Berechnung dieses Integrals im gesamten Vektorraum \mathbb{R}^{n_i} ist allerdings zu aufwendig und überdies auch nicht notwendig. Das KNN soll nur in einem, von den Punkten $\{\mathbf{x}_k\}_{k=1}^{P_T}$ der Trainingsdatenpaare definierten *Trainingsraum* angewendet werden. Die numerisch sinnvolle Berücksichtigung von Nebenbedingungen erfordert daher die präzise Quantifizierung dieses Trainingsraums $\Omega \subset \mathbb{R}^{n_i}$, um den Fehler (4.155) über

$$\mathcal{L}_C^{PM} = \frac{\epsilon}{2} \int_{\Omega} h(\mathbf{z}(\mathbf{x}, \mathbf{w}))^2 \mathrm{d}\mathbf{x} \quad (4.156)$$

definieren zu können. Im einfachsten Fall wäre der Trainingsraum bspw. das kleinstmögliche, alle Punkte $\{\mathbf{x}_k\}_{k=1}^P$ umschließende Hyperrechteck. Die Berechnung des Integrals (4.156) kann nur näherungsweise erfolgen. Es wird mit zusätzlichen NB-Datenpunkten

$$C := \{\mathbf{x}_k\}, \quad k = 1, \dots, P_C \quad (4.157)$$

als Monte-Carlo-Integration⁵³ durchgeführt, über

$$\mathcal{L}_C^{PM} \approx \frac{\epsilon}{2P_C} \sum_{k=1}^{P_C} h(\mathbf{z}(\mathbf{x}_k^C, \mathbf{w}))^2. \quad (4.158)$$

An dieser Stelle sei ausdrücklich hervorgehoben, dass die zusätzlichen Datenpunkte (4.157) keine Information über zugehörige Zielwerte \mathbf{t} benötigen. Sie können folglich in beliebiger Stückzahl innerhalb Ω generiert werden.

Für die Ungleichheits-Zwangsbedingung $g \leq 0$ wird die *aktive Menge*

$$\mathbb{I}_a := \{k \mid g(\mathbf{x}_k^C) > 0\} \quad (4.159)$$

eingeführt, welche die Indizes der Punkte enthält, auf denen sie nicht erfüllt ist. Damit lässt sich die Maximumsfunktion in den Gleichungen (4.150) und (4.152) ersetzen. Letztlich folgen jeweils für eine Gleichheits-Nebenbedingung $h(\mathbf{w}) = 0$ und eine Ungleichheits-Nebenbedingung $g(\mathbf{w}) \leq 0$ die approximierten Fehlerfunktionen der drei gegebenen Penalty-Methoden zu

$$\mathcal{L}_C^{PM}(\mathbf{w}) \approx \frac{\epsilon}{2P_C} \left[\sum_{k=1}^{P_C} h(\mathbf{x}_k^C, \mathbf{w})^2 + \sum_{k \in \mathbb{I}_a} g(\mathbf{x}_k^C, \mathbf{w})^2 \right] \quad (4.160)$$

$$\mathcal{L}_C^{EP}(\mathbf{w}) \approx \frac{\epsilon}{P_C} \left[\sum_{k=1}^{P_C} |h(\mathbf{x}_k^C, \mathbf{w})| + \sum_{k \in \mathbb{I}_a} g(\mathbf{x}_k^C, \mathbf{w}) \right] \quad (4.161)$$

$$\begin{aligned} \mathcal{L}_{Cm}^{AL}(\mathbf{w}) \approx & \frac{1}{P_C} \sum_{k=1}^{P_C} \left(\mu_{km} h(\mathbf{x}_k^C, \mathbf{w}) + \frac{\epsilon}{2} h(\mathbf{x}_k^C, \mathbf{w})^2 \right) \\ & + \frac{1}{2\epsilon P_C} \sum_{k=1}^{P_C} \left(\max \left(0, \lambda_{km} + \epsilon g(\mathbf{x}_k^C, \mathbf{w}) \right)^2 - \lambda_{km}^2 \right). \end{aligned} \quad (4.162)$$

Bei der ALM sind jedem Nebenbedingungs-Punkt \mathbf{x}_k^C zugehörige LAGRANGE-Multiplikatoren λ_k und μ_k zuzuordnen, die, wie in Abschnitt 4.5.3 beschrieben, initialisiert und in den m -Schritten aktualisiert werden müssen. Zur Darstellung

⁵³ Normalerweise muss bei der Monte-Carlo-Integration das Volumen $|\Omega|$ noch als Faktor hinzugefügt werden. Hier wurde das Volumen mit dem Penalty-Parameter $\bar{\epsilon}$ vereinigt zu $\epsilon = |\Omega|\bar{\epsilon}$, da zur Minimierung dieses Terms nicht der absolute Wert relevant ist.

der zugehörigen Gradienten wird mit z. B. $h_k = h(\mathbf{x}_k^C, \mathbf{w})$ die Abhängigkeiten von \mathbf{x}_k^C und \mathbf{w} zur besseren Lesbarkeit nicht mehr angetragen. Sie lauten

$$\nabla \mathcal{L}_C^{PM} \approx \frac{\epsilon}{P_C} \left[\sum_{k=1}^{P_C} h_k \quad \nabla h_k + \sum_{k \in \mathbb{I}_a} g_k \quad \nabla g_k \right] \quad (4.163)$$

$$\nabla \mathcal{L}_C^{EP} \approx \frac{\epsilon}{P_C} \left[\sum_{k=1}^{P_C} \frac{h_k}{|h_k|} \quad \nabla h_k + \sum_{k \in \mathbb{I}_a} \frac{g_k}{|g_k|} \quad \nabla g_k \right] \quad (4.164)$$

$$\nabla \mathcal{L}_{Cm}^{AL} \approx \frac{1}{P_C} \left[\sum_{k=1}^{P_C} \underbrace{(\mu_{km} + \epsilon h_k)}_{=: \rho_{h_k}} \quad \nabla h_k + \sum_{k=1}^{P_C} \underbrace{\max(0, \lambda_{km} + \epsilon g_k)}_{=: \rho_{g_k}} \quad \nabla g_k \right]. \quad (4.165)$$

Die Vorfaktoren ρ_{h_k} und ρ_{g_k} der jeweiligen NB-Gradienten ∇h_k und ∇g_k hängen somit von der verwendeten Penalty-Methode ab. Sie gehen – je nach Nebenbedingung – in die Vorfaktoren d_{jk} des Gradienten bzgl. KNN-Ausgängen aus Abschnitt 4.2.4 ein oder in die Vorfaktoren f_{ijk} des Gradienten bzgl. KNN-Ableitungen aus Abschnitt 4.2.5.

4.6 Die 2D-Sinc-Funktion als Prinzipbeispiel

Ziel dieses Abschnitts ist es, die KNN-Approximation im Allgemeinen und die Handhabung der Nebenbedingungen im Speziellen anhand eines einfachen Beispiels zu zeigen. Daher wird im Folgenden die Funktion des zweidimensionalen Sinus cardinalis (Sinc)

$$f(\mathbf{x}) = f(x, y) = \frac{\sin(\sqrt{x^2 + y^2})}{\sqrt{x^2 + y^2}} \quad (4.166)$$

im Bereich $x, y \in [-10, 10]$ mit dem FNN aus Abschnitt 4.2 approximiert. Die Funktion ist rotationssymmetrisch; der Gradient $df/d\mathbf{x}$ zeigt stets zum Ursprung. Dies kann auch durch die Symmetrieeigenschaft

$$\frac{df}{d\mathbf{x}} \cdot \begin{bmatrix} -y \\ x \end{bmatrix} = x f_{,y} - y f_{,x} = 0 \quad (4.167)$$

ausgedrückt werden. Diese Eigenschaft sowie der Graph der Sinc-Funktion (4.166) sind in Abbildung 4.11 dargestellt.

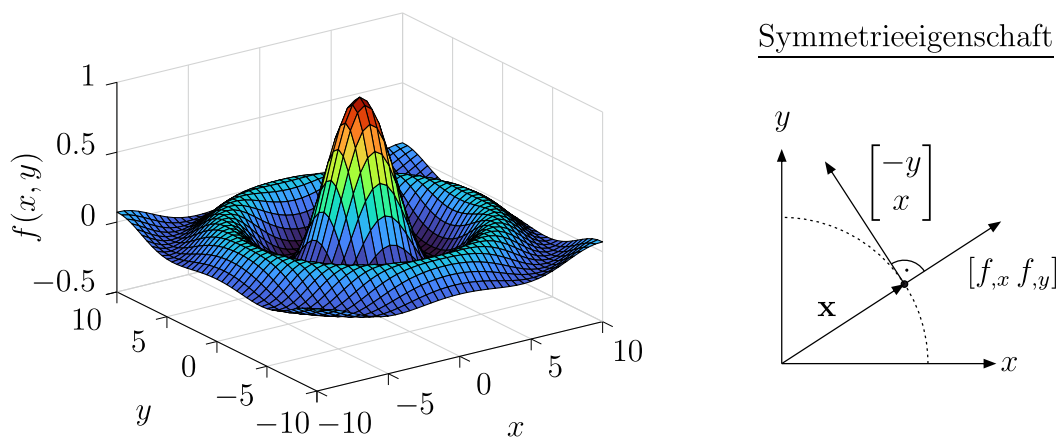


Abbildung 4.11: Zweidimensionale Sinc-Funktion in den Intervallen $x, y \in [-10, 10]$ (links) und grafische Darstellung der Rotationssymmetrie-Bedingung (rechts)

Der Definitionsbereich

$$\Omega = [-10, 10]^2 = \{(x, y) \in \mathbb{R}^2 \mid -10 \leq x, y \leq 10\} \quad (4.168)$$

ist gleichzeitig als Trainingsbereich des KNN zu verstehen. Innerhalb dieses Bereichs soll das KNN die gegebene Funktion approximieren. Im Folgenden wird stets das QN-Trainingsverfahren nach Abschnitt 4.4.2 genutzt.

4.6.1 Generierung der Trainings-, Validierungs- und NB-Punkte

Die Eingangsvariablen \mathbf{x}_k der Trainings- und Validierungsdaten sowie – falls notwendig – die NB-Punkte \mathbf{x}_k^C werden jeweils unabhängig voneinander im Trainingsraum (4.168) generiert. Dazu wird das *Latin-Hypercube-Stichprobenverfahren* (LHS, vgl. MCKAY ET AL. [129]) genutzt, womit im Vergleich zu unabhängig voneinander erzeugten gleichverteilten Zufallszahlen eine gleichmäßigere Verteilung der Punkte möglich ist, gerade für $n_i \gg 1$. Zur Generierung von P Punkten mittels LHS wird der Trainingsraum Ω in $P^{n_i} = P^2$ gleich große, rechteckige Unterräume unterteilt. Die Punkte $\{\mathbf{x}_k\}_{k=1}^P$ werden zufallsverteilt so erzeugt, dass in jeder *Zeile* und in jeder *Spalte* jeweils nur ein Punkt liegt. Lagen alle Punkte auf einer Diagonalen, wäre die Bedingung auch erfüllt, die räumliche Verteilung aber nicht brauchbar. Daher wird innerhalb einer iterativen Datengenerierung der minimale Abstand zwischen zwei Punkten maximiert. Dieses Prinzip ist ohne Einschränkung auf $n_i > 2$ Dimensionen erweiterbar. Es wird auch zur Generierung zufälliger Datenpunkte in den nachfolgenden Kapiteln verwendet. In Abbildung 4.12 ist eine LHS-Realisation für $P = 100$ Datenpunkte dargestellt. Die Sinc-Funktion ist relativ sensibel bzgl. der Verfügbarkeit von Information am

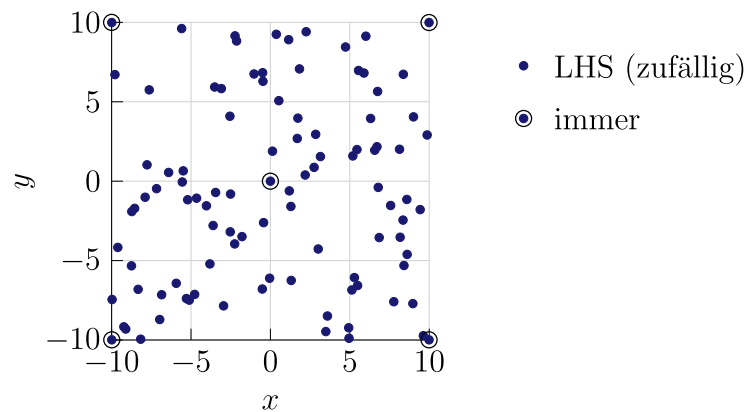


Abbildung 4.12: Realisation einer LHS-Datengenerierung für 100 Punkte $\mathbf{x} = [x, y]$. Die Ecken des Trainingsraums Ω sowie die Mitte $\mathbf{x} = \mathbf{0}$ sind stets in der Menge der Trainingsdaten enthalten.

Ursprung $\mathbf{x} = \mathbf{0}$. Dieser signifikante Punkt sowie die vier Eckpunkte des Trainingsraums Ω werden deshalb stets gegen fünf zufällig ausgewählte LHS-Punkte ausgetauscht und sind somit Teil der Trainingsdatenpunkte.

4.6.2 Parameterstudien und Untersuchungen zum Training

Das KNN-Training ist ein stochastischer Prozess. Durch die Gewichtsinitialisierung (4.131) mit Pseudo-Zufallszahlen, die unter Umständen zufällige Auswahl der Trainings-Batches sowie ggf. das Erzeugen von Datenpunkten nach Abschnitt 4.6.1 ist der Ausgang eines einzelnen KNN-Trainings allein nicht aussagekräftig. Daher sind im Rahmen von Parameterstudien immer mehrere Trainingsdurchläufe bei gleichbleibenden Trainingseinstellungen durchzuführen. Die jeweils niedrigsten Validierungsfehler $\mathcal{L}_{V(i)}^{min}$ (vgl. Definition (4.143)) werden über n_{MCS} Trainingsdurchläufe $i = 1, \dots, n_{MCS}$ gemittelt zu

$$\bar{\mathcal{L}}_V^{min} = \frac{1}{n_{MCS}} \sum_{i=1}^{n_{MCS}} \mathcal{L}_{V(i)}^{min}, \quad (4.169)$$

um Trends zu identifizieren und Effekte einzelner Maßnahmen bewerten zu können. Auf die Darstellung der zugehörigen Streuung, bspw. in Form der Standardabweichung, wird hier verzichtet. Zur Festlegung einer Parameterstudie sind folglich fünf Parameter zu wählen: Die Anzahl der Trainingsdurchläufe mit konstanten Einstellungen n_{MCS} , die Anzahl der zufällig erzeugten Validierungspunkte P_V sowie die drei Parameter zum Trainingsabbruch δ^{abs} , δ^{rel} und epo_{max} . Diese Einstellungen sind für alle Untersuchungen dieses Abschnitts in Tabelle 4.3 gegeben.

Trainingsmittel n_{MCS}	Epochenmittel P_V	Abs. Krit. δ^{abs}	Rel. Krit. δ^{rel}	Ende Training epo_{max}
20	10^3	10^{-3}	10^{-3}	10^4

Tabelle 4.3: Statistische Parameter für die Untersuchungen an der Sinc-Funktion

Zur Verdeutlichung dieser intrinsischen Variabilität sind in Abbildung 4.13 die Verläufe des einheitenbereinigten Trainingsfehlers $\sqrt{2\mathcal{L}_T}$ und des Validierungsfehlers \mathcal{L}_V für 20 Trainingsdurchläufe dargestellt. Dabei wurde eine Topologie [2-10-10-1] gewählt mit $P_T = 200$ Trainingsdatenpunkten. Während durch die Abstiegsbedingung des QN-Verfahrens der Trainingsdatenfehler mit jeder Epoche epo kleiner wird, stagniert der Validierungsfehler teilweise oder steigt sogar. Die jeweiligen minimalen Validierungsfehler \mathcal{L}_V^{min} können sich durchaus um eine dezimale Größenordnung unterscheiden. Einer der Trainingsdurchläufe wurde vor Erreichen der maximalen Anzahl an Epochen abgebrochen, wodurch das relative Abbruchkriterium noch einmal dargestellt ist. Das absolute Abbruchkriterium, mit hier $\mathcal{L}_V^{abs} \approx 10^{-4}$, wurde in keinem der Trainingsdurchläufe erreicht.

Eine weitere wichtige Untersuchung klärt die Frage nach der ausreichenden Anzahl an Freiheitsgraden und Datenpunkten. In Abbildung 4.14 ist links eine Studie

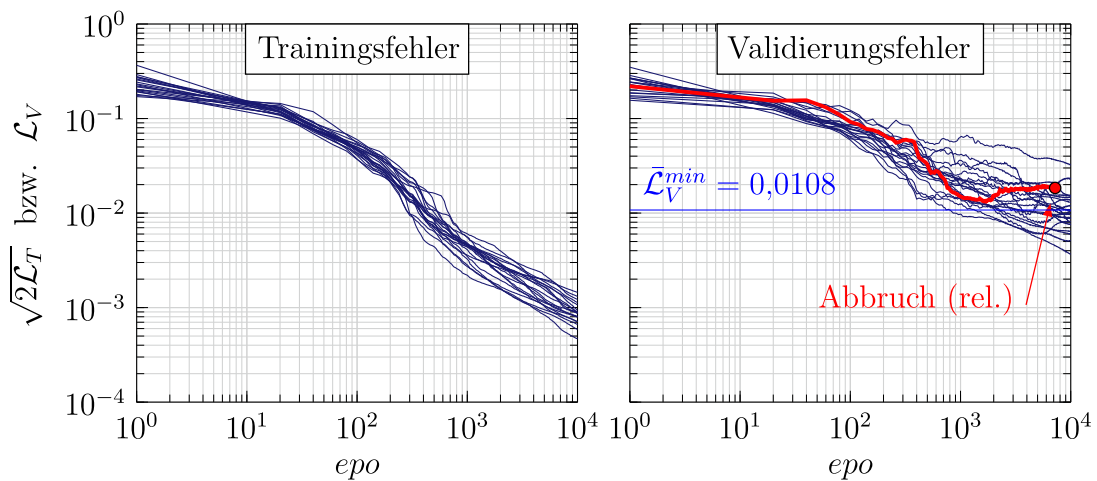


Abbildung 4.13: Darstellung der intrinsischen Variabilität des KNN-Trainings anhand von 20 Durchläufen mit Topologie [2-10-10-1] und $P_T = 200$ Trainingspaaren

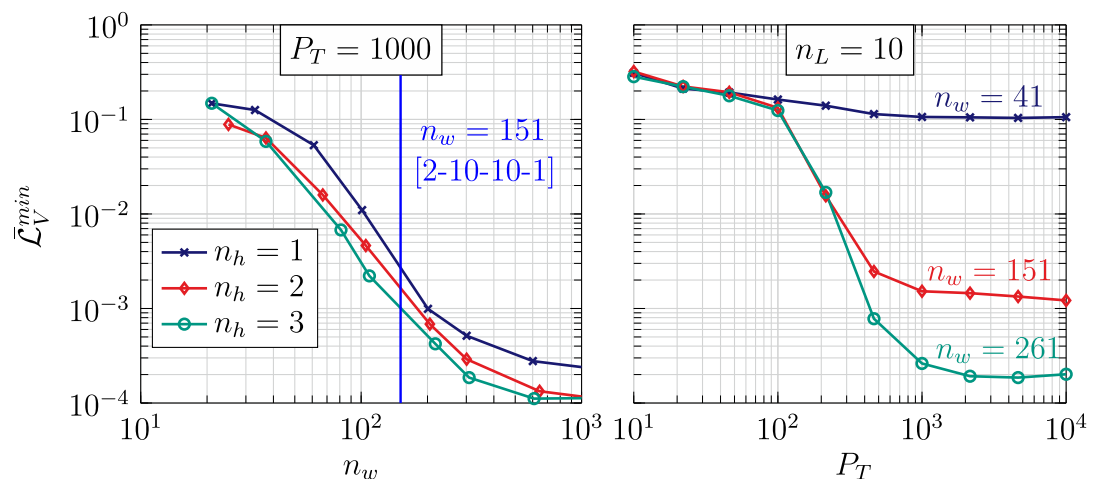


Abbildung 4.14: Parameterstudien bzgl. Topologie (links) und Anzahl der Trainingsdatenpaare P_T (rechts). Jeder Marker steht für eine Untersuchung.

bzgl. der Netztopologie (n_h : Anzahl der Zwischenschichten, n_w : Anzahl Gewichte) und rechts eine Studie bzgl. Trainingsdaten abgebildet. Dabei sind die wechselseitigen Einflüsse von vorhandenen Informationen und Freiheitsgraden gut zu erkennen. Eine aussagekräftige Topologiestudie sollte in Abhängigkeit der Gewichtsanzahl n_w gemacht werden, wodurch sich infolge der ganzzahligen Menge an Neuronen verschiedene Auswertungsstellen ergeben. Das augenscheinliche Konvergenzverhalten hin zu $\bar{\mathcal{L}}_V^{min} \approx 10^{-4}$ liegt zum einen an der festen Anzahl an Trainingsdatenpaaren mit $P_T = 1\,000$, zum andern am absoluten Abbruchkriterium. Mit beliebig vielen Trainingsdaten und Freiheitsgraden ließe sich der Fehler theoretisch bis zur Maschinengenauigkeit heruntertreiben. An der Studie zur An-

zahl der Trainingsdatenpaare lässt sich erkennen, dass bei einer festen Topologie ab einer gewissen Informationsschwelle keine Verbesserung mehr erzielt werden kann, auch wenn die Menge an Trainingsdaten weiter steigt. Dieses Konvergenzverhalten tritt umso später ein, je mehr Freiheitsgrade zur Verfügung stehen. Für die folgenden Untersuchungen wird meist die Topologie [2-10-10-1] gewählt, die in Abbildung 4.14 gekennzeichnet ist. Außerdem ist zu erkennen, dass für eine auskonvergierte Approximation ca. $P_T = 1000$ Trainingspaare notwendig sind.

4.6.3 Training mit der Nebenbedingung Rotationssymmetrie

Die Symmetrieeigenschaft (4.167) wird in Form einer Gleichheits-NB im KNN-Training berücksichtigt. Dazu wird sie in der KNN-Ausgangsgröße $z(\mathbf{x}, \mathbf{w})$ und deren Ableitungen formuliert. Für einen einzelnen NB-Punkt \mathbf{x}_k^C lautet sie

$$h_k = h(\mathbf{x}_k^C, \mathbf{w}) = x_k^C \left(\frac{\partial z_k}{\partial y} \right) - y_k^C \left(\frac{\partial z_k}{\partial x} \right) = 0, \quad (4.170)$$

mit der Abkürzung $z_k = z(\mathbf{x}_k^C, \mathbf{w})$. Eingesetzt in die Gleichungen (4.160), (4.161) oder (4.162) lässt sich die zugehörige Fehlerfunktion \mathcal{L}_C für eine der Penalty-Methoden berechnen. Der NB-Gradient $\nabla h_k = \nabla h(\mathbf{x}_k^C, \mathbf{w})$ hat unter Berücksichtigung der Ableitungstransformation (4.34) die Form

$$\nabla h_k = \frac{x_k^C s_z}{s_{x2}} \frac{\partial}{\partial \mathbf{w}} \left(\frac{\partial \hat{z}_k}{\partial \hat{x}_2} \right) - \frac{y_k^C s_z}{s_{x1}} \frac{\partial}{\partial \mathbf{w}} \left(\frac{\partial \hat{z}_k}{\partial \hat{x}_1} \right), \quad (4.171)$$

mit $s_z = s_{z1}$, da nur eine Ausgangsvariable vorhanden ist. In Verbindung mit den von den Penalty-Methoden abhängigen Vorfaktoren ρ_{h_k} in den Gradienten (4.163) bis (4.165) lässt sich der gesamte Fehlergradient $\nabla \mathcal{L}_C$ in der gleichen Form schreiben, wie der allgemeine Fehlergradient (4.68) aus Abschnitt 4.2.5. Dabei lassen sich die Vorfaktoren f_{ijk} aus der Matrix (4.69) über

$$f_{ijk} = \begin{cases} -\frac{y_k^C s_z \rho_{h_k}}{s_{x1}} & i = 1 \\ \frac{x_k^C s_z \rho_{h_k}}{s_{x2}} & i = 2 \end{cases} \quad (4.172)$$

berechnen. Weiter ist $\mathbb{I} = \{1, 2\}$ und $\mathbb{J}(1) = \mathbb{J}(2) = \{1\}$. Die Berechnung des Fehlergradienten $\nabla \mathcal{L}_C$ folgt mit dem Algorithmus aus Abschnitt 4.2.5. Somit sind alle Werkzeuge zur Verfügung gestellt, um mit einer der drei Penalty-Methoden die Nebenbedingung der Rotationssymmetrie im KNN-Training zu berücksichtigen. Im Folgenden werden einzelne Untersuchungen bezogen auf das Validierungsmittel (4.169) gezeigt.

In Abbildung 4.15 ist für die Topologien [2-10-10-1] und [2-10-10-10-1] je eine Studie bzgl. Penalty-Parameter ϵ und Anzahl NB-Punkten P_C dargestellt. Dabei wurde stets die klassische L2-Penalty-Methode nach Abschnitt 4.5.1 benutzt.

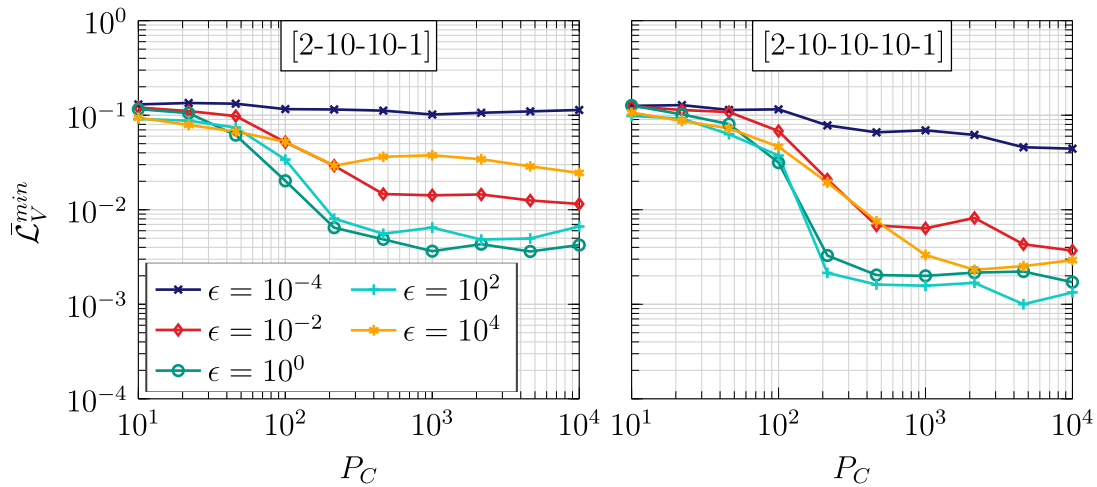


Abbildung 4.15: Parameterstudie zum Penalty-Parameter ϵ und der Anzahl von NB-Punkten P_C , für $n_h = 2$ bzw. $n_h = 3$ Zwischenschichten und $P_T = 100$ Trainingspaaren

Wie auch bzgl. der Anzahl an Trainingsdatenpaaren P_T ist ein Konvergenzverhalten zu erkennen, welches auf die Monte-Carlo-Integration (4.158) zurückzuführen ist. Mit steigendem Penalty-Parameter ϵ wird die Nebenbedingung gegenüber dem Trainingsdatenfehler stärker gewichtet. Dies spiegelt sich bei $\epsilon = 10^4$ durch ein Verschlechtern der Ergebnisse wider. Diese Verschlechterung verschwindet wiederum, wenn die Anzahl der Freiheitsgrade erhöht wird (Abbildung 4.15, rechts). Grundsätzlich ist zu beobachten, dass trotz der unzureichenden Anzahl an Trainingsdaten eine Verbesserung der KNN-Approximationsfähigkeit um mehrere Größenordnungen möglich ist. Diese *zentrale Erkenntnis* motiviert in den folgenden Kapiteln die Berücksichtigung physikalisch motivierter Nebenbedingungen im Rahmen der KNN-Materialmodellierung.

In Abbildung 4.16 sind Studien bzgl. des Penalty-Parameters abgebildet, in welchen die drei eingeführten Penalty-Methoden für die Fälle weniger Daten ($P_T = 100$) und ausreichend vieler Daten ($P_T = 1000$) verglichen werden. Die Topologie ist stets $[2-10-10-1]$ und die Anzahl der NB-Punkte ist mit $P_C = 1000$ ausreichend hoch. Grundsätzlich verhalten sich alle drei Verfahren ähnlich, auch wenn das L1-Penalty-Verfahren im Bereich hoher Penalty-Parameter wesentlich schlechtere Ergebnisse liefert. Im Fall weniger Trainingsdaten gibt es einen großen Bereich, in dem durch die Nebenbedingungen eine Verbesserung erzielt werden kann. Der Anstieg für $\epsilon = 10^4$ ist schon in Abbildung 4.15 beobachtet worden. Er resultiert aus der zu geringen Anzahl an Freiheitsgraden. Diese Verschlechterung könnte durch eine größere Topologie entschärft werden.

Trainingsdaten im Rahmen der Materialmodellierung können nicht nur in ihrer Anzahl gering, sondern auch nicht gleichmäßig im Trainingsraum Ω verteilt sein. Im Rahmen des Modellproblems wurden für die Ergebnisse in Abbildung 4.17 nur

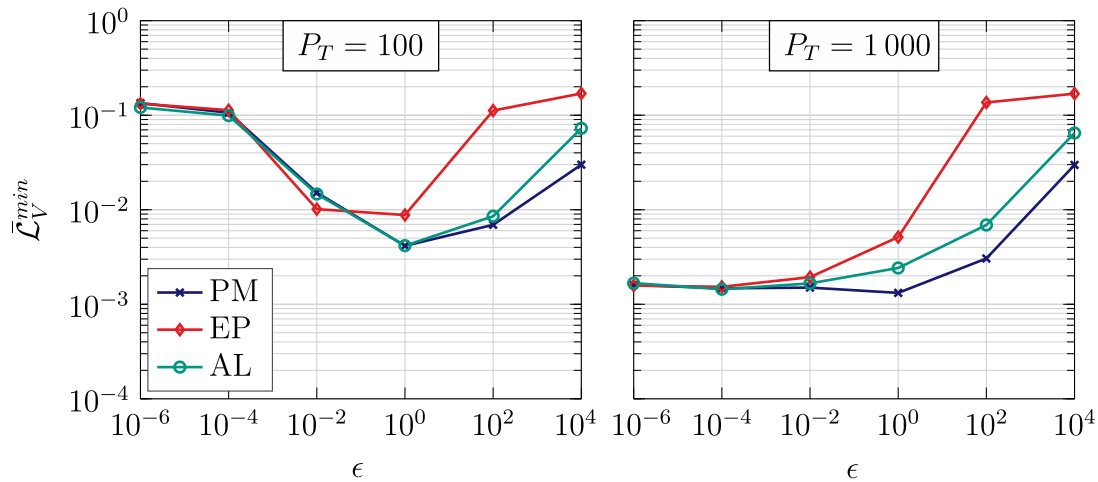


Abbildung 4.16: Parameterstudie zum Penalty-Parameter ϵ und Vergleich der L2-Penalty-Methode (PM), der L1-Penalty-Methode (EP) und der Augmented-LAGRANGE-Methode (AL) für den Fall weniger Trainingsdaten (links) und ausreichender Trainingsdaten (rechts)

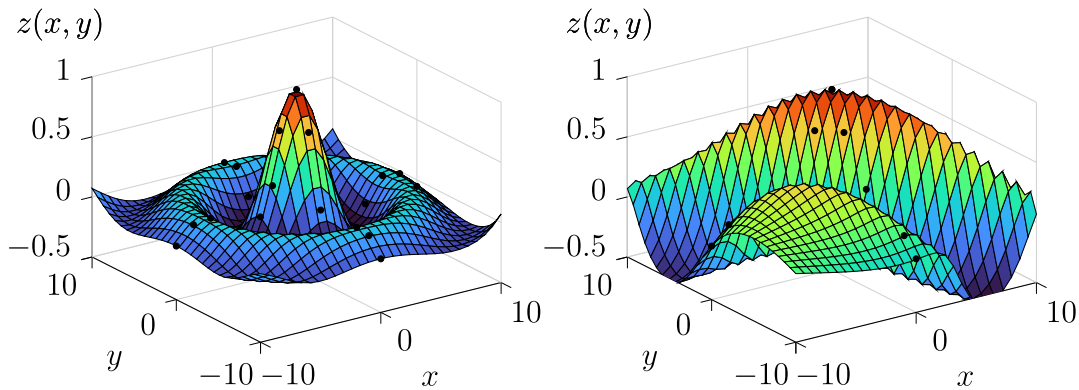


Abbildung 4.17: Graphen der KNN-Funktionen mit insgesamt $P_T = 29$ Trainingsdatenpaaren, die nur auf den Achsen $x = 0$ und $y = 0$ generiert sind. Links mit NB ($\epsilon = 100$, $P_C = 1000$), rechts ohne.

$P_T = 29$ Trainingsdatenpaare generiert, und das nur auf den Koordinatenachsen $x = 0$ und $y = 0$. Die Topologie ist jeweils $[2-10-10-10-1]$. Anhand der beiden KNN-Antworten ist der Effekt der Nebenbedingung (mit $\epsilon = 100$ und $P_C = 1000$) eindrucksvoll zu erkennen. Trotz unzureichender Informationsgrundlage ist das KNN in der Lage, den Verlauf der Sinc-Funktion sinnvoll wiederzugeben. Der zugehörige Validierungsfehler ist $\bar{\mathcal{L}}_V^{min} \approx 0,01$. Ohne Nebenbedingungen bricht das Training nach kurzer Zeit ab, bei $\bar{\mathcal{L}}_V^{min} \approx 0,5$.

4.6.4 Schlussfolgerungen über das Training mit Nebenbedingungen

Zum Abschluss dieses Kapitels werden hier die durch das Prinzipbeispiel deutlich gemachten Erkenntnisse über das KNN-Training mit Nebenbedingungen (NB) stichwortartig zusammengefasst. Die Erfahrung gelten ebenfalls bei der KNN-Materialmodellierung mit Nebenbedingung.

- **Verbesserung der Approximationsgüte:** Die Einforderung von Nebenbedingungen kann die Approximationsgüte des KNN um viele dezimale Größenordnungen verbessern, vor allem dann, wenn vergleichsweise wenige Trainingsdaten vorhanden sind.
- **Konvergenz der NB-Punkte:** Ab einer gewissen Anzahl an NB-Punkten ist die Monte-Carlo-Integration (4.158) auskonvergiert, sodass keine Verbesserung mehr erzielt werden kann. Eine große Anzahl von NB-Punkten verschlechtert das generelle Approximationsverhalten nicht grundsätzlich.
- **Penalty-Methode ausreichend:** Die klassische Penalty-Methode aus Abschnitt 4.5.1 erzielt im Mittel die besten Ergebnisse und ist dabei numerisch stabil. Daher werden die anderen beiden Penalty-Methoden im Folgenden nicht weiter berücksichtigt.
- **Mehr Freiheitsgrade für stärkere NB:** Wenn die NB mit einem hohen Penalty-Parameter eingefordert werden soll, dann muss die Anzahl der Gewichte erhöht werden. Dies ist insbesondere dann wichtig, wenn nur sehr wenige und nicht gleichmäßig verteilte Trainingsdaten zur Verfügung stehen.

Diese grundsätzlichen Erkenntnisse werden in den folgenden Kapiteln um spezifischere Untersuchungen rund um die KNN-Materialmodellierung mit physikalischen Nebenbedingungen erweitert. Dazu werden bspw. neben dem Validierungsfehler andere Indikatoren der Approximationsgüte angewendet, wie bspw. das Verhalten der KNN-Materialmodelle am Materialpunkt oder innerhalb der FEM.

Darüber hinaus hätte man das KNN zur Approximation der Sinc-Funktion auch in der einzelnen Eingangsvariable $r = \sqrt{x^2 + y^2}$ definieren können. Dadurch wäre die Rotationssymmetrie per Konstruktion des KNN erfüllt, und es würden weniger Trainingsdaten sowie Freiheitsgrade benötigt. Dies ist in der KNN-Materialmodellierung allerdings nicht immer möglich und auch nicht immer zweckmäßig. Auf diese beiden Möglichkeiten der Einforderung von physikalischen Eigenschaften, nämlich *schwach* oder *stark*, wird in den nachfolgenden Kapiteln eingegangen.

5 Modellierung von elastischem Materialverhalten mit KNN

Aufbauend auf den grundsätzlichen Eigenschaften elastischer Materialien aus Kapitel 3 und den KNN-Architekturen aus Kapitel 4 wird hier die KNN-Materialmodellierung für elastisches Materialverhalten beschrieben und untersucht. Dieses zeichnet sich durch den in Gleichung (3.1) formulierten direkten funktionalen Zusammenhang zwischen den Spannungen \mathbf{S} und den Verzerrungen \mathbf{E} aus. Jener Zusammenhang soll nachfolgend durch ein KNN definiert werden:

$$\mathbf{S} = \mathcal{S}^{\text{KNN}}(\mathbf{E}) , \quad \mathbf{C}_T = \frac{d\mathcal{S}^{\text{KNN}}(\mathbf{E})}{d\mathbf{E}} . \quad (5.1)$$

In Abschnitt 5.1 wird zunächst die Generierung von Trainings-, Validierungs- und NB-Punkten in den kartesischen Verzerrungskomponenten beschrieben. Das schließt einen Algorithmus ein, mit dessen Hilfe numerische Materialpunkt-Versuche durchgeführt werden können. Verschiedene KNN-Materialmodelle für elastisches Materialverhalten folgen daraufhin in Abschnitt 5.2. Physikalisch motivierte Nebenbedingungen für elastisches Materialverhalten werden in Abschnitt 5.3 vorgestellt. In den Abschnitten 5.4 und 5.5 werden die beschriebenen Modelle am Materialpunkt untersucht und als Materialmodelle innerhalb der FEM angewendet. Die Erkenntnisse aus diesen Untersuchungen bereiten den Weg zur Anwendung an echten Versuchsdaten in Kapitel 7. Innerhalb dieses Kapitels werden ausschließlich dreidimensionale Materialmodelle betrachtet. Eine Reduktion auf bspw. den ebenen Spannungszustand ist ohne Weiteres möglich.

Ziel dieses Kapitels ist die Beschreibung grundlegender Prinzipien der KNN-Materialmodellierung. Dazu werden einige verschiedene KNN-Materialmodelle vorgestellt und miteinander verglichen. Dies schließt ein neuartiges Modell für isotrope Materialien mit ein. Da einige dieser Modelle grundlegende physikalische Eigenschaften schon a priori erfüllen, liegt der Fokus des Kapitels weniger auf dem Vorteil der Nebenbedingungen im Allgemeinen, sondern auf dem Vergleich verschiedener Herangehensweisen zur Berücksichtigung physikalischer Eigenschaften. Außerdem werden einige Herausforderungen im Zusammenhang mit der Modellierung großer Verzerrungen isoliert beschrieben und Lösungsvorschläge bereitgestellt.

5.1 Datengenerierung mit Datenpunkt-Materialversuchen

Für die Parameterstudien auf Basis analytischer Materialmodelle $\mathbf{S} = \mathcal{S}(\mathbf{E})$ müssen Trainings- bzw. Validierungsdatenpaare $\{(\mathbf{x}, \mathbf{t})\}_{k=1}^P$ generiert werden, mit $P = P_T$ oder $P = P_V$. Die Eingangsvektoren \mathbf{x} stehen mit den Verzerrungen

\mathbf{E} in Zusammenhang, die Zielwerte \mathbf{t} mit den Spannungen \mathbf{S} . Für etwaige Nebenbedingungen werden nur die Eingangsvariablen in Form der NB-Punkte $\{\mathbf{x}_k^C\}_{k=1}^{PC}$ benötigt, ohne Berechnung zugehöriger Zielwerte. Dabei wird im Folgenden nur die Generierung kartesischer Verzerrungs- und Spannungskomponenten E_{ij} und S_{ij} beschrieben, aus denen ggf. im Anschluss Folgegrößen, bspw. Invarianten, und dementsprechend die Eingangsvektoren \mathbf{x}_k und die Zielwerte \mathbf{t}_k berechnet werden können. Alle generierten Verzerrungen \mathbf{E} müssen konsistent mit $\det \mathbf{C} = J^2 > 0$ sein, was ggf. zu prüfen ist.

Da die aktuellen Spannungen \mathbf{S}_k elastischer Materialmodelle nur von den aktuellen Verzerrungen \mathbf{E}_k abhängen, können letztere unabhängig voneinander generiert werden. Dies kann bspw. durch das Latin-Hypercube-Stichprobenverfahren (LHS) geschehen, welches in Abschnitt 4.6.1 skizziert wurde. Falls nötig, werden die zugehörigen Spannungsvektoren im Anschluss mithilfe des analytischen Materialmodells $\mathbf{S}_k = \mathcal{S}(\mathbf{E}_k)$ berechnet.

5.1.1 Ein gemischt gesteuerter Materialpunkt-Versuch

Eine aus Sicht des KNN-Trainings optimale Generierung von Trainingsdaten mithilfe des LHS spiegelt jedoch nicht die realen Gegebenheiten wider. In standardisierten Materialversuchen werden i. d. R. spezifische Spannungs- und Dehnungspfade abgetastet. Die zugehörigen Verzerrungen $\{\mathbf{E}_k\}_{k=1}^P$ sind folglich nicht gleichmäßig in einem Hyperrechteck verteilt, wie es durch LHS der Fall wäre. Daher werden nachfolgend analytische Materialpunktversuche beschrieben, mit denen Daten in Bezug auf echte Experimente realitätsgetreuer generiert werden können. Dabei wird von einem beliebigen analytischen Materialmodell

$$\mathbf{S} = \mathcal{S}(\mathbf{E}) \quad , \quad \text{mit} \quad \mathbf{C}_T(\mathbf{E}) = \frac{d\mathbf{S}(\mathbf{E})}{d\mathbf{E}} \quad , \quad (5.2)$$

in reduzierter Notation ausgegangen, vgl. Abschnitt 2.1.6.

Die Verzerrungskomponenten werden in einen bekannten (vorgegebenen) Anteil $\mathbf{E}^b \in \mathbb{R}^{n_b}$ und einen unbekanntem Anteil $\mathbf{E}^u \in \mathbb{R}^{n_u}$ aufgeteilt. Mit $\mathbf{E} = [\mathbf{E}^b, \mathbf{E}^u]^T$ wird das Stoffgesetz (5.2) zur Berechnung der ‚inneren‘ Spannungen \mathbf{S} ausgewertet. Von diesem Stoffgesetz zunächst unabhängige ‚äußere‘ Spannungen $\tilde{\mathbf{S}} = [\tilde{\mathbf{S}}^u, \tilde{\mathbf{S}}^b]^T$ werden in die gleichen Anteile zerlegt. Die zu den bekannten Verzerrungen \mathbf{E}^b zugehörigen Spannungskomponenten $\tilde{\mathbf{S}}^b$ sind frei, während die $\tilde{\mathbf{S}}^u$ -Anteile vorgegeben werden. Zwischen beiden Spannungsgrößen wird über das Residuum

$$\mathbf{R}(\mathbf{E}^u, \tilde{\mathbf{S}}^b) := \begin{bmatrix} \mathbf{S}^b(\mathbf{E}^u) \\ \mathbf{S}^u(\mathbf{E}^u) \end{bmatrix} - \begin{bmatrix} \tilde{\mathbf{S}}^b \\ \tilde{\mathbf{S}}^u \end{bmatrix} = \begin{bmatrix} \mathbf{R}^b(\mathbf{E}^u, \tilde{\mathbf{S}}^b) \\ \mathbf{R}^u(\mathbf{E}^u, \tilde{\mathbf{S}}^b) \end{bmatrix} \stackrel{!}{=} \mathbf{0} \quad (5.3)$$

Äquivalenz eingefordert. Dies geschieht iterativ durch das NEWTON-Verfahren.

Vorgabe	$\{\mathbf{E}_n^b\}_{n=1}^N$ und $\{\tilde{\mathbf{S}}_n^u\}_{n=1}^N$
Startwerte	$\mathbf{E}_0 = [\mathbf{E}_0^b, \mathbf{E}_0^u]$ und $\tilde{\mathbf{S}}_0 = [\tilde{\mathbf{S}}_0^b, \tilde{\mathbf{S}}_0^u]$
Lastschleife	for $n = 1, \dots, N$ do $i = 0, \mathbf{E}^{(0)} = \mathbf{E}_{n-1}, \tilde{\mathbf{S}}^{(0)} = \tilde{\mathbf{S}}_{n-1}$ $\mathbf{E}^{b(0)} = \mathbf{E}_n^b, \tilde{\mathbf{S}}^{u(0)} = \tilde{\mathbf{S}}_n^u$ $\mathbf{S}^{(0)} = \mathcal{S}(\mathbf{E}^{(0)})$ und $\mathbf{C}_T^{(0)}(\mathbf{E}^{(0)})$ (ggf. <u>KNN</u>) $\mathbf{R}^{(0)} = \mathbf{S}^{(0)} - \tilde{\mathbf{S}}^{(0)}$
NEWTON-Schleife	while $\ \mathbf{R}^{(i)}\ > TOL$ do $\Delta \mathbf{E}^{u(i)} = (\mathbf{C}_T^{uu(i)})^{-1}(\mathbf{S}^{u(i)} - \tilde{\mathbf{S}}^{u(i)})$ $\Delta \tilde{\mathbf{S}}^{b(i)} = \mathbf{C}_T^{bu(i)} \Delta \mathbf{E}^{u(i)} + \mathbf{S}^{b(i)} - \tilde{\mathbf{S}}^{b(i)}$ $\mathbf{E}^{u(i+1)} = \mathbf{E}^{u(i)} + \Delta \mathbf{E}^{u(i)}, \mathbf{E}^{(i+1)} = [\mathbf{E}_n^b, \mathbf{E}^{u(i+1)}]$ $\tilde{\mathbf{S}}^{b(i+1)} = \tilde{\mathbf{S}}^{b(i)} + \Delta \tilde{\mathbf{S}}^{b(i)}, \tilde{\mathbf{S}}^{(i+1)} = [\tilde{\mathbf{S}}^{b(i+1)}, \tilde{\mathbf{S}}_n^u]$ $\mathbf{S}^{(i+1)} = \mathcal{S}(\mathbf{E}^{(i+1)})$ und $\mathbf{C}_T^{(i+1)}(\mathbf{E}^{(i+1)})$ (ggf. <u>KNN</u>) $\mathbf{R}^{(i+1)} = \mathbf{S}^{(i+1)} - \tilde{\mathbf{S}}^{(i+1)}$ $i \leftarrow i + 1$ end while $\mathbf{E}_n^u = \mathbf{E}^{u(i)}, \mathbf{E}_n = [\mathbf{E}_n^b, \mathbf{E}_n^u]$ $\tilde{\mathbf{S}}_n^u = \tilde{\mathbf{S}}^{u(i)}, \mathbf{S}_n = \tilde{\mathbf{S}}_n = [\tilde{\mathbf{S}}_n^b, \tilde{\mathbf{S}}_n^u]$
	end for

Tafel 5.1: Ein Algorithmus für einen numerischen Materialpunkt-Versuch

Die zugehörige, allgemein linearisierte Gleichung

$$L[\mathbf{R}] = \mathbf{R} + \Delta \mathbf{R} = \begin{bmatrix} \mathbf{S}^b - \tilde{\mathbf{S}}^b \\ \mathbf{S}^u - \tilde{\mathbf{S}}^u \end{bmatrix} + \begin{bmatrix} \mathbf{C}_T^{bb} & \mathbf{C}_T^{bu} \\ \mathbf{C}_T^{ub} & \mathbf{C}_T^{uu} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{E}^b \\ \Delta \mathbf{E}^u \end{bmatrix} - \begin{bmatrix} \Delta \tilde{\mathbf{S}}^b \\ \Delta \tilde{\mathbf{S}}^u \end{bmatrix} = \mathbf{0} \quad (5.4)$$

kann durch die vorgeschlagene gemischte Steuerung mit $\Delta \mathbf{E}^b = \mathbf{0}$ und $\Delta \tilde{\mathbf{S}}^u = \mathbf{0}$ in zwei Schritten gelöst werden. Zunächst werden die Verzerrungsinkremente

$$\Delta \mathbf{E}^u = -(\mathbf{C}_T^{uu})^{-1} \mathbf{R}^u \quad (5.5)$$

und anschließend die Spannungsinkremente

$$\Delta \tilde{\mathbf{S}}^b = \mathbf{C}_T^{bu} \Delta \mathbf{E}^u + \mathbf{R}^b . \quad (5.6)$$

Mit diesen Gleichungen lässt sich ein numerischer Versuch am Materialpunkt durchführen, dessen Algorithmus in Tafel 5.1 gegeben ist. Ausgehend von den Startwerten $\mathbf{E}_0 = [\mathbf{E}_0^b, \mathbf{E}_0^u]$ und $\mathbf{S}_0 = [\mathbf{S}_0^b, \mathbf{S}_0^u] = \tilde{\mathbf{S}}_0$ werden die bekannten Verzerrungskomponenten $\{\mathbf{E}_n^b\}_{n=1}^N$ und die bekannten Spannungskomponenten $\{\tilde{\mathbf{S}}_n^u\}_{n=1}^N$ für jeden Lastschritt $n = 1, \dots, N$ vorgegeben. Pro Schritt werden die unbekannt GröÙen \mathbf{E}_n^u und $\tilde{\mathbf{S}}_n^b$ mithilfe der Gleichungen (5.5) und (5.6) über das NEWTON-Verfahren berechnet.

Das Resultat eines solchen numerischen Materialpunkt-Versuchs sind kartesische Verzerrungs- und Spannungskomponenten, die als sortierter Pfad

$$\mathcal{P} = \{(\mathbf{E}_0, \mathbf{S}_0), (\mathbf{E}_1, \mathbf{S}_1), \dots, (\mathbf{E}_n, \mathbf{S}_n), \dots, (\mathbf{E}_N, \mathbf{S}_N)\} \quad (5.7)$$

die Grundlage zur Berechnung der KNN-Trainings- und Validierungsdatenpaare $\{(\mathbf{x}_k, \mathbf{t}_k)\}_{k=1}^P$ sein können. Die Sortierung wird allerdings erst im Fall von geschichtsabhängigen Materialverhalten relevant. Algorithmus 5.1 ist ebenfalls für KNN-Materialmodelle geeignet, welche wie in Definition (5.2) die Spannungen \mathbf{S} und die Materialtangente \mathbf{C}_T berechnen, vgl. Definition (5.1).

5.1.2 Anreicherung der Trainingsdaten

Falls materielle Symmetrien bekannt sind, können gegebene kartesische Datenpaare $\{(\mathbf{E}_k, \mathbf{S}_k)\}_{k=1}^P$ unter Ausnutzung der zugehörigen Symmetriebedingungen angereichert werden. Dies wird auch als *Data Augmentation* bezeichnet. Es ist vor allem dann sinnvoll, wenn die Daten in begrenzter Zahl vorliegen, wie bspw. bei realen Experimenten, oder in Fällen bei denen die Generierung solcher Daten viel Zeit in Anspruch nimmt, wie bspw. auf Basis numerischer Homogenisierung. Dazu wird die zugehörige Symmetriebedingung (3.37) ausgenutzt. Gegeben sei ein Datenpaar $(\mathbf{E}_k, \mathbf{S}_k)$ in Matrixnotation, sowie die Transformationsmatrix \mathbf{Q} von einer Symmetrieklasse wie aus Abschnitt 3.2.3. Dann ist

$$(\mathbf{E}_k^*, \mathbf{S}_k^*) = (\mathbf{Q}^T \mathbf{E}_k \mathbf{Q}, \mathbf{Q}^T \mathbf{S}_k \mathbf{Q}) \quad (5.8)$$

ein neues Datenpaar, welches zum Training oder zur Validierung des KNN verwendet werden kann. Diese Umformung basiert auf der Eigenschaft aus Gleichung (3.39).

Im Fall von Isotropie können die Matrizen \mathbf{Q} zufällig in der $SO(3)$ -Gruppe generiert werden. Bei transversaler Isotropie müssen sie nach Definition (3.41) erzeugt werden, mit fester materieller Achse $\boldsymbol{\omega}^{\text{TS}}$. Für orthotropes Materialverhalten können maximal drei Spiegelungsmatrizen genutzt werden, vgl. Definition (3.40). Welche materielle Symmetrie als Annahme für das zu modellierende Material gültig ist, muss jeweils vom Anwender bestimmt werden. Dabei können u. a. die Mikrostruktur und der Herstellprozess herangezogen werden.

Falls die materielle Objektivität durch die Wahl anderer Verzerrungs- und Spannungsmaße nicht explizit erfüllt ist, kann die beschriebene Anreicherung der Trainingsdaten auch zur Rotation der Ortsvektoren der Momentankonfiguration genutzt werden. Dies wurde bspw. von LEFIK und SCHREFLER [112] durchgeführt.

5.2 KNN-Materialmodelle für elastisches Materialverhalten (KNN-E)

Die Spezifikation eines KNN-Materialmodells erfolgt maßgeblich durch die Definition der Eingangs- und Ausgangsvariablen \mathbf{x} und \mathbf{z} . Dies muss in solcher Weise erfolgen, dass die zugehörige Abbildungsvorschrift $\mathbf{z} = \mathbf{f}^{\text{KNN}}(\mathbf{x}, \mathbf{w})$ eine Funktion darstellt. Im Fall elastischen Materialverhaltens ist das durch die Zuweisung von Verzerrungs- und Spannungsgrößen zum Ein- und Ausgang i. d. R. erfüllt⁵⁴. Als KNN-Architekturen für elastisches Materialverhalten eignen sich das FNN aus Abschnitt 4.2 und das ICNN bzw. NCNN aus Abschnitt 4.3.1.

Die Wahl der Eingangs- und Ausgangsvariablen bestimmt und limitiert gleichzeitig die Anwendungsmöglichkeiten eines KNN-Materialmodells. In den Abschnitten 5.2.1 und 5.2.2 werden jeweils alle unabhängigen Verzerrungs- und Spannungs-komponenten verwendet. Dadurch ist Materialverhalten mit beliebiger Anisotropie abbildbar. Durch die Wahl von Invarianten als Eingangs- und Ausgangsvariablen sind die KNN-Materialmodelle der Abschnitte 5.2.3 und 5.2.4 nur für isotropes Materialverhalten gültig.

Alle folgenden Modelle basieren auf Spannungs- und Verzerrungsgrößen und sind somit nur zur Beschreibung von Körpern 1. Grades⁵⁵ gültig. Darüber hinaus sind die folgenden physikalischen Prinzipien a priori erfüllt, vgl. Abschnitt 3.2.

- **Massen- und Impulserhaltung:** Durch die zu lösenden Gleichungen bzgl. der Referenzkonfiguration erfüllt.
- **Drehimpulserhaltung:** Durch Berücksichtigung von maximal sechs unabhängigen Spannungskomponenten ($\mathbf{S} = \mathbf{S}^T$) erfüllt.
- **Materielle Objektivität:** Durch die Wahl rein auf der Referenzkonfiguration definierter Spannungs- und Verzerrungsmaße erfüllt.

Die Konsistenz zum ersten und zweiten Hauptsatz der Thermodynamik, vgl. 3.2.1, ist nicht ohne Weiteres gegeben. Sie wird an gegebener Stelle thematisiert. Ferner sind die Stabilitätsbedingungen der Polykonvexität nach Abschnitt 3.3.5 nicht automatisch erfüllt, können aber durch spezielle Wahl der Eingangs- und Ausgangsvariablen sowie der ICNN-Architektur erzwungen werden.

⁵⁴ Im Fall inkompressiblen Materialverhaltens muss der hydrostatische Druck gesondert betrachtet werden, vgl. Abschnitt 5.2.5.

⁵⁵ Eine Erweiterung auf andere Kontinuumstheorien wäre durchaus möglich, wenn in den Ein- und Ausgangsvariablen dementsprechend Verzerrungs- und Spannungsmaße höherer Ordnung berücksichtigt würden.

5.2.1 Anisotropes KNN-Materialmodell FNN-E-Ani

Im einfachsten Fall werden alle Verzerrungs- und Spannungskomponenten direkt als Eingangs- und Ausgangsvariablen des klassischen FNN definiert. Die Eingangsvariablen

$$\mathbf{x} := \mathbf{E} \in \mathbb{R}^6 \quad (5.9)$$

sind die sechs unabhängigen Verzerrungskomponenten. Die Spannungen \mathbf{S} werden folglich direkt als KNN-Ausgangsvektor

$$\mathbf{z} := \mathbf{S} \in \mathbb{R}^6 \quad (5.10)$$

berechnet. Die Materialtangente

$$\mathbf{C}_T = \mathbf{j}^{\text{KNN}} = \frac{\partial \mathbf{S}(\mathbf{E}, \mathbf{w})}{\partial \mathbf{E}} \neq \mathbf{C}_T^T \quad (5.11)$$

ist die JACOBI-Matrix der KNN-Funktion (5.10) und kann durch die in Abschnitt 4.2.2 gegebenen Algorithmen für das FNN berechnet werden. Dieses Standard-Modell wurde bspw. von LEFIK ET AL. [111] verwendet.

Mit diesem KNN-Materialmodell ist beliebiges anisotropes elastisches Materialverhalten abbildbar, da alle möglichen unabhängigen Verzerrungskomponenten berücksichtigt werden. Die Materialtangente (5.11) ist im Allgemeinen nicht symmetrisch. Daher ist die Existenz einer freien Energiedichte $\Psi(\mathbf{E})$ und folglich Energieerhaltung als Forderung des 1. Hauptsatzes der Thermodynamik nicht gewährleistet.

5.2.2 Anisotropes KNN-Materialmodell NCNN-E-Ani

Die Eingangs- und Ausgangsvariablen werden wie im vorausgegangenen Abschnitt durch die Gleichungen (5.9) und (5.10) definiert. Statt des FNN wird allerdings das ICNN aus Abschnitt 4.3.1 eingesetzt, ohne Einforderung der Konvexität von $\Psi(\mathbf{E})$. Dies stünde im Widerspruch zu physikalischen Anforderungen an das Materialgesetz (vgl. Abschnitt 3.3.5) und ist daher nicht zweckmäßig. Deshalb wird diese KNN-Architektur im Rahmen der vorliegenden Arbeit als *NCNN* bezeichnet. Die Direktverbindungen werden nicht benötigt.

Nach Gleichung (4.87) existiert das Potential

$$\mathbf{S} = \mathcal{S}^{\text{KNN}}(\mathbf{E}) = \frac{\partial \Psi(\mathbf{E}, \mathbf{w})}{\partial \mathbf{E}}, \quad \text{mit} \quad \dot{\Psi} = \mathbf{S} \dot{\mathbf{E}}. \quad (5.12)$$

Somit erfüllt das KNN-Materialmodell NCNN-E-Ani die CLAUSIUS-PLANCK-UNGleichung (3.25) und ist per Konstruktion energieerhaltend. Für die Tangente

$$\mathbf{C}_T = \frac{\partial^2 \Psi}{\partial \mathbf{E}^2} \quad (5.13)$$

müssen die zweifachen Ableitungen des KNN-Ausgangs nach den Eingangsvariablen berechnet werden. Dies ist aufwendig, aber analytisch möglich und wird im Algorithmus von Tafel 4.6 beschrieben. Dieses Modell funktioniert in Verbindung mit SOBOLEV-Training, vgl. CZARNECKI ET AL. [32] und wurde bspw. von VLASSIS ET AL. [181] verwendet.

Dieses KNN-Materialmodell ist konsistent mit den Hauptsätzen der Thermodynamik und erfüllt daher alle physikalischen Anforderungen, die zu Beginn von Abschnitt 3.2 aufgelistet wurden. Die Forderung der Polykonvexität ist jedoch nicht erfüllt.

5.2.3 Isotropes KNN-Materialmodell FNN-E-Iso

Für isotrope Materialgesetze besteht ein funktionaler Zusammenhang zwischen den Invarianten der Verzerrungen und denjenigen der Spannungen. Die Modellierung eines solchen Zusammenhangs ist i. d. R. einfacher, als die Abbildung der letzten beiden KNN-Materialmodelle. Dadurch kann die Menge nötiger Trainingsdaten stark reduziert werden. Ziel dieses Abschnitts ist es daher, ein KNN-Materialmodell in diesen reduzierten Eingangs- und Ausgangsgrößen zu formulieren. Als Verzerrungsmaß wird der Rechts-CAUCHY-GREEN-Tensor \mathbf{C} genutzt, da dessen Invarianten leichter zu handhaben und physikalisch interpretierbarer sind. Das nachfolgend formulierte elastische KNN-Materialmodell ist nach bestem Wissen des Autors so noch nicht in der Literatur veröffentlicht worden.

Durch entsprechende Umrechnung der Verzerrungen \mathbf{C} und Spannungen \mathbf{S} in ihre Invarianten könnte ein KNN problemlos dazu trainiert werden, diesen Zusammenhang zu approximieren. Die Herausforderung liegt in der Formulierung des zugehörigen KNN-Materialmodells, welches die *kartesischen* Spannungen \mathbf{S} und die zugehörige kartesische Materialtangente \mathbf{C}_T berechnen muss. Dabei können die folgenden Probleme auftreten:

- Bei direkter Formulierung in den Hauptinvarianten mittels FNN, z. B.

$$[I_C \ II_C \ III_C] \mapsto [I_S \ II_S \ III_S] ,$$

geht die Information der kartesischen Basis verloren. Nach Berechnung der Invarianten des Spannungstensors \mathbf{S} mittels KNN wäre demzufolge eine Berechnung der Hauptspannungen $\{S_a\}_{a=1}^3$ sowie der Eigenvektoren $\{\mathbf{n}_a\}_{a=1}^3$ von \mathbf{C} nötig, um unter Ausnutzung der Koaxialität von \mathbf{C} und \mathbf{S} über die spektrale Zerlegung (3.12) die kartesischen Spannungen \mathbf{S} zu berechnen. Dies ist prinzipiell möglich. In der Nähe mehrfacher Hauptspannungen ($S_a \approx S_b$) können durch die Approximationsfehler des KNN allerdings negative Wurzelargumente und daher komplexe Zahlen auftreten, die ‚händisch‘

behalten werden müssten. Darüber hinaus muss bei der Berechnung der Materialtangente \mathbf{C}_T die Ableitung des Spannungstensors \mathbf{S} nach seinen Invarianten gebildet werden, was nach Auffassung des Autors nur mit der Pseudoinversen nach MOORE [134] und PENROSE [143] möglich ist. Aus den genannten Gründen wird diese Möglichkeit hier nicht weiterverfolgt.

- Ebenfalls denkbar ist eine direkte Formulierung

$$\left[\lambda_1^2 \quad \lambda_2^2 \quad \lambda_3^2 \right] \mapsto \left[S_1 \quad S_2 \quad S_3 \right] ,$$

mit den Eigenwerten $\{\lambda_a^2\}_{a=1}^3$ von \mathbf{C} und den Hauptspannungen $\{S_a\}_{a=1}^3$, den Eigenwerten von \mathbf{S} . Die erste Problematik ist die fehlende natürliche Sortierung von Eigenwerten, was bspw. mit der Festlegung $S_1 \leq S_2 \leq S_3$ und $\lambda_1^2 \leq \lambda_2^2 \leq \lambda_3^2$ behoben werden könnte. Die Berechnung der kartesischen Spannungen \mathbf{S} erfolgt über die spektrale Zerlegung (3.12), für welche die Eigenvektoren von \mathbf{C} verwendet werden können. Auch hier liegt die Problematik in der Berechnung der Materialtangente \mathbf{C}_T , und zwar im Fall mehrfacher Eigenwerte. An diesen Stellen ist die übergeordnete Tensorfunktion $\mathbf{S}(\mathbf{C})$ nur noch in Richtung der mehrfachen Eigenwerte differenzierbar, jedoch nicht mehr allgemein. MAGNUS [123] hat dieses Problem anschaulich dargestellt. Die für analytische Materialmodelle angewendete Grenzwertbildung nach DE L'HOSPITAL, vgl. Gleichung (3.58), ist beim KNN nicht möglich, da durch dessen Approximationsfehler der Grenzwert $\lim_{\lambda_b \rightarrow \lambda_a} S_b = S_a$ nicht eingehalten ist. Dadurch zeigt dieses Materialmodell im Bereich mehrfacher Eigenwerte schlechte Konvergenzeigenschaften⁵⁶, vor allem innerhalb der FEM. Daher wird dieser Weg hier ebenfalls nicht weiterverfolgt.

Zur erfolgreichen Formulierung eines in den quadratischen Hauptstreckungen $\{\lambda_a^2\}_{a=1}^3$ und den Hauptspannungen $\{S_a\}_{a=1}^3$ definierten KNN-Materialmodells wird das nach RIVLIN und ERICKSEN [149] benannte Repräsentationstheorem für isotrope Tensorfunktionen herangezogen, vgl. auch TRUESDELL und NOLL [175]. Formuliert in den 2. PIOLA-KIRCHHOFF'schen Spannungen \mathbf{S} und dem Rechts-CAUCHY-GREEN-Tensor \mathbf{C} , lautet es

$$\mathbf{S} = \alpha_0 \mathbf{I} + \alpha_1 \mathbf{C} + \alpha_2 \mathbf{C}^2 , \quad (5.14)$$

mit den von den Hauptinvarianten I_C , II_C und III_C abhängigen Koeffizienten α_0 , α_1 und α_2 und dem Einheitstensor 2. Stufe \mathbf{I} . Das Quadrat von \mathbf{C} lässt sich mithilfe des CAYLEY-HAMILTON-Theorems über

$$\mathbf{C}^2 = -III_C \mathbf{I} + I_C \mathbf{C} + III_C \mathbf{C}^{-1} \quad (5.15)$$

⁵⁶ Außerhalb dieser kritischen Stellen verhält sich das KNN-Materialmodell exzellent. Da die in Versuchen gefahrenen Dehnungszustände meist mehrfache Eigenwerte aufweisen, sind aber gerade diese problematischen Zustände relevant.

darstellen. Eingesetzt in Gleichung (5.14) folgt

$$\mathbf{S} = (\alpha_0 - \alpha_2 II_{\mathbf{C}}) \mathbf{I} + (\alpha_1 + I_{\mathbf{C}} \alpha_2) \mathbf{C} + (\alpha_2 III_{\mathbf{C}}) \mathbf{C}^{-1} . \quad (5.16)$$

Die Tensoren \mathbf{S} und \mathbf{C} sind koaxial, sie besitzen dieselben Eigenvektoren $\{\mathbf{n}_a\}_{a=1}^3$, vgl. Abschnitt 3.3.1. Das Eigenwertproblem von \mathbf{C}^{-1}

$$\mathbf{C}^{-1} \mathbf{n}_a = \lambda_a^{-2} \mathbf{n}_a , \quad a = 1, 2, 3 \quad (5.17)$$

resultiert durch Multiplikation des Eigenwertproblems von \mathbf{C} mit \mathbf{C}^{-1} und Division mit λ_a^2 , den quadratischen Hauptstreckungen. Daraus folgt, dass auch \mathbf{C}^{-1} die Eigenvektoren $\{\mathbf{n}_a\}_{a=1}^3$ teilt, die Eigenwerte sind die Kehrwerte der quadratischen Hauptstreckungen. Der Einheitstensor 2. Stufe $\mathbf{I} = \mathbf{1} \mathbf{n}_a \otimes \mathbf{n}_a$ kann ebenfalls in diesen Eigenvektoren formuliert werden, wenn diese als Orthonormalbasis definiert sind mit $\mathbf{n}_a \cdot \mathbf{n}_b = \delta_{ab}$. Werden für \mathbf{S} , \mathbf{I} , \mathbf{C} und \mathbf{C}^{-1} die besprochenen Spektralformen in Gleichung (5.16) eingesetzt, folgt

$$\sum_{a=1}^3 S_a \mathbf{n}_a \otimes \mathbf{n}_a = \sum_{a=1}^3 \left((\alpha_0 - \alpha_2 II_{\mathbf{C}}) + \lambda_a^2 (\alpha_1 + I_{\mathbf{C}} \alpha_2) + \lambda_a^{-2} \alpha_2 III_{\mathbf{C}} \right) \mathbf{n}_a \otimes \mathbf{n}_a . \quad (5.18)$$

Daraus lässt sich schließen, dass für jedes beliebige isotrope Materialgesetz eine Funktion \hat{S} existieren muss, mit der die Hauptspannungen

$$S_a = \hat{S}(\lambda_a^2, I_{\mathbf{C}}, II_{\mathbf{C}}, III_{\mathbf{C}}) , \quad a = 1, 2, 3 , \quad (5.19)$$

jeweils in Abhängigkeit von den zugehörigen quadratischen Hauptstreckung λ_a^2 und den drei Hauptinvarianten $I_{\mathbf{C}}$, $II_{\mathbf{C}}$ und $III_{\mathbf{C}}$ berechnet werden können⁵⁷.

Die Spur $I_{\mathbf{C}}$ lässt sich über

$$I_{\mathbf{C}} = \frac{II_{\mathbf{C}}}{\lambda_a^2} - \frac{III_{\mathbf{C}}}{(\lambda_a^2)^2} + \lambda_a^2 , \quad a = 1, 2, 3 \quad (5.20)$$

stets in Abhängigkeit der anderen Eingangsvariablen formulieren⁵⁸. Sie wird daher im Folgenden nicht berücksichtigt. Letztlich ist die Idee des folgenden KNN-Materialmodells die Approximation der Funktion $\hat{S}(\lambda_a^2, II_{\mathbf{C}}, III_{\mathbf{C}})$.

Somit folgt für die Definition des KNN-Eingangsvektors

$$\mathbf{x}_a := \left[\lambda_a^2 \quad II_{\mathbf{C}} \quad J^2 \right] \in \mathbb{R}^3 , \quad a = 1, 2, 3 , \quad (5.21)$$

⁵⁷ Eine solche Separation kann bspw. im inkompressiblen Grenzfall auch für die Verzerrungsenergiegedichtefunktion gemacht werden. Dies wird Hypothese nach VALANIS und LANDEL [179] genannt. Sie lässt sich nach OGDEN [141] an Versuchen von JONES und TRELOAR [92] bestätigen, gilt jedoch nur für inkompressibles Materialverhalten und führt nicht zu einer solchen Separation der Gesamtspannungen.

⁵⁸ Bspw. $II_{\mathbf{C}} = \lambda_1^2 \lambda_2^2 + \lambda_1^2 \lambda_3^2 + \lambda_2^2 \lambda_3^2 = \lambda_1^2 (\lambda_2^2 + \lambda_3^2 + \lambda_2^2 \lambda_3^2 \lambda_1^2 / \lambda_1^4) = \lambda_1^2 (I_{\mathbf{C}} - \lambda_1^2 + III_{\mathbf{C}} / \lambda_1^4)$

mit $J^2 = III_{\mathbf{C}}$. Die einzige Ausgangsvariable ist die jeweils zu $x_1 = \lambda_a^2$ gehörige Hauptspannung

$$z_a := S_a \in \mathbb{R}, \quad a = 1, 2, 3. \quad (5.22)$$

Jede Hauptspannung S_a wird somit über eine von insgesamt drei separaten KNN-Vorwärtsberechnungen mit je $\mathbf{x}_a = [\lambda_a^2, II_{\mathbf{C}}, J^2]$ ermittelt. An dieser Stelle sei nochmals hervorgehoben, dass dabei stets dasselbe KNN verwendet wird. Durch diese Definitionen ist das Zusammenfallen zweier Hauptspannungen bei zugehörigen mehrfachen Hauptstreckungen gewährleistet. Der Grenzwert

$$\lim_{\lambda_b \rightarrow \lambda_a} S_b = S_a \quad (5.23)$$

ist erfüllt und die Regeln von DE L'HOSPITAL zur Berechnung der Materialtangente können angewendet werden. Die Probleme der weiter oben beschriebenen alternativen Strategien treten hier somit nicht auf.

Die kartesischen Spannungen \mathbf{S} können nach Berechnung der Hauptspannungen über die Transformation (3.65)₁ bestimmt werden, mit der von den Eigenvektoren von \mathbf{C} abhängigen Transformationsmatrix nach Gleichung (3.63). Zur Berechnung der Materialtangente \mathbf{C}_T können ebenfalls die in Abschnitt 3.3.1 angegebenen Beziehungen genutzt werden. Für jeden Eingangsvektor $\mathbf{x}_a = [\lambda_a^2, II_{\mathbf{C}}, J^2]$ wird die JACOBI-Matrix

$$\mathbf{j}^a := \mathbf{j}(\mathbf{x}_a) = \begin{bmatrix} \frac{\partial S_a}{\partial x_1} & \frac{\partial S_a}{\partial x_2} & \frac{\partial S_a}{\partial x_3} \end{bmatrix} = \begin{bmatrix} j_{11}^a & j_{12}^a & j_{13}^a \end{bmatrix} \in \mathbb{R}^{1 \times 3} \quad (5.24)$$

berechnet. Mit ihr werden die Komponenten $\{\hat{C}_{aabb}\}_{b=1}^3$ der spektralen Matrix (3.62)₁, analog zu Gleichung (3.57), mit

$$\hat{C}_{aabb} = 2 \frac{\partial S_a}{\partial \lambda_b^2} = \begin{cases} 2j_{11}^a + 2j_{12}^a(I_{\mathbf{C}} - \lambda_a^2) + 2j_{13}^a \left(\frac{J^2}{\lambda_a^2} \right) & b = a \\ 2j_{12}^a(I_{\mathbf{C}} - \lambda_b^2) + 2j_{13}^a \left(\frac{J^2}{\lambda_b^2} \right) & b \neq a \end{cases} \quad (5.25)$$

ermittelt, unter Berücksichtigung der Kettenregel sowie der Unabhängigkeit der drei KNN-Berechnungen ($\partial S_a / \partial \lambda_b^2 = 0$). Im Anschluss an die drei KNN-Berechnungen werden die drei Komponenten \hat{C}_{abab} der Matrix (3.62)₂, mit $(a, b) \in \{(1, 2), (1, 3), (2, 3)\}$ äquivalent zu Gleichung (3.58) über

$$\hat{C}_{abab} = \frac{S_b - S_a}{\lambda_b^2 - \lambda_a^2} \xrightarrow{\lambda_a = \lambda_b} \frac{\partial S_b}{\partial \lambda_b^2} - \frac{\partial S_a}{\partial \lambda_b^2} = \frac{1}{2} (\hat{C}_{bbbb} - \hat{C}_{aabb}) \quad (5.26)$$

berechnet. In diesem Fall darf der Satz von DE L'HOSPITAL angewendet werden. Nach Einsetzen der Komponenten \hat{C}_{aabb} und \hat{C}_{abab} in die Matrizen $\hat{\mathbf{C}}_1$ und $\hat{\mathbf{C}}_2$ aus Definition (3.62) kann die kartesische Materialtangente \mathbf{C}_T mithilfe der Transformationsmatrizen (3.63) und (3.64) mit der Vorschrift (3.65) berechnet werden. In

Eingang	$\mathbf{E} = [E_{11}, E_{22}, E_{33}, 2E_{12}, 2E_{13}, 2E_{23}]$
Umrechnung	$\mathbf{C} = 2\mathbf{E} + \mathbf{I}$ in Matrixnotation
spez. EWP	$\mathbf{C}\mathbf{n} = \lambda^2\mathbf{n}$ für die Eigenwerte $\{\lambda_a^2\}_{a=1}^3$ und die Eigenvektoren $\{\mathbf{n}_a\}_{a=1}^3$
KNN-Schleife	for $a = 1, 2, 3$ do $\mathbf{x}_a = [\lambda_a^2, II_{\mathbf{C}}, J^2]$, mit $J^2 = \lambda_1^2\lambda_2^2\lambda_3^2$ und $II_{\mathbf{C}} = \lambda_1^2\lambda_2^2 + \lambda_1^2\lambda_3^2 + \lambda_2^2\lambda_3^2$ $S_a = f^{\text{KNN}}(\mathbf{x}_a, \mathbf{w})$, $\mathbf{j}^a = \partial f^{\text{KNN}}(\mathbf{x}_a, \mathbf{w})/\partial \mathbf{x}$ (<u>KNN</u>) $\hat{C}_{aa11}, \hat{C}_{aa22}, \hat{C}_{aa33}$ nach Gleichung (5.25) end for
Spektr. Felder	$\hat{C}_{1212}, \hat{C}_{1313}, \hat{C}_{2323}$ nach Gleichung (5.26), $\hat{\mathbf{S}} = [S_1, S_2, S_3]^T$
	$\hat{\mathbf{C}}_1 = \begin{bmatrix} \hat{C}_{1111} & \hat{C}_{1122} & \hat{C}_{1133} \\ \hat{C}_{2211} & \hat{C}_{2222} & \hat{C}_{2233} \\ \hat{C}_{3311} & \hat{C}_{3322} & \hat{C}_{3333} \end{bmatrix}$, $\hat{\mathbf{C}}_2 = \begin{bmatrix} \hat{C}_{1212} & 0 & 0 \\ 0 & \hat{C}_{1313} & 0 \\ 0 & 0 & \hat{C}_{2323} \end{bmatrix}$
Trafo	Matrizen $\mathbf{Q}_1, \mathbf{Q}_2$ nach Gleichungen (3.63) und (3.64)
Ausgang	$\mathbf{S} = \mathbf{Q}_1\hat{\mathbf{S}}$ und $\mathbf{C}_T = \mathbf{Q}_1\hat{\mathbf{C}}_1\mathbf{Q}_1^T + \mathbf{Q}_2\hat{\mathbf{C}}_2\mathbf{Q}_2^T$

Tafel 5.2: Übersicht zum isotropen KNN-Materialmodell FNN-E-Iso

Tafel 5.2 sind die nötigen Schritte zur Anwendung dieses KNN-Materialmodells noch einmal zusammengefasst.

Zum Training des KNN müssen die (\mathbf{E}, \mathbf{S}) -Paare aus Pfad (5.7) zunächst in die (\mathbf{x}, \mathbf{t}) -Paare umgerechnet werden. Dazu wird der GREEN-LAGRANGE'sche Verzerrungstensor \mathbf{E} in den Rechts-CAUCHY-GREEN-Tensor $\mathbf{C} = 2\mathbf{E} + \mathbf{I}$ umgeformt, mit dem Einheitstensor 2. Stufe \mathbf{I} . Anschließend wird jeweils das spezielle Eigenwertproblem (3.11) für \mathbf{C} und \mathbf{S} gelöst, um die Eigenwerte $\{\lambda_a^2\}_{a=1}^3$ und $\{S_a\}_{a=1}^3$ sowie die normierten Eigenvektoren $\{\mathbf{n}_a\}_{a=1}^3$ zu erhalten. Die Eigenwerte müssen nun jeweils drei Paaren $\{(\lambda_a^2, S_a)\}_{a=1}^3$ zugeteilt werden. Dabei ist zu beachten, dass im Fall großer Verzerrungen die größte (kleinste) Hauptstreckung nicht unbedingt zur größten (kleinsten) Hauptspannung gehört. Die korrekte Zuweisung muss über die Dyaden $\{\mathbf{n}_a \otimes \mathbf{n}_a\}_{a=1}^3$ erfolgen, welche im Gegensatz zu den Eigenvektoren eindeutig sind, vgl. Gleichung (5.18). Mit den Invarianten $II_{\mathbf{C}}$ und $J^2 = III_{\mathbf{C}}$ werden letztlich für jedes (\mathbf{E}, \mathbf{S}) -Paar drei Trainingsdatenpaare

$$\left([\lambda_1^2, II_{\mathbf{C}}, J^2], S_1\right), \quad \left([\lambda_2^2, II_{\mathbf{C}}, J^2], S_2\right) \quad \text{und} \quad \left([\lambda_3^2, II_{\mathbf{C}}, J^2], S_3\right) \quad (5.27)$$

berechnet. Diese können unabhängig voneinander zum Training oder zur Validierung genutzt werden.

Das hier beschriebene KNN-Materialmodell ist ausschließlich für isotropes Ma-

terialverhalten geeignet. Die Materialtangente \mathbf{C}_T ist nicht symmetrisch, daher ist Energieerhaltung nicht gewährleistet. Dementsprechend sind dadurch automatisch auch die Bedingungen der Polykonvexität nicht erfüllt. Durch Nutzung des ICNN als KNN-Architektur könnte die Symmetrie der Materialtangente \mathbf{C}_T ebenfalls nicht erzwungen werden, da die drei Hauptspannungen S_1 , S_2 und S_3 in voneinander unabhängigen KNN-Vorwärtsberechnungen ermittelt werden. Die Existenz von drei Potentialfunktionen

$$\Psi(\lambda_a^2, II_{\mathbf{C}}, J^2) \quad \text{mit} \quad S_a = 2 \left(\frac{\partial \Psi}{\partial \lambda_a^2} + \frac{\partial \Psi}{\partial III_{\mathbf{C}}} (I_{\mathbf{C}} - \lambda_a^2) + \frac{\partial \Psi}{\partial J^2} \frac{J^2}{\lambda_a^2} \right) \quad (5.28)$$

für jeweils *eine* Hauptspannung impliziert nicht die Existenz einer Potentialfunktion für *alle* Hauptspannungen. Energieerhaltung könnte jedoch über eine Nebenbedingung im Trainingsprozess eingefordert werden, was in Abschnitt 7.1 getan wird.

5.2.4 Polykonvexes isotropes KNN-Materialmodell ICNN-E-Poly

Durch Nutzung des ICNN haben KLEIN ET AL. [103] polykonvexe, anisotrope KNN-Materialmodelle definiert und auf Daten aus numerischer Homogenisierung angewendet. Im vorliegenden Abschnitt wird eine rein isotrop formulierte Version dieses Modells erläutert, um sie später mit den zuvor eingeführten KNN-Materialmodellen zu vergleichen.

Der Eingangsvektor

$$\mathbf{x} := [I_{\mathbf{C}} \quad II_{\mathbf{C}} \quad III_{\mathbf{C}} \quad -2\sqrt{III_{\mathbf{C}}}]^T \in \mathbb{R}^4 \quad (5.29)$$

ist in den Invarianten des Rechts-CAUCHY-GREEN-Tensors \mathbf{C} definiert. Die vierte Eingangsvariable $x_4 = -2\sqrt{III_{\mathbf{C}}} = -2J$ erhöht die Flexibilität des polykonvexen KNN-Materialmodells. Dies wird unten näher erläutert. Im Ausgangsvektor

$$\mathbf{z} := \mathbf{S} \in \mathbb{R}^6 \quad (5.30)$$

befinden sich die sechs unabhängigen kartesischen Spannungskomponenten des 2. PIOLA-KIRCHHOFF'schen Spannungstensors \mathbf{S} .

Ein direkter funktionaler Zusammenhang $\mathbf{z}(\mathbf{x})$ besteht nicht. Die Nutzung des FNN als KNN-Architektur ist daher aussichtslos. Dieses KNN-Materialmodell ist somit untrennbar an die ICNN-Architektur gebunden, über die mithilfe der Ableitungen der Potentialfunktion $\Psi(\mathbf{x})$ die Spannungen

$$\mathbf{S} = 2 \frac{\partial \Psi}{\partial \mathbf{C}} = 2 \left(\left(\frac{\partial \Psi}{\partial x_1} + I_{\mathbf{C}} \frac{\partial \Psi}{\partial x_2} \right) \mathbf{I} - \left(\frac{\partial \Psi}{\partial x_2} \right) \mathbf{C} + \left(J^2 \frac{\partial \Psi}{\partial x_3} - J \frac{\partial \Psi}{\partial x_4} \right) \mathbf{C}^{-1} \right) \quad (5.31)$$

eindeutig in Abhängigkeit der Invarianten definiert werden können. Für die reduzierten Notationen der Tensoren 2. Stufe \mathbf{I} , \mathbf{C} und \mathbf{C}^{-1} , vgl. Anhang B. Die kartesische Materialtangente \mathbb{C}_T lässt sich durch weitere Differentiation ebenfalls über

$$\begin{aligned} \mathbf{C}_T = 4 \frac{\partial^2 \Psi}{\partial \mathbf{C}^2} = 4 \left(s_1 (\mathbf{I} \otimes \mathbf{C} + \mathbf{C} \otimes \mathbf{I}) + s_2 (\mathbf{I} \otimes \mathbf{C}^{-1} + \mathbf{C}^{-1} \otimes \mathbf{I}) \right. \\ \left. + s_3 (\mathbf{C} \otimes \mathbf{C}) + s_4 (\mathbf{C} \otimes \mathbf{C}^{-1} + \mathbf{C}^{-1} \otimes \mathbf{C}) \right. \\ \left. + s_5 (\mathbf{C}^{-1} \otimes \mathbf{C}^{-1}) + s_6 \frac{\partial \mathbf{C}^{-1}}{\partial \mathbf{C}} + s_7 \mathbf{I}^S + s_8 (\mathbf{I} \otimes \mathbf{I}) \right) \end{aligned} \quad (5.32)$$

angeben. Die reduzierten Notationen dieser darin enthaltenen elf Tensoren 4. Stufe sind mithilfe von Anhang B berechenbar. Die acht Koeffizienten sind

$$\begin{aligned} s_1 &= -\Psi_{,12} - I_{\mathbf{C}} \Psi_{,22} , \\ s_2 &= J^2 \Psi_{,13} - J \Psi_{,14} + I_{\mathbf{C}} J^2 \Psi_{,23} - I_{\mathbf{C}} J \Psi_{,24} , \\ s_3 &= \Psi_{,22} , \\ s_4 &= J \Psi_{,24} - J^2 \Psi_{,23} , \\ s_5 &= J^2 (\Psi_{,3} + \Psi_{,44}) + J^4 \Psi_{,33} - 2J^3 \Psi_{,34} - \frac{1}{2} J \Psi_{,4} , \\ s_6 &= J^2 \Psi_{,3} - J \Psi_{,4} , \\ s_7 &= -\Psi_{,2} \quad \text{und} \\ s_8 &= \Psi_{,11} + 2I_{\mathbf{C}} \Psi_{,12} + I_{\mathbf{C}}^2 \Psi_{,22} + \Psi_{,2} , \end{aligned} \quad (5.33)$$

mit $I_{\mathbf{C}} = \text{Sp}(\mathbf{C})$ und $J = \sqrt{\det \mathbf{C}}$. Für die ersten und zweiten partiellen Ableitungen der Potentialfunktion wurden die Abkürzungen $\Psi_{,i} = \partial \Psi / \partial x_i$ und $\Psi_{,ij} = \partial^2 \Psi / \partial x_i \partial x_j$ verwendet. Sie müssen über die nichttransformierte, also physikalische Potentialfunktion

$$\Psi(\mathbf{x}) = \bar{s}_z \bar{s}_x \hat{\Psi}(\mathbf{x}) \quad (5.34)$$

berechnet werden, vgl. Gleichung (4.87). Dabei ist zur Existenz jener Potentialfunktion nicht nur die Mittelung der Streckungsparameter \bar{s}_x und \bar{s}_z nach Definition (4.86) nötig, sondern auch das Entfernen etwaiger Mittelwertbereinigungen der Ausgangsvariablen mit $\mathbf{m}_z = \mathbf{0}$, vgl. Definition (4.9).

Mit den dargelegten Sachverhalten ist das KNN-Materialmodell durch die Existenz der physikalischen Potentialfunktion (5.34) zwar energieerhaltend, es erfüllt allerdings noch nicht die Forderungen nach Polykonvexität aus Abschnitt 3.3.5. Nach diesen Forderungen muss die Potentialfunktion Ψ konvex und monoton wachsend in den Invarianten von \mathbf{C} sein. Diese Einschränkungen sind schärfer als die des ICNN in Abschnitt 4.3.1, welche nur die Konvexität bzgl. der Eingangsvariablen einfordern. Daher müssen zur Einforderung der Polykonvexität insgesamt die folgenden Bedingungen erfüllt sein:

- a) Alle Gewichte müssen größer null sein, auch die der ersten Zwischenschicht und die der Direktverbindungen. Die Bias-Gewichte sind davon befreit.
- b) Alle Aktivierungsfunktionen müssen monoton wachsend sein.

Ein in den drei Invarianten von \mathbf{C} definiertes KNN-Materialmodell, welches innerhalb der ICNN-Architektur diese Einschränkungen erfüllt, ist sicher polykonvex. Allerdings sind die Bedingungen schärfer als diejenigen, die aus der Theorie nach Abschnitt 3.3.5 folgen. Wie dort beschrieben, muss die Potentialfunktion nur konvex in $J = \sqrt{III_{\mathbf{C}}}$ sein, nicht aber in $III_{\mathbf{C}}$. Außerdem wären negative Spannungen nur durch eine direkte Abhängigkeit von $II_{\mathbf{C}}$ möglich, vgl. Gleichung (5.31). Daher haben KLEIN ET AL. [103] die vierte Eingangsvariable $x_4 = -2\sqrt{III_{\mathbf{C}}} = -2J$ eingeführt. Alternativ könnte statt $III_{\mathbf{C}}$ auch J direkt als Eingangsvariable definiert werden. Dann könnten die zugehörigen Gewichte zwischen J und der ersten Zwischenschicht von der Nichtnegativitäts-Regel a) befreit werden, um den gleichen Effekt zu erzielen.

5.2.5 Exakte Einforderung der Inkompressibilitätsbedingung

In Abschnitt 3.3.4 wurde beschrieben, wie die Zwangsbedingung der Inkompressibilität durch Einführung des unabhängigen hydrostatischen Drucks p exakt innerhalb einer numerischen Berechnung eingefordert werden kann. Um dies mit KNN-Materialmodellen durchführen zu können, wird der Ausgangsvektor \mathbf{z} so umformuliert, dass ausschließlich die deviatorischen Spannungen

$$\mathbf{S}_{\text{dev}} = \mathbf{S} + p\mathbf{C}^{-1} \quad (5.35)$$

mittels KNN berechnet werden. Bei einem gegebenem Datenpaar $(\mathbf{E}_k, \mathbf{S}_k)$ wird der zugehörige hydrostatische Druck mit

$$p_k = -\frac{1}{3}\text{Sp}(\mathbf{C}\mathbf{S}) \quad (5.36)$$

berechnet. Für kleine Verzerrungen wäre $p_k \approx -\text{Sp}(\mathbf{S})/3$ ebenfalls möglich. Besagtes Datenpaar kann folglich mit Gleichung (5.35) zu $(\mathbf{E}_k, \mathbf{S}_{\text{dev}k})$ umgeformt und anschließend zum Training genutzt werden. Alle zuvor definierten KNN-Materialmodelle können mit dieser Methode für beliebig große Verzerrungen inkompressibel erweitert werden. Zur Anwendung dieser Modelle ist jedoch eine Modifikation des Materialpunkt-Versuchs nach Abschnitt 5.1.1 notwendig. Dies wird genauer in Anhang D beschrieben.

Eine Anpassung der Eingangsvariablen \mathbf{x}_k ist im Allgemeinen nicht notwendig, verringert aber ggf. die Anzahl der notwendigen Freiheitsgrade und Trainingsdaten. Für die isotropen Modelle aus den Abschnitten 5.2.3 und 5.2.4 würde

es sich anbieten, für den Fall exakter Inkompressibilität das Volumenverhältnis J^2 aus dem Eingangsvektor zu entfernen. An dieser Stelle sei nochmal erwähnt, dass die so modifizierten KNN-Materialmodelle nur dann nutzbar sind, wenn der hydrostatische Druck p über die äußeren Randbedingungen berechnet wird, entweder durch spezielle finite Elemente, oder durch den in Anhang D beschriebenen Materialpunkt-Versuch. Im Rahmen der vorliegenden Arbeit wird der KNN-Eingangsvektor nicht in dieser Hinsicht angepasst.

5.2.6 Erfüllung der Wachstumsbedingungen

Die Wachstumsbedingungen (3.51) und (3.52) für die Verzerrungsenergiedichtefunktion werden im Allgemeinen nicht von KNN-Approximationen erfüllt. Die Wachstumsbedingungen können dadurch erzwungen werden, dass die KNN-Abbildung nicht mehr die Spannungen \mathbf{S} direkt, sondern die Differenz zwischen den Spannungen und einem volumetrischen Anteil \mathbf{S}_{koe} darstellt⁵⁹. Da die Wachstumsbedingungen auch Koerzivitätsbedingungen genannt werden, wird der Index wie beschrieben gewählt. Dieser kann über beliebige kompressible Funktionen $\Psi_{\text{koe}}(J)$ hergeleitet werden, welche die genannten Grenzwerte erfüllen. Eine Übersicht solcher Funktionen ist bspw. in DOLL und SCHWEIZERHOF [36] zu finden. An dieser Stelle wird ein anderer Ansatz

$$\Psi_{\text{koe}}(J) = \frac{1}{4} \left((J-1)^2 + (\ln J)^2 \right) \xrightarrow{J=0, J=\infty} \infty \quad (5.37)$$

gewählt als der des OGDEN-Materialgesetzes aus Gleichung (3.77), da dieses zur Generierung der synthetischen Trainingsdaten und schließlich zur Verifizierung der KNN-Materialmodelle verwendet wird. Das KNN-Materialmodell soll keinen Bias in Richtung des analytischen Materialmodells haben.

Durch Differenzieren von Gleichung (5.37) nach \mathbf{E} resultieren die volumetrischen Spannungen

$$\mathbf{S}_{\text{koe}} = \frac{1}{2} \left(J(J-1) + \ln J \right) \mathbf{C}^{-1} \quad (5.38)$$

und die volumetrische Materialtangente

$$\mathbf{C}_{T_{\text{koe}}} = \frac{1}{2} \left(2J^2 - J + 1 \right) \left(\mathbf{C}^{-1} \otimes \mathbf{C}^{-1} \right) + \left(J^2 - J + \ln J \right) \frac{\partial \mathbf{C}^{-1}}{\partial \mathbf{C}}. \quad (5.39)$$

Die reduzierten Notationen des Tensors 2. Stufe \mathbf{C}^{-1} , des dyadischen Produkts $\mathbf{C}^{-1} \otimes \mathbf{C}^{-1} = \mathbf{C}^{-1} \mathbf{C}^{-T}$ sowie des Tensors 4. Stufe $\partial \mathbf{C}^{-1} / \partial \mathbf{C}$ sind in Anhang B beschrieben. Für $J = 1$ sind sowohl die Spannungen \mathbf{S}_{koe} als auch die Tangente $\mathbf{C}_{T_{\text{koe}}}$ null.

⁵⁹ Dieser Anteil ist nicht zu verwechseln mit dem gesamten volumetrischen Anteil \mathbf{S}_{vol} .

Zur Berücksichtigung der Wachstumsbedingungen innerhalb der Trainingsdaten müssen die volumetrischen Spannungen (5.38) von denen aus den Materialversuchen (5.7) subtrahiert werden. Während der Anwendung als Materialmodell müssen die Anteile (5.38) und (5.39) jeweils auf die Spannungen \mathbf{S} und die Materialtangente \mathbf{C}_T aus den Abschnitten 5.2.1 bis 5.2.3 addiert werden.

5.3 Physikalische Nebenbedingungen für E-KNN

In Abschnitt 4.6 wurde anhand eines Prinzipbeispiels gezeigt, dass die Berücksichtigung von Nebenbedingungen im Rahmen des KNN-Trainings einen positiven Effekt auf das Trainingsresultat haben kann. Im Rahmen der KNN-Materialmodellierung kann dies ebenfalls genutzt werden. Die physikalischen Prinzipien aus Kapitel 3 dienen dabei als Grundlage zur Formulierung zusätzlicher Fehlerterme, die innerhalb des Optimierungsprozesses minimiert werden. Die explizite Form der Fehlerterme hängt dabei neben der Nebenbedingung selbst auch von der Definition des KNN-Materialmodells ab. In diesem Kapitel werden die Nebenbedingungen ausschließlich für das ‚Standard‘-Modell FNN-E-Ani mit $\mathbf{x} = \mathbf{E}$ und $\mathbf{z} = \mathbf{S}$ formuliert und angewendet. Eine Übertragung auf andere Modelle ist sinngemäß möglich, bspw. um Energieerhaltung beim isotropen Materialmodell FNN-E-Iso einzufordern, vgl. Abschnitt 7.1.

Die Fehlerterme $h(\mathbf{x}_k^C, \mathbf{w})$ sowie ihre Gradienten $\nabla h(\mathbf{x}_k^C, \mathbf{w})$ werden nachfolgend für einen festen NB-Punkt \mathbf{x}_k^C angegeben. Es handelt sich dabei im Fall der Elastizität stets um Gleichheits-Nebenbedingungen ($h = 0$). An dieser Stelle sei nochmals angemerkt, dass dabei nur die Information der Eingangsvariablen \mathbf{x}^C benötigt wird und nicht die Information über mögliche Zielwerte \mathbf{t}^C , wie bei den Trainingsdatenpaaren. Die Berechnung der Fehlerterme \mathcal{L}_C sowie der Gradienten $\nabla \mathcal{L}_C$ geschieht für alle NB-Punkte $\{(\mathbf{x}_k^C)\}_{k=1}^{P_C}$. Sie hängt ferner von der verwendeten Penalty-Methode ab. Dies wurde bereits in Abschnitt 4.5.4 beschrieben.

5.3.1 Spannungsfreie Referenzkonfiguration

Meist wird von einer spannungsfreien Referenzkonfiguration ausgegangen. Diese Normalisierung wird von KNN-Materialmodellen nicht a priori erfüllt. Daher ist es zweckmäßig, dies als zusätzliche Nebenbedingung

$$\mathbf{S}_0 := \mathcal{S}^{\text{KNN}}(\mathbf{E} = \mathbf{0}) \stackrel{!}{=} \mathbf{0} \quad (5.40)$$

im Trainingsprozess zu berücksichtigen. Die Normalisierung wird für jede der n_s unabhängigen Spannungskomponenten einzeln⁶⁰ eingefordert. Unter Berücksich-

⁶⁰ Alternativ könnte bspw. auch die Norm von \mathbf{S}_0 bestraft werden.

tigung der Daten-Normalisierung aus Abschnitt 4.1 folgen die n_s Bedingungen

$$h_j = h_j(\mathbf{x}^0, \mathbf{w}) = \hat{z}_j(\mathbf{x}^0, \mathbf{w}) + \frac{m_{zj}}{s_{zj}} = 0, \quad j = 1, \dots, n_s. \quad (5.41)$$

Der Punkt $\mathbf{x}^C =: \mathbf{x}^0 = \mathbf{0}$ ist der einzige Punkt, an dem diese Nebenbedingung eingefordert werden kann⁶¹. Sie kann mit $t_j^0 = 0$ und $\hat{t}_j^0 = -m_{zj}/s_{zj}$ in Form eines zusätzlichen Trainingsdatenpunkts verwendet werden, weil in Gleichung (5.41) durch $s_{zj} > 0$ geteilt wurde. Das ist nicht für beliebige Nebenbedingungen möglich. Außerdem würde die Wichtung des Penalty-Parameters vernachlässigt. Der Gradient folgt mit

$$\nabla h_j = \frac{\partial \hat{z}_j}{\partial \mathbf{w}}. \quad (5.42)$$

Die Vorfaktoren zur Berechnung des Gradienten $\nabla \mathcal{L}_C$ nach Abschnitt 4.2.4 sind $d_{j1} = \rho_{h_j1}$, abhängig von den drei Penalty-Methoden, vgl. Definitionen (4.163) bis (4.165).

Die zusätzlichen Terme aus Abschnitt 5.2.6 zur Berücksichtigung der Wachstumsbedingungen haben keinen Einfluss auf diese Nebenbedingung. Die volumetrischen Spannungen \mathbf{S}_{koe} und die Anteile der Materialtangente $\mathbf{C}_{T\text{koe}}$ sind für $\mathbf{E} = \mathbf{0}$ null.

5.3.2 Energieerhaltung bzw. Symmetrie der Materialtangente

Ohne explizite Definition einer Verzerrungsenergiedichtefunktion $\Psi(\mathbf{E})$ kann nicht gewährleistet werden, dass die Hauptsätze der Thermodynamik erfüllt sind. Die Existenz einer Funktion $\Psi(\mathbf{E})$ ist äquivalent zur Hauptsymmetrie der Materialtangente $C_{ijkl} = C_{klij}$, vgl. Abschnitt 3.3. Die daraus folgende Symmetrie der Tangente in reduzierter Notation

$$\mathbf{C}_T \stackrel{!}{=} \mathbf{C}_T^T \quad (5.43)$$

kann innerhalb des Trainingsprozesses als Nebenbedingung eingefordert werden. Sie wird direkt in den partiellen Ableitungen der JACOBI-Matrix \mathbf{j} definiert. Jedes Paar (u, v) aus Nebendiagonalelementen $u \neq v$ entspricht dabei einem Term. Unter Berücksichtigung der Ableitungs-Transformation (4.34) lauten die Terme für jeden NB-Punkt \mathbf{x}_k und für jedes Indexpaar (u, v) mit $v = 1, \dots, n_s - 1$ und $u = v + 1, \dots, n_s$ wie folgt:

$$h_{vuk} = h_{vu}(\mathbf{x}_k^C, \mathbf{w}) = \left(\frac{s_{zv}}{\alpha_{vu}s_{xu}} \right) \frac{\partial \hat{z}_{vk}}{\partial \hat{x}_u} - \left(\frac{s_{zu}}{\alpha_{vu}s_{xv}} \right) \frac{\partial \hat{z}_{uk}}{\partial \hat{x}_v}. \quad (5.44)$$

⁶¹ Durch Erweiterung des Eingangsvektors \mathbf{x} kann das KNN-Materialmodell parametrisiert werden. Der Raum zur Einforderung dieser Nebenbedingung vergrößert sich damit auf den gesamten Parameterraum.

Dabei wurde für $\hat{z}_v(\mathbf{x}_k, \mathbf{w})$ die Abkürzung \hat{z}_{vk} geschrieben. Durch die Normalisierung der Trainingsdatenpaare ist die Trainingsdatenfehlerfunktion \mathcal{L}_T stets in einer ähnlichen Größenordnung, vgl. Abschnitt 4.1. Die Fehlerfunktion dieser Ableitungen kann jedoch durch die Faktoren s_{zv}/s_{xu} und s_{zu}/s_{xv} sehr große Werte annehmen. Diese Faktoren sind nicht kürzbar, da sie im Allgemeinen nicht gleich sind⁶². Daher empfiehlt es sich, die Nebenbedingungen zusätzlich zu normalisieren, was hier durch den Faktor

$$\alpha_{vu} := \max \left\{ \left| \frac{s_{zv}}{s_{xu}} \right|, \left| \frac{s_{zu}}{s_{xv}} \right| \right\} \quad (5.45)$$

durchgeführt wurde. Die Penalty-Parameter aus Abschnitt 4.5 sind dadurch besser zu interpretieren, mit einer gleichwertigen Fehlerterm-Größe bei $\epsilon \approx 1$. Der Gradient für jedes Indexpaar (u, v) und für jeden NB-Punkt \mathbf{x}_k^C ist

$$\nabla h_{vuk} = \left(\frac{s_{zv}}{\alpha_{vu} s_{xu}} \right) \frac{\partial}{\partial \mathbf{w}} \left(\frac{\partial \hat{z}_{vk}}{\partial \hat{x}_u} \right) - \left(\frac{s_{zu}}{\alpha_{vu} s_{xv}} \right) \frac{\partial}{\partial \mathbf{w}} \left(\frac{\partial \hat{z}_{uk}}{\partial \hat{x}_v} \right). \quad (5.46)$$

Zur Berechnung des Gradienten $\nabla \mathcal{L}_C$ nach Abschnitt 4.2.5 werden die Ableitungen umsortiert. Mit $u > v$ werden in Gleichung (5.46) Gradienten von Ableitungen der unteren Dreiecksmatrix von denen der oberen Dreiecksmatrix abgezogen. Mit dieser Definition sind die Vorfaktoren f_{ijk} wie folgt zu definieren:

$$f_{ijk} = \begin{cases} \frac{s_{zj} \rho_{h_{vuk}}}{\alpha_{vu} s_{xi}} & j > i \\ -\frac{s_{zj} \rho_{h_{vuk}}}{\alpha_{vu} s_{xi}} & i < j \end{cases}. \quad (5.47)$$

Die Faktoren $\rho_{h_{vuk}}$ und α_{vu} sind jeweils mit den zugehörigen Termen (5.44) zu berechnen und korrekt zuzuordnen. Die Index-Mengen für die Berechnung des Gradienten sind $\mathbb{I} = \{1, \dots, n_s\}$ und $\mathbb{J}(i) = \mathbb{I} \setminus \{i\}$ für $i \in \mathbb{I}$.

Für die Symmetriebedingung des FNN-E-Iso-Materialmodells reicht es, die Symmetrie der Matrix $\hat{\mathbf{C}}_1$ einzufordern, mit

$$\begin{aligned} h_k^{12} &:= h^{12}(\mathbf{x}_k) = \hat{C}_{1122} - \hat{C}_{2211} \stackrel{!}{=} 0, \\ h_k^{13} &:= h^{13}(\mathbf{x}_k) = \hat{C}_{1133} - \hat{C}_{3311} \stackrel{!}{=} 0 \quad \text{und} \\ h_k^{23} &:= h^{23}(\mathbf{x}_k) = \hat{C}_{2233} - \hat{C}_{3322} \stackrel{!}{=} 0. \end{aligned} \quad (5.48)$$

Bei der Herleitung der zugehörigen Terme ist darauf zu achten, dass die jeweiligen Komponenten \hat{C}_{ijj} aus verschiedenen KNN-Berechnungen resultieren. Auf die genaue Darstellung wird an dieser Stelle verzichtet.

⁶² Die Transformationsparameter könnten auch gleichgesetzt werden, ähnlich dem Fall des ICNN.

5.3.3 Isotropie

Ist Isotropie nicht direkt in der Formulierung des KNN-Materialmodells berücksichtigt, kann sie mit einer Nebenbedingung schwach eingefordert werden. Dazu wird die Existenz einer *fiktiven* Verzerrungsenergiedichtefunktion Ψ^{KNN} angenommen, mit $\mathbf{S} = \partial\Psi^{\text{KNN}}/\partial\mathbf{E}$. Die Verzerrungen \mathbf{E} werden über $\mathbf{E}^* = \mathbf{Q}^T\mathbf{E}\mathbf{Q}$ mithilfe einer Rotationsmatrix $\mathbf{Q} \in \text{SO}(3)$ nach der EULER-RODRIGUES-Formel (3.8) passiv transformiert. Im Fall von Isotropie muss für einen beliebigen, aber festen Rotationsvektor $\boldsymbol{\omega} \in \mathbb{R}^3$ die Änderung der Verzerrungsenergiedichtefunktion infolge Rotation verschwinden. Die an der Stelle $\theta = 0$ ausgewertete Gleichung

$$\left. \frac{d\Psi^{\text{KNN}}}{d\theta} \right|_{\theta=0} = \left. \frac{\partial\Psi^{\text{KNN}}}{\partial\mathbf{E}^*} \frac{\partial\mathbf{E}^*}{\partial\theta} \right|_{\theta=0} = \mathbf{S} : \tilde{\mathbf{E}} \stackrel{!}{=} 0 \quad (5.49)$$

kann ausschließlich in den Verzerrungen und Spannungen definiert und dadurch als Nebenbedingung genutzt werden. Die fiktive Funktion Ψ^{KNN} geht dabei nicht direkt in die Gleichung ein. Sie dient lediglich zur Herleitung des Nebenbedingungsterms. Der Tensor

$$\tilde{\mathbf{E}} := \left. \frac{\partial}{\partial\theta} (\mathbf{Q}(\theta)^T \mathbf{E} \mathbf{Q}(\theta)) \right|_{\theta=0} = \boldsymbol{\Omega} \mathbf{E} - \mathbf{E} \boldsymbol{\Omega} \quad (5.50)$$

hängt von der Kreuzproduktmatrix $\boldsymbol{\Omega}$ des Rotationsvektors $\boldsymbol{\omega}$ (vgl. Definition (3.9)) und vom Verzerrungstensor \mathbf{E} ab. Unter Nutzung der Symmetrie⁶³ von $\tilde{\mathbf{E}}$ lässt sich letztlich die Nebenbedingung (5.49) mithilfe der Matrix

$$\hat{\boldsymbol{\Omega}} := \begin{bmatrix} 0 & 0 & 0 & -\omega_3 & \omega_2 & 0 \\ 0 & 0 & 0 & \omega_3 & 0 & -\omega_1 \\ 0 & 0 & 0 & 0 & -\omega_2 & \omega_1 \\ 2\omega_3 & -2\omega_3 & 0 & 0 & -\omega_1 & \omega_2 \\ -2\omega_2 & 0 & 2\omega_2 & \omega_1 & 0 & -\omega_3 \\ 0 & 2\omega_1 & -2\omega_1 & -\omega_2 & \omega_3 & 0 \end{bmatrix} \quad (5.51)$$

in reduzierter Notation formulieren zu

$$\mathbf{S}^T \hat{\boldsymbol{\Omega}} \mathbf{E} = \mathbf{S}^T \tilde{\mathbf{E}} = 0. \quad (5.52)$$

Der Vektor $\tilde{\mathbf{E}} = \hat{\boldsymbol{\Omega}} \mathbf{E} \in \mathbb{R}^6$ ist ab hier in reduzierter Notation geschrieben.

Für das Materialmodell FNN-E-Ani kann die Nebenbedingung direkt bzgl. der KNN-Ausgangsvariablen $\mathbf{z} = \mathbf{S}$ definiert werden. Gegeben sei eine beliebige Menge Rotationsvektoren $\{\boldsymbol{\omega}_m\}_{m=1}^M$. Der Fehlerterm bzgl. eines NB-Punkts \mathbf{x}_k^C kann unter Berücksichtigung der Transformation mit

$$h_k = h(\mathbf{x}_k^C, \mathbf{w}) = \frac{1}{M\alpha} \sum_{m=1}^M \sum_{j=1}^{n_s} \tilde{E}_{mj}(\mathbf{x}_k^C) (s_{zj} \hat{z}_j(\mathbf{x}_k^C, \mathbf{w}) + m_{zj}) \quad (5.53)$$

⁶³ Es gilt $\mathbf{E}^T = \mathbf{E}$ und $\boldsymbol{\Omega}^T = -\boldsymbol{\Omega}$.

berechnet werden, wobei die \tilde{E}_{mj} -Terme jeweils mit der Kreuzproduktmatrix des Rotationsvektors $\boldsymbol{\omega}_m$ zu ermitteln sind. Die Normalisierung dieses Skalarprodukts wird durch den Faktor

$$\alpha := \max_{j=1, \dots, n_s} \{s_{xj}s_{zj}\} \quad (5.54)$$

durchgeführt. Der Gradient dieses Fehlerterms lässt sich durch Vertauschen der Summen über

$$\nabla h_k = \frac{1}{M\alpha} \sum_{j=1}^{n_s} \sum_{m=1}^M \tilde{E}_{mj}(\mathbf{x}_k^C) s_{zj} \frac{\partial \hat{z}_{jk}}{\partial \mathbf{w}} \quad (5.55)$$

berechnen. Dies hat den Vorteil, dass zur Berechnung des Fehlergradienten $\nabla \mathcal{L}_C$ nach Abschnitt 4.2.4 die Summe über die Rotationsvektoren in den Faktoren

$$d_{jk} = \frac{1}{M\alpha} \sum_{m=1}^M \rho_{h_k} E_{mj}(\mathbf{x}_k^C) s_{zj} \quad (5.56)$$

zusammengefasst werden kann. Die Faktoren ρ_{h_k} sind abhängig von der Penalty-Methode, vgl. Definitionen (4.163) bis (4.165). Nach dem gleichen Prinzip könnte man auch andere materielle Symmetrieklassen einfordern.

5.4 Untersuchungen am Materialpunkt für E-KNN

Alle vorgestellten KNN-Materialmodelle FNN-E-Ani, NCNN-E-Ani, FNN-E-Iso und ICNN-E-Poly sind grundsätzlich für elastisches Materialverhalten geeignet. Sie werden einander im Folgenden anhand von diversen Untersuchungen gegenübergestellt. Dabei ist das ausgesprochene Ziel nicht, das *beste KNN-Materialmodell* zu finden, sondern signifikante Unterschiede in den einzelnen Herangehensweisen herauszuarbeiten. Für die Modelle FNN-E-Ani, NCNN-E-Ani und FNN-E-Iso wird in den Zwischenschichten der Tangens hyperbolicus als Aktivierungsfunktion gewählt. Für die Ausgangsschicht ist die Identitätsfunktion ausreichend. Beim Modell NCNN-E-Ani sind die Direktverbindungen deaktiviert. Zur Einforderung der Polykonvexitätsbedingung des Modells ICNN-E-Poly wird für alle Schichten die SoftPlus-Aktivierungsfunktion gewählt. Außerdem werden die Gewichte mit der Modifikation (4.105) versehen, vgl. Abschnitt 5.2.4. Die Direktverbindungen sind in diesem Fall aktiviert. Darüber hinaus divergiert das Training des ICNN-E-Poly-Materialmodells zu Beginn des öfteren. Daher wird dieses Modell in den ersten 10% der Epochen mit dem stabileren aber langsameren Adam-Verfahren aus Abschnitt 4.4.1 trainiert, bevor auf das QN-Verfahren nach Abschnitt 4.4.2 gewechselt wird. Die anderen Modelle werden stets mit dem QN-Verfahren trainiert. Alle Untersuchungen erfolgen in MATLAB [170].

5.4.1 Funktionsfähigkeit aller KNN-Materialmodelle

Zunächst wird das Verhalten der vier vorgestellten KNN-Materialmodelle in Bezug auf die Anzahl der Trainingsdaten P_T sowie auf die Anzahl der Gewichte n_w für moderate Verzerrungen $\mathbf{E} \in \Omega$ untersucht, mit

$$\Omega = [-0,1; 0,1]^6 \subset \mathbb{R}^6 . \quad (5.57)$$

Die synthetischen Daten für alle Berechnungen werden mit dem OGDEN-Materialmodell aus Abschnitt 3.3.3 generiert. Die zugehörigen Materialparameter sind in Tabelle 5.1 gegeben. Sie sind an TRELOAR's [172] Daten von vulkanisiertem Gummi aus dem Jahr 1944 angelehnt, vgl. auch HOLZAPFEL [81]. Der

r	μ_r [N/mm ²]	α_r [-]	Λ [N/mm ²]
1	0,63	1,3	1
2	0,0012	5,0	
3	-0,01	-2,0	

Tabelle 5.1: Materialparameter für das OGDEN-Modell mit $m = 3$ Summengliedern

Schubmodul ist $\mu = 0,4225$ N/mm², vgl. Gleichung (3.79). Im Gegensatz zur Eigenschaft der Quasi-Inkompressibilität des echten Gummimaterials, werden die synthetischen Daten mit $\Lambda = 1$ N/mm² kompressibel modelliert. Schwierigkeiten bei der Modellierung $\Lambda \gg 1$ werden in Abschnitt 5.4.3 thematisiert und sind hier zunächst ausgeklammert. Die Wahl dieses LAMÉ-Parameters Λ entspricht im Bereich infinitesimaler Verzerrungen einer Querdehnzahl $\nu \approx 0,35$ und einem E-Modul $E \approx 1,14$ N/mm².

Die Komponenten des GREEN-LAGRANGE-Verzerrungstensors $\mathbf{E} \in \mathbb{R}^6$ werden mithilfe des LHS innerhalb des rechteckigen Trainingsraums Ω aus Gleichung (5.57) generiert. Durch die moderate Größe der Verzerrungen und das definierte kompressible Materialverhalten ist das LHS in diesem Fall gültig⁶⁴ und bietet eine praktische Basis für die folgenden Untersuchungen.

Zur Validierung der Ergebnisse und zur Steuerung der Abbruchkriterien nach Abschnitt 4.4.3 wird mittels LHS eine Menge $\{(\mathbf{E}_k, \mathbf{S}_k)\}_{k=1}^{P_V}$ von zusätzlichen Validierungsdatenpaaren generiert, welche nicht im Trainingsprozess beteiligt sind. Innerhalb eines einzelnen Trainingsprozesses wird der minimale Fehler

$$\mathcal{L}_V^{\min}(\mathbf{S}) = \min_{epo} \left\{ \sqrt{\frac{1}{P_V} \sum_{k=1}^{P_V} \|\mathbf{S}(\mathbf{E}_k, \mathbf{w}_{epo}) - \mathbf{S}_k\|^2} \right\} \quad (5.58)$$

⁶⁴ Die Verzerrungskomponenten müssen stets $\det \mathbf{C} > 0$ erfüllen.

als Maß für die Generalisierungsfähigkeit des KNN-Materialmodells definiert. Durch die Formulierung in den kartesischen Spannungen \mathbf{S} sind die Ergebnisse der verschiedenen KNN-Materialmodelle miteinander vergleichbar. Dieses Fehlermaß wird über n_{MCS} Trainingsdurchläufe gemittelt zu

$$\bar{\mathcal{L}}_V^{\min}(\mathbf{S}) = \frac{1}{n_{\text{MCS}}} \sum_{i=1}^{n_{\text{MCS}}} \mathcal{L}_V^{\min}(\mathbf{S}) , \quad (5.59)$$

um der inhärenten Variabilität des KNN-Trainingsprozesses Rechnung zu tragen, vgl. Abschnitt 4.6. Die Parameter zur Berechnung dieses Validierungsfehlers und diejenigen zur Einstellung der Abbruchkriterien sind in Tabelle 5.2 angegeben. Sie werden für alle nachfolgenden Untersuchungen dieses Kapitels verwendet.

Trainingsmittel n_{MCS}	Epochenmittel P_V	Abs. Krit. δ^{abs}	Rel. Krit. δ^{rel}	Ende Training epo_{\max}
5(20*)	10^3	10^{-5}	10^{-3}	10^4

Tabelle 5.2: Statistische Parameter für die Untersuchungen der KNN-E-Materialmodelle. *In Abschnitt 5.4.2 gilt $n_{\text{MCS}} = 20$, sonst $n_{\text{MCS}} = 5$.

Für die Untersuchung in Abhängigkeit der Trainingsdaten P_T wurde stets eine Topologie mit 15 Neuronen in je zwei Zwischenschichten gewählt. Dies führt für die verschiedenen KNN-Materialmodelle zu einer unterschiedlichen Anzahl von Gewichten $n_w \in \{316, 331, 361, 441\}$. Für die Untersuchung in Abhängigkeit der Anzahl der Gewichte n_w wurden stets $P_T = 500$ Trainingsdatenpaare verwendet. Durch die ungleiche Anzahl von Eingangs- und Ausgangsvariablen unterscheiden sich die Auswertungsstellen der einzelnen KNN-Materialmodelle. Die Untersuchungsergebnisse sind in Abbildung 5.1 dargestellt.

Die anisotropen KNN-Materialmodelle benötigen wesentlich mehr Daten und Gewichte, um auf ähnliche Fehler-Niveaus zu gelangen, als diejenigen Modelle, die per Konstruktion isotropes Verhalten abbilden. Die anisotropen Modelle müssen die Eigenschaft der Isotropie erst mit erlernen und benötigen dazu mehr Information aus den Trainingsdaten. Dabei ist das NCNN-Modell stets etwas effizienter, als dasjenige mit der klassischen FNN-Architektur. Das durch die Hauptspannungs-Separation definierte FNN-E-Iso KNN-Materialmodell zeigt in nahezu allen Untersuchungsfällen in Bezug auf den Validierungsfehler bessere Ergebnisse. Das liegt daran, dass neben der verringerten Anzahl von Eingangs- und Ausgangsvariablen hinzukommt, dass aus einem Spannungs-Dehnungs-Paar bis zu drei Trainingsdatenpaare generiert werden können, vgl. Gleichung (5.27).

Das polykonvexe KNN-Materialmodell ICNN-E-Poly zeichnet sich durch eine bemerkenswerte Konstanz aus, auch im Bereich weniger Trainingsdaten und Freiheitsgrade. Tatsächlich reicht bereits eine Zwischenschicht mit vier Neuronen (29

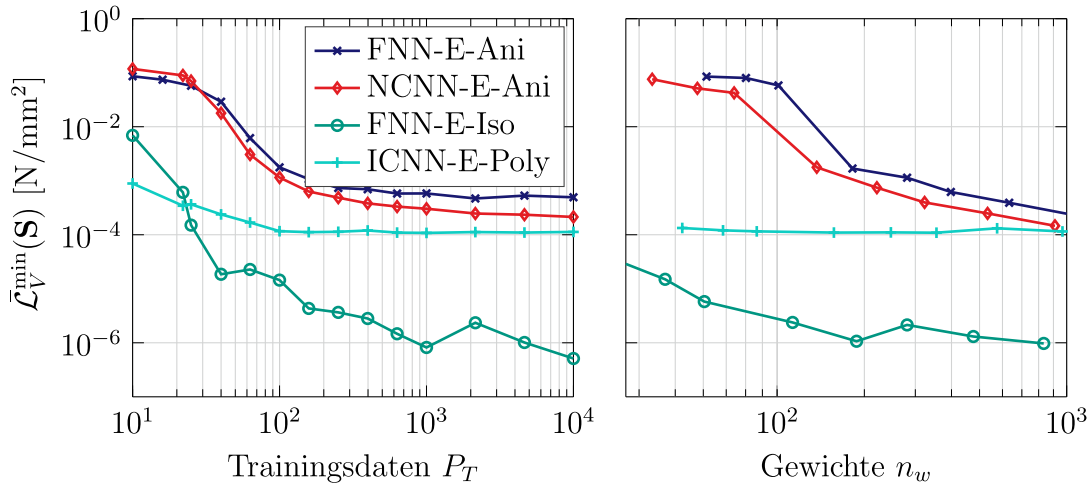


Abbildung 5.1: Verhalten der vier elastischen KNN-Materialmodelle in Abhängigkeit der Anzahl der Trainingsdaten (links, zwei Zwischenschichten à 15 Neuronen), sowie das Verhalten in Abhängigkeit der Anzahl der Gewichte (rechts, $P_T = 500$)

Gewichte) aus, um einen Fehler von $\bar{\mathcal{L}}_V^{\min}(\mathbf{S}) \approx 10^{-4}$ zu erreichen. Allerdings kommt das Trainingsverfahren des ICNN-E-Poly-Modells an dieser Stelle stets zum Stehen und verbessert sich kaum. Dies betrifft auch den Trainingsdatenfehler \mathcal{L}_T . Die Gründe dafür könnten u. a. mit der Gewichtstransformation (4.105) im Zusammenhang stehen⁶⁵. Ferner könnte es daran liegen, dass das in den Invarianten definierte polykonvexe KNN-Materialmodell das in den Hauptstreckungen definierte polykonvexe OGDEN-Modell nicht beliebig genau abbilden kann. Ähnliche Untersuchungen an einem kompressiblen Neo-HOOKE-Modell oder einem kompressiblen MOONEY-RIVLIN-Modell haben dieses Grenzwert-Verhalten nicht gezeigt. Dieser Sachverhalt sollte in Zukunft näher beleuchtet werden.

5.4.2 Untersuchungen zur Energieerhaltung

Hyperelastische Materialmodelle sind energieerhaltend. Daraus folgt u. a., dass auf einem beliebigen geschlossenen Verzerrungspfad $\mathbf{E}(t)$, $t \in [t_1, t_2]$, mit $\mathbf{E}(t_1) = \mathbf{E}(t_2)$, die geleistete (volumenbezogene) spezifische Arbeit

$$\delta\Pi = \int_{t_1}^{t_2} \mathbf{S}(\mathbf{E}(t)) : \dot{\mathbf{E}}(t) dt = \int_{\mathbf{E}(t_1)}^{\mathbf{E}(t_2)} \mathbf{S}(\mathbf{E}(t)) : d\mathbf{E} \quad (5.60)$$

null ist. Dies lässt sich auch für KNN-Materialmodelle numerisch überprüfen. Für das aktuelle Beispiel wird dazu ein geschlossener Verzerrungspfad als Polygonzug

⁶⁵ Mehrere solcher Transformationsfunktionen wurden getestet, ohne dass sich Änderungen des beschriebenen Umstandes ergeben hätten.

definiert. Die Ecken sind über die Spalten der folgenden Matrix festgelegt:

$$\begin{bmatrix} \hat{\mathbf{E}}_1 & \hat{\mathbf{E}}_2 & \hat{\mathbf{E}}_3 & \hat{\mathbf{E}}_4 & \hat{\mathbf{E}}_5 & \hat{\mathbf{E}}_6 \end{bmatrix} = \begin{bmatrix} 0 & 0 & -0,1 & 0,02 & 0,1 & 0 \\ 0 & 0,05 & -0,1 & 0 & 0,1 & 0 \\ 0 & 0,1 & -0,1 & -0,03 & 0,1 & 0 \\ 0 & 0,02 & 0 & 0,02 & -0,01 & 0 \\ 0 & 0 & -0,02 & 0,05 & 0,04 & 0 \\ 0 & 0,07 & 0 & -0,08 & 0,06 & 0 \end{bmatrix}. \quad (5.61)$$

Die Direktverbindungen zwischen den Ecken $\{\hat{\mathbf{E}}_i\}_{i=1}^6$ werden jeweils mit 25 Zwischenschritten äquidistant diskretisiert. Somit ergibt sich ein durch insgesamt 125 vorgegebene Verzerrungswerte $\{\mathbf{E}_i\}_{i=1}^{125}$ definierter geschlossener Pfad (mit $\mathbf{E}_1 = \mathbf{E}_{125}$). Der Polygonzug befindet sich vollständig im Trainingsraum nach Definition (5.57). Auf diesem Polygonzug lässt sich die Energiedifferenz (5.60) mithilfe der Trapezregel in reduzierter Notation numerisch mit

$$\delta\Pi \approx \frac{1}{2} \sum_{n=1}^{124} (\mathbf{S}(\mathbf{E}_n) + \mathbf{S}(\mathbf{E}_{n+1}))^T (\mathbf{E}_{n+1} - \mathbf{E}_n) \quad (5.62)$$

ermitteln. Durch den Fehler der numerischen Integration ist diese Energiedifferenz – auch bei Verwendung des hyperelastischen OGDEN-Materialmodells – nicht exakt null. Daher wird im Folgenden die durch das KNN-Materialmodell berechnete Residual-Energie $\delta\Pi^{\text{KNN}}$ mit

$$\delta\Pi^{\text{rel}} := \frac{|\delta\Pi^{\text{KNN}}|}{|\delta\Pi^{\text{ana}}|} \quad (5.63)$$

auf diejenige des analytischen Materialmodells $\delta\Pi^{\text{ana}}$ bezogen. Ein Wert von $\delta\Pi^{\text{rel}} \approx 1$ suggeriert somit Energieerhaltung.

Nachfolgend wird die Nebenbedingung der Energieerhaltung nach Abschnitt 5.3.2 für das FNN-E-Ani-Modell untersucht. Die Ergebnisse werden mit dem NCNN-E-Ani-Modell verglichen, welches per Konstruktion energieerhaltend ist. Die Topologie wird jeweils zu [6-20-20-6] bzw. [6-20-20-1] gewählt. Die Generierung der Trainings- und Validierungsdaten erfolgt mittels LHS im Trainingsraum (5.57) mit den Materialparametern nach Tabelle 5.1. Die NB-Punkte werden ebenfalls per LHS erzeugt. In einer Voruntersuchung wurde festgestellt, dass $P_C = 1000$ Punkte ausreichend sind für die Monte-Carlo-Integration (4.158), vgl. Abbildung 4.15. Die Nebenbedingung wird über die klassische Penalty-Methode eingefordert. Die relative Energiedifferenz $\delta\Pi^{\text{rel},\text{min}}$, die sich bei Erreichen des innerhalb eines Trainingsprozesses minimalen Validierungsfehlers $\mathcal{L}_V^{\text{min}}(\mathbf{S})$ nach Definition (5.58) berechnet, wird abgespeichert. Sie wird ebenfalls über $n_{\text{MCS}} = 20$ Trainingsdurchläufe zu $\delta\bar{\Pi}^{\text{rel},\text{min}}$ gemittelt, vgl. Gleichung (5.59). Die im Vergleich zur vorherigen Untersuchung hohe Anzahl an Trainingsdurchläufen ist nötig, da die

Energiedifferenz stärker streut als der Validierungsfehler, welcher in den Spannungen definiert ist. Die restlichen Einstellungen für die Trainingsprozesse sind Tabelle 5.2 zu entnehmen.

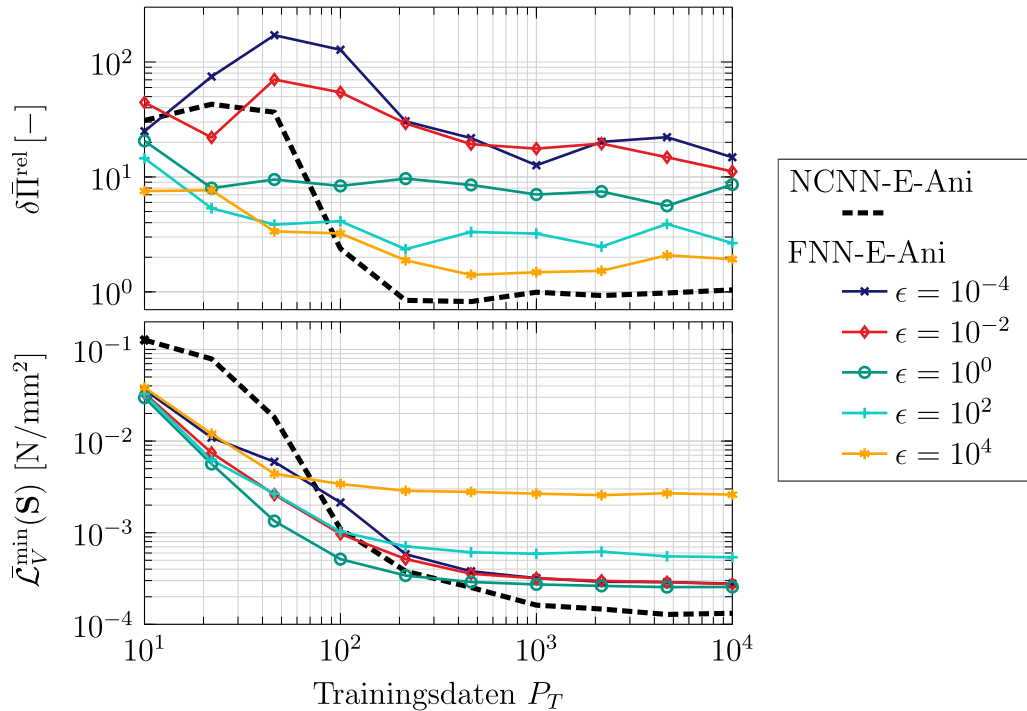


Abbildung 5.2: Gemittelte relative Energiedifferenz (oben) und gemittelter Validierungsfehler (unten) in Abhängigkeit des Penalty-Parameters ϵ und der Anzahl der Trainingsdaten P_T

In Abbildung 5.2 sind sowohl der gemittelte Validierungsfehler als auch die gemittelte Energiedifferenz über die Anzahl der Trainingsdaten und in Abhängigkeit des Penalty-Parameters ϵ aufgetragen. Das energieerhaltende NCNN erreicht ab ca. $P_T = 200$ Trainingsdaten den zu erwartenden Wert $\delta\bar{\Pi}^{\text{rel},\text{min}} \approx 1$. Die Abweichungen für $P_T < 200$ ist dem noch nicht ausreichend trainierten KNN zuzuordnen, da dessen Spannungspfad sich zu stark von demjenigen des analytischen Modells unterscheidet. Für das Modell FNN-E-Ani verbessert sich die Energiedifferenz stetig mit höherem Penalty-Parameter. Für $\epsilon = 10^4$ ist das Ergebnis ähnlich gut wie beim NCNN-Modell. Allerdings verschlechtert sich in diesem Fall wieder der Validierungsfehler erheblich. Dieser Umstand wurde schon in Abschnitt 4.6 angesprochen. Er wäre bspw. durch mehr Gewichte behebbar. Bemerkenswert ist allerdings, dass es einen großen Bereich für den Penalty-Parameter $\epsilon \in [0, 10^2]$ gibt, bei dem sich der Validierungsfehler nicht ändert. In diesem Fall wird die Generalisierungsfähigkeit nicht verschlechtert. Darüber hinaus ist im Bereich kleiner

Mengen von Trainingsdaten $P_T < 100$ eine wesentliche Verbesserung der Ergebnisse des FNN-Modells mit Nebenbedingungen gegenüber dem NCNN-Modell zu erkennen. Dies lässt an dieser Stelle den Schluss zu, dass schwaches Einfordern von Nebenbedingungen durchaus Vorteile haben kann gegenüber Modellen, die diese Eigenschaften per Konstruktion erfüllen. Da die Approximation des Spannungspfades mit Nebenbedingungen bei wenigen Daten besser ist, was am Validierungsfehler zu erkennen ist, ist in dieser Region auch augenscheinlich der Fehler der Energiedifferenz besser. Dies liegt aber nicht an einer vermeintlich besseren Einhaltung der Energieerhaltung, sondern an der höheren Übereinstimmung der Spannungspfade. Zuletzt sei erwähnt, dass ein Vergleich der beiden Architekturen nicht unbedingt objektiv ist, da einmal die Spannungen direkt und einmal die Verzerrungsenergiedichtefunktion approximiert wird. Daher sind die Ergebnisse vor allem qualitativ zu interpretieren.

5.4.3 KNN-Materialmodelle bei Quasi-Inkompressibilität

Die meisten Festkörper leisten gegen eine Volumenänderung wesentlich mehr Widerstand als gegen eine isochore Deformation. In Abschnitt 3.3.4 wurde die Modellierung dieses Verhaltens im Rahmen klassischer Materialmodellierung beschrieben. Die synthetischen Daten für die vorangegangenen zwei Untersuchungen der Abschnitte 5.4.1 und 5.4.2 wurden mit $\Lambda = 1 \text{ N/mm}^2$ vergleichsweise kompressibel⁶⁶ modelliert, was nur dem Zweck der Parameterstudien diene. Insbesondere bei der KNN-Materialmodellierung stellt Quasi-Inkompressibilität eine gesonderte Herausforderung dar, die in diesem Abschnitt untersucht und erläutert wird.

Der KNN-Trainingsraum Ω aus Definition (5.57) beschränkt sich weiterhin auf moderate Verzerrungen mit $\mathbf{E} \in [-0,1; 0,1]^6$. Die Datenpunkte werden allerdings nicht mehr mittels LHS generiert. Mit wachsendem Λ nimmt der Wertebereich der Spannungen linear zu, und die Bereiche großer volumetrischer Dehnungen werden energetisch immer ungünstiger. Im Folgenden wird daher der Materialpunkt-Versuch aus Abschnitt 5.1.1 zur Generierung aller Trainings- und Validierungsdatenpunkte genutzt.

Jeder Pfad startet bei $(\mathbf{E}_0, \mathbf{S}_0) = (\mathbf{0}, \mathbf{0})$. Grundsätzlich sind stets alle zwölf uniaxialen und alle 60 biaxialen Kombinationen zu den Ecken des Hyperrechtecks Ω beinhaltet. Dabei wird im uniaxialen Fall ein linearer Pfad zu $E_{\text{end}}^b = \pm 0,1$ bzw. im biaxialen Fall zu $\mathbf{E}_{\text{end}}^b = [E^{b1}, E^{b2}] = [\pm 0,1, \pm 0,1]$ mit jeweils zehn Schrit-

⁶⁶ OGDEN [139] gibt für den Parameter Λ Werte von 2600 N/mm^2 und 3600 N/mm^2 für zwei verschiedene Elastomere an, allerdings in Verbindung mit einer anderen Funktion $\Psi_{\text{vol}}(J)$. Die Größenordnung $\Lambda/\mu > 1000$ kann jedoch als grober Anhaltswert für reale Elastomere herangezogen werden.

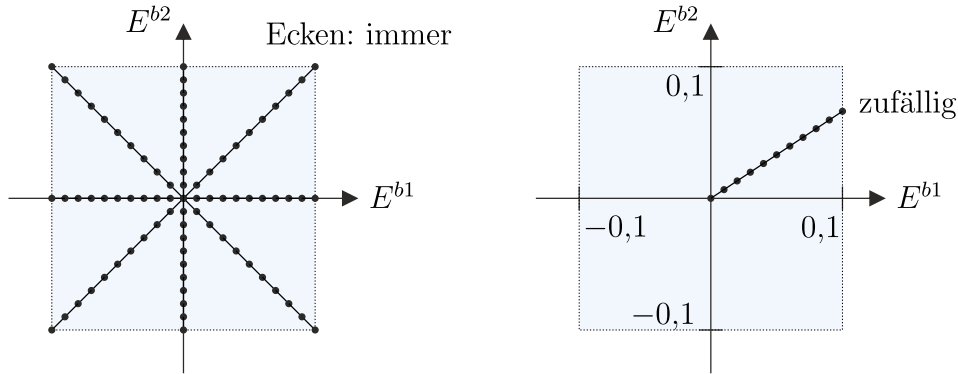


Abbildung 5.3: Zufällige biaxiale Pfadgenerierung: die Ecken des Hyperrechtecks Ω sind immer inkludiert, andere Verzerrungsverhältnisse werden zufällig generiert.

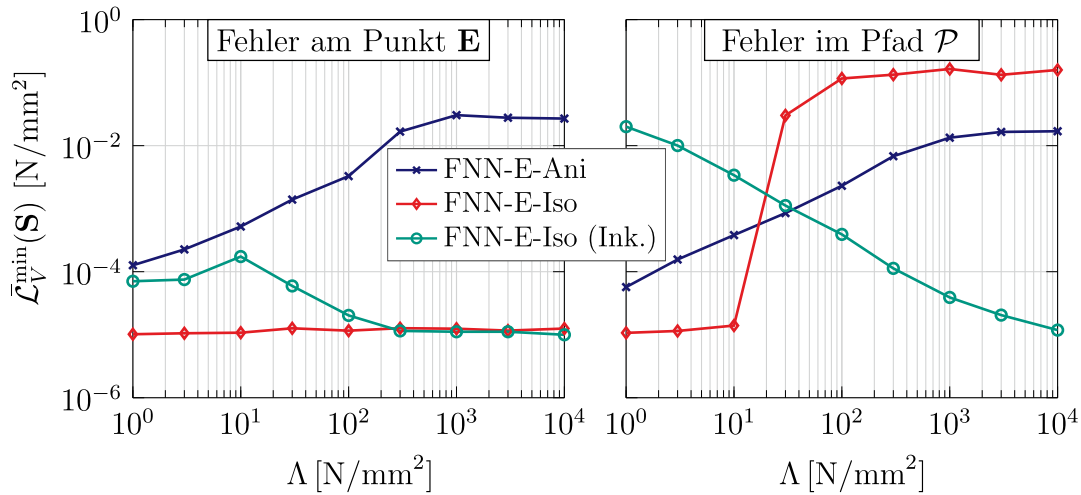


Abbildung 5.4: Gemittelte Fehler einzelner Datenpunkte (links) und der Fehler bzgl. ganzer, gemischt gesteuerter Materialpunkt-Versuche (rechts) in Abhängigkeit des LAMÉ-Parameters Λ

ten diskretisiert, vgl. Abbildung 5.3 links. Darüber hinaus werden 50 zusätzliche, mit zufälligen Verzerrungsverhältnissen generierte biaxiale⁶⁷ Pfade erzeugt, die stets bis zum Rand des Trainingsraums reichen, vgl. Abbildung 5.3 rechts. Auch sie werden jeweils mit zehn Schritten diskretisiert. Somit folgen für jeden Pfad die vorgegebenen Verzerrungen $\{\mathbf{E}_n^b\}_{n=1}^{10}$. Die Versuche werden jeweils rein verzerungsgesteuert gerechnet, daher gilt $\tilde{\mathbf{S}}_n^u \equiv \mathbf{0}$. Die so generierten Pfade (5.7) dienen als Grundlage zur Berechnung der Trainings- und Validierungsdatenpaare.

In Abbildung 5.4 ist der jeweils über fünf Trainingsdurchläufe gemittelte Fehler (5.59) in Abhängigkeit der LAMÉ-Parameters Λ dargestellt. Dabei werden die Mo-

⁶⁷ Durch die Beschränkung auf biaxiale Versuche wird garantiert, dass sich die restlichen, nicht vorgegebenen Verzerrungskomponenten z. B. näherungsweise isochor einstellen können.

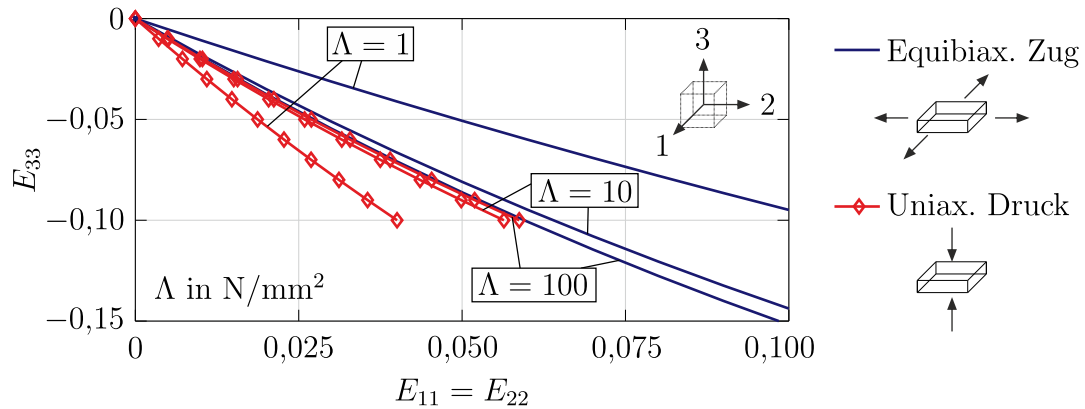


Abbildung 5.5: Beispiel für die schwindende Unterscheidbarkeit verschiedener Spannungszustände bei quasi-inkompressiblen Werkstoffen

delle FNN-E-Ani und FNN-E-Iso miteinander verglichen. Letzteres ist einmal mit Modellierung der exakten Inkompressibilitätsbedingung (Ink.) nach Abschnitt 5.2.5 modelliert und einmal ohne. Im linken Bild wurde der Fehler mit einzelnen voneinander unabhängigen Verzerrungsdatenpaaren $\{(\mathbf{E}_k, \mathbf{S}_k)\}_{k=1}^{P_V}$ berechnet, die dem KNN nicht während des Trainingsprozesses zur Verfügung standen. Sie wurden zufällig aus der Menge der Trainingsdatenpaare gewählt. Im rechten Bild wurde der Fehler an 25 neuen, vollständigen Pfaden $\mathcal{P} = \{(\mathbf{E}_0, \mathbf{S}_0), (\mathbf{E}_1, \mathbf{S}_1) \dots\}$ berechnet, welche mit dem KNN-Materialmodell mittels biaxialem Materialpunkt-Versuch nach Abschnitt 5.1.1 durchgeführt wurden. Der entscheidende Unterschied in beiden Fehlerberechnungen liegt darin, dass im Fall der Pfade nur zwei Verzerrungskomponenten vorgegeben und die restlichen aus der NEWTON-Iteration berechnet werden. Dadurch kann sich der vollständige Pfad stärker von dem ursprünglich vorgegebenen Pfad unterscheiden, gerade im Hinblick auf Volumenänderungen. Diese unterschiedlichen Fehlerberechnungen beleuchten zwei völlig unterschiedliche Herausforderungen, die bei der Modellierung von Quasi-Inkompressibilität mit KNN in Erscheinung treten und in Abbildung 5.4 zu erkennen sind. Sie werden nachfolgend beschrieben.

Schwindende Unterscheidbarkeit verschiedener Spannungszustände

Im ideal-inkompressiblen Grenzfall kann aus einem gegebenen Verzerrungszustand nicht eindeutig auf einen Spannungszustand geschlossen werden, da durch den hydrostatischen Druck keine Verzerrungen resultieren, vgl. Abschnitt 3.3.4. Im Fall kompressibler Materialien erlauben die volumetrischen Dehnungen eine Unterscheidung der einzelnen Spannungszustände. Je näher das Materialverhalten am ideal-inkompressiblen Grenzfall liegt, desto geringer ist der Anteil der volumetrischen Dehnungen und desto näher liegen einzelne Verzerrungszustände

beieinander, obwohl sie zu völlig unterschiedlichen Spannungszuständen gehören. Diese Problematik ist in Abbildung 5.5 am Vergleich von equibiaxialem Zug und uniaxialem Druck dargestellt. Die Daten wurden bei steigendem Λ [N/mm²] mit dem OGDEN-Materialgesetz erzeugt.

Λ [N/mm ²]	Intervall λ_a^2	Intervall $II_{\mathbf{C}}$	Intervall J^2
1	[0,769, 1,248]	[2,637, 3,385]	[0,799, 1,167]
10	[0,712, 1,501]	[2,990, 3,150]	[0,961, 1,026]
100	[0,696, 1,556]	[3,000, 3,129]	[0,996, 1,003]

Tabelle 5.3: Intervalle der Eingangsvariablen λ_a^2 , $II_{\mathbf{C}}$ und J^2 in Abhängigkeit des LAMÉ-Parameters Λ

Die immer schlechtere Unterscheidbarkeit der Spannungszustände macht es für KNN besonders schwer, den zugrundeliegenden funktionalen Zusammenhang zu erlernen. Dies ist in Abbildung 5.4 am FNN-E-Ani-Modell in beiden Diagrammen zu erkennen. Bezüglich des Datenpunkt-Fehlers (links) zeigt das FNN-E-Iso-Modell allerdings keine Probleme, obwohl der Umstand der zusammenfallenden Verzerrungszustände dort ebenfalls vorliegt. Das liegt daran, dass der Trainingsraum Ω für das Invarianten-Modell mit wachsendem Λ kleiner wird. Dies ist in Tabelle 5.3 gezeigt. Im normalisierten Trainingsraum (vgl. Abschnitt 4.1) muss das FNN stets die gleiche Antwortfläche erlernen. Daher ändert sich der Fehler nicht merklich. Wie in Abbildung 5.5 zu erkennen ist, folgt der Trainingsraum des FNN-E-Ani-Modells nicht diesem Verhalten.

Verlassen des Trainingsbereichs bei Verzerrungssteuerung

Bemerkenswert ist das schlechte Abschneiden des FNN-E-Iso-Modells bzgl. des über vollständige Pfade gemittelten Fehlers, vgl. Abbildung 5.4 rechts. Das liegt an der Verzerrungssteuerung. Im ersten NEWTON-Schritt werden die vorgegebenen Verzerrungskomponenten \mathbf{E}^b festgelegt, während die restlichen zu null initialisiert sind. Dadurch erfüllt der gesamte Verzerrungstensor \mathbf{E} im ersten Iterationsschritt keine Anforderungen an Quasi-Inkompressibilität. Wie zuvor beschrieben, wird der Trainingsraum des FNN-E-Iso-Modells mit wachsendem LAMÉ-Parameter Λ kleiner. Ein solcher, im ersten Iterationsschritt vorgegebener Verzerrungstensor liegt mit seinen Komponenten folglich weit außerhalb des Trainingsraums. Daher divergiert das gesamte Materialpunkt-Verfahren meist schon zu Beginn, was den Verlauf des Fehlers erklärt. Das FNN-E-Ani-Modell hat mit der Verzerrungssteuerung weniger Schwierigkeiten als mit den unabhängigen Datenpunkten, da – wie oben bereits beschrieben – der Trainingsraum nicht kleiner wird.

Unter Berücksichtigung der exakten Inkompressibilitätsbedingung nach Abschnitt 5.2.5 in Verbindung mit dem inkompressiblen Materialpunkt-Versuch aus Anhang D zeigt das FNN-E-Iso-Modell hervorragende Ergebnisse für größere Werte von Λ . Grundsätzlich lässt sich daraus schließen, dass es für nahezu inkompressibles Materialverhalten numerisch sinnvoll ist, dies im KNN-Materialmodell zu berücksichtigen. An dieser Stelle sollte eine gewisse Ironie nicht unerwähnt bleiben: Die Lösung der idealen Inkompressibilität steht in starkem Widerspruch zum Ziel der von sämtlichen Modellen befreiten, rein auf Daten basierenden KNN-Materialmodellierung. Alternativ könnte dieses Problem für das FNN-E-Iso-Modell mit einer Spannungssteuerung⁶⁸ gelöst werden. Ebenfalls zu erwähnen ist hier, dass die Wachstumsbedingung nach Abschnitt 5.2.6 zu keiner Verbesserung der beschriebenen Probleme führt, da erst für sehr große bzw. sehr kleine J die Anteile maßgebend werden⁶⁹.

5.4.4 Verhalten bei großen Verzerrungen

Verzerrungen mit $|E_{ij}| \leq 0,1$, $i, j = 1, 2, 3$ sind als moderat anzusehen. Unter uniaxialem Zug mit $\lambda_{\max} = \sqrt{2E_{ii,\max} + 1}$ war bisher eine maximale Längenänderung von ca 9,5% möglich. Die elastischen KNN-Materialmodelle werden nachfolgend für größere Verzerrungen miteinander verglichen. Der größere Trainingsraum

$$\Omega = [-0.25, 1.5]^3 \times [-1.0, 1.0]^3 \in \mathbb{R}^6 \quad (5.64)$$

beinhaltet bspw. bei uniaxialem Zug mit $E_{ii} \in [-0.25, 1.5]$ Längenänderungen bis 100%. Je größer die inbegriffenen Deformationen sind, desto komplizierter wird der funktionale Zusammenhang, den das KNN erlernen muss.

Um die Effekte großer Verzerrungen von denen der Inkompressibilität zu isolieren, werden Parameterstudien für $\Lambda = 1 \text{ N/mm}^2$ durchgeführt. Die (\mathbf{E}, \mathbf{S}) -Pfade zur Generierung der Trainingsdatenpaare werden wie in Abschnitt 5.4.3 mit den 72 festen Materialpunkt-Versuchen für die Ecken und n_p zufälligen, zusätzlichen biaxialen Materialpunkt-Versuchen generiert. Darüber hinaus werden 25 solcher zufälliger Pfade zur Berechnung des Validierungsfehlers (5.59) herangezogen. In Abbildung 5.6 sind für die KNN-Materialmodelle FNN-E-Ani, NCNN-E-Ani und FNN-E-Iso die Untersuchungen bzgl. der Anzahl der zufälligen Pfade n_p sowie bzgl. der Anzahl der Gewichte n_w dargestellt. Dabei wurde für die Untersuchung der Anzahl der Trainingsdaten die Topologie stets mit drei Zwischenschichten à 15 Neuronen gewählt. Für die Topologieuntersuchung wurden jeweils $n_p = 100$ zufällige Pfade generiert.

⁶⁸ Im Hinblick auf eine FE-Implementierung ist das allerdings keine Lösung.

⁶⁹ Theoretisch könnte ein Vorfaktor eingeführt werden, der mit dem Kompressionsmodul korreliert. Dieser müsste dazu im Trainingsprozess angepasst werden.

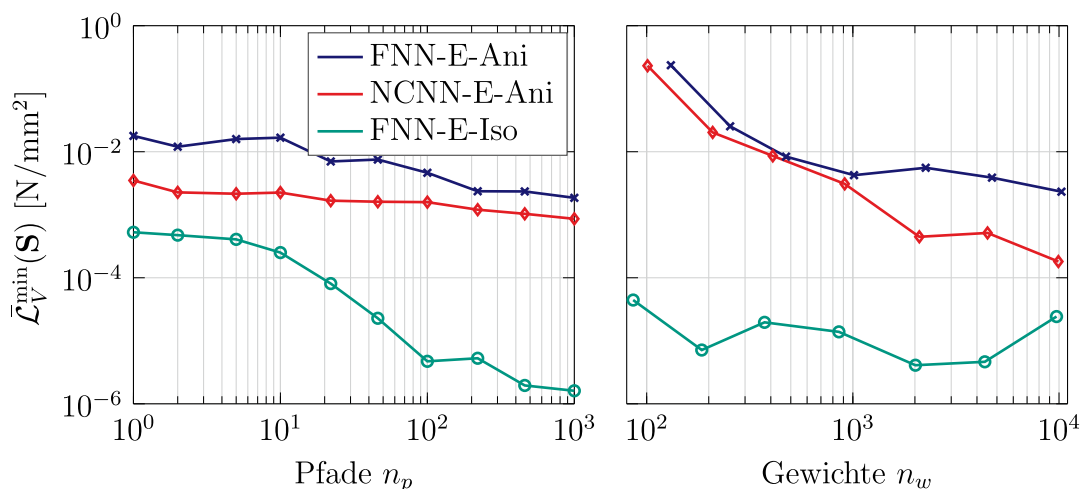


Abbildung 5.6: Verhalten von drei elastischen KNN-Materialmodellen für große Verzerrungen in Abhängigkeit der Anzahl der Trainingsdaten (links, drei Zwischenschichten à 15 Neuronen) sowie das Verhalten in Abhängigkeit der Anzahl der Gewichte (rechts, $n_p = 100$)

Auffallend gegenüber den Untersuchungen mit moderaten Verzerrungen aus Abschnitt 5.4.1 sind das grundsätzlich etwas höhere Fehlerniveau und die tendenziell spätere Konvergenz bzgl. Daten und Freiheitsgraden. Das ICNN-E-Poly-Modell wurde hier nicht aufgeführt, da das schon oben beschriebene Konvergenzverhalten für große Verzerrungen stärker ausgeprägt ist, und zwar bereits bei einem Fehler von $\bar{\mathcal{L}}_V^{\min}(\mathbf{S}) \approx 10^{-2}$.

5.5 Anwendung der KNN-E-Materialmodelle in der FEM

Die Festkörpersimulation mittels FEM erfordert ein robustes Zusammenspiel aus Elementformulierung, Materialmodell und Lösungsverfahren. KNN-Materialmodelle, welche an einzelnen Datenpunkten oder in Materialpunkt-Versuchen brauchbare Ergebnisse liefern, können in der Anwendung am Gesamtsystem durchaus scheitern. Daher werden die vorgestellten KNN-Materialmodelle nachfolgend in der numerischen Struktursimulation eingesetzt und miteinander verglichen.

Im Rahmen des inkrementell-iterativen Lösungsverfahrens nach Abschnitt 2.2.3 sind innerhalb eines Elements die Knoten-Verschiebungsgrößen $\{\mathbf{v}_I^{(i)}\}_{I=1}^{nel}$ im Iterationsschritt i fest. Mit diesen und den zugehörigen Ansatzfunktionen werden die Verzerrungen $\mathbf{E}_l = \mathbf{E}(\boldsymbol{\xi}_l)$ für jeden Integrationspunkt ermittelt, mit den Koordinaten $\boldsymbol{\xi}_l, l = 1, \dots, n_g$. Für ein Volumenelement geschieht dies bspw. mithilfe der Verschiebungsgradienten (2.72) nach Gleichung (2.18) bzw. (2.19). Abhängig vom gewählten KNN-Materialmodell werden die Verzerrungen \mathbf{E}_l zunächst in die

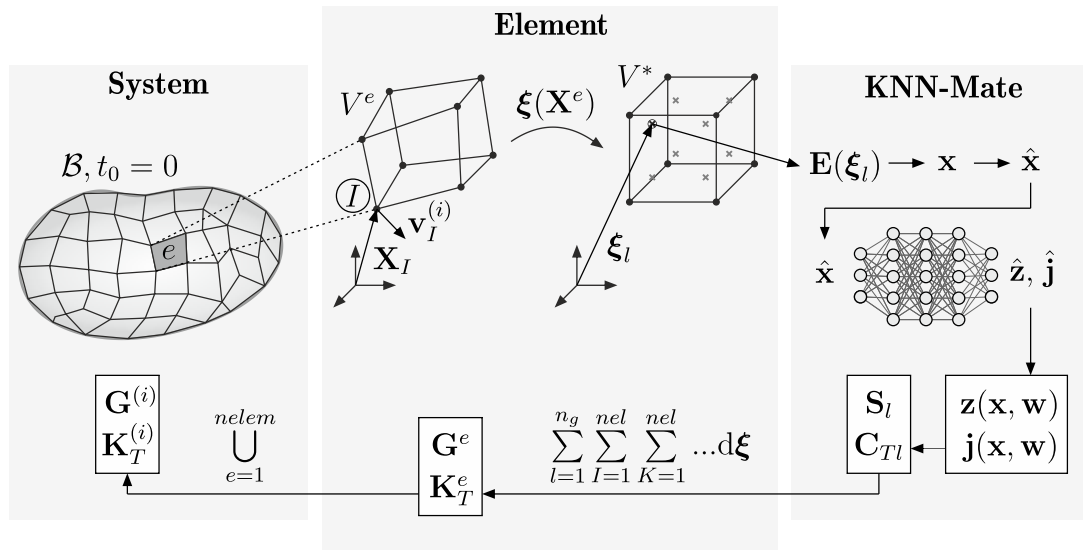


Abbildung 5.7: Schematische Darstellung der Informationsübertragung innerhalb eines Iterationsschritts i der FEM

KNN-EingangsvARIABLEN \mathbf{x} umgeformt und zu $\hat{\mathbf{x}}$ transformiert. Nach Ausführung der KNN-Algorithmen zur Berechnung der transformierten AusgangsvARIABLEN $\hat{\mathbf{z}}$ und der JACOBI-Matrix $\hat{\mathbf{j}}$ werden diese zu \mathbf{z} und \mathbf{j} rücktransformiert. Anschließend folgt die Berechnung der Integrationspunkt-Spannungen \mathbf{S}_l sowie der zugehörigen Materialtangente \mathbf{C}_{Tl} nach den jeweils in Abschnitt 5.2 beschriebenen Gleichungen. Innerhalb des Elements werden damit die Elementsteifigkeitsmatrix \mathbf{K}_T^e und der Elementresidualvektor \mathbf{G}^e berechnet, welche im Anschluss zu den Systemgrößen $\mathbf{K}_T^{(i)}$ und $\mathbf{G}^{(i)}$ assembliert werden. Dieser Zusammenhang ist in Abbildung 5.7 schematisch dargestellt.

Das KNN ist als Materialmodell in die 1D-, 2D- und 3D-Materialbibliothek einer erweiterten Version des FE-Programms FEAP (*Finite Element Analysis Programm*) [168] implementiert worden, welches am Institut für Baustatik (KIT) verwendet wird. Dabei stehen das FNN, das ICNN und das NCNN als KNN-Architekturen zur Verfügung und können flexibel mit den in diesem Kapitel vorgestellten KNN-Materialmodellen kombiniert werden. Das Training der KNN erfolgt in MATLAB [170]. Über ein Transferdatei werden alle notwendigen Informationen wie die Topologie, die Gewichte, das KNN-Materialmodell usw. an FEAP übergeben.

5.5.1 Scheibe mit Loch: Systembeschreibung und Referenzlösung

Als numerisches Beispiel dient das Auseinanderziehen einer quadratischen Scheibe, das bspw. von PARISCH [142] oder BETSCH ET AL. [15] zur Verifizierung von

Schalenelementformulierungen untersucht wurde. Die Scheibe hat in dieser Arbeit die Seitenlänge $L = 200$ mm und die Dicke $d = 1$ mm. In ihrer Mitte befindet sich ein Loch mit Durchmesser $2R = 60$ mm. Durch die vorhandene Doppelsymmetrie wird das FE-Modell an einem Viertel der Scheibe aufgestellt. Das Modell ist in Abbildung 5.8 dargestellt, inklusive der Symmetrie- und Systemrandbedingungen. Das FE-Netz besteht aus $16 \times 16 = 256$ gemischten Schalenelementen von KLINKEL ET AL. [106], welche in Abschnitt 2.2.4 beschrieben wurden. Der Fehler in der Reaktionskraft ist gegenüber einem FE-Netz mit sehr vielen Elementen weniger als 3%. Dabei wird eine einzige numerische Schicht über die Höhe berücksichtigt. Die EAS-Parameter werden zu $\beta_m = \beta_b = \beta_z = 2$ gewählt, da lineare Dickenverzerrungen für das vorliegende Beispiel ausreichend sind, vgl. KLINKEL ET AL. [106] und die Definition der Parameter in Abschnitt 2.2.4. Die Belastung erfolgt durch Vorgabe der Verschiebung \bar{u} an der gesamten rechten Kante. Diese kann sich dadurch nicht verdrehen. Die Kraft $F(\bar{u})$ wird durch die Summe der Reaktionskräfte an jener Kante zurückgerechnet.

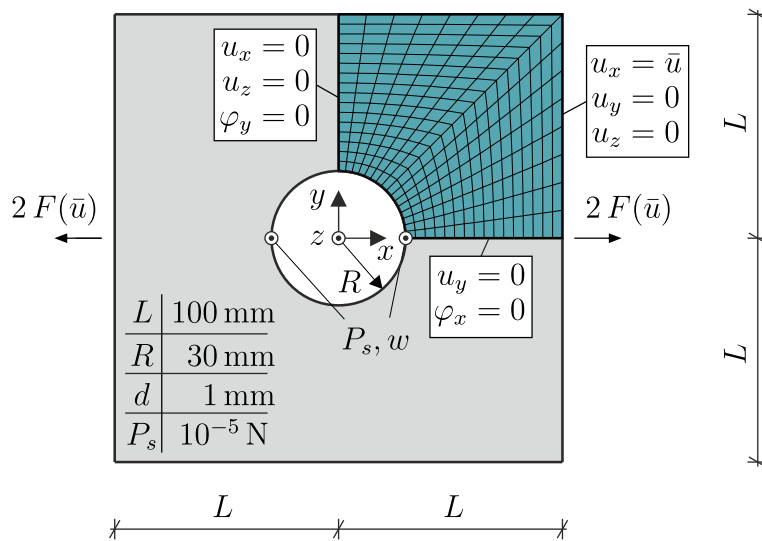


Abbildung 5.8: Scheibe mit Loch: Referenzkonfiguration, FE-Modell am Viertelsystem inklusive Randbedingungen und Definition der Reaktionskraft $F(\bar{u})$ in Abhängigkeit der vorgegebenen Verschiebung \bar{u}

Durch die Belastung entstehen am Lochrand senkrecht zur Belastungsrichtung Druckspannungen, welche schon bei vergleichsweise geringer Last zu einem Stabilitätsproblem führen. Neben dem rein ebenen Zustand existieren somit energetisch äquivalente Nachbarzustände, die Verschiebungen senkrecht zur Ebene enthalten. Um dem zugehörigen sekundären Gleichgewichtspfad zu folgen, wird eine Störlast $P_s = 10^{-5}$ an dieser Stelle aufgebracht, welche ebenfalls Abbildung 5.8 gekennzeichnet ist. Alternativ wäre ein direktes Wechseln in den Sekundärpfad

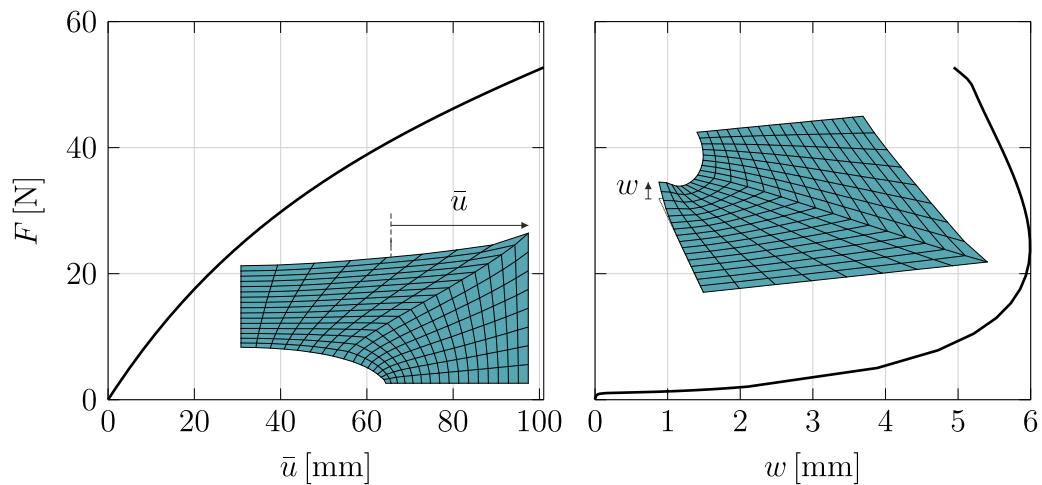


Abbildung 5.9: Scheibe mit Loch: Referenzlösung mit dem OGDEN-Materialgesetz

ohne Störlast möglich, vgl. WAGNER und WRIGGERS [187]. Als Referenzlösung wird das OGDEN-Materialmodell aus Abschnitt 3.3.3 mit den Materialparametern aus Tabelle 5.1 genutzt. Der 2. LAMÉ-Parameter wird zur Modellierung geringer Kompressibilität zu $\Lambda = 10 \text{ N/mm}^2$ gewählt. Die Scheibe wird durch die beschriebene Verschiebungssteuerung auf ca. das Doppelte ihrer Ausgangslänge gedehnt. Dabei werden zunächst 20 Schritte mit $\Delta u = 0,2 \text{ mm}$ zu $\bar{u} = 2 \text{ mm}$ ausgeführt, um den Stabilitätspunkt zu überschreiten. Im Anschluss folgen 33 Schritte mit $\Delta u = 3 \text{ mm}$ zur Endverschiebung $\bar{u} = 101 \text{ mm}$. In Abbildung 5.9 ist die Reaktionskraft in Abhängigkeit der vorgegebenen Längsverschiebung \bar{u} (links) sowie in Abhängigkeit der Verschiebung w senkrecht zur Systemebene (rechts) dargestellt.

5.5.2 Vergleich der vier KNN-Materialmodelle an der Lochscheibe

Die vier in Abschnitt 5.2 vorgestellten elastischen KNN-Materialmodelle werden zur Berechnung des Lochscheiben-Beispiels herangezogen. Der Fokus der folgenden Untersuchung liegt in der prinzipiellen Anwendung der KNN-Materialmodelle innerhalb der FEM. Daher werden mit drei Zwischenschichten à 20 Neuronen jeweils vergleichsweise viele Gewichte zum Training zur Verfügung gestellt. Darüber hinaus werden im Trainingsraum (5.64) die 72 festen, biaxialen Materialpunkt-Versuche und zusätzlich $n_p = 1000$ weitere zufällige biaxiale Materialpunkt-Versuche durchgeführt, vgl. Abschnitt 5.4.4. Von den resultierenden (\mathbf{E}, \mathbf{S}) -Datenpaaren werden 80 % für das Training und 20 % zur Validierung und zum Abbruch des Trainingsprozesses genutzt. Im Folgenden werden keine Nebenbedingungen berücksichtigt. Wie zu Beginn des Kapitels motiviert, werden hier hauptsächlich verschiedene KNN-Materialmodelle miteinander verglichen. In Kapitel 6 liegt der Fokus auf den Nebenbedingungen. Für jedes der vier KNN-Materialmodel-

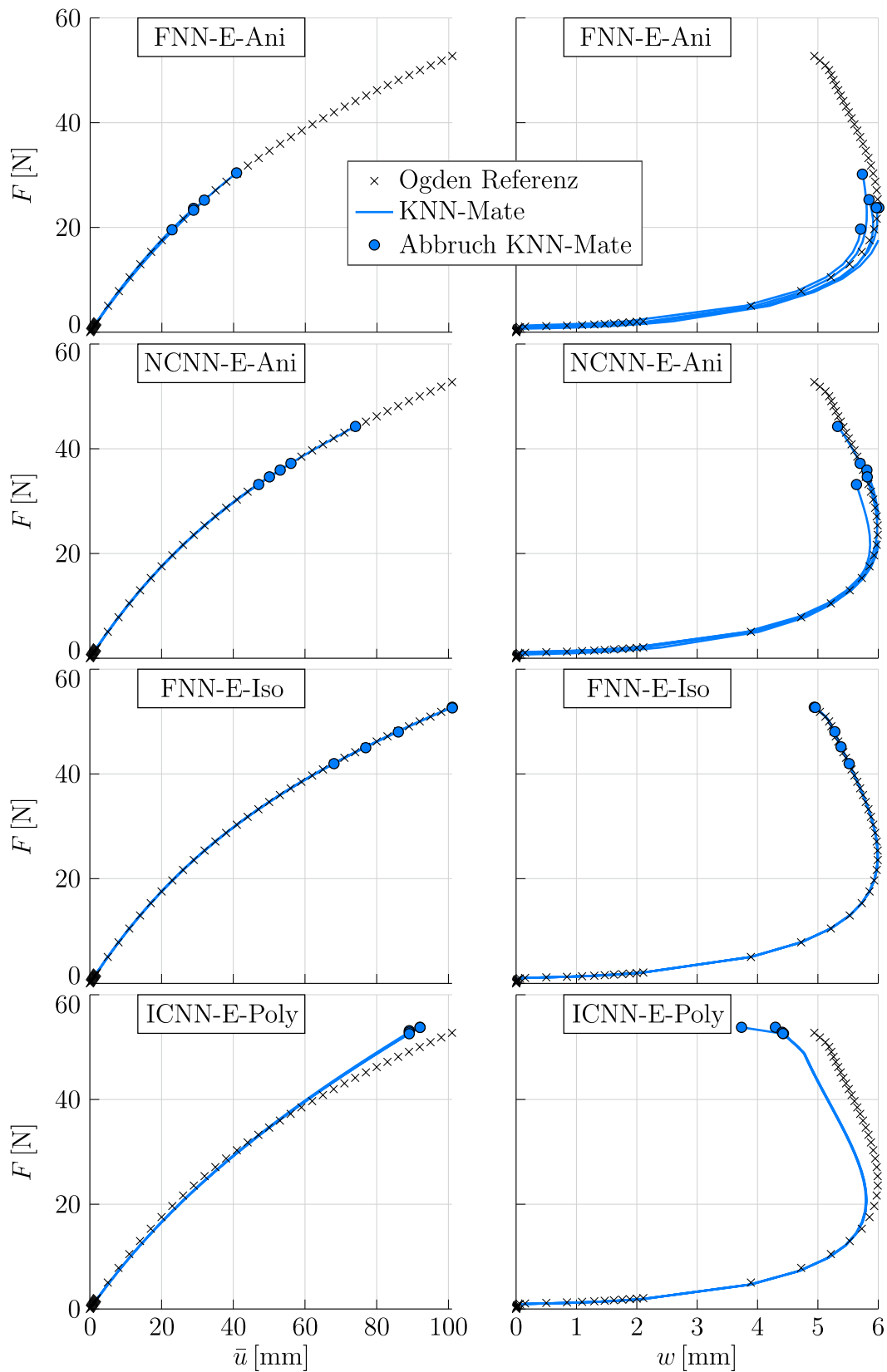


Abbildung 5.10: Scheibe mit Loch: Ergebnisse von jeweils fünf FE-Berechnungen mit den KNN-Materialmodellen im Vergleich zur Referenzlösung

le FNN-E-Ani, NCNN-E-Ani, FNN-E-Iso und ICNN-E-Poly werden jeweils fünf Trainingsdurchläufe bei gleichbleibenden Einstellungen durchgeführt.

Im Rahmen aller Beispiele dieser Arbeit wird die Materialtangente innerhalb von FEAP [168] symmetrisiert:

$$\mathbf{C}_T^{\text{sym}} = \frac{1}{2}(\mathbf{C}_T + \mathbf{C}_T^T). \quad (5.65)$$

Da alle KNN unter Hinzunahme vieler Daten ein hyperelastisches Materialmodell approximieren sollen, greift diese Maßnahme nicht zu stark in die Berechnung ein, erlaubt jedoch die Nutzung von Gleichungslösern für symmetrische Matrizen.

In Abbildung 5.10 sind die Ergebnisse der insgesamt 20 FE-Berechnungen abgebildet. Der Abbruch der globalen Gleichgewichts-Iteration ist jeweils mit einem Kreis gekennzeichnet. Zunächst sei noch einmal erwähnt, dass alle KNN-Materialmodelle mit biaxialen Versuchen trainiert wurden. Innerhalb des vorliegenden Beispiels liegen jedoch komplexe dreidimensionale Spannungszustände vor, gerade im Bereich von Verschiebungen senkrecht zur Referenzebene. Da die Eingangsvariablen in den Verzerrungen definiert sind, müssen die KNN zwar nicht extrapolieren, doch deren Interpolationsfähigkeit wird damit auf die Probe gestellt.

Die inhärente Variabilität, die dem KNN-Trainingsprozess zugrunde liegt, spiegelt sich auch in den Ergebnissen der FE-Berechnungen wider, vor allem bei den allgemein anisotropen Modellen. Das mit NCNN formulierte Materialmodell ist demjenigen des FNN dabei stets überlegen im Bezug auf die Stabilität im Rahmen der äußeren Gleichgewichtsiteration. Das polykonvexe ICNN-E-Poly-Modell sticht durch eine beachtliche Konstanz der Ergebnisse heraus. Die FE-Berechnungen brechen immer erst innerhalb der letzten beiden Lastschritte ab. Allerdings kann an der Diskrepanz zur Vergleichslösung das in Abschnitt 5.4.1 beschriebene Verhalten erkannt werden: Die in den Invarianten definierte Polykonvexitätsbedingung scheint den Approximationsraum des KNN zu stark einzuschränken. Demgegenüber zeigt das isotrope FNN-E-Iso-Modell nahezu deckungsgleiche Ergebnisse in allen Berechnungen. In diesem Fall bricht jedoch die Berechnung früher ab.

5.5.3 Abschließende Bemerkungen

- Da keines der Modelle die Bedingung einer spannungsfreien Referenzkonfiguration $\mathbf{S}(\mathbf{0}) = \mathbf{0}$ erfüllt, befindet sich das gesamte Tragwerk zu Beginn der FE-Berechnung nicht im Gleichgewicht. Daher empfiehlt es sich, vor dem ersten Lastschritt den aus diesem Umstand resultierenden Eigenspannungszustand zu berechnen, bevor mit den eigentlichen Lastschritten begonnen

wird. In Abbildung 5.11 ist ein Ausschnitt der fünf FNN-E-Ani-Berechnungen abgebildet, der diesen Sachverhalt verdeutlicht.

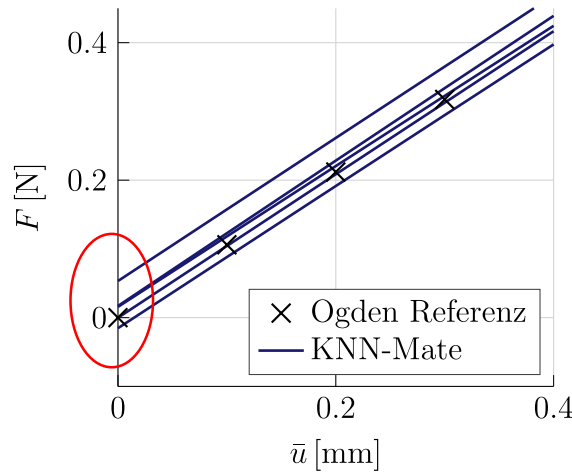


Abbildung 5.11: Eigenspannungszustand durch nicht exaktes Erfüllen der spannungsfreien Referenzkonfiguration

- Für größere LAMÉ-Parameter $\Lambda \gg 10 \text{ N/mm}^2$ konvergiert das FNN-E-Iso-Modell nicht mehr innerhalb der FEM. Dies liegt wahrscheinlich an den in Abschnitt 5.4.3 herausgearbeiteten Gründen bei Quasi-Inkompressibilität. In diesem Fall wäre es zu empfehlen, exakte Inkompressibilität nach Abschnitt 5.2.5 einzufordern und im Rahmen der FEM mit einem gemischten Element zu verknüpfen, vgl. SIMO ET AL. [162]. Eine symptomatische Linderung dieser Problematik ist auch durch eine Verringerung der Schrittweite möglich, da somit die Verzerrungen weniger weit vom Trainingsraum des KNN entfernt sind.

Das ICNN-E-Poly-Modell zeigt bei der Quasi-Inkompressibilität weniger Probleme und konvergiert weiterhin stabil. Dahinter steckt zum einen, dass nicht die Spannungen direkt approximiert werden, sondern indirekt über die Verzerrungsenergiedichtefunktion. Zum anderen gilt die Polykonvexitätsbedingung auch außerhalb des Trainingsraums, was die Gefahr eines vorzeitigen Divergierens abschwächt.

- Die Trainingsdatenpaare $\{(\mathbf{x}, \mathbf{t})\}_{k=1}^{P_T}$ bestehen u. a. aus einheitenbehafteten Größen, wie bspw. die Spannungen. Daher ist ein fertiges KNN-Materialmodell an die Einheiten gebunden, die dem Trainingsprozess zugrunde liegen. Daher ist es zweckmäßig mithilfe einer Einheitenumrechnung am Integrationspunkt die Flexibilität des KNN-Materialmodells zu erhöhen. Dies ist ebenfalls in FEAP implementiert.

- Die Zwangssymmetrisierung (5.65) kann durchaus positive Effekte auf das Konvergenzverhalten haben, selbst wenn dadurch die NEWTON-Tangente nicht mehr exakt zum Residuum passt, vgl. WEBER ET AL. [189].
- Die Symmetrisierung der Lochscheibe ist im Fall der grundsätzlich anisotropen KNN-Materialmodelle nicht ganz korrekt. Da die Isotropie nur indirekt über die Daten erlernt und somit nicht exakt garantiert wird, ist die resultierende Struktur theoretisch nicht symmetrisch. Würde man das Beispiel vollständig modellieren, wären folglich leicht andere Ergebnisse zu erwarten.

Nach diesen FE-Berechnungen lässt sich ein erstes Zwischenfazit ziehen. Alle vorgestellten KNN-Materialmodelle sind prinzipiell geeignet, um elastisches Materialverhalten abbilden zu können, auch innerhalb der FEM. Dabei lassen sich die Erkenntnisse über Unterschiede zwischen den KNN-Materialmodellen sowie ihre Stärken und Schwächen vom Materialpunkt auf die Gesamtstruktur übertragen. Außerdem müssen bei KNN-Materialmodellen andere Sachverhalte beachtet werden als bei herkömmlichen Materialmodellen, wie bspw. die Größe und Form der Trainingsräume im Fall der Quasi-Inkompressibilität. Im nachfolgenden Kapitel 6 wird der Fokus mehr auf den Nebenbedingungen liegen sowie der Umgang mit ihnen.

6 Modellierung von plastischem Materialverhalten mit KNN

Aufbauend auf den grundsätzlichen Eigenschaften plastischer Materialien aus Kapitel 3 und den KNN-Architekturen aus Kapitel 4, wird im Folgenden die Modellierung plastischen Materialverhaltens mit KNN beschrieben. Der entscheidende Unterschied zu elastischem Materialverhalten ist die Abhängigkeit der Spannungsantwort von der Belastungsgeschichte. Dies wurde bereits durch die Beschreibung der Spannungen \mathbf{S} als Funktional der Verzerrungsgeschichte $\mathbf{E}_{(\tilde{t})}$, $\tilde{t} \in [0, t]$ in Gleichung (3.2) deutlich gemacht. Im vorliegenden Kapitel soll diese Abbildungsvorschrift durch ein KNN definiert werden:

$$\mathbf{S}_{(t)} = \mathcal{S}^{\text{KNN}}(\mathbf{E}_{(\tilde{t})}) \quad \mathbf{C}_{T(t)} = \frac{d\mathcal{S}^{\text{KNN}}(\mathbf{E}_{(\tilde{t})})}{d\mathbf{E}_{(t)}}, \quad \tilde{t} \in [0, t]. \quad (6.1)$$

Durch die Pfadabhängigkeit ist die Anwendung des Latin-Hypercube-Stichprobenverfahrens (LHS), wie im Fall elastischer Materialmodelle, nicht sachgemäß. Daher werden innerhalb dieses Kapitels die synthetischen Trainings- und Validierungsdatenpaare ausschließlich durch numerische Materialpunkt-Versuche generiert. Dies kann ohne Einschränkungen mit dem in Abschnitt 5.1 beschriebenen, gemischt gesteuerten Verfahren durchgeführt werden. Aus jedem dieser numerischen Materialpunkt-Versuche resultiert ein Pfad

$$\mathcal{P} = \{(\mathbf{E}_0, \mathbf{S}_0), (\mathbf{E}_1, \mathbf{S}_1), \dots, (\mathbf{E}_n, \mathbf{S}_n), \dots, (\mathbf{E}_N, \mathbf{S}_N)\}, \quad (6.2)$$

bestehend aus diskreten Gleichgewichtspunkten $(\mathbf{E}_n, \mathbf{S}_n)$. Diese werden im Folgenden als „GGW-Punkte“ abgekürzt. Abhängig vom definierten KNN-Materialmodell werden diese Verzerrungs-Spannungs-Paare in die Menge der jeweiligen Trainings- und Validierungsdatenpaare $\{(\mathbf{x}_k, \mathbf{t}_k)\}_{k=1}^P$ umgerechnet. Als KNN-Architekturen für plastisches Materialverhalten eignen sich das FNN aus Abschnitt 4.2, das NCNN aus Abschnitt 4.3.1 sowie das RNN aus Abschnitt 4.3.2.

In Abschnitt 6.1 werden verschiedene plastische KNN-Materialmodelle beschrieben. Physikalisch motivierte Nebenbedingungen für plastisches Materialverhalten werden in Abschnitt 6.2 vorgestellt, inklusive einer effektiven Strategie zur Generierung der NB-Punkte. In den Abschnitten 6.3 bis 6.5 folgen die Untersuchungen der beschriebenen KNN-Materialmodelle am Materialpunkt und im Rahmen der FEM. Die Erkenntnisse aus diesen Untersuchungen bereiten den Weg zur Anwendung an echten Versuchsdaten in Kapitel 7. Der Fokus dieses Kapitels liegt auf den physikalisch Motivierten Nebenbedingungen und ihrem Einfluss auf das Verhalten der KNN-Materialmodelle, gerade im Fall einer geringen Datenbasis.

6.1 KNN-Materialmodelle für plastisches Materialverhalten (KNN-P)

Im Folgenden wird die Definition der Eingangsvariablen \mathbf{x} und der Ausgangsvariablen \mathbf{z} zur Spezifizierung eines plastischen KNN-Materialmodells beschrieben. Dies muss derart erfolgen, dass die zugehörige Abbildungsvorschrift $\mathbf{z} = \mathbf{f}^{\text{KNN}}(\mathbf{x}, \mathbf{w})$ eine Funktion darstellt. Durch die Pfadabhängigkeit plastischen Materialverhaltens ist kein funktionaler Zusammenhang $\mathbf{S}(\mathbf{E})$ möglich. Daher müssen im Rahmen eines Feedforward-KNN (FNN oder NCNN) neben den aktuellen Verzerrungen \mathbf{E} weitere Größen im Eingangsvektor berücksichtigt werden, welche Informationen über die vergangene Belastungshistorie beinhalten. Klassischerweise sind dies Verzerrungs- und Spannungsgrößen oder deren Inkremente, bspw.

$$\mathbf{x} := [\Delta \mathbf{E}, \mathbf{E}_l, \mathbf{S}_l] \quad \text{und} \quad \mathbf{z} := \Delta \mathbf{S}, \quad (6.3)$$

die zum aktuellen (\mathbf{E}, \mathbf{S}) bzw. letzten GGW-Punkt $(\mathbf{E}_l, \mathbf{S}_l)$ gehören. Dabei ist $\Delta \mathbf{E} = \mathbf{E} - \mathbf{E}_l$ das üblicherweise vorgegebene Verzerrungsinkrement und $\Delta \mathbf{S} = \mathbf{S} - \mathbf{S}_l$ das gesuchte Spannungsinkrement. Dies ist in Abbildung 6.1 dargestellt.

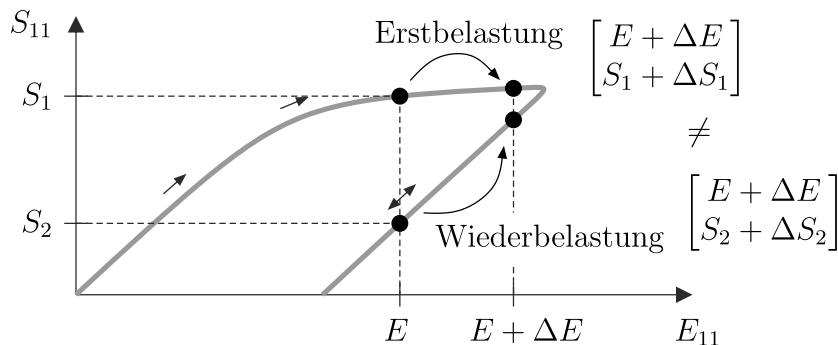


Abbildung 6.1: Illustration der Pfadabhängigkeit anhand von einachsigen plastischem Materialverhalten: Die GGW-Punkte von Erst- und Wiederbelastungspfad stimmen nicht überein, trotz gleicher Verzerrungen.

Im Fall des RNN müssen neben den Verzerrungen \mathbf{E} keine weiteren Eingangsvariablen berücksichtigt werden. Die rekurrenten Gewichte entsprechen den Geschichtsvariablen und werden im Trainingsprozess eingestellt. Sie sind allerdings nicht mehr physikalisch interpretierbar.

Die ab Abschnitt 6.1.2 aufgeführten KNN-Materialmodelle erfüllen alle physikalischen Eigenschaften, die ebenfalls zu Beginn von Abschnitt 5.2 angesprochen wurden: Die Beschränkung auf Körper 1. Grades, die materielle Objektivität sowie die Massen-, Impuls und Drehimpulsbilanz. Allerdings erfüllt keines der im Folgenden vorgestellten KNN-Materialmodelle die Dissipationsungleichung (3.25).

Daher kann die Konsistenz zu den thermodynamischen Hauptsätzen nicht garantiert werden. Darüber hinaus basieren die Eingangs- und Ausgangsvariablen ausschließlich auf der Information von Spannungen und Verzerrungen. Die Definition von Energie- oder Dissipationspotentialen sowie von Fließflächen und Fließregeln ist nicht notwendig. Anders ausgedrückt: Das in Abschnitt 3.4.1 beschriebene, eng geschnürte thermodynamische Korsett kann im Folgenden nicht erfüllt werden und wird daher nicht näher untersucht. Die Modelle sind rein auf Basis von in Versuchen gemessenen Verzerrungen und berechneten Spannungen kalibrierbar.

Nach Auftreten einer kinematischen Verfestigung verhält sich ein zuvor isotroper Werkstoff anisotrop bzgl. der Spannungsantwort. Daher werden im Folgenden stets alle Verzerrungs- und Spannungskomponenten bei der Definition der KNN-Materialmodelle berücksichtigt. Eine Reduktion auf skalare Invarianten wäre zwar grundsätzlich denkbar, diese Strategie wird hier allerdings nicht weiter verfolgt.

6.1.1 Geschichtsvariablen zur Modellierung von Verfestigung

Für das FNN können neben den Verzerrungen \mathbf{E}_l und den Spannungen \mathbf{S}_l des letzten GGW-Punkts zusätzliche Geschichtsvariablen \mathbf{h}_l definiert werden, welche die vollständige Belastungsgeschichte in integraler Form berücksichtigen. Solche Variablen sind vor allem zur Modellierung von Verfestigungsverhalten notwendig. Für eine Anwendung auf reale Daten ist es wichtig, dass diese Variablen ausschließlich auf Basis der Verzerrungen und der Spannungen und somit ausgehend von den (\mathbf{E}, \mathbf{S}) -Pfad (6.2) bestimmt werden können.

Bei klassischer kinematischer Verfestigung sind neben den Verzerrungsgrößen $\Delta\mathbf{E}$ und \mathbf{E}_l keine weiteren Geschichtsvariablen notwendig. Die Verschiebung der Fließfläche im Spannungsraum, vgl. Abbildung 3.6, kann explizit in den Verzerrungen $\mathbf{E} = \mathbf{E}_l + \Delta\mathbf{E}$ angegeben werden. Dies wird nachfolgend an einem einfachen, eindimensionalen Beispiel gezeigt. In Abbildung 6.2 ist eine stückweise lineare elastoplastische Spannungs-Dehnungskurve gegeben, mit der elastischen Steifigkeit C , der tangentialen Steifigkeit während des Fließens C_T sowie der initialen Fließspannung Y_0 . Durch Belastung bis zum GGW-Punkt (E_1, S_1) verschiebt sich der elastische Bereich von $\mathbb{E}_0 = [-Y_0, Y_0]$ zu $\mathbb{E}_1 = [-Y_0 + q_1, Y_0 + q_1]$, ohne seine Ausdehnung zu verändern. Bei entgegengesetzter Belastung tritt bereits am GGW-Punkt (E_2, S_2) Fließen ein. Somit muss die von der relativen Spannung q_1 abhängige Gleichung

$$F(S_2, q_1) = S_2 - q_1 - Y_0 = 0 \quad (6.4)$$

erfüllt sein. Mithilfe der in Abbildung 6.2 gegebenen grafischen Zusammenhänge lässt sich zeigen, dass die relative Spannung

$$q_1 = C_T \left(E_2 + \frac{Y_0}{C} \right) \quad (6.5)$$

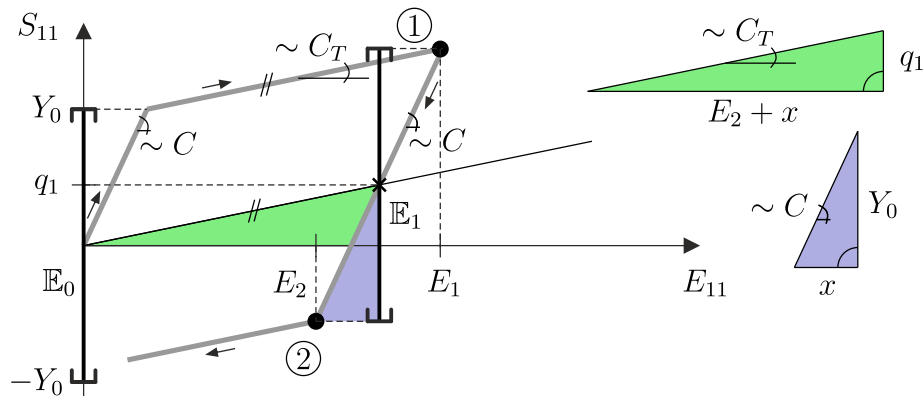


Abbildung 6.2: Uniaxiale Interpretation der Formulierung kinematischer Verfestigung in den Verzerrungen

explizit in Abhängigkeit der aktuellen Verzerrung E_2 sowie der Materialparameter C , C_T und Y_0 berechnet werden kann. Folglich reicht die Information über die aktuelle Verzerrung aus, um den Zeitpunkt des Fließens zu bestimmen, trotz zuvor auftretender kinematischer Verfestigung.

Klassische isotrope Verfestigung muss über eine zusätzliche skalare Geschichtsvariable h berücksichtigt werden. Dabei ist die Wahl dieser Variable generell beliebig, solange sie den gewünschten Effekt berücksichtigt. Eine Möglichkeit wäre bspw. die maximal bisher erreichte VON MISES'sche Vergleichsspannung

$$h(t) := \max_{\tau \leq t} \{ (S_V(\tau)) \} , \quad (6.6)$$

mit der Vergleichsspannung $S_V(\tau)$ zum Zeitpunkt $\tau \leq t$ nach Gleichung (3.111). Der Nachteil einer solchen monoton wachsenden Variable ist, dass etwaige Approximationsfehler nicht ausgeglichen werden können. KNN-Materialmodelle mit solchen Geschichtsvariablen neigen dazu, die Verfestigung zu überschätzen⁷⁰. Daher wird die energiebasierte, über die Belastungsgeschichte integrierte Variable

$$h^E(t) := \int_0^t \mathbf{S}^T \dot{\mathbf{E}} dt \quad (6.7)$$

empfohlen, welche in reduzierter Notation angegeben ist. Sie ist im CAUCHY-elastischen Bereich vollständig reversibel. Wird für die Verzerrungsrate $\dot{\mathbf{E}}$ der Differenzenquotient $(\mathbf{E}_{n+1} - \mathbf{E}_n)/(t_{n+1} - t_n)$ eingesetzt, kann das Integral im Intervall $t \in [t_n, t_{n+1}]$ eines Lastschritts mithilfe der Trapezregel numerisch integriert

⁷⁰ Dies hat sich im Rahmen verschiedener numerischer Studien im Kontext dieser Arbeit herauskristallisiert. Theoretisch sind monoton wachsende Geschichtsvariablen selbstverständlich möglich.

werden. Daraus folgt die Aktualisierungsvorschrift

$$h_{n+1}^E = h_n^E + \frac{1}{2}(\mathbf{S}_{n+1} + \mathbf{S}_n)^T (\mathbf{E}_{n+1} - \mathbf{E}_n) . \quad (6.8)$$

Mit Vorgabe von h_0^E können die Werte h_1^E, h_2^E, \dots lastschrittweise aktualisiert werden. Neben der hier vorgestellten skalaren Geschichtsvariable ist die Definition anderer Variablen prinzipiell beliebig. Verfestigungsverhalten, bei dem die Fließfläche ihre Gestalt komplizierter und anisotrop ändert, sich also nicht nur verschiebt oder vergrößert, würde nicht nur skalare sondern auch tensorwertige Geschichtsvariablen erfordern. Die Wahl, welche Geschichtsvariablen im Eingangsvektor berücksichtigt werden müssen, kann durch Erfahrung oder Trial-and-Error getroffen werden und ist nicht eindeutig.

Falls zusätzliche Geschichtsvariablen \mathbf{h} als KNN-Eingangsgrößen definiert wurden, müssen diese zunächst aus den gegebenen Verzerrungs-Spannungs-Pfaden (6.2) für jeden Lastschritt n berechnet werden. Für allgemeine Geschichtsvariablen in Form des Vektors h wird dies, ausgehend von einem Startwert \mathbf{h}_0 , über die inkrementelle Aktualisierungsvorschrift

$$\mathbf{h}_{n+1} = \mathbf{h}_{n+1}(\mathbf{h}_n, \mathbf{S}_{n+1}, \mathbf{S}_n, \mathbf{E}_{n+1}, \mathbf{E}_n) \quad (6.9)$$

durchgeführt, vgl. Beispiel (6.8). Über diese können die in Definition (6.2) gegebenen Pfade um die zusätzlichen Geschichtsvariablen \mathbf{h}_n erweitert werden zu

$$\mathcal{P} = \{(\mathbf{E}_0, \mathbf{S}_0, \mathbf{h}_0), (\mathbf{E}_1, \mathbf{S}_1, \mathbf{h}_1), \dots, (\mathbf{E}_n, \mathbf{S}_n, \mathbf{h}_n), \dots, (\mathbf{E}_N, \mathbf{S}_N, \mathbf{h}_N)\} . \quad (6.10)$$

Die so ergänzten Pfade sind die Basis für die folgenden FNN-Materialmodelle.

Wie im vorherigen Abschnitt dargelegt, sind diese zusätzlichen Variablen \mathbf{h} bei Verwendung eines RNN nicht notwendig.

6.1.2 Inkrementelles KNN-Materialmodell FNN-P-Ink

Die Definition (6.3) zur Motivation der angereicherten Eingangsvektoren wird um die zuvor eingeführten Geschichtsvariablen \mathbf{h} erweitert zu

$$\mathbf{x} := [\Delta\mathbf{E}, \mathbf{E}_l, \mathbf{S}_l, \mathbf{h}_l] \quad \text{und} \quad \mathbf{z} := \Delta\mathbf{S} . \quad (6.11)$$

Die Verzerrungen \mathbf{E}_l und die Spannungen \mathbf{S}_l des letzten GGW-Punkts sowie die Inkremente $\Delta\mathbf{E}$ und $\Delta\mathbf{S}$ enthalten jeweils n_s Komponenten. Im dreidimensionalen Fall gilt $n_s = 6$ und im ebenen Spannungszustand $n_s = 3$. Auch die Geschichtsvariablen \mathbf{h}_l beziehen sich auf den letzten GGW-Punkt und enthalten n_h Komponenten. Ihre Wahl hängt vom Anwendungsfall ab und ist nicht eindeutig. Damit ist im Allgemeinen die Größe des KNN-Eingangsvektors $n_i = 3n_s + n_h$ und

die des KNN-Ausgangsvektors $n_o = n_s$. In Spezialfällen können bspw. die Verzerrungen \mathbf{E}_l des letzten GGW-Punktes vernachlässigt werden, wie bei idealer Elastoplastizität. Gleiches gilt für den Vektor der zusätzlichen Geschichtsvariablen \mathbf{h}_l .

Die Spannungen

$$\mathbf{S} = \mathcal{S}^{\text{KNN}}(\mathbf{E}, \mathbf{E}_l, \mathbf{S}_l, \mathbf{h}_l) = \mathbf{S}_l + \Delta\mathbf{S} = \mathbf{S}_l + \mathbf{z} \quad (6.12)$$

können mit $\mathbf{E} = \mathbf{E}_l + \Delta\mathbf{E}$ direkt berechnet werden. Dazu ist es notwendig die Variablen \mathbf{E}_l , \mathbf{S}_l , \mathbf{h}_l als lokale Geschichtsvariablen vorzuhalten. Sie sind während der äußeren Gleichgewichtsiteration konstant. Im Rahmen der FEM wird der Umgang mit solchen Geschichtsvariablen in Abschnitt 6.5 dargestellt. Die Materialtangente in reduzierter Notation

$$\mathbf{C}_T = \frac{\partial\mathbf{S}}{\partial\mathbf{E}} = \frac{\partial(\mathbf{S}_l + \Delta\mathbf{S})}{\partial\Delta\mathbf{E}} \underbrace{\frac{\partial\Delta\mathbf{E}}{\partial\mathbf{E}}}_{\mathbf{I}} = \frac{\partial\Delta\mathbf{S}}{\partial\Delta\mathbf{E}} = \mathbf{j}_{\Delta\mathbf{S}(\Delta\mathbf{E})} \neq \mathbf{C}_T^T \quad (6.13)$$

enthält Anteile $\mathbf{j}_{\Delta\mathbf{S}(\Delta\mathbf{E})}$ der KNN-JACOBI-Matrix \mathbf{j} , welche den partiellen Ableitungen von $\Delta\mathbf{S}$ nach $\Delta\mathbf{E}$ entsprechen. Sie können bspw. durch die in Abschnitt 4.2.2 gegebenen Algorithmen für das FNN berechnet werden, mit $\mathbb{I} = \{1, \dots, n_s\}$. Die Materialtangente ist im Allgemeinen nicht symmetrisch. Dieses Standard-KNN-Materialmodell wurde bereits von GHABOUSSI ET AL. [58] verwendet, allerdings ohne zusätzliche Geschichtsvariablen \mathbf{h}_l .

Um das KNN für dieses Materialmodell trainieren zu können, müssen die GGW-Punkte des Pfades (6.10) in die zugehörigen $(\mathbf{x}_k, \mathbf{t}_k)$ -Paare nach Definition (6.11) transformiert werden. Dabei ist auf die Prozessrichtung bei Bildung der Verzerrungs- und Spannungsinkremente zu achten: die Inkremente dürfen nicht „rückwärts“ gebildet werden. Darüber hinaus dürfen zwischen verschiedenen *Belastungsrichtungen* keine Inkremente gebildet werden, um das Kausalitätsprinzip nicht zu verletzen. Dies ist in Abbildung 6.3 dargestellt. Unter Beachtung dieser Regeln können inkrementelle Datenpaare durch Subtraktion von GGW-Punkten generiert werden. Die Eingangs- und Zielvektoren

$$\mathbf{x} = \begin{bmatrix} \mathbf{E}_n - \mathbf{E}_{n-\Delta n} \\ \mathbf{E}_n \\ \mathbf{S}_n \\ \mathbf{h}_n \end{bmatrix} \quad \text{und} \quad \mathbf{t} = \mathbf{S}_n - \mathbf{S}_{n-\Delta n} \quad (6.14)$$

hängen dabei vom Indexversatz $\Delta n \geq 0$ ab. Für $\Delta n = 0$ wird ein Datenpunkt ohne Verzerrungszuwachs generiert. Der zugehörige Spannungszuwachs ist ebenfalls null, was bspw. für viskoses Materialverhalten nicht gültig wäre. Wird $\Delta n = 1$ gewählt, findet die Inkrementbildung zwischen benachbarten Versuchspunkten bzw.

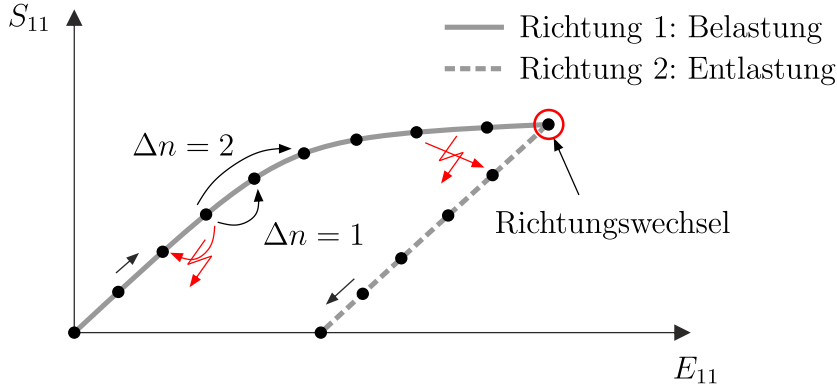


Abbildung 6.3: Transformation der GGW-Punkte zu inkrementellen Datenpaaren anhand einachsigem plastischem Materialverhalten unter Beachtung der korrekten Prozessrichtung

GGW-Punkten statt. Zur Abdeckung einer großen Breite von Schrittweiten ist es zweckmäßig, für die Versätze $\Delta n \in \{0, 1, 2, \dots\} = \mathbb{I}_{\Delta n}$ mehrere Werte zu definieren. Dies stabilisiert das Verhalten der KNN-Materialmodelle innerhalb einer äußeren Gleichgewichtssiteration.

Gegeben seien eine Belastungsrichtung $\mathcal{P}_d = \{(\mathbf{E}, \mathbf{S}, \mathbf{h})_n\}_{n=1}^{M_d}$ mit M_d GGW-Punkten und eine Menge $\mathbb{I}_{\Delta n}$ von Indexversätzen. Die damit berechenbare Anzahl an Trainingsdatenpaaren ist

$$P_d = \sum_{i=1}^{|\mathbb{I}_{\Delta n}|} (M_d - \Delta n_i) = |\mathbb{I}_{\Delta n}| M_d - \sum_{i=1}^{|\mathbb{I}_{\Delta n}|} \Delta n_i, \quad (6.15)$$

mit der Anzahl von Indexversätzen $|\mathbb{I}_{\Delta n}|$. Für das Beispiel aus Abbildung 6.3 ergäben sich mit $\mathbb{I}_{\Delta n} = \{0, 1, 2, 3, 4\}$ für die Belastungsrichtung $d = 1$ mit $M_1 = 10$ folglich $P_1 = 40$ Datenpaare, für die Entlastungsrichtung $d = 2$ mit $M_2 = 6$ gälte $P_2 = 20$.

6.1.3 Sequenzielles KNN-Materialmodell FNN-P-Seq

Alternativ zur Formulierung in den Verzerrungs- und Spannungsincrementen $\Delta \mathbf{E}$ und $\Delta \mathbf{S}$ kann das KNN-Materialmodell auch direkt in den Verzerrungen \mathbf{E} und den Spannungen \mathbf{S} des neuen GGW-Punkts definiert werden.

$$\mathbf{x} := [\mathbf{E}, \mathbf{E}_l, \mathbf{S}_l, \mathbf{h}_l] \quad \text{und} \quad \mathbf{z} := \mathbf{S} \quad (6.16)$$

Die Spannungen

$$\mathbf{S} = \mathcal{S}^{\text{KNN}}(\mathbf{E}, \mathbf{E}_l, \mathbf{S}_l, \mathbf{h}_l) = \mathbf{z} \quad (6.17)$$

entsprechen den Ausgangsvariablen \mathbf{z} des KNN. Die Materialtangente

$$\mathbf{C}_T = \frac{\partial \mathbf{S}}{\partial \mathbf{E}} = \mathbf{j}_{\mathbf{S}(\mathbf{E})} \neq \mathbf{C}_T^T \quad (6.18)$$

entspricht den Anteilen der KNN-JACOBI-Matrix \mathbf{j} , welche die partiellen Ableitungen von \mathbf{S} nach \mathbf{E} enthalten. Sie ist im Allgemeinen nicht symmetrisch. Auch dieses Modell ist häufiger in der Literatur anzutreffen und wurde bspw. von YUN ET AL. [201] verwendet.

Im Vergleich zum inkrementellen Modell des Abschnitts 6.1.2 ist die gleiche Information in den Eingangs- und Ausgangsvektoren enthalten. Sie ist mit $\mathbf{E} = \mathbf{E}_l + \Delta \mathbf{E}$ und $\mathbf{S} = \mathbf{S}_l + \Delta \mathbf{S}$ nur anders verteilt. Daher können die in Abschnitt 6.1.2 berechneten inkrementellen Datenpaare ohne Weiteres in die sequenziellen umgerechnet werden.

6.1.4 Inkrementelles KNN-Materialmodell NCNN-P-Ink

Das KNN-Materialmodell FNN-P-Ink kann auch mit der NCNN-Architektur aus Abschnitt 4.3.1 formuliert werden. Dabei bestünde kein Grund, die Konvexität der Potentialfunktion einzufordern. Dies stand schon im Fall elastischen Materialverhaltens im Widerspruch zu einigen physikalischen Prinzipien, vgl. Abschnitt 3.3.5.

Die Definition der Eingangsvariablen ist identisch zu Abschnitt 6.1.2. Dadurch können die dort beschriebenen Trainingsdaten uneingeschränkt zum Training dieses Materialmodells genutzt werden. Das Spannungsinkrement wird als partielle Ableitung

$$\Delta \mathbf{S} = \frac{\partial \Psi(\Delta \mathbf{E}, \mathbf{E}_l, \mathbf{S}_l, \mathbf{h}_l, \mathbf{w})}{\partial \Delta \mathbf{E}} \quad (6.19)$$

der nicht-transformierten Potentialfunktion Ψ definiert. Die Materialtangente

$$\mathbf{C}_T = \frac{\partial^2 \Psi}{\partial \Delta \mathbf{E}^2} = \mathbf{j}_{\Delta \mathbf{S}(\Delta \mathbf{E})} = \mathbf{C}_T^T \quad (6.20)$$

entspricht der zweifachen Ableitung dieses Potentials und ist daher stets symmetrisch. In dieser Form ist das KNN-Materialmodell nach bestem Wissen des Autors noch nicht verwendet worden, allerdings verwenden bspw. MASI ET AL. [125] ein ähnliches Modell, welches aus mehreren gekoppelten KNN besteht.

Dies steht hier nicht direkt im Zusammenhang mit Energieerhaltung oder Entropiezunahme, sondern mit der bloßen Existenz einer skalaren Funktion Ψ , über deren partielle Ableitung nach dem Verzerrungsinkrement das Spannungsinkrement berechnet werden kann. Die Einbettung in den thermodynamischen Rahmen des Abschnitts 3.4.1 ist nach Einschätzung des Autors nicht ohne Weiteres möglich. Die Konsistenz mit der Dissipationsungleichung kann folglich nicht gewährleistet werden.

6.1.5 RNN-Materialmodell RNN-P

Wird das rekurrente neuronale Netz (RNN) aus Abschnitt 4.3.2 verwendet, müssen neben den aktuellen Verzerrungen \mathbf{E} keine weiteren Eingangsvariablen definiert werden. Durch die intrinsische ‚zeitliche‘ Komponente des RNN werden die aktuellen Spannungen $\mathbf{S} = \mathbf{z}_{(t)}$ ausschließlich mit den aktuellen Verzerrungen $\mathbf{E} = \mathbf{x}_{(t)}$ als Eingangsvariablen berechnet⁷¹. Die Information über die vergangene Belastungshistorie steckt vollständig in den Neuronen-Zuständen $\{\mathbf{y}_{(t-1)}\}_{L=1}^{n_h}$ der Zwischenschichten. Die Eingangs- und Ausgangsvariablen für einen ‚Zeit‘-Schritt lassen sich folglich mit

$$\mathbf{x}_{(t)} = \mathbf{E} \quad \text{und} \quad \mathbf{z}_{(t)} = \mathbf{S} \quad (6.21)$$

angeben. Die Neuronen-Zustände $\{\mathbf{y}_{(t-1)}\}_{L=1}^{n_h}$ müssen sich im Laufe der Belastungsgeschichte zunächst aufbauen. Daher ist es beim RNN nicht möglich, einen Schritt inmitten eines Belastungsprozesses isoliert zu berechnen, wie es beim FNN durch die Angabe des zugehörigen GGW-Punkts $(\mathbf{E}_t, \mathbf{S}_t)$ der Fall ist. In dieser Form werden RNN häufig verwendet, bspw. von ABUEIDDA ET AL. [1] oder WU ET AL. [199].

Darüber hinaus reagiert das RNN auf den folgenden Umstand empfindlicher als das FNN: Die Zeitreihe der Eingangsvariablen $\mathcal{X} = \{\mathbf{E}_{(t)}\}_{t=0}^{\tau}$ ist während des Trainings stets vollständig gegeben. Der Fehler wird bzgl. der Zeitreihe der Ausgangsvariablen $\mathcal{Z} = \{\mathbf{S}_{(t)}\}_{t=0}^{\tau}$ berechnet. Bei der Anwendung als Materialmodell ist das mitnichten der Fall. Die Verzerrungen werden im Laufe der äußeren Gleichgewichtsiteration mithilfe der RNN-Ausgabe erst ermittelt. Dadurch können sich rasch Abweichungen in den resultierenden Zeitreihen der Verzerrungen und vor allem in den zugehörigen Zuständen der Zwischenneuronen ergeben gegenüber denjenigen, die während des Trainings erlernt wurden.

U. a. aus den oben genannten Gründen ist das RNN als KNN-Materialmodell auf eine vergleichsweise große Datenmenge angewiesen. Dies ist auch aus der Sicht des Autors bei den aus echten Versuchen nicht zu realisierenden Datenmengen der Veröffentlichungen von bspw. WU ET AL. [199] und ABUEIDDA ET AL. [1] zu sehen. Die Anwendung im Rahmen der Mehrskalenmodellierung ist dagegen realistischer.

⁷¹ Die Variable t wurde in Abschnitt 4.3.2 zur Kennzeichnung aufeinanderfolgender Ereignisse eingeführt. Sie hat nichts mit der physikalischen Zeit gemein.

6.2 Physikalische Nebenbedingungen für KNN-P

Auch für die Modellierung plastischen Materialverhaltens mit KNN lassen sich physikalische Nebenbedingungen formulieren, um den Trainingsprozess zu regularisieren und letztlich in der Anwendung als Materialmodell mit der Physik verträglichere Ergebnisse zu erzielen. Die Prinzipien aus Kapitel 3 dienen dabei als Grundlage zur Formulierung zusätzlicher Fehlerterme, die innerhalb des Optimierungsprozesses minimiert werden. Im Folgenden werden ausschließlich Nebenbedingungen für das explizite inkrementelle KNN-Materialmodell FNN-P-Ink aus Abschnitt 6.1.2 vorgestellt, mit $\mathbf{x} = [\Delta \mathbf{E}, \mathbf{E}_l, \mathbf{S}_l, \mathbf{h}_l]$ und $\mathbf{z} = \Delta \mathbf{S}$. Eine Übertragung auf andere Materialmodelle ist sinngemäß möglich.

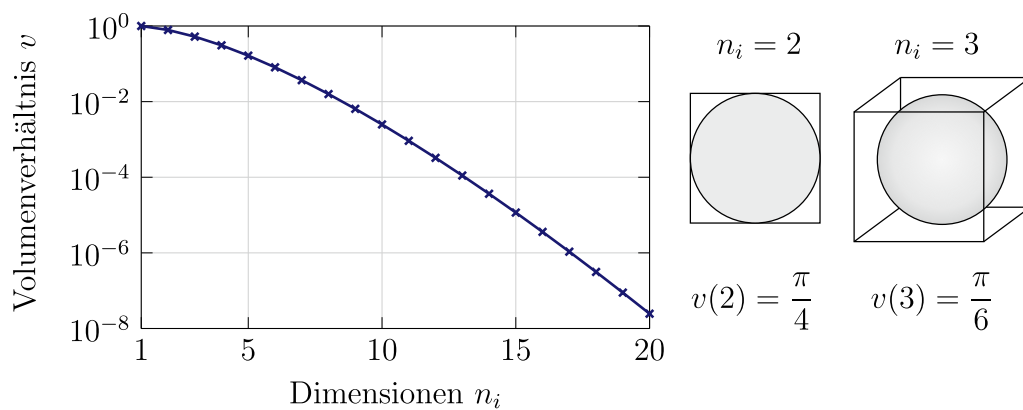


Abbildung 6.4: Verhältnis von Hyperkugel-Volumen zum Volumen des einschließenden Hyperwürfels in Abhängigkeit der Dimension n_i

An dieser Stelle sei noch einmal hervorgehoben, dass zur Einforderung etwaiger Nebenbedingungen nur die Eingangsvariablen – die NB-Punkte \mathbf{x}^C – im Trainingsraum $\Omega \subset \mathbb{R}^{n_i}$ generiert werden müssen. Im Fall der plastischen KNN-Materialmodelle ist das allerdings eine größere Herausforderung als bei den elastischen Modellen aus Kapitel 5. Die Anzahl der Eingangsvariablen n_i kann beträchtlich werden. Würden in diesem hochdimensionalen Raum die NB-Punkte mittels LHS generiert werden, wäre ein Großteil der Daten im physikalisch nicht sinnvollen Bereich zu erwarten. Zur Untermauerung dieses Auswuchses des „Fluchs der Dimensionalität“ (engl.: *curse of dimensionality*⁷²) wird folgende Analogie herangezogen: Der physikalisch sinnvolle Trainingsraum $\Omega \subset \mathbb{R}^{n_i}$ sei durch die

⁷² Dieser Begriff wurde von RICHARD BELLMAN [12] im Rahmen der Regelungstheorie eingeführt. Seither ist *curse of dimensionality* ein geflügeltes Wort im Zusammenhang mit der Handhabung von Daten in hochdimensionalen Räumen.

n_i -dimensionale Kugel mit Radius R definiert. Ihr Volumen lässt sich mit

$$V^{\text{Kugel}}(n_i) = \frac{2 \pi^{\frac{n_i}{2}} R^{n_i}}{n_i \Gamma(\frac{n_i}{2})} \quad (6.22)$$

in Abhängigkeit der Dimension n_i definieren. Die transzendente Gamma-Funktion $\Gamma(n_i)$ entspricht für $n_i \in \mathbb{N}$ gerade der Fakultät $\Gamma(n_i) = (n_i - 1)!$ der nächst kleineren natürlichen Zahl. Wird das Volumen der Hyperkugel (6.22) auf das des sie einschließenden n_i -dimensionalen Hyperwürfels mit Kantenlänge $2R$ bezogen, so resultiert der in Abbildung 6.4 dargestellte Zusammenhang des Volumenverhältnisses

$$v(n_i) := \frac{V^{\text{Kugel}}(n_i)}{(2R)^{n_i}} = \frac{\pi^{\frac{n_i}{2}}}{2^{n_i-1} n_i \Gamma(\frac{n_i}{2})}. \quad (6.23)$$

Beim KNN-Materialmodell FNN-P-Ink liegen ohne zusätzliche Geschichtsvariablen $n_i = 18$ Eingangsvariablen vor. Dies entspricht einem Volumenverhältnis von $\nu \approx 3,13 \cdot 10^{-7}$. Anders ausgedrückt: Wird auf Basis einer Gleichverteilung in diesem 18-dimensionalen Hyperwürfel, bspw. mittels LHS, zufällig ein Datenpunkt generiert, so wäre die Chance ca. 1 zu 3,2 Millionen, dass er auch innerhalb der Hyperkugel liegt⁷³. Diese Analogie soll zeigen, dass eine Strategie zur Datengenerierung mittels LHS nicht zielführend ist. Daher wird im folgenden Abschnitt zunächst eine Alternative vorgeschlagen, die sich im Rahmen der plastischen KNN-Materialmodellierung als zweckmäßig herausgestellt hat.

6.2.1 Datengenerierung innerhalb einer konvexen Hülle

Als Trainingsraum Ω wird im Folgenden die konvexe Hülle der Menge der Trainingsdatenpunkte $\{\mathbf{x}_k\}_{k=1}^{P_T}$ definiert. Eine konvexe Hülle ist die kleinstmögliche konvexe Menge, die diese Punkte umschließt. Ein zweidimensionales Beispiel ist in Abbildung 6.5 dargestellt. Eine konvexe Menge hat die folgende wichtige Eigenschaft: Die Verbindungslinie zwischen zwei Punkten innerhalb einer konvexen Menge ist vollständig in dieser enthalten.

Um zufällig einen NB-Punkt \mathbf{x}^C zu generieren, der nicht Teil der Trainingsdatenpunkte $\{\mathbf{x}_k\}_{k=1}^{P_T}$ ist, wird sich diese Eigenschaft zunutze gemacht. Das Vorgehen ist in Abbildung 6.5 links dargestellt. Insgesamt werden zufällig vier unterschiedliche Trainingsdatenpunkte $\{\mathbf{x}_{a1}, \mathbf{x}_{a2}, \mathbf{x}_{b1}, \mathbf{x}_{b2}\} \subset \{\mathbf{x}_k\}_{k=1}^{P_T}$ aus der Menge aller Trainingsdatenpunkte gewählt. Auf der Verbindungslinie zwischen den Punkten \mathbf{x}_{a1} und \mathbf{x}_{a2} sowie \mathbf{x}_{b1} und \mathbf{x}_{b2} wird jeweils zufällig ein neuer Punkt $\mathbf{x}_a \in \Omega$ bzw.

⁷³ Die Chance entspricht der Wahrscheinlichkeit des Eintretens eines Ereignisses bezogen auf die Wahrscheinlichkeit des Nicht-Eintretens jenes Ereignisses. Die Chance beim klassischen „6 aus 49 - Lotto“ ohne Superzahl zu gewinnen, ist ca. 1 zu 14 Millionen. Sie ist dementsprechend nur etwas schlechter als die Chance die 18-dimensionale Hyperkugel zu treffen.

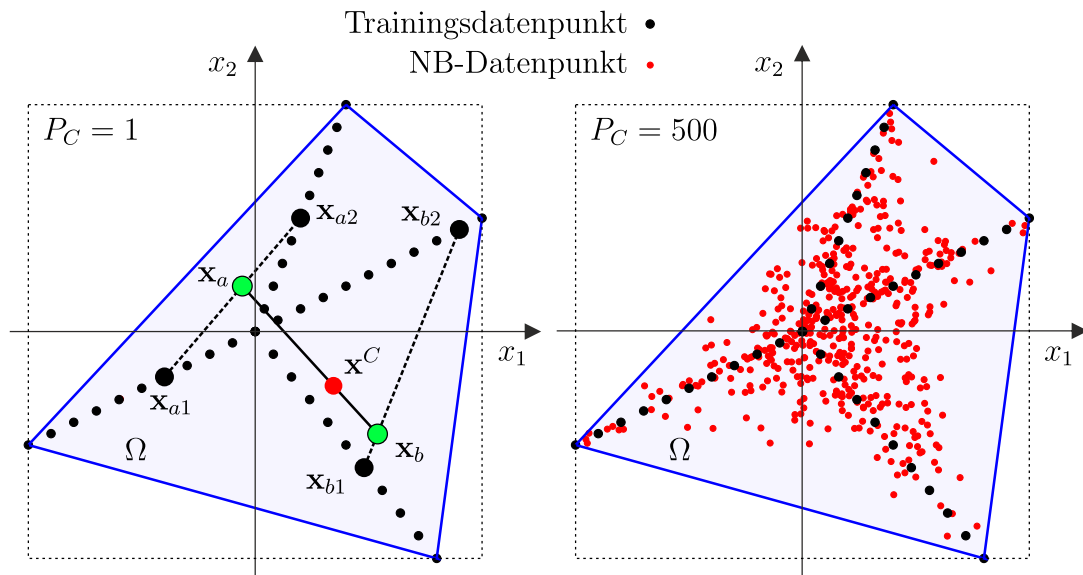


Abbildung 6.5: Zufällige Datengenerierung innerhalb einer konvexen Hülle Ω

$\mathbf{x}_b \in \Omega$ erzeugt. Wiederum auf der Verbindungslinie zwischen diesen neuen Punkten \mathbf{x}_a und \mathbf{x}_b wird zufällig der Punkt \mathbf{x}^C generiert. Mithilfe dreier gleichverteilter Pseudozufallszahlen $\theta_a, \theta_b, \theta_c \sim \mathcal{U}(0, 1)$ kann \mathbf{x}^C über die explizite Formel

$$\mathbf{x}^C = (1 - \theta_a)(1 - \theta_c) \mathbf{x}_{a1} + \theta_a(1 - \theta_c) \mathbf{x}_{a2} + \theta_c(1 - \theta_b) \mathbf{x}_{b1} + \theta_c\theta_b \mathbf{x}_{b2} \quad (6.24)$$

berechnet werden. In Abbildung 6.5 rechts sind $P_C = 500$ solcher Punkte in einer konvexen Hülle generiert worden. Es ist zu erkennen, dass die globale Wahrscheinlichkeitsverteilung nicht gleichverteilt ist. Die zufällig generierten Punkte tendieren zu Bereichen, die viele Trainingsdatenpunkte enthalten. Diese Eigenschaft hat sich bei der Generierung von NB-Punkten nicht als Nachteil herausgestellt.

An dieser Stelle sei nochmal angemerkt, dass es im Fall elastischer Materialien nicht möglich wäre, die Spannungen \mathbf{S} unabhängig von den Verzerrungen \mathbf{E} zufällig zu generieren, da für jeden Verzerrungszustand ausschließlich ein Spannungszustand gültig ist. Im Fall plastischer Materialien können durch Be- und Entlastungsvorgänge unendlich viele, allerdings an Grenzen gebundene Spannungszustände erreicht werden. Aus diesen Grenzen resultiert ebenfalls die Motivation der konvexen Hülle.

6.2.2 Inkrementelle Normalisierung

Bei ratenunabhängigem Materialverhalten ist bei einem Verzerrungsinkrement $\Delta \mathbf{E} = \mathbf{0}$ ein null-wertiges Spannungsinkrement $\Delta \mathbf{S} = \mathbf{0}$ zu erwarten.

$$\Delta \mathbf{S}_0 := \mathcal{S}^{\text{KNN}}(\mathbf{E}) - \mathbf{S}_l \stackrel{\text{Ink.}}{=} \Delta \mathbf{S}(\Delta \mathbf{E} = \mathbf{0}, \mathbf{E}_l, \mathbf{S}_l, \mathbf{h}_l) \stackrel{!}{=} \mathbf{0} \quad (6.25)$$

Dabei wird im Folgenden angenommen, dass die Komponenten des Verzerrungsin-
krementes $\Delta \mathbf{E}$ die ersten n_s KNN-Eingangsvariablen sind. Diese Nebenbedingung
ähmelt derjenigen der Normalisierung aus Abschnitt 5.3.1. Der entscheidende Un-
terschied liegt in den NB-Punkten

$$\mathbf{x}^0 = [\mathbf{0}, \mathbf{E}_l, \mathbf{S}_l, \mathbf{h}_l] \in \Omega^0 := \{\Omega \mid \Delta \mathbf{E} = \mathbf{0}\} , \quad (6.26)$$

auf denen sie einzufordern ist, nämlich für jeden GGW-Punkt $(\mathbf{E}_l, \mathbf{S}_l, \mathbf{h}_l)$ inner-
halb des Trainingsraums Ω . Ein mit der Strategie aus Abschnitt 6.2.1 generierter
NB-Punkt mit allgemein $\Delta \mathbf{E} \neq \mathbf{0}$ kann dahingehend angepasst werden, indem
das Verzerrungsincrement im Anschluss zu $\mathbf{0}$ gesetzt wird.

Diese inkrementelle Normalisierung wird ebenfalls für jede der n_s unabhängi-
gen Spannungskomponenten einzeln eingefordert. Für jeden NB-Punkt \mathbf{x}_k^0 lauten
folglich die n_s Bedingungen

$$h_{jk} = h_j(\mathbf{x}_k^0, \mathbf{w}) = \hat{z}_j(\mathbf{x}_k^0, \mathbf{w}) + \frac{m_{zj}}{s_{zj}} = 0 , \quad j = 1, \dots, n_s , \quad (6.27)$$

mit ihren Gradienten

$$\nabla h_{jk} = \frac{\partial \hat{z}_{jk}}{\partial \mathbf{w}} , \quad j = 1, \dots, n_s . \quad (6.28)$$

Die Nebenbedingung ist ebenfalls mit $t_{jk}^0 = 0$ und $\hat{t}_{jk}^0 = -m_{zj}/s_{zj}$ als zusätzliches
Trainingsdatenpaar interpretierbar. Im Kontext der Penalty-Methoden ist der
Faktor zur Berechnung des Gradienten nach Abschnitt 4.2.4 gerade $d_{jk} = \rho_{h_{jk}}$.
Eine zusätzliche Normalisierung des Fehlerterms im Vergleich zum Trainingsda-
tenfehler ist nicht notwendig.

6.2.3 Stationarität der Normalisierung

Die inkrementelle Normalisierung aus Abschnitt 6.2.2 gilt für jeden GGW-Punkt
 $(\mathbf{E}_l, \mathbf{S}_l, \mathbf{h}_l)$. Eine weitere Möglichkeit dies zu berücksichtigen ist, für $\Delta \mathbf{E} = \mathbf{0}$ die
partiellen Ableitungen des Spannungsincrementes nach den restlichen Eingangs-
variablen zu null einzufordern.

$$\frac{\partial \Delta \mathbf{S}(\mathbf{x}^0)}{\partial \mathbf{E}_l} \stackrel{!}{=} \mathbf{0} \quad \frac{\partial \Delta \mathbf{S}(\mathbf{x}^0)}{\partial \mathbf{S}_l} \stackrel{!}{=} \mathbf{0} \quad \frac{\partial \Delta \mathbf{S}(\mathbf{x}^0)}{\partial \mathbf{h}_l} \stackrel{!}{=} \mathbf{0} \quad (6.29)$$

Diese Nebenbedingung gilt ebenfalls auf der Untermenge Ω^0 aus Definition (6.26).
Unter der Annahme, dass die Komponenten des Verzerrungsincrementes $\Delta \mathbf{E}$ die
ersten n_s KNN-Eingangsvariablen sind, folgen die $(n_i - n_s)n_s$ Terme der Neben-
bedingung zu

$$h_{jik} = h_{ji}(\mathbf{x}_k^0, \mathbf{w}) = \frac{\partial \hat{z}_{jk}}{\partial \hat{x}_i} , \quad \begin{cases} j = 1, \dots, n_s \\ i = n_s + 1, \dots, n_i \end{cases} . \quad (6.30)$$

Die Fehlerterm-Normalisierung wurde mit $\alpha_{ji} = s_{zj}/s_{xi}$ berücksichtigt, vgl. Abschnitt 5.3.2. Dadurch kürzen sich die Terme der Ableitungstransformation heraus. Die Gradienten der Terme (6.30) dieser Nebenbedingung lauten jeweils

$$\nabla h_{jik} = \frac{\partial}{\partial \mathbf{w}} \left(\frac{\partial \hat{z}_{jk}}{\partial \hat{x}_i} \right). \quad (6.31)$$

Der Fehlergradient $\nabla \mathcal{L}_C$ muss folglich mit dem Algorithmus aus Abschnitt 4.2.5 berechnet werden, mit den Faktoren $f_{ijk} = \rho_{h_{ijk}}$, vgl. Abschnitt 4.5.4.

6.2.4 Symmetrie der Materialtangente

Analog zu Abschnitt 5.3.2 kann die Symmetrie der Materialtangente mit

$$\mathbf{C}_T \stackrel{!}{=} \mathbf{C}_T^T \quad (6.32)$$

eingefordert werden. Für eine Diskussion zum thermodynamischen Rahmen vgl. Abschnitt 6.1.4.

Die Verzerrungsinkremente $\Delta \mathbf{E}$ sind die ersten n_s Eingangsvariablen, die Spannungsinkremente $\Delta \mathbf{S}$ die n_s Ausgangsvariablen. Für jedes Indexpaar (u, v) mit $v = 1, \dots, n_s - 1$ und $u = v + 1, \dots, n_s$ lassen sich die normalisierten Fehlerterme $h_{vuk} = h_{vu}(\mathbf{x}_k^C, \mathbf{w})$, deren Gradienten ∇h_{vuk} sowie die Vorfaktoren f_{ijk} zur Berechnung des Fehlergradienten $\nabla \mathcal{L}_C$ analog zu Abschnitt 5.3.2 mit den Gleichungen (5.44), (5.46) und (5.47) berechnen.

6.2.5 Inkrementelle Stabilität

Für inelastisches, ratenunabhängiges Materialverhalten kann für kleine Verzerrungen das Stabilitätskriterium

$$\dot{\boldsymbol{\sigma}} : \dot{\boldsymbol{\varepsilon}} \approx \dot{\mathbf{S}} : \dot{\mathbf{E}} \geq 0 \quad (6.33)$$

nach HILL [79] bzw. DRUCKER [37] herangezogen werden. Durch den Ausschluss der Inkrementbildung zwischen verschiedenen Belastungsrichtungen, vgl. Abbildung 6.3, ist es möglich, dieses Stabilitätskriterium in den Inkrementen

$$\Delta \mathbf{S}^T \Delta \mathbf{E} \stackrel{!}{\geq} 0 \quad (6.34)$$

zu definieren. Diese Ungleichung kann für einen NB-Punkt $\mathbf{x}_k^C \in \Omega$ durch den Fehlerterm

$$g_k = g(\mathbf{x}_k^C, \mathbf{w}) = \frac{1}{\alpha} \sum_{j=1}^{n_s} (m_{zj} + s_{zj} \hat{z}_{jk}) (m_{xj} + s_{xj} \hat{x}_{jk}) \quad (6.35)$$

mithilfe der Penalty-Methoden berücksichtigt werden. Der Normalisierungsfaktor α wird analog zur Isotropienebenbedingung mit Gleichung (5.54) definiert. Der Gradient dieses Fehlerterms ist

$$\nabla g_k = \frac{1}{\alpha} \sum_{j=1}^{n_s} s_{zj} (m_{xj} + s_{xj} \hat{x}_{jk}^C) \frac{\partial \hat{z}_{jk}}{\partial \mathbf{w}}. \quad (6.36)$$

Daraus folgen die Faktoren

$$d_{jk} = \frac{1}{\alpha} \rho_{h_k} s_{zj} (m_{xj} + s_{xj} \hat{x}_{jk}^C), \quad (6.37)$$

die zur Berechnung des Fehlergradienten $\nabla \mathcal{L}_C$ nach Abschnitt 4.2.4 benötigt werden. Die ρ_{h_k} -Werte sind abhängig von der gewählten Penalty-Methode, vgl. Abschnitt 4.5.4.

Zur Einforderung dieser Ungleichheits-Nebenbedingungen ist zuvor die Berechnung der aktiven Menge (4.159) der NB-Datenpunkte zu ermitteln.

6.2.6 Energiedissipation in direktem geschlossenen Verzerrungszyklus

Ausgehend von einem GGW-Punkt $(\mathbf{E}_l, \mathbf{S}_l, \mathbf{h}_l)$ wird ein Schritt $\Delta \mathbf{E}$ zum Punkt $(\mathbf{E}^I = \mathbf{E}_l + \Delta \mathbf{E}, \mathbf{S}^I, \mathbf{h}^I)$ durchgeführt. Die Spannungen

$$\mathbf{S}^I := \mathbf{S}_l + \Delta \mathbf{S}^I = \mathbf{S}_l + \mathbf{z}(\mathbf{x}^I) \quad (6.38)$$

werden mit dem Eingangsvektor $\mathbf{x}^I := [\Delta \mathbf{E}, \mathbf{E}_l, \mathbf{S}_l, \mathbf{h}_l]$ berechnet und die Geschichtsvariablen \mathbf{h}^I anschließend aktualisiert. Direkt danach folgt ein Schritt in die entgegengesetzte Richtung mit $-\Delta \mathbf{E}$. Die Spannungen

$$\mathbf{S}^{II} := \mathbf{S}^I + \Delta \mathbf{S}^{II} = \mathbf{S}^I + \mathbf{z}(\mathbf{x}^{II}) \quad (6.39)$$

werden mit dem Eingangsvektor $\mathbf{x}^{II} := [-\Delta \mathbf{E}, \mathbf{E}_l + \Delta \mathbf{E}, \mathbf{S}^I, \mathbf{h}^I]$ berechnet, welcher direkt von den Spannungen \mathbf{S}^I und den Geschichtsvariablen \mathbf{h}^I des vorherigen Schritts abhängt. Die Bilanz der spezifischen Energie $\Delta \Pi$ innerhalb dieses geschlossenen Verzerrungspfads darf nicht größer null sein, sonst würde das KNN-Materialmodell gegen den 2. Hauptsatz der Thermodynamik verstoßen⁷⁴. Mithilfe der Trapezregel lässt sich die Energiebilanz mit

$$\Delta \Pi = \frac{1}{2} (\mathbf{S}_l + \mathbf{S}^I)^T \Delta \mathbf{E} - \frac{1}{2} (\mathbf{S}^I + \mathbf{S}^{II})^T \Delta \mathbf{E} \quad (6.40)$$

berechnen. Dies ist im uniaxialen Fall in Abbildung 6.6 dargestellt. Nach Einset-

⁷⁴ Die Einforderung dieser Nebenbedingung garantiert im Umkehrschluss aber nicht, dass der 2. Hauptsatz der Thermodynamik immer erfüllt ist.

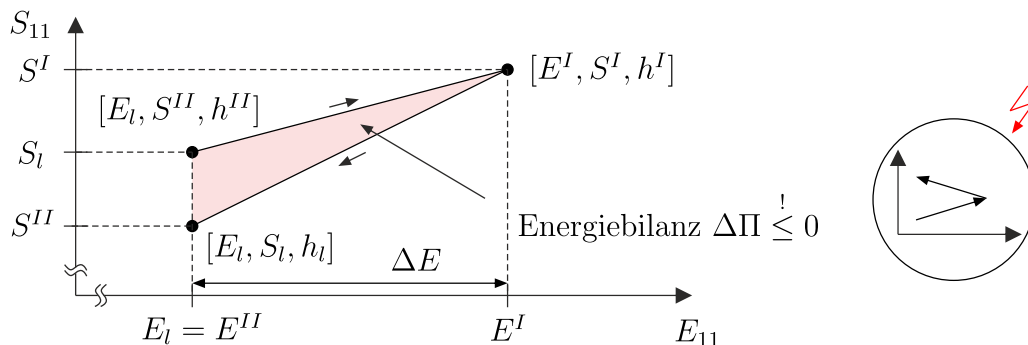


Abbildung 6.6: Illustration an einem uniaxialen Beispiel, dass die Energie innerhalb eines geschlossenen Dehnungspfades nicht größer null sein darf

zen der Spannungen \mathbf{S}^I aus Gleichung (6.38) und \mathbf{S}^{II} aus Gleichung (6.39) in die Energiebilanz (6.40) resultiert die Nebenbedingung in der Form

$$(\Delta \mathbf{S}^I + \Delta \mathbf{S}^{II})^T \Delta \mathbf{E} \stackrel{!}{\leq} 0, \quad (6.41)$$

mit den KNN-Ausgangsvektoren $\Delta \mathbf{S}^I = \mathbf{z}(\mathbf{x}^I)$ und $\Delta \mathbf{S}^{II} = \mathbf{z}(\mathbf{x}^{II})$. Diese Ungleichung kann für einen NB-Punkt $\mathbf{x}_k^C \in \Omega$ durch den Fehlerterm

$$g_k = g(\mathbf{x}_k^C, \mathbf{w}) = \frac{1}{\alpha} \sum_{j=1}^{n_s} \left(2m_{zj} + s_{zj} \hat{z}_j(\mathbf{x}_k^I, \mathbf{w}) + s_{zj} \hat{z}_j(\mathbf{x}_k^{II}, \mathbf{w}) \right) (m_{xj} + s_{xj} \hat{x}_{jk}^C) \quad (6.42)$$

festgelegt werden. Dazu sind somit zwei konsekutive KNN-Auswertungen notwendig – je eine für \mathbf{z}^I und für \mathbf{z}^{II} . Der Normalisierungsfaktor α wird analog zur Isotropie-Nebenbedingung mit Gleichung (5.54) definiert. Die Eingangsvariablen

$$\mathbf{x}_k^I = \mathbf{x}_k^C = [\Delta \mathbf{E}, \mathbf{E}_l, \mathbf{S}_l, \mathbf{h}_l] \quad \text{und} \quad (6.43)$$

$$\mathbf{x}_k^{II} = [-\Delta \mathbf{E}, \mathbf{E}_l + \Delta \mathbf{E}, \mathbf{S}_l + \Delta \mathbf{S}^I, \mathbf{h}^I(\Delta \mathbf{S}^I)] \quad (6.44)$$

zeigen die Besonderheit dieser Nebenbedingung auf: \mathbf{x}_k^{II} ist über die Spannungen \mathbf{S}^I und die Geschichtsvariablen \mathbf{h}^I vom KNN-Ausgang $\mathbf{z}(\mathbf{x}_k^I) = \Delta \mathbf{S}^I$ des vorherigen Schritts abhängig. Somit hängt \mathbf{x}_k^{II} indirekt von den Gewichten \mathbf{w} ab. Die zugehörige Ableitung lautet

$$\frac{d\mathbf{x}_k^{II}}{d\mathbf{w}} = \left[\mathbf{0}_{n_w \times n_s} \quad \mathbf{0}_{n_w \times n_s} \quad \left(\frac{\partial \Delta \mathbf{S}^I}{\partial \mathbf{w}} \right)^T \quad \left(\frac{\partial \mathbf{h}^I}{\partial \Delta \mathbf{S}} \frac{\partial \Delta \mathbf{S}^I}{\partial \mathbf{w}} \right)^T \right]^T \in \mathbb{R}^{(3n_s + n_h) \times n_w}. \quad (6.45)$$

Dabei wurde der Anordnung der Verzerrungszinkremente $\Delta \mathbf{E}$ und der Verzerrungen \mathbf{E}_l als die ersten $2n_s$ Komponenten im Eingangsvektor berücksichtigt. Diese Verkettung muss bei der Berechnung des Gradienten

$$\nabla g_k = \sum_{j=1}^{n_s} \left(\beta_{jk}^I \frac{\partial \hat{z}_{jk}^I}{\partial \mathbf{w}} + \beta_{jk}^{II} \frac{\partial \hat{z}_{jk}^{II}}{\partial \mathbf{w}} \right) \quad (6.46)$$

beachtet werden, mit $\hat{z}_j^I := \hat{z}_{jk}(\mathbf{x}_k^I, \mathbf{w})$ und $\hat{z}_{jk}^{II} := \hat{z}_j(\mathbf{x}_k^{II}, \mathbf{w})$ und

$$\beta_{jk}^I := \frac{s_{zj}(m_{xj} + s_{xj}\hat{x}_{jk}^C)}{\alpha} \left(1 + \sum_{m=1}^{n_s} \frac{\partial \hat{z}_{jk}^{II}}{\partial \hat{x}_{2n_s+m}} + \sum_{m=1}^{n_h} \sum_{n=1}^{n_s} \frac{\partial \hat{z}_{jk}^{II}}{\partial \hat{x}_{3n_s+m}} \frac{\partial h_m^I}{\partial \hat{z}_n} \right) \quad (6.47)$$

$$\beta_{jk}^{II} := \frac{s_{zj}(m_{xj} + s_{xj}\hat{x}_{jk}^C)}{\alpha} . \quad (6.48)$$

Die partiellen Ableitungen $\partial \hat{z}_{jk}^I / \partial \hat{x}_i$ und $\partial \hat{z}_{jk}^{II} / \partial \hat{x}_i$ sind jeweils den transformierten JACOBI-Matrizen $\hat{\mathbf{j}}_k^I := \hat{\mathbf{j}}(\mathbf{x}_k^I)$ und $\hat{\mathbf{j}}_k^{II} := \hat{\mathbf{j}}(\mathbf{x}_k^{II})$ zu entnehmen. Mit

$$\frac{\partial h^E}{\partial \tilde{\mathbf{S}}} = \frac{1}{2} \Delta \mathbf{E}^T . \quad (6.49)$$

kann die Ableitung der Geschichtsvariablen $\partial h_m^I / \partial \hat{z}_n$ nach den Spannungen für die in Gleichung (6.7) definierte Beispielvariable berechnet werden. Die Faktoren zur Berechnung des Fehlergradienten $\nabla \mathcal{L}_C$ sind $d_{jk}^I = \beta_{jk}^I \rho_{h_{jk}}$ und $d_{jk}^{II} = \beta_{jk}^{II} \rho_{h_{jk}}$. Zur Einforderung dieser Ungleichheits-Nebenbedingung ist zuvor die Berechnung der aktiven Menge (4.159) der NB-Datenpunkte zu ermitteln.

6.3 Untersuchungen am Materialpunkt für P-KNN

Alle vorgestellten plastischen KNN-Materialmodelle FNN-P-Ink, FNN-P-Seq sowie NCNN-P-Ink und RNN-P sind grundsätzlich für plastisches Materialverhalten geeignet. Sie werden einander im Folgenden gegenübergestellt. Wie in Abschnitt 5.4 liegt auch hier der Fokus auf Untersuchungen in Bezug auf signifikante Unterschiede der einzelnen Herangehensweisen. Für die FNN- und NCNN-Modelle wird in den Zwischenschichten die SoftPlus-Aktivierungsfunktion gewählt. Mit ihr wurden im Rahmen der Parameterstudien bessere Ergebnisse erzielt als mit dem Tangens hyperbolicus. Für die Ausgangsschicht ist die Identitätsfunktion ausreichend. Für das RNN-P-Materialmodell wird der Tangens hyperbolicus als Aktivierungsfunktion gewählt. Die SoftPlus-Aktivierungsfunktion führt beim RNN häufig zum Problem des explodierenden Gradienten (engl.: *exploding gradient problem*), vgl. Abschnitt 4.3.2. Alle KNN-Materialmodelle werden mit dem QN-Verfahren aus Abschnitt 4.4.2 trainiert. Die Untersuchungen erfolgen in MATLAB [170].

Die nachfolgenden Parameterstudien und Detailuntersuchungen werden im ebenen Spannungszustand (ESZ) durchgeführt, mit $n_s = 3$ arbeitswirksamen Verzerungskomponenten $\mathbf{E} = [E_{11}, E_{22}, 2E_{12}]$ und den drei zugehörigen Spannungskomponenten $\mathbf{S} = [S_{11}, S_{22}, S_{12}]$. Dabei lassen sich die in Abschnitt 3.4 vorgestellten dreidimensionalen analytischen Stoffgesetze mit der Projektionsmatrix

$$\mathbf{P}^{\text{ESZ}} = \begin{bmatrix} 2/3 & -1/3 & 0 \\ -1/3 & 2/3 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad (6.50)$$

ohne Weiteres auf den ESZ übertragen, vgl. Gleichung (3.113). Die aus dem ESZ gewonnenen, qualitativen Erkenntnisse zum Verhalten von plastischen KNN-Materialmodellen sind vollständig auf den dreidimensionalen Fall übertragbar und benötigen dafür wesentlich weniger Rechenleistung. Alle KNN werden mit dem QN-Verfahren aus Abschnitt 4.4.2 trainiert.

6.3.1 Zufällige Generierung von Verzerrungspfaden

In diesem Abschnitt wird zunächst die Generierung der Trainings- und Validierungsdatenpaare erläutert. Diese werden jeweils aus den Pfaden definierter Materialpunkt-Versuche berechnet, vgl. Gleichung (6.2). Die Umrechnung in die inkrementellen oder sequenziellen Datenpaare geschieht nach Abschnitt 6.1.2.

Die Materialpunkt-Versuche aus Abschnitt 5.1.1 werden grundsätzlich verzerrungsgesteuert durchgeführt. Eine, zwei oder alle drei Verzerrungskomponenten \mathbf{E}^b werden vorgegeben, während die jeweils anderen Spannungskomponenten $\tilde{\mathbf{S}}^u$ konstant null sind. Zunächst muss ein Verzerrungsraum $\Omega_{\mathbf{E}} \subset \mathbb{R}^3$ definiert werden, mit $\mathbf{E} \in \Omega_{\mathbf{E}}$. Innerhalb dieses Raums werden nachfolgend zwei verschiedene Arten von Pfaden vorgegeben, vgl. Abbildung 6.7:

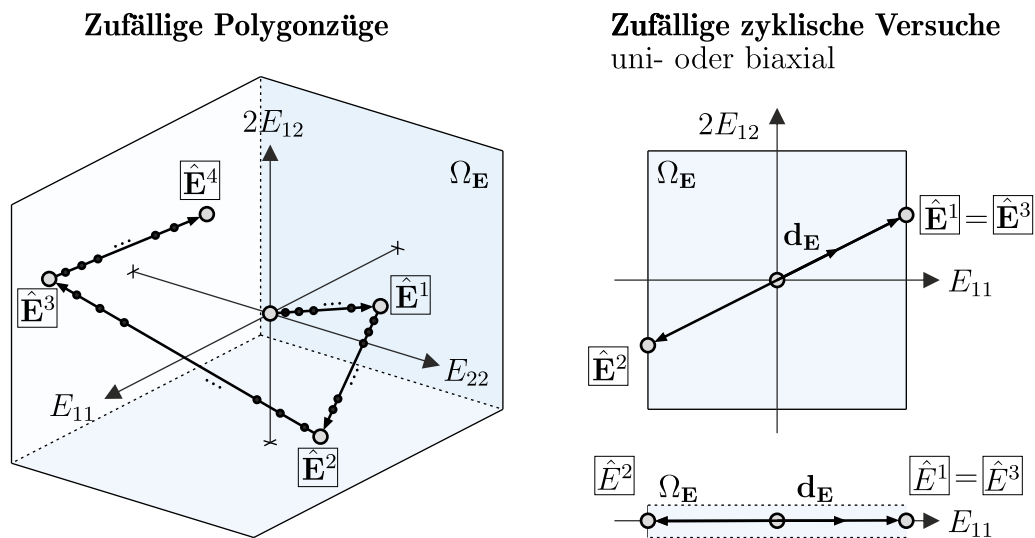


Abbildung 6.7: Definition von Verzerrungspfaden zur Generierung von Trainings- und Validierungsdatenpaaren

- Ein zufälliger Polygonzug, der stets bei $\mathbf{E} = \mathbf{0}$ startet und nacheinander n_{Eck} zufällige Eckpunkte $\{\hat{\mathbf{E}}_1, \dots, \hat{\mathbf{E}}_{n_{\text{Eck}}}\}$ abfährt. Jede lineare Verbindungsstrecke wird mit n_{disk} Zwischenschritten diskretisiert. Dabei sind stets alle n_s Verzerrungskomponenten vorgegeben.

- Ein zufälliger zyklischer Versuch, der uni- oder biaxial sein kann. Durch eine vorgegebene Richtung $\mathbf{d}_{\mathbf{E}}$ werden die Eckpunkte $\{\hat{\mathbf{E}}_1, \hat{\mathbf{E}}_2, \hat{\mathbf{E}}_3 = \hat{\mathbf{E}}_1\}$ auf einer Geraden definiert. Sie liegen stets auf dem durch den Verzerrungsraum $\mathbf{E} \in \Omega_{\mathbf{E}}$ definierten Rand. Die nicht vorgegebenen Verzerrungskomponenten werden über die äußere Gleichgewichtsiteration bestimmt.

Beide Arten der numerischen Materialpunkt-Versuche können am Ende mit oder ohne Entlastung auf $\mathbf{S} = \mathbf{0}$ abgeschlossen werden. Die Entlastung ist dabei stets vollständig spannungsgesteuert, mit $\hat{\mathbf{S}}^u = \mathbf{0}$ (bei entsprechender Inkrementierung). Im dreidimensionalen Fall sei angemerkt, dass eine sinngemäße Datengenerierung mit zufälligen Polygonzügen nicht zweckmäßig ist, da ggf. zu große Volumendehnungen und dementsprechend große hydrostatische Spannungen generiert werden, die im Fall echter Versuchsdaten nicht aufträten.

6.3.2 Funktionsfähigkeit aller KNN-Materialmodelle

Zunächst wird das Verhalten der vier KNN-Materialmodelle FNN-P-Ink, FNN-P-Seq, NCNN-P-Ink und RNN-P in Bezug auf die Anzahl der Trainingsdaten sowie auf die Anzahl der Gewichte n_w untersucht.

Die synthetischen Daten für alle folgenden Untersuchungen werden mit der VON MISES'schen Fließbedingung und kinematischer Verfestigung aus Abschnitt 3.4.5 generiert. Die zugehörigen Materialparameter sind in Tabelle 6.1 gegeben, mit dem E-Modul C , der Querdehnzahl ν , der initialen Fließspannung Y_0 , dem plastischen Verfestigungsmodul C_P sowie dem Parameter θ zur Einstellung der gemischten Verfestigung ($\theta = 0$: reine kinematische Verfestigung, $\theta = 1$: reine isotrope Verfestigung). Die Werte C , ν und Y_0 entsprechen denjenigen von Baustahl, vgl. EUROCODE 3 [35]. Der Verfestigungsmodul C_P wurde etwas größer definiert, um einen merklichen Effekt der Verfestigung zu simulieren.

E-Modul C [kN/cm ²]	Querd. ν [-]	Fließgrenze Y_0 [kN/cm ²]	Verfestigungsmodul C_P [kN/cm ²]	Iso./Kin. θ [-]
21 000	0,3	23,5	$C/50 = 420$	0

Tabelle 6.1: Materialparameter für das analytische VON MISES'sche Materialmodell mit kinematischer Verfestigung für die Parameterstudien

Die Verzerrungen $\mathbf{E} = [E_{11}, E_{22}, 2E_{12}]$ werden auf den Raum

$$\Omega_{\mathbf{E}} = [-0,01; 0,01]^3 \subset \mathbb{R}^3 \quad (6.51)$$

begrenzt. Die Pfade zur Berechnung der Trainings- und Validierungsdatenpaare werden mit n_p Polygonzügen definiert, vgl. Abschnitt 6.3.1. Jeder Polygonzug besteht aus $n_{Eck} = 5$ zufällig in $\Omega_{\mathbf{E}}$ erzeugten Eckpunkten, die mit $n_{\text{disk}} = 30$ Zwischenschritten verbunden sind. Zur Generierung der Validierungsdatenpaare werden stets 30 dieser Pfade zusätzlich erzeugt. Die Umrechnung in die inkrementellen Daten erfolgt für die zwei FNN-Materialmodelle und das NCNN-Materialmodell mit den Indexversätzen $\Delta n \in \{0, 1, \dots, 10\}$. Darüber hinaus wird die Menge aller inkrementellen Trainingsdatenpaare im Anschluss auf 20% gekürzt. Dies hat keinen sichtbaren Effekt auf die Ergebnisse, beschleunigt aber den Trainingsprozess.

Das über n_{MCS} Trainingsdurchläufe gemittelte Fehlermaß $\bar{\mathcal{L}}_V^{\min}(\mathbf{S})$ wird so berechnet, wie es in Abschnitt 5.4.1 für die Untersuchungen der elastischen KNN-Materialmodelle beschrieben wurde, vgl. Gleichungen (5.58) und (5.59). Allerdings basiert die Berechnung des Fehlermaßes $\bar{\mathcal{L}}_V^{\min}(\mathbf{S})$ im Folgenden auf den in Abschnitt 6.3.1 definierten Pfaden, welche mit dem Materialpunkt-Versuch aus Abschnitt 5.1.1 abgefahren werden. Diese Pfade werden pro Untersuchung spezifiziert. Daher kann eine feste Zahl P_V von Validierungsdatenpaaren nicht mehr pauschal vorgegeben werden. Die restlichen Hyperparameter für die Parameterstudien sind in Tabelle 6.2 angegeben.

Trainingsmittel n_{MCS}	Epochenmittel P_V	Abs. Krit. δ^{abs}	Rel. Krit. δ^{rel}	Ende Training epo_{max}
5	abh. von Pfaden	10^{-3}	10^{-3}	10^4

Tabelle 6.2: Statistische Parameter für die Untersuchungen der KNN-P-Modelle

Im linken Diagramm von Abbildung 6.8 ist die Untersuchung der genannten KNN-Materialmodelle in Abhängigkeit der Anzahl von Trainingspfaden n_p dargestellt. Dabei wurde für die FNN-Modelle eine Topologie mit drei Zwischenschichten à 20 Neuronen gewählt. Dies entspricht $n_w = 1166$ bzw. $n_w = 1061$ Gewichten. Für das RNN-Modell waren es drei Zwischenschichten à 13 Neuronen, folglich $n_w = 1004$ Gewichte. Im rechten Diagramm von Abbildung 6.8 ist das Fehlermaß $\bar{\mathcal{L}}_V^{\min}(\mathbf{S})$ über die Anzahl der Gewichte n_w aufgetragen. Dabei wurden stets $n_p = 50$ Trainingspfade erzeugt.

Im Vergleich zu den Untersuchungen zu elastischen KNN-Materialmodellen aus Abschnitt 5.4 fällt eines auf: Die Fehlerwerte sind deutlich höher. Die von den FNN-Materialmodellen erreichten besten Fehlerwerte liegen bei $\bar{\mathcal{L}}_V^{\min}(\mathbf{S}) \approx Y_0/100$, um die Werte in ein interpretierbares Verhältnis zu setzen. Dabei wurde mit den Polygonzügen eine vergleichsweise gute Verteilung der Trainingsdaten ermöglicht.

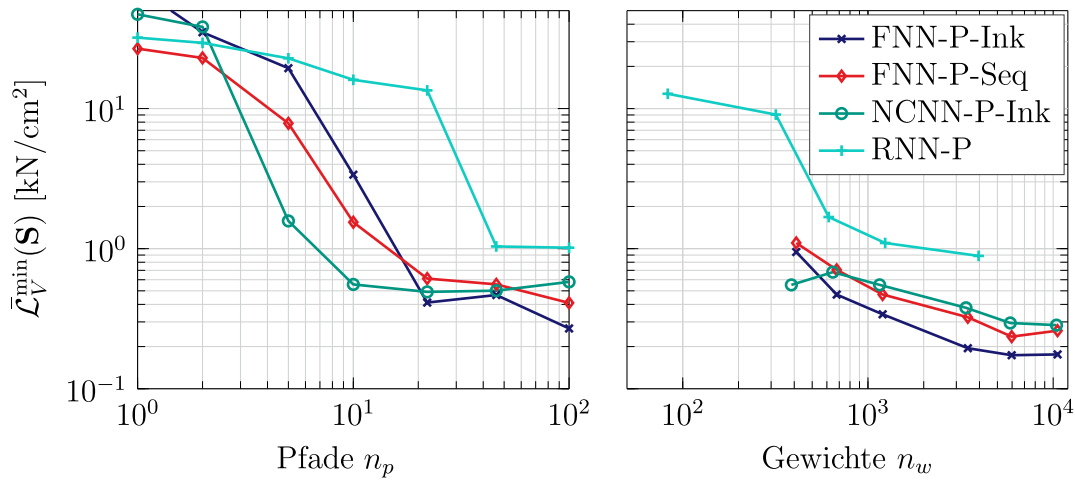


Abbildung 6.8: Verhalten der vier plastischen KNN-Materialmodelle in Abhängigkeit der Anzahl der Trainingspfade (links) sowie das Verhalten in Abhängigkeit der Anzahl der Gewichte (rechts)

Dies macht deutlich, dass pfadabhängiges Materialverhalten wesentlich mehr Daten und Freiheitsgrade benötigt.

Weiter zeigt das NCNN-Materialmodell bessere Konvergenzeigenschaften im Fall weniger Daten als die FNN-Modelle. Dieser Vorteil wird allerdings schnell wieder ausgeglichen. Darüber hinaus ist es fraglich, ob die Existenz eines Potentials (6.19) immer gegeben ist. Dies ist bei Materialien, die üblicherweise mit nicht-assoziierten Fließregeln modelliert werden, nicht zu erwarten. Das NCNN-Materialmodell wird im Rahmen von plastischen Materialverhalten nachfolgend nicht weiter berücksichtigt.

6.3.3 Diskussion zum RNN-Materialmodell

In den Untersuchungen aus Abschnitt 6.3.2 hat sich gezeigt, dass das RNN bei vergleichbarer Anzahl von Freiheitsgraden und Trainingsdaten zu schlechteren Resultaten in Bezug auf den hier definierten Fehler führt. Noch deutlicher wird dieser Umstand, wenn der Validierungsfehler nicht auf Pfaden ermittelt wird, deren Verzerrungskomponenten vollständig vorgegeben sind. Die restlichen Komponenten müssen über die äußere Gleichgewichtsiteration mitberechnet werden. Daraus resultieren stärkere Abweichungen in den Pfaden, was beim RNN erfahrungsgemäß zu größeren Problemen führt.

Darüber hinaus kommt beim RNN erschwerend hinzu, dass die Trainingsdaten nur mit den auskonvergierten GGW-Punkten (\mathbf{E}, \mathbf{S}) berechnet werden. Innerhalb der NEWTON-Iteration sind somit Pfad-Schritte verlangt, die das RNN nie gelernt

haben kann. Daher empfehlen bspw. GHAVAMIAN und SIMONE [60] die nicht auskonvergierten Zwischenschritte des NEWTON-Verfahrens ebenfalls im Trainingsprozess mitzubersichtigen, was aus offensichtlichen Gründen nur im Rahmen einer Mehrskalenmodellierung möglich ist.

Aus diesen und anderen Gründen benötigt das RNN wesentlich mehr Trainingsdaten als das FNN, um eine vergleichbare Leistung in der Simulation zu zeigen. Daher wird an dieser Stelle das RNN ebenfalls nicht weiter untersucht. Nach Auffassung des Autors sind RNN prinzipiell geeignet zur Modellierung plastischen Materialverhaltens. Allerdings stehen die eigentlichen Vorteile des RNN in der vorliegenden Form und im Rahmen dieser Arbeit nicht heraus. Dies ist nämlich die Identifikation von Geschichtsabhängigkeit, ohne diese explizit im Eingangsvektor definieren zu müssen.

6.3.4 Diskussion zur KNN-Materialtangente

Im Rahmen der klassischen Elastoplastizitätstheorie wird zwischen der analytischen Kontinuumtangente $\mathbf{C}_T^{\text{kont}}$ und der algorithmischen Tangente $\mathbf{C}_T^{\text{alg}}$ unterschieden, vgl. Abschnitt 3.4. Außerdem trennt die Fließbedingung (3.90) scharf zwischen einem elastischen und einem plastischen Bereich. Die Spannungen sind beim Übergang zwischen den beiden Bereichen zwar stetig, aber nicht mehr stetig differenzierbar. Im Gegensatz dazu ist die KNN-Abbildungsvorschrift $\mathbf{z} = \mathbf{f}^{\text{KNN}}(\mathbf{x}, \mathbf{w})$ durch die Wahl der glatten Aktivierungsfunktionen eine zweimal stetig differenzierbare Funktion. Die zugehörige Materialtangente, innerhalb des vorliegenden Abschnitts mit $\mathbf{C}_T^{\text{KNN}}$ bezeichnet, ist folglich ebenfalls stetig. Außerdem gibt es keine Unterscheidung der Tangenten infolge eines Projektionsalgorithmus.

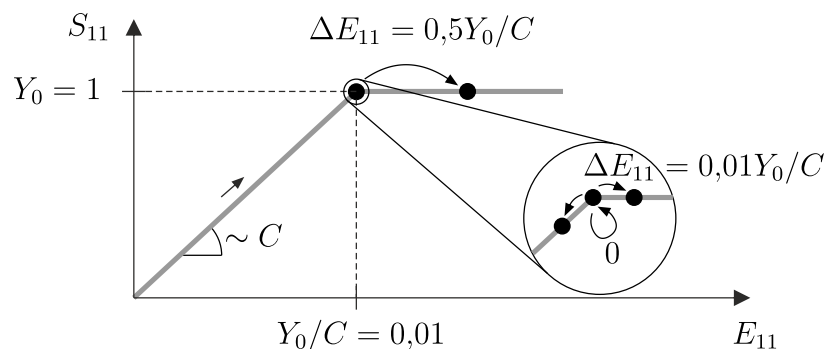


Abbildung 6.9: Drei Verzerrungsschritte $\Delta E_{11} > 0$ und einer mit $\Delta E_{11} = 0$ ausgehend vom GGW-Punkt $(\mathbf{E}, \mathbf{S}) = ([Y_0/C, 0, 0], [Y_0, 0, 0])$

Um diesen Sachverhalt näher zu beleuchten, werden in diesem Abschnitt die

drei genannten Materialtangente anhand eines einfachen Beispiels einander gegenübergestellt. Gegeben sei ein elastoplastisches Materialmodell mit VON MISES'scher Fließbedingung und ohne Verfestigung, vgl. Abschnitt 3.4.4. Die Materialparameter sind der E -Modul $C = 100$, die Querdehnzahl $\nu = 0$ und die Fließspannung $Y_0 = 1$. Die Einheiten sind für die folgende Betrachtung nicht relevant und werden daher nicht beigefügt. Daraus folgen die elastische Materialtangente \mathbf{C} und die Kontinuumstangente während des Plastifizierens⁷⁵ $\mathbf{C}_T^{\text{kont}}$ im ESZ zu

$$\mathbf{C} = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 100 & 0 \\ 0 & 0 & 50 \end{bmatrix} \quad \text{und} \quad \mathbf{C}_T^{\text{kont}} = \begin{bmatrix} 20 & 40 & 0 \\ 40 & 80 & 0 \\ 0 & 0 & 50 \end{bmatrix}. \quad (6.52)$$

Mit besagtem analytischen Materialmodell wird ein FNN für das FNN-P-Exp-Ink-Materialmodell nach Abschnitt 6.1.2 trainiert. Die Anzahl der Datenpunkte und Gewichte wurde dabei so groß gewählt, dass das KNN-Materialmodell das gewünschte Materialverhalten gut nachbilden kann. Die jeweilige genaue Anzahl ist hier im Detail nicht essenziell.

Ausgehend vom GGW-Punkt $(\mathbf{E}, \mathbf{S}) = ([Y_0/C, 0, 0], [Y_0, 0, 0])$ werden vier verschiedene Verzerrungsschritte $\Delta \mathbf{E} = [\Delta E_{11}, -\Delta E_{11}/2, 0]$ vorgegeben: ein kürzerer Entlastungsschritt mit $\Delta E_{11} = -0,01 Y_0/C$, ein Nullschritt mit $\Delta E_{11} = 0$, ein kürzerer Belastungsschritt mit $\Delta E_{11} = +0,01 Y_0/C$ und ein größerer Belastungsschritt mit $\Delta E_{11} = +0,5 Y_0/C$. Diese sind ebenfalls in Abbildung 6.9 dargestellt. Zunächst wird die algorithmische Tangente des analytischen Materialmodells $\mathbf{C}_T^{\text{alg}}$ mit der des KNN-Materialmodells $\mathbf{C}_T^{\text{KNN}}$ im Fall des größeren Belastungsschritts verglichen:

$$\mathbf{C}_T^{\text{alg}} = \begin{bmatrix} 15,4 & 30,8 & 0 \\ 30,8 & 61,5 & 0 \\ 0 & 0 & 28,6 \end{bmatrix} \quad \text{und} \quad \mathbf{C}_T^{\text{KNN}} = \begin{bmatrix} 16,3 & 33,4 & 0,6 \\ 31,8 & 68,4 & -0,6 \\ 0,1 & -0,2 & 32,4 \end{bmatrix}. \quad (6.53)$$

Aus der Ähnlichkeit beider Tangenten lässt sich schließen, dass die KNN-Materialtangente ebenfalls die Schrittweite mitberücksichtigt. Das ist ersichtlich, da das Verzerrungsinkrement $\Delta \mathbf{E}$ Teil des KNN-Eingangsvektors \mathbf{x} ist.

In Tabelle 6.3 sind die beiden Tangente für die restlichen drei Verzerrungsschritte einander gegenübergestellt. Für $\Delta E_{11} = 0$ entspricht die algorithmische Tangente $\mathbf{C}_T^{\text{alg}}$ der elastischen Tangente \mathbf{C} , weil die *neutrale Belastung* (auf der Fließfläche) dem elastischen Bereich zugewiesen wird, vgl. Tafel 3.1. Sobald $\Delta E_{11} > 0$ gilt, *springt* die algorithmische Tangente $\mathbf{C}_T^{\text{alg}}$ auf diejenige des plastischen Bereichs

⁷⁵ Der deutlich zu erkennende Rangabfall der Matrix $\mathbf{C}_T^{\text{kont}}$ resultiert aus der Annahme idealer Plastizität. Um kinematische Gesamtsysteme in der FEM zu vermeiden, kann eine kleine numerische Verfestigung berücksichtigt werden.

ΔE_{11}	$-0,01 Y_0/C$	0	$+0,01 Y_0/C$
$\mathbf{C}_T^{\text{alg}}$	$\begin{bmatrix} 100 & 0 & 0 \\ 0 & 100 & 0 \\ 0 & 0 & 50 \end{bmatrix}$	$\begin{bmatrix} 100 & 0 & 0 \\ 0 & 100 & 0 \\ 0 & 0 & 50 \end{bmatrix}$	$\begin{bmatrix} 19,9 & 39,8 & 0 \\ 39,8 & 79,5 & 0 \\ 0 & 0 & 49,3 \end{bmatrix}$
$\mathbf{C}_T^{\text{KNN}}$	$\begin{bmatrix} 81,2 & 9,6 & -0,1 \\ 10,1 & 95,5 & -0,3 \\ 1,4 & -0,6 & 50,2 \end{bmatrix}$	$\begin{bmatrix} 66,8 & 16,7 & -0,1 \\ 18,4 & 91,1 & -0,3 \\ 2,2 & 1,1 & 50,1 \end{bmatrix}$	$\begin{bmatrix} 47,9 & 26,1 & -0,1 \\ 28,9 & 85,6 & -0,3 \\ 2,8 & -1,4 & 49,8 \end{bmatrix}$

Tabelle 6.3: Vergleich der algorithmischen Materialtangente $\mathbf{C}_T^{\text{alg}}$ und der KNN-Materialtangente $\mathbf{C}_T^{\text{KNN}}$ bei verschiedenen Verzerrungsschritten

und ist daraufhin von der Schrittweite abhängig. Für $\Delta E_{11} \rightarrow 0$ gilt $\mathbf{C}_T^{\text{alg}} \rightarrow \mathbf{C}_T^{\text{kont}}$. Dieses Verhalten ist bei der KNN-Materialtangente $\mathbf{C}_T^{\text{KNN}}$ nicht zu beobachten. Vielmehr ist durch die weiter oben beschriebene stetige Differenzierbarkeit ein Übergangsbereich zu erkennen, in dem $\mathbf{C}_T^{\text{KNN}}$ vom elastischen zum plastischen Bereich übergeht. Für $\Delta E_{11} = 0$ ist daher zu erwarten, dass $\mathbf{C}_T^{\text{KNN}}$ gegen

$$\frac{1}{2}(\mathbf{C} + \mathbf{C}_T^{\text{kont}}) = \begin{bmatrix} 60 & 20 & 0 \\ 20 & 90 & 0 \\ 0 & 0 & 50 \end{bmatrix} \quad (6.54)$$

konvergiert, was sich beim Vergleich der Zahlwerte zu bestätigen scheint.

6.3.5 Einfluss der physikalischen Nebenbedingungen

Nachfolgend wird der Einfluss der in Abschnitt 6.2 beschriebenen Nebenbedingungen (NB) für plastisches Materialverhalten untersucht. Die Stationaritäts-NB aus Abschnitt 6.2.3 wird dabei ausgeklammert. Durch das explizite Vorgeben der Komponenten der Materialtangente verhält sich diese NB vergleichsweise restriktiv, was eine vorsichtiger Wahl des Penalty-Parameters erfordert. Die vier NB werden über die klassische Penalty-Methode eingefordert, vgl. Abschnitt 4.5.1. Dabei werden für die Normalisierungs-NB $P_C = 500$ NB-Punkte und für die restlichen drei NB jeweils 5000 NB-Punkte mit der Methode der konvexen Hülle aus Abschnitt 6.2.1 generiert. Die Studie wird mit den in Tabelle 6.2 gegebenen Hyperparametern für den Trainingsprozess und die Berechnung des Validierungsfehlers $\bar{\mathcal{L}}_V^{\min}(\mathbf{S})$ durchgeführt, welcher mittels Materialpunkt-Versuchen aus Abschnitt 5.1.1 berechnet wird.

In Abbildung 6.10 ist die Parameterstudie bzgl. des Penalty-Parameters ϵ abgebildet. Dabei wurden die KNN der Topologie [9-25-25-25-3] stets mit sechs

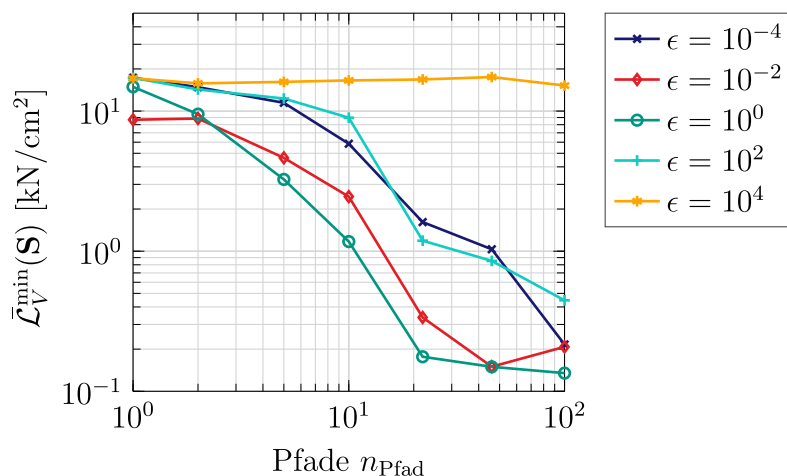


Abbildung 6.10: Validierungsfehler $\bar{\mathcal{L}}_V^{\min}(\mathbf{S})$ in Abhängigkeit der Anzahl biaxialer Trainingspfade n_p sowie des Penalty-Parameters ϵ für die Nebenbedingungen

uniaxialen zyklischen Versuchen sowie n_p zufällig erzeugten biaxialen zyklischen Versuchen trainiert, vgl. Abschnitt 6.3.1. Die Berechnung der inkrementellen Trainings- und Validierungsdatenpaare erfolgt analog zu Abschnitt 6.3.2. Für $\epsilon = 1$ und $\epsilon = 100$ lassen sich die Werte von $\bar{\mathcal{L}}_V^{\min}(\mathbf{S})$ um bis zu eine 10er-Potenz verbessern. Dies lässt sich vor allem dadurch erklären, dass die Datenbasis mit den festen, uniaxialen zyklischen Versuchen ohne Hinzuziehen vieler zusätzlicher, zufälligen biaxialen Versuche zu wenig ist, um das KNN ausreichend gut zu trainieren. Die NB regularisieren den Trainingsprozess und füllen die Bereiche des Trainingsraums Ω ohne Datenpunkte mit Information. In Abbildung 6.11 sind die Verläufe der einzelnen Fehlerterme \mathcal{L} der Gleichheits-NB während des Trainingsprozesses für $n_p = 50$ dargestellt. Dabei wird einmal der Fall ohne NB und einmal mit NB ($\epsilon = 1$) betrachtet. Darüber hinaus ist die Anzahl der aktiven NB-Punkte $|\mathbb{I}_a|$ für die beiden Ungleichheits-NB abgebildet, vgl. Definition (4.159).

Auch ohne Einforderung der zugehörigen NB sinkt der Normalisierungsfehler parallel zum Trainingsdatenfehler. Das ist schlüssig, da die zugehörigen NB-Terme denen der Trainingsdaten ähnlich sind und auch als solche interpretiert werden können, vgl. Abschnitt 6.2.2. Allerdings kann der Fehlerwert mit Einforderung der NB um fast zwei 10er-Potenzen verringert werden. Der Fehlerterm bzgl. der Symmetrie der Materialtangente, vgl. Abschnitt 6.2.4, steigt im Laufe des Trainings ohne NB immer weiter an. Für das Optimierungsverfahren ist es scheinbar effizienter, die Spannungsantwort mit einer unsymmetrischen Tangente zu approximieren. Erst durch Einforderung der Symmetrie als NB wird dieser Fehlerterm kleiner, und das um vier 10er-Potenzen. Dabei sollte beachtet werden, dass sich der Daten-Fehlerterm dadurch kaum ändert. Diese Beobachtung wird in Abschnitt 7.2 nochmal aufgegriffen und an echten Versuchsdaten diskutiert.

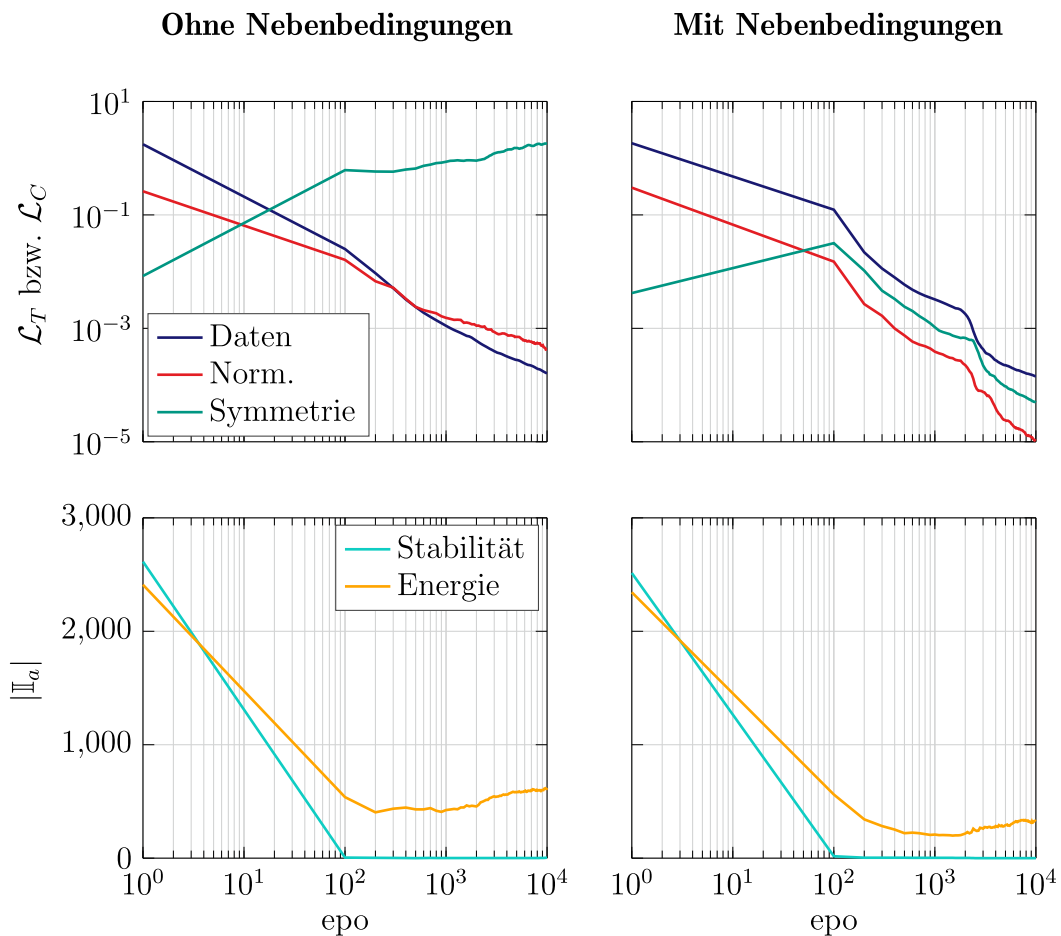


Abbildung 6.11: Fehlerterme bzgl. Datenpaaren für NB der Normalisierung und NB der Tangentensymmetrie (oben) und Anzahl der aktiven Punkte für die Ungleichheits-NB der materiellen Stabilität und dem Ausschluss der Energieproduktion (unten)

Bemerkenswert ist das Verhalten der Anzahl der aktiven NB-Punkte für die Ungleichheits-NB der Energiedissipation nach Abschnitt 6.2.6. Während sich der zugehörige Fehler (nicht abgebildet) ähnlich verhält wie derjenige der Symmetrie-NB, ändert sich die Anzahl der NB-Punkte, auf denen er berechnet wird, kaum. Das liegt womöglich daran, dass die Verletzung dieser NB meist von geringem Ausmaß ist. Trotz der im Vergleich zur Stabilitäts-NB hohen Anzahl an aktiven NB-Punkten ist der Fehlerterm vergleichsweise gering. Für Ungleichheits-NB wird daher empfohlen, einen höheren Penalty-Parameter zu wählen.

6.3.6 Diskussion zur Schrittweitenabhängigkeit

Die hier vorgestellten KNN-Materialmodelle sind abhängig von der Lastschrittweite. Dies ist nicht gleichzusetzen mit Ratenabhängigkeit. Beim KNN-Mate-

rialmodell FNN-P-Ink geht die Schrittweite $\Delta \mathbf{E}$ direkt in die Berechnung des Spannungsinkrements $\Delta \mathbf{S}$ und damit in die Berechnung der Spannungen \mathbf{S} ein. In Bezug auf große Lastschritte sollte daher nach Möglichkeit der Raum aller Verzerrungsinkremente, die in der späteren Anwendung zu erwarten sind, im Trainingsprozess berücksichtigt werden.

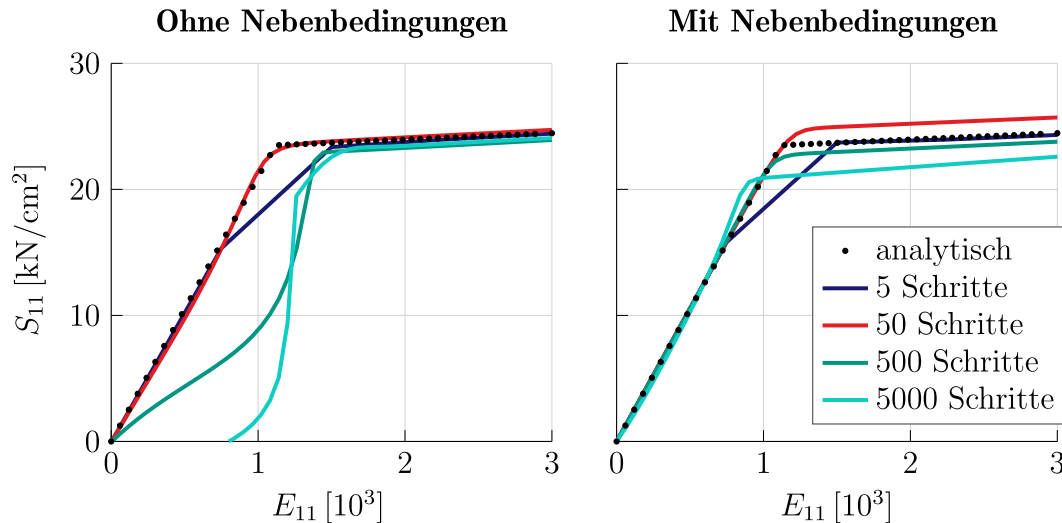


Abbildung 6.12: Uniaxialer Zugversuch mit verschiedenen Schrittweiten. Die KNN sind diejenigen aus Abbildung 6.11.

Werden die Verzerrungsschritte $\Delta \mathbf{E}$ allerdings beliebig klein, tritt ein anderes Phänomen auf: Approximationsfehler bei der Spannungsberechnung können sich akkumulieren und letztlich zur Divergenz der Berechnung führen. Eine Verfeinerung der Lastschrittweite führt daher nicht zwangsläufig zu einer Verbesserung der Ergebnisse. In Abbildung 6.12 ist dieser Umstand am Beispiel der beiden KNN-Materialmodelle verdeutlicht, deren Trainingsprozesse in Abbildung 6.11 dargestellt sind. Die Normalisierungs-NB schwächt diesen Effekt deutlich ab, beseitigt ihn aber nicht gänzlich. Dieses Verhalten steht in Widerspruch zu einem grundlegenden numerischen Selbstverständnis, nämlich dass mit feinerer Diskretisierung eine Näherungslösung besser werden sollte⁷⁶.

⁷⁶ In der klassischen Numerik treten solche Effekte ebenfalls auf, haben dann allerdings etwas mit der Darstellungsgenauigkeit von Fließkommazahlen zutun: Die Berechnung einer numerischen Tangente mittels Differenzenquotienten ist bspw. nicht beliebig genau möglich.

6.4 Approximation einer Fließfläche

Bei der plastischen KNN-Materialmodellierung ist die Definition einer Fließfläche nicht notwendig, geschweige denn die Definition von Verfestigungsregeln, welche die Änderung der Fließfläche während des Fließens beschreiben. Zum besseren Verständnis des den Daten zugrunde liegenden Materialverhaltens kann es trotzdem zweckmäßig sein, mithilfe eines trainierten KNN-Materialmodells eine Quasi-Fließfläche zu approximieren. Dies wird im Folgenden beschrieben und anhand eines Beispiels im ESZ demonstriert.

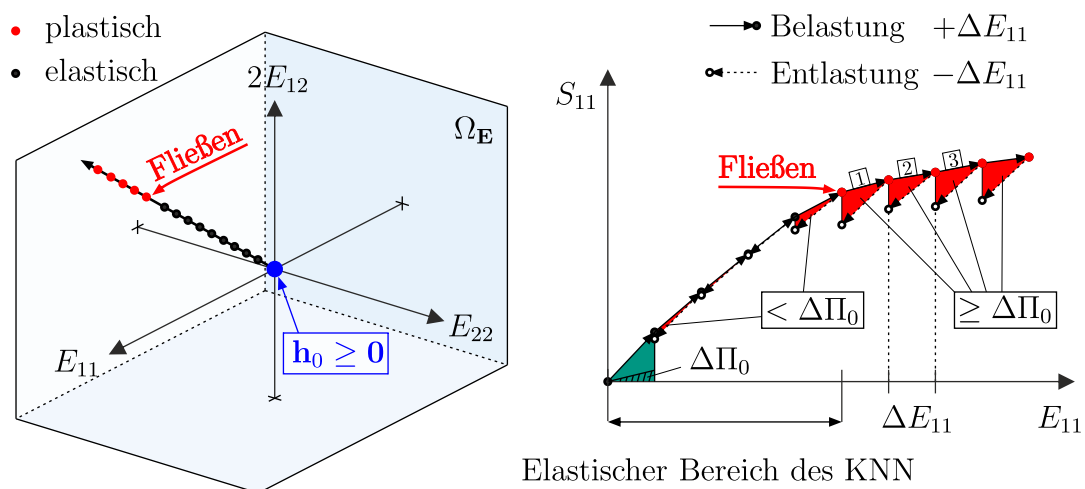


Abbildung 6.13: Approximation einer KNN-Fließfläche: Identifikation eines Schwellenwerts auf einem beliebigen Verzerrungspfad (links) und vorgeschlagene Identifikationsregel im Fall eines uniaxialen Zugversuchs (rechts)

Dazu ist es zunächst notwendig, den Begriff *Fließen* rein auf Basis vorhandener Information über Verzerrungen und Spannungen quantitativ zu erfassen. Gegeben sei ein vorgegebener Pfad $\{\mathbf{E}_0, \dots, \mathbf{E}_N\}$ im Verzerrungsraum $\Omega_{\mathbf{E}}$. Gesucht ist der Punkt \mathbf{E}_Y dieses Pfades, der eine im Folgenden definierte, datenbasierte Fließbedingung verletzt. Die zugehörige Spannung \mathbf{S}_Y kann anschließend im Spannungsraum $\Omega_{\mathbf{S}}$ eingetragen werden. Nach Auswertung mehrerer solcher Pfade lässt sich eine Punktwolke erzeugen, die als Approximation der Fließfläche dient. Die Visualisierung eines solchen Verzerrungspfades sowie die nachfolgend vorgestellte Regel zur Identifikation besagten Fließens ist in Abbildung 6.13 dargestellt.

Zur Definition von *Fließen* wird ein Kriterium herangezogen, welches auf der Energiebilanz $\Delta \Pi$ konsekutiver Be- und Entlastungsschritte basiert. Vom GGW-Punkt $(\mathbf{E}_n, \mathbf{S}_n)$ wird ein Schritt mit $\Delta \mathbf{E}$ vorwärts zu \mathbf{E}_{n+1} durchgeführt. Die zugehörigen Spannungen \mathbf{S}^I werden berechnet. Darauf folgt ein Entlastungsschritt mit $-\Delta \mathbf{E}$, um im Anschluss die Spannungen \mathbf{S}^{II} zu erhalten. Mit Gleichung (6.40)

kann die Energiebilanz $\Delta\Pi$ dieses geschlossenen Verzerrungspfades aufgestellt werden. Anschließend wird $\Delta\Pi$ einem Referenzwert $\Delta\Pi_0$ gegenübergestellt. Gilt $\Delta\Pi \geq \Delta\Pi_0$, tritt per Definition der Fall des Fließens ein. Um etwaige Streuungen zu glätten, wird überdies vorausgesetzt, dass die Bedingung $\Delta\Pi \geq \Delta\Pi_0$ drei mal hintereinander erfüllt sein muss. Der zugehörige erste Punkt entspricht in diesem Fall der Fließgrenze. Dieses Vorgehen ist in Abbildung 6.13 rechts am Beispiel eines uniaxialen Zugversuchs dargestellt. Im Rahmen des nachfolgenden Beispiels wird $\Delta\Pi_0$ als 20% der im ersten Lastschritt geleisteten Spannungsarbeit $\mathbf{S}_1^T \Delta\mathbf{E}/2$ definiert. Diese Werte haben sich in numerischen Untersuchungen als brauchbar herausgestellt, sind aber prinzipiell vom Anwender zu wählen.

Ziel dieses Abschnitts ist die Rückrechnung der VON MISES'schen Fließfläche durch ein KNN-Materialmodell, welches mit zugehörigen Daten trainiert wurde. Darüber hinaus soll das selbstähnliche Aufblähen der Fließfläche infolge isotroper Verfestigung nachgebildet werden. Dazu werden die Materialparameter für das zugehörige Materialmodell aus Abschnitt 3.4.5 nach Tabelle 6.4 gewählt. Zur besseren Sichtbarkeit der isotropen Verfestigung wurde der Wert des Verfestigungsmoduls $C_P = C/10$, verglichen zum realistischen Verhalten von Baustahl, groß gewählt.

E-Modul	Querd.	Fließgrenze	Verfestigungsmodul	Iso./Kin.
C [kN/cm ²]	ν [-]	Y_0 [kN/cm ²]	C_P [kN/cm ²]	θ [-]
21 000	0,3	23,5	$C/10 = 2100$	1

Tabelle 6.4: Materialparameter für das analytische VON MISES'sche Materialmodell mit isotroper Verfestigung zur Rückrechnung der Fließfläche

Der Verzerrungsraum wird zu $\Omega_{\mathbf{E}} = [-0,003; 0,003]^3$ festgelegt. Als synthetische Datengrundlage dienen ein uniaxialer Zugversuch, ein uniaxialer Druckversuch sowie zugehörige biaxiale Zug- und Druckversuche, die mittels Materialpunkt-Versuch aus Abschnitt 5.1.1 berechnet wurden. Diese sind in Abbildung 6.14 mit ihren jeweiligen GGW-Punkten dargestellt. Insgesamt stehen somit 168 (\mathbf{E}, \mathbf{S}) -Paare zur Verfügung. Beim Richtungswechsel werden hier die GGW-Punkte doppelt gezählt.

Durch die rein isotrope Verfestigung ist das Materialverhalten insgesamt isotrop. Daher kann die in Abschnitt 5.1.2 beschriebene Anreicherung der Trainingsdaten durchgeführt werden. Die Rotationsmatrizen

$$\mathbf{Q} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad (6.55)$$

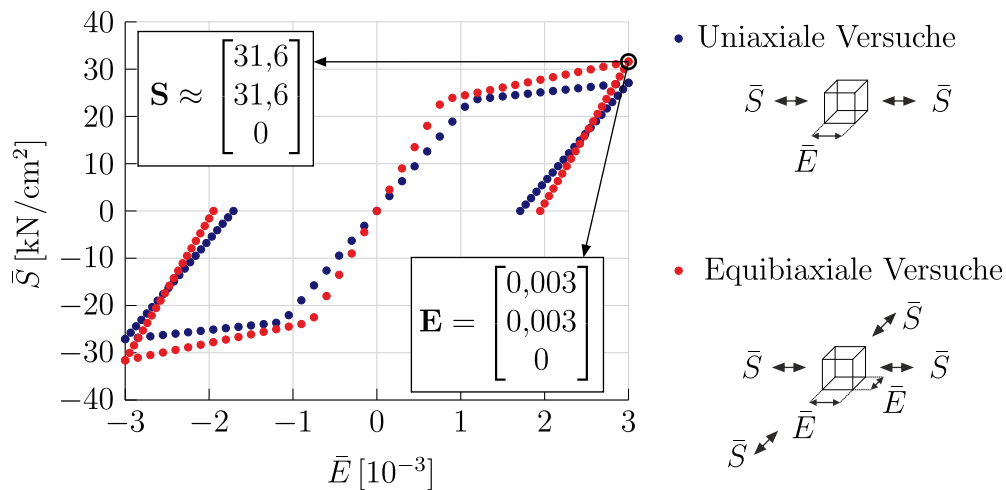


Abbildung 6.14: Uni- und equibiaxiale numerische Versuchsdaten zur Rückrechnung der von MISES'schen Fließfläche

hängen im ESZ⁷⁷ nur vom Rotationswinkel $\theta \in [0, 2\pi)$ ab. Mit 20 zufällig generierten Winkeln θ werden die vier vorgegebenen Pfade auf insgesamt 84 erweitert, mit 3528 GGW-Punkten in Summe. Dabei ist darauf zu achten, dass ein Pfad als Ganzes mit einem festen Winkel rotiert wird, um die korrekten Inkremente zum Training des KNN bilden zu können.

Die Umwandlung der GGW-Punkte in die Trainingsdatenpaare erfolgt für das FNN-P-Ink-Materialmodell nach Abschnitt 6.1.2 mit

$$\mathbf{x} = [\Delta \mathbf{E}, \mathbf{S}_l, h_l^E]^T \in \mathbb{R}^7 \quad \text{und} \quad \mathbf{z} = \Delta \mathbf{S} \in \mathbb{R}^3. \quad (6.56)$$

Dabei werden die Verzerrungen \mathbf{E}_l nicht als Eingangsvariablen berücksichtigt. Dies ist im Fall reiner isotroper Verfestigung zulässig. Die energiebasierte Geschichtsvariable $h^E(t)$ wurde in Abschnitt 6.1.1 vorgestellt. Sie soll das isotrope Verfestigungsverhalten abbilden. Nach der Definition der Eingangs- und Ausgangsvariablen werden die (\mathbf{E}, \mathbf{S}) -Pfade zunächst um die zusätzliche Geschichtsvariable erweitert, vgl. Gleichung (6.10), um anschließend mit Gleichung (6.14) die inkrementellen Trainingsdatenpunkte zu erzeugen. Mit $\Delta n \in \{0, 1, 2, 3, 4\}$ werden so insgesamt 15 960 Trainingsdatenpaare generiert, welche im Anschluss zufällig auf 10 000 gekürzt werden. Dabei sei an dieser Stelle nochmal erwähnt, dass nach Bildung der Inkremente die einzelnen Trainingsdatenpaare unabhängig voneinander sind; sie sind nicht mehr an die zugehörigen Pfade gebunden. Das

⁷⁷ Streng genommen fehlen in \mathbf{Q} die Komponenten in die 3-Richtung, mit $Q_{33} = 1$ und $Q_{3i} = Q_{i3} = 0, i = 1, 2$. Da die Dickenverzerrungen $E_{33} \neq 0$ allerdings nicht bekannt sind, bietet es sich an, hier diese Version von \mathbf{Q} in Verbindung mit den 2×2 -Matrizen \mathbf{E} und \mathbf{S} zu verwenden.

Nebenbed.?	Mit. $\mathcal{L}_Y(\mathbf{S})$	Std. $\mathcal{L}_Y(\mathbf{S})$	Min. $\mathcal{L}_Y(\mathbf{S})$	Max. $\mathcal{L}_Y(\mathbf{S})$
nein	6,97 kN/cm ²	3,24 kN/cm ²	2,75 kN/cm ²	14,29 kN/cm ²
ja	1,49 kN/cm ²	0,38 kN/cm ²	0,46 kN/cm ²	2,47 kN/cm ²

Tabelle 6.5: Der über $n_{\text{MCS}} = 15$ Trainingsprozesse gemittelte Fehler $\mathcal{L}_Y(\mathbf{S})$ (Mit.), die zugehörige Standardabweichung (Std.) sowie der jeweilige kleinste (Min.) und größte (Max.) Wert für den Fall mit Nebenbedingungen und den Fall ohne Nebenbedingungen

Training wird stets nach 10 000 Epochen beendet. Validierungsdatenpaare zum Abbruch des Trainings werden in diesem Beispiel nicht genutzt⁷⁸.

Im Folgenden wird die Approximation der Fließfläche mit Berücksichtigung der physikalischen Nebenbedingungen mit derjenigen ohne Nebenbedingungen verglichen. Alle fünf Nebenbedingungen aus Abschnitt 6.2 werden mit der klassischen L2-Penalty-Methode mit $\epsilon = 1$ berücksichtigt. Für die Normalisierungs- und Stationaritäts-Nebenbedingung werden jeweils $P_C = 200$ NB-Punkte generiert, für die anderen drei Nebenbedingungen jeweils $P_C = 5000$. Als Fehlermaß eines einzelnen Trainingsprozesses wird die über 50 Verzerrungspfade definierte, mittlere Abweichung

$$\mathcal{L}_Y(\mathbf{S}) := \frac{1}{50} \sum_{s=1}^{50} \left\| \mathbf{S}_Y^{\text{KNN}} - \mathbf{S}_Y^{\text{von Mises}} \right\| \quad (6.57)$$

der KNN-Fließzustände $\mathbf{S}_Y^{\text{KNN}}$ von denjenigen der initialen VON MISES'schen Fließfläche $\mathbf{S}_Y^{\text{von Mises}}$ festgelegt. Jene analytische Fließfläche lässt sich im ESZ als Ellipse

$$F(S_1, S_2) = S_1^2 + S_2^2 - S_1 S_2 - Y_0^2 = 0 \quad (6.58)$$

im Hauptspannungsraum darstellen, vgl. Abbildung 3.7. In Tabelle 6.5 sind der über 15 Trainingsprozesse gemittelte Fehler $\mathcal{L}_Y(\mathbf{S})$ (Mit.), die zugehörige Standardabweichung (Std.) sowie der jeweilige kleinste (Min.) und größte (Max.) Wert für den Fall mit Nebenbedingungen und den Fall ohne Nebenbedingungen angegeben. Dabei ist eine deutliche Verbesserung bei Nutzung der genannten Nebenbedingungen zu erkennen, und zwar sowohl beim mittleren Fehlerniveau als auch bei der Streuung. Die zurückgerechneten Fließflächen der jeweiligen besten Realisationen der 15 Trainingsprozesse sind in Abbildung 6.15 dargestellt.

Neben der initialen Fließfläche mit $h_0^E = 0$ kN/cm² kann das KNN ebenfalls erfolgreich das selbstähnliche Aufblähen der Fließfläche approximieren. Die Fließflächen unter Berücksichtigung der Verfestigung können mittels KNN simuliert werden, indem der Anfangswert für die Geschichtsvariable $h_0^E > 0$ gewählt wird. Da diese Geschichtsvariable bei rein elastischer Be- und Entlastung vollständig

⁷⁸ Validierungsdatenpaare hätten hier auch keinen positiven Effekt. Dies hat sich in Voruntersuchungen ergeben.

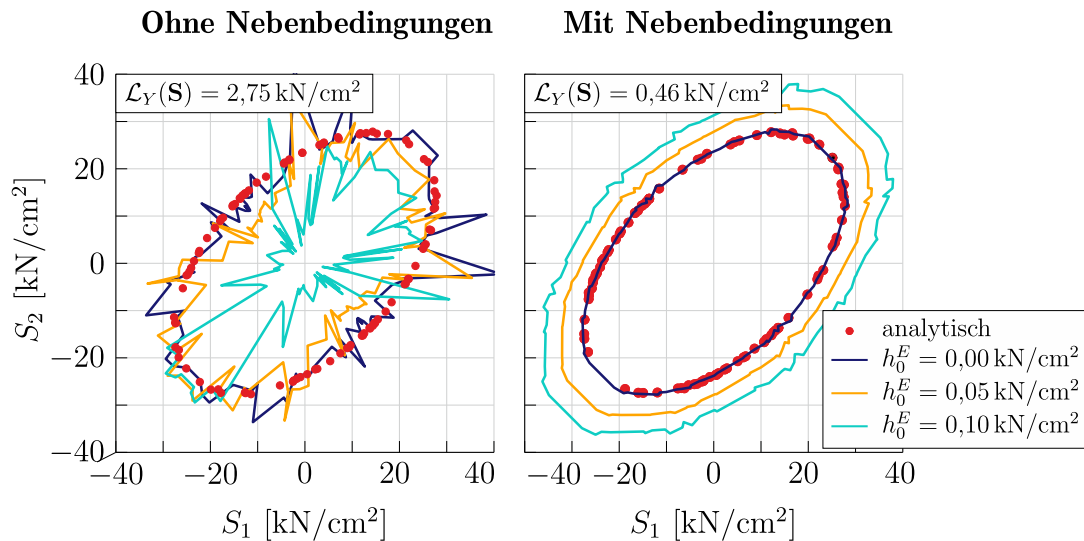


Abbildung 6.15: Approximation der VON MISES'schen Fließfläche mit KNN-Materialmodellen. Aus den 15 Trainingsprozessen ist jeweils das beste KNN in Bezug auf den $\mathcal{L}_Y(\mathbf{S})$ -Fehler abgebildet, links ohne Nebenbedingungen und rechts mit.

reversibel ist, kann somit eine in der Vergangenheit liegende Belastung simuliert werden, ohne dass sie explizit durchgeführt werden musste. Diese Eigenschaft hängt von der gewählten Geschichtsvariable ab.

Zusammenfassend kann festgehalten werden, dass es möglich ist, selbst mit einer vergleichsweise geringen Menge an Daten eine Fließfläche mit KNN zu approximieren. Allerdings ist diese Aufgabenstellung wesentlich einfacher als das Erlernen beliebigen Be- und Entlastungsverhaltens.

Bisher wurden alle Untersuchungen am Materialpunkt durchgeführt. Dabei hat sich gezeigt, dass im Fall von geschichtsabhängigem Materialverhalten die Einforderung physikalisch motivierter Nebenbedingungen einen fabelhaften Effekt in Bezug auf das KNN-Training sowie die Anwendung als KNN-Materialmodell hat. Wie bereits in Kapitel 5, soll nun der Schritt hin zur Struktursimulation getätigt werden. Dabei liegt der Fokus in diesem Kapitel mehr auf den Nebenbedingungen und deren Effekt im Rahmen der Anwendung in der FEM und nicht auf dem Vergleich verschiedener KNN-Architekturen bzw. KNN-Materialmodelle. Daher wird nachfolgend ausschließlich das FNN-P-Ink-Materialmodell verwendet.

6.5 Anwendung der KNN-P-Materialmodelle in der FEM

Grundsätzlich gelten die in Abschnitt 5.5 beschriebenen Zusammenhänge für elastische KNN-Materialmodelle ebenfalls für diejenigen aus Kapitel 6. Im Fall der FNN-Modelle müssen darüber hinaus die im Eingangsvektor \mathbf{x} definierten Geschichtsvariablen des letzten Lastschritts für jeden Integrationspunkt abgespeichert werden. Eine Übersicht des gesamten inkrementell-iterativen Algorithmus, inklusive der Handhabung besagter Geschichtsvariablen, ist in Abbildung 6.16 für das FNN-P-Ink-Materialmodell dargestellt. Das Training der nachfolgend verwendeten KNN erfolgt in MATLAB [170] mit dem QN-Verfahren aus Abschnitt 4.4.2. Über eine Transferdatei werden alle notwendigen Informationen wie die Topologie, die Gewichte, das KNN-Materialmodell usw. an FEAP [168] übergeben.

6.5.1 Aluminiumscheibe mit zentralem Loch

Als Beispiel des ebenen Spannungszustands wird nachfolgend eine uniaxial gezogene quadratische Aluminiumscheibe mit zentralem Loch untersucht. Die Scheibe ist 100 cm breit und 2 cm dick. Das Loch hat einen Radius von 10 cm. Das FE-Netz besteht aus insgesamt 200 isoparametrischen Scheibenelementen mit bilinearen Ansatzfunktionen, vgl. Abschnitt 2.2.4. Durch die vorhandene Doppelsymmetrie wird das FE-Netz nur an einem Viertel des Systems definiert, vgl. Beispiel 5.5.2. Außerdem ist das Netz zur besseren Approximation der Spannungsspitze zum Loch hin verfeinert. Die Scheibe wird mit einer zyklischen Last $q(t)$ uniaxial beansprucht, sodass am Lochrand plastische Effekte auftreten, inklusive Entlastung und Wiederbelastung. Das FE-Modell sowie die Lastfunktion sind in Abbildung 6.17 dargestellt. Die Pseudo-Zeitachse $t \in [0, 8]$ wird zur Definition der Lastschritte in 80 äquidistante Schritte diskretisiert.

E-Modul	Querd.	Fließgrenze	Verfestigungsmodul	Iso./Kin.
C [kN/cm ²]	ν [-]	Y_0 [kN/cm ²]	C_P [kN/cm ²]	θ [-]
7000	0,34	24	$C/100 = 70$	0

Tabelle 6.6: Materialparameter für das analytische VON MISES'sche Materialmodell mit kinematischer Verfestigung zur Modellierung von Aluminium

In Tabelle 6.6 sind die Parameter für das VON MISES'sche Materialmodell nach Abschnitt 3.4.5 zur Modellierung von Aluminium gegeben. Dabei wird reine kinematische Verfestigung ($\theta = 0$) angenommen. Im Jahr 1899 hat GUSTAV KIRSCH herausgefunden, dass bei einer unendlich ausgedehnten Scheibe mit Loch unter uniaxialer Zugbelastung am Lochrand eine Spannungsspitze auftritt, die

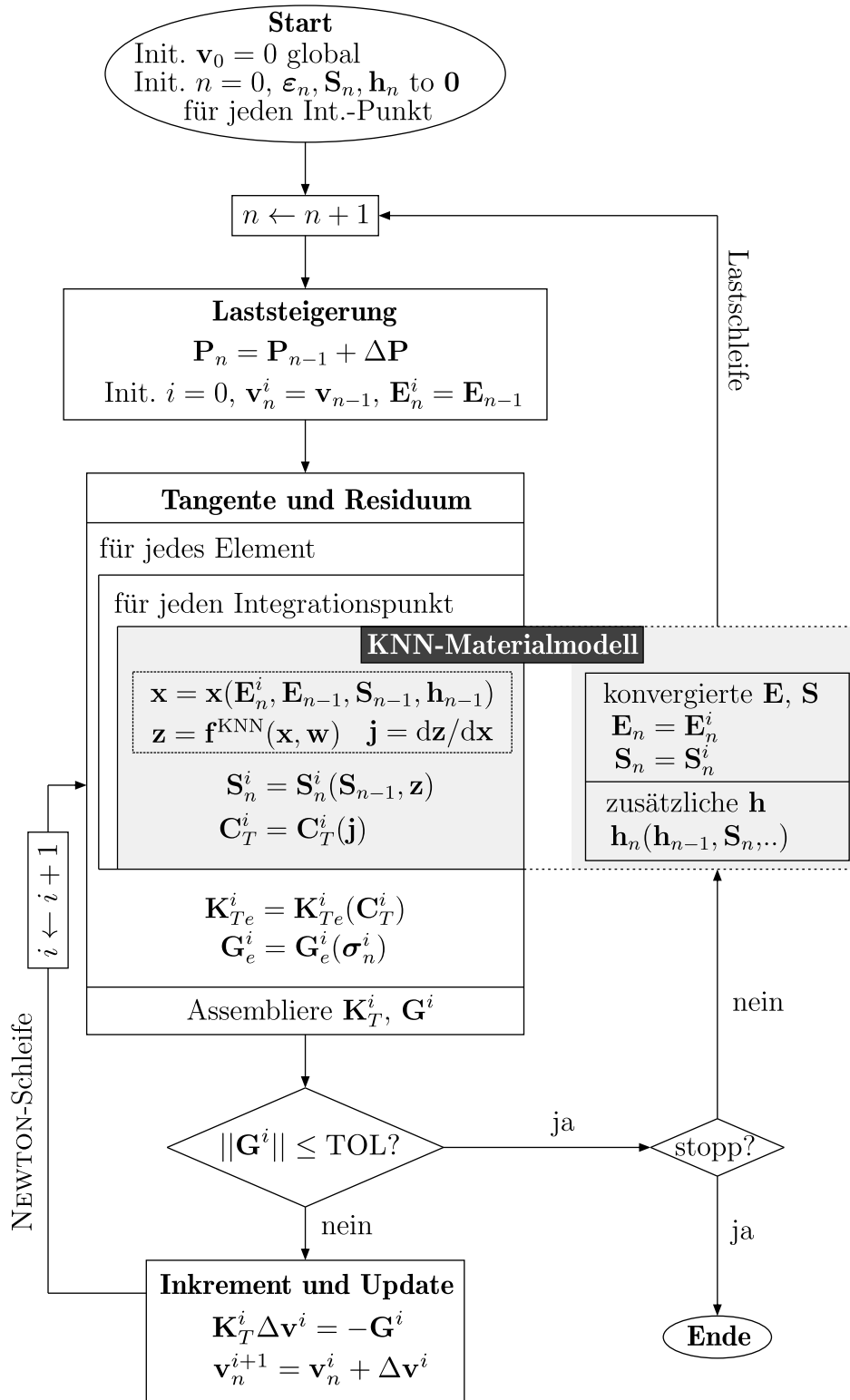


Abbildung 6.16: Ablaufschema des inkrementell-iterativen Lösungsverfahrens der FEM mit dem KNN-Materialmodell FNN-P-Exp-Ink

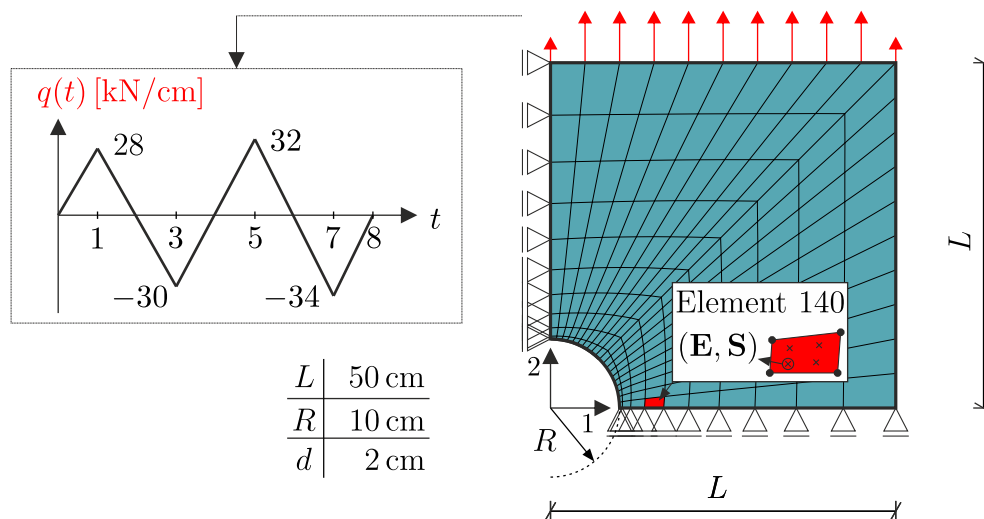


Abbildung 6.17: Scheibe mit Loch: Abmessungen und Diskretisierung des FE-Netzes sowie die Lastfunktion

dem Dreifachen der aufgetragenen Last entspricht, vgl. GIRMANN [61]. Mit $Y_0 = 24 \text{ kN/cm}^2$ lässt sich folglich abschätzen, dass ab ca. $q = 16 \text{ kN/cm}$ Fließen am Lochrand zu erwarten ist.

Da durch die Plastifizierung am Lochrand auch größere Verzerrungen zu erwarten sind, wird der Verzerrungsraum zu $\Omega_{\mathbf{E}} = [-0,02; 0,02]^3$ definiert. Dieser ist folglich wesentlich größer als derjenige in Abschnitt 6.4 zur Berechnung der Fließfläche. Innerhalb dieses Verzerrungsraums werden insgesamt 22 Materialpunktversuche durchgeführt, um die (\mathbf{E}, \mathbf{S}) -Pfade zur Berechnung der Datenpaare zu generieren, vgl. Abschnitt 5.1.1. Dabei handelt es sich um zyklische Versuche mit jeweils einer einzigen Wiederholung und einer abschließenden Entlastung, vgl. Abschnitt 6.3.1. Maximal zwei Spannungskomponenten sind jeweils ungleich null. Es handelt sich um uni- und äquibiaxiale Zug- und Druckversuche, Schubversuche, diverse Kombinationen aus diesen fünf Basisversuchen sowie Versuche mit behinderter Querkontraktion. In Tabelle 6.7 sind diese 22 Pfade beschrieben. Dabei sind für jeden Versuch ebenfalls die zugehörigen Spannungsintervalle derjenigen Spannungen angegeben, die ungleich null sind.

Zur Erfassung der kinematischen Verfestigung wird das FNN-P-Ink-Materialmodell nach Abschnitt 6.1.2 mit

$$\mathbf{x} = [\Delta \mathbf{E}, \mathbf{E}_l, \mathbf{S}_l]^T \in \mathbb{R}^9 \quad \text{und} \quad \mathbf{z} = \Delta \mathbf{S} \in \mathbb{R}^3 \quad (6.59)$$

gewählt. Die Topologie [9-25-20-15-10-3] enthält eine abgestufte Anzahl von Neuronen in den Zwischenschichten. Jeder der 22 Pfade aus Tabelle 6.7 besteht aus vier Belastungsrichtungen, die mit jeweils 40 Schritten diskretisiert sind. Zur Berechnung der Trainings- und Validierungsdatenpaare werden mit $\Delta n \in$

Nr.	Beschreibung	Spannungsintervalle [kN/cm ²]
1/2	Zug 1/2	$S_{11}/S_{22} \in [-25,47; 25,51]$
3/4	Druck 1/2	$S_{11}/S_{22} \in [-25,51; 25,47]$
5	Schub 1	$S_{12} \in [-14,86; 14,87]$
6	Schub 2	$S_{12} \in [-14,87; 14,86]$
7	Äquibiax. Zug	$S_{11} = S_{22} \in [-25,36; 25,45]$
8	Äquibiax. Druck	$S_{11} = S_{22} \in [-25,45; 25,36]$
9	Zug + Druck 1	$S_{11} \in [-15,08; 15,11]$ $S_{22} \in [-15,11; 15,08]$
10	Zug + Druck 2	$S_{11} \in [-15,11; 15,08]$ $S_{22} \in [-15,08; 15,11]$
11/12	Zug + Schub 1	$S_{11}/S_{22} \in [-22,37; 22,41]$ $S_{12} \in [-8,04; 8,05]$
13/14	Zug + Schub 2	$S_{11}/S_{22} \in [-22,37; 22,41]$ $S_{12} \in [-8,05; 8,04]$
15/16	Druck + Schub 1	$S_{11}/S_{22} \in [-22,41; 22,37]$ $S_{12} \in [-8,04; 8,05]$
17/18	Druck + Schub 2	$S_{11}/S_{22} \in [-22,41; 22,37]$ $S_{12} \in [-8,05; 8,04]$
19	Behinderte Querkontraktion ($E_{22} = 0$) Zug	$S_{11} \in [-28,94; 29,00]$ $S_{22} \in [-13,93; 13,96]$
20	Behinderte Querkontraktion ($E_{22} = 0$) Druck	$S_{11} \in [-29,00; 28,94]$ $S_{22} \in [-13,96; 13,93]$
21	Behinderte Querkontraktion ($E_{11} = 0$) Zug	$S_{11} \in [-13,93; 13,96]$ $S_{22} \in [-28,94; 29,00]$
22	Behinderte Querkontraktion ($E_{11} = 0$) Druck	$S_{11} \in [-13,96; 13,93]$ $S_{22} \in [-29,00; 28,94]$

Tabelle 6.7: 22 Spannungs-Dehnungspfade zum KNN-Training für das Beispiel der Aluminiumlochscheibe

$\{0, 1, \dots, 40\}$ alle möglichen Indexversätze gewählt, um für eine möglichst große Menge an möglichen Verzerrungsschritten innerhalb der FEM gewappnet zu sein. Daraus folgen pro Belastungsrichtung $P_d = 861$ inkrementelle Datenpaare $(\mathbf{x}, \mathbf{t}) = ([\Delta \mathbf{E}, \mathbf{E}_t, \mathbf{S}_t], \Delta \mathbf{S})$, vgl. Gleichung (6.15), und letztlich 75 768 Datenpaare, die für das Training zur Verfügung stehen. Diese werden zufällig auf 20 000 gekürzt. Von diesen 20 000 Paaren werden 80% für das Training und 20% zur Berechnung des Validierungsfehlers und damit zur Steuerung des Trainingsabbruchs genutzt. Die dafür nötigen Einstellungen sind $\delta_{rel} = \delta_{abs} = 10^{-3}$ und $epo_{max} = 10^4$.

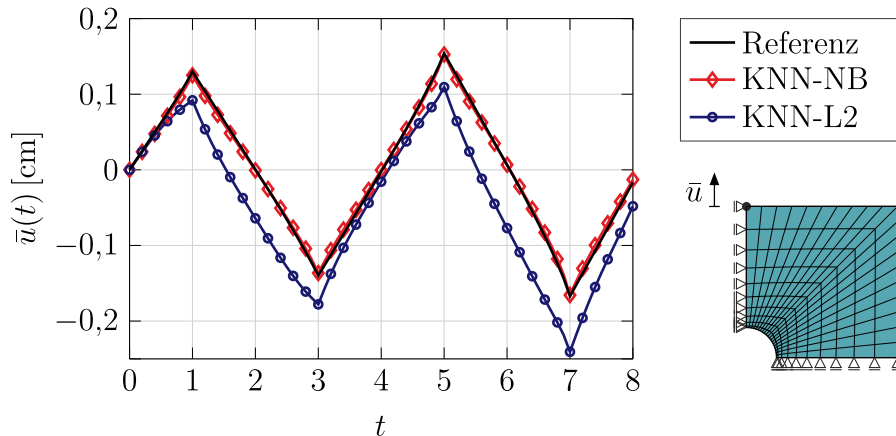


Abbildung 6.18: Scheibe mit Loch: vertikale Verschiebung \bar{u} des oberen linken FE-Knotens in Abhängigkeit von der Zeit t

Das KNN wird einmal mit den Nebenbedingungen aus Abschnitt 6.2 trainiert und einmal ohne Nebenbedingungen. Die Ergebnisse werden nachfolgend im Hinblick auf die physikalische Plausibilität und die Ergebnisgenauigkeit miteinander verglichen. Dabei war das Ziel die Einstellungen dieses Beispiels so zu wählen, dass beide KNN das Beispiel der zyklisch gezogenen, plastifizierenden Lochscheibe vollständig berechnen können. Das KNN ohne Nebenbedingungen wurde dazu zusätzlich mit einer L2-Regularisierung während des Trainingsprozesses stabilisiert, vgl. Abschnitt 4.5.1. Dabei wurde durch ausprobieren der Wert $\epsilon = 10^{-5}$ ermittelt. Es wird im Folgenden KNN-L2 bezeichnet. Das mit Nebenbedingungen trainierte KNN wird KNN-NB genannt. Dabei wurde im Rahmen der klassischen Penalty-Methode nach Abschnitt 4.5.1 der Penalty-Parameter $\epsilon = 1$ gewählt. Für die Nebenbedingung der Normalisierung nach Abschnitt 6.2.2 werden $P_C = 2000$ NB-Punkte mit der Strategie der konvexen Hülle generiert, vgl. Abschnitt 6.2.1. Die Stationarität der Normalisierung aus Abschnitt 6.2.3 wird nicht berücksichtigt, vgl. Abschnitt 6.3.5. Für die restlichen drei Nebenbedingungen werden jeweils $P_C = 20\,000$ NB-Punkte generiert.

In Abbildung 6.18 ist der Verlauf der Verschiebung \bar{u} des oberen linken FE-Knotens für beide KNN über die Zeit t dargestellt. Während die Lösung des KNN-NB nahezu deckungsgleich mit der Referenzlösung ist, weicht die KNN-L2-Lösung schon bei der ersten Lastspitze signifikant von der Referenzlösung ab. Stellenweise wird im weiteren Verlauf der Berechnung der Fehler wieder ausgeglichen oder verstärkt. In Abbildung 6.19 ist der Verlauf der Spannung S_{22} in Abhängigkeit von der Dehnung E_{22} dargestellt, welche am ersten Integrationspunkt von Element 140 aufgezeichnet wurde, vgl. Abbildung 6.17. Die KNN-NB-Lösung liegt dabei deutlich näher an der Referenzlösung, auch wenn es einige Abweichungen gibt. Außerdem bilden beide KNN die kinematische Verfestigung in diesem Fall

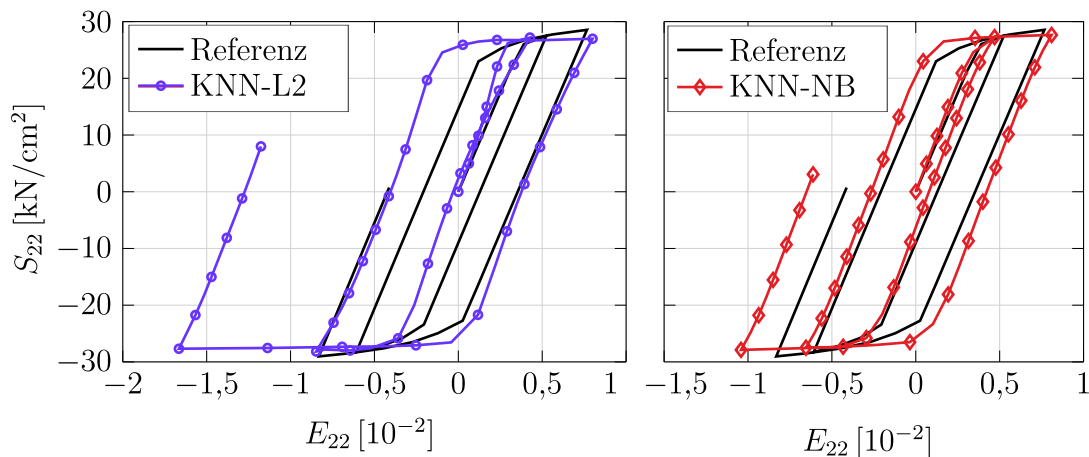


Abbildung 6.19: Scheibe mit Loch: Hysterese der Spannung S_{22} in Abhängigkeit der Verzerrung E_{22} am ersten Integrationspunkt von Element 140, vgl. Abbildung 6.17

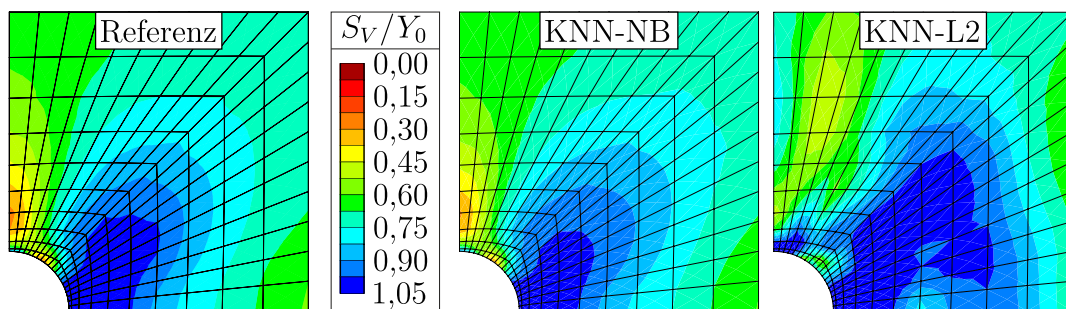


Abbildung 6.20: Scheibe mit Loch: Verlauf der Vergleichsspannung S_V bezogen auf die initiale Fließspannung Y_0 zum Zeitpunkt $t = 7$

lokal nicht exakt ab. Das Verhältnis der Vergleichsspannung S_V (vgl. (3.111) mit $S_{i3} = 0, i = 1, 2, 3$) zur initialen Fließspannung Y_0 ist in Abbildung 6.20 zum Zeitpunkt $t = 7$ der höchsten Lastspitze über die gesamte Struktur aufgetragen. Der qualitative Verlauf der KNN-Lösung unter Berücksichtigung der physikalischen Nebenbedingungen stimmt dabei mit der Referenzlösung überein. Das ausschließlich mit L2-Regularisierung trainierte KNN ist nicht in der Lage, den Verlauf der plastischen Zone korrekt wiederzugeben.

Die KNN wurden sowohl an der globalen Strukturantwort als auch in Bezug auf die lokale Spannungsantwort miteinander verglichen. Dabei haben die Nebenbedingungen in allen Fällen die Ergebnisse verbessert. Neben der besseren lokalen Approximation der Spannungen ist vor allem die plausible Abbildung der Fließzone relevant. Auch wenn lokal kleinere Abweichungen zur Referenzlösung vorhanden sind, kann global das physikalisch korrekte Verhalten wiedergegeben werden. An dieser Stelle soll noch einmal hervorgehoben werden, dass die Men-

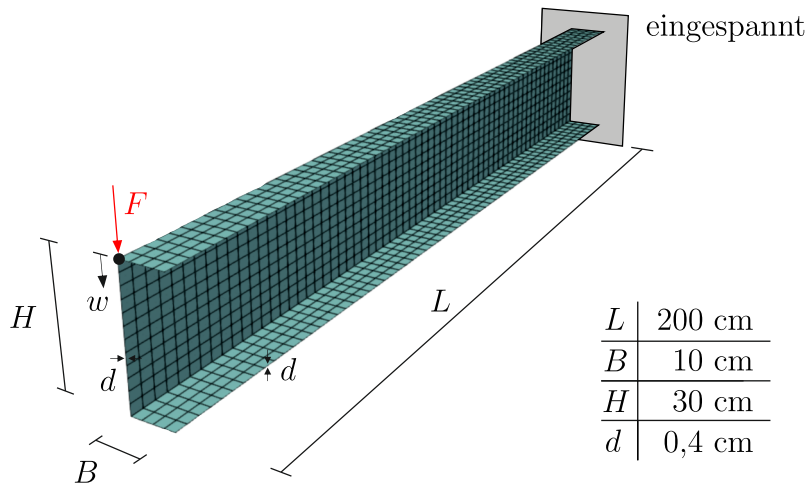


Abbildung 6.21: U-Profil-Beispiel: Abmessungen, Belastung und FE-Diskretisierung

ge der Trainingspfade mit 22 Versuchen sehr gering ist zur Approximation der neundimensionalen Antwortfläche. Die Berücksichtigung von Nebenbedingungen erlaubt trotz dieser begrenzten Informationsquelle eine qualitativ gute und physikalisch sinnvolle Approximation des Materialverhaltens. Aufgrund der geringen Anzahl an Trainingsszenarien sollte im Übrigen auch nicht erwartet werden, dass die Verläufe der Spannungen exakt deckungsgleich sind.

6.5.2 Eingespanntes U-Profil

In diesem Beispiel soll die Be- und Entlastung eines eingespannten U-Profils mithilfe eines dreidimensionalen KNN-Materialmodells untersucht werden. Der Kragarm ist 200 cm lang und 30 cm hoch. Die Flansche sind 10 cm breit. Das Blech ist überall 0,4 cm dick. Die FE-Modellierung wird mit den Schalenelementen aus Abschnitt 2.2.4 durchgeführt. Zur Berücksichtigung des durch die Plastifizierung nichtlinearen Spannungsverlaufs über die Höhe werden drei numerische Schichten mit insgesamt sechs Integrationspunkten eingestellt. Die EAS-Parameter werden zu $\beta_m = \beta_b = 4$ und $\beta_z = 2$ gewählt. Die Struktur wird insgesamt mit 1280 Schalenelementen diskretisiert: 64 Elemente in Balkenlängsrichtung, zwölf in Höhenrichtung und jeweils vier in Richtung der Flansche. Eine Einzellast an der äußeren Kante zwischen Oberflansch und Steg ist die einzige Belastung. Die Berechnung erfolgt verschiebungsgesteuert durch das Bogenlängenverfahren, vgl. WAGNER und WRIGGERS [187]. Dabei wird am Lastangriffspunkt pro Lastschritt ein Verschiebungssinkrement von $\Delta w = 0,2$ cm vorgegeben und das zugehörige Lastinkrement ΔF berechnet. Nach 30 Lastschritten ist die maximale Durchbiegung von 6 cm erreicht und der Kragarm wird wieder entlastet. Das System, die Last

E-Modul C [kN/cm ²]	Querd. ν [-]	Fließgrenze Y_0 [kN/cm ²]	Verfestigungsmodul C_P [kN/cm ²]	Iso./Kin. θ [-]
21 000	0,3	23,5	$C/100 = 210$	1

Tabelle 6.8: Materialparameter für das analytische VON MISES'sche Materialmodell mit isotroper Verfestigung für das U-Profil-Beispiel

und das FE-Netz sind in Abbildung 6.21 dargestellt.

Das FNN-P-Iso-Materialmodell soll ein VON MISES'sches Materialmodell mit isotroper Verfestigung für Stahl approximieren. Die zugehörigen Materialparameter sind in Tabelle 6.8 gegeben. Die Eingangs- und Ausgangsvariablen

$$\mathbf{x} = [\Delta \mathbf{E}, \mathbf{S}_t, h_t^E]^T \in \mathbb{R}^{13} \quad \text{und} \quad \mathbf{z} = \Delta \mathbf{S} \in \mathbb{R}^6 \quad (6.60)$$

entsprechen denjenigen aus Definition (6.56) für den dreidimensionalen Fall. Die energiebasierte Geschichtsvariable $h^E(t)$ wurde in Abschnitt 6.1.1 vorgestellt und berücksichtigt das isotrope Verfestigungsverhalten. In Voruntersuchungen hat sich die Topologie [13-30-25-20-15-6] als ausreichend herausgestellt, mit insgesamt 2126 Gewichten.

Lediglich 14 Pfade werden mittels Materialpunkt-Versuchen nach Abschnitt 5.1.1 generiert und zum Training des dreidimensionalen KNN-Materialmodells verwendet. Wie im vorangegangenen Beispiel sind dabei jeweils nur zwei Spannungskomponenten ungleich null. Alle Pfade sind als Hysteresen mit drei Belastungsrichtungen und einer Entlastungsrichtung definiert, vgl. Abschnitt 6.3.2. Die Verzerrungskomponenten sind dabei durch $\Omega_{\mathbf{E}} = [-0,01; 0,01]^6$ begrenzt. Die Pfade sind in Tabelle 6.9 angegeben. Diese 14 Pfade werden unter Annahme von Isotropie durch 20 zufällige Rotationen auf insgesamt 294 Pfade erweitert, vgl. Abschnitt 5.1.2. Die zufälligen Rotationen werden über die EULER-RODRIGUES-Formel (3.8) mit jeweils vier gleichverteilten Zufallszahlen generiert: Drei definieren den Rotationsvektor und die vierte den Winkel.

Jeder Pfad ist mit $M = 41$ GGW-Punkten diskretisiert. Zur Berechnung der inkrementellen Datenpunkte nach Abschnitt 6.1.2 werden alle möglichen Indexversätze $\Delta n \in \{0, 1, \dots, 40\}$ verwendet. Insgesamt resultieren so 1 012 536 inkrementelle Datenpunkte, die zufällig auf 20 000 reduziert werden. Davon werden 80% zum Training und 20% zur Validierung verwendet. Der Trainingsprozess wird nach 10 000 Epochen beendet. Diejenigen Gewichte, die zum kleinsten Validierungsfehler führen, werden abgespeichert.

Wie im vorangegangenen Abschnitt wird ein KNN-Materialmodell mit Nebenbedingungen (KNN-NB) mit einem KNN verglichen, welches nur eine L2-Regularisierung (KNN-L2) berücksichtigt. Für die L2-Regularisierung wird $\epsilon = 10^{-5}$ per

Nr.	Beschreibung	Spannungsintervalle [kN/cm ²]
1	Zug	$S_{11} \in [-28,95; 32,46]$
2	Druck	$S_{11} \in [-32,46; 28,59]$
3	Schub 1	$S_{12} \in [-15, 29; 16, 39]$
4	Schub 2	$S_{12} \in [-16, 39; 15, 29]$
5	Äquibiax. Zug	$S_{11} = S_{22} \in [-34,73; 41,91]$
6	Äquibiax. Druck	$S_{11} = S_{22} \in [-41,91; 34,73]$
7	Zug + Druck 1	$S_{11} \in [-17,32; 19,75]$ $S_{22} \in [-19,75; 17,32]$
8	Zug + Druck 2	$S_{11} \in [-19,75; 17,32]$ $S_{22} \in [-17,32; 19,75]$
9	Zug + Schub 1	$S_{11} \in [-25,91; 29,50]$ $S_{12} \in [-8,71; 10,23]$
10	Zug + Schub 2	$S_{11} \in [-25,91; 29,50]$ $S_{12} \in [-10,23; 8,71]$
11	Druck + Schub 1	$S_{11} \in [-29,50; 25,91]$ $S_{12} \in [-8,71; 10,23]$
12	Druck + Schub 2	$S_{11} \in [-29,50; 25,91]$ $S_{12} \in [-10,23; 8,71]$
13	Behinderte Querkontraktion ($E_{22} = 0$) Zug	$S_{11} \in [-34,29; 38,89]$ $S_{22} \in [-17,08; 19,36]$
14	Behinderte Querkontraktion ($E_{22} = 0$) Druck	$S_{11} \in [-38,89; 34,29]$ $S_{22} \in [-19,36; 17,08]$

Tabelle 6.9: 14 Spannungs-Dehnungspfade zum KNN-Training für das Beispiel des eingespannten U-Profiles

Trial-and-Error-Prinzip gewählt. Für das KNN-NB-Modell werden alle Nebenbedingungen (NB) außer die der Stationarität nach Abschnitt 6.2.3 verwendet. Für die Normalisierungs-NB gilt $\epsilon = 1$ und $P_C = 1000$. Für die Symmetrie-NB ist $\epsilon = 1$ und $P_C = 10\,000$. Die beiden Ungleichheits-NB werden mit $\epsilon = 100$ und $P_C = 50\,000$ stärker gewichtet.

In Abbildung 6.22 ist die Last-Verschiebungskurve des Lastangriffspunkts für beide KNN-Materialmodelle dargestellt. Da die äußere Last nicht im Schubmittelpunkt des Trägers angreift, rotiert dieser direkt zu Beginn der Belastung. Dadurch entstehen große Verschiebungen über die ganze Struktur, was die Nutzung von \mathbf{E} und \mathbf{S} als Verzerrungs- und Spannungsmaß motiviert. Am Oberflansch entsteht durch die aus Rotation resultierenden Druckspannungen eine Beule. Das mit Nebenbedingungen angereicherte KNN ist in der Lage die vollständige Be-

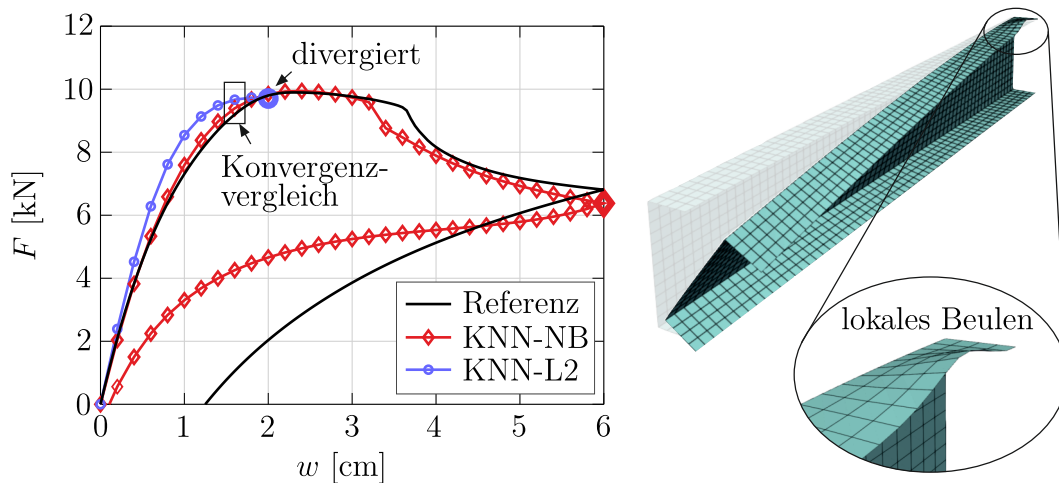


Abbildung 6.22: U-Profil-Beispiel: Last-Verschiebungskurve des Lastangriffspunkts (links) und deformierte Konfiguration bei $w = 6$ cm (rechts). Eine Konvergenzstudie wird von Lastschritt 7 zu Lastschritt 8 durchgeführt.

und Entlastung des Trägers zu simulieren. Das ausschließlich mit L2-Regularisierung trainierte KNN divergiert in der Nähe der Traglast. Dies zeigt erneut den positiven Einfluss der physikalisch motivierten Nebenbedingungen im Hinblick auf die Stabilität innerhalb der FEM. Allerdings unterscheidet sich die Entlastungskurve stark von derjenigen des VON MISES'schen Referenzmaterials. Dies liegt u. a. an der geringen Datenbasis. Nur 14 Pfade wurden dazu genutzt, wobei die Entlastung ausschließlich am Rand des Trainingsraums durchgeführt wurde. Dass KNN muss also alle möglichen Entlastungspunkte dazwischen interpolieren. Eine andere Erklärung wird in den abschließenden Bemerkungen aus Abschnitt 6.5.3 gegeben. Zwischen Lastschritt 7 und 8 wird eine Konvergenzstudie für beide KNN-Materialmodelle durchgeführt. Diese ist in Tabelle 6.10 dargestellt. Hier zeigt sich ebenfalls der positive Einfluss der Nebenbedingungen durch die wesentlich bessere, quadratisch anmutende Konvergenzrate.

6.5.3 Abschließende Bemerkungen

- Das Volumen des dreidimensionalen Ellipsoids, der die VON MISES'sche Fließfläche des ESZ im dreidimensionalen Komponentenraum darstellt, lässt sich mit

$$V_{\mathbf{E}} = \frac{16}{9} \pi \left(\frac{Y_0}{C} \right)^3 (1 + \nu)^2 (1 - \nu) \quad (6.61)$$

berechnen. Mit den Werten aus Tabelle 6.6 folgt für das Beispiel der Lochscheibe $V_{\mathbf{E}} = 2,67 \cdot 10^{-7}$. Der Würfel $[-0,02; 0,02]^3$ hat das Volumen $V_{\Omega} =$

Iterationsschritt	$\ \mathbf{G}\ $ für KNN-NB	$\ \mathbf{G}\ $ KNN-L2
1	$2,7358558 \cdot 10^0$	$4,3467223 \cdot 10^0$
2	$7,1441342 \cdot 10^1$	$7,6390431 \cdot 10^1$
3	$9,7346244 \cdot 10^{-1}$	$3,1685351 \cdot 10^0$
4	$1,2218827 \cdot 10^{-1}$	$4,2916863 \cdot 10^{-1}$
5	$6,1137981 \cdot 10^{-2}$	$4,4903747 \cdot 10^{-2}$
6	$1,3957943 \cdot 10^{-4}$	$2,3330237 \cdot 10^{-2}$
7	$5,6976029 \cdot 10^{-7}$	$4,8986588 \cdot 10^{-3}$
8		$2,4455923 \cdot 10^{-3}$
...		...
16		$7,5131456 \cdot 10^{-7}$

Tabelle 6.10: Konvergenzstudie über den Verlauf der Residualnorm $\|\mathbf{G}\|$ (vgl. Gleichung (2.83)) von Lastschritt 7 zu Lastschritt 8

$6,40 \cdot 10^{-5}$. Der elastische Bereich nimmt folglich nicht einmal 0,5% dieses Trainingsraums ein. Dies hat zur Folge, dass gerade der elastische Bereich beim Training von KNN für plastische Materialgesetze häufig unterrepräsentiert ist und dadurch schlechter trainiert wird. Wenn durch etwaige Lastumlagerungen die plastischen Bereiche in größeren Systemen im Vergleich zur Gesamtgröße klein sind, dann fällt diese Problematik bei der numerischen Anwendung stark ins Gewicht. Dieser Umstand wird bei dreidimensionalen Materialgesetzen und einem größeren abzubildenden Bereich plastischen Fließens noch verstärkt. Auch wenn im elastischen Bereich das Verhalten gleich bleibt, muss das FNN für jeden möglichen GGW-Punkt dieses Verhalten erlernen. Dies stellt einen fundamentalen Nachteil der FNN-Approximation von geschichtsabhängigem Materialverhalten dar und ist womöglich der Grund für die Diskrepanz der Entlastungskurven des U-Profil-Beispiels.

- Im Bereich von Singularitäten können die Eingangsvariablen eines KNN-Materialmodells schnell außerhalb des Trainingsbereichs liegen. Da FNN-Materialmodelle alle möglichen Verzerrungszustände erlernen müssen, sollte diese Problematik am besten auf Strukturebene entgegengewirkt werden, indem bspw. die Singularitäten durch geeignete Randbedingungen entschärft werden.
- In jedem Fall ist eine hohe Anzahl von Indexversätzen Δn zu empfehlen, um innerhalb der GGW-Iteration alle möglichen Verzerrungsschritte bestmöglich abdecken zu können, vgl. Abschnitt 6.1.2. Gegebenenfalls ist die Lastschrittweite anzupassen.

- Auch, wenn KNN-Materialmodelle im Rahmen der Mehrskalenmodellierung häufig zur Modellreduktion verwendet werden, kann die Rechenzeit unter Umständen erheblich sein. Diese steht in direktem Zusammenhang mit der Anzahl der Gewichte. Zum Vergleich: Das KNN-NB Materialmodell des U-Profil-Beispiels benötigt für die ersten fünf Lastschritte ca. das 60-fache der Zeit des analytischen VON MISES'SCHEN Materialmodells. Mit zugeschnittenen FE-Formulierungen könnte die Rechenzeit weiter reduziert werden, vgl. FRANKE ET AL. [48]. Wichtig ist, dass die Nebenbedingungen nur die Trainingszeit beeinflussen und nicht die Zeit der KNN-Ausführung als Materialmodell.

Bei der Berechnung von Strukturen kommt der Nachteil der plastischen FNN-Materialmodelle deutlich hervor: Alle möglichen GGW-Punkte, die innerhalb der Berechnung erreicht werden könnten, müssen in den Trainingsdaten beinhaltet sein. Die physikalischen Nebenbedingungen verbessern diesen Umstand außerordentlich, räumen den Nachteil allerdings nicht gänzlich weg. Bei der Anwendung von plastischen FNN-Materialmodellen muss daher ein besonderes Augenmerk auf der Datendichte liegen. Außerdem sollten mögliche Singularitäten entschärft werden, die lokal den Trainingsraum des KNN verletzen.

7 Anwendung auf experimentelle Daten

Die in den Kapiteln 5 und 6 herausgearbeiteten Erkenntnisse zur Materialmodellierung mit künstlichen neuronalen Netzen (KNN) werden in diesem Kapitel an realen Versuchsdaten in die Anwendung gebracht. In Abbildung 7.1 sind die drei Materialien dargestellt, welche nachfolgend untersucht werden, inklusive der zugehörigen Herausforderungen bei der Modellierung mit KNN.

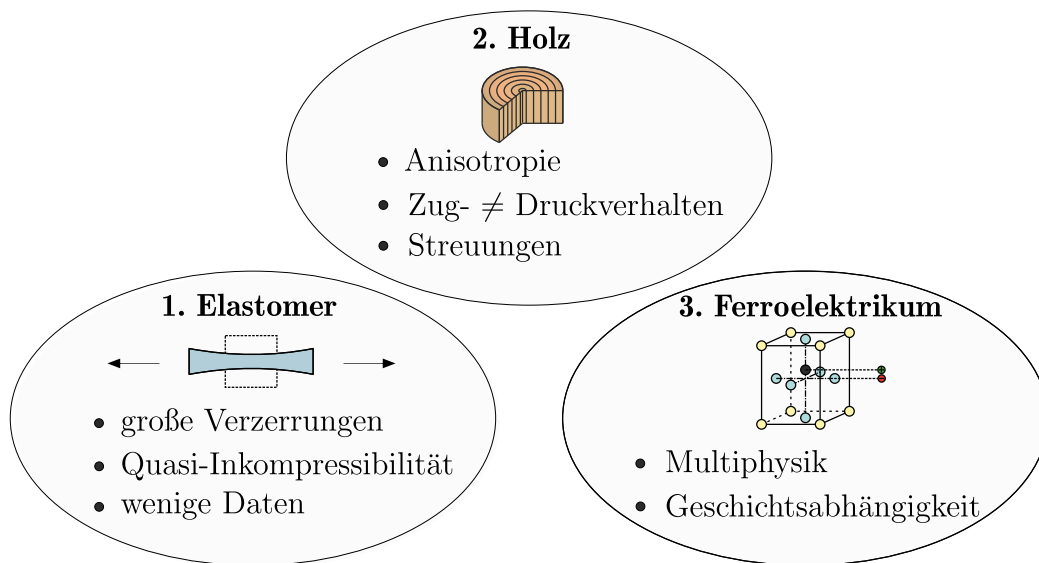


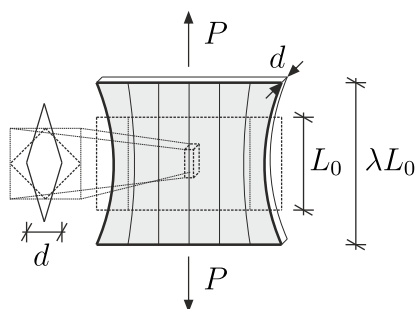
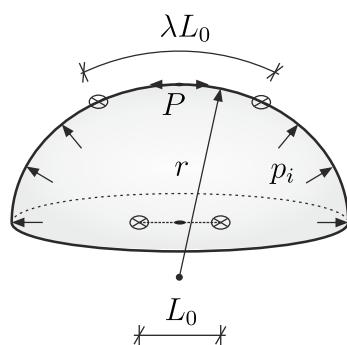
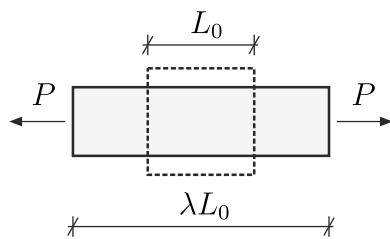
Abbildung 7.1: Übersicht über die in Kapitel 7 untersuchten Materialien

Im Jahr 1944 hat TRELOAR [172] die bis heute als essenziell geltenden Daten zu vulkanisiertem Gummi veröffentlicht. Sie werden in Abschnitt 7.1 im Hinblick auf elastisches, quasi-inkompressibles Materialverhalten mit großen Verzerrungen verwendet. In Abschnitt 7.2 wird auf Basis von biaxialen Versuchen von EBERHARDSTEINER [40] das nichtlineare und anisotrope Verhalten von Fichtenholz unter monotoner Belastungssteigerung modelliert. Als Ausblick in multiphysikalische Anwendungen wird in Abschnitt 7.3 das geschichtsabhängige Verhalten einer ferroelektrischen PZT-Keramik untersucht, für das von ZHOU ET AL. [202, 203] Versuche durchgeführt wurden.

Ziel dieses Kapitels ist es, einige über die Anwendung synthetischer Daten hinausgehenden Herausforderungen bei der KNN-Materialmodellierung herauszuarbeiten und ggf. Lösungen vorzuschlagen. Dabei ist bspw. die Datenaufbereitung stets ein wichtiger Punkt. Sie wird in den jeweiligen Abschnitten detailliert beschrieben. An dieser Stelle sei angemerkt, dass alle genannten Materialien und deren Daten die Möglichkeit zu tiefschürfenderen Untersuchungen bieten, als im Folgenden dargestellt.

7.1 Treloars Daten von vulkanisiertem Gummi

Gummis werden u. a. durch Vulkanisation von Natur- oder Synthekautschuken hergestellt. Durch Änderung der molekularen Bindungen erhalten sie infolge dieses Prozesses ihre charakteristischen elastischen Eigenschaften. Gummis sind isotrop, eine Teilgruppe der Elastomere und erlauben bspw. in Schubauflagern bei Brückenbauwerken relativ große horizontale Verformungen, ohne dabei ihre elastischen Eigenschaften zu verlieren. TRELOAR [172] untersuchte in seiner Veröffentlichung aus dem Jahr 1944 zwei Gummiwerkstoffe, von denen einer mit 8% und der andere mit 2% Schwefel vulkanisiert wurde. Während letzterer aus Sicht der industriellen Anwendung Vorzüge besitzt, bspw. eine höhere Festigkeit, verhält sich der mit 8% vulkanisierte Werkstoff auch bei größeren Streckungen reversibel. Daher werden die Versuchsdaten desjenigen Gummis mit 8% Schwefel in der Literatur häufig als Maßstab eines idealen elastischen Festkörpers angesehen, vgl. HOLZAPFEL [81]. Sie werden ebenfalls für die nachfolgenden Untersuchungen herangezogen. TRELOAR führte insgesamt vier Versuche durch:



Einen uniaxialen Zugversuch bis zum Versagen bei $\lambda \approx 7,6$: Die Proben zeigen bis $\lambda \approx 5,5$ ein elastisches Verhalten, bevor sich Hysteresen mit nicht deckungsgleicher Be- und Entlastungskurve einstellen (Energiedissipation).

Einen äquibiaxialen Zugversuch durch Aufblasen einer kreisförmigen Membran: In einer Region um den Pol wird eine kugelförmige Konfiguration angenommen. Mit zugehörigem Radius r und dem aufgebrachtten Innendruck p_i folgt die 1. PIOLA-KIRCHHOFF'sche Spannung zu $P = \lambda p_i r / d / 2$. Das Material wird bis $\lambda \approx 4,5$ belastet und verhält sich bis $\lambda \approx 3,0$ elastisch. d ist die initiale Dicke.

Einen reinen Schubversuch durch Streckung einer breiten Membran: Dadurch ist in der Mitte die Dehnung in Breitenrichtung näherungsweise null. Unter Annahme von Volumenkonstanz deformieren sich die mittleren Bereiche der Membran in Dickenrichtung d äquivalent zu einem reinen Schubzustand. Das Material wird bis $\lambda \approx 5,0$ belastet und verhält sich durchgehend elastisch.

Darüber hinaus führte TRELOAR einen vierten Versuch durch, bei dem er einen uniaxialen Zugzustand mit einem anschließenden reinen Schubzustand überlagerte. Dieser Versuch wird nachfolgend nicht berücksichtigt. Alle Versuchsdaten basieren auf der Wiederbelastungskurve, da sich bei Gummi die Erstbelastungskurve durch inelastisches Verhalten wesentlich von den nachfolgenden Belastungskurven unterscheidet. Die Dimensionen der Prüfkörper lagen alle im zweistelligen Millimeterbereich.

Die Aufstellung eines KNN-Materialmodells auf Basis der vorliegenden Versuche birgt die folgenden Herausforderungen: sehr große Verzerrungen, nahezu inkompressibles Materialverhalten sowie eine geringe Anzahl von Versuchsdaten. Die Daten der einzelnen Versuche werden in Abschnitt 7.1.1 zunächst aufbereitet, bevor sie zum Training von KNN-Materialmodellen in Abschnitt 7.1.2 genutzt werden. Die Auswertung der erstellten Materialmodelle sowie eine abschließende Diskussion folgen in den Abschnitten 7.1.3 und 7.1.4.

7.1.1 Aufbereitung der Versuchsdaten

Aus den Messungen von TRELOAR lassen sich pro Lastschritt jeweils die maßgebende Streckung λ sowie die zugehörige 1. PIOLA-KIRCHHOFF'sche Spannung P ermitteln. STEINMANN ET AL. [163] haben die Werte grafisch ausgelesen und tabellarisch zusammengefasst. Von diesen Daten werden nachfolgend ausschließlich diejenigen zugrunde gelegt, bei denen sich das Material noch elastisch verhalten hat. Die Grenzen sind jeweils bei den Versuchsbeschreibungen angegeben.

Bei den drei Versuchen fehlt jeweils die Information über die Querdehnungen. Zur Definition der Trainingsdatenpaare für das Training eines dreidimensionalen KNN-Materialmodells müssen jedoch alle Verzerrungsgrößen vorhanden sein. Unter der Annahme einer isochoren Deformation mit $J = 1$ könnten die übrigen Verzerrungsgrößen leicht ermittelt werden. Allerdings kann das KNN-Materialmodell dann ausschließlich auf die deviatorischen Spannungen trainiert werden, vgl. Abschnitte 3.3.4 und 5.4.3. Daher wird die notwendige Rückrechnung der übrigen Streckungen unter Annahme einer *Quasi*-Inkompressibilität durchgeführt, basierend auf der Veröffentlichung von BLATZ und KO [17]. Für den uniaxialen Zugversuch (UZ) wird dabei nicht $J = 1$, sondern

$$J_{\text{UZ}} = (\lambda_{\text{UZ}})^{1-2\bar{\nu}} \xrightarrow{\bar{\nu}=0,5} 1 \quad (7.1)$$

angenommen, mit der bekannten Streckung in Zugrichtung λ_{UZ} sowie einem Parameter $\bar{\nu} \leq 0,5$. Dieser Parameter entspricht für infinitesimale Verzerrungen gerade der Querdehnzahl ν . Für $\bar{\nu} = 0,5$ gilt $J_{\text{UZ}} = 1$ und die Umrechnung geschieht unter perfekter Inkompressibilität. Mit $J_{\text{UZ}} = \lambda_{\text{UZ}}\lambda_q^2$ und Definition (7.1) folgt für

die beiden, aus den Versuchsdaten nicht direkt hervorgehenden Querstreckungen jeweils

$$\lambda_q = (\lambda_{UZ})^{-\bar{\nu}} \xrightarrow{\bar{\nu}=0,5} \frac{1}{\sqrt{\lambda_{UZ}}} . \quad (7.2)$$

Im Fall des äquibiaxialen Zugversuchs (ÄZ) gilt für die unbekannte Querstreckung

$$\lambda_q = (\lambda_{\text{ÄZ}})^{-2\bar{\nu}/(1-\bar{\nu})} \xrightarrow{\bar{\nu}=0,5} \frac{1}{(\lambda_{\text{ÄZ}})^2} , \quad (7.3)$$

mit der aus dem Versuch bekannten Streckung $\lambda_{\text{ÄZ}}$. Für den reinen Schubversuch (RS) gilt für die unbekannte Streckung in Dickenrichtung

$$\lambda_q = (\lambda_{\text{RS}})^{(-\bar{\nu}/(1-\bar{\nu}))} \xrightarrow{\bar{\nu}=0,5} \frac{1}{\lambda_{\text{RS}}} , \quad (7.4)$$

mit der bekannten Streckung in Zugrichtung λ_{RS} . An dieser Stelle sei hervor- gehoben, dass diese Umrechnung rein mathematischer Natur ist und ggf. das reale Materialverhalten verzerrt. In der Nähe des inkompressiblen Grenzfalls mit $\bar{\nu} \rightarrow 0,5$ ist die Umrechnung konsistent zum tatsächlichen Materialverhalten. Dies ist aus Sicht der KNN-Materialmodellierung allerdings schwer zu realisieren, wenn nicht nur die deviatorischen Anteile trainiert werden sollen, vgl. Abschnitt 5.4.3. Daher wird für die folgenden Untersuchungen der Wert $\bar{\nu} = 0,45$ festgelegt. Folglich ist in Bereichen mit großem hydrostatischen Druck nicht zu erwarten, dass das KNN-Materialmodell das tatsächliche Materialverhalten deckungsgleich widerspiegelt.

Darüber hinaus wurden die Versuche nur im Zugbereich gefahren, sodass keine Information über das Verhalten unter Druckbelastung vorhanden ist. Im inkompressiblen Grenzfall kann über die differentielle Arbeitsäquivalenz

$$P_{\text{UD}} d\lambda_{\text{UD}} \stackrel{!}{=} 2 P_{\text{ÄZ}} d\lambda_{\text{ÄZ}} \quad (7.5)$$

aus den Daten des äquibiaxialen Zugversuchs auf einen äquivalenten uniaxialen Druckversuch (UD) geschlossen werden, vgl. TRELOAR [173]. Dieses Prinzip wird nachfolgend auf die Quasi-Inkompressibilität mit $\bar{\nu} < 0,5$ übertragen⁷⁹. Die Streckung λ_{UD} entspricht der Querstreckung aus Gleichung (7.3). Diese lässt sich nach $\lambda_{\text{ÄZ}}$ ableiten und in die nach P_{UD} umgestellte Gleichung (7.5) einsetzen für

$$P_{\text{UD}} = \left[-2\bar{\nu} (\lambda_{\text{ÄZ}})^{(1+\bar{\nu})/\bar{\nu}} \right] P_{\text{ÄZ}} \xrightarrow{\bar{\nu}=0,5} \left[-(\lambda_{\text{ÄZ}})^3 \right] P_{\text{ÄZ}} . \quad (7.6)$$

Aus den Daten des uniaxialen Zugs lassen sich sinngemäß die Spannungen

$$P_{\text{ÄD}} = \left[-\left(\frac{\bar{\nu}}{1-\bar{\nu}}\right) (\lambda_{\text{UZ}})^{(1+\bar{\nu})/2\bar{\nu}} \right] P_{\text{UT}} \xrightarrow{\bar{\nu}=0,5} \left[-\sqrt{(\lambda_{\text{UZ}})^3} \right] P_{\text{UT}} \quad (7.7)$$

⁷⁹ Dabei gilt die Arbeitsäquivalenz nur noch näherungsweise, da sich außerhalb des inkompressiblen Grenzfalls mit $\bar{\nu} < 0,5$ die Deformationszustände von uniaxialem Druck und äquibiaxialem Zug unterscheiden.

des äquibiaxialen Drucks (ÄD) herleiten, mit den zugehörigen Streckungen $\lambda_{\text{ÄD}}$ nach Gleichung (7.2). Die Spannungen P_{UD} und $P_{\text{ÄD}}$ fallen relativ schnell gegen $-\infty$ und sorgen somit für eine schlechte Verteilung des Wertebereichs im Rahmen des KNN-Trainings, was gerade bei einer so geringen Anzahl von Trainingsdaten die Approximationsgüte im Zugbereich verschlechtert. Daher wird zur Berechnung der uni- und äquibiaxialen Druckspannungen nur ein Teil der verfügbaren Zugdaten mit $\lambda \leq 1,3$ verwendet. Darüber hinaus ist die Umrechnung über die Arbeitsäquivalenz (7.5) ohnehin nur eine Näherung im Fall der Quasi-Inkompressibilität. Dies würde die Nutzung weiterer Druckdaten zusätzlich in Frage stellen. Demzufolge werden die Druckdaten hier nur zur Stabilisierung von in der Berechnung möglicherweise auftretenden Volumendehnungen $J < 1$ genutzt.

Insgesamt folgen die in Tabelle 7.1 gegebenen Datenpaare als Basis zum Training des KNN-Materialmodells. Dabei ist λ stets die Streckung in Lastrichtung und λ_q diejenige quer dazu. Die Kennzeichnungen für die einzelnen Versuche werden nachfolgend vernachlässigt. Durch die Umrechnung der Quasi-Inkompressibilität gilt $J \in [0,83; 1,47]$. Dieser Bereich ist unrealistisch hoch, durch die Umrechnung aber unabdingbar. Daher ist es sinnvoller davon zu sprechen, das Materialverhalten von Daten *angelehnt* an die von TRELOAR durchgeführten Versuche im Folgenden approximieren zu wollen.

7.1.2 Training des KNN-Materialmodells

Das elastische Materialverhalten soll mithilfe des isotropen KNN-Materialmodells FNN-E-Iso aus Abschnitt 5.2.3 modelliert werden. Dazu werden die gegebenen Datenpaare aus Tabelle 7.1 in die Eingangs- und Ausgangsvariablen

$$\mathbf{x} = [\lambda_a^2, II_C, J^2] \quad \text{und} \quad z = S_a \quad (7.8)$$

des KNN-Materialmodells umgeformt, vgl. Gleichung (5.27). Dies geschieht jeweils in Belastungsrichtung und quer dazu. Daraus ergeben sich für die drei Spalten aus Tabelle 7.1 jeweils zwei unabhängige Trainingsdatenpaare, die nachfolgend aufgelistet sind.

$$\begin{aligned} \text{uniaxial:} & \quad ([\lambda^2, 2\lambda^2\lambda_q^2 + \lambda_q^4, \lambda^2\lambda_q^4], P/\lambda) , & \quad ([\lambda_q^2, 2\lambda^2\lambda_q^2 + \lambda_q^4, \lambda^2\lambda_q^4], 0) \\ \text{äquibiaxial:} & \quad ([\lambda^2, 2\lambda^2\lambda_q^2 + \lambda^4, \lambda^4\lambda_q^2], P/\lambda) , & \quad ([\lambda_q^2, 2\lambda^2\lambda_q^2 + \lambda^4, \lambda^4\lambda_q^2], 0) \\ \text{reiner Schub:} & \quad ([\lambda^2, \lambda^2\lambda_q^2 + \lambda^2 + \lambda_q^2, \lambda^2\lambda_q^2], P/\lambda) , & \quad ([\lambda_q^2, \lambda^2\lambda_q^2 + \lambda^2 + \lambda_q^2, \lambda^2\lambda_q^2], 0) \end{aligned}$$

Zur Kalibrierung des KNN-Materialmodells stehen somit insgesamt $P_T = 92$ Trainingsdatenpaare zur Verfügung. Validierungsdatenpaare werden nicht vorgehalten. Die Generalisierungsfähigkeit wird über Materialpunkt-Versuche geprüft.

Wie durch die Untersuchungen in Kapitel 5 gezeigt wurde, wächst die Anzahl nötiger Freiwerte mit dem Wertebereich der abzubildenden Verzerrungen und

Uniaxiale Daten			Äquibiaxiale Daten			Daten reiner Schub		
λ	λ_q	P	λ	λ_q	P	λ	λ_q	P
0,67	1,20	-0,53						
0,75	1,14	-0,36						
0,78	1,12	-0,31	0,88	1,24	-0,27			
0,84	1,08	-0,18	0,93	1,12	-0,14			
0,92	1,04	-0,09	0,99	1,01	-0,02			
1,00	1,00	0,00	1,00	1,00	0,00	1,00	1,00	0,00
1,01	1,00	0,03	1,04	0,94	0,09	1,06	0,95	0,07
1,12	0,95	0,14	1,08	0,88	0,16	1,14	0,90	0,16
1,24	0,91	0,23	1,12	0,83	0,24	1,21	0,86	0,24
1,39	0,86	0,32	1,14	0,81	0,26	1,32	0,80	0,33
1,61	0,81	0,41	1,20	0,74	0,33	1,46	0,73	0,42
1,89	0,75	0,50	1,31	0,64	0,44	1,87	0,60	0,59
2,17	0,71	0,58	1,42	0,56	0,51	2,40	0,49	0,76
2,42	0,67	0,67	1,69	0,42	0,65	2,98	0,41	0,93
3,01	0,61	0,85	1,94	0,34	0,77	3,48	0,36	1,11
3,58	0,56	1,04	2,49	0,22	0,96	3,96	0,32	1,28
4,03	0,53	1,21	3,03	0,16	1,24	4,36	0,30	1,46
4,76	0,50	1,58				4,69	0,28	1,62
5,36	0,47	1,94				4,96	0,27	1,79

Tabelle 7.1: Aufbereitete Versuchsdaten von TRELOAR [172], basierend auf STEINMANN ET AL. [163], mit $\bar{\nu} = 0,45$. Die Rückrechnung der Daten $\lambda < 1$ ist approximativ und basiert nicht direkt auf TRELOAR's Daten. P ist in N/mm^2 angegeben.

der Ausprägung nahezu inkompressiblen Materialverhaltens. Aus einer Reihe von Voruntersuchungen wurde festgestellt, dass eine Topologie von [3-15-15-15-1] mit $n_w = 556$ Gewichten ausreichend ist. Dabei reicht eine wesentlich geringere Anzahl von Gewichten bereits aus, um die gegebenen Kurven abzubilden – vorausgesetzt alle Verzerrungskomponenten werden vorgegeben. Die nötige Flexibilität im Rahmen gemischt gesteuerter Materialpunkt-Versuche bzw. bei der Anwendung innerhalb der FEM wird erst durch eine höhere Anzahl von Neuronen und in Kombination mit den nachfolgend beschriebenen Nebenbedingungen erzielt.

Darüber hinaus ist eine Regularisierung nötig, um ein Auswendiglernen⁸⁰ zu verhindern, da insbesondere hier $n_w \gg P_T$ gilt. Dazu wird die spannungsfreie Referenzkonfiguration mit $S_a(1, 3, 1) = 0$ (vgl. Abschnitt 5.3.1) und Energieerhaltung eingefordert, vgl. Abschnitt 5.3.2. Für letztere werden $P_C = 2000$ NB-Datenpunk-

⁸⁰ Ohne zusätzliche Information ist das Optimierungsverfahren unterbestimmt.

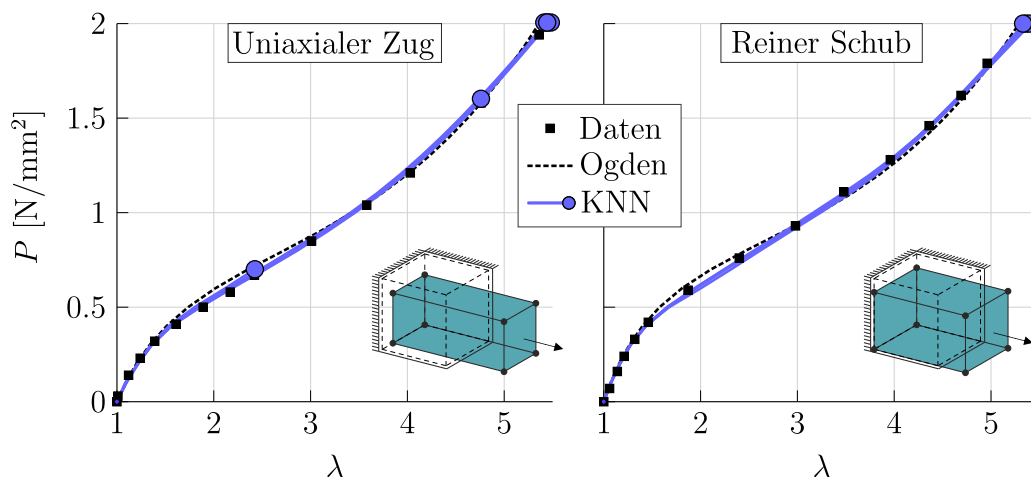


Abbildung 7.2: Zug- und Schubversuch: Vergleich der fünf FNN-E-Iso-Materialmodelle mit den Versuchsdaten sowie einem OGDEN-Modell mit $\Lambda = 3,8$

te zufällig im Trainingsdatenraum Ω erzeugt⁸¹, mit

$$[\lambda_a^2, II_C, J^2] \in \Omega = [0,03; 28,73] \times [2,94; 84,78] \times [0,91; 2,24] . \quad (7.9)$$

Ferner gilt für den Penalty-Parameter $\epsilon = 10$, der durch Trial-and-Error gewählt wurde. Insgesamt werden fünf Trainingsdurchläufe bis maximal 10 000 Epochen durchgeführt. Die Gewichte derjenigen Epoche werden gewählt, die im Rahmen des folgenden Testszenarios den kleinsten Fehler bzgl. der Trainingsdaten aufweisen. Der Versuch des reinen Schubs wird mit dem Materialpunkt-Versuch aus Abschnitt 5.1.1 nachgestellt. Die Spannungen unterscheiden sich durch die gemischte Steuerung von denjenigen Spannungen, die durch direktes Einsetzen der Eingangsvektoren erzeugt werden. Dieser Test liefert dadurch einen guten Anhaltswert, ob das KNN-Materialmodell stabil genug ist, innerhalb der FEM eingesetzt zu werden. Das Training der KNN-Materialmodelle erfolgt in MATLAB [170] mithilfe des QN-Verfahrens, vgl. Abschnitt 4.4.2.

7.1.3 Anwendung des KNN-Materialmodells auf die Daten

In Abbildung 7.2 sind die Resultate von uniaxialen Zugversuchen sowie reinen Schubversuchen abgebildet. Die Versuche wurden nicht mit dem Materialpunkt-Versuch nach Abschnitt 5.1.1 durchgeführt, sondern mit einem entsprechend gelagerten einzelnen finiten Volumenelement in FEAP [168], welches lastgesteuert⁸²

⁸¹ Die Werte in Tabelle 7.1 sind gerundet, daher können sich kleine Abweichungen ergeben.

⁸² Strenggenommen verwölben sich die Querschnitte der Einheitswürfel durch das nicht perfekte KNN-Materialmodell. Dadurch ist die Berechnung der Streckungen und der Kräfte nicht eindeutig. In diesem Fall ist der Effekt jedoch vernachlässigbar gewesen. Eine sauberere Verschiebungssteuerung wäre wiederum einfacher für das KNN-Materialmodell.

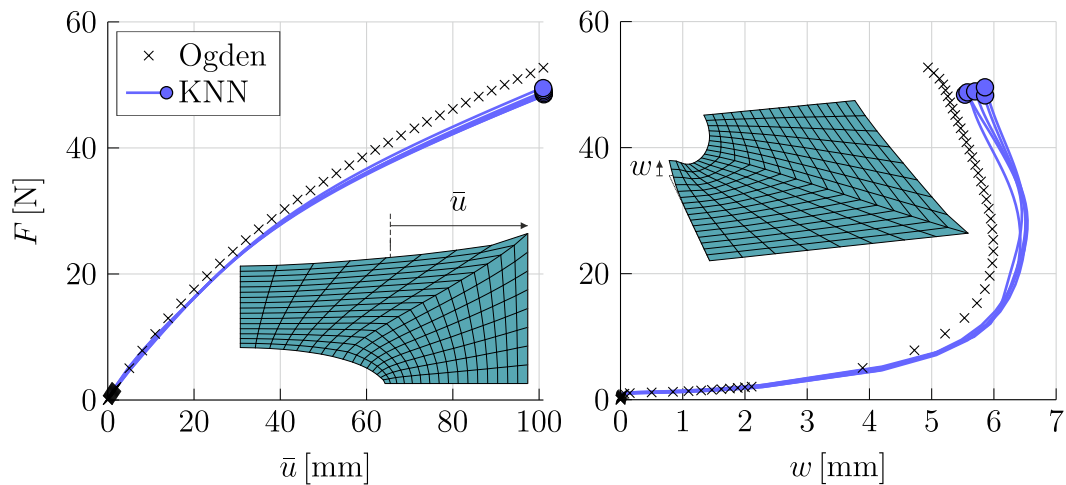


Abbildung 7.3: Scheibe mit Loch: Vergleich der fünf FNN-E-Iso-Materialmodelle mit einem OGDEN-Modell mit $\Lambda = 3,8$

beansprucht wurde. Dabei werden die fünf FNN-E-Iso-Materialmodelle mit den Versuchsdaten sowie einem OGDEN-Modell mit den Parametern nach Tabelle 5.1 verglichen. Der 1. LAMÉ-Parameter wurde zu $\Lambda = 3,8$ gewählt, um mit $\nu = 0,45$ konsistent zu sein.

Wie zu erkennen ist, liegen alle Modelle nahe beieinander und approximieren die gegebenen Daten zufriedenstellend. Die Streuung unter den KNN-Materialmodellen ist dabei vergleichsweise gering. Allerdings divergiert im Fall des uniaxialen Zugversuchs die FE-Berechnung bei zwei von fünf Modellen vor Erreichen der Maximallast. Dies ist mit ausgefüllten Kreisen gekennzeichnet.

In Abbildung 7.3 sind die fünf FNN-E-Iso-Materialmodelle dem OGDEN-Modell am Beispiel der Lochscheibe aus Abschnitt 5.5.1 gegenübergestellt. Dabei wurden alle Einstellungen so übernommen, wie sie in besagtem Abschnitt beschrieben sind. Keines der fünf KNN-Materialmodelle divergiert vor Erreichen der Maximallast. Zwar streuen sie nicht stark untereinander, zeigen aber im $F(\bar{u})$ -Diagramm eine Abweichung von ca. 6% gegenüber der Lösung mittels OGDEN-Modell. Das $F(\bar{u})$ -Diagramm ist dabei eine globale Tragwerksantwort, während das $F(w)$ -Diagramm einen lokalen Effekt repräsentiert.

Auf der einen Seite sind beide Modelle zunächst gleichwertig, da es zum Problem der Scheibe mit Loch keine Versuchsdaten mit TRELOARS Gummimaterial gibt. Beide Modelle sagen dementsprechend das Verhalten der Struktur vorher, auf Basis der zur Verfügung stehenden Versuchsdaten von einfachen Experimenten. Auf der anderen Seite ist das KNN-Materialmodell sehr sensibel bzgl. der gewählten Hyperparameter wie bspw. der Topologie. Gerade im Fall weniger Trainingsdaten ist alles, was außerhalb der trainierten Lastpfade passiert, schwierig zu interpretieren.

tieren. Die Nebenbedingungen beeinflussen ebenfalls das Ergebnis, auch wenn sie hier nur als Regularisierung genutzt wurden. Ohne Nebenbedingungen konnte kein numerisch stabiles KNN-Materialmodell mit den gegebenen Daten trainiert werden.

7.1.4 Abschließende Bemerkungen

- Um ein auf Invarianten basierendes KNN-Materialmodell zu trainieren, müssen die Daten des reinen Schubs nicht in die eines einfachen Schubs umgerechnet werden, um Informationen über Gleitungen und die zugehörigen Schubspannungen zu enthalten⁸³.
- Durch die geringe Anzahl an Trainingsdatenpaaren reagiert das KNN sehr sensibel auf eine Änderung der Hyperparameter. Außerdem ist stets eine Form von Regularisierung notwendig, entweder durch Nebenbedingungen oder durch eine L2-Regularisierung. Das polykonvexe KNN-Materialmodell aus Abschnitt 5.2.4 verhält sich dabei im Übrigen weniger anfällig als die anderen Modelle, da die Polykonvexitätsbedingung recht einschränkend ist. Das Modell leidet aber an den in Kapitel 5 angesprochenen Grenzen der Approximationsgüte.
- Bei Nutzung der SoftPlus-Aktivierungsfunktion konnten wesentlich kleinere Topologien gewählt werden. Allerdings verhalten sich diese häufig numerisch instabiler im Rahmen einer Anwendung innerhalb der FEM.
- Die Umrechnung in quasi-inkompressible Daten und die anschließende Approximation eines vollständigen Materialmodells stellen eine größere Herausforderung dar als das Training auf die rein deviatorischen Anteile. Das ist womöglich auch der Grund, wieso SHEN ET AL. [156] sowie LINKA ET AL. [118] augenscheinlich weniger Probleme bei der Approximation von TRELOARS Daten haben: Sie trainieren ausschließlich auf die deviatorischen Anteile der Verzerrungen. Zur Nutzung in der FEM ist dazu ein gemischt formuliertes Element notwendig, welches den hydrostatischen Druck als unabhängige Feldgröße beschreibt. Vor dem Hintergrund, dass sich Gummi nahezu ideal inkompressibel verhält, ist diese Vorgehensweise sicherlich besser, aber auch einschränkender im Hinblick auf andere Materialien.

⁸³ WEBER ET AL. [189] haben zum Training eines in den kartesischen Verzerrungs- und Spannungskomponenten \mathbf{E} und \mathbf{S} definierten KNN-Materialmodells (vgl. Abschnitt 5.2.1) diese Umformung durchgeführt. Dazu müssen allerdings bei großen Verzerrungen die Normalspannungen abgeschätzt werden. Dies ist mit einer Annahme von HORGAN und MURPHY [84] gemacht worden. Dies führte zu plausiblen Resultaten, ist jedoch ein weiterer Eingriff im Rahmen der Datenaufbereitung.

7.2 Biaxial beanspruchtes Fichtenholz

Seit Jahrtausenden wird Holz als Werkstoff u. a. in den Bereichen des Gebäude- und Schiffbaus genutzt. Auf der Klaviatur der Konstruktionsbaustoffe ist es im Bauingenieurwesen nicht mehr wegzudenken. Darüber hinaus hat Holz als Baumaterial aus dem Blickwinkel des modernen Klimaschutzes Vorteile gegenüber Stahl und Stahlbeton: Holz ist ein nachwachsender Rohstoff. Zudem speichert es CO_2 und hilft somit bei der Eindämmung des Treibhauseffekts.

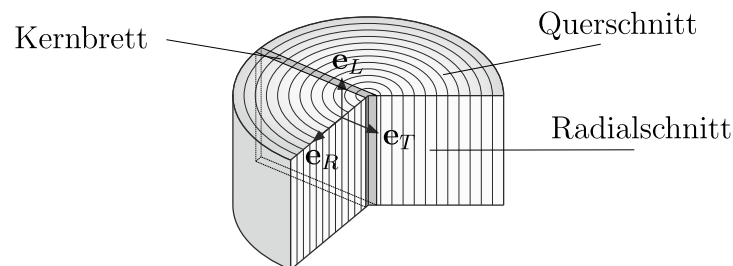


Abbildung 7.4: Kennzeichnung der Materialhauptrichtungen

Auf der Mikroebene setzt sich Holz aus längs zur Stammachse ausgerichteten Zellen zusammen, den sog. Tracheiden. Dieser u. a. zum Wasser- und Nährstofftransport dienenden Struktur verdankt der Werkstoff Holz ebenfalls seine makroskopischen Festigkeits- und Steifigkeitseigenschaften. Diese sind in Stammlängsrichtung (L) wesentlich höher als in radialer (R) und tangentialer (T) Richtung, vgl. Abbildung 7.4 und BLASS ET AL. [18]. Daher wird Holz i. d. R. orthotrop oder vereinfacht als transversalisotrop modelliert. Zusätzlich unterscheidet sich das Festigkeitsverhalten unter Zugbelastung von dem unter Druckbelastung. Weiter haben vor allem technologische Parameter wie Rohdichte, Holzfeuchtigkeit, Ästigkeit usw. einen großen Einfluss auf die letztlich resultierenden Eigenschaften des Holzes als Konstruktionswerkstoff, vgl. GLOS [62]. Zusammenfassend ist Holz ein anisotroper, natürlich gewachsener und von Umwelteinflüssen abhängiger Werkstoff und stellt folglich eine Herausforderung bei der Erstellung leistungsfähiger Materialmodelle dar.

In den 1990er Jahren wurde am Institut für Festigkeitslehre der TU Wien schwerpunktmäßig an den Steifigkeits- und Festigkeitseigenschaften von Holz geforscht. Dabei wurden insgesamt 439 proportionale⁸⁴ und 16 nichtproportionale Versuche durchgeführt, bei denen Fichtenholz biaxial und schräg zur Faserrichtung beansprucht wurde. Darauf aufbauend haben u. a. MACKENZIE-HELNWEIN ET AL. [122] und MUELLNER ET AL. [136] Ein- und Mehrflächenmodelle entwickelt,

⁸⁴ Monotone Belastungssteigerung bis zum Versagen der Probe

um das anisotrope elastoplastische Verhalten von Fichtenholz zu beschreiben, inklusive der Berücksichtigung von Schädigung. Die Versuchsdaten der biaxial beanspruchten Holzscheiben, bestehend aus Bruchspannungszuständen und Spannungs-Verzerrungs-Kurven, sowie die zugehörigen Auswertungen hat EBERHARDSTEINER [40] im Rahmen seiner Habilitationsschrift veröffentlicht. Darüber hinaus hat EBERHARDSTEINER die Dateien mit den zugehörigen Rohdaten zur Verfügung gestellt. Dafür möchte ich mich an dieser Stelle noch einmal herzlich bedanken. Diese Daten bilden die Grundlagen für die folgenden Ausführungen.

In Abschnitt 7.2.1 werden die Versuche sowie die Aufbereitung und Auswahl der zugehörigen Daten beschrieben. Die grundsätzliche Unterscheidung zwischen dem, was man modellieren *kann* und was man modellieren *möchte*, folgt am Beispiel einer linearen Regression in Abschnitt 7.2.2. In Abschnitt 7.2.3 wird ein nichtlineares KNN-Materialmodell zur Approximation von Fichtenholz beschrieben und angewendet. Dabei beschränken sich alle Abschnitte auf den ESZ, was durch die Informationen aus den Versuchen limitiert ist⁸⁵. Da alle Verzerrungen klein sind und keine großen Verschiebungen berücksichtigt werden müssen, werden nachfolgend die infinitesimalen Spannungs- und Verzerrungsgrößen σ und ϵ genutzt. Außerdem werden ausschließlich die proportionalen Versuche zur Erstellung eines Materialmodells für proportionale Laststeigerung verwendet.

7.2.1 Versuchsdaten: Beschreibung und Aufbereitung

In Abbildung 7.5 sind die kreuzförmigen Probekörper der biaxialen Versuche dargestellt. Über insgesamt zwölf Lasteinleitungspunkte wurde der Probekörper über Bolzen verschiebungsgesteuert beansprucht. Details zur servohydraulischen Belastungseinrichtung sind EBERHARDSTEINER [40] zu entnehmen. Zur Vermeidung ungewollter Schädigungen hatte der Messbereich eine wesentlich geringere Wandstärke als die Bereiche der Lasteinleitung, die ferner durch Stahlplättchen verstärkt worden sind. Durch Schlitze wurden ungewollte Lastumlagerungen vermieden. Die Verschiebungen orthogonal zu den Probekörperrändern sind linear und antimetrisch, die Verschiebungen tangential zu den Rändern sind jeweils konstant. Dadurch ergeben sich insgesamt periodische Verschiebungsrandbedingungen⁸⁶. Das Verhältnis $\kappa = \bar{u}/\bar{v}$ der jeweils mittigen Verschiebungen \bar{u} und \bar{v}

⁸⁵ Wäre zusätzlich die Dehnung der Scheiben in Dickenrichtung gemessen worden, könnten unter Annahme von Transversalisotropie die Datensätze zur Erstellung eines dreidimensionalen Stoffgesetzes erweitert werden.

⁸⁶ Die genauen Verläufe der aufgetragenen Verschiebungen sowie die Geometrie des Probekörpers wurden mit numerischen Methoden so optimiert, dass im Messbereich möglichst homogene Verzerrungs- und Spannungsfelder vorherrschen. Ohne diese Homogenität ließe sich vom makroskopischen Probekörper nicht auf den theoretischen Materialpunkt schließen.

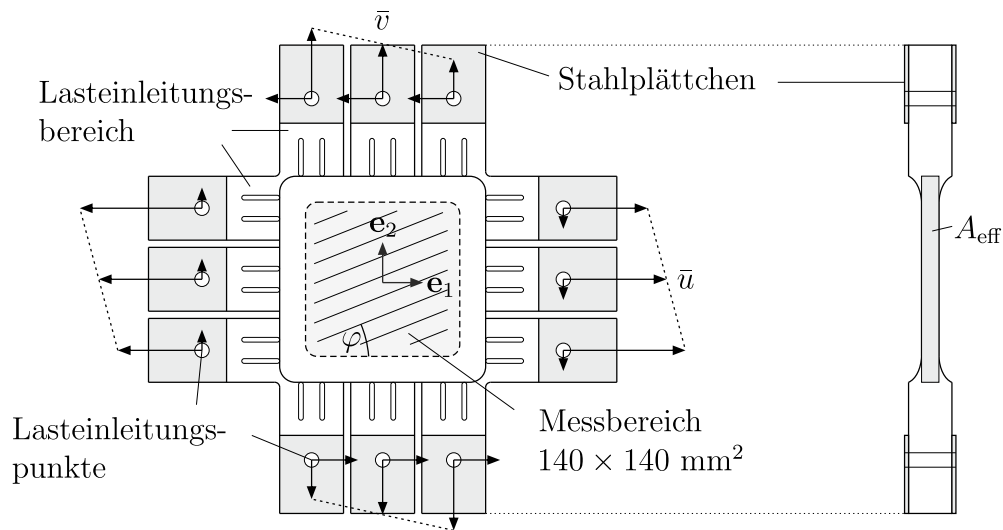


Abbildung 7.5: Skizze der biaxial geprüften Fichtenholzproben, vgl. EBERHARDSTEINER [40]

dient neben dem Winkel zur Faserrichtung φ als Charakterisierung des Versuchs.

Die Probekörper waren Kernbretter⁸⁷. Kernbretter werden aus der L - R -Ebene herausgeschnitten, vgl. Abbildung 7.4. Darüber hinaus bestanden die Probekörper aus technologisch fehlerfreiem, homogenen Fichtenholz und wiesen durch kontrollierte Klimatisierung eine Holzfeuchte von 12% auf. Die Jahresringverläufe im $140 \times 140 \text{ mm}^2$ großen Messbereich waren parallel und äquidistant. Damit ist theoretisch nur noch die Dichte ρ der Probekörper als Einflussfaktor für die Steifigkeits- und Festigkeitseigenschaften übrig. Diese wurde vor den Versuchen gemessen und dokumentiert, ihr Einfluss wird im Rahmen dieses Abschnitts allerdings in erster Näherung vernachlässigt.

Ein Versuch besteht aus der Einrichtungphase, der Vorspannphase sowie der Belastungsphase. Die ersten beiden Phasen dienen der Justierung des Probekörpers, welcher dadurch eine geringe, allseitig konstante Vorspannung erfährt. Erst in der Belastungsphase werden die Verschiebungen dann mit konstanter Geschwindigkeit aufgebracht. Dabei wird die Dehnung gemessen. Diese Messung erfolgte über eine dreidimensionale Deformationsanalyse mittels elektronischer Speckle-Interferometrie. Eine ausführliche Beschreibung der Methode sowie der nötigen Auswertungsschritte ist in EBERHARDSTEINER [40] dokumentiert. Die Spannungen werden über die an den Belastungspunkten gemessenen Kräfte des hydraulischen Belastungssystems und der effektiven Fläche A_{eff} des Probekörpers ermittelt, vgl. Abbildung 7.5. Dabei wurde die Form der Verschiebungsrandbedingungen so er-

⁸⁷ Die Probekörper wurden durch Verleimung zweier Einzelbretter erzeugt.

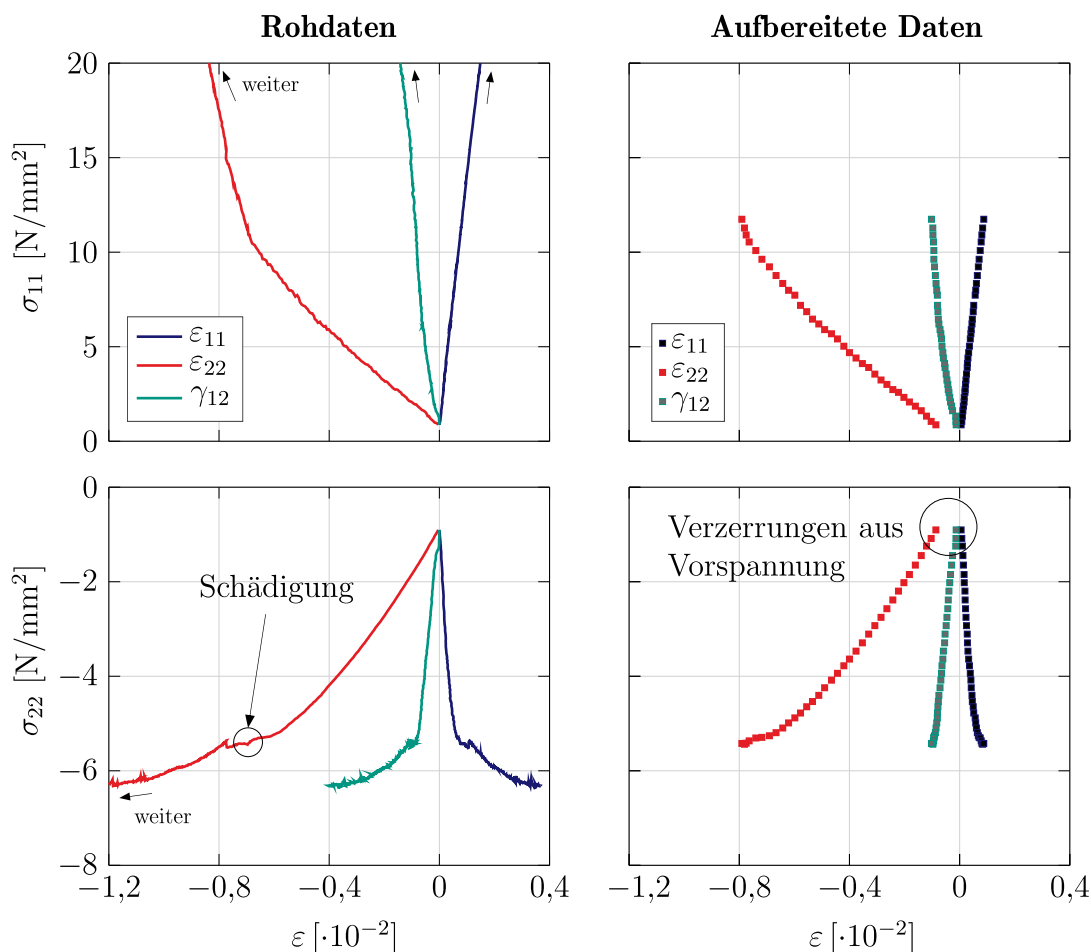


Abbildung 7.6: Aufbereitung der Rohdaten am Beispiel von Versuch 380, mit $\varphi = 7,5^\circ$ und $\kappa = 3/-10$

mittelt, dass Schubdeformationen geringstmöglich behindert wurden: Die Schubspannung τ_{12} wird daher zu konstant null angenommen.

Für die insgesamt 455 Versuche wurden verschiedene Belastungsverhältnisse $\kappa = \bar{u}/\bar{v}$ und Winkel $\varphi \in \{0^\circ; 7,5^\circ; 15^\circ; 30^\circ; 45^\circ\}$ zwischen Faserrichtung L und der \mathbf{e}_1 -Achse eingestellt. Jede Konfiguration wurde dabei mindestens sechs Mal durchgeführt. Für jeden Versuch wurden die globalen linearen Verzerrungen $\boldsymbol{\varepsilon}_G = [\varepsilon_{11}, \varepsilon_{22}, \gamma_{12}]^T$ sowie die globalen Spannungen $\boldsymbol{\sigma}_G = [\sigma_{11}, \sigma_{22}, 0]^T$ dokumentiert, mit der Annahme verschwindender globaler Schubspannungen. Dabei entspricht $\gamma_{12} = 2\varepsilon_{12}$ der Gleitung. Daher ist das 1-2-Koordinatensystem stets das Hauptachsensystem.

Diese im globalen Koordinatensystem definierten Rohdaten müssen nun zunächst aufbereitet werden. Dies ist in Abbildung 7.6 am Beispiel von Versuch 380 dargestellt, mit $\varphi = 7,5^\circ$ und $\kappa = 3/-10$. Dazu sind die drei folgenden Schritte notwendig, die ebenfalls in Abbildung 7.6 dargestellt sind:

1. Identifikation der ersten Schädigung und Abschneiden der restlichen Daten. Dies wird hier automatisiert und rein geometrisch über die relative Spannungsänderung der einzelnen GGW-Punkte durchgeführt. Dabei wird in erster Näherung keine Unterscheidung in Schädigung infolge Zug und Schädigung infolge Druck gemacht, anders als von EBERHARDSTEINER.
2. Abschätzung der Verzerrung durch die Vorspannphase durch eine lokale lineare Regression und anschließende Verschiebung aller Verzerrungen um diesen Wert. Die Spannungen werden nicht geändert, da diese bereits in den wahren Spannungen angegeben sind, inklusive Vorspannung. Infolge etwaiger Ungenauigkeiten gehen dabei die die initialen Daten mittelnden Geraden nicht unbedingt für beide Spannungen σ_{11} und σ_{22} durch den Ursprung.
3. Reduktion der Datenpunkte pro Versuch auf eine vorher festgelegte mittlere Menge. Die Anzahl der Datenpunkte unterscheidet sich zwischen den Versuchen teilweise erheblich, vor allem nach dem automatisierten Abschneiden. Dies führt dazu, dass das KNN Kurven mit mehr Datenpunkten favorisiert lernt. Diese Problematik ist in Abbildung 7.7 an einem Modellbeispiel mit linearer Regression dargestellt. Dabei wird dieses Phänomen immer dann relevant, wenn durch material- oder versuchsbedingte Streuungen Kurven bei gleichen Einstellungen weiter auseinander liegen. Daher ist es wichtig, dass alle Kurven mit einer ähnlichen Anzahl von Punkten diskretisiert sind⁸⁸. Für die vorliegenden Versuche wurden stets ca. 30 Datenpunkte pro Kurve gewählt.

Die beschriebene Aufbereitung der Daten ist durchaus subjektiv. Sie kann bspw. beim automatisierten Abschneiden sicherlich noch verfeinert werden. Nach Aufbereitung der globalen Daten müssen die im globalen Koordinatensystem definierten Größen ϵ_G und σ_G in die des lokalen Materialhauptachsensystems umgerechnet werden⁸⁹. Im Fall der vorliegenden Daten ist dies das \mathbf{e}_L - \mathbf{e}_R -Koordinatensystem, das in Abbildung 7.4 eingeführt wurde. Ohne dazu die in Kapitel 3.1 beschriebenen tensoriellen Zusammenhänge bedienen zu müssen, kann die Transformation der Verzerrungen mithilfe der in Abbildung 7.8 dargestellten Beziehungen hergeleitet werden. So folgt bspw. für die Verzerrung in lokale Faserrichtung mithilfe

⁸⁸ Strenggenommen wäre es noch vorteilhafter, wenn alle Datenpunkte gleich weit von ihren Nachbarn entfernt wären. Eine solche Datenaufbereitung könnte im Rahmen weiterer Forschungsaktivitäten entwickelt werden.

⁸⁹ Theoretisch könnte zum Training des KNN jedes Referenzsystem gewählt werden. In diesem Fall bietet sich dasjenige System an, das sich an der Mikrostruktur orientiert.

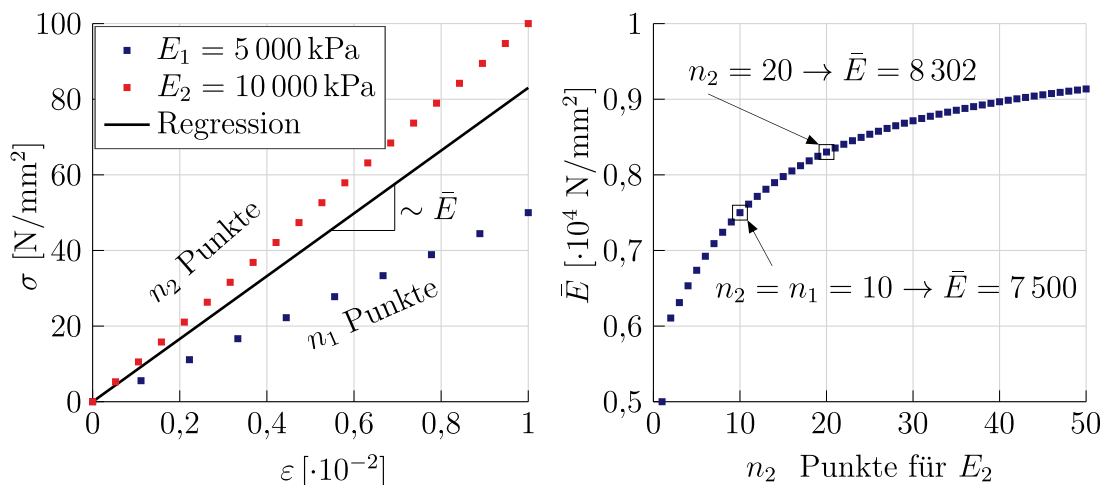


Abbildung 7.7: Problematik von Pfaden mit nicht ausgewogener Anzahl von Datenpunkten: Lineare Regression zwischen zwei Modellpfaden (links) und Verhalten des E-Moduls der Regression in Abhängigkeit von n_2 mit $n_1 \equiv 10$ (rechts)

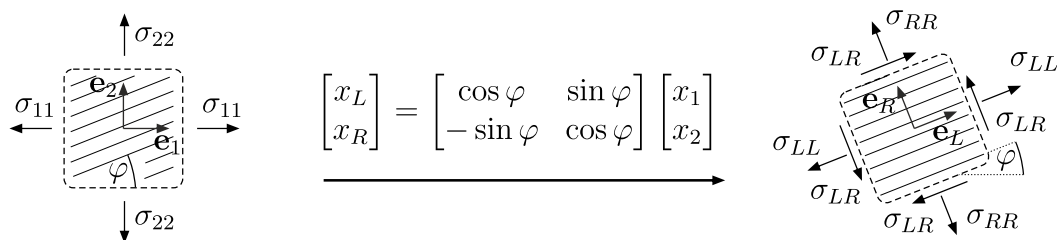


Abbildung 7.8: Transformation von Vektoren zwischen globalem 1-2 und lokalem L - R -Koordinatensystem. An den Holzelementen sind aus Übersichtsgründen nur die Spannungskomponenten abgebildet.

der Kettenregel zunächst

$$\varepsilon_{LL} = \frac{\partial u_L}{\partial x_L} = \frac{\partial(\cos \varphi u_1 + \sin \varphi u_2)}{\partial x_1} \frac{\partial x_1}{\partial x_L} + \frac{\partial(\cos \varphi u_1 + \sin \varphi u_2)}{\partial x_2} \frac{\partial x_2}{\partial x_L}. \quad (7.10)$$

Wird die Gleichung weiter umgeformt und das Vorgehen ebenfalls für ε_{RR} und $\gamma_{LR} = 2\varepsilon_{LR}$ durchgeführt, resultiert die Transformationsbeziehung

$$\underbrace{\begin{bmatrix} \varepsilon_{LL} \\ \varepsilon_{RR} \\ \gamma_{LR} \end{bmatrix}}_{=: \boldsymbol{\varepsilon}_L} = \underbrace{\begin{bmatrix} \cos^2(\varphi) & \sin^2(\varphi) & \sin \varphi \cos \varphi \\ \sin^2(\varphi) & \cos^2(\varphi) & -\sin \varphi \cos \varphi \\ -2 \sin \varphi \cos \varphi & 2 \sin \varphi \cos \varphi & \cos^2(\varphi) - \sin^2(\varphi) \end{bmatrix}}_{=: \mathbf{T}_\varepsilon} \underbrace{\begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \gamma_{12} \end{bmatrix}}_{=: \boldsymbol{\varepsilon}_G} \quad (7.11)$$

zur Berechnung der lokalen Verzerrungen $\boldsymbol{\varepsilon}_L$ ⁹⁰. Die Transformationsmatrix \mathbf{T}_ε gilt für die Verzerrungen. Die Transformationsmatrix für die Spannungen kann

⁹⁰Der Index L für die lokalen Verzerrungs- und Spannungsmaße ist nicht zu verwechseln mit dem Index L für die lokale Faserrichtung.

bspw. durch Forderung der inkrementellen Arbeitsäquivalenz

$$\boldsymbol{\sigma}_L^T d\boldsymbol{\varepsilon}_L \stackrel{!}{=} \boldsymbol{\sigma}_G^T d\boldsymbol{\varepsilon}_G \quad (7.12)$$

hergeleitet werden. Wird dort Gleichung (7.11) eingesetzt, folgt letztlich für die Transformation der Spannungen

$$\boldsymbol{\sigma}_L = (\mathbf{T}_\varepsilon)^{-T} \boldsymbol{\sigma}_G . \quad (7.13)$$

Diese Zusammenhänge können jedem Textbuch für orthotropes Materialverhalten entnommen werden. Somit resultiert für jeden Versuch ein Pfad

$$\mathcal{P} = \{(\boldsymbol{\varepsilon}_L, \boldsymbol{\sigma}_L)_1, (\boldsymbol{\varepsilon}_L, \boldsymbol{\sigma}_L)_2, \dots, (\boldsymbol{\varepsilon}_L, \boldsymbol{\sigma}_L)_M\} , \quad (7.14)$$

bestehend aus $M \approx 30$ lokalen GGW-Punkten. Von den insgesamt 439 proportionalen Versuchen werden 394 für die nachfolgenden Untersuchungen verwendet. Die restlichen Pfade werden visuell herausgefiltert, da sie durch die automatisierte Normalisierung nicht korrekt erfasst oder nicht korrekt gekürzt wurden. Dies ist nachfolgend kein Nachteil, da dessen ungeachtet die meisten Winkel φ und Beanspruchungsverhältnisse κ in den Daten vorhanden sind. Mit ausgefeilteren Methoden könnten sicherlich weitere Rohdaten zum Training des KNN aufbereitet werden.

7.2.2 Eine Grundsatzdiskussion anhand der linearen Approximation

Ohne Zwischenschichten fungiert die KNN-Abbildung

$$\boldsymbol{\sigma}_L = \mathbf{f}^{\text{KNN}}(\boldsymbol{\varepsilon}_L, \mathbf{w}) = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_{LL} \\ \varepsilon_{RR} \\ \gamma_{LR} \end{bmatrix} - \begin{bmatrix} w_{10} \\ w_{20} \\ w_{30} \end{bmatrix} \quad (7.15)$$

als mehrdimensionale lineare Regression⁹¹. Die Trainingsdatennormalisierung im Rahmen des KNN-Trainings wird hier zunächst vernachlässigt. Anhand dieser linearen Regression soll nachfolgend eine Grundsatzdiskussion geführt werden: Ist die *mathematisch* beste Lösung auch gleichzeitig *physikalisch* erstrebenswert? Dazu werden die Versuche 116 ($\varphi = 15^\circ$, $\kappa = 5/-4$), 308 ($\varphi = 30^\circ$, $\kappa = -3/-5$) und 335 ($\varphi = 30^\circ$, $\kappa = 5/-4$) genutzt, um ein linear elastisches Materialmodell zu erstellen⁹². Von den Versuchen werden dabei jeweils nur diejenigen Datenpunkte berücksichtigt, die optisch auf einer Geraden liegen. Die Normierung erfolgt als

⁹¹ Sind die linearen Daten der Spatz, ist das KNN in diesem Fall redensartlich die Kanone.

⁹² Ein einzelner Versuch bietet drei Gleichungen. Da das KNN ohne Zwischenschichten zwölf Gewichte hat, reichen drei Versuche mit unterschiedlichen Konfigurationen (φ, κ) und die Nebenbedingung $\boldsymbol{\sigma}_L(\mathbf{0}) = \mathbf{0}$ aus, um das Training sauber zu definieren.

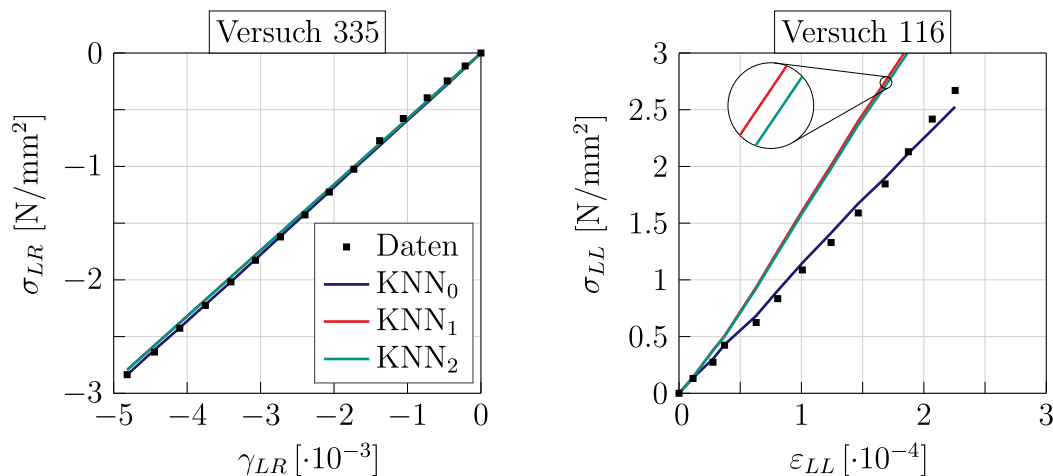


Abbildung 7.9: Auswirkung der Nebenbedingungen auf die lineare Regression am Beispiel zweier Spannungs-Dehnungs-Kurven

Näherung durch Abzug des ersten GGW-Punkts von allen Datenpaaren. Somit starten alle drei Pfade bei $(\mathbf{0}, \mathbf{0})$.

Die folgenden KNN-Konfigurationen werden miteinander verglichen:

- **KNN₀**: Ausschließlich die Normalisierung $\sigma_L(\mathbf{0}) = \mathbf{0}$ wird eingefordert über $w_{10} = w_{20} = w_{30} \stackrel{!}{=} 0$. Somit bleiben neun freie Gewichte zum Training.
- **KNN₁**: Zusätzlich zu KNN₀ wird Orthotropie eingefordert mit $w_{13} = w_{23} = w_{31} = w_{32} \stackrel{!}{=} 0$ über die Unabhängigkeit von Gleitungen und Dehnungen in den Materialhaupttrichtungen. Dies kann mit den in Abschnitt 5.3 vorgestellten Methoden hergeleitet und implementiert werden. Eine sinngemäße Entkopplung per Konstruktion wurde auch von FREITAG ET AL. [50] durchgeführt. Somit bleiben fünf freie Gewichte, was einem orthotropen Stoffgesetz entspricht.
- **KNN₂**: Zusätzlich zu KNN₁ wird Energieerhaltung eingefordert mit $w_{12} \stackrel{!}{=} w_{21}$, wodurch vier freie Gewichte zum Training verbleiben.

Alle Nebenbedingungen werden mit $P_C = 1000$ NB-Punkten eingefordert, welche mittels LHS in $\Omega = [-10^{-2}; 10^{-2}]^3$ erzeugt werden. Der Penalty-Parameter ist $\epsilon = 10^4$.

In Abbildung 7.9 werden die drei KNN-Materialmodelle anhand des $\sigma_{LR}(\gamma_{LR})$ -Verlaufs von Versuch 335 und des $\sigma_{LL}(\varepsilon_{LL})$ -Verlaufs von Versuch 116 miteinander verglichen. Da zur Auswertung der KNN jeweils die gemessenen Verzerrungspunkte der Versuche verwendet wurden, entspricht der Graph nicht einer

perfekten Geraden. In Tabelle 7.2 sind die linearen Materialtangente der jeweiligen KNN dargestellt. Im Fall gültiger Orthotropiebedingungen sind ebenfalls die Materialparameter angegeben, welche mit der Nachgiebigkeitsmatrix

$$\mathbf{C}^{-1} = \begin{bmatrix} 1/E_L & -\nu_{RL}/E_R & 0 \\ -\nu_{LR}/E_L & 1/E_R & 0 \\ 0 & 0 & 1/G_{LR} \end{bmatrix} \quad (7.16)$$

direkt ermittelt werden können, vgl. Abschnitt 3.3.2. E_L und E_R sind dabei die E-Moduln in Faserrichtung und quer dazu, G_{LR} der Schubmodul und ν_{LR} sowie ν_{RL} die Querdehnzahlen. Die Werte sollten dabei aufgrund der geringen Anzahl von Versuchen, der starken Streuung der Werkstoffeigenschaften zwischen den einzelnen Versuchen und der Vernachlässigung des Einflusses der Rohdichte nur relativ zueinander und nicht absolut interpretiert werden.

	Lineare Materialtangente \mathbf{C}	Parameter
KNN ₀	$\begin{bmatrix} 7696 & 1187 & -621 \\ -1796 & 83 & -129 \\ 641 & 148 & 599 \end{bmatrix}$ N/mm ²	nicht orthotrop
KNN ₁	$\begin{bmatrix} 17\,662 & 2161 & 0 \\ 278 & 347 & 0 \\ 0 & 0 & 580 \end{bmatrix}$ N/mm ²	$E_L = 15\,695$ N/mm ² $E_R = 309$ N/mm ² $G_{LR} = 580$ N/mm ² $\nu_{LR} = 0,80$ $\nu_{RL} = 0,14$
KNN ₂	$\begin{bmatrix} 17\,399 & 2417 & 0 \\ 2417 & 618 & 0 \\ 0 & 0 & 580 \end{bmatrix}$ N/mm ²	$E_L = 7951$ N/mm ² $E_R = 282$ N/mm ² $G_{LR} = 580$ N/mm ² $\nu_{LR} = 3,91$ $\nu_{RL} \approx 0,14 = \nu_{LR}E_R/E_L$

Tabelle 7.2: Lineare Materialtangente und zugehörige orthotrope Materialparameter für die drei KNN-Materialmodelle. Die Orthotropie- und Symmetriebedingungen sind bis auf die fünfte Nachkommastelle erfüllt.

Die nur auf Basis der Versuchsdaten definierte „beste“ Regression wird durch KNN₀ ohne zusätzliche Nebenbedingungen erreicht. Dabei wird vom KNN nur auf Basis der Daten eine Kopplung zwischen Dehnungen und Gleitungen interpretiert, die bei Holz so nicht zu erwarten und auch im Rahmen eines interpretierbaren Materialmodells nicht erwünscht ist. Das kann u. a. am großen Steifigkeitsunterschied zwischen den beiden Materialhaupttrichtungen liegen. Dadurch

wirken sich schon geringe Messfehler sehr stark auf die Komponenten der linearen Materialtangente aus. Dies ist auch der Grund, wieso zur Bestimmung der orthotropen Materialparameter E_L , E_R und ν_{LR} von EBERHARDSTEINER zusätzliche uniaxiale Versuche durchgeführt wurden, anstatt die Parameter direkt aus den biaxialen Versuchen zu berechnen. Die besagte Kopplung wird beim KNN₁ durch eine Nebenbedingung unterdrückt. Die restlichen Werte der linearen Materialtangente ändern sich dabei erheblich. Außerdem weicht in Abbildung 7.9 die KNN-Lösung stärker von den Datenpunkten ab. Dieses Modell wäre aus physikalischer Sicht erstrebenswerter, auch wenn die Regression in Bezug auf die Daten nicht mehr optimal ist. Wird darüber hinaus auch Energieerhaltung eingefordert, verändern sich die in Abbildung 7.9 gezeigten Regressionen kaum, während sich der E-Modul in Faserrichtung sowie die Querdehnzahlen dagegen wesentlich ändern, vgl. Tabelle 7.2. Dabei sticht vor allem die unrealistisch hohe Querdehnzahl $\nu_{LR} = 3,91$ ins Auge.

Dieser Vergleich hat gezeigt, dass eine optimale Regression und eine physikalisch sinnhafte Regression nicht unbedingt deckungsgleich sind. Der Trainingsprozess des KNN wird dabei ohne zusätzliche Informationen stets auf die mathematisch beste Regression abzielen. Dabei gilt nicht unbedingt die Regel, dass grundlegende physikalische Eigenschaften automatisch erlernt werden, wenn die Datenbasis nur breit genug ist. Dies wird in Abschnitt 7.2.3 anhand der nichtlinearen Daten ebenfalls bestätigt. Gründe dafür können die starke Streuungen der Daten und mathematisch nahezu gleichwertige Lösungen sein. Auch deswegen müssen wünschenswerte physikalische Eigenschaften berücksichtigt werden, ob mit Nebenbedingungen oder mit spezifischen KNN-Architekturen.

7.2.3 Anwendung des nichtlinearen KNN-Materialmodells

Insgesamt stehen 14 478 GGW-Punkte ($\boldsymbol{\varepsilon}_L, \boldsymbol{\sigma}_L$) zur Verfügung, von denen 90% zum Training und 10% zur Validierung genutzt werden. Dabei dienen die Validierungsdatenpunkte hier weniger zum Abbruch des Trainingsprozesses als vielmehr zur Einschätzung, ob das KNN die stark streuungsbehafteten Daten gut generalisiert. Das KNN-Materialmodell FNN-E-Ani aus Abschnitt 5.2.1 mit

$$\mathbf{x} = \boldsymbol{\varepsilon}_L \quad \text{und} \quad \mathbf{z} = \boldsymbol{\sigma}_L \quad (7.17)$$

wird im ESZ mit den drei zugehörigen Verzerrungs- und Spannungskomponenten verwendet. Aus Voruntersuchungen hat sich eine Topologie von [3-10-10-3] als ausreichend flexibel herausgestellt, auch wenn die Sensitivität im Rahmen dieses Beispiels vergleichsweise gering ist. Das Training erfolgt mittels QN-Verfahren aus Abschnitt 4.4.2 in MATLAB [170]. Neben den Trainingsdatenpaaren werden die nachfolgenden vier Nebenbedingungen im Trainingsprozess berücksichtigt.

- **Spannungsfreie Referenzkonfiguration:** $\boldsymbol{\sigma}_L(\mathbf{0}) \stackrel{!}{=} \mathbf{0}$, vgl. Abschnitt 5.3.1
- **Symmetrie der Materialtangente:** $\mathbf{C}_T \stackrel{!}{=} \mathbf{C}_T^T$, vgl. Abschnitt 5.3.2. Da die nichtlinearen Effekte ebenfalls durch plastische Deformationen hervorgerufen werden, wäre der Begriff *Energieerhaltung* hier unpassend, obwohl im Rahmen der KNN-Materialmodellierung ein „elastisches“ Modell unter Annahme monotoner Belastung trainiert wird.
- **Orthotropie:** $C_{T13} = C_{T23} = C_{T31} = C_{T32} \stackrel{!}{=} 0$, vgl. Abschnitt 7.2.2.
- **Positive Hauptdiagonalelemente:** $C_{Tii} \stackrel{!}{>} 0, i = 1, 2, 3$. Diese Nebenbedingung wurde bisher noch nicht eingeführt. Sie lässt sich jedoch mit den in Abschnitt 5.3 beschriebenen Methoden herleiten und implementieren. Dabei ist zu beachten, dass es sich um eine Ungleichheits-Nebenbedingung handelt, vgl. Abschnitt 6.2.5. Man könnte diese Nebenbedingung als minimalinvasive Form der positiven Definitheit von \mathbf{C}_T interpretieren. Eigentlich impliziert die positive Definitheit einer Matrix positive Diagonalelemente und nicht umgekehrt. Allerdings wäre die Einforderung der positiven Definitheit von \mathbf{C}_T direkt wesentlich komplizierter.

Für die ersten drei Nebenbedingungen gilt $\epsilon = 10^4$, für die Nebenbedingung der positiven Diagonalelemente $\epsilon = 10^6$. Für die letzten drei Nebenbedingungen wurden $P_C = 10^3$ NB-Punkte mittels LHS im rechteckigen Trainingsraum

$$\Omega = [-0,0053; 0,0166] \times [-0,0244; 0,0286] \times [-0,0174; 0,0248] \quad (7.18)$$

generiert. Dieser entspricht der kleinsten rechteckigen Einhüllenden aller Verzerrungen der genutzten Trainingsdaten.

Der Trainingsprozess wird nach 10 000 Epochen abgebrochen, ohne zusätzliche Abbruchkriterien zu berücksichtigen. In Abbildung 7.10 sind die fünf im Training beteiligten Fehlerterme über die Epochen dargestellt, und zwar einmal mit aktiver Berücksichtigung der Nebenbedingungen und einmal ohne. Der Penalty-Parameter wurde dabei herausgerechnet, um die Werte miteinander vergleichen zu können. Es ist eindeutig zu erkennen, dass trotz der sehr großen Datenbasis physikalische Eigenschaften wie bspw. die Orthotropie *nicht* implizit über die Daten erlernt werden. Aus Sicht des Optimierungsalgorithmus ist es günstiger, diese Bedingungen zu verletzen, um den Trainingsdatenfehler weiter zu reduzieren. Erst mit Einforderung der Nebenbedingungen werden alle erstrebenswerten Eigenschaften⁹³ des Materials vom KNN erlernt, und das, ohne den Trainingsdatenfehler maßgeblich zu verschlechtern.

⁹³ Die Frage, welche Eigenschaften überhaupt erstrebenswert sind, sollte durch den Anwender vorgegeben und u. U. im Rahmen von Voruntersuchungen ermittelt werden.

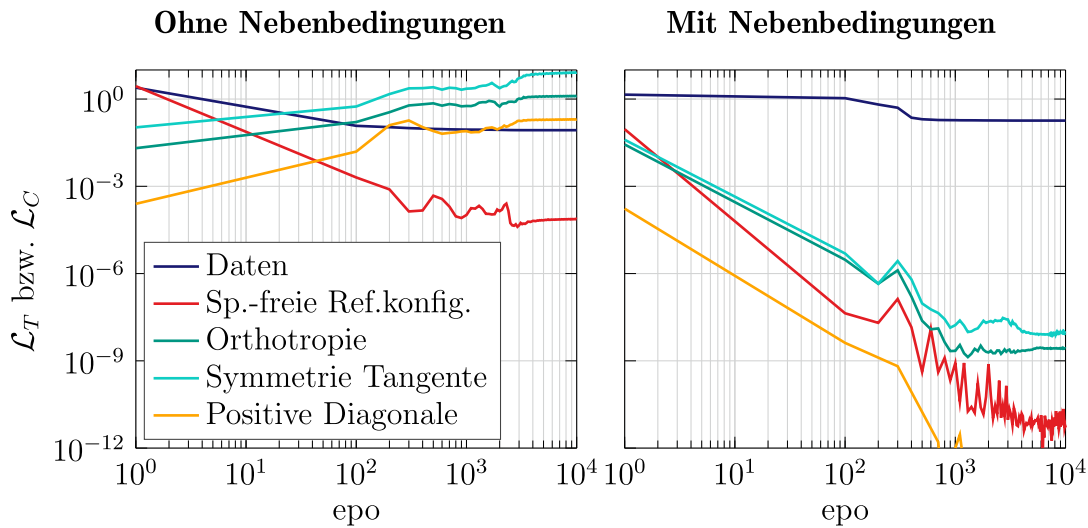


Abbildung 7.10: Fehlerterme bzgl. Trainingsdaten und bzgl. der vier Nebenbedingungen, jeweils mit und ohne Berücksichtigung der Nebenbedingungen im Trainingsprozess

In Abbildung 7.11 sind für den Fall mit Nebenbedingungen und den Fall ohne Nebenbedingungen die KNN-Spannungsfunktionen $\sigma_{LL}(\varepsilon_{LL}, \varepsilon_{RR})$ und $\sigma_{RR}(\varepsilon_{LL}, \varepsilon_{RR})$ für eine schubfreie, biaxiale Beanspruchung in Materialhaupttrichtung gegeben. Dabei sind die jeweils relevanten uniaxialen Kurven schwarz markiert. Die Verzerrungen wurden auf den Bereich $(\varepsilon_{LL}, \varepsilon_{RR}) \in [-0,002; 0,005] \times [-0,01; 0,01]$ begrenzt. Das KNN ohne Nebenbedingungen ist durch die sehr große Datenmenge ebenfalls in der Lage, den prinzipiellen Verlauf der Spannungsfunktionen wiederzugeben. Allerdings sind in diesem Fall wesentlich mehr Unregelmäßigkeiten in der Fläche zu erkennen als beim KNN mit berücksichtigten Nebenbedingungen. Diese unregelmäßige Spannungsfläche lässt auf schlechte numerische Eigenschaften innerhalb der FEM schließen.

Abschließend wird das nichtlineare KNN-Materialmodell für Fichtenholz zur Berechnung eines Kreisbogenträgers im ebenen Spannungszustand innerhalb der FEM eingesetzt. Das System, die Lagerungsbedingungen sowie die Belastung sind in Abbildung 7.12 dargestellt. Der Träger ist zum Zweck der nachfolgenden Untersuchung ausschließlich tangential gelagert, um Spannungsspitzen zu vermeiden. Die konstante Streckenlast q wirkt vertikal und greift an der Oberseite des Trägers an. Die lokalen Materialhauptachsen sind bzgl. eines Polarkoordinatensystems um den Kreismittelpunkt definiert. Die Holzfasern verlaufen somit tangential. Das FE-Netz besteht aus $120 \times 36 = 4320$ isoparametrischen Scheibenelementen mit bilinearen Ansatzfunktionen. In Abbildung 7.12 ist nur ein Drittel der Elemente in radiale Richtung dargestellt. Die hohe Feinheit in radiale Richtung wird zur genaueren Darstellung der Spannungen benötigt, da gerade an den Rändern des Trägers hohe Spannungen auftreten. Die Last q wird monoton gesteigert. Die

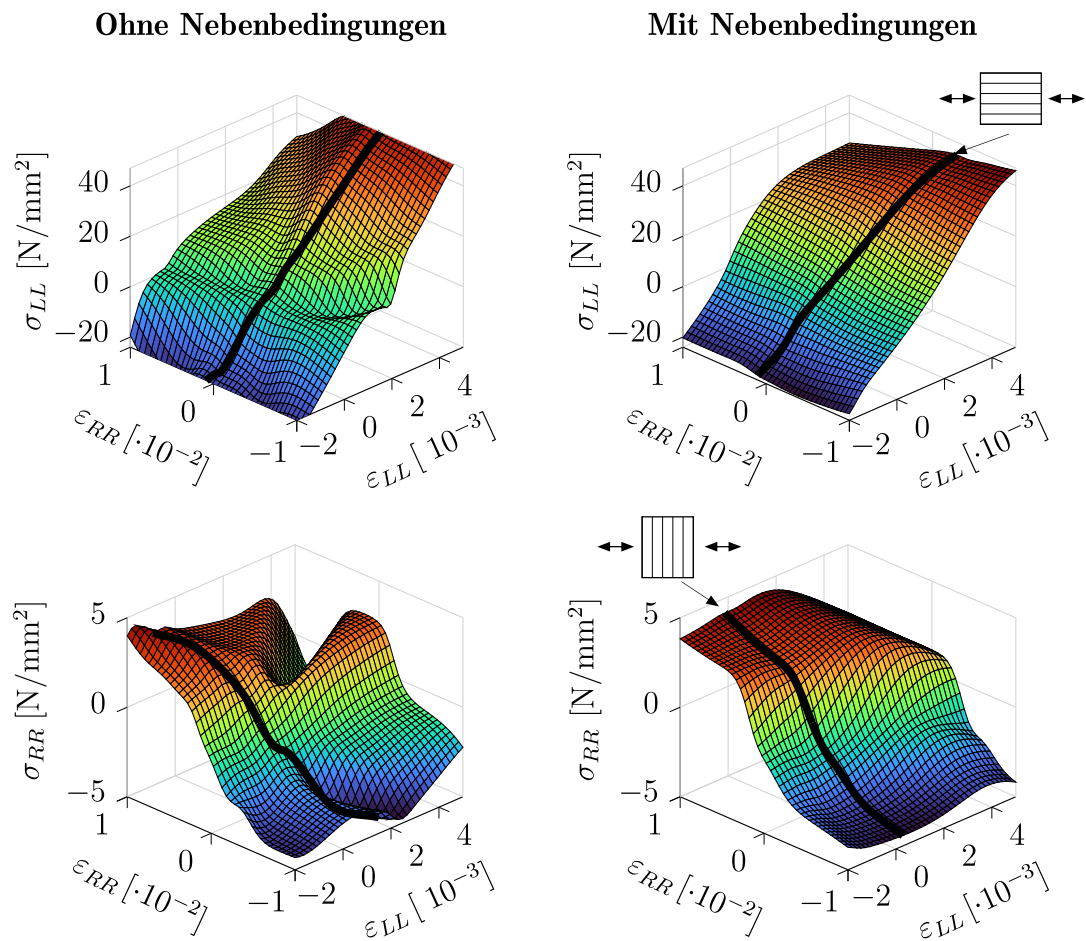


Abbildung 7.11: KNN-Spannungsfunktionen $\sigma_{LL}(\varepsilon_{LL}, \varepsilon_{RR})$ und $\sigma_{RR}(\varepsilon_{LL}, \varepsilon_{RR})$ für den Fall mit Nebenbedingungen und den Fall ohne Nebenbedingungen. Die jeweils relevanten uniaxialen Kurven sind schwarz markiert.

Berechnung wird in FEAP [168] durchgeführt und ist geometrisch linear.

In Abbildung 7.13 sind für beide KNN-Materialmodelle die Last-Verschiebungskurven für die Verschiebungen $w_o(q)$ und $w_u(q)$ an der Ober- und Unterseite des Trägers dargestellt. Darüber hinaus sind für den Fall des KNN mit Nebenbedingungen jeweils nach dem fünften und zwanzigsten Lastschritt die zehnfach überhöhten Verschiebungsfiguren abgebildet. Das KNN ohne Nebenbedingungen divergiert – wie erwartet – sehr früh. Das KNN mit Nebenbedingungen erlaubt die vollständige Berechnung des Versuchs und kann das weicher werdende Verhalten quer zur Faserrichtung wiedergeben. Dies kann sowohl an der Verschiebungsfigur erkannt werden, als auch daran, dass die Verschiebung an der Oberseite überproportional anwächst gegenüber derjenigen der Unterseite. Neben den KNN ist ferner eine linear elastische Lösung in Abbildung 7.13 dargestellt. Mithilfe der mittleren Dichte $\rho = 0,438 \text{ kg/cm}^3$ der im KNN-Training genutzten Versuche

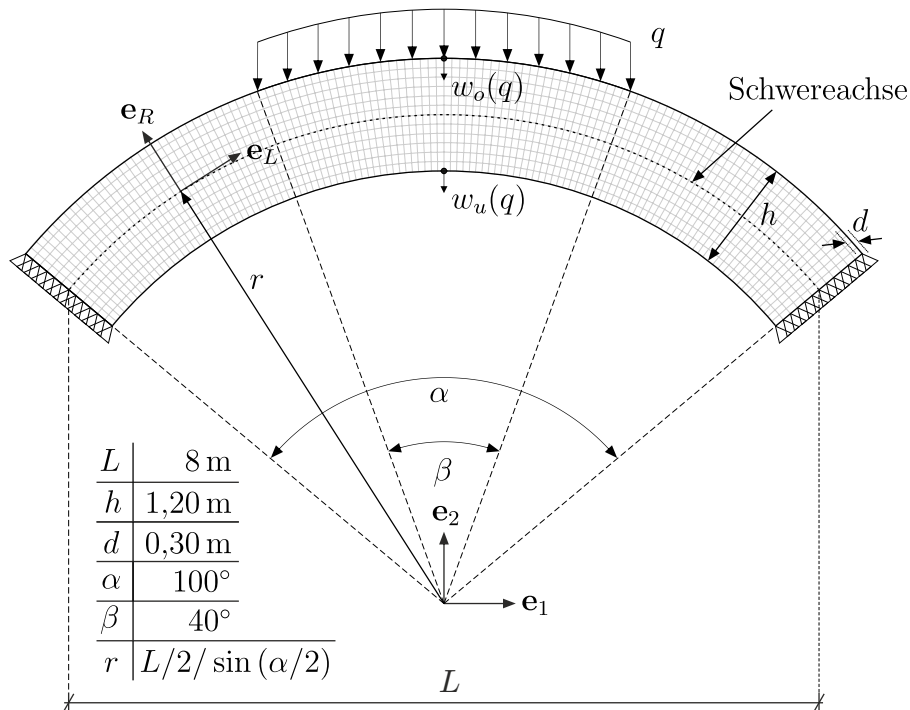


Abbildung 7.12: Bogenträger aus Holz: System, Lagerungsbedingungen, Belastung und Definition des lokalen Koordinatensystems

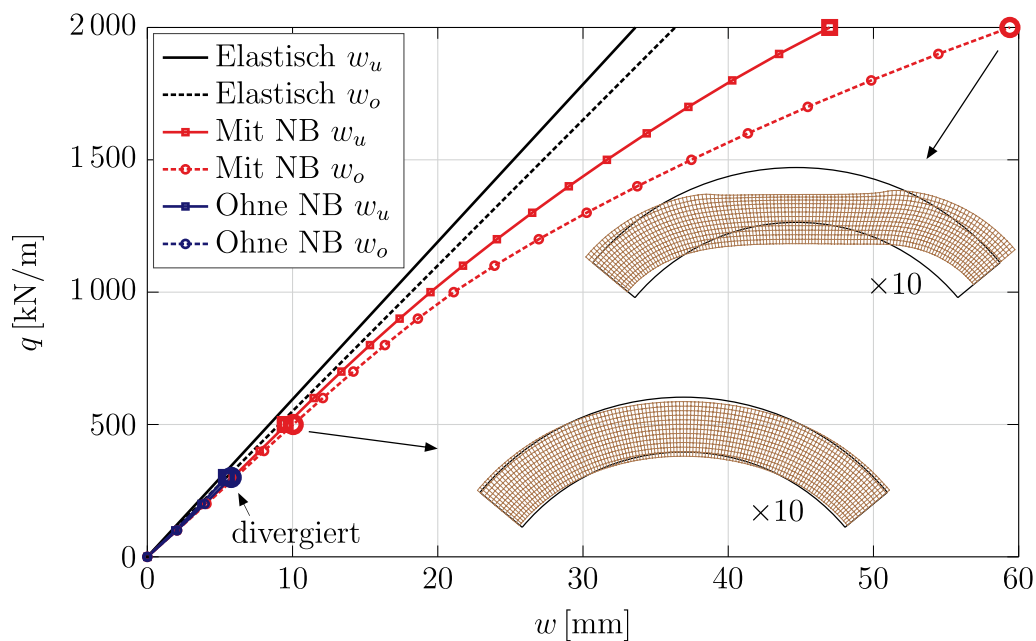


Abbildung 7.13: Bogenträger aus Holz: Last-Verschiebungskurven für die Verschiebungen $w_o(q)$ und $w_u(q)$ an der Ober- und Unterseite des Trägers sowie zwei Verschiebungsfiguren (zehnfach überhöht dargestellt)

und den Regressionskurven für E_L , E_R und ν_{LR} von EBERHARDSTEINER [40] wurden dazu die Parameter

$$E_L = 12\,561 \text{ N/mm}^2, \quad E_R = 744 \text{ N/mm}^2 \quad \text{und} \quad \nu_{LR} = 0,499 \quad (7.19)$$

ermittelt. Der Schubmodul wird über die Formel

$$G_{LR} = \frac{E_L E_R}{E_L + E_R + 2\nu_{LR} E_R} \quad (7.20)$$

berechnet, welche EBERHARDSTEINER ebenfalls verwendet hat⁹⁴. Die Kurven des KNN liegen nicht ganz auf den Kurven des linear elastischen Referenzmaterials. Letzteres basiert auf der Mittelung von Materialparametern, welche aus Versuchen mit anderen Holzscheiben ermittelt wurden. Die KNN-Lösung basiert auf der direkten Mittelung der Daten, ohne vorher den Weg über die Ingenieurkonstanten gegangen zu sein. Daher ist dieser vergleichsweise hohe Steifigkeitsunterschied von 10-20% zu erklären.

In Abbildung 7.14 sind die Verläufe der Spannungen σ_{LL} , σ_{RR} und σ_{LR} über die gesamte Struktur nach dem zwanzigsten und damit letzten Lastschritt für das KNN mit Nebenbedingungen dargestellt. Das aus ästhetischen Gründen hinzugefügte FE-Netz besitzt nur ein Drittel der tatsächlichen Feinheit in radialer Richtung. Das KNN ist in der Lage, einen physikalisch sinnvollen Verlauf der Spannungen vorherzusagen. Dabei wird durch die nicht perfekte Orthotropie des KNN die Symmetrie des Systems nicht exakt abgebildet. Die minimale Längsdruckspannung liegt dabei knapp über der von EBERHARDSTEINER angegebenen uniaxialen Bruchspannung von $f_{c,0} = 49,98 \text{ N/mm}^2$. Tatsächlich divergiert die FE-Berechnung zwei Lastschritte später. Da außerhalb der Bruchgrenzen keine Trainingsdaten mehr vorhanden sind, verhält sich das KNN einzig aus mathematischen Gründen dort instabil, ohne physikalisch eine Schädigung implementiert zu haben. Dies ist aber allenfalls ein Nebeneffekt und sollte nicht zu stark interpretiert werden. Zuletzt muss erwähnt werden, dass die maximale Last $q = 2000 \text{ kN/m}$ vergleichsweise hoch ist.

Die Verläufe der Spannungen σ_{RR} und σ_{LR} in den in Abbildung 7.14 angedeuteten Schnitten sind in Abbildung 7.15 dargestellt. Sie werden den Ergebnissen des linear elastischen Vergleichsmaterials gegenübergestellt. Die prinzipiellen Verläufe stimmen dabei überein. Auf der Unterseite des Trägers entsteht Querkzug, was für Holz besonders gefährlich ist. Die KNN-Lösung liefert eine um 30% höhere maximale Zugspannung. Die am Ende der Streckenlast entstehende maximale Schubspannung ist beim KNN-Materialmodell um 10% geringer. Beide Unterschiede können durch Lastumlagerungen innerhalb des Trägers resultieren, die

⁹⁴ Die Formel wurde im Rahmen von Gesteinsmodellierung von BATUGIN und NIRENBURG [11] aufgestellt und von LEKHNITSKII [113] aufgegriffen. Sie ist eine Näherung für den eigentlich unabhängigen Schubmodul.

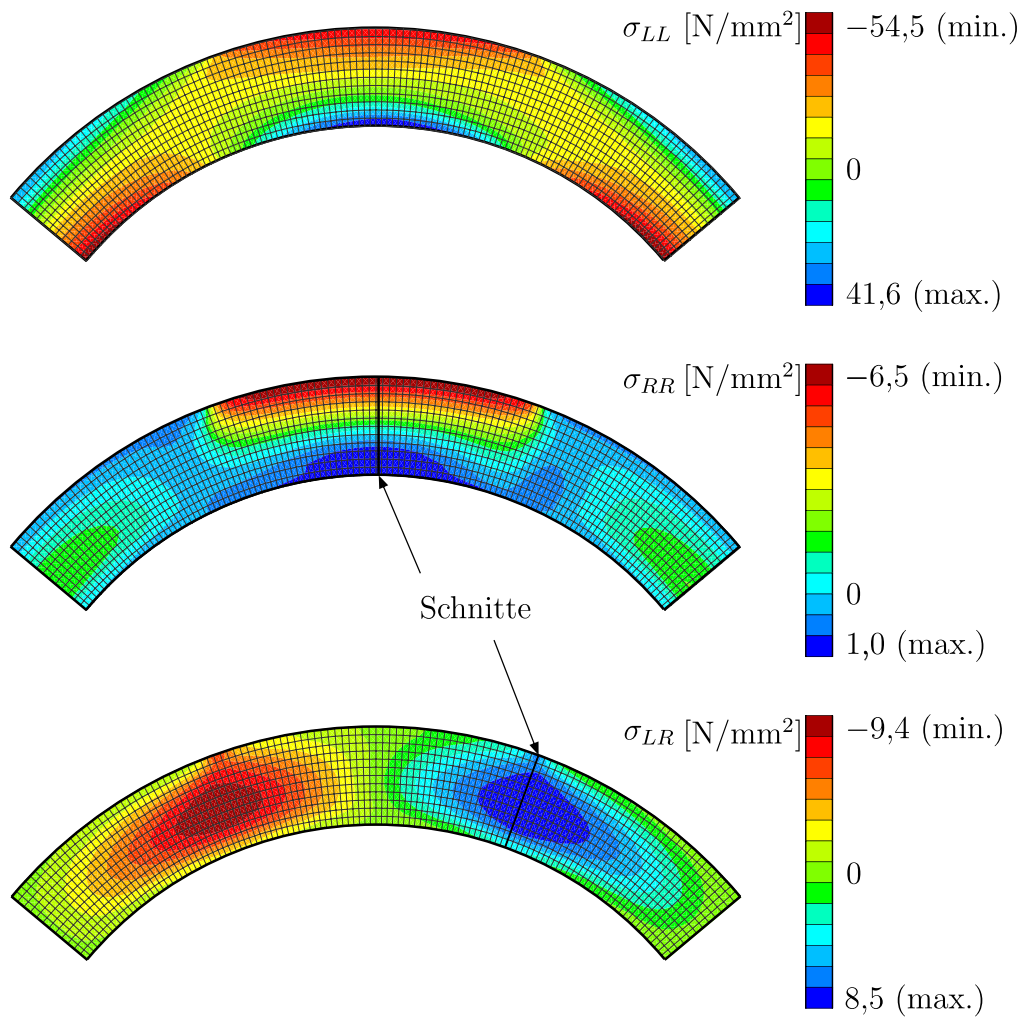


Abbildung 7.14: Verläufe der Spannungen σ_{LL} , σ_{RR} und σ_{LR} über die gesamte Struktur mit dem KNN-Materialmodell mit Nebenbedingungen nach dem zwanzigsten und damit letzten Lastschritt mit $q = 2000 \text{ kN/m}$. Die Verläufe in den Schnitten sind in Abbildung 7.15 dargestellt.

infolge plastischer Verformungen entstehen. An der Unterseite des Trägers sollten sowohl σ_{RR} als auch σ_{LR} exakt null sein. Die Diskrepanz der analytischen Lösung kommt durch die Extrapolation von den Integrationspunkten auf die Knoten im Zuge der Spannungsglättung.

7.2.4 Abschließende Bemerkungen

- Dieses Berechnungsbeispiel hat noch einmal demonstriert, dass die Erkenntnisse, welche mit synthetischen Daten gewonnen wurden, nicht ohne Ergänzungen auf echte Daten übertragen werden können. Hinzu kommen u. a.

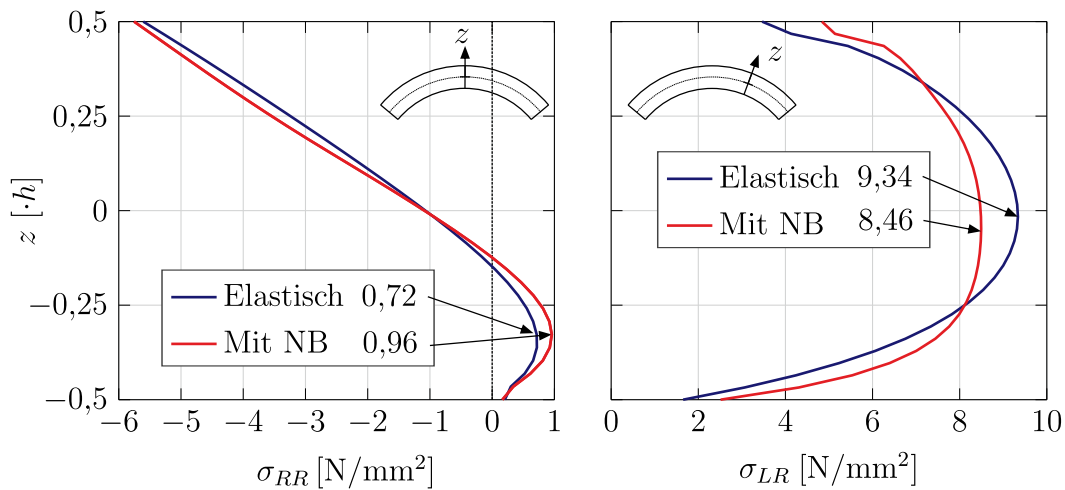


Abbildung 7.15: Verläufe der Spannungen σ_{RR} und σ_{LR} in den in Abbildung 7.14 angedeuteten Schnitten nach dem zwanzigsten und damit letzten Lastschritt mit $q = 2000 \text{ kN/m}$

- eine aufwendige Datenaufbereitung, die ebenfalls einen großen Einfluss auf die nachfolgende Güte der KNN-Approximation hat,
- die Frage, welche physikalischen Eigenschaften dem Material zugesprochen werden sollen, und schließlich
- die Frage, bis zu welchem Grad physikalische Inkonsistenzen vertretbar sind, um eine mathematisch gute Regression zu erhalten.

Daher ist es in Zukunft unabdingbar, dass das Forschungsgebiet der KNN-Materialmodellierung den nächsten Schritt hin zu mehr echten Daten geht, um die bereits so zahlreich entwickelten Algorithmen anhand der zugehörigen Herausforderungen zu untersuchen.

- Eine Übertragung des hier aufgestellten KNN-Materialmodells auf reale Bauteile ist nur eingeschränkt möglich, da hier technisch fehlerfreie Holzproben verwendet wurden. Die Eigenschaften realer Bauteile werden durch Ästigkeit, Risse, Umwelteinflüsse und andere Faktoren beeinflusst. Außerdem werden selbstverständlich nicht nur Kernbretter verbaut, was zu einer bauteilspezifischen Anisotropie führt.
- Weitere Untersuchungen zu den genutzten Fichtenholzdaten könnten sich mit dem Einfluss der Rohdichte und dem zyklischen Verhalten befassen.
- Die Orthotropie von eigentlich radialsymmetrischem Holz ist ebenfalls nur ein idealisiertes Modell, da selbst unter den hier beschriebenen Versuchsbedingungen eine perfekt parallele und äquidistante Verteilung der Jahres-

ringe nicht gegeben ist. Daher könnte die Nebenbedingung der Orthotropie ebenfalls vernachlässigt werden.

- Nicht alle Trainingsprozesse resultieren in einem brauchbaren Ergebnis. Trotz Nebenbedingungen und vergleichsweise vielen Daten divergieren manche KNN-Materialmodelle ebenso schnell wie diejenigen ohne Nebenbedingungen. Von den letzteren ist allerdings kein einziges brauchbar. An dieser Stelle könnten weitere Untersuchungen zur Streuung der KNN-Materialmodelle zielführend sein.

7.3 Ferroelektrische Funktionskeramik

Als Abschlussbeispiel und in der Absicht einen Ausblick in multiphysikalische Problemstellungen zu bieten, wird in diesem Abschnitt das eindimensionale geschichtsabhängige Verhalten einer ferroelektrischen Funktionskeramik mithilfe eines KNN approximiert. Ziel dabei ist die Flexibilität der KNN-Materialmodellierung im Allgemeinen und die der physikalischen Nebenbedingungen im Speziellen hervorzuheben. Zur Modellierung des besagten Materialverhaltens werden nämlich die gleichen Architekturen und Nebenbedingungen genutzt wie für rein mechanisches plastisches Materialverhalten.

Dazu wird das Verhalten der kommerziell verfügbaren Piezokeramik PIC 151 der Firma PI Ceramic [144] aus Lederhose (Deutschland) approximiert, welche aus Blei-Zirkonat-Titanat (PZT) besteht. ZHOU ET AL. [202, 203] haben an diesem Material Anfang der 2000er Jahre am Forschungszentrum Karlsruhe (mittlerweile Karlsruher Institut für Technologie) Versuche durchgeführt.

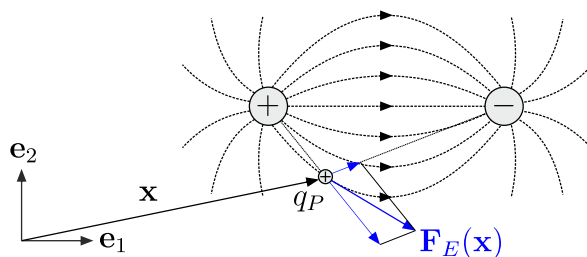


Abbildung 7.16: Elektrisches Feld zweier Punktladungen

Zur Beschreibung von PZT ist ein kurzer Exkurs (vgl. MAUGIN [127]) in die dazu nötigen Grundlagen der Elektrostatik von Dielektrika notwendig, um die Größen *elektrische Feldstärke*, *elektrische Polarisierung* und *dielektrische Verschiebung* einzuführen. Dabei werden quasi-statische Prozesse vorausgesetzt. Außerdem sind elektromagnetische Wechselwirkungen ausgeschlossen. Die Basis bildet

die elektrische Ladung q , die in der Einheit Coulomb⁹⁵ (C) beschrieben wird und entweder positiv oder negativ ist. Ruhende Ladungen wechselwirken über COULOMB'sche Kräfte \mathbf{F}_E miteinander, wobei sich gleichnamige Ladungen abstoßen und ungleichnamige Ladungen anziehen. Die Auswirkung vorhandener Ladungen auf deren Umgebung wird mithilfe des *elektrischen Feldes* beschrieben, was am Beispiel zweier Punktladungen durch die sog. elektrischen Feldlinien in Abbildung 7.16 dargestellt ist⁹⁶. Die elektrische Feldstärke kann durch die COULOMB'sche Kraft auf eine hinreichend kleine⁹⁷ Probeladung q_P charakterisiert werden, mit

$$\mathbf{E}(\mathbf{x}) = \frac{\mathbf{F}_E(\mathbf{x})}{q_P} . \quad (7.21)$$

Da nachfolgend nur infinitesimale Verzerrungen $\boldsymbol{\varepsilon}$ vorausgesetzt werden, wird das Symbol \mathbf{E} verwendet, ohne Verwechslungen mit dem GREEN-LAGRANGE'schen Verzerrungstensor zu riskieren. Häufig werden elektrische Potentiale $\phi(\mathbf{x})$ definiert, mit denen $\mathbf{E} = -\partial\phi/\partial\mathbf{x}$ als deren Gradient berechnet wird. Die Differenz der elektrischen Potentiale zwischen zwei Punkten wird als elektrische Spannung bezeichnet.

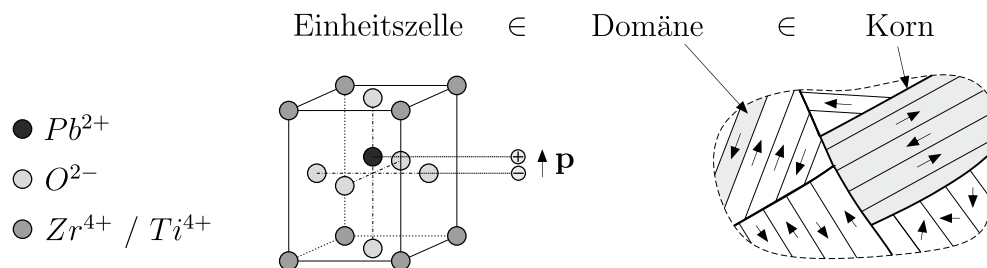


Abbildung 7.17: Tetragonal verzerrte Einheitszelle von PZT und die daraus aufgebaute Struktur aus Domänen gleicher Polarisation und Körnern gleicher Gitterorientierung

Als Dielektrika werden jene Materialien bezeichnet, die eine verschwindend geringe elektrische Leitfähigkeit haben und somit als elektrischer Isolator modelliert werden können. Etwaige Ladungen sind im Material gebunden, wodurch kein elektrischer Strom fließen kann. Allerdings ist es möglich, dass die gebundenen Ladungen Dipole bilden, die zu einer Polarisation des Materials führen können. Zur Veranschaulichung der Polarisation solcher Dielektrika ist es zweckmäßig, in Abbildung 7.17 eine tetragonale Einheitszelle von PZT zu betrachten. In dieser

⁹⁵ Benannt nach dem französischen Physiker CHARLES AUGUSTIN DE COULOMB. Coulomb ist keine SI-Einheit und wird mit $1\text{ C} = 1\text{ As}$ in die Einheiten der elektrischen Stromstärke (Ampere nach ANDRÉ-MARIE AMPÈRE) und der Zeit (Sekunde) umgerechnet.

⁹⁶ Als mechanische Analogie kann das Gravitationsfeld herangezogen werden, wobei die Kräfte zwischen Massen ausschließlich anziehender Natur sind.

⁹⁷ Die Probeladung selbst verändert sonst ebenfalls das elektrische Feld.

ferroelektrischen Phase⁹⁸ liegen die Ladungsschwerpunkte der positiven und negativen Ionen nicht aufeinander, wodurch ein sog. *Dipolmoment* \mathbf{p} [Cm] resultiert, welches bezogen auf das Zellgitter in sechs Richtungen orientiert sein kann. Im Sinne der Kontinuumsmechanik ist die *elektrische Polarisation*

$$\mathbf{P}(\mathbf{x}) = \frac{d\mathbf{p}(\mathbf{x})}{dV} \quad (7.22)$$

die Mittelung dieses Moments auf ein Einheitsvolumen. Makroskopisch besteht PZT aus Körnern gleicher kristalliner Ausrichtung. Innerhalb solcher Körner werden Domänen mit unterschiedlicher Orientierung der Polarisation unterschieden, die stets um 90° oder 180° zueinander ausgerichtet sind. Mehrere solcher Domänen sind Körnern zuzuordnen, in denen das Ionengitter gleich ausgerichtet ist. Dies ist ebenfalls in Abbildung 7.17 dargestellt. Als Antwort auf ein von außen einwirkendes elektrisches Feld verschieben sich die Ladungsschwerpunkte der einzelnen Zellen. Als Maß dieser Reaktion wird die dielektrische Verschiebung oder elektrische Flussdichte \mathbf{D} [C/m²] definiert, welche die wesentliche Änderung des inneren Polarisationszustandes charakterisiert. Ausführlichere Beschreibungen der hier nur knapp eingeführten Zusammenhänge sind bspw. in MAUGIN [127] zu finden.

Piezoelektrische Materialien zeichnen sich durch eine Kopplung mechanischer und elektrischer Eigenschaften aus. Eine durch die beschriebene Mikrostruktur gegebene elektrische Polarisation kann durch elastische Verformungen verstärkt oder abgeschwächt werden, was als *direkter Piezoeffekt* bekannt ist. In Gegenwart eines elektrischen Feldes deformieren sich piezoelektrische Materialien. Dies wird *indirekter Piezoeffekt* genannt. Die Eigenschaft, durch mechanische Deformation eine elektrische Spannungsquelle zu bilden oder vice versa, wird in zahlreichen technischen Anwendungen genutzt. So werden piezoelektrische Bauelemente etwa in der Sensorik oder als Mikroaktoren verwendet. Im Bauingenieurwesen werden piezoelektrische Seismiksensoren eingesetzt, um Schwingungen von Gebäuden oder Brücken zu messen.

Die beschriebenen piezoelektrischen Eigenschaften weisen die Materialien nicht unbedingt von Natur aus auf. Nach dem Sintervorgang im Zuge des Herstellungsprozesses sind die Dämonen selbst zwar polarisiert, allerdings auch willkürlich verteilt. Somit resultiert zunächst ein makroskopisch gemittelter unpolarisierter Zustand. Daher ist ein Polungsprozess mit hinreichend großen elektrischen Feldern notwendig, um die Domänen in Richtung des aufgebrauchten Feldes auszurichten. Erst dann ist das Material für die technische Anwendung brauchbar. Ist eine Polarisierung in dieser Form möglich, wird von ferroelektrischen Materialien

⁹⁸ Für PZT liegen je nach molekularer Zusammensetzung und Temperatur auch andere Phasen vor, die paraelektrisch sein können, sprich ohne Dipolmoment.

gesprochen. Ferroelektrische Stoffe sind somit stets auch piezoelektrisch. Darüber hinaus kann durch mechanische Einwirkung ein Umklappen der Domänen induziert werden, was zu einer Depolarisation des Materials führen kann.

Während die genannten Anwendungen meist im sog. Kleinsignal-Bereich liegen und mit linearen Stoffgesetzen beschrieben werden können, sind der Polungs- und Depolarisationsprozess durch die beschriebenen Umklappprozesse nichtlinear und geschichtsabhängig. Dieser Großsignal-Bereich ist Gegenstand der nachfolgenden Untersuchungen.

7.3.1 Aufbereitung der Versuchsdaten

ZHOU ET AL. [202, 203] haben das kombinierte ferroelektrische und ferroelastische Verhalten von PIC 151 untersucht. Dazu wurden an $5 \times 5 \times 15 \text{ mm}^3$ große Probekörper zwei gegenüberliegende Elektroden angebracht, über die das elektrische Feld gesteuert werden konnte. Gleichzeitig war eine mechanische Belastung in Richtung des elektrischen Feldes möglich. An einer Elektrode konnte durch Messung der Oberflächenladungsdichte die dielektrische Verschiebung indirekt bestimmt werden. Die Verzerrung wurde an der Probe mittels Dehnungsmessstreifen ermittelt. Eine stark vereinfachte, für die folgende Anwendung jedoch ausreichende Darstellung dieses Versuchsaufbaus ist in Abbildung 7.18 dargestellt. Zusammengefasst: Die Spannung σ und die elektrische Feldstärke E wurden vorgegeben, während die Verzerrung ε und die dielektrische Verschiebung D gemessen wurden. Alle Größen sind nachfolgend Skalare und beziehen sich auf die Richtung der längeren Probenabmessung.

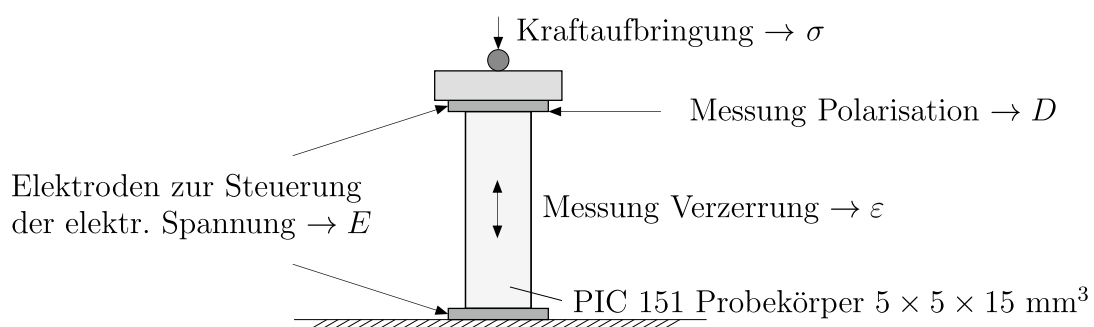


Abbildung 7.18: Stark vereinfachter Versuchsaufbau für die gekoppelte mechanische und elektrische Beanspruchung der PIC 151 Probekörper, vgl. ZHOU ET AL. [202, 203]

Bei jeweils konstantem Druck $\sigma \in \{-400, -200, -125, -75, -50, -25, 0\} \text{ N/mm}^2$ resultieren unterschiedliche Hysteresen der Verzerrung und der dielektrischen Verschiebung durch zyklische Variation des elektrischen Feldes $E \in [-2, 2] \text{ kN/mm}$.

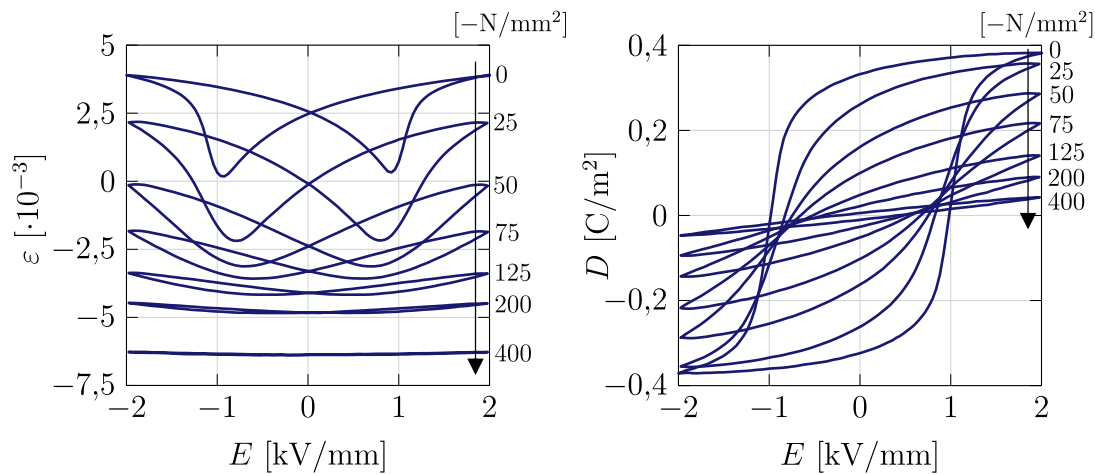


Abbildung 7.19: Stationäre Schmetterlingshysterese (links) und dielektrische Hysterese (rechts) von PIC 151, entnommen aus ZHOU ET AL. [203]. Die Druckspannungsniveaus sind jeweils an den Graphen angegeben.

Die zugehörigen stationären Hysteresen⁹⁹ sind graphisch aus ZHOU ET AL. [203] entnommen und in Abbildung 7.19 dargestellt. Die Druckspannungsniveaus sind jeweils gekennzeichnet. Die Proben waren zu Beginn der Untersuchungen ungepolt. Die sog. Schmetterlingshysterese (links) beschreibt das zyklische Verhalten der Verzerrung in Abhängigkeit der elektrischen Feldstärke, während die dielektrische Hysterese (rechts) die dielektrische Verschiebung in Abhängigkeit der elektrischen Feldstärke zeigt. Durch Anlegen einer hinreichend großen elektrischen Feldstärke werden die Domänen im Material entlang der elektrischen Feldlinien ausgerichtet. Dies macht sich auf makroskopischer Ebene einerseits durch eine Erhöhung der Polarisierung und andererseits durch eine Änderung des Verzerrungszustandes bemerkbar. Diese irreversiblen Domänenenumklappprozesse sind maßgeblich für das nichtlineare Verhalten beider Hysteresen verantwortlich. Durch Überlagerung mit einer konstanten Spannung wird dieses Verhalten jedoch unterdrückt, was an den immer flacheren Verläufen zu erkennen ist. Außerdem verschieben sich die Schwerpunkte der Schmetterlingshysterese nach unten infolge der überlagerten elastischen Verzerrung. Details zur Interpretation der Verläufe können aus ZHOU ET AL. [203] entnommen werden.

Zur Untersuchung des ferroelastischen Verhaltens von PIC 151 haben ZHOU ET AL. [202] die Proben mit einer mechanischen Spannung von -400 N/mm^2 belastet und wieder entlastet. Dies wurde für eine initial ungepolte Probe sowie für initial gepolte Proben durchgeführt. Bei den gepolten Proben wurde jeweils ein konstantes elektrisches Feld mit $E \in \{0; 0,25; 0,50; 0,75; 1; 1,2; 1,5; 2\} \text{ kV/mm}$

⁹⁹ Stationär heißt hier ohne Berücksichtigung der Erstbelastungskurve. Außerdem wurde von vier durchgeführten Zyklen jeweils nur der letzte abgedruckt.

überlagert. Aus Gründen der Übersicht sind nur vier der insgesamt neun Kurven in Abbildung 7.20 dargestellt. Die sog. mechanische Depolarisation ist links abgebildet. Eine durch die initiale Polung erzeugte Polarisation mit $D > 0$ wird ohne überlagertes elektrisches Feld fast vollständig vernichtet. Liegt ein elektrisches Feld vor, welches in Richtung der initialen Polung zeigt, wird der spannungsinduzierte Umklappprozess behindert. Für $E = 2 \text{ kV/mm}$ wird die Polarisation beinahe vollständig zurückgewonnen. Das Spannungs-Verzerrungsverhalten der Proben ist im rechten Bild zu sehen. Selbst die ungepolte Probe zeigt ein rein mechanisch nichtlineares und geschichtsabhängiges Verhalten, mit bleibender irreversibler Verformung. Eine durch Polung hervorgerufene initiale Verzerrung $\varepsilon > 0$ wird ohne überlagertes elektrisches Feld durch die Druckspannung vollständig wieder abgebaut. Es wird ein vollständig depolarisierter Zustand erreicht, in dem die Domänen in eine Ebene senkrecht zur Belastungsrichtung mit regelloser Dipolausrichtung umgeklappt sind. Dieser Zustand wird auch bei der ungepolten Probe erreicht, sodass sich für beide Proben eine identische Entlastungskurve ergibt. Durch die überlagerten elektrischen Felder wird zum einen die initiale Dehnung erhöht, zum anderen der Depolarisationsprozess gehemmt. Dies spiegelt sich in geringeren bleibenden Verzerrungen nach Entlastung wider. Der Verlauf der ferroelastischen Kurven lässt sich ebenfalls durch Umklappprozesse erklären. Details zur Interpretation der Verläufe können aus ZHOU ET AL. [202] entnommen werden.

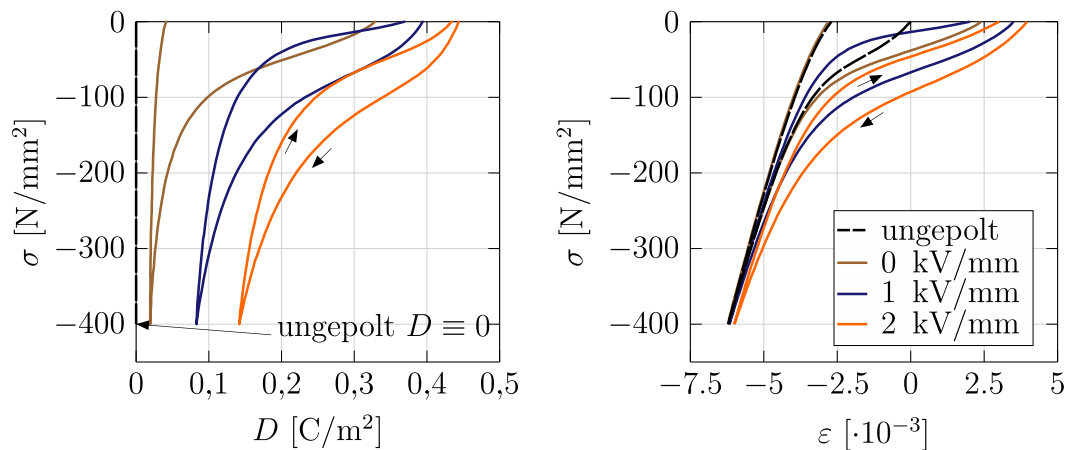


Abbildung 7.20: Vier Graphen der mechanischen Depolarisation (links) und des Spannungs-Verzerrungsverhaltens (rechts), entnommen aus ZHOU ET AL. [202]

Die Aufbereitung der Daten zum Training eines KNN verläuft äquivalent zum Vorgehen bei rein mechanischen plastischen Daten. Dazu müssen zunächst aus den Kurven diskrete Punkte gewonnen werden. Da die vier Zustandsgrößen ε , D , σ und E in jeweils zwei Bildern verteilt sind, muss darauf geachtet werden,

dass die graphisch entnommenen Ordinaten jeweils zum gleichen GGW-Punkt $(\varepsilon, D, \sigma, E)$ gehören. Dies ist möglich, da im Fall der Schmetterlingshysterese und der dielektrischen Hysterese beide Bilder die elektrische Feldstärke E enthalten. Im Fall der ferroelastischen Hysterese und der mechanischen Depolarisation enthalten beide Bilder die Spannung σ . Somit folgt für jeden der insgesamt 16 Versuche ein geordneter Pfad

$$\mathcal{P} = \{(\varepsilon_0, D_0, \sigma_0, E_0), (\varepsilon_1, D_1, \sigma_1, E_1), \dots, (\varepsilon_N, D_N, \sigma_N, E_N)\} \quad (7.23)$$

mit $N + 1$ GGW-Punkten. Für die nachfolgenden Untersuchungen wurde stets $N = 100$ gewählt. Im Gegensatz zu den Angaben in ZHOU ET AL. [202, 203] wurde das Vorzeichen von D und E von negativ auf positiv geändert (positiv entspricht in Richtung der initialen Polung) sowie die Definition der Spannung konsistent zur restlichen Arbeit gewählt (Druck negativ). Im vorliegenden Fall hat dies keine Auswirkung, solange beide Vorzeichen konsistent sind.

7.3.2 Anwendung eines KNN-Materialmodells

Zur Approximation des beschriebenen ferroelektrischen und ferroelastischen Materialverhaltens wird das inkrementelle KNN-Materialmodell FNN-P-Ink aus Abschnitt 6.1.2 verwendet. Der Eingangsvektor

$$\mathbf{x} = [\Delta\varepsilon, \Delta D, \varepsilon_l, D_l, \sigma_l, E_l]^T \in \mathbb{R}^6 \quad (7.24)$$

enthält neben dem letzten GGW-Punkt $(\varepsilon_l, D_l, \sigma_l, E_l)$ die Inkremente der Verzerung $\Delta\varepsilon$ und der dielektrischen Verschiebung ΔD . Der Ausgangsvektor

$$\mathbf{z} = [\Delta\sigma, \Delta E]^T \in \mathbb{R}^2 \quad (7.25)$$

enthält das Inkrement der Spannung $\Delta\sigma$ und das Inkrement der elektrischen Feldstärke ΔE . Die Umrechnung der Pfade (7.23) in die inkrementellen Datenpaare erfolgt äquivalent zum Fall plastischer KNN-Materialmodelle in Abschnitt 6.1.2. Dabei orientieren sich die dazu nötigen Belastungsrichtungen an den Umkehrpunkten der Spannung σ oder der elektrischen Feldstärke E der jeweiligen Versuche, vgl. Abbildung 6.3. Zur Umrechnung werden die Versätze $\Delta n \in \{0, 1, 2, \dots, 10\}$ festgelegt.

In Voruntersuchungen hat sich die Topologie [6-20-20-2] als brauchbar herausgestellt. Von den insgesamt 16 Pfaden werden 14 wie beschrieben für das Training mittels QN-Verfahren aufbereitet. 80% der resultierenden 27 585 Datenpaare werden zum Training genutzt. Die restlichen 5517 Datenpaare dienen zur Berechnung des Validierungsfehlers. Das Training wird nach 10 000 Epochen abgebrochen. Die Gewichte, die bis zur letzten Epoche zum kleinsten Validierungsfehler

fürten, werden als Resultat des Trainingsprozesses abgespeichert. Die zwei zurückgehaltenen Pfade dienen als nachträglicher Test der Generalisierungsfähigkeit des KNN-Materialmodells. Diese zwei Versuche sind der ferroelektrische Versuch bei konstanter Druckspannung $\sigma = -75 \text{ N/mm}^2$ und der ferroelastische Versuch bei konstantem elektrischen Feld mit $E = 1 \text{ kV/mm}$. Das Training erfolgt mittels QN-Verfahren aus Abschnitt 4.4.2 in MATLAB [170].

Mit $\boldsymbol{\xi} := [\varepsilon, D]^T$ und $\boldsymbol{\zeta} := [\sigma, E]^T$ werden die multiphysikalischen Eingangs- und Ausgangsgrößen zusammengefasst¹⁰⁰. Neben den Daten werden die folgenden vier physikalischen Nebenbedingungen im Trainingsprozess berücksichtigt:

- Inkrementelle Normalisierung: $\Delta\boldsymbol{\zeta}(\Delta\boldsymbol{\xi} = \mathbf{0}) \stackrel{!}{=} \mathbf{0}$, vgl. Abschnitt 6.2.2,
- Symmetrie der Materialtangente: $\mathbf{C}_T = \frac{\partial\Delta\boldsymbol{\zeta}}{\partial\Delta\boldsymbol{\xi}} \stackrel{!}{=} \mathbf{C}_T^T$, vgl. Abschnitt 6.2.4,
- Inkrementelle Stabilität: $\Delta\boldsymbol{\zeta}^T \Delta\boldsymbol{\xi} \stackrel{!}{\geq} 0$, vgl. Abschnitt 6.2.5,
- Positive Hauptdiagonalelemente: $C_{T11} \stackrel{!}{\geq} 0, C_{T22} \stackrel{!}{\geq} 0$, vgl. Abschnitt 7.2.3.

Die ersten beiden Nebenbedingungen werden mit $\epsilon = 1$ eingefordert, die letzten beiden mit $\epsilon = 10^5$. Die zugehörigen NB-Punkte werden in der konvexen Hülle der Trainingsdaten generiert, vgl. Abschnitt 6.2.1. Dabei werden für die inkrementelle Normalisierung $P_C = 10^3$ Punkte generiert und für die anderen drei jeweils $P_C = 10^4$. Im Folgenden werden drei verschiedene Kombinationen von Nebenbedingungen miteinander verglichen: für **KNN₀** werden keine Nebenbedingungen aktiviert, für **KNN₃** alle bis auf die der symmetrischen Materialtangente, und für **KNN₄** sind alle vier Nebenbedingungen aktiv. Die drei KNN-Materialmodelle sollen die beiden als Testpfade zurückgehaltenen und nie im Trainingsprozess verwendeten Versuche nachbilden, indem sie von den bekannten Daten interpolieren.

In Abbildung 7.21 sind die Lösungen der drei genannten KNN-Materialmodelle den Versuchsdaten gegenübergestellt. Die KNN-Lösung wurde dabei mit dem Materialpunkt-Versuch aus Abschnitt 5.1.1 berechnet, wobei sinngemäß ε und D als die „Verzerrungen“ sowie σ und E als die „Spannungen“ zu interpretieren sind. Die Materialpunkt-Versuche wurden dabei zur Nachbildung der realen Versuche durch Vorgabe von σ und E durchgeführt. Dies entspricht dem Äquivalent

¹⁰⁰ An dieser Stelle sei erwähnt, dass die Zuordnung ebenfalls getauscht werden könnte, um die dielektrische Verschiebung als Ausgangsvariable zu definieren. Die Zuordnung wurde hier so gewählt, da in den Versuchen σ und E vorgegeben wurden. In analytischen Materialmodellen hat die Definition einen Einfluss auf die Struktur des lokal zu lösenden konstitutiven Problems, vgl. MIEHE ET AL. [130], KLINKEL [105] oder SUTTER und KAMLAH [166].

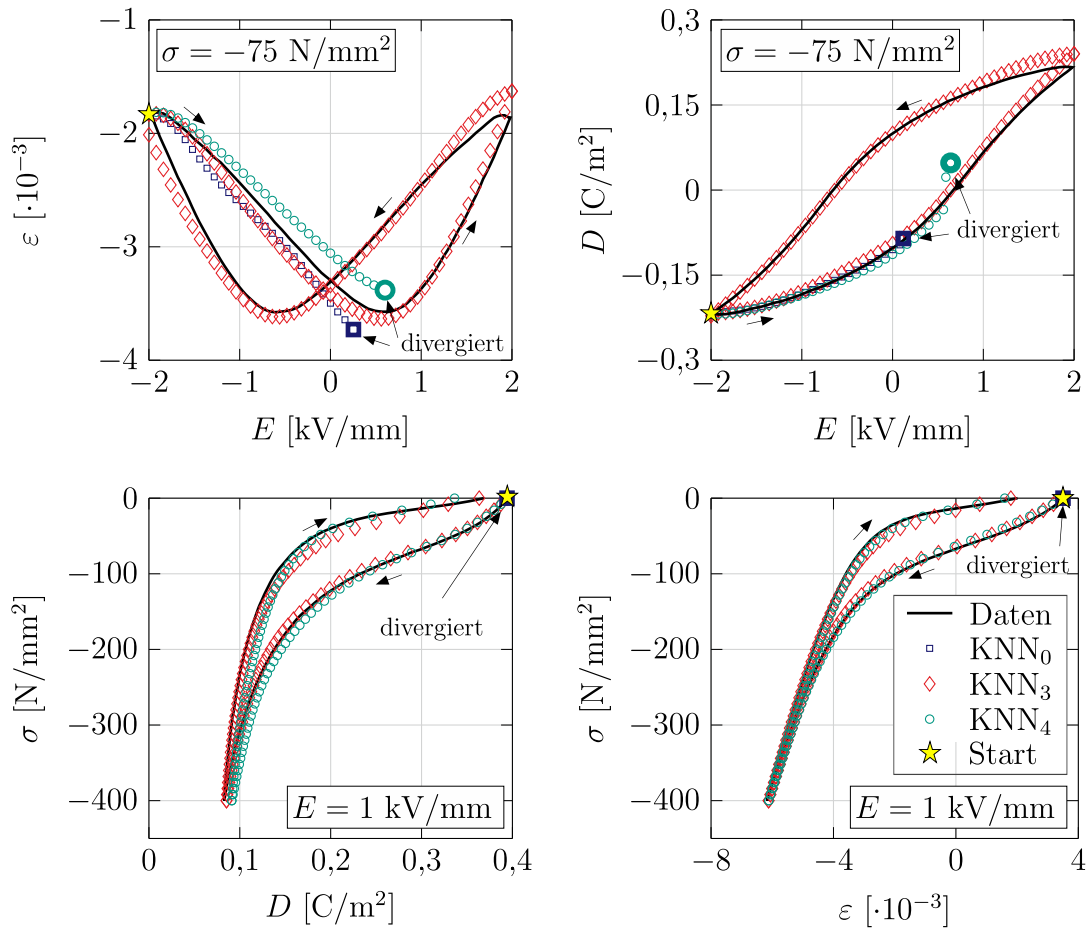


Abbildung 7.21: Vergleich der KNN-Materialmodelle KNN_0 (ohne NB), KNN_3 (alle NB außer Symmetrie der Materialtangente) und KNN_4 (alle NB) am Beispiel der zwei nicht zum Training verwendeten Versuche

einer reinen Spannungssteuerung. Somit ist für die lokale NEWTON-Iteration die Information über die Materialtangente eminent.

Das ohne zusätzliche Nebenbedingungen trainierte KNN_0 divergiert im Fall des ferroelektrischen Verhaltens vergleichsweise früh, während das NEWTON-Verfahren bei den ferroelastischen Be- und Entlastungsversuchen unmittelbar nach dem Start abbricht. Das KNN ist ohne zusätzliche physikalische Informationsquelle nicht in der Lage, auf die nicht trainierten Versuch zu interpolieren. Das mit allen Nebenbedingungen trainierte KNN_4 verhält sich wesentlich stabiler und ist in der Lage, die ferroelastischen Versuche vollständig wiederzugeben. Allerdings hat sich in zuvor durchgeführten Parameterstudien der Eindruck ergeben, dass die Nebenbedingung der Symmetrie der Materialtangente widersprüchlich ist zu den gegebenen Daten. Dies zeigt sich vor allem in den ferroelektrischen Versuchen, die in Abbildung 7.21 ebenfalls divergieren. Diese Beobachtungen lassen den Schluss

zu, dass es schwieriger ist für das vorliegende Materialverhalten ferroelektrischer Funktionskeramiken, ein auf einer symmetrischen Materialtangente basierendes Materialmodell zu entwickeln, zumindest mit dem Anspruch, alle Effekte korrekt wiedergeben zu können. Durch diese Widersprüche und die fehlenden Daten im Bereich der in Abbildung 7.21 dargestellten Tests resultieren vermutlich die beobachteten Instabilitäten. Wird die zugehörige Nebenbedingung deaktiviert, verhalten sich die Netze wesentlich stabiler und sind in der Lage, alle Versuche brauchbar wiederzugeben. Dies ist ebenfalls an den Ergebnissen des KNN₃ in Abbildung 7.21 zu erkennen.

7.3.3 Abschließende Bemerkungen

- Die eigentlich für elastisches und plastisches Materialverhalten implementierten KNN-Algorithmen und Nebenbedingungen waren ohne Anpassung auf die Daten der ferroelektrischen Funktionskeramik anwendbar. Der größte Aufwand liegt in der Datenaufbereitung. Dies zeigt ein weiteres Mal die Flexibilität als große Stärke der KNN-Materialmodellierung.
- Rein mathematisch sind die gewählten Einheiten nicht von Belang, da die Eingangs- und Ausgangsvariablen unabhängig voneinander normiert werden, vgl. Abschnitt 4.1. Generell ist es allerdings sinnvoll, dass die Einheitensysteme der verschiedenen physikalischen Größen zusammenpassen. Für $\varepsilon \cdot \sigma$ ist hier die Einheit $\text{N}/\text{mm}^2 = 10^6(\text{kgm}^2/\text{s}^2)/\text{m}^3$ gleich wie die Einheit $\text{kV}/\text{mm} \cdot \text{C}/\text{m}^2 = 10^6(\text{kgm}^2/\text{s}^2)/\text{m}^3$ für $D \cdot E$.
- Die Erkenntnis, dass die Einforderung einer symmetrischen Materialtangente scheinbar widersprüchlich zum Materialverhalten ist, hat sich im Rahmen mehrerer Vorstudien ergeben. Ganz nach dem Trial-and-Error-Prinzip wurden dabei verschiedene Kombinationen überprüft.
- Im vorliegenden Beispiel wurden ausschließlich die stationären Hysteresen zur Generierung der Trainingsdaten genutzt. Auf dem Weg hin zu einem flexibleren KNN-Materialmodell müssten die Erstbelastungskurven ebenfalls hinzugezogen werden. Dadurch sind höchstwahrscheinlich zusätzliche Geschichtsvariablen im Eingangsvektor notwendig. Außerdem ist eine Erweiterung auf ein dreidimensionales KNN-Materialmodell denkbar, da von ZHOU ET AL. [202, 203] ebenfalls die Querdehnungen angegeben sind.
- Die Überlegung, dass der Eingangsvektor (7.24) ausreichend zur Beschreibung dieses Materialverhaltens ist, basiert auf der Interpretation der $D-E$ -Kurven als nichtlineare kinematische Verfestigung, vgl. Abschnitt 6.1.1 und Abbildung 7.19.

8 Zusammenfassung und Ausblick

Die vorliegende Arbeit befasst sich mit der physikalisch konsistenten Modellierung von elastischem und plastischem Materialverhalten mit künstlichen neuronalen Netzen (KNN). Die besagte Konsistenz wird hauptsächlich durch Einforderung physikalisch motivierter Nebenbedingungen innerhalb des KNN-Trainingsprozesses erzielt. Die wesentlichen Aspekte, die im Rahmen der vorliegenden Arbeit diskutiert und analysiert werden, lauten wie folgt:

- Das Training künstlicher neuronaler Netze wird im Kontext restringierter Optimierungsstrategien formuliert. Dadurch ist die Erweiterung des Trainingsprozesses mit Nebenbedingungen über verschiedene Verfahren möglich. In diesem Zusammenhang hat sich die klassische Penalty-Methode als zweckmäßig herausgestellt. Über eine Monte-Carlo-Integration werden die zugehörigen Fehlerterme formuliert.
- Für elastisches und plastisches Materialverhalten werden diverse physikalisch motivierte Nebenbedingungen eingeführt, die mit den vorgestellten Methoden in den Trainingsprozess integriert werden können. Im Zuge der Formulierung von Nebenbedingungen für plastisches Materialverhalten wird eine effiziente Methode zur Generierung der Monte-Carlo-Integrationspunkte vorgestellt.
- Mit synthetisch generierten Trainingsdaten werden ausführliche Untersuchungen zu elastischen und plastischen KNN-Materialmodellen durchgeführt. Neben dem positiven Einfluss der Nebenbedingungen werden dabei auch grundlegende Herausforderungen der KNN-Materialmodellierung behandelt wie bspw. Quasi-Inkompressibilität oder eine geringe Datenbasis. Die am Materialpunkt gewonnenen Erkenntnisse werden anschließend mit Finite-Element-Rechnungen auf das Gesamttragwerk übertragen.
- Die physikalischen Nebenbedingungen werden mit alternativen, auf spezifischen KNN-Architekturen basierenden Methoden aus der Literatur verglichen, welche per Konstruktion einige dieser Eigenschaften exakt erfüllen.
- Mit echten Versuchsdaten von drei unterschiedlichen Materialien wird das Konzept des KNN-Trainings mit Nebenbedingungen in die praktische Anwendung gebracht. Dabei werden Gemeinsamkeiten und Unterschiede zu den Erkenntnissen auf Basis synthetischer Daten herausgearbeitet. Die aufkommenden zusätzlichen Herausforderungen werden mit Lösungsvorschlägen adressiert. Das Verhalten dieser drei Materialien kann mittels KNN erfolgreich approximiert werden. Die Einforderung physikalisch motivierter Nebenbedingungen spielt dabei eine zentrale Rolle.

Anhand der Beispiele mit synthetisch generierten Daten in den Kapiteln 5 und 6 sowie nicht zuletzt durch die Übertragung auf echte Versuchsdaten in Kapitel 7 können die folgenden Schlussfolgerungen gezogen werden:

- Die Einforderung physikalischer Nebenbedingungen verbessert die Eigenschaften der resultierenden KNN-Materialmodelle außerordentlich. Der Trainingsprozess wird reguliert. Streuungen in den Trainingsdaten werden ausgeglichen. Geringe Datenmengen sind möglich zum Training eines KNN-Materialmodells. Die numerische Stabilität innerhalb einer äußeren Gleichgewichtsiteration wird bemerkenswert gesteigert. Falls möglich, ist es daher aus Sicht des Autors zu empfehlen, physikalische Nebenbedingungen zu berücksichtigen.
- Die Nebenbedingungen werden zwar eingefordert, jedoch nicht exakt erfüllt. Ein Vergleich mit solchen Methoden, die per KNN-Konstruktion physikalische Eigenschaften gewährleisten, zeigt jedoch ein vergleichbar gutes Verhalten, zumindest innerhalb des Trainingsraums. Darüber hinaus sind die Nebenbedingungen flexibel ein- und ausschaltbar und so beliebig miteinander kombinierbar. So ist bspw. die Vernachlässigung der Symmetrie der Tangente in Abschnitt 7.3 als Nebenbedingung problemlos möglich, während die i. d. R. auf Potentialen basierenden Methoden der Literatur dort zu starr wären. Im Fall der Nebenbedingungen ist dazu allerdings eine zusätzliche Strategie zur Erzeugung der Monte-Carlo-Integrationspunkte notwendig.
- Die Anwendung auf echte Versuchsdaten hat aus Sicht des Autors ein grundlegendes Credo der KNN-Materialmodellierung erschüttert, welches lautet: Wenn genügend Versuchsdaten vorliegen, so erlernt das KNN die zugrunde liegenden physikalischen Eigenschaften implizit. Daran ist nach den Erkenntnissen aus Abschnitt 7.2 stark zu zweifeln. Ohne physikalischen Input findet der Optimierungsalgorithmus die mathematisch beste Lösung. Durch die Komplexität der zu minimierenden Fehlerfläche liegen womöglich viele gleichwertige Lösungen nebeneinander, wobei nur manche von ihnen ebenfalls die physikalischen Eigenschaften erfüllen. Erst die explizite Einforderung grundlegender physikalischer Prinzipien lenkt den Trainingsprozess verlässlich zu den erstrebenswerteren Lösungen.

Generell ist das Konzept der KNN-Materialmodellierung ein vielversprechender Weg hin zur Beschreibung immer komplexeren Materialverhaltens. Gleichwohl sind einige Nachteile nicht von der Hand zu weisen. Im Gegensatz zu klassischen Materialmodellen können die Gewichte nicht interpretierbar geändert werden, wie bspw. der Elastizitätsmodul („größerer Elastizitätsmodul heißt größere

Steifigkeit“). Dies trifft jedoch bei komplexeren analytischen Materialmodellen ebenfalls nicht mehr zu, wie bspw. schon beim OGDEN-Modell (vgl. STEINMANN ET AL. [163]). Ein weiterer Nachteil ist die große Menge nötiger Trainingsdaten, die diejenige von analytischen Materialmodellen weit übersteigt. Eine Extrapolation außerhalb der gegebenen Daten ist gleichfalls nicht möglich. Analytische Modelle sind dazu allerdings ebenfalls nur bedingt geeignet. Daher ist es unwahrscheinlich, dass KNN in Zukunft auf Basis gegebener Daten Materialverhalten außerhalb jener Daten vorhersagen können, was ein grundsätzliches Ziel klassischer Modellbildung ist.

Dafür sind KNN-Materialmodelle vergleichsweise schnell zu erstellen. Dazu wird an dieser Stelle in Erinnerung gerufen, dass im Prinzip das gleiche KNN-Materialmodell, welches in Kapitel 6 zur Modellierung plastischen Materialverhaltens beschrieben wurde, in Kapitel 7 das hysteretische Verhalten von ferroelektrischen Funktionskeramiken approximiert hat. Während die Erstellung analytischer Materialmodelle unter Umständen Jahre dauern kann, war das KNN-Materialmodell direkt einsetzbar. Als universelle Funktionsapproximatoren sind KNN darüber hinaus nicht an die Korsette der klassischen Materialmodellierung gebunden, die häufig aus pragmatischen Gründen geschnürt worden sind. Außerdem ist es durch Einbezug physikalischer Informationen möglich, den Nachteil der großen Datenmenge auf ein brauchbares Maß zu reduzieren.

Mit den genannten Vor- und Nachteilen im Kontext der KNN-Materialmodellierung und auf Basis der vorliegenden Arbeit können die folgenden Punkte Gegenstand weiterer Forschungsaktivitäten sein:

- Das Konzept der physikalischen Nebenbedingungen sollte weiter untersucht werden. Ein Algorithmus, der automatisch zu den Daten konforme Nebenbedingungen charakterisiert, wäre gerade im Bereich echter Versuchsdaten vorteilhaft.
- Wie in Kapitel 1.1 beschrieben, wird aktuell vergleichsweise viel zur KNN-Materialmodellierung veröffentlicht. In Zukunft muss eine Konsolidierung stattfinden, bei der die so zahlreich entwickelten und vielversprechenden Modelle miteinander verglichen werden. Dazu sind vor allem materialspezifische Benchmarks notwendig, um die Modelle miteinander objektiv zu vergleichen. Solche Vergleichsbeispiele sind nach bestem Wissen des Autors noch nicht vorhanden.
- Die Zuverlässigkeit der KNN-Materialmodelle ist noch relativ gering, auch aufgrund der inhärenten Variabilität des Trainingsprozesses. Je weniger Daten vorliegen, desto sensibler reagiert darüber hinaus das KNN auf die Einstellung der Hyperparameter, bspw. die Anzahl der Neuronen. Daher ist es

in Zukunft wichtig, einheitliche Abbruchkriterien für den Trainingsprozess zu entwickeln. Außerdem sollten Methoden angewendet werden, welche die Hyperparameter automatisch einstellen, etwa die von FUCHS ET AL. [51].

- Für geschichtsabhängiges Materialverhalten müssen FNN-Architekturen alle möglichen Inkremente für alle möglichen GGW-Punkte erlernen. Dies erfordert eine große Anzahl von Daten und induziert eine Schrittweitenabhängigkeit. Diese Problematik könnte in Zukunft durch neue KNN-Architekturen gelöst werden. Rekurrente neuronale Netze sind vielversprechend, weisen aktuell allerdings die gleichen Nachteile auf.
- Zur Kalibrierung analytischer Materialmodelle werden oft standardisierte Versuche durchgeführt, gerade wenn die zugehörigen Materialparameter nicht nur mathematische Koeffizienten, sondern auch bemessungsrelevante Grenzwerte darstellen. Im Fall der KNN-Materialmodellierung fällt letzteres weg. Daher sollten Strategien zur Versuchsdurchführung entwickelt werden, aus denen ein optimaler Informationsgehalt zum Training der KNN resultiert.

Es folgt ein subjektives Fazit des Autors: Die KNN-Materialmodellierung ist durch die Berücksichtigung physikalischer Information – ob per Nebenbedingungen schwach oder per Konstruktion stark eingefordert – einen großen Schritt in Richtung der praktischen Anwendung gegangen. Die klassische analytische Materialmodellierung wird dadurch jedoch nicht ersetzt werden. Wahrscheinlicher ist, dass beide Ansätze koexistieren und sich gegenseitig unterstützen. So kann die KNN-Materialmodellierung durch ihre Flexibilität beim Finden und Kalibrieren analytischer Materialmodelle helfen. Letztlich ist Materialmodellierung nicht nur eine Kurvenapproximation.

Abschließend wird die Brücke zu den ersten Sätzen der Einleitung geschlagen: Der Weg vom Verständnis der Materialstruktur über die daraus resultierenden Eigenschaften, die in einem Materialmodell repräsentiert werden sollen, bis hin zur ingenieurmäßigen Anwendung bei der Modellierung vollständiger Strukturen reicht über das Approximieren von Datenpunkten hinaus und muss gepflastert sein mit tieferen wissenschaftlichen und ingenieurmäßigen Erkenntnissen. KNN können auf diesem Weg sicherlich unterstützen.

A Elemente der Differentialrechnung

In diesem Kapitel werden die in der vorliegenden Arbeit verwendeten Konventionen der Matrix-Differentialrechnung angegeben und denjenigen der Tensor-Differentialrechnung gegenübergestellt. Die Verwendung von Matrizen bietet sich dabei im Rahmen einer reduzierten Tensornotation an. Dabei ist jedoch strikt auf die im folgenden festgelegten Konventionen zu achten. Neu definierte Vektoren werden stets als Spaltenvektoren festgelegt. In Tensornotation wird die EINSTEIN'sche Summenkonvention bei doppelt auftretenden Indizes angewendet.

$$\mathbf{a} = \sum_{i=1}^n a_i \mathbf{e}_i = a_i \mathbf{e}_i \quad \mathbf{a} = [a_1, \dots, a_i, \dots, a_n]^T \in \mathbb{R}^{n \times 1} \quad (\text{A1.1})$$

Die Ableitung eines Skalars $\rho(\mathbf{a})$ nach einem Vektor \mathbf{a} ist ein Zeilenvektor.

$$\frac{\partial \rho}{\partial \mathbf{a}} = \frac{\partial \rho}{\partial a_i} \mathbf{e}_i \quad \frac{\partial \rho}{\partial \mathbf{a}} = \left[\frac{\partial \rho}{\partial a_1}, \dots, \frac{\partial \rho}{\partial a_i}, \dots, \frac{\partial \rho}{\partial a_n} \right] \in \mathbb{R}^{1 \times n} \quad (\text{A1.2})$$

Die Ableitung eines Vektors $\mathbf{a}(\rho)$ nach einem Skalar ρ ist ein Spaltenvektor.

$$\frac{\partial \mathbf{a}}{\partial \rho} = \frac{\partial a_i}{\partial \rho} \mathbf{e}_i \quad \frac{\partial \mathbf{a}}{\partial \rho} = \left[\frac{\partial a_1}{\partial \rho}, \dots, \frac{\partial a_i}{\partial \rho}, \dots, \frac{\partial a_n}{\partial \rho} \right]^T \in \mathbb{R}^{n \times 1} \quad (\text{A1.3})$$

Die zweifache Ableitung eines Skalars $\rho(\mathbf{a}, \mathbf{b})$ nach $\mathbf{a} \in \mathbb{R}^{n \times 1}$ und anschließend nach $\mathbf{b} \in \mathbb{R}^{m \times 1}$ ist eine Matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$.

$$\begin{aligned} \frac{\partial^2 \rho}{\partial \mathbf{a} \partial \mathbf{b}} &= \frac{\partial^2 \rho}{\partial a_i \partial b_j} \mathbf{e}_i \otimes \mathbf{e}_j \\ \frac{\partial^2 \rho}{\partial \mathbf{a} \partial \mathbf{b}} &= \begin{bmatrix} \frac{\partial^2 \rho}{\partial a_1 \partial b_1} & \cdots & \frac{\partial^2 \rho}{\partial a_1 \partial b_m} \\ \vdots & \frac{\partial^2 \rho}{\partial a_i \partial b_j} & \vdots \\ \frac{\partial^2 \rho}{\partial a_n \partial b_1} & \cdots & \frac{\partial^2 \rho}{\partial a_n \partial b_m} \end{bmatrix} \in \mathbb{R}^{n \times m} \end{aligned} \quad (\text{A1.4})$$

B Reduzierte Notation von Tensoren

Der korrekte Umgang mit der reduzierten Tensornotation ermöglicht eine numerisch effiziente Implementierung. An dieser Stelle wird das grundlegende Prinzip beschrieben sowie in der Arbeit häufig genutzte Beispiele aufgelistet. Dabei betrifft dies i. d. R. das Tensorkalarprodukt

$$b = \mathbf{A} : \mathbf{E} = A_{ij} E_{ij} \quad (\text{A2.5})$$

und die doppelte Überschiebung

$$\mathbf{B} = B_{ij} \mathbf{e}_i \otimes \mathbf{e}_j = \mathbb{A} \mathbf{E} = A_{ijkl} E_{kl} \mathbf{e}_i \otimes \mathbf{e}_j, \quad (\text{A2.6})$$

mit dem Skalar b , den symmetrischen Tensoren 2. Stufe \mathbf{A} und \mathbf{B} sowie dem Tensor 4. Stufe \mathbb{A} . Letzterer besitzt die Symmetrieeigenschaften $A_{ijkl} = A_{klij} = A_{jikl} = A_{ijlk}$. Die reduzierten Notationen von \mathbf{A} , \mathbf{B} und \mathbb{A} werden im Rahmen dieser Arbeit so gewählt, dass sie mit der Anordnung der Komponenten des GREEN-LAGRANGE'schen Verzerrungstensors

$$\mathbf{E} = \begin{bmatrix} E_{11} & E_{22} & E_{33} & 2E_{12} & 2E_{13} & 2E_{23} \end{bmatrix}^T \quad (\text{A2.7})$$

verwendet werden können. Der Faktor 2 auf den Elementen $E_{ij}, i \neq j$ zielt auf die Konsistenz mit den infinitesimalen Ingenieurgleitungen $\gamma_{ij} = 2\varepsilon_{ij}$ ab.

Unter den genannten Vorgaben lautet die allgemeine reduzierte Notation eines symmetrischen Tensors 2. Stufe

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{22} & A_{33} & A_{12} & A_{13} & A_{23} \end{bmatrix}^T \in \mathbb{R}^{6 \times 1}, \quad (\text{A2.8})$$

wodurch das Skalarprodukt (A2.5) über

$$b = \mathbf{A}^T \mathbf{E} = \mathbf{E}^T \mathbf{A} \quad (\text{A2.9})$$

berechnet werden kann. Die allgemeine Form des Tensors 4. Stufe \mathbb{A} lässt sich in reduzierter Notation mit

$$\mathbf{A} = \begin{bmatrix} A_{1111} & A_{1122} & A_{1133} & A_{1112} & A_{1113} & A_{1123} \\ A_{2211} & A_{2222} & A_{2233} & A_{2212} & A_{2213} & A_{2223} \\ A_{3311} & A_{3322} & A_{3333} & A_{3312} & A_{3313} & A_{3323} \\ A_{1211} & A_{1222} & A_{1233} & A_{1212} & A_{1213} & A_{1223} \\ A_{1311} & A_{1322} & A_{1333} & A_{1312} & A_{1313} & A_{1323} \\ A_{2311} & A_{2322} & A_{2333} & A_{2312} & A_{2313} & A_{2323} \end{bmatrix} \in \mathbb{R}^{6 \times 6} \quad (\text{A2.10})$$

angeben, was sich über

$$\begin{aligned} B_{ij} = A_{ijkl} E_{kl} = & A_{ij11} E_{11} + A_{ij22} E_{22} + A_{ij33} E_{33} + \frac{1}{2} (A_{ij12} + A_{ij21}) (2E_{12}) \\ & + \frac{1}{2} (A_{ij13} + A_{ij31}) (2E_{13}) + \frac{1}{2} (A_{ij23} + A_{ij32}) (2E_{23}) \end{aligned} \quad (\text{A2.11})$$

zeigen lässt. Dabei wurde in Definition (A2.10) die kleine Symmetriebedingung $A_{ijkl} = A_{ijlk}$ berücksichtigt. Andernfalls ändert sich die Form (A2.10) sinngemäß in den Spalten vier bis sechs. Die doppelte Überschiebung (A2.6) lässt sich folglich durch die Vektor-Matrix-Multiplikation

$$\mathbf{B} = \mathbf{A} \mathbf{E} \quad (\text{A2.12})$$

in reduzierter Notation berechnen, wobei \mathbf{B} nach Gleichung (A2.8) definiert ist. Entsteht der Tensor $\mathbb{A} = \mathbf{B} \otimes \mathbf{C}$ durch ein dyadisches Produkt zweier Tensoren 2. Stufe \mathbf{B} und \mathbf{C} , so ist dieses ebenfalls in reduzierter Notation berechenbar mit

$$\mathbf{A} = \mathbf{B} \mathbf{C}^T . \quad (\text{A2.13})$$

In Abschnitt 2.1.6 wurden bereits die Spannungen \mathbf{S} und die Materialtangente \mathbf{C}_T in reduzierter Notation angegeben. Im Folgenden werden weitere spezielle Tensoren aufgelistet:

- Einheitstensor 2. Stufe: $\mathbf{I} = \delta_{ij} \mathbf{e}_i \otimes \mathbf{e}_j$

$$\mathbf{I} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}^T \in \mathbb{R}^{6 \times 1} . \quad (\text{A2.14})$$

- Symmetrischer Einheitstensor 4. Stufe: $\mathbb{I}^S = \frac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k \otimes \mathbf{e}_l$

$$\mathbb{I}^S = \text{diag} \left[1 \quad 1 \quad 1 \quad \frac{1}{2} \quad \frac{1}{2} \quad \frac{1}{2} \right]^T \in \mathbb{R}^{6 \times 6} . \quad (\text{A2.15})$$

- Die Ableitung der Inversen \mathbf{A}^{-1} eines Tensors 2. Stufe \mathbf{A} nach sich selbst ist ein Tensor 4. Stufe und lautet

$$\frac{\partial \mathbf{A}^{-1}}{\partial \mathbf{A}} = -\frac{1}{2} \left(A_{ik}^{-1} A_{lj}^{-1} + A_{il}^{-1} A_{kj}^{-1} \right) \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k \otimes \mathbf{e}_l , \quad (\text{A2.16})$$

vgl. HOLZAPFEL [81]. Mit Definition der Untermatrizen

$$\mathbf{A}_1 = \begin{bmatrix} (A_{11}^{-1})^2 & (A_{12}^{-1})^2 & (A_{13}^{-1})^2 \\ (A_{12}^{-1})^2 & (A_{22}^{-1})^2 & (A_{23}^{-1})^2 \\ (A_{13}^{-1})^2 & (A_{23}^{-1})^2 & (A_{33}^{-1})^2 \end{bmatrix} , \quad \mathbf{A}_2 = \begin{bmatrix} A_{11}^{-1} A_{12}^{-1} & A_{11}^{-1} A_{13}^{-1} & A_{12}^{-1} A_{13}^{-1} \\ A_{12}^{-1} A_{22}^{-1} & A_{12}^{-1} A_{23}^{-1} & A_{22}^{-1} A_{23}^{-1} \\ A_{13}^{-1} A_{23}^{-1} & A_{13}^{-1} A_{33}^{-1} & A_{23}^{-1} A_{33}^{-1} \end{bmatrix}$$

und

$$\mathbf{A}_3 = \frac{1}{2} \begin{bmatrix} A_{11}^{-1} A_{22}^{-1} + (A_{12}^{-1})^2 & A_{11}^{-1} A_{23}^{-1} + A_{13}^{-1} A_{12}^{-1} & A_{12}^{-1} A_{23}^{-1} + A_{13}^{-1} A_{22}^{-1} \\ A_{11}^{-1} A_{23}^{-1} + A_{13}^{-1} A_{12}^{-1} & A_{11}^{-1} A_{33}^{-1} + (A_{13}^{-1})^2 & A_{12}^{-1} A_{33}^{-1} + A_{13}^{-1} A_{23}^{-1} \\ A_{12}^{-1} A_{23}^{-1} + A_{13}^{-1} A_{22}^{-1} & A_{12}^{-1} A_{33}^{-1} + A_{13}^{-1} A_{23}^{-1} & A_{22}^{-1} A_{33}^{-1} + (A_{23}^{-1})^2 \end{bmatrix}$$

kann die reduzierte Notation für diese Ableitung wie folgt definiert werden.

$$\frac{\partial \mathbf{A}^{-1}}{\partial \mathbf{A}} = (-1) \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_2^T & \mathbf{A}_3 \end{bmatrix} \in \mathbb{R}^{6 \times 6} \quad (\text{A2.17})$$

C Strategien der inexakten Liniensuche

Die globale Minimierung einer Funktion $\mathcal{L}(\mathbf{w})$ erfolgt durch iterative Aktualisierung der Variablen \mathbf{w} über

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \alpha_n \mathbf{d}_n . \quad (\text{A3.18})$$

Während die Suchrichtung \mathbf{d}_n durch das Optimierungsverfahren vorgegeben wird, kann die Berechnung der Schrittweite α_n durch eine *inexakte Liniensuche* (LS) bestimmt werden. Dabei wird eine Schrittweite α_n gesucht, welche die Funktion $\mathcal{L}(\mathbf{w})$ mindestens weiter verringert und dabei gewisse Eigenschaften erfüllt. Die Funktion muss auf der gegebenen Linie nicht exakt minimiert werden. Nachfolgend werden die ARMIJO-Regel und die WOLFE-POWELL-Regel beschrieben, inklusive zugehöriger Algorithmen. Dabei werden nachfolgend die Abkürzungen

$$\varphi(\alpha) := \mathcal{L}(\mathbf{w}_n + \alpha \mathbf{d}_n) \quad \text{mit} \quad \varphi'(\alpha) = \nabla \mathcal{L}(\mathbf{w}_n + \alpha \mathbf{d}_n)^T \mathbf{d}_n \quad (\text{A3.19})$$

und

$$\Psi(\alpha) := \varphi(\alpha) - \varphi(0) - \sigma \alpha \varphi'(0) \quad (\text{A3.20})$$

für den Funktionswert und die Richtungsableitung auf der Linie $\mathbf{w}_n + \alpha \mathbf{d}_n$ eingeführt. Beide Regeln sind in Abbildung 3.1 grafisch dargestellt.

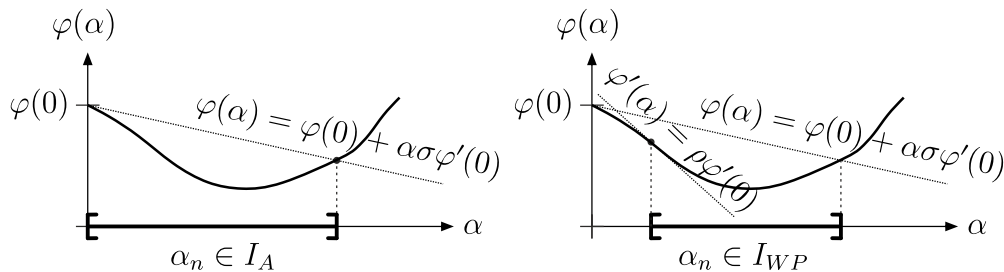


Abbildung 3.1: Die ARMIJO-Regel (links) und die WOLFE-POWELL-Regel (rechts)

C.1 Armijo-Liniensuche

Liegt ein Funktionswert $\varphi(\alpha)$ unter der sog. ARMIJO-GOLDSTEIN-Geraden, mit

$$\varphi(\alpha) \leq \varphi(0) + \sigma \alpha \varphi'(0) \quad \text{bzw} \quad \Psi(\alpha) \leq 0 , \quad (\text{A3.21})$$

dann genügt die Schrittweite α der ARMIJO-Regel, vgl. Abbildung 3.1 links. Mit dem Parameter $\sigma \in [0, 1)$ wird die Abstiegsbedingung $\varphi(\alpha) < \varphi(0)$ weiter verschärft. Der nachfolgende Algorithmus in Tafel C.1 verfolgt die Strategie eine möglichst große Schrittweite zu suchen, die gerade noch die Regel (A3.21) erfüllt. Die Parameter $\beta \in (0, 1)$ und $i_{max} > 1$ müssen gewählt werden. Die für das KNN-Training festgelegten Werte sind in Tabelle 4.2 aus Abschnitt 4.4 gegeben.

Parameter	$\sigma \in [0, 1), \beta \in (0, 1), i_{max} > 1$
Gegeben	$\mathbf{w}_n, \varphi(0) = \mathcal{L}_n, \nabla \mathcal{L}_n, \mathbf{d}_n$
Abstieg?	$\varphi'(0) = \nabla \mathcal{L}_n^T \mathbf{d}_n > 0? \rightarrow$ Exit LS (kein Abstieg)
Init.	$i = 0, \alpha_i = \alpha_0 = \alpha_{n-1}$ (Schrittweite der letzten Liniensuche)
Iteration	while ($i < i_{max}$) do Berechne $\mathcal{L}^{trial} = \mathcal{L}(\mathbf{w}_n + \alpha_i \mathbf{d}_n)$ und $\nabla \mathcal{L}^{trial}$ (<u>KNN</u>) Damit: $\Psi(\alpha_i)$, siehe Def. (A3.20) if $\Psi(\alpha_i) \leq 0$ then $\alpha_n = \alpha_i \rightarrow$ go to Erfolg! else $\alpha_{i+1} = \alpha_i \cdot \beta^i$ $i \leftarrow i + 1$ end if end while $i = i_{max}? \rightarrow$ Exit LS (keine Schrittweite gefunden)
Erfolg!	Ausgabe von $\alpha_n, \mathcal{L}_{n+1} = \mathcal{L}^{trial}$ und $\nabla \mathcal{L}_{n+1} = \nabla \mathcal{L}^{trial}$

Tafel C.1: Algorithmus der inexakten ARMIJO-LS zur Bestimmung von α_n

C.2 Wolfe-Powell-Liniensuche

Da die ARMIJO-Regel auch sehr kleine Schrittweiten α zulässt, kann das globale Liniensuchverfahren verlangsamen und schließlich zum Stillstand kommen. Die WOLFE-POWELL-Liniensuche (vgl. WOLFE [196, 197]) fordert daher die weitere Bedingung

$$\varphi'(\alpha) \geq \rho \varphi'(0) \quad (\text{A3.22})$$

ein, vgl. Abbildung 3.1 rechts, mit $\rho \in [\sigma, 1)$. Der in dieser Arbeit genutzte Algorithmus zur Suche von α_n ist in Tafel C.2 dargestellt. Er teilt sich in zwei Phasen auf: In Phase A wird durch Expansion zunächst ein Hilfsintervall $I_H = [a, b]$ gesucht, welches in Phase B wiederum solange verkleinert wird, bis die Bedingungen (A3.21) und (A3.22) erfüllt sind. Innerhalb von Phase B wird dabei jeweils das Minimum einer Parabel gesucht, welche von den Funktionswerten $\varphi(a)$, $\varphi(b)$ und der Ableitung $\varphi'(a)$ konstruiert werden kann. Für die genaue Herleitung und Motivation, siehe GEIGER und KANZOW [56]. Innerhalb des Algorithmus müssen neben ρ und σ vier weitere Parameter γ , τ_1 , τ_2 und i_{max} gewählt werden. In Tabelle 4.2 sind die Einstellungen gegeben, die hier für das KNN-Training gewählt werden.

Parameter	$\sigma \in [0, \frac{1}{2}), \rho \in [\sigma, 1), \gamma > 0, \tau_1, \tau_2 \in (0, \frac{1}{2}], i_{max} > 1$
Gegeben	$\mathbf{w}_n, \mathcal{L}_n, \nabla \mathcal{L}_n, \mathbf{d}_n$
Abstieg?	$\nabla \mathcal{L}_n^T \mathbf{d}_n = \varphi'(0) > 0? \rightarrow$ Exit LS (kein Abstieg)
Phase A	Expansionsphase: Finde Intervall $I_H = [a, b]$
Init	$i = 0, \alpha_i = \alpha_0 = \alpha_{n-1}$ (Schrittweite der letzten Liniensuche)
Iteration	while ($i < i_{max}$) do Berechne $\mathcal{L}^{trial} = \mathcal{L}(\mathbf{w}_n + \alpha_i \mathbf{d}_n)$ und $\nabla \mathcal{L}^{trial}$ (<u>KNN</u>) Damit: $\Psi(\alpha_i), \varphi'(\alpha_i)$ siehe Defs. (A3.20) und (A3.19) if ($\Psi(\alpha_i) \geq 0$) then $a = 0, b = \alpha_i \rightarrow$ go to Phase B elseif ($\Psi(\alpha_i) < 0$) and ($\varphi'(\alpha_i) \geq \rho \varphi'(0)$) then $\alpha_n = \alpha_i, \rightarrow$ go to Erfolg! elseif ($\Psi(\alpha_i) < 0$) and ($\varphi'(\alpha_i) < \rho \varphi'(0)$) then $\alpha_{i+1} = \gamma \cdot \alpha_i, i \leftarrow i + 1$ end if end while $i = i_{max}?$ \rightarrow Exit LS (keine Schrittweite gefunden)
Phase B	Kontraktionsphase: Verkleinere Hilfsintervall I_H zu I_{WP}
Init	$i = 0, a_i = a_0 = a, b_i = b_0 = b$ ($= \alpha$ von Phase A)
Iteration	while ($i < i_{max}$) do
Intervall	$\hat{a} = a_i + \tau_1(b_i - a_i), \hat{b} = b_i - \tau_2(b_i - a_i)$
Parabel	Identifiziere $\varphi(a_i), \varphi'(a_i)$ und $\varphi(b_i)$ $c_1 = (\varphi(b_i) - \varphi(a_i) - \varphi'(a_i)(b_i - a_i))/(b_i - a_i)^2, c_2 = \varphi'(a_i)$ $\alpha_i^{trial} = a_i - c_2/(2c_1)$ if $\alpha_i \notin [\hat{a}, \hat{b}]$ or $\varphi'(a_i) \geq 0$ then $\alpha_i = (\hat{a} + \hat{b})/2$ Berechne $\mathcal{L}^{trial} = \mathcal{L}(\mathbf{w}_n + \alpha_i \mathbf{d}_n)$ und $\nabla \mathcal{L}^{trial}$ (<u>KNN</u>) Damit: $\Psi(\alpha_i), \varphi'(\alpha_i)$ siehe Defs. (A3.20) und (A3.19) if ($\Psi(\alpha_i) \geq 0$) then $a_{i+1} = a_i, b_{i+1} = \alpha_i, i \leftarrow i + 1$ elseif ($\Psi(\alpha_i) < 0$) and ($\varphi'(\alpha_i) \geq \rho \varphi'(0)$) then $\alpha_n = \alpha_i, \rightarrow$ go to Erfolg! elseif ($\Psi(\alpha_i) < 0$) and ($\varphi'(\alpha_i) < \rho \varphi'(0)$) then $a_{i+1} = \alpha_i, b_{i+1} = b_i, i \leftarrow i + 1$ end if end while $i = i_{max}?$ \rightarrow Exit LS (keine Schrittweite gefunden)
Erfolg!	Ausgabe von $\alpha_n, \mathcal{L}_{n+1} = \mathcal{L}^{trial}$ und $\nabla \mathcal{L}_{n+1} = \nabla \mathcal{L}^{trial}$

Tafel C.2: Algorithmus einer inexakten WP-LS zur Bestimmung von α_n

D Ein inkompressibler Materialpunkt-Versuch

Der in Abschnitt 5.1.1 beschriebene, numerische Materialpunkt-Versuch wird hier um die Nebenbedingung der Inkompressibilität erweitert. Dadurch ist gewährleistet, dass das Volumenverhältnis $J = \sqrt{\det \mathbf{C}}$ während der gesamten Belastung konstant eins ist.

Der Versuch wird in den GREEN-LAGRANGE'schen Verzerrungen \mathbf{E} und den 2. PIOLA-KIRCHHOFF'schen Spannungen \mathbf{S} durchgeführt. Letztere werden im Grenzfall der Inkompressibilität nach Gleichung (3.84) über $\mathbf{S}(p, \mathbf{E}) = -p\mathbf{C}^{-1} + \mathbf{S}_{\text{dev}}(\mathbf{E})$ aufgeteilt. Dabei ist zu beachten, dass der hydrostatische Druck p , als auf der Momentankonfiguration definierte Spannungsgröße, nicht der Spur von \mathbf{S} entspricht. Von den Verzerrungen \mathbf{E} hängt ausschließlich der deviatorische Anteil $\mathbf{S}_{\text{dev}}(\mathbf{E})$ direkt ab. Die im Tensor \mathbf{E} formulierte Nebenbedingung

$$h(\mathbf{E}) = 4 \det \mathbf{E} + \text{Sp} \mathbf{E} + (\text{Sp} \mathbf{E})^2 - \text{Sp}(\mathbf{E}^2) \stackrel{!}{=} 0, \quad (\text{A4.23})$$

vgl. Gleichung (3.44), ist Teil des erweiterten Residuums

$$\hat{\mathbf{R}} := \begin{bmatrix} \mathbf{R} \\ h \end{bmatrix} = \begin{bmatrix} \mathbf{S}(\mathbf{E}, p) - \tilde{\mathbf{S}} \\ h(\mathbf{E}) \end{bmatrix}, \quad (\text{A4.24})$$

mit den ‚inneren‘ Spannungen \mathbf{S} und den ‚äußeren‘ Spannungen $\tilde{\mathbf{S}}$. Wie in Abschnitt 5.1.1 beschrieben, werden die Verzerrungskomponenten $\mathbf{E} = [\mathbf{E}^b, \mathbf{E}^u]$ und die Spannungskomponenten $\tilde{\mathbf{S}} = [\tilde{\mathbf{S}}^b, \tilde{\mathbf{S}}^u]$ aufgeteilt. Die Linearisierung lautet

$$L[\hat{\mathbf{R}}] = \begin{bmatrix} \mathbf{R}^b \\ \mathbf{R}^u \\ h \end{bmatrix} + \begin{bmatrix} \hat{\mathbf{C}}_T^{bb} & \hat{\mathbf{C}}_T^{bu} & -(\mathbf{C}^{-1})^b \\ \hat{\mathbf{C}}_T^{ub} & \hat{\mathbf{C}}_T^{uu} & -(\mathbf{C}^{-1})^u \\ dh/d\mathbf{E}^b & dh/d\mathbf{E}^u & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{E}^b \\ \Delta \mathbf{E}^u \\ \Delta p \end{bmatrix} - \begin{bmatrix} \tilde{\mathbf{S}}^b \\ \tilde{\mathbf{S}}^u \\ 0 \end{bmatrix}. \quad (\text{A4.25})$$

Die Ableitung der Inkompressibilitäts-Nebenbedingung resultiert mit

$$\frac{dh}{d\mathbf{E}} = 4(\det \mathbf{E})(\mathbf{E}^{-1})^T + (1 + 2\text{Sp} \mathbf{E}) \mathbf{I}^T - 2\mathbf{E}^T \quad (\text{A4.26})$$

in reduzierter Notation und kann mit den Zeilenvektoren $(\mathbf{E}^{-1})^T$, \mathbf{E}^T und \mathbf{I}^T formuliert werden, vgl. Definition (A2.8). Die Vektoren $(\mathbf{E}^{-1})^T$ und \mathbf{E}^T sind dabei nicht mit dem üblichen Faktor 2 zu bilden. Für $\det(\mathbf{E}) \approx 0$ muss der erste Term $(4(\det \mathbf{E})\mathbf{E}^{-1})$ aus Ableitung (A4.26) in einer Fallunterscheidung zu null gesetzt werden. Diese Problematik würde bei Definition anderer Verzerrungsmaße nicht auftreten. Die erweiterte Materialtangente

$$\hat{\mathbf{C}}_T = \mathbf{C}_T - 2p \frac{\partial \mathbf{C}^{-1}}{\partial \mathbf{C}} \quad (\text{A4.27})$$

berücksichtigt den hydrostatischen Druck p , der nicht Teil des deviatorischen Materialgesetzes $\mathbf{S}_{\text{dev}}(\mathbf{E})$ ist, mit $\mathbf{C}_T = \partial \mathbf{S}_{\text{dev}} / \partial \mathbf{E}$. Die reduzierte Notation des Tensors 4. Stufe $\partial \mathbf{C}^{-1} / \partial \mathbf{C}$ kann nach Gleichung (A2.17) bestimmt werden.

Unter Berücksichtigung der gemischten Steuerung mit $\Delta \mathbf{E}^b = \mathbf{0}$ und $\Delta \tilde{\mathbf{S}}^u = \mathbf{0}$ lässt sich die lineare Gleichung $L[\hat{\mathbf{R}}] = \mathbf{0}$ schrittweise folgendermaßen lösen:

$$\Delta p = \frac{\frac{dh}{d\mathbf{E}^u} (\hat{\mathbf{C}}_T^{uu})^{-1} \mathbf{R}^u - h}{\frac{dh}{d\mathbf{E}^u} (\hat{\mathbf{C}}_T^{uu})^{-1} (\mathbf{C}^{-1})^u} \quad (\text{A4.28})$$

$$\Delta \mathbf{E}^u = (\hat{\mathbf{C}}_T^{uu})^{-1} (\mathbf{C}^{-1})^u \Delta p - (\hat{\mathbf{C}}_T^{uu})^{-1} \mathbf{R}^u \quad (\text{A4.29})$$

$$\Delta \tilde{\mathbf{S}}^b = \mathbf{R}^b + \hat{\mathbf{C}}_T^{bu} \Delta \mathbf{E}^u - (\mathbf{C}^{-1})^b \Delta p . \quad (\text{A4.30})$$

Diese Aktualisierungsvorschriften können im NEWTON-Verfahren innerhalb des in Tafel 5.1 beschriebenen Materialpunkt-Versuchs sinngemäß eingepflegt werden. Die Gleichungen (A4.29) und (A4.30) entsprechen mit $p = 0$ und $\Delta p = 0$ den nicht-inkompressiblen Gleichungen (5.5) und (5.6).

Literatur

- [1] ABUEIDDA, D.W. / KORIC, S. / SOBH, N.A. / SEHITOGLU, H.: *Deep learning for plasticity and thermo-viscoplasticity*. International Journal of Plasticity 136 (2021), S. 102852.
- [2] ADELI, H.: *Neural Networks in Civil Engineering: 1989–2000*. Computer-Aided Civil and Infrastructure Engineering 16(2) (2001), S. 126–142.
- [3] AGRAWAL, A. / CHOUDHARY, A.: *Perspective: Materials informatics and big data: Realization of the “fourth paradigm” of science in materials science*. APL Materials 4(5) (2016), S. 053208.
- [4] ALTENBACH, H.: *Kontinuumsmechanik*. Springer Vieweg Berlin / Heidelberg, 2015.
- [5] ALTENBACH, H. / ALTENBACH, J. / RIKARDS, R.B.: *Einführung in die Mechanik der Laminat- und Sandwichtragwerke: Modellierung und Berechnung von Balken und Platten aus Verbundwerkstoffen*. Dt. Verlag für Grundstoffindustrie, 1996.
- [6] AMOS, B. / XU, L. / KOLTER, J.Z.: *Input Convex Neural Networks*. In: *Proceedings of the 34th International Conference on Machine Learning*, S. 146–155, 2017.
- [7] ARGYRIS, J.: *The computer shapes the theory*. Paper read to the Royal Aeronautical Society, Mai 1965 Lecture Summary Aeronautical Journal of the Royal Society 1965; 69: XXXII).
- [8] ARGYRIS, J.H.: *Energy Theorems and Structural Analysis: A Generalized Discourse with Applications on Energy Principles of Structural Analysis Including the Effects of Temperature and Non-Linear Stress-Strain Relations*. Aircraft Engineering and Aerospace Technology 26(10) (1954), S. 347–356.
- [9] AS’AD, F. / AVERY, P. / FARHAT, C.: *A mechanics-informed artificial neural network approach in data-driven constitutive modeling*. International Journal for Numerical Methods in Engineering 123(12) (2022), S. 2738–2759.
- [10] BALL, J.M.: *Convexity conditions and existence theorems in nonlinear elasticity*. Archive for Rational Mechanics and Analysis 63(4) (1976), S. 337–403.
- [11] BATUGIN, S.A. / NIRENBURG, R.K.: *Approximate relationship among the elastic constants of anisotropic rocks and anisotropy parameters*. Fiz-Tekhn. Probl. Razrabotki Polezn. Iskopaemykh 7(1) (1972), S. 7–11 (auf russisch).
- [12] BELLMAN, R.: *Dynamic Programming*. Science 153(3731) (1966), S. 34–37.
- [13] BERTRAM, A.: *Elasticity and Plasticity of Large Deformations*. Springer Berlin / Heidelberg, 2012.

- [14] BESSA, M.A. / BOSTANABAD, R. / LIU, Z. / HU, A. / APLEY, D.W. / BRINSON, C. / CHEN, W. / LIU, W.K.: *A framework for data-driven analysis of materials under uncertainty: Countering the curse of dimensionality*. Computer Methods in Applied Mechanics and Engineering 320 (2017), S. 633–667.
- [15] BETSCH, P. / GRUTTMANN, F. / STEIN, E.: *A 4-node finite shell element for the implementation of general hyperelastic 3D-elasticity at finite strains*. Computer Methods in Applied Mechanics and Engineering 130(1-2) (1996), S. 57–79.
- [16] BISHOP, C.M.: *Curvature-driven smoothing: a learning algorithm for feedforward networks*. IEEE Transactions on Neural Networks 4(5) (1993), S. 882–884.
- [17] BLATZ, P.J. / KO, W.L.: *Application of Finite Elastic Theory to the Deformation of Rubbery Materials*. Transactions of the Society of Rheology 6(1) (1962), S. 223–252.
- [18] BLASS, H.J. / GÖRLACHER, R. / STECK, G.: *Holzbauwerke nach Eurocode 5-Bemessung und Baustoffe-Step 1*. Düsseldorf: Fachverlag Holz, 1995.
- [19] BOCK, F.E. / AYDIN, R.C. / CYRON, C.J. / HUBER, N. / KALIDINDI, S.R. / KLUSEMANN, B.: *A Review of the Application of Machine Learning and Data Mining Approaches in Continuum Materials Mechanics*. Frontiers in Materials 6 (2019).
- [20] BONATTI, C. / BERISHA, B. / MOHR, D.: *From CP-FFT to CP-RNN: Recurrent neural network surrogate model of crystal plasticity*. International Journal of Plasticity 158 (2022), S. 103430.
- [21] BONATTI, C. / MOHR, D.: *On the importance of self-consistency in recurrent neural network models representing elasto-plastic solids*. Journal of the Mechanics and Physics of Solids 158 (2022), S. 104697.
- [22] BORKOWSKI, L. / SORINI, C. / CHATTOPADHYAY, A.: *Recurrent neural network-based multiaxial plasticity model with regularization for physics-informed constraints*. Computers and Structures 258 (2022), S. 106678.
- [23] BRANNON, R.M.: *Rotation, Reflection, and Frame Changes: Orthogonal tensors in computational engineering mechanics*. IOP Publishing, 2018.
- [24] BROYDEN, C.G.: *The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations*. IMA Journal of Applied Mathematics 6(1) (1970), S. 76–90.
- [25] CARRARA, P. / DE LORENZIS, L. / STAINIER, L. / ORTIZ, M.: *Data-driven fracture mechanics*. Computer Methods in Applied Mechanics and Engineering 372 (2020), S. 113390.
- [26] CIARLET, P.G.: *Three-dimensional elasticity*. Elsevier, 1988.

- [27] COLEMAN, B.D. / GURTIN, M.E.: *Thermodynamics with Internal State Variables*. The Journal of Chemical Physics 47(2) (1967), S. 597–613.
- [28] COLEMAN, B.D. / NOLL, W.: *On the thermostatics of continuous media*. Archive for Rational Mechanics and Analysis 4(1) (1959), S. 97–128.
- [29] COLEMAN, B.D. / NOLL, W.: *The thermodynamics of elastic materials with heat conduction and viscosity*. Archive for Rational Mechanics and Analysis 13(1) (1963), S. 167–178.
- [30] COURANT, R.: *Variational methods for the solution of problems of equilibrium and vibrations*. Bulletin of the American Mathematical Society 49(1) (1943), S. 1–23.
- [31] CYBENKO, G.: *Approximation by superpositions of a sigmoidal function*. Mathematics of Control, Signals, and Systems 2(4) (1989), S. 303–314.
- [32] CZARNECKI, W.M. / OSINDERO, S. / JADERBERG, M. / SWIRSZCZ, G. / PASCANU, R.: *Sobolev Training for Neural Networks*. In: *Advances in Neural Information Processing Systems*, Band 30, Curran Associates, Inc., 2017.
- [33] DE BOER, R.: *Vektor- und Tensorrechnung für Ingenieure*. Springer Berlin / Heidelberg, 1982.
- [34] DE SOUZA NETO, E.A. / PERIĆ, D. / OWEN, D.R.J.: *Computational Methods for Plasticity*. Wiley, 2008.
- [35] DIN EN 1993-1-1: *Eurocode 3: Bemessung und Konstruktion von Stahlbauten - Teil 1-1: Allgemeine Bemessungsregeln und Regeln für den Hochbau*. DIN Deutsches Institut für Normung e.V., August 2020.
- [36] DOLL, S. / SCHWEIZERHOF, K.: *On the Development of Volumetric Strain Energy Functions*. Journal of Applied Mechanics 67(1) (1999), S. 17–21.
- [37] DRUCKER, D.C.: *A Definition of Stable Inelastic Material*. Journal of Applied Mechanics 26(1) (1959), S. 101–106.
- [38] DRUCKER, D.C. / PRAGER, W. / GREENBERG, H.J.: *Extended limit design theorems for continuous media*. Quarterly of applied mathematics 9(4) (1952), S. 381–389.
- [39] DVORKIN, E.N. / BATHE, K.-J.: *A continuum mechanics based four-node shell element for general non-linear analysis*. Engineering Computations 1(1) (1984), S. 77–88.
- [40] EBERHARDSTEINER, J.: *Mechanisches Verhalten von Fichtenholz*. Springer Vienna, 2002.
- [41] ELMAN, J.L.: *Finding Structure in Time*. Cognitive Science 14(2) (1990), S. 179–211.

- [42] FERNÁNDEZ, M. / JAMSHIDIAN, M. / BÖHLKE, T. / KERSTING, K. / WEEGER, O.: *Anisotropic hyperelastic constitutive models for finite deformations combining material theory and data-driven approaches with application to cubic lattice metamaterials*. Computational Mechanics 67(2) (2020), S. 653–677.
- [43] FLASCHEL, M. / KUMAR, S. / DE LORENZIS, L.: *Unsupervised discovery of interpretable hyperelastic constitutive laws*. Computer Methods in Applied Mechanics and Engineering 381 (2021), S. 113852.
- [44] FLASCHEL, M. / KUMAR, S. / DE LORENZIS, L.: *Discovering plasticity models without stress data*. npj Computational Materials 8(91) (2022).
- [45] FLETCHER, R.: *A new approach to variable metric algorithms*. The Computer Journal 13(3) (1970), S. 317–322.
- [46] FLOOD, I. / KARTAM, N.: *Neural Networks in Civil Engineering. I: Principles and Understanding*. Journal of Computing in Civil Engineering 8(2) (1994), S. 131–148.
- [47] FLOOD, I. / KARTAM, N.: *Neural Networks in Civil Engineering. II: Systems and Application*. Journal of Computing in Civil Engineering 8(2) (1994), S. 149–162.
- [48] FRANKE, M. / KLEIN, D.K. / WEEGER, O. / BETSCH, P.: *Advanced discretization techniques for hyperelastic physics-augmented neural networks*. arXiv preprint arXiv:2306.09866, 2023.
- [49] FREITAG, S. / GRAF, W. / KALISKE, M.: *Recurrent neural networks for fuzzy data*. Integrated Computer-Aided Engineering 18(3) (2011), S. 265–280.
- [50] FREITAG, S. / GRAF, W. / KALISKE, M.: *A material description based on recurrent neural networks for fuzzy data and its application within the finite element method*. Computers and Structures 124 (2013), S. 29–37.
- [51] FUCHS, A. / HEIDER, Y. / WANG, K. / SUN, W.C. / KALISKE, M.: *DNN2: A hyper-parameter reinforcement learning game for self-design of neural network based elasto-plastic constitutive descriptions*. Computers and Structures 249 (2021), S. 106505.
- [52] FURUKAWA, T. / YAGAWA, G.: *Implicit constitutive modelling for viscoplasticity using neural networks*. International Journal for Numerical Methods in Engineering 43(2) (1998), S. 195–219.
- [53] GAJEK, S. / SCHNEIDER, M. / BÖHLKE, T.: *On the micromechanics of deep material networks*. Journal of the Mechanics and Physics of Solids 142 (2020), S. 103984.
- [54] GAJEK, S. / SCHNEIDER, M. / BÖHLKE, T.: *An FE-DMN method for the multiscale analysis of thermomechanical composites*. Computational Mechanics 69(5) (2022), S. 1087–1113.

- [55] GALERKIN, B.G.: *Rods and Plates, Series Occuring in Various Questions Concerning the Elastic Equilibrium of Rods and Plates*. Engineers Bulletin (Vestnik Inzhenerov) 19 (1915), S. 897–908.
- [56] GEIGER, C. / KANZOW, C.: *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*. Springer Berlin / Heidelberg, 1999.
- [57] GEIGER, C. / KANZOW, C.: *Theorie und Numerik restringierter Optimierungsaufgaben*. Springer Berlin / Heidelberg, 2002.
- [58] GHABOUSSI, J. / GARRETT, J.H. / WU, X.: *Knowledge-Based Modeling of Material Behavior with Neural Networks*. Journal of Engineering Mechanics 117(1) (1991), S. 132–153.
- [59] GHABOUSSI, J. / SIDARTA, D.E.: *New nested adaptive neural networks (NANN) for constitutive modeling*. Computers and Geotechnics 22(1) (1998), S. 29–52.
- [60] GHAVAMIAN, F. / SIMONE, A.: *Accelerating multiscale finite element simulations of history-dependent materials using a recurrent neural network*. Computer Methods in Applied Mechanics and Engineering 357 (2019), S. 112594.
- [61] GIRKMANN, K.: *Flächentragwerke*. Springer Berlin / Heidelberg, 1959.
- [62] GLOS, P.: *Zur Bestimmung des Festigkeitsverhaltens von Brettschichtholz bei Druckbeanspruchung aus Werkstoff- und Einwirkungskenngrößen*. Dissertation, Technische Universität München.
- [63] GOLDFARB, D.: *A family of variable-metric methods derived by variational means*. Mathematics of Computation 24(109) (1970), S. 23–26.
- [64] GOODFELLOW, I. / Y.B. / COURVILLE, A.: *Deep Learning*. MIT Press, 2016 <http://www.deeplearningbook.org> (besucht am 07.03.23).
- [65] GORJI, M.B. / MOZAFFAR, M. / HEIDENREICH, J.N. / CAO, J. / MOHR, D.: *On the potential of recurrent neural networks for modeling path dependent plasticity*. Journal of the Mechanics and Physics of Solids 143 (2020), S. 103972.
- [66] GRAY, J.: *Change and Variations*. Springer Cham, 2021.
- [67] GRUTTMANN, F. / WAGNER, W.: *An advanced shell model for the analysis of geometrical and material nonlinear shells*. Computational Mechanics 66(6) (2020), S. 1353–1376.
- [68] HADAMARD, J.: *Leçons sur la propagation des ondes et les équations de l'hydrodynamique*. A. Hermann, 1903.
- [69] HAGAN, M.T. / MENHAJ, M.B.: *Training feedforward networks with the Marquardt algorithm*. IEEE Transactions on Neural Networks 5(6) (1994), S. 989–993.

- [70] HAJ-ALI, R. / PECKNOLD, D.A. / GHABOUSSI, J. / VOYIADJIS, G.Z.: *Simulated Micromechanical Models Using Artificial Neural Networks*. Journal of Engineering Mechanics 127(7) (2001), S. 730–738.
- [71] HALPHEN, B. / SON NGUYEN, Q.: *Sur les matériaux standard généralisés*. Journal de Mécanique 14(1) (1975), S. 39–63.
- [72] HASHASH, Y.M.A. / JUNG, S. / GHABOUSSI, J.: *Numerical implementation of a neural network based material model in finite element analysis*. International Journal for Numerical Methods in Engineering 59(7) (2004), S. 989–1005.
- [73] HEBB, D.O.: *The Organization of Behavior*. Psychology Press, 2005.
- [74] HEIDER, Y. / WANG, K. / SUN, W.: *SO(3)-invariance of informed-graph-based deep neural network for anisotropic elastoplastic materials*. Computer Methods in Applied Mechanics and Engineering 363 (2020), S. 112875.
- [75] HENKES, A. / WESSELS, H. / MAHNKEN, R.: *Physics informed neural networks for continuum micromechanics*. Computer Methods in Applied Mechanics and Engineering 393 (2022), S. 114790.
- [76] HERCULANO-HOUZEL, S.: *The Human Advantage: How Our Brains Became Remarkable*. MIT Press, 2017.
- [77] HEY, A.J.G. / TANSLEY, S. / TOLLE, K.M. / OTHERS: *The fourth paradigm: data-intensive scientific discovery*. Band 1, Microsoft research Redmond, WA, 2009.
- [78] HILL, R.: *On uniqueness and stability in the theory of finite elastic strain*. Journal of the Mechanics and Physics of Solids 5(4) (1957), S. 229–241.
- [79] HILL, R.: *A general theory of uniqueness and stability in elastic-plastic solids*. Journal of the Mechanics and Physics of Solids 6(3) (1958), S. 236–249.
- [80] HOCHREITER, S. / SCHMIDHUBER, J.: *Long Short-Term Memory*. Neural Computation 9(8) (1997), S. 1735–1780.
- [81] HOLZAPFEL, G.A.: *Nonlinear Solid Mechanics: A Continuum Approach for Engineering*. John Wiley & Sons, 2000.
- [82] HOLZAPFEL, G.A. / GASSER, T.C. / OGDEN, R.W.: *A New Constitutive Framework for Arterial Wall Mechanics and a Comparative Study of Material Models*. Journal of Elasticity 61(1/3) (2000), S. 1–48.
- [83] HOPFIELD, J.J.: *Neural networks and physical systems with emergent collective computational abilities*. Proceedings of the national academy of sciences 79(8) (1982), S. 2554–2558.

-
- [84] HORGAN, C.O. / MURPHY, J.G.: *Simple Shearing of Incompressible and Slightly Compressible Isotropic Nonlinearly Elastic Materials*. Journal of Elasticity 98(2) (2009), S. 205–221.
- [85] HOULSBY, G.T. / PUZRIN, A.M.: *Principles of hyperplasticity: an approach to plasticity theory based on thermodynamic principles*. Springer Science & Business Media, 2007.
- [86] HUANG, D. / FUHG, J.N. / WEISSENFELS, C. / WRIGGERS, P.: *A machine learning based plasticity model using proper orthogonal decomposition*. Computer Methods in Applied Mechanics and Engineering 365 (2020), S. 113008.
- [87] HUGHES, T.J.R., SIMO J.C.: *Computational Inelasticity*. Springer New York, 2000.
- [88] HUGHES, T.J.R.: *The Finite Element Method*. Dover Publications Inc., 2000.
- [89] HUGHES, T.J.R. / COTTRELL, J.A. / BAZILEVS, Y.: *Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement*. Computer Methods in Applied Mechanics and Engineering 194(39-41) (2005), S. 4135–4195.
- [90] IBRAHIMBEGOVIC, A.: *Nonlinear Solid Mechanics*. Springer Dordrecht, 2009.
- [91] JOHANSSON, E.M. / DOWLA, F.U. / GOODMAN, D.M.: *Backpropagation learning for multilayer feed-forward neural networks using the conjugate gradient method*. International Journal of Neural Systems 02(04) (1991), S. 291–301.
- [92] JONES, D.F. / TRELOAR, L.R.G.: *The properties of rubber in pure homogeneous strain*. Journal of Physics D: Applied Physics 8(11) (1975), S. 1285–1304.
- [93] JORDAN, M.: *Attractor dynamics and parallelism in a connectionist sequential machine*. In: *Eighth Annual Conference of the Cognitive Science Society*, S. 513–546, 1986.
- [94] JUNG, S. / GHABOUSSI, J.: *Neural network constitutive model for rate-dependent materials*. Computers and Structures 84(15-16) (2006), S. 955–963.
- [95] KALINA, K.A. / LINDEN, L. / BRUMMUND, J. / KÄSTNER, M.: *FE²: an efficient data-driven multiscale approach based on physics-constrained neural networks and automated data mining*. Computational Mechanics 71(5) (2023), S. 827–851.
- [96] KALINA, K.A. / LINDEN, L. / BRUMMUND, J. / METSCH, P. / KÄSTNER, M.: *Automated constitutive modeling of isotropic hyperelasticity based on artificial neural networks*. Computational Mechanics 69(1) (2021), S. 213–232.
- [97] KARAPIPERIS, K. / STAINIER, L. / ORTIZ, M. / ANDRADE, J.E.: *Data-Driven multiscale modeling in mechanics*. Journal of the Mechanics and Physics of Solids 147 (2021), S. 104239.

- [98] KHAN, A.S. / HUANG, S.: *Continuum Theory of Plasticity*. John Wiley & Sons, 1995.
- [99] KIDGER, P. / LYONS, T.: *Universal approximation with deep narrow networks*. In: *Conference on learning theory*, PMLR, S. 2306–2327, 2020.
- [100] KINGMA, D.P. / BA, J.: *Adam: A method for stochastic optimization*. arXiv preprint arXiv:1412.6980, 2014.
- [101] KIRCHDOERFER, T. / ORTIZ, M.: *Data-driven computational mechanics*. *Computer Methods in Applied Mechanics and Engineering* 304 (2016), S. 81–101.
- [102] KIRCHDOERFER, T. / ORTIZ, M.: *Data-driven computing in dynamics*. *International Journal for Numerical Methods in Engineering* 113(11) (2017), S. 1697–1710.
- [103] KLEIN, D.K. / FERNÁNDEZ, M. / MARTIN, R.J. / NEFF, P. / WEEGER, O.: *Polyconvex anisotropic hyperelasticity with neural networks*. *Journal of the Mechanics and Physics of Solids* 159 (2022), S. 104703.
- [104] KLINKEL, S.: *Theorie und Numerik eines Volumen-Schalen-Elementes bei finiten elastischen und plastischen Verzerrungen*. Dissertation, Universität Karlsruhe, 2000.
- [105] KLINKEL, S.: *A phenomenological constitutive model for ferroelastic and ferroelectric hysteresis effects in ferroelectric ceramics*. *International Journal of Solids and Structures* 43(22-23) (2006), S. 7197–7222.
- [106] KLINKEL, S. / GRUTTMANN, F. / WAGNER, W.: *A mixed shell formulation accounting for thickness strains and finite strain 3d material models*. *International Journal for Numerical Methods in Engineering* 74(6) (2008), S. 945–970.
- [107] KOLLMANNBERGER, S. / D'ANGELLA, D. / JOKEIT, M. / HERRMANN, L.: *Deep Learning in Computational Mechanics*. Springer International Publishing, 2021.
- [108] KRIESEL, D.: *Ein kleiner Überblick über Neuronale Netze*. Erhältlich auf <http://www.dkriesel.com> (besucht am 10.03.23), 2007.
- [109] KURRER, K.-E.: *Geschichte der Baustatik*. Wiley, 2016.
- [110] LECUN, Y. / BOTTOU, L. / ORR, G.B. / MÜLLER, K.-R.: *Efficient BackProp*. In: *Lecture Notes in Computer Science*, S. 9–50, Springer Berlin / Heidelberg, 1998.
- [111] LEFIK, M. / BOSO, D.P. / SCHREFLER, B.A.: *Artificial Neural Networks in numerical modelling of composites*. *Computer Methods in Applied Mechanics and Engineering* 198(21-26) (2009), S. 1785–1804.

-
- [112] LEFIK, M. / SCHREFLER, B.A.: *Artificial neural network as an incremental non-linear constitutive model for a finite element code*. Computer Methods in Applied Mechanics and Engineering 192(28-30) (2003), S. 3265–3283.
- [113] LEKHNITSKII, S.G. / FERN, P. / BRANDSTATTER, J.J. / DILL, E.H.: *Theory of Elasticity of an Anisotropic Elastic Body*. Physics Today 17(1) (1964), S. 84–84.
- [114] LEMAITRE, J. / CHABOCHE, J.-L.: *Mechanics of Solid Materials*. Cambridge University Press, 1990.
- [115] LI, B. / ZHUANG, X.: *Multiscale computation on feedforward neural network and recurrent neural network*. Frontiers of Structural and Civil Engineering 14(6) (2020), S. 1285–1298.
- [116] LIANG, G. / CHANDRASHEKHARA, K.: *Neural network based constitutive model for elastomeric foams*. Engineering Structures 30(7) (2008), S. 2002–2011.
- [117] LINDEN, L. / KLEIN, D.K. / KALINA, K.A. / BRUMMUND, J. / WEEGER, O. / KÄSTNER, M.: *Neural networks meet hyperelasticity: A guide to enforcing physics*. 2023.
- [118] LINKA, K. / HILLGÄRTNER, M. / ABDOLAZIZI, K.P. / AYDIN, R.C. / ITSKOV, M. / CYRON, C.J.: *Constitutive artificial neural networks: A fast and general approach to predictive data-driven constitutive modeling by deep learning*. Journal of Computational Physics 429 (2021), S. 110010.
- [119] LIU, I.S.: *Continuum Mechanics*. Springer, 2002.
- [120] LIU, Z. / WU, C.T.: *Exploring the 3D architectures of deep material network in data-driven multiscale mechanics*. Journal of the Mechanics and Physics of Solids 127 (2019), S. 20–46.
- [121] LIU, Z. / WU, C.T. / KOISHI, M.: *A deep material network for multiscale topology learning and accelerated nonlinear modeling of heterogeneous materials*. Computer Methods in Applied Mechanics and Engineering 345 (2019), S. 1138–1168.
- [122] MACKENZIE-HELNWEIN, P. / EBERHARDSTEINER, J. / MANG, H.A.: *A multi-surface plasticity model for clear wood and its application to the finite element analysis of structural details*. Computational Mechanics 31(1-2) (2003), S. 204–218.
- [123] MAGNUS, J.R.: *On Differentiating Eigenvalues and Eigenvectors*. Econometric Theory 1(2) (1985), S. 179–191.
- [124] MÁRQUEZ-NEILA, P. / SALZMANN, M. / FUA, P.: *Imposing hard constraints on deep networks: Promises and limitations*. arXiv preprint arXiv:1706.02025, 2017.

- [125] MASI, F. / STEFANOU, I.: *Multiscale modeling of inelastic materials with Thermodynamics-based Artificial Neural Networks (TANN)*. Computer Methods in Applied Mechanics and Engineering 398 (2022), S. 115190.
- [126] MASI, F. / STEFANOU, I. / VANNUCCI, P. / MAFFI-BERTHIER, V.: *Thermodynamics-based Artificial Neural Networks for constitutive modeling*. Journal of the Mechanics and Physics of Solids 147 (2021), S. 104277.
- [127] MAUGIN, G.A.: *Continuum mechanics of electromagnetic solids*. Elsevier, 2013.
- [128] MCCULLOCH, W.S. / PITTS, W.: *A logical calculus of the ideas immanent in nervous activity*. The Bulletin of Mathematical Biophysics 5(4) (1943), S. 115–133.
- [129] MCKAY, M.D. / BECKMAN, R.J. / CONOVER, W.J.: *A comparison of three methods for selecting values of input variables in the analysis of output from a computer code*. Technometrics 42(1) (2000), S. 55–61.
- [130] MIEHE, C. / ROSATO, D. / KIEFER, B.: *Variational principles in dissipative electro-magneto-mechanics: A framework for the macro-modeling of functional materials*. International Journal for Numerical Methods in Engineering 86(10) (2011), S. 1225–1276.
- [131] MINSKY, MARVIN / PAPERT, SEYMOUR: *An introduction to computational geometry*. Cambridge trass., HIT 479(480) (1969), S. 104.
- [132] MONTÁNS, F.J. / CHINESTA, F. / GÓMEZ-BOMBARELLI, R. / KUTZ, J.N.: *Data-driven modeling and learning in science and engineering*. Comptes Rendus Mécanique 347(11) (2019), S. 845–855.
- [133] MOONEY, M.: *A Theory of Large Elastic Deformation*. Journal of Applied Physics 11(9) (1940), S. 582–592.
- [134] MOORE, E.H.: *On the Reciprocal of the General Algebraic Matrix*. Bulletin of American Mathematical Society 26 (1920), S. 394–395.
- [135] MOZAFFAR, M. / BOSTANABAD, R. / CHEN, W. / EHMANN, K. / CAO, J. / BESSA, M.A.: *Deep learning predicts path-dependent plasticity*. Proceedings of the National Academy of Sciences 116(52) (2019), S. 26414–26420.
- [136] MÜLLNER, H.W. / EBERHARDSTEINER, J. / MACKENZIE-HELNWEIN, P.: *Konstitutives Modellieren von Fichtenholz unter biaxialer Beanspruchung mittels eines orthotropen Einflächensmodells unter Berücksichtigung von Ver- und Entfestigung*. PAMM 4(1) (2004), S. 264–265.
- [137] NYE, J.F.: *Physical Properties of Crystals*. Oxford University Press, 1985.
- [138] OESER, M. / FREITAG, S.: *Modeling of materials with fading memory using neural networks*. International Journal for Numerical Methods in Engineering 78(7) (2008), S. 843–862.

- [139] OGDEN, R.W.: *Large deformation isotropic elasticity - on the correlation of theory and experiment for compressible rubberlike solids*. Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences 328(1575) (1972), S. 567–583.
- [140] OGDEN, R.W.: *Large deformation isotropic elasticity – on the correlation of theory and experiment for incompressible rubberlike solids*. Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences 326(1567) (1972), S. 565–584.
- [141] OGDEN, R.W.: *Recent Advances in the Phenomenological Theory of Rubber Elasticity*. Rubber Chemistry and Technology 59(3) (1986), S. 361–383.
- [142] PARISCH, H.: *Efficient non-linear finite element shell formulation involving large strains*. Engineering Computations 3(2) (1986), S. 121–128.
- [143] PENROSE, R.: *A generalized inverse for matrices*. Mathematical Proceedings of the Cambridge Philosophical Society 51(3) (1955), S. 406–413.
- [144] PI CERAMIC GMBH: *Piezoelectric Ceramic Products: Fundamentals, characteristics and applications*. 2016 <https://www.piceramic.de/de/> (besucht am 01.07.23).
- [145] RAISSI, M. / PERDIKARIS, P. / KARNIADAKIS, G.E.: *Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*. Journal of Computational Physics 378 (2019), S. 686–707.
- [146] RAOULT, A.: *Non-polyconvexity of the stored energy function of a Saint Venant-Kirchhoff material*. Aplikace matematiky 31(6) (1986), S. 417–419.
- [147] REESE, S.: *Theorie und Numerik des Stabilitätsverhaltens hyperelastischer Festkörper*. Dissertation, Technische Universität Darmstadt, 1994.
- [148] RIVLIN, R.S.: *Large elastic deformations of isotropic materials IV. further developments of the general theory*. Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences 241(835) (1948), S. 379–397.
- [149] RIVLIN, R.S. / ERICKSEN, J.L.: *Stress-Deformation Relations for Isotropic Materials*. Journal of Rational Mechanics and Analysis 4 (1955), S. 323–425.
- [150] ROBINSON, A.J. / FALLSIDE, F.: *The utility driven dynamic error propagation network*. Band 1, University of Cambridge Department of Engineering Cambridge, 1987.
- [151] ROSENBLATT, F.: *The perceptron: A probabilistic model for information storage and organization in the brain*. Psychological Review 65(6) (1958), S. 386–408.

- [152] RUMELHART, D.E. / HINTON, G.E. / WILLIAMS, R.J.: *Learning internal representations by error propagation*. California Univ San Diego La Jolla Inst for Cognitive Science, 1985.
- [153] SELBY, S.M.: *Standard Mathematical Tables, Cleveland*. 1975.
- [154] SETTGAST, C. / ABENDROTH, M. / KUNA, M.: *Constitutive modeling of plastic deformation behavior of open-cell foam structures using neural networks*. Mechanics of Materials 131 (2019), S. 1–10.
- [155] SHANNO, D.F.: *Conditioning of quasi-Newton methods for function minimization*. Mathematics of Computation 24(111) (1970), S. 647–656.
- [156] SHEN, Y. / CHANDRASHEKHARA, K. / BREIG, W.F. / OLIVER, L.R.: *Neural Network Based Constitutive Model for Rubber Material*. Rubber Chemistry and Technology 77(2) (2004), S. 257–277.
- [157] SHEPHERD, A.J.: *Second-Order Methods for Neural Networks*. Springer London, 1997.
- [158] SIMO, J.C.: *Topics on the Numerical Analysis and Simulation of Plasticity*. In: CIARLET, R. S. / LIONS, J. L., (HRSG.), *Handbook of Numerical Analysis*, Elsevier, Amsterdam, 1994.
- [159] SIMO, J.C. / RIFAI, M.S.: *A class of mixed assumed strain methods and the method of incompatible modes*. International Journal for Numerical Methods in Engineering 29(8) (1990), S. 1595–1638.
- [160] SIMO, J.C. / TAYLOR, R.L.: *Consistent tangent operators for rate-independent elastoplasticity*. Computer Methods in Applied Mechanics and Engineering 48(1) (1985), S. 101–118.
- [161] SIMO, J.C. / TAYLOR, R.L.: *Quasi-incompressible finite elasticity in principal stretches. continuum basis and numerical algorithms*. Computer Methods in Applied Mechanics and Engineering 85(3) (1991), S. 273–310.
- [162] SIMO, J.C. / TAYLOR, R.L. / PISTER, K.S.: *Variational and projection methods for the volume constraint in finite deformation elasto-plasticity*. Computer Methods in Applied Mechanics and Engineering 51(1-3) (1985), S. 177–208.
- [163] STEINMANN, P. / HOSSAIN, M. / POSSART, G.: *Hyperelastic models for rubber-like materials: consistent tangent operators and suitability for Treloar's data*. Archive of Applied Mechanics 82(9) (2012), S. 1183–1217.
- [164] STEPHAN, P. / SCHABER, K. / STEPHAN, K. / MAYINGER, F.: *Thermodynamik*. Springer Berlin / Heidelberg, 2013.
- [165] STIERSTADT, K.: *Thermodynamik*. Springer Berlin / Heidelberg, 2010.

-
- [166] SUTTER, F. / KAMLAH, M.: *Variational free energy based macroscopical modeling of ferroelectroelasticity*. Journal of the Mechanics and Physics of Solids 178 (2023), S. 105341.
- [167] SZABÓ, I.: *Geschichte der Mechanischen Prinzipien*. Birkhäuser Basel, 1977.
- [168] TAYLOR, R.L.: *FEAP - Finite Element Analysis Program*. University of California, Berkeley, 2023 <http://projects.ce.berkeley.edu/feap/> (besucht am 10.03.23).
- [169] TENSORFLOW: *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015 Autoren, siehe: <https://www.tensorflow.org/> (besucht am 02.07.23).
- [170] THE MATHWORKS INC.: *MATLAB version: 9.11.0 (R2021b)*. 2021.
- [171] TRELOAR, L.R.G.: *The elasticity of a network of long-chain molecules—II*. Transactions of the Faraday Society 39 (1943), S. 241–246.
- [172] TRELOAR, L.R.G.: *Stress-strain data for vulcanised rubber under various types of deformation*. Transactions of the Faraday Society 40 (1944), S. 59–70.
- [173] TRELOAR, L.R.G.: *The Physics of Rubber Elasticity*. Oxford Classic Texts in the Physical Sciences OUP Oxford, 1975.
- [174] TRESCA, H.E.: *Sur l'écoulement des corps solides soumis a de fortes pressions*. Imprimerie de Gauthier-Villars, successeur de Mallet-Bachelier, rue de Seine-Saint-Germain, 10, près l'Institut, 1864.
- [175] TRUESDELL, C. / NOLL, W.: *The Non-Linear Field Theories of Mechanics*. Springer Berlin / Heidelberg, 2004.
- [176] TURNER, M.J. / CLOUGH, R.W. / MARTIN, H.C. / TOPP, L.J.: *Stiffness and Deflection Analysis of Complex Structures*. Journal of the Aeronautical Sciences 23(9) (1956), S. 805–823.
- [177] UNGER, J.F. / KÖNKE, C.: *Coupling of scales in a multiscale simulation using neural networks*. Computers and Structures 86(21-22) (2008), S. 1994–2003.
- [178] UNGER, J.F. / KÖNKE, C.: *Neural networks as material models within a multiscale approach*. Computers and Structures 87(19-20) (2009), S. 1177–1186.
- [179] VALANIS, K.C. / LANDEL, R.F.: *The Strain-Energy Function of a Hyperelastic Material in Terms of the Extension Ratios*. Journal of Applied Physics 38(7) (1967), S. 2997–3002.
- [180] VAN DER SMAGT, P.: *Minimisation methods for training feedforward neural networks*. Neural Networks 7(1) (1994), S. 1–11.

- [181] VLASSIS, N.N. / MA, R. / SUN, W.: *Geometric deep learning for computational mechanics Part I: anisotropic hyperelasticity*. Computer Methods in Applied Mechanics and Engineering 371 (2020), S. 113299.
- [182] VOIGT, W.: *Die fundamentalen Physikalischen Eigenschaften der Krystalle in elementarer Darstellung*. De Gruyter, 1898.
- [183] VOIGT, W.: *Lehrbuch der Kristallphysik (mit Ausschluss der Kristalloptik)*. Vieweg+Teubner Verlag, 1928.
- [184] WAGNER, W. / GRUTTMANN, F.: *A simple finite rotation formulation for composite shell elements*. Engineering Computations 11(2) (1994), S. 145–176.
- [185] WAGNER, W. / GRUTTMANN, F.: *A robust non-linear mixed hybrid quadrilateral shell element*. International Journal for Numerical Methods in Engineering 64(5) (2005), S. 635–666.
- [186] WAGNER, W. / GRUTTMANN, F.: *An improved quadrilateral shell element based on the Hu–Washizu functional*. Advanced Modeling and Simulation in Engineering Sciences 7(1) (2020).
- [187] WAGNER, W. / WRIGGERS, P.: *A simple method for the calculation of postcritical branches*. Engineering Computations 5(2) (1988), S. 103–109.
- [188] WANG, K. / SUN, W.: *A multiscale multi-permeability poroplasticity model linked by recursive homogenizations and deep learning*. Computer Methods in Applied Mechanics and Engineering 334 (2018), S. 337–380.
- [189] WEBER, P. / GEIGER, J. / WAGNER, W.: *Constrained neural network training and its application to hyperelastic material modeling*. Computational Mechanics 68(5) (2021), S. 1179–1204.
- [190] WEBER, P. / WAGNER, W. / FREITAG, S.: *Physically enhanced training for modeling rate-independent plasticity with feedforward neural networks*. Computational Mechanics, 2023.
- [191] WERBOS, P.J.: *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Dissertation, Harvard University, 1974.
- [192] WERBOS, P.J.: *Generalization of backpropagation with application to a recurrent gas market model*. Neural Networks 1(4) (1988), S. 339–356.
- [193] WEX, C. / ARNDT, S. / STOLL, A. / BRUNS, C. / KUPRIYANOVA, Y.: *Isotropic incompressible hyperelastic models for modelling the mechanical behaviour of biological tissues: a review*. Biomedical Engineering / Biomedizinische Technik 60(6) (2015).
- [194] WIDROW, B. / HOFF, M.E.: *Adaptive switching circuits*. Stanford Univ Ca Stanford Electronics Labs, 1960.

-
- [195] WILLIAMS, R.J. / ZIPSER, D.: *A Learning Algorithm for Continually Running Fully Recurrent Neural Networks*. *Neural Computation* 1(2) (1989), S. 270–280.
- [196] WOLFE, P.: *Convergence Conditions for Ascent Methods*. *SIAM Review* 11(2) (1969), S. 226–235.
- [197] WOLFE, P.: *Convergence Conditions for Ascent Methods. II: Some Corrections*. *SIAM Review* 13(2) (1971), S. 185–188.
- [198] WRIGGERS, P.: *Nonlinear Finite Element Methods*. Springer Berlin / Heidelberg, 2008.
- [199] WU, L. / NGUYEN, V.D. / KILINGAR, N.G. / NOELS, L.: *A recurrent neural network-accelerated multi-scale model for elasto-plastic heterogeneous materials subjected to random cyclic and non-proportional loading paths*. *Computer Methods in Applied Mechanics and Engineering* 369 (2020), S. 113234.
- [200] XU, K. / HUANG, D.Z. / DARVE, E.: *Learning constitutive relations using symmetric positive definite neural networks*. *Journal of Computational Physics* 428 (2021), S. 110072.
- [201] YUN, G.J. / GHABOUSSI, J. / ELNASHAI, A.S.: *A new neural network-based model for hysteretic behavior of materials*. *International Journal for Numerical Methods in Engineering* 73(4) (2008), S. 447–469.
- [202] ZHOU, D. / KAMLAH, M. / MUNZ, D.: *Effects of Bias Electric Fields on the Non-linear Ferroelastic Behavior of Soft Lead Zirconate Titanate Piezoceramics*. *Journal of the American Ceramic Society* 88(4) (2005), S. 867–874.
- [203] ZHOU, D. / KAMLAH, M. / MUNZ, D.: *Effects of uniaxial prestress on the ferroelectric hysteretic response of soft PZT*. *Journal of the European Ceramic Society* 25(4) (2005), S. 425–432.
- [204] ZIENKIEWICZ, O.C. / CHEUNG, Y.K.: *The Finite Element Method in Structural and Continuum Mechanics: Numerical Solution of Problems in Structural and Continuum Mechanics*. McGraw-Hill, 1967.
- [205] ZIENKIEWICZ, O.C. / TAYLOR, R.L.: *The Finite Element Method: Solid Mechanics*. Butterworth-Heinemann, 2000.
- [206] ZIENKIEWICZ, O.C. / TAYLOR, R.L.: *The Finite Element Method: Its Basis and Fundamentals*. Butterworth-Heinemann, 2013.
- [207] ZOHDI, T.I. / WRIGGERS, P.: *An Introduction to Computational Micromechanics*. Springer Berlin / Heidelberg, 2005.

Berichte des Instituts für Baustatik, Karlsruher Institut für Technologie

Verzeichnis der bisher erschienenen Berichte:

- 51/1: Beitrag zur Ermittlung der zeitabhängigen Eigenspannungen in Verbundträgern, Dissertation, Hansjürgen Sonntag, 1951.
- 54/2: Beiträge zur Ermittlung der zeitabhängigen Spannungen in Verbundträgern, Dissertation, Eugen Müller, 1954.
- 54/3: Zur Statik räumlich gekrümmter Träger, Dissertation, Heinrich Bechert, 1954.
- 55/4: Bemessungsverfahren für Verbundträger, Dissertation, Günter Utescher, 1955.
- 57/5: Einflußfelder für die Hauptträgerschnittkräfte zweistegiger Plattenbalkensysteme, Dissertation, Otto Köller, 1957.
- 60/6: Beitrag zur numerischen Berechnung der Schnittkräfte von rechteckigen und schiefen randversteiften Platten, Dissertation, Klaus Stiglat, 1960.
- 61/7: Allgemeines und einheitliches Berechnungsverfahren für Verbundkonstruktionen aus Stahl und Beton, Dissertation, Herbert Wippel, 1961.
- 62/8: Beitrag zur Bemessung vorgespannter stählerner Fachwerkträger mit besonderer Berücksichtigung der Tragsicherheit, Dissertation, Wolfgang Heidecker, 1962.
- 63/9: Beitrag zur Berechnung der Schnittkräfte und Durchbiegungen von umfanggelagerten Rechteckplatten mit veränderlichem Querschnitt in einer Spannrichtung, Dissertation, Ernst Buchholz, 1963.
- 63/10: Der winkelrecht zu seiner Ebene belastete, kreisförmig gekrümmte Träger mit starrer Torsionseinspannung an den Auflagerpunkten, Dissertation, Hans Wittfoht, 1963.
- 64/11: Ein Verfahren zur Berechnung biegsamer rechteckiger Gründungsplatten mit elektronischen Rechengeralten, Dissertation, Artur Deninger, 1964.
- 64/12: Anwendung des Momentenausgleichsverfahrens für durchlaufende schiefwinklige Platten, Dissertation, Jahidul Alum, 1964.

- 67/13: Beitrag zur Berechnung des Tragvermögens von Rahmentragwerken aus Stahl nach der Plastizitätstheorie II. Ordnung, Dissertation, Mieng-Siong Wang, 1967.
- 67/14: Zusammenwirken von biegesteifem Skelett und aussteifenden Wandscheiben in Hochhauskonstruktionen, Dissertation, Rolf Zickendraht, 1967.
- 68/15: Experimentelle Untersuchung von beliebig gelagerten Plattentragwerken mit Hilfe von elektrischen Analogieschaltungen, Dissertation, Werner Sitzler, 1968.
- 69/16: Operatorenkalkül zur Berechnung von Stab- und Flächentragwerken, Dissertation, Günter Eisenbiegler, 1969.
- 72/17: Umlagerungen der Schnittkräfte und Verformungen von biegebeanspruchten Stahlbetonkonstruktionen im Zustand II infolge Kriechens und Schwindens, Dissertation, Walter H. Schneider, 1972.
- 73/18: Berechnungssystematik für Rechteckscheiben, Habilitation, Günter Eisenbiegler, 1973.
- 1/75 : Berechnung einfeldriger und durchlaufender Kastenträger unter Berücksichtigung der Querschnittsdeformation, Dissertation, Jürgen Hanisch, 1975.
- 2/75 : Der Plattenbalken mit Lastangriff innerhalb des Steges, Dissertation, Hans-Erich Heining, 1975.
- 3/76 : Das Tragverhalten längsversteifter, vorverformter Rechteckplatten unter Axialbelastung nach der nichtlinearen Beultheorie (unter Berücksichtigung der orthotropen Struktur sowie einer elastischen Lagerung am Längsrand), Habilitation, Helmut Rubin, 1976.
- 4/78 : Traglastuntersuchungen ebener Rahmen aus Stahl unter Berücksichtigung teilplastischer Zonen, Dissertation, Ülgen Korkut, 1978.
- 5/79 : Traglastermittlung von räumlich belasteten Durchlaufträgern mit offenem, dünnwandigem Querschnitt bei beliebigem Werkstoffgesetz, Dissertation, Wolfgang Heil, 1979.
- 6/79 : Schnittgrößen im Verschneidungsbereich eines kreiszylindrischen Druckbehälters mit einem Stutzen beliebiger Länge, Dissertation, Kurt Schumer, 1979.
- 7/81 : Traglastberechnung räumlicher Rahmen aus Stahl- oder Leichtmetallprofilen mit dünnwandigen offenen Querschnitten, Dissertation, Theodor Ackermann, 1981.
- 8/83 : Berechnung orthotroper Rechteckplatten nach der nichtlinearen Elastizitätstheorie für beliebige Randbedingungen, Dissertation, Karl Morgen, 1983.

- 9/86 : Traglastermittlung räumlicher Stabtragwerke aus Stahl und Leichtmetall unter Berücksichtigung der Schubweichheit, Dissertation, Dietmar H. Maier, 1986.
- 10/89 : Bemessung von Raumfachwerken bezüglich ihrer Grenztragfähigkeit mit Hilfe der mathematischen Optimierung, Dissertation, Richard Harich, 1989.
- 11/89 : Quasi-Newton Verfahren und Kurvenverfolgungsalgorithmen für die Lösung nichtlinearer Gleichungssysteme in der Strukturmechanik, Forschungsbericht, Karl Schweizerhof, 1989.
- 12/90 : Finite Element Konzepte für schubelastische Schalen mit endlichen Drehungen, Dissertation, Heinrich Gebhardt, 1990.
- 13/91 : Theorie II. Ordnung des eben gekrümmten Stabes bei nicht richtungstreuen Kräften, Dissertation, Peter Brković, 1991.
- 14/91 : Nichtlineare Finite Element Berechnungen von Schalentragwerken mit geschichtetem anisotropen Querschnitt, Dissertation, Rudolf Klarmann, 1991.
- 15/94 : Adaptive Finite Element Konzepte zur Analyse von Schalentragwerken, Dissertation, Markus Baumann, 1994.
- 1 (1995) : Optimierung von Stockwerkrahmen aus Stahl mit biegesteifen und biegeweichen Rahmenknoten, Dissertation, Petra Höß, 1994, ISBN: 978-3-935322-00-3.
- 2 (1996) : Theorie und Numerik dünnwandiger Faserverbundstrukturen, Forschungsbericht, Friedrich Gruttmann, 1996, ISBN: 978-3-935322-01-1.
- 3 (1997) : Stabilität von Stahlzylindern mit unilateralen Randbedingungen bei statischen und dynamischen Beanspruchungen, Forschungsbericht, Klaus Knebel, 1997, ISBN: 978-3-935322-02-7.
- 4 (1998) : Eine einheitliche Finite-Element-Formulierung für Stab- und Schalentragwerke mit endlichen Rotationen, Dissertation, Roland Sauer, 1998, ISBN: 978-3-935322-03-4.
- 5 (1999) : Finite-Element-Modellierung von Starrkörper- und Übergangsbedingungen in der Statik, Dissertation, Joachim Kugler, 1999, ISBN: 978-3-935322-04-1.
- 6 (2000) : Delaminationsanalysen von Faserverbundstrukturen mit der Methode der finiten Elemente, Dissertation, Wolfgang Sprenger, 2000, ISBN: 978-3-935322-05-8.
- 7 (2000) : Theorie und Numerik eines Volumen-Schalen-Elementes bei finiten elastischen und plastischen Verzerrungen, Dissertation, Sven Klinkel, 2000, ISBN: 978-3-935322-06-5.

- 8 (2001) : Theorie und Numerik für den Entwurf von vorgespannten Membrantragwerken, Dissertation, René Ziegler, 2001, ISBN: 978-3-935322-07-2.
- 9 (2005) : Ein inelastisches 3D-Versagensmodell für Beton und seine Finite-Element-Implementierung, Dissertation, Jan Schütt, 2005, ISBN: 978-3-935322-08-9.
- 10 (2006) : Formfindung der Fließliniengeometrie für polygonale Platten im Traglastzustand, Dissertation, Jochen Wüst, 2006, ISBN: 978-3-935322-09-6.
- 11 (2006) : Nichtlineare Formulierung piezoelektrischer 3D-Stabstrukturen - Theorie und Finite-Element-Modellierung, Dissertation, Alexander Butz, 2006, ISBN: 978-3-935322-10-2.
- 12 (2007) : Nichtlineare Modellierung ferroelektrischer Keramiken und piezoelektrischer Strukturen – Analyse und Finite-Element-Formulierung, Habilitationsschrift, Sven Klinkel, 2007, ISBN: 978-3-935322-11-9.
- 13 (2007) : Ein geometrisch und materiell nichtlineares Cosserat-Modell – Theorie, Numerik und Anwendungsmöglichkeiten, Dissertation, Ingo Münch, 2007, ISBN: 978-3-935322-12-6.
- 14 (2008) : Magnetostriktive und piezoelektrische Materialien – Konstitutive Modellierung und Finite-Element-Formulierung, Dissertation, Konrad Linnemann, 2008, ISBN: 978-3-935322-13-3.
- 15 (2008) : Theorie und Numerik unidirektional verstärkter Faserverbundwerkstoffe: 3D Finite-Elemente-Untersuchungen der Faser-Matrix Mikroinstabilitäten, Dissertation, Jens Harich, 2008, ISBN: 978-3-935322-15-7.
- 16 (2009) : Finite Element Modeling of Intra- and Interlaminar Damage Growth in Composite Laminates, Dissertation, Claudio Balzani, 2009, ISBN: 978-3-935322-16-4.
- 17 (2010) : Theorie und Finite-Elemente-Modellierung geometrisch und materiell nichtlinearer piezoelektrischer Schalenstrukturen, Dissertation, Katrin Schulz, 2010, ISBN: 978-3-935322-17-1.
- 18 (2011) : Finite-Elemente-Formulierung mit abgestimmten Approximationsräumen für die Modellierung piezoelektrischer Stab- und Schalenstrukturen, Dissertation, Dieter Legner, 2011, ISBN: 978-3-935322-18-8.
- 19 (2014) : Nichtlineare elektromechanisch gekoppelte Viskoelastizität bei großen Verformungen am Beispiel dielektrischer Elastomere – Konstitutive Modellierung und Finite-Elemente-Formulierung, Dissertation, Alexander Büschel, 2014, ISBN: 978-3-935322-19-5.

- 20 (2014) : Duroplastische Kunststoffe mit Elastomerzusätzen - Konstitutive Modellierung und Finite-Elemente-Formulierung, Dissertation, Stefan Lacher, 2014, ISBN: 978-3-935322-20-1.
- 21 (2014) : Numerische Stabilitätsuntersuchung dünnwandiger Tragwerke unter dem Einfluss stochastischer Imperfektionen, Dissertation, Stefan Lauterbach, 2014, ISBN: 978-3-935322-21-8.
- 22 (2017) : Entwicklung robuster Ein- und Mehrskalen-Modelle zur Simulation von Delaminationsprozessen faserverstärkter Laminats, Dissertation, Tillmann Herwig, 2017, ISBN: 978-3-935322-22-5.
- 23 (2017) : Entwurf und Finite-Elemente-Modellierung ferroelektrischer Nanostrukturen zur Transformation mechanischer Deformationen in elektrische Energie, Dissertation, Matthias Krauß, 2017, ISBN: 978-3-935322-23-2.
- 24 (2018) : Entwicklung eines gradientenbasierten Modells zur netzunabhängigen Schädigungssimulation geschichteter Faserverbundstrukturen, Dissertation, Jonas Läufer, 2018, ISBN: 978-3-935322-24-9.
- 25 (2018) : Ein gemischtes FE-Modell zur Untersuchung des nichtlinearen Deformations- und Spannungsverhaltens schubweicher Laminats, Dissertation, Max Köpple, 2018, ISBN: 978-3-935322-25-6.
- 26 (2020) : Metaheuristische Form- und Systemoptimierung von Stabtragwerken mit Evolutionsstrategien, Dissertation, Alexander Keller, 2020, ISBN: 978-3-935322-26-3.
- 27 (2020) : Polymorphe Unschärfemodellierung in der nichtlinearen Strukturmechanik – Stabilität von Schalentragwerken, räumliche Variabilität und Metamodellierung, Dissertation, Marc Fina, 2020, ISBN: 978-3-935322-27-0.

Wissenschaftlicher Lebenslauf

Persönliche Daten

Name	Patrick Thomas Weber
Geburtsjahr	1994
Geburtsort	Saarlouis

Ausbildung

2004–2012	Schulbildung Peter-Wust-Gymnasium, Merzig Abitur
2012–2015	Bachelorstudium Bauingenieurwesen Karlsruher Institut für Technologie (KIT) Bachelorarbeit: Institut für Baustatik
2015	Vollzeit-Praktikum als Tragwerksplaner RSP Remmel + Sattler Ingenieurgesellschaft mbH, Frankfurt a. M.
2016–2018	Masterstudium Funktionaler und Konstruktiver Ingenieurbau – Engineering Structures Karlsruher Institut für Technologie (KIT) Profil Modellierung und Simulation im Ingenieurbau Masterarbeit: Institut für Baustatik
2018–2024	Wissenschaftlicher Mitarbeiter Karlsruher Institut für Technologie (KIT) Institut für Baustatik

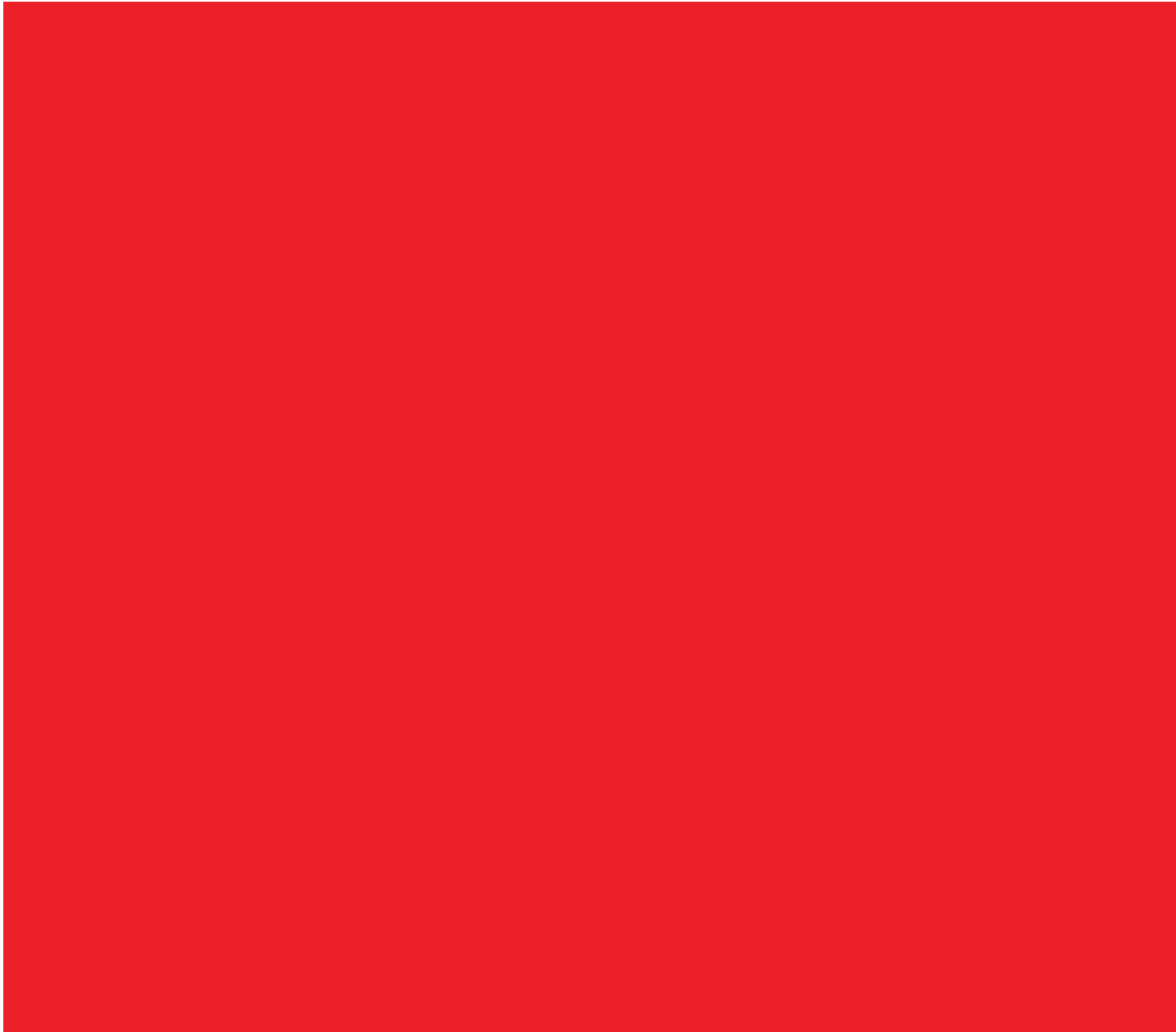
Preise und Auszeichnungen

2015	Bilfinger SE 1. Preis für die Bachelorarbeit
2013-2018	Deutschlandstipendium des KIT
2018	Ludwig-Lenz-Züblin-Preis für den Masterabschluss

Veröffentlichungen

- Zeitschriftenartikel (Peer Review)
 - **Weber, P.**, Wagner, W., Freitag S.: Physically enhanced training for modeling rate-independent plasticity with feedforward neural networks, *Computational Mechanics*, 72, S. 827-857, 2023, <https://doi.org/10.1007/s00466-023-02316-9>
 - **Weber, P.**, Geiger, J., Wagner, W.: Constrained neural network training and its application to hyperelastic material modeling, *Computational Mechanics*, 68, S. 1179–1204, 2021, <https://doi.org/10.1007/s00466-021-02064-8>
 - Fina, M., Panther, L., **Weber, P.**, Wagner, W.: Shell buckling with polymorphic uncertain surface imperfections and sensitivity analysis, *ASME J. Risk Uncertainty Part B*. 7(2), 020909, 2021, <https://doi.org/10.1115/1.4050165>
 - Fina, M., **Weber, P.**, Wagner, W.: Polymorphic uncertainty modeling for the simulation of geometric imperfections in probabilistic design of cylindrical shells, *Structural Safety* (82), 101894, 2020, <https://doi.org/10.1016/j.strusafe.2019.101894>
- Konferenzbeiträge (Artikel)
 - **Weber, P.**, Wagner, W., Freitag S.: Daten-basierte Materialmodellierung und Simulation von Holztragwerken mit künstlichen neuronalen Netzen, 15. Fachtagung Baustatik-Baupraxis (im Druck), 04.-05. März 2024, Hamburg, Deutschland
 - Fina, M., **Weber, P.**, Wagner, W.: Modeling of aleatory and epistemic uncertainties in probabilistic design of cylindrical shells, in M. Beer, E. Zio (Eds.): *Proceedings of the 29th European Safety and Reliability Conference (ESREL)*, S. 2243-2250, 22.-26. September 2019, Hannover, Deutschland, <http://rpsonline.com.sg/proceedings/9789811127243/html/0143.xml>
 - **Weber, P.**, Fina, M., Wagner, W.: Time domain simulation of earthquake excited buildings using a fuzzy stochastic approach, in M. Beer, E. Zio (Eds.): *Proceedings of the 29th European Safety and Reliability Conference (ESREL)*, S. 2127-2134, 22.-26. September 2019, Hannover, Deutschland, <http://rpsonline.com.sg/proceedings/9789811127243/html/0155.xml>
 - **Weber, P.**, Fina, M., Wagner, W.: Zur Fuzzy-stochastischen Analyse in der Strukturdynamik mittels Metamodellen, in J. Schneider, N. Kiziltoprak (Hrsg.) *Forschungskolloquium 2018 Grasellenbach Baustatik-Baupraxis e.V.*, S. 91-93, Springer Vieweg, 03.-06. Oktober 2018, Grasellenbach, Deutschland, <https://doi.org/10.1007/978-3-658-23627-4>.

- Weitere Konferenzbeiträge
 - **Weber, P.**, Wagner, W., Freitag, S.: Energy conservation for ANN material models: Comparing a weak enforcement with the penalty method to a strong enforcement utilizing a specific data loss function, 9th GAMM AG Data Workshop, 20.-21. März 2023, Stuttgart, Deutschland, <https://www.mib.uni-stuttgart.de/dae/ag-data/>
 - **Weber, P.**, Wagner, W., Freitag, S.: Considering physical restrictions in the training process of feedforward neural networks for modeling rate-independent plasticity, 5th African Conference on Computational Mechanics, 02.-04. November 2022, Kapstadt, Südafrika
 - **Weber, P.**, Wagner, W., Freitag, S.: Enhancing Data-based ANN Material Modeling with Physical Restrictions using Constraint Optimization Techniques, 9th GACM Colloquium on Computational Mechanics for Young Scientists from Academia and Industry, 21.-23. September 2022, Essen, Deutschland
 - **Weber, P.**, Wagner, W., Freitag, S.: Zur Berücksichtigung physikalischer Nebenbedingungen bei der Materialmodellierung mit künstlichen neuronalen Netzen, Forschungskolloquium FE ohne Schnee, 16.-19. Juli 2022, Hirschegg, Österreich
 - **Weber, P.**, Geiger, J., Wagner, W., Freitag, S.: Incorporating physical restrictions by constraint optimization techniques for ANN-based hyperelastic material modeling, 7th GAMM AG Data Workshop, 15. Dezember 2021, online, <https://www.mib.uni-stuttgart.de/dae/ag-data/>
 - **Weber, P.**, Fina, M., Wagner, W.: Polymorphic uncertainty in time-domain dynamic analysis of buildings under earthquake excitation using metamodels, 3rd International Conference on Uncertainty Quantification in Computational Sciences and Engineering, 24.-26. Juni 2019, Kreta, Griechenland, <https://2019.uncecomp.org/proceedings/pdf/18456.pdf>



ISBN 978-3-935322-28-7