Mathematical Optimization of Industrial Sheet Metal Forming Processes

Von der Fakultät für Elektrotechnik der
Helmut-Schmidt-Universität / Universität der Bundeswehr Hamburg
zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs genehmigte DISSERTATION
vorgelegt von **Marco Rozgić** aus Herdecke
Hamburg 2018

Vorsitzender:	Univ. Prof. Dr. -Ing. Joachim Horn
Referenten:	Univ. Prof. Dr. rer. nat habil. Marcus Stiemer
	Univ. Prof. Dr. -Ing. Stefanie Reese
Tag der mündlichen Prüfung:	24.04.2018

Marco Rozgič

Lebenslauf, Düsseldorf im August 2018

Berufliche Laufbahn

seit 01/2018 **Data Scientist**, ***Wurm GmbH & Co. KG Elektronische Systeme***.

01/2012 – 12/2017 **Wissenschaftlicher Mitarbeiter**, ***Lehrstuhl für theoretische Elektrotechnik und numerische Feldsimulation***, Helmut-Schmidt-Universität der Bundeswehr Hamburg.

Identifikation optimaler Prozessparameter bei der gekoppelten, elektromagnetischen, quasi-statischen Blechumformung im Rahmen des DFG Projekts PAK343.
Weitere Projekte:

- Entwicklung und Implementierung einer effizienten und schnellen FEM in MATLAB.
- Simulation und statistische Auswertung elektromagnetischer Felddaten für EMV Messungen mittels Modenverwirbelungskammern.
- Anwendung maschineller Lernverfahren zur optimalen, sensorgestützten Steuerung von Industrieanlagen.
- Betreuung von Bachelor-, Studien- und Masterarbeiten.

02/2010 – 12/2011 **Wissenschaftlicher Mitarbeiter**, ***ICAMS Interdisciplinary Centre for Advanced Materials Simulation***, Ruhr-Universität Bochum.

Atomistische Simulationen (MD) zum Studium der kinetischen Eigenschaften bei fest-fest Phasenübergängen in Stählen.

09/2009 – 01/2010 **Wissenschaftlicher Mitarbeiter**, ***Lehstuhl X für wissenschaftliches Rechnen***, Technische Universität Dortmund.

Implementation eines auf CMake basierenden Installationsprogramms für die Finite Elemente Software SOFAR.

Ausbildung

2002 – 2009 **Diplom in Mathematik**, *Technische Universität Dortmund*.

Die Studienschwerpunkte umfassten die nichtlineare Optimierung, das Gebiet der numerischen Mathematik und die komplexe Funktionentheorie. Als Nebenfach wurde Statistik belegt.

1992 – 2001 **Abitur**, *Geschwister-Scholl-Gymnasium Wetter (Ruhr)*.

Abiturfächer waren Mathematik, Informatik, Geschichte und Deutsch.

Diplomarbeit

Titel *Designoptimierung mechanischer Strukturen unter dynamischer Last*

Betreuer Prof. Dr. W. Achtziger

Inhalt Herleitung einer passenden Formulierung für die Topologieoptimierung mechanischer Strukturen unter zeitlich variierenden Lasten. Die aufgestellte Problemformulierung wurde für verschiedene Testszenarien gelöst. Weiter wurden Optimalitätskriterien der Lösungen nachgewiesen.

For the love of my life Julia,
for my parents Emmi and Stanimir,
for my brother Ivo and his family,
and for metal, of course.

Acknowledgment

Before turning to the formal part I get the chance to address some personal words to some of the many people who helped me and were there along the way. There are the members of the research group of PAK 343, in which I was involved. Thanks to the constructive efforts of Koray Demir from the IFUM at TU Dortmund the coupled forming process, which is theoretically discussed here, became something real, something touchable. The results delivered by Marcus Engelhardt from the IW at Leibniz University in Hannover provided the in depth understanding of the material's behavior. And in particular I have to thank Yalin Kiliclar from the IFAM at RWTH Aachen, who provided the code and model for the simulations. Without his tireless efforts we still would not be able to perform these multi-physical simulations in such a comfortable manor.

Next, there are the members of the department of the Theory of Electrical Engineering at the Helmut Schmidt University in Hamburg I have to thank. Each one of them helped me in their very own way, but some individuals deserve a special mention: Thank you Michael for always pushing me over my boundaries and never letting a weak argument stand. Thank you Inés for showing me practical solutions, which I often oversaw. And last but not least, thank you Manuel for teaching me thoroughness. My students taught me patience, you guys were fun, at least most of the times.

Most of all I have to thank my advisor Marcus Stiemer. In Germany the word "Doktorvater" is used for the word advisor, literally this translates to "doctoral father", which is quite stronger. But, that Marcus was (or is) to me, indeed. He taught me to be curious and open towards new things. He showed me how tough problems can only be solved with a relaxed state of mind. He was always there to listen to the wildest

theories and always saw something good and useful in it. I will forever be grateful that he invited me into his group. and showed me how fulfilling science can be. Thank you Marcus!

Needless to say that such a work can not be done without the necessary support from all the people outside of the scientific world, be it friends, family or my wonderful wife. Thank you all for being there when I wasn't and thank you all for carrying parts of my weight for me.

Abstract

In this work two methods for virtual process design in the context of coupled quasi-static electromagnetic impulse sheet metal forming applications are introduced. The first half of this thesis is concerned with solving the two main parameter identification problems arising when simulating and designing such a forming process: The identification of the employed material model and the identification of optimal process parameters. For both problems a similar black box framework based on LS-DYNA FEM/BEM simulations and IPOPT, an implementation of the inner point method is used.

A thermodynamically consistent material model developed by Vladimirov et al. (2014), which is tailored for the efficient numerical treatment of anisotropic hyperelastic-plastic materials is considered. This model is capable of representing rate dependent viscoplastic as well as rate independent elastoplastic materials. To account for material failure it is coupled to a scalar Lemaitre-type damage model. A framework for the reliable identification of the model's parameters for the simulation of the behavior of the aluminum alloy EN AW-5083 at various strain rates based on uniaxial tensile tests and a corresponding LS-DYNA simulation coupled to IPOPT is introduced. For the comparison of data obtained by simulation and data coming from experiments a distance measure based on the L^2-distance of functions is used. To this end an Akima spline developed by Akima (1970) is adapted to the experimental data. These special splines have the advantage of producing smooth interpolates, that are just as smooth as the according experimental curves, as long as no damage occurs. It is hence possible to interpolate the experimental data at every point needed and thus compare it efficiently to the data from simulation. It will be shown how a set of parameters suitable for the modeling of the high-speed situation can be derived from data obtained by

quasi-static experiments via a linear extrapolation of the damage threshold parameter. The identified parameter set is suitable to represent a coupled forming process very precisely, as is shown in Kiliclar et al. (2016).

The developed method for material parameter identification is adapted to perform a process parameter identification in a coupled sheet metal forming process. The task is to find a set of parameters describing a double exponential current pulse yielding a sharp cup radius. Double exponential pulses are used as a prototype for general mono-directional pulses. The material's behavior is simulated by the previously identified material model, thus constraints for the damage can be employed efficiently. In contrast to the usual optimal shape determination, an objective function completely independent of the geometry parametrization is derived. It will turn out that a maximization of the major strain in the area of question can be traced back to a lower cup radius, thus yielding a sharper form. The subsequent application of electromagnetic forming reduces the drawing radius by roughly 5 mm, compared to quasi-static deep drawing alone. The presented method can be seen as an extension to previous approaches by Taebi et al. (2012), since it uses a unified simulation framework and exploits a material model that incorporates damage.

The second part of this work is concerned with treating the identification of process parameters as a PDE constrained optimization problems and deriving a fitting SAND formulation. The solution of the PDE, modeling the physical behavior, is seen as a state, and the parameters of the process are seen as controls. The first discretize, then optimize approach, is pursued in this work. Regularization methods keeping the controls in check can be avoided in this particular case, since the process parameters are already represented as a real valued finite dimensional vector. A class of problems mimicking a sheet metal forming problem in the linear elastic regime is introduced. This simplification is picked, because the required FEM can easily be implemented but the insights gained are useful for the development of a general optimization scheme for those types of technologically relevant problems. In the course of this work an efficient FEM solver for MATLAB, *o*FEM, was developed by Dudzinski et al. (2016) to provide the necessary discretizations of the PDEs. With *o*FEM and IPOPT the family of problems is thoroughly numerically analyzed. The derived SAND formulation is compared to a black box optimization problem based on earlier works by Taebi et al. (2012) and Rozgič et al. (2012). Compared to this so called NAND approach the SAND approach shows a potential gain in efficiency by two orders of magnitude.

Zusammenfassung

In dieser Arbeit werden zwei Methoden für die virtuelle Prozessauslegung im Kontext der gekoppelten quasi-statischen elektromagnetischen Impulsblechumformung eingeführt. Die erste Hälfte dieser Arbeit beschäftigt sich mit der Lösung zweier Identifikationsprobleme, wie sie hauptsächlich bei der Simulation und Gestaltung eines solchen Umformprozesses entstehen: die Identifikation des eingesetzten Materialmodells und die Identifizierung optimaler Prozessparameter. Bei beiden Problemen wurde ein ähnliches Verfahren auf Basis von LS-DYNA FEM/BEM Simulationen und IPOPT, eine Implementierung der inneren Punktmethode, verwendet.

Ein thermodynamisch konsistentes Materialmodell, das von Vladimirov et al. (2014) entwickelt wurde und für die effiziente numerische Behandlung von anisotropen hyperelastisch-plastischen Materialien besonders geeignet ist, wird verwendet. Dieses Modell ist dazu in der Lage, ratenabhängige viskoplastische sowie ratenunabhängige elastoplastische Materialien abzubilden. Um Materialversagen zu berücksichtigen, ist es an ein skalares, lemaitreartiges Schadensmodell gekoppelt. Mit diesem Modell soll das Verhalten der Aluminiumlegierung EN AW-5083 bei verschiedenen Dehnraten simuliert werden. Dazu müßen entsprechende Modellparameter identifiziert werden. Diese geschieht in dieser Arbeit durch die systematische Anpassung einer LS-DYNA Simulation an einen entsprechenden uniaxial Zugversuch mittels IPOPT. Um die durch Simulation gewonnen Daten möglichst exakt mit Daten aus dem Experiment vergleichen zu können, wird ein Abstandsmaß auf der Grundlage des L^2- Abstands von Funktionen verwendet. Zu diesem Zweck wird ein Akima-Spline, der von Akima (1970) entwickelt wurde, an die experimentellen Daten angepasst. Diese speziellen Splines haben den Vorteil, glatte Funktionen ohne Überschwinger zu erzeugen. Es ist daher

möglich, die experimentellen Daten an jedem gewünschten Punkt zu interpolieren und damit effizient mit den Daten aus der Simulation zu vergleichen. Es wird gezeigt, wie durch eine lineare Extrapolation des Schadensschwellenwerts ein Satz von Parametern, der für die Modellierung der Hochgeschwindigkeitssituation geeignet ist, aus Daten erzeugt wird, die aus quasi-statische Experimente stammen. Der identifizierte Parametersatz ist geeignet, um einen gekoppelten Umformprozess sehr genau darzustellen, wie in Kiliclar et al. (2016) gezeigt ist.

Das entwickelte Verfahren zur Materialparameteridentifikation wird so modifiziert, dass eine Prozessparameteridentifikation in einem gekoppelten Blechumformprozess durchgeführt werden kann. Die Aufgabe besteht darin, einen Satz von Parametern zu finden, die einen doppelten exponentiellen Stromimpuls beschreiben, sodass sich ein möglichst scharfer Becherradius ergibt. Doppelt Exponentialpulse werden dabei als Prototypen für allgemeine mono-direktionale Pulse verwendet. Das Verhalten des Materials wird durch das zuvor identifizierte Materialmodell simuliert, so kann die d die Materialschädigung effizient als Nebenbedingungen genutzt werden. Im Gegensatz zur üblichen optimalen Formbestimmung wird eine von der Geometrie Parametrisierung völlig unabhängige Zielfunktion abgeleitet. Es wird sich herausstellen, dass eine Maximierung der Hauptdehnung im fraglichen Bereich auf eine Minimierung des Blecherradius zurückgeführt werden kann, was eine schärfere Form ergibt. Durch die anschließende Anwendung der elektromagnetischen Formgebung konnte der Ziehradius, im Vergleich zum konventionellen Tiefziehen, um etwa 5 mm reduziert werden. Die vorgestellte Methode kann als eine Erweiterung zu früheren Ansätzen von Taebi et al. (2012) gesehen werden, da sie ein einheitliches Simulations-Framework verwendet und ein Materialmodell nutzt, das Schädigung direkt berücksichtigt.

Der zweite Teil dieser Arbeit beschäftigt sich mit der Behandlung der Identifikation von Prozessparametern im Sinne von PDE-beschränkten Optimierungsproblemen und der Herleitung einer passenden SAND-Formulierung. Die Lösung der PDE, die das physikalische Verhalten modelliert, wird dabei als Zustand, und die Prozessesparameter werden als Kontrollen betrachtet. Der „erst diskretisieren, dann optimieren “Ansatz, wird in dieser Arbeit verfolgt. Regularisierungsmethoden, die die Kontrollen in Schach halten, konnten in diesem speziellen Fall vermieden werden, da die Prozessparameter bereits als ein reellwertiger endlicher dimensionaler Vektor dargestellt sind. Eine Klasse von Problemen, die ein Blechumformungsproblem im linear elastischen Fall nachahmen, wird eingeführt. Diese Vereinfachung wird gewählt, da die zugrundeliegende FEM leicht umgesetzt werden kann, aber die gewonnenen Erkenntnisse für die Entwicklung eines allgemeinen Optimierungsschemas für solche technologisch relevanten

Probleme genutzt werden können. Im Laufe dieser Arbeit wurde ein effizienter FEM-Löser für MATLAB, *o*FEM, von Dudzinski et al. (2016) entwickelt, um die notwendigen Diskretisierungen der PDEs zu liefern. Mit *o*FEM und IPOPT wird die Familie von Problemen gründlich numerisch analysiert. Die abgeleitete SAND-Formulierung wird mit einer Black-Box-Methode verglichen, welche auf früheren Werken von Taebi et al. (2012) und Rozgič et al. (2012) basiert. Im Vergleich zu diesem so genannten NAND-Ansatz zeigt der SAND-Ansatz einen potenziellen Effizienzsteigerung um zwei Größenordnungen.

Contents

1 Motivation and Introduction

With growing computational power virtual process planing and design is more and more part of industrially relevant technology development. The key ingredients for successful virtual design are, on the one hand efficient and reliable simulation tools, and on the other, versatile and robust optimization strategies. In the past years the quality and capability of commercially available simulation tools grew in such a manner that their use in production cycles is unquestioned.

Thanks to constant development, tools, like the here used LS-DYNA, are able to solve multi physical problems. Those problems arise, when different physical regimes, usually expressed by partial differential equations, have to be coupled for the description of a single phenomenon. In this work equations describing the mechanical situation have to be coupled to the magneto quasi-static approximation of Maxwell's equation describing the occurring magnetic fields. With these a combined electromagnetic and classical quasi-static sheet metal forming process can be modeled.

By this combined process the effectivity and sustainability of conventional deep drawing can be increased. Electromagnetic forming allows for faster process times and, since it is a contact free method, reduces the amount of required lubricants. The strain rates occurring during electromagnetic forming are significantly higher than those at quasi-static forming. These high strain rates yield an increase of the material's formability, hence allowing for sharper drawing radii. However, the application of electromagnetic forming alone is inefficient, the overall process efficiency is reported to be around 2%. Therefore this combination offers an alternative, where the advantages of one process obliterate the drawbacks of the other.

Simulation techniques aiming to couple these two regimes and special material models to reproduce the changing material behavior at low and high strain rates are

needed. During the course of a long research project both, novel simulation methods based on ALE descriptions and effective material models, incorporating the strain rate change and reproducing the material's behavior at failure, were developed. This thesis is part of this research project and focuses on the optimization methods required for the incorporation of the developed methods into virtual process design.

In the first part of this work a methodology for the successful identification of a suitable material model will be introduced. It will be shown, how the results obtained by fitting the material model to a quasi-static uniaxial tensile test can easily be transferred to the high-speed situation for the aluminum alloy under consideration. The identified material model will be validated by suitable experiments. The developed method will be based on simulations by the commercial toolbox LS-DYNA and the optimization problem solver IPOPT, an implementation of an interior point algorithm. This special class of optimization algorithms allows for the effective treatment of non-linear problems with a huge number of non-linear constraints.

After the material model is identified, the developed method will be adapted, such that it can be used for process parameter optimization. In this case parameters influencing the coupled sheet metal forming process have to be found, yielding an as good as possible deep drawing result. A mono-directional current pulse will be adjusted, such that a sharp drawing radius of a cup is achieved. A novel approach, independent of geometry parameterizations, is introduced to compute a set of parameters yielding this sharp radius. Instead of minimizing distances between the edge radius and the die, the major stain is taken into account. It will turn out that a maximization of the major strain at the bottom of the cup will yield a sharp radius.

This unified approach to virtual process design beginning with the identification of a suitable material model can be used in industrial relevant applications, since it relies on those black boxes already in use for the numerical simulation of such applications. It will turn out that the main limiting factor is computational time and power.

The second part of this thesis aims for the development of a method for virtual design that remedies the previous limitations. An approach based on ideas from topology optimization is tailored, rendering the governing partial differential equations to constraints of an optimization problem. It will be shown how this idea can be used for process optimization and, based on a simple example, the potential benefits of the approach will be demonstrated.

As this method requires direct access to the discrete representations of the partial differential equations, i.e., the system matrices, it is not yet applicable in industrial relevant design, since these matrices are usually not accessible in commercially available

software tools. However, the results of this thesis show the benefits software toolkit developers can gain by directly incorporating the optimization procedures into their packages.

The work is structured as follows: Chapter 2 gives a brief introduction to the finite element method. There the relevant terminologies and concepts for the numerical treatment of partial differential equations are discussed.

The fundamentals of constraint optimization are covered in Chapter 3. There a basic interior point algorithm is formulated. At the end of Chapter 3 a first result of this thesis is presented: the equivalence of various formulations relevant for the treatment of optimization problems with interior points is shown.

Chapter 4 is dedicated to the solution of the virtual design task for a coupled electromagnetic quasi-static forming process. An industrial relevant method based on LS-DYNA and IPOPT for the effective treatment of the material model and process parameter identification is introduced. It is followed by a first synopsis in Chapter 5.

In Chapter 6 an alternative approach exploiting the idea of simultaneous analysis and design is developed. It is shown how the new formulation can be used for process parameter identification and how it scales compared to the previously introduced algorithm. The work closes with a conclusion in Chapter 7.

Each chapter holds an exhaustive review of the literature about the according topic, as each chapter is based on either textbooks or previous publications.

2 A Short Introduction into the Finite Element Method

In this work the finite element method will be the main tool to solve the partial differential equations arising in the mathematical description of forming processes. A huge number of textbooks on this subject is available, e.g., Braess (2013), Larson and Bengzon (2013), or Zienkiewicz et al. (2013). To create a common basis the finite element formulation will be derived in an abstract setting, allowing for a broad range of generalizations, which are used throughout this work. Readers familiar with vector spaces, variational formulations and the FEM in general can easily skip the upcoming sections.

2.1. Introduction

The finite element method (FEM) originated from the need to solve problems in structural, civil, and aerospace engineering, where complex mechanical scenarios arise. The early developments can be traced back to the works of Hrennikoff (1941) from civil engineering and Courant (1943), who set the mathematical foundation. These approaches differ in their perspective, but share the same fundamental characteristic of today's finite element analysis (FEA): A continuous domain is discretized into a mesh of discrete sub domains, usually called elements, and a solution is determined by piecewise approximations on those domains.

With the rise of computers the development of modern days FEA frameworks gained momentum in the 1960s and 1970s, fuelled by the works of Argyris and Kelsey (1960) and Clough (1958) : The first implementations like NASTRAN or SAP IV were released and became available to the scientific community. With growing computational power and rising interest in the theory behind FEM, algorithms to solve problems from

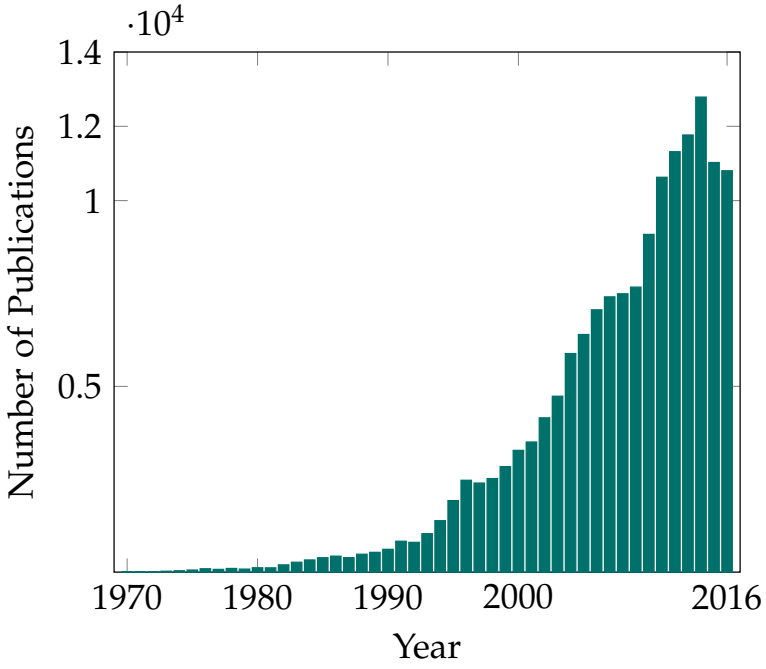


Figure 2.1.: The number of publications with the term 'Finite Element Simulation' in the title, abstract or keyword list from 1970 till 2016 according to the SCOPUS database.

structural mechanics, electromagnetics, fluid dynamics and heat and energy transfer, and so on, were developed.

Modern design cycles employ the FEM at any phase of the work flow to analyze and optimize processes. Todays toolboxes like ANSYS, LS-DYNA, or COMSOL, to name just a few, offer pre and post processing options, as well as a broad range of applications, including multiphysical problems. Non commercial tools like deal.II by Bangerth et al. (2007) or the FEniCS project by Alnæs et al. (2015) also offer rich functionalities and are exploited by scientists from different fields to tailor solutions to particular problems arising in their distinct fields. As can be seen in Fig. 2.1 the number of publications using or discussing the FEM has been growing exponentially since 1970 with a slight decay since 2015.

2.2. Vector Spaces, Hilbert Spaces, Banach Spaces and Sobolev Spaces

Vector spaces are a basic mathematical construct, therefore this topic will only be discussed briefly to lay the foundation of the upcoming sections. The field over which the vector space will be constructed is $\mathbb{R}$. All of the following can be done over $\mathbb{C}$ as well.

Definition 2.2.1 (Vector Space)
Let V be a nonempty set equipped with the operations $+ : V \times V \to V$ and $\cdot : \mathbb{R} \times V \to V$.

a) The set V is called *real valued vector space* over $\mathbb{R}$ equipped with the operations $+ : V \times V \to V$ and $\cdot : \mathbb{R} \times V \to V$ if for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ and all $\lambda, \mu \in \mathbb{R}$ the following holds:

$$\begin{aligned} \mathbf{0} &\in V, \\ (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} &= \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}), \\ \boldsymbol{u} + \boldsymbol{v} &= \boldsymbol{v} + \boldsymbol{u}, \\ \lambda \cdot (\boldsymbol{u} + \boldsymbol{v}) &= \lambda \cdot \boldsymbol{u} + \lambda \cdot \boldsymbol{v}, \\ (\lambda + \mu) \cdot \boldsymbol{v} &= \lambda \cdot \boldsymbol{v} + \mu \cdot \boldsymbol{v}. \end{aligned} \tag{2.1}$$

As usual the $\cdot$ symbol is mostly omitted if scalar multiplication is performed.

b) The mapping $\|.\| : V \to \mathbb{R}$ is called *norm* if it satisfies

$$\begin{aligned} \|\boldsymbol{u} + \boldsymbol{v}\| &= \|\boldsymbol{u}\| + \|\boldsymbol{v}\|, \\ \|\lambda \boldsymbol{v}\| &= |\lambda| \|\boldsymbol{v}\|, \\ \|\boldsymbol{v}\| &\geq 0, \\ \|\boldsymbol{v}\| &= 0 \Leftrightarrow \boldsymbol{v} = \mathbf{0} \end{aligned} \tag{2.2}$$

for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$.

c) The set of vectors $\{\boldsymbol{\phi}_i\}_{i=1}^n \subset V$ is called *basis* of V, if the set is minimal and if all $\boldsymbol{v} \in V$ can be written as

$$\boldsymbol{v} = \sum_{i=1}^{n} c_i \boldsymbol{\phi}_i,$$

with some $c_i \in \mathbb{R}$ depending on $\boldsymbol{v}$, and if from $\boldsymbol{v} = \boldsymbol{0}$ it follows $c_i = 0$, for all $i = 1, \ldots, n$ (linear independency). The number of basis vectors is called *dimension* of V and is not necessarily finite.

d) A *linear functional* or *linear form* is a mapping $l : V \to \mathbb{R}$, such that

$$\begin{aligned} l(\boldsymbol{u} + \boldsymbol{v}) &= l(\boldsymbol{u}) + l(\boldsymbol{v}), \\ l(\lambda \boldsymbol{v}) &= \lambda l(\boldsymbol{v}), \end{aligned} \tag{2.3}$$

for all $\boldsymbol{u}, \boldsymbol{v} \in V$ and $\lambda \in \mathbb{R}$. With

$$\|l\| = \sup_{\boldsymbol{v} \in V} \frac{|l(\boldsymbol{v})|}{\|\boldsymbol{v}\|}$$

the space of all linear functional, called *dual space* V^* of V, becomes a *normed vector space*.

e) A mapping $a : V \times V \to \mathbb{R}$ is a *bilinear form*, if

$$\begin{aligned} a(\boldsymbol{u} + \boldsymbol{v}, \boldsymbol{w}) &= a(\boldsymbol{u}, \boldsymbol{w}) + a(\boldsymbol{v}, \boldsymbol{w}), \\ a(\boldsymbol{u}, \boldsymbol{v} + \boldsymbol{w}) &= a(\boldsymbol{u}, \boldsymbol{v}) + a(\boldsymbol{u}, \boldsymbol{w}), \\ a(\lambda \boldsymbol{u}, \mu \boldsymbol{v}) &= \lambda \mu a(\boldsymbol{u}, \boldsymbol{v}), \end{aligned} \tag{2.4}$$

for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ and $\lambda, \mu \in \mathbb{R}$. The bilinear form a is called *symmetric*, if $a(\boldsymbol{u}, \boldsymbol{v}) = a(\boldsymbol{v}, \boldsymbol{u})$, for all $\boldsymbol{u}, \boldsymbol{v} \in V$.

f) Let a be a bilinear form. If for all $\boldsymbol{u} \in V$ the estimate $a(\boldsymbol{u}, \boldsymbol{u}) \geq 0$ holds, and if $a(\boldsymbol{u}, \boldsymbol{u}) = 0$ if and only if $\boldsymbol{u} = \boldsymbol{0}$, then a is called *inner product*. A shorthand notation for this case is $(\boldsymbol{u}, \boldsymbol{v})$.

g) The *dual pairing* $\langle , \rangle : V^* \times V \to \mathbb{R}$ is a bilinear mapping, such that for all $\boldsymbol{v}^* \in V^*$ there exists a $\boldsymbol{v} \in V$ with $\langle \boldsymbol{v}^*, \boldsymbol{v} \rangle \neq 0$ and that for all $\boldsymbol{v} \in V$ there exists a $\boldsymbol{v}^* \in V^*$ with $\langle \boldsymbol{v}^*, \boldsymbol{v} \rangle \neq 0$. The mapping $\langle , \rangle : V^* \times V \to \mathbb{R} : (\boldsymbol{v}^*, \boldsymbol{v}) \mapsto \boldsymbol{v}^*(\boldsymbol{v})$ is called *natural pairing*.

h) The *angle* between two vectors $\boldsymbol{u}, \boldsymbol{v} \in V$ is given as

$$\theta = \arccos \frac{(\boldsymbol{u}, \boldsymbol{v})}{\|\boldsymbol{u}\| \|\boldsymbol{v}\|}.$$

Two vectors are called *orthogonal* to each other, if and only if

$$(\boldsymbol{u}, \boldsymbol{v}) = 0.$$

▲

Remark 2.2.2

a) A mapping $|.| : V \to \mathbb{R}$ is called *semi norm* if it satisfies Eq. (2.2), except that $|v| = 0$ does not imply $v = \mathbf{0}$.

b) A linear form is called *continuous*, or *bounded*, if there exists a constant $C \in \mathbb{R}$, such that

$$|l(v)| \leq C \|v\|,$$

for all $v \in V$.

c) The inner product induces a norm by

$$\|v\|^2 = (v, v).$$

d) The *Cauchy-Schwarz inequality*

$$(u, v) \leq \|u\| \|v\|$$

holds in all vector spaces equipped with an inner product (*inner product spaces*). Further, the *parallelogram law* holds in all such spaces:

$$\|u + v\|^2 + \|u - v\|^2 = 2\|u\|^2 + 2\|v\|^2.$$

Finally, if two vectors are orthogonal, the *Pythagorean Theorem* holds:

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2. \qquad \blacktriangle$$

Next, two special types of vector spaces, Banach and Hilbert spaces, are defined.

Definition 2.2.3 (Banach Spaces and Hilbert Spaces)
Let V be a normed vector space with norm $\|.\|$.

a) A sequence $\{v_i\}_{i=1}^{\infty}$ of elements $v_i \in V$, for all i, is called *Cauchy sequence*, if for all $\varepsilon > 0$ there exists one $n \in \mathbb{N}$, such that

$$\|v_i - v_j\| \leq \varepsilon, \quad \text{for all } i, j \geq n.$$

b) The sequence $\{v_i\}_{i=1}^{\infty}$ of elements $v_i \in V$, for all i, is called *convergent*, if there exists a $v \in V$, such that for all $\varepsilon > 0$ there exists one $n \in \mathbb{N}$ with:

$$\|v - v_i\| \leq \varepsilon, \quad \text{for all } i \geq n.$$

c) The vector space V is called *complete* or a *Banach space*, if every Cauchy sequence is a convergent sequence. Note that only normed and complete spaces are called Banach spaces.

d) If the norm of the Banach space V is induced by an inner product, it is called *Hilbert space*.

e) Let V be a Banach space. A linear subspace $V_0 \subset V$ is called *closed* if every convergent sequence in V_0 has its limit in V_0. ▲

Remark 2.2.4

a) Every convergent sequence is a Cauchy sequence.

b) Let V be a Hilbert space with inner product $(.,.)$ and let V_0 be a closed subspace of V. The *orthogonal complement* $V_0^{\perp} = \{\boldsymbol{v} \in V : (\boldsymbol{v}, \boldsymbol{v}_0) = 0, \text{ for all } \boldsymbol{v}_0 \in V_0\}$ is also a closed subspace of V. Further, V can be decomposed as a direct sum, i. e., $V = V_0 \oplus V_0^{\perp}$. Consequently, for all $\boldsymbol{v} \in V$ there exists a $\boldsymbol{v}_0 \in V_0$, which is the best approximation of $\boldsymbol{v}$ with respect to the induced norm:

$$\|\boldsymbol{v} - \boldsymbol{v}_0\| \leq \|\boldsymbol{v} - \boldsymbol{u}_0\|, \quad \text{for all } \boldsymbol{u}_0 \in V_0. \qquad ▲$$

For the definition of Sobolev spaces, which are vital for the abstract discussion of the theory of finite elements, special vector spaces of integrable functions are needed. These are constructed over domains with a certain regularity.

Definition 2.2.5 (L^p-Spaces and Weak Derivatives)

a) Let $\Omega \subset \mathbb{R}^d$ be a domain, and let $B_r(\boldsymbol{x}) = \{\boldsymbol{u} \in \mathbb{R}^d : \|\boldsymbol{x} - \boldsymbol{u}\| < r\}$ be the ball of radius r around $\boldsymbol{x}$. The domain Ω has a *Lipschitz boundary*, or is called a *Lipschitz domain*, if for every $\boldsymbol{x} \in \partial\Omega$, there exist a radius $r > 0$ and a mapping $h : B_r(\boldsymbol{x}) \to B_1(\boldsymbol{0})$, such that:

   1) The mapping h is a bijection,

   2) both, h and h^{-1} are Lipschitz continuous,

   3) $h(\partial\Omega \cap B_r(\boldsymbol{x})) = \{(x_1, \ldots, x_d)^{\top} \in B_1(\boldsymbol{0}) : x_d = 0\}$,

   4) $h(\Omega \cap B_r(\boldsymbol{x})) = \{(x_1, \ldots, x_d)^{\top} \in B_1(\boldsymbol{0}) : x_d > 0\}$.

b) Let $\Omega \subset \mathbb{R}^d$ be a domain with Lipschitz boundary. The function space $L^p(\Omega)$ is given as

$$L^p(\Omega) = \left\{ v : \Omega \to \mathbb{R} : \|v\|_{L^p(\Omega)} \leq \infty \right\},$$

where the norm $\|.\|_{L^p}$ is given as

$$\|v\|_{L^p} = \left(\int_\Omega |v|^p \, \mathrm{d}x \right)^{1/p}, \quad \text{for } 1 \leq p < \infty,$$

and

$$\|v\|_{L^\infty} = \sup_{\boldsymbol{x} \in \Omega} |v(\boldsymbol{x})|.$$

The integral used here is the Lebesgue integral.

c) Let $\Omega \subset \mathbb{R}^d$ be a domain. The space of *locally integrable* functions $L^1_{\text{loc}}(\Omega)$ is given as

$$L^1_{\text{loc}}(\Omega) = \left\{ v : \, v \in L^1(K), \text{ for all } K \subset \Omega, \, K \text{ compact} \right\}.$$

d) Let $\Omega \subset \mathbb{R}^d$ be a domain. Let $u \in L^1_{\text{loc}}(\Omega)$ and let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^\top$, with $\alpha_j \in \mathbb{N} \cup \{0\}$ be a $d-$tuple. The function $v \in L^1_{\text{loc}}(\Omega)$ is called *weak derivative of degree* $\boldsymbol{\alpha}$ of u, if for all *test functions* $\Psi \in C_0^\infty$

$$\int_\Omega v(\boldsymbol{x}) \Psi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = (-1)^{|\boldsymbol{\alpha}|} \int_\Omega u(\boldsymbol{x}) \mathrm{D}^{\boldsymbol{\alpha}} \Psi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}, \tag{2.5}$$

holds, where:

$$|\boldsymbol{\alpha}| = \sum_{i=1}^d \alpha_i$$

and

$$\mathrm{D}^{\boldsymbol{\alpha}} \Psi(\boldsymbol{x}) = \prod_{i=1}^d \left(\frac{\partial}{\partial x_i} \right)^{\alpha_i} \Psi(\boldsymbol{x}).$$

The space C_0^∞ is the space of arbitrarilly often differentiable functions with compact *support*, i.e., the set where Ψ does not map to zero: $\text{supp}(\Psi) = \{\boldsymbol{x} \in \Omega : \, \Psi(\boldsymbol{x}) \neq 0\}$. ▲

Remark 2.2.6

The space $L^2(\Omega)$ is a Hilbert space with the inner product

$$(u, v)_{L^2(\Omega)} = \int_\Omega uv \, \mathrm{d}x.$$

For $p \neq 2$ the norm $\|.\|_{L^p}$ is not induced by an inner product, thus making $L^p(\Omega)$ only a Banach space. ▲

With these preparations the Sobolev spaces can be defined.

Definition 2.2.7 (Sobolev Space)
Let $u \in L^p(\Omega)$ for a Lipschitz domain $\Omega \subset \mathbb{R}^d$, and $1 \leq p \leq \infty$. Further, let for $k \in \mathbb{N}$ all weak derivatives $D^{\boldsymbol{\alpha}} u$ for $|\boldsymbol{\alpha}| \leq k$ exist.

a) The *Sobolev norm* of u is given by

$$\|u\|_{W_k^p(\Omega)} = \left(\sum_{|\boldsymbol{\alpha}| \leq k} \|D^{\boldsymbol{\alpha}} u\|_{L^p(\Omega)}^p \right)^{1/p}, \quad \text{for } 1 \leq p < \infty$$

and

$$\|u\|_{W_k^\infty(\Omega)} = \max_{|\alpha| \leq k} \|D^{\boldsymbol{\alpha}} u\|_{L^\infty(\Omega)}.$$

b) The *Sobolev space* $W_k^p(\Omega)$ is defined as the space of functions in $L^p(\Omega)$ whose weak derivatives of degree k are also in $L^p(\Omega)$, i.e.:

$$W_k^p(\Omega) = \left\{ u \in L^p(\Omega) : \|u\|_{W_k^p(\Omega)} < \infty \right\}.$$

c) The *Sobolev semi norm* is defined by:

$$|u|_{W_k^p(\Omega)} = \left(\sum_{|\alpha| = k} \|D^{\boldsymbol{\alpha}} u\|_{L^p(\Omega)}^p \right)^{1/p}, \quad \text{for } 1 \leq p < \infty$$

and

$$|u|_{W_k^\infty(\Omega)} = \max_{|\alpha| = k} \|D^{\boldsymbol{\alpha}} u\|_{L^\infty(\Omega)}.$$

d) For $p = 2$ the Sobolev space $W_k^2(\Omega)$ is a Hilbert space with inner product

$$(u, v)_{W_k^2(\Omega)} = \sum_{|\boldsymbol{\alpha}| \leq k} (D^{\boldsymbol{\alpha}} u, D^{\boldsymbol{\alpha}} v)_{L^2(\Omega)},$$

and is henceforth denoted by $H^k(\Omega)$. ▲

Remark 2.2.8
Sobolev functions are defined by the Lebesgue integral, therefore it is somewhat meaningless to speak about the restriction of a Sobolev function to the boundary of a domain (Lebesgue integrals are fixed up to a set of measure zero, and the boundary of a Lipschitz domain $\partial\Omega$ has exactly Lebesgue measure zero). However, with the *trace operator* this can still be achieved, by evaluating a fitting function in $L^p(\Omega)$ instead. To this end let $\Omega \subset \mathbb{R}^d$ be a Lipschitz domain. Then there exists a continuous linear operator $T : W_1^p(\Omega) \to L^q(\partial\Omega)$, such that:

a) $Tu = u|_{\partial\Omega}$, if $u \in C^\infty\left(\overline{\Omega}\right)$,

b) $\|Tu\|_{L^q(\partial\Omega)} \leq C\,\|u\|_{W_1^p(\Omega)}$, for all $u \in W_1^p\left(\Omega\right)$.

Here,

$$q \begin{cases} = \frac{(d-1)p}{(d-p)}, & \text{if } p < d, \\ < \infty, & \text{if } p = d, \\ = \infty, & \text{if } p > d, \end{cases}$$

and C is a constant depending on p, Ω and q, only. The proof for this statement, known as the *Trace Theorem*, can be found for example in Dobrowolski (2010), where also the construction of $C^\infty\left(\overline{\Omega}\right)$ by embedding is discussed.

With the trace operator the space of functions in $H^1\left(\Omega\right)$ which vanish at the boundary, in the sense of traces, can be defined as

$$H_0^1\left(\Omega\right) = \left\{v \in H^1\left(\Omega\right) :\ Tv = 0\right\},$$

with norm

$$\|u\|_{H_0^1}^2 = \|\nabla u\|_{L^2}^2 + \|u\|_{L^2}^2.$$

If it is clear over which domain the space is constructed the domain Ω is omitted. ▲

These preparations allow for establishing the variational problem.

2.3. Variational Problem

Definition 2.3.1 (Coercive and Continuous Bilinear Form)
Let V be a Hilbert space and let $a : V \times V \to \mathbb{R}$ be a bilinear form.

a) The bilinear form a is called *coercive (elliptic)*, if there exists a constant $m > 0$, such that for all $v \in V$

$$m\|v\|^2 \leq a\left(v, v\right).$$

b) The bilinear form a is called *continuous*, if there exists a constant C, such that for all $u, v \in V$

$$a\left(u, v\right) \leq C\|u\|\|v\|.$$ ▲

For a bilinear form a as in Definition 2.3.1 and a continuous linear form $l : V \to \mathbb{R}$ the variational problem is as follows:

Find $u \in V$ such that

$$|a(u,v)| = l(v), \text{ for all } v \in V. \tag{2.6}$$

Two essential theorems ensure that Problem (2.6) has a unique solution.

Theorem 2.3.2 (Riesz' Representation Theorem)
Let V be a Hilbert space and let $v \in V$. Every linear form $l \in V^*$ can be uniquely represented as

$$l(v) = (u,v), \tag{2.7}$$

with some $u \in V$.

Proof See, e. g., (Larson and Bengzon, 2013, Chap. 7.3.2, Theorem 7.1). ■

Theorem 2.3.2 ensures uniqueness and solvability of Problem (2.6), if the bilinear form is symmetric. Lax-Milgram's Lemma is the generalization for non symmetric bilinear forms.

Theorem 2.3.3 (Lax-Milgram Lemma)
Let V be a Hilbert space, and let $a : V \times V \to \mathbb{R}$ be a coercive and continuous bilinear form, let further $l \in V^*$ be continuous. Then there exists a unique $u \in V$, such that for all $v \in V$

$$a(u,v) = l(v). \tag{2.8}$$

Proof The proof can be found for example in (Larson and Bengzon, 2013, Chap. 7.3.4, Theorem 7.3). ■

So called elliptic differential problems will play a prominent role throughout this work, hence, the application of the Lax-Milgram Lemma shall be demonstrated for such a problem.

To this end let $\Omega \subset \mathbb{R}^d$ be a Lipschitz bounded domain, and let $L : \mathcal{C}^2(\Omega) \to \mathcal{C}^0(\Omega)$ be a second order linear elliptic differential operator in divergence form, i. e., there exist a bounded and continuously differentiable field

$$\underbrace{\boldsymbol{A}}_{=(a_{ij})} : \Omega \to \mathbb{R}^{d\times d},$$

and bounded continuous fields

$$\underbrace{\boldsymbol{b}}_{=(b_i)} : \Omega \to \mathbb{R}^d, \text{ and } \quad c : \Omega \to \mathbb{R},$$

where $\boldsymbol{A}(\boldsymbol{x})$ is symmetric and positive definite for all $\boldsymbol{x} \in \Omega$, such that

$$\begin{aligned} Lu &= -\sum_{i,j=1}^{d} \partial_i \left(a_{ij}\partial_j u\right) + \sum_{i=1}^{d} b_i \partial_i u + cu \\ &= -\nabla \cdot (\boldsymbol{A}\nabla u) + \boldsymbol{b} \cdot \nabla u + cu. \end{aligned} \tag{2.9}$$

Further, let there exist constants $a_0, c_0 > 0$, such that

$$\begin{aligned} a_0 \sum_{i=1}^{d} x_i^2 &\leq \sum_{i,j=1}^{d} a_{ij} x_i x_j, \\ c_0 &< c - \frac{1}{2} \sum_{i=1}^{d} \frac{\partial b_i}{\partial x_i}. \end{aligned} \tag{2.10}$$

The *strong* form of the homogenous Dirichlet boundary value problem is seeking $u \in \mathcal{C}^0(\bar{\Omega}) \cap \mathcal{C}^2(\Omega)$ for a bounded $f \in L^2(\Omega)$, such that

$$\begin{aligned} Lu &= f, \text{ in } \Omega, \\ u &= 0, \text{ on } \partial\Omega. \end{aligned} \tag{2.11}$$

General boundary conditions can be established via the trace operator. Instead of trying to find such strong solutions, the problem is relaxed to functions in H_0^1, where a so called *weak solution* shall be established.

Multiplying Eq. (2.11) with an arbitrary test function $\varphi \in H_0^1(\Omega)$, integrating it over the whole domain and applying Green's Formula yields:

$$\int_\Omega \boldsymbol{A}\nabla u \cdot \nabla \varphi + \boldsymbol{b} \cdot \nabla u \varphi + cu\varphi \, \mathrm{d}\boldsymbol{x} = \int_\Omega f\varphi \, \mathrm{d}\boldsymbol{x}, \text{ for all } \varphi \in H_0^1.$$

Above equation can, by using the L^2 inner product, be compactly written as

$$\underbrace{(\boldsymbol{A}\nabla u, \nabla\varphi) + (\boldsymbol{b} \cdot \nabla u, \varphi) + (cu, \varphi)}_{=a(u,\varphi)} = \underbrace{(f, \varphi)}_{=l(\varphi)}.$$

Here, $\cdot$ now denotes the usual inner product in $\mathbb{R}^d$, i.e, $\boldsymbol{x} \cdot \boldsymbol{y} := \boldsymbol{x}^\top \boldsymbol{y} = \sum_{i=1}^{d} x_i y_i$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$.

First, continuity of the bilinear form is shown. From the Cauchy-Schwarz inequality the following can be deduced:

$$\begin{aligned} a(u,\varphi) &\leq \|\boldsymbol{A}\nabla u\|_{L^2}\|\nabla\varphi\|_{L^2} + \|\boldsymbol{b}\cdot\nabla u\|_{L^2}\|\varphi\|_{L^2} + \|cu\|_{L^2}\|\varphi\|_{L^2} \\ &\leq \|\boldsymbol{A}\|_\infty\|\nabla u\|_{L^2}\|\nabla\varphi\|_{L^2} + \|\boldsymbol{b}\|_\infty\|\nabla u\|_{L^2}\|\varphi\|_{L^2} + \|c\|_{L^\infty}\|u\|_{L^2}\|\varphi\|_{L^2}, \end{aligned} \tag{2.12}$$

where

$$\begin{aligned} \|\boldsymbol{A}\|_\infty &= \max_{1\leq i,j\leq d}\|a_{ij}\|_{L^\infty}, \\ \|\boldsymbol{b}\|_\infty &= \max_{1\leq i\leq d}\|b_i\|_{L^\infty}. \end{aligned}$$

By recalling the definition of the norm on H_0^1 from Remark 2.2.8 and overestimating all constants with $C \in \mathbb{R}$, Eq. (2.12) is further assessed against the bilinear form:

$$\begin{aligned} a(u,\varphi) &\leq \|\boldsymbol{A}\|_\infty\|\nabla u\|_{L^2}\|\nabla\varphi\|_{L^2} + \|\boldsymbol{b}\|_\infty\|\nabla u\|_{L^2}\|\varphi\|_{L^2} + \|c\|_{L^\infty}\|u\|_{L^2}\|\varphi\|_{L^2} \\ &\leq C\left(\|\nabla u\|_{L^2}\|\nabla\varphi\|_{L^2} + \|\nabla u\|_{L^2}\|\varphi\|_{L^2} + \|u\|_{L^2}\|\varphi\|_{L^2}\right) \\ &\leq C\|u\|_{H_0^1}\|\varphi\|_{H_0^1}. \end{aligned}$$

To show that the bilinear form is coercive, on the one hand the divergence theorem together with $u = 0$ on $\partial\Omega$ and, on the other the product rule for the divergence operator $(\nabla\cdot)$ has to be used as follows:

$$\begin{aligned} 0 = \left(\boldsymbol{n}\cdot\boldsymbol{b}u^2, 1\right)_{L^2(\partial\Omega)} &= \left(\nabla\cdot\left(\boldsymbol{b}u^2\right), 1\right) \\ &= \left((\nabla\cdot\boldsymbol{b})u, u\right) + 2\left(\boldsymbol{b}\cdot\nabla u, u\right). \end{aligned}$$

Where $\boldsymbol{n}$ denotes the normal vector of Ω and 1 denotes the constant function mapping to 1. Consequently, with Eq. (2.10) the bilinear form a can be gauged by

$$\begin{aligned} a(u,u) &= (\boldsymbol{A}\nabla u, \nabla u) - \frac{1}{2}\left((\nabla\cdot\boldsymbol{b})u, u\right) + (cu, u) \\ &\geq a_0\|\nabla u\|^2 + c_0\|u\|^2 \\ &\geq m\|u\|_{H_0^1}^2, \end{aligned}$$

yielding coercivity, with $m = \min\{a_0, c_0\}$.

Continuity of the linear form is also obtained by the Cauchy-Schwarz inequality and from $f \in L^2$:

$$\begin{aligned} l(\varphi) = (f,\varphi) &\leq \|f\|_{L^2}\|\varphi\|_{L^2} \\ &\leq F\|\varphi\|_{H_0^1}. \end{aligned}$$

Hence, by Theorem 2.3.3 the weak form has a solution in H_0^1.

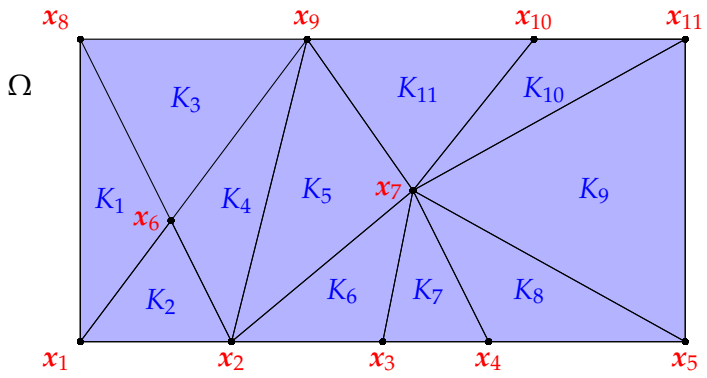


Figure 2.2.: An exemplary mesh, where the elements K_j are triangles.

2.4. Discrete Form

A solution to the variational Problem (2.6) is approximated by using finite elements, i.e., a triplet consisting of a collection of polygons, used to construct a mesh, a polynomial space, used as a finite dimensional approximation to the according Hilbert space V, and a set of functionals, defining the degrees of freedom. Let $V_h \subset V$ be a finite dimensional subspace, typically the space of piecewise polynomials on a mesh $\mathcal{T}_h$ of Ω, i.e: $\mathcal{T}_h = \{K_m\}_{1 \leq m \leq N_{\text{el}}}$, where K_m are compact and connected Lipschitz sets, with non-empty interior, such that

$$\overline{\Omega} = \bigcup_{m=1}^{N_{\text{el}}} K_m, \text{ and } \quad \mathring{K}_m \cap \mathring{K}_n = \emptyset.$$

Each $K_m \subset \mathbb{R}^d$ is usually a polygon, for example a triangle like in Fig. 2.2. The index h refers to the degree of refinement of the mesh:

$$h = \max_{K \in \mathcal{T}_h} (\operatorname{diam}(K)) = \max_{K \in \mathcal{T}_h} \left(\max_{x_1, x_2 \in K} \|x_1 - x_2\| \right).$$

The approximation to Problem (2.6) on V_h takes the form: Find $u_h \in V_h$, such that

$$a(u_h, v) = l(v), \text{ for all } v \in V_h. \tag{2.13}$$

At this point it becomes clear that the subspace V_h is chosen such, that $u_h \to u \in V$ for $h \to 0$. For a basis $\{\varphi_i\}_{i=1}^n$ of V_h Eq. (2.13) is equivalent to

$$a(u_h, \varphi_i) = l(\varphi_i), \text{ for } i = 1 \ldots, n. \tag{2.14}$$

Obviously, every solution u_h to Eq. (2.14) can also be represented by the basis:

$$u_h = \sum_{j=1}^{n} u_j \varphi_j,$$

with unknown coefficients $u_j \in \mathbb{R}$. If this identity is plugged into Eq. (2.14), it reduces to

$$b_i = l(\varphi_i) = \sum_{j=1}^{n} u_j a(\varphi_j, \varphi_i) = \sum_{j=1}^{n} s_{ij} u_j, \text{ for } i = 1 \ldots, n,$$

which can be compactly written as a matrix vector product

$$\boldsymbol{S}\boldsymbol{u} = \boldsymbol{b}. \tag{2.15}$$

The matrix $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ is the *system*, or *stiffness*, matrix and the vector $\boldsymbol{b} \in \mathbb{R}^n$ is called *load vector*. From the solution (which, due to corresponding boundary conditions, exists uniquely) of the linear system in Eq. (2.15) the finite element approximation $u_h \in V_h$ to Problem (2.6) is obtained. The discussed conceptual idea is exerted on each element of the mesh, yielding an efficient strategy to assemble the global quantities $\boldsymbol{S}$ and $\boldsymbol{b}$ element wise. The outlined construction, where the space of test functions is the same as the ansatz space is known as *conformal finite elements*.

A comprehensive discussion about the efficient assembling of all matrices and vectors necessary to solve the general elliptic boundary problem is for example given in Dudzinski et al. (2016). Further readings are given in, e.g., Ern and Guermond (2013); Larson and Bengzon (2013).

3 Interior Point Methods in a Nutshell

Within this work optimization problems in various forms occur. The aim of the following chapter is to give a mathematical description of those problems and describe an efficient solution strategy, which is exploited within the scope of this work . After a short introduction in Section 3.1 the basic mathematical definitions and major theorems in the field of constrained optimization are recapitulated in Section 3.2. In Section 3.3 a basic version of an interior point algorithm is given. The chapter is concluded by a proof of equivalence of different formulations of a non-linear constrained optimization problem in Section 3.4.

3.1. Motivation and Introduction

Optimization plays a prominent role in a vast number of disciplines. In many applications from physics, for example, some energy term has to be minimized (see, e.g., Sheppard et al., 2008), in certain engineering problems optimal shapes are desired (see, e.g., Schwarz et al., 2001), in finance some cost need to be minimized, where on the other hand sales need to be maximized (see, e.g., Schaible, 1981), to just call a few. Mostly the figure of merit that has to be optimized is subject to constraints that are expressed as equations or inequalities. Further, the participating functions are often non-linear, which adds more difficulty to finding a solution and yields non-linear constrained optimization problems. Since these problems appear in many different disciplines, an accessible and complete theory (see, e.g., Nocedal and Wright, 2006; Bazaraa and Shetty, 1979; Fletcher, 2000), as well as easy applicable algorithms were tailored in the last decades, like ALGENCAN (see Andreani et al., 2008, 2007), KNITRO (see Byrd et al., 2000, 1999, 2006), LANCELOT (see Conn et al., 2010), filterSQP (see

Fletcher and Leyffer, 1998), NPSOL (see Gill et al., 1984), SNOPT (see Gill et al., 2005), PENNON (see Kočvara and Stingl, 2003), MINOS (see Murtagh and Saunders, 1983), LOQO (see Vanderbei and Shanno, 1999) or IPOPT (see Wächter and Biegler, 2006).

Mainly two branches of optimization methods can be distinguished, derivative-free methods and gradient based methods. Whereas derivative free methods based on surrogates for the objective function or heuristics for the direction of descent, like Kriging (see, e.g., Jones et al., 1998), simulated annealing (Van Laarhoven and Aarts, 1987) or the Nelder-Mead method proposed by Nelder and Mead (1965), require no further information about the function, the number of—under certain circumstances expensive—objective calculations is higher than when applying gradient based methods.

A very efficient class of gradient based methods are the so called *Interior Point* (IP) methods. They, as will be shown later, generate a sequence of strictly feasible points, which, under certain conditions, converges towards a local optimum of a given, usually non-linear, optimization problem. This feature makes them very attractive for the type of problems covered in this work. Firstly, feasibility is usually defined by the region where the model assumptions, or the assumed physical laws, hold. Secondly, even if the algorithm fails to converge towards an optimal solution, for various reasons, the identified point generally improves the objective and is feasible, thus yields a desired enhancement.

IP methods are well understood and efficient implementations are available (see, e.g., Wächter and Biegler, 2006; Byrd et al., 2000). For the sake of completeness a short introduction into constrained optimization along with the basic concepts of these methods are recapitulated, such that a basic algorithm can be formulated. Next, a result showing the equivalence of the first order necessary conditions for constrained optimization for several problem (re-) formulations is presented.

3.2. Constrained Optimization

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable real valued function. Further, let the functions $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^k$ also be twice continuously differentiable. This particular smoothness is only needed to formulate second order conditions, mostly continuous differentiability is sufficient. With these functions the following optimization

problem is formulated:

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & f(\boldsymbol{x}), \\ \text{subject to} \quad & g(\boldsymbol{x}) \leq \mathbf{0}, \\ & h(\boldsymbol{x}) = \mathbf{0}, \end{aligned} \tag{3.1}$$

where all inequalities have to be understood component wise. The functions f, g and h are allowed to be non-linear. For the time being no box constraints, i.e. inequalities of the type $\boldsymbol{x} \leq (\geq)\mathbf{0}$, to the variables are considered explicitly. They can, however, be expressed with the inequalities given by the function g.

As far as possible all vectors are denoted by lowercase bold symbols, matrices as upper case bold symbols and scalars are small italic letters. The gradient is denoted by ∇, and $\nabla_{\boldsymbol{s}}$ indicates that only those partial derivatives with respect to $\boldsymbol{s}$ are taken into account. Similar notations are used for the Hessian, denoted by ∇^2, or $\nabla^2_{\boldsymbol{xx}}$ if only special variables are considered.

Within this section necessary and sufficient conditions for the characterization of local solutions to Problem (3.1) will be presented. These will be guidelines to construct efficient algorithms to find approximations to those solutions. All given results can also be found in Nocedal and Wright (2006) or other textbooks about constrained optimization.

Definition 3.2.1 (Feasible Set)
The *feasible set* Ω_f is given as set of points $\boldsymbol{x} \in \mathbb{R}^n$ at which the constraints hold, i.e.,

$$\Omega_f = \{\boldsymbol{x} : h(\boldsymbol{x}) = \mathbf{0} \text{ and } g(\boldsymbol{x}) \leq \mathbf{0}\}.$$

Further, let $\mathcal{I}$ and $\mathcal{E}$ be the following sets of indices $\mathcal{E} = \{1, \ldots, k\}$, and $\mathcal{I} = \{1, \ldots, m\}$.▲

Definition 3.2.2 (Active Set)
The set of *active inequality constraints* $\mathcal{A}^{\mathcal{I}}(\boldsymbol{x})$ at a given point $\boldsymbol{x} \in \Omega_f$ is given as

$$\mathcal{A}^{\mathcal{I}}(\boldsymbol{x}) = \{j \in \mathcal{I} : g_j(\boldsymbol{x}) = 0\}. \tag{3.2}$$

The *active set* $\mathcal{A}(\boldsymbol{x})$ at an arbitrary $\boldsymbol{x} \in \Omega_f$ is the set of indices from $\mathcal{E}$ and those from $\mathcal{A}^{\mathcal{I}}(\boldsymbol{x})$, i.e.,

$$\mathcal{A}(\boldsymbol{x}) = \left\{(i,j) : i \in \mathcal{E},\, j \in \mathcal{A}^{\mathcal{I}}(\boldsymbol{x})\right\}. \tag{3.3}$$

Inequalities that hold strictly at $\boldsymbol{x} \in \Omega_f$, i.e., $g_j(\boldsymbol{x}) < 0$, $j \in \mathcal{I}$, are called *inactive*. ▲

With Definition 3.2.1 Problem (3.1) can be written compactly as

$$\min_{\boldsymbol{x}\in\Omega_f} \quad f(\boldsymbol{x}).$$

The active set will be useful to characterize solutions, which are often found at the boundary of the feasible set. A point $\boldsymbol{x}^* \in \Omega_f$ is called *local solution to*, or *local optimum of* Problem (3.1), if there exists a neighborhood $\mathcal{N} \subset \Omega_f$ of $\boldsymbol{x}^*$, such that $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{N}$.

However, when aiming to develop necessary conditions for constrained problems, it is clear, that the constraints must be taken into account carefully. The first order condition known from unconstrained optimization

$$\nabla f(\boldsymbol{x}) = 0,$$

is not enough anymore; the linearizations of the constraints have to be considered, too. Further, conditions which ensure that these linearizations still contain enough information about the original problem are needed. Such conditions are called constraint qualifications and they aim to formulate requirements under which the linearization of the active set is similar enough to the geometry of the feasible set, in a neighborhood of a given point $\boldsymbol{x} \in \Omega_f$.

Definition 3.2.3 (Tangent Vector and Tangent Cone)
The vector $\boldsymbol{d} \in \mathbb{R}^n$ is a *tangent vector* to Ω_f at a point $\boldsymbol{x} \in \Omega_f$, if there exist a sequence $\{z_k\}$, with $z_k \in \Omega_f$, for all $k \in \mathbb{N}$, and $z_k \xrightarrow[k\to\infty]{} \boldsymbol{x}$, and a sequence $\{t_k\}$, with $t_k \in \mathbb{R}$, $t_k > 0$, and $t_k \xrightarrow[k\to\infty]{} 0$, such that

$$\lim_{k\to\infty} \frac{z_k - \boldsymbol{x}}{t_k} = \boldsymbol{d}. \tag{3.4}$$

The set of all tangent vectors to Ω_f at a point $\boldsymbol{x}$ is the *tangent cone*, $T_{\Omega_f}(\boldsymbol{x})$. The limit of the expression in Eq. (3.4) is often referred to as the *limiting direction*. ▲

The tangent cone $T_{\Omega_f}(\boldsymbol{x})$ is a cone in the following sense:

A *cone* $\mathcal{C}$ is a set, such that for all $\boldsymbol{x} \in \mathcal{C}$ also follows that $\alpha\boldsymbol{x} \in \mathcal{C}$, for all $\alpha > 0$.

To see that this property holds for $\mathcal{C}$, let $\boldsymbol{d} \in T_{\Omega_f}(\boldsymbol{x})$, then there exist corresponding sequences $\{z_k\}$ and $\{t_k\}$. For every $\alpha > 0$ the sequence $\{\alpha^{-1}t_k\}$ still converges towards 0, thus $\alpha\boldsymbol{d} \in T_{\Omega_f}(\boldsymbol{x})$. When setting $z_k \equiv x$, it follows that $\mathbf{0} \in T_{\Omega_f}(\boldsymbol{x})$. The definition of the tangent cone is independent of the algebraic representation of the set Ω_f, it only depends on the feasible set's geometry.

Definition 3.2.4 (The Set of Linearized Directions)
Let $\boldsymbol{x} \in \Omega_f$ and let $\mathcal{A}(\boldsymbol{x})$ be the according active set, then the set

$$\begin{aligned} \mathcal{F}(\boldsymbol{x}) = \{\boldsymbol{d} : \boldsymbol{d}^\top \nabla h_i(\boldsymbol{x}) = 0, \text{ for all } i \in \mathcal{E}, \\ \boldsymbol{d}^\top \nabla g_j(\boldsymbol{x}) \leq 0, \text{ for all } j \in \mathcal{A}^{\mathcal{I}}(\boldsymbol{x})\Big\}, \end{aligned} \tag{3.5}$$

is the *set of feasible linearized directions* at $\boldsymbol{x}$. ▲

Opposed to $T_{\Omega_f}(\boldsymbol{x})$, the set $\mathcal{F}(\boldsymbol{x})$ depends on the algebraic formulation of the constraints.

The aforementioned constraint qualifications aim to formulate prerequisites under which $\mathcal{F}(\boldsymbol{x})$ is similar to $T_{\Omega_f}(\boldsymbol{x})$, or even identical. Wang et al. (2013) give a list of some constraint qualifications and show how they are connected, since some are stronger and imply other qualifications. In this work only one will be quoted, the linear independency constraint qualification, since it can conveniently be checked algebraically.

Definition 3.2.5 (LICQ)
Let $\boldsymbol{x} \in \Omega_f$ and let $\mathcal{A}(\boldsymbol{x})$ be the according active set, the *linear independence constraint qualification (LICQ)* holds if and only if all gradients of the active constraints are linearly independent, i.e, $\{\nabla h_i(\boldsymbol{x}), \nabla g_j(\boldsymbol{x}) : (i,j) \in \mathcal{A}(\boldsymbol{x})\}$ consists of linearly independent vectors. ▲

Prior to formulating the first order necessary conditions, the Lagrangian to Problem (3.1) is introduced.

Definition 3.2.6 (Lagrangian)
Given an optimization problem as in Problem (3.1), then the *Lagragian* to it reads as

$$\begin{aligned} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\vartheta}) &= f(\boldsymbol{x}) + \sum_{i \in \mathcal{E}} \lambda_i h_i(\boldsymbol{x}) + \sum_{j \in \mathcal{I}} \vartheta_j g_j(\boldsymbol{x}) \\ &= f(\boldsymbol{x}) + \sum_{i=1}^{k} \lambda_i h_i(\boldsymbol{x}) + \sum_{j=1}^{m} \vartheta_j g_j(\boldsymbol{x}) \\ &= f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top h(\boldsymbol{x}) + \boldsymbol{\vartheta}^\top g(\boldsymbol{x}). \end{aligned} \tag{3.6}$$

The vectors $\boldsymbol{\lambda} \in \mathbb{R}^k$, and $\boldsymbol{\vartheta} \in \mathbb{R}^m$ are the *Lagrange multipliers* or *dual variables*. Note that the sign in front of $\boldsymbol{\vartheta}$ depends on the type of inequality, if the inequality read $g(\boldsymbol{x}) \geq 0$ it would be a minus. ▲

With this the following theorem is formulated.

Theorem 3.2.7 (First Order Necessary Conditions)
Let $\boldsymbol{x}^* \in \mathbb{R}^n$ be a local solution to Problem (3.1), and let the LICQ hold at $\boldsymbol{x}^*$. Then there exist Lagrange multipliers $\boldsymbol{\lambda}^* \in \mathbb{R}^k$ and $\boldsymbol{\vartheta}^* \in \mathbb{R}^m$ such that the following conditions hold:

$$\nabla_x \mathcal{L}\left(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\vartheta}^*\right) = \mathbf{0}, \tag{3.7a}$$
$$h\left(\boldsymbol{x}^*\right) = \mathbf{0}, \tag{3.7b}$$
$$g\left(\boldsymbol{x}^*\right) \leq \mathbf{0}, \tag{3.7c}$$
$$\vartheta_j^* g_j\left(\boldsymbol{x}^*\right) = 0, \text{ for } j = 1, \ldots, k, \tag{3.7d}$$
$$\boldsymbol{\vartheta}^* \geq \mathbf{0}. \tag{3.7e}$$

Proof The proof to this theorem is given in (Nocedal and Wright, 2006, Chap. 12.4). It makes use of the fact that if the LICQ holds at $\boldsymbol{x}^*$, $\mathcal{F}\left(\boldsymbol{x}^*\right) = T_{\Omega_f}\left(\boldsymbol{x}^*\right)$ follows. Further, Farkas' Lemma is exploited and it is shown that for all local solutions $\boldsymbol{x}^*$ the tangent vectors yield a non-negative inner product with the gradient of the objective, i.e., $\nabla f\left(\boldsymbol{x}^*\right)^\top \boldsymbol{d} \geq 0$, for all $\boldsymbol{d} \in T_{\Omega_f}\left(\boldsymbol{x}^*\right)$. ■

The conditions in Eq. (3.7) are called *Karush-Kuhn-Tucker (KKT)* conditions. The KKT conditions are a generalization of the usual first order necessary conditions from unconstrained optimization to constrained optimization. They have been developed by Kuhn and Tucker (1951) and independently earlier by Karush (1939). Equation (3.7a) is referred to as the *stationarity condition*, Eqs. (3.7b) and (3.7c) are the *primal feasibility* conditions, Eq. (3.7d) is called *complementary slackness condition* and Eq. (3.7e) is the so called *dual feasibility*. From Eqs. (3.7d) and (3.7e) it can be deduced, that $\vartheta_j^* = 0$ for $j \in \mathcal{I} \setminus \mathcal{A}^{\mathcal{I}}\left(\boldsymbol{x}^*\right)$, that is, the multipliers for the inactive constraints are 0.

In addition to the first order conditions discussed above, second order conditions can be formulated. These are, as in the unconstrained case, needed for the formulation of sufficient conditions. Their role can be described as the one of a decision maker in the case, when a direction $\boldsymbol{d} \in \mathcal{F}\left(\boldsymbol{x}^*\right)$ is encountered, for which $\nabla f\left(\boldsymbol{x}^*\right)^\top \boldsymbol{d} = 0$ holds. Both, a necessary and a sufficient condition can be stated. The characterization of those directions which yield the above mentioned tie is done via the critical cone.

Definition 3.2.8 (Critical Cone)
Let $\boldsymbol{x}^* \in \mathbb{R}^n$ be a local solution to Problem (3.1), and let $\boldsymbol{\lambda}^* \in \mathbb{R}^k$ and $\boldsymbol{\vartheta}^* \in \mathbb{R}^m$ be Lagrange multipliers, such that the KKT conditions Eq. (3.7) hold. The *critical cone* $\mathcal{C}\left(\boldsymbol{x}^*, \boldsymbol{\vartheta}^*\right)$ is given by

$$\mathcal{C}\left(\boldsymbol{x}^*, \boldsymbol{\vartheta}^*\right) = \left\{ \boldsymbol{w} \in \mathcal{F}\left(\boldsymbol{x}^*\right) : \boldsymbol{w}^\top \nabla g_j\left(\boldsymbol{x}^*\right) = 0, \text{ for all } j \in \mathcal{A}^{\mathcal{I}}\left(\boldsymbol{x}^*\right) \text{ with } \vartheta_j^* > 0 \right\}. \quad ▲$$

From Definitions 3.2.4 and 3.2.8 the following can easily be deduced:

$$w \in \mathcal{C}(x^*, \vartheta^*) \Leftrightarrow \begin{cases} w^\top \nabla h_i(x^*) & = 0, \text{ for all } i \in \mathcal{E}, \\ w^\top \nabla g_j(x^*) & = 0, \text{ for all } j \in \mathcal{A}^{\mathcal{I}}(x^*) \text{ with } \vartheta_j^* > 0, \\ w^\top \nabla g_j(x^*) & \leq 0, \text{ for all } j \in \mathcal{A}^{\mathcal{I}}(x^*) \text{ with } \vartheta_j^* = 0. \end{cases}$$

Further, from $\vartheta_j^* = 0$, for $j \in \mathcal{I} \setminus \mathcal{A}^{\mathcal{I}}(x^*)$ it follows:

$$w \in \mathcal{C}(x^*, \vartheta^*) \Rightarrow \vartheta_j^* w^\top \nabla g_j(x^*) = 0, \text{ for all } j \in \mathcal{I}.$$

Hence, with the KKT conditions and the definition of the gradient of the Lagrangian it follows

$$w \in \mathcal{C}(x^*, \vartheta^*) \Rightarrow w \nabla f(x^*) = -\sum_{j \in \mathcal{I}} \vartheta_j^* w^\top \nabla g_j(x^*) = 0.$$

This shows that $\mathcal{C}(x^*, \vartheta^*)$ gathers all directions, which neither yield descent nor ascent of the objective function.

Theorem 3.2.9 (Second Order Necessary Condition)
Let $x^* \in \mathbb{R}^n$ be a local solution to Problem (3.1), and let the LICQ hold at x^*. Let $\lambda^* \in \mathbb{R}^k$ and $\vartheta^* \in \mathbb{R}^m$ be Lagrange multipliers, such that the KKT conditions, Eq. (3.7), hold. Then

$$w^\top \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*, \vartheta^*) w \geq 0, \text{ for all } w \in \mathcal{C}(x^*, \vartheta^*). \tag{3.8}$$

Proof The proof to this is found in (Nocedal and Wright, 2006, Chap. 12.5, Theorem 12.5). It is done by constructing a feasible sequence $\{z_k\}$ converging towards x^* with the property $f(z_k) \geq f(x^*)$ from which the stated follows for the limiting direction w belonging to this sequence (see Definition 3.2.3).

Theorem 3.2.10 (Second Order Sufficient Condition)
Assume that for some feasible point $x^* \in \Omega_f$ there are Lagrange multipliers $\lambda^* \in \mathbb{R}^k$ and $\vartheta^* \in \mathbb{R}^m$ for which the KKT conditions as in Eq. (3.7) hold. Assume further, that for all $w \in \mathcal{C}(x^*, \vartheta^*) \setminus \{0\}$ the following holds:

$$w^\top \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*, \vartheta^*) w > 0. \tag{3.9}$$

Then x^* is a strict local solution to Problem (3.1), that is, there exists a neighborhood $\mathcal{N}$ of x^*, such that $f(x) > f(x^*)$ for all $x \in \mathcal{N} \cap \Omega_f$.

Proof The proof is found in (Nocedal and Wright, 2006, Chap. 12.5, Theorem 12.6). It is shown by contradiction, that every feasible sequence $\{z_k\}$ converging towards x^* has the property $f(z_k) > f(x^*)$ following from the assumptions. ■

Equipped with these basic fundamentals about constrained optimization of smooth functions, a simple interior point algorithm can be developed.

3.3. An Interior Point Algorithm

For convenience, Problem (3.1) is stated once again, with a slight modification on the objective variables, by including box constraints. For twice partially differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^k$ the problem reads:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x), \\ \text{subject to} \quad & g(x) \leq 0, \\ & h(x) = 0, \\ & x \geq 0. \end{aligned} \tag{3.10}$$

For the time beeing all components of the objective variable x are assumed to be positive, i.e., $x_i \geq 0$, for all $i = 1, \ldots, n$. It will become clear in Section 3.4 why this assumption is justified.

In the context of IP methods, Problem (3.10) is reformulated into a problem with only equality and box constraints on the variables. In order to do so, a component-wise positive *slack variable* $s \in \mathbb{R}^m$ is introduced and added to the inequalities, yielding the following optimization problem:

$$\begin{aligned} \min_{x \in \mathbb{R}^n, s \in \mathbb{R}^m} \quad & \tilde{f}(x, s), \\ \text{subject to} \quad & \tilde{c}_s(x, s) = 0, \\ & x, s \geq 0. \end{aligned} \tag{3.11}$$

Here, the participating functions $\tilde{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, and $\tilde{c}_s : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{m+k}$ are defined as

$$\begin{aligned} \tilde{f}(x, s) &= f(x), \\ \tilde{c}_s(x, s) &= \begin{pmatrix} g(x) + s \\ h(x) \end{pmatrix}. \end{aligned}$$

The extension of f, as a function defined on $\mathbb{R}^n$ to the function $\tilde{f}$ defined on $\mathbb{R}^n \times \mathbb{R}^m$, makes it possible to compute gradients with respect to the slack variable $\boldsymbol{s}$, which is treated as a *primal variable*. However, Nocedal and Wright (2006) formulate the following without taking the slack variable into consideration. This can be done and will yield the same principal algorithm, it just lacks some strictness, which will be needed next, in Section 3.4.

A strategy to find a solution to Problem (3.11) is to find critical points of the associated Lagrangian function, given by

$$\mathcal{L}\left(\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{\lambda}, \boldsymbol{\vartheta}\right) = \tilde{f}\left(\boldsymbol{x}, \boldsymbol{s}\right) + \boldsymbol{\lambda}^\top \tilde{c}_s\left(\boldsymbol{x}, \boldsymbol{s}\right) - \boldsymbol{\vartheta}^\top \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{s} \end{pmatrix}, \tag{3.12}$$

which also fulfil the KKT conditions, particularly given as:

$$\nabla \tilde{f}\left(\boldsymbol{x}, \boldsymbol{s}\right) + \boldsymbol{J}_{\tilde{c}_s}^\top \boldsymbol{\lambda} - \boldsymbol{\vartheta} = \boldsymbol{0}, \tag{3.13a}$$
$$\tilde{c}_s\left(\boldsymbol{x}, \boldsymbol{s}\right) = \boldsymbol{0}, \tag{3.13b}$$
$$\boldsymbol{x}, \boldsymbol{s} \geq \boldsymbol{0}, \tag{3.13c}$$
$$\boldsymbol{X}\boldsymbol{\vartheta} = \boldsymbol{0}, \tag{3.13d}$$
$$\boldsymbol{\vartheta} \geq \boldsymbol{0}. \tag{3.13e}$$

The matrix $\boldsymbol{J}_{\tilde{c}_s}$ denotes the Jacobian of the function $\tilde{c}_s$. The vector $\boldsymbol{\lambda} \in \mathbb{R}^{m+k}$ is the Lagrangian multiplier associated to the equality constraints, $\boldsymbol{\vartheta} \in \mathbb{R}^{n+m}$ is the multiplier belonging to the inequality constraints, and $\boldsymbol{X} = \operatorname{diag}\left(\boldsymbol{x}, \boldsymbol{s}\right)$ is a diagonal matrix whose diagonal entries are given by the vectors $\boldsymbol{x}$ and $\boldsymbol{s}$. This approach is motivated by Theorem 3.2.7, since together with the LICQ these are the necessary conditions for a local optimum.

It is generally a hard combinatorial problem to identify the active set $\mathcal{A}(\boldsymbol{x}, \boldsymbol{s})$ close to an optimum to check for sufficient conditions. To overcome this particular problem a regularization is introduced, yielding the *perturbed KKT* conditions, given as:

$$\nabla \tilde{f}\left(\boldsymbol{x}, \boldsymbol{s}\right) + \boldsymbol{J}_{\tilde{c}_s}^\top \boldsymbol{\lambda} - \boldsymbol{\vartheta} = \boldsymbol{0}, \tag{3.14a}$$
$$\tilde{c}_s\left(\boldsymbol{x}, \boldsymbol{s}\right) = \boldsymbol{0}, \tag{3.14b}$$
$$\boldsymbol{x}, \boldsymbol{s} \geq \boldsymbol{0}, \tag{3.14c}$$
$$\boldsymbol{X}\boldsymbol{\vartheta} - \mu \mathbb{1} = \boldsymbol{0}, \tag{3.14d}$$
$$\boldsymbol{\vartheta} \geq \boldsymbol{0}. \tag{3.14e}$$

Here, $\mathbb{1}$ is a vector of ones of appropriate size, i.e., $\mathbb{1} = (1, \ldots, 1)^\top$ and $\mu \in \mathbb{R}$, $\mu > 0$ is the regularization parameter, also denoted as *barrier parameter*. With this parameter

the problem of identifying the active set can be omitted, instead the primal and dual variables are strictly away form $\mathbf{0}$ as long as $\mu > 0$. This property contributes to the name *inner point method*, since all solutions to Eq. (3.14) are strictly inside the region given by the inequalities.

System (3.14) are the KKT conditions to another optimization problem, this, for now, is referred to as the *perturbed problem* or *barrier problem*, and is discussed in detail in Section 3.4.

However, the link between a solution to the perturbed problem and Problem (3.11) can be established by the following line of thoughts: If $(\boldsymbol{x}^*, \boldsymbol{s}^*, \boldsymbol{\lambda}^*, \boldsymbol{\vartheta}^*)$ is a solution to Problem (3.11), such that in a neighborhood the LICQ (Definition 3.2.5) holds, and $\vartheta_j^* > 0$, for all $j \in \mathcal{A}^{\mathcal{I}}(\boldsymbol{x}^*, \boldsymbol{s}^*)$, and the second order sufficient conditions stated in Theorem 3.2.10 hold, then Eq. (3.14) has a local solution $(\boldsymbol{x}(\mu), \boldsymbol{s}(\mu), \boldsymbol{\lambda}(\mu), \boldsymbol{\vartheta}(\mu))$ for all sufficiently small $\mu > 0$. The sequence of those solutions is called *central path*, and it converges towards $(\boldsymbol{x}^*, \boldsymbol{s}^*, \boldsymbol{\lambda}^*, \boldsymbol{\vartheta}^*)$, for $\mu \to 0$. The proof to this is similar to the proof of Theorem 3.3.2.

To solve Eq. (3.14) Newton's method is used, yielding the following linear system:

$$\begin{bmatrix} \nabla^2_{(\boldsymbol{x},\boldsymbol{s}),\boldsymbol{x}}\mathcal{L} & \nabla^2_{(\boldsymbol{x},\boldsymbol{s}),\boldsymbol{s}}\mathcal{L} & \boldsymbol{J}_{\tilde{c}_s}^\top & -\boldsymbol{I} \\ \boldsymbol{I}\boldsymbol{\Theta}_{\boldsymbol{x}} & \boldsymbol{I}\boldsymbol{\Theta}_{\boldsymbol{s}} & \mathbf{0}_{\mathbb{R}^{(n+m)\times(m+k)}} & \boldsymbol{X} \\ (\nabla_{\boldsymbol{x}}\tilde{c}_s)^\top & (\nabla_{\boldsymbol{s}}\tilde{c}_s)^\top & \mathbf{0}_{\mathbb{R}^{(m+k)\times(m+k)}} & \mathbf{0}_{\mathbb{R}^{(m+k)\times(n+m)}} \\ \boldsymbol{I}_{\boldsymbol{x}} & \boldsymbol{I}_{\boldsymbol{s}} & \mathbf{0}_{\mathbb{R}^{(n+m)\times(m+k)}} & \mathbf{0}_{\mathbb{R}^{(n+m)\times(n+m)}} \end{bmatrix} \begin{bmatrix} \boldsymbol{d}_{\boldsymbol{x}} \\ \boldsymbol{d}_{\boldsymbol{s}} \\ \boldsymbol{d}_{\boldsymbol{\lambda}} \\ \boldsymbol{d}_{\boldsymbol{\vartheta}} \end{bmatrix} = - \begin{bmatrix} \nabla_{(\boldsymbol{x},\boldsymbol{s})}\mathcal{L} \\ \boldsymbol{X}\boldsymbol{\vartheta} - \mu\mathbb{1} \\ \tilde{c}_s(\boldsymbol{x}, \boldsymbol{s}) \\ \boldsymbol{X} \end{bmatrix}. \tag{3.15}$$

Where, $\nabla^2_{(\boldsymbol{x},\boldsymbol{s}),\boldsymbol{x}}\mathcal{L}$ means, that first the gradient with respect to both primal variables $\boldsymbol{x}$ and $\boldsymbol{s}$ has to be computed (the same as $\nabla_{(\boldsymbol{x},\boldsymbol{s})}\mathcal{L}$) and afterwards the second derivative with respect to $\boldsymbol{x}$ has to be taken, or for $\boldsymbol{s}$, respectively. The matrix $\boldsymbol{I} \in \mathbb{R}^{(n+m)\times(n+m)}$ denotes the identity matrix. The matrices $\boldsymbol{\Theta}_{\boldsymbol{x}} \in \mathbb{R}^{(n+m)\times n}$ and $\boldsymbol{\Theta}_{\boldsymbol{s}} \in \mathbb{R}^{(n+m)\times m}$ are given

as follows:

$$\boldsymbol{\Theta}_{\boldsymbol{x}} = \begin{pmatrix} \vartheta_1 & 0 & 0 & \dots & 0 \\ 0 & \vartheta_2 & 0 & \dots & 0 \\ 0 & 0 & \vartheta_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \vartheta_n \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix},$$

$$\boldsymbol{\Theta}_{\boldsymbol{s}} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \\ \vartheta_{n+1} & 0 & 0 & \dots & 0 \\ 0 & \vartheta_{n+2} & 0 & \dots & 0 \\ 0 & 0 & \vartheta_{n+3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \vartheta_{n+m} \end{pmatrix}.$$

In a similar fashion the matrices $\boldsymbol{I}_{\boldsymbol{x}} \in \mathbb{R}^{(n+m)\times n}$ and $\boldsymbol{I}_{\boldsymbol{s}} \in \mathbb{R}^{(n+m)\times m}$ are defined, only with ones at the according positions.

The linear system given in Eq. (3.15) is often referred to as the *primal-dual system*, since it couples the descent directions of the primal variables $\left(\boldsymbol{d}_{\boldsymbol{x}}^{\top}, \boldsymbol{d}_{\boldsymbol{s}}^{\top}\right)$ and the those belonging to the dual variables $\left(\boldsymbol{d}_{\boldsymbol{\lambda}}^{\top}, \boldsymbol{d}_{\boldsymbol{\vartheta}}^{\top}\right)$. To complete a step into the determined directions, given as solutions of Eq. (3.15), the following step sizes are proposed:

$$\alpha_{\boldsymbol{s}}^{\max} = \max\left\{\alpha \in (0,1] : \boldsymbol{s} + \alpha \boldsymbol{d}_{\boldsymbol{s}} \geq (1-\tau)\,\boldsymbol{s}\right\}, \tag{3.16a}$$

$$\alpha_{\boldsymbol{\vartheta}}^{\max} = \max\left\{\alpha \in (0,1] : \boldsymbol{\vartheta} + \alpha \boldsymbol{d}_{\boldsymbol{\vartheta}} \geq (1-\tau)\,\boldsymbol{\vartheta}\right\}. \tag{3.16b}$$

Here τ is a parameter with $\tau \in (0,1)$. It is included to ensure that all variables are strictly positive, i.e., away from the boundary, hence, the name *fraction to the boundary rule* for this special kind of step size heuristic.

The Newton step in the variables is consequently given as the following update:

$$\boldsymbol{x}^{+} = \boldsymbol{x} + \alpha_{s}^{\max} \boldsymbol{d}_{\boldsymbol{x}}, \tag{3.17a}$$
$$\boldsymbol{s}^{+} = \boldsymbol{s} + \alpha_{s}^{\max} \boldsymbol{d}_{\boldsymbol{s}}, \tag{3.17b}$$
$$\boldsymbol{\lambda}^{+} = \boldsymbol{\lambda} + \alpha_{\vartheta}^{\max} \boldsymbol{d}_{\boldsymbol{\lambda}}, \tag{3.17c}$$
$$\boldsymbol{\vartheta}^{+} = \boldsymbol{\vartheta} + \alpha_{\vartheta}^{\max} \boldsymbol{d}_{\vartheta}. \tag{3.17d}$$

The last basic ingredient the IP method needs is a rule to decrease the barrier parameter μ. There are various methods on how to adjust this parameter throughout the optimization procedure. Since the algorithmic implementation given by Wächter and Biegler (2006) is used in this work, the rule applied is given by

$$\mu_{j+1} = \max\left\{\frac{\epsilon_{\text{tol}}}{10}, \min\left\{\kappa_{\mu}\mu_{j}, \mu_{j}^{\theta_{\mu}}\right\}\right\}, \tag{3.18}$$

with user provided constants $\kappa_{\mu} \in (0,1)$, $\epsilon_{\text{tol}} \in \mathbb{R}$, and $\theta_{\mu} \in (1,2)$. The small number ϵ_{tol} is the solution tolerance of the overall problem. A problem is regarded as solved, if the condition

$$\max\left\{\left\|\nabla \mathcal{L}\right\|, \left\|\boldsymbol{X}\boldsymbol{\vartheta}\right\|, \left\|\tilde{c}_{s}\left(\boldsymbol{x}, \boldsymbol{s}\right)\right\|\right\} \leq \epsilon_{\text{tol}} \tag{3.19}$$

is met—note, that this comes form setting $\mu = 0$ in the KKT conditions Eq. (3.14)—, for some vector norm $\|.\|$.

The barrier parameter is not adjusted in every step: The sub problems generated by this procedure are solved for fixed μ_j, $j \in \mathbb{N}$ to a certain tolerance (not ϵ_{tol}), then μ_{j+1} is updated according to Eq. (3.18).

With the above discussion a simple inner point algorithm can be stated as follows:

Algorithm 3.3.1

Choose initial primal points $\boldsymbol{x}_0, \boldsymbol{s}_0$ and initial dual points $\boldsymbol{\lambda}_0, \boldsymbol{\vartheta}_0$. Set the constants $\epsilon_{\text{tol}}, \kappa_{\mu}, \theta_{\mu}$, and τ. Select an initial barrier parameter $\mu_0 > 0$ and set $k = 0$.

1. Calculate a directions $(\boldsymbol{d}_{\boldsymbol{x}}, \boldsymbol{d}_{\boldsymbol{s}}, \boldsymbol{d}_{\boldsymbol{\lambda}}, \boldsymbol{d}_{\vartheta})$ by solving Eq. (3.15).
2. Perform a step according to Eq. (3.17).
3. If the sub problem, with the computed values, is solved to the desired tolerance, decrease μ_k according to Eq. (3.18), else set $\mu_{k+1} = \mu_k$ and set $k = k + 1$.
4. If condition Eq. (3.19) holds for $\mu = 0$, **STOP**.

Further, to prevent an infinite loop, a $k_{\max} \in \mathbb{N}$ can be set as the maximum number of iterations that should be performed.

Note, that Algorithm 3.3.1 represents the simplest possible version of the IP scheme. A robust and fast implementation of an IP algorithm requires a lot more effort and its discussion is out of the scope of this work. For more details refer to the work of Wächter and Biegler (2006). However, even for this rather simple algorithm the following convergence theorem holds:

Theorem 3.3.2
Assume Algorithm 3.3.1 does not loop infinitely long. Let $\{x_k\}$ be the sequence generated by Algorithm 3.3.1 for $\mu_k \to 0$ and let the functions f, g and h be twice continuously differentiable. Then all limit points $\hat{x}$ of $\{x_k\}$ are feasible points. If, further, any limit point $\hat{x}$ satisfies the LICQ, then the first order optimality conditions for Problem (3.10) hold at $\hat{x}$.

Proof The proof is presented in (Nocedal and Wright, 2006, Chap. 19, Theorem 19.1). A slightly weaker result will be proofed in the next section, concerning the KKT conditions of the various reformulations of Problem (3.10) needed during the IP scheme. ■

3.4. On the Necessary Conditions of Various Reformulations of a Constrained Optimization Problem

The following results were published earlier in Rozgič and Stiemer (2014b).

For convenience and readability once again the prototype of a constrained optimization problem, Problem (3.1), is recalled. For $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^k$ being twice continuously differentiable functions it reads as:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x), \\ \text{subject to} \quad & g(x) \leq 0, \\ & h(x) = 0. \end{aligned}$$

The variable $x \in \mathbb{R}^n$ is decomposed as $x = x^+ - x^-$, where $x^+, x^- \in \mathbb{R}^n$ are component wise non-negative vectors, further as in Problem (3.11) a slack variable $s \in \mathbb{R}^m$ is introduced to eliminate the, possibly non-linear, inequality constraints given by g.

Together this yields:

$$\begin{aligned} \min_{x^+, x^- \in \mathbb{R}^n, s \in \mathbb{R}^m} \quad & \tilde{f}\left(x^+, x^-, s\right), \\ \text{subject to} \quad & \tilde{c}_s\left(x^+, x^-, s\right) = \mathbf{0}, \\ & x^+, x^-, s \geq \mathbf{0}. \end{aligned} \tag{3.20}$$

Similar to the discussion made in Section 3.3 the participating functions $\tilde{f} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, and $\tilde{c}_s : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{m+k}$ are defined by

$$\begin{aligned} \tilde{f}\left(x^+, x^-, s\right) &= f\left(x^+ - x^-\right), \\ \tilde{c}_s\left(x^+, x^-, s\right) &= \begin{pmatrix} g\left(x^+ - x^-\right) + s \\ h\left(x^+ - x^-\right) \end{pmatrix}. \end{aligned}$$

The logarithmic barrier problem to Problem (3.20) is given by the optimization program

$$\begin{aligned} \min_{x^+, x^- \in \mathbb{R}^n, s \in \mathbb{R}^m} \quad & \tilde{f}\left(x^+, x^-, s\right) - \mu \left(\sum_{i=1}^{n} \ln\left(x_i^+\right) + \sum_{i=1}^{n} \ln\left(x_i^-\right) + \sum_{i=1}^{m} \ln\left(s_i\right) \right), \\ \text{subject to} \quad & \tilde{c}_s\left(x^+, x^-, s\right) = \mathbf{0}. \end{aligned} \tag{3.21}$$

Here $\mu \in \mathbb{R}$, $\mu \geq 0$ is the barrier parameter. This is the aforementioned problem to which Eq. (3.14) are the KKT conditions, as will also be seen momentarily.

The Lagrangian of Problem (3.20) is given by

$$\mathcal{L}\left(x^+, x^-, s, \lambda, \vartheta\right) = \tilde{f}\left(x^+, x^-, s\right) + \lambda^\top \left(\tilde{c}_s\left(x^+, x^-, s\right)\right) - \vartheta^\top \begin{pmatrix} x^+ \\ x^- \\ s \end{pmatrix}, \tag{3.22}$$

with Lagrangian multipliers $\lambda \in \mathbb{R}^{m+k}$ and $\vartheta \in \mathbb{R}^{\tilde{m}}$, where $\tilde{m} = n + n + m$. Further the multipliers belonging to the respective contributions are denoted by $\lambda_g, \vartheta_s \in \mathbb{R}^m$, $\lambda_h \in \mathbb{R}^k$, $\vartheta_{x^+}, \vartheta_{x^-} \in \mathbb{R}^n$, yielding $\lambda = (\lambda_g^\top, \lambda_h^\top)^\top$ and $\vartheta = \left(\vartheta_{x^+}^\top, \vartheta_{x^-}^\top, \vartheta_s^\top\right)^\top$. The gradient of the Lagrangian with respect to the primal variables x^+, x^-, s, is for simplicity denoted by $\nabla\mathcal{L}\left(x^+, x^-, s, \lambda, \vartheta\right)$, is used to define the parametric function $\Psi_\mu : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{2\tilde{m}+m+k}$ by

$$\Psi_\mu\left(x^+, x^-, s\right) = \begin{pmatrix} \nabla\mathcal{L}\left(x^+, x^-, s, \lambda, \vartheta\right) \\ \tilde{c}_s\left(x^+, x^-, s\right) \\ X\vartheta - \mu\mathbb{1} \end{pmatrix}, \tag{3.23}$$

where $X \in \mathbb{R}^{\tilde{m} \times \tilde{m}}$ is a diagonal matrix with the target variables of Problems (3.20) and (3.21) on the diagonal, $X = \operatorname{diag}\left(x^+, x^-, s\right)$ and $\mathbb{1} \in \mathbb{R}^{\tilde{m}}$ is again a vector of ones of appropriate size. With these declarations the following theorem can be formulated:

Theorem 3.4.1
Let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^k$ be continuously differentiable functions. Let further $\Psi_\mu : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{2\tilde{m}+m+k}$ be defined as in Eq. (3.23). Then the following holds:

a) Let $\Psi_\mu \left(x^+, x^-, s\right) = \mathbf{0}$ and $\left(x^+, x^-, s\right)^\top > \mathbf{0}$ hold, then $\left(x^+, x^-, s\right)^\top$ is a KKT point of the barrier Problem (3.21).

b) Let $\Psi_{\mu=0} \left(x^+, x^-, s\right) = \mathbf{0}$ and $\vartheta \geq \mathbf{0}$ as well as $\left(x^+, x^-, s\right)^\top \geq \mathbf{0}$ hold, then $\left(x^+, x^-, s\right)^\top$ is a KKT point of Problem (3.20) with multipliers (λ, ϑ).

c) Let $\left(x^+, x^-, s\right)^\top$ be a KKT point of Problem (3.20) with multipliers $(\lambda, \vartheta)^\top$, then $x = x^+ - x^-$ is a KKT point of Problem (3.1) with multiplier λ.

Proof a) The KKT conditions of the barrier Problem (3.21), are given by the stationarity condition

$$\nabla \tilde{f} \left(x^+, x^-, s\right) - \mu X^{-1} \mathbb{1} + J_{\tilde{c}_s}^\top \lambda = \mathbf{0}, \tag{3.24}$$

with the Jacobian of $\tilde{c}_s \left(x^+, x^-, s\right)$ denoted by $J_{\tilde{c}_s}$, and the primal feasibility condition

$$\tilde{c}_s \left(x^+, x^-, s\right) = \mathbf{0}. \tag{3.25}$$

Since the prerequisite $\Psi_\mu \left(x^+, x^-, s\right) = \mathbf{0}$ equals

$$\begin{pmatrix} \nabla \mathcal{L} \left(x^+, x^-, s, \lambda, x^+\right) \\ \tilde{c}_s \left(x^+, x^-, s\right) \\ X\vartheta - \mu \mathbb{1} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} \tag{3.26}$$

primal feasibility is fulfilled. The matrix X is regular since $\left(x^+, x^-, s\right)$ are strictly away from zero. Therefore ϑ can be written as

$$\vartheta = \mu X^{-1} \mathbb{1}. \tag{3.27}$$

This identity is plugged into the first line of Eq. (3.26) yielding

$$\nabla \tilde{f} \left(x^+, x^-, s\right) - \mu X^{-1} \mathbb{1} + J_{\tilde{c}_s}^\top \lambda = \mathbf{0}, \tag{3.28}$$

which is the desired stationarity condition.

b) Let (x^+, x^-, s) be a KKT point of Problem (3.20), then $(x^+, x^-, s, \lambda, \vartheta)$ fulfil

$$\nabla \tilde{f}\left(x^+, x^-, s\right) + J_{\tilde{c}_s}^\top \lambda - \vartheta = 0, \tag{3.29a}$$
$$\tilde{c}_s\left(x^+, x^-, s\right) = 0, \tag{3.29b}$$
$$x^+, x^-, s \geq 0, \tag{3.29c}$$
$$X\vartheta = 0, \tag{3.29d}$$
$$\vartheta \geq 0. \tag{3.29e}$$

The equations

$$\Psi_{\mu=0}\left(x^+, x^-, s\right) = \begin{pmatrix} \nabla \mathcal{L}\left(x^+, x^-, s, \lambda, \vartheta\right) \\ \tilde{c}_s\left(x^+, x^-, s\right) \\ X\vartheta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{3.30}$$

yield the stationarity Eq. (3.29a), primal feasibility Eq. (3.29b) and complementary slackness conditions Eq. (3.29d) of Problem (3.20). Dual feasibility Eq. (3.29e) holds, since $\vartheta \geq 0$ is prescribed. Further Eq. (3.29c) follows since $x^+, x^-, s \geq 0$ is assumed.

c) The KKT conditions at a point $x^* \in \mathbb{R}^n$ for Problem (3.1) are given by

$$\nabla f\left(x^*\right) + \nabla g\left(x^*\right)^\top \lambda_g + \nabla h\left(x^*\right)^\top \lambda_h = 0, \tag{3.31a}$$
$$g\left(x^*\right) \leq 0, \tag{3.31b}$$
$$h\left(x^*\right) = 0, \tag{3.31c}$$
$$G\left(x^*\right) \lambda_g = 0, \tag{3.31d}$$
$$\lambda_g \geq 0, \tag{3.31e}$$

where $G\left(x^*\right) = \operatorname{diag}\left(g\left(x^*\right)\right) \in \mathbb{R}^{m \times m}$ is the diagonal matrix with the components of the inequality constraint evaluated at x^* on its diagonal. To prove that indeed the KKT conditions Eq. (3.31) of Problem (3.1), hold the components of the equations given in Eq. (3.29) have to be investigated closer. The gradient of the objective of Problem (3.20) in Eq. (3.29a) is given as

$$\nabla \tilde{f}\left(x^+, x^-, s\right) = \begin{pmatrix} \nabla f\left(x\right) \\ -\nabla f\left(x\right) \\ 0_{\mathbb{R}^m} \end{pmatrix}, \tag{3.32}$$

where $\nabla f(\boldsymbol{x})$ is the gradient of the objective in Problem (3.1) and $\mathbf{0}_{\mathbb{R}^m}$ is the vector of zeros in $\mathbb{R}^m$. The second contribution in Eq. (3.29a) is given by

$$\boldsymbol{J}_{\tilde{c}_s}^{\top}\boldsymbol{\lambda} = \begin{pmatrix} \nabla g_1(\boldsymbol{x}) & \dots & \nabla g_m(\boldsymbol{x}) & \nabla h_1(\boldsymbol{x}) & \dots & \nabla h_k(x) \\ -\nabla g_1(\boldsymbol{x}) & \dots & -\nabla g_m(\boldsymbol{x}) & -\nabla h_1(\boldsymbol{x}) & \dots & -\nabla h_k(\boldsymbol{x}) \\ \boldsymbol{e}_1 & \dots & \boldsymbol{e}_m & \mathbf{0}_{\mathbb{R}^m} & \dots & \mathbf{0}_{\mathbb{R}^m} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \\ \lambda_{m+1} \\ \vdots \\ \lambda_{m+k} \end{pmatrix},$$

where $\boldsymbol{e}_j$ corresponds to the $j-$th unit vector in $\mathbb{R}^m$, and $\nabla g_j(\boldsymbol{x})$ denotes the gradient of the $j-$th inequality constraint function in Problem (3.1), respectively $\nabla h_i(\boldsymbol{x})$ denotes the gradient of the $i-$th equality constraint function. The first line of above equation can be written in more detail as

$$\begin{aligned} &\begin{pmatrix} \nabla g_1(\boldsymbol{x}) & \dots & \nabla g_m(\boldsymbol{x}) & \nabla h_1(\boldsymbol{x}) & \dots & \nabla h_k(\boldsymbol{x}) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \\ \lambda_{m+1} \\ \vdots \\ \lambda_{m+k} \end{pmatrix} \\ &= \begin{pmatrix} (\nabla g_1(\boldsymbol{x}))_1 & \dots & (\nabla g_m(\boldsymbol{x}))_1 & (\nabla h_1(\boldsymbol{x}))_1 & \dots & (\nabla h_k(\boldsymbol{x}))_1 \\ (\nabla g_1(\boldsymbol{x}))_2 & \dots & (\nabla g_m(\boldsymbol{x}))_2 & (\nabla h_1(\boldsymbol{x}))_2 & \dots & (\nabla h_k(\boldsymbol{x}))_2 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ (\nabla g_1(\boldsymbol{x}))_n & \dots & (\nabla g_m(\boldsymbol{x}))_n & (\nabla h_1(\boldsymbol{x}))_n & \dots & (\nabla h_k(\boldsymbol{x}))_n \\ \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \\ \lambda_{m+1} \\ \vdots \\ \lambda_{m+k} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^{m} (\nabla g_i(\boldsymbol{x}))_1 \lambda_i + \sum_{j=1}^{k} (\nabla h_j(\boldsymbol{x}))_1 \lambda_{m+j} \\ \vdots \\ \sum_{i=1}^{m} (\nabla g_i(\boldsymbol{x}))_n \lambda_i + \sum_{j=1}^{k} (\nabla h_j(\boldsymbol{x}))_n \lambda_{m+j} \end{pmatrix} = \boldsymbol{J}_c^{\top}\boldsymbol{\lambda}, \end{aligned}$$

where $\boldsymbol{J}_c$ denotes the Jacobian of the function $c : \mathbb{R}^n \to \mathbb{R}^{m+k}$, given by

$$c(\boldsymbol{x}) = \begin{pmatrix} g(\boldsymbol{x}) \\ h(\boldsymbol{x}) \end{pmatrix}.$$

Note that c is representing the constraints of Problem (3.1). The product $J_{\tilde{c}_s}^\top \boldsymbol{\lambda}$ in Eq. (3.29a) can therefore be written as

$$J_{\tilde{c}_s}^\top \boldsymbol{\lambda} = \begin{pmatrix} J_c^\top \boldsymbol{\lambda} \\ -J_c^\top \boldsymbol{\lambda} \\ \boldsymbol{M}\boldsymbol{\lambda} \end{pmatrix}, \tag{3.33}$$

where the matrix $\boldsymbol{M} \in \mathbb{R}^{m \times m+k}$ is given as

$$M = \begin{pmatrix} \boldsymbol{e}_1 & \dots & \boldsymbol{e}_m & \mathbf{0}_{\mathbb{R}^m} & \dots & \mathbf{0}_{\mathbb{R}^m} \end{pmatrix}.$$

Finally Eq. (3.29a) can be written as

$$\nabla \tilde{f}\left(\boldsymbol{x}^+, \boldsymbol{x}^-, \boldsymbol{s}\right) + J_{\tilde{c}_s}^\top \boldsymbol{\lambda} - \boldsymbol{\vartheta} = \begin{pmatrix} \nabla f\left(\boldsymbol{x}\right) \\ -\nabla f\left(\boldsymbol{x}\right) \\ \mathbf{0}_{\mathbb{R}^m} \end{pmatrix} + \begin{pmatrix} J_c^\top \boldsymbol{\lambda} \\ -J_c^\top \boldsymbol{\lambda} \\ \boldsymbol{M}\boldsymbol{\lambda} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\vartheta}_{\boldsymbol{x}^+} \\ \boldsymbol{\vartheta}_{\boldsymbol{x}^-} \\ \boldsymbol{\vartheta}_{\boldsymbol{s}} \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{\mathbb{R}^n} \\ \mathbf{0}_{\mathbb{R}^n} \\ \mathbf{0}_{\mathbb{R}^m} \end{pmatrix}.$$

From this the stationarity condition Eq. (3.31a) for Problem (3.1) can be concluded, since from

$$\nabla f\left(\boldsymbol{x}\right) + J_c^\top \boldsymbol{\lambda} = \boldsymbol{\vartheta}_{\boldsymbol{x}^+} \geq \mathbf{0}$$

and

$$\nabla f\left(\boldsymbol{x}\right) + J_c^\top \boldsymbol{\lambda} = -\boldsymbol{\vartheta}_{\boldsymbol{x}^-} \leq \mathbf{0}$$

the equation

$$\nabla f\left(\boldsymbol{x}\right) + J_c^\top \boldsymbol{\lambda} = \mathbf{0} \tag{3.34}$$

follows. Further, from $\boldsymbol{M}\boldsymbol{\lambda} = \boldsymbol{\lambda}_g$ the identity $\boldsymbol{\lambda}_g = \boldsymbol{\vartheta}_{\boldsymbol{s}}$ can be deduced, and, hence, $\boldsymbol{\lambda}_g \geq 0$ holds, which is the dual feasibility condition, Eq. (3.31e), of Problem (3.1). To obtain the complementary slackness condition, Eq. (3.31d), Eq. (3.29d) is exploited:

$$\begin{aligned} \boldsymbol{X}\boldsymbol{\vartheta} = 0 \Rightarrow s_i \vartheta_j = 0 &\underset{g(\boldsymbol{x})+\boldsymbol{s}=0}{\Longleftrightarrow} \left(-g\left(\boldsymbol{x}\right)\right)_i \vartheta_j = 0 \\ \underset{\boldsymbol{\lambda}_g = \boldsymbol{\vartheta}_{\boldsymbol{s}}}{\Longleftrightarrow} \left(g\left(\boldsymbol{x}\right)\right)_i \boldsymbol{\lambda}_i = 0 &\Longleftrightarrow \boldsymbol{G}\left(\boldsymbol{x}\right) \boldsymbol{\lambda}_g = \mathbf{0}, \end{aligned}$$

for all $i = 1, \dots, m$ and $j = 2n+1, \dots, \widetilde{m}$. The primal feasibility conditions Eq. (3.31b) and Eq. (3.31c) for Problem (3.1) can be deduced by exploiting Eq. (3.29b) and Eq. (3.29c). This concludes the proof. ■

As pointed out in Nocedal and Wright (2006) the inner point method and finding roots of the parametric function Ψ_μ combined with reducing μ to 0 are equivalent. With this in mind it can be concluded that solutions found by the inner point method are at least KKT-points and equipped with a constraint qualification, for example the LICQ, therefore fulfill the necessary optimality conditions for non-linear constrained optimization problems.

4

Industrial Process Parameter Optimization in Coupled Sheet Metal Applications

Sheet metal forming as an industrial process is important in numerous applications where the automotive industry is to be named as probably the biggest beneficiary of this technology. Although one might have an intuitive idea about forming, it becomes vital to study a particular sheet metal forming process systematically in order to optimize it. Regardless of the special process, Duncan and Panton propose the following four common points in their overview in Bhattacharya (1997):

1. The sheet is deformed by traction transmitted through the sheet.
2. Of all the principle stresses, at least one is a tensile stress, which means that the stress leads to expansion into that principle direction.
3. The stresses belonging to the thickness of the sheet are generally zero, therefore a state of plane stress exists.
4. The whole forming process is limited by wrinkling or necking.

Above points are sufficient to describe most sheet metal forming processes, such as deep drawing or stamping. The points 1 – 3 refer to the force and therefore stress situation. Traction yields tensile stress and tensile expansion yields plane stress.

Optimization of industrial processes usually aims for cost reduction and sustainability. Both can be reached, if the parameters defining the process are adjusted carefully. To efficiently explore the space of possible configurations virtual process design is a suitable tool, reducing the number of costly experiments. But by introducing virtual process design, new problems have to be accounted for. The simulations have to match

the physical reality as good as possible, which mostly depends on the choice of the used material model. Numerical challenges, like time and space discretization have to be considered as well as the computation time. In this particular situation, where virtual design is used to optimize a coupled forming process, different physical equations have to be coupled, resulting in a multi-physics simulation.

As commercial tools like LS-DYNA became more accessible and are under permanent development, due to the growing demand of simulation driven innovation, they also became a suitable tool to face the numerical challenges. However, the identification of optimal process parameters requires a deeper understanding of the process and its modeling. The basic mechanics and physics of a sheet metal forming process are discussed during the course of Section 4.1. In Sections 4.2 and 4.3 a unified approach to determine a set of optimal parameters for the governing material model as well as a set of optimal forming process parameters is introduced.

4.1. A Brief Review of the Mechanics of Sheet Metal Forming

In continuum mechanics the deformation of a body is described in terms of stresses and strains. Stresses are linked to the forces that cause the deformation of a body. These forces may be external loads or body forces. Strains describe the deformations in relative terms, usually in terms of displacement from the reference configuration of the body, wich is often assumed to be the unperturbed state. For a more comprehensive and mathematical definition of the used terminology, the behavior of material under load will be examined by the example of infinitesimal, linear strain theory. The basic notions and concepts are the same as in more complex material models, such as finite strain models. Nevertheless, the linear theory offers an easy way to define the used tensors and introduce the wording. It will also be exploited in the second half of this work to establish a novel way of tackling process optimization problems.

In Section 4.1.2 the experimental determination of material properties via uniaxial tensile tests is covered. These kind of tests are used to establish a physically valid source of information to which a tailored material model is fitted, by means of simulation based optimization. Within this work damage is accounted for via a Lemaitre-type parametric model, where the parameters have to be determined via comparison to experiments. In experiments damage is usually described through forming limit curves, discussed in Section 4.1.3. Finally, electromagnetic forming processes are outlined in Section 4.1.4.

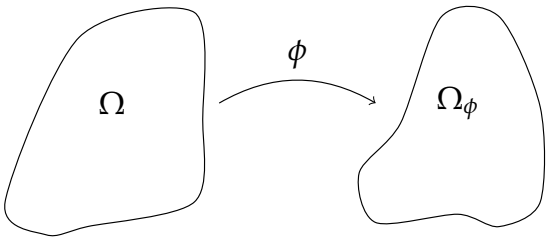


Figure 4.1.: The domain Ω is deformed by the mapping ϕ into the domain Ω_ϕ.

4.1.1. The Linear Theory

Consider a resting body B occupying a domain $\Omega \subset \mathbb{R}^d$, with $d \in \{2,3\}$. In the following the deformation of B is investigated. Materials that allow for self-penetration are excluded and only small deformations are taken into account.

Let $\phi : \Omega \to \mathbb{R}^d$ be a differentiable mapping, describing the deformation of B as follows:

$$\phi(z) = z + u(z). \tag{4.1}$$

Further, $\det(\nabla\phi) \neq 0$ is assumed, since deformations are assumed to map material points one to one on material points. The mapping $u : \Omega \to \mathbb{R}^d$ denotes the displacement of a point $z \in \Omega$. To study the local action of ϕ, Taylor's theorem is used, yielding

$$\phi(z+y) - \phi(z) = \nabla\phi(z)y + o(\|y\|), \text{ for } \|y\| \to 0. \tag{4.2}$$

The Euclidean distance between $\phi(z+y)$ and $\phi(z)$ is according to Eq. (4.2) given by

$$\begin{aligned} \|\phi(z+y) - \phi(z)\|^2 &= \|\nabla\phi(z)y\|^2 + o\left(\|y\|^2\right) \\ &= y^\top \nabla\phi(z)^\top \nabla\phi(z)y + o\left(\|y\|^2\right), \text{ for } \|y\| \to 0. \end{aligned} \tag{4.3}$$

If the distance between the inspected points is small, the main contribution of the deformation is given by $\nabla\phi(z)^\top\nabla\phi(z)$, as can be concluded from Eq. (4.3). Consequently, the Cauchy-Green-strain tensor is defined as:

$$V := \frac{1}{2}(\nabla\phi^\top\nabla\phi - I), \tag{4.4}$$

where I denotes the identity in $\mathbb{R}^d$. It describes the geometrical action of the deformation. In particular, the Cauchy-Green strain tensor for Eq. (4.1) reads as

$$V_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial z_j} + \frac{\partial u_j}{\partial z_i}\right) + \frac{1}{2}\sum_{k=1}^{d}\left(\frac{\partial u_k}{\partial z_i}\cdot\frac{\partial u_k}{\partial z_j}\right), \tag{4.5}$$

for $i, j = 1, \ldots, d$. In infinitesimal strain theory only small deformations are considered, therefore the non-linear contributions in Eq. (4.5) can be neglected. In finite strain theory the non-linear contributions are accounted for. Taking only the linear contributions into account, the Cauchy-Green strain tensor reduces to:

$$\varepsilon(u) := \frac{1}{2}\left(\nabla u + \nabla u^{\top}\right). \tag{4.6}$$

The stresses within the body under deformation can be described by the Cauchy stress tensor σ. The existence of σ follows from Cauchy's Axiom, also known as the balance of internal and external forces and is not discussed in more detail at this point, see Ciarlet (1988) for an exhaustive treatment of this topic.

Both tensors are second order tensors and can conveniently be identified with 3×3 matrices, after an appropriate basis is chosen:

$$\sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix}, \; \varepsilon = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{pmatrix}. \tag{4.7}$$

The eigenvalues of these tensors are called principle components, whereas the corresponding eigenvectors are the principle directions. When considering a small cubic element, with its deformed state as shown in Fig. 4.2, the principle strain increments are given by:

$$\Delta\varepsilon_a = \frac{\Delta a}{a}, \; \Delta\varepsilon_b = \frac{\Delta b}{b}, \; \Delta\varepsilon_c = \frac{\Delta c}{c}. \tag{4.8}$$

Stresses and strains can be linked by so called constitutive equations. These are equations that relate material specific physical quantities (like stress, strain, Young's modulus, etc.) to the material's response to external excitation (like external force fields, temperature fields, electric fields, etc.). Hooke's law is one of the most famous examples of a constitutive equation. It relates the stress and strain tensor by a fourth order tensor S:

$$\varepsilon = \mathsf{S}\sigma. \tag{4.9}$$

The *compliance tensor* S consists of $3^4 = 81$ material constants. By taking into account isotropic and homogeneous materials only, the total number of material constants reduces to 2, namely the Lamé constants λ and μ (shear modulus). They are related to the elastic (Young's) modulus, E, and Poisson's ratio, ν, as follows:

$$\begin{aligned} \lambda &= \frac{\nu E}{(1+\nu)(1-2\nu)}, \\ \mu &= \frac{E}{2(1+\nu)}. \end{aligned} \tag{4.10}$$

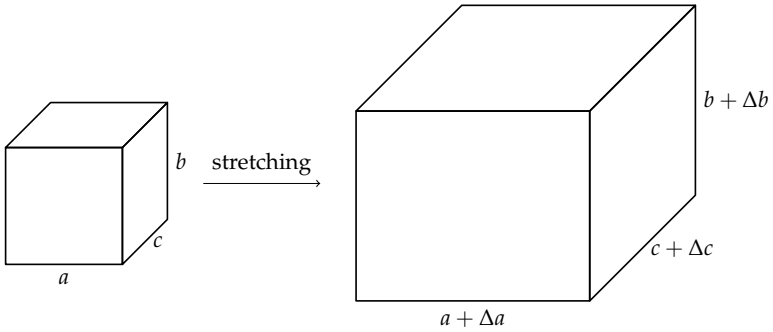


Figure 4.2.: From a reference cube (left) the principle strain can be computed from the change of the three principle components of the stretched cube (right).

Note that ν has no unit and E, λ, and μ have units of pressure, i.e., $\mathrm{Pa} = \mathrm{N\,m^{-2}}$. Szekeres (1999) gives a nice historical view on Hooke's law.

A special stress situation arises, when the stress contribution across a single surface of the material is significantly smaller than the stresses in the remaining planes. This happens for example, when a non torsion load acts on a thin but long sheet. This case is denoted as plane stress. The stress tensor simplifies, since one of the principle stresses is assumed to be zero. Let w.l.o.g. this principle stress be σ_3. The full plane stress tensor in matrix representation reads:

$$\sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & 0 \\ \sigma_{yx} & \sigma_{yy} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{4.11}$$

it can be reduced to a 2×2 matrix, by omitting the zero lines and columns

$$\sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix}. \tag{4.12}$$

A similar notion is connected to the strain tensor, if the length change in one particular direction is much smaller then in the other, for example the thickness change under compression for a very thick structure. This situation is called plane strain, and the strain tensor can be reduced just like the stress tensor in the case of plane stress. However, the plane stress regime is of bigger interest for this work, since in sheet metal forming the metal sheets under consideration are assumed to be thin and the acting loads are not used to cause torsion.

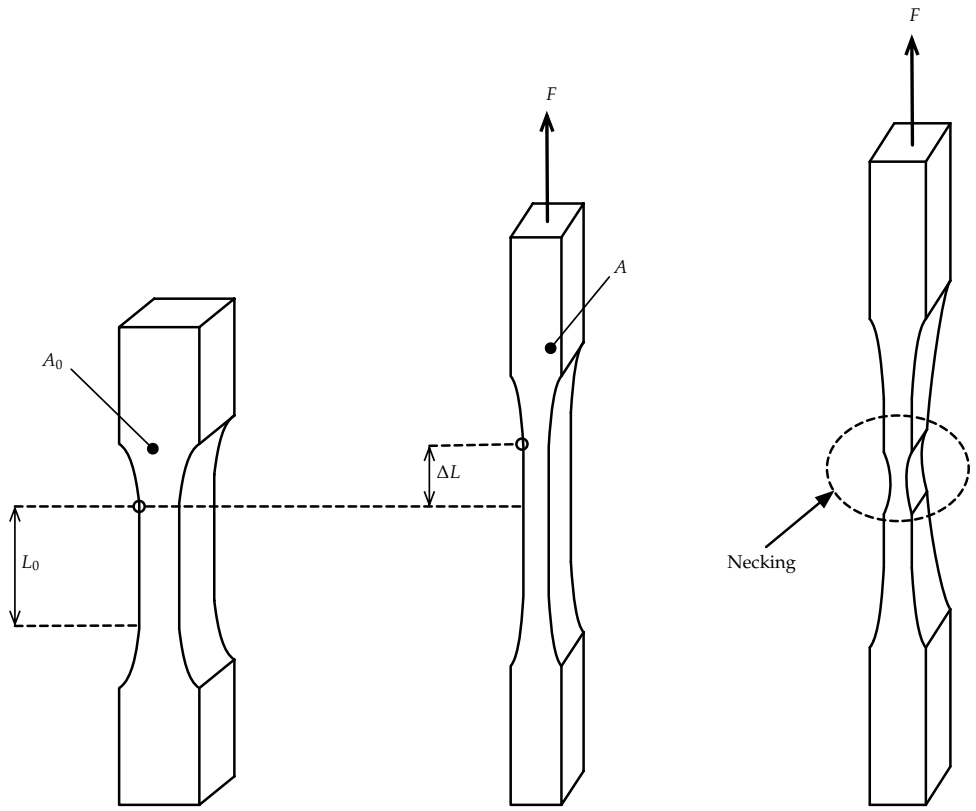


Figure 4.3.: Schematic of a tensile test specimen. The specimen has a defined surface A_0 and gauge length L_0. Typically it is fixed at one end an a force is applied at the other, stretching the specimen. The thickness change, the elongation of the gauge length (ΔL) and the change of the cross section (A) are recorded and used for the calculation of the strain and stress curve.

4.1.2. Material Properties by Uniaxial Tensile Testing

Stresses and strains, as well as other material properties, such as the bulk and shear moduli, are usually determined by experiments. A convenient way to obtain the information about the basic material properties that are important for sheet metal forming are strain vs stress curves, like in Fig. 4.4. Although stress and strain are tensors, they can be reduced to one dimensional quantities in certain load scenarios. Uniaxial tensile tests are such a scenario.

In a uniaxial tensile test, the surface A_0 and the gauge length L_0 of the test specimen

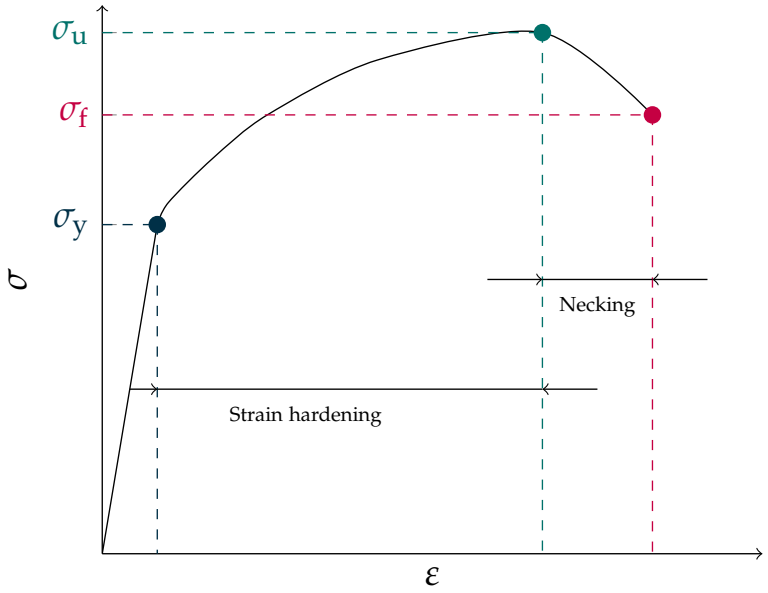


Figure 4.4.: Typical stress vs strain curve. The areas where strain hardening takes place as well as the necking region are displayed along with the special stress points σ_y, σ_u and σ_f.

are previously known. The typical situation is depicted in Fig. 4.3. A load F is applied into exactly one direction, yielding an expansion towards that direction. The volume does not change, but the gauge length increases, and the surface changes. The tensile load is increased until the specimen fails. From the extension of the gauge length, ΔL, and the applied force F, the engineering strain and engineering stress can be calculated, they read:

$$\begin{aligned} \varepsilon_{en} &= \frac{\Delta L}{L_0}, \\ \sigma_{en} &= \frac{F}{A_0}. \end{aligned} \tag{4.13}$$

During the test σ_{en} and ε_{en} are computed and can then be represented in a strain vs stress curve, which offers some basic material information. The first part of the curve given in Fig. 4.4 characterizes the elastic part, i.e., the region where the stress strain relation is linear. This region covers all stress up to σ_y, which is called the yield stress or yield strength. From the rise of the linear part, the elastic modulus E can be deduced. Ductile materials, such as forged metals, exhibit strain hardening. During that phase yield is enhanced, but imperfections are evolving in the crystal lattice. When the ultimate strength (also the terminology tensile strength is used), or

ultimate stress σ_u of a material is reached, the material's ability to respond to stress by yield is exhausted. From this point on material failure is likely to occur when force is applied. This behavior is called necking, see Fig. 4.3 for a typical necking during tensile tests. During necking lower stresses are required to elongate the material as before. Eventually, the point of material failure is reached, at σ_f the material fails, and damage occurs.

When using engineering stress, the stress state is not represented in terms of deformed structure, but in terms of the reference structure A_0.The so called *true stress* relates to the deformed state and is given by

$$\sigma_t = \frac{F}{A}, \tag{4.14}$$

for the current cross sectional area A. While σ_{en} is decaying during the necking phase, as is seen in Fig. 4.4, true stress σ_t increases until the specimen fails. Analogous to the true stress the quantity true strain is used. It is given by:

$$\varepsilon_t = \ln\left(\frac{L_0 + \Delta L}{L_0}\right). \tag{4.15}$$

The engineering curve is usually used for material characterization, whereas the true strain stress curve is used in the context of simulation and modeling, since the data relates to the momentary configuration.

4.1.3. Characterization of the Forming Limit

To quantify the limitations of sheet metal forming the so called forming limit curve (FLC), sometimes also referred to as the forming limit diagram (FLD) is used. In an FLC basically the first two principle strains, i.e., the largest and second largest eigenvalue of the relative, i.e., true strain tensor, w.l.o.g. ε_1 and ε_2, are mapped against each other for proportional strain paths. For those strain paths the materials response to load is monitored until necking occurs. This concept of material failure characterization in the space of strains was first introduced by Keeler and Backofen (1963) and Goodwin (1968). A schematic of an FLC and the typical strain paths is depicted in Fig. 4.5. For positive ε_2 the tension-tension load regime holds, for negative ε_2 the governing load situation is tension-compression. The zone beneath the FLC is called the safe zone, since no material failure should occur there, above the FLC damage is likely to happen. FLCs are used to control and characterize sheet metal forming processes, by analyzing the typical strain paths and making sure that the major- minor strain pairs stay below the limiting curve.

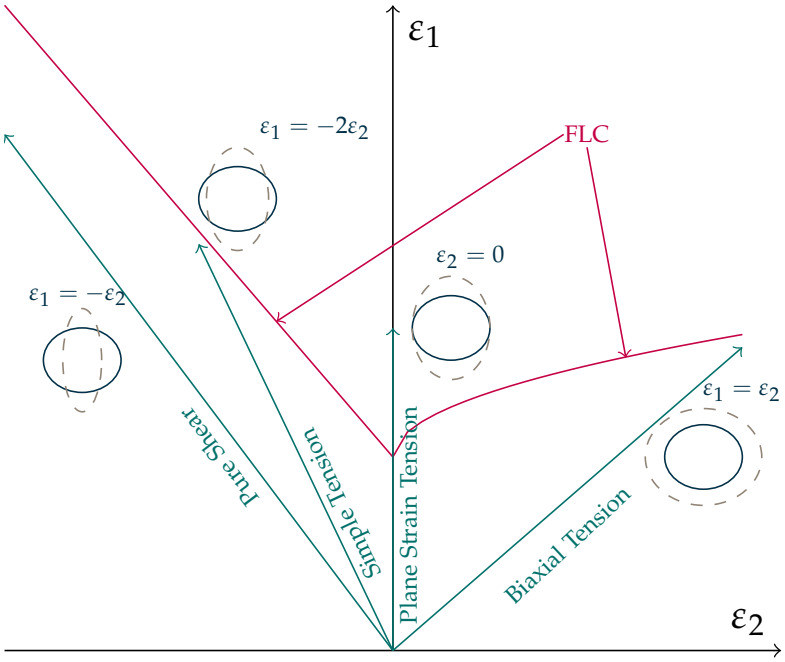


Figure 4.5.: A typical forming limit curve (red). The green arrows represent the proportional strain path, together with a pictogram of the corresponding major and minor strain relation.

As argued by Stoughton (2000) the FLC strongly depend on the use of proportional strain paths, to overcome this limitation they proposed the use of stress based FLCs. These are path independent and can be exploited in numerical investigations, as shown by Alsos et al. (2008). However, strain based FLC are still widely used, and the definition of a common ground is still subject of active research, as can be seen in a later work by Stoughton and Yoon (2012).

The experimental determination of FLCs is usually performed by testing so called Nakazima test specimens as shown in Fig. 4.6. These specimens have a special predefined geometry and their outward surface is marked with circles. From the deformation of the circles into ellipses, the principle strains ε_1 and ε_2 can be determined. Similar to the characterization of damage by FLCs the experimental determination is subject to active investigation, see e.g., Karadogan and Tamer (2015). Especially for higher strain rates the experimental setup is not entirely clear. Advances in this area have been made in recent works by Kiliclar et al. (2012b).

More details on strain and stress based FLCs and their respective experimental and numerical treatments can be found in the monograph by Banabic (2010), and in research

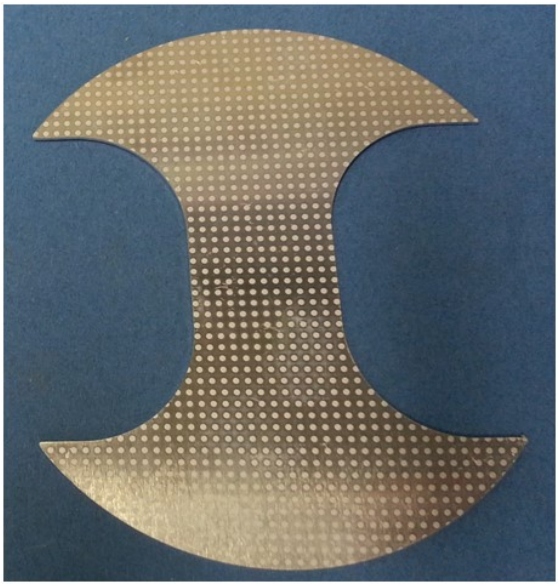


Figure 4.6.: A typical Nakazima test specimen for the determination of FLCs. Small circles are attached to the surface, after deformation the circles turn to ellipses characterizing ε_1 and ε_2. The picture was taken at the "Institut für Werkstoffkunde (IW)", Hannover.

articles form Butuc et al. (2006), Banabic et al. (2013), Hasan et al. (2011) and Wu et al. (2005), to name only a small selection.

4.1.4. Electromagnetic Forming

A special class of sheet metal forming processes are the electromagnetic forming (EMF) processes. There forming is achieved by Lorentz forces, resulting from eddy currents induced in a metal sheet (workpiece) by a time varying (transient) magnetic field formed around electrically triggered coils. In the following a characterization of EMFs is discussed briefly. A comprehensive review is given in Psyk et al. (2011).

EMF is a contact free high-speed forming technology, mainly used for tubular forming (compression and expansion) and sheet metal forming , where this work focuses on the latter. Since it is contact free, no lubricants are needed, thus making it an environment friendly technology. Further advantages comprise, among other, high production rates paired with a high reproducibility that can be achieved; high flexibility, since it can be applied in areas that are hard to reach by a conventional deep drawing tools and the ability to form combinations of materials that could not be formed otherwise (coated

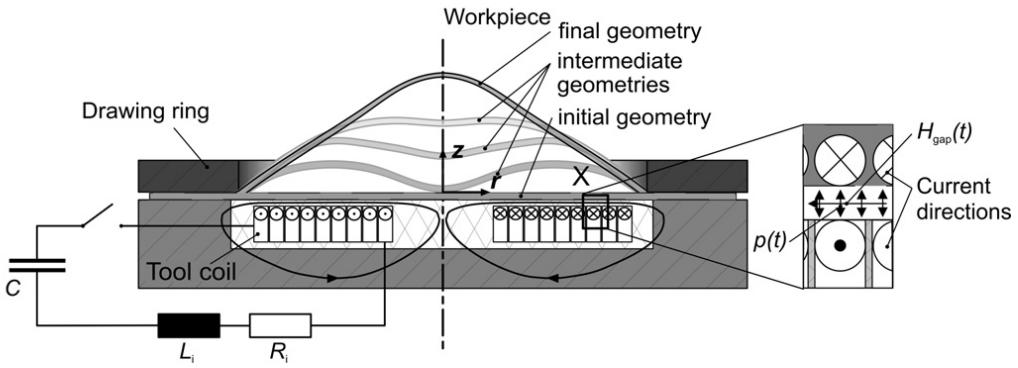


Figure 4.7.: Schematic of a typical EMF process setup. The forming machine is represented by an equivalent circuit, the current directions, as well as the magnetic field lines and the resulting Maxwell pressure are depicted. In this particular case, the forming task is axis symmetric. The figure is taken from Psyk et al. (2011).

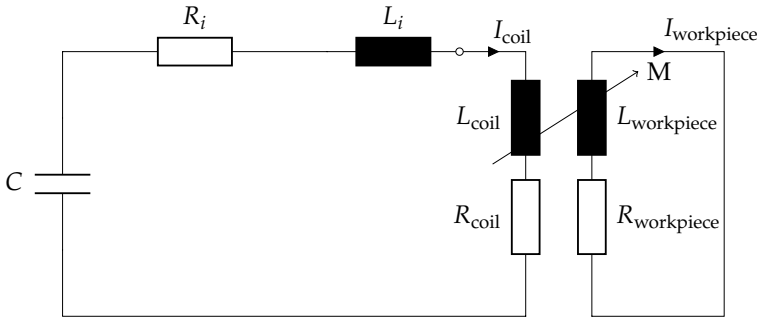


Figure 4.8.: Diagram of an equivalent circuit representing the forming machine, the tool coils, and the workpiece, motivated by Bauer (1967).

or laminated materials); and last but not least the favorable change of mechanical properties at high strain rates. However, disadvantages are the limitation to materials with a high electrical conductivity; the low overall process efficiency ($\sim 20 - -30\%$), reported by Risch (2009); and among others the mechanical and thermal load acting on the tool coils, limiting their lifetime significantly.

A typical EMF configuration, as depicted in Fig. 4.7, can be represented by a corresponding resonant circuit. The forming machine itself is described by a capacitor C, an inductance L_i and a resistance R_i. The associated tool coil is represented by its resistance R_{coil} and inductance L_{coil}, see Fig. 4.8. When both are assumed to be

connected in series to a pulse power generator the stored charging energy is given as

$$E_c\,(t) = \frac{1}{2} C U^2\,(t)\,,$$

where $U(t)$ is the charging voltage. The resulting current $I(t)$ can be determined by summarizing the different resistances and inductances and representing them by an equivalent reduced resonant circuit, as in Fig. 4.9. The current is then given as the solution to

$$L_{\mathrm{rc}} \frac{\mathrm{d}}{\mathrm{d}t} I(t) + R_{\mathrm{rc}} I(t) = -\frac{1}{C} \int I(t)\,\mathrm{d}t. \tag{4.16}$$

With initial conditions $I(0) = 0$ and $U(0) = U_0$ Eq. (4.16) is solved by the following damped sinusoidal current

$$I(t) = \frac{U_0}{\omega L_{\mathrm{rc}}} \mathrm{e}^{-\beta t} \sin \omega t,$$

where

$$\beta = \frac{R_{\mathrm{rc}}}{2 L_{\mathrm{rc}}}$$

is the damping coefficient, and

$$\omega = \sqrt{\omega_0^2 - \beta^2}$$

is the angular frequency with

$$\omega_0 = \frac{1}{\sqrt{L_{\mathrm{rc}} C}}.$$

The corresponding magnetic field induces eddy currents located close to the workpiece's surface and flowing into its opposite direction. The depth of penetration into the workpiece is the skin depth, given by

$$\delta = \frac{1}{\sqrt{\pi f \kappa \mu_0 \mu_{\mathrm{r}}}}, \tag{4.17}$$

which mainly depends on the frequency $f = \frac{\omega}{2\lambda}$ of the circuit and the electrical conductivity κ of the workpiece. The magnetic permeability in vacuum and the material's relative magnetic permeability are denoted by μ_0 and μ_{r}, respectively. The involved electromagnetic field can be modeled by the magneto quasi-static approximation of

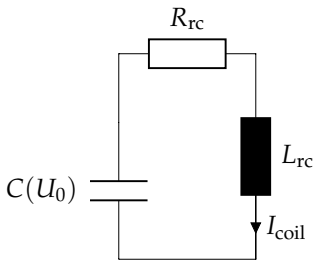


Figure 4.9.: The reduced equivalent resonant circuit as proposed by Jablonski (1976).

Maxwell's equations (see, e.g., Jackson (1999)), i.e.:

$$\operatorname{curl} \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t}, \tag{4.18a}$$

$$\operatorname{div} \boldsymbol{B} = 0, \tag{4.18b}$$

$$\operatorname{curl} \boldsymbol{H} = \boldsymbol{J}, \tag{4.18c}$$

$$\boldsymbol{J} = \kappa \boldsymbol{E} + \boldsymbol{J}_0, \tag{4.18d}$$

$$\boldsymbol{B} = \mu \boldsymbol{H}. \tag{4.18e}$$

As the magnetic flux density $\boldsymbol{B}$ is divergence free (Eq. (4.18b)), there exits a continuous vector potential $\boldsymbol{A} : \mathbb{R}^3 \to \mathbb{R}^3$, such that

$$\operatorname{curl} \boldsymbol{A} = \boldsymbol{B}.$$

The curl of the electric field (Eq. (4.18a)) $\boldsymbol{E}$ can thus be written as

$$\operatorname{curl} \boldsymbol{E} = \operatorname{curl} \left(-\frac{\partial \boldsymbol{A}}{\partial t} \right),$$

yielding

$$\boldsymbol{E} = -\frac{\partial \boldsymbol{A}}{\partial t} - \nabla \phi, \tag{4.19}$$

where $\phi : \mathbb{R}^3 \to \mathbb{R}$ is a scalar potential.

When no imposed current density is assumed, i.e., $\boldsymbol{J}_0 = 0$ in Eq. (4.18d), the curl of the magnetic field $\boldsymbol{H}$ (Eq. (4.18c)) is given by:

$$\operatorname{curl} \boldsymbol{H} = \kappa \boldsymbol{E} = -\kappa \left(\frac{\partial \boldsymbol{A}}{\partial t} + \nabla \phi \right) \tag{4.20}$$

Finally, the *eddy current equation* follows from rewriting Eq. (4.20) in terms of the vector potential $\boldsymbol{A}$, with Eq. (4.18e) ($\mu = \mu_0 \mu_r$):

$$\operatorname{curl} \frac{1}{\mu} \operatorname{curl} \boldsymbol{A} + \kappa \frac{\partial \boldsymbol{A}}{\partial t} = -\kappa \nabla \phi. \tag{4.21}$$

At this point a closer look has to be taken onto Eq. (4.19): if the vector potential $\boldsymbol{A}$ is transformed into $\boldsymbol{A} + \nabla\psi$, then the magnetic flux density $\boldsymbol{B}$ still remains unchanged, since:

$$\boldsymbol{B} = \operatorname{curl}(\boldsymbol{A} + \nabla\psi) = \operatorname{curl}\boldsymbol{A}.$$

The electric field, on the other hand, is affected by that change, i.e.:

$$\boldsymbol{E} = -\frac{\partial \boldsymbol{A}}{\partial t} - \nabla\left(\phi + \frac{\partial \psi}{\partial t}\right).$$

Thus, for the electric field to be also unaffected by the change of the vector potential, the scalar potential has to be transformed as $\phi \to \phi - \frac{\partial \psi}{\partial t}$, simultaneously. Consequently, the participating potentials may only be altered simultaneously by a function ψ. A function with this property is called *gauge function* and a pair of potentials $\boldsymbol{A}$ and ϕ fulfilling all of the above is constituting a *gauge*. This so called *gauge fixing* can be realized in numerous ways, here the *Coulomb gauge* is taken into account, reading:

$$\operatorname{div}\boldsymbol{A} = 0. \tag{4.22}$$

Applying the divergence operator to Eq. (4.21) and exploiting the Coulomb gauge condition from Eq. (4.22) it can be found, that

$$\Delta\phi = 0, \tag{4.23}$$

which, with fitting boundary conditions, yields an equation for the scalar potential. Further, the gauging condition yields a decoupling of the vector potential $\boldsymbol{A}$ and the scalar potential ϕ in Eq. (4.21). By solving above Laplace equation the scalar potential can be determined and then plugged into Eq. (4.21), which can then be solved for $\boldsymbol{A}$.

Due to

$$\begin{aligned}\rho &= \operatorname{div}\boldsymbol{D} \\ &= \operatorname{div}\left(\frac{\varepsilon}{\kappa}(\boldsymbol{J} - \boldsymbol{J}_0)\right) \\ &= -\frac{\varepsilon}{\kappa}\operatorname{div}\boldsymbol{J}_0\end{aligned}$$

with constant ϵ and constant $\kappa > 0$ the assumption $\boldsymbol{J}_0 = 0$ implies that no free charges are present. Consequently, the Lorentz force $\boldsymbol{F}_{\text{workpiece}}$ responsible for the forming is given by:

$$\begin{aligned}\boldsymbol{F}_{\text{workpiece}} &= \boldsymbol{J} \times \boldsymbol{B} \\ &= -\kappa\frac{\partial \boldsymbol{A}}{\partial t} \times \operatorname{curl}\boldsymbol{A} - \kappa\nabla\phi \times \operatorname{curl}\boldsymbol{A}.\end{aligned}$$

In fact, this is the Lorentz force everywhere, not only in the workpiece, but the particular part related to the workpiece is the cause of forming. Inside the workpiece no voltages are applied, and, as $\Delta\phi = 0$, it is constant at the workpiece's boundary. Hence, the potential is constant throughout the workpiece, therefore $\nabla\phi = 0$ in the workpiece, yielding

$$\boldsymbol{F}_{\text{workpiece}} = -\kappa \frac{\partial \boldsymbol{A}}{\partial t} \times \operatorname{curl} \boldsymbol{A}.$$

Above derived equations are fundamental for the numerical treatment of EMF processes, considered within this work.

4.2. Virtual Process Design I: Material Model Parameter Identification

Next, a method for virtual process design focusing on the determination of material model parameters is introduced. It is based on so called black box simulation methods, i.e., commercial simulation tools are exploited. Parts of Sections 4.2 and 4.3 can be found in earlier publications, Kiliclar et al. (2015a, 2016). In Section 4.2.1 a general introduction into the problem of identifying material model parameters is given, followed by the formulation of the problem at hand in Section 4.2.2. There, a new way of computing the distance between data sets originating from experiments and those obtained by simulation is introduced. A material model tailored to the forming applications discussed in this work in shortly outlined in Section 4.2.3. Finally, to conclude Section 4.2, the results are presented in Section 4.2.4.

4.2.1. Model Parameter Identification

Determining material parameters in constitutive material models generally gives rise to inverse problems. Solving inverse problems is of broad interest in a wide number of fields, see, e.g., Bates and Watts (1988). It is hence not surprising that several monographs deal with different solution techniques, an overview can for example be found in the works of Vogel (2002) or Tarantola (2005). The characteristics of the obtained solutions are discussed in Rieder (2003). Remarks on the experimental design to perform parameter identification can be found in Atkinson et al. (2007). Inverse problems are usually *ill-posed* problems, in the sense of Hadamard, therefore regularization methods may have to be applied. Engl et al. (1996) offer a deeper discussion about those.

In a constitutive material model two types of parameters can be distinguished; the ones that are directly measurable, often referred to as *physical* parameters, and those that are not directly accessible. The former are factors like the yield stress or the elastic modulus. The latter usually result from the constitutive assumptions or particular phenomenological approaches in the thermodynamic derivation of the model, such as hardening parameters and damage variables. Since these are not directly measurable, their identification usually requires *fitting* the model to experimental data. Mahnken et al. (1998) introduced methods which employ the material model within a FEM based simulation and adjust the parameters of interest systematically in order to reproduce experimentally obtained reference data.

This approach requires a function of merit, that depends on the parameters, and a measure for the deviation of the FEM simulation results from the experimental data. Thus, identifying parameters becomes equivalent to minimizing the function of merit under certain constraints, given by physics, limits of the simulation, and limits of the proposed model.

Systematic exploration of the parameter space requires the use of mathematical optimization methods, where an important aspect, particularly when using black box FEM simulation packages, is maintaining feasibility of the parameters during the optimization process, as discussed by Schnur and Zabaras (1992). To address this, a numerical realization of the inner point method, IPOPT by Wächter and Biegler (2006), is used in this work. Inner point methods as introduced in Chapter 3 enjoy the advantage of generating a sequence of feasible iteration points throughout the optimization procedure. Further IPOPT is relatively stable against a large number of variables and constraints, in fact a large number of constraints usually improves the convergence behavior. Further, Mahnken et al. (1998) point out that due to a potentially lower number of function evaluations, gradient based methods are to be preferred over derivative free methods.

The material considered in this work is the aluminum alloy EN AW-5083. The assumed material model is proposed by Vladimirov et al. (2014) and Kiliclar et al. (2012a). They already investigated the particular alloy and the material model in the context of combined quasi-static and electromagnetic impulse forming. Also parameters to identify the model under consideration have already been reported. These values, however, were obtained by manual parameter variation and not by means of mathematical optimization. The aim of this work is to overcome the need to manually adjust parameters and give rise to a systematic parameter space exploration method.

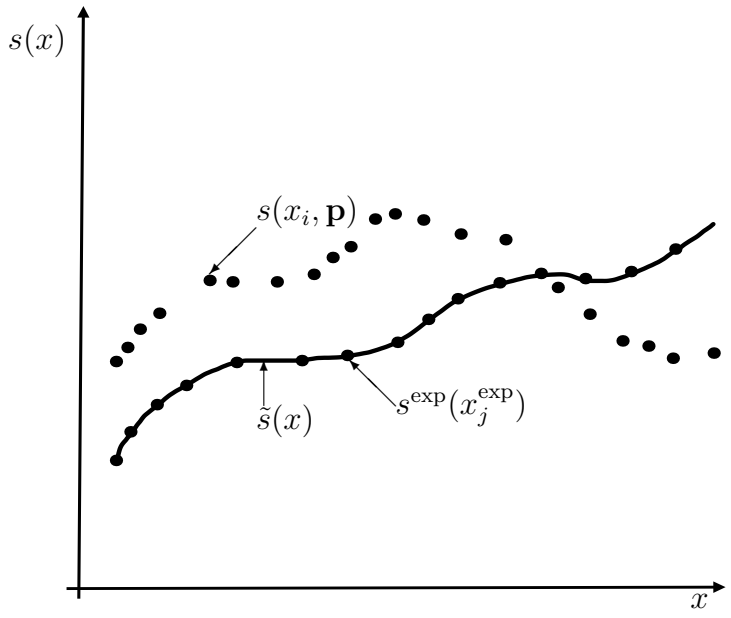


Figure 4.10.: A schematic of a typical relation between the simulated values $s(x_i, \mathbf{p})$ and the experimentally determined data $s^{\exp}(x_j^{\exp})$. The points at which the data is evaluated usually differs, as does the number of function evaluations.

4.2.2. Problem Formulation

The deviation between the experimental data and the data obtained by FEM simulations can be measured in different ways. Mahnken et al. (1998) proposed the least squares distance of simulated and measured stresses. They pointed out that problems may arise if simulated and measured stresses are not available for the same process time. Later Kleinermann and Ponthot (2003) used a slightly modified version, where they introduced weighted least squares distances for displacement over time curves. Still the question how to compare outputs that are not accessible for the same input remained unanswered. The scheme introduced in the following proposes a solution to this. Mathematically the problem boils down to measuring the distance between two functions. The remainder of this section will discuss how this observation can be used to formulate a novel distance measure, overcoming the problem of comparing outputs for different inputs.

Assume that input and output quantities are real valued. Then the set of all possible inputs is denoted by $X \subset \mathbb{R}$. Further, assume that the relation between input and

output of the experiment can be described by an unknown but continuous function $s^{\text{exp}} : X \to \mathbb{R}$. An experiment will yield a finite number of evaluations of the function s^{exp}, i.e., pairs $\left(x_j^{\text{exp}}, s^{\text{exp}}(x_j^{\text{exp}})\right)$, with $x_j^{\text{exp}} \in X$ for $j = 1, \ldots, M$. In the context of metal forming this can for example be be strain and stress pairs $(\varepsilon_j, \sigma_j)$ (both assumed as scalar values, as in, e.g., an uniaxial situation). Another example are displacement and force pairs (u_j, F_j) (again assumed as scalar values). Let $s : X \times \mathbb{R}^n \to \mathbb{R}$ be the mapping for the corresponding simulated quantity, additionally depending on some model parameters $\mathbf{p} = (p_1, \ldots, p_n)^\top \in \mathbb{R}^n$. Then the function of merit $f : \mathbb{R}^n \to \mathbb{R}$ is given as the L^2-distance between the functions s^{exp} and s, namely:

$$f(\mathbf{p}) = \frac{1}{|X|} \int_X |s(x, \mathbf{p}) - s^{\text{exp}}(x)|^2 \, \mathrm{d}x. \tag{4.24}$$

Here, x is the variable on which the quantities s and s^{exp} directly depend, e.g., $s^{\text{exp}}(\varepsilon) = \sigma$ or $s^{\text{exp}}(u) = F$. The total Lebesgue measure of X is denoted as $|X|$, i.e., $|X| = \int_X \mathrm{d}x$. Above formalism and distance definition can of course be expanded to more complex situations and are not restricted to scalar, real valued inputs and outputs, however, the assumption of an at least quadratic integrable functional relation is vital.

Results from simulations are only available at $N \in \mathbb{N}$ discrete points x_i $(i = 1, \ldots, N)$, therefore the formulation in Eq. (4.24) is brought into a discrete form, by applying a numerical integration scheme. Here, the trapezoidal rule is used (see, e.g., Schwarz and Köckler, 2009):

$$\begin{aligned} f(\mathbf{p}) \approx \frac{1}{2\,(x_N - x_1)} \sum_{i=1}^{N-1} (x_{i+1} - x_i) \Big[&\left(s\left(x_{i+1}, \mathbf{p}\right) - s^{\text{exp}}\left(x_{i+1}\right)\right)^2 + \\ &\left(s\left(x_i, \mathbf{p}\right) - s^{\text{exp}}\left(x_i\right)\right)^2 \Big]. \end{aligned} \tag{4.25}$$

As mentioned earlier, simulated and experimentally obtained data is not always accessible at the same points x_i, as depicted in Fig. 4.10. Not even the total number of data points needs to be the same. To overcome this problem values of s^{exp} are approximated for those x_i, for which only simulated data is available. To this end interpolation techniques (see, e.g., Schwarz and Köckler, 2009) can be used to interpolate the unknown function s^{exp}. This yields a function $\tilde{s} : X \to \mathbb{R}$ with

$$\tilde{s}\left(x_j^{\text{exp}}\right) = s^{\text{exp}}\left(x_j^{\text{exp}}\right), \tag{4.26}$$

for all data points of the experiment x_j^{exp}, $j = 1, \ldots, M$. The function $\tilde{s}$ is known and can be evaluated at all points $x \in X$, especially at all x_i for which exclusively simulated data

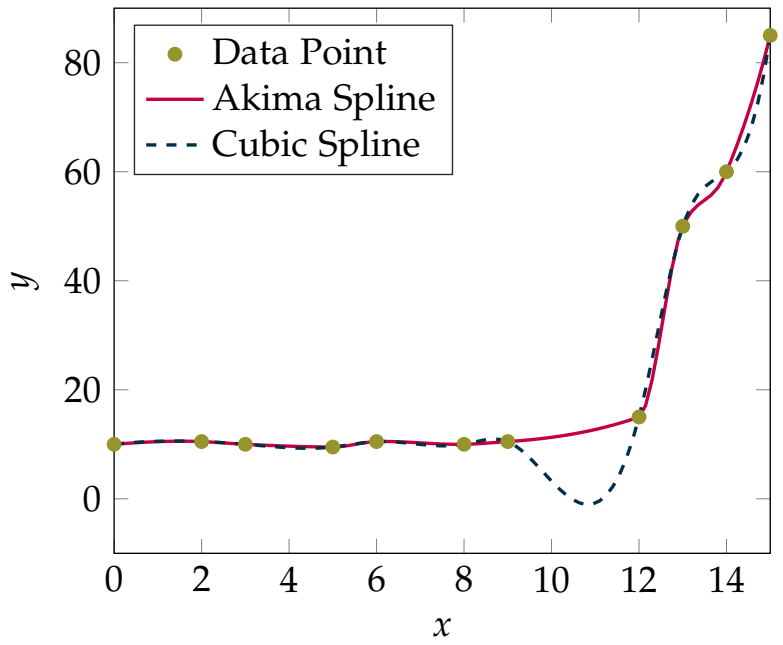


Figure 4.11.: Difference of the approximation with a cubic spline and an Akima spline. The Akima spline does not oscillate between two data points.

is available. In this work the Akima spline interpolation, introduced by Akima (1970), is applied. A general advantage of splines is that they do not oscillate at the boundary, as polynomials would. Additionally, the Akima spline does not oscillate between two data points x_j and x_{j+1}, yielding a curve that appears "smooth and natural" (Akima, 1970), as can be seen in Fig. 4.11. Inserting this interpolation into Eq. (4.25) yields:

$$F(\mathbf{p}) = \frac{1}{2\left(x_N - x_1\right)} \sum_{i=1}^{N-1} \left(x_{i+1} - x_i\right) \Big[\left(s\left(x_{i+1}, \mathbf{p}\right) - \tilde{s}\left(x_{i+1}\right)\right)^2 + \left(s\left(x_i, \mathbf{p}\right) - \tilde{s}\left(x_i\right)\right)^2 \Big] . \tag{4.27}$$

In addition to minimizing the difference between experiment and simulation the parameter space has to be confined appropriately. The constraints can represent physical and technical considerations. They are expressed as $m \in \mathbb{N}$ real valued constraints by a vector valued function $c : \mathbb{R}^n \to \mathbb{R}^m$. Collectively the parameter identification problem can be written as a constrained optimization problem:

$$\begin{aligned} \min_{\boldsymbol{p} \in \mathbb{R}^n} \quad & F\left(\boldsymbol{p}\right), \\ \text{subject to} \quad & c\left(\boldsymbol{p}\right) \geq 0. \end{aligned} \tag{4.28}$$

The inequalities have to be understood component wise.

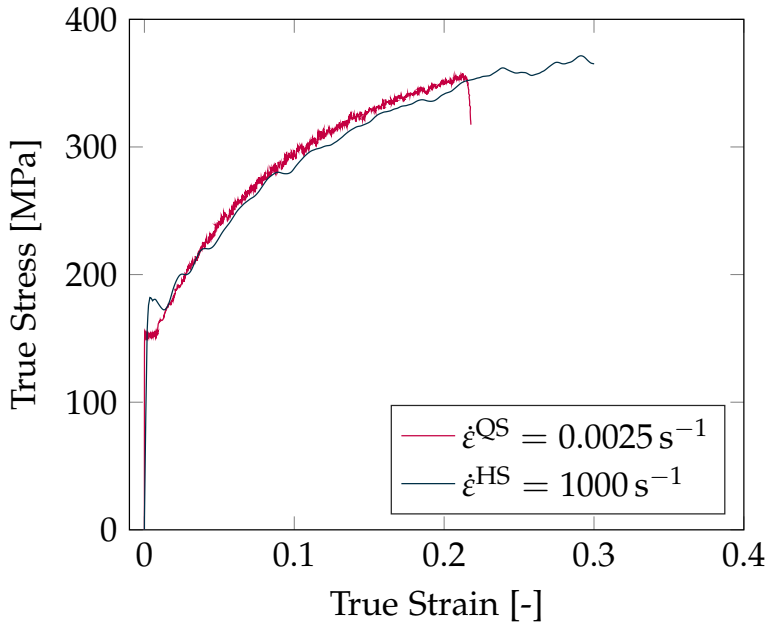


Figure 4.12.: Experimentally determined strain-stress curves for EN AW-5083 for different strain rates $\dot{\varepsilon}^{QS}$ and $\dot{\varepsilon}^{HS}$. The superscript QS denotes the quasi-static strain rate, and HS the high-speed rate, respectively.

4.2.3. Material Model

In this work an anisotropic hyperelastic-plastic material model with damage is assumed. The constitutive equations of the coupled damage-plasticity model are derived and discussed in full detail by Vladimirov et al. (2014). Including the main material characteristics of sheet metals, such as non-linear kinematic and isotropic hardening, anisotropy and damage, leads in the context of continuum damage mechanics to a certain set of parameters which have to be identified.

For the readers convenience in the following some aspects of the thermodynamically consistent constitutive material model are discussed in more detail.

The initial derivation of the model is carried out in the intermediate configuration and is based on a twofold multiplicative split of the deformation gradient tensor F. First, F is decomposed into an elastic and into a plastic contribution,

$$\mathsf{F} = \mathsf{F}_e \mathsf{F}_p.$$

Secondly, in order to take non-linear kinematic hardening into account, the plastic deformation is further split into a plastic-elastic and plastic-inelastic part, yielding an

extension of the rheological model by Frederick and Armstrong (2007):

$$F_p = F_{p_e} F_{p_i}.$$

Above split consequently yields a similar split of the right Couchy-Green stress tensor C into an elastic and a plastic contribution:

$$\begin{aligned} C_e &= F_e^\top F_e = F_p^{-\top} C F_p^{-1}, \\ C_{p_e} &= F_{p_e}^\top F_{p_e} = F_{p_i}^{-\top} C_p F_{p_i}^{-1}, \\ C_p &= F_p^\top F_p. \end{aligned}$$

On the thermodynamic level the Helmholtz free energy is, due to the principle of material objectivity and material isomorphism, additively decomposed into

$$\Psi = \Psi_e\left(C_e\right) + \Psi_{kin}\left(C_p\right) + \Psi_{iso}\left(\kappa\right).$$

Here, Ψ_e describes the macroscopic elastic material properties, which are assumed to be isotropic. The elastic energy stored in dislocation fields associated to the kinematic hardening is given by Ψ_{kin}, and the elastic energy storage referring to the isotropic hardening is modeled by Ψ_{iso}, which is a mapping of the accumulated plastic strain κ.

Thermodynamic consistency of the model is achieved by defining the second Piola-Kirchhoff tensor S such that the entropy inequality in the Clausius-Dunhem form for isothermal processes holds:

$$-\dot{\Psi} + S \cdot \frac{1}{2}\dot{C} \geq 0.$$

Following the work of Colemann and Gurtin (1967) and assuming isotropy of the Helmholtz free energy, the second Piola-Kirchhoff stress tensor is then given by

$$S = 2F_p^{-1}\frac{\partial \Psi_e}{\partial C_e} F_p^{-\top}.$$

Let Φ denote the yield function. The evolution equation for the plastic deformation, i.e., the symmetric part of the plastic velocity gradient $l_p = \dot{F}_p F_p^{-1}$, reads

$$\dot{d}_p = \dot{\lambda}\frac{\partial \Phi}{\partial \Sigma},$$

where

$$\begin{aligned} \Sigma &= M - \chi \\ &= 2C_e\frac{\partial \Psi_e}{\partial C_e} - 2F_{p_e}\frac{\partial \Psi_{kin}}{\partial C_{p_e}} F_{p_e}^\top \end{aligned}$$

corresponds to a relative stress in the intermediate configuration, given as the difference between the Mandel stress tensor M and the back stress tensor χ. Evolution equations for the kinematic and isotropic hardening are given by

$$\mathsf{d}_{\mathrm{p_i}} = \dot{\lambda}\,\frac{b}{c}\,\mathsf{M}^{\mathrm{D}}_{\mathrm{kin}}, \text{ and } \dot{\kappa} = \dot{\lambda}\,\frac{\partial \Phi}{\partial R},$$

respectively. Here, $R = -\frac{\partial \Psi_{\mathrm{iso}}}{\partial \kappa} = -Q\,(1 - e^{-\beta\,\kappa})$ is a stress-like quantity, and $\mathsf{M}_{\mathrm{kin}} = 2\mathsf{C}_{\mathrm{p_e}}\frac{\partial \Psi_{\mathrm{kin}}}{\partial \mathsf{C}_{\mathrm{p_e}}}$ is a symmetric Mandel-type stress tensor, related to the kinematic hardening. Further, b and c are non-linear kinematic hardening parameters, which have to be identified, like the isotropic hardening parameters β and Q. The slip rate is given by $\dot{\lambda}$. As presented in Vladimirov et al. (2010), $\mathsf{d}_{\mathrm{p_i}}$ is the symmetric part of the inelastic plastic velocity gradient $\mathsf{l}_{\mathrm{p_i}} = \dot{\mathsf{F}}_{\mathrm{p_i}}\mathsf{F}^{-1}_{\mathrm{p_i}}$. The superscript D indicates the deviator of a second order tensor, given as $\mathsf{T}^{\mathrm{D}} = \mathsf{T} - \frac{1}{3}\,\mathrm{tr}\,(\mathsf{T})\,\mathsf{I}$.

With the Hill-type yield function, cf. Hill (1948),

$$\Phi = \sqrt{\Sigma^{\mathrm{D}} \cdot (\tilde{\mathcal{A}}\,[(\Sigma^{\mathrm{D}})^{\top}])} - \sqrt{\frac{2}{3}}\,(\sigma_{\mathrm{y}} - R),$$

plastic anisotropy is accounted for, where the anisotropy in the intermediate configuration is described by the fourth order Hill tensor $\tilde{\mathcal{A}}$. The brackets $[\,]$ indicate that the fourth order anisotropy tensor $\tilde{\mathcal{A}}$ acts as a linear mapping on the second order tensor Σ.

Above formulated constitutive equations are now carried over to the undeformed or reference configuration, for the effective numerical implementation. To this end Vladimirov et al. (2014) employ the pull-back of the corresponding tensors, $\mathsf{M}, \mathsf{M}_{\mathrm{kin}}$, and χ and exploit the relation $\dot{\mathsf{C}}_{\mathrm{p}} = 2\mathsf{F}^{\top}_{\mathrm{p}}\mathsf{d}_{\mathrm{p}}\mathsf{F}_{\mathrm{p}}$. The resulting constitutive model in the reference configuration is given by the stress tensors

$$\begin{aligned}
\mathsf{S} &= \mu\left(\mathsf{C}^{-1}_{\mathrm{p}} - \mathsf{C}^{-1}\right) + \frac{\Lambda}{2}\left(\det \mathsf{C}\,(\det \mathsf{C}_{\mathrm{p}})^{-1} - 1\right)\mathsf{C}^{-1},\\
\mathsf{X} &= c\left(\mathsf{C}^{-1}_{\mathrm{p}_i} - \mathsf{C}^{-1}_{\mathrm{p}}\right),\\
\mathsf{Y} &= \mathsf{C}\mathsf{S} - \mathsf{C}_{\mathrm{p}}\mathsf{X},\\
\mathsf{Y}_{\mathrm{kin}} &= \mathsf{C}_{\mathrm{p}}\mathsf{X},
\end{aligned}$$

the evolution equations

$$\dot{\mathsf{C}}_{\mathrm{p}} = \dot{\lambda}\frac{\operatorname{sym}\left(\mathsf{C}_{\mathrm{p}}\left(\tilde{\mathcal{A}}\left[\left(\mathsf{Y}^{\mathrm{D}}\right)^{\top}\right] + \left(\tilde{\mathcal{A}}^{\top}\left[\mathsf{Y}^{\mathrm{D}}\right]\right)^{\top}\right)^{\mathrm{D}}\right)}{\sqrt{\mathsf{Y}^{\mathrm{D}} \cdot \left(\tilde{\mathcal{A}}\left[\left(\mathsf{Y}^{\mathrm{D}}\right)^{\top}\right]\right)}},$$

$$\dot{\mathsf{C}}_{\mathrm{p}_i} = 2\dot{\lambda}\frac{b}{c}\mathsf{Y}^{\mathrm{D}}_{\mathrm{kin}}\mathsf{C}_{\mathrm{p}_i},$$

$$\dot{\kappa} = \sqrt{\frac{2}{3}}\dot{\lambda},$$

the yield function

$$\Phi = \sqrt{\mathsf{Y}^{\mathrm{D}} \cdot \left(\tilde{\mathcal{A}}\left[\left(\mathsf{Y}^{\mathrm{D}}\right)^{\top}\right]\right)} - \sqrt{\frac{2}{3}}\left(\sigma_{\mathrm{y}} + Q\left(1 - e^{-\beta\kappa}\right)\right),$$

the Kuhn-Tucker conditions for the elastoplastic, rate independent format of the model

$$\dot{\lambda} \geq 0 \quad \Phi \leq 0, \quad \dot{\lambda}\Phi = 0,$$

and the Perzyna formulation for the modelling of viscoplastic, rate dependent materials, as are used here,

$$\dot{\lambda} = \frac{\langle\Phi\rangle^{m}}{\eta}.$$

The extension of above material model with ductile damage is performed by including Lemaitre-type isotropic damage. The evolution equation describing the scalar damage variable according to Lemaitre (1985):

$$\dot{D} = \dot{\lambda}\sqrt{\frac{2}{3}}\frac{1}{1-D}\left(\frac{Y}{s}\right)^{k} H\left(\kappa - p_D\right), \tag{4.29}$$

where, the unknown parameters (s, k, p_D) are controlling the rate of damage and the damage threshold. The latter part $H(\kappa - p_D)$ represents a Heaviside step function depending on the accumulated plastic strain κ. The quantity Y is the scalar density release rate as introduced by Lemaitre and Chaboche (1995). The Kachanov-Rabotnov concept is applied to compute the stress tensor which is divided by $(1 - D)$ and yields a softening behavior obtained with $\mathsf{S} = \bar{\mathsf{S}}\,(1 - D)$. It defines the relation between the continuum mechanical second Piola-Kirchhoff stress tensor S and the corresponding local effective stress tensor $\bar{\mathsf{S}}$. The damage variable in Eq. (4.29) is positive and smaller than 1, i.e., $0 \leq D \leq 1$, where $D = 0$ refers to an undamaged state and $D = 1$ to a fully damaged state.

So far, from above equations, seven parameters have to be identified, $\{Q, \beta, k, s, p_D, b, c\}$. This set is expanded with the elastic modulus E and the yield stress σ_{y} . Altogether this yields nine parameters to be identified, i.e., a vector $\mathbf{p} = \left(Q, \beta, k, s, p_D, E, \sigma_{\mathrm{y}}, b, c\right)^{\top}$.

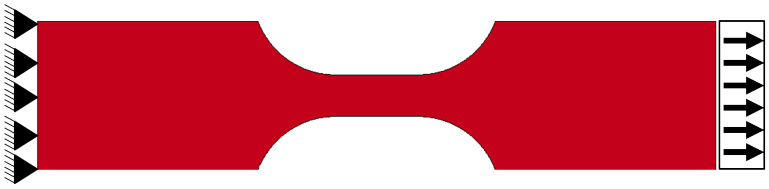


Figure 4.13.: The FEM simulation setup of the tensile test, picture is a courtesy of Yalin Kiliclar.

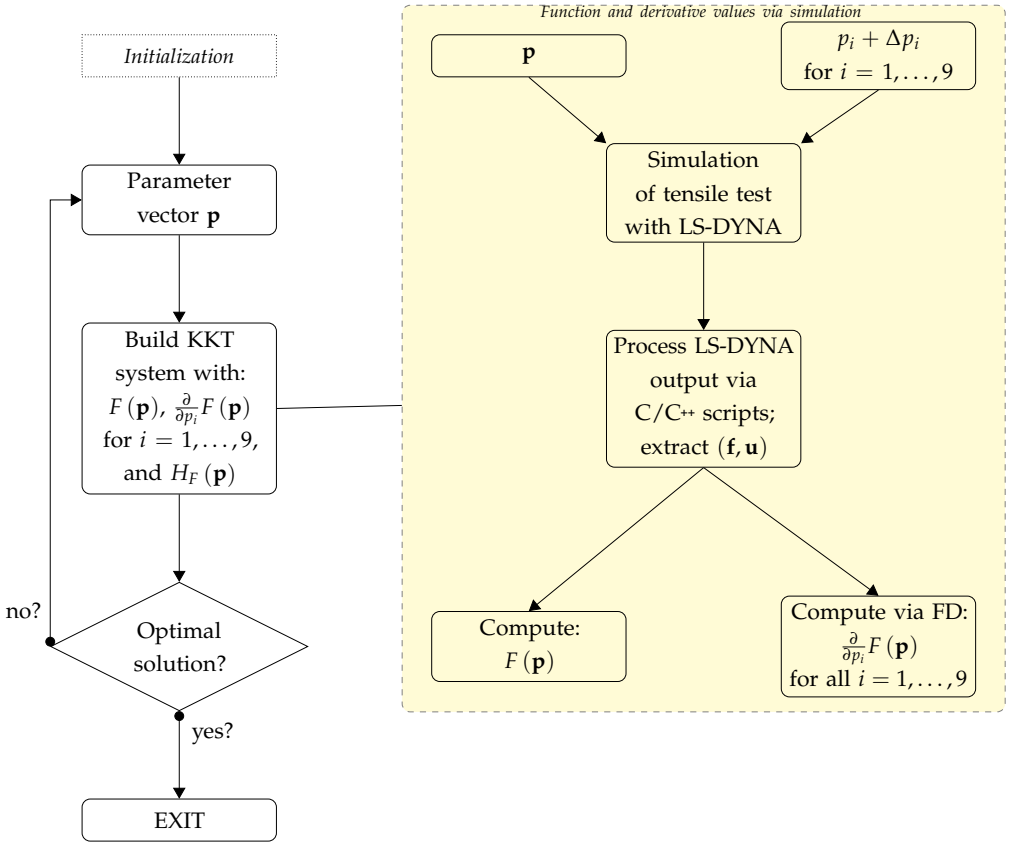


Figure 4.14.: A flow chart of the model parameter identification algorithm.

4.2.4. Results

The scheme introduced above will be used to identify the model parameters assumed in Section 4.2.3 for the aluminum alloy EN AW-5083. The experimental data used in this work is obtained by uniaxial tensile tests, conducted by Nordmetall GmbH. The experimentally obtained strain and stress curves for two different strain rates,

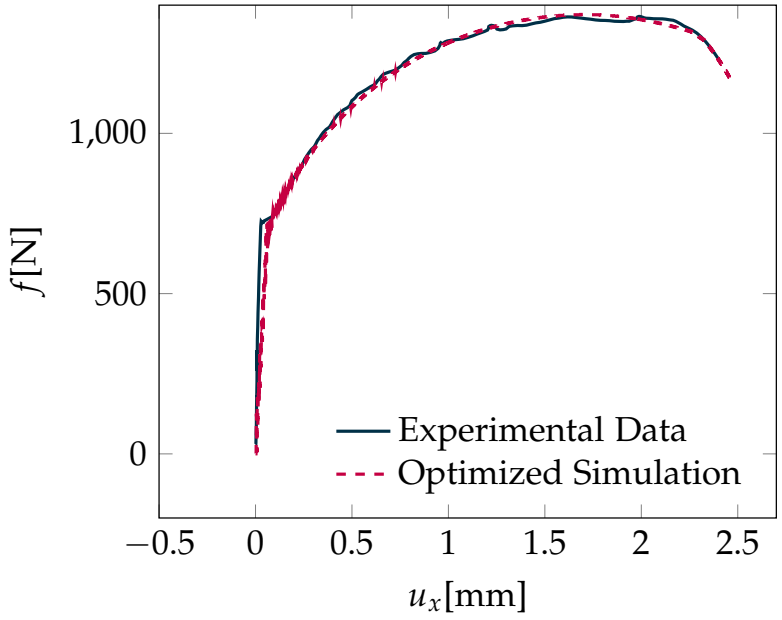


Figure 4.15.: The experimental data after fitting with an Akima spline and the results of the optimization procedure.

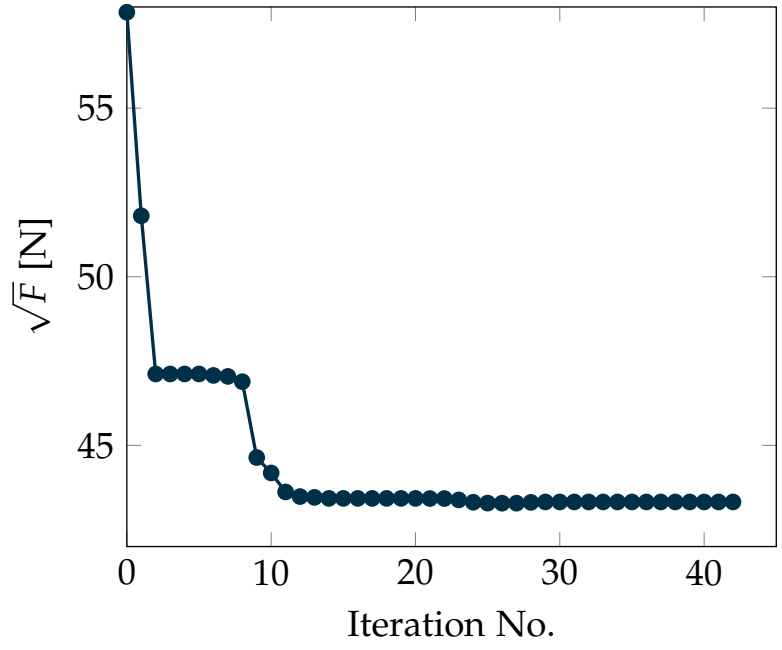
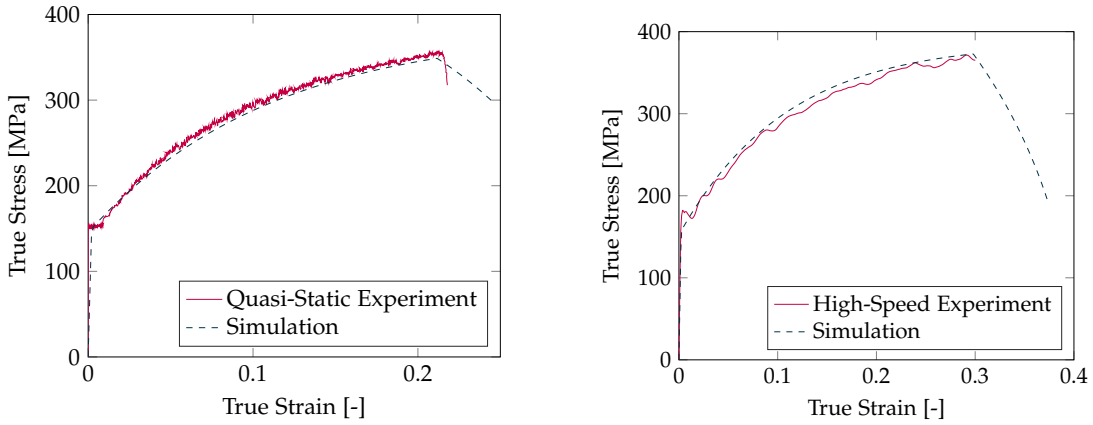


Figure 4.16.: The iteration progress during the optimization.



(a) Quasi-static stress-strain curves.

(b) High-speed stress-strain curves.

Figure 4.17.: Stress-strain curves obtained by high-speed tensile tests Fig. 4.17(a), with a strain rate of $\dot{\varepsilon}^{\mathrm{QS}} = 0.0025\,\mathrm{s}^{-1}$, and by high-speed tensile tests Fig. 4.17(b), with $\dot{\varepsilon}^{\mathrm{HS}} = 1000\,\mathrm{s}^{-1}$ compared to curves from FEM simulations using the parameters from Table 4.1 and the evolution of the damage threshold described in Eq. (4.30) for the higher strain rate.

$\dot{\varepsilon}^{\mathrm{QS}} = 0.0025\,\mathrm{s}^{-1}$ and $\dot{\varepsilon}^{\mathrm{HS}} = 1000\,\mathrm{s}^{-1}$, are given in Fig. 4.12. Opposed to the behavior found by Huh et al. (2009) for some steel alloys, no increase of the hardening under high strain rates is observed for this particular aluminum alloy. Winzer and Glinicka (2011) observed an increase of the flow stress for very high strain rates up to $5500\,\mathrm{s}^{-1}$ for this aluminium alloy. Although in the results at hand a higher yield stress σ_{y} can be seen for higher strain rates, the hardening behavior for both strain rates yields, that the overall shape of the curves coincides. Since the focus of this work lies ion formability rather than on material strength, the increase of σ_{y} can be neglected. The higher tensile stress, which is typically observed at high strain rates and is also observed here, is of bigger interest for the application at hand.

These observations and the results reported by Winzer and Glinicka (2011) and Clausen et al. (2004) suggest the following procedure for the determination of the damage threshold p_D for strain rates below $1450\,\mathrm{s}^{-1}$: Perform the model parameter identification with the data obtained with $\dot{\varepsilon}^{\mathrm{QS}}$. The damage threshold p_D can afterwards be adjusted to the high-speed case by linear extrapolation. As it will turn out, this approach work nearly perfect, thus reducing the number of high-speed tensile tests that have to be conducted.

Let the damage threshold determined by the optimization procedure be denoted

by p_D^{QS}. Motivated by the linear relation of accumulated plastic strains and equivalent strain, the threshold p_D^{HS} for the higher strain rate is extrapolated by

$$p_D^{\mathrm{HS}} := p_D^{\mathrm{QS}} + \frac{\kappa^{\mathrm{HS}} - \kappa^{\mathrm{QS}}}{\dot{\varepsilon}^{\mathrm{HS}} - \dot{\varepsilon}^{\mathrm{QS}}} \dot{\varepsilon}^{\mathrm{EQ}}, \tag{4.30}$$

where κ^{HS} and κ^{QS} denote the accumulated plastic strains, which can be computed by the model, and $\dot{\varepsilon}^{\mathrm{QS}}$ and $\dot{\varepsilon}^{\mathrm{HS}}$ the strain rate, for the respective regime. The equivalent strain rate $\dot{\varepsilon}^{\mathrm{EQ}}$ is computed using Green's strain tensor E:

$$\dot{\varepsilon}^{\mathrm{EQ}} = \sqrt{\frac{2}{3} \dot{\mathsf{E}} \cdot \dot{\mathsf{E}}}.$$

Earlier works of Taebi et al. (2012) made similar approaches to model the damage in high-speed regimes. There the usage of a so called forming limit surface is proposed. This surface is derived from a quasi-static forming limit curve and by assuming a rate depended Johnson-Cook model. They used this surface to construct constraints to their process parameter identification problem.

The data obtained by uniaxial tensile tests is compared to a corresponding FEM simulation. The FEM simulation of the tensile test is set up using the software package LS-DYNA, Fig. 4.13 shows the basic setup. For the simulation 1 130 Belytschko-Tsay shell elements are used, within the mechanics problem solver of LS-DYNA, see Hallquist et al. (2006) for more information. A prescribed motion curve is assumed and the FEM problem is explicitly solved in time. The material model is added via the USERMAT interface by Yalin Kiliclar. The simulation is used to provide the objective function as well as derivative information for the optimization algorithm.

To construct an objective function according to Eq. (4.27) an appropriate quantity that depends on the parameters has to be selected. In this work the force-displacement curve, given in Fig. 4.15, is chosen. It provides information about the elastic behavior of the material from which E and σ_{y} can be derived. From the plastic loading phase the kinematic and isotropic hardening parameters (β, Q, b, c) can be estimated. The damage parameters (s, k, p_D) are derived from the part of the curve, where an increase of displacement yields a decreasing force. It turned out, that by using stress-strain curves especially these parameters can not be reliably determined. The resolution at damage is higher when using force-displacement data.

In the simulation the displacement u in x-direction is monitored for $N = 153\,515$ time steps at one element at the right hand side of the specimen, i.e, where the force is applied in the experiment. The corresponding force component is computed by adding the partial forces from the fixed nodes at the left hand side (the fixed part) of the work

piece. Form this simulated data the force-displacement curve is constructed. An Akima spline is fitted to the obtained experimental data, to this end some basic data post processing had to be performed: multiple force values for a single displacement were discarded and the data was ordered. None of theses steps changed the experimental data significantly.

Although the parameters of interest are coupled to different stages of the the tensile test, the identification task can be solved by a single objective function, cf.Eq. (4.27):

$$F(\boldsymbol{p}) = \frac{1}{2(u_N - u_1)} \sum_{i=1}^{N-1} (u_{i+1} - u_i) \left[\left(f_{i+1}(\boldsymbol{p}) - \tilde{f}_{i+1} \right)^2 + \left(f_i(\boldsymbol{p}) - \tilde{f}_i \right)^2 \right].$$

Here, $\boldsymbol{p} = \left(Q, \beta, k, s, p_D, E, \sigma_y, b, c\right)^\top$ is the vector of parameters as described in Section 4.2.2, $\tilde{f}$ are the reference forces from the experiment, after fitting an Akima spline to the raw data. The forces obtained by simulation are denoted by f. The variables are constrained as follows:

$$\begin{aligned} 190 \leq Q[\text{MPa}] \leq 250, \quad \varepsilon_{\text{tol}} \leq \beta \leq 20, \\ 1 \cdot 10^{-4} \leq b \leq 1 \cdot 10^{-2}, \quad 1 \cdot 10^{-5} \leq c[\text{MPa}] \leq 1 \cdot 10^{-3}, \\ \varepsilon_{\text{tol}} \leq k \leq 1, \quad \varepsilon_{\text{tol}} \leq s[\text{MPa}] \leq 1, \quad 0 \leq p_D \leq 1, \\ 70\,000 \leq E[\text{MPa}] \leq 90\,000, \quad 100 \leq \sigma_y[\text{MPa}] \leq 200. \end{aligned} \tag{4.31}$$

Here, $\varepsilon_{\text{tol}} = 1 \cdot 10^{-8}$ is a small number to keep the values away from zero. The boundaries assumed for E and σ_y arise from the material under consideration. The elastic modulus for EN AW-5083 is reported to be somewhere between 70 GPa and 90 GPa, the tensile yield stress is reported to be between 100 MPa and 200 MPa. Both strongly depend on the manufacturing conditions like rolling, tempering and so on. The constraints to the other parameters result from earlier approaches to identify the material model parameters for EN AW-5083 made by Vladimirov et al. (2014). These results serve as starting values and are used to construct upper and lower bounds.

In order to make use of the inner point method as described in Section 3.4, derivative information is needed. The finite difference (FD) (see, e.g., Schwarz and Köckler, 2009) method is used to approximate the partial derivatives $\frac{\partial}{\partial p_i} F(\boldsymbol{p})$, for $i = 1, \ldots, 9$. The use of the FD method increases the number of objective evaluations, but makes the proposed scheme versatile and suitable for a wide range of material models, especially when using commercial black box solvers, as pointed out by Kreißig et al. (2007).

When introducing a new model, the implementation of the optimization method needs no changes, only the underlying FEM simulation has to be adjusted to the

new model. The Hessian H_F, as needed to solve the linear system in Eq. (3.15), is approximated using the LBFGS-update scheme following the work of Byrd et al. (1994).

To establish communication between LS-DYNA and IPOPT throughout the optimization procedure, both programs had to be coupled by interface scripts. These scripts manage the reconstruction of the forces f and displacements u from the LS-DYNA output, and construct the objective function F and its gradient for the optimization. The gradient calculation via the FD method is carried out in parallel, yielding a significant speed improvement. The overall algorithm is visualized in Fig. 4.14.

The force-displacement curve for the quasi-static strain rate $\dot{\varepsilon}^{\mathrm{QS}} = 0.0025\,\mathrm{s}^{-1}$ after optimization is given in Fig. 4.15. The algorithm needed 42 steps to find the optimal value. Fig. 4.16 shows the development of the objective function during the optimization procedure. It can be seen that after 10 iteration steps almost the final value is reached. The following steps are made to improve the dual problem as is discussed in Section 3.4. The optimal parameters are given in Table 4.1.

To obtain parameters characterizing the material behavior for the higher strain rate, $\dot{\varepsilon}^{\mathrm{HS}} = 1000\,\mathrm{s}^{-1}$, the extrapolation method introduced in Eq. (4.30) is used. Fig. 4.17 shows the stress and strain curves for the respective experiments and the corresponding simulation results after computing p_D^{HS} according to Eq. (4.30). It can be seen how the application of this formula yields nearly perfect matching for the higher strain rate, without computing a new parameter set by means of optimization. Further, once a parameter set has been determined, the damage threshold can easily be adjusted to the forming task at hand.

Parameter	Identified value
Q	$1.604 \cdot 10^2$MPa
β	$1.265 \cdot 10^1$
k	$4.694 \cdot 10^{-1}$
s	$2.680 \cdot 10^{-1}$
p_D	$6.306 \cdot 10^{-1}$
E	$8.089 \cdot 10^4$ MPa
σ_{y}	$1.185 \cdot 10^2$ MPa
b	$5.124 \cdot 10^{-3}$
c	$4.598 \cdot 10^{-4}$MPa

Table 4.1.: The identified parameters values after optimization.

It should be pointed out, that the presented method is not restricted to parameter identification problems based on force-displacement analysis. The proposed algorithm works for any given quantity that is experimentally measurable and reproducible by simulation. Moreover, the proposed method works for arbitrary identification problems with underlying phenomenological models, since the optimization makes no further assumptions on the model. This makes the method applicable to a wide range of material models, as well as to other types of parameter identification problems.

The determined parameters can be used to simulate more complicated setups with a high accuracy. It was shown by Kiliclar et al. (2016) that the parameters determined by the presented algorithm can be used to predict the outcome in coupled high-speed and electromagnetic round cup forming. The fact that model parameters obtained by relatively cheap simulations can be applied to more complex situations while retaining a high level of agreement to experimental results makes the proposed method very attractive for technologically relevant applications. The implementation can be used with other black box FEM software packages with little effort, which offers even more flexibility.

In the next section the presented method will be modified and applied to process parameter identification.

4.3. Virtual Process Design II: Process Parameter Identification

Section 4.3 is the extension of the introduced optimization procedure to the determination of optimal process parameters. A short overview of process parameters of interest is given in Section 4.3.1. The simulation of the coupled high-speed, electromagnetic sheet metal forming, as well as the formulation of an adequate optimization problem to identify optimal double exponential current pulses is the topic of Section 4.3.2. In Section 4.3.3 the numerical results are given.

4.3.1. Process Parameter Identification in Coupled Forming Applications

In sheet metal drawing applications process parameter control is important for designing an economic and efficient process. These parameters comprise the blank holder force, punch speed, lubrication conditions and the die geometry, in the case of quasi-static deep drawing. Controlling and modifying these parameters was already the subject of a vast number of investigations, only a few of which are discussed here.

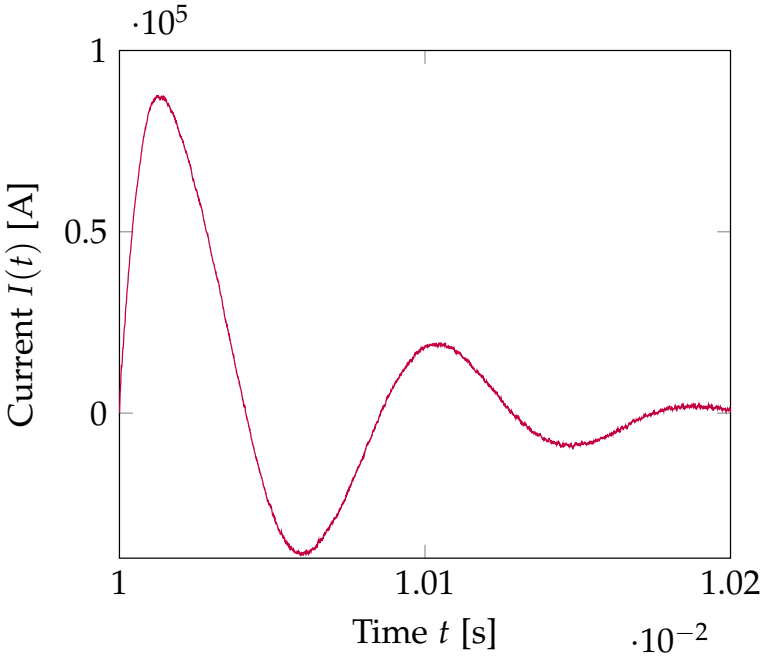


Figure 4.18.: Typical decaying sinusoidal current curve used for pulse generation in impulse forming.

Wang et al. (2007) proposed a method for the optimal setting of the blank holder force. The effects of a variating punch speed is discussed by, e.g., Vollertsen and Hu (2010). The process studied in the scope of this work differs from the classical one by the fact, that a subsequent electromagnetic forming step is incorporated. Hence, next to the above mentioned parameters also the control of this new phase of the forming process is of particular interest.

In electromagnetic forming a coil is triggered by a current released from a previously charged capacitor bank. A transient magnetic field forms around the tool coils and, by this, eddy currents are induced in the workpiece. Accelerated by the mutually repulsive force, the metal sheet moves away from the fixed coils towards the surface of the die. The strain rates in this phase of the process are usually so much higher than during high-speed forming, that the kinetic energy becomes a dominant driving force of the deformation. Therefore, electromagnetic forming falls under the class of high-speed forming methods, cf. Section 4.1.4.

One parameter playing a major role in this process is the employed triggering current. Usually the current generating the pulses in electromagnetic forming has a decaying sinusoidal form, as shown in Fig. 4.18. Because of the speed of the process, only the

energy generated by the first half wave contributes to forming. Consequently, the fields generated by the other half waves only interact with the coil itself, yielding fast wear of the coil. Gies et al. (2014) observed, that the polymer matrix, where the coils are embedded in, heats up, leading to thermal wear. Additionally, the resulting Lorentz force will cause the coil windings to drift apart.

In order to reduce this quickly occurring and, when using sinusoidal pulses, unavoidable defects, the use of mono-directed pulses is of major interest. Methods to realize these pulses in technical applications comprise techniques that cut the subsequent half waves or at least dampen them to a level that causes no further damage in the coil, as is discussed in Bodenseher et al. (1973).

A certain class of mono-directed pulses, which can be treated rather conveniently, from a mathematical viewpoint, are double exponential pulses. The technical realization of those pulses is quite complicated and out of the scope of this work. However, they can be used as a triggering current by the proposed tool for virtual process design. A coupled forming simulation of a cup is set up, and the proposed method is used to find an optimal current, such that the coupled process yields a cup with a rather sharp bottom edge radius.

4.3.2. Combined Quasi-Static and Electromagnetic Forming Simulation

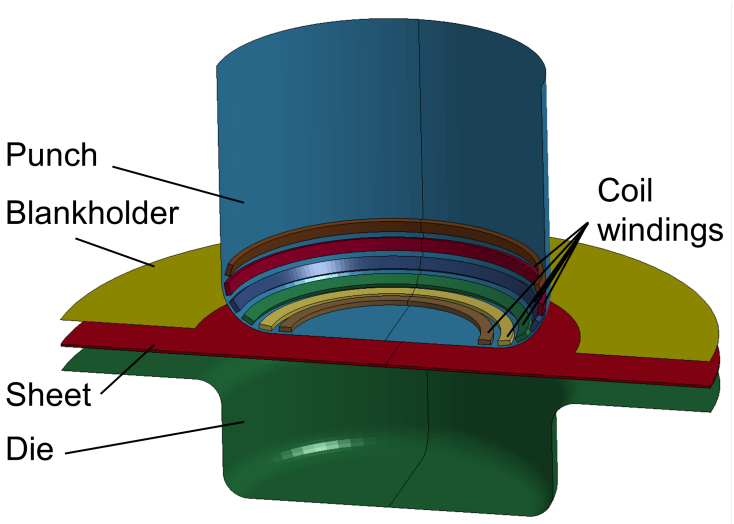


Figure 4.19.: The geometrical setup which is used for the simulation of the coupled forming of a cup, the picture is a courtesy of Yalin Kiliclar.

By utilizing the material model introduced in Section 4.2.3 and using the parameters

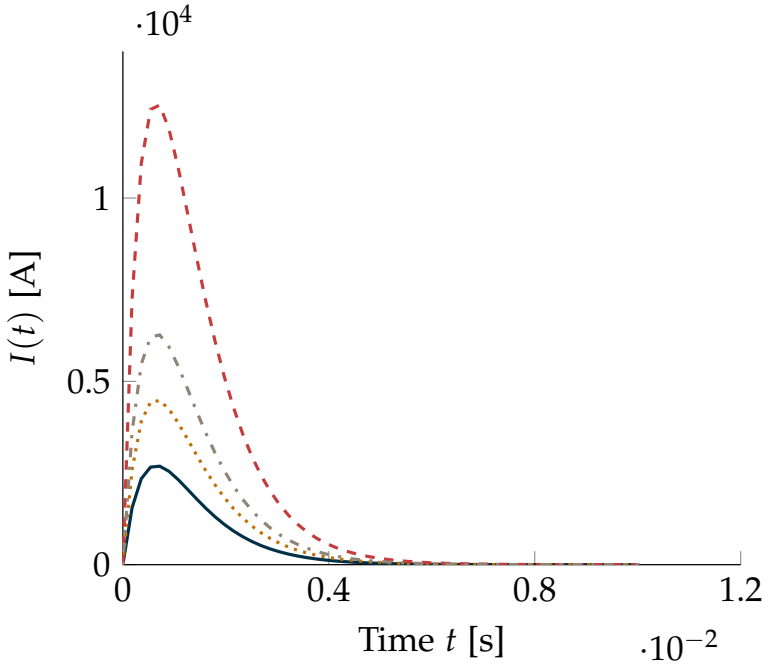


Figure 4.20.: Different current curves resulting from double exponential pulses as described in Eq. (4.32), for various $I_\alpha, I_\beta, \alpha,$ and β.

obtained by optimization and given in Section 4.2.4 a simulation of the combined forming process is set up. Again the simulation software LS-DYNA is used. The geometrical model is given in Fig. 4.19. It consists of the coils, the punch, the blank holder, the metal sheet and the forming die. Due to symmetry considerations only a quarter of the total cup needs to be simulated, thereby reducing the computational efforts significantly. The overall simulated process time is 10.2 ms, automatic time step control for the explicit time integration scheme is used. The mechanical part of the process is simulated by the FEM. In total 10 146 Belytschko-Tsay shell elements and 42 445 constant stress solid elements are used, see Hallquist et al. (2006) for details on the elements used in LS-DYNA.

Since the wave lengths associated with the typical frequencies occurring in impulse forming are longer then the distances under consideration, Maxwell's Equations are solved in their eddy current approximation, i.e., contributions from Maxwell's displacement currents are neglected, as discussed in Section 4.1.4. The electromagnetic field inside the coils is also simulated by the FEM, using 3D Nédélec edge elements, see, e.g., Monk (2003). A CG iterative solver without preconditioner is applied to solve the resulting linear system. The last known solution is used as a starting point and the

iteration is limited to 50 steps, or a relative tolerance of 0.005.

Within the air gap between work piece and coil the field is computed using the boundary element method (BEM). To solve the linear system arising from the BEM once more the CG iterative method without preconditioner is used. The maximum iteration number is set to 10 000 and the relative tolerance is set to 0.001.

The coupling between the mechanical and electromagnetic system is realized using the arbitrary Lagrangian Eulerian (ALE) method. A thorough mathematical discussion of the coupled mechanical and electrical forming processes simulation using the ALE method can be found in the work of Stiemer et al. (2009). Details on the implementation of the electromagnetic coupling within LS-DYNA are given by Caldichoury and L'Eplattenier (2008).

The developed simulation tool allows for studying various pulse families, independent of their technical realization for real world applications. Double exponential pulses are a class of mono-directed pulses that can be described mathematically in a convenient fashion, as will be seen later. These pulses are well studied in the field of electromagnetic compatibility (EMC), since they naturally occur during lightning strikes. The technical realization of these pulses is quite expensive and complicated, nevertheless this class can be used to study the effects and control of mono-directed pulses in general. They can be used as a prototype for other, more conveniently producible, pulses, as features like the maximum amplitude, rise time and damping can be derived.

A double exponential pulse can be described as

$$I(t) = I_\alpha e^{-\alpha t} + I_\beta e^{-\beta t}, \tag{4.32}$$

with parameters $I_\alpha, I_\beta, \alpha, \beta \in \mathbb{R}$. In particular the optimization will focus on these four parameters, fully describing the pulse. The effects of various parameter constellations can be seen in Fig. 4.20. Camp and Garbe (2004) showed how the parameters used in the mathematical description, given in Eq. (4.32), can be related to technical parameters like wave length, rise time and maximum amplitude, which are more commonly used in the engineering context and can be transferred to other pulse forms.

To make use of the algorithm discussed earlier in the context of material model parameter identification, an appropriate problem formulation has to be derived. To this end the following optimization target is formulated:

Find current parameters of the pulse, such that the bottom radius of the cup is maximal after electromagnetic forming and no damage occurs.

In contrast to the problem discussed in Section 4.2.2, this optimization target needs

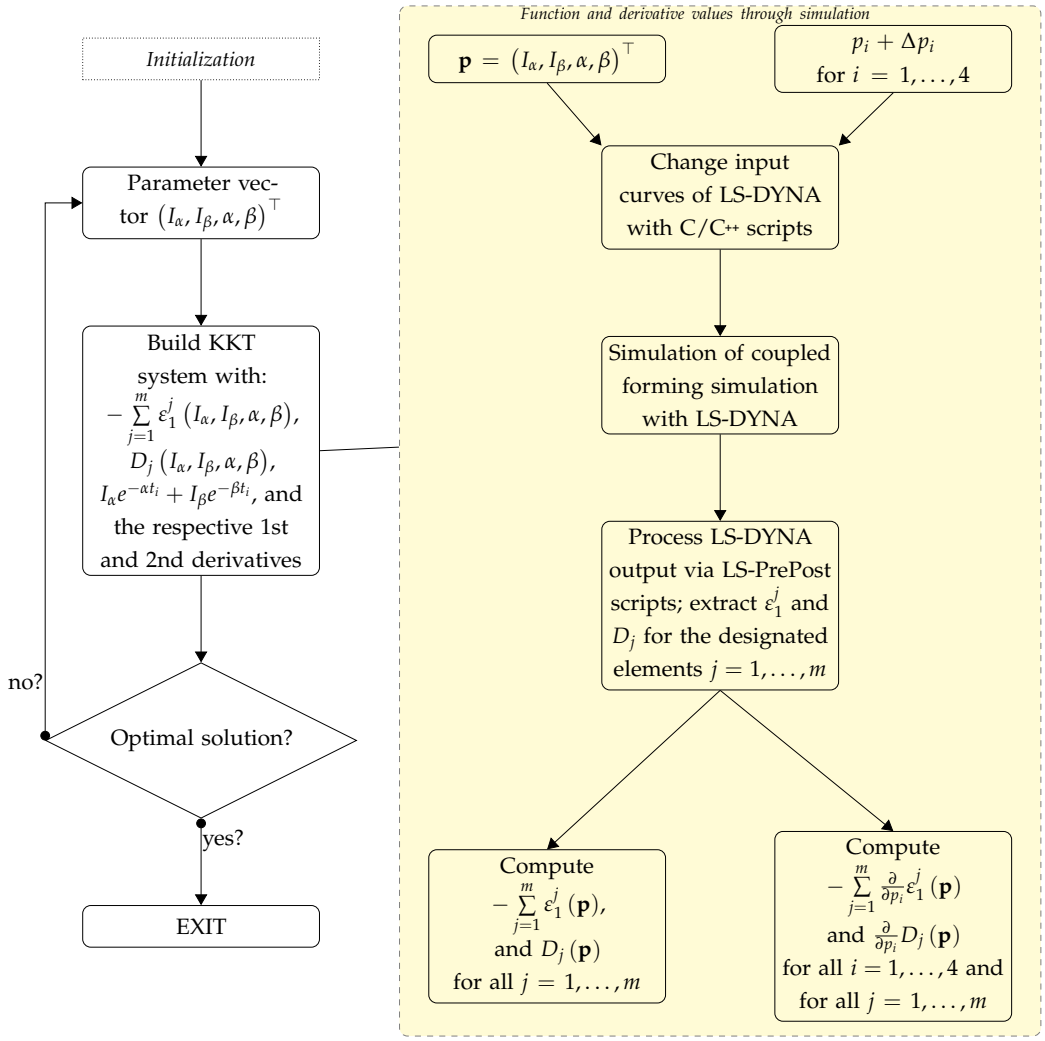
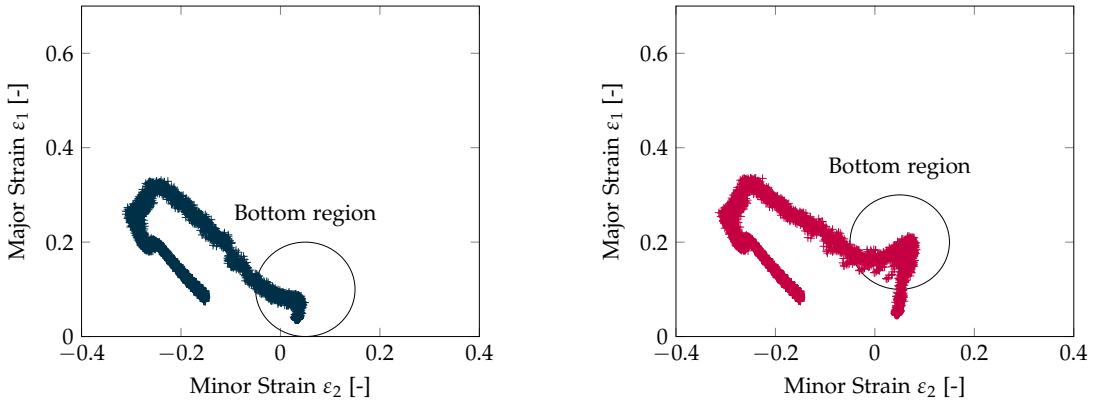


Figure 4.21.: A flow chart of the process parameter identification algorithm. Only some small changes have to be made compared to the algorithm in Fig. 4.14.

no comparison to experimental data for its determination. However, a comparison to an experiment with the numerically identified parameters may serve as a validation of the proposed method. The formulation of the target can be seen as a classical shape optimization problem, as was done by Taebi et al. (2012) in the context of combined forming. In shape optimization, the objective function is usually formulated as a distance between die and sheet. To construct and evaluate this distance, the sheet's and die's surfaces have to be parametrized and an appropriate distance measure has to be set. The choice of the parametrization and the distance measure have a similar effect

on the optimization, as the choice of the norm earlier in Section 4.2. Again problems will arise, when the distance measure is transferred to the discrete level.



(a) Minor vs major strain after quasi-static phase of the process.

(b) Minor vs major strain after applying the electromagnetic impulse.

Figure 4.22.: Minor vs major strain diagram obtained by simulation. Fig. 4.22(a) shows the relation between major and minor strain after the high-speed deep drawing phase of the process. In Fig. 4.22(b) their relation is depicted, after the subsequent electromagnetic impulse forming is finished. It can be seen, that the major strain ε_1 rises at the bottom of the cup, in the circled area, after the electromagnetic forming step. This data can be compared to an appropriate FLC.

To avoid these problems, this work proposes a new way of achieving and measuring the maximal bottom radius. It is based on the observation, that the major strain ε_1, i.e., the first eigenvalue of Green's strain tensor, tends to become larger with decreasing bottom edge radius , as is seen in Fig. 4.22. Other simulations and experiments, as conducted by Taebi et al. (2012), showed the same behavior of ε_1 at the cup's bottom radius. Therefore, if the bottom edge radius is to be decreased, the following choice for the objective function is justified:

$$F\left(I_\alpha, I_\beta, \alpha, \beta\right) = -\sum_{j=1}^{m} \varepsilon_1^j\left(I_\alpha, I_\beta, \alpha, \beta\right), \tag{4.33}$$

where the sum is taken over $m \in \mathbb{N}$ previously assigned elements at the bottom of the cup. The superscript j relates ε_1 to the $j-$th element of the chosen selection. The advantage of this objective function is that it neither depends on a parametrization of the sheet, nor on some distance measure. Moreover, the first principle strain is a

quantity that is intrinsic, and computed in every Gauss point during the simulation and is therefore easily accessible.

To formulate non-linear constraints, the damage variable D introduced in Eq. (4.29) is used. The cup is most likely to be damaged at the radius, D is thus computed in the same selection of elements as ε_1. In order to prevent the optimization algorithm from choosing a current curve that would produce a result where elements have values of D too close to one, a threshold parameter $0 \leq p \leq 1$ is introduced. Further, the current is restricted at all time steps (coming from the explicit time integration scheme) t_i $(i = 1, \ldots, N)$, by a maximum current $I_{\max}$. Together this yields

$$\begin{aligned} \min_{(I_\alpha, I_\beta, \alpha, \beta)^\top \in \mathbb{R}^4} \quad & -\sum_{j=1}^{m} \varepsilon_1^j \left(I_\alpha, I_\beta, \alpha, \beta\right), \\ \text{subject to} \quad & D_j \left(I_\alpha, I_\beta, \alpha, \beta\right) \leq 1 - p, \quad \forall\, j = 1, \ldots, m, \\ & I_\alpha e^{-\alpha t_i} + I_\beta e^{-\beta t_i} \leq I_{\max}, \quad \forall\, i = 1, \ldots, N. \end{aligned} \tag{4.34}$$

4.3.3. Results of the Process Parameter Optimization

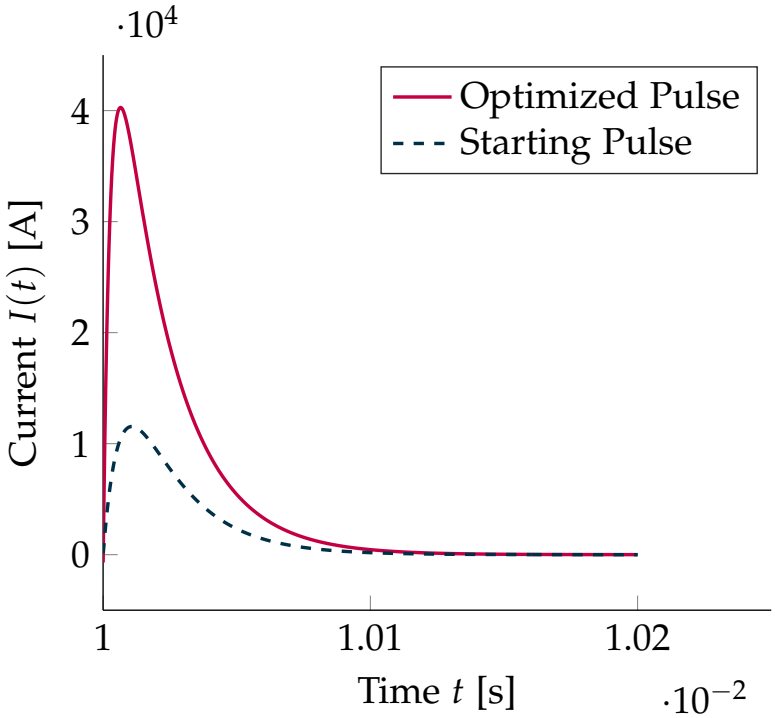


Figure 4.23.: Current $I(t)$ for the identified parameters $I_\alpha = -65\,570.2\,\text{A}$, $I_\beta = 64\,867.8\,\text{A}$, $\alpha = 6878.78\,\text{s}^{-1}$, and $\beta = 973.021\,\text{s}^{-1}$.

The non-linear optimization problem in Eq. (4.34) can again be solved by the algorithm proposed earlier. Only a few changes in the interface between IPOPT and LS-DYNA have to be made, as can be seen in Fig. 4.21. The evaluation of ε_1 and D requires the use of the pre- and post-processing software LSPrePost, which can be controlled by macros to evaluate the requested quantities. Again, as for the model parameter identification in Section 4.2.4, the computation of the derivative's approximation by the FD method is performed in parallel. The derivatives with respect to the current parameters of the double exponential pulse from Eq. (4.32) can be computed analytically and are passed to IPOPT via C/C++ functions.

The optimization was performed with $I_{\max} = 125\,000\,\mathrm{A}$ and $p = 0.1$. The values identified after optimization are:

$$I_\alpha = -65\,570.2\,\mathrm{A},\ I_\beta = 64\,867.8\,\mathrm{A},\ \alpha = 6878.78\,\mathrm{s}^{-1},\ \beta = 973.021\,\mathrm{s}^{-1}. \tag{4.35}$$

The objective function takes the values $F = -5.7903$ after 17 iteration steps of the inner point algorithm. As can be seen in Figs. 4.24 and 4.25, the cup's bottom edge radius is enhanced. The drawing radius $r = 20\,\mathrm{mm}$ achieved by classical high-speed deep drawing alone was changed by a subsequent application of electromagnetic forming to $r = 15.35\,\mathrm{mm}$. It should be stressed, that other parameters, like blank holder force, punch speed and lubrication were kept fix. The computed parameter set does not lead to material damage. The initial sheet thickness of 1 mm evolves to a minimum thickness of 0.91 mm after deep drawing and 0.85 mm after electromagnetic forming. The optimized pulse form yields a triggering current curve which is considerably higher and sharper than the initial one, as can be seen in Fig. 4.23. The identified maximum amplitude is $I^*_{\max} = 40\,279\mathrm{A}$, which is lower than the prescribed $I_{\max}$. The rather fast decay is in good agreement with the observation, that only the first half wave contributes to forming.

The presented idea is not restricted to the optimization of pulses, it can be further pursued to find optimal parameters that describe a non constant blank holder force or an optimal time varying punch speed during the quasi-static forming. Those variations may eventually yield better forming results at lower cost. Opposed to earlier approaches by Taebi et al. (2012) the presented approach works within a unified simulation environment. This reduces the number of scripts that have to be written in order to perform function evaluations. It further makes the method more accessible and flexible. Contrary to the typical shape optimization approach the new method does not require parametrizations of certain geometry parts or distance measures. Still, the obtained results are purely numerical, it will be vital to carry out experiments for

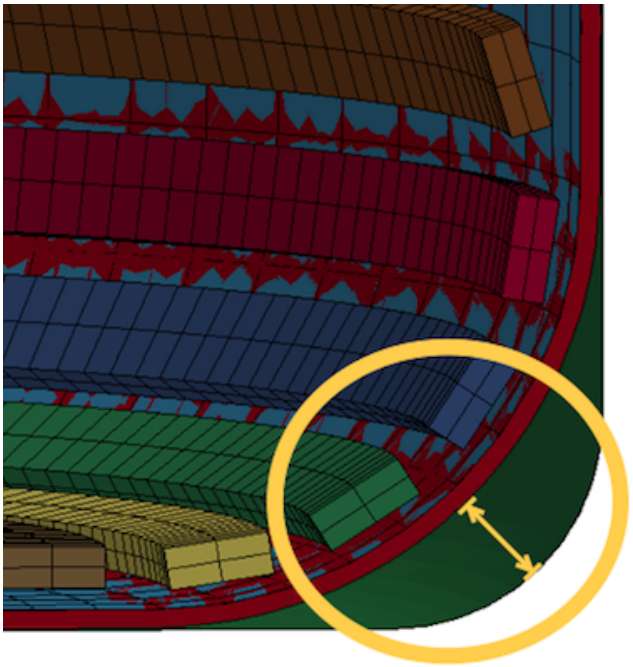


Figure 4.24.: Radius at the bottom of the cup after quasi-static forming.

the verification of the computed results. However, insights regarding the pulse form and triggering current gained by simulations with a double exponential pulse can then be transferred to technologically more accessible pulses.

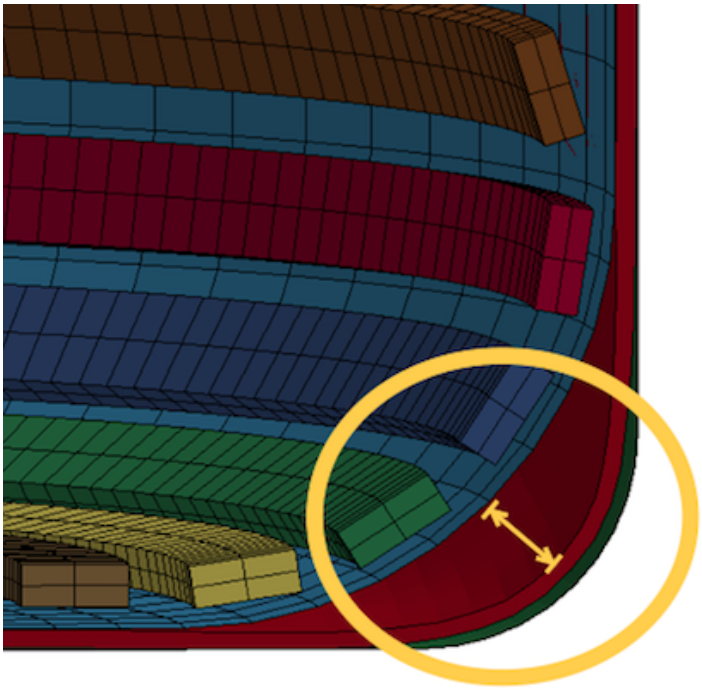


Figure 4.25.: Radius at the bottom of the cup after electromagnetic forming.

5 A first Synopsis

In the preceding chapter a unified approach to tackle the problems arising in virtual process design for sheet metal forming applications has been presented. The challenges to determine material model parameters as well as the difficulties resulting from process parameter identification were addressed. The latter was solved using a novel way of formulating the objective function, which, in contrast to the usual shape optimization approach, relies solely on quantities that are independent of the geometry parametrization, i.e., the major strain. The major strain is usually accessible in all black box mechanics FEM solvers, and can thus be utilized by the employed optimization algorithm. Here, an inner point method was used, as those methods generate a strictly feasible sequence of trail points. By ensuring this any result during the Newton iteration can become useful for practical purposes. If for example computational costs or time are limiting factors points that just yield an improvement of the objective function may be enough. These points of course do not fulfil the KKT conditions, but are still feasible, e.g., solve the underlying PDEs.

The proposed framework is not restricted to the optimization of forming processes, it can, by modifying the underlying simulation, be applied to a broad range of process optimization tasks, since it exploits black box simulation tools. The problem to be solved only needs to be formulated in terms of solutions of the underlying PDE, which then can be provided by any solver capable of solving such PDEs. Changes have to be made in the C/C++ scripts managing the data transfer between the simulation software and the optimization tool, because every commercial program has its own way to store the required quantities.

The same holds for the presented algorithm to identify material model parameters. By using Akima splines for the representation of the experimental results and for-

mulating the distance in terms of a distance between functions, a formulation for the identification problem was found, that can be applied to numerous situations, as long as the experiment can be reproduced by a corresponding simulation. For example, material parameters for describing hyperelastic plastics can be identified by the same method. Or parameters determining a relaxation model for said plastics, by simulating a relaxation test.

At this point the abstract formulation of the FEM in Chapter 2 can be used, to formulate the previously stated more precisely. To this end, let $\Omega \subset \mathbb{R}^d$ be an open, bounded set with Lipschitz boundary $\partial\Omega$. Suppose the parameter depending elliptic second order boundary value problem (cf. Eq. (2.11))

$$\begin{aligned} Lu &= f(\boldsymbol{p}) \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \Gamma_D \subseteq \partial\Omega, \qquad \frac{\partial u}{\partial n} = 0 \quad \text{on } \partial\Omega \setminus \Gamma_D \end{aligned}$$

is given, with $L : H^1_{\Gamma_D} \to L^2(\Omega)$ as in Eq. (2.9) and a function $f(\boldsymbol{p}) \in L^2(\Omega)$, which is assumed to depend twice differentiably on the parameter $\boldsymbol{p} \in \mathbb{R}^m$ for $x \in \Omega$ almost everywhere. For simplicity, all occurring functions are considered real valued.

The weak formulation (cf. Section 2.3) is equivalent to finding $u \in H^1_{\Gamma_D}(\Omega)$ that solves

$$a(u, w) = (f(\boldsymbol{p}), w)_{L^2(\Omega)}, \tag{5.1}$$

for all test functions $w \in H^1_{\Gamma_D}(\Omega)$. The bilinear form a is as in Eq. (2.12) and the space $H^1_{\Gamma_D}(\Omega)$ denotes the Sobolev space of all square integrable functions with weak first order derivatives in L^2 and a zero trace on Γ_D. To avoid technical discussions, Γ_D is assumed to be a sufficiently regular closed subset of $\partial\Omega$. As before in Chapter 2, $(.,.)_{L^2(\Omega)}$ denotes the inner product on $L^2(\Omega)$. Obviously,

$$H^1_{\Gamma_D}(\Omega) = H^1_0(\Omega), \text{ if } \Gamma_D = \partial\Omega.$$

With this abstract setting, the previously discussed framework is suited to solve optimization problems of the form

$$\begin{aligned} \min_{\boldsymbol{p} \in \mathbb{R}^m} \quad & \frac{1}{2} ||B(\boldsymbol{p}) - \bar{u}||^2_{L^2(\Omega)}, \\ \text{subject to} \quad & CB(\boldsymbol{p}) \geq 0. \end{aligned} \tag{5.2}$$

Here, the mapping $B : \mathbb{R}^m \to L^2(\Omega)$, $\boldsymbol{p} \mapsto u$ maps the parameters $\boldsymbol{p} \in \mathbb{R}^m$ to the corresponding state, e.g., a solution of Eq. (5.1), $\bar{u} \in H^1_{\Gamma_D}(\Omega)$ is an optimal, or desired,

state, and $C : H^1_{\Gamma_D}(\Omega) \to H^{-1}_{\Gamma_D}(\Omega)$ is a bounded linear operator mapping into the dual space of $H^1_{\Gamma_D}(\Omega)$ with respect to the L^2 inner product, denoted as $H^{-1}_{\Gamma_D}(\Omega)$. The mapping C expresses that a state u not only is a solution the underlying PDE, but is also amenable to certain restraining laws, e.g., material damage laws in the context of mechanical applications. The extension of Problem (5.2) to a vector-valued problem, as the are needed in the context of mechanics, is just of formal nature, as will be seen in Chapter 6. Of course, the inequality $CB(\boldsymbol{p}) \geq 0$ has to be understood in a weak sense: For any $w \in H^1_{\Gamma_D, w \geq 0, \text{ a.e.}}$ the inequality $\langle CB(\boldsymbol{p}), w \rangle \geq 0$ for the dual pairing holds. The set $H^1_{\Gamma_D, w \geq 0, \text{ a.e.}}$ is a cone, cf. Chapter 3.

It becomes clear that the cost of solving Problem (5.2) depends on the cost of evaluating B, i.e., of solving Eq. (5.1) numerically for u. When formulated as a material parameter identification problem, as in Section 4.2, the costs are nearly admissible, but in the latter case, when used to optimize a whole production process, the computation time is the main limiting factor. In the application discussed in Section 4.3 the simulation of a single forming process took nearly one day ($\sim 18\,\text{h}$), on an adequate computer mainframe. The evaluation of the constraints in (5.2) again needs a complete simulation, as does the approximation of the derivatives via the FD method for each dimension of the parameter vector.

A potential solution to this problem are model order reduction (MOR) methods. Benner et al. (2013) give a survey on MOR methods especially for parametric systems. The key is, that during the so called *offline* phase a huge number of simulation results is acquired, allowing for an efficient harvesting of the solution space during the *online* phase. Applying MOR to the introduced method is a promising perspective.

Another possible approach to overcome the problem of high computational costs is inspired by topology optimization, as is discussed in Bendsøe and Sigmund (2013). In this approach the underlying PDE becomes a constraint, and by appropriate design parametrization, there is no need to simulate the problem over and over again. A popular method to obtain solutions to a topology optimization problem is the *simultaneous analysis and design* (SAND) method, as introduced by Haftka (1984) and Haftka and Kamat (1989).

In a more general context the SAND method is a variant of approaching the solution of a *PDE constrained optimization problem*, see, e.g., Hinze et al. (2010). These problems arise in numerous applications, all of which are governed by a single PDE, or many coupled PDEs. In fact Problem (5.2) can be seen as a PDE constrained optimization problem, in the sense, that the objective is a cost function, and the constraints include the equilibrium conditions given in Eq. (5.1), as will be seen in the following section.

In this framework the solution strategy employed in the preceding chapter is called *nested analysis and design* (NAND) or *back-box* method (BB-method), cf. Biegler et al. (2003). As argued by Choi (2012) the BB-method is usually the method of choice if the PDE solver and the optimization problem solver are only weakly coupled. It also offers a broad range of flexibility, as discussed earlier. However, due to its higher efficiency, and, among other advantages, relative straightforward implementation the SAND method is of growing interest for applications, see, e.g., Arora and Wang (2005). It has been coupled to IP methods by Canelas et al. (2008), in the context of boundary element methods for the discretization of the underlying PDE. Biegler and Wächter (2003) discuss the general coupling of the SAND to so called SQP methods to solve the resulting optimization problem. The remainder of this work is concerned with the reformulation of Problem (5.2) as a SAND problem and solving it with the IP method. Potential benefits when applying this formulation to sheet metal forming are discussed.

6 The SAND Method

In this chapter the SAND approach for the identification of process parameters in sheet metal applications is introduced. An adequate problem formulation in terms of L^2 functions is derived in Section 6.1. The resulting problem is solved by pursuing the the *first discretize, then optimize* framework and exploiting the advantages of IP methods to treat a huge number of constraints efficiently. A family of problems for the analysis of a model problem is discussed in Section 6.2. For a linear elastic material law a problem similar to the optimal shape problem in Section 4.3 is introduced with various discretizations and different prescribed forms. The solutions to the SAND formulations are thoroughly discussed. Before closing with a short conclusion in Section 6.4 the derived approach is compared to the black box algorithm in Section 6.3. It will be shown that the SAND approach is potentially 100 times more efficient than the approach discussed in the preceding chapter.

6.1. Problem Formulation

Following Haftka and Kamat (1989) the SAND formulation to solve a structural design problem is given as follows. Let $\boldsymbol{u} \in \mathbb{R}^n$ be a vector of displacements, and let $\boldsymbol{p} \in \mathbb{R}^m$ be a vector of design variables. Then the simultaneous problem reads:

$$\begin{aligned} \min_{\boldsymbol{u}\in\mathbb{R}^n, \boldsymbol{p}\in\mathbb{R}^m} \quad & f(\boldsymbol{p}), \\ \text{subject to} \quad & g_j(\boldsymbol{u}, \boldsymbol{p}) \geq 0, \quad j = 1, \ldots, k, \\ & F(\boldsymbol{u}, \boldsymbol{b}, \boldsymbol{p}) = 0. \end{aligned} \tag{6.1}$$

Here, f is assumed to be an objective function depending on the design variables, g_j are constraints, like stress and displacement constraints, and the function F represents

the (under circumstances non linear) system of algebraic equations the displacements must fulfill, with a load vector $\boldsymbol{b}$. Naturally, the objective can also depend on the displacements $\boldsymbol{u}$, without changing the overall structure if the optimization task.

This early problem formulation was focused in solving structural design problems, but in recent years the SAND method has been classified as a method to tackle general PDE constrained optimization problems, as argued by Arora and Wang (2005). In the following this should be highlighted. To this end Problem (6.1) will be reformulated akin to Problem (5.2). Suppose, as in Chapter 5, the following elliptic boundary value problem

$$\begin{aligned} Lu &= f(\boldsymbol{p}) \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \Gamma_D \subseteq \partial\Omega, \qquad \frac{\partial u}{\partial n} = 0 \quad \text{on } \partial\Omega \setminus \Gamma_D, \end{aligned}$$

on a Lipschitz domain $\Omega \in \mathbb{R}^d$. Given the weak formulation to the boundary value problem (cf. Eq. (5.1))

$$a(u, w) = (f(\boldsymbol{p}), w)_{L^2(\Omega)},$$

for all test functions $w \in H^1_{\Gamma_D}(\Omega)$, an optimal state $\bar{u} \in H^1_{\Gamma_D}(\Omega)$, and the bounded linear operator mapping $C : H^1_{\Gamma_D}(\Omega) \to H^{-1}_{\Gamma_D}(\Omega)$ Problem (5.2) as a SAND problem is:

$$\begin{aligned} \min_{u \in H^1_{\Gamma_D}(\Omega), \boldsymbol{p} \in \mathbb{R}^n} \quad & \frac{1}{2} ||u - \bar{u}||^2_{L^2(\Omega)}, \\ \text{subject to} \quad & a(u, w) - (f(\boldsymbol{p}), w)_{L^2(\Omega)} = 0 \quad \forall\, w \in H^1_{\Gamma_D}(\Omega), \\ & \langle Cu, w \rangle_{H^{-1}_{\Gamma_D}(\Omega), H^1_{\Gamma_D}(\Omega)} \geq 0 \quad \forall\, w \in H^1_{\Gamma_D, w \geq 0, \text{ a.e.}}(\Omega). \end{aligned} \tag{6.2}$$

By Riesz' Representation Theorem, Theorem 2.3.2, there exists a linear operator $A : H^1_{\Gamma_D} \to H^{-1}_{\Gamma_D}$ such that problem (6.2) can be written compactly as:

$$\begin{aligned} \min_{u \in H^1_{\Gamma_D}(\Omega), \boldsymbol{p} \in \mathbb{R}^n} \quad & \frac{1}{2} ||u - \bar{u}||^2_{L^2(\Omega)} \\ \text{subject to} \quad & Au - f(\boldsymbol{p}) = 0, \\ & Cu \geq 0. \end{aligned} \tag{6.3}$$

Here, in order to simplify the inequality Riesz' Representation Theorem for Gel'fand triplets has been applied, see Gelfand et al. (1964).

In the framework of PDE constrained optimization problems u is usually referred to as the state variable and $\boldsymbol{p}$ as the control, or design variable, see e.g., Hinze et al. (2010).

As pointed out by Choi (2012) the distinction between design and control depends on the application at hand and is rather of historical nature then of formal. In problems originating from structural design $\boldsymbol{p}$ is called design variable, in problems coming from control theory, e.g., temperature or flow control, $\boldsymbol{p}$ is denoted a control variable. Hence, the expression optimal control is also often used as a synonym for PDE constrained optimization.

Contrary to usual optimal control problems the control, or design, variable $\boldsymbol{p}$ in the presented formulation (6.3) is already discrete, i.e., it is of finite dimension. If that is not the case, regularization techniques as introduced in Tikhonov and Arsenin (1977), have to be employed in order to identify a solution the optimization problem. Here, this is not needed and the objective function can remain unchanged, as $\boldsymbol{p}$ does not need to be approximated by a finite dimensional counterpart.

At this point a crucial decision has to be made, whether the approach *first optimize, then discretize* or the concept *first discretize, then optimize* will be applied. In this work the latter approach is pursued. As is argued by Hinze et al. (2010) there is, up to now, no general criterion which of the two approaches should be preferred. The *first discretize, then optimize* is chosen here, since on the one hand the original SAND formulation is given in discrete terms and on the other hand, it seems reasonable with regard to applications in sheet metal forming: the technologically relevant processes are mostly given as multiphysyical problems, where the structure of the underlying PDEs is so complex, that the resulting discretization is usually aligned with simplifications motivated by experiments. Thus, a deeper analysis of the governing operators and their adjoints, as is needed for the derivation of appropriate optimality conditions for the continuous problem, is often cumbersome, if not completely out of question. Moreover, the *first discretize, then optimize* approach can potentially be incorporated into existing FEM modeling tools used in engineering applications, as argued by Arora and Wang (2005). Further discussions on the drawbacks and advantages of both approaches can be found in Hinze and Tröltzsch (2010). A comparison of different solution strategies, like the SAND method and BB-method can be found in Arora and Wang (2005).

Applying an adequate discretization scheme, e.g. Ritz-Galerkin finite elements as in Section 2.4, renders Problem (6.3) into a variant of (6.1). When mechanical problems are approached, the equations of course become vectorial, the transition to a vector valued formulation is canonical. In the resulting discrete version the objective function, the distance to a desired state, becomes the usual euclidean distance for the displacements. The constraints are replaced by the corresponding algebraic representations, these may be non-linear functions, which usually appear when the mechanical situation is

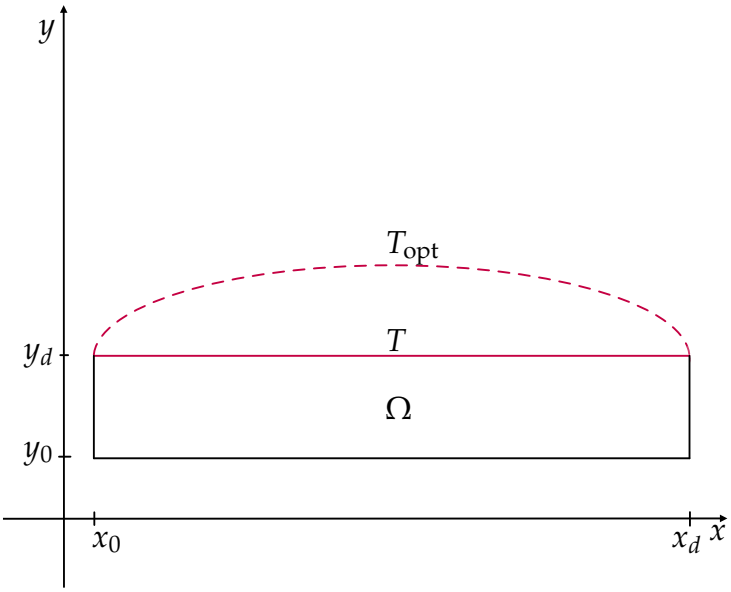


Figure 6.1.: A schematic of the model problem. A bar with dimension $[x_0, x_d] \times [y_0, y_d]$ is to be deformed such, that the top T is as close to the prescribed curve T_{opt} as possible.

governed by plasticity. However, since a candidate for a solution $\boldsymbol{u}_j$ is proposed in the $j-$th step by the optimization procedure (cf. Chapter 3), there is no need to solve the constraints for $\boldsymbol{u}_j$. In case of a linear material law, as in Section 4.1.1, the constrain evaluation reduces to a multiplication of a matrix and a vector, as will be seen in Section 6.2. The earlier introduced linear elastic material law will be used to show how above derived problem formulation can be applied to process parameter identification in a forming application. The outlined approach is sometimes also referred to as *Full Newton SAND*, since Newton's method is used to find a KKT point to Problem (6.1), cf. Hazra (2009).

6.2. A 2D Test Problem

6.2.1. The Model geometry and Optimization Problem

The 2D model problem under consideration is given as follows. Let $\Omega = [x_0, x_d] \times [y_0, y_d] \subset \mathbb{R}^2$, with $0 \leq x_0, y_0 < y_d < x_d$ be a rectangular domain as shown in Fig. 6.1. Assuming isotropic linear elasticity with plane stress and small deformations, as discussed in Section 4.1.1, yields the following PDE

$$\text{div } \mathsf{C}\varepsilon(u) = -f(\boldsymbol{p}), \text{ with } \varepsilon(u) = \frac{1}{2}\left(\nabla u + \nabla u^\top\right), \tag{6.4}$$

where u is the deformation field, ε the second order symmetric strain tensor, as in Eq. (4.6), and f the force density depending on an $m-$dimensional parameter vector $\boldsymbol{p} \in \mathbb{R}^m$.

The fourth order *stiffness tensor* is denoted by C. It is given in direct tensor notation as

$$\mathsf{C}\tau = 2\mu\left(\tau + \lambda \text{tr}\tau \boldsymbol{I}\right),$$

where μ and λ are the Lamé constants, see Eq. (4.10), tr denotes the trace operator, $\boldsymbol{I}$ is the unity matrix and τ is a symmetric second order tensor. Comparing to Hooke's Law $\varepsilon = \mathsf{S}\sigma$ yields,

$$\mathsf{C} = \mathsf{S}^{-1}.$$

For all boundaries perpendicular to the $x-$axis homogeneous Dirichlet boundary conditions are prescribed, at all boundaries parallel to the $x-$axis homogeneous Neumann conditions are employed, i.e., $\Gamma_N = \{(x,y)^\top \in \partial\Omega : \ x = x_0, \text{ or } x = x_d\}$ and $\Gamma_D = \{(x,y)^\top \in \partial\Omega : \ y = y_0, \text{ or } y = y_d\}$. The aim is to find an optimal parameter vector $\boldsymbol{p}^* \in \mathbb{R}^m$, such that the distance between the top of Ω given by $T = \{(x, y_d) : x \in [x_0, x_d]\}$ and the prescribed bounding curve T_{opt} is minimal.

According to Section 2.3, the weak formulation of the linear elasticity problem can be written as

$$\int_\Omega \mathsf{C}\epsilon(u) : \epsilon(w)\, \mathrm{d}x = \int_\Omega f(\boldsymbol{p}) \cdot w \, \mathrm{d}x \quad \forall\, w \in H^1_{\Gamma_D}(\Omega), \tag{6.5}$$

where

$$\mathsf{C}\epsilon(u) : \epsilon(w) := \sum_{i,j=1}^{2} \left(\mathsf{C}\epsilon(u)\right)_{ij} \left(\epsilon(w)\right)_{ij}.$$

Here, $H^1_{\Gamma_D}(\Omega)$ denotes the Sobolev space of all square integrable functions with weak first order derivatives in L^2 and a zero trace on Γ_D. Further, $(.,.)_{L^2(\Omega)}$ denotes the inner product on $L^2(\Omega)$. Note that $H^1_{\Gamma_D}(\Omega) = H^1_0(\Omega)$ if $\Gamma_D = \partial\Omega$.

In order to treat Eq. (6.5) numerically a Ritz-Galerkin finite element method is applied to Eq. (6.5). Piecewise linear basis functions on a triangular mesh are chosen and $N \in \mathbb{N}$ degrees of freedom are located at the vertices of the elements, as discussed in Section 2.4.

The weak formulation Eq. (6.5) can thus be written in discretized form as a system of linear equations

$$\boldsymbol{S}_h \boldsymbol{u}_h = \boldsymbol{b}_h(\boldsymbol{p}), \tag{6.6}$$

where $\boldsymbol{S}_h \in \mathbb{R}^{N \times N}$ denotes the global stiffness matrix, $\boldsymbol{u}_h \in \mathbb{R}^N$ denotes the discretized displacement field, i.e., the displacement vector and $\boldsymbol{b}_h \in \mathbb{R}^N$ is the load vector, which here depends on the finite set of parameters $\boldsymbol{p}$ to be identified. For the sake of simplicity the subscript h is omitted in the following. However Eq. (6.7) strongly depends on h, the level of refinement of the mesh (see Section 2.4). Following the concept in Section 6.1, the optimization problem can be written as:

$$\begin{aligned} \min_{\boldsymbol{u} \in \mathbb{R}^N, \boldsymbol{p} \in \mathbb{R}^m} \quad & \frac{1}{2} \sum_{i \in \mathcal{T}} \|\boldsymbol{u}_i - \bar{\boldsymbol{u}}_i\|_2^2, \\ \text{subject to} \quad & \boldsymbol{S}\boldsymbol{u} = \boldsymbol{b}(\boldsymbol{p}). \end{aligned} \tag{6.7}$$

In the above objective function $\mathcal{T}$ denotes the index set that takes all the displacements belonging to nodes at the top of the structure.

Optimization Problem (6.7) is a quadratic problem in the vector of displacements $\boldsymbol{u}$. The equality constraints are linear in $\boldsymbol{u}$ but not necessarily in $\boldsymbol{p}$. Depending on $\boldsymbol{b}(\boldsymbol{p})$ the equality constraints are not even convex.

In the following numerical studies solutions to Eq. (6.7) will be computed for various boundary curves T_{opt} and for a collection of discretizations of the geometry.

6.2.2. Numerical Studies

For the numerical analysis Ω is assumed to be a metal sheet with length 100 mm and thickness 1.5 mm with parameters given in Table 6.1. Further, six different discretizations are taken into account. The mesh of the domain is created by using the meshing and computer aided design (CAD) software Trelis. The total number of elements and nodes for the different instances is given in Table 6.2. Since the optimization problem Eq. (6.7) has only to be solved for those displacements belonging to nodes which are not equipped with Dirichlet boundary conditions, the number of these *free nodes* is also given in Table 6.2.

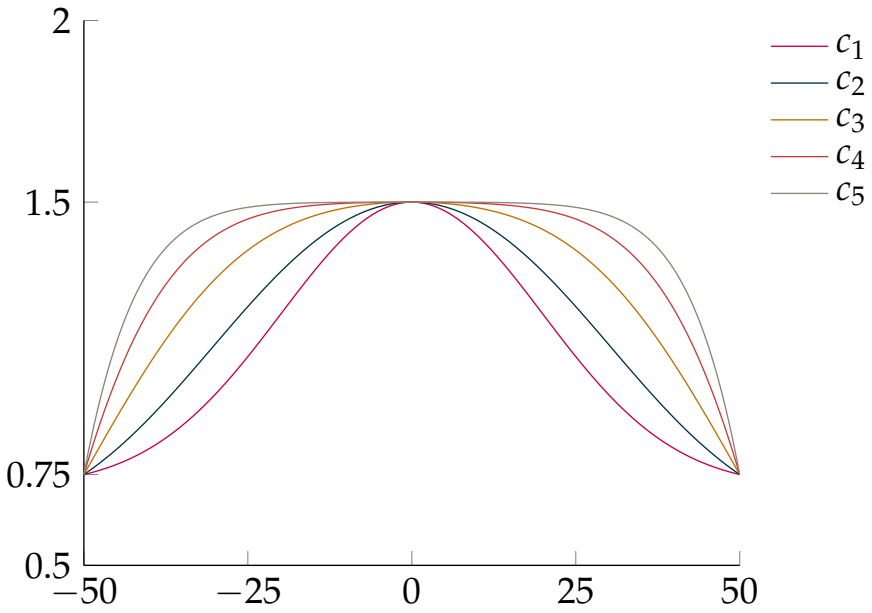


Figure 6.2.: The five different optimal shape curves resulting from Eq. (6.8).

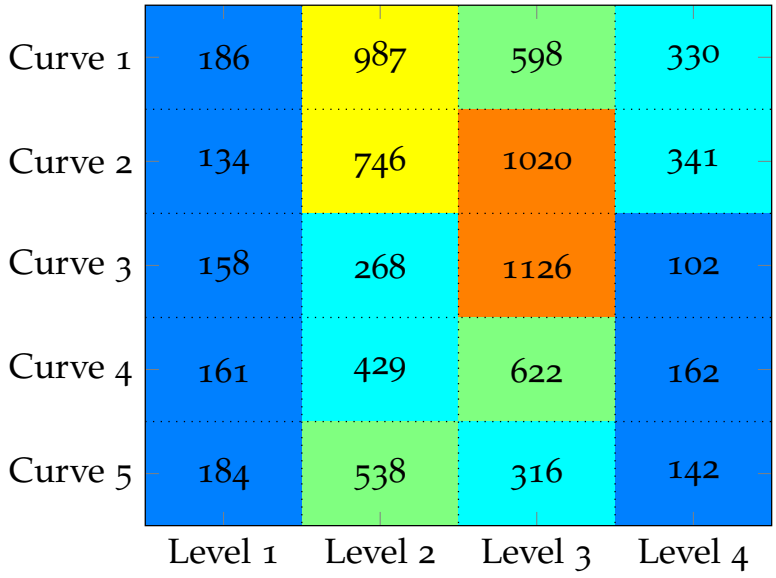


Figure 6.3.: The total number of iterations for every computed instance. The numbering on the x-axis refers to the discretization levels introduced in Table 6.2, the y-axis indicates the ideal bounding curves given in Fig. 6.2.

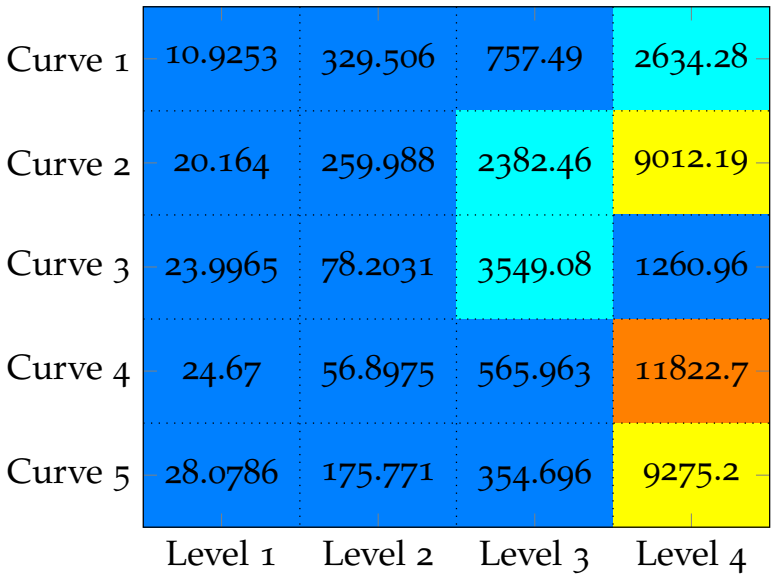


Figure 6.4.: The total time in s IPOT needed to solve every considered problem instance. The numbering on the x-axis refers to the discretization levels introduced in Table 6.2, the y-axis indicates the ideal bounding curves given in Fig. 6.2.

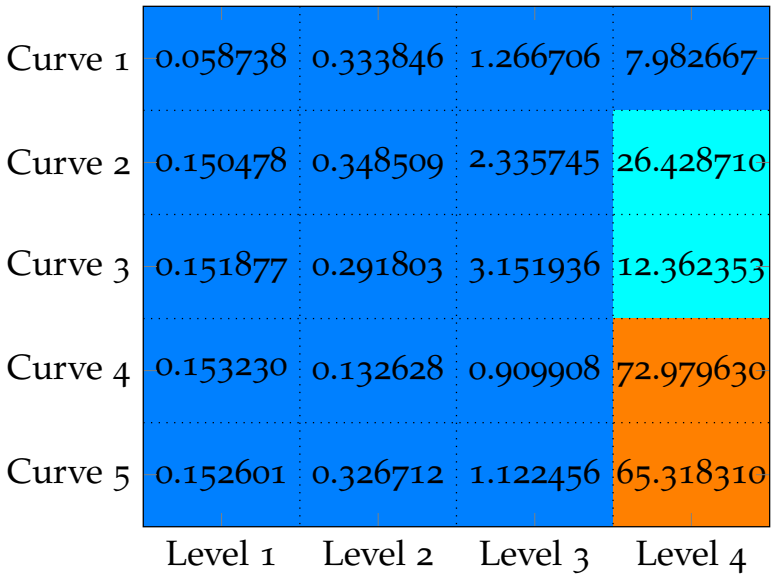


Figure 6.5.: Above figure shows the mean time per iteration step for every considered problem instance in s.

Five different functions $c_i : [x_0, x_d] \to \mathbb{R}$, for $i = 1, \ldots, 5$, each yielding an optimal shape T_{opt} are taken into account for the numerical analysis. They are constructed as sums of Gaussian functions, which are transformed to the appropriate value range.

To this end let $n : [x_0, x_d] \to \mathbb{R}$ be the normal distribution's probability density function with parameters μ and σ, i.e.,

$$n(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

The transformation function T_f, yielding values for this particular situation is:

$$T_f(\zeta) = 0.75 + 0.75 \frac{\zeta - n(50|\mu,\sigma)}{n(0|\mu,\sigma) - n(50|\mu,\sigma)}.$$

With this the bounding functions c_i, $i = 1 \ldots, 5$, read:

$$\begin{aligned}
c_1(x) &= T_f\left(n(x|\mu_1, 20)\right), \\
c_2(x) &= T_f\left(\sum_{j=1}^{3} n(x|\mu_j, 20)\right), \\
c_3(x) &= T_f\left(\sum_{j=1}^{5} n(x|\mu_j, 20)\right), \\
c_4(x) &= T_f\left(\sum_{j=1}^{7} n(x|\mu_j, 20)\right), \\
c_5(x) &= T_f\left(\sum_{j=1}^{9} n(x|\mu_j, 20)\right).
\end{aligned} \tag{6.8}$$

The different expectation values μ_j are the components of the vector

$$\boldsymbol{\mu} = (0, 20, -20, 40, -40, 60, -60, 80, -80)^\top.$$

The functions resulting from Eq. (6.8) are depicted in Fig. 6.2.

For every possible upper boundary configuration given by the functions c_i, $i = 1 \ldots, 5$ the x and y component of the desired optimal deformation vector $\bar{\boldsymbol{u}}$ in problem Eq. (6.7)

x_0	x_d	y_0	y_d	λ	μ
$-50\,\text{mm}$	$50\,\text{mm}$	$-0.75\,\text{mm}$	$0.75\,\text{mm}$	$115.3846\,\text{GPa}$	$76.9231\,\text{GPa}$

Table 6.1.: Geometry and material parameters for the sheet Ω assumed for the numerical analysis.

are constructed as:

$$\bar{\boldsymbol{u}} = \begin{pmatrix} x \\ c_i(x) \end{pmatrix}. \tag{6.9}$$

The considered volume force density $\mathrm{f}(\boldsymbol{p})$, depending on a parameter vector $\boldsymbol{p} \in \mathbb{R}^{20}$, has two components for every point $(x, y)^\top \in \Omega$. The $x-$component is zero everywhere, i.e., $f_x \equiv 0$, whereas the $y-$component is given in terms of the parameter vector's components

$$\boldsymbol{p} = (\alpha_1, \beta_1, \gamma_1, \delta_1, \alpha_2, \beta_2, \gamma_2, \delta_2, \alpha_3, \beta_3, \gamma_3, \delta_3, \alpha_4, \beta_4, \gamma_4, \delta_4)^\top$$

and reads:

$$f_y(\boldsymbol{p}) = \sum_{i=1}^{4} \alpha_i \mathrm{e}^{\beta_i (x-\delta_i)} \mathrm{e}^{-\gamma_i y}. \tag{6.10}$$

A similar problem was studied in an earlier work, see Rozgič and Stiemer (2014a). There, only a single optimal shape was considered and the force density was only depending on four parameters:

$$f_y(\alpha, \beta, \gamma, \delta) = \alpha \mathrm{e}^{\beta (x-\delta)^2} \mathrm{e}^{-\gamma y}.$$

The number of target variables is higher than in Rozgič and Stiemer (2014a), which is negligible, since the number of force parameters is still considerably smaller than twice the number of free nodes. However, the density defined in Eq. (6.10) is able to produce broader shapes and is therefore more suitable for the task at hand. The introduced function can be differentiated analytically, therefore no numerical approximation schemes for the computation of the Jacobian or the Hessian are needed.

Discretization level	# Elements	# Nodes	# Free Nodes
1	170	172	168
2	680	513	507
3	2720	1705	1695
4	10880	6129	6111

Table 6.2.: The number of elements, nodes and free nodes for six different levels discretization, which are considered for the numerical analysis.

The first order partial derivatives of f_y at any point $(x, y)^\top \in \Omega$ for every $i = 1, \dots, 4$ read as:

$$\begin{aligned}
\frac{\partial f_y}{\partial \alpha_i} &= \mathrm{e}^{-\gamma_i y}\, \mathrm{e}^{\beta_i (\delta_i - x)}, \\
\frac{\partial f_y}{\partial \beta_i} &= -\alpha_i\, \mathrm{e}^{-\gamma_i y}\, \mathrm{e}^{\beta_i (\delta_i - x)}\, (\delta_i - x), \\
\frac{\partial f_y}{\partial \gamma_i} &= -\alpha_i\, y\, \mathrm{e}^{-\gamma_i y}\, \mathrm{e}^{-\beta_i (\delta_i - x)}, \\
\frac{\partial f_y}{\partial \delta_i} &= \alpha_i\, \beta_i\, \mathrm{e}^{-\gamma_i y}\, \mathrm{e}^{\beta_i (\delta_i - x)}.
\end{aligned} \tag{6.11}$$

The second order partial derivatives are consequently given by:

$$\begin{aligned}
\frac{\partial^2 f_y}{\partial \alpha_i \partial \alpha_i} &= 0, \\
\frac{\partial^2 f_y}{\partial \alpha_i \partial \beta_i} &= \mathrm{e}^{-\gamma_i y} \mathrm{e}^{\beta_i (\delta_i - x)} (\delta_i - x), \\
\frac{\partial^2 f_y}{\partial \alpha_i \partial \gamma_i} &= -y \mathrm{e}^{-\gamma_i y} \mathrm{e}^{\beta_i (\delta_i - x)}, \\
\frac{\partial^2 f_y}{\partial \alpha_i \partial \delta_i} &= \beta_i \mathrm{e}^{-\gamma_i y} \mathrm{e}^{\beta_i (\delta_i - x)}, \\
\frac{\partial^2 f_y}{\partial \beta_i \partial \beta_i} &= \alpha_i \mathrm{e}^{-\gamma_i y} \mathrm{e}^{\beta_i (\delta_i - x)} (\delta_i - x)^2, \\
\frac{\partial^2 f_y}{\partial \beta_i \partial \gamma_i} &= -\alpha_i y \mathrm{e}^{-\gamma_i y} \mathrm{e}^{\beta_i (\delta_i - x)} (\delta_i - x), \\
\frac{\partial^2 f_y}{\partial \beta_i \partial \delta_i} &= \alpha_i \mathrm{e}^{-\gamma_i y} \mathrm{e}^{\beta_i (\delta_i - x)} \left(1 + \beta_i \left(\delta_i - x\right)\right) \\
\frac{\partial^2 f_y}{\partial \gamma_i \partial \gamma_i} &= \alpha_i y^2 \mathrm{e}^{-\gamma_i y} \mathrm{e}^{\beta_i (\delta_i - x)}, \\
\frac{\partial^2 f_y}{\partial \gamma_i \partial \delta_i} &= -\alpha_i \beta_i y \mathrm{e}^{-\gamma_i y} \mathrm{e}^{\beta_i (\delta_i - x)}, \\
\frac{\partial^2 f_y}{\partial \delta_i \partial \delta_i} &= \alpha_i {\beta_i}^2 \mathrm{e}^{-\gamma_i y} \mathrm{e}^{\beta_i (\delta_i - x)}.
\end{aligned} \tag{6.12}$$

Here, only 40 terms have to be computed due to symmetry. The contribution to the systems Hessian matrix, however, consists of a 20×20 matrix accumulating these 40 values evaluated at integration points coming from the finite element discretization.

For all instances, resulting from the different discretizations and the desired optimal shapes a solution shall be computed. All computations are performed with IPOPT,

specifically the Matlab interface has been used. To this end *callback functions* were written. These provide routines to evaluate the objective function, the constraints, and all needed derivatives (first and second order) at points $(\boldsymbol{u}_i, \boldsymbol{p}_i)$ determined by the optimization algorithm. To obtain the needed stiffness matrices and load vectors, the assembly routines of the software package *o*FEM by Dudzinski et al. (2016) have been used.

Further, to ensure numerical stability, box constraints were imposed on the displacements $\boldsymbol{u}$ and the components of the parameter vector $\boldsymbol{p}$:

$$\begin{aligned} -1 &\leq \boldsymbol{u} \leq 1, \\ -10 &\leq \alpha_i \leq 10, \quad i = 1, \ldots, 4, \\ 0 &\leq \beta_i \leq 50, \quad i = 1, \ldots, 4, \\ 0 &\leq \gamma_i \leq 1, \quad i = 1, \ldots, 4, \\ -1 &\leq \delta_i \leq 1, \quad i = 1, \ldots, 4. \end{aligned}$$

Almost all problem instances, resulting from the different meshes (see Table 6.2) and bounding curves introduced in Eq. (6.8), were solved up to a tolerance of 10^{-8}, only the coarsest mesh in combination with the first boundary curve could only be solved up to a tolerance level of 10^{-4}. The maximum number of iterations was set to 4 000. Consequently, IPOPT was invoked with the following options:

```
%% IPOPT options used
options.ipopt.mu_strategy = 'adaptive';
options.ipopt.tol         = 1e-8; %1e-4 for mesh 1 curve 1
options.ipopt.max_iter    = 4000;
```

For all problems the initial points were set to 0, i.e., no displacement, and no acting force.

IPOPT found optimal solutions for all given problems. Only a few are shown and discussed here, a thorough overview of the computed results can be found in Appendix A.

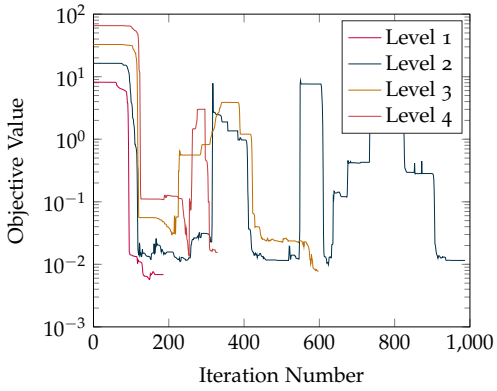
In Fig. 6.3 the number of iterations IPOPT needed to find an optimal solution is shown for every considered problem. It can be seen that all optimization programs were solved within less than half of the maximum of the allowed steps. Especially the bigger instances, i.e. Level 4, were solved with a relative small number of iterations. The total computation time for each example is found in Fig. 6.4. Naturally the time is increasing with the number of free nodes of the FEM model, as the number of constraints grows with it. In Fig. 6.5 the average time per iteration step is shown. It stays within the

range of 1 minute per step (the maximum is $\approx 73\,\mathrm{s}$), mostly the average time per step is below 1 s.

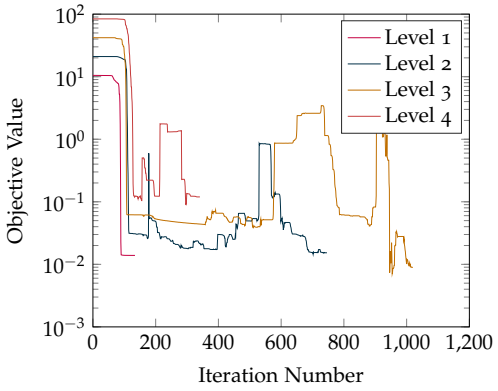
Although, the combination Level 1 with Curve 1 could only be solved to a relative tolerance of 10^{-4}, the result looks fine, see Fig. 6.7(a). The final objective value is $6.8451 \cdot 10^{-3}$. Figs. 6.7(b) and 6.7(c) suggest that a possible reason for the relatively low solution precision is that there exists no combination with a primal and dual feasible point, i.e., a saddle point of the Lagrangian, satisfying a tolerance lower than 10^{-4}. Because all other problem instances could be solved to a precision of 10^{-8}, the reason for this event is unlikely to be a structural one. This would be, if the general problem definition would not allow for a common point of the primal and dual cone. It seems more likely that the combination of the used mesh and the particular boundary doesn't allow for a better solution.

When looking at the more extreme combinations, e.g. Level 4 with Curve 5, as shown in Fig. 6.8(a), it can be seen that an identified optimal solution, does not necessarily yield a smoothly fitting form. Due to the nature of the applied force as a superposition of pulses, structures with waves are likely to be formed. That makes it difficult to obtain sharp edged cups like Curve 5. From Figs. 6.8(b) and 6.8(c), however, it can be seen that the obtained solutions are indeed a local minimum of the optimization problem, to a high precision. It is further interesting to see that the objective function value drops in the last third of the iteration and remains nearly constant and only the primal and dual feasibility drop to the desired tolerance.

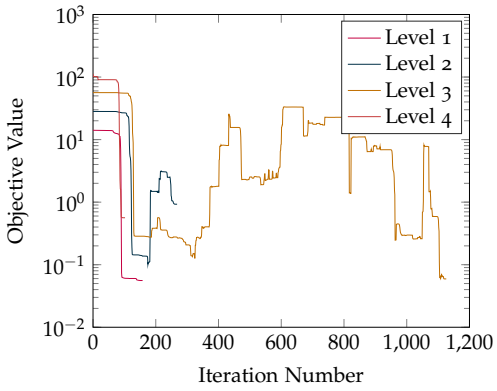
A similar behavior can be seen when looking at the development of the objective function during the iterations for each boundary curve over the different meshes, as shown in Fig. 6.6. In most cases the objective value also drops during the last third of the iterations. For those instances where the number of iterations is high, see also Fig. 6.3, the objective value fluctuates indicating that primal and dual feasibility were not reached, yet. A convergence for a growing number of elements can not be seen. It may be possible that the number of iterations can be reduced, by carefully adjusting the according tolerance values.



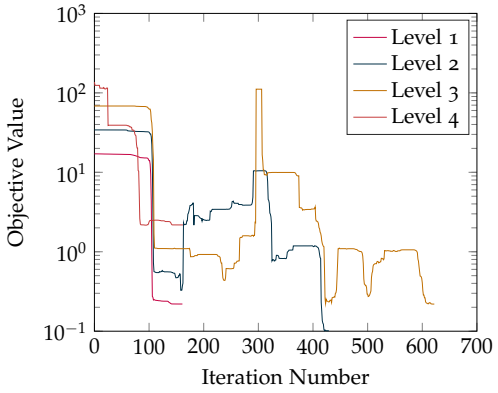
(a) Objective functions decrease during iteration for Curve 1.



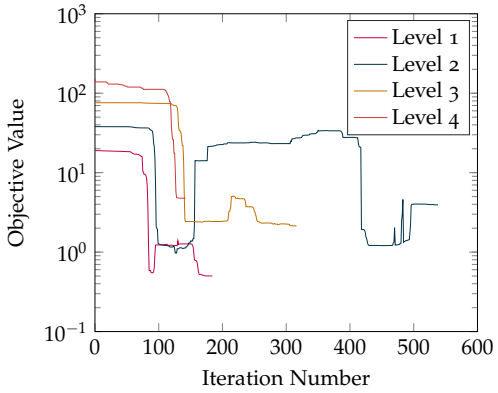
(b) Objective functions decrease during iteration for Curve 2.



(c) Objective functions decrease during iteration for Curve 3.

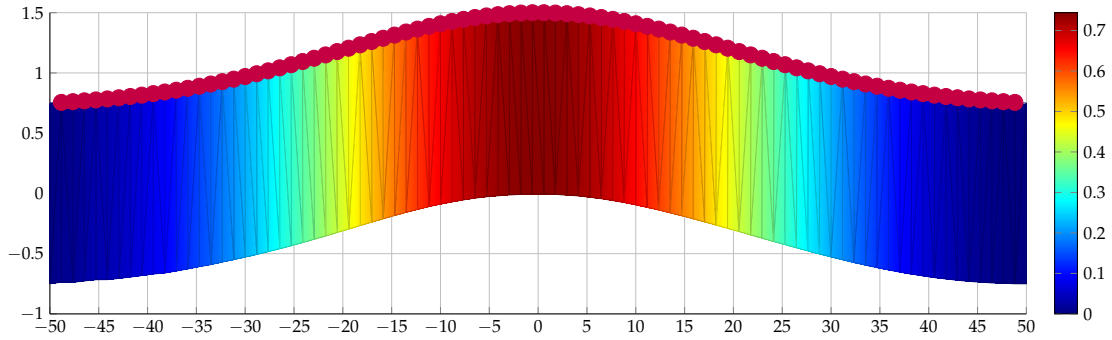


(d) Objective functions decrease during iteration for Curve 4.

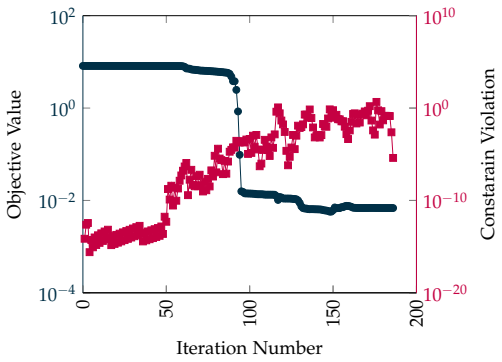


(e) Objective functions decrease during iteration for Curve 5.

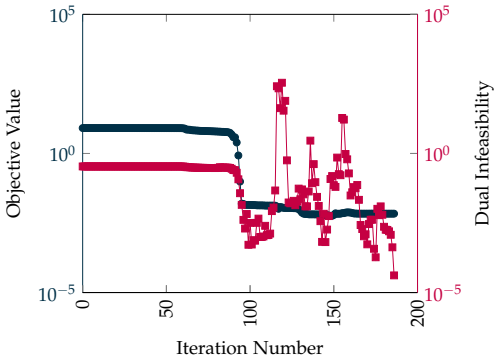
Figure 6.6.: The course of the objective function value during the iteration of IPOPT for each boundary curve and the respective meshes is shown.



(a) Final configuration

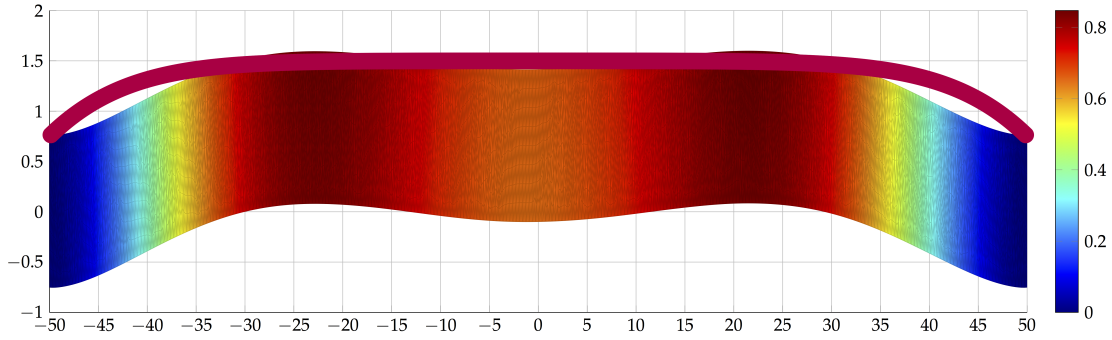


(b) Objective value and constraint violation.

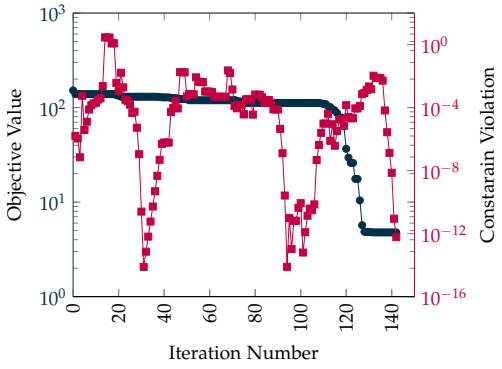


(c) Objective value and dual infeasibility.

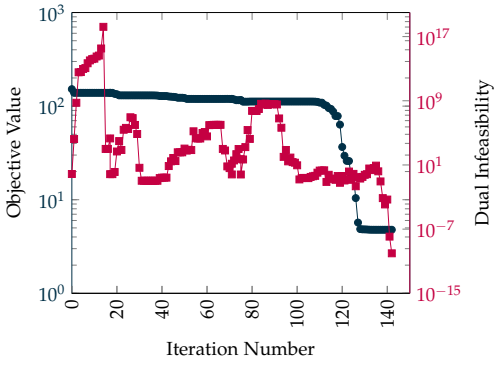
Figure 6.7.: The final configuration computed with IPOPT for Level 1 and Curve 1 is depicted in Fig. 6.7(a). The coloring gives the absolute displacement. Also the change of the objective function value (blue line, left axis), the constraint violation (red line, right axis, Fig. 6.7(b)) and the dual infeasibility (red line, right axis Fig. 6.7(c)) during the iteration for Level 1, Curve 1 are shown. For convenience all values are given logarithmically.



(a) Final configuration



(b) Objective value and constraint violation.



(c) Objective value and dual infeasibility.

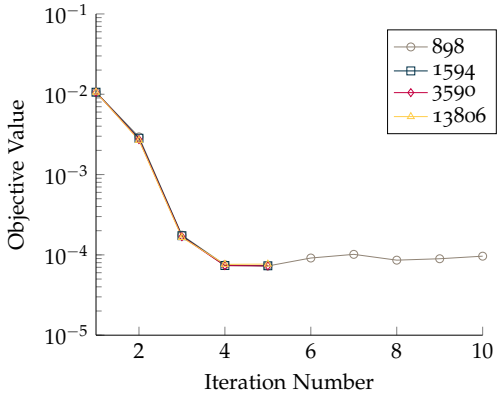
Figure 6.8.: The final configuration computed with IPOPT for Level 4 and Curve 5 is depicted in Fig. 6.8(a). The coloring gives the absolute displacement. Also the change of the objective function value (blue line, left axis), the constraint violation (red line, right axis, Fig. 6.8(b)) and the dual infeasibility (red line, right axis Fig. 6.8(c)) during the iteration for Level 4, Curve 5 are shown. For convenience all values are given logarithmically.

6.3. SAND vs the Black Box Method

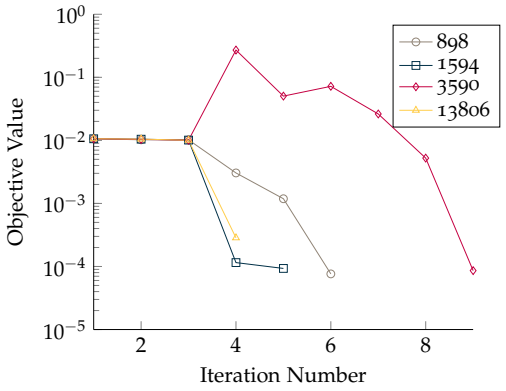
In this section the derived SAND formulation should be compared to the previously discussed black box method (BB-method). To this end, similar to Section 6.2, a two dimensional sheet deformation problem under a linear elastic material law is considered. The sheet has a thickness of 1 mm and a length of 150 mm. The same material parameters as in Table 6.1 are assumed. The prescribed optimal shape is constructed analogously to Eq. (6.8) by a linear combination of five Gaussians, linearly transformed to the appropriate domain. For the sake of simplicity the $y-$ component of the force density during this investigations has only four parameters, following Rozgič and Stiemer (2014a), i.e, $f_y(\alpha, \beta, \gamma, \delta) = \alpha e^{\beta(x-\delta)^2} e^{-\gamma y}$.

# Elements	α	β	γ	δ	f^*_{SAND}	No. Iter.
898	1.1873	-1.9814	-0.8184	0.8966	$4.5 \cdot 10^{-3}$	5
1594	1.4815	-0.3466	-0.1318	1.0396	$6.8 \cdot 10^{-3}$	6
3590	0.3252	-9.5692	-1.8476	0.7580	$5.7 \cdot 10^{-3}$	8
13806	0.369	-9.3043	-0.3697	0.7595	$1.1 \cdot 10^{-2}$	7
# Elements	α	β	γ	δ	$f^*_{\text{BB-num}}$	No.Iter
898	1.5413	-0.1537	-0.1046	0.7204	$7.6 \cdot 10^{-5}$	6
1594	1.5403	-0.2247	-0.1379	0.7513	$9.3 \cdot 10^{-5}$	5
3590	2.5649	0.9319	0.6873	$-2 \cdot 10^{-3}$	$8.6 \cdot 10^{-5}$	9
13806	1.6373	-0.2082	-0.1274	0.75	$2.8 \cdot 10^{-4}$	4
# Elements	α	β	γ	δ	$f^*_{\text{BB-ana}}$	No.Iter
898	2.202	4.2886	1.5457	0.722	$9.6 \cdot 10^{-5}$	10
1594	1.4808	-0.2157	-0.1137	0.7983	$7.3 \cdot 10^{-5}$	5
3590	1.4675	-0.152	-0.0752	0.8073	$7.3 \cdot 10^{-5}$	5
13806	1.496	$6.3 \cdot 10^{-3}$	0.0182	0.8297	$7.7 \cdot 10^{-5}$	5

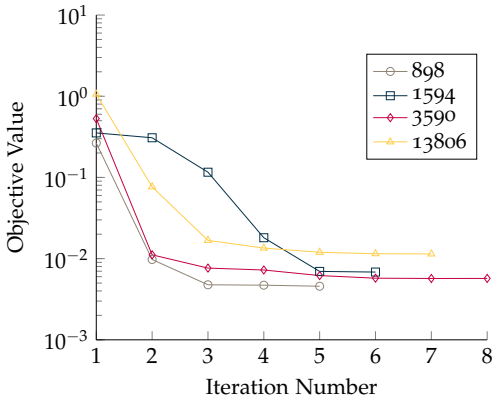
Table 6.3.: The table shows the final values of the load parameters and the respective objective function values for the SAND method (f^*_{SAND}), the BB-method with derivatives of the constraints obtained with FD and LBFGS ($f^*_{\text{BB-num}}$), and the BB-method with analytic constraint derivatives ($f^*_{\text{BB-ana}}$) for the four different discretizations together with the number of iterations IPOPT needed to find the solutions.



(a) BB-method with analytic derivatives.



(b) BB-method with numerical derivatives.



(c) SAND method.

Figure 6.9.: The values of the objective functions during the optimization for all three methods and all four meshes is shown.

Following the works of Taebi et al. (2012) and Rozgič et al. (2012) the identification formulation associated to the BB-method reads

$$\begin{aligned} \min_{\boldsymbol{p}\in\mathbb{R}^4} \quad & \frac{1}{\operatorname{meas} T}\int_T |s(\boldsymbol{p},x)-s_{\text{opt}}(x)|^2\,\mathrm{d}x, \\ \text{subject to} \quad & \varepsilon_1 - G(\varepsilon_2) \leq 0. \end{aligned} \tag{6.13}$$

In this formulation $s : \mathbb{R}^4 \times T \to \mathbb{R}^2$ is a parametrization of the sheet's surface depending on the values of the parameter vector $\boldsymbol{p}$. The function $s_{\text{opt}} : T \to \mathbb{R}^2$ maps the values from the surface to the prescribed optimal shape (cf. Fig. 6.1). The area of the sheet's surface is denoted by meas T. The constraints in Problem (6.13) are a model to an FLC, discussed earlier in Section 4.1.3, where ε_1 and ε_2 are the major and minor

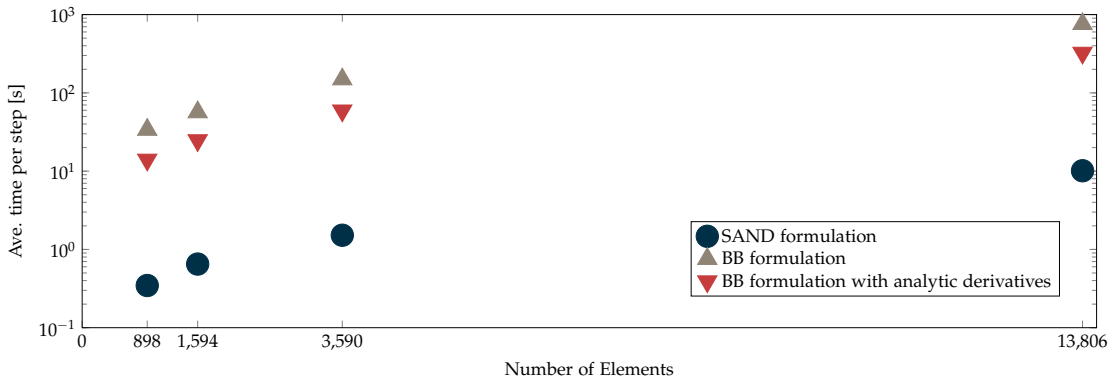


Figure 6.10.: Average time per iteration of IPOPT for the three different methods against the number of elements. The time is logarithmically scaled, for convenience.

strain. The function $G : \mathbb{R} \to \mathbb{R}$ was fitted to experimentally obtained data and reads

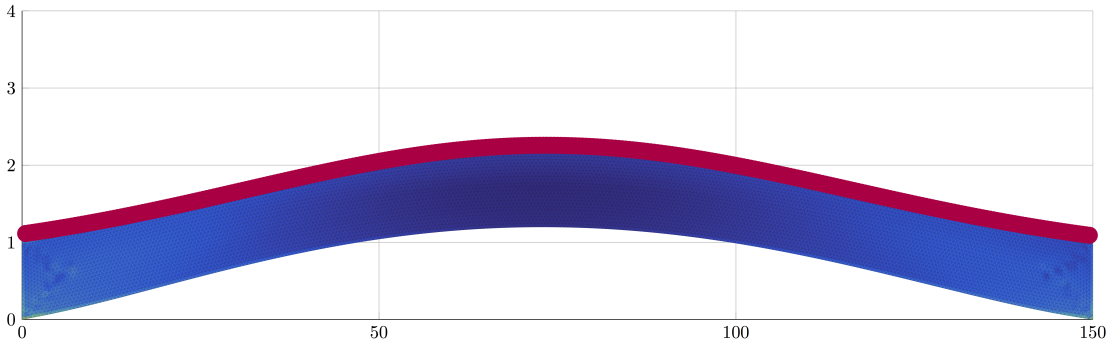
$$G(\varepsilon_2) = \begin{cases} 3.14416 \cdot t^3(\varepsilon_2) - 2.35404 \cdot t^2(\varepsilon_2) + 0.72462 \cdot t(\varepsilon_2) + 0.00213, & \text{if } t(\varepsilon_2) \geq 0, \\ 0.075283 \cdot t^2(\varepsilon_2) - 0.676572 \cdot t(\varepsilon_2) + 0.00223, & \text{if } t(\varepsilon_2) < 0, \end{cases}$$

with $t(\varepsilon_2) = \left(\frac{\varepsilon_2}{1.1}\right)^{0.126}$, further details are discussed in Taebi et al. (2012).

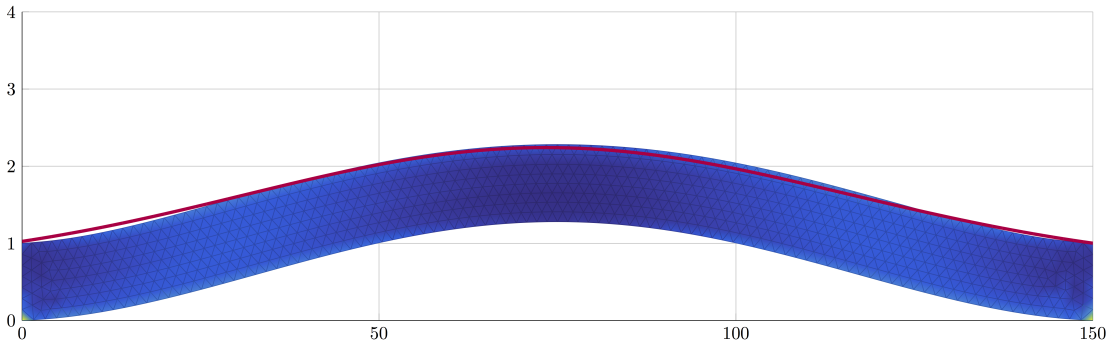
As can be seen, every evaluation of the objective function and the constraints requires a solution of the underlying PDE, describing the mechanical situation and given in Eq. (6.4). Similar to the problems discussed in Chapter 4 first and second order derivative information can be obtained via the FD method and the LBFGS update strategy. However, in this case, since $\varepsilon(u) = \frac{1}{2}\left(\nabla u + \nabla u^\top\right)$, the gradient and Hessian of the constraints can also be computed analytically by differentiating the force density. Therefore, both variants, with numerical derivatives and analytical derivatives, are used and matched against the SAND method. The SAND formulation remains unchanged and is the same as in (6.7), no further box or damage constraints were assumed.

Again, the FEM is used to obtain a numerical solution to Eq. (6.4). Four different discretizations were employed, with 898, 1 594, 3 590, and 13 806 triangular elements. The solutions obtained by IPOPT with all three variants (SAND, BB-method with numerical derivatives, and BB-method with analytical derivatives) are given in Table 6.3. Three examples for identified configurations are shown in Fig. 6.11.

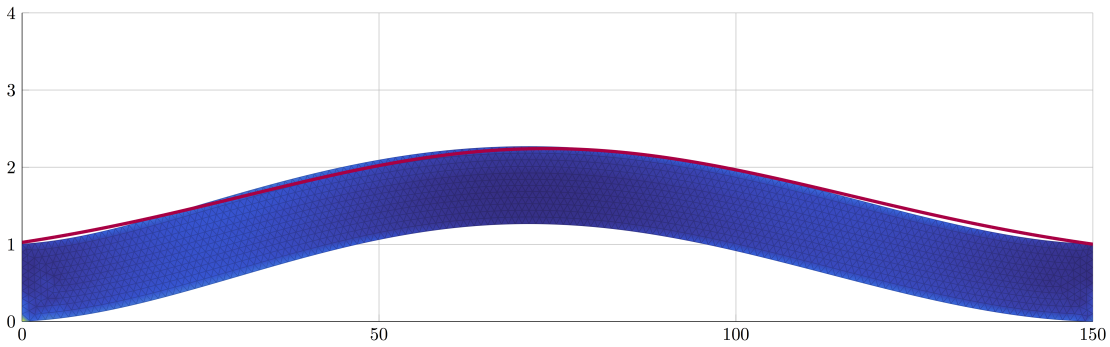
It can be seen that solutions were obtained after a relatively low number of iterations, between five and ten. Further, the objective values of the BB-method are always lower, this is due to the fact that there the objective is scaled by the area of T, whereas in in the



(a) Final configuration with SAND and 13806 elements



(b) Final configuration with analytic BB and 1594 elements



(c) Final configuration with numeric BB and 3590 elements

Figure 6.11.: The final configurations computed with IPOPT obtained by the SAND method Fig. 6.11(a), by the BB-method with analytic derivatives of the constraints, Fig. 6.11(b), and by the BB-method with numerical derivative approximations to both, the constraints and the objective, Fig. 6.11(c). Exemplarily three different discretizations are depicted.

SAND formulation it remains unscaled. All identified parameter vectors differ, since the force density function is not convex, and thus it becomes unlikely that a global optimum can be found. As can be seen in Fig. 6.9(a) and Table 6.3 the final objective values, as well as the values during the iteration remain constant regardless of the used mesh, for the BB-method with analytic derivatives. This stable behavior can not be preserved when numerical first and second order derivatives are employed, as depicted in Fig. 6.9(b). Contrary to the earlier observations, in Fig. 6.9(c) convergence can be seen. But, since this is not seen in Fig. 6.6, this might be a singular event.

In Fig. 6.10 the average time per step of IPOPT for the three methods can be seen. It is remarkable that, although all three methods needed roughly the same number of steps, the times per step differ by two orders of magnitude. The SAND method is 10^2 faster, on average. Employing analytical derivatives for the constraints only yields a slight speed up compared to the purely numerical derivative computation. The superiority of the SAND approach can be explained by the fact, that the multiplication of a vector with a matrix is mostly cheaper than solving a linear system.

The BB-method yields smaller KKT systems (cf. Eq. (3.15)) for the optimization, since the number of primal variables (here 4) is significantly smaller compared to the number of variables in the SAND method (here $N + 4$, with N being the number of free nodes). But solving the corresponding KKT systems seems to be cheaper than simulating the FEM system. This is due to the fact that reduced forms of the linear systems are exploited during the optimization loop, see Wächter and Biegler (2006) for details. However, this example demonstrates the potential gain of efficiency the application of the SAND methodology holds, when used for process parameter optimization problems.

6.4. Conclusion

A framework to utilize the classical SAND approach to process parameter optimization in sheet metal applications has been introduced. The novel approach was analyzed by a 2D test example, where the underlying PDE is governed by linear elastic material behavior. It was shown how the method is capable of finding optimal control parameter sets for various mesh sizes and for different prescribed forms. Within the linear elastic regime a model comparison to the previously discussed BB-method has been made. Following earlier works by Taebi et al. (2012)and Rozgič et al. (2012) a simplified BB-problem with physical constraints, mimicking damage constraints, has been compared against the SAND formulation. The latter turned out to be 100 times faster, on average,

at every step of the iteration. This promising gain in efficiency alone makes the further pursue of the SAND approach worthwhile.

Still, the BB-method is proved to be reliable and robust, especially against changes in the physics of the process. Regardless of the underlying PDEs, as long as they can be solved by some (commercial) solver, the BB-method can be applied. The SAND method, on the other hand, requires access to the discretizations of the PDEs. Most commercial tools do not have such an interface. However, with growing computational power the demand of solving PDE constrained optimization problems grows fast. Approaches to automatically solve such problems were introduced lately by Funke (2013). An exhaustive survey about the advantages of incorporating the SAND method directly into PDE solvers has been presented by van Bloemen Waanders et al. (2002). Their research motivated the development of the Sundance software by Long et al. (2012), which can be incorporated into Trilinos suite Heroux et al. (2005) to solve PDE constrained optimization problems efficiently. Eldred et al. (2002) point out that the combination of the SAND approach with methods of surrogate-based optimization under uncertainty can potentially be used to solve problems governed by stochastic PDEs.

7 Conclusion and Outlook

In this work two methods for virtual process design in the context of sheet metal forming applications have been introduced. The first half of this thesis was concerned with solving the two main parameter identification problems arising when simulating and designing a coupled, high-speed electromagnetic forming process: The identification of the employed material model and the identification of optimal process parameters. For both problems a similar black box framework based on LS-DYNA FEM/BEM simulations and IPOPT, an implementation of the inner point method by Wächter and Biegler (2006), has been used.

A thermodynamically consistent material model developed by Vladimirov et al. (2014) which is tailored for the efficient numerical treatment of anisotropic hyperelastic-plastic materials has been considered in this work. This model is capable of representing rate dependent viscoplastic as well as rate independent elastoplastic materials by switching between a Kuhn-Tucker and a Perzyna formulation for the yield function. Vladimirov et al. (2014) coupled the material model to a scalar Lemaitre-type damage model. After the identification of the model's parameters it proved to be able to simulate the behavior of the aluminum alloy EN AW-5083 at various strain rates. A framework for the reliable identification of these model parameters was introduced, based on experimental data obtained by uniaxial tensile tests and a corresponding LS-DYNA simulation coupled to IPOPT. For this particular aluminum alloy the flow stress rather than the hardening increases at high strain rates, as shown by, e.g., Winzer and Glinicka (2011). Thus, it was possible to derive a set of parameters suitable for the modeling of the high-speed situation from the data obtained by quasi-static experiments. This could be achieved with a linear extrapolation of the damage threshold as introduced by Kiliclar et al. (2015b).

To deal with the problems coming from the comparison of data obtained by simulation and data coming from experiments a distance measure based on the L^2-distance of functions was used. To this end an Akima spline developed by Akima (1970) was adapted to the experimental data. These special splines have the advantage of producing smooth interpolates, just as the according curves are assumed to be smooth, as long as no damage occurs. It was hence possible to interpolate the experimental data at every point needed and thus compare it efficiently to the data from simulation. It was shown that the identified parameter set is suitable to represent a coupled forming process very precisely, see Kiliclar et al. (2016). The presented framework is not restricted to either the employed material model, nor to the identification via uniaxial tensile tests. It can be used in a broad range of applications, since it only needs an adequate simulation as a provider for values of the objective, the constraints and their respective derivatives. Other software toolboxes like Abaqus or COMSOL can be used by performing minor changes in the communication scripts with IPOPT.

A main problem in the industrial use of electromagnetic forming is the life time of the used coils, as discussed by Psyk et al. (2011). The high strain rates and the resulting fast process times yield that only the first half wave, resulting form a sinusoidal pulse, is involved in the forming. The remaining field causes damage in the coils, as they are pushed apart by the resulting induced Lorentz forces. Mono-directional pulses offer a solution to that particular problem and can make the technology more appealing to commercial use. The previously developed method for material parameter identification was therefore adapted in Kiliclar et al. (2015a) to perform a process parameter identification in a coupled sheet metal forming process. The task was to find a set of parameters yielding a sharp cup radius. Of particular interest was the adjustment of a double exponential pulse as a prototype for mono-directional pulses.

The material's behavior was simulated by the previously identified material model, thus constraints for the damage could be employed efficiently. In contrast to the usual optimal shape problems, an objective function completely independent of the geometry parametrization has been derived. It turned out that a maximization of the major strain in the area of question can be traced back to a lower cup radius, thus yielding a sharper form. The subsequent application of electromagnetic forming could reduce the drawing radius by roughly 5 mm, compared to quasi-static deep drawing alone. The presented method can be seen as an extension to previous approaches by Taebi et al. (2012), since it uses a unified simulation framework and exploits a material model that incorporates damage. Again, the presented algorithm is not restricted to parameters defining the current pulse. Parameters defining the blankholder forces, or the punch forces, or

simulating the lubrication conditions can also be treated by the proposed method.

It should be stressed, that the obtained results are purely numerical. An experimental validation is still subject of active research. However, the tools developed in this work can be exploited for further design of industrial relevant forming applications, given an efficient simulation, since the time needed to find a suitable solution to the design problem scales with the time needed to simulate the process at hand.

The second part of this work was concerned with finding a solution to this performance problem. Inspired by algorithmic methods to tackle PDE constrained optimization problems a fitting SAND formulation for the identification of process parameters was introduced. The solution of the PDE, modeling the physical behavior, is seen as a state, and the parameters of the process are seen as controls. The solution of such optimization problems is an active field of research, and mainly two approaches are distinguished. One is the first optimize, then discretize approach, which is based on the derivation of optimality conditions in the infinite dimensional spaces used. The optimality conditions have to be derived carefully, and the subsequent discretization of the problem needs to hold controls and states in check, usually by employing regularization techniques, cf. Hinze et al. (2010). The second one is the first discretize, then optimize approach, which is chosen in this work. The use of regularization methods for the controls could be avoided in this particular work, since the process parameters are already represented as a real valued finite dimensional vector.

A class of problems mimicking a sheet metal forming problem in the linear elastic regime was introduced. This simplification was picked, since, the underlying FEM can easily be implemented but the insights gained are useful for the development of a general optimization scheme for those types of technologically relevant problems. In the course of this work an efficient FEM solver for MATLAB, *o*FEM, has been developed by Dudzinski et al. (2016) to provide the necessary discretizations of the PDEs, which is needed for the optimization procedure. Again the IP implementation IPOPT was used for the solution of the resulting optimization problems. Here, the ability of IP methods to treat a huge number of constraints proved to be extraordinarily useful for the numerical solution of SAND problems. With *o*FEM and IPOPT the family of problems could be thoroughly numerically analyzed.

The derived SAND formulation was compared to a black box optimization problem based on earlier works by Taebi et al. (2012) and Rozgič et al. (2012). Compared to this NAND approach the SAND approach showed a potential gain in efficiency by two orders of magnitude. It was shown, that when solving a SAND problem, every iteration step is on average nearly 100 times faster. This will, of course, be less when non-linear

material laws are employed. The trend however will be retained. A major drawback of the SAND method is that the system matrices of the discretized PDEs need to be employed in the optimization procedure. For the time being, this makes the method unattractive for the design of industrial processes, since the simulations are usually based on the commercial tools. However, this work clearly shows the benefit that can be achieved, when incorporating the SAND optimization directly into commercial PDE solvers. Thus, no open interface needs to be created and the user can still benefit from an efficient optimization strategy.

Among others, two questions for future research arise immediately:

1. *In view of the computational costs, how efficient can the simulation based process parameter identification method be used to find more parameters in the coupled forming process*?

   The answer to this question will become vital for the commercial use of the proposed framework. It is clear that the growing availability of parallel computers and the growth of available memory are beneficial for the introduced scheme. In this work it was already shown how parallel computations can be exploited to determine necessary derivative information efficiently. Also, the underlying simulations are optimized in terms of speed. This can be done by exploiting symmetry, as is done here in the numerical treatment of cup forming. Instead of using one quarter one could resort to even smaller fractions. Then, the coupling at the center of the bottom has to be done carefully. However, as argued by Arora and Wang (2005) the use of the FD method to determine approximations to the derivative can become a problem during the optimization. The step sizes have to be adjusted carefully and for every parameter the sensitivity has to be determined, which creates further problems. Clearly, the work on this type of solution methods will go on, as computational power grows and PDE solvers become faster.

2. *How good can the SAND approach deal with time dependent problems with non-linear material descriptions?*
   Of course, one part of the answer to this problem again depends on the growth of computational power. If a time dependent problem, where every time step matters, is considered, then, in the SAND approach, the number of constraints multiplies with the number of time steps. This huge number of constraints has to be stored, making it a memory problem, but it also has to be efficiently treated by the employed optimization strategy. Modern IP methods can, in theory, deal with

many constraints, making them a promising candidate as the optimization solver. Since the resulting KKT system needs to be solved as well, suitable linear algebra solvers are needed. The development of those solvers is ongoing, and one can be pretty confident that large scale linear systems will be of particular interest.

The incorporation on non-linear material laws, in particular viscoplasticity laws is of huge interest for the use of the introduced SAND approach in technologically relevant applications. In general the SAND method is capable of treating those problems, since it allows any kind of discretized PDEs as constraint. Thus, the resulting non-linear equation can simply be used. Then, the introduced formulation of the SAND would not change and only the discretization of the viscoplaticity equations has to be considered. Another possibility to tackle such a problem is by incorporating the thermodynamical equations defining the stress tensors as constraints and treating the stress tensor as another variable in the optimization problem. Thus, the state variable would comprise of the displacements and the stresses, while the controls are still the process parameters. Following the notation of Han and Reddy (2012) a possible version of such a problem could look like this:

Let u denote the displacements, and let the space of displacements be $V = \left[H_0^1(\Omega)\right]^d$, further let the principle stress be $\Sigma = (\sigma, \chi)$, with conjugate forces χ. Let the stress space be $S = \left\{\tau = \left(\tilde{\tau}_{ij}\right) : \ \tilde{\tau}_{ij} = \tilde{\tau}_{ji}, \ \tilde{\tau}_{ij} \in H(\mathrm{div}, \Omega)\right\}$, and the space of conjugate forces be $M = \left\{\mu = \left(\tilde{\mu}_j\right) : \ \tilde{\mu}_j \in L^2(\Omega)\right\}$, further let the combined space be $\mathcal{M} = S \times M$. Then appropriate bilinear forms are given as:

$$
\begin{aligned}
a : S \times S \to \mathbb{R}, \quad a(\sigma, \tau) &= \int_\Omega \sigma : C^{-1}\tau \, \mathrm{d}x, \\
b : V \times S \to \mathbb{R}, \quad b(v, \tau) &= -\int_\Omega \epsilon(v) : \tau \, \mathrm{d}x, \\
c : M \times M \to \mathbb{R}, \quad c(\chi, \mu) &= \int_\Omega \chi : H^{-1}\mu \, \mathrm{d}x, \\
A : \mathcal{M} \times \mathcal{M} \to \mathbb{R}, \quad A(\Sigma, T) &= a(\sigma, \tau) + c(\chi, \mu).
\end{aligned}
$$

With these bilinear forms the weak formulation reads:
Given $f(t, p) \in L^2(\Omega)$ and appropriate initial conditions, find $(u, \Sigma) \in V \times \mathcal{M}$ such that for almost all $t \in [0, T_{\max}]$:

$$
\begin{aligned}
b(v, \sigma(t)) &= -\int_\Omega f(t, p) \cdot v \, \mathrm{d}x, \quad \forall v \in V, \\
A\left(\dot{\Sigma}(t), T - \Sigma(t)\right) + b\left(\dot{u}(t), \tau - \sigma(t)\right) &\geq 0, \quad \forall T \in \mathcal{M}, \\
\phi(\Sigma) &\leq 0,
\end{aligned}
$$

where ϕ denotes the yield function.

The resulting optimization problem is consequently:

$$\begin{aligned}
&\min_{\substack{u\in V(\Omega)\\ \Sigma\in\mathcal{M}\\ p\in\mathbb{R}^n}} \frac{1}{2}\left|\left|u-\bar{u}\right|\right|^2_{L^2(\mathcal{T})},\\
&\quad \text{s. t. } b\left(v,\sigma\left(t\right)\right) = -\int_\Omega f(t,p)\cdot v\,\mathrm{d}x, \quad \forall v\in V,\\
&\qquad A\left(\dot{\Sigma}\left(t\right),T-\Sigma\left(t\right)\right)+b\left(\dot{u}\left(t\right),\tau-\sigma\left(t\right)\right)\geq 0, \quad \forall T\in\mathcal{M},\\
&\qquad \phi\left(\Sigma\right)\leq 0.
\end{aligned}$$

As can be seen both questions are closely tied to the development of new hardware and software. In general the following should be considered when working on the development of new algorithms to challenge hard technological problems: *Today's supercomputers are tomorrow's desktops, are the mobile devices of the day after tomorrow.* With this in mind algorithm development should not only aim for the present but also for the future.

Bibliography

Akima, H., 1970. A new method of interpolation and smooth curve fitting based on local procedures. Journal of the ACM (JACM) 17, 589–602.

Alnæs, M.S., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M.E., Wells, G.N., 2015. The FEniCS project version 1.5. Archive of Numerical Software 3.

Alsos, H.S., Hopperstad, O.S., Törnqvist, R., Amdahl, J., 2008. Analytical and numerical analysis of sheet metal instability using a stress based criterion. International Journal of Solids and Structures 45, 2042–2055.

Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L., 2007. On augmented lagrangian methods with general lower-level constraints. SIAM Journal on Optimization 18, 1286–1309.

Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L., 2008. Augmented lagrangian methods under the constant positive linear dependence constraint qualification. Mathematical Programming 111, 5–32.

Argyris, J.H., Kelsey, S., 1960. Energy theorems and structural analysis. volume 960. Springer.

Arora, J., Wang, Q., 2005. Review of formulations for structural and mechanical system optimization. Structural and Multidisciplinary Optimization 30, 251–272.

Atkinson, A.C., Donev, A.N., Tobias, R.D., 2007. Optimum experimental designs, with SAS. volume 34. Oxford University Press Oxford.

Banabic, D., 2010. Sheet metal forming processes: constitutive modelling and numerical simulation. Springer Science & Business Media.

Banabic, D., Lazarescu, L., Paraianu, L., Ciobanu, I., Nicodim, I., Comsa, D., 2013. Development of a new procedure for the experimental determination of the forming limit curves. {CIRP} Annals - Manufacturing Technology 62, 255 – 258.

Bangerth, W., Hartmann, R., Kanschat, G., 2007. deal.II – a general purpose object oriented finite element library. ACM Trans. Math. Softw. 33, 24/1–24/27.

Bates, D.M., Watts, D.G., 1988. Nonlinear regression: iterative estimation and linear approximations. Wiley Online Library.

Bauer, D., 1967. Ein neuartiges Messverfahren zur Bestimmung der Kräfte, Arbeiten, Formänderungen, Formänderungsgeschwindigkeiten und Formänderungsfestigkeiten beim Aufweiten zylindrischer Werkstüke durch schnellveränderliche magnetische Felder. Technische Hochschule, Hannover.

Bazaraa, M.S., Shetty, C., 1979. Nonlinear Programming. John Wiley & Sons.

Bendsøe, M.P., Sigmund, O., 2013. Topology optimization: theory, methods, and applications. Springer Science & Business Media.

Benner, P., Gugercin, S., Willcox, K., 2013. A survey of model reduction methods for parametric systems .

Bhattacharya, D., 1997. Composite Sheet Forming. Composite Materials Series, Elsevier Science.

Biegler, L.T., Ghattas, O., Heinkenschloss, M., van Bloemen Waanders, B., 2003. Large-scale pde-constrained optimization: an introduction, in: Large-Scale PDE-Constrained Optimization. Springer, pp. 3–13.

Biegler, L.T., Wächter, A., 2003. Sqp sand strategies that link to existing modeling systems. Large-Scale PDE-Constrained Optimization , 199–217.

van Bloemen Waanders, B., Bartlett, R., Long, K., Boggs, P., Salinger, A., 2002. Large scale non-linear programming for pde constrained optimization. SAND Report, SAND 3198, 2002.

Bodenseher, H., Holtschmit, H., Siewerth, G., 1973. Einrichtung zum elektromagnetischen Umformen von Werkstuecken. DE Patent App. DE19,722,205,981.

Braess, D., 2013. Finite Elemente: Theorie, schnelle Löser und Anwendungen in der Elastizitätstheorie. Springer-Verlag.

Butuc, M., Gracio, J., Da Rocha, A.B., 2006. An experimental and theoretical analysis on the application of stress-based forming limit criterion. International Journal of Mechanical Sciences 48, 414–429.

Byrd, R.H., Gilbert, J.C., Nocedal, J., 2000. A trust region method based on interior point techniques for nonlinear programming. Mathematical Programming 89, 149–185.

Byrd, R.H., Hribar, M.E., Nocedal, J., 1999. An interior point algorithm for large-scale nonlinear programming. SIAM Journal on Optimization 9, 877–900.

Byrd, R.H., Nocedal, J., Schnabel, R.B., 1994. Representations of quasi-newton matrices and their use in limited memory methods. Mathematical Programming 63, 129–156.

Byrd, R.H., Nocedal, J., Waltz, R.A., 2006. Knitro: An integrated package for nonlinear optimization, in: Large-scale nonlinear optimization. Springer, pp. 35–59.

Caldichoury, I., L'Eplattenier, P., 2008. Introduction of an electromagnetism module in ls-dyna for coupled mechanical thermal electromagnetic simulations. ICHSF 2008 , 85–96.

Camp, M., Garbe, H., 2004. Parameter estimation of double exponential pulses (emp, uwb) with least squares and nelder mead algorithm. Electromagnetic Compatibility, IEEE Transactions on 46, 675–678.

Canelas, A., Herskovits, J., Telles, J., 2008. Shape optimization using the boundary element method and a sand interior point algorithm for constrained optimization. Computers & Structures 86, 1517–1526.

Choi, Y.S., 2012. Simultaneous analysis and design in PDE-constrained optimization. Ph.D. thesis. Stanford University.

Ciarlet, P.G., 1988. Three-dimensional elasticity. volume 1. Elsevier.

Clausen, A.H., Børvik, T., Hopperstad, O.S., Benallal, A., 2004. Flow and fracture characteristics of aluminium alloy AA5083–H116 as function of strain rate, temperature and triaxiality. Materials Science and Engineering: A 364, 260–272.

Clough, R., 1958. Structural analysis by means of a matrix algebra program, in: Proceedings of the First Conference on Electronic Computation, American Society of Civil Engineers, Kansas City, MO, pp. 109–132.

Colemann, B.D., Gurtin, M., 1967. Thermodynamics with internal variables. Journal of Chemical Physics 47 , 597–613.

Conn, A.R., Gould, G., Toint, P.L., 2010. LANCELOT: a Fortran package for large-scale nonlinear optimization (Release A). Springer Publishing Company, Incorporated.

Courant, R., 1943. Variational methods for the solution of problems of equilibrium and vibrations. Bull. Amer. Math. Soc. 49, 1–23.

Dobrowolski, M., 2010. Angewandte Funktionalanalysis: Funktionalanalysis, Sobolev-Räume und elliptische Differentialgleichungen. Springer-Verlag.

Dudzinski, M., Rozgič, M., Stiemer, M., 2016. *o*FEM: An object oriented finite element package for MATLAB. Submitted to Applied Mathematics and Computation.

Eldred, M.S., Giunta, A.A., Wojtkiewicz, S., Trucano, T.G., 2002. Formulations for surrogate-based optimization under uncertainty, in: 9th AIAA/ISSMO symposium on multidisciplinary analysis and optimization, Atlanta, GA.

Engl, H.W., Hanke, M., Neubauer, A., 1996. Regularization of inverse problems. volume 375. Springer.

Ern, A., Guermond, J.L., 2013. Theory and practice of finite elements. volume 159. Springer Science & Business Media.

Fletcher, R., 2000. Practical Methods of Optimization. Wiley.

Fletcher, R., Leyffer, S., 1998. User manual for filtersqp. University of Dundee Numerical Analysis Report NA-181 .

Frederick, C.O., Armstrong, P.J., 2007. A mathematical representation of the multiaxial bauschinger effect. Materials at High Temperatures 24, 1–26.

Funke, S.W., 2013. The automation of PDE-constrained optimisation and its applications. Ph.D. thesis. Department of Earth Science and Engineering, Imperial College London.

Gelfand, I.M., Vilenkin, N.Â., Feinstein, A., 1964. Generalized functions. vol. 4. .

Gies, S., Löbbe, C., Weddeling, C., Tekkaya, A., 2014. Thermal loads of working coils in electromagnetic sheet metal forming. Journal of Materials Processing Technology 214, 2553 – 2565.

Gill, P.E., Murray, W., Saunders, M.A., 2005. Snopt: An sqp algorithm for large-scale constrained optimization. SIAM review 47, 99–131.

Gill, P.E., Murray, W., Saunders, M.A., Wright, M.H., 1984. User's guide for NPSOL (Version 2. 1): a Fortran package for nonlinear programming. Technical Report. Stanford Univ., CA (USA). Systems Optimization Lab.

Goodwin, G.M., 1968. Application of strain analysis to sheet metal forming problems in the press shop. Technical Report. SAE Technical Paper.

Haftka, R., Kamat, M., 1989. Simultaneous nonlinear structural analysis and design. Computational Mechanics 4, 409–416.

Haftka, R.T., 1984. Simultaneous analysis and design. Virginia Polytechnic Institute and State University.

Hallquist, J.O., et al., 2006. Ls-dyna theory manual. Livermore software Technology corporation 3.

Han, W., Reddy, B.D., 2012. Plasticity: mathematical theory and numerical analysis. volume 9. Springer Science & Business Media.

Hasan, R.Z., Kinsey, B.L., Tsukrov, I., 2011. Effect of element types on failure prediction using a stress-based forming limit curve. Journal of Manufacturing Science and Engineering 133, 061002.

Hazra, S.B., 2009. Large-scale PDE-constrained Optimization in Applications. volume 49. Springer Science & Business Media.

Heroux, M.A., Bartlett, R.A., Howle, V.E., Hoekstra, R.J., Hu, J.J., Kolda, T.G., Lehoucq, R.B., Long, K.R., Pawlowski, R.P., Phipps, E.T., et al., 2005. An overview of the trilinos project. ACM Transactions on Mathematical Software (TOMS) 31, 397–423.

Hill, R., 1948. A theory of the yielding and plastic flow of anisotropic metals. Proceedings of the Royal Society of London A193, 281–297.

Hinze, M., Pinnau, R., Ulbrich, M., Ulbrich, S., 2010. Optimization with PDE Constraints. Springer.

Hinze, M., Tröltzsch, F., 2010. Discrete concepts versus error analysis in pde-constrained optimization. GAMM-Mitteilungen 33, 148–162.

Hrennikoff, A., 1941. Solution of problems of elasticity by the framework method. Journal of applied mechanics 8, 169–175.

Huh, H., Lim, J., Park, S., 2009. High speed tensile test of steel sheets for the stress-strain curve at the intermediate strain rate. International Journal of Automotive Technology 10, 195–204.

Jablonski, F., 1976. Analyse der elektromagnetischen Umformung rohrförmiger Werkstücke. Dr.-Ing. Ph.D. thesis. Dissertation, Akademie der Wissenschaften der DDR, Berlin.

Jackson, J.D., 1999. Classical electrodynamics. 3 ed., Wiley New York etc.

Jones, D.R., Schonlau, M., Welch, W.J., 1998. Efficient global optimization of expensive black-box functions. Journal of Global optimization 13, 455–492.

Karadogan, C., Tamer, M., 2015. Development of a new and simplified procedure for the experimental determination of forming limit curves. {CIRP} Annals - Manufacturing Technology 64, 265 – 268.

Karush, W., 1939. Minima of functions of several variables with inequalities as side constraints. Ph.D. thesis. Master's thesis, Dept. of Mathematics, Univ. of Chicago.

Keeler, S.P., Backofen, W.A., 1963. Plastic instability and fracture in sheets stretched over rigid punches. Asm Trans Q 56, 25–48.

Kiliclar, Y., Demir, O., Engelhardt, M., Rozgič, M., Vladimirov, I., Wulfinghoff, S., Weddeling, C., Gies, S., Klose, C., Reese, S., et al., 2016. Experimental and numerical investigation of increased formability in combined quasi-static and high-speed forming processes. Journal of Materials Processing Technology .

Kiliclar, Y., Demir, O., Vladimirov, I., Kwiatkowski, L., Brosius, A., Reese, S., Tekkaya, A., 2012a. Combined simulation of quasi-static deep drawing and electromagnetic forming by means of a coupled damage–viscoplasticity model at finite strains.

Kiliclar, Y., Engelhardt, M., Vladimirov, I.N., Pietryga, M.P., von Senden genannt Haverkamp, H., Reese, S., Bach, F.W., 2012b. On the improvement of formability and the prediction of forming limit diagrams at fracture by means of constitutive modelling, in: Key Engineering Materials, Trans Tech Publ. pp. 29–34.

Kiliclar, Y., Rozgič, M., Reese, S., Stiemer, M., 2015a. Numerical optimization of current parameters in combined electromagnetic-classical forming processes. Key Engineering Materials .

Kiliclar, Y., Vladimirov, I., Wulfinghoff, S., Reese, S., Demir, O., Weddeling, C., Tekkaya, A., Engelhardt, M., Klose, C., Maier, H., Rozgič, M., Stiemer, M., 2015b. Finite element analysis of combined forming processes by means of rate dependent ductile damage modelling. International Journal of Material Forming , 1–12.

Kleinermann, J.P., Ponthot, J.P., 2003. Parameter identification and shape/process optimization in metal forming simulation. Journal of Materials Processing Technology 139, 521–526.

Kočvara, M., Stingl, M., 2003. Pennon: A code for convex nonlinear and semidefinite programming. Optimization methods and software 18, 317–333.

Kreißig, R., Benedix, U., Görke, U.J., Lindner, M., 2007. Identification and estimation of constitutive parameters for material laws in elastoplasticity. GAMM-Mitteilungen 30, 458–480.

Kuhn, H.W., Tucker, A.W., 1951. Nonlinear programming, in: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950, University of California Press, Berkeley and Los Angeles. pp. 481–492.

Larson, M.G., Bengzon, F., 2013. The finite element method: Theory, implementation, and applications. volume 10. Springer Science & Business Media.

Lemaitre, J., 1985. A continuous damage mechanics model for ductile fracture. Journal of Engineering Materials and Technology 107, 83.

Lemaitre, J., Chaboche, J.L., 1995. Mechanics of Solid Materials. Cambridge University Press.

Long, K., Boggs, P.T., van Bloemen Waanders, B.G., 2012. Sundance: high-level software for pde-constrained optimization. Scientific Programming 20, 293–310.

Mahnken, R., Johansson, M., Runesson, K., 1998. Parameter estimation for a viscoplastic damage model using a gradient-based optimization algorithm. Engineering Computations 15, 925–955.

Monk, P., 2003. Finite element methods for Maxwell's equations. Oxford University Press.

Murtagh, B.A., Saunders, M.A., 1983. MINOS 5. 0 user's guide. Technical Report. Stanford Univ., CA (USA). Systems Optimization Lab.

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. The Computer Journal 7, 308–313.

Nocedal, J., Wright, S., 2006. Numerical optimization, series in operations research and financial engineering. Springer, New York.

Psyk, V., Risch, D., Kinsey, B., Tekkaya, A., Kleiner, M., 2011. Electromagnetic forming—a review. Journal of Materials Processing Technology 211, 787–829.

Rieder, A., 2003. Keine Probleme mit Inversen Problemen. volume 1. Vieweg Wiesbaden.

Risch, D., 2009. Energietransfer und Analyse der Einflussparameter bei der formgebundenen elektromagnetischen Blechumformung. Dr.-Ing. Ph.D. thesis. Dissertation, TU Dortmund.

Rozgič, M., Stiemer, M., 2014a. Numerical comparison of optimization methods for process parameters identification in forming technology. PAMM 14, 797–798.

Rozgič, M., Taebi, F., Stiemer, M., 2012. Numerical optimisation of process parameters for combined quasi-static and impulse forming. PAMM 12, 625–626.

Rozgič, M.and Jaraczewski, M., Stiemer, M., 2014b. Inner point methods: On necessary optimality conditions of various reformulations of a constrained optimization problem .

Schaible, S., 1981. Generalized concavity in optimization and economics. Academic Press, New York.

Schnur, D.S., Zabaras, N., 1992. An inverse method for determining elastic material properties and a material interface. International Journal for Numerical Methods in Engineering 33, 2039–2057.

Schwarz, H.R., Köckler, N., 2009. Numerische Mathematik. Springer DE.

Schwarz, S., Maute, K., Ramm, E., 2001. Topology and shape optimization for elastoplastic structural response. Computer Methods in Applied Mechanics and Engineering 190, 2135–2155.

Sheppard, D., Terrell, R., Henkelman, G., 2008. Optimization methods for finding minimum energy paths. The Journal of Chemical Physics 128, –.

Stiemer, M., Unger, J., Svendsen, B., Blum, H., 2009. An arbitrary lagrangian eulerian approach to the three-dimensional simulation of electromagnetic forming. Computer Methods in Applied Mechanics and Engineering 198, 1535–1547.

Stoughton, T.B., 2000. A general forming limit criterion for sheet metal forming. International Journal of Mechanical Sciences 42, 1–27.

Stoughton, T.B., Yoon, J.W., 2012. Path independent forming limits in strain and stress spaces. International Journal of Solids and Structures 49, 3616–3625.

Szekeres, A., 1999. The wonderful Hooke's law: history, theory, and story, in: Proceedings-Estonian Academy of Sciences Physics Mathematics, Tallin Technical University and the the Estonain Academy of Sciences Physics Mathematics. pp. 296–305.

Taebi, F., Demir, O., Stiemer, M., Psyk, V., Kwiatkowski, L., Brosius, A., Blum, H., Tekkaya, A., 2012. Dynamic forming limits and numerical optimization of combined quasi-static and impulse metal forming. Computational Materials Science 54, 293 – 302.

Tarantola, A., 2005. Inverse problem theory and methods for model parameter estimation. siam.

Tikhonov, A.N., Arsenin, V.Y., 1977. Solutions of ill-posed problems. V. H. Winston & Sons.

Van Laarhoven, P.J., Aarts, E.H., 1987. Simulated annealing. Springer.

Vanderbei, R.J., Shanno, D.F., 1999. An interior-point algorithm for nonconvex nonlinear programming. Computational Optimization and Applications 13, 231–252.

Vladimirov, I.N., Pietryga, M.P., Kiliclar, Y., Tini, V., Reese, S., 2014. Failure modelling in metal forming by means of an anisotropic hyperelastic-plasticity model with damage. International Journal of Damage Mechanics 23, 1096–1132.

Vladimirov, I.N., Pietryga, M.P., Reese, S., 2010. Anisotropic finite elastoplasticity with nonlinear kinematic and isotropic hardening and application to sheet metal forming. International Journal of Plasticity 26 , 659–687.

Vogel, C.R., 2002. Computational methods for inverse problems. volume 23. Siam.

Vollertsen, F., Hu, Z., 2010. Analysis of punch velocity dependent process window in micro deep drawing. Production Engineering 4, 553–559.

Wächter, A., Biegler, L.T., 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical Programming 106, 25–57.

Wang, W., Chen, G., Lin, Z., Li, S., 2007. Determination of optimal blank holder force trajectories for segmented binders of step rectangle box using pid closed-loop fem simulation. The International Journal of Advanced Manufacturing Technology 32, 1074–1082.

Wang, Z., Fang, S.C., Xing, W., 2013. On constraint qualifications: motivation, design and inter-relations. Journal of Industrial and Management Optimization 9, 983–1001.

Winzer, R., Glinicka, A., 2011. The static and dynamic compressive behaviour of selected aluminium alloys. Engineering transactions 59, 85–100.

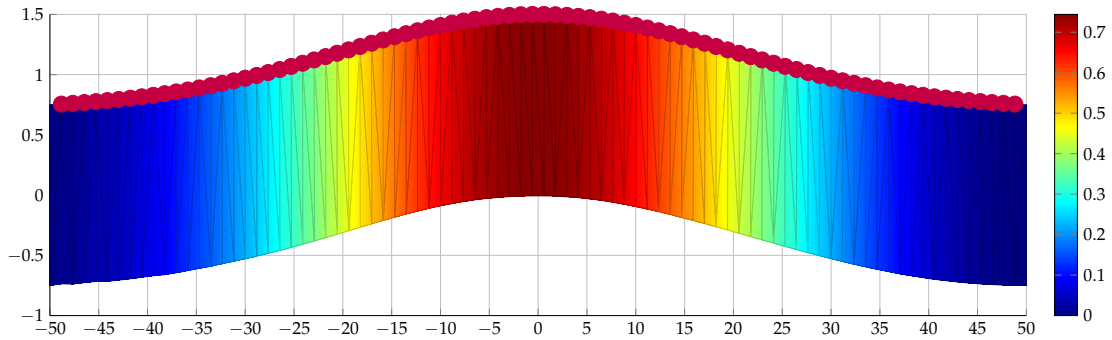
Wu, P., Graf, A., MacEwen, S., Lloyd, D., Jain, M., Neale, K., 2005. On forming limit stress diagram analysis. International journal of solids and structures 42, 2225–2241.

Zienkiewicz, O., Taylor, R., Zhu, J. (Eds.), 2013. The Finite Element Method: Its Basis and Fundamentals. Seventh edition ed., Butterworth-Heinemann, Oxford.

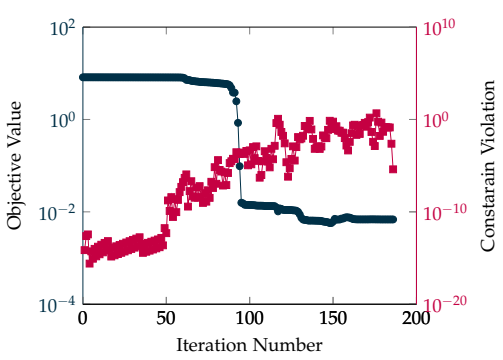
A

Numerical Results of the SAND Analysis

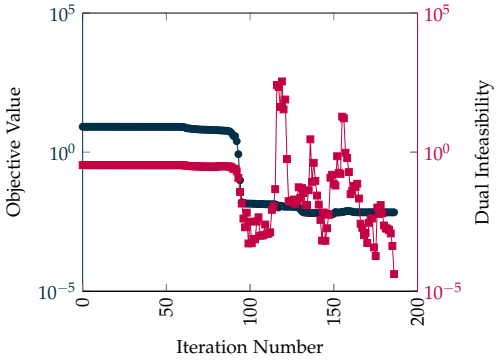
In the following all computed configurations of the 2D SAND problem introduced in Section 6.2 are depicted.



(a) Final configuration

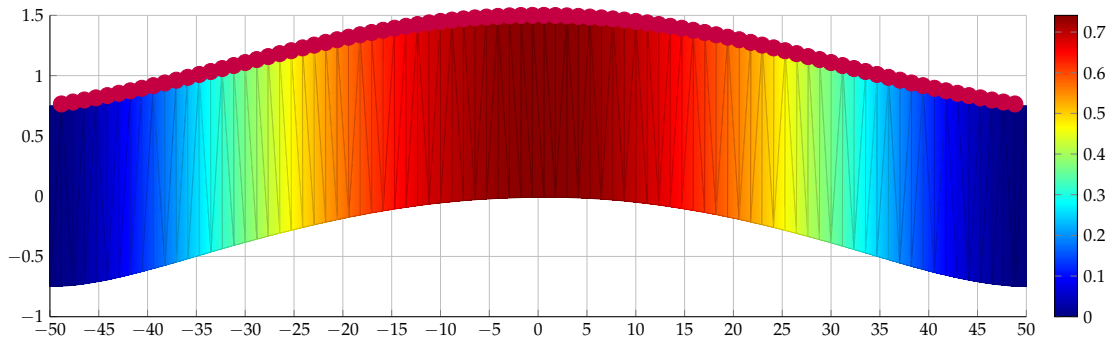


(b) Objective value and constrain violation.

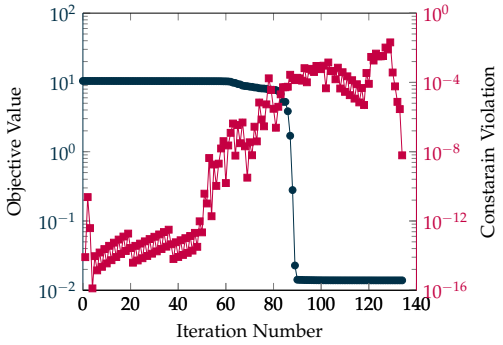


(c) Objective value and dual infeasibility.

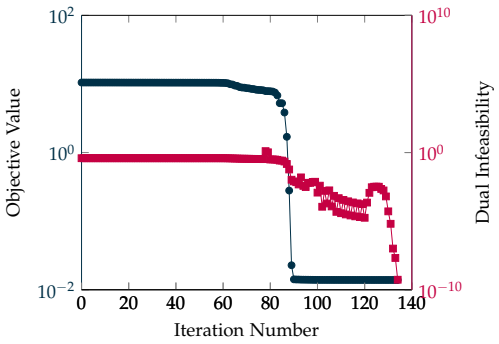
Figure A.1.: The final configuration computed with IPOPT for Level 1 and Curve 1 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 1, Curve 1 are shown. For convenience all values are given logarithmically.



(a) Final configuration

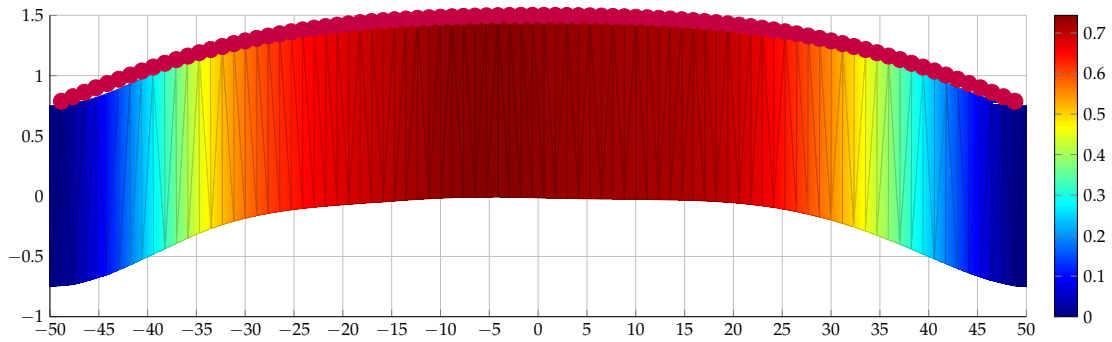


(b) Objective value and constrain violation.

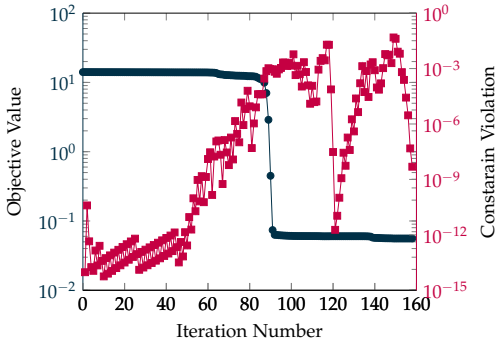


(c) Objective value and dual infeasibility.

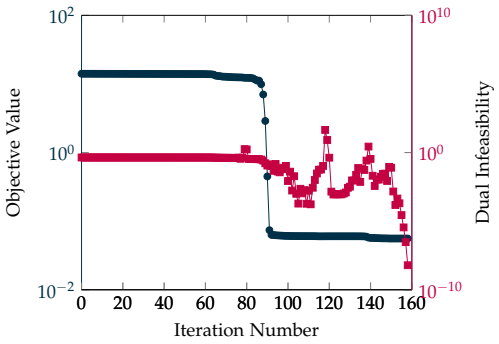
Figure A.2.: The final configuration computed with IPOPT for Level 1 and Curve 2 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 1, Curve 2 are shown. For convenience all values are given logarithmically.



(a) Final configuration

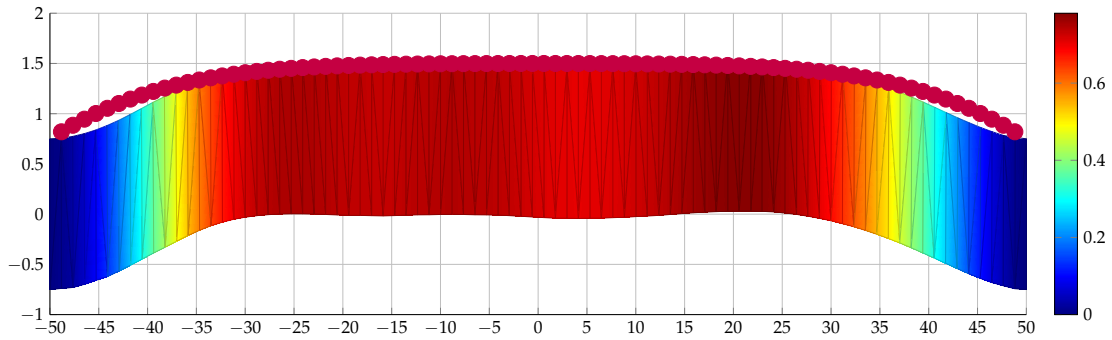


(b) Objective value and constrain violation.

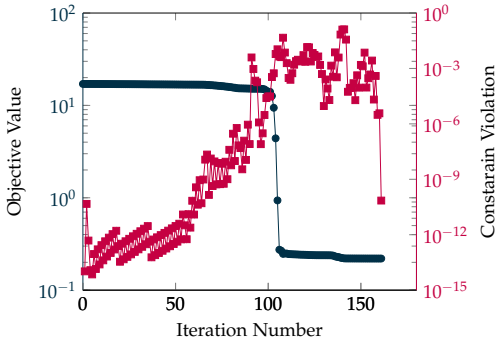


(c) Objective value and dual infeasibility.

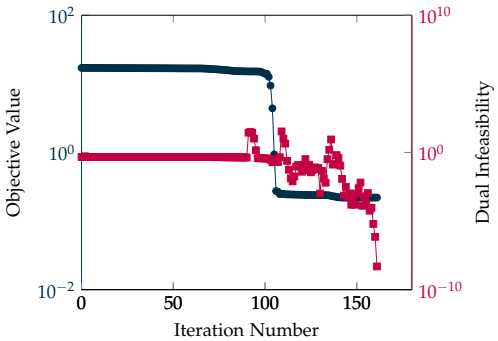
Figure A.3.: The final configuration computed with IPOPT for Level 1 and Curve 3 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 1, Curve 3 are shown. For convenience all values are given logarithmically.



(a) Final configuration

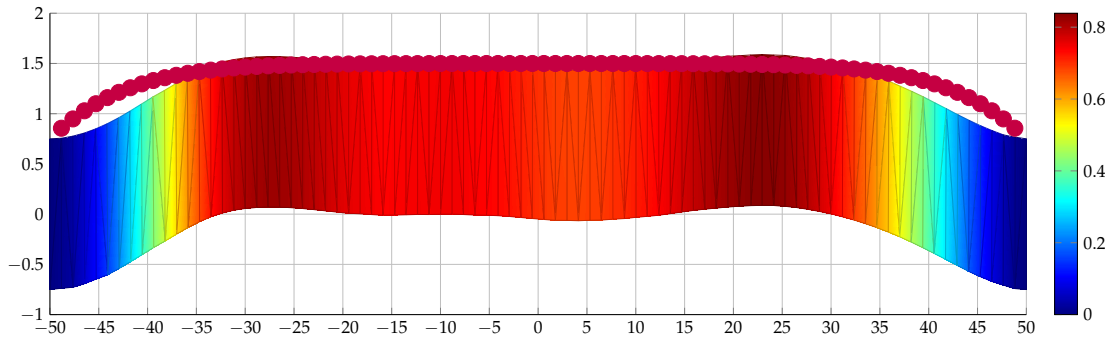


(b) Objective value and constrain violation.

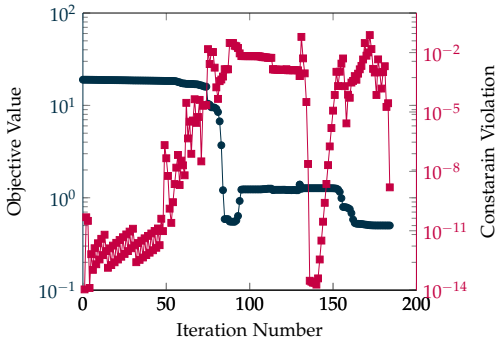


(c) Objective value and dual infeasibility.

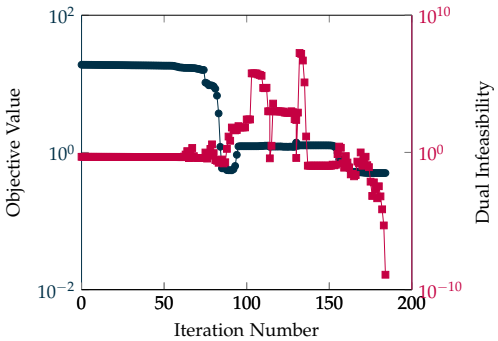
Figure A.4.: The final configuration computed with IPOPT for Level 1 and Curve 4 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 1, Curve 4 are shown. For convenience all values are given logarithmically.



(a) Final configuration

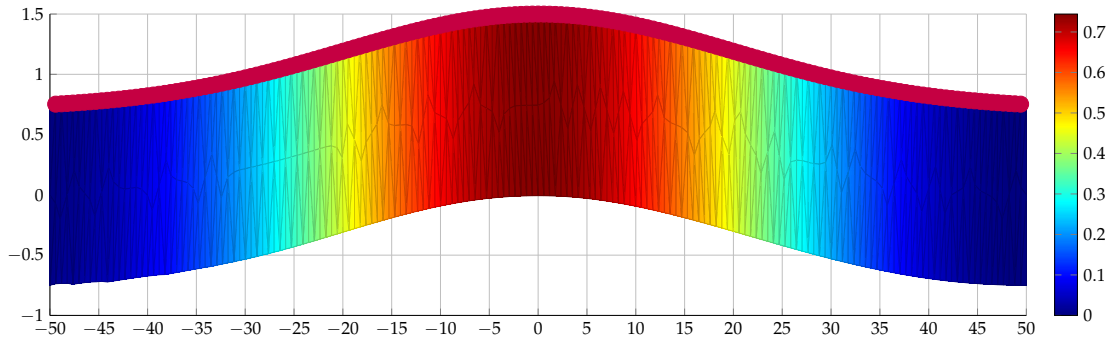


(b) Objective value and constrain violation.

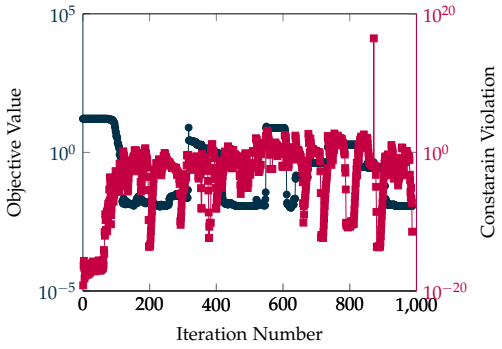


(c) Objective value and dual infeasibility.

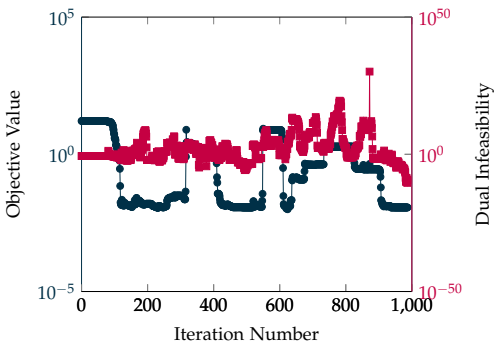
Figure A.5.: The final configuration computed with IPOPT for Level 1 and Curve 5 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 1, Curve 5 are shown. For convenience all values are given logarithmically.



(a) Final configuration

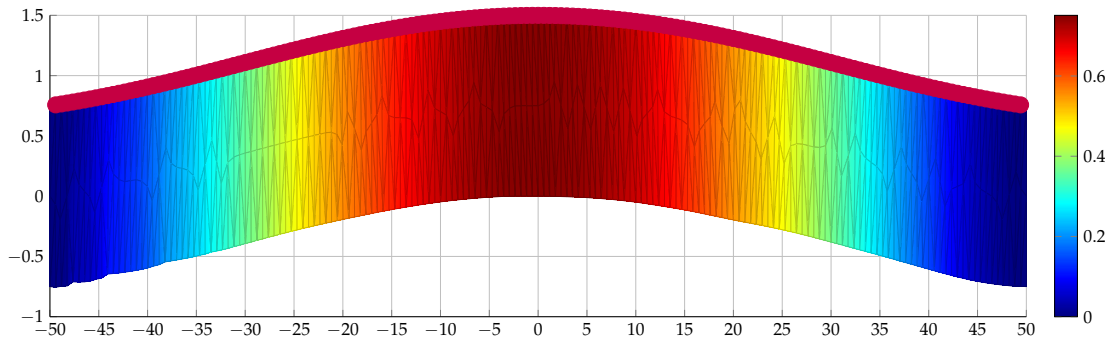


(b) Objective value and constrain violation.

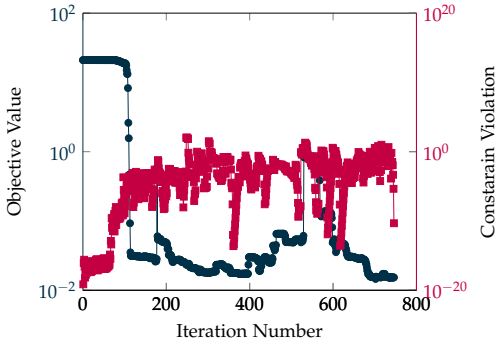


(c) Objective value and dual infeasibility.

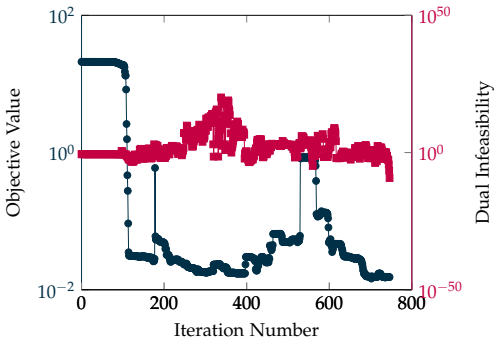
Figure A.6.: The final configuration computed with IPOPT for Level 2 and Curve 1 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 2, Curve 1 are shown. For convenience all values are given logarithmically.



(a) Final configuration

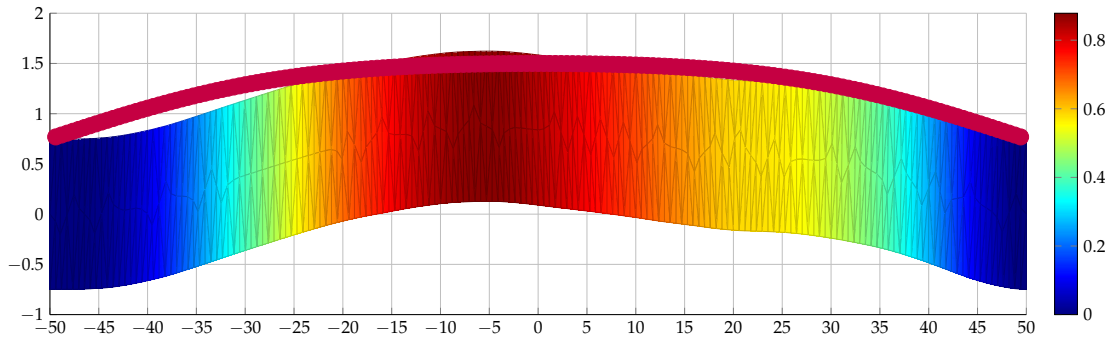


(b) Objective value and constrain violation.

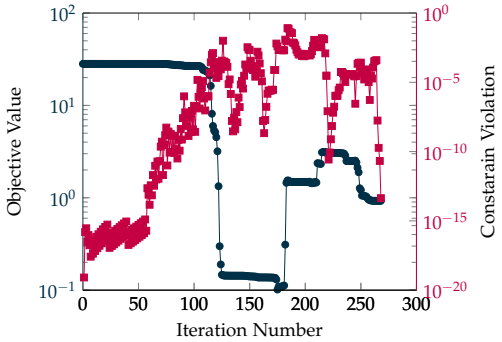


(c) Objective value and dual infeasibility.

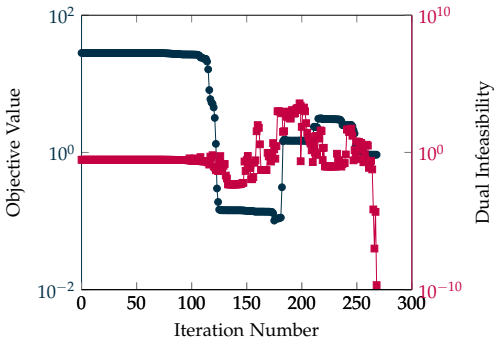
Figure A.7.: The final configuration computed with IPOPT for Level 2 and Curve 2 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 2, Curve 2 are shown. For convenience all values are given logarithmically.



(a) Final configuration

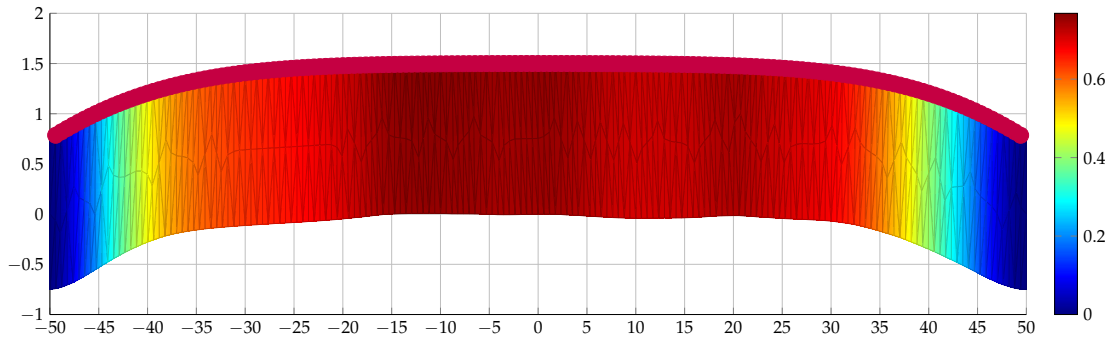


(b) Objective value and constrain violation.

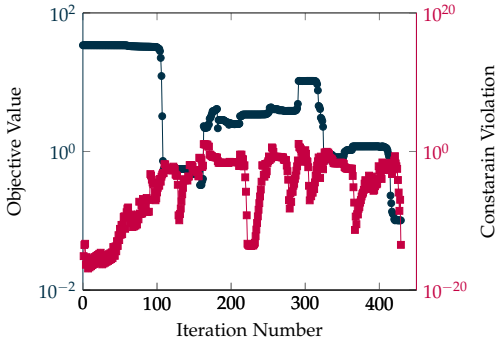


(c) Objective value and dual infeasibility.

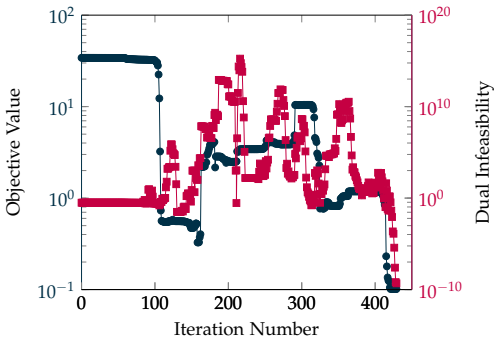
Figure A.8.: The final configuration computed with IPOPT for Level 2 and Curve 3 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 2, Curve 3 are shown. For convenience all values are given logarithmically.



(a) Final configuration

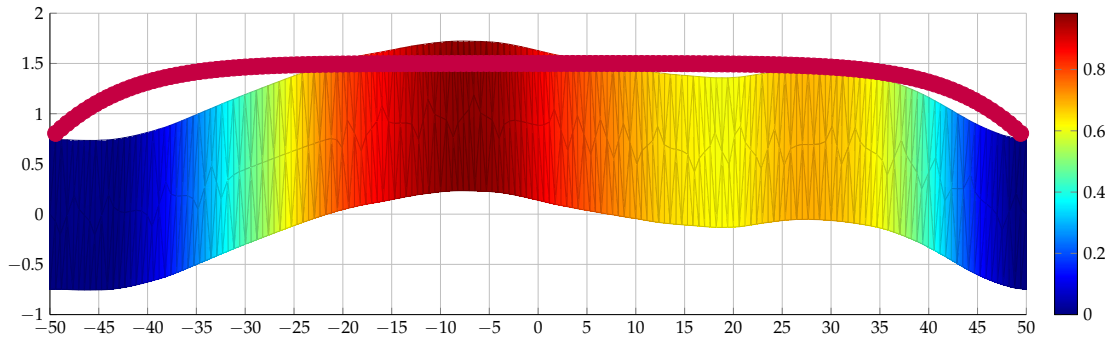


(b) Objective value and constrain violation.

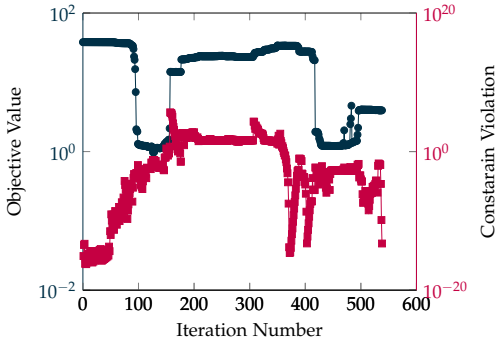


(c) Objective value and dual infeasibility.

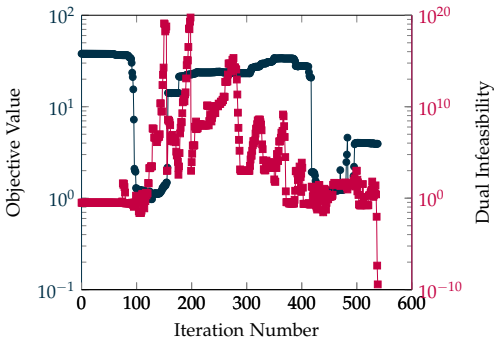
Figure A.9.: The final configuration computed with IPOPT for Level 1 and Curve 4 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 1, Curve 4 are shown. For convenience all values are given logarithmically.



(a) Final configuration

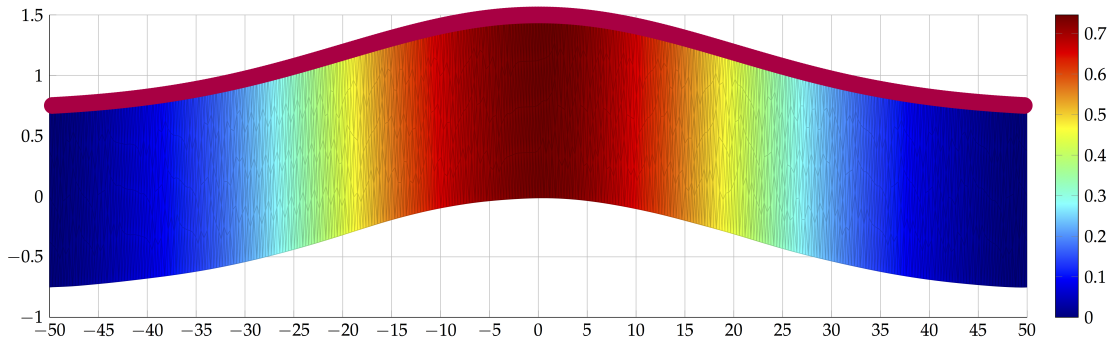


(b) Objective value and constrain violation.

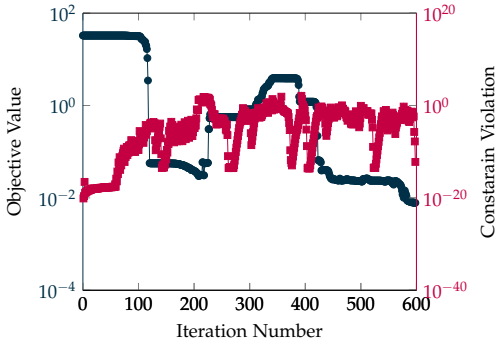


(c) Objective value and dual infeasibility.

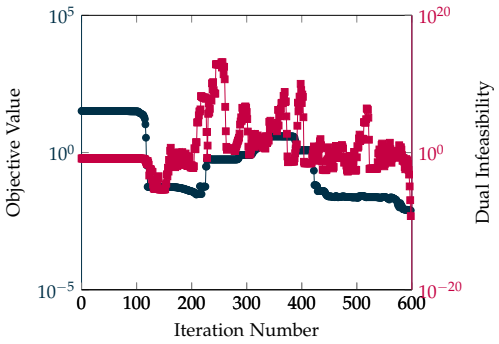
Figure A.10.: The final configuration computed with IPOPT for Level 2 and Curve 5 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 2, Curve 5 are shown. For convenience all values are given logarithmically.



(a) Final configuration

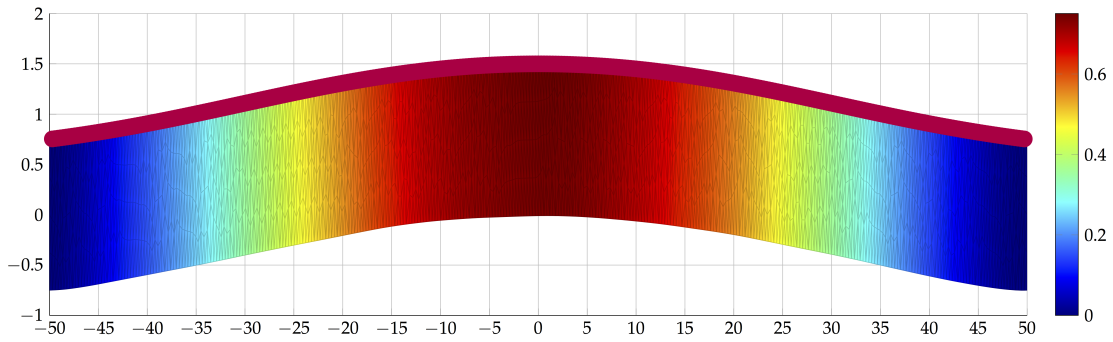


(b) Objective value and constrain violation.

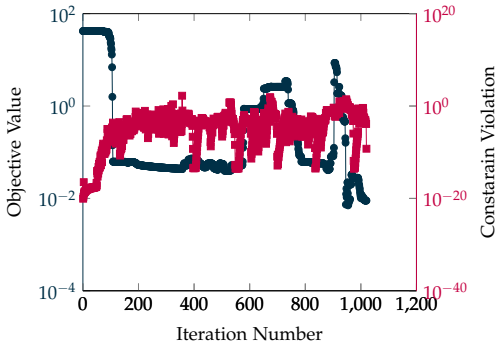


(c) Objective value and dual infeasibility.

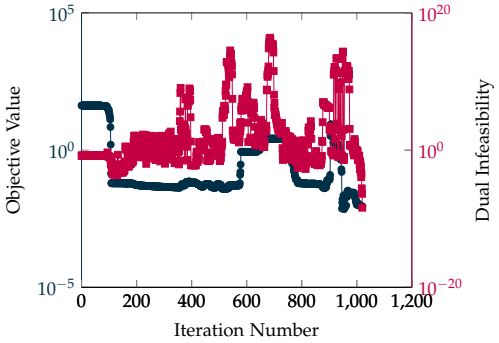
Figure A.11.: The final configuration computed with IPOPT for Level 3 and Curve 1 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 3, Curve 1 are shown. For convenience all values are given logarithmically.



(a) Final configuration

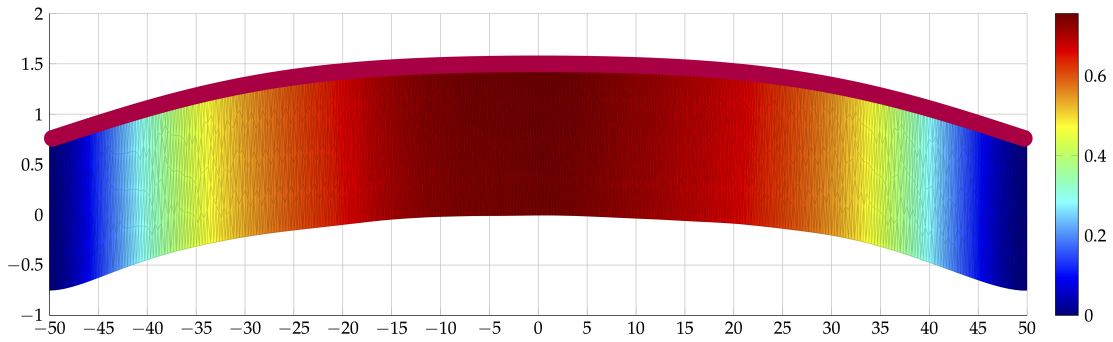


(b) Objective value and constrain violation.

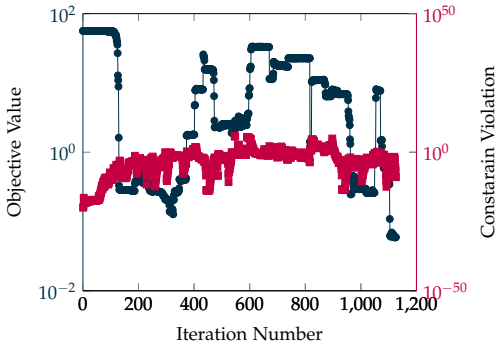


(c) Objective value and dual infeasibility.

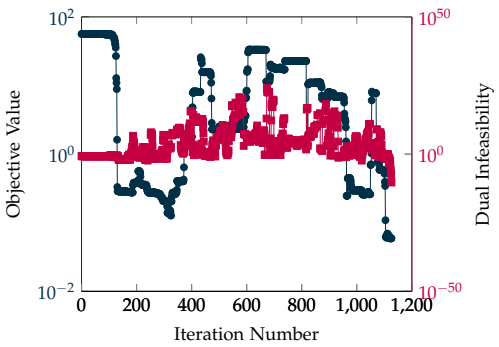
Figure A.12.: The final configuration computed with IPOPT for Level 3 and Curve 2 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 3, Curve 2 are shown. For convenience all values are given logarithmically.



(a) Final configuration

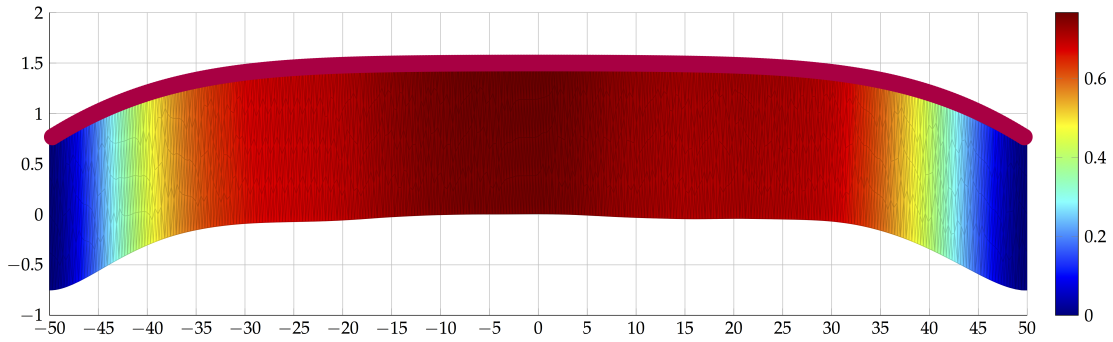


(b) Objective value and constrain violation.

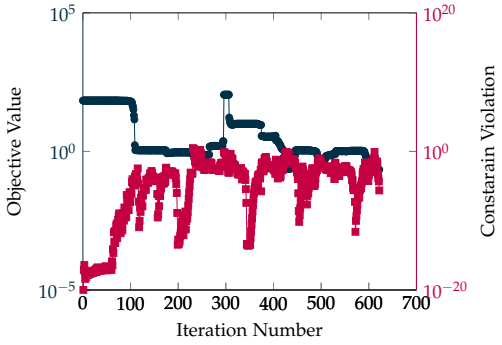


(c) Objective value and dual infeasibility.

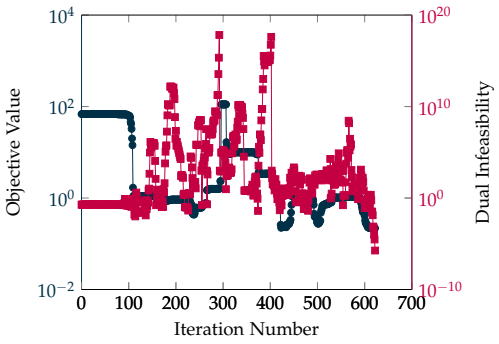
Figure A.13.: The final configuration computed with IPOPT for Level 3 and Curve 3 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 3, Curve 3 are shown. For convenience all values are given logarithmically.



(a) Final configuration

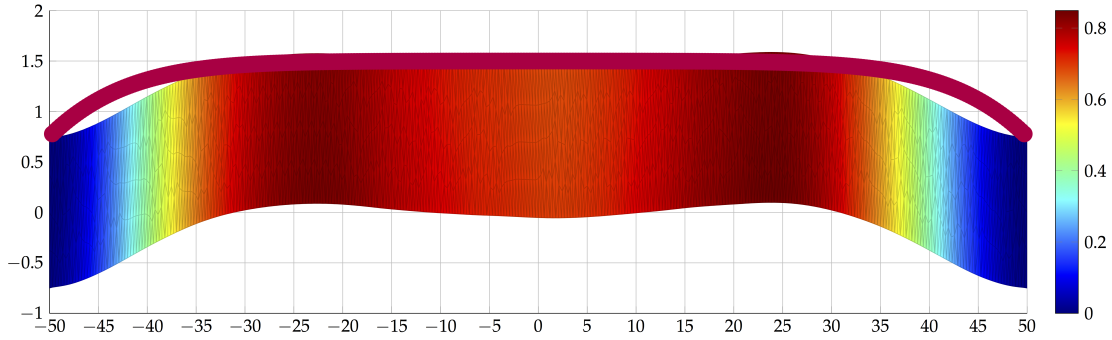


(b) Objective value and constrain violation.

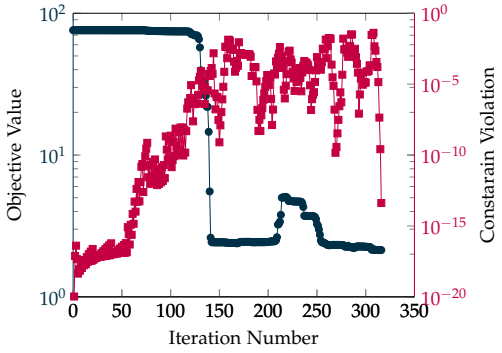


(c) Objective value and dual infeasibility.

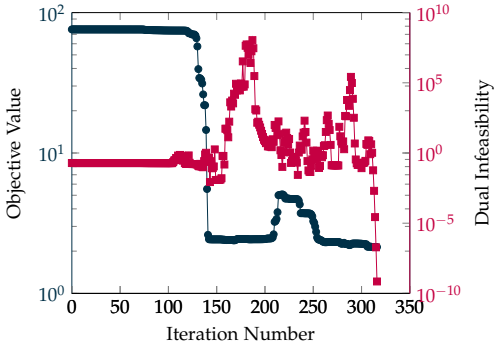
Figure A.14.: The final configuration computed with IPOPT for Level 3 and Curve 4 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 3, Curve 4 are shown. For convenience all values are given logarithmically.



(a) Final configuration

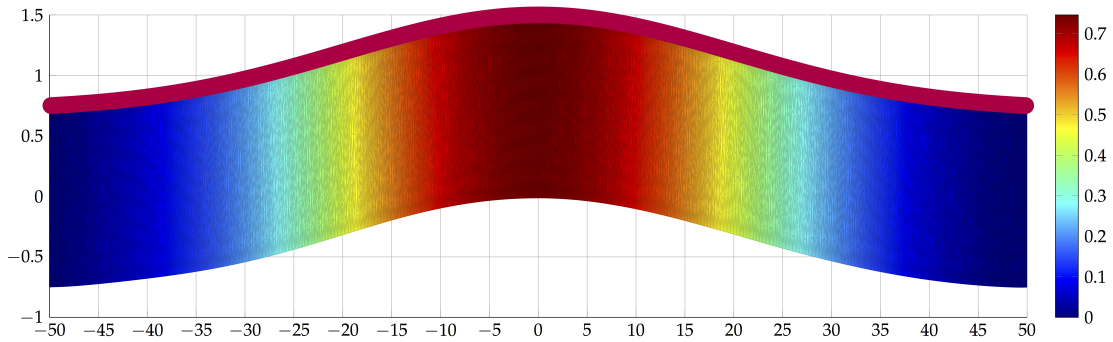


(b) Objective value and constrain violation.

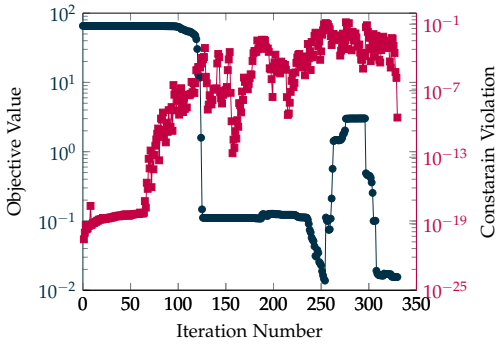


(c) Objective value and dual infeasibility.

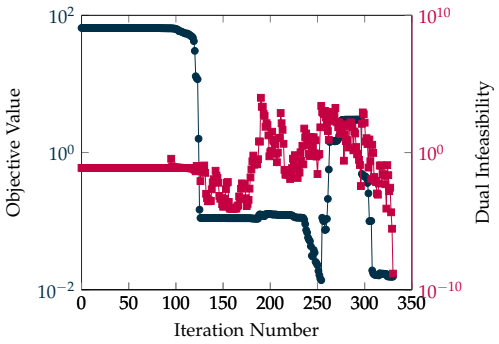
Figure A.15.: The final configuration computed with IPOPT for Level 3 and Curve 5 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 3, Curve 5 are shown. For convenience all values are given logarithmically.



(a) Final configuration

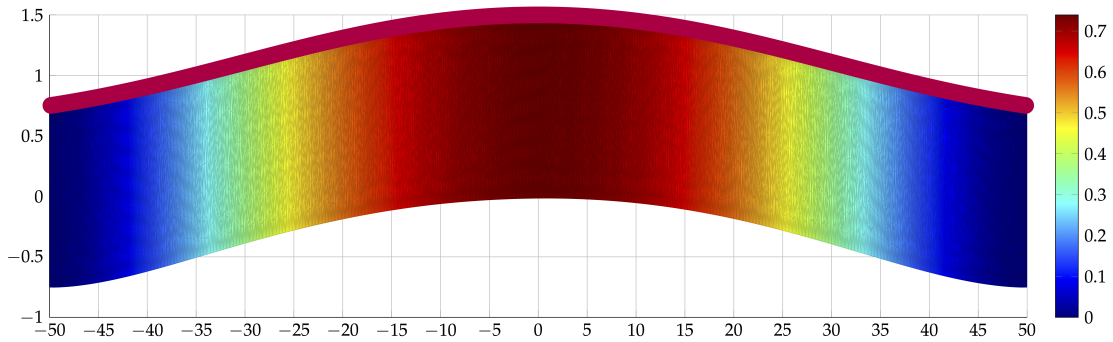


(b) Objective value and constrain violation.

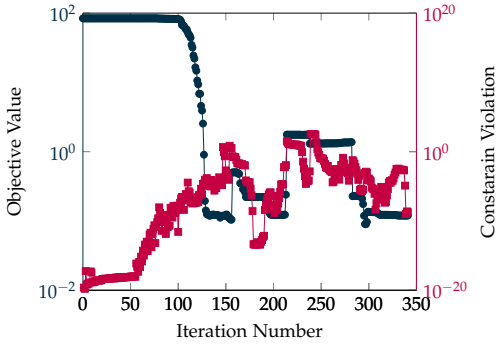


(c) Objective value and dual infeasibility.

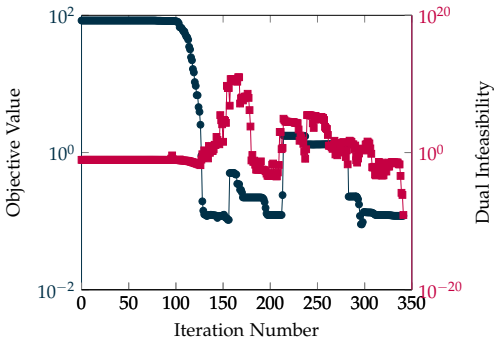
Figure A.16.: The final configuration computed with IPOPT for Level 4 and Curve 1 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 4, Curve 1 are shown. For convenience all values are given logarithmically.



(a) Final configuration

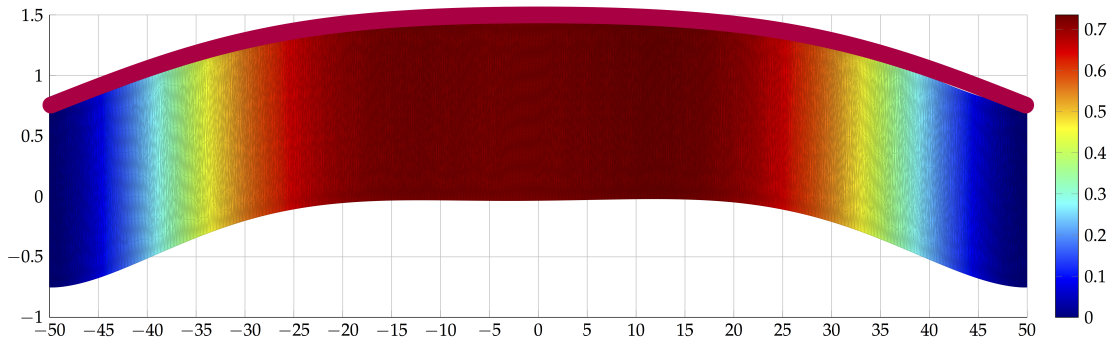


(b) Objective value and constrain violation.

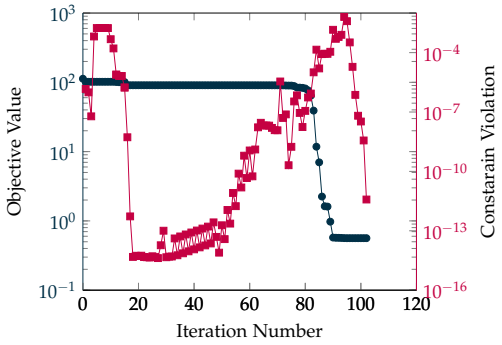


(c) Objective value and dual infeasibility.

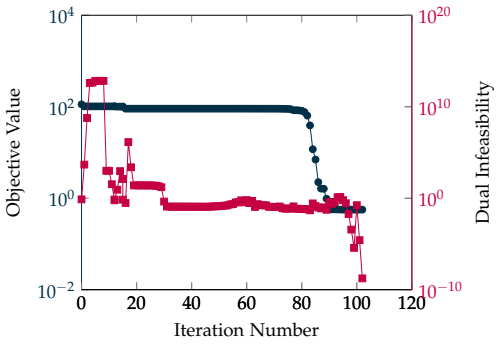
Figure A.17.: The final configuration computed with IPOPT for Level 4 and Curve 2 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 4, Curve 2 are shown. For convenience all values are given logarithmically.



(a) Final configuration

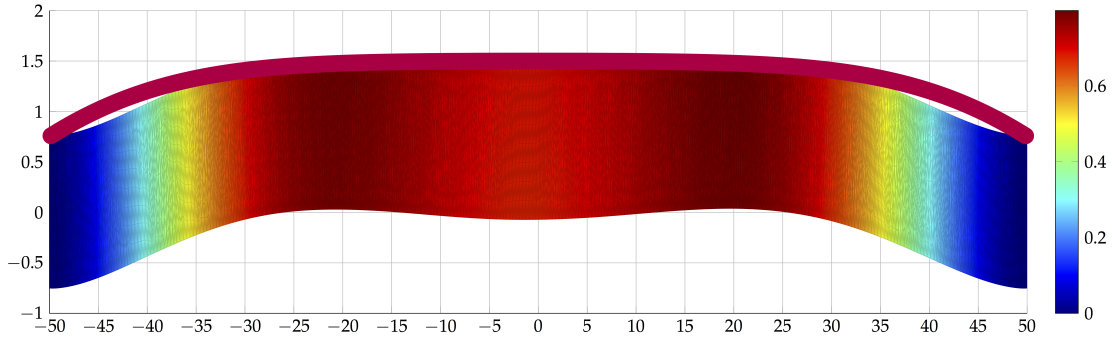


(b) Objective value and constrain violation.

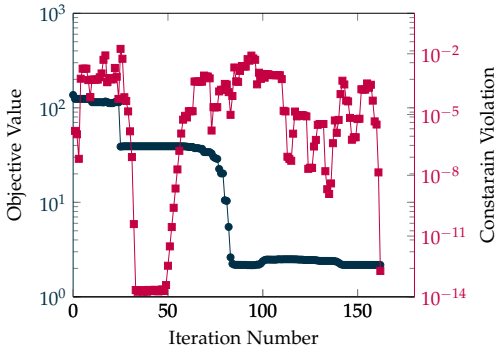


(c) Objective value and dual infeasibility.

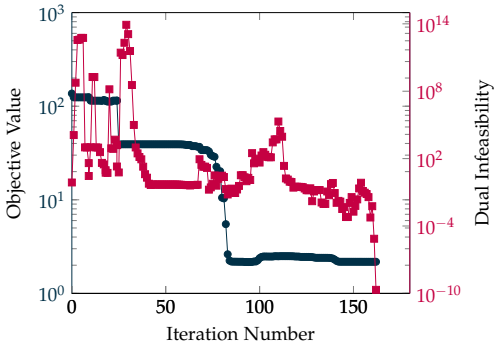
Figure A.18.: The final configuration computed with IPOPT for Level 4 and Curve 3 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 4, Curve 3 are shown. For convenience all values are given logarithmically.



(a) Final configuration

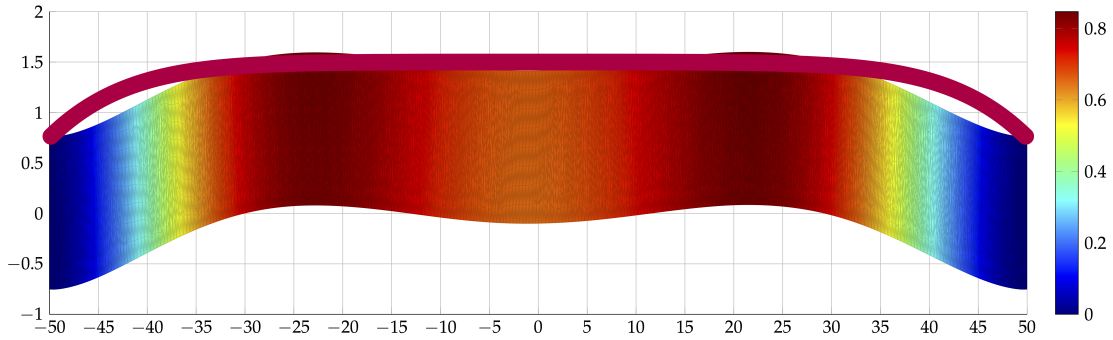


(b) Objective value and constrain violation.

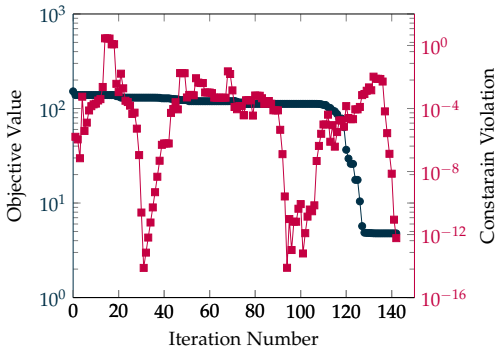


(c) Objective value and dual infeasibility.

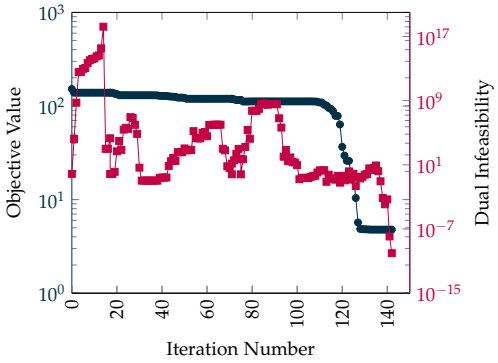
Figure A.19.: The final configuration computed with IPOPT for Level 4 and Curve 4 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 4, Curve 4 are shown. For convenience all values are given logarithmically.



(a) Final configuration



(b) Objective value and constrain violation.



(c) Objective value and dual infeasibility.

Figure A.20.: The final configuration computed with IPOPT for Level 4 and Curve 5 is depicted. The coloring gives the absolute displacement. Further, the change of the objective function value, the constraint violation, and the dual infeasibility during the iteration for Level 4, Curve 5 are shown. For convenience all values are given logarithmically.

List of Figures

List of Tables