

# Optimal Thickness of Shells with an Application to Cylindrical Bodies

**Inauguraldissertation**

**zur**

**Erlangung des akademischen Grades eines**

**Doktors der Naturwissenschaften  
(Dr. rer. nat.)**

**der**

**Mathematisch-Naturwissenschaftlichen Fakultät**

**der**

**Ernst-Moritz-Arndt-Universität Greifswald**

vorgelegt von

Paul Ziemann

geboren am 12.11.1987

in Stralsund

Greifswald, den 17. Mai 2016

Dekan: Prof. Dr. Werner Weitschies

1. Gutachter: Prof. Dr. Bernd Kugelmann

2. Gutachter: Prof. Dr. Jaan Lellep

3. Gutachter: Prof. Dr. René Pinnau

Tag der Promotion: 8. September 2016

# Contents

<b>1. Introduction</b>	<b>1</b>
<b>2. Modeling</b>	<b>5</b>
2.1. Continuum mechanics . . . . .	5
2.1.1. Basic equations . . . . .	5
2.1.2. Linearization . . . . .	10
2.1.3. Summary . . . . .	11
2.2. Shell models . . . . .	12
2.2.1. Geometrical modeling . . . . .	12
2.2.2. Kinematic modeling . . . . .	18
2.2.3. Stress tensor . . . . .	21
2.2.4. Steady-state equilibrium conditions . . . . .	23
2.2.5. Comparison with other shell models . . . . .	24
2.2.6. Smoothness of midsurface and thickness . . . . .	27
2.2.7. Dynamic equations . . . . .	28
2.2.8. Summary . . . . .	30
2.3. Optimization problem . . . . .	31
<b>3. Steady-state problem</b>	<b>33</b>
3.1. Analysis of the state equation . . . . .	33
3.1.1. Existence and uniqueness of solutions . . . . .	33
3.1.2. Continuous dependence on the thickness . . . . .	36
3.1.3. Differentiable dependence on the thickness . . . . .	39
3.2. Analysis of the optimization problem . . . . .	45
3.3. Summary . . . . .	51
3.4. Methods from Shape Optimization . . . . .	51
3.4.1. Basic definitions . . . . .	52
3.4.2. Application to the steady-state problem . . . . .	54
3.4.3. Discussion of shape optimization techniques . . . . .	63
<b>4. Dynamic problem</b>	<b>65</b>
4.1. Dynamic state equation . . . . .	65
4.1.1. Weak formulation . . . . .	65
4.1.2. Existence and uniqueness of a solution . . . . .	69
4.1.3. Continuous dependence on the thickness . . . . .	75
4.1.4. Differentiable dependence on the thickness . . . . .	79
4.2. Analysis of the optimization problem . . . . .	88

4.3. Summary . . . . .	93
<b>5. Numerical solution</b>	<b>95</b>
5.1. Solution of the steady-state equation . . . . .	95
5.1.1. Shell elements and finite element displacement . . . . .	96
5.1.2. Evaluating the quantities in the variational problem . . . . .	99
5.1.3. Solution of the linear system . . . . .	103
5.2. Solution of the dynamic equation . . . . .	104
5.3. Solution of the optimization problem . . . . .	105
5.3.1. Creating a finite-dimensional problem . . . . .	106
5.3.2. Solution of the finite-dimensional problem . . . . .	107
5.3.3. Optimality test . . . . .	111
5.4. Some words on locking . . . . .	111
5.5. Summary . . . . .	112
<b>6. Results for the steady-state problem</b>	<b>113</b>
6.1. General problem data . . . . .	113
6.2. Rotational symmetric force . . . . .	115
6.2.1. Optimization results . . . . .	115
6.2.2. Comparison with results from Nestler . . . . .	118
6.3. Discontinuous force . . . . .	119
6.4. Loaded beam . . . . .	123
6.4.1. Optimization results . . . . .	123
6.4.2. Comparison with analytic solution . . . . .	125
6.5. Asymmetric force . . . . .	127
6.6. Summary . . . . .	130
<b>7. Results for the dynamic problem</b>	<b>131</b>
7.1. Force with time-dependent intensity . . . . .	131
7.2. Moving load . . . . .	135
7.3. Impact on partially clamped shell . . . . .	139
7.4. Summary . . . . .	143
<b>8. Summary and outlook</b>	<b>145</b>
<b>Bibliography</b>	<b>149</b>
<b>List of Figures</b>	<b>155</b>
<b>List of Tables</b>	<b>157</b>
<b>List of Algorithms</b>	<b>157</b>
<b>A. Appendix</b>	<b>167</b>
<b>B. Source code overview</b>	<b>173</b>

# 1. Introduction

The optimization of structures has been an ongoing and lively field of research for decades. Various problem formulations give rise for investigations. Amongst these, finding an optimal *shape* of an object is a popular task. In particular the distribution of a given amount of material across a fixed surface is an interesting assignment.

Structural optimization problems are often settled in the context of linear elasticity. This theory describes (small) deformations of an object under an applied load. Starting from a general three dimensional body, several specializations are present for different types of objects: The theory of plates and shells for objects with one thin dimension or the theory of beams for objects with two thin dimensions compared to the remaining one(s). Shells occur in various places in nature as well as in technical applications. Objects as complex as seashells, as well as roofs or domes or simpler objects like tubes are all around. The theory of the deformation of shells gained attention with contributions from Koiter ([Koi66]) and Naghdi ([Nag72]) whose models are still widely used today. However, more general models like the basic shell model ([CB03]) or nonlinear ones are available nowadays. These models also extend from the steady-state case to time-dependent deformations caused by dynamic forces or an initial momentum.

A common task is to find an optimal thickness – i.e. the expansion of the thin dimension – of a loaded shell in order to minimize its displacement under an applied load. The volume of the shell is prescribed and lower and upper bounds for the thickness are given. Those bounds arise from practical issues: The shell should not be too thin or too heavy and a maximal thickness is necessary to stay in the context of shells. Results for such kind of problem formulations can be found for different particular cases: Nestler investigated full tubes ([Nes12; Nes13]), Azhmyakov and Schmidt ([AS04]) and Lepikult et al. ([LSW99]) considered beams and the book from Neittaanmäki, Sprekels, and Tiba ([NST06]) provides results using a Naghdy-like model. In [ST09] Sprekels and Tiba collect various results on thickness and shape optimization problems in linear elasticity. Moreover, the publications from Lellep et al. ([Lel+10; LP10; LP12]) involving different types of shells with piecewise constant thickness as well as both elastic and plastic behavior are worth mentioning. In particular the work from Nestler ([Nes10a]), which itself was inspired by a dynamic problem formulation from Lellep, gave rise for an abstraction from full tubes to more general objects.

This thesis deals with thickness optimization for a broader range of shells. The basic shell model serves as the underlying theory from linear elasticity and can be viewed as middle course between full three dimensional models and the Naghdy shell model. The theoretical results are provided for arbitrary shells made of isotropic

homogeneous elastic material while the numerical results focus on parabolic ones, in particular those with a cylindrical (tube-like) shape. Both the steady-state and the dynamic case are investigated, whereas the dynamic equations are derived from scratch based on the basic shell model. The models are studied with the aim of analytical expressions for the sensitivity of the displacement of the shell with respect to its thickness. This allows not only a numerical optimization which is superior to the use of finite difference gradients but also the formulation of necessary conditions for an optimal thickness. Finally, the efficient implementation of the optimization algorithms in combination with available libraries is a noteworthy part of the thesis.

Thickness optimization can be put into the more general context of shape optimization. While the former one considers only the extension in one dimension to be optimized, the latter one tries to find an optimal shape of an object as a whole. An overview of modern techniques for shape optimization can be found in the book from Delfour and Zolésio ([DZ11]). The application of these techniques to the actual thickness optimization problem is discussed and compared to the particular approach chosen in this thesis. A related field of research is topology optimization which considers the optimal placement of holes into structures. Beside the view of the thickness as a (continuously) distributed parameter also discretized values (e.g. on finite element based nets) can be chosen as design variables as shown by Arnout, Firl, and Bletzinger ([AFB12]).

From a mathematical point of view, the models from linear elasticity provide a system of partial differential equations whose solution is the displacement of the shell under a given load. The thickness can be viewed as a distributed control of this system. It does not occur as an explicit parameter in the system, but rather defines certain integral boundaries in the weak formulation of the state equation. The aim is to find an optimal control to minimize a given objective functional – here, the objective rates the displacement of the shell. The additional constraints restrict the space of admissible thickness distributions. Optimal control of partial differential equations in general is an ongoing field of research in mathematics. A crucial factor in optimal control theory is the type of the underlying differential equation. An overview for elliptic systems is given in the book from Neittaanmäki, Sprekels, and Tiba ([NST06]) or Tröltzsch ([Trö10]) where also results for the parabolic case can be found. The dynamic equation is a hyperbolic one with less results available.

For the numerical solution of partial differential equations and nonlinear optimization problems a wide range of libraries is available. With growing computational power and the widespread use of parallelization techniques discretized optimal control problems can be solved in reasonable time on fine grids. Remarkable modern tools are the Pardiso solver for linear equation systems ([Par15]), the interior point optimization Ipopt ([Kaw+15]) and the Intel Math Kernel Library ([MKL15]). The application of these tools and the effective combination of their powers is the base for a fast and precise calculation of an optimal thickness distribution.

Altogether, this thesis provides several contributions to thickness optimization of shells in linear elasticity: Starting from the use of the basic shell model and the derivation of the dynamic model equations, via the thorough analysis of the

displacement models regarding their sensitivity with respect to the thickness of the shell and the formulation of necessary conditions for optimal solutions in a certain case, heading to the efficient numerical implementation of the optimization process and finally the discussion of several examples involving cylindrical shells.

Following this introduction, in Chapter 2 the basic equations from continuum mechanics will be shortly reintroduced. Afterwards the geometrical modeling and the derivation of shell models will be discussed. The basic shell model will be introduced and compared to other well-known models. Moreover, the dynamic equations will be presented. The chapter will be closed with the introduction of the steady-state optimization problem which will be one of the main subjects of investigation.

The third chapter is dedicated to the steady-state case. After a thorough investigation of the displacement model, in particular its sensitivity with respect to the thickness, the first main result will be presented: An expression for the Gâteaux-derivative of the displacement with respect to the thickness. This result will be applied to the reduced optimization problem and a formula for the sensitivity of the objective with respect to the thickness as well as necessary conditions for an optimal solution will be provided. The second part of this chapter deals with the embedding of the thickness optimization problem into the field of shape optimization. The basic terms such as shape functions and shape derivatives will be introduced and the shape derivative for the particular problem treated in this thesis will be calculated. The result will show that both Gâteaux- and shape derivative coincide, however the latter one needs higher regularity assumptions.

After that, the fourth chapter considers the dynamic problem. First, the dynamic state equation will be restated in a setting of rigged Hilbert spaces and investigated regarding existence and uniqueness of a solution. Under tightened regularity assumptions a sensitivity theorem for the Gâteaux-derivative of the displacement with respect to the thickness will be provided. The result and the proof of this theorem and its preceding lemmas are other main results of this thesis. Following, the dynamic optimization problem will be stated and the sensitivity of the reduced objective with respect to the thickness will be investigated. In addition, necessary conditions for an optimal solution will be provided.

The fifth chapter deals with the numerical implementation of the optimization problem. The discretization of both the steady-state and the dynamic optimization problem with help of the finite element method will be presented. The required tools for the solution of the state equation and for the actual optimization will be introduced and their efficient combination in the implementation will be explained.

In Chapter 6 and 7 the numerical results are presented. Different examples involving cylindrical shells will be investigated. Those examples consist of various types of loads and boundary conditions. In addition to the optimal thickness the corresponding displacements will be compared to non-optimized ones and the differences will be commented. Partly, the optimal thickness will be compared to available solutions from different shell models. Every example will be completed with a convergence study for the optimal thickness on successively refined grids.



## 2. Modeling

This chapter is devoted to the derivation of the state equation and the optimization problem. The first section contains an overview of the basic equations in linear elasticity. In the second part an introduction to the theory of shells is presented. This includes both the geometrical modeling of the shell body as well as the kinematical modeling and further assumptions special to shells. In addition the particular shell model used in this thesis is compared to other well known shell models from literature. The use of balance laws leads to the state equation which describes the displacement of a shell that is subject to an external load. Both the steady state (time-independent) case as well as the dynamic (time-dependent) case are investigated.

In the third part of the chapter the actual optimization problem is presented. The aim is to minimize the displacement of the shell (which is modeled by the state equation) in a proper sense. The thickness of the shell serves as a distributed control function which is restricted by lower and upper bounds together with a constraint on the volume of the shell body.

### 2.1. Continuum mechanics

This section gives an overview on continuum mechanics and the equations evolving from this theory. These equations are the basis for further work. Central terms like strains and stresses as well as balance equations are introduced. In particular the focus is on concepts from linear elasticity which allow considerable simplifications of the general equations in case of small deformations of a material body. This overview is based on the books of Altenbach ([Alt12]), Gould ([Gou94]), Liu ([Liu02]) and Temam and Miranville ([TM05]).

#### 2.1.1. Basic equations

The theory of *continuum mechanics* deals with material bodies undergoing motion and deformation ([Liu02, p. 1]). It provides equations to describe such processes in a consistent way. Here, a three dimensional *material body*  $\mathcal{B}$  is identified with a subset of the Euclidean space  $\mathbb{R}^3$  via an injective mapping

$$\kappa : \mathcal{B} \rightarrow \mathbb{R}^3$$

called *reference configuration*. Let the image of  $\kappa$  be denoted by  $\mathcal{B}_\kappa \subset \mathbb{R}^3$ . When undergoing a motion or deformation the configuration changes, i.e. the body is described by another mapping

$$\chi : \mathcal{B} \rightarrow \mathbb{R}^3.$$

The mapping  $\chi_\kappa := \chi \circ \kappa^{-1} : \mathcal{B}_\kappa \rightarrow \mathbb{R}^3$  is called the *deformation* of the body. Often, the reference configuration is fixed so that there is no need to mention it explicitly. Let this be the case from now on and let  $\mathcal{B}_\kappa$  be denoted by  $\mathcal{B}$  and  $\chi_\kappa$  by  $\chi$ .

To describe a quantity at specific points inside of the deformed body  $\chi(\mathcal{B})$  there are two options now: First, the coordinates of the deformed configuration can be used. Then a quantity  $f : \chi(\mathcal{B}) \rightarrow W$  defined on the deformed body  $\chi(\mathcal{B})$  is written in the form  $f = f(x)$ . This is known as the *Eulerian view* and often used in fluid mechanics. Second, the coordinates of the reference configuration can be used, which means a point  $x \in \chi(\mathcal{B})$  is actually described as  $x(a) = \chi(a)$  for  $a \in \mathcal{B}$ . Then the quantity  $f$  is written in the form  $f = f(x(a)) = f(a)$ . This concept is known as the *Lagrangian view* (or *material description*) and commonly used in mechanics of solids. In the following the focus is on the Lagrangian view. However, this fixing is not a decisive factor as will be seen later, because in linear elasticity there is no distinction between both views.

If the deformation is viewed as a time dependent process, there is a family of deformations  $\chi_t := \chi(a; t)$  and deformed bodies  $\mathcal{B}_t$ . The velocity of a point  $a \in \mathcal{B}$  is then given by

$$v(a) = \frac{\partial \chi(a, t)}{\partial t}.$$

Regarding time derivatives of quantities  $f_t : \mathcal{B}_t \rightarrow W$  (Eulerian view) that are viewed in material description  $f(a, t) = f_t(\chi(a; t), t)$  the notation

$$\dot{f} = \frac{df}{dt} = \frac{\partial f}{\partial t} + (\nabla \chi f)v$$

means the *total* derivative of  $f$  with respect to time.

**Kinematics** It is important to separate rigid-body transformations from local deformations (*strains*) of the body. The time dependence of the deformation is not needed in this paragraph, so the time parameter is suppressed.

Local deformations at a point  $x(a) = \chi(a) \in \chi(\mathcal{B})$  are characterized as follows: Let  $\varepsilon > 0$  and  $\gamma : (-\varepsilon, \varepsilon) \rightarrow \mathcal{B}$ ,  $\gamma = \gamma(s)$  be a continuously differentiable curve in  $\mathcal{B}$  passing through  $a$ ,  $\gamma(0) = a$ , and  $\gamma'(0) = w$ . Then in the deformed body  $\chi(\mathcal{B})$  it holds  $\chi(\gamma(0)) = x(a)$  and

$$\frac{d}{ds} \chi(\gamma(0)) = \nabla \chi(a) \gamma'(0) = \nabla \chi(a) w.$$

So the local change of tangential vectors at a point  $a \in \mathcal{B}$  under the deformation  $\chi$  (local deformation) is described by the derivative of  $\chi$ .

**Definition 2.1** (Deformation gradient, [Liu02, p. 2]). The derivative  $F := \nabla \chi$  is called *deformation gradient*.

When there is no local deformation at  $a \in \mathcal{B}$  (only rigid-body transformation) then the deformation gradient at  $a$  is equal to the unit tensor  $I$ . The term strain is

justified by the observation of *local changes of lengths*. Let the vector  $w \in \mathbb{R}^3$  be given, then it holds

$$\|\nabla\chi(a)w\|^2 = (\nabla\chi(a)w) \cdot (\nabla\chi(a)w) = (\nabla\chi(a)^T \nabla\chi(a)w) \cdot w$$

and the change of the norm of  $w$  can be described by the tensor  $\nabla\chi(a)^T \nabla\chi(a)$ .

**Definition 2.2** (Right Cauchy-Green strain tensor, [Liu02, p. 5]). The tensor field  $C := \nabla\chi^T \nabla\chi = F^T F$  is called *right Cauchy-Green strain tensor*.

It is often more convenient to describe the deformed configuration in terms of a *displacement*  $U$ , i.e. the deviation of  $\chi$  from the identity transformation,

$$\chi(a) = a + U(a).$$

The derivative of  $U$ , called *displacement gradient*, is then given by

$$H := \nabla U = \nabla\chi - I,$$

and it can be seen that in case of no local deformation  $\nabla U = 0$ , which is more intuitive. The right Cauchy-Green strain tensor can be expressed as

$$C = (I + H)^T (I + H) = I + H + H^T + H^T H.$$

Another interesting quantity is the *difference* of lengths after a deformation. It is calculated by

$$\|\nabla\chi(a)w\|^2 - \|w\|^2 = Cw \cdot w - w \cdot w = ((C - I)w) \cdot w =: (2Ew) \cdot w.$$

**Definition 2.3** (Green-St. Venant strain tensor, [Liu02, p. 8]). The tensor field

$$E = \frac{1}{2}(C - I) = \frac{1}{2}(H + H^T + H^T H)$$

is called *Green-St. Venant strain tensor* (*Green strain tensor*).

Like the displacement gradient, the Green strain tensor at  $a \in \mathcal{B}$  equals zero in case there is no local deformation at  $a$ .

The tensors introduced in this paragraph are the basic ones in the Lagrangian description of kinematics. The (linearized) Green strain tensor from Definition 2.7 is the strain tensor which will play the most important role in further text.

**Stresses** As a consequence of external forces acting on a body, its interior is in a state of stress ([Alt12, p. 140]). Internal forces arise and different parts of the body interact with each other. It is assumed that this interaction is only due to contact forces (*surface traction*, [Liu02, p. 43]). To describe this, the principle of *stress* is introduced:

At a point  $x$  inside a body  $\mathcal{B}$  the body can be cut into two parts by a surface  $\partial\mathcal{P}$ . One part of the body then exerts a force on the other part through  $\partial\mathcal{P}$  at  $x$ . This force is called the surface traction  $t(x, \partial\mathcal{P})$ . To simplify the general dependence of  $t$  on  $\partial\mathcal{P}$ , in classical continuum mechanics *Cauchy's Postulate* is used which states that the traction  $t$  only depends on the normal  $n$  of  $\partial\mathcal{P}$  in  $x$ , i.e.  $t = t(x, n)$ . In other words, through two surfaces with the same (oriented) normal at  $x$  the same surface traction is exerted at  $x$ . This dependence can be even further simplified: *Cauchy's Theorem* ([Liu02, p. 45]) states that the dependence of  $t$  on  $n$  is a linear one. That is, there exists a second-order tensor field  $T$ , such that

$$t(x, n) = T(x)n. \quad (2.1.1)$$

**Definition 2.4** (Cauchy stress tensor, [Alt12, p. 144]). The second order tensor field  $T$  given in (2.1.1) is called *Cauchy stress tensor*.

**Balance equations** Balance equations can be used to express basic laws of mechanics ([Liu02, p.31]). In general, balance equations describe the temporal change of a quantity inside a region  $\mathcal{P} \subseteq \mathcal{B}$  of a body due to inflow through the boundary of  $\mathcal{P}$  (*flux*) and due to the growth of the quantity inside  $\mathcal{P}$  (*supply*). A special balance equation is known as *Euler's first law*: Here the quantity under investigation is the linear momentum of a body.

**Definition 2.5** (Linear Momentum, [Liu02, p. 41]). Let  $\chi_t$  be a (time-dependent) deformation of a body  $\mathcal{B}$ . The *linear momentum*  $P$  of a part  $\mathcal{P} \subseteq \mathcal{B}_t$  is given by

$$P(\mathcal{P}, t) = \int_{\mathcal{P}} \rho \dot{x} dV,$$

where  $\rho$  is the density of the body.

Euler's first law formulates the balance equation

$$\dot{P}(\mathcal{P}, t) = f(\mathcal{P}, t),$$

where  $f$  is the total force acting on  $\mathcal{P}$  at time  $t$ . Because of the *conservation of mass* ([Liu02, p. 40, ex. 2.2.2]) it holds

$$\dot{P}(\mathcal{P}, t) = \frac{d}{dt} \int_{\mathcal{P}} \rho \dot{x} dV = \int_{\mathcal{P}} \rho \ddot{x} dV. \quad (2.1.2)$$

In addition it is assumed that the total force acting on  $\mathcal{P}$  is given by a sum of body forces with density  $b$  and contact forces with traction  $t$ ,

$$f(\mathcal{P}, t) = \int_{\mathcal{P}} \rho b dV + \int_{\partial\mathcal{P}} t dS. \quad (2.1.3)$$

Putting (2.1.2), (2.1.3) and (2.1.1) together leads to the momentum equation

$$\int_{\mathcal{P}} \rho \ddot{x} dV = \int_{\mathcal{P}} \rho b dV + \int_{\partial\mathcal{P}} t dS = \int_{\mathcal{P}} \rho b dV + \int_{\partial\mathcal{P}} Tn dS, \quad (2.1.4)$$

which is also known as *Cauchy's first law* ([Liu02, p. 50]). In order to put (2.1.4) into a local formulation, the second summand is rewritten using Gauss's theorem for second order tensors (see Theorem A.1)

$$\int_{\partial\mathcal{P}} Tn \, dS = \int_{\mathcal{P}} \operatorname{div} T \, dV.$$

Using the fact that (2.1.4) is valid for arbitrary parts  $\mathcal{P} \subseteq \mathcal{B}_t$  it follows from the localization theorem

$$\rho\ddot{x} - \operatorname{div} T = \rho b \quad \text{in } \mathcal{B}_t, \quad (2.1.5)$$

which is the commonly known form of Cauchy's first law. Furthermore, there is also *Cauchy's second law* ([Liu02, p. 51]) which states that the stress tensor  $T$  is symmetric. This is a consequence of the balance of angular momentum (Euler's second law, [Liu02, p. 50]).

Equation (2.1.5) shall now be transformed from an equation on  $\mathcal{B}_t$  to an equation on the undeformed body  $\mathcal{B}$  (which corresponds to the Lagrangian view). Let  $\mathcal{P}_t = \chi_t(\mathcal{P}) \subseteq \mathcal{B}_t$ , then by transforming the integrals in (2.1.2) back from  $\mathcal{P}_t$  to  $\mathcal{P}$  it holds

$$\begin{aligned} \int_{\mathcal{P}} (\rho b) \circ \chi_t |\det F| \, dV + \int_{\partial\mathcal{P}} (T \circ \chi_t) |\det F| (F^{-T}n) \, dS \\ = \int_{\mathcal{P}} (\rho \circ \chi_t) \frac{d^2}{dt^2} (x \circ \chi_t) |\det F| \, dV, \end{aligned} \quad (2.1.6)$$

where the transformation of the volume and surface element are taken from [Liu02, p. 3, ex. 1.1.1]. Writing

$$\begin{aligned} \rho_\kappa(a) &:= (\rho \circ \chi_t(a)) |\det F(a)|, & T_\kappa(a) &:= (T \circ \chi_t(a)) |\det F(a)| F^{-T}(a), \\ \ddot{x}(a) &= \frac{d^2}{dt^2} (x \circ \chi_t)(a), & b(a) &= b \circ \chi_t(a), \end{aligned}$$

and using the divergence theorem and the fact that  $\mathcal{P} \subseteq \mathcal{B}$  is arbitrary, equation (2.1.6) can be rewritten as

$$\rho_\kappa \ddot{x} - \operatorname{div} T_\kappa = \rho_\kappa b \quad \text{in } \mathcal{B}. \quad (2.1.7)$$

**Definition 2.6** (First Piola-Kirchhoff stress tensor, [Liu02, p. 56]). The tensor field  $T_\kappa := (T \circ \chi_t) |\det F| F^{-T}$  is called *first Piola-Kirchhoff stress tensor*.

Note that unlike the Cauchy-stress tensor, the first Piola-Kirchhoff stress tensor is not symmetric, but it holds

$$T_\kappa F^T = F T_\kappa^T.$$

Finally, enforcing  $\ddot{x} = 0$  the steady-state equation (in contrast to the dynamic ones (2.1.5) and (2.1.7)) can be stated as

$$\begin{aligned} -\operatorname{div} T &= \rho b \quad \text{in } \mathcal{B}_t, \\ \text{resp. } -\operatorname{div} T_\kappa &= \rho_\kappa b \quad \text{in } \mathcal{B}. \end{aligned}$$

### 2.1.2. Linearization

In the following calculations only *small* displacements are of interest. This gives rise to a linearization of the above defined quantities and will lead to a considerable simplification of the model equations. From now on, the displacement  $U$  is supposed to be small, i.e.  $\|U\| \ll 1$ . This leads to

$$x(a) = U(a) + a \approx a,$$

so that a distinction between Eulerian and Lagrangian view is no longer necessary. Moreover, the Green strain tensor

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

is linearized by canceling the  $H^T H$ -term.

**Definition 2.7** (Linearized Green strain tensor, [Liu02, p. 10]). The tensor field

$$e := \frac{1}{2}(H + H^T) = \frac{1}{2}(\nabla U + \nabla U^T)$$

is called *linear Green strain tensor (strain tensor)*.

Because it holds for the deformation  $\chi(a) = x(a) \approx a$ , it is  $F = \nabla \chi \approx I$ , and hence

$$T_\kappa = (T \circ \chi) |\det F| F^{-T} \approx T.$$

**Definition 2.8** (Stress tensor). In case of small displacements, denote  $\sigma := T \approx T_\kappa$  (*stress tensor*).

Moreover, for the density it holds  $\rho_\kappa = \rho \circ \chi_t |\det F| \approx \rho$ . Finally, the equations of motion and steady-state can be rewritten in the form

$$\rho \ddot{x} - \operatorname{div} \sigma = \rho b, \quad \text{resp.} \quad -\operatorname{div} \sigma = \rho b \quad \text{in } \mathcal{B}, \quad \sigma = \sigma^T.$$

**Constitutive Equations** In the past paragraphs, deformations were related to strains on the one side and stresses were related to external forces on the other side. Now, the missing link is the one between stresses and strains, which then allows to calculate displacements caused by external forces. This linkage between stresses and strains is modeled by *constitutive equations*. Constitutive equations, in difference to kinematical and balance equations, are the only ones which are material dependent ([Liu02, p. 63]). *Elastic* materials are characterized by their ability to change their shapes under external forces and to return to their original shape, once the force is taken away ([Gou94, p. 55]). This is expressed in the dependence of the stress tensor  $\sigma = \sigma(e)$  only on the strain tensor  $e$ .

It is assumed, that the undeformed configuration is a stress-free one, i.e.  $\sigma(0) = 0$ . In the context of small deformations (*linear elasticity*) the dependence is linearized ([Liu02, p. 95])

$$\sigma(e) = \sigma(0) + D\sigma(0)e =: Le,$$

which is also known as *Hooke's Law*.

**Definition 2.9** (Elasticity tensor, [Liu02, p. 108]). In linear elasticity, the tensor field  $L$  which relates strain and stress through  $\sigma = Le$  is called *elasticity tensor*.

Note that  $L$  is a three-dimensional fourth-order tensor field. Hence it has 81 independent components. Due to the symmetry of  $\sigma$  and  $e$ , the number of independent components reduces to 36. In addition, it is assumed that the work done during a deformation only depends on the start and end configuration but not on the specific path taken (*conservative system*). This is equivalent to the symmetry of  $L$  itself, so the independent components reduce further to 21 ([Gou94, pp. 56–57]).

In further work only isotropic materials are considered. In isotropic material there is no preferred direction (in contrast to e.g. fibered materials like wood, where one direction is distinguished). Then even more symmetry conditions arise and only two independent components will be left (see (2.2.17)).

### 2.1.3. Summary

At the end of this section the basic equations in linear elasticity are collected: Let  $\mathcal{B} \subset \mathbb{R}^3$  be a body with density  $\rho$  and  $b : \mathcal{B} \rightarrow \mathbb{R}^3$  be an applied external force density. Let the resulting displacement  $U : \mathcal{B} \rightarrow \mathbb{R}^3$  be small. Then in the theory of linear elasticity it holds for the balance equations

$$\left\{ \begin{array}{lll} -\operatorname{div} \sigma = \rho b & \text{in } \mathcal{B} & \text{(steady-state)} \quad (2.1.8a) \\ \rho \ddot{x} - \operatorname{div} \sigma = \rho b & \text{in } \mathcal{B} \times [0, T] & \text{(dynamic)} \quad (2.1.8b) \\ \sigma = \sigma^T & \text{in } \mathcal{B} & \text{(angular momentum)} \quad (2.1.8c) \end{array} \right.$$

and additionally

$$\text{Constitutive equations: } \sigma = Le, \quad L \text{ symmetric}$$

$$\text{Kinematic equations: } e = \frac{1}{2} (\nabla U + \nabla U^T).$$

Together with proper boundary conditions these equations determine the resulting displacement caused by the applied force in a unique way. In this system there are 21 equations for 21 unknowns. In further work instead of this full formulation a displacement based one will be used: The second and third group of equations will be inserted into the first one, so that only the displacement will be left as unknown. This results in a system of three equations to be solved.

**Weak formulation** The steady-state equation (2.1.8a) is turned into a weak formulation. The stress tensor  $\sigma = \sigma(U) = Le(U)$  is understood as a quantity dependent on the displacement  $U$ . Let all mappings be equipped with appropriate regularity. For a detailed discussion of this topic see Section 2.2.4. Then (2.1.8a) is multiplied by a function  $V : \mathcal{B} \rightarrow \mathbb{R}^3$  and integrated over  $\mathcal{B}$ :

$$\int_{\mathcal{B}} -\operatorname{div} \sigma(U) \cdot V \, dV = \int_{\mathcal{B}} \rho b \cdot V \, dV.$$

Now, Green's formula for symmetric second-order tensors (see Theorem A.2) is used for the left-hand-side and the weak formulation reads

$$\int_{\mathcal{B}} \sigma(U) : e(V) \, dV + \int_{\partial \mathcal{B}} (\sigma(U) \cdot V) \cdot n \, dS = \int_{\mathcal{B}} \rho b \cdot V \, dV. \quad (2.1.9)$$

By imposing suitable boundary conditions on  $U$  and  $V$ , the boundary integral vanishes and the equation

$$\int_{\mathcal{B}} \sigma(U) : e(V) \, dV = \int_{\mathcal{B}} \rho b \cdot V \, dV \quad (2.1.10)$$

is left.

A weak formulation of the dynamic equation is derived and discussed later in Section 4.1.

## 2.2. Shell models

This section deals with special geometrical objects called shells. The consideration of shells in relation with linear elasticity leads to simplifications of the equations derived in the last section. First, a shell is described from a geometrical point of view and cylindrical shells are characterized. These are the test objects in later numerical calculations. After that, a particular ansatz for the displacement is made and the corresponding strain and stress tensors are presented. Moreover, the balance equations for the steady-state and dynamic case are investigated. Finally, the shell model used in this thesis is compared to other ones present in literature. The following sections about modeling and the steady-state balance equation are mainly based on the books from Chapelle and Bathe ([CB03]) and Ciarlet ([Cia00]).

### 2.2.1. Geometrical modeling

The most important characteristic of a three dimensional object called *shell* is that it is *thin* in one dimension compared to the other ones. A common starting point for the derivation of shell models is a surface in  $\mathbb{R}^3$ , called midsurface, which is then equipped with a thickness to obtain the whole shell body. The geometrical properties of the midsurface characterize the shell, and in order to describe them in a convenient way, some instruments from differential geometry are introduced here.

### Differential geometry on the midsurface

Let a bounded domain  $\omega \subset \mathbb{R}^2$  and an injective mapping  $z : \omega \rightarrow \mathbb{R}^3$ ,  $z = z(\xi^1, \xi^2)$  (where the smoothness of  $z$  is discussed later in Subsection 2.2.6) be given. Then  $\mathcal{S} := z(\omega)$  is called the *midsurface* of the shell. At first, a local *covariant* basis is introduced on the midsurface by

$$a_1(\xi^1, \xi^2) := \frac{\partial z(\xi^1, \xi^2)}{\partial \xi^1}, \quad a_2(\xi^1, \xi^2) := \frac{\partial z(\xi^1, \xi^2)}{\partial \xi^2}. \quad (2.2.1)$$

The injectivity of  $z$  ensures that these two vectors are linear independent and that they span a two-dimensional tangential plane

$$T_{(\xi^1, \xi^2)} := \text{span}\{a_1(\xi^1, \xi^2), a_2(\xi^1, \xi^2)\}.$$

Additionally, let in every point on  $\mathcal{S}$  an orthonormal vector

$$a_3(\xi^1, \xi^2) := \frac{a_1(\xi^1, \xi^2) \times a_2(\xi^1, \xi^2)}{\|a_1(\xi^1, \xi^2) \times a_2(\xi^1, \xi^2)\|_2}$$

be given. The corresponding *contravariant* basis vectors  $a^1$  and  $a^2$  of  $T_{(\xi^1, \xi^2)}$  are uniquely defined by

$$a_\alpha(\xi^1, \xi^2) \cdot a^\beta(\xi^1, \xi^2) = \delta_{\alpha, \beta}, \quad \alpha, \beta = 1, 2$$

where  $\delta$  denotes the Kronecker delta.

To investigate geometrical properties of the midsurface it is useful to introduce three second-order tensors called fundamental forms.

**Definition 2.10** (Fundamental forms, [CB03, pp. 23–27]). Let  $\mathcal{S}$  be a surface with local covariant basis  $a_\alpha$  and local contravariant basis  $a^\beta$ ,  $\alpha, \beta = 1, 2$ . The two-dimensional tensor

$$\mathbf{a} = \sum_{\alpha, \beta=1}^2 (a_\alpha \cdot a_\beta) a^\alpha \otimes a^\beta = \sum_{\alpha, \beta=1}^2 a_{\alpha\beta} a^\alpha \otimes a^\beta$$

is called *first fundamental form*. The tensor

$$\begin{aligned} \mathbf{b} &= \sum_{\alpha, \beta=1}^2 (a_3 \cdot \frac{\partial a_\alpha}{\partial \xi^\beta}) a^\alpha \otimes a^\beta = \sum_{\alpha, \beta=1}^2 b_{\alpha\beta} a^\alpha \otimes a^\beta \\ &= \sum_{\alpha, \beta=1}^2 (a_3 \cdot \frac{\partial a^\alpha}{\partial \xi^\beta}) a_\alpha \otimes a^\beta = \sum_{\alpha, \beta=1}^2 b_\alpha^\beta a_\alpha \otimes a^\beta \end{aligned}$$

is called *second fundamental form*. Furthermore, the tensor

$$\mathbf{c} = \sum_{\alpha, \beta=1}^2 \sum_{\lambda=1}^2 (b_\alpha^\lambda b_{\lambda\beta}) a^\alpha \otimes a^\beta = \sum_{\alpha, \beta=1}^2 c_{\alpha\beta} a^\alpha \otimes a^\beta$$

is called *third fundamental form*.

The second fundamental form carries geometrical characteristics of the midsurface, in particular the local *mean curvature*

$$H = \frac{1}{2}(b_1^1 + b_2^2)$$

and the *Gaussian curvature*

$$K = b_1^1 b_2^2 - b_2^1 b_1^2.$$

Cylindrical midsurfaces, as a special case of parabolic ones, have the property of a vanishing Gaussian curvature. For plane surfaces both quantities are zero.

To state the kinematic model for the deformation of shells it is useful to express the derivative  $\nabla u$  of a vector field  $u = u_1 a^1 + u_2 a^2$  as a second order tensor with covariant-covariant coordinates in a special way.

**Definition 2.11** (Christoffel symbols and covariant derivative, [CB03, pp. 28–29]). The *Christoffel-symbols* are defined by

$$\Gamma_{\beta\alpha}^\lambda := \frac{\partial a_\beta}{\partial \xi^\alpha} \cdot a^\lambda, \quad \alpha, \beta, \lambda = 1, 2.$$

Let  $u = u_1 a^1 + u_2 a^2$  be a vector field on  $\mathcal{S}$ . The *covariant derivative* of  $u$  is given by

$$\nabla u = \sum_{\alpha, \beta=1}^2 u_{\beta|\alpha} a^\alpha \otimes a^\beta, \quad u_{\beta|\alpha} = \frac{\partial u_\beta}{\partial \xi^\alpha} - \sum_{\lambda=1}^2 \Gamma_{\beta\alpha}^\lambda u_\lambda.$$

Moreover, the surface element for integrals over the midsurface can be stated explicitly. Let  $\mathbf{a}$  be the first fundamental form on  $\mathcal{S}$  with covariant-covariant coordinates  $a_{\alpha\beta}$ ,  $\alpha, \beta = 1, 2$ . Define

$$a := \det(a_{\alpha\beta}) := a_{11} a_{22} - a_{12}^2$$

and for a function  $f : \mathcal{S} \rightarrow \mathbb{R}$  it holds

$$\int_{\mathcal{S}} f \, dS := \int_{\omega} (f \circ z) \sqrt{a} \, d\xi^1 \, d\xi^2 = \int_{\omega} f \sqrt{a} \, d\xi^1 \, d\xi^2$$

where  $f = f \circ z$  is written if the context is clear.

In addition, Sobolev spaces on the midsurface  $\mathcal{S}$  can be defined:

**Definition 2.12** (Sobolev space on the midsurface, [Heb96, p. 10]). The Sobolev space  $W^{k,p}(\mathcal{S})$  on the midsurface  $\mathcal{S}$  is given by

$$W^{k,p}(\mathcal{S}) := \{f : \mathcal{S} \rightarrow \mathbb{R} \mid f \circ z \in W^{k,p}(\omega)\}$$

and the usual abbreviations  $L^p(\mathcal{S}) := W^{0,p}(\mathcal{S})$  and  $H^k(\mathcal{S}) := W^{k,2}(\mathcal{S})$  apply.

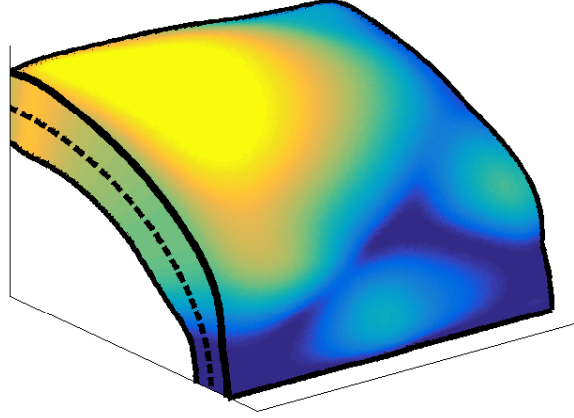


Figure 2.1.: Example of a shell body with varying thickness based on a cylindrical midsurface

### Modeling of the shell body

Now that the midsurface  $\mathcal{S}$  is modeled, the *thickness*  $\tau : \mathcal{S} \rightarrow \mathbb{R}_{>0}$  of the shell can be introduced. As for the midsurface parametrization  $z$ , the smoothness of  $\tau$  is discussed in Subsection 2.2.6. When the context is clear,  $\tau(\xi^1, \xi^2)$  will be written as a short notation for  $(\tau \circ z)(\xi^1, \xi^2)$ . Next, the parameter domain

$$\Omega_{(\tau)} := \left\{ (\xi^1, \xi^2, \xi^3) \in \mathbb{R}^3 \mid (\xi^1, \xi^2) \in \omega, \xi^3 \in \left( -\frac{\tau(\xi^1, \xi^2)}{2}, \frac{\tau(\xi^1, \xi^2)}{2} \right) \right\} \quad (2.2.2)$$

depending on the thickness  $\tau$  is introduced. The  $\xi^3$ -direction shall be the “thin” dimension, allowing to work in the context of shells. Now the mapping

$$Z_{(\tau)} : \Omega_{(\tau)} \rightarrow \mathbb{R}^3, \quad Z_{(\tau)}(\xi^1, \xi^2, \xi^3) = z(\xi^1, \xi^2) + \xi^3 a_3(\xi^1, \xi^2) \quad (2.2.3)$$

can be defined. Note that  $Z_{(\tau)}$  depends on  $\tau$  only via its domain, so this index on  $Z$  will be omitted in further text. The *shell body* is defined by

$$\mathcal{B}_{(\tau)} = Z(\Omega_{(\tau)}),$$

see for example Figure 2.1.

Once again, some geometrical quantities are derived – this time for the three-dimensional shell body  $\mathcal{B}_{(\tau)}$ . First, a local covariant basis

$$\begin{aligned} g_\alpha(\xi^1, \xi^2, \xi^3) &= \frac{\partial Z(\xi^1, \xi^2, \xi^3)}{\partial \xi^\alpha} = a_\alpha(\xi^1, \xi^2) + \xi^3 \frac{\partial a_3}{\partial \xi^\alpha}, \quad \alpha = 1, 2 \\ g_3(\xi^1, \xi^2, \xi^3) &= a_3(\xi^1, \xi^2) \end{aligned} \quad (2.2.4)$$

is defined. These three vectors are linear independent as long as

$$\tau(\xi^1, \xi^2) < 2 |R_{\min}(\xi^1, \xi^2)|, \quad (2.2.5)$$

where  $R_{\min}(\xi^1, \xi^2)$  is the radius of curvature of smallest modulus of the midsurface  $\mathcal{S}$  at  $z(\xi^1, \xi^2)$  ([CB03, p. 37]). In the cylindrical case it is always  $R_{\min} = \text{const.}$  Remember that the  $\xi^3$ -direction bounded by  $\tau$  shall be the thin dimension allowing to speak from  $\mathcal{B}_{(\tau)}$  as a shell, so in general it will even be  $\tau(\xi^1, \xi^2) \ll 2 |R_{\min}(\xi^1, \xi^2)|$ . The corresponding contravariant basis vectors  $g^1, g^2$  and  $g^3$  are uniquely defined by the equations

$$g_i \cdot g^j = \delta_{ij}, \quad i, j = 1, 2, 3.$$

Next, the metric tensor is introduced which is the three-dimensional counterpart to the first fundamental form.

**Definition 2.13** (Metric tensor, [CB03, p. 11]). Let  $\mathcal{B}_{(\tau)}$  be the shell body with local covariant and contravariant basis vectors  $g_i$  and  $g^i$ ,  $i = 1, 2, 3$ , resp. Then the second order tensor given by

$$\begin{aligned} \mathbf{g} &= \sum_{i,j=1}^3 (g_i \cdot g_j) g^i \otimes g^j = \sum_{i,j=1}^3 g_{ij} g^i \otimes g^j \\ &= \sum_{i,j=1}^3 (g^i \cdot g^j) g_i \otimes g_j = \sum_{i,j=1}^3 g^{ij} g_i \otimes g_j \end{aligned}$$

is called *metric tensor*.

Note that the matrix of contravariant components,  $(g^{ij})_{i,j=1,2,3}$  is actually the inverse of the matrix of covariant components  $(g_{ij})_{i,j=1,2,3}$ . In particular, the relations

$$g^{33} = 1, \quad g^{\alpha 3} = g^{3\alpha} = 0, \quad \alpha = 1, 2$$

hold.

With help of the metric tensor, the integral over the shell body  $\mathcal{B}_{(\tau)}$  can be rewritten. Let  $\mathbf{g}$  be the metric tensor of the shell body  $\mathcal{B}_{(\tau)}$  with covariant-covariant components  $g_{ij}$ ,  $i, j = 1, 2, 3$ . Then, according to [CB03, p. 17]

$$g := |g_1 \cdot (g_2 \times g_3)|^2 = \det(g_{ij}) \quad (2.2.6)$$

is defined and for a function  $f : \mathcal{B}_{(\tau)} \rightarrow \mathbb{R}$  it holds

$$\int_{\mathcal{B}_{(\tau)}} f \, dV := \int_{\Omega_{(\tau)}} (f \circ Z) \sqrt{g} \, d\xi^1 \, d\xi^2 \, d\xi^3 = \int_{\Omega_{(\tau)}} f \sqrt{g} \, d\xi^1 \, d\xi^2 \, d\xi^3,$$

where  $f = f \circ Z$  is written if the context is clear.

**Remark 2.1** (Calculation of  $g$ , [CB03, p. 36]). *For the quantity  $g$  appearing in the volume element it holds*

$$g = a (1 - 2H\xi^3 + K(\xi^3)^2)^2. \quad (2.2.7)$$

### Cylindrical case

Now, a closer look into the case of a cylindrical shell is taken. Due to the parabolic character of this surface many of the above geometrical quantities vanish or take a simple form. Start with the parameter domain

$$\omega = (0, l) \times (\varphi_a, \varphi_b) \quad l > 0, \quad 0 \leq \varphi_a < \varphi_b < 2\pi \quad (2.2.8)$$

and the mapping  $z : \omega \rightarrow \mathbb{R}^3$

$$z(x, \varphi) = \begin{pmatrix} x \\ R \cos(\varphi) \\ R \sin(\varphi) \end{pmatrix}, \quad (2.2.9)$$

where  $R \in \mathbb{R}$ ,  $R > 0$  is the radius of the midsurface. Note that the naming of the coordinates changed in order to reflect a length-coordinate  $x$  and an angle-coordinate  $\varphi$ . The local co- and contravariant basis vectors are given by

$$a_1(x, \varphi) = a^1(x, \varphi) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad a_2(x, \varphi) = R \begin{pmatrix} 0 \\ -\sin(\varphi) \\ \cos(\varphi) \end{pmatrix}, \quad a^2(x, \varphi) = \frac{1}{R} \begin{pmatrix} 0 \\ -\sin(\varphi) \\ \cos(\varphi) \end{pmatrix},$$

as well as an orthonormal vector

$$a_3(x, \varphi) = \begin{pmatrix} 0 \\ \cos(\varphi) \\ \sin(\varphi) \end{pmatrix}.$$

The three fundamental forms also take a quite simple form, the covariant-covariant coordinates  $a_{\alpha\beta}$ ,  $b_{\alpha\beta}$ ,  $c_{\alpha\beta}$  and the covariant-contravariant coordinates  $b_\alpha^\beta$  are given by

$$(a_{\alpha\beta}) = \begin{pmatrix} 1 & 0 \\ 0 & R^2 \end{pmatrix}, \quad (b_{\alpha\beta}) = \begin{pmatrix} 0 & 0 \\ 0 & -R \end{pmatrix}, \quad (b_\alpha^\beta) = \begin{pmatrix} 0 & 0 \\ 0 & -\frac{1}{R} \end{pmatrix}, \quad (c_{\alpha\beta}) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

From the second fundamental form, it can also be seen that the two main curvatures are constant in every point, in particular

$$\kappa_1 = 0 \quad \kappa_2 = -\frac{1}{R},$$

and that the mean and Gaussian curvatures are

$$H = -\frac{1}{2R}, \quad K = 0,$$

resp. All the Christoffel symbols vanish, i.e.  $\Gamma_{\alpha\beta}^\lambda = 0$ ,  $\alpha, \beta, \lambda = 1, 2$  and so the covariant derivative of  $u = u_1 a^1 + u_2 a^2$  simplifies to

$$\nabla u = \sum_{\alpha, \beta=1}^2 u_{\beta|\alpha} a^\alpha \otimes a^\beta, \quad u_{\beta|\alpha} = \frac{\partial u_\beta}{\partial \xi^\alpha}.$$

The determinant of the coordinate matrix of the first fundamental form is given by

$$a = \det(a_{\alpha\beta}) = R^2,$$

so the midsurface integral of a function  $f : \mathcal{S} \rightarrow \mathbb{R}$  reads

$$\int_{\mathcal{S}} f \, dS = \int_{\omega} f R \, dx \, d\varphi.$$

The shell body  $\mathcal{B}_{(\tau)}$  is parametrized using the domain  $\Omega_{(\tau)}$  from (2.2.2) and the mapping

$$Z(x, \varphi, h) = \begin{pmatrix} x \\ (R+h) \cos \varphi \\ (R+h) \sin \varphi \end{pmatrix}.$$

The covariant basis vectors for the shell body can be found as

$$g_1(x, \varphi, h) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad g_2(x, \varphi, h) = (R+h) \begin{pmatrix} 0 \\ -\sin \varphi \\ \cos \varphi \end{pmatrix}, \quad g_3(x, \varphi, h) = a_3(x, \varphi)$$

and the contravariant basis vectors as

$$g^1(x, \varphi, h) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad g^2(x, \varphi, h) = \frac{1}{R+h} \begin{pmatrix} 0 \\ -\sin \varphi \\ \cos \varphi \end{pmatrix}, \quad g^3(x, \varphi, h) = a_3(x, \varphi). \quad (2.2.10)$$

From this, the covariant-covariant components of the metric tensor are calculated

$$(g_{ij}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & (R+h)^2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

together with its determinant  $g = (R+h)^2$ .

### 2.2.2. Kinematic modeling

Now that there is a geometric concept of what a body called shell should be, the focus is back on continuum mechanics to model deformations of such bodies. Remember the initial configuration  $\mathcal{B}_{(\tau)} = Z(\Omega_{(\tau)})$  and introduce a displacement  $U : \mathcal{B}_{(\tau)} \rightarrow \mathbb{R}^3$ . As for the thickness  $\tau$ ,  $U(\xi^1, \xi^2, \xi^3) := U \circ Z(\xi^1, \xi^2, \xi^3)$  will be written. The *deformed* configuration is then given by  $(Z+U)(\Omega_{(\tau)})$ . Note that the starting point is from general three dimensional continuum mechanics, in particular the displacement  $U$  can depend nonlinearly on all three coordinates which would lead to expensive numerical calculations later on.

### Displacement ansatz and strain tensor

In the context of shells, *kinematic assumptions* are posed on the displacement in order to simplify further calculations. A well-known set of such examples are the *Reissner-Mindlin kinematic assumptions*. These state that normals to the midsurface remain *straight* and *unstretched* during deformation ([CB03, p. 81]). More precise, the term “straight” means that the points on a line segment normal to midsurface in point  $z(\xi^1, \xi^2)$ , i.e. the points contained in

$$L_{(\xi^1, \xi^2)} := \left\{ z(\xi^1, \xi^2) + \xi^3 a_3(\xi^1, \xi^2) \mid \xi^3 \in \left( -\frac{\tau(\xi^1, \xi^2)}{2}, \frac{\tau(\xi^1, \xi^2)}{2} \right) \right\}$$

will still be contained in a straight line after deformation, i.e.

$$L + U(L) = \{ z(\tilde{\xi}^1, \tilde{\xi}^2) + \tilde{\xi}^3 v \mid v \in \mathbb{R}^3, \tilde{\xi}^3 \in (a, b), a, b \in \mathbb{R}, a < b \}.$$

The term “unstretched” means that the displacement at point  $z(\xi^1, \xi^2)$  in direction  $a_3(\xi^1, \xi^2)$  is independent of the  $\xi^3$ -coordinate. Note that “unstretched” in this context does not mean that the line segments  $L$  and  $U(L)$  have the same length, i.e. in general

$$(b - a) \|v\| \neq \tau(\xi^1, \xi^2).$$

By applying these assumptions, the displacement  $U$  can be written in the simpler form

$$U(\xi^1, \xi^2, \xi^3) = u(\xi^1, \xi^2) + \xi^3 \theta(\xi^1, \xi^2). \quad (2.2.11)$$

Here,  $u = u_1 a^1 + u_2 a^2 + u_3 a^3$  describes a translational displacement of all points contained in the line segment  $L_{(\xi^1, \xi^2)}$ . It can be decomposed into a first order tensor field on the midsurface,  $\underline{u}(\cdot) : \mathcal{S} \rightarrow T_{(\cdot)}\mathcal{S}$  and the real-valued function  $u_3 : \mathcal{S} \rightarrow \mathbb{R}$ , i.e.  $u = (\underline{u}, u_3)$ . The first order tensor field  $\theta = \theta_1 a^1 + \theta_2 a^2$  describes a field of rotational vectors and the line segment  $L_{(\xi^1, \xi^2)}$  is subject to a rotation along  $\theta(\xi^1, \xi^2)$  in a linearized sense. Because the object of interest is a shell,  $\xi^3$  varies only between very tight bounds in comparison to the other coordinates and the linearization of the actual rotational displacement is justified.

By using the Reissner-Mindlin assumptions, the general three dimensional displacement  $U$  is decomposed into a translational component  $u$  and a rotational component  $\theta$  both only depending on the two coordinates parametrizing the midsurface, i.e. the identification  $U = (u, \theta) = (\underline{u}, u_3, \theta)$  is made. This is very advantageous for later numerical calculations and reflects the shell characteristics of  $\mathcal{B}_{(\tau)}$ .

Next, the linearized three dimensional strain tensor  $e$  is introduced (see Definition 2.7) whose covariant-covariant components are given by

$$e_{ij} = \frac{1}{2} \left( g_i \cdot \frac{\partial U}{\partial \xi^j} + g_j \cdot \frac{\partial U}{\partial \xi^i} \right), \quad i, j = 1, 2, 3.$$

To underline the dependence of the strain tensor on the displacement, the notation  $e_{ij} = e_{ij}(U) = e_{ij}(u, \theta)$  is introduced. A direct calculation using the displacement

ansatz (2.2.11) (see [CB03, p. 83]) leads to the components

$$e_{\alpha\beta}(u, \theta) = \gamma_{\alpha\beta}(u) + \xi^3 \chi_{\alpha\beta}(u, \theta) - (\xi^3)^2 \kappa_{\alpha\beta}(\theta), \quad \alpha, \beta = 1, 2 \quad (2.2.12a)$$

$$e_{\alpha 3}(u, \theta) = \zeta_{\alpha}(u, \theta), \quad \alpha = 1, 2 \quad (2.2.12b)$$

$$e_{33}(u, \theta) = 0. \quad (2.2.12c)$$

Here, four new first order midsurface tensor fields appear, where three of them have a specific physical meaning. The first one is the  $\xi^3$ -independent part of  $e_{\alpha\beta}$ ,

$$\gamma_{\alpha\beta}(u) = \frac{1}{2} \left( u_{\alpha|\beta} + u_{\beta|\alpha} \right) - b_{\alpha\beta} u_3,$$

which corresponds to the *membrane strain*.

The second one can be found in the  $\xi^3$ -linear part of  $e_{\alpha\beta}$ ,

$$\chi_{\alpha\beta}(u, \theta) = \frac{1}{2} \left( \theta_{\alpha|\beta} + \theta_{\beta|\alpha} - \sum_{\lambda=1}^2 b_{\beta}^{\lambda} u_{\lambda|\alpha} - \sum_{\lambda=1}^2 b_{\alpha}^{\lambda} u_{\lambda|\beta} \right) + c_{\alpha\beta} u_3,$$

which reflects the *bending strain*.

The third one appears in  $e_{\alpha 3}$ ,

$$\zeta_{\alpha}(u, \theta) = \frac{1}{2} \left( \theta_{\alpha} + \frac{\partial u_3}{\partial \xi^{\alpha}} + b_{\alpha}^1 u_1 + b_{\alpha}^2 u_2 \right), \quad (2.2.13)$$

and stands for the *shear strain*.

The fourth tensor represents the part of  $e_{\alpha\beta}$  quadratic in  $\xi^3$ ,

$$\kappa_{\alpha\beta}(\theta) = \frac{1}{2} \left( b_{\beta}^1 \theta_{1|\alpha} + b_{\beta}^2 \theta_{2|\alpha} + b_{\alpha}^1 \theta_{1|\beta} + b_{\alpha}^2 \theta_{2|\beta} \right). \quad (2.2.14)$$

Additionally, note that the normal strain  $e_{33}$  always equals zero.

### Cylindrical case

In the case of a cylindrical shell with the midsurface  $\mathcal{S}$  given by (2.2.8) and (2.2.9), the expressions for the tensors (strains) appearing in (2.2.12) can be calculated explicitly. For symmetry reasons, only the  $(\cdot)_{11}$ ,  $(\cdot)_{22}$  and  $(\cdot)_{12}$  components are stated and rearranged in Voigt-notation, see Section A.2. Using the results for the components of the three fundamental forms and the covariant derivative in the cylindrical case, it is

$$\gamma := \begin{pmatrix} \gamma_{11} \\ \gamma_{22} \\ \sqrt{2}\gamma_{12} \end{pmatrix} = \begin{pmatrix} \frac{\partial u_1}{\partial x} \\ \frac{\partial u_2}{\partial \varphi} + R u_3 \\ \frac{\sqrt{2}}{2} \left( \frac{\partial u_1}{\partial \varphi} + \frac{\partial u_2}{\partial x} \right) \end{pmatrix} \quad (2.2.15)$$

$$\begin{aligned}
\boldsymbol{\chi} &:= \begin{pmatrix} \chi_{11} \\ \chi_{22} \\ \sqrt{2}\chi_{12} \end{pmatrix} = \begin{pmatrix} \frac{\partial\theta_1}{\partial x} \\ \frac{\partial\theta_2}{\partial\varphi} + \frac{1}{R}\frac{\partial u_2}{\partial\varphi} + u_3 \\ \frac{\sqrt{2}}{2}\left(\frac{\partial\theta_1}{\partial\varphi} + \frac{\partial\theta_2}{\partial x} + \frac{1}{R}\frac{\partial u_2}{\partial x}\right) \end{pmatrix} \\
\boldsymbol{\zeta} &:= \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\left(\theta_1 + \frac{\partial u_3}{\partial x}\right) \\ \frac{1}{2}\left(\theta_2 + \frac{\partial u_3}{\partial\varphi} - \frac{1}{R}u_2\right) \end{pmatrix} \\
\boldsymbol{\kappa} &:= \begin{pmatrix} \kappa_{11} \\ \kappa_{22} \\ \sqrt{2}\kappa_{12} \end{pmatrix} = \begin{pmatrix} 0 \\ -\frac{1}{R}\frac{\partial\theta_2}{\partial\varphi} \\ -\frac{\sqrt{2}}{2R}\frac{\partial\theta_2}{\partial x} \end{pmatrix}.
\end{aligned} \tag{2.2.15 cont.}$$

### 2.2.3. Stress tensor

The relation between stress and strain tensor is given via constitutive equations, see Subsection 2.1.2. In linear elasticity, this relation is a linear one known as Hooke's law

$$\boldsymbol{\sigma} = L\boldsymbol{e}, \tag{2.2.16}$$

where  $L$  is called elasticity tensor. In general, there are 21 independent components. In view of isotropic materials where no direction is distinguished, this number reduces to two, and the covariant components of  $L$  are

$$L^{ijkl} = \tilde{L}_1 g^{ij} g^{kl} + \tilde{L}_2 (g^{ik} g^{jl} + g^{il} g^{jk}), \quad i, j, k, l = 1, 2, 3. \tag{2.2.17}$$

The numbers  $\tilde{L}_1, \tilde{L}_2$  are called *Lamé constants*.

For the stress tensor  $\boldsymbol{\sigma}$  another simplification suitable for shells is made ([CB03, p. 87]): The normal stress  $\sigma^{33}$  is neglectable, so let

$$\sigma^{33} = 0. \tag{2.2.18}$$

By exploiting this additional assumption, the components of  $\boldsymbol{\sigma}$  can be decomposed into

$$\begin{aligned}
\sigma^{\alpha\beta} &= \sum_{\lambda,\mu=1}^2 C^{\alpha\beta\lambda\mu} e_{\lambda\mu}, \quad \alpha, \beta = 1, 2 \\
\sigma^{\alpha 3} &= \sum_{\lambda=1}^2 \frac{1}{2} D^{\alpha\lambda} e_{\lambda 3}, \quad \alpha = 1, 2
\end{aligned} \tag{2.2.19}$$

where the tensors  $C$  and  $D$  are given by

$$\begin{aligned} C^{\alpha\beta\lambda\mu} &= \frac{E}{2(1+\nu)} \left( g^{\alpha\lambda} g^{\beta\mu} + g^{\alpha\mu} g^{\beta\lambda} + \frac{2\nu}{1-\nu} g^{\alpha\beta} g^{\lambda\mu} \right), \quad \alpha, \beta, \lambda, \mu = 1, 2 \\ D^{\alpha\lambda} &= \frac{2E}{1+\nu} g^{\alpha\lambda}, \quad \alpha, \lambda = 1, 2. \end{aligned} \tag{2.2.20}$$

The numbers

$$L_1 = \frac{E\nu}{(1+\nu)(1-\nu)}, \quad L_2 = \frac{E}{2(1+\nu)}$$

are called *Lamé constants of plane stress* and can be calculated from the measurable quantities *Young's modulus*  $E$  and *Poisson's ratio*  $\nu$ .

**Remark 2.2** ([CB03, p. 88]). *The vanishing normal stress assumption is not without contradiction in this case. Using (2.2.16) together with (2.2.17) leads to*

$$\begin{aligned} 0 = \sigma^{33} &= \sum_{k,l=1}^3 \tilde{L}^1 g^{33} g^{kl} + \tilde{L}^2 (g^{3k} g^{j3} + g^{3l} g^{3k}) e_{kl} \\ \Leftrightarrow e_{33} &= -\frac{\tilde{L}^1}{\tilde{L}^1 + 2\tilde{L}^2} \sum_{\alpha,\beta=1}^2 g^{\alpha\beta} e_{\alpha\beta}, \end{aligned}$$

which is  $\neq 0$  in general. This is in contradiction to the vanishing normal strain resulting from (2.2.12c). However, the component  $e_{33}$  will not be present in the later stated equilibrium formulation (2.2.25).

### Cylindrical case

In order to represent the fourth order tensors  $C$  and  $D$  in a convenient way for later calculations, their components are rearranged in a special matrix notation (see *Voigt Notation*, Section A.2). For the cylindrical shell with the basis defined in (2.2.10), this reads

$$\begin{aligned} \mathbf{C} &:= \begin{pmatrix} C^{1111} & C^{1122} & \sqrt{2}C^{1112} \\ C^{1122} & C^{2222} & \sqrt{2}C^{2212} \\ \sqrt{2}C^{1112} & \sqrt{2}C^{2212} & 2C^{1212} \end{pmatrix} = \begin{pmatrix} L_1 + 2L_2 & \frac{L_1}{(R+h)^2} & 0 \\ \frac{L_1}{(R+h)^2} & \frac{L_1 + 2L_2}{(R+h)^4} & 0 \\ 0 & 0 & \frac{2L_2}{(R+h)^2} \end{pmatrix}, \\ \mathbf{D} &:= \begin{pmatrix} D^{11} & D^{12} \\ D^{12} & D^{22} \end{pmatrix} = \begin{pmatrix} 4L_2 & 0 \\ 0 & \frac{4L_2}{(R+h)^2} \end{pmatrix}. \end{aligned} \tag{2.2.21}$$

### 2.2.4. Steady-state equilibrium conditions

The next thing to discuss are the equilibrium conditions. First, the focus is on the steady-state case. In Section 2.1 they were derived in weak form as (see (2.1.9))

$$\int_{\mathcal{B}} \sigma(U) : e(V) \, dV = \int_{\mathcal{B}} \rho b \cdot V \, dV \quad (2.2.22)$$

for all suitable test functions  $V : \mathcal{B} \rightarrow \mathbb{R}^3$ . Note that the missing boundary integral term implicitly imposes boundary conditions on  $U$  and/or  $V$ . At first, the proper *displacement space* for the test functions is discussed.

#### Displacement space

In the displacement ansatz (2.2.11) the general displacement  $U : \mathcal{B}_{(\tau)} \rightarrow \mathbb{R}^3$  was decomposed into a translational part  $u : \mathcal{S} \rightarrow \mathbb{R}^3$  and a rotational part  $\theta : \mathcal{S} \rightarrow \mathbb{R}^3$ , where  $\theta$  was a first order midsurface tensor, i.e.  $\theta(\xi^1, \xi^2) \in T_{(\xi^1, \xi^2)}\mathcal{S}$ . Furthermore,  $u$  can be written as a sum of a first order midsurface tensor  $\underline{u} = u_1 a^1 + u_2 a^2$  and  $u_3 a^3$ . For the integral arising in the weak formulation of the steady-state equation (2.2.22) to be well-defined, all components have to be equipped with  $H^1$ -regularity. A suitable choice of the displacement space is

$$\mathcal{V} := \{(v, \eta) \mid v = (v_1, v_2, v_3) = (\underline{v}, v_3) \in H^1(\mathcal{S})^2 \times H^1(\mathcal{S}), \eta \in H^1(\mathcal{S})^2\} \cap \mathcal{BC}, \quad (2.2.23)$$

where  $\mathcal{BC}$  introduces proper boundary conditions.

Two types of boundary conditions are investigated here: First, the *hardclamped* conditions. Here, the body is clamped over the whole boundary, so that neither translational nor rotational displacements are possible. Suitable for this situation is the space

$$\mathcal{V}_{\text{HC}} := \{(v, \psi) \mid v = (\underline{v}, v_3) \in H_0^1(\mathcal{S})^2 \times H_0^1(\mathcal{S}), \psi \in H_0^1(\mathcal{S})^2\}.$$

Second, there are *softclamped* (or *simply supported*) conditions. In this case, the edges of the body are supported, but not clamped. This means, that translational displacements are not possible, but rotational ones are. The proper space in this case is

$$\mathcal{V}_{\text{SC}} := \{(v, \eta) \mid v = (\underline{v}, v_3) \in H_0^1(\mathcal{S})^2 \times H_0^1(\mathcal{S}), \eta \in H^1(\mathcal{S})^2\}. \quad (2.2.24)$$

Boundary conditions for  $\theta$  are implicitly imposed by the vanishing boundary integral term in the equilibrium equation (2.2.22) (in contrast to (2.1.9)).

Unless otherwise stated, the identification  $\mathcal{V} := \mathcal{V}_{\text{SC}}$  is given in further work.

#### Variational formulation

For the cylindrical test problems, the incoming force is applied directly to the midsurface in transverse direction. Moreover, the density  $\rho$  is taken constant (homogeneous material) and the force density is taken per unit area. The left-hand-side of

(2.2.22) is thus replaced by

$$\int_{\mathcal{S}} f v_3 \, dS.$$

Now, the variational problem can be stated in the steady-state case: Let a force density  $f \in L^2(\mathcal{S})$  be given. Then, find  $(u, \theta) \in \mathcal{V}$  such that

$$\int_{\mathcal{B}(\tau)} \sigma(u, \theta) : e(v, \psi) \, dV = \int_{\mathcal{S}} f v_3 \, dS$$

for all  $(v, \psi) \in \mathcal{V}$ . Let  $\varepsilon$  be the second order tensor with components  $\varepsilon_{\alpha\beta} := e_{\alpha\beta}$ ,  $\alpha, \beta = 1, 2$  and  $\zeta$  be the tensor given in (2.2.13). With equation (2.2.19) the left-hand-side becomes

$$\int_{\mathcal{B}(\tau)} \sigma(u, \theta) : e(v, \psi) \, dV = \int_{\mathcal{B}(\tau)} (C\varepsilon(u, \theta)) : \varepsilon(v, \psi) + (D\zeta(u, \theta)) : \zeta(v, \psi) \, dV.$$

This model, consisting of the variational problem: Find  $(u, \theta) \in \mathcal{V}$  such that

$$\int_{\mathcal{B}(\tau)} (C\varepsilon(u, \theta)) : \varepsilon(v, \psi) + (D\zeta(u, \theta)) : \zeta(v, \psi) \, dV = \int_{\mathcal{S}} f v_3 \, dS \quad \forall (v, \psi) \in \mathcal{V}, \quad (2.2.25)$$

is introduced as the *basic shell model* in [CB03, p. 87f] and is the underlying model in further work.

Using Voigt notation and the representations as in (2.2.15), (2.2.21) and

$$\varepsilon(u, \theta) := \begin{pmatrix} e_{11}(u, \theta) \\ e_{22}(u, \theta) \\ \sqrt{2}e_{12}(u, \theta) \end{pmatrix},$$

the integrand can be expressed as matrix-vector multiplication in the convenient form

$$\int_{\mathcal{B}(\tau)} \sigma(u, \theta) : e(v, \psi) \, dV = \int_{\mathcal{B}(\tau)} \varepsilon^T(u, \theta) \mathbf{C} \varepsilon(v, \psi) + \zeta(u, \theta)^T \mathbf{D} \zeta(v, \psi) \, dV,$$

which is very useful for later numerical calculations.

### 2.2.5. Comparison with other shell models

The basic shell model introduced above is a quite general model in linear elasticity theory of shells. The key simplifications in comparison to the general three dimensional model (2.1.10) for isotropic homogeneous materials include

1. Kinematical ansatz (2.2.11),
2. Vanishing normal stress assumption (2.2.18).

Now, further simplifications can be made and lead to well-known shell models investigated in detail. The following survey is aligned to the books by Chapelle and Bathe ([CB03]) and Ciarlet ([Cia00]) and includes additional approaches by Nestler ([Nes10a]) and Lelley ([Lel12]).

**Naghdi model** In the *Naghdi* shell model [Nag63; Nag72] the components of the elasticity tensor are replaced by their restrictions to the midsurface, i.e. they are assumed to be constant along the “thin” dimension of the shell. The same is done with  $g$  appearing in  $dV$ . Moreover, the part of the strain tensor given by  $\kappa$  (see (2.2.14)) is neglected. Then the integration in (2.2.25) with respect to the thickness direction can be carried out explicitly, leading to the simplified left-hand-side ([CB03, p.89])

$$\int_S \tau(C|_{\xi^3=0} \gamma(u)) : \gamma(v) + \frac{\tau^3}{12} (C|_{\xi^3=0} \chi(u, \theta)) : \chi(v, \psi) + \tau(D|_{\xi^3=0} \zeta(u, \theta)) : \zeta(v, \psi) dS$$

with the tensors  $\gamma$ ,  $\chi$  and  $\zeta$  as in (2.2.12). Note that this model still includes membrane, bending and shear strain. A formulation as a classical boundary value problem can be found in [Cia00, p. 365].

The corresponding plate model for a midsurface with zero (mean and Gaussian) curvature is known as the *Reissner-Mindlin* model. In this case, the Reissner-Mindlin model and the basic shell model are equivalent ([CB03, p.94]).

**Koiter model** The *Koiter* shell model [Koi66] goes one step further: A stronger kinematical assumption is made, such that the shear strain  $\zeta$  from (2.2.13) vanishes. This introduces a relation between the translational and the rotational part of the displacement. This additional assumption is called *Kirchhoff-Love* assumption and the geometrical interpretation says, that normals to the midsurface remain straight and *normal to the midsurface* during deformation. Inserting the relation between  $u$  and  $\theta$  into the bending strain tensor  $\chi$  leads to  $\tilde{\chi}(u) := -\chi(u, \theta(u))$  with

$$\tilde{\chi}_{\alpha\beta}(u) = u_{3|\alpha\beta} + \sum_{\mu=1}^2 b_{\alpha|\beta}^{\mu} u_{\mu} + \sum_{\mu=1}^2 b_{\alpha}^{\mu} u_{\mu|\beta} + \sum_{\mu=1}^2 b_{\beta}^{\mu} u_{\mu|\alpha} - c_{\alpha\beta} u_3$$

and yields the left-hand-side

$$\int_S \tau(C|_{\xi^3=0} \gamma(u)) : \gamma(v) + \frac{\tau^3}{12} (C|_{\xi^3=0} \tilde{\chi}(u)) : \tilde{\chi}(v) dS \quad (2.2.26)$$

and the variational problem only depends on the translational displacements  $u$  and  $v$ . A boundary value problem formulation can be found in [Cia00, p. 338].

The corresponding plate model is known as the *Kirchhoff-Love* model. It is suggested to use the Koiter model only in case of very *thin* shells, while the Naghdi model is also suited for moderate thick shells ([Cia00, p. 362]).

**Model from Tartu University** The model for cylindrical shells communicated by Jaan Lellep from the Mechanic Department of the University of Tartu (Estonia) ([Lel12]) was the first one the author got in contact with. It can also be found in a book by Vinson ([Vin89, p. 121]). This model contains the application of elasticity theory for shells with a cylindrical midsurface. Starting from general three

dimensional elasticity, the Kirchhoff-Love displacement ansatz is employed together with the vanishing normal stress assumption. Unlike in the Naghdi and Koiter model explained above, where only the components of the elasticity tensor and the metric tensor were assumed to be constant through the thickness, here also the basis vectors  $g_i$ ,  $i = 1, 2, 3$  are approximated to 0-th order, i.e.  $g_i = a_i$ , which is actually more related to plate models. In particular, this affects the covariant differentiation while deriving the strain tensor. This leads to slightly different equations in comparison to the Koiter model (which apart from the 0-th order approximation of the basis vectors consists of the same assumptions). To be precise, the following terms are different concerning the individual parts of the strain tensor (2.2.12):

- In  $\chi_{\alpha\beta}$  the part

$$\frac{1}{2} \left( - \sum_{\lambda=1}^2 b_{\beta}^{\lambda} u_{\lambda|\alpha} - \sum_{\lambda=1}^2 b_{\alpha}^{\lambda} u_{\lambda|\beta} \right) + c_{\alpha\beta} u_3$$

is missing and

- the tensor  $\kappa$  is missing completely.

While the second point has no influence because  $\kappa$  is neglected in the Koiter model anyway, the first point influences  $\tilde{\chi}$ , which is in the Tartu model given by

$$\tilde{\chi}_{\alpha\beta}(u) = u_{3|\alpha\beta} + \sum_{\mu=1}^2 b_{\alpha|\beta}^{\mu} u_{\mu} + \frac{1}{2} \left( \sum_{\mu=1}^2 b_{\alpha}^{\mu} u_{\mu|\beta} + \sum_{\mu=1}^2 b_{\beta}^{\mu} u_{\mu|\alpha} \right).$$

The missing part containing  $c$  as well as the half weighted part with  $b$  carries information about the curvature of the midsurface, which is an important property of the shell. Hence, the difference between the Koiter and Tartu model can become relevant in certain cases. In the particular application to a cylindrical midsurface, this reveals in the influence on  $\tilde{\chi}_{12}$  and  $\tilde{\chi}_{22}$ ; compare

$$\begin{aligned} \tilde{\chi}_{12}(u) &= \partial_{x\varphi} u_3 - \frac{1}{R} \partial_x u_2 \\ \tilde{\chi}_{22}(u) &= \partial_{\varphi\varphi} u_3 - \frac{2}{R} \partial_{\varphi} u_2 - u_3 \end{aligned}$$

from the Koiter model to

$$\begin{aligned} \tilde{\chi}_{12}(u) &= \partial_{x\varphi} u_3 - \frac{1}{2R} \partial_x u_2 \\ \tilde{\chi}_{22}(u) &= \partial_{\varphi\varphi} u_3 - \frac{1}{R} \partial_{\varphi} u_2 \end{aligned}$$

from the Tartu model. The comparison of results for these two different models is not part of this work because the basic shell model (without such kind of simplifications) was chosen anyway.

**Further models** The Koiter model can be simplified even further: Depending on the geometry of the midsurface and the boundary conditions, for a very small thickness in (2.2.26) either the part linear in  $\tau$  or the part cubic in  $\tau$  becomes dominant. Hence the respective other part is neglected in the integral. This is known as the *membrane* model (in case the part consisting of the membrane strain  $\gamma$  becomes dominant) or the *flexural* model (in case the part with the bending strain  $\chi$  dominates).

Another option, this time not simplifying the model but making it more complex, is a higher order approximation in the kinematic ansatz (2.2.11). In the basic shell model, only a first order expansion is present. Higher order expansion are needed when large deformations transverse to the midsurface are expected [CB03, p. 97].

When considering a related problem about cylindrical shells, Nestler ([Nes10a; Nes12]) combined the Kirchhoff-Love assumption with an expansion of the integrand in (2.2.25) up to the second order and then carried out the integration in thickness direction explicitly.

To summarize, the underlying model is settled between well-known and extensively investigated classical shell models and full three dimensional elasticity models. The only two assumptions made, namely the Reissner-Mindlin kinematic ansatz together with the vanishing normal stress assumption, allow for a displacement space which consists of functions defined on the midsurface and for a decomposition of the elasticity tensor which considerably simplifies the numerical calculations. On the other hand, generality is kept by not simplifying the models further like in Naghdi or Koiter models. Moreover, the basic shell model has indeed been investigated in detail in [CB03] with respect to mathematical analysis of the equation itself, but not much has been done in relation to *thickness optimization*, which is the goal of this work. In addition, a dynamic (time-dependent) formulation of (2.2.25) has not been introduced nor investigated in [CB03] and little can be found in other literature. Hence, the basic shell model is a well suited object for investigation in this work.

### 2.2.6. Smoothness of midsurface and thickness

Up to this point, nothing has been said about the regularity assumptions to be put on the midsurface parametrization  $z$  and the thickness  $\tau$ . The derivatives appearing in the fundamental forms from Definition 2.10 shall be considered as weak derivatives, hence in the strain tensor  $e$  (2.2.12) as well as in the three-dimensional basis vectors  $g_i, g^i$  (2.2.4) and in the determinant of the metric tensor  $g$  (2.2.7) second weak derivatives of  $z$  appear. So for the integral in the variational equation (2.2.25) to be well-defined the assumption  $z \in W^{2,\infty}(\omega)$  in conjunction with the displacement space  $\mathcal{V}$  is needed. From now on, this assumption shall hold if not mentioned otherwise. Note that this implicates  $z \in C^1(\bar{\omega})$  by the Sobolev embedding theorem (see Theorem A.3), and because of the injectivity of  $z : \omega \rightarrow \mathcal{S}$  the parametrization is even a diffeomorphism. Further, it is remarked that this regularity is also sufficient for the Naghdi model. In contrast, the Koiter model involves third weak derivatives, hence  $z \in W^{3,\infty}(\omega)$  is needed. However, for both the Koiter and the Naghdi model reformulations exist, where only  $z \in W^{2,\infty}(\omega)$  ([BL99; BL01]) or even  $z \in W^{1,\infty}(\omega)$

([Tam14]) is needed. These are known as models for shells with little regularity.

In the following paragraph  $\tau \circ z$  and  $\tau$  are distinguished explicitly to avoid confusion. The discussion of the thickness regularity has two facets: For the integral in (2.2.25) to be well-defined it is enough to have  $\tau \circ z \in L^\infty(\omega)$ . This assumption is also made in [Nes12]. Nevertheless, the derivation of the model started from full three dimensional elasticity and included the application of the divergence theorem in (2.1.7) and Green's formula for second order tensors in (2.1.9) on the shell body  $\mathcal{B}_{(\tau)}$ , which has to have a Lipschitz boundary then, see Theorem A.2. The upper boundary of  $\mathcal{B}_{(\tau)}$  is given by

$$\partial\mathcal{B}_{(\tau),\text{upper}} = \left\{ z(\xi^1, \xi^2) + \frac{\tau \circ z(\xi^1, \xi^2)}{2} \mid (\xi^1, \xi^2) \in \omega \right\},$$

hence  $\tau \circ z$  has to be Lipschitz-continuous itself. The condition  $\tau \circ z \in C^{0,1}(\omega)$  is equivalent to  $\tau \circ z \in W^{1,\infty}(\omega)$ , as long as  $\omega$  is (quasi-)convex (see [Hei05, Thm. 4.1] or [EG15, Thm. 4.5] for a local version). So to stay in the context of Sobolev spaces, the latter regularity assumptions are chosen for  $\tau \circ z$  and  $\omega$  and apply in further text if not stated otherwise. This is also equivalent to  $\tau \in W^{1,\infty}(\mathcal{S})$ . However, it is interesting to note that the results from Section 3.1 and Chapter 4 stay valid also for  $\tau \in L^\infty(\mathcal{S})$ . Note that the application to the cylindrical case indeed allows for a convex  $\omega$  (see (2.2.8)).

### 2.2.7. Dynamic equations

In Subsection 2.2.4 steady-state equilibrium conditions were introduced. Another interesting subject is to view the deformation as a time-dependent process. Moreover, this enables the model to account for time-dependent loads as well as an initial momentum. The time-dependent basic shell model is derived from scratch in this work because it could not be found in available literature.

In contrast to the steady-state equations, the displacement  $U$  is equipped with an additional time parameter, i.e.  $U = U(\xi^1, \xi^2, \xi^3; t)$ . The same is done with the translational and rotational parts evolving from the kinematic ansatz (2.2.11),  $u = u(\xi^1, \xi^2; t)$  and  $\theta = \theta(\xi^1, \xi^2; t)$ . The starting point will be the equation of motion from general three dimensional elasticity (2.1.8b),

$$-\operatorname{div} \sigma(U(t)) + \rho \ddot{U}(t) = \rho b(t) \text{ in } \mathcal{B}_{(\tau)} \text{ for all } t \in [0, T].$$

First, a weak formulation is derived. Let  $t \in [0, T]$  be fixed, multiply by a (time-independent) displacement  $V$  from a suitable test space and integrate over  $\mathcal{B}_{(\tau)}$ . Then the equation reads

$$\underbrace{\int_{\mathcal{B}_{(\tau)}} \rho \ddot{U}(t) \cdot V \, dV}_{(a)} - \underbrace{\int_{\mathcal{B}_{(\tau)}} \operatorname{div} \sigma(U(t)) \cdot V \, dV}_{(b)} = \underbrace{\int_{\mathcal{B}_{(\tau)}} \rho b(t) \cdot V \, dV}_{(c)}.$$

Now, inserting the (time-dependent) displacement ansatz (2.2.11),

$$U(\xi^1, \xi^2, \xi^3; t) = u(\xi^1, \xi^2; t) + \xi^3 \theta(\xi^1, \xi^2; t),$$

applying the vanishing normal stress assumption and considering a transverse force acting on the midsurface, the parts (b) and (c) can be transformed like in the steady-state case and result in the variational equation of the basic shell model (2.2.25),

$$\int_{\mathcal{B}(\tau)} (C\varepsilon(u(t), \theta(t))) : \varepsilon(v, \psi) + (D\zeta(u(t), \theta(t))) : \zeta(v, \psi) dV = \int_{\mathcal{S}} f(t)v_3 dS.$$

The term labeled (a) is of greater interest: Inserting the displacement ansatz (2.2.11) leads to

$$\int_{\mathcal{B}(\tau)} \rho \partial_t^2 (u(t) + \xi^3 \theta(t)) \cdot (v + \xi^3 \psi) dV$$

where  $\partial_t^2$  is written for the second time-derivative to enhance readability. Let the density  $\rho$  be constant through the body from now on. Expanding the scalar product and carrying out the integration with respect to  $\xi^3$  gives

$$\begin{aligned} \int_{\mathcal{B}(\tau)} \rho \ddot{U}(t) \cdot V dV &= \rho \int_{\Omega(\tau)} \partial_t^2 (u(t) + \xi^3 \theta(t)) \cdot (v + \xi^3 \psi) \sqrt{g} d\xi^3 d\xi^2 d\xi^1 \\ &= \int_{\omega} \rho \int_{-\tau/2}^{\tau/2} \left[ \left( \ddot{u}(t) + \xi^3 \ddot{\theta}(t) \right) \cdot v + \left( \ddot{u}(t) + \xi^3 \ddot{\theta}(t) \right) \cdot (\xi^3 \psi) \right] \\ &\quad \sqrt{a} (1 - 2\xi^3 H + K(\xi^3)^2) d\xi^3 d\xi^2 d\xi^1 \\ &= \int_{\omega} \left[ \left( \Phi_1(\tau) \ddot{u}(t) + \Phi_2(\tau) \ddot{\theta}(t) \right) \cdot v + \left( \Phi_2(\tau) \ddot{u}(t) + \Phi_3(\tau) \ddot{\theta}(t) \right) \cdot \psi \right] \sqrt{a} d\xi^2 d\xi^1 \\ &= \int_{\mathcal{S}} \left[ \left( \Phi_1(\tau) \ddot{u}(t) + \Phi_2(\tau) \ddot{\theta}(t) \right) \cdot v + \left( \Phi_2(\tau) \ddot{u}(t) + \Phi_3(\tau) \ddot{\theta}(t) \right) \cdot \psi \right] dS, \end{aligned}$$

where the operators  $\Phi_i : L^\infty(\mathcal{S}) \rightarrow L^\infty(\mathcal{S})$  are defined as follows

$$\Phi_1(\tau) = \rho\tau + \rho K \frac{\tau^3}{12}, \quad \Phi_2(\tau) = -2\rho H \frac{\tau^3}{12}, \quad \Phi_3(\tau) = \rho \frac{\tau^3}{12} + \rho K \frac{\tau^5}{80}.$$

The actual state equation will be stated and discussed later (see Subsection 4.1.1, equation (4.1.5)). The displacements and their derivatives will be understood as *vector-valued functions* and the equation will be viewed in a setting of a *Gelfand triple* (or *rigged Hilbert space*). For now, it shall be enough to state the resulting equation without deeper discussion of the proper function spaces and initial values: Let a time-dependent force density  $f$  be given. Then, find  $(u, \theta)$  such that

$$\begin{aligned} \int_{\mathcal{S}} \left[ \left( \Phi_1(\tau) \ddot{u}(t) + \Phi_2(\tau) \ddot{\theta}(t) \right) \cdot v + \left( \Phi_2(\tau) \ddot{u}(t) + \Phi_3(\tau) \ddot{\theta}(t) \right) \cdot \psi \right] dS + \\ \int_{\mathcal{B}(\tau)} (C\varepsilon(u(t), \theta(t))) : \varepsilon(v, \psi) + (D\zeta(u(t), \theta(t))) : \zeta(v, \psi) dV = \int_{\mathcal{S}} f(t)v_3 dS \end{aligned} \quad (2.2.27)$$

for all  $(v, \psi) \in \mathcal{V}$ ,  $t \in [0, T]$ .

### 2.2.8. Summary

In this section the steady-state as well as the dynamic equations to describe the deformation of a shell under load were introduced. The derivation of the steady-state equation was oriented on the book by Chapelle and Bathe ([CB03]), where the *basic shell model* was chosen from. This model unites the advantage of having the unknown displacements as functions defined on the midsurface only, while staying as close as possible to three dimensional linear elasticity. In fact, only the Reissner-Mindlin displacement ansatz (2.2.11) and the vanishing normal stress assumption were used to simplify the full three dimensional equation (2.1.9) from Section 2.1. Then the model equations were applied to the specific cylindrical case.

The basic shell model was comparatively rated to other well-known shell models (Naghdi, Koiter) as well as to the model from Tartu University.

In the following the corresponding dynamic equation was derived starting from three dimensional linear elasticity, only using the Reissner-Mindlin displacement ansatz and the vanishing normal stress assumption. To summarize, both state equations are stated once again at this point:

**Steady-state equation** Let a force density  $f \in L^2(\mathcal{S})$  be given. Then, find  $(u, \theta) \in \mathcal{V}$  such that

$$\int_{\mathcal{B}(\tau)} (C\varepsilon(u, \theta)) : \varepsilon(v, \psi) + (D\zeta(u, \theta)) : \zeta(v, \psi) dV = \int_{\mathcal{S}} f v_3 dS \quad \text{f.a. } (v, \psi) \in \mathcal{V}, \quad (2.2.28)$$

where  $\mathcal{V} := \{(v, \eta) \mid v = (v_1, v_2, v_3) \in H_0^1(\mathcal{S})^3 \times H_0^1(\mathcal{S}), \eta \in H^1(\mathcal{S})\}$  (see (2.2.25)).

**Dynamic equation** Let a (time-dependent) force density  $f$  be given. Then, find  $(u, \theta)$  such that

$$\int_{\mathcal{S}} \left[ \left( \Phi_1(\tau)\ddot{u}(t) + \Phi_2(\tau)\ddot{\theta}(t) \right) \cdot v + \left( \Phi_2(\tau)\ddot{u}(t) + \Phi_3(\tau)\ddot{\theta}(t) \right) \cdot \psi \right] dS + \int_{\mathcal{B}(\tau)} (C\varepsilon(u(t), \theta(t))) : \varepsilon(v, \psi) + (D\zeta(u(t), \theta(t))) : \zeta(v, \psi) dV = \int_{\mathcal{S}} f(t)v_3 dS \quad (2.2.29)$$

for all  $(v, \psi) \in \mathcal{V}$ ,  $t \in [0, T]$  (see (2.2.27)), where

$$\Phi_1(\tau) = \rho\tau + \rho K \frac{\tau^3}{12}, \quad \Phi_2(\tau) = -2\rho H \frac{\tau^3}{12}, \quad \Phi_3(\tau) = \rho \frac{\tau^3}{12} + \rho K \frac{\tau^5}{80}$$

and the proper displacement space is discussed in Subsection 4.1.1.

## 2.3. Optimization problem

After discussing and choosing a proper shell model, now the optimization problem is introduced. At this point only the steady-state problem is stated, because the dynamic state equation is fully discussed later and both problems share the same structure.

The object of investigation is a loaded shell. The aim is to minimize the corresponding displacement (according to the shell model (2.2.28)). This minimization can be done subject to various objectives:

- Square-norm of the displacement  $\frac{1}{2} \|(u, \theta)\|^2$  (where a suitable norm must be chosen, e.g.  $\|\cdot\|_{\mathcal{V}}$ , or  $L^2(\mathcal{S})$ -norms)
- Compliance, where the displacement is weighted with the incoming force, and the objective matches the right-hand-side of the state equations, i.e.  $\int_{\mathcal{S}} u_3 f \, dS$
- $u_3$ -component of the displacement integrated over the midsurface, i.e.  $\int_{\mathcal{S}} u_3 \, dS$ , as done in [Nes10a]
- Maximum of the  $u_3$ -component (if it exists), as proposed by Lellep [Lel12]

This itemization is referenced as “Itemization 2.3” in the following. Proposal four suffers from missing differentiability and needs additional regularity assumptions to be meaningful. The third proposal bears the danger of changing signs, and misses differentiability if the absolute value is used instead. The first list item is a commonly used objective in optimal control problems, in which the goal is to minimize  $\|y - y_d\|^2$ , where  $y$  solves the state equation and  $y_d$  is a reference state [Trö10]. However, the second entry in this list is mainly used in shape optimization, where this work is closest to. Hence, this objective is used as the minimization subject in later examples. Moreover, an  $H^1(\mathcal{S})$ -regularization term for the thickness is added to the objective in order to smoothen the solution.

Some additional limitations have to be taken into account:

- The thickness is bounded from above and below, i.e.

$$\tau_{\min} \leq \tau(\xi^1, \xi^2) \leq \tau_{\max} \quad \text{in } \mathcal{S}.$$

- The volume of the shell body is prescribed, i.e.

$$\begin{aligned} \int_{\mathcal{B}(\tau)} dV &= \int_{\Omega(\tau)} \sqrt{a} (1 - 2H\xi^3 + K(\xi^3)^2) \, d\xi^3 \, d\xi^2 \, d\xi^1 \\ &= \int_{\omega} \left( \tau + \frac{K\tau^3}{12} \right) \sqrt{a} \, d\xi^2 \, d\xi^1 = \int_{\mathcal{S}} \left( \tau + \frac{K\tau^3}{12} \right) \, dS = C. \end{aligned} \quad (2.3.1)$$

Now, the steady-state optimization problem can be stated:

$$\min_{\substack{\tau \in W^{1,\infty}(\mathcal{S}) \\ (u,\theta) \in \mathcal{V}}} J(u, \theta; \tau) := \int_{\mathcal{S}} u_3 f \, dS + \frac{\lambda}{2} \|\tau\|_{H^1(\mathcal{S})}^2 \quad (2.3.2a)$$

$$\text{s.t. : } \int_{\mathcal{B}(\tau)} \sigma(u, \theta) : e(v, \psi) \, dV = \int_{\mathcal{S}} f v_3 \, dS \quad \text{f.a. } (v, \psi) \in \mathcal{V} \quad (2.3.2b)$$

$$\tau_{\min} \leq \tau(\xi^1, \xi^2) \leq \tau_{\max} \text{ in } \mathcal{S} \quad (2.3.2c)$$

$$\int_{\mathcal{S}} \left( \tau + \frac{K\tau^3}{12} \right) \, dS = C. \quad (2.3.2d)$$

Another interesting problem formulation was given by Lellep [Lel12]: In a certain dual way to the above stated problem, the objective is to minimize the volume of the shell given by (2.3.1) according to the restriction of a prescribed maximal displacement (see Itemization 2.3, point 4). This introduces *state constraints* into the problem formulation which need careful investigation.

**Cylindrical case** The actual application to the cylindrical case does not change much of the above stated formulation (2.3.2). Only the volume condition can be reformulated, because the cylindrical midsurface entails  $K = 0$ . Hence, (2.3.2d) can be simplified to

$$\int_{\mathcal{S}} \tau \, dS = \int_{\omega} \tau R \, d\varphi \, dx = C. \quad (2.3.3)$$

## 3. Steady-state problem

This chapter is devoted to the analytical discussion of the steady-state optimization problem. The overall aim is to derive necessary conditions for an optimal solution of problem (2.3.2). First, the state equation given through the basic shell model (2.2.25) is investigated in relation to existence and uniqueness of a solution and, even more important, regarding continuity and differentiability with respect to the thickness (Theorem 3.3 and 3.4). Followed by this, the results are applied to the optimization problem, and necessary conditions for an optimal solution in case of a parabolic midsurface are derived (Theorem 3.6 and 3.7). Some results for this can be found in the book by Neittaanmäki, Sprekels, and Tiba ([NST06]) for a generalized Naghdi model. The line of action in the following sections has been inspired by the *control into parameter* approach from chapter six of the above book. The content of this chapter is a new contribution to the field of thickness optimization in linear elasticity and has been published in [Zie13] and [Zie14].

### 3.1. Analysis of the state equation

As a reminder, the steady-state equation derived in the previous chapter is stated again here. It is a variational problem: Let a force density  $f \in L^2(\mathcal{S})$  be given. Then, find  $(u, \theta) \in \mathcal{V}$  such that

$$\int_{\mathcal{B}(\tau)} (C\varepsilon(u, \theta)) : \varepsilon(v, \psi) + (D\zeta(u, \theta)) : \zeta(v, \psi) \, dV = \int_{\mathcal{S}} f v_3 \, dS \quad \text{f.a. } (v, \psi) \in \mathcal{V}, \quad (3.1.1)$$

where  $\mathcal{S}$  is the midsurface of the shell,  $\tau$  is its thickness,  $\mathcal{B}(\tau)$  is the shell body and  $C$  and  $D$  are the parts of the elasticity tensor due to the vanishing normal stress assumption. The displacement space  $\mathcal{V}$  contains pairs of translational and rotational displacements arising from the Reissner-Mindlin kinematical ansatz and  $\varepsilon$  and  $\zeta$  represent the parts of the strain tensor. The colon denotes the inner product in the space of second order tensors.

#### 3.1.1. Existence and uniqueness of solutions

First, the existence and uniqueness of a solution to the steady-state equation is discussed. To shorten the notation, for the left-hand-side of (3.1.1)

$$\begin{aligned} a_{(\tau)}(u, \theta; v, \psi) &:= \int_{\mathcal{B}(\tau)} (C\varepsilon(u, \theta)) : \varepsilon(v, \psi) + (D\zeta(u, \theta)) : \zeta(v, \psi) \, dV \\ &= \int_{\mathcal{B}(\tau)} \sigma(u, \theta) : e(v, \psi) \, dV \end{aligned} \quad (3.1.2)$$

is defined where the subscript  $(\tau)$  in  $a_{(\tau)}$  emphasizes the dependence on the thickness  $\tau$ . The representation in the first line will be used for the proof of continuity in Subsection 3.1.2 while the latter form will be used for the investigation of the differentiability in Subsection 3.1.3 (to shorten the notation). For the right-hand-side

$$F(v, \psi) := \int_{\mathcal{S}} f v_3 \, dS$$

is defined. Clearly,  $a_{(\tau)}$  is a bilinear mapping  $\mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$  and  $F$  is a continuous linear mapping  $\mathcal{V} \rightarrow \mathbb{R}$ . In [CB03], the mapping  $a_{(\tau)}$  has been investigated in detail, and it holds:

**Theorem 3.1** ([CB03, p. 111]). The bilinear mapping  $a_{(\tau)} : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$  is continuous, i.e.

$$a_{(\tau)}(u, \theta; v, \psi) \leq C_{(\tau)} \|(u, \theta)\|_{\mathcal{V}} \|(v, \psi)\|_{\mathcal{V}}.$$

Moreover, if the boundary conditions imposed on  $\mathcal{V}$  from (2.2.23) are such that no rigid body motion is possible, then  $a_{(\tau)}$  is also coercive, i.e.

$$a_{(\tau)}(u, \theta; u, \theta) > c_{(\tau)} \|(u, \theta)\|_{\mathcal{V}}^2.$$

*Proof.* See [CB03, pp. 109–113]. □

Note that both the hardclamped and the softclamped conditions (over the whole boundary) assure that no rigid body motion is possible. Hence, there is a unique solution to the state equation:

**Theorem 3.2** ([CB03, pp. 113f.]). Let  $f \in L^2(\mathcal{S})$  be given and softclamped boundary conditions imposed on  $\mathcal{V}$ . There is a unique solution to the variational problem: Find  $(u, \theta)$  in  $\mathcal{V}$  such that

$$a_{(\tau)}(u, \theta; v, \psi) = F(v, \psi) \quad \text{f.a. } (v, \psi) \in \mathcal{V},$$

and for the solution it holds

$$\|(u, \theta)\|_{\mathcal{V}} \leq C_{(\tau)} \|f\|_{L^2(\mathcal{S})}.$$

*Proof.* By use of the Lax-Milgram lemma and the above theorem. □

The last theorem gives rise to the definition of the *control-to-state operator*  $G$ , which maps the thickness to the corresponding solution of (3.1.1). Remember the overall restriction on  $\tau$  from (2.2.5),  $\tau(\xi^1, \xi^2) < 2 |R_{\min}(\xi^1, \xi^2)|$ , which must be imposed on the domain of  $G$ . However, the constraint (2.3.2c),

$$\tau_{\min} \leq \tau(\xi^1, \xi^2) \leq \tau_{\max} \quad \text{in } \mathcal{S},$$

is an even stronger requirement. So, let  $\tau_{\min} > 0$  and  $\tau_{\max} < 2 |R_{\min}(\xi^1, \xi^2)|$ . Note the implicit assumption that  $|R_{\min}|$  does not get arbitrarily close to zero. This is

a reasonable assumption in practice, in particular for the cylindrical case where  $R_{\min} = \text{const.}$

A suitable domain for  $G$  would therefore be the set

$$U_b := \{\tau \in W^{1,\infty}(\mathcal{S}) \mid 0 < \tau_{\min} \leq \tau(\xi^1, \xi^2) \leq \tau_{\max} < 2 |R_{\min}(\xi^1, \xi^2)| \text{ in } \mathcal{S}\}. \quad (3.1.3)$$

**Definition 3.1** (Control-to-state operator). The operator  $G : U_b \rightarrow \mathcal{V}$  which maps the thickness  $\tau$  to the solution  $(u, \theta)$  of (3.1.1) is called *control-to-state operator*.

The properties of  $G$  are of special interest. In particular, continuity and (Gâteaux-) differentiability of the operator will be investigated. First, the continuity and coercivity theorem from above is generalized.

**Lemma 3.1.** There exist common constants  $c, C > 0$  so that

$$\begin{aligned} a_{(\tau)}(u, \theta; u, \theta) &\geq c \|(u, \theta)\|_{\mathcal{V}}^2, \\ a_{(\tau)}(u, \theta; v, \psi) &\leq C \|(u, \theta)\|_{\mathcal{V}} \|(v, \psi)\|_{\mathcal{V}} \end{aligned}$$

holds for all  $(u, \theta), (v, \psi) \in \mathcal{V}$  and all  $\tau \in U_b$  with  $U_b$  from (3.1.3).

*Proof.* The proof of Theorem 3.1 (for fixed  $\tau$ ) from [CB03] relies essentially on the estimations

$$\hat{c}_{(\tau)} \sqrt{a}(\xi^1, \xi^2) \leq \sqrt{g}(\xi^1, \xi^2, \xi^3) \leq \hat{C}_{(\tau)} \sqrt{a}(\xi^1, \xi^2) \quad \forall (\xi^1, \xi^2, \xi^3) \in \Omega_{(\tau)}$$

and

$$\begin{aligned} \hat{c}_{(\tau)} \sum_{\alpha, \beta=1}^2 a^{\alpha\beta} X_{\alpha} X_{\beta} &\leq \sum_{\alpha, \beta=1}^2 g^{\alpha\beta} X_{\alpha} X_{\beta} \leq \hat{C}_{(\tau)} \sum_{\alpha, \beta=1}^2 a^{\alpha\beta} X_{\alpha} X_{\beta} \\ &\quad \forall (\xi^1, \xi^2, \xi^3) \in \Omega_{(\tau)}, (X_1, X_2) \in \mathbb{R}^2. \end{aligned}$$

Those are the only places (beside using the boundedness of  $\tau$  from above and below) where a dependence of  $c_{(\tau)}$  and  $C_{(\tau)}$  from Theorem 3.1 on the thickness is introduced. The proof of both statements is based on the compactness of  $\bar{\Omega}_{(\tau)}$  and the continuous dependence of  $g$  and  $g^{\alpha\beta}$  from  $\xi^3$ . Indeed, the arguments stay valid if  $\Omega_{(\tau)}$  is replaced by

$$\Omega_{(\tau_{\max})} := \left\{ (\xi^1, \xi^2, \xi^3) \in \mathbb{R}^3 \mid (\xi^1, \xi^2) \in \omega, \xi^3 \in \left( -\frac{\tau_{\max}}{2}, \frac{\tau_{\max}}{2} \right) \right\},$$

which has also compact closure. The constants  $c$  and  $C$  obtained by using  $\Omega_{(\tau_{\max})}$  for the above statement (independent of  $\tau$ ) are also suitable for all  $\tau \in U_b$ , since  $\Omega_{(\tau)} \subseteq \Omega_{(\tau_{\max})}$ .  $\square$

This result also generalizes the dependence of the displacement on the applied load:

**Corollary 3.1.** For the solution  $(u, \theta)$  from (3.1.1) it holds

$$\|(u, \theta)\|_{\mathcal{V}} \leq C \|f\|_{L^2(\mathcal{S})} \quad \forall \tau \in U_b$$

with a constant  $C > 0$  independent of  $\tau$ .

### 3.1.2. Continuous dependence on the thickness

In this subsection the continuity of the control-to-state operator is discussed. In other words, the continuous dependence of the displacement on the thickness of the shell is investigated. Some results for this can be found in [CB03] for the basic shell model and a given thickness profile. Here, the restriction to a thickness profile is not needed.

The sketch of the subsection will be as follows: First, given a converging thickness sequence, show that the corresponding sequence of solutions has a weakly convergent subsequence. Second, prove that the weak limit is solution to the limit thickness. Third, verify that the convergence is also strong.

Let a sequence  $\tau_n \in U_b$  with  $\tau_n \rightarrow \bar{\tau} \in U_b$  and the set  $U_b$  as in (3.1.3) be given. Denote the sequence of the corresponding solutions to (3.1.1) by  $(u_n, \theta_n) := G(\tau_n)$ .

**Lemma 3.2.** The sequence  $(u_n, \theta_n)$  is bounded.

*Proof.* From the above Corollary 3.1 and for a fixed load  $f \in L^2(\mathcal{S})$ , it holds

$$\|(u_n, \theta_n)\|_{\mathcal{V}} \leq C \|f\|_{L^2(\mathcal{S})} =: \hat{C}$$

for all  $n$ . □

Hence, by use of Theorem A.4 there is a weakly convergent subsequence

$$(u_{n_k}, \theta_{n_k}) \rightharpoonup (\bar{u}, \bar{\theta}). \quad (3.1.4)$$

This limit is investigated further. For this purpose, a proposition is stated in front of the next lemma.

**Proposition 3.1.** The mappings

$$(u, \theta) \mapsto C\varepsilon(u, \theta) \quad \text{and} \quad (u, \theta) \mapsto D\zeta(u, \theta)$$

are continuous linear mappings  $\mathcal{V} \rightarrow L^2(\mathcal{B}_{(\tau_{\max})})$ .

*Proof.* The linearity is clear from the definition of  $\varepsilon_{\alpha\beta} = e_{\alpha\beta}$ ,  $\alpha, \beta = 1, 2$  (see (2.2.12)) and  $\zeta$  (see (2.2.13)). Let a sequence  $(u_n, \theta_n) \rightarrow (\bar{u}, \bar{\theta})$  in  $\mathcal{V}$  be given. Since the displacement space  $\mathcal{V}$  is a composition of  $H^1(\mathcal{S})$ -spaces (see (2.2.23)), all components of  $u$  and  $\theta$  together with their derivatives converge in  $L^2(\mathcal{S})$  to the corresponding parts of  $\bar{u}$  and  $\bar{\theta}$ .

Because all components of  $\varepsilon$  and  $\zeta$  only contain  $u_n, \theta_n$  and their first derivatives multiplied by bounded functions, it follows

$$\varepsilon(u_n, \theta_n) \rightarrow \varepsilon(\bar{u}, \bar{\theta}) \quad \text{and} \quad \zeta(u_n, \theta_n) \rightarrow \zeta(\bar{u}, \bar{\theta}),$$

both in  $L^2(\mathcal{B}_{(\tau_{\max})})$ . The application of the fourth-order tensors  $C$  and  $D$  does not destroy the continuity since their components are all bounded functions, so

$$C\varepsilon(u_n, \theta_n) \rightarrow C\varepsilon(\bar{u}, \bar{\theta}) \quad \text{and} \quad D\zeta(u_n, \theta_n) \rightarrow D\zeta(\bar{u}, \bar{\theta})$$

in  $L^2(\mathcal{B}_{(\tau_{\max})})$ . □

**Lemma 3.3.** The weak limit  $(\bar{u}, \bar{\theta})$  from (3.1.4) is the solution of the steady-state equation with thickness  $\bar{\tau}$ , i.e.

$$a_{(\bar{\tau})}(\bar{u}, \bar{\theta}; v, \psi) = F(v, \psi) \quad \forall (v, \psi) \in \mathcal{V}.$$

*Proof.* First, for  $\tau \in U_b$  rewrite  $a_{(\tau)}$  in the form

$$a_{(\tau)}(u, \theta; v, \psi) := \int_{\mathcal{B}_{(\tau_{\max})}} ((C\varepsilon(u, \theta)) : \varepsilon(v, \psi) + (D\zeta(u, \theta)) : \zeta(v, \psi)) \chi_{(\tau)} \, dV,$$

where  $\chi_{(\tau)} \in L^\infty(\mathcal{B}_{(\tau_{\max})})$  is given through

$$\chi_{(\tau)}(\xi^1, \xi^2, \xi^3) := \begin{cases} 1, & \text{if } -\frac{\tau(\xi^1, \xi^2)}{2} \leq \xi^3 \leq \frac{\tau(\xi^1, \xi^2)}{2}, \\ 0, & \text{else.} \end{cases}$$

In terms of the  $L^2(\mathcal{B}_{(\tau_{\max})})$  inner product this reads

$$a_{(\tau)}(u, \theta; v, \psi) = (C\varepsilon(u, \theta), \varepsilon(v, \psi)\chi_{(\tau)})_{L^2(\mathcal{B}_{(\tau_{\max})})} + (D\zeta(u, \theta), \zeta(v, \psi)\chi_{(\tau)})_{L^2(\mathcal{B}_{(\tau_{\max})})}$$

where the multiplication of a second-order tensor field with a scalar-valued function is interpreted in the natural component-wise form.

Again considering the sequence  $\tau_n$ , it holds

$$\chi_{(\tau_n)} \rightarrow \chi_{(\bar{\tau})} \quad \text{pointwise,}$$

since for fixed  $(\xi^1, \xi^2, \xi^3)$  it is

$$|\chi_{(\tau_n)}(\xi^1, \xi^2, \xi^3) - \chi_{(\bar{\tau})}(\xi^1, \xi^2, \xi^3)| = \begin{cases} 1, & \text{if } |\xi^3| \text{ between } \frac{\tau_n(\xi^1, \xi^2)}{2} \text{ and } \frac{\bar{\tau}(\xi^1, \xi^2)}{2}, \\ 0, & \text{else,} \end{cases}$$

and for sufficient big  $k$  only the latter case can occur because of  $\tau_n \rightarrow \bar{\tau}$ .

From this it follows (for fixed  $(v, \psi) \in \mathcal{V}$ )

$$\varepsilon_{\alpha\beta}(v, \psi)\chi_{(\tau_n)} \rightarrow \varepsilon_{\alpha\beta}(v, \psi)\chi_{(\bar{\tau})} \quad \text{pointwise}$$

for  $\alpha, \beta = 1, 2$  and also

$$\varepsilon(v, \psi)\chi_{(\tau_n)} \rightarrow \varepsilon(v, \psi)\chi_{(\bar{\tau})} \quad \text{strong in } L^2(\mathcal{B}_{(\tau_{\max})})$$

due to the above pointwise convergence and the boundedness of  $\|\varepsilon(v, \psi)\|_{L^2(\mathcal{B}_{(\tau_{\max})})}$ , assuring that the components of  $\varepsilon(v, \psi)$  can be used as integrable majorants.

Now, consider the sequence of solutions  $(u_{n_k}, \theta_{n_k}) \rightharpoonup (\bar{u}, \bar{\theta})$ . The mapping

$$(u, \theta) \mapsto C\varepsilon(u, \theta)$$

is a continuous linear operator  $\mathcal{V} \rightarrow L^2(\mathcal{B}_{(\tau_{\max})})$  (see Proposition 3.1) and is therefore weakly sequentially continuous. So

$$C\varepsilon(u_{n_k}, \theta_{n_k}) \rightharpoonup C\varepsilon(\bar{u}, \bar{\theta}) \quad \text{in } L^2(\mathcal{B}_{(\tau_{\max})}).$$

Analogously, the convergence

$$\begin{aligned} \zeta(v, \psi)\chi_{(\tau_{n_k})} &\rightarrow \zeta(v, \psi)\chi_{(\bar{\tau})}, \\ D\zeta(u_{n_k}, \theta_{n_k}) &\rightarrow D\zeta(\bar{u}, \bar{\theta}), \end{aligned}$$

both in  $L^2(\mathcal{B}_{(\tau_{\max})})$  can be shown. From the weak and strong convergence of its arguments, it follows that also the inner product converges, i.e.

$$\begin{aligned} (C\varepsilon(u_{n_k}, \theta_{n_k}), \varepsilon(v, \psi)\chi_{(\tau_{n_k})}) &\rightarrow (C\varepsilon(\bar{u}, \bar{\theta}), \varepsilon(v, \psi)\chi_{(\bar{\tau})}), \\ (D\zeta(u_{n_k}, \theta_{n_k}), \zeta(v, \psi)\chi_{(\tau_{n_k})}) &\rightarrow (D\zeta(\bar{u}, \bar{\theta}), \zeta(v, \psi)\chi_{(\bar{\tau})}). \end{aligned}$$

Hence it holds for all  $(v, \psi) \in \mathcal{V}$

$$F(v, \psi) = a_{(\tau_{n_k})}(u_{n_k}, \theta_{n_k}; v, \psi) \rightarrow a_{(\bar{\tau})}(\bar{u}, \bar{\theta}; v, \psi)$$

and  $(\bar{u}, \bar{\theta})$  is the solution to  $\bar{\tau}$ . □

**Remark 3.1.** *Because of the uniqueness of the state equation solution, every convergent subsequence of  $(u_n, \theta_n)$  converges weakly to the above found  $(\bar{u}, \bar{\theta})$ . So the whole sequence converges weakly in  $\mathcal{V}$ , i.e.*

$$(u_n, \theta_n) \rightharpoonup (\bar{u}, \bar{\theta}) \quad \text{in } \mathcal{V}.$$

Now that the first two steps are done – finding a weakly convergent subsequence and showing that the weak limit is the solution to the limit thickness – the last step proves the strong convergence of the sequence.

**Lemma 3.4.** The convergence established in Lemma 3.3 and Remark 3.1 is also strong, i.e.

$$\|(u_n, \theta_n) - (\bar{u}, \bar{\theta})\|_{\mathcal{V}} \rightarrow 0.$$

*Proof.* For every  $(v, \psi) \in \mathcal{V}$  it holds

$$\begin{aligned} a_{(\tau_{n_k})}(u_{n_k}, \theta_{n_k}; v, \psi) &= F(v, \psi), \\ a_{(\bar{\tau})}(\bar{u}, \bar{\theta}; v, \psi) &= F(v, \psi), \end{aligned}$$

and by subtraction of the equations and insertion of  $a_{(\tau_{n_k})}(\bar{u}, \bar{\theta}; v, \psi)$  it follows

$$0 = a_{(\tau_{n_k})}(u_{n_k} - \bar{u}, \theta_{n_k} - \bar{\theta}; v, \psi) + a_{(\tau_{n_k})}(\bar{u}, \bar{\theta}; v, \psi) - a_{(\bar{\tau})}(\bar{u}, \bar{\theta}; v, \psi). \quad (3.1.5)$$

The last two terms are of interest now. By inserting  $(v, \psi) = (u_n - \bar{u}, \theta_n - \bar{\theta})$  they get

$$a_{(\tau_{n_k})}(\bar{u}, \bar{\theta}; u_n - \bar{u}, \theta_n - \bar{\theta}) - a_{(\bar{\tau})}(\bar{u}, \bar{\theta}; u_n - \bar{u}, \theta_n - \bar{\theta})$$

which (also using the symmetry of  $a_{(\tau)}$ ) translates to

$$\int_{\mathcal{B}_{(\tau_{\max})}} ((C\varepsilon(u_n - \bar{u}, \theta_n - \bar{\theta})) : \varepsilon(\bar{u}, \bar{\theta}) + (D\zeta(u_n - \bar{u}, \theta_n - \bar{\theta})) : \zeta(\bar{u}, \bar{\theta})) (\chi_{(\tau_n)} - \chi_{(\bar{\tau})}) \, dV.$$

By taking the limit, the same situation as in Lemma 3.3 occurs:

$$\begin{aligned} C\varepsilon(u_n - \bar{u}, \theta_n - \bar{\theta}) &\rightharpoonup 0 \text{ and } D\zeta(u_n - \bar{u}, \theta_n - \bar{\theta}) \rightharpoonup 0, \\ \varepsilon(\bar{u}, \bar{\theta})(\chi_{(\tau_n)} - \chi_{(\bar{\tau})}) &\rightarrow 0 \text{ and } \zeta(\bar{u}, \bar{\theta})(\chi_{(\tau_n)} - \chi_{(\bar{\tau})}) \rightarrow 0, \end{aligned}$$

all in  $L^2(\mathcal{B}_{(\tau_{\max})})$ . Therefore also the inner product converges to 0 and

$$a_{(\tau_n)}(\bar{u}, \bar{\theta}; u_n - \bar{u}, \theta_n - \bar{\theta}) - a_{(\bar{\tau})}(\bar{u}, \bar{\theta}; u_n - \bar{u}, \theta_n - \bar{\theta}) \rightarrow 0.$$

Looking again at (3.1.5), it follows

$$0 = \lim_{n \rightarrow \infty} a_{(\tau_n)}(u_n - \bar{u}, \theta_n - \bar{\theta}; u_n - \bar{u}, \theta_n - \bar{\theta}).$$

Hence, by using the (uniform) coercivity of  $a_{(\tau)}$

$$0 \geq \lim_{n \rightarrow \infty} c \|(u_n - \bar{u}, \theta_n - \bar{\theta})\|_{\mathcal{V}}^2,$$

which is the desired statement.  $\square$

The above lemmas are merged to the actual continuity theorem:

**Theorem 3.3.** The control-to-state operator  $G : U_b \rightarrow \mathcal{V}$  is continuous.

*Proof.* From the above lemmas it follows for a sequence of thicknesses  $\tau_n \rightarrow \bar{\tau}$  in  $U_b$  that for the corresponding solutions  $G(\tau_n) = (u_n, \theta_n)$  and  $G(\bar{\tau}) = (\bar{u}, \bar{\theta})$

$$(u_n, \theta_n) \rightarrow (\bar{u}, \bar{\theta})$$

strong in  $\mathcal{V}$ .  $\square$

### 3.1.3. Differentiable dependence on the thickness

After clarifying the continuity of the control-to-state operator, now the differentiability comes in place. In view of the optimization problem, this is even more important, since necessary conditions as well as numerical methods rely on a first derivative. In infinite-dimensional spaces a whole series of derivation concepts exists; the Gâteaux-derivative is suited for this work, its definition is recapitulated first.

**Definition 3.2** (Gâteaux-differential, [Trö10, p. 56]). Let  $X, Y$  be Banach-spaces,  $U \subseteq X$  open,  $U \neq \emptyset$  and  $f : U \rightarrow Y$ . If the limit

$$\delta f(x, h) := \lim_{\lambda \rightarrow 0} \frac{f(x + \lambda h) - f(x)}{\lambda}$$

exists in  $Y$ , then  $\delta f(x, h)$  is called *directional derivative* of  $f$  in direction  $h$ . If the mapping

$$h \mapsto \delta f(x, h)$$

is linear and continuous, i.e.  $f'(x) := \delta f(x, \cdot) \in L(X, Y)$ , then  $f'(x)$  is called the *Gâteaux-derivative* of  $f$  at  $x$ .

Note that some authors do not demand the linearity and the continuity of the Gâteaux-derivative (e.g. Berger, [Ber77, p. 68]).

The sketch of this subsection is to first investigate the directional derivative, in particular to show the existence of the limit, and then to clarify the linearity and the continuity of the derivative. Remember the domain of  $G$  was the *closed* set  $U_b$ , but as  $\tau_{\min}$  and  $\tau_{\max}$  could be chosen freely in the interval  $(0, 2|R_{\min}|)$ , there is always an *open* subset of  $W^{1,\infty}(\mathcal{S})$  containing  $U_b$ , where all the results established above stay valid.

Start by considering a fixed thickness  $\tau \in U_b$  and a direction  $q \in W^{1,\infty}(\mathcal{S})$ . For the solutions  $(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}) := G(\tau + \lambda q)$  and  $(u_{(\tau)}, \theta_{(\tau)}) := G(\tau)$  it holds

$$\begin{aligned} a_{(\tau+\lambda q)}(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}; v, \psi) &= F(v, \psi), \\ a_{(\tau)}(u_{(\tau)}, \theta_{(\tau)}; v, \psi) &= F(v, \psi) \end{aligned}$$

for all  $(v, \psi) \in \mathcal{V}$ , and therefore by subtraction of the above equations, division by  $\lambda$  and by exploiting the linearity of  $a_{(\tau)}$  it follows

$$\begin{aligned} 0 = \int_{\mathcal{B}(\tau_{\max})} \left( \sigma \left( \frac{(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}) - (u_{(\tau)}, \theta_{(\tau)})}{\lambda} : e(v, \psi) \right) \chi_{(\tau)} \right) dV + \quad (3.1.6) \\ \int_{\mathcal{B}(\tau_{\max})} (\sigma(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}) : e(v, \psi)) \left( \frac{\chi_{(\tau+\lambda q)} - \chi_{(\tau)}}{\lambda} \right) dV. \end{aligned}$$

For the last summand from (3.1.6), the mapping  $Z_\lambda : \mathcal{V} \times \mathcal{V} \times W^{1,\infty}(\mathcal{S}) \rightarrow \mathbb{R}$  is defined as

$$Z_\lambda(u, \theta; v, \psi; q) := \int_{\mathcal{B}(\tau_{\max})} (\sigma(u, \theta) : e(v, \psi)) \left( \frac{\chi_{(\tau+\lambda q)} - \chi_{(\tau)}}{\lambda} \right) dV, \quad (3.1.7)$$

so the above equation (3.1.6) reads in short

$$a_{(\tau)} \left( \frac{(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}) - (u_{(\tau)}, \theta_{(\tau)})}{\lambda}; v, \psi \right) = -Z_\lambda(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}; v, \psi; q) \quad (3.1.8)$$

for all  $(v, \psi) \in \mathcal{V}$ .

For the mapping  $Z_\lambda$ , the limit case  $\lambda \rightarrow 0$  is of interest.

**Lemma 3.5.** The pointwise limit

$$\lim_{\lambda \rightarrow 0} Z_\lambda(u, \theta; v, \psi; q)$$

exists.

*Proof.* It holds that  $Z_\lambda(\cdot; \cdot; q) : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$  is bounded for fixed  $\lambda$ , because

$$\begin{aligned} & |Z_\lambda(u, \theta; v, \psi; q)| \\ & \leq \frac{1}{\lambda} \int_{\mathcal{B}(\tau_{\max})} |\sigma(u, \theta) : e(v, \psi)| |\chi(\tau + \lambda q) - \chi(\tau)| \, dV \\ & \leq \frac{1}{\lambda} \int_{\mathcal{B}(\tau_{\max})} |\sigma(u, \theta) : e(v, \psi)| \, dV \\ & \leq C \frac{1}{\lambda} \|(u, \theta)\|_{\mathcal{V}} \|(v, \psi)\|_{\mathcal{V}}. \end{aligned}$$

The last inequality is due to the boundedness of  $a_{(\tau_{\max})}$ . Moreover,  $Z_\lambda$  is linear in both arguments  $(u, \theta)$  and  $(v, \psi)$ . So, given a sequence  $\lambda_n \rightarrow 0$ ,  $Z_{\lambda_n}(\cdot; \cdot; q)$  can be considered as a sequence in  $\mathcal{V}^* \times \mathcal{V}^*$ .

Next, the pointwise limit

$$\lim_{\lambda \rightarrow 0} Z_\lambda(u, \theta; v, \psi; q)$$

(for fixed  $(u, \theta)$ ,  $(v, \psi)$  and  $q$ ) is determined. First, switch to local coordinates and rewrite (3.1.7) as

$$Z_\lambda(u, \theta; v, \psi; q) := \int_{\omega} \int_{-\frac{\tau_{\max}}{2}}^{\frac{\tau_{\max}}{2}} (\sigma(u, \theta) : e(v, \psi)) \left( \frac{\chi(\tau + \lambda q) - \chi(\tau)}{\lambda} \right) \sqrt{a} (1 - 2H\xi^3 + K(\xi^3)^2) \, d\xi^3 \, d\xi^2 \, d\xi^1.$$

Note that the innermost integral is bounded by an  $L^1(\mathcal{S})$ -function because of the continuous dependence on the  $\xi^3$ -coordinate, more precise

$$\begin{aligned} & \left| \int_{-\frac{\tau_{\max}}{2}}^{\frac{\tau_{\max}}{2}} (\sigma(u, \theta) : e(v, \psi)) \left( \frac{\chi(\tau + \lambda q) - \chi(\tau)}{\lambda} \right) \sqrt{a} (1 - 2H\xi^3 + K(\xi^3)^2) \, d\xi^3 \right| \\ & \leq b(\xi^1, \xi^2) \int_{-\tau_{\max}/2}^{\tau_{\max}/2} \left| \frac{\chi(\tau + \lambda q) - \chi(\tau)}{\lambda} \right| \, d\xi^3 \\ & \leq b(\xi^1, \xi^2) |q(\xi^1, \xi^2)|, \end{aligned}$$

where  $b(\xi^1, \xi^2)$  is chosen as the maximum

$$b(\xi^1, \xi^2) := \max_{\xi^3 \in [-\tau_{\max}/2, \tau_{\max}/2]} |(\sigma(u, \theta) : e(v, \psi)) \sqrt{a} (1 - 2H\xi^3 + K(\xi^3)^2)|.$$

Furthermore, note

$$\|b\|_{L^1(\mathcal{S})} \leq C \|(u, \theta)\|_{\mathcal{V}} \|(v, \psi)\|_{\mathcal{V}}, \quad (3.1.9)$$

because of the continuity of  $\sigma : e$  (see Proposition 3.1) and the boundedness of all additional factors.

So the limit can be put past the midsurface-integral. If, in addition, the notation is returned from the characteristic function  $\chi(\tau)$  to variable integral bounds, the

limit reads

$$\begin{aligned} & \lim_{\lambda \rightarrow 0} Z_\lambda(u, \theta; v, \psi; q) \\ &= \int_{\mathcal{S}} \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \left( \int_{-\frac{\tau+\lambda q}{2}}^{\frac{\tau+\lambda q}{2}} (\sigma(u, \theta) : e(v, \psi)) (1 - 2H\xi^3 + K(\xi^3)^2) d\xi^3 dS \right. \\ & \quad \left. \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} (\sigma(u, \theta) : e(v, \psi)) (1 - 2H\xi^3 + K(\xi^3)^2) d\xi^3 dS \right). \end{aligned}$$

The integrand is continuous with respect to  $\xi^3$ , so this limit reduces to the derivative of the integral bounds with respect to  $\lambda$  at  $\lambda = 0$ . According to Leibniz's rule ([RHB06, p.178]), this limit *exists* and evaluates to

$$\begin{aligned} & \lim_{\lambda \rightarrow 0} Z_\lambda(u, \theta; v, \psi; q) \\ &= \int_{\mathcal{S}} \sum_{\xi^3 \in \{\pm \frac{\tau}{2}\}} [(\sigma(u, \theta) : e(v, \psi)) (1 - 2H\xi^3 + K(\xi^3)^2)] \Big|_{\xi^3 = \xi^3} \frac{q}{2} dS \quad (3.1.10) \\ &=: Z_0(u, \theta; v, \psi; q). \end{aligned}$$

Moreover, the newly defined  $Z_0$  keeps the linearity and continuity properties for the arguments  $(u, \theta)$  and  $(v, \psi)$ .  $\square$

Up to now, a pointwise limit of  $Z_\lambda(\cdot; \cdot; q)$  is established, which corresponds to the weak-\* convergence in  $\mathcal{V}^* \times \mathcal{V}^*$ . The next step is to insert the solutions  $(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)})$  into the first argument of  $Z_\lambda$  and consider the limit case  $\lambda \rightarrow 0$  again.

**Lemma 3.6.** The limit

$$\lim_{\lambda \rightarrow 0} Z_\lambda(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}; \cdot; q) = Z_0(u_{(\tau)}, \theta_{(\tau)}; \cdot; q) =: Z_q(\cdot)$$

exists pointwise, i.e. with respect to the weak-\* topology in  $\mathcal{V}^*$ .

*Proof.* In the last lemma the pointwise convergence for  $Z_\lambda(\cdot; \cdot; q)$  in  $\mathcal{V}^* \times \mathcal{V}^*$  was established. So, by the Banach-Steinhaus theorem (uniform boundedness theorem [Ber77, p. 34]), the mappings are uniformly bounded, i.e.

$$|Z_\lambda(u, \theta; v, \psi; q)| \leq C \| (u, \theta) \|_{\mathcal{V}} \| (v, \psi) \|_{\mathcal{V}}$$

for a constant  $C$  independent of  $\lambda$ . For fixed  $(v, \psi) \in \mathcal{V}$  it holds

$$\begin{aligned} & |Z_\lambda(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}; v, \psi; q) - Z_0(u_{(\tau)}, \theta_{(\tau)}; v, \psi; q)| \\ & \leq |Z_\lambda(u_{(\tau+\lambda q)} - u_{(\tau)}, \theta_{(\tau+\lambda q)} - \theta_{(\tau)}; v, \psi; q)| + |(Z_\lambda - Z_0)(u_{(\tau)}, \theta_{(\tau)}; v, \psi; q)| \\ & \leq C \| (u_{(\tau+\lambda q)} - u_{(\tau)}, \theta_{(\tau+\lambda q)} - \theta_{(\tau)}) \|_{\mathcal{V}} \| (v, \psi) \|_{\mathcal{V}} + |(Z_\lambda - Z_0)(u_{(\tau)}, \theta_{(\tau)}; v, \psi; q)| \\ & \xrightarrow[\lambda \rightarrow 0]{} 0, \end{aligned}$$

where the first term tends to zero because of the continuity of  $G$  (Theorem 3.3) and the second term tends to zero because of the pointwise convergence of  $Z_\lambda$  (Lemma 3.5). Hence it holds

$$\begin{aligned} & \lim_{\lambda \rightarrow 0} Z_\lambda(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}; v, \psi; q) \\ &= Z_0(u_{(\tau)}, \theta_{(\tau)}; v, \psi) \\ &= \int_{\mathcal{S}} \sum_{\xi^3 \in \{\pm \frac{\tau}{2}\}} [(\sigma(u_{(\tau)}, \theta_{(\tau)}) : e(v, \psi)) (1 - 2H\xi^3 + K(\xi^3)^2)] \Big|_{\xi^3 = \xi^3} \frac{q}{2} dS \\ &=: Z_q(v, \psi) \end{aligned}$$

and  $Z_\lambda(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}; \cdot; q)$  converges pointwise to  $Z_q \in \mathcal{V}^*$ .  $\square$

Now, the quotient

$$\frac{(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}) - (u_{(\tau)}, \theta_{(\tau)})}{\lambda}$$

from Definition 3.2 of the directional derivative is ready to be investigated.

**Lemma 3.7.** For fixed thickness  $\tau \in U_b$  and direction  $q \in W^{1,\infty}(\mathcal{S})$ , the norm of

$$\frac{G(\tau + \lambda q) - G(\tau)}{\lambda} = \frac{(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}) - (u_{(\tau)}, \theta_{(\tau)})}{\lambda}$$

is bounded for  $\lambda \rightarrow 0$ .

*Proof.* Inserting

$$(v, \psi) = \frac{(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}) - (u_{(\tau)}, \theta_{(\tau)})}{\lambda}$$

into equation (3.1.8) and using the uniform coercivity of  $a_{(\tau)}$  and the uniform boundedness of  $Z_\lambda$  leads to

$$\begin{aligned} & \left\| \frac{(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}) - (u_{(\tau)}, \theta_{(\tau)})}{\lambda} \right\|_{\mathcal{V}}^2 \\ & \leq ca_{(\tau)} \left( \frac{(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}) - (u_{(\tau)}, \theta_{(\tau)})}{\lambda}; \frac{(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}) - (u_{(\tau)}, \theta_{(\tau)})}{\lambda} \right) \\ & = -Z_\lambda \left( (u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}) - (u_{(\tau)}, \theta_{(\tau)}); \frac{(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}) - (u_{(\tau)}, \theta_{(\tau)})}{\lambda}; q \right) \\ & \leq C \left\| (u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}) - (u_{(\tau)}, \theta_{(\tau)}) \right\|_{\mathcal{V}} \left\| \frac{(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}) - (u_{(\tau)}, \theta_{(\tau)})}{\lambda} \right\|_{\mathcal{V}}, \end{aligned}$$

and division by  $\left\| \frac{(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}) - (u_{(\tau)}, \theta_{(\tau)})}{\lambda} \right\|_{\mathcal{V}}$  shows

$$\left\| \frac{(u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}) - (u_{(\tau)}, \theta_{(\tau)})}{\lambda} \right\|_{\mathcal{V}} \leq C \left\| (u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}) - (u_{(\tau)}, \theta_{(\tau)}) \right\|_{\mathcal{V}},$$

so the expression is bounded because of the continuity of  $G$ .  $\square$

Since the norm of solutions is bounded for any sequence  $\lambda_n \rightarrow 0$ , by use of Theorem A.4 there is a weakly convergent subsequence  $\lambda_{n_k}$  with

$$\frac{(u_{(\tau+\lambda_{n_k}q)}, \theta_{(\tau+\lambda_{n_k}q)}) - (u_{(\tau)}, \theta_{(\tau)})}{\lambda_{n_k}} \rightharpoonup (w, \nu) \in \mathcal{V}.$$

Now go back to equation (3.1.8). Passing to the limit on both sides leads to

$$\begin{aligned} \lim_{\lambda_{n_k} \rightarrow 0} a_{(\tau)} \left( \frac{(u_{(\tau+\lambda_{n_k}q)}, \theta_{(\tau+\lambda_{n_k}q)}) - (u_{(\tau)}, \theta_{(\tau)})}{\lambda_{n_k}}; v, \psi \right) = \\ - \lim_{\lambda_{n_k} \rightarrow 0} Z_{\lambda_{n_k}} (u_{(\tau+\lambda_{n_k}q)}, \theta_{(\tau+\lambda_{n_k}q)}; v, \psi; q), \end{aligned}$$

and using the continuity of  $a_{(\tau)}$  and the result from Lemma 3.6 yields

$$a_{(\tau)}(w, \nu; v, \psi) = -Z_q(v, \psi) \quad (3.1.11)$$

for arbitrary  $(v, \psi) \in \mathcal{V}$ . This is a variational problem to be solved for  $(w, \nu)$ . Since the right-hand-side  $Z_q$  is in  $\mathcal{V}^*$ , there is a unique solution to this problem and hence the whole sequence

$$\frac{G(\tau + \lambda_n q) - G(\tau)}{\lambda_n} \rightharpoonup (w, \nu) \quad (3.1.12)$$

converges weakly in  $\mathcal{V}^*$  for  $\lambda_n \rightarrow 0$ .

The next thing to tackle is the question if the strong convergence also holds in the above equation (3.1.12).

**Lemma 3.8.** It holds

$$\frac{G(\tau + \lambda q) - G(\tau)}{\lambda} \rightarrow (w, \nu)$$

strong in  $\mathcal{V}$  with  $(w, \nu)$  given from (3.1.11).

*Proof.* It is

$$\begin{aligned} & \left\| \frac{G(\tau + \lambda q) - G(\tau)}{\lambda} - (w, \nu) \right\|_{\mathcal{V}}^2 \\ & \leq ca_{(\tau)} \left( \frac{G(\tau + \lambda q) - G(\tau)}{\lambda} - (w, \nu); \frac{G(\tau + \lambda q) - G(\tau)}{\lambda} - (w, \nu) \right) \\ & = -cZ_{\lambda} \left( G(\tau + \lambda q) - G(\tau); \frac{G(\tau + \lambda q) - G(\tau)}{\lambda} - (w, \nu); q \right) + \\ & \quad cZ_q \left( \frac{G(\tau + \lambda q) - G(\tau)}{\lambda} - (w, \nu) \right) \\ & \leq C \|q\|_{L^\infty(\mathcal{S})} \|G(\tau + \lambda q) - G(\tau)\|_{\mathcal{V}} \left\| \frac{G(\tau + \lambda q) - G(\tau)}{\lambda} - (w, \nu) \right\|_{\mathcal{V}}, \end{aligned}$$

where the first “ $\leq$ ” is due to the coercivity of  $a_{(\tau)}$ , the following “ $=$ ” is from (3.1.8) and (3.1.11), and the last “ $\leq$ ” follows from an estimation in the style of (3.1.9). Therefore

$$\left\| \frac{G(\tau + \lambda q) - G(\tau)}{\lambda} - (w, \nu) \right\|_{\mathcal{V}} \leq C \|q\|_{L^\infty(\mathcal{S})} \|G(\tau + \lambda q) - G(\tau)\|_{\mathcal{V}} \xrightarrow{\lambda \rightarrow 0} 0,$$

since  $q$  is fixed and  $G$  is continuous.  $\square$

Now everything is prepared to state the essential theorem of this subsection:

**Theorem 3.4.** The control-to-state operator  $G : U_b \rightarrow \mathcal{V}$  is Gâteaux-differentiable. For a thickness  $\tau \in U_b$  and direction  $q \in W^{1,\infty}(\mathcal{S})$ , it holds

$$G'(\tau)q = (w, \nu),$$

where  $(w, \nu) \in \mathcal{V}$  is the solution to the variational problem: Find  $(w, \nu) \in \mathcal{V}$  such that

$$a_{(\tau)}(w, \nu; v, \psi) = -Z_q(v, \psi) \quad \text{for all } (v, \psi) \in \mathcal{V}, \quad (3.1.13)$$

with

$$Z_q(v, \psi) = \int_{\mathcal{S}} \sum_{\xi^3 \in \{\pm \frac{\tau}{2}\}} [(\sigma(u_{(\tau)}, \theta_{(\tau)}) : e(v, \psi)) (1 - 2H\xi^3 + K(\xi^3)^2)] \Big|_{\xi^3 = \xi^3} \frac{q}{2} dS$$

and  $(u_{(\tau)}, \theta_{(\tau)}) = G(\tau)$ .

*Proof.* The existence of the (strong) limit

$$\lim_{\lambda \rightarrow 0} \frac{G(\tau + \lambda q) - G(\tau)}{\lambda}$$

in  $\mathcal{V}$  has been shown in the above Lemmas 3.5 through 3.8. Since the problem (3.1.13) is linear and  $q$  appears linear in the right-hand-side  $Z_q$ , the derivative  $(w, \nu)$  depends linear on  $q$ . Moreover,  $\|Z_q\|_{\mathcal{V}^*}$  is bounded by  $C \|q\|_{W^{1,\infty}(\mathcal{S})}$  (actually even by  $C \|q\|_{L^\infty(\mathcal{S})}$ ), so the same is valid for  $\|(w, \nu)\|_{\mathcal{V}}$  by the Lax-Milgram-Lemma, hence  $G'(\tau) : W^{1,\infty}(\mathcal{S}) \rightarrow \mathcal{V}$  is a continuous linear operator – the sought Gâteaux-derivative.  $\square$

Note the importance of the above theorem, which is necessary for the differentiation of the reduced objective in Theorem 3.6.

## 3.2. Analysis of the optimization problem

This section addresses the analysis of the steady-state optimization problem (2.3.2). The continuity and differentiability results from the previous section can be applied to this problem and lead to related conclusions for the reduced objective. Finally, necessary conditions for an optimal solution can be stated.

As a reminder, the steady-state problem (2.3.2) is stated again at this point:

$$\min_{\substack{\tau \in W^{1,\infty}(\mathcal{S}) \\ (u,\theta) \in \mathcal{V}}} J(u, \theta; \tau) := \int_{\mathcal{S}} u_3 f \, dS + \frac{\lambda}{2} \|\tau\|_{H^1(\mathcal{S})}^2 \quad (3.2.1a)$$

$$\text{s.t. : } \int_{\mathcal{B}(\tau)} \sigma(u, \theta) : e(v, \psi) \, dV = \int_{\mathcal{S}} f v_3 \, dS \quad \text{f.a. } (v, \psi) \in \mathcal{V} \quad (3.2.1b)$$

$$\tau_{\min} \leq \tau(\xi^1, \xi^2) \leq \tau_{\max} \text{ in } \mathcal{S} \quad (3.2.1c)$$

$$\int_{\mathcal{S}} \left( \tau + \frac{K\tau^3}{12} \right) \, dS = C. \quad (3.2.1d)$$

First, the corresponding (state-)reduced problem is stated. All the restrictions (except the state equation) are collected in the set

$$U_{\text{ad}} := \left\{ \tau \in W^{1,\infty}(\mathcal{S}) \mid \tau_{\min} \leq \tau(\xi^1, \xi^2) \leq \tau_{\max}, \int_{\mathcal{S}} \left( \tau + \frac{K\tau^3}{12} \right) \, dS = C \right\}. \quad (3.2.2)$$

Since  $U_{\text{ad}} \subset U_{\text{b}}$ , the control-to-state operator is well-defined on  $U_{\text{ad}}$ , and the objective in conjunction with the state equation can be rewritten as

$$J(u, \theta; \tau) = J(G(\tau); \tau) =: J_s(\tau), \quad (3.2.3)$$

where  $J_s : U_{\text{ad}} \rightarrow \mathbb{R}$  only depends on the thickness  $\tau$ .

**Definition 3.3** (Reduced objective). The functional  $J_s : U_{\text{ad}} \rightarrow \mathbb{R}$  defined in (3.2.3) is called *reduced objective*.

Combining both equations (3.2.2) and (3.2.3), the *reduced problem* is stated as

$$\min_{\tau \in U_{\text{ad}}} J_s(\tau). \quad (3.2.4)$$

The first thing that can be established is the continuity of the reduced objective  $J_s$ .

**Lemma 3.9.** The reduced objective  $J_s : U_{\text{ad}} \rightarrow \mathbb{R}$  is continuous.

*Proof.* The original objective  $J$  is continuous, since

$$(u, \theta) \mapsto \int_{\mathcal{S}} u_3 f \, dS$$

is a continuous linear functional and  $W^{1,\infty}(\mathcal{S})$  is continuously embedded into  $H^1(\mathcal{S})$ . Moreover, the control-to-state operator is continuous, and so is the composition  $J_s = J \circ (G, \text{id})$ .  $\square$

The continuity of the reduced objective gives the possibility for an existence theorem for an optimal solution of (3.2.4), as long as  $U_{\text{ad}}$  is further restricted.

**Theorem 3.5** (cf. [HM03, p. 16]). Consider the set

$$U_{\text{ad}}^M := U_{\text{ad}} \cap \left\{ \tau \in W^{1,\infty}(\mathcal{S}) \left\| \left\| \frac{\partial \tau}{\partial \xi^1} \right\|_{L^\infty(\mathcal{S})} < M, \left\| \frac{\partial \tau}{\partial \xi^2} \right\|_{L^\infty(\mathcal{S})} < M \right\},$$

where  $M > 0$  is fixed. Then the problem

$$\min_{\tau \in U_{\text{ad}}^M} J_s(\tau)$$

has an optimal solution.

*Proof.* With help of the embedding Theorem A.3, the set  $U_{\text{ad}}$  belongs to  $C(\bar{\omega})$  and  $\bar{\omega} \subset \mathbb{R}^2$  is compact. Moreover, the functions in  $U_{\text{ad}}$  are equibounded by  $\tau_{\max}$ . With the additional restriction introduced in  $U_{\text{ad}}^M$  (the equiboundedness of the derivatives), the functions therein are even equicontinuous on  $\bar{\omega}$  (not to be confused with uniform continuous), i.e.

$$\left| \tau(\xi^1, \xi^2) - \tau(\tilde{\xi}^1, \tilde{\xi}^2) \right| \leq CM \left\| (\xi^1, \xi^2) - (\tilde{\xi}^1, \tilde{\xi}^2) \right\| \quad \forall \tau \in U_{\text{ad}}^M, \forall (\xi^1, \xi^2), (\tilde{\xi}^1, \tilde{\xi}^2) \in \omega.$$

From the Arzela-Ascoli theorem (see [Yos80, p. 85]) it is known that  $U_{\text{ad}}^M$  is relatively compact in  $C(\bar{\omega})$ . In addition, the set  $U_{\text{ad}}^M$  is closed in  $C(\bar{\omega})$ , so it is even a compact subset. Since  $C(\bar{\omega})$  is a normed space, it is also sequentially compact. For the closedness of  $U_{\text{ad}}^M$  see [Nes10b], where an analogous proof has been given.

The objective  $J_s$  is continuous on  $U_{\text{ad}}$  (see Lemma 3.9), so in particular on  $U_{\text{ad}}^M$ . Therefore by use of a generalization of the (Weierstraß) extreme value theorem (see [LS74, p. 152]),  $J_s$  attains its minimum on  $U_{\text{ad}}^M$ .  $\square$

**Remark 3.2.** *Certainly, the above Theorem 3.5 is valid for any continuous objective  $J$ , e.g. the norm minimization proposed in Itemization 2.3, point 2, or the minimization of the maximum displacement (under higher smoothness assumptions) suggested in point 4. A deeper discussion on less restricted admissible sets for the thickness can be found in [HM03, pp. 16–18].*

The next step is to consider the differentiability of the reduced objective and to derive necessary conditions for an optimal solution of (3.2.1) afterwards. While the necessary conditions open the possibility for indirect optimization methods, the expression of a directional derivative which does not rely on finite differences is invaluable for direct numerical optimization methods and sometimes, which is the case here, allows for a numerical solution at all. First, the objective is inspected regarding differentiability. For this, a stronger differentiability concept than the Gâteaux-derivative is needed.

**Definition 3.4** (Fréchet-derivative [Trö10, pp. 58f.]). Let  $X, Y$  be Banach-spaces,  $U \subseteq X$  open,  $U \neq \emptyset$  and  $f : U \rightarrow Y$ . Then  $f$  is called *Fréchet-differentiable* at  $x \in U$  if there exist an operator  $A \in L(X, Y)$  and a mapping  $r(x, \cdot) : U \rightarrow Y$  with

$$F(x + h) = F(x) + Ah + r(x, h)$$

and

$$\frac{\|r(x, h)\|_Y}{\|h\|_X} \rightarrow 0 \quad \text{as } \|h\|_X \rightarrow 0.$$

The operator  $A =: F'(x) =: D_x F(x)$  is called *Fréchet derivative*. If  $f$  is Fréchet-differentiable at every  $x \in U$ , then it is called *Fréchet-differentiable in  $U$* .

**Remark 3.3** (Fréchet and Gâteaux-differential [Trö10, p. 59]). *If a function  $f : U \subseteq X \rightarrow Y$  is Fréchet-differentiable at  $x \in U$ , then it is also Gâteaux-differentiable and both derivatives coincide, i.e.*

$$f'_G(x) = f'_F(x) \in L(X, Y).$$

**Lemma 3.10.** The objective  $J$  from (3.2.1a) is a Fréchet-differentiable mapping  $\mathcal{V} \times W^{1,\infty}(\mathcal{S}) \rightarrow \mathbb{R}$ . The derivative with respect to  $(u, \theta)$ , denoted by  $D_{(u,\theta)} J(u, \theta; \tau)$ , is given by

$$D_{(u,\theta)} J(u, \theta; \tau)(v, \psi) = \int_{\mathcal{S}} v_3 f \, dS,$$

and the derivative with respect to  $\tau$ , denoted by  $D_\tau J(u, \theta; \tau)$  is given by

$$D_\tau J(u, \theta; \tau)q = \lambda(\tau, q)_{H^1(\mathcal{S})}.$$

*Proof.* As a continuous linear functional  $\mathcal{V} \rightarrow \mathbb{R}$ , the first part

$$(u, \theta) \mapsto \int_{\mathcal{S}} u_3 f \, dS$$

is Fréchet-differentiable (see [Trö10, p. 59, ex. (vi)]), and the derivative is the operator itself, i.e.

$$D_{(u,\theta)} J(u, \theta; \tau)(v, \psi) = \int_{\mathcal{S}} v_3 f \, dS.$$

The second part containing the  $H^1(\mathcal{S})$ -norm is also Fréchet-differentiable (see [Trö10, p. 59, ex. (vii)]), and since  $H^1(\mathcal{S})$  is a Hilbert-space, the derivative with respect to  $\tau$  is given by

$$D_\tau J(u, \theta; \tau)q = \lambda(\tau, q)_{H^1(\mathcal{S})}.$$

□

**Remark 3.4.** *The above Lemma 3.10 could have simplified the proof of Lemma 3.9, since the continuity of the objective  $J$  follows from its Fréchet-differentiability [Jah07, p. 41].*

**Remark 3.5.** *The objectives proposed in Itemization 2.3, point 1 and 3 are also Fréchet-differentiable, since they consist of an  $H^1(\mathcal{S})$ -norm and linear operators, resp.*

Now that the existence of the Fréchet-derivative of the objective function is assured, an *adjoint state* can be defined as the solution to a related problem.

**Definition 3.5** (Adjoint state, [NST06, p. 434]). The adjoint state  $(p, \eta)$  to a thickness  $\tau$  with corresponding displacement  $(u_{(\tau)}, \theta_{(\tau)})$  is defined as the solution to: Find  $(p, \eta) \in \mathcal{V}$  such that

$$a_{(\tau)}(p, \eta; v, \psi) = D_{(u, \theta)}J(u_{(\tau)}, \theta_{(\tau)})(v, \psi) \quad (3.2.5)$$

for all  $(v, \psi) \in \mathcal{V}$ .

First, note that the above equation (3.2.5) indeed has a unique solution, because  $D_{(u, \theta)}J(u_{(\tau)}, \theta_{(\tau)})$  is a continuous linear operator by definition. Further, note that in case of the objective (3.2.1a), its derivative  $D_{(u, \theta)}J$  coincides with the original right-hand-side of the state equation, so that the adjoint state and the actual displacement are equal. However, the general notation was kept in order to allow for other objectives (like in Itemization 2.3).

**Theorem 3.6.** The reduced objective is Gâteaux-differentiable. For the directional derivative  $J'_s(\tau)q$  at point  $\tau$  with corresponding displacement  $(u_{(\tau)}, \theta_{(\tau)})$ , it holds

$$J'_s(\tau)q = -Z_q(u_{(\tau)}, \theta_{(\tau)}) + \lambda(\tau, q)_{H^1(\mathcal{S})}.$$

*Proof.* The proof is based on [NST06, p. 434]. Let  $(p, \eta)$  be the adjoint state defined in (3.2.5). By using the chain rule for Gâteaux- and Fréchet-derivatives (see [KA82, p. 499, example IV]), it is

$$\begin{aligned} J'_s(\tau)q &= D_{(u, \theta)}J(u_{(\tau)}, \theta_{(\tau)}; \tau)G'(\tau)q + D_\tau J(u_{(\tau)}, \theta_{(\tau)}; \tau)q \\ &= a_{(\tau)}(p, \eta; G'(\tau)q) + D_\tau J(u_{(\tau)}, \theta_{(\tau)}; \tau)q \\ &= a_{(\tau)}(G'(\tau)q; p, \eta) + D_\tau J(u_{(\tau)}, \theta_{(\tau)}; \tau)q \\ &= -Z_q(p, \eta) + D_\tau J(u_{(\tau)}, \theta_{(\tau)}; \tau)q. \end{aligned}$$

Here, the second “=” is by definition from the adjoint state, the third one uses the symmetry of  $a_{(\tau)}$ , and the fourth one is from the Gâteaux-derivative of  $G$  (see Theorem 3.4). Now using the fact that for  $J$  from (3.2.1a) it is  $(p, \eta) = (u_{(\tau)}, \theta_{(\tau)})$  (see the remark after Definition 3.5) and that from Lemma 3.10

$$D_\tau J(u, \theta; \tau)q = \lambda(\tau, q)_{H^1(\mathcal{S})},$$

the equation from the theorem is proven.  $\square$

**Remark 3.6.** As in the definition of the adjoint state, the general notation has been kept in the above proof to allow for other objectives. The norm minimization from Itemization 2.3, point 2 is Gâteaux-differentiable too, and the derivative reads

$$J'_s(\tau)q = -Z_q(p, \eta) + \lambda(\tau, q)_{H^1(\mathcal{S})},$$

where  $(p, \eta)$  is solution to: Find  $(p, \eta) \in \mathcal{V}$  such that

$$a_{(\tau)}(p, \eta; v, \psi) = ((u_{(\tau)}, \theta_{(\tau)}), (v, \psi))_{\mathcal{V}}$$

for all  $(v, \psi) \in \mathcal{V}$ .

Again, the importance of an analytical expression for the directional derivative of the reduced objective is emphasized here. The first things to collect from the above work are necessary conditions for an optimal solution of (3.2.1).

**Theorem 3.7.** Let  $\tau^* \in U_{\text{ad}}$  be a (locally) optimal solution for the problem (3.2.1) with corresponding state  $(u_{(\tau^*)}, \theta_{(\tau^*)})$ , and let  $K = 0$  everywhere on  $\mathcal{S}$ . Then it holds

$$J'_s(\tau^*)(q - \tau^*) = -Z_{(q-\tau^*)}(u_{(\tau^*)}, \theta_{(\tau^*)}) + \lambda(\tau^*, q - \tau^*)_{H^1(\mathcal{S})} \geq 0 \quad (3.2.6)$$

for all directions  $q \in U_{\text{ad}}$ .

*Proof.* According to [Trö10, p. 63], if  $U_{\text{ad}}$  is a convex subset of the real Banach-space  $W^{1,\infty}(\mathcal{S})$ , and  $J_s$  is Gâteaux-differentiable on an open subset of  $W^{1,\infty}(\mathcal{S})$  containing  $U_{\text{ad}}$ , then: If  $\tau^*$  is an optimal solution to the reduced problem (3.2.4), then it solves

$$J'_s(\tau^*)(q - \tau^*) \geq 0 \quad \forall q \in U_{\text{ad}}.$$

Since  $G$  is Gâteaux-differentiable on  $U_b$  (where  $\tau_{\min}$  and  $\tau_{\max}$  can be chosen arbitrarily from an open interval), there is an open set  $U$  with

$$U_{\text{ad}} \subset U \subset U_b,$$

where all the above statements concerning the differentiability of  $J_s$  stay valid. So the only thing left to show is the convexity of  $U_{\text{ad}}$ . Indeed,  $U_{\text{ad}}$  is convex, since

- $\tau_{\min} = \lambda\tau_{\min} + (1 - \lambda)\tau_{\min} \leq \lambda\tau_1 + (1 - \lambda)\tau_2 \leq \lambda\tau_{\max} + (1 - \lambda)\tau_{\max} = \tau_{\max}$
- $\int_{\mathcal{S}} \lambda\tau_1 + (1 - \lambda)\tau_2 \, dS = \lambda \int_{\mathcal{S}} \tau_1 \, dS + (1 - \lambda) \int_{\mathcal{S}} \tau_2 \, dS = C$

for all  $\tau_1, \tau_2 \in U_{\text{ad}}$  and  $\lambda \in [0, 1]$ . The actual expression for  $J'_s(\tau^*)(q - \tau^*)$  was derived in Theorem 3.6.  $\square$

This necessary condition gives rise to indirect numerical optimization methods. However, those will not be used in later work, instead the necessary conditions will be used to check, how “far away” the solution calculated by a direct method is from an optimal solution. This will be explained later in Subsection 5.3.3.

A final remark regarding the smoothness of the thickness shall close this section:

**Remark 3.7.** *All the above calculations could have also been done for a thickness  $\tau$  that only lies in  $L^\infty(\mathcal{S})$  (except the use of the  $H^1(\mathcal{S})$  in the objective). In view of the discussion from Subsection 2.2.6, the stronger assumption  $\tau \in W^{1,\infty}(\mathcal{S})$  was retained. However, the lower smoothness requirement of the “optimal control approach” in contrast to the more general shape derivative approach following in Section 3.4 is one of the major differences between the two methods, which is surely an advantage of the approach taken in this section.*

### 3.3. Summary

In this section, the steady-state equation (3.1.1) and the corresponding optimization problem (3.2.1) were investigated in detail. After stating the unique solvability of the state equation, the existence of a well-defined control-to-state operator had been assured. This operator was studied with respect to continuity and Gâteaux-differentiability, which is a key component of this work and a new contribution to optimization in linear elasticity. Next, the results were applied to the reduced objective, and analogous results could be shown. On the one hand, this ensures the existence of a solution to the optimization problem and on the other hand gives an expression for the directional derivative of the reduced objective and for necessary conditions for an optimal solution.

### 3.4. Methods from Shape Optimization

Shape optimization describes a broad range of problems and is a field of current research. In contrast to real calculus, the *geometry* is considered as a variable and a whole new concept of continuity, differentiability, etc. is established. According to Delfour and Zolésio ([DZ11]), in the context of continuum mechanics, shape optimization contributes statements about the optimal structure of beams, plates or shells. Optimal in this sense means to minimize the compliance of a body, i.e. the work of applied loadings.

In [DZ11] it is also emphasized to distinguish between *shape optimization problems* and *distributed parameter problems*. While shape optimization includes problems such as finding an optimal shaped midsurface of a shell (which is fixed in this work), distributed parameter problems deal with tasks like thickness optimization (where the thickness is viewed as a distributed parameter over the midsurface). Hence, the more general techniques from shape optimization are not suited as well for thickness optimization as the optimal control techniques from the previous section, but nevertheless they can be applied to this problem. A related introduction to shape optimization with applications to linear elasticity can be found in [HM03].

The field of *topology optimization* is also closely related and shall be mentioned here. It deals with the optimal placement of *holes* inside of structures and has been an ongoing topic of research for decades, as key contributions by Cheng and Olhoff ([KO82], for plates) from 1982 and Bendsoe ([Ben95]) from 1995 show. Example 7.3 goes in this direction, suggesting the optimal topology (a truss-like structure) of a partly clamped cylindrical shell.

The application of shape optimization techniques is the aim of the following section, and certainly the strong presumption is that this approach leads to the same results as in the previous section. This is answered in Theorem 3.10.

### 3.4.1. Basic definitions

The first definition is the one of a shape functional itself. As said in the introduction, it does not depend on real variables but on a certain geometry.

**Definition 3.6** (Shape functional [DZ11, p. 170]). Let  $D \subseteq \mathbb{R}^n$  be a set called *holdall* and  $\mathcal{A} \subseteq \mathcal{P}(D)$  be a family of subsets of  $D$ . A real valued *shape functional*

$$J : \mathcal{A} \rightarrow \mathbb{R}, \quad \Omega \mapsto J(\Omega)$$

maps sets in  $\mathcal{A}$  to real values.

In general, the set  $D$  is described by mathematical or physical constraints (like a fixed midsurface for shells or volume constraints). In applications it should be chosen large enough and smoothly bounded to allow for all required calculations.

The general task in shape optimization is to find an optimal subset  $\Omega^*$  such that

$$J(\Omega^*) = \min_{\Omega \in \mathcal{A}} J(\Omega).$$

In order to make statements about the existence and necessary conditions for a minimum and to provide methods for finding such an optimal subset, a concept of continuity and differentiability of shape functionals must be derived. A general way would be to define  $\mathcal{A}$  as a set of certain transformations of a fixed set  $\Omega_0$  and to define a metric (with a corresponding topology) on  $\mathcal{A}$ , see [DZ11, chapters 4–8]. This can be done in various ways and the deep investigation of all the methods would lead too far away from the core of this work.

The needs in this section are completely fulfilled by using the so called *velocity method* (see [DZ11, p. 159ff.]). Remember that the Definition 3.2 of the Gâteaux-derivative, of which a counterpart for shape functionals will be given, requires a vector space structure on the underlying space. Such a structure is not inherent in families of sets.

Here the velocity method helps out: Consider families of vector fields

$$V(s) = V(s, \cdot) : \bar{D} \rightarrow \mathbb{R}^n, \quad 0 \leq s \leq s_0$$

and the ordinary differential equation

$$\frac{dx}{ds}(s; X) = V(s, x(s; X)), \quad x(0; X) = X. \quad (3.4.1)$$

To ensure the existence of a solution and certain smoothness properties, some conditions must be stated on the vector field  $V(s)$ . In the following, only vector fields which fulfill

$$\begin{aligned} &\forall X \in \bar{D} : V(\cdot, X) \in C([0, s_0]; \mathbb{R}^n), \\ &\exists c > 0, \forall X, Y \in \bar{D} : \|V(\cdot, Y) - V(\cdot, X)\|_{C([0, s_0]; \mathbb{R}^n)} \leq c \|Y - X\|_2 \end{aligned}$$

are considered. In other words, the continuity with respect to the “time parameter”  $s$  and a uniform Lipschitz-condition is required. This is written shortly as

$$V \in C([0, s_0]; C^{0,1}(\bar{D}; \mathbb{R}^n)). \quad (3.4.2)$$

In addition, a tangential condition is required:

$$\forall s \in [0, s_0], X \in \bar{D} : \pm V(s, X) \in T_{\bar{D}}(X) \quad (3.4.3)$$

where  $T_{\bar{D}}(X)$  is the (Bouligand-)tangential cone in  $X$ , i.e.

$$T_{\bar{D}}(X) = \{v \in \mathbb{R}^n \mid \exists \alpha_k \searrow 0, h_k \rightarrow v : X + \alpha_k h_k \in \bar{D}\}.$$

The theory of ODEs ([Mei07, pp. 84–98]) then states that there is a unique solution to equation (3.4.1). This solution defines a family of transformations  $T_s : \bar{D} \rightarrow \bar{D}$  by

$$T_s(X) := T(s, X) := x(s; X) \quad (3.4.4)$$

and transformed sets  $\Omega_s := T_s(\Omega)$ . Moreover, the transformations have certain smoothness properties (referenced as List 3.4.1 in the following):

- $T(\cdot, X) : [0, s_0] \rightarrow \mathbb{R}^n$  is continuously differentiable for all  $X \in \bar{D}$ ,
- $T(s, \cdot) : \bar{D} \rightarrow \bar{D}$  is Lipschitz-continuous and invertible for all  $s \in [0, s_0]$ .

With help of the above transformations (3.4.4), a useful structure is defined on the sets in  $\mathcal{A}$  and a concept of (semi-)differentiability can be stated:

**Definition 3.7** (Shape semi-derivative [DZ11, p. 473]). Consider a topological vector subspace

$$\Theta \subseteq \{v \in C^{0,1}(\bar{D}; \mathbb{R}^n) \mid \forall X \in \bar{D} : \pm v(X) \in T_{\bar{D}}(X)\}.$$

Let  $v \in \Theta$  be given. Then the shape functional  $J : \mathcal{A} \rightarrow \mathbb{R}$  has a *shape semi-derivative* (Hadamard semi-derivative) in  $\Omega \in \mathcal{A}$  in direction  $v$ , if the limit

$$dJ(\Omega; v) := \lim_{\substack{V \in C^{0,1}([0, s_0]; \Theta) \\ V(0) = v \\ s \searrow 0}} \frac{J(T_s(\Omega)) - J(\Omega)}{s}$$

exists and only depends on  $V(0) = v$  for all  $V$  which fulfill (3.4.3).

**Definition 3.8** (Shape derivative, [DZ11, p. 473]). If  $J$  has a shape semi-derivative in  $\Omega$  for all  $v \in \Theta$  and if the mapping

$$v \mapsto dJ(\Omega; v) : \Theta \rightarrow \mathbb{R}$$

is linear and continuous in  $v$ , then  $J$  has a *shape derivative* and the above mapping is called *gradient* of  $J$ .

The term ‘‘Hadamard’’ semi-derivative refers to a concept that must be distinguished from the Gâteaux-derivative introduced in Definition 3.2. The key difference is that the Hadamard semi-derivative does not depend on the ‘‘history’’ of  $V$ . The corresponding counterpart for functionals on Banach spaces is given here for completeness.

**Definition 3.9** (Hadamard (semi-)derivative [DZ11, p. 459]). Let  $X$  be a Banach space,  $U \subseteq X$  open,  $U \neq \emptyset$  and  $f : U \rightarrow \mathbb{R}$ . If the limit

$$d_H f(x; h) := \lim_{\substack{s \rightarrow 0 \\ w \rightarrow h}} \frac{f(x + sw) - f(x)}{s}$$

exists in  $Y$ , then  $d_H f(x; v)$  is called Hadamard semi-derivative of  $f$  at  $x$  in direction  $h$ . If  $d_H f(x; h)$  exists for all  $h \in X$  and the mapping

$$h \mapsto d_H f(x; h)$$

is linear and continuous, then  $d_H f(x; \cdot)$  is called *Hadamard derivative* of  $f$  at  $x$ .

The advantage of the Hadamard derivative is that there is a chain rule valid for compositions of Hadamard-differentiable functions, in contrast to Gâteaux-differentiable ones. Furthermore, Hadamard-differentiable functions are continuous [DZ11, p. 458].

### 3.4.2. Application to the steady-state problem

Now that the basic definitions are stated, the techniques of shape optimization can be applied to the steady-state optimization problem (3.2.1). In [DZ11] a formula is given to calculate shape derivatives for a special type of shape functionals, more precise for domain integrals. Similar formulas can also be found in [SZ92].

**Theorem 3.8** (cf. [DZ11, p. 482], [SZ92, p.113]). Let  $\varphi$  be a function in  $W_{\text{loc}}^{1,1}(\mathbb{R}^n)$ , i.e.  $\varphi|_K \in W^{1,1}(K)$  for every compact subset  $K \subset \mathbb{R}^n$ . Let the vector field  $v \in C_{\text{loc}}^1(\mathbb{R}^n, \mathbb{R}^n)$  be given. Then the shape functional

$$J(\Omega) := \int_{\Omega} \varphi \, dx$$

has the shape semi-derivative

$$dJ(\Omega; v) = \int_{\Omega} \text{div}(\varphi v) \, dx.$$

If  $\Omega$  has a Lipschitz-boundary  $\Gamma$ , then it holds

$$dJ(\Omega; v) = \int_{\Gamma} \varphi v \cdot n \, dS.$$

*Proof.* The proof is a straight-forward calculation and given in [DZ11, p.482f.]. It is reproduced in short here, because a similar proof will be given later in Lemma 3.13. The idea is to rewrite the domain integrals over transformed subsets  $T_s(\Omega)$  as integrals over  $\Omega$  by a change of variables, i.e.

$$J(T_s(\Omega)) = \int_{T_s(\Omega)} \varphi \, dx = \int_{\Omega} (\varphi \circ T_s) J_s \, dx,$$

where  $J_s$  denotes the Jacobian determinant of the transformation  $T_s$ . The transformations  $T_s$  (see (3.4.4)) are induced by an arbitrary but fixed vector field  $V \in C^0([0, s_0]; C_{\text{loc}}^1(\mathbb{R}^n, \mathbb{R}^n))$  with  $V(0) = v$  that satisfies condition (3.4.2). Then, it can be shown that

$$\begin{aligned} \frac{d}{ds}(\varphi \circ T_s) &= (\nabla \varphi \cdot V(s)) \circ T_s, \\ \frac{d}{ds} J_s &= (\text{div } V(s)) \circ T_s J_s \end{aligned}$$

and finally

$$\begin{aligned} dJ(\Omega; v) &= \left. \frac{d}{ds} J(T_s(\Omega)) \right|_{s=0} = \\ &= \int_{\Omega} \nabla \varphi \cdot V(0) + \varphi \text{div}(V(0)) \, dx = \int_{\Omega} \nabla \varphi \cdot v + \varphi \text{div}(v) \, dx. \end{aligned}$$

□

Certainly, the optimization problem (3.2.1) does not only consist of a domain integral. There are the state equation and additional side constraints that must be considered. In order to do this, the problem is first reformulated as a min-min-problem. The compliance objective function without the regularization term is best suited for this approach. As a reminder, it is stated once again here:

$$\min_{\substack{\tau \in W^{1,\infty}(\mathcal{S}) \\ (u, \theta) \in \mathcal{V}}} J(u, \theta) := \int_{\mathcal{S}} u_3 f \, dS \quad (3.4.5a)$$

$$\text{s.t. : } \int_{\mathcal{B}(\tau)} \sigma(u, \theta) : e(v, \psi) \, dV = \int_{\mathcal{S}} f v_3 \, dS \quad \text{f.a. } (v, \psi) \in \mathcal{V} \quad (3.4.5b)$$

$$\tau_{\min} \leq \tau(\xi^1, \xi^2) \leq \tau_{\max} \quad \text{in } \mathcal{S} \quad (3.4.5c)$$

$$\int_{\mathcal{S}} \left( \tau + \frac{K\tau^3}{12} \right) \, dS = C. \quad (3.4.5d)$$

To substitute the state equation (3.4.5b), an energy functional is introduced,

$$\Pi(u, \theta; \tau) := \frac{1}{2} a_{(\tau)}(u, \theta; u, \theta) - F(u, \theta) = \frac{1}{2} \int_{\mathcal{B}(\tau)} \sigma(u, \theta) : e(u, \theta) \, dV - \int_{\mathcal{S}} f u_3 \, dS.$$

**Lemma 3.11.** The problem (3.4.5) with the additional thickness restrictions collected in  $U_{\text{ad}}$  is equivalent to: Find  $\tau^* \in U_{\text{ad}}$  and  $(u^*, \theta^*) \in \mathcal{V}$  with

$$\Pi(u^*, \theta^*; \tau^*) = \min_{\tau \in U_{\text{ad}}} -2 \min_{(u, \theta) \in \mathcal{V}} \Pi(u, \theta; \tau)$$

*Proof.* First note from the theory of PDE ([Eva10, ch. 8]) that the solution of the variational equation (3.4.5b) corresponds to finding a minimizer for  $\Pi$ , i.e.

$$(u, \theta) = \arg \min_{(v, \psi) \in \mathcal{V}} \Pi(v, \psi; \tau) \quad \Leftrightarrow \quad (u, \theta) \text{ solves (3.4.5b).}$$

Hence, if  $(u_{(\tau)}, \theta_{(\tau)})$  solves the steady-state equation, the objective value  $J(u_{(\tau)}, \theta_{(\tau)})$  can be written as

$$\begin{aligned} J(u_{(\tau)}, \theta_{(\tau)}) &= \int_{\mathcal{S}} u_{(\tau),3} f \, dS = -a_{(\tau)}(u_{(\tau)}, \theta_{(\tau)}; u_{(\tau)}, \theta_{(\tau)}) + 2 \int_{\mathcal{S}} u_{(\tau),3} f \, dS \\ &= -2\Pi(u_{(\tau)}, \theta_{(\tau)}; \tau) = -2 \min_{(u, \theta) \in \mathcal{V}} \Pi(u, \theta; \tau). \end{aligned}$$

The minimization of  $J_s(\tau) = J(u_{(\tau)}, \theta_{(\tau)})$  with respect to  $\tau$  then leads to

$$\min_{\tau \in U_{\text{ad}}} J_s(\tau) = \min_{\tau \in U_{\text{ad}}} -2 \min_{(u, \theta) \in \mathcal{V}} \Pi(u, \theta; \tau).$$

□

The thickness  $\tau$  only influences the *domain* where the variational equation (3.4.5b) is solved. So in the energy functional the dependence on  $\tau$  can be replaced by a dependence on the shell body  $\mathcal{B}_{(\tau)}$ . This fact together with the above lemma gives the idea to define the shape functional

$$J_{\text{Sh}}(\mathcal{B}_{(\tau)}) := \min_{(u, \theta) \in \mathcal{V}} \Pi(u, \theta; \mathcal{B}_{(\tau)}) := \min_{(u, \theta) \in \mathcal{V}} \Pi(u, \theta; \tau). \quad (3.4.6)$$

An appropriate holdall  $D$  would be a shell of arbitrary “thickness”

$$D := \{Z(\xi^1, \xi^2, \xi^3) \mid (\xi^1, \xi^2) \in \omega\} \quad (3.4.7)$$

and the family of subsets  $\mathcal{A} \subset \mathcal{P}(D)$ , where  $J_{\text{Sh}}$  is defined on, is given as shells with specific thicknesses

$$\mathcal{A} := \{\mathcal{B}_{(\tau)} \mid 0 < \tau(\xi^1, \xi^2) < 2 \mid R_{\min}(\xi^1, \xi^2)\}.$$

The actual shape optimization problem is to find  $\mathcal{B}_{(\tau^*)}$  with

$$J_{\text{Sh}}(\mathcal{B}_{(\tau^*)}) = \max_{\mathcal{B}_{(\tau)} \in \mathcal{A}} J_{\text{Sh}}(\mathcal{B}_{(\tau)}). \quad (3.4.8)$$

Next, the subspace  $\Theta \subset C^{0,1}(\bar{D}, \mathbb{R}^3)$  from Definition 3.7 has to be defined, i.e. the space of directions in which  $J_{\text{Sh}}$  shall be (semi-)derived. For this purpose,

$$\Theta := \{v \in C^{0,1}(\bar{D}, \mathbb{R}^n) \mid v = \xi^3 q(\xi^1, \xi^2) a_3(\xi^1, \xi^2), q \in C^{0,1}(\mathcal{S})\}$$

is suited.

**Lemma 3.12.** The vector fields  $V \in C^0([0, s_0]; \Theta)$  fulfill the tangential condition (3.4.3) for  $D$  given from (3.4.7), i.e.

$$\forall s \in [0, s_0], \forall X \in \bar{D} : \pm V(s, X) \in T_{\bar{D}}(X) \quad (3.4.9)$$

where  $T_{\bar{D}}(X)$  is the (Bouligand-)tangential cone in  $X$ . Moreover, the transformations  $T_s$  defined in (3.4.4) fulfill

$$T_s : \mathcal{A} \rightarrow \mathcal{A}$$

for sufficient small  $s \leq s_0$ .

*Proof.* For the first part of the lemma, consider the remark in [DZ11, p. 195], where it is stated that (3.4.9) is fulfilled if for all  $X \in \partial D$

$$V(s, X) \cdot n(X) = 0$$

if the outward normal  $n(X)$  exists. Since  $\partial D$  is given by

$$\partial D = \{z(\xi^1, \xi^2) + \xi^3 a_3(\xi^1, \xi^2) \mid (\xi^1, \xi^2) \in \partial\omega\},$$

the normal  $n(X)$  for  $X \in \partial D$ , where it exists, is perpendicular to  $a_3$  and therefore  $V(s, X) \cdot n(X) = 0$  for  $V \in C^0([0, s_0]; \Theta)$ .

For the second part, let  $\mathcal{B}_{(\tau)} \in \mathcal{A}$  be fixed and  $V \in C^0([0, s_0]; \Theta)$ . In the local basis only the third component of  $V(s)$  is different from zero. So for any  $X = Z(\xi^1, \xi^2, \xi^3) \in \mathcal{B}_{(\tau)}$  under the transformation  $T_s$  induced by  $V$  only its  $\xi^3$ -coordinate changes, i.e.

$$T_s(X) = z(\xi^1, \xi^2) + a_3(\xi^1, \xi^2)k(\xi^3) = Z(\xi^1, \xi^2, k(\xi^3)).$$

Since  $V(s)$  is linear in  $\xi^3$ , the same holds for  $T_s$ , in particular for  $\tilde{X} = Z(\xi^1, \xi^2, -\xi^3)$

$$T_s(\tilde{X}) = z(\xi^1, \xi^2) - a_3(\xi^1, \xi^2)k(\xi^3) = Z(\xi^1, \xi^2, -k(\xi^3)).$$

Hence, the transformed shell body  $T_s(\mathcal{B}_{(\tau)})$  can be described as a shell body with transformed thickness  $\mathcal{B}_{(\tilde{\tau}_s)}$ . The smoothness for  $\tilde{\tau}$  is also kept, i.e.  $\tilde{\tau} \in W^{1,\infty}(\mathcal{S})$ , since the transformation  $T_s$  is Lipschitz-continuous (see List 3.4.1, point 2). So  $\mathcal{B}_{(\tilde{\tau})} \in \mathcal{A}$  for sufficient small  $s$ .  $\square$

Now, the limit

$$\lim_{s \searrow 0} \frac{J_{\text{Sh}}(T_s(\mathcal{B}_{(\tau)})) - J_{\text{Sh}}(\mathcal{B}_{(\tau)})}{s}$$

for transformations  $T_s$  induced by  $V \in C^0([0, s_0]; \Theta)$  has to be investigated. The above Lemma 3.12 ensures that  $J_{\text{Sh}}(T_s(\mathcal{B}_{(\tau)}))$  is well-defined for  $s \leq s_0$ .

A look at the definition of  $J_{\text{Sh}}$  (see (3.4.6)) reveals that this shape functional is defined as a minimum. The differentiation of a minimum can be quite difficult, but fortunately in [DZ11, p. 524] the following Theorem 3.9 can be found. It is suited for mappings of type

$$s \mapsto \inf_{x \in X} G(s, x), \quad G : [0, s_0] \times X \rightarrow \mathbb{R},$$

which shall be derived with respect to  $s$ .

**Theorem 3.9** ([DZ11, p. 524]). Let  $X$  be an arbitrary set,  $s_0 > 0$  and  $G : X \times [0, s_0] \rightarrow \mathbb{R}$  be a functional. Let  $X(s) = \arg \min_{x \in X} G(s, x)$ . Assume that the following conditions are satisfied:

1. For all  $s \in [0, s_0]$ :  $X(s) \neq \emptyset$
2. For all  $x \in \bigcup_{s \in [0, s_0]} X(s)$ :  $\frac{\partial G}{\partial s}(x, s)$  exists everywhere in  $[0, s_0]$
3. There exists a topology  $\mathcal{T}_X$  on  $X$  such that for any sequence  $s_n \subset (0, s_0]$  with  $s_n \rightarrow 0$  there exist  $x^0 \in X(0)$  and a subsequence  $s_{n_k}$  of  $s_n$  and for each  $k \geq 1$  there exists  $x_{n_k} \in X(s_{n_k})$  such that
  - a)  $x_{n_k} \rightarrow x^0$  in the  $\mathcal{T}_X$ -topology
  - b)  $\liminf_{\substack{k \rightarrow \infty \\ s_{n_k} \searrow 0}} \frac{\partial G}{\partial s}(x_{n_k}, s) \geq \frac{\partial G}{\partial s}(x^0, 0)$
4. For all  $x$  in  $X(0)$ : the map  $s \mapsto \frac{\partial G}{\partial s}(x, s)$  is upper semicontinuous at  $s = 0$ .

Then there exists  $x^0 \in X(0)$  such that

$$\lim_{s \searrow 0} \frac{1}{s} \left( \inf_{x \in X} G(s, x) - \inf_{x \in X} G(0, x) \right) = \frac{\partial G}{\partial s}(s, x^0).$$

The above Theorem 3.9 is applied to obtain the shape semi-derivative of  $J_{\text{Sh}}$  from (3.4.6) which is given by

$$\lim_{s \searrow 0} \frac{1}{s} \left( \min_{(u, \theta) \in \mathcal{V}} \Pi(u, \theta; T_s(\mathcal{B}(\tau))) - \min_{(u, \theta) \in \mathcal{V}} \Pi(u, \theta; \mathcal{B}(\tau)) \right),$$

i.e. the conditions of the theorem have to be fulfilled by

$$\tilde{\Pi}(u, \theta; s) := \Pi(u, \theta; T_s(\mathcal{B}(\tau))) : \mathcal{V} \times [0, s_0] \rightarrow \mathbb{R}.$$

In points two, three and four of the conditions in Theorem 3.9, the partial derivative of  $\tilde{\Pi}$  with respect to  $s$  occurs. This is calculated in advance:

**Lemma 3.13.** Let  $\tau \in U_b$  and  $V \in C^0([0, s_0]; \Theta)$  with corresponding transformations  $T_s$ ,  $s \in [0, s_0]$  be given. Then the mapping  $\tilde{\Pi} : \mathcal{V} \times [0, s_0] \rightarrow \mathbb{R}$ ,

$$\tilde{\Pi}(u, \theta; s) = \Pi(u, \theta; T_s(\mathcal{B}(\tau)))$$

is partially differentiable with respect to  $s$ . In  $s = 0$  it holds

$$\frac{\partial \tilde{\Pi}}{\partial s}(u, \theta; 0) = \int_{\mathcal{B}(\tau)} \frac{\partial \varphi}{\partial \xi^3} V_3(0) + \varphi \frac{\partial V_3(0)}{\partial \xi^3} dV,$$

where  $V_3$  denotes the third component with respect to the local basis of  $\mathcal{B}(\tau)$  and

$$\varphi := \sigma(u, \theta) : e(u, \theta)$$

is written to shorten the notation.

*Proof.* Remember the mapping  $\Pi : \mathcal{V} \times \mathcal{A} \rightarrow \mathbb{R}$  is given by a domain and surface integral,

$$\Pi(u, \theta; \mathcal{B}_{(\tau)}) = \frac{1}{2} \int_{\mathcal{B}_{(\tau)}} \sigma(u, \theta) : e(u, \theta) \, dV - \int_{\mathcal{S}} f u_3 \, dS.$$

Since the midsurface  $\mathcal{S}$  is not affected by the transformations  $T_s$  at all, the latter part can be neglected when calculating the derivative. The partial derivative of  $\tilde{\Pi}$  at  $s = 0$  corresponds to the Hadamard semi-derivative of  $\Pi$  in direction  $V(0) = v$ . The derivative of domain integrals was discussed in Theorem 3.8, but a direct application is not possible here, since it requires a  $W_{\text{loc}}^{1,1}$ -regularity of the integrand, which would force  $(u, \theta)$  to be in  $H^2(\mathcal{S})$ -spaces. This regularity cannot be achieved and is a too strict requirement. However, the proof of Theorem 3.8 can be reproduced using the special structure of the vector fields in  $\Theta$ , namely

$$v = \xi^3 q(\xi^1, \xi^2) a_3(\xi^1, \xi^2),$$

to achieve a proper result.

Then, by transforming the integral over  $T_s(\mathcal{B}_{(\tau)})$  back to  $\mathcal{B}_{(\tau)}$ ,

$$\Pi(u, \theta; T_s(\mathcal{B}_{(\tau)})) = \int_{\mathcal{B}_{(\tau)}} (\varphi \circ T_s) J_s \, dV.$$

Since the transformation  $T_s$  only influences the  $\xi^3$ -coordinate (see Lemma 3.12), the derivative of the integrand with respect to  $s$  reads (cf. the proof of Theorem 3.8)

$$\begin{aligned} \frac{d}{ds}(\varphi \circ T_s) &= \left( \frac{\partial \varphi}{\partial \xi^3} V_3(s) \right) \circ T_s, \\ \frac{dJ_s}{ds} &= (\text{div } V(s)) \circ T_s J_s = \left( \frac{\partial V_3(s)}{\partial \xi^3} \right) \circ T_s J_s, \end{aligned}$$

Note that indeed  $\varphi = \sigma(u, \theta) : e(u, \theta)$  is differentiable with respect to  $\xi^3$  for  $(u, \theta) \in \mathcal{V}$ . Then it holds

$$\frac{\partial \tilde{\Pi}}{\partial s}(u, \theta; s) = \int_{\mathcal{B}_{(\tau)}} \left( \frac{\partial \varphi}{\partial \xi^3} V_3(s) \right) \circ T_s J_s + \varphi \left( \frac{\partial V_3(s)}{\partial \xi^3} \right) \circ T_s J_s \, dV$$

which shows that the partial derivative exists and especially for  $s = 0$

$$\frac{\partial \tilde{\Pi}}{\partial s}(u, \theta; 0) = \int_{\mathcal{B}_{(\tau)}} \frac{\partial \varphi}{\partial \xi^3} V_3(0) + \varphi \frac{\partial V_3(0)}{\partial \xi^3} \, dV.$$

□

Now the conditions of Theorem 3.9 are checked for  $\tilde{\Pi}$ . In particular, these are:

1. It holds  $X(s) \neq \emptyset$  for all  $s \in [0, s_0]$ , because  $T_s(\mathcal{B}_{(\tau)}) = \mathcal{B}_{(\tilde{\tau}_s)}$  with  $\tilde{\tau}_s \in U_b$  for all  $s \leq s_0$  with sufficient small  $s_0$  (see Lemma 3.12). For those  $\tilde{\tau}_s$  there is a unique solution to the state equation (see Theorem 3.2). Moreover,  $X(s) = \{(u_s, \theta_s)\}$  is a singleton.

2. According to Lemma 3.13, the partial derivative  $\frac{\partial \tilde{\Pi}}{\partial s}(u, \theta; s)$  exists for all  $(u, \theta) \in \mathcal{V}$  and  $s \in [0, s_0]$ .
3. The first part is due to the continuity of the control-to-state operator  $G$ , because a sequence  $s_n \rightarrow 0$  corresponds to a sequence of thicknesses  $\tilde{\tau}_{s_n} \in U_b$  for which  $\Pi(u, \theta; s_n) = \Pi(u, \theta; \mathcal{B}_{(\tilde{\tau}_{s_n})})$ . The corresponding sets of minimizers contain the (unique) solutions  $X(s_n) = \{(u_{(\tilde{\tau}_{s_n})}, \theta_{(\tilde{\tau}_{s_n})})\} = \{G(\tilde{\tau}_{s_n})\}$ . For  $s_n \rightarrow 0$  it holds  $\tilde{\tau}_{s_n} \rightarrow \tau$  and  $G(\tilde{\tau}_{s_n}) \rightarrow G(\tau) =: (u_0, \theta_0)$  in  $\mathcal{V}$ -strong.
- For the second part, set

$$\begin{aligned}\varphi_{n_k} &:= \sigma(u_{n_k}, \theta_{n_k}) : e(u_{n_k}, \theta_{n_k}) \\ \varphi &:= \sigma(u_0, \theta_0) : e(u_0, \theta_0)\end{aligned}$$

and note that

$$\varphi_{n_k} \rightarrow \varphi \tag{3.4.10}$$

in  $L^1(\mathcal{B}_{(\tau)})$ -strong because of the continuity of  $\sigma$  and  $e$ . The same is true for the derivative

$$\frac{\partial \varphi_{n_k}}{\partial \xi^3} \rightarrow \frac{\partial \varphi}{\partial \xi^3} \tag{3.4.11}$$

because  $(u_{n_k}, \theta_{n_k})$  is independent of  $\xi^3$ . Consider the difference

$$\begin{aligned}& \frac{\partial \tilde{\Pi}}{\partial s}(u_{n_k}, \theta_{n_k}; s) - \frac{\partial \tilde{\Pi}}{\partial s}(u_0, \theta_0; 0) \\ &= \int_{\mathcal{B}_{(\tau)}} \left( \frac{\partial \varphi_{n_k}}{\partial \xi^3} V_3(s) \right) \circ T_s J_s + \varphi_{n_k} \frac{\partial V_3(s)}{\partial \xi^3} \circ T_s J_s - \frac{\partial \varphi}{\partial \xi^3} V_3(0) + \varphi \frac{\partial V_3(0)}{\partial \xi^3} dV \\ &= \int_{\mathcal{B}_{(\tau)}} \frac{\partial \varphi_{n_k}}{\partial \xi^3} (V_3(s) \circ T_s J_s - V_3(0)) + \varphi_{n_k} \left( \frac{\partial V_3(s)}{\partial \xi^3} \circ T_s J_s - \frac{\partial V_3(0)}{\partial \xi^3} \right) dV \\ & \quad + \int_{\mathcal{B}_{(\tau)}} \left( \frac{\partial \varphi_{n_k}}{\partial \xi^3} - \frac{\partial \varphi}{\partial \xi^3} \right) V_3(0) + (\varphi_{n_k} - \varphi) \frac{\partial V_3(0)}{\partial \xi^3} dV\end{aligned}$$

The first part of the last term converges to zero since the  $\varphi_{n_k}$  and  $\frac{\partial \varphi_{n_k}}{\partial \xi^3}$  are bounded and

$$\begin{aligned}V_3(s) \circ T_s J_s - V_3(0) &\rightarrow 0 \\ \left( \frac{\partial V_3(s)}{\partial \xi^3} \right) \circ T_s J_s - \frac{\partial V_3(0)}{\partial \xi^3} &\rightarrow 0\end{aligned}$$

in  $L^\infty(\mathcal{B}_{(\tau)})$ . The second part converges to zero because of (3.4.10) and (3.4.11).

4. The fourth point is also satisfied, because for fixed  $(u, \theta)$  the mapping

$$s \mapsto \int_{\mathcal{B}_{(\tau)}} \left( \frac{\partial \varphi}{\partial \xi^3} V_3(s) \right) \circ T_s J_s + \varphi \left( \frac{\partial V_3(s)}{\partial \xi^3} \right) \circ T_s J_s dV$$

is continuous in  $s$ .

So all conditions are satisfied and the conclusion of Theorem 3.9 can be used. It states that the shape semi-derivative can be expressed as

$$\begin{aligned}
& dJ_{\text{Sh}}(\mathcal{B}(\tau); V(0)) \\
&= \lim_{s \searrow 0} \frac{1}{s} \left( \min_{(u, \theta) \in \mathcal{V}} \Pi(u, \theta; T_s(\mathcal{B}(\tau))) - \min_{(u, \theta) \in \mathcal{V}} \Pi(u, \theta; \mathcal{B}(\tau)) \right) \\
&= \frac{\partial \tilde{\Pi}}{\partial s}(u(\tau), \theta(\tau); 0) \\
&= \int_{\mathcal{B}(\tau)} \frac{\partial(\sigma(u(\tau), \theta(\tau)) : e(u(\tau), \theta(\tau)))}{\partial \xi^3} V_3(0) + \sigma(u(\tau), \theta(\tau)) : e(u(\tau), \theta(\tau)) \frac{\partial V_3(0)}{\partial \xi^3} dV.
\end{aligned} \tag{3.4.12}$$

Note that the expression for  $dJ_{\text{Sh}}(\mathcal{B}(\tau); V(0))$  indeed does not depend on  $V(s)$ ,  $s > 0$ , but only on  $v = V(0)$ , which is requested in the definition of the Hadamard shape semi-derivative 3.7.

The last term from (3.4.12) can be simplified even further. It holds

$$\begin{aligned}
dJ_{\text{Sh}}(\mathcal{B}(\tau); v) &= \int_{\mathcal{B}(\tau)} \frac{\partial((\sigma(u(\tau), \theta(\tau)) : e(u(\tau), \theta(\tau))) v_3)}{\partial \xi^3} dV \\
&= \int_{\partial \mathcal{B}(\tau)} \sigma(u(\tau), \theta(\tau)) : e(u(\tau), \theta(\tau)) v_3 a_3 \cdot n d\Gamma.
\end{aligned} \tag{3.4.13}$$

This boundary integral can be evaluated in detail.

**Lemma 3.14.** It holds

$$\begin{aligned}
& dJ_{\text{Sh}}(\mathcal{B}(\tau); v) \\
&= \int_{\partial \mathcal{B}(\tau)} \sigma(u(\tau), \theta(\tau)) : e(u(\tau), \theta(\tau)) v_3 a_3 \cdot n d\Gamma \\
&= \int_{\mathcal{S}} \sum_{\xi^3 \in \{\pm \frac{\tau}{2}\}} [(\sigma(u(\tau), \theta(\tau)) : e(u(\tau), \theta(\tau))) (1 - 2H\xi^3 + K(\xi^3)^2)] \Big|_{\xi^3 = \xi^3} \tau \frac{q}{2} dS.
\end{aligned}$$

*Proof.* The evaluation of the boundary integral reduces to integration over

$$\begin{aligned}
\Gamma_1 &:= \left\{ Z \left( \xi^1, \xi^2, \frac{\tau(\xi^1, \xi^2)}{2} \right) = z(\xi^1, \xi^2) + \frac{\tau(\xi^1, \xi^2)}{2} a_3 \mid (\xi^1, \xi^2) \in \omega \right\}, \\
\Gamma_2 &:= \left\{ Z \left( \xi^1, \xi^2, -\frac{\tau(\xi^1, \xi^2)}{2} \right) = z(\xi^1, \xi^2) - \frac{\tau(\xi^1, \xi^2)}{2} a_3 \mid (\xi^1, \xi^2) \in \omega \right\},
\end{aligned}$$

since on the other part of the boundary  $a_3 \cdot n = 0$ , see the proof of the first part of Lemma 3.12. Considering  $\Gamma_1$ , the outward normal  $n$  is given by

$$\begin{aligned}
\tilde{n} &= \left( \frac{\partial z}{\partial \xi^1} + \frac{1}{2} \left( \frac{\partial \tau}{\partial \xi^1} a_3 + \tau \frac{\partial a_3}{\partial \xi^1} \right) \right) \times \left( \frac{\partial z}{\partial \xi^2} + \frac{1}{2} \left( \frac{\partial \tau}{\partial \xi^2} a_3 + \tau \frac{\partial a_3}{\partial \xi^2} \right) \right) \\
n &= \frac{\tilde{n}}{\|\tilde{n}\|}.
\end{aligned}$$

The expression for  $\tilde{n}$  can be rewritten using the covariant basis of  $\mathcal{B}_{(\tau)}$  from (2.2.4)

$$\tilde{n} = \left( g_1 \left( \xi^1, \xi^2, \frac{\tau(\xi^1, \xi^2)}{2} \right) + \frac{1}{2} \frac{\partial \tau}{\partial \xi^1} a_3 \right) \times \left( g_2 \left( \xi^1, \xi^2, \frac{\tau(\xi^1, \xi^2)}{2} \right) + \frac{1}{2} \frac{\partial \tau}{\partial \xi^2} a_3 \right),$$

and using  $a_3 = g_3$  it follows

$$\begin{aligned} a_3 \cdot \tilde{n} &= g_3 \cdot \left( g_1 \left( \xi^1, \xi^2, \frac{\tau(\xi^1, \xi^2)}{2} \right) \times g_2 \left( \xi^1, \xi^2, \frac{\tau(\xi^1, \xi^2)}{2} \right) \right) \\ &= \sqrt{a} \left( 1 - 2H \frac{\tau(\xi^1, \xi^2)}{2} + K \left( \frac{\tau(\xi^1, \xi^2)}{2} \right)^2 \right) \end{aligned}$$

where the first equation only includes the non-vanishing parts of the product and the second equation comes from (2.2.6) and (2.2.7). The integral over  $\Gamma_1$  then reads

$$\begin{aligned} &\int_{\Gamma_1} \sigma(u_{(\tau)}, \theta_{(\tau)}) : e(u_{(\tau)}, \theta_{(\tau)}) v_3 a_3 \cdot n \, d\Gamma \\ &= \int_{\Gamma_1} \sigma(u_{(\tau)}, \theta_{(\tau)}) : e(u_{(\tau)}, \theta_{(\tau)}) \frac{\tau}{2} q \sqrt{a} \left( 1 - 2H \frac{\tau}{2} + K \left( \frac{\tau}{2} \right)^2 \right) \frac{1}{\|\tilde{n}\|} \, d\Gamma \\ &= \int_{\omega} \sigma(u_{(\tau)}, \theta_{(\tau)}) : e(u_{(\tau)}, \theta_{(\tau)}) \Big|_{\xi^3 = \frac{\tau}{2}} \frac{\tau}{2} q \sqrt{a} \left( 1 - 2H \frac{\tau}{2} + K \left( \frac{\tau}{2} \right)^2 \right) \, d\xi^1 \, d\xi^2 \\ &= \int_{\mathcal{S}} [\sigma(u_{(\tau)}, \theta_{(\tau)}) : e(u_{(\tau)}, \theta_{(\tau)}) (1 - 2H\xi^3 + K(\xi^3)^2)] \Big|_{\xi^3 = \frac{\tau}{2}} \tau \frac{q}{2} \, dS. \end{aligned}$$

An analogous computation gives

$$\begin{aligned} &\int_{\Gamma_2} \sigma(u_{(\tau)}, \theta_{(\tau)}) : e(u_{(\tau)}, \theta_{(\tau)}) v_3 a_3 \cdot n \, d\Gamma \\ &= \int_{\mathcal{S}} [\sigma(u_{(\tau)}, \theta_{(\tau)}) : e(u_{(\tau)}, \theta_{(\tau)}) (1 - 2H\xi^3 + K(\xi^3)^2)] \Big|_{\xi^3 = -\frac{\tau}{2}} \tau \frac{q}{2} \, dS. \end{aligned}$$

Addition of both integrals then leads to the statement of the lemma.  $\square$

Taking a closer look at the reformulated expression for  $dJ_{\text{Sh}}(\mathcal{B}_{(\tau)}; v)$  reveals that it matches with the expression for the Gâteaux-derivative of the reduced objective from Theorem 3.6. This is the main result of this section:

**Theorem 3.10.** Let  $J_s : U_{\text{ad}} \rightarrow \mathbb{R}$  be the reduced objective from (3.2.3) using the compliance functional (Itemization 2.3, point 2) with no regularization term. Then the Gâteaux-derivative of  $J_s$  at point  $\tau \in U_{\text{ad}}$  in direction  $q \in W^{1,\infty}(\mathcal{S})$  and the shape semi-derivative of  $-2J_{\text{Sh}}$  from (3.4.6) at point  $\mathcal{B}_{(\tau)}$  in direction

$$v = \xi^3 \frac{q(\xi^1, \xi^2)}{2\tau(\xi^1, \xi^2)} a_3(\xi^1, \xi^2)$$

coincide.

*Proof.* The expression from  $dJ_{\text{Sh}}(\mathcal{B}(\tau); v)$  obtained in (3.4.12) and simplified in Lemma 3.14 reads

$$\begin{aligned} & -2 dJ_{\text{Sh}}(\mathcal{B}(\tau); v) = \\ & = -2 \int_{\mathcal{S}} \sum_{\bar{\xi}^3 \in \{\pm \frac{\tau}{2}\}} [(\sigma(u(\tau), \theta(\tau)) : e(u(\tau), \theta(\tau))) (1 - 2H\xi^3 + K(\xi^3)^2)] \Big|_{\xi^3 = \bar{\xi}^3} \tau \frac{q}{4\tau} dS \\ & = -2 \int_{\mathcal{S}} \sum_{\bar{\xi}^3 \in \{\pm \frac{\tau}{2}\}} [(\sigma(u(\tau), \theta(\tau)) : e(u(\tau), \theta(\tau))) (1 - 2H\xi^3 + K(\xi^3)^2)] \Big|_{\xi^3 = \bar{\xi}^3} \frac{q}{2} dS \end{aligned}$$

which is the same as the expression from Theorem 3.6 where

$$\begin{aligned} & J'_s(\tau)q \\ & = -Z_q(u(\tau), \theta(\tau)) \\ & = - \int_{\mathcal{S}} \sum_{\bar{\xi}^3 \in \{\pm \frac{\tau}{2}\}} [(\sigma(u(\tau), \theta(\tau)) : e(u(\tau), \theta(\tau))) (1 - 2H\xi^3 + K(\xi^3)^2)] \Big|_{\xi^3 = \bar{\xi}^3} \frac{q}{2} dS. \end{aligned}$$

Both expression coincide, i.e.

$$-2 dJ_{\text{Sh}} \left( \mathcal{B}(\tau); \xi^3 \frac{q(\xi^1, \xi^2)}{2\tau(\xi^1, \xi^2)} a_3(\xi^1, \xi^2) \right) = J'_s(\tau)q.$$

□

**Remark 3.8.** *The factor -2 in front of  $J_{\text{Sh}}$  was reintroduced to keep the equivalence between the optimal control problem (3.4.5) and the shape optimization problem (3.4.8).*

### 3.4.3. Discussion of shape optimization techniques

The above theorem shows that the derivative obtained by using techniques from shape optimization is the same as the derivative obtained by the optimal control/distributed parameter approach from the previous section. However, the smoothness assumption for the thickness, namely  $\tau \in W^{1,\infty}(\mathcal{S})$ , is really needed here (e.g. in equation (3.4.13) when the divergence theorem is used), while in the previous section  $\tau \in L^\infty(\mathcal{S})$  would have been sufficient. Moreover, the calculations made above do not allow for other objective functionals than the compliance functional, because using this particular objective enabled the min-min-reformulation of the problem. Both points are a drawback and show that the thickness optimization problem should really be viewed as a distributed parameter optimization problem. The full power of the shape optimization techniques would come in place if the midsurface would not be left fixed, because the approach from the previous section relies fundamentally on a fixed midsurface.

However, the coincidence of both derivatives is a verification of the results obtained in the last section and shows that essentially both techniques lead to the desired result.



## 4. Dynamic problem

This chapter is devoted to the statement and the analysis of the dynamic state equation and the corresponding optimization problem. The structure is as follows: First, the statement of a suitable weak formulation for the dynamic case is established, since this was not given in Section 2.2. This equation is investigated regarding existence and uniqueness of a solution. After obtaining such a result, the control-to-state operator is analyzed and continuity and differentiability properties are formulated. Finally, the actual dynamic optimization problem is stated and the reduced objective is investigated with respect to the above named properties.

Note that neither a dynamic formulation based on the basic shell model nor its investigation could be found in available literature. The results of this chapter were published by the author in [Zie15] and are a new contribution to optimization in dynamic linear elasticity.

### 4.1. Dynamic state equation

The analysis of the dynamic state equation which was introduced in Subsection 2.2.7 is more complex than the steady-state case. With the introduction of the time-parameter  $t$  an additional dimension is added and the type of the equation is changed from elliptic to hyperbolic. This case needs a different treatment which however relies on results of the steady-state analysis. Those results are necessary to state the dynamic state equation in a proper weak sense at all. The necessary mathematical tools are mainly taken from the books by Evans ([Eva10]), Tröltzsch ([Trö10]) and Wloka ([Wlo87]), while the derivation of the equation is done from scratch, starting from full three-dimensional equations in linear elasticity.

#### 4.1.1. Weak formulation

In order to continue the derivation of the weak formulation which was started in Subsection 2.2.7, the formulation stated in (2.2.29) is repeated here:

Let a time-dependent force density  $f$  be given. Then, find  $(u, \theta)$  such that

$$\int_{\mathcal{S}} \left[ (\Phi_1(\tau)\ddot{u}(t) + \Phi_2(\tau)\ddot{\theta}(t)) \cdot v + (\Phi_2(\tau)\ddot{u}(t) + \Phi_3(\tau)\ddot{\theta}(t)) \cdot \psi \right] dS + \int_{\mathcal{B}(\tau)} (C\varepsilon(u(t), \theta(t))) : \varepsilon(v, \psi) + (D\zeta(u(t), \theta(t))) : \zeta(v, \psi) dV = \int_{\mathcal{S}} f(t)v_3 dS \quad (4.1.1)$$

for all  $(v, \psi) \in \mathcal{V}$ ,  $t \in [0, T]$ . The operators  $\Phi_i : L^\infty(\mathcal{S}) \rightarrow L^\infty(\mathcal{S})$  are given by

$$\Phi_1(\tau) = \rho\tau + \rho K \frac{\tau^3}{12}, \quad \Phi_2(\tau) = -2\rho H \frac{\tau^3}{12}, \quad \Phi_3(\tau) = \rho \frac{\tau^3}{12} + \rho K \frac{\tau^5}{80}. \quad (4.1.2)$$

Clearly, a lot of questions remain open from that starting point, mainly how to choose a proper displacement space for  $(u, \theta)$  and how to interpret this weak formulation. These questions will be clarified in the following.

**Displacement space** Since  $(u, \theta)$  are considered as time-dependent displacements, for a fixed time  $t$  the pair  $(u(t), \theta(t))$  should lie in the space  $\mathcal{V}$  already known from the steady-state case (see (2.2.23)). This concept is expressed with the help of vector-valued functions. A concept of  $L^p$ -spaces of vector-valued functions as well as differentiation of such functions can be found in [Trö10, pp. 141–147] or [Wlo87, §§ 24–25] and is also presented in the Appendix A.4.

In order to define a proper displacement space, first the concept of rigged Hilbert spaces or Gelfand triples must be introduced. For this, let a Hilbert space  $V$  be given which is continuously and densely embedded into another Hilbert space  $H$ . The spaces  $V = H^1(\Omega)$  and  $H = L^2(\Omega)$  for an open subset  $\Omega \in \mathbb{R}^n$  may serve as an example. Since  $H$  is a Hilbert space, it can be identified with  $H^*$ . This identification is avoided for the Hilbert space  $V$ , instead it can be shown that the embedding  $H \hookrightarrow V^*$  is also dense and continuous ([Wlo87, p. 253]), so that the scalar product from  $H$  can be continuously extended to elements in  $V^* \times V$ . This extension can be viewed as a new representation formula for functionals from  $V^*$  ([Wlo87, p. 263]).

**Definition 4.1** (Gelfand triple [Trö10, p. 147]). For two Hilbert spaces  $H$  and  $V$ , the chain of dense and continuous embeddings

$$V \hookrightarrow H \hookrightarrow V^*$$

is called a *Gelfand triple*, the space  $H$  is called *rigged Hilbert space*.

For the steady-state displacement space  $\mathcal{V}$ , a suitable rigged Hilbert space is

$$\mathcal{H} = L^2(\mathcal{S})^2 \times L^2(\mathcal{S}) \times L^2(\mathcal{S})^2. \quad (4.1.3)$$

In addition, a suitable displacement space is  $W(0, T)$ ,

$$W(0, T) := \{(u, \theta) \mid (u, \theta) \in L^2(0, T; \mathcal{V}), (\dot{u}, \dot{\theta}) \in L^2(0, T; \mathcal{H})\},$$

equipped with the norm

$$\|(u, \theta)\|_{W(0, T)}^2 = \int_0^T \|(u(t), \theta(t))\|_{\mathcal{V}}^2 + \|(\dot{u}(t), \dot{\theta}(t))\|_{\mathcal{H}}^2 dt,$$

whereas the  $L^2(0, T; \cdot)$ -spaces are as in Definition A.5 and the time derivative as in Definition A.8. It is reasonable not to include an assumption on  $(\ddot{u}, \ddot{\theta})$ , since it is

achieved automatically from the fact that  $(u, \theta)$  solves the dynamic state equation (4.1.5).

For the applied force  $f$ , the space  $H^1(0, T; L^2(\mathcal{S}))$  is suitable. Actually, for existence and uniqueness of a solution to the state equation  $f \in L^2(0, T; L^2(\mathcal{S}))$  would be sufficient, but the stronger regularity is needed for the differentiability of the control-to-state operator (see Lemma 4.8 and Theorem 4.4).

**Weak formulation** Back to the formulation (4.1.1), the second integral in the equation is (for fixed time  $t$ ) the bilinear form introduced in (3.1.2),

$$a_{(\tau)}(u(t), \theta(t); v, \psi) = \int_{\mathcal{B}(\tau)} \sigma(u(t), \theta(t)) : e(v, \psi) \, dV.$$

According to Theorem 3.1, this bilinear form is coercive and continuous on  $\mathcal{V}$ . Hence, by the Lax-Milgram theorem (see [Yos80, p. 92]), there is a unique operator  $L_{(\tau)} : \mathcal{V} \rightarrow \mathcal{V}^*$  with

$$L_{(\tau)}(u(t), \theta(t))(v, \psi) = a_{(\tau)}(u(t), \theta(t); v, \psi).$$

The right hand side can be interpreted as a continuous functional  $F(t) \in \mathcal{V}^*$ , namely

$$F(t)(v, \psi) = \int_{\mathcal{S}} f(t) v_3 \, dS,$$

as long as  $f(t) \in L^2(\mathcal{S})$ .

The missing part is how to interpret the first integral from (4.1.1) as an element in  $\mathcal{V}^*$ . To do this, it is rewritten using  $L^2(\mathcal{S})$  scalar products,

$$\begin{aligned} & \int_{\mathcal{S}} \left[ (\Phi_1(\tau) \ddot{u}(t) + \Phi_2(\tau) \ddot{\theta}(t)) \cdot v + (\Phi_2(\tau) \ddot{u}(t) + \Phi_3(\tau) \ddot{\theta}(t)) \cdot \psi \right] \, dS \\ &= \left( \Phi_1(\tau) \ddot{u}(t) + \Phi_2(\tau) \ddot{\theta}(t), v \right)_{L^2(\mathcal{S})^2} + (\Phi_1(\tau) \ddot{u}_3(t), v_3)_{L^2(\mathcal{S})} + \\ & \quad \left( \Phi_2(\tau) \ddot{u}(t) + \Phi_3(\tau) \ddot{\theta}(t), \psi \right)_{L^2(\mathcal{S})^2}. \end{aligned} \quad (4.1.4)$$

In view of the above equation an operator  $M_{(\tau)} : \mathcal{H} \rightarrow \mathcal{H}$  is defined via

$$M_{(\tau)}(v, \psi) := \begin{pmatrix} \Phi_1(\tau) v + \Phi_2(\tau) \psi \\ \Phi_1(\tau) v_3 \\ \Phi_2(\tau) v + \Phi_3(\tau) \psi \end{pmatrix}.$$

Because  $\Phi_i(\tau) \in L^\infty(\mathcal{S})$ , the operator is well-defined, linear and continuous, i.e.  $M_{(\tau)} \in L(\mathcal{H}, \mathcal{H})$ . Since  $\mathcal{H}$  is dense and continuous in  $\mathcal{V}^*$ , it can also be continuously extended to an operator  $M_{(\tau)} \in L(\mathcal{V}^*, \mathcal{V}^*)$ . With the help of  $M_{(\tau)}$ , equation (4.1.4) can be rewritten as

$$\begin{aligned} & \left( \Phi_1(\tau) \ddot{u}(t) + \Phi_2(\tau) \ddot{\theta}(t), v \right)_{L^2(\mathcal{S})^2} + (\Phi_1(\tau) \ddot{u}_3(t), v_3)_{L^2(\mathcal{S})} + \\ & \quad \left( \Phi_2(\tau) \ddot{u}(t) + \Phi_3(\tau) \ddot{\theta}(t), \psi \right)_{L^2(\mathcal{S})^2} = \left( M_{(\tau)}(\ddot{u}, \ddot{\theta}), (v, \psi) \right)_{\mathcal{H}}. \end{aligned}$$

Since the first term is written in the scalar product of  $\mathcal{H}$ , and  $L(u(t), \theta(t))$  and  $F(t)$  can be represented by this scalar product due to the properties of the Gelfand triple  $\mathcal{V} \hookrightarrow \mathcal{H} \hookrightarrow \mathcal{V}^*$ , the dynamic state equation (4.1.1) can be rewritten as

$$\left( M_{(\tau)}(\ddot{u}(t), \ddot{\theta}(t)), (v, \psi) \right)_{\mathcal{H}} + (L(u(t), \theta(t)), (v, \psi))_{\mathcal{H}} = (F(t), (v, \psi))_{\mathcal{H}}$$

for all  $(v, \psi) \in \mathcal{V}$ ,  $t \in [0, T]$ . Back as an equation in  $\mathcal{V}^*$ , this reads:

$$M_{(\tau)}(\ddot{u}(t), \ddot{\theta}(t)) + L_{(\tau)}(u(t), \theta(t)) = F(t) \quad \text{in } \mathcal{V}^* \text{ for } t \in [0, T].$$

What is still missing are *initial conditions* for the displacement and its derivative. While it is reasonable to demand no initial displacement at  $t = 0$  (corresponding to a stress-free initial state), it should be possible to give an initial momentum to the shell body. This momentum depends on the thickness and is expressed as

$$(\dot{u}(0), \dot{\theta}(0)) = \Phi_g(\tau),$$

where the continuous operator  $\Phi_g : L^\infty(\mathcal{S}) \rightarrow \mathcal{V}$  depends on a prescribed function  $g \in \mathcal{V}$ . The whole problem then reads: Given  $F \in H^1(0, T; \mathcal{H})$ ,  $F(t) = (0, f(t), 0)$  and  $g \in \mathcal{V}$ , find  $(u, \theta) \in W(0, T)$  such that

$$\begin{aligned} M_{(\tau)}(\ddot{u}(t), \ddot{\theta}(t)) + L_{(\tau)}(u(t), \theta(t)) &= F(t) \quad \text{in } \mathcal{V}^* \text{ for } t \in [0, T] \\ (u(0), \theta(0)) &= 0, \quad (\dot{u}(0), \dot{\theta}(0)) = \Phi_g(\tau). \end{aligned} \tag{4.1.5}$$

This will serve as the *dynamic state equation* from now on.

**Remark 4.1.** *To ensure the existence and uniqueness of a solution to the state equation as well as the continuity of this solution with respect to the thickness, it would be sufficient to have*

$$F \in L^2(0, T; \mathcal{H}), \quad \Phi_g(\tau) \in \mathcal{H} \text{ depending on } g \in \mathcal{H},$$

*see Theorem 4.1. But for the differentiable dependence on the thickness, which is essential for a successful optimization, the higher regularity claim from above is needed, see Theorem 4.4.*

**Choice of initial conditions** As stated above, there should be no displacement at time  $t = 0$ , i.e. the shell body starts in an undeformed configuration. However, to model some sort of initial momentum, for example a body “falling” onto the shell, there could be a non-vanishing time derivative of the displacement at  $t = 0$ . In the last paragraph, a quite general form was chosen,

$$(\dot{u}(0), \dot{\theta}(0)) = \Phi_g(\tau),$$

where the operator  $\Phi_g : L^\infty(\mathcal{S}) \rightarrow \mathcal{H}$  (or  $\mathcal{V}$ ) depends on a prescribed function  $g \in \mathcal{H}$  (or  $\mathcal{V}$ ). In the numerical test cases in Chapter 7, the operator

$$\Phi_g(\tau) = \left( 0, \frac{g}{\tau}, 0 \right)$$

will be used, so the momentum is perpendicular to the midsurface.

This can be regarded as if a fixed momentum  $g$  is given to the shell body: If  $\dot{u}_3(0)(\xi^1, \xi^2) \neq 0$ , this actually means that the whole line segment perpendicular to the midsurface through  $(\xi^1, \xi^2)$  attains this “velocity”. The mass of this line segment depends linearly on the thickness  $\tau$ , and so does the momentum. Since the momentum shall be a prescribed function  $g$ , the initial velocity has to be given as  $\frac{g}{\tau}$ .

If  $g \in \mathcal{V}$ , then  $\Phi_g|_{U_{\text{ad}}}$  is well-defined (because  $\tau \geq \tau_{\min} > 0$  for  $\tau \in U_{\text{ad}}$ ) and continuously Fréchet-differentiable with the derivative given as

$$\Phi'_g(\tau)q = -\frac{g}{\tau^2}q,$$

see the discussion on Nemytskii operators in Lemma 4.6.

#### 4.1.2. Existence and uniqueness of a solution

The first question to deal with when regarding the dynamic state equation (4.1.5) is the existence and uniqueness of a solution. Luckily, the theory of hyperbolic PDE provides a theorem for this purpose which can be found in [Wlo87, p. 437]. The theorem itself is stated for equations involving a time-dependent bilinear form (which provides more generality), but does not include operators in front of  $(\ddot{u}, \ddot{\theta})$ . This can either be bypassed by showing that  $M_{(\tau)}$  is invertible and that  $M_{(\tau)}^{-1}L_{(\tau)}$  and  $M_{(\tau)}^{-1}F(t)$  fulfill the conditions from theorem 29.1 in [Wlo87] – or the theorem is restated and the proof is adapted for the purposes of equation (4.1.5). Here the latter way is chosen.

In preparation, some properties of the operator  $M_{(\tau)}$  are shown.

**Lemma 4.1.** Let  $\tau \in U_b$  be given. Then the operator  $M_{(\tau)} \in L(\mathcal{H}, \mathcal{H})$  is self-adjoint and if in addition there is  $c_{(\tau)}$  such that

$$\begin{aligned} \Phi_1(\tau) &\geq c_{(\tau)} > 0, \\ \Phi_1(\tau)\Phi_3(\tau) - \Phi_2(\tau)^2 &\geq c_{(\tau)} > 0 \end{aligned}$$

a.e. in  $\mathcal{S}$ , then it holds

$$(M_{(\tau)}(v, \psi), (v, \psi))_{\mathcal{H}} \geq c_{(\tau)} \|(v, \psi)\|_{\mathcal{H}}^2$$

for all  $(v, \psi) \in \mathcal{H}$ .

*Proof.* First, the expression for  $M_{(\tau)}$  is reformulated in local coordinates. For this, fix a covariant basis  $\tilde{a}_i(\xi^1, \xi^2)$ ,  $i = 1, \dots, 5$ , for  $(\xi^1, \xi^2) \in \mathcal{S}$  of the space

$$T_{(\xi^1, \xi^2)}(\mathcal{S}) \oplus \text{span}\{a_3(\xi^1, \xi^2)\} \oplus T_{(\xi^1, \xi^2)}(\mathcal{S}),$$

namely

$$\begin{aligned} \tilde{a}_1 &= a_1 \oplus 0 \oplus 0, & \tilde{a}_2 &= a_2 \oplus 0 \oplus 0, & \tilde{a}_3 &= 0 \oplus a_3 \oplus 0, \\ \tilde{a}_4 &= 0 \oplus 0 \oplus a_1, & \tilde{a}_5 &= 0 \oplus 0 \oplus a_2, \end{aligned}$$

and let  $\tilde{a}^i(\xi^1, \xi^2)$ ,  $i = 1, \dots, 5$  be the corresponding contravariant basis. Then define the second order tensor field

$$\tilde{M}_{(\tau)} = \sum_{i,j=1}^5 (\tilde{M}_{(\tau)})_i^j \tilde{a}^i \otimes \tilde{a}_j,$$

where its components are given by

$$(\tilde{M}_{(\tau)})_i^j := \begin{pmatrix} \Phi_1(\tau) & 0 & 0 & \Phi_2(\tau) & 0 \\ 0 & \Phi_1(\tau) & 0 & 0 & \Phi_2(\tau) \\ 0 & 0 & \Phi_1(\tau) & 0 & 0 \\ \Phi_2(\tau) & 0 & 0 & \Phi_3(\tau) & 0 \\ 0 & \Phi_2(\tau) & 0 & 0 & \Phi_3(\tau) \end{pmatrix}.$$

Note that  $\tilde{M}_{(\tau)}$  is symmetric and that

$$M_{(\tau)}(u, \theta)(\xi^1, \xi^2) = \tilde{M}_{(\tau)}(\xi^1, \xi^2) \begin{pmatrix} u(\xi^1, \xi^2) \\ u_3(\xi^1, \xi^2) \\ \theta(\xi^1, \xi^2) \end{pmatrix}.$$

It holds for all  $(u, \theta), (v, \psi) \in \mathcal{H}$

$$\begin{aligned} (M_{(\tau)}(u, \theta), (v, \psi))_{\mathcal{H}} &= \int_{\mathcal{S}} \tilde{M}_{(\tau)} \begin{pmatrix} u \\ u_3 \\ \theta \end{pmatrix} \cdot \begin{pmatrix} v \\ v_3 \\ \psi \end{pmatrix} dS = \int_{\mathcal{S}} \begin{pmatrix} u \\ u_3 \\ \theta \end{pmatrix} \cdot \tilde{M}_{(\tau)} \begin{pmatrix} v \\ v_3 \\ \psi \end{pmatrix} dS \\ &= ((u, \theta), M_{(\tau)}(v, \psi))_{\mathcal{H}}, \end{aligned}$$

hence  $M_{(\tau)}$  is self-adjoint. For the second claim of the lemma, look at the eigenvalues  $\lambda_i$ ,  $i = 1, \dots, 5$ , of  $\tilde{M}_{(\tau)}$  which are given by

$$\begin{aligned} \lambda_1 &= \Phi_1(\tau), \\ \lambda_{2/3} &= \frac{1}{2} \left( \Phi_1(\tau) + \Phi_3(\tau) + \sqrt{(\Phi_1(\tau) - \Phi_3(\tau))^2 + 4\Phi_2(\tau)^2} \right), \\ \lambda_{4/5} &= \frac{1}{2} \left( \Phi_1(\tau) + \Phi_3(\tau) - \sqrt{(\Phi_1(\tau) - \Phi_3(\tau))^2 + 4\Phi_2(\tau)^2} \right). \end{aligned}$$

From the conditions of the lemma, it holds  $\lambda_1 \geq c_{(\tau)} > 0$  a.e. in  $\mathcal{S}$ . From the claim

$$\Phi_1(\tau)\Phi_3(\tau) - \Phi_2(\tau)^2 > 0 \quad \text{a.e. in } \mathcal{S}$$

and the fact  $\Phi_1(\tau) > 0$ , it follows  $\Phi_3(\tau) > 0$  and therefore

$$\Phi_1(\tau) + \Phi_3(\tau) > 0 \quad \text{a.e. in } \mathcal{S},$$

with the consequence that  $\lambda_{2/3} \geq c_{(\tau)} > 0$  a.e. in  $\mathcal{S}$ . Finally, a short calculation gives

$$\begin{aligned} \Phi_1(\tau)\Phi_3(\tau) - \Phi_2(\tau)^2 &\geq c > 0 \\ \implies \Phi_1(\tau) + \Phi_3(\tau) &\geq \sqrt{(\Phi_1(\tau) - \Phi_3(\tau))^2 + 4\Phi_2(\tau)^2} + c_{(\tau)}, \end{aligned}$$

hence  $\lambda_{4/5} \geq c_{(\tau)} > 0$  a.e. in  $\mathcal{S}$ . This shows that the matrix  $(\tilde{M}_{(\tau)})_i^j(\xi^1, \xi^2)$  is positive definite a.e. in  $\mathcal{S}$  and that the eigenvalues do not get arbitrarily close to zero. It follows

$$\begin{aligned} (M_{(\tau)}(v, \psi), (v, \psi)) &= \int_{\mathcal{S}} \tilde{M}_{(\tau)} \begin{pmatrix} v \\ v_3 \\ \psi \end{pmatrix} \cdot \begin{pmatrix} v \\ v_3 \\ \psi \end{pmatrix} dS \geq c_{(\tau)} \int_{\mathcal{S}} \begin{pmatrix} v \\ v_3 \\ \psi \end{pmatrix} \cdot \begin{pmatrix} v \\ v_3 \\ \psi \end{pmatrix} dS \\ &= c_{(\tau)} \|(v, \psi)\|_{\mathcal{H}}^2. \end{aligned}$$

□

**Remark 4.2** (Discussion and existence of a common constant). *The claims from the above Lemma 4.1,*

$$\begin{aligned} \Phi_1(\tau) &\geq c_{(\tau)} > 0, \\ \Phi_1(\tau)\Phi_3(\tau) - \Phi_2(\tau)^2 &\geq c_{(\tau)} > 0 \end{aligned}$$

a.e. in  $\mathcal{S}$  are not a significant restriction. They translate to

$$\begin{aligned} \rho\tau + \rho K \frac{\tau^3}{12} &\geq c_{(\tau)} > 0, \\ \frac{1}{12}\rho^2\tau^4 + \left(\frac{7}{360}K - \frac{H^2}{36}\right)\rho^2\tau^6 + \frac{K^2}{960}\rho^2\tau^8 &\geq c_{(\tau)} > 0 \end{aligned} \tag{4.1.6}$$

a.e. in  $\mathcal{S}$ . In both cases the leading coefficient in front of the smallest power is positive, so for sufficient small values of  $\tau_{\max}$  the terms will be positive for all  $\tau \in U_b$ . Moreover, they do not get arbitrarily close to zero since  $\tau \geq \tau_{\min} > 0$ . It is assumed in further text that the above condition (4.1.6) is always fulfilled.

Since the above result is valid for arbitrary  $\tau \in U_b$ , there is also a common constant  $c$  such that

$$\Phi_1(\tau) \geq c > 0, \quad \Phi_1(\tau)\Phi_3(\tau) - \Phi_2(\tau)^2 \geq c > 0$$

a.e. in  $\mathcal{S}$  for all  $\tau \in U_b$ .

**Remark 4.3** (Cylindrical case). *The above inequality (4.1.6) simplifies further in case of a parabolic midsurface where  $K = 0$ : Then it reads*

$$\frac{1}{12}\rho^2\tau^4 - \frac{H^2\rho^2\tau^6}{36} \geq c > 0$$

and is fulfilled if

$$0 < \tau_{\min} \leq \tau \leq \tau_{\max} < \frac{\sqrt{3}}{|H|}.$$

In case of a cylindrical midsurface with  $H = -\frac{1}{2R}$ , this reads

$$\frac{\tau}{2} \leq \sqrt{3}R$$

and is always fulfilled since it is assumed that  $\tau \ll R$  (see discussion in Subsection 2.2.1).

**Remark 4.4** (Square root of  $M_{(\tau)}$ ). *From the above lemma it follows in particular that  $M_{(\tau)}$  is a positive operator, because*

$$(M_{(\tau)}(v, \psi), (v, \psi))_{\mathcal{H}} \geq 0$$

for all  $(v, \psi) \in \mathcal{H}$ . This ensures the existence of a unique “square root” of  $M_{(\tau)}$ , i.e. an operator  $S_{(\tau)} =: \sqrt{M_{(\tau)}}$  with

$$M_{(\tau)} = S_{(\tau)}^2,$$

see [KA82, p. 170].

Now that the required properties of the operator  $M_{(\tau)}$  have been introduced and proven, a theorem about the existence and uniqueness of the solution to the dynamic state equation can be stated.

**Theorem 4.1.** Let  $\tau \in U_b$ ,  $F \in L^2(0, T; \mathcal{H})$  and  $\Phi_g(\tau) \in \mathcal{H}$  be given. Then there is a unique solution  $(u, \theta) \in W(0, T)$  to the dynamic state equation (4.1.5),

$$\begin{aligned} M_{(\tau)}(\ddot{u}(t), \ddot{\theta}(t)) + L_{(\tau)}(u(t), \theta(t)) &= F(t) \quad \text{in } \mathcal{V}^* \text{ for } t \in [0, T] \\ (u(0), \theta(0)) &= 0 \quad (\dot{u}(0), \dot{\theta}(0)) = \Phi_g(\tau). \end{aligned}$$

Furthermore, it holds

$$\|(u, \theta)\|_{W(0, T)}^2 \leq CT \left( \|\Phi_g(\tau)\|_{\mathcal{H}}^2 + \|f\|_{L^2(0, T; \mathcal{H})}^2 \right) \quad (4.1.7)$$

with a constant  $C$  independent of  $\tau \in U_b$ .

*Proof.* The proof is adapted from [Wlo87, pp. 437–442] and suited to the case of the operator  $M_{(\tau)}$  in front of  $(\ddot{u}(t), \ddot{\theta}(t))$ . So only the parts where adjustments are necessary are stated here, the rest is referenced to the above citation. The proof of the existence of a solution starts by supposing a set

$$\{w_n \mid n \in \mathbb{N}\}$$

which is linearly independent and total in  $\mathcal{V}$ . This is possible since  $\mathcal{V}$  is separable as a composition of  $H^1$  Sobolev spaces ([AF03, p. 61]). To shorten the notation, write  $y := (u, \theta)$ . The idea of the proof is to approximate the initial conditions and the solution by sequences in  $\mathcal{H}$  and  $\mathcal{V}$  and to show that a weak limit exists which solves equation (4.1.5).

In order to approximate the initial conditions, choose coefficients  $\alpha_{im}$  with

$$y_{1m} = \sum_{i=1}^m \alpha_{im} w_i$$

and  $y_{1m} \rightarrow \Phi_g(\tau)$  in  $\mathcal{H}$  as  $m \rightarrow \infty$ . This is possible since  $\mathcal{V}$  is continuous and dense in  $\mathcal{H}$ .

A sequence  $y_m(t)$  is sought which approximates the solution and is given in the form

$$y_m(t) = \sum_{i=1}^m \lambda_{im}(t) w_i$$

where the functions  $\lambda_{im}(t)$  are determined by the equations

$$\begin{aligned} \frac{d^2}{dt^2} (M_{(\tau)} y_m(t), w_j)_{\mathcal{H}} + a_{(\tau)}(y_m(t), w_j) &= (F(t), w_j)_{\mathcal{H}} \quad \text{for } 1 \leq j \leq m, \\ y_m(0) = 0 \quad \dot{y}_m(0) &= y_{1m}. \end{aligned} \quad (4.1.8)$$

This is a system of  $m$  linear ODEs for the unknown functions  $\lambda_{im}(t)$ , namely

$$\hat{M}_{(\tau)} \frac{d^2}{dt^2} \lambda_m(t) + \hat{A}_{(\tau)} \lambda_m(t) = \hat{F}(t), \quad (4.1.9)$$

where  $\lambda_m(t) := (\lambda_{1m}(t), \dots, \lambda_{mm}(t))^T$  and

$$\begin{aligned} (\hat{M}_{(\tau)})_{ij} &= (M_{(\tau)} w_j, w_i)_{\mathcal{H}}, \\ (\hat{A}_{(\tau)})_{ij} &= a_{(\tau)}(w_j, w_i), \\ (\hat{F}(t))_j &= (F(t), w_j)_{\mathcal{H}}, \end{aligned}$$

for  $i, j = 1, \dots, m$ . The matrix  $\hat{M}_{(\tau)}$  is positive definite, because for any  $v \in \mathbb{R}^m$ ,  $v \neq 0$  it is  $\sum_{i=1}^m v_i w_i =: \hat{v} \in \mathcal{V}$  and

$$v^T \hat{M}_{(\tau)} v = (M_{(\tau)} \hat{v}, \hat{v})_{\mathcal{H}} \geq c \|\hat{v}\|_{\mathcal{H}} > 0,$$

which follows from Lemma 4.1. The same is true for the matrix  $\hat{A}_{(\tau)}$ , because

$$v^T \hat{A}_{(\tau)} v = a_{(\tau)}(\hat{v}, \hat{v}) > c \|\hat{v}\|_{\mathcal{V}} > 0$$

which follows from the coercivity of  $a_{(\tau)}$ . Since both matrices are positive definite, they are in particular invertible and the system (4.1.9) has a unique solution ([Wal98, p. 162]). So the functions  $\lambda_{im}(t)$  and also  $y_m(t)$  are well-defined.

If the above equations (4.1.8) are multiplied by  $\dot{\lambda}_{jm}(t)$  and summed over  $j$ , they read

$$(M_{(\tau)} \dot{y}_m(t), \dot{y}_m(t))_{\mathcal{H}} + a_{(\tau)}(y_m(t), \dot{y}_m(t)) = (F(t), \dot{y}_m(t))_{\mathcal{H}}.$$

This is equivalent to

$$\frac{d}{dt} \left( \left\| \sqrt{M_{(\tau)}} \dot{y}_m(t) \right\|_{\mathcal{H}}^2 + a_{(\tau)}(y_m(t), y_m(t)) \right) = 2 (F(t), \dot{y}_m(t))_{\mathcal{H}},$$

where the existence of  $\sqrt{M_{(\tau)}}$  was discussed in Remark 4.4. Integration with respect to  $t$  (from 0 to  $t$ ) then leads to

$$\left\| \sqrt{M_{(\tau)}} \dot{y}_m(t) \right\|_{\mathcal{H}}^2 + a_{(\tau)}(y_m(t), y_m(t)) = \left\| \sqrt{M_{(\tau)}} \dot{y}_m(0) \right\|_{\mathcal{H}}^2 + 2 \int_0^t (F(s), \dot{y}_m(s))_{\mathcal{H}} ds. \quad (4.1.10)$$

The left-hand-side can be bounded from below by using the coercivity of  $a_{(\tau)}$  and Lemma 4.1 with Remark 4.2,

$$\begin{aligned} & \left\| \sqrt{M_{(\tau)}} \dot{y}_m(t) \right\|_{\mathcal{H}}^2 + a_{(\tau)}(y_m(t), y_m(t)) = \\ & (M_{(\tau)} \dot{y}_m(t), \dot{y}_m(t))_{\mathcal{H}} + a_{(\tau)}(y_m(t), y_m(t)) > c \left( \|\dot{y}_m(t)\|_{\mathcal{H}}^2 + \|y_m(t)\|_{\mathcal{V}}^2 \right). \end{aligned} \quad (4.1.11)$$

Note that the constant  $c$  is independent from  $\tau \in U_b$  following Remark 4.2. The right-hand-side is bounded from above via

$$\begin{aligned} & \left\| \sqrt{M_{(\tau)}} \dot{y}_m(0) \right\|_{\mathcal{H}}^2 + 2 \int_0^t (F(s), \dot{y}_m(s))_{\mathcal{H}} \, ds \\ & \leq \left\| \sqrt{M_{(\tau)}} y_{1m} \right\|_{\mathcal{H}}^2 + 2 \int_0^t \|F(s)\|_{\mathcal{H}} \|\dot{y}_m(s)\|_{\mathcal{H}} \, ds \\ & \leq C_1 \|y_{1m}\|_{\mathcal{H}}^2 + \int_0^t \|F(s)\|_{\mathcal{H}}^2 \, ds + \int_0^t \|\dot{y}_m(s)\|_{\mathcal{H}}^2 \, ds \end{aligned}$$

where the second last “ $\leq$ ” is due to inequality  $2|ab| \leq a^2 + b^2$ .

By setting

$$\beta_m(t) =: \|\dot{y}_m(t)\|_{\mathcal{H}}^2 + \|y_m(t)\|_{\mathcal{V}}^2,$$

both estimates can be combined which leads to

$$c\beta_m(t) \leq C_1 \|y_{1m}\|_{\mathcal{H}}^2 + \int_0^t \|F(s)\|_{\mathcal{H}}^2 \, ds + \int_0^t \beta_m(s) \, ds.$$

Now, Grönwall’s lemma is applied ([Wlo87, p. 436]) which shows that  $\beta_m(t)$  is bounded. This ensures that the sequence  $y_m$  is bounded in  $L^2(0, T; \mathcal{V})$  and  $\dot{y}_m$  is bounded in  $L^2(0, T; \mathcal{H})$  and weak limits of both sequences can be obtained. The next steps include showing that those weak limits indeed solve equation (4.1.5) as well as obtaining the estimate (4.1.7) and can be taken directly from [Wlo87, pp. 441–442] (only adding the continuity of  $M_{(\tau)}$ ), so they are not reproduced here. The same holds for the proof of the uniqueness of the solution which can be reproduced from [Wlo87, pp. 437–439] making the same modifications as above.  $\square$

The proof of existence and uniqueness of a solution to the dynamic state equation is the essential step for further considerations. As in the steady-state case, it is possible to define a control-to-state operator.

**Definition 4.2** (Control-to-state operator). The operator  $G : U_b \rightarrow W(0, T)$  which maps the thickness  $\tau$  to the solution  $(u, \theta)$  of (4.1.5) is called *control-to-state operator*.

Note that the notation does not distinguish between the control-to-state operator from the steady-state case and the dynamic case as long as it is clear from the context which operator is meant.

The above defined operator is the object of detailed investigation now. It is studied with regard to its continuity and differentiability properties.

### 4.1.3. Continuous dependence on the thickness

The aim of this subsection is to show the continuity of the solution to the dynamic state equation with respect to the thickness as it was done in Subsection 3.1.2 for the steady-state case. The sketch will be as follows: First, given a converging thickness sequence, show that the corresponding sequence of solutions has a weakly convergent subsequence. Second, prove that the weak limit is solution to the thickness limit. Third, verify that the convergence is also strong.

The first lemma establishes some additional properties of  $M_{(\tau)}$ :

**Lemma 4.2.** It holds:

1. For all  $\tau \in U_b$  there exist common constants  $c > 0$  and  $C > 0$  such that

$$\begin{aligned} (M_{(\tau)}(v, \psi), (v, \psi))_{\mathcal{H}} &\geq c \|(v, \psi)\|_{\mathcal{H}}^2, \\ \|M_{(\tau)}(v, \psi)\|_{\mathcal{H}} &\leq C \|(v, \psi)\|_{\mathcal{H}} \end{aligned}$$

for all  $(v, \psi) \in \mathcal{H}$ .

2.  $M_{(\tau)}$  depends continuously on  $\tau$ , i.e.

$$\|M_{(\tau_n)} - M_{(\tau)}\|_{L(\mathcal{H}, \mathcal{H})} \rightarrow 0$$

as  $\tau_n \rightarrow \tau$  in  $L^\infty(\mathcal{S})$ .

3.  $M_{(\tau)}$  is invertible with a continuous inverse  $M_{(\tau)}^{-1}$ . There exists  $c > 0$  independent of  $\tau \in U_b$  such that

$$\|M_{(\tau)}^{-1}(v, \psi)\|_{\mathcal{H}} \leq c \|(v, \psi)\|_{\mathcal{H}}.$$

*Proof.* 1. The existence of the common continuity constant can be established via the common bound on  $\|\Phi_i(\tau)\|_{L^\infty(\mathcal{S})}$ . It is

$$\begin{aligned} \|\Phi_1(\tau)\|_{L^\infty(\mathcal{S})} &= \left\| \rho\tau + \rho K \frac{\tau^3}{12} \right\|_{L^\infty(\mathcal{S})} \leq \rho\tau_{\max} + \|\rho K\|_{L^\infty(\mathcal{S})} \frac{\tau_{\max}^3}{12}, \\ \|\Phi_2(\tau)\|_{L^\infty(\mathcal{S})} &= \left\| -2\rho H \frac{\tau^3}{12} \right\|_{L^\infty(\mathcal{S})} \leq 2\|\rho H\|_{L^\infty(\mathcal{S})} \frac{\tau_{\max}^3}{12}, \\ \|\Phi_3(\tau)\|_{L^\infty(\mathcal{S})} &= \left\| \rho \frac{\tau^3}{12} + \rho K \frac{\tau^5}{80} \right\|_{L^\infty(\mathcal{S})} \leq \rho \frac{\tau_{\max}^3}{12} + \|\rho K\|_{L^\infty(\mathcal{S})} \frac{\tau_{\max}^5}{80}. \end{aligned} \tag{4.1.12}$$

By taking the maximum of all three bounds and using the fact that the norm of  $M_{(\tau)}$  only depends on  $\|\Phi_i(\tau)\|_{L^\infty(\mathcal{S})}$ , a common constant  $C > 0$  with

$$\|M_{(\tau)}(v, \psi)\|_{\mathcal{H}} \leq C \|(v, \psi)\|_{\mathcal{H}}$$

for all  $\tau \in U_b$  and  $(v, \psi) \in \mathcal{H}$  is established. The existence of a common coercivity constant follows from Lemma 4.1 and the subsequent Remark 4.2.

2. The continuous dependence of  $M_{(\tau)}$  on  $\tau$  is due to the continuous dependence of the operators  $\Phi_i$  on  $\tau$ , which follows from the discussion on Nemytskii operators in Lemma 4.6.
3. A direct calculation gives

$$M_{(\tau)}^{-1}(v, \psi) = \frac{1}{\Phi_1(\tau)\Phi_3(\tau) - \Phi_2(\tau)^2} \begin{pmatrix} \Phi_3(\tau)v - \Phi_2(\tau)\psi \\ \frac{\Phi_1(\tau)\Phi_3(\tau) - \Phi_2(\tau)^2}{\Phi_1(\tau)} v_3 \\ -\Phi_2(\tau)v + \Phi_1(\tau)\psi \end{pmatrix}.$$

In view of the uniform bounds on  $\Phi_i(\tau)$  from (4.1.12), this is again a linear continuous operator  $M_{(\tau)}^{-1} \in L(\mathcal{H}, \mathcal{H})$  with

$$\left\| M_{(\tau)}^{-1}(v, \psi) \right\|_{\mathcal{H}} \leq c \|(v, \psi)\|_{\mathcal{H}}$$

where  $c > 0$  is independent from  $\tau \in U_b$ . Because of the continuous embedding  $\mathcal{H} \hookrightarrow \mathcal{V}^*$ , the inverse  $M_{(\tau)}^{-1}$  can be extended to a continuous operator in  $L(\mathcal{V}^*, \mathcal{V}^*)$ , i.e.

$$\left\| M_{(\tau)}^{-1}(v, \psi) \right\|_{\mathcal{V}^*} \leq c \|(v, \psi)\|_{\mathcal{V}^*}.$$

□

Let a sequence  $\tau_n \in U_b$  with  $\tau_n \rightarrow \bar{\tau} \in U_b$  be given while the applied load  $f$  and the continuous operator  $\Phi_g : L^\infty(\mathcal{S}) \rightarrow \mathcal{H}$  from the initial conditions is fixed. Denote the corresponding sequence of solutions to (4.1.5) by

$$y_n := (u_n, \theta_n) := G(\tau_n) \in W(0, T).$$

The abbreviation using  $y$  is to shorten the notation.

**Lemma 4.3.** The sequence of solutions  $y_n$  is bounded in  $W(0, T)$ . Moreover,  $\ddot{y}_n$  is bounded in  $L^2(0, T; \mathcal{V}^*)$ .

*Proof.* By using the estimate from Theorem 4.1, it is

$$\begin{aligned} \|y_n\|_{W(0, T)}^2 &\leq C \left( \|\Phi_g(\tau_n)\|_{\mathcal{H}}^2 + \|F\|_{L^2(0, T; \mathcal{H})}^2 \right) \\ &\leq C \left( \|\tau_n\|_{L^\infty(\mathcal{S})}^2 + \|F\|_{L^2(0, T; \mathcal{H})}^2 \right) \\ &\leq C \left( \tau_{\max}^2 + \|F\|_{L^2(0, T; \mathcal{H})}^2 \right). \end{aligned} \tag{4.1.13}$$

The bound for  $\ddot{y}_n$  can be achieved using the inverse of  $M_{(\tau)}$ ,

$$\begin{aligned} \|\ddot{y}_n(t)\|_{\mathcal{V}^*} &= \left\| M_{(\tau_n)}^{-1} (F(t) - L_{(\tau_n)} y_n(t)) \right\|_{\mathcal{V}^*} \\ &\leq \left\| M_{(\tau_n)}^{-1} \right\|_{L(\mathcal{V}^*, \mathcal{V}^*)} \|L_{(\tau_n)}\|_{L(\mathcal{V}, \mathcal{V}^*)} \|y_n(t)\|_{\mathcal{V}} + \left\| M_{(\tau_n)}^{-1} \right\|_{L(\mathcal{H}, \mathcal{H})} \|F(t)\|_{\mathcal{H}}, \end{aligned}$$

so

$$\begin{aligned} & \|\ddot{y}_n\|_{L^2(0,T;\mathcal{V}^*)} \leq \\ & 2 \left\| M_{(\tau_n)}^{-1} \right\|_{L(\mathcal{V}^*,\mathcal{V}^*)}^2 \|L_{(\tau_n)}\|_{L(\mathcal{V},\mathcal{V}^*)}^2 \|y_n\|_{L^2(0,T;\mathcal{V})}^2 + 2 \left\| M_{(\tau_n)}^{-1} \right\|_{L(\mathcal{H},\mathcal{H})}^2 \|F\|_{L^2(0,T;\mathcal{H})}^2. \end{aligned}$$

All terms on the right-hand-side can be bounded independently of  $n$ :  $\left\| M_{(\tau_n)}^{-1} \right\|$  due to Lemma 4.2 point 3,  $\|L_{(\tau_n)}\|$  because of Lemma 3.1 and  $\|y_n\|$  from equation (4.1.13).  $\square$

By use of Theorem A.4 there are weakly convergent subsequences

$$\begin{aligned} y_{n_k} & \rightharpoonup \bar{y} \text{ in } W(0,T), \\ \dot{y}_{n_k} & \rightharpoonup \bar{y}_2 \text{ in } L^2(0,T;\mathcal{V}^*). \end{aligned} \tag{4.1.14}$$

Similar to [Wlo87, pp. 392-393], it can be shown that

$$\ddot{y} = \bar{y}_2, \tag{4.1.15}$$

which relies on the Gelfand-triple property of  $\mathcal{V} \hookrightarrow \mathcal{H} \hookrightarrow \mathcal{V}^*$ .

The next thing to show is that the weak limit  $\bar{y}$  is the solution of (4.1.5) to  $\bar{\tau}$ .

**Lemma 4.4.** The weak limit  $\bar{y}$  is solution of (4.1.5) to  $\bar{\tau}$ , i.e.

$$\begin{aligned} M_{(\bar{\tau})}\ddot{\bar{y}}(t) + L_{(\bar{\tau})}\bar{y}(t) & = F(t) \quad \text{in } \mathcal{V}^* \text{ for } t \in [0,T] \\ \bar{y}(0) = 0 \quad \dot{\bar{y}}(0) & = \Phi_g(\bar{\tau}). \end{aligned}$$

*Proof.* Since  $y_{n_k}(0) = 0$ , this is also true for  $\bar{y}(0)$ . For the initial condition on the first derivative, it holds

$$\dot{y}_{n_k}(0) = \Phi_g(\tau_{n_k}).$$

Because  $\Phi_g$  is continuous, it holds  $\Phi_g(\tau_{n_k}) \rightarrow \Phi_g(\bar{\tau})$  in  $\mathcal{H}$ -strong. Taking the limit in the above equation yields

$$\dot{\bar{y}}(0) = \Phi_g(\bar{\tau}),$$

so the initial conditions are fulfilled by  $\bar{y}$ . For a fixed time  $t \in [0,T]$  consider

$$\left( M_{(\tau_{n_k})}\ddot{y}_{n_k}(t), z \right)_{\mathcal{H}} + a_{(\tau_{n_k})}(y_{n_k}(t); z) = (F(t), z)_{\mathcal{H}}, \tag{4.1.16}$$

which is valid for all  $z = (v, \psi) \in \mathcal{V}$  because  $y_{n_k}$  is solution to  $\tau_{n_k}$ . In Lemma 3.3 from the investigation of the steady-state case, it was shown that

$$a_{(\tau_{n_k})}(y_{n_k}(t); z) \rightarrow a_{(\bar{\tau})}(\bar{y}(t); z).$$

For the first part of (4.1.16) it is

$$M_{(\tau_{n_k})}\ddot{y}_{n_k}(t) = \left( M_{(\tau_{n_k})} - M_{(\bar{\tau})} \right) \ddot{y}_{n_k}(t) + M_{(\bar{\tau})} (\ddot{y}_{n_k}(t) - \ddot{\bar{y}}(t)) + M_{(\bar{\tau})}\ddot{\bar{y}}(t).$$

As  $n_k \rightarrow \infty$ , the first summand converges to 0 because of Lemma 4.2, point 2, and the second one converges weakly to zero because of (4.1.14) and (4.1.15). So it holds

$$\left( M_{(\tau_{n_k})} \ddot{y}_{n_k}(t), z \right)_{\mathcal{H}} \rightarrow \left( M_{(\bar{\tau})} \ddot{y}(t), z \right)_{\mathcal{H}}.$$

Taking the limit in (4.1.16) then leads to

$$\left( M_{(\bar{\tau})} \ddot{y}(t), z \right)_{\mathcal{H}} + a_{(\bar{\tau})}(\bar{y}(t); z) = (F(t), z)_{\mathcal{H}}$$

for all  $z \in \mathcal{V}$  and  $t \in [0, T]$ , which means that  $\bar{y} = (\bar{u}, \bar{\theta})$  is solution to  $\bar{\tau}$ .  $\square$

**Remark 4.5.** *Because of the uniqueness of the state equation solution, every convergent subsequence of  $y_n$  converges to the above found  $\bar{y}$ . So the whole sequence converges weakly in  $W(0, T)$ , i.e.*

$$(u_n, \theta_n) = y_n \rightharpoonup \bar{y} = (\bar{u}, \bar{\theta}) \quad \text{in } W(0, T).$$

Now, only the last step is missing – showing that the weak limit  $\bar{y}$ , which solves the state equation for the “limit thickness”, is also a strong limit.

**Lemma 4.5.** The convergence established in the above Lemma 4.4 and Remark 4.5 is also strong, i.e.

$$\|y_n - \bar{y}\|_{W(0, T)} \rightarrow 0.$$

The same is true for the second time derivatives, i.e.

$$\|\ddot{y}_n - \ddot{y}\|_{L^2(0, T; \mathcal{V}^*)} \rightarrow 0.$$

*Proof.* Consider the equations for  $y_n$  and  $\bar{y}$ :

$$\begin{aligned} M_{(\tau_n)} \ddot{y}_n(t) + L_{(\tau_n)} y_{n_k}(t) &= F(t), \\ M_{(\bar{\tau})} \ddot{y}(t) + L_{(\bar{\tau})} \bar{y}(t) &= F(t) \end{aligned}$$

for all  $t \in [0, T]$ . Subtraction of both equations and insertion of

$$M_{(\bar{\tau})} \ddot{y}_n(t) + L_{(\bar{\tau})} y_n(t)$$

leads to

$$0 = M_{(\bar{\tau})}(\ddot{y}_n(t) - \ddot{y}(t)) + (M_{(\tau_n)} - M_{(\bar{\tau})})\ddot{y}_n(t) + L_{(\bar{\tau})}(y_n(t) - \bar{y}(t)) + (L_{(\tau_n)} - L_{(\bar{\tau})})y_n(t).$$

From the proof of Lemma 3.4 it follows that the last summand converges to 0 as  $n \rightarrow \infty$ . Looking at the second summand reveals

$$\|(M_{(\tau_n)} - M_{(\bar{\tau})})\ddot{y}_n(t)\|_{\mathcal{V}^*} \leq \|M_{(\tau_n)} - M_{(\bar{\tau})}\|_{L(\mathcal{V}^*, \mathcal{V}^*)} \|\ddot{y}_n(t)\|_{\mathcal{V}^*} \rightarrow 0$$

by the extension of Lemma 4.2, point 2, to the  $\mathcal{V}^*$ -case. Taking the limit  $n \rightarrow \infty$  and using the continuity of  $M_{(\bar{\tau})}$  and  $L_{(\bar{\tau})}$  shows

$$0 = \lim_{n \rightarrow \infty} M_{(\bar{\tau})}(\ddot{y}_n(t) - \ddot{y}(t)) + \lim_{n \rightarrow \infty} L_{(\bar{\tau})}(y_n(t) - \bar{y}(t)).$$

This gives

$$\lim_{n \rightarrow \infty} \|y_n - \bar{y}\|_{W(0,T)} = 0. \quad (4.1.17)$$

To obtain the result for the second time derivative, consider

$$\begin{aligned} & \|\ddot{y}_n - \ddot{\bar{y}}\|_{L^2(0,T;\mathcal{V}^*)}^2 = \\ &= \int_0^T \left\| M_{(\bar{\tau})}^{-1} L_{(\bar{\tau})} y_{(\bar{\tau})}(t) - M_{(\tau_n)}^{-1} L_{(\tau_n)} y_{(\tau_n)}(t) \right\|_{\mathcal{V}^*}^2 dt \\ &= \int_0^T \left\| M_{(\bar{\tau})}^{-1} L_{(\bar{\tau})} (y_{(\bar{\tau})}(t) - y_{(\tau_n)}(t)) + (M_{(\bar{\tau})}^{-1} L_{(\bar{\tau})} - M_{(\tau_n)}^{-1} L_{(\tau_n)}) y_{(\tau_n)}(t) \right\|_{\mathcal{V}^*}^2 dt \\ &\leq 2 \left\| M_{(\bar{\tau})}^{-1} L_{(\bar{\tau})} \right\|_{L(\mathcal{V},\mathcal{V}^*)}^2 \|y_{(\bar{\tau})} - y_{(\tau_n)}\|_{L^2(0,T;\mathcal{V})}^2 + \\ &\quad 2 \left\| M_{(\bar{\tau})}^{-1} L_{(\bar{\tau})} - M_{(\tau_n)}^{-1} L_{(\tau_n)} \right\|_{L(\mathcal{V},\mathcal{V}^*)}^2 \|y_{(\tau_n)}\|_{L^2(0,T;\mathcal{V})}^2 \\ &\rightarrow 0, \end{aligned}$$

where the first summand converges due to (4.1.17) and the second term due to the continuous dependence of  $L_{(\tau)}$  and  $M_{(\tau)}^{-1}$  on  $\tau$  (see Subsection 3.1.2 and the expression for  $M_{(\tau)}^{-1}$  in the proof of Lemma 4.2, point 3).  $\square$

Now the main result of this subsection can be formulated:

**Theorem 4.2.** The control-to-state operator  $G : U_b \rightarrow W(0, T)$  is continuous.

*Proof.* From the above Lemma 4.3 through 4.5, it follows for a sequence of thicknesses  $\tau_n \rightarrow \bar{\tau}$  in  $U_b$  that for the corresponding solutions  $G(\tau_n) = (u_n, \theta_n)$  and  $G(\bar{\tau}) = (\bar{u}, \bar{\theta})$

$$(u_n, \theta_n) \rightarrow (\bar{u}, \bar{\theta})$$

strong in  $W(0, T)$ .  $\square$

#### 4.1.4. Differentiable dependence on the thickness

After the continuity of the control-to-state operator is clarified, its differentiability comes into focus. As in the steady-state case, the differentiability is even more important for numerical calculations, since those methods rely on an accurate first derivative. Furthermore, necessary conditions are stated with help of the reduced objective gradient.

The sketch of this subsection is as in Subsection 3.1.3: First, investigate the directional derivative, in particular show the existence of the limit, and then clarify the linearity and the continuity of the derivative.

In preparation, the differentiability of  $M_{(\tau)}$  with respect to  $\tau$  is shown. For this purpose, it is helpful to consider the operators  $\Phi_i$  from (4.1.2) as Nemytskii operators.

**Definition 4.3** (Nemytskii operator [Trö10, p. 197]). Let  $E \subset \mathbb{R}^m$  be a bounded and measurable set, and let  $\varphi = \varphi(x, y) : E \times \mathbb{R} \rightarrow \mathbb{R}$  be a function. The mapping  $\Phi$  given by

$$\Phi(y) := \varphi(\cdot, y(\cdot)),$$

which assigns to a function  $y : E \rightarrow \mathbb{R}$  the function  $z : E \rightarrow \mathbb{R}$ ,  $z(x) = \varphi(x, y(x))$ , is called a *Nemytskii operator*.

The operators  $\Phi_i$  from (4.1.2) are Nemytskii operators that are associated in the above sense with

$$\begin{aligned} \varphi_1(\xi^1, \xi^2, y) &= \rho y + \rho K(\xi^1, \xi^2) \frac{y^3}{12}, \\ \varphi_2(\xi^1, \xi^2, y) &= -2\rho H(\xi^1, \xi^2) \frac{y^3}{12}, \\ \varphi_3(\xi^1, \xi^2, y) &= \rho \frac{y^3}{12} + \rho K(\xi^1, \xi^2) \frac{y^5}{80}. \end{aligned} \tag{4.1.18}$$

The differentiability of Nemytskii operators has been well studied in the literature. Under proper conditions on  $\varphi$ , they are continuously Fréchet-differentiable.

**Definition 4.4** (Continuously Fréchet-differentiable [Trö10, p. 203]). Let  $X, Y$  be Banach spaces,  $U \subseteq X$  open,  $U \neq \emptyset$  and  $f : U \rightarrow Y$ . If the Fréchet derivative of  $f$  exists in an open neighborhood  $V$  of  $x$  and if the mapping  $x \mapsto f'(x)$ ,  $V \rightarrow L(X, Y)$  is continuous at  $x$ , i.e.

$$\|\bar{x} - x\|_X \rightarrow 0 \Rightarrow \|f'(\bar{x}) - f'(x)\|_{L(X, Y)} \rightarrow 0,$$

then  $f$  is *continuously Fréchet-differentiable* at  $x$ .

**Lemma 4.6** (Differentiability of Nemytskii operators [Trö10, pp. 199–203]). Suppose that the function  $\varphi$  is measurable with respect to  $x \in E$  for every  $y \in \mathbb{R}$  and differentiable with respect to  $y$  for almost every  $x \in E$ . Moreover, let the boundedness condition

$$\frac{\partial \varphi}{\partial y}(x, 0) \leq K \text{ for a.e. } x \in E$$

for a constant  $K$  and the local Lipschitz condition

$$\left| \frac{\partial \varphi}{\partial y}(x, y_1) - \frac{\partial \varphi}{\partial y}(x, y_2) \right| \leq L(M) |y_1 - y_2|$$

for all  $y_1, y_2 \in \mathbb{R}$  with  $|y_i| \leq M$ ,  $i = 1, 2$  be fulfilled. Then the Nemytskii operator  $\Phi : L^\infty(E) \rightarrow L^\infty(E)$  associated with  $\varphi$  is continuously Fréchet-differentiable and it is

$$(\Phi'(y)q)(x) = \frac{\partial \varphi}{\partial y}(x, y(x))q(x) \tag{4.1.19}$$

for almost every  $x \in E$  and all  $q \in L^\infty(E)$ .

The above lemma can be used to make a differentiability statement on  $M_{(\tau)}$ :

**Lemma 4.7.** The operator  $M_{(\tau)} \in L(\mathcal{H}, \mathcal{H})$  is continuously Fréchet-differentiable with respect to  $\tau$ , i.e. the mapping

$$U_b \rightarrow L(\mathcal{H}, \mathcal{H}), \quad \tau \mapsto M_{(\tau)}$$

has a Fréchet-derivative for all  $\tau \in U_b$ . The derivative  $M'_{(\tau)} : U_b \rightarrow L(\mathcal{H}, \mathcal{H})$  is given by

$$(M'_{(\tau)}q)(u, \theta) = \begin{pmatrix} \Phi'_1(\tau)qu + \Phi'_2(\tau)q\theta \\ \Phi'_1(\tau)qu_3 \\ \Phi'_2(\tau)qu + \Phi'_3(\tau)q\theta \end{pmatrix},$$

where

$$\Phi'_1(\tau)q = \left( \rho + \rho K \frac{\tau^2}{4} \right) q, \quad \Phi'_2(\tau)q = -2\rho H \frac{\tau^2}{4} q, \quad \Phi'_3(\tau)q = \left( \rho \frac{\tau^2}{4} + \rho K \frac{\tau^4}{16} \right) q. \quad (4.1.20)$$

*Proof.* The functions  $\varphi_i$ ,  $i = 1, 2, 3$  from (4.1.18) that are associated with  $\Phi_i$ ,  $i = 1, 2, 3$  are polynomials in  $y$ . So they are differentiable with respect to  $y$  and fulfill the local Lipschitz condition. Moreover, they are measurable in  $(\xi^1, \xi^2)$ , because so are the Gaussian and mean curvature  $K$  and  $H$ . Also, the boundedness condition is fulfilled because  $\varphi_i(\xi^1, \xi^2; 0) = 0$ . Applying Lemma 4.6 yields that the  $\Phi_i$  are continuously Fréchet-differentiable, where the derivative is given by (4.1.19). A direct calculation gives the expressions for  $\Phi'_i(\tau)q$  from (4.1.20).

To obtain the differentiability of  $M_{(\tau)}$ , consider the mapping  $\tau \mapsto M_{(\tau)}$  as a composition

$$L^\infty(\mathcal{S}) \rightarrow L^\infty(\mathcal{S})^3 \rightarrow L(\mathcal{H}, \mathcal{H}), \quad \tau \mapsto (\Phi_1(\tau), \Phi_2(\tau), \Phi_3(\tau)) \mapsto M_{(\tau)}.$$

The first mapping is continuously Fréchet-differentiable because so are the  $\Phi_i$ ,  $i = 1, 2, 3$ . The second mapping is linear, which follows directly from the definition of  $M_{(\tau)}$ . Hence it is also continuously Fréchet-differentiable with the derivative given by the mapping itself (see [Trö10, p. 59]). By use of the chain rule ([Trö10, p. 60]), the composition is also continuously Fréchet-differentiable and the derivative  $M'_{(\tau)}$  is given by

$$(M'_{(\tau)}q)(u, \theta) = \begin{pmatrix} \Phi'_1(\tau)qu + \Phi'_2(\tau)q\theta \\ \Phi'_1(\tau)qu_3 \\ \Phi'_2(\tau)qu + \Phi'_3(\tau)q\theta \end{pmatrix}.$$

□

For the Gâteaux-differential of the control-to-state operator itself, the limit

$$\lim_{\lambda \rightarrow 0} \frac{G(\tau + \lambda q) - G(\tau)}{\lambda}$$

has to be investigated in  $W(0, T)$ . The sketch is as for the steady-state case: Show the boundedness of the above quotient when a sequence  $\lambda_n \rightarrow 0$  is inserted, find an equation that is fulfilled by the weak limit, and finally show the strong convergence.

Start by considering a fixed thickness  $\tau \in U_b$  and a direction  $q \in W^{1,\infty}(\mathcal{S})$ . For the solutions  $y_{(\tau)} := (u_{(\tau)}, \theta_{(\tau)}) := G(\tau)$  and  $y_{(\tau+\lambda q)} := (u_{(\tau+\lambda q)}, \theta_{(\tau+\lambda q)}) := G(\tau + \lambda q)$ , it holds

$$\begin{aligned} M_{(\tau+\lambda q)} \ddot{y}_{(\tau+\lambda q)}(t) + L_{(\tau+\lambda q)} y_{(\tau+\lambda q)}(t) &= F(t), \\ M_{(\tau)} \ddot{y}_{(\tau)}(t) + L_{(\tau)} y_{(\tau)}(t) &= F(t) \end{aligned} \quad (4.1.21)$$

in  $\mathcal{V}^*$  for all  $t \in [0, T]$  together with the initial conditions

$$y_{(\tau)}(0) = y_{(\tau+\lambda q)}(0) = 0, \quad \dot{y}_{(\tau)}(0) = \Phi_g(\tau), \quad \dot{y}_{(\tau+\lambda q)}(0) = \Phi_g(\tau + \lambda q). \quad (4.1.22)$$

Subtraction of both equations from (4.1.21), division by  $\lambda$ , insertion of

$$M_{(\tau)} \ddot{y}_{(\tau+\lambda q)}(t) + L_{(\tau)} y_{(\tau+\lambda q)}(t)$$

and using the linearity of  $L_{(\tau)}$  and  $M_{(\tau)}$  leads to

$$\begin{aligned} M_{(\tau)} \partial_t^2 \left( \frac{y_{(\tau+\lambda q)} - y_{(\tau)}}{\lambda} \right) (t) + L_{(\tau)} \left( \frac{y_{(\tau+\lambda q)} - y_{(\tau)}}{\lambda} \right) (t) &= \\ - \left( \frac{M_{(\tau+\lambda q)} - M_{(\tau)}}{\lambda} \right) \ddot{y}_{(\tau+\lambda q)}(t) - Z_\lambda(y_{(\tau+\lambda q)}(t); q), \end{aligned} \quad (4.1.23)$$

where  $Z_\lambda(y_{(\tau+\lambda q)}(t); q) \in \mathcal{V}^*$  is the linear form defined by

$$Z_\lambda(y_{(\tau+\lambda q)}(t); q)z := Z_\lambda(y_{(\tau+\lambda q)}(t); z; q)$$

with  $Z_\lambda : \mathcal{V} \times \mathcal{V} \times W^{1,\infty}(\mathcal{S}) \rightarrow \mathbb{R}$  on the right-hand-side given from the steady-state case (3.1.7). The reformulated initial conditions from (4.1.22) read

$$\left( \frac{y_{(\tau+\lambda q)} - y_{(\tau)}}{\lambda} \right) (0) = 0, \quad \partial_t \left( \frac{y_{(\tau+\lambda q)} - y_{(\tau)}}{\lambda} \right) (0) = \frac{\Phi_g(\tau + \lambda q) - \Phi_g(\tau)}{\lambda}. \quad (4.1.24)$$

Equation (4.1.23) together with the initial conditions (4.1.24) state a hyperbolic PDE. Now it is tempting to use the estimate from Theorem 4.1 to obtain an upper bound for the norm of the solutions

$$y_\lambda := \frac{y_{(\tau+\lambda q)} - y_{(\tau)}}{\lambda} \quad (4.1.25)$$

for  $\lambda \rightarrow 0$ . The problem with this is that the right-hand-side of (4.1.23) does not have the necessary regularity, because it is only

$$\ddot{y}_{(\tau+\lambda q)}(t), Z_\lambda(y_{(\tau+\lambda q)}; q(t)) \in L^2(0, T; \mathcal{V}^*),$$

whereas the theorem requires it to be in  $L^2(0, T; \mathcal{H})$ . This is the point mentioned in the previous Subsection 4.1.1, where additional regularity must be posed on the incoming force  $f$  and on the initial condition  $\Phi_g$ . Moreover, an extended estimate for the solution of the dynamic state equation (4.1.5) must be stated, which could not be found in the standard references [Wlo87; Eva10; Trö10], but should actually be known in the theory of hyperbolic PDE. However, since it cannot be cited, the proof is given in full length here.

**Theorem 4.3.** Let  $\tau \in U_b$ ,  $f \in H^1(0, T; \mathcal{V}^*)$  and  $\Phi_g(\tau) \in \mathcal{H}$  be given. Then it holds for the unique solution  $(u, \theta) \in W(0, T)$  to the dynamic state equation (4.1.5)

$$\|(u, \theta)\|_{W(0, T)}^2 \leq C \left( \|\Phi_g(\tau)\|_{\mathcal{H}}^2 + \|f\|_{L^2(0, T; \mathcal{V}^*)}^2 + \|f\|_{C(0, T; \mathcal{V}^*)}^2 \right) \quad (4.1.26)$$

with  $C$  independent of  $\tau \in U_b$ .

*Proof.* As a first remark, note that the space  $H^1(0, T; \mathcal{V}^*)$  is continuously embedded into  $C(0, T; \mathcal{V}^*)$  (see [Hun14, p. 205]), so the use of the norm of the latter space in (4.1.26) is meaningful.

The actual proof starts the same way as the proof of Theorem 4.1. All claims from there stay valid until the intermediate step

$$c \left( \|\dot{y}_m(t)\|_{\mathcal{H}}^2 + \|y_m(t)\|_{\mathcal{V}}^2 \right) < \|y_{1m}\|_{\mathcal{H}}^2 + 2 \int_0^t (f(s), \dot{y}_m(s))_{\mathcal{H}} \, ds \quad (4.1.27)$$

appears as a combination of (4.1.10) and (4.1.11). The next step taken in the proof of Theorem 4.1, namely

$$(f(s), \dot{y}_m(s))_{\mathcal{H}} \leq \|f(s)\|_{\mathcal{H}} \|\dot{y}_m(s)\|_{\mathcal{H}}$$

cannot be carried out here, since  $f(s)$  is only in  $\mathcal{V}^*$ . Instead, a partial integration with respect to  $t$  is done which leads to

$$\begin{aligned} 2 \int_0^t (f(s), \dot{y}_m(s))_{\mathcal{H}} \, ds = \\ 2 (f(t), y_m(t))_{\mathcal{H}} - 2 (f(0), y_m(0))_{\mathcal{H}} - 2 \int_0^t \left( \dot{f}(s), y_m(s) \right)_{\mathcal{H}} \, ds. \end{aligned}$$

This term is bounded from above by

$$\begin{aligned} & \left| 2 (f(t), y_m(t)) - 2 (f(0), y_m(0)) - 2 \int_0^t \left( \dot{f}(s), y_m(s) \right)_{\mathcal{H}} \, ds \right| \\ & \leq 2 \|f(t)\|_{\mathcal{V}^*} \|y_m(t)\|_{\mathcal{V}} + 2 \int_0^t \left\| \dot{f}(s) \right\|_{\mathcal{V}^*} \|y_m(s)\|_{\mathcal{V}} \, ds \\ & \leq 2 \|f\|_{C(0, T; \mathcal{V}^*)} \|y_m(t)\|_{\mathcal{V}} + \int_0^t \left\| \dot{f}(s) \right\|_{\mathcal{V}^*}^2 \, ds + \int_0^t \|y_m(s)\|_{\mathcal{V}}^2 \, ds \end{aligned}$$

where the factor in front of  $\|y_m(t)\|_{\mathcal{V}}$  is a constant  $C_f < \infty$ . Again introducing

$$\beta_m(t) := \|\dot{y}_m(t)\|_{\mathcal{H}}^2 + \|y_m(t)\|_{\mathcal{V}}^2,$$

equation (4.1.27) becomes

$$c\beta_m(t) < C_f \sqrt{\beta_m(t)} + \|y_{1m}\|_{\mathcal{H}}^2 + \int_0^t \left\| \dot{f}(s) \right\|_{\mathcal{V}^*}^2 \, ds + \int_0^t \beta_m(s) \, ds \quad (4.1.28)$$

with  $C_f = 2 \|f\|_{C(0,T;\mathcal{V}^*)}$ . There are two cases now: It either holds

$$\beta_m(t) \leq 4 \frac{C_f^2}{c^2} = \frac{16}{c^2} \|f\|_{C(0,T;\mathcal{V}^*)}^2, \quad (4.1.29)$$

or it follows

$$\beta_m(t) \geq 4 \frac{C_f^2}{c^2} \implies \frac{c}{2} \beta_m(t) \geq C_f \sqrt{\beta_m(t)} \implies c\beta_m(t) - C_f \sqrt{\beta_m(t)} \geq \frac{c}{2} \beta_m(t), \quad (4.1.30)$$

which in combination with (4.1.28) yields

$$\frac{c}{2} \beta_m(t) < \|y_{1m}\|_{\mathcal{H}}^2 + \int_0^t \|\dot{f}(s)\|_{\mathcal{V}^*}^2 ds + \int_0^t \beta_m(s) ds.$$

Applying Grönwall's lemma as in the proof of the original existence Theorem 4.1 shows the boundedness of the sequence  $\beta_m(t)$  and the existence of a weak limit  $y \in W(0,T)$ . Analogously to the proof from Theorem 4.1, it can be shown that the weak limit is indeed solution to the state equation and that the solution is unique. Moreover, in the case of (4.1.29) the estimate

$$\|y\|_{W(0,T)}^2 \leq \frac{16}{c^2} \|f\|_{C(0,T;\mathcal{V}^*)}^2$$

holds while in the case of (4.1.30)

$$\|y\|_{W(0,T)}^2 \leq C_2 \left( \|y_{1m}\|_{\mathcal{H}}^2 + \int_0^T \|\dot{f}(s)\|_{\mathcal{V}^*}^2 ds \right)$$

is valid. Both estimates can be combined to

$$\|y\|_{W(0,T)}^2 \leq C \left( \|y_{1m}\|_{\mathcal{H}}^2 + \int_0^T \|\dot{f}(s)\|_{\mathcal{V}^*}^2 ds + \|f\|_{C(0,T;\mathcal{V}^*)}^2 \right)$$

where  $C > 0$  is a constant independent of  $\tau \in U_b$ .  $\square$

It is time to put some more regularity onto the solution of the dynamic state equation. For this, let the applied force and the initial condition

$$F \in H^1(0,T;\mathcal{H}), \quad \Phi_g : L^\infty(\mathcal{S}) \rightarrow \mathcal{V} \text{ differentiable}, \quad (4.1.31)$$

in contrast to the previous assumptions  $F \in L^2(0,T;\mathcal{H})$  and  $\Phi_g(\tau) : L^\infty(\mathcal{S}) \rightarrow \mathcal{H}$ . By an abstract regularity theorem from Wloka ([Wlo87, pp. 442–443]), it follows for the solution  $y \in W(0,T)$  of the dynamic state equation

$$y \in L^2(0,T;\mathcal{V}), \quad \dot{y} \in L^2(0,T;\mathcal{V}), \quad \ddot{y} \in L^2(0,T;\mathcal{H}), \quad \ddot{y} \in L^2(0,T;\mathcal{V}^*). \quad (4.1.32)$$

An analogous proof as for the continuity Theorem 4.2 shows that all derivatives depend continuously on the thickness in their respective norms.

With help of this additional regularity, the above Theorem 4.3 can be applied to show that the norm of the solutions from (4.1.25) is bounded for  $\lambda \rightarrow 0$ .

**Lemma 4.8.** Let the assumptions from (4.1.31) hold for the dynamic state equation. Then the norm of the solutions  $y_\lambda$  to equation (4.1.23) with initial condition (4.1.24) is bounded for  $\lambda \rightarrow 0$ .

*Proof.* Take a look at the right-hand-side of (4.1.23), namely

$$\begin{aligned} \text{rhs} &:= -M_\lambda \ddot{y}_{(\tau+\lambda q)}(t) - Z_\lambda(y_{(\tau+\lambda q)}(t); q) \\ &:= -\left(\frac{M_{(\tau+\lambda q)} - M_{(\tau)}}{\lambda}\right) \ddot{y}_{(\tau+\lambda q)}(t) - Z_\lambda(y_{(\tau+\lambda q)}(t); q), \end{aligned} \quad (4.1.33)$$

where  $y_{(\tau+\lambda q)}$  solves the dynamic state equation for thickness  $\tau + \lambda q$ .

From the stronger regularity assumptions on the load and initial data, it follows

$$y \in L^2(0, T; \mathcal{V}), \quad \dot{y} \in L^2(0, T; \mathcal{V}), \quad \ddot{y} \in L^2(0, T; \mathcal{H}), \quad \ddot{\ddot{y}} \in L^2(0, T; \mathcal{V}^*),$$

hence the right-hand-side itself *and its first time derivative* are in  $L^2(0, T; \mathcal{V}^*)$ . For the first summand of the right-hand-side, it holds

$$M_\lambda \ddot{y}_{(\tau+\lambda q)} \xrightarrow{\lambda \rightarrow 0} (M'(\tau)q) \ddot{y}_{(\tau)} \in L^2(0, T; \mathcal{V}^*), \quad (4.1.34)$$

because both parts converge in their respective norm from Lemma 4.7 and 4.5. In particular, the  $L^2(0, T; \mathcal{V}^*)$ -norm is bounded.

For the second summand it was shown in Lemma 3.6

$$Z_\lambda(y_{(\tau+\lambda q)}(t); q) \xrightarrow{\lambda \rightarrow 0} Z_0(y_{(\tau)}(t); \cdot; q) =: Z_q(t) \quad (4.1.35)$$

in  $\mathcal{V}^*$ , where  $Z_0$  was defined in (3.1.10). This also ensures the boundedness of  $Z_\lambda$  in the  $L^2(0, T; \mathcal{V}^*)$ -norm for  $\lambda \rightarrow 0$ . The same is true for the first time-derivative of the right-hand-side, i.e.

$$\begin{aligned} \|M_\lambda \ddot{y}_{(\tau+\lambda q)}(t)\|_{L^2(0, T; \mathcal{V}^*)} &< C, \\ \|Z_\lambda(\dot{y}_{(\tau+\lambda q)}(t); q)\|_{L^2(0, T; \mathcal{V}^*)} &< C. \end{aligned}$$

Since  $H^1(0, T; \mathcal{V}^*)$  is continuously embedded into  $C(0, T; \mathcal{V}^*)$  (see [Hun14, p. 205]), it holds

$$\begin{aligned} \|M_\lambda \ddot{y}_{(\tau+\lambda q)}(t)\|_{C(0, T; \mathcal{V}^*)} &< C, \\ \|Z_\lambda(\dot{y}_{(\tau+\lambda q)}(t); q)\|_{C(0, T; \mathcal{V}^*)} &< C. \end{aligned}$$

A look at the initial conditions

$$\dot{y}_\lambda(0) = \frac{\Phi_g(\tau + \lambda q) - \Phi_g(\tau)}{\lambda}$$

shows that they also converge (in particular they are bounded),

$$\frac{\Phi_g(\tau + \lambda q) - \Phi_g(\tau)}{\lambda} \xrightarrow{\lambda \rightarrow 0} \Phi'_g(\tau)q$$

in  $\mathcal{V}$  according to the differentiability of  $\Phi_g$ .

Now the estimate from the extended solution Theorem 4.3 can be applied. It states that for the solutions  $y_\lambda$  to (4.1.23), it holds

$$\|y_\lambda\|_{W(0,T)}^2 \leq C \left( \left\| \frac{\Phi_g(\tau + \lambda q) - \Phi_g(\tau)}{\lambda} \right\|_{\mathcal{H}}^2 + \|\text{rhs}\|_{L^2(0,T;\mathcal{V}^*)}^2 + \|\text{rhs}\|_{C(0,T;\mathcal{V}^*)}^2 \right) < C_2,$$

where all terms were bounded in the above considerations.  $\square$

Since the norm of  $y_{\lambda_n}$  is bounded in  $W(0,T)$  for  $\lambda_n \rightarrow 0$ , there is a weakly convergent subsequence

$$y_{\lambda_{n_k}} \rightharpoonup z =: (w, \nu), \quad (4.1.36)$$

which is the candidate for the Gâteaux-derivative. From the proof of the above lemma, it also follows which equation is solved by  $(w, \nu)$ : The right-hand-side rhs from (4.1.33) converges in  $\mathcal{V}^*$  to

$$(M'(\tau)q)\ddot{y}_{(\tau)} + Z_q(t),$$

see equations (4.1.34) and (4.1.35). The weak convergence of  $y_{\lambda_{n_k}}$  is also sufficient to take the limit on the left-hand-side in (4.1.23) which becomes

$$M_{(\tau)}z(t) + L_{(\tau)}z(t).$$

It follows that the Gâteaux-derivative  $G'(\tau)q = (w, \nu)$  must solve the following problem: Find  $z = (w, \nu) \in W(0,T)$  such that

$$\begin{aligned} M_{(\tau)}\ddot{z}(t) + L_{(\tau)}z(t) &= -((M'(\tau)q)\ddot{y}_{(\tau)} + Z_q(t)) \\ z(0) &= 0, \quad \dot{z}(0) = \Phi'_g(\tau)q \end{aligned}$$

in  $\mathcal{V}^*$  for all  $t \in [0, T]$ . Because of the uniqueness of the limit, the whole sequence  $y_{\lambda_n}$  converges weakly to  $z$  from (4.1.36) for any  $\lambda_n \rightarrow 0$ . The last step is to investigate if this convergence is also strong.

**Lemma 4.9.** The convergence of  $y_\lambda$  to  $z$  from (4.1.36) is strong in  $W(0,T)$ , i.e.

$$\|y_\lambda - z\|_{W(0,T)} \rightarrow 0.$$

*Proof.* The difference  $y_\lambda - z$  is a solution to the problem

$$\begin{aligned} M_{(\tau)}(\ddot{y}_\lambda(t) - \ddot{z}(t)) + L_{(\tau)}(\ddot{y}_\lambda(t) - \ddot{z}(t)) \\ = -(M_\lambda \ddot{y}_{(\tau+\lambda q)}(t) - (M'_{(\tau)}q)\ddot{y}_{(\tau)}(t)) - (Z_\lambda(y_{(\tau+\lambda q)}(t); q) - Z_q(t)), \\ y_\lambda(0) - z(0) = 0, \quad \dot{y}_\lambda(0) - \dot{z}(0) = \frac{\Phi_g(\tau + \lambda q) - \Phi_g(\tau)}{\lambda} - \Phi'_g(\tau)q, \end{aligned}$$

so the norm  $\|y_\lambda - z\|_{W(0,T)}$  can be bounded from above with the help of the extended solution Theorem 4.3.

Actually, the main work has been done in the previous lemma: In (4.1.34) it was shown that the first part of the right-hand-side converges to 0 in the norm of  $L^2(0, T; \mathcal{V}^*)$ , which is also valid for its first time derivative

$$M_\lambda \ddot{y}_{(\tau+\lambda q)} - (M'_{(\tau)} q) \ddot{y}_{(\tau)} \xrightarrow{\lambda \rightarrow 0} 0$$

strong in  $L^2(0, T; \mathcal{V}^*)$  because of the differentiability of  $M_{(\tau)}$  and the continuity of  $\ddot{y}_{(\tau)}$  with respect to  $\tau$  (under the stronger assumptions on the initial data from Lemma 4.8). The considerations from the steady-state case of Lemma 3.4 also show that

$$Z_\lambda(y_{(\tau+\lambda q)}) - Z_q \xrightarrow{\lambda \rightarrow 0} 0$$

strong in  $L^2(0, T; \mathcal{V}^*)$  together with its first time derivative. Furthermore, the initial condition converges to 0 in  $\mathcal{H}$  because of the differentiability of  $\Phi_g$ . Using the estimate from Theorem 4.3 then yields

$$\|y_\lambda - z\|_{W(0, T)} \xrightarrow{\lambda \rightarrow 0} 0,$$

which shows the desired strong convergence.  $\square$

So everything is prepared to state the essential theorem of this subsection:

**Theorem 4.4.** Consider the dynamic state equation (4.1.5) with applied force and initial condition

$$F \in H^1(0, T; \mathcal{H}), \quad \Phi_g : L^\infty(\mathcal{S}) \rightarrow \mathcal{V} \text{ differentiable.}$$

Then the control-to-state operator  $G : U_b \rightarrow W(0, T)$  is Gâteaux-differentiable. For a thickness  $\tau \in U_b$  and a direction  $q \in W^{1, \infty}(\mathcal{S})$ , it holds

$$G'(\tau)q = (w, \nu),$$

where  $(w, \nu) \in W(0, T)$  is the solution to the variational problem: Find  $z = (w, \nu) \in W(0, T)$  such that

$$\begin{aligned} M_{(\tau)} \ddot{z}(t) + L_{(\tau)} z(t) &= -((M'_{(\tau)} q) \ddot{y}_{(\tau)}(t) + Z_q(t)) \\ z(0) &= 0, \quad \dot{z}(0) = \Phi'_g(\tau)q \end{aligned} \tag{4.1.37}$$

in  $\mathcal{V}^*$  for all  $t \in [0, T]$  with

$$\begin{aligned} (M'_{(\tau)} q)(u, \theta) &= \begin{pmatrix} \Phi'_1(\tau)qu + \Phi'_2(\tau)q\theta \\ \Phi'_1(\tau)qu_3 \\ \Phi'_2(\tau)qu + \Phi'_3(\tau)q\theta \end{pmatrix}, \\ \Phi'_1(\tau)q &= \left( \rho + \rho K \frac{\tau^2}{4} \right) q, \quad \Phi'_2(\tau)q = -2\rho H \frac{\tau^2}{4} q, \quad \Phi'_3(\tau)q = \left( \rho \frac{\tau^2}{4} + \rho K \frac{\tau^4}{16} \right) q \end{aligned}$$

and  $Z_q(t) \in \mathcal{V}^*$  given by

$$Z_q(t)(v, \psi) = \int_{\mathcal{S}} \sum_{\xi^3 \in \{\pm \frac{\tau}{2}\}} [(\sigma(u_{(\tau)}(t), \theta_{(\tau)}(t)) : \epsilon(v, \psi)) (1 - 2H\xi^3 + K(\xi^3)^2)] \Big|_{\xi^3 = \xi^3} \frac{q}{2} dS$$

where  $(u_{(\tau)}, \theta_{(\tau)}) = G(\tau)$ .

*Proof.* The existence of the (strong) limit

$$\lim_{\lambda \rightarrow 0} \frac{G(\tau + \lambda q) - G(\tau)}{\lambda}$$

in  $W(0, T)$  has been shown in the above Lemma 4.7 through 4.9. Since the problem (4.1.37) is linear, and  $q$  appears linearly in the right-hand-side in  $M'(\tau)q$  and  $Z_q(t)$ , the derivative  $(w, \nu)$  depends linearly on  $q$ . Moreover, the right-hand-side can be bounded by  $C \|q\|_{L^\infty(\mathcal{S})}$  and the same is valid for  $\|(w, \nu)\|_{W(0, T)}$  by Theorem 4.3, hence  $G'(\tau) : W^{1, \infty}(\mathcal{S}) \rightarrow W(0, T)$  is a continuous linear operator – the sought Gâteaux-derivative.  $\square$

As in the steady-state case, note the importance of the above theorem, which is necessary for the differentiation of the reduced objective in Theorem 4.6.

## 4.2. Analysis of the optimization problem

This subsection is devoted to the analysis of the dynamic optimization problem. It has a similar structure as the steady-state problem and is stated as follows: Given an applied force  $f \in H^1(0, T; L^2(\mathcal{S}))$  and an initial condition by  $\Phi_g : L^\infty(\mathcal{S}) \rightarrow \mathcal{V}$  differentiable, then with  $F := (0, f, 0) \in H^1(0, T; \mathcal{H})$ ,

$$\min_{\substack{\tau \in W^{1, \infty}(\mathcal{S}) \\ (u, \theta) \in W(0, T)}} J(u, \theta; \tau) := \int_0^T \int_{\mathcal{S}} u_3(t) f(t) dS dt + \frac{\lambda}{2} \|\tau\|_{H^1(\mathcal{S})}^2 \quad (4.2.1a)$$

$$\text{s.t. : } M_{(\tau)}(\ddot{u}(t), \ddot{\theta}(t)) + L_{(\tau)}(u(t), \theta(t)) = F(t) \text{ in } \mathcal{V}^*, t \in [0, T] \quad (4.2.1b)$$

$$(u(0), \theta(0)) = 0, \quad (\dot{u}(0), \dot{\theta}(0)) = \Phi_g(\tau) \quad (4.2.1c)$$

$$\tau_{\min} \leq \tau(\xi^1, \xi^2) \leq \tau_{\max} \text{ in } \mathcal{S} \quad (4.2.1d)$$

$$\int_{\mathcal{S}} \left( \tau + \frac{K\tau^3}{12} \right) dS = C. \quad (4.2.1e)$$

The sketch of this subsection is to apply the continuity and differentiability results from the previous subsection and make respective conclusions for the reduced objective. Finally, necessary conditions for an optimal solution in case of a parabolic midsurface can be stated.

First, the corresponding reduced problem is stated. All the restrictions (except the state equation) are collected in the set  $U_{\text{ad}}$  already known from the steady-state case,

$$U_{\text{ad}} := \left\{ \tau \in W^{1,\infty}(\mathcal{S}) \mid \tau_{\min} \leq \tau(\xi^1, \xi^2) \leq \tau_{\max}, \int_{\mathcal{S}} \left( \tau + \frac{K\tau^3}{12} \right) dS = C \right\}. \quad (4.2.2)$$

Since  $U_{\text{ad}} \subset U_{\text{b}}$ , the control-to-state operator is well-defined on  $U_{\text{ad}}$  (see Theorem 4.1), and the objective in conjunction with the state equation can be rewritten as

$$J(u, \theta; \tau) = J(G(\tau); \tau) =: J_s(\tau), \quad (4.2.3)$$

where the reduced objective  $J_s : U_{\text{ad}} \rightarrow \mathbb{R}$  only depends on the thickness  $\tau$ . The notation is not distinguished from the steady-state case as long as it is clear from the context which objective is meant. Combining both equations (4.2.2) and (4.2.3), the *reduced problem* is stated as

$$\min_{\tau \in U_{\text{ad}}} J_s(\tau).$$

This reduced objective is subject to investigation. The first thing that can be established is its continuity.

**Lemma 4.10.** The reduced objective  $J_s : U_{\text{ad}} \rightarrow \mathbb{R}$  is continuous.

*Proof.* The original objective  $J : W(0, T) \times W^{1,\infty}(\mathcal{S})$  is continuous since

$$(u, \theta) \mapsto \int_0^T \int_{\mathcal{S}} u_3(t) f(t) dS dt \leq \int_0^T \|(u(t), \theta(t))\|_{\mathcal{H}} \|f(t)\|_{L^2(\mathcal{S})} dt \leq C \|(u, \theta)\|_{L^2(0, T; \mathcal{V})}$$

is a continuous linear functional and  $W^{1,\infty}(\mathcal{S})$  is continuously embedded into  $H^1(\mathcal{S})$ . Moreover, the control-to-state operator is continuous, and so is the composition  $J_s = J \circ (G, \text{id})$ .  $\square$

By using the continuity of the reduced objective an existence theorem for an optimal solution can be stated for further restricted  $U_{\text{ad}}$ .

**Theorem 4.5.** Consider the set

$$U_{\text{ad}}^M := U_{\text{ad}} \cap \left\{ \tau \in W^{1,\infty}(\mathcal{S}) \left| \left\| \frac{\partial \tau}{\partial \xi^1} \right\|_{L^\infty(\mathcal{S})} < M, \left\| \frac{\partial \tau}{\partial \xi^2} \right\|_{L^\infty(\mathcal{S})} < M \right\},$$

where  $M > 0$  is fixed. Then the dynamic optimization problem

$$\min_{\tau \in U_{\text{ad}}^M} J_s(\tau)$$

has an optimal solution.

*Proof.* Analogous to the steady-state case in Theorem 3.5.  $\square$

The next step is to consider the differentiability of the reduced objective and in the sequel derive necessary conditions for an optimal solution of (4.2.1). As in the steady-state case, the expression of a directional derivative which does not rely on finite differences is invaluable for direct numerical optimization methods and sometimes, as it is the case here, makes a numerical solution possible at all.

**Lemma 4.11.** The objective functional  $J$  from (4.2.1a) is a Fréchet-differentiable mapping  $W(0, T) \times W^{1,\infty}(\mathcal{S}) \rightarrow \mathbb{R}$ . The derivative with respect to  $(u, \theta)$ , denoted by  $D_{(u,\theta)}J(u, \theta; \tau)$ , is given by

$$D_{(u,\theta)}J(u, \theta; \tau)(v, \psi) = \int_0^T \int_{\mathcal{S}} v_3(t) f(t) \, dS \, dt,$$

and the derivative with respect to  $\tau$ , denoted by  $D_{(\tau)}J(u, \theta; \tau)$ , is given by

$$D_{\tau}J(u, \theta; \tau)q = \lambda(\tau, q)_{H^1(\mathcal{S})}.$$

*Proof.* As a continuous linear functional  $W(0, T) \rightarrow \mathbb{R}$  the first part

$$(u, \theta) \mapsto \int_0^T \int_{\mathcal{S}} u_3(t) f(t) \, dS \, dt$$

is Fréchet-differentiable (see [Trö10, p. 59, ex. (vi)]), and the derivative is the operator itself, i.e.

$$D_{(u,\theta)}J(u, \theta; \tau)(v, \psi) = \int_0^T \int_{\mathcal{S}} v_3(t) f(t) \, dS \, dt.$$

The second part containing the  $H^1(\mathcal{S})$ -norm is also Fréchet-differentiable (see [Trö10, p. 59, ex. (vii)]) and since  $H^1(\mathcal{S})$  is a Hilbert space, the derivative with respect to  $\tau$  is given by

$$D_{\tau}J(u, \theta; \tau)q = \lambda(\tau, q)_{H^1(\mathcal{S})}.$$

$\square$

**Remark 4.6.** *The objectives proposed in Itemization 2.3, point 1 and 3 can be adapted to the dynamic case and are also Fréchet-differentiable:*

1. *Square-norm of the displacement  $\frac{1}{2} \|(u, \theta)\|^2$  (where a suitable norm must be chosen, e.g. the  $L^2(0, T; \mathcal{V})$  or  $L^2(0, T; \mathcal{H})$  norm),*
2.  *$u_3$ -component of the displacement integrated over the midsurface, i.e.*

$$\int_0^T \int_{\mathcal{S}} u_3(t) \, dS \, dt,$$

*as done in [Nes10a].*

After the existence of the Fréchet-derivative of the objective functional is assured, an *adjoint state* can be defined as the solution to a related problem:

**Definition 4.5** (Adjoint state). The *adjoint state*  $p = (r, \eta)$  to a thickness  $\tau$  with corresponding displacement  $(u_{(\tau)}, \theta_{(\tau)})$  is defined as the solution to: Find  $(r, \eta) \in W(0, T)$  such that

$$\begin{aligned} M_{(\tau)}(\ddot{r}(t), \ddot{\eta}(t)) + L_{(\tau)}(r(t), \eta(t)) &= F(t) \\ (r(T), \eta(T)) &= (\dot{r}(T), \dot{\eta}(T)) = 0 \end{aligned} \quad (4.2.4)$$

in  $\mathcal{V}^*$  for all  $t \in [0, T]$ .

First, note that the above equation (4.2.4) does not have end conditions instead of initial conditions, so it is an equation “backwards” in time. For this type of problem, the solvability must be clarified first.

**Lemma 4.12.** The adjoint equation (4.2.4) has a unique solution.

*Proof.* Consider the auxiliary function  $(\tilde{r}, \tilde{\eta}) \in W(0, T)$ ,

$$(\tilde{r}(t), \tilde{\eta}(t)) := (r(T - t), \eta(T - t)).$$

If  $(r, \eta)$  solves the adjoint equation (4.2.4), this is equivalent to  $(\tilde{r}, \tilde{\eta})$  solving

$$\begin{aligned} M_{(\tau)}(\ddot{\tilde{r}}(t), \ddot{\tilde{\eta}}(t)) + L_{(\tau)}(\tilde{r}(t), \tilde{\eta}(t)) &= F(T - t) \\ (\tilde{r}(0), \tilde{\eta}(0)) &= (\dot{\tilde{r}}(0), \dot{\tilde{\eta}}(0)) = 0, \end{aligned} \quad (4.2.5)$$

in  $\mathcal{V}^*$  for all  $t \in [0, T]$ . Looking at the right-hand-side of the equations shows that  $F(T - t) \in L^2(0, T; \mathcal{H})$ , so equation (4.2.5) has a unique solution (see Theorem 4.5) and the same is true for equation (4.2.4).  $\square$

Note that, given the initial data  $F \in H^1(0, T; \mathcal{H})$ , it follows like in (4.1.32) that the adjoint has the same smoothness as the displacement, in particular  $p \in H^1(0, T; \mathcal{V})$ .

**Remark 4.7.** The adjoint state can also be defined for the other objectives mentioned in Itemization 4.6, where the right-hand-side then reads

1.  $((u(t), \theta(t)), (v, \psi))_{\mathcal{H}}$  or
2.  $(1, v_3)_{L^2(\mathcal{S})}$ .

It follows the main result of this subsection:

**Theorem 4.6.** The reduced objective  $J_s$  from (4.2.3) is Gâteaux-differentiable. For the directional derivative  $J'_s(\tau)q$  at point  $\tau$  with corresponding displacement  $y_{(\tau)} = (u_{(\tau)}, \theta_{(\tau)})$  and adjoint  $p \in W(0, T)$ , it holds

$$\begin{aligned} J'_s(\tau)q &= \\ &= - \int_0^T \left( (M'_{(\tau)}q)\ddot{y}_{(\tau)}(t) + Z_q(t), p(t) \right)_{\mathcal{H}} dt + (M_{(\tau)}p(0), \Phi'_g(\tau)q)_{\mathcal{H}} + \lambda(\tau, q)_{H^1(\mathcal{S})}. \end{aligned}$$

*Proof.* By using the chain rule for Gâteaux- and Fréchet-derivatives (see [KA82, p. 499, example IV]), it is

$$J'_s(\tau)q = D_{(u,\theta)}J(u(\tau), \theta(\tau); \tau)(G'(\tau)q) + D_\tau J(u(\tau), \theta(\tau); \tau)q. \quad (4.2.6)$$

Consider the first summand on the right-hand-side, define  $z = (w, \nu) := G'(\tau)q$ , let  $p \in W(0, T)$  be the solution to the adjoint equation (4.2.4), then

$$\begin{aligned} & D_{(u,\theta)}J(u(\tau), \theta(\tau); \tau)(w, \nu) \\ &= \int_0^T \int_S w_3(t) f(t) \, dS \, dt = \int_0^T (F(t), z(t))_{\mathcal{H}} \, dt \\ &= \int_0^T (M_{(\tau)}\ddot{p}(t), z(t))_{\mathcal{H}} + (L_{(\tau)}p(t), z(t))_{\mathcal{H}} \, dt \\ &= (M_{(\tau)}\dot{p}(T), z(T))_{\mathcal{H}} - (M_{(\tau)}\dot{p}(0), z(0))_{\mathcal{H}} + \\ &\quad \int_0^T - (M_{(\tau)}\dot{p}(t), \dot{z}(t))_{\mathcal{H}} + (L_{(\tau)}p(t), z(t))_{\mathcal{H}} \, dt \\ &= \int_0^T - (M_{(\tau)}\dot{z}(t), \dot{p}(t))_{\mathcal{H}} + (L_{(\tau)}z(t), p(t))_{\mathcal{H}} \, dt \\ &= (M_{(\tau)}\dot{z}(0), p(0))_{\mathcal{H}} - (M_{(\tau)}\dot{z}(T), p(T))_{\mathcal{H}} + \\ &\quad \int_0^T (M_{(\tau)}\ddot{z}(t), p(t))_{\mathcal{H}} + (L_{(\tau)}z(t), p(t))_{\mathcal{H}} \, dt \\ &= (M_{(\tau)}p(0), \dot{z}(0))_{\mathcal{H}} + \int_0^T (M_{(\tau)}\ddot{z}(t), p(t))_{\mathcal{H}} + (L_{(\tau)}z(t), p(t))_{\mathcal{H}} \, dt \\ &= (M_{(\tau)}p(0), \Phi'_g(\tau)q)_{\mathcal{H}} + \int_0^T - ((M'(\tau)q)\ddot{y}_{(\tau)}(t) + Z_q(t), p(t))_{\mathcal{H}} \, dt \end{aligned}$$

which gives the first part of the derivative. The derivative of the squared norm in the second summand of (4.2.6) is given by

$$D_\tau J(u(\tau), \theta(\tau); \tau)q = \lambda(\tau, q)_{H^1(S)}.$$

Combining both terms yields the statement of the theorem.  $\square$

The importance of an analytical expression for the directional derivative of the reduced objective is emphasized once again. The first thing to collect from the above work are necessary conditions for an optimal solution of (4.2.1).

**Theorem 4.7.** Let  $\tau^* \in U_{\text{ad}}$  be a (locally) optimal solution for the problem (4.2.1) with corresponding state  $y_{(\tau^*)}$  and adjoint  $p$ , and let  $K = 0$  everywhere on  $\mathcal{S}$ . Then it holds

$$\begin{aligned} J'_s(\tau^*)(q - \tau^*) &= - \int_0^T \left( M'_{(\tau^*)}(q - \tau^*)\ddot{y}_{(\tau^*)}(t) + Z_{(q-\tau^*)}(t), p(t) \right)_{\mathcal{H}} \, dt + \\ &\quad (M_{(\tau^*)}p(0), \Phi'_g(\tau^*)(q - \tau^*))_{\mathcal{H}} + \lambda(\tau^*, q - \tau^*)_{H^1(S)} \geq 0 \end{aligned} \quad (4.2.7)$$

for all directions  $q \in U_{\text{ad}}$ .

*Proof.* Analogous to the steady-state case using the derivative of  $J_s$  and the convexity of  $U_{\text{ad}}$  together with the theorem from [Trö10, p. 63].  $\square$

### 4.3. Summary

In this chapter the dynamic state equation (4.1.5) and the corresponding optimization problem (4.2.1) were investigated in detail. After stating the unique solvability of the state equation, the existence of a well-defined control-to-state operator had been assured. This operator was studied then with respect to continuity and Gâteaux-differentiability which is a key component of this work and a new contribution to optimization in (dynamic) linear elasticity. Since the type of the state equation changed from elliptic to hyperbolic, considerably more work had to be done to show the existence of all desired expressions. Optimal control with hyperbolic state equations is however a more complex and not so deep investigated field of research (take for example the work of Tröltzsch [Trö10] which ends up with parabolic equations) and standard works on hyperbolic equations by Wloka [Wlo87] or Evans [EG15] do not always provide the needed theorems (like Theorem 4.3). Nevertheless, all needed results could be achieved.

Afterwards, those results were applied to the reduced objective and continuity and differentiability could be shown. On the one hand, this ensures the existence of a solution to the optimization problem (under additional restrictions) and on the other hand gives an expression for the directional derivative of the reduced objective and for necessary conditions for an optimal solution.



## 5. Numerical solution

This chapter is devoted to the numerical solution of the state equation and the optimization problem. The type of the steady-state and dynamic state equation are fundamentally different (elliptic vs. hyperbolic) and so is the numerical treatment. This will be discussed in the first two sections. The actual optimization is done using an interior point method provided by the package Ipopt ([Ipo15]) which is explained in the third section.

The numerical solution of the state equation relies on the finite element method and is done in Fortran where self-written and already available code are combined to achieve a time-efficient solution. While the numerical methods (e.g. the particular elements used or the interior point optimization) are well-known, the efficient implementation is pronounced as a substantial own contribution to this work. It consists of the combination of fast linear algebra using Intel MKL ([MKL15]), the parallel solution of sparse linear equation systems using Pardiso ([Par15]) and parallelized element calculations using OpenMP ([OMP15]).

### 5.1. Solution of the steady-state equation

The first section deals with the numerical solution of the steady-state equation (2.2.25). It is stated once again using the (implementation-friendly) Voigt notation: Find  $(u, \theta) \in \mathcal{V}$  such that

$$\int_{\mathcal{B}(\tau)} \boldsymbol{\varepsilon}^T(u, \theta) \mathbf{C} \boldsymbol{\varepsilon}(v, \psi) + \boldsymbol{\zeta}(u, \theta)^T \mathbf{D} \boldsymbol{\zeta}(v, \psi) \, dV = \int_{\mathcal{S}} f v_3 \, dS \quad (5.1.1)$$

for all  $(v, \psi) \in \mathcal{V}$ . The vector  $\boldsymbol{\varepsilon}(u, \theta)$  (representing a part of the strain tensor  $e$  given in (2.2.12)) is defined by

$$\boldsymbol{\varepsilon}(u, \theta) := \begin{pmatrix} e_{11}(u, \theta) \\ e_{22}(u, \theta) \\ \sqrt{2}e_{12}(u, \theta) \end{pmatrix},$$

while  $\boldsymbol{\zeta}$  represents the shear strain

$$\boldsymbol{\zeta}(u, \theta) := \begin{pmatrix} \zeta_1(u, \theta) \\ \zeta_2(u, \theta) \end{pmatrix}.$$

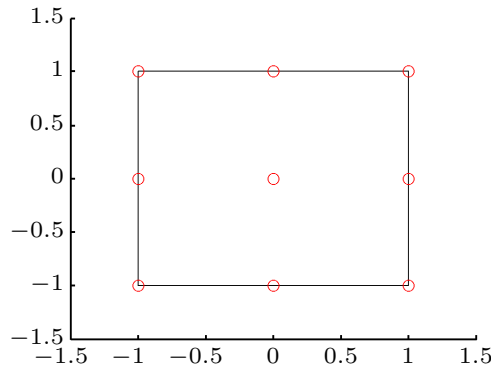


Figure 5.1.: 9-node reference element

The matrices  $\mathbf{C}$  and  $\mathbf{D}$  are the representations for the parts  $C$  and  $D$  of the elasticity tensor from (2.2.20) and consist of

$$\mathbf{C} := \begin{pmatrix} C^{1111} & C^{1122} & \sqrt{2}C^{1112} \\ C^{1122} & C^{2222} & \sqrt{2}C^{2212} \\ \sqrt{2}C^{1112} & \sqrt{2}C^{2212} & 2C^{1212} \end{pmatrix}, \quad \mathbf{D} := \begin{pmatrix} D^{11} & D^{12} \\ D^{12} & D^{22} \end{pmatrix}.$$

The numerical solution of the variational problem (5.1.1) is done as in the book by Chapelle and Bathe ([CB03, sec. 6.3]) and the paper by Lee and Bathe ([LB05]). In the above literature, *general shell elements* are suggested which use an isoparametric approach that interpolates the geometry and the displacement using the same shape functions. However, for the cylindrical test case the parametrization  $Z$  is readily available from (2.2.3), so the parameter region  $\Omega_{(\tau)}$  is discretized rather than the shell body  $\mathcal{B}_{(\tau)}$ . This is the main difference to the above method. Indeed, for more general test cases, general shell elements should be used which directly discretize  $\mathcal{B}_{(\tau)}$ . An overview of various other finite element approaches in shell theory is given in [Hug00, chp. 5–6].

### 5.1.1. Shell elements and finite element displacement

**Discretization using shell elements** The spatial discretization uses shell elements based on 9-node biquadratic rectangle elements, see Figure 5.1. Given the coordinates  $\xi^{1,(j)}$  and  $\xi^{2,(j)}$  at the nine local nodes  $j = 1, \dots, 9$  together with the nodal values of the thickness  $\tau^{(j)}$ , the global coordinates  $(\xi^1, \xi^2, \xi^3)$  (in the parameter domain  $\Omega_{(\tau)}$ ) depending on the element coordinates  $(r, s, z) \in [-1, 1]^3$  read

$$\begin{pmatrix} \xi^1 \\ \xi^2 \\ \xi^3 \end{pmatrix} (r, s, z) = \sum_{j=1}^9 \lambda_j(r, s) \begin{pmatrix} \xi^{1,(j)} \\ \xi^{2,(j)} \\ z \frac{\tau^{(j)}}{2} \end{pmatrix}.$$

The ansatz functions  $\lambda_j$ ,  $j = 1, \dots, 9$  on the reference element are given as

$$\begin{aligned}\lambda_1(r, s) &= \frac{1}{4}(r+1)r(s+1)s, & \lambda_5(r, s) &= \frac{1}{2}(1-r^2)(s+1)s, \\ \lambda_2(r, s) &= \frac{1}{4}(r-1)r(s+1)s, & \lambda_6(r, s) &= \frac{1}{2}(r-1)r(1-s^2), \\ \lambda_3(r, s) &= \frac{1}{4}(r-1)r(s-1)s, & \lambda_7(r, s) &= \frac{1}{2}(1-r^2)(s-1)s, \\ \lambda_4(r, s) &= \frac{1}{4}(r+1)r(s-1)s, & \lambda_8(r, s) &= \frac{1}{2}(r+1)r(1-s^2), \\ & & \lambda_9(r, s) &= (1-r^2)(1-s^2).\end{aligned}\tag{5.1.2}$$

With the use of such shell elements, a three dimensional mesh is obtained using two dimensional shape functions. The existence of only one layer through the thickness reflects the idea of degenerating a fully three dimensional solid shell element ([CB03, p. 180f.]).

**Finite element displacement** The modeling of the displacement involves an additional coordinate transformation as suggested in [LB05]. The values  $u^{(i)}$  of the translational displacement at the nodes  $i = 1, \dots, N$  are represented in the standard basis  $i_1, i_2, i_3$  of  $\mathbb{R}^3$ ,

$$u^{(i)} = \hat{u}_1^{(i)} i_1 + \hat{u}_2^{(i)} i_2 + \hat{u}_3^{(i)} i_3.$$

The same is done with the nodal values of the rotational displacement  $\theta$ ,

$$\theta^{(i)} = \hat{\theta}_1^{(i)} i_1 + \hat{\theta}_2^{(i)} i_2 + \hat{\theta}_3^{(i)} i_3.$$

Note that this leads to an additional degree of freedom per node (six in standard basis, five in local basis) but shows more accurate numerical results according to [LB05]. However, the rotational part  $\theta^{(i)}$  must be enforced to be in the tangent space of node  $i$ , i.e.

$$\theta^{(i)} \cdot a_3^{(i)} = 0,$$

in order to be compliant with the displacement ansatz from (2.2.11). Altogether, the finite element displacements are given as

$$U_h(\xi^1, \xi^2, \xi^3) = \sum_{i=1}^N h_i(\xi^1, \xi^2)(u^{(i)} + \xi^3 \theta^{(i)}), \quad \theta^{(i)} \cdot a_3^{(i)} = 0,$$

where the  $h_i$  are chosen as the shape functions arising from the biquadratic ansatz functions (5.1.2). The index  $h$  refers to the mesh-size and indicates a quantity that is interpolated from discrete nodal values. In local element coordinates this reads

$$U_h \Big|_E(r, s, z) = \sum_{j=1}^9 \lambda_j(r, s) \left( u^{n(j)} + z \frac{\tau^{n(j)}}{2} \theta^{n(j)} \right), \tag{5.1.3}$$

where  $n(j)$  denotes the global node corresponding to the local node  $j$ . This also reflects the isoparametric ansatz. The division into translational and rotational parts leads to the finite element displacement space

$$\mathcal{V}_h := \left\{ (u_h, \theta_h) \mid u_h = \sum_{i=1}^N h_i u^{(i)}, \theta_h = \pi \left( \sum_{i=1}^N h_i \theta^{(i)} \right) \right\} \cap \mathcal{BC},$$

where  $\pi$  denotes the projection operator onto the tangent space at point  $z(\xi^1, \xi^2) \in \mathcal{S}$  (see [CB03, p. 183]). The part denoted  $\mathcal{BC}$  imposes proper boundary conditions on the displacement. Accordingly, the nodal thickness  $\tau^{(i)}$  is interpolated using the shape functions  $h_i$  which leads to

$$\tau_h = \sum_{i=1}^N h_i \tau^{(i)}. \quad (5.1.4)$$

The actual variational problem to solve on the finite element displacement space is: Find  $(u_h, \theta_h) \in \mathcal{V}_h$  such that

$$\int_{\mathcal{B}(\tau_h)} \boldsymbol{\varepsilon}^T(u_h, \theta_h) \mathbf{C} \boldsymbol{\varepsilon}(v_h, \psi_h) + \boldsymbol{\zeta}(u_h, \theta_h)^T \mathbf{D} \boldsymbol{\zeta}(v_h, \psi_h) dV = \int_{\mathcal{S}} f v_{3,h} dS \quad (5.1.5)$$

for all  $(v_h, \psi_h) \in \mathcal{V}_h$ . In short notation this reads

$$a_{(\tau_h)}(u_h, \theta_h; v_h, \psi_h) = F(v_h, \psi_h).$$

The linear structure of the finite element displacements with respect to the nodal values  $(u^{(i)}, \theta^{(i)})$  shows that the above equation has to be evaluated numerically only for the shape functions  $h_i$ ,  $i = 1, \dots, N$ . The actual calculation is presented in Subsection 5.1.2.

**Coordinate transformation** In order to transform between the  $\hat{\cdot}$ -accented Cartesian coordinates and the coordinates in local basis, the vectors  $a_i$ ,  $i = 1, 2, 3$  from (2.2.1) are used. It is

$$\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} (a_1)_1 & (a_1)_2 & (a_1)_3 \\ (a_2)_1 & (a_2)_2 & (a_2)_3 \\ (a_3)_1 & (a_3)_2 & (a_3)_3 \end{pmatrix} \begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \end{pmatrix} =: S_u \begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \end{pmatrix}.$$

Moreover, the coordinates of the derivatives must be transformed, so

$$\frac{\partial}{\partial \xi^\alpha} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = S_u \frac{\partial}{\partial \xi^\alpha} \begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \end{pmatrix} + \frac{\partial S_u}{\partial \xi^\alpha} \begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \end{pmatrix}.$$

For the coordinates of the rotational displacement, only the first two rows of  $S_u$  are used due to the constraint  $a_3 \cdot \theta^{(i)} = 0$ , i.e.

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} (a_1)_1 & (a_1)_2 & (a_1)_3 \\ (a_2)_1 & (a_2)_2 & (a_2)_3 \end{pmatrix} \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \end{pmatrix} =: S_\theta \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \end{pmatrix}.$$

The combination of  $S_u$  and  $S_\theta$  gives the transformation

$$\begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} S_u & 0 \\ 0 & S_\theta \end{pmatrix} \begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \end{pmatrix} =: S \begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \end{pmatrix}.$$

For the cylindrical case, the matrix  $S_u(x, \varphi)$  reads

$$S_u(x, \varphi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -R \sin \varphi & R \cos \varphi \\ 0 & \cos \varphi & \sin \varphi \end{pmatrix} \quad (5.1.6)$$

and its derivatives are

$$\frac{\partial S_u}{\partial x} = 0, \quad \frac{\partial S_u}{\partial \varphi} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -R \cos \varphi & -R \sin \varphi \\ 0 & -\sin \varphi & \cos \varphi \end{pmatrix}. \quad (5.1.7)$$

The matrix  $S_\theta$  and its derivatives consist of the first two rows of  $S_u$  and its respective derivatives.

### 5.1.2. Evaluating the quantities in the variational problem

The time-efficient evaluation of the quantities appearing in equation (5.1.5) is a key feature of a successive optimization in reasonable time.

**Strain-displacement matrices** The strain vector  $\varepsilon$  is related to the nodal displacement via *strain-displacement matrices*. Remember from (2.2.12) that  $\varepsilon$  is actually given as

$$\varepsilon = \gamma + \xi^3 \chi - (\xi^3)^2 \kappa.$$

To obtain the components of the vector

$$\gamma = \begin{pmatrix} \gamma_{11} \\ \gamma_{22} \\ \sqrt{2}\gamma_{12} \end{pmatrix} = \begin{pmatrix} u_{1|1} - b_{11}u_3 \\ u_{2|2} - b_{22}u_3 \\ \frac{\sqrt{2}}{2}(u_{1|2} + u_{2|1}) - \sqrt{2}b_{12}u_3 \end{pmatrix}$$

from the nodal displacement values on the element  $(\hat{u}^{(j)}, \hat{\theta}^{(j)})$ ,  $j = 1, \dots, 9$ , the following calculation is done exemplarily for  $u_{1|1}$  from Definition 2.11:

$$\begin{aligned} u_{1|1} &= \frac{\partial u_1}{\partial \xi^1} - \sum_{\alpha=1}^2 \Gamma_{11}^\alpha u_\alpha \\ &= \frac{\partial}{\partial \xi^1} \left( \sum_{j=1}^9 \lambda_j u_1^{(j)} \right) - \Gamma_{11}^1 \sum_{j=1}^9 \lambda_j u_1^{(j)} - \Gamma_{11}^2 \sum_{j=1}^9 \lambda_j u_2^{(j)} \\ &= \sum_{j=1}^9 \left( \frac{\partial}{\partial \xi^1} \left( \lambda_j \left[ S \begin{pmatrix} \hat{u}^{(j)} \\ \hat{\theta}^{(j)} \end{pmatrix} \right]_1 \right) - \Gamma_{11}^1 \lambda_j \left[ S \begin{pmatrix} \hat{u}^{(j)} \\ \hat{\theta}^{(j)} \end{pmatrix} \right]_1 - \Gamma_{11}^2 \lambda_j \left[ S \begin{pmatrix} \hat{u}^{(j)} \\ \hat{\theta}^{(j)} \end{pmatrix} \right]_2 \right) \\ &= \sum_{j=1}^9 \left( \left( \frac{\partial \lambda_j}{\partial \xi^1} - \Gamma_{11}^1 \lambda_j, -\Gamma_{11}^2 \lambda_j, 0, \dots, 0 \right) S \begin{pmatrix} \hat{u}^{(j)} \\ \hat{\theta}^{(j)} \end{pmatrix} + (\lambda_j, 0, \dots, 0) \frac{\partial S}{\partial \xi^1} \begin{pmatrix} \hat{u}^{(j)} \\ \hat{\theta}^{(j)} \end{pmatrix} \right) \end{aligned}$$

Altogether, for the membrane strain the strain-displacement matrix on the reference element is calculated as follows

$$\begin{aligned}
\boldsymbol{\gamma} &= \sum_{j=1}^9 \left[ \begin{pmatrix} \frac{\partial \lambda_j}{\partial \xi^1} - \Gamma_{11}^1 \lambda_j & -\Gamma_{11}^2 \lambda_j & -b_{11} \lambda_j & 0 & 0 & 0 \\ -\Gamma_{22}^1 \lambda_j & \frac{\partial \lambda_j}{\partial \xi^2} - \Gamma_{22}^2 \lambda_j & -b_{22} \lambda_j & 0 & 0 & 0 \\ \frac{\sqrt{2}}{2} \left( \frac{\partial \lambda_j}{\partial \xi^2} - (\Gamma_{12}^1 + \Gamma_{21}^1) \lambda_j \right) & \frac{\sqrt{2}}{2} \left( \frac{\partial \lambda_j}{\partial \xi^1} - (\Gamma_{21}^2 + \Gamma_{12}^2) \lambda_j \right) & -\sqrt{2} b_{12} \lambda_j & 0 & 0 & 0 \end{pmatrix} S \right. \\
&\quad + \begin{pmatrix} \lambda_j & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{\sqrt{2}}{2} \lambda_j & 0 & 0 & 0 & 0 \end{pmatrix} \frac{\partial S}{\partial \xi^1} \\
&\quad \left. + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_j & 0 & 0 & 0 & 0 \\ \frac{\sqrt{2}}{2} \lambda_j & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \frac{\partial S}{\partial \xi^2} \right] \begin{pmatrix} \hat{u}_1^{(j)} \\ \hat{u}_2^{(j)} \\ \hat{u}_3^{(j)} \\ \hat{\theta}_1^{(j)} \\ \hat{\theta}_2^{(j)} \\ \hat{\theta}_3^{(j)} \end{pmatrix} \\
&=: \sum_{j=1}^9 \left( B_\gamma^{(j),0} S + B_\gamma^{(j),1} \frac{\partial S}{\partial \xi^1} + B_\gamma^{(j),2} \frac{\partial S}{\partial \xi^2} \right) (\hat{u}^{(j)}, \hat{\theta}^{(j)}) \\
&=: \sum_{j=1}^9 B_\gamma^{(j)} (\hat{u}^{(j)}, \hat{\theta}^{(j)}).
\end{aligned}$$

Corresponding matrices  $B_\chi^{(j)}$ ,  $B_\kappa^{(j)}$  and  $B_\zeta^{(j)}$  can be calculated for the strain vectors  $\boldsymbol{\chi}$ ,  $\boldsymbol{\kappa}$  and  $\boldsymbol{\zeta}$ . For the cylindrical test case, the matrices  $B_\gamma^{(j),k}$ ,  $k = 0, 1, 2$  take the simple form

$$\begin{aligned}
B_\gamma^{(j),0} S &= \begin{pmatrix} \frac{\partial \lambda_j}{\partial x} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{\partial \lambda_j}{\partial \varphi} & R \lambda_j & 0 & 0 & 0 \\ \frac{\sqrt{2}}{2} \frac{\partial \lambda_j}{\partial \varphi} & \frac{\sqrt{2}}{2} \frac{\partial \lambda_j}{\partial x} & 0 & 0 & 0 & 0 \end{pmatrix} S, \quad B_\gamma^{(j),1} \frac{\partial S}{\partial x} = 0, \\
B_\gamma^{(j),2} \frac{\partial S}{\partial \varphi} &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_j & 0 & 0 & 0 & 0 \\ \frac{1}{\sqrt{2}} \lambda_j & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \frac{\partial S}{\partial \varphi}.
\end{aligned}$$

Note that the derivatives of the ansatz functions have to be transformed from element coordinates  $(r, s)$  to global coordinates  $(\xi^1, \xi^2)$  which is a straight-forward procedure in finite element calculations.

**Evaluation of the bilinear form** The matrices  $\boldsymbol{C}$  and  $\boldsymbol{D}$  representing the elasticity tensor can be calculated directly from (2.2.20). To evaluate the bilinear form on the

left-hand-side of (5.1.1), the matrix-vector-product

$$\boldsymbol{\varepsilon}^T \mathbf{C} \boldsymbol{\varepsilon}$$

must be considered which boils down to the calculation of

$$B_\gamma^{(i)T} \mathbf{C} B_\gamma^{(i)} \quad (5.1.8)$$

and corresponding matrix products. To enable fast and efficient linear algebra calculations, the Intel Math Kernel Library (MKL) comes in place. It provides an efficient implementation of the well-known Basic Linear Algebra System (BLAS) routines optimized for Intel processors. However, the use of the general matrix-matrix product routine `gemm` ([MKL12, p. 124] is not a good guess here: It should be noted that for the cylindrical test case, the matrix

$$\mathbf{C} = \begin{pmatrix} L_1 + 2L_2 & \frac{L_1}{(R+h)^2} & 0 \\ \frac{L_1}{(R+h)^2} & \frac{L_1+2L_2}{(R+h)^4} & 0 \\ 0 & 0 & \frac{2L_2}{(R+h)^2} \end{pmatrix}$$

can be decomposed into a product  $\mathbf{C}_R^T \mathbf{C}_R$  with

$$\mathbf{C}_R = \begin{pmatrix} \sqrt{L_1 + 2L_2} & \frac{L_1}{(R+h)^2 \sqrt{L_1+2L_2}} & 0 \\ 0 & 2\sqrt{\frac{L_1 L_2 + L_2^2}{(R+h)^4 (L_1+2L_2)}} & 0 \\ 0 & 0 & \frac{\sqrt{2L_2}}{R+h} \end{pmatrix}.$$

Therefore, the product in (5.1.8) can be rewritten as

$$\left( \mathbf{C}_R B_\gamma^{(i)} \right)^T \left( \mathbf{C}_R B_\gamma^{(i)} \right).$$

This can be calculated as a matrix-matrix-product with one matrix being triangular (`trmm`, [MKL12, p. 147]) followed by a symmetric rank-k update (`syrk`, [MKL12, p. 141]), which is much more efficient than using the general matrix-matrix product. The same is possible for the other products containing  $\mathbf{C}$  and also for the products containing  $\mathbf{D}$ . Moreover, the calculation is done simultaneously for all  $i = 1, \dots, 9$  by concatenating the matrices  $B_\gamma^{(i)}$  and using the fact that  $\mathbf{C}$  is independent of  $i$ . Further matrix additions and scalar multiplications (e.g. to construct  $\boldsymbol{\varepsilon}$ ) can be executed efficiently using `daxpy` ([MKL12, p. 54]).

The integral itself is evaluated using Gauss quadrature with three points in each direction. This is sufficient to integrate the ansatz functions exactly and for an accurate quadrature of the entries in  $\mathbf{C}$ .

**Parallelization** The single element matrices can be calculated independently of each other, so this step is suited well for parallelization. To obtain this, the OpenMP library is used. It provides an easy way to carry out independent loop iterations in parallel by adding `!$omp parallel do` and `!$omp end parallel do` statements around the loop ([CJP08, p. 58]).

In addition, other spacial integrations appearing in the evaluation of the objective function (see Subsection 5.3.1) or in the calculation of the finite element gradient (see Subsection 5.3.2) can be carried out element-wise in parallel. To provide thread-safe additions during the calculation (instead of using an additional assembly step after the loop), the `!$omp reduction` command is used ([CJP08, pp. 105ff.]).

**Assembly of the stiffness matrix** The element stiffness matrices calculated by evaluating the bilinear form in the last paragraph are assembled and build up the left-hand-side of a linear equation system for the nodal values of the displacement. The stiffness matrix is usually large, sparse and – due to the elliptic character of the steady-state equation – positive definite. The solver Pardiso (Parallel Direct Sparse Solver) contained in the MKL is constructed to solve large sparse systems in an efficient ([GSH07]) way. As input it needs the stiffness matrix in a special format for sparse matrices ([MKL12, p. 2339]), called *Compressed Sparse Row* (CSR) (see e.g. [Bar+94, p. 57]):

Given a sparse matrix  $A \in \mathbb{R}^{N \times N}$  with `#nnz` non-zero entries. To represent  $A$ , three vectors are constructed:  $\mathbf{rInd} \in \mathbb{R}^{N+1}$ ,  $\mathbf{c} \in \mathbb{R}^{\#nnz}$  and  $\mathbf{v} \in \mathbb{R}^{\#nnz}$ . The entries inside  $\mathbf{rInd}$ , for example

$$\mathbf{rInd}(k) = i_k \quad \text{and} \quad \mathbf{rInd}(k+1) = i_{k+1}$$

mean that the  $k$ -th row of  $A$  is represented by the values from  $\mathbf{c}(i_k)$  through  $\mathbf{c}(i_{k+1}-1)$  and corresponding entries in  $\mathbf{v}$ . The value in  $\mathbf{c}(i)$  stands for the column of  $A$  in which the value  $\mathbf{v}(i)$  occurs. In other words,

$$\mathbf{c}(i) = j \text{ and } \mathbf{v}(i) = a \rightsquigarrow A_{kj} = a, \quad i = \mathbf{rInd}(k), \dots, \mathbf{rInd}(k+1) - 1.$$

In addition,  $\mathbf{rInd}(N+1) := \#nnz + 1$  is set and  $\mathbf{c}$  must be in strictly ascending order inside a row of  $A$ . Although the MKL provides methods to create a sparse matrix in CSR format out of a full matrix ([MKL12, p. 307]), it is time- and memory-consuming to construct the full stiffness matrix from the single element matrices before switching to CSR format. Since no other tools could be found to construct the CSR stiffness matrix directly from the element matrices, this had to be implemented by hand.

The position of the non-zero elements in  $A$  is determined by the element-to-node assignment of the discretization. By traversing each element stiffness matrix with its corresponding nodes, the vectors  $\mathbf{c}$  and  $\mathbf{v}$  are build gradually. Given the matrix  $E_i$  to element  $i$  with element nodes  $n(1)$  to  $n(9)$ , each entry of  $E_{jk}$  must be added to the stiffness matrix  $A$  at position  $(n(j), n(k))$ . The main idea is to use a linked (ordered) list to manage the entries for each row in  $A$ . The entries in those lists are

used later to create the vectors  $\mathbf{c}$  and  $\mathbf{v}$ . When adding an entry to  $A$  at position  $(n(j), n(k))$ , the list of row  $n(j)$  is traversed to find the correct position of  $n(k)$ : If the entry  $n(k)$  does not exist, a new one is created between the next bigger and smaller entry to keep the list ordered. The value  $E_{jk}$  is stored at the same position inside the list. Alternatively, if the entry  $n(k)$  already exists,  $E_{jk}$  is added to the value already stored at this position. After having processed all element matrices, the vectors  $\mathbf{c}$  and  $\mathbf{v}$  are created by concatenating the lists for each row. The fact that the single lists were ordered ensures an ascending order of the vector  $\mathbf{c}$  inside each row, as required for the CSR format. The vector  $\mathbf{rInd}$  is finally created by counting the entries in each single list.

The algorithm used is shown in Listing 5.1. It provides an efficient way to create the stiffness matrix in CSR format. Further adjustments were made in the implementation to account for the symmetry of the stiffness matrix and for multiple degrees of freedom per node resulting from the six coordinates that must be determined. The implementation of the linked list is done using [Ble09] and [Ola16].

---

**Algorithm 5.1** Pseudo-code for CSR format (one degree of freedom per node)

---

```

1: for  $i = 1$  to  $\#Elements$  do
2:   for  $j = 1$  to 9 do
3:     for  $k = 1$  to 9 do
4:       Search the correct position for the  $(j, k)$ -entry of the  $i$ -th element matrix
         in the ascending ordered list of row  $n(j)$ 
5:       Create a new list entry or add to an existing entry
6:     end for
7:   end for
8: end for
9: for  $i = 1$  to  $N$  do
10:  Save the entries in the list of node  $i$  consecutively in  $\mathbf{c}$  and  $\mathbf{v}$ 
11:  Count the entries and adjust  $\mathbf{rInd}$ 
12: end for

```

---

**Evaluation of the linear form** The right-hand-side of (5.1.5) is calculated using standard Gauss quadrature with three points in each direction. The resulting element load vectors are assembled simultaneously with the element stiffness matrices.

### 5.1.3. Solution of the linear system

The discretization of (5.1.1) finally leads to a linear system of equations

$$K_h Y_h = F_h \quad (5.1.9)$$

with a large sparse positive definite left-hand-side  $K_h \in \mathbb{R}^{6N \times 6N}$  and the vector  $Y_h$  collecting the (six) nodal values  $(\hat{u}^{(i)}, \hat{\theta}^{(i)})$ ,  $i = 1, \dots, N$ . The solution using

Pardiso involves three phases ([MKL12, p. 2344]): First, a reordering phase, where a permutation of  $K_h$  is created to reduce the fill-in. Second, the actual factorization of  $K_h$  into a product  $L_h L_h^T$  and at last a forward/backward-substitution.

During the optimization process, the above system (5.1.9) has to be solved several times for different values of  $\tau_h$ . Note that the matrix  $K_h$  shares the same structural pattern for all  $\tau_h$ , so the results from the reordering phase can be reused which saves roughly a third of the time for the solution.

In addition, Pardiso offers the possibility to use an indirect pcg-method instead of the above described direct factorization method ([MKL12, p. 2364]). While indirect methods are in general a good choice for large sparse systems ([Saa03, Preface]), they rely on a good preconditioner ([Saa03, ch. 9]), which is not always available. From the multiple solution of (5.1.9), benefits can be gained from a combination of direct and indirect methods which is described in Subsection 5.3.2.

## 5.2. Solution of the dynamic equation

The dynamic state equation has a hyperbolic character and thus needs a different numerical treatment than the steady-state equation. The dynamic equation to be solved is (see (4.1.5))

$$\begin{aligned} M_{(\tau)} \dot{y}(t) + L_{(\tau)} y(t) &= F(t) \quad \text{in } \mathcal{V}^* \text{ for } t \in [0, T] \\ y(0) &= 0, \quad \dot{y}(0) = \Phi_g(\tau) \end{aligned} \quad (5.2.1)$$

for  $y = (u, \theta) \in W(0, T)$ . Such a hyperbolic equation can be solved by reducing it to a system of first order equations and using a Crank-Nicolson scheme. This approach is taken in [Nes10b] and is also followed in this work. By introducing the new vector-valued function  $w := \dot{y}$ , equation (5.2.1) reads

$$\begin{aligned} \dot{y}(t) - w(t) &= 0, \quad y(0) = 0 \\ M_{(\tau)} \dot{w}(t) + L_{(\tau)} y(t) &= F(t), \quad w(0) = \Phi_g(\tau) \end{aligned} \quad (5.2.2)$$

in  $\mathcal{V}^*$  for all  $t \in [0, T]$ . This equation system is discretized on an equidistant time-grid

$$0 = t_1 < t_2 < \dots < t_M = T$$

with step size  $s = t_m - t_{m-1}$ . Let the functions

$$y^m := y(t_m) \in \mathcal{V}, \quad w^m := w(t_m) \in \mathcal{H}, \quad F^m := F(t_m) \in \mathcal{H}$$

be defined. Note that from the stronger regularity assumptions on the initial data (4.1.31), it actually holds  $w^m \in \mathcal{V}$  if  $F \in H^1(0, T; \mathcal{H})$ . The time-derivative in (5.2.2) is approximated using a Crank-Nicolson scheme,

$$\begin{aligned} y^m &= y^{m-1} + \frac{s}{2} (w^m + w^{m-1}) \\ M_{(\tau)} w^m &= M_{(\tau)} w^{m-1} + \frac{s}{2} (-L_{(\tau)} (y^m + y^{m-1}) + F^m + F^{m-1}). \end{aligned} \quad (5.2.3)$$

The spatial discretization involves the evaluation of  $\langle L_{(\tau_h)}(\cdot), \cdot \rangle_{\mathcal{V}^* \times \mathcal{V}} = a_{(\tau_h)}(\cdot; \cdot)$  which is done in the same manner as in Subsection 5.1 leading to the stiffness matrix  $K_h$ . The discretization of  $M_{(\tau_h)}$  is done by inserting the ansatz functions into  $(M_{(\tau_h)} \cdot, \cdot)_{\mathcal{H}}$ . This gives the mass matrix  $M_h \in \mathbb{R}^{6N \times 6N}$ . Let the vectors  $Y_h^m$  and  $W_h^m$  collect the (six) components of  $y^m$  and  $w^m$  at the  $N$  nodes, i.e.

$$Y_h^m := (y_1^m, \dots, y_N^m)^T \in \mathbb{R}^{6N}, \quad W_h^m := (w_1^m, \dots, w_N^m)^T \in \mathbb{R}^{6N}.$$

Then, for every time-step  $t_{m-1} \rightarrow t_m$  the system

$$Y_h^m = Y_h^{m-1} + \frac{s}{2} (W_h^{m-1} + W_h^m) \quad (5.2.4a)$$

$$M_h W_h^m = M_h W_h^{m-1} - \frac{s}{2} K_h (Y_h^m + Y_h^{m-1}) + \frac{s}{2} (F_h^m + F_h^{m-1}) \quad (5.2.4b)$$

has to be solved. Multiplication of (5.2.4a) with  $M_h$  and inserting (5.2.4b) leads to

$$\begin{aligned} \left( M_h + \frac{s^2}{4} K_h \right) Y_h^m &= \left( M_h - \frac{s^2}{4} K_h \right) Y_h^{m-1} + s M_h W_h^{m-1} + \frac{s^2}{4} (F_h^m + F_h^{m-1}) \\ W_h^m &= \frac{2}{s} (Y_h^m - Y_h^{m-1}) - W_h^{m-1}. \end{aligned} \quad (5.2.5)$$

The above equation has to be solved for every time-step, in total  $(M - 1)$ -times. Note that the matrices  $M_h$  and  $K_h$  do not differ at all between any two time-steps which saves a considerably amount of time during the calculation, which is done as follows:

---

**Algorithm 5.2** Calculating the dynamic solution

---

- 1: Calculate the element matrices of  $M_h$  and  $K_h$  for thickness  $\tau_h$
  - 2: Assemble  $\left( M_h + \frac{s^2}{4} K_h \right)$  in CSR format
  - 3: Use Pardiso to calculate the factorization of the above matrix
  - 4: **for**  $m = 1$  to  $M - 1$  **do**
  - 5:   Calculate force  $F_h^m$  and right-hand-side of (5.2.5)
  - 6:   Calculate  $Y_h^m$  by forward/backward-substitution
  - 7:   Calculate  $W_h^m$
  - 8: **end for**
- 

Note that the reuse of the factorization in line 6 of Algorithm 5.2 saves about 95 % of the time compared to an additional (redundant) factorization.

### 5.3. Solution of the optimization problem

In order to solve the optimization problem, the main part of the work has been done in the previous sections by solving the state equation. There are the objective functional and the remaining constraints left to be evaluated, which is discussed

first. Since those evaluations are similar for steady-state and dynamic case, they are explained simultaneously in this section. Afterwards, the finite-dimensional approximation to the reduced problem is presented and its solution using Ipopt is discussed.

### 5.3.1. Creating a finite-dimensional problem

In this subsection the creation of a finite-dimensional approximation to the optimization problems (3.2.1) and (4.2.1) is discussed.

**Evaluation of the objective functional** Having calculated the nodal values  $Y_h$  of the steady-state solution, the evaluation of the objective functional

$$J(u_h, \theta_h; \tau_h) = \int_{\mathcal{S}} u_{3,h} f \, dS + \frac{\lambda}{2} \|\tau_h\|_{H^1(\mathcal{S})}^2$$

is done using Gauss quadrature with the same Gauss points as in the evaluation of the bilinear form. The interpolated functions  $(u_h, \theta_h)$  as well as  $\tau_h$  are constructed with help of the shape functions as in (5.1.3) and (5.1.4). The evaluation of the dynamic objective

$$J(u_h, \theta_h; \tau_h) = \int_0^T \int_{\mathcal{S}} u_{3,h}(t) f(t) \, dS \, dt + \frac{\lambda}{2} \|\tau_h\|_{H^1(\mathcal{S})}^2 \quad (5.3.1)$$

from the nodal values  $Y_h^m$  is done in the same way using Gauss quadrature in the spatial domain and the trapezoidal rule for the time domain. Since both integrals only depend on the nodal values, this evaluation induces functions

$$J_h : \mathbb{R}^{6N} \times \mathbb{R}^N \rightarrow \mathbb{R} \quad \text{resp.} \quad J_h : \mathbb{R}^{6N \times M} \times \mathbb{R}^N \rightarrow \mathbb{R}, \quad (5.3.2)$$

which map the nodal values to the corresponding objective value.

**Evaluation of the remaining constraints** Let the nodal values of the thickness be given together with  $\tau_h$  as in (5.1.4). The volume constraint

$$\int_{\mathcal{S}} \left( \tau_h + \frac{K\tau_h^3}{12} \right) \, dS = C$$

is evaluated using Gauss quadrature leading to restrictions on the nodal values  $\tau^{(i)}$ ,  $i = 1, \dots, N$ . In the case of a parabolic midsurface, those restrictions are even linear because the integral reads

$$\int_{\mathcal{S}} \tau_h \, dS = C \quad (5.3.3)$$

as discussed in (2.3.3). The minimum and maximum thickness constraint is also enforced only on the nodal values leading to  $2N$  additional constraints. All those constraints can be collected in the set  $U_{\text{ad,h}}$  which is defined as

$$U_{\text{ad,h}} := \{T_h \in \mathbb{R}^N \mid f(T_h) - C = 0, \tau_{\min} \leq T_h \leq \tau_{\max}\}$$

where “ $\leq$ ” has to be understood componentwise.

**Creating the finite-dimensional approximation** When collecting the nodal values of the thickness in the vector

$$\mathbf{T}_h := \left( \tau^{(1)}, \dots, \tau^{(N)} \right)^T$$

and solving the linear equation(s) (5.1.9) for the steady-state case or (5.2.5) for the dynamic case, there is a function

$$G_h : \mathbb{R}^N \rightarrow \mathbb{R}^{6N} \quad \text{resp.} \quad G_h : \mathbb{R}^N \rightarrow \mathbb{R}^{6N \times M}$$

which maps  $\mathbf{T}_h$  to the nodal values of the approximated solution of the state equation. The concatenation of  $G_h$  and  $J_h$  from (5.3.2) builds an discretized reduced objective

$$J_{s,h} : \mathbb{R}^N \rightarrow \mathbb{R}, \quad J_{s,h}(\mathbf{T}_h) := J_h(G_h(\mathbf{T}_h), \mathbf{T}_h)$$

which serves as an approximation to the reduced objective from (3.2.3) or (4.2.3). Adding the remaining side constraints leads to the finite-dimensional approximation to the problems (3.2.1) or (4.2.1),

$$\min_{\mathbf{T}_h \in U_{\text{ad},h}} J_{s,h}(\mathbf{T}_h). \quad (5.3.4)$$

### 5.3.2. Solution of the finite-dimensional problem

Problem (5.3.4) is a “standard” finite-dimensional nonlinear optimization problem. Since  $U_{\text{ad},h}$  describes a compact subset of  $\mathbb{R}^n$  and the objective  $J_{s,h}$  is continuous, it has a solution by the Weierstrass theorem. This solution can be approximated by various methods: active-set strategies, SQP methods or trust-region algorithms are only a small part of the manifold methods available. Amongst these options, the interior-point method has gained much attention over the last decades ([FGW02]) and it is chosen to solve the above optimization problem.

A sophisticated implementation of the interior-point method can be found in the package Ipopt ([Kaw+15]). The library is equipped with a Fortran interface which is used inside the program. The key concept of interior-point optimization and the details of the implementation are explained in detail in [WB06]. In short, the interior-point method creates a sequence of points (vectors  $\mathbf{T}_h^{(i)}$  of thickness values) inside the feasible region ( $U_{\text{ad},h}$ ) that approximates an optimal solution – so it is an iterative method. In each iteration, the following information is provided to Ipopt for a fixed thickness vector  $\mathbf{T}_h^{(i)}$ :

- value of the objective  $J_{s,h}(\mathbf{T}_h^{(i)})$ ,
- gradient of the objective  $\nabla J_{s,h}(\mathbf{T}_h^{(i)})$ ,
- value of the constraints  $f(\mathbf{T}_h^{(i)}) - C$ ,
- gradient of the constraint  $\nabla f(\mathbf{T}_h^{(i)})$ .

The information about upper and lower bounds on  $T_h$  as well as general problem data (such as dimension and optimization parameters) are provided once at the beginning. The calculation of points one and three was discussed in the previous subsection. In the particular case of a parabolic midsurface, the constraint  $f(T_h^{(i)}) - C$  becomes a linear one,

$$c_{\text{vol}}^T T_h^{(i)} - C = 0,$$

so the gradient is the vector  $c_{\text{vol}}^T \in \mathbb{R}^N$  itself. In addition, Ipopt provides special treatment for linear constraints ([Kaw+15, p. 72]) which avoids the query of their (constant) gradient in every iteration.

The repeated calculation of the objective value involves the solution of the linear system(s) (5.1.9) or (5.2.5). After calculating a search direction for the next iterate  $T_h^{(i+1)}$ , a line search is performed in Ipopt. Between two subsequent line search steps, the nodal thickness values are not “too far away” from each other, so by continuity the corresponding left-hand-sides of the equation systems do not “differ much” either. As mentioned at the end of Section 5.1, the solver Pardiso offers – beside of a direct solution via factorization – also an indirect (iterative) pcg-method [MKL12, p. 2364f.]. Indirect methods rely heavily on a good preconditioner of the system. Once the factorization for  $T_h^{(i)}$  is done, the pcg-method is used for subsequent line search steps involving the thickness vector  $T_h^{(i)} + \alpha_k d$ . The previously calculated factorization  $L_{(T_h^{(i)})} L_{(T_h^{(i)})}^T$  serves as a well-suited preconditioner, since only one forward/backward-substitution is needed to approximate the inverse of  $K_{(T_h^{(i)} + \alpha_k d)}$ .

Although the Pardiso pcg-method works well, there is an additional possibility to save computation time: As communicated by Nestler ([Nes13]), the pcg-method can be implemented without the need to assemble the stiffness matrix explicitly. So an own implementation of the pcg-method is used in the program instead of the one provided by Pardiso. This saves about 20 % of the computation time for the solution of the linear systems during the line search. Clearly, if the pcg-method uses too much iterations (more than about 5), the program switches back to a direct factorization.

It can be asked whether the factorization can also be reused during subsequent optimization iterations, i.e. between  $T_h^{(i)}$  and  $T_h^{(i+1)}$ . While this is working principally, at the beginning of each iteration the gradient  $\nabla J_{s,h}(T_h^{(i)})$  must be provided. This is a crucial point of the optimization process, so the accurate solution using direct factorization is preferred in comparison to an iterative solution of the system.

The point which is left to be discussed is the actual calculation of the gradient  $\nabla J_{s,h}$ :

**Providing the objective gradient** The straight-forward way of providing the objective gradient is to approximate it by finite differences. This method does not need any additional information other than the value of the objective itself at certain

disturbed points in a neighborhood of  $T_h^{(i)}$ . Serious drawbacks are the missing accuracy and multiple evaluations of the objective function (and hence solutions of the state equation). The missing accuracy of the gradient leads to unsatisfying optimization results as shown in [Nes12, p. 201]. This is the reason why some effort has been put into the calculation of the reduced objective gradient. The steady-state gradient from Theorem 3.6 involves the evaluation of

$$J'_s(\tau_h)q = -Z_q(u_{(\tau_h)}, \theta_{(\tau_h)}) + \lambda(\tau_h, q)_{H^1(\mathcal{S})}$$

from the nodal values of  $\tau_h$  collected in  $T_h$ . Because of the special choice of the objective as the compliance functional, an additional solution of the adjoint system is avoided. The directions  $q$  are chosen as the ansatz functions  $h_i$ . The evaluation of  $Z_q$  as well as the  $H^1(\mathcal{S})$  scalar product are done using Gauss quadrature. The nodal values of  $u_{(\tau_h)}$  and  $\theta_{(\tau_h)}$  are available from the solution of the state equation. The so called *finite element gradient* reads

$$\nabla J_{s,h}(T_h) \approx (J'_s(\tau_h)h_1, \dots, J'_s(\tau_h)h_n) =: J'_{s,h}(T_h)$$

and is a much better approximation to  $\nabla J_{s,h}(T_h)$  than finite differences would create.

The situation is a bit more complex in the dynamic case. First, an additional adjoint system has to be solved backwards in time, namely: Find  $p \in W(0, T)$  such that

$$\begin{aligned} M_{(\tau)}\ddot{p}(t) + L_{(\tau)}p(t) &= F(t) \\ p(T) = \dot{p}(T) &= 0 \end{aligned} \tag{5.3.5}$$

in  $\mathcal{V}^*$  for all  $t \in [0, T]$ . The reformulation as a first order system and the discretization is done as for the dynamic state equation in (5.2.3) except that  $p^m = p(t_m)$  is calculated from  $p^{m+1} = p(t_{m+1})$ . Let the nodal values of  $p^m$  and its first derivative  $r^m = \dot{p}(t_m)$  be collected in

$$P_h^m := (p_1^m, \dots, p_N^m) \in \mathbb{R}^{6N}, \quad R_h^m := (r_1^m, \dots, r_N^m) \in \mathbb{R}^{6N}.$$

Then, in every time-step  $t_{m+1} \rightarrow t_m$  the equation system

$$\begin{aligned} \left( M_h + \frac{s^2}{4} K_h \right) P_h^m &= \left( M_h - \frac{s^2}{4} K_h \right) P_h^{m+1} - sM_h R_h^{m+1} + \frac{s^2}{4} (F_h^m + F_h^{m+1}) \\ R_h^m &= \frac{2}{s} (P_h^{m+1} - P_h^m) - R_h^{m+1} \end{aligned} \tag{5.3.6}$$

has to be solved. Note that in case of other objective functionals the right-hand-side  $F$  of the adjoint equation would be replaced by another  $\tilde{F}$  which could also depend on the solution  $y$  (see Remark 4.7). Additionally, note that the factorization of  $M_h + \frac{s^2}{4} K_h$  has already been calculated during the solution of the state equation (see Algorithm 5.2, line 3), so it can be reused here. The calculation of the adjoint boils down to:

**Algorithm 5.3** Calculating the adjoint solution

- 
- 1: **for**  $m = M - 1$  to 1 **do**
  - 2:   Calculate  $F_h^m$  (resp.  $\tilde{F}_h^m$  which could depend on  $Y_h^m$ )
  - 3:   Calculate right-hand-side of (5.3.6)
  - 4:   Calculate  $P_h^m$  by forward/backward-substitution
  - 5:   Calculate  $R_h^m$
  - 6: **end for**
- 

The forward/backward-substitutions are carried out efficiently by the Pardiso solver. The gradient of the reduced objective, as calculated in Theorem 4.6, reads

$$J'_s(\tau)q = - \int_0^T \left( (M'_{(\tau)}q)\ddot{y}_{(\tau)}(t) + Z_q(t), p(t) \right)_{\mathcal{H}} dt + (M_{(\tau)}p(0), \Phi'_g(\tau)q)_{\mathcal{H}} + \lambda(\tau, q)_{H^1(\mathcal{S})}.$$

However, the second derivative of  $y$  is not readily available from the values  $Y_h^m$  and  $W_h^m$ . To overcome this, a partial integration is carried out which leads to

$$J'_s(\tau)q = \int_0^T \left( (M'_{(\tau)}q)\dot{y}_{(\tau)}(t), \dot{p}(t) \right)_{\mathcal{H}} - (Z_q(t), p(t))_{\mathcal{H}} dt + (M_{(\tau)}p(0), \Phi'_g(\tau)q)_{\mathcal{H}} + \left( (M'_{(\tau)}q)\dot{y}_{(\tau)}(0), p(0) \right)_{\mathcal{H}} + \lambda(\tau, q)_{H^1(\mathcal{S})}. \quad (5.3.7)$$

This can be evaluated by using the values of  $Y_h^m, W_h^m, P_h^m, R_h^m$  and  $T_h$ . As in the steady-state case, the direction  $q$  is chosen as the ansatz functions  $h_i$ , so the finite element gradient reads

$$\nabla J_{s,h}(T_h) \approx (J'_s(\tau_h)h_1, \dots, J'_s(\tau_h)h_n) =: J'_{s,h}(T_h)$$

and is a good approximation to  $\nabla J_{s,h}(T_h)$ .

**Providing a start solution** Another crucial point for optimization methods is providing a suitable initial estimate of the optimal solution. A good starting solution reduces the necessary iterations to only a few while a bad estimate can prevent the optimization process from succeeding. On coarse grids this is not as deciding as on fine grids because one iteration takes less time. So it is reasonable to solve the optimization problem on successively refined grids, where the optimal solution on the coarser grid is used as an estimate on the finer grid.

In the examples presented in the next chapter, the calculation started on a coarse grid consisting of only 289 nodes with the constant thickness as a very general estimate of the optimal solution. The grid is refined to half the step size (leading to a fourfold number of nodes) and the optimization is restarted with the interpolated solution of the coarse grid as the initial estimate. This process is repeated three (dynamic case) to four (steady-state case) times and the optimization even succeeded on fine grids consisting of around 16000 resp. 66000 nodes in reasonable time. This would not be possible with a constant thickness as an initial estimate on the fine grids.

**Saving the result** The optimization results, i.e. the nodal values of the optimal thickness as well as the optimal displacement and adjoint together with general problem information must be saved for further postprocessing. The plain save of the values in a text file would cause a huge amount of disk space, since those formats are not optimized for keeping large numbers of numerical values.

The “mat” file format known from Matlab is considerably better suited for such tasks. It enables much smaller result files and also time-efficient read and write processes. A Fortran library for reading and writing mat-files ([Mat15b]) as well as for creating suitable structures to be saved in mat-files ([Mat15a]) is used in the implementation.

### 5.3.3. Optimality test

While the finite element gradient accelerates the optimization significantly, it can also be used to evaluate the necessary conditions for an optimal solution. This evaluation can be used as an “optimality test” to measure how “far away” the calculated solution is from an exact solution to the finite-dimensional problem. In detail, the necessary conditions read

$$J'_s(\tau^*)(q - \tau^*) \geq 0 \quad \text{for all } q \in U_{\text{ad}} \quad (5.3.8)$$

if  $\tau^*$  is an optimal solution, see Theorem 3.7 and 4.7. Let  $\mathbf{T}_h^*$  collect the nodal values of  $\tau^*$ , then by using the discretized version of  $J'_s$ , equation (5.3.8) reads

$$J'_{s,h}(\mathbf{T}_h^*)(Q_h - \mathbf{T}_h^*) \geq 0 \quad \text{for all } Q_h \in U_{\text{ad},h}. \quad (5.3.9)$$

As done in [Nes12], this equation can be reformulated as an optimization problem,

$$(5.3.9) \iff J'_{s,h}(\mathbf{T}_h^*)\mathbf{T}_h^* = \min_{Q_h \in U_{\text{ad},h}} J'_{s,h}(\mathbf{T}_h^*)Q_h. \quad (5.3.10)$$

In case of a parabolic midsurface, the set  $U_{\text{ad},h}$  only consists of linear constraints, so the problem on the right-hand-side of the above equation is a linear program. Those can be solved in short time, e.g. using Matlab’s `linprog` procedure. In numerical calculations the optimal solution  $\mathbf{T}_h^*$  will not be reached in general, so equation (5.3.10) will evaluate to

$$\min_{Q_h \in U_{\text{ad},h}} J'_{s,h}(\mathbf{T}_h^*)Q_h - J'_{s,h}(\mathbf{T}_h^*)\mathbf{T}_h^* = -\varepsilon < 0. \quad (5.3.11)$$

The value  $\varepsilon$  can be used as an “optimality measure” to specify how “far away” the calculated  $\mathbf{T}_h^*$  is from a solution of the finite-dimensional problem (5.3.4), and hence from a solution of the continuous problems (3.2.1) or (4.2.1).

## 5.4. Some words on locking

The phenomenon of locking describes the overestimation of the stiffness of the shell by certain finite elements ([CB03, p. 211]). This results in inexact numerically

calculated displacements, in particular they become too small. Clearly, an inaccurate solution of the state equations can lead to wrong optimization results, as is described in the paper [CBB04].

However, locking mainly appears in very *thin* structures ([CB03, p. 212f.]). In the presented examples in Chapter 6 and 7, the thickness of the cylindrical shell and its length have a ratio of about 0.1 which counts rather to *moderately thick* shells. This is also the reason why the basic shell model was chosen prior to the Koiter model or the model that Nestler ([Nes12]) used (see also the discussion in Subsection 2.2.5).

On the other hand, to overcome locking different elements could be used. The so called MITC elements (Mixed Interpolation of Tensorial Components) avoid the locking phenomenon and are described in connection with general shell elements in the paper [BIC00]. The use of other shell elements can be an implementation feature for future work and provides the basis for an additional investigation of locking in thickness optimization.

## 5.5. Summary

In this chapter the numeric solution of the optimization problems (3.2.1) and (4.2.1) was discussed. First, the solution of the respecting state equations with help of the finite element method was investigated. Shell elements were introduced and implementation details for an time-efficient calculation of the element matrices and solution of the resulting sparse linear system were provided. For this, the Intel MKL and the solver Pardiso were used. It was also shown how benefits can be gained from the reuse of matrix factorizations in the dynamic case, where the finite element method and the Crank-Nicolson scheme were combined.

Afterwards, a finite-dimensional approximation to the continuous optimization problem was provided. It was shown how to solve it with the interior point method using the package Ipopt. Special treatment was given in providing a finite element gradient to the solver instead of using finite differences. Moreover, initial estimates were created by solving the optimization problem on successively refined grids. Finally, it was shown how to use the necessary conditions that were derived in the previous chapters as an optimality measure to the numerically calculated optimal solution.

While the algorithms and methods described in the chapter are known in general, the efficient implementation of the single process steps and the effective combination of the various available libraries are an own contribution to the numerical solution of optimal control problems in linear elasticity.

## 6. Results for the steady-state problem

In this chapter the numerical results for the steady-state optimization problem (3.2.1) are presented. Different test cases are chosen involving a shell with a cylindrical midsurface. Although the theoretical considerations in the previous chapters are valid for shells with a more general midsurface, the program files are optimized to account for cylindrical ones. Four different examples are presented, including not only the calculated optimal thickness but also convergence studies and the investigation of the “optimality measure” discussed in (5.3.11). In addition, the optimized displacement is compared to the initial displacement involving a constant thickness. Moreover, the results from Example 6.2 and 6.4 are verified by using available solutions for similar problems with more specialized shell models. The examples presented in Section 6.2 and 6.3 were also published in [Zie14].

### 6.1. General problem data

As mentioned above, shells with a cylindrical midsurface are investigated. The midsurface is modeled as presented in Subsection 2.2.1 by the parameter domain

$$\omega = (0, l) \times (\varphi_a, \varphi_b) \quad l > 0, \quad 0 \leq \varphi_a < \varphi_b < 2\pi,$$

and the mapping  $z : \omega \rightarrow \mathbb{R}^3$ ,

$$z(x, \varphi) = \begin{pmatrix} x \\ R \cos(\varphi) \\ R \sin(\varphi) \end{pmatrix},$$

where  $R \in \mathbb{R}$ ,  $R > 0$  is the radius of the midsurface. The length parameter is chosen to be  $l = 1$  over all examples; it can be considered to be in meters. The radius is taken as  $R = 1$  m and the angle  $\varphi$  varies between 0 and  $\pi/2$ , except in the second part of Example 6.2 where a full tube is considered and in Example 6.4 involving a beam model.

The material parameters are the same over all examples: Poisson’s ratio is  $\nu = 0.3$  and Young’s modulus is  $E = 210 \text{ kN mm}^{-2}$ . This corresponds to steel ([Kuc11, p. 624]). The incoming force density is in units of  $\text{MN m}^{-2} = \text{MPa}$ . This yields displacements in units of mm.

In all examples (except Example 6.4) the thickness  $\tau$  (also given in meters) is restricted to be within

$$0.05 \leq \tau(x, \varphi) \leq 0.15.$$

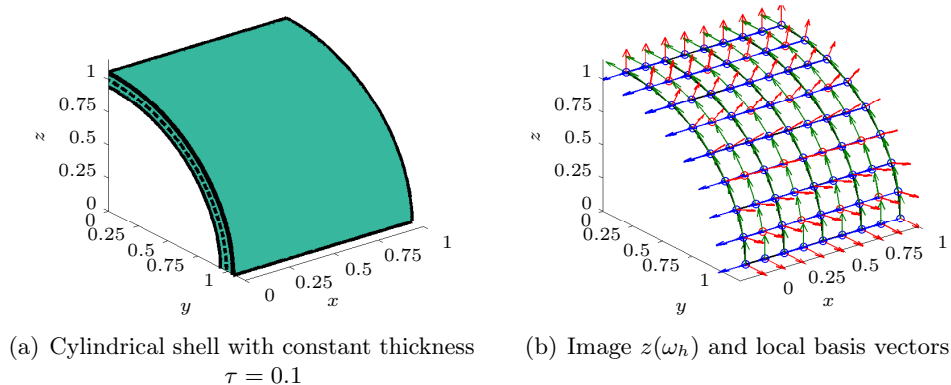


Figure 6.1.: Initial geometry

The optimization starts with a constant thickness of 0.1 which corresponds to a volume of  $C = 0.1(\varphi_b - \varphi_a)\text{m}^3$ . This volume is kept constant during the optimization. In Figure 6.1 the quarter tube with constant thickness is shown which serves as the initial geometry for the examples presented in Section 6.2, 6.3 and 6.5. Besides, the image  $z(\omega_h)$  of the discretized parameter region  $\omega_h$  is shown together with the local basis vectors at the nodes. The regularization parameter  $\lambda$  is chosen in a way that the regularization term is about two to three orders of magnitude smaller than the objective value.

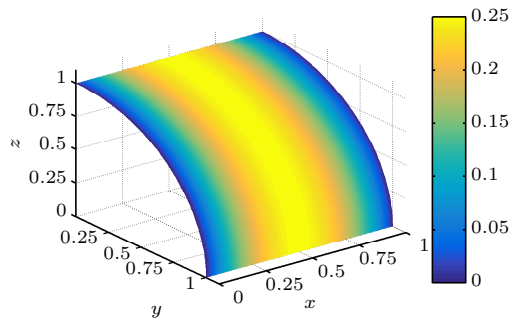


Figure 6.2.: Applied force for Example 6.2 as colored midsurface

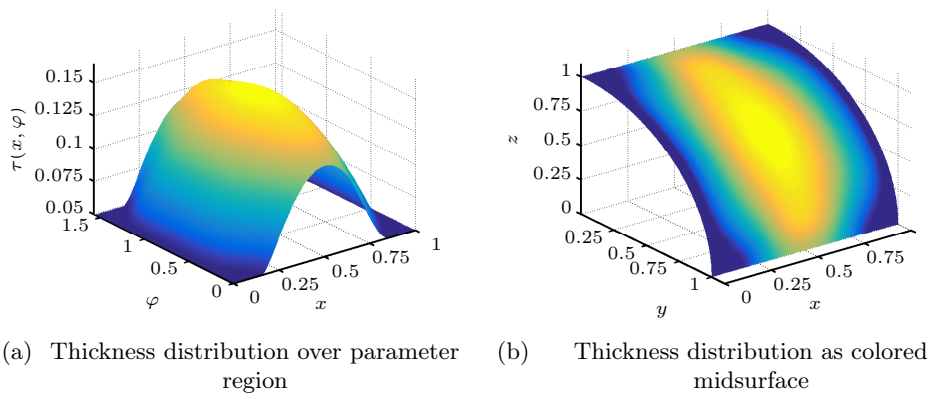


Figure 6.3.: Optimal thickness for Example 6.2

## 6.2. Rotational symmetric force

### 6.2.1. Optimization results

The first example considers the quarter of a tube which is subject to the transverse force

$$f(x, \varphi) = x(1 - x).$$

Its distribution is shown in Figure 6.2 as a colored midsurface. The panel is softclamped over its whole boundary, i.e. the translational displacement is forced to be zero (see the displacement space from (2.2.24)).

The optimal thickness distribution is shown in Figure 6.3. This is the result on a very fine grid consisting of about 66,000 nodes, which corresponds to a step size of  $2^{-8}$  in  $x$ -direction. It can be seen that the thickness follows approximately the incoming force. Nevertheless it is not independent from the angle  $\varphi$ . On the other hand, the symmetry with respect to the  $x = 0.5$ -axis is reflected well and also the symmetry with respect to the  $\varphi = 0.25\pi$ -axis can be observed.

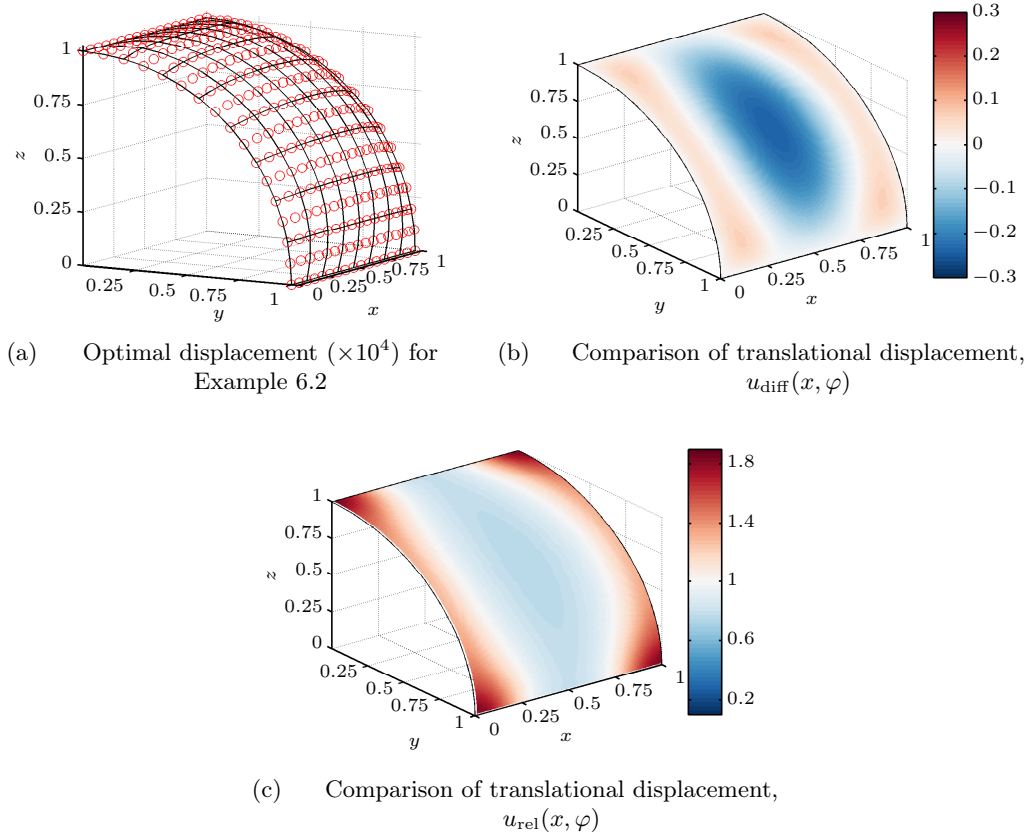


Figure 6.4.: Displacement for optimal thickness of Example 6.2

An investigation of the optimal displacement is shown in Figure 6.4: On the upper left-hand side the displacement for the optimal thickness  $\tau_{\text{opt}}$  is shown. The upper right picture illustrates the comparison of the translational displacement norm for the initial thickness  $\tau_{\text{const}} = 0.1$  and for the calculated optimal thickness. The difference is normalized with respect to a “characteristic displacement” taken as the maximal displacement norm, i.e.

$$u_{\text{diff}}(x, \varphi) := \frac{\|u_{h,\text{opt}}(x, \varphi)\|_2 - \|u_{h,\text{const}}(x, \varphi)\|_2}{\sup \|u_{h,\text{const}}(x, \varphi)\|_2}. \quad (6.2.1)$$

It can be noticed that the optimal thickness results in a displacement more than 20 % smaller compared to the constant thickness at the region where the force has biggest magnitude. In addition, a slightly increased displacement near the boundary can be observed. This increasing is even more obvious if the relative displacement change

$$u_{\text{rel}}(x, \varphi) := \frac{\|u_{h,\text{opt}}(x, \varphi)\|_2}{\|u_{h,\text{const}}(x, \varphi)\|_2} \quad (6.2.2)$$

step size $h_x$	$\ \tau_h - \tau_{\text{opt}}\ _{\infty, h}$
$2^{-4}$	$1.16 \cdot 10^{-2}$
$2^{-5}$	$1.23 \cdot 10^{-3}$
$2^{-6}$	$9.92 \cdot 10^{-4}$
$2^{-7}$	$6.87 \cdot 10^{-4}$

Table 6.1.: Comparison of optimal thickness on different grids for Example 6.2

is considered. This is presented in Figure 6.4(c), where the increased displacement in particular near the corners is visible. Nevertheless, the absolute displacement values near the boundary are nearly neglectable and so are the absolute differences. Hence, the overall smaller displacement is a clear gain from thickness optimization.

Furthermore, it is interesting to investigate the rate of convergence between successively refined grids. For this, the solution on the finest grid is taken as the “exact” solution. The comparison of the norms of the optimal thickness  $\tau_h$  from a coarser grid and  $\tau_{\text{opt}}$  restricted to that grid is shown in Table 6.1. From those results, convergence towards the exact solution can be presumed. Finally, the solution of the (linear) program (5.3.11) leads to the optimality measure  $\varepsilon = 4.5 \cdot 10^{-7}$ , which suggests a good approximation of the real solution.

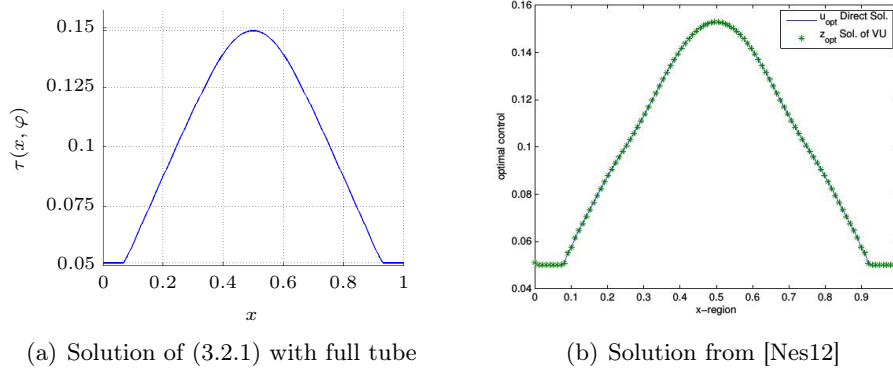


Figure 6.5.: Optimal thickness for Example 6.2 with a full tube

### 6.2.2. Comparison with results from Nestler

A similar problem was investigated by Nestler ([Nes12]). In his work the same force was chosen but applied to a full tube. A Koiter-like model was used (see discussion in 2.2.5) for the calculations. The objective was chosen as the compliance functional without a regularization term. Moreover, the thickness was modeled to be independent from the angle  $\varphi$ .

To enable comparable calculations, some modifications to the program were necessary: An additional introduction of  $N$  constraints forced the thickness to be independent of  $\varphi$ . Moreover, the boundaries  $\varphi = 0$  and  $\varphi = 2\pi$  had to be coupled to account for a full tube.

The comparison of the optimal thickness distributions is shown in Figure 6.5. Since the thickness is independent of  $\varphi$ , only the behavior along the  $x$ -axis is shown. Differences between the solution of the steady-state problem (3.2.1) and the optimal thickness calculated by Nestler are hardly visible. This shows that the model and its implementation can also be used for full tubes and serves as a verification of the results. On the other hand, although the model in this work does not suffer from the restrictions of specialized models, a “simpler” model (like in [Nes12]) seems to be sufficient in certain cases.

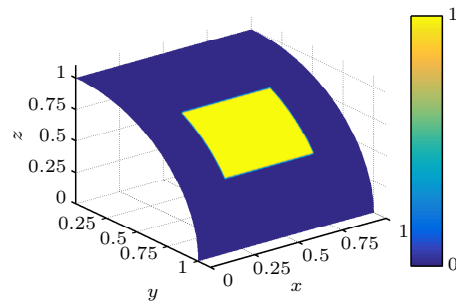
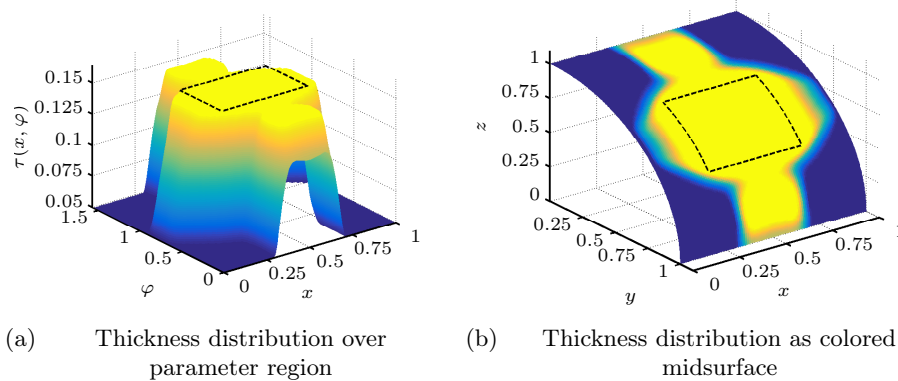


Figure 6.6.: Region (yellow) of applied force for Example 6.3



(a) Thickness distribution over parameter region

(b) Thickness distribution as colored midsurface

Figure 6.7.: Optimal thickness for Example 6.3, region of applied force (dashed-line)

### 6.3. Discontinuous force

The second example deals with a discontinuous force

$$f(x, \varphi) = \begin{cases} 1, & \text{if } (x, \varphi) \in \left[ \frac{1}{4}, \frac{3}{4} \right] \times \left[ \frac{\pi}{4} - \frac{1}{4}, \frac{\pi}{4} + \frac{1}{4} \right] \\ 0, & \text{else} \end{cases}$$

applied to the quarter of a tube. This setup is shown in Figure 6.6. Note that the weak solution theory in particular allows for discontinuous forces. As in the former example, the panel is softclamped over its whole boundary.

The optimal thickness distribution is shown in Figure 6.7. This is the result on a fine grid with a step size of  $2^{-8}$  in  $x$ -direction. It can be seen that the thickness follows approximately the incoming force. Certainly, it is interesting to observe that the shell is thickest in a cross-shaped area, whereas the force is applied on a rectangular region.

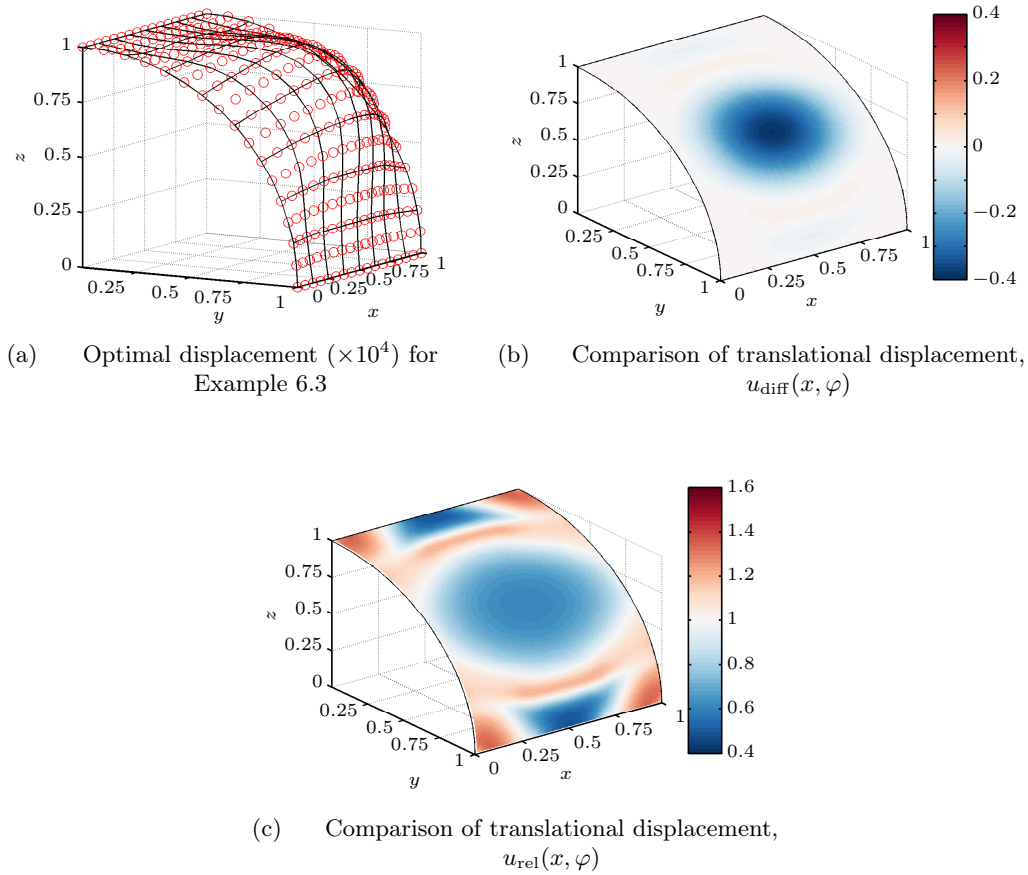


Figure 6.8.: Displacement for optimal thickness of Example 6.3

The displacement corresponding to the optimal thickness (referred to as *optimal displacement* in the following) and the comparison of the translational displacement norm is shown in Figure 6.8. In the upper left part, the displaced midsurface for the optimal thickness  $\tau_{\text{opt}}$  is shown. The upper right part presents the quantity  $u_{\text{diff}}(x, \varphi)$  from (6.2.1). It can be noticed from Figure 6.8(b) that the optimal thickness results in an up to 35 % smaller displacement compared to the constant thickness in the region where the force is applied. Outside this area, the displacement hardly shows (absolute) differences between constant and optimized thickness. The relative differences  $u_{\text{rel}}(x, \varphi)$  can be observed in Figure 6.8(c): Here, the bigger displacement outside the loaded region is clearly visible – up to 135 % near the corners of the shell. As in Example 6.2, the absolute displacement values near the boundary are nearly neglectable and so are the absolute differences.

The increased displacement surrounding the loaded region is due to the compliance objective: Outside the loaded area the displacement is not weighted at all, so the increasing displacement has no influence on the objective value.

Altogether, the overall smaller displacement resulting from thickness optimization

step size $h_x$	$\ \tau_h - \tau_{\text{opt}}\ _{\infty, h}$
$2^{-4}$	$4.45 \cdot 10^{-2}$
$2^{-5}$	$1.96 \cdot 10^{-2}$
$2^{-6}$	$7.04 \cdot 10^{-3}$
$2^{-7}$	$1.88 \cdot 10^{-3}$

Table 6.2.: Comparison of optimal thickness on different grids for Example 6.3

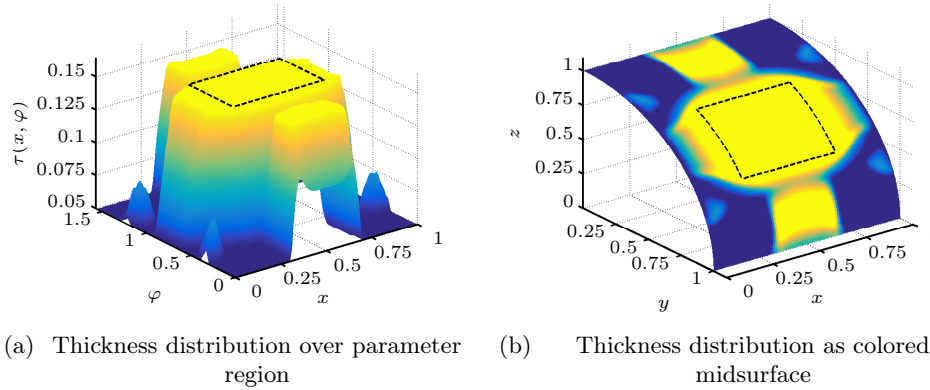


Figure 6.9.: Optimal thickness for Example 6.3 with unweighted objective

is a satisfying result. The difference between the solution on successively refined grids is shown in Table 6.2. As in Example 6.2, convergence towards the “exact” solution is noticeable, although the differences are about one order of magnitude bigger than in the previous example. This could be caused by the discontinuous load that is applied. Finally, the solution of the (linear) program (5.3.11) leads to the optimality measure  $\varepsilon = 6.2 \cdot 10^{-7}$ , which suggests a good approximation of the solution from (5.3.4).

The focus on the displacement minimization around the area of the incoming force is even more evident if the objective is changed to an unweighted one: As discussed in Itemization 2.3, the use of  $L^2(\mathcal{S})$ -norms for the displacement is another candidate for the objective function. In detail, it is given as

$$J_{\|\cdot\|}(u, \theta; \tau) = \|(u, \theta)\|_{\mathcal{H}}^2 + \frac{\lambda}{2} \|\tau\|_{H^1(\mathcal{S})}$$

with the  $L^2$ -space  $\mathcal{H}$  as in (4.1.3).

The optimal thickness profile and distribution is shown in Figure 6.9. The cross-shaped structure can be recognized as in the compliance objective case. Certainly, there are four small additional spikes near the  $x = 0$  and  $x = 1$ -boundary. As an offset to these spikes, the central thick region is separated from the outer part by a small “valley”.

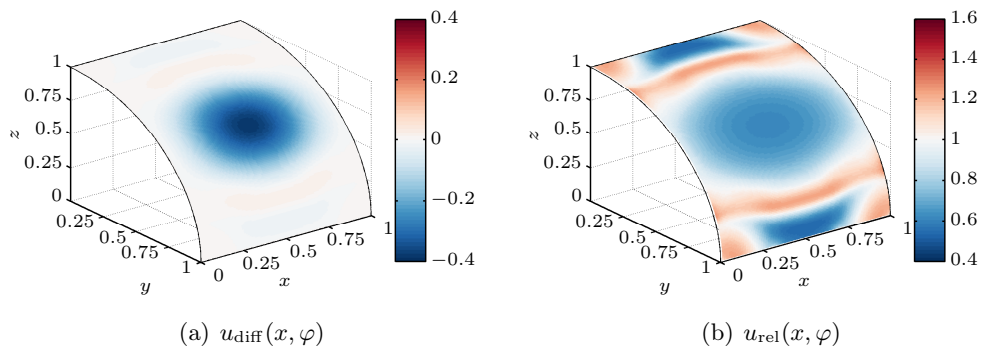


Figure 6.10.: Comparison of translational displacement for Example 6.3 with unweighted objective

The influence on the displacement can be seen in Figure 6.10. Overall, the difference is distributed more evenly: The gain at the loaded region is not as big as for the compliance objective. On the other hand, the increase of the displacement at the corners and around the central area is smaller. It depends on the particular application, which objective should be preferred.

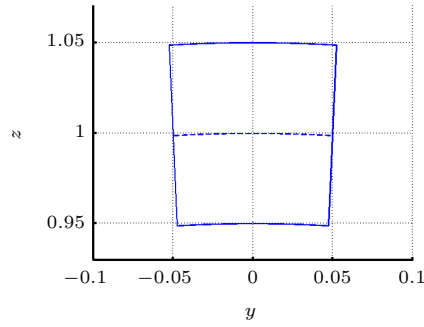


Figure 6.11.: Cross section of the beam from Example 6.4

## 6.4. Loaded beam

### 6.4.1. Optimization results

The problem of a loaded beam was communicated by Leliep ([Lel15]). Beams are a further specialization of shells, where not only one but *two* dimensions are thin compared to the third one. This leads to further simplifications in the model equations: A well-known one amongst these models is the Euler–Bernoulli beam, see for example [Ger04]. To account for the shell model in this work, the beam is approximated with help of a cylindrical midsurface with a very narrow interval given for the angle  $\varphi$ . In this example, the parameter domain is chosen as

$$\omega = (0, 1) \times \left( \frac{\pi}{2} - 0.05, \frac{\pi}{2} + 0.05 \right)$$

with an initial thickness of 0.1. The cross-section of the beam is shown in Figure 6.11. Note that the curvature of the midsurface is hardly visible, hence this parametrization serves as a good approximation to the beam. Additionally, the upper and lower bound for the thickness are expanded to 0.2 and  $10^{-4}$ , resp. This is due to the comparison between the numerical and the exact solution (for the Euler–Bernoulli model) presented in Subsection 6.4.2. Moreover, the thickness is forced to be independent from  $\varphi$  like in Subsection 6.2.2.

The beam is hardclamped at  $x = 0$ , i.e. there are neither translational nor rotational displacements allowed. The other three edges are free. Note that these boundary conditions forbid rigid body motions, as required for the existence and uniqueness of the state equation solution (see discussion in Subsection 3.1.1).

The beam is subject to a point load applied on the opposite of the clamped edge. Point loads are not included in the class  $L^2(\mathcal{S})$ , so an approximation must be used. This is done by using a discontinuous load applied on a small rectangular region,

$$f(x, \varphi) = \begin{cases} 1, & \text{if } x \geq 0.99, \\ 0, & \text{else.} \end{cases}$$

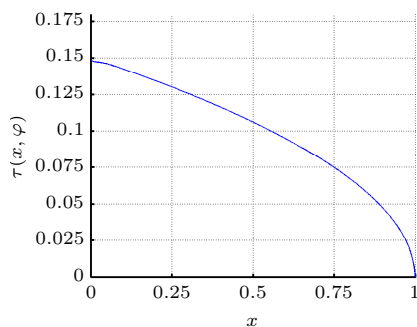
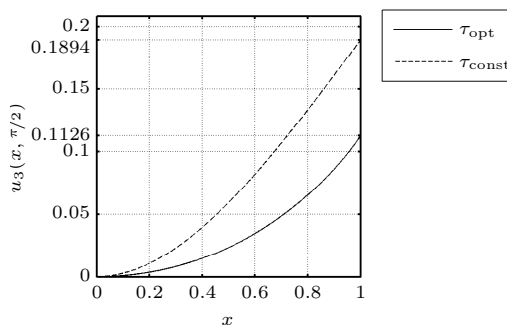


Figure 6.12.: Optimal thickness for Example 6.4

Figure 6.13.: Comparison of displacement component  $u_3$  for Example 6.4

The optimal thickness distribution is shown in Figure 6.12. This is the result on a fine grid with a step size of  $2^{-8}$  in  $x$ -direction. It can be seen that the beam is thickest at the hardclamped edge while the thickness tends to zero at the edge where the load is applied.

The comparison of the dominating displacement component  $u_3$  for constant and optimal thickness can be seen in Figure 6.13. The maximum of the optimized displacement component ( $\approx 0.1126$ ) is about 40 % smaller than in the case of a constant thickness ( $\approx 0.1894$ ). The evaluation of the linear program (5.3.10) (including additional restrictions on the thickness to be independent from  $\varphi$ ) gives  $\varepsilon = 3.77 \cdot 10^{-6}$ .

### 6.4.2. Comparison with analytic solution

The use of the Euler–Bernoulli beam model does not only simplify the model equations, but even enables an analytic solution of the problem. This is a great possibility to verify the numerical results and to study the difference of optimization results between the general basic shell model and the Euler-Bernoulli model. A related problem was studied in [HM03, p. 13]. The simplified optimization problem reads as follows:

$$\min_{\substack{\tau \in C(0,l) \\ y_1, y_2 \in C^2(0,l)}} J(y_1, y_2; \tau) = y_1(l) \quad (6.4.1a)$$

$$\text{s.t. : } y_1'(x) = y_2(x) \quad (6.4.1b)$$

$$y_2'(x) = -\frac{12P}{E\tau^3}(x-l) \quad (6.4.1c)$$

$$y_1(0) = 0, \quad y_2(0) = 0, \quad y_1(l), \quad y_2(l) \text{ free} \quad (6.4.1d)$$

$$\int_0^l \tau(x) dx = C. \quad (6.4.1e)$$

All quantities in the above equations only depend on the variable  $x$ , which makes it a one-dimensional problem. This is characteristic for beam models. In detail, there is  $y_1$ , which stands for the transverse displacement component, and its derivative  $y_2$ , which can be understood as a rotational displacement. The coupling in (6.4.1b) is due to the assumption of a vanishing shear strain like in the Koiter model or in the model from Nestler discussed in Subsection 2.2.5. The value of  $P$  is the amount of force applied in transverse direction and  $E$  is Young's modulus. The initial conditions at  $x = 0$  correspond to the hardclamped case, since neither translational nor rotational displacements are permitted. The edge at  $x = l$  can move freely. The volume condition is the same as in the steady-state model except that the constant  $C$  has to be multiplied by 0.1 in contrast to the original formulation, since the thickness  $\tau$  is independent from  $\varphi$  and the integration in  $\varphi$ -direction has already been carried out.

With help of the methods of variations, an analytic solution to (6.4.1) can be calculated. All constraints are coupled to  $J$  which leads to the extended objective

$$J_* = \int_0^l \lambda(\tau(x) - C) + \psi_1(y_1' - y_2) + \psi_2 \left( y_2' + \frac{12P}{E\tau^3}(x-l) \right) dx + y_1(l).$$

The variation of  $J_*$  leads to necessary conditions for an optimal solution  $\tau^*$ . The adjoint equations are given as

$$\begin{aligned} \psi_1' &= 0, & \psi_1(l) &= -1 \\ \psi_2' &= -\psi_1, & \psi_2(l) &= 0. \end{aligned}$$

These are solved by

$$\psi_1(x) = -1, \quad \psi_2(x) = x - l.$$

The variation with respect to  $\tau$  gives

$$\lambda - \frac{36P}{E(\tau^*)^4}(x-l)^2 = 0 \implies \tau^* = \sqrt{6} \left( \frac{P}{\lambda E} \right)^{1/4} \sqrt{l-x}. \quad (6.4.2)$$

Inserting into the volume constraint (6.4.1e) leads to

$$\int_0^l \tau^*(x) dx = \int_0^l \sqrt{6} \left( \frac{P}{\lambda E} \right)^{1/4} \sqrt{l-x} dx = \frac{2\sqrt{6}}{3} l^{3/2} \left( \frac{P}{\lambda E} \right)^{1/4} \stackrel{!}{=} C,$$

so the expression for  $\lambda$  reads

$$\lambda = \frac{64P}{9E} l^6 C^{-4},$$

and inserting into (6.4.2) yields

$$\tau^*(x) = \frac{3}{2} C l^{-3/2} \sqrt{l-x}.$$

It is interesting to see that the optimal thickness does neither depend on the amount of force applied to the beam nor on Young's modulus (which characterizes the material). The corresponding displacement is given as the solution to the state equation (6.4.1b) and (6.4.1c) and reads

$$\begin{aligned} y_2^*(x) &= -\frac{64}{9} \frac{P}{EC^3} l^{9/2} \left( \sqrt{l-x} - \sqrt{l} \right) \\ y_1^*(x) &= \frac{128}{27} \frac{P}{EC^3} l^{9/2} \left( (l-x)^{3/2} - l^{3/2} - \frac{3}{2} \sqrt{l}x \right). \end{aligned}$$

In order to compare with the results of Example 6.4,  $l = 1$  m,  $P = 0.01$  MN,  $E = 210$  kN mm<sup>-2</sup> and  $C = 0.01$  m<sup>3</sup> are chosen. This gives the optimal thickness and displacements

$$\begin{aligned} \tau^*(x) &= \frac{3}{200} \sqrt{1-x} \\ y_2^*(x) &= -\frac{64}{189} \left( \sqrt{1-x} - \sqrt{1} \right) \\ y_1^*(x) &= \frac{128}{567} \left( (1-x)^{3/2} - 1 + \frac{3}{2}x \right), \end{aligned} \quad (6.4.3)$$

where  $[\tau^*] = \text{m}$  and  $[y_i^*] = \text{mm}$  according to the numerical quantities. When the displacement

$$y_1^*(1) = \frac{64}{567} \approx 0.1129$$

is compared to the value of  $u_3(1, \pi/2)$  denoted as 0.1126 in Figure 6.13 it can be observed that the first three significant digits coincide, which shows a great accuracy of the numerically calculated solution. This is underlined by the comparison of  $\tau^*$  from (6.4.3) and the numerical result in Figure 6.14. The error is below  $0.5 \cdot 10^{-3}$  in the region  $x \in (0.1, 0.9)$ . Outside this interval the error grows because the lower bound  $10^{-4}$  becomes active for the numerical solution while the exact solution tends to 0 at  $x = 1$ .

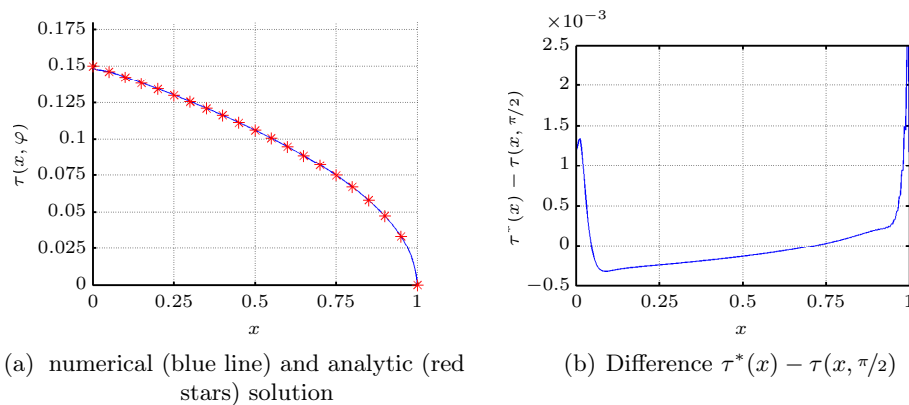


Figure 6.14.: Comparison of analytical solution of (6.4.1) and numerical solution of Example 6.4

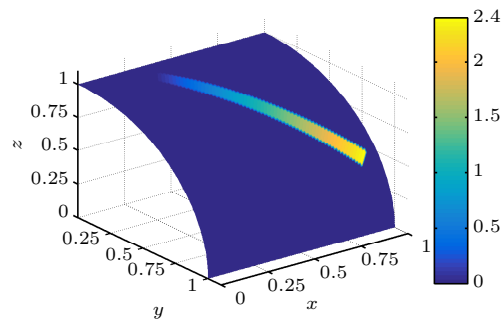


Figure 6.15.: Applied force for Example 6.5 as colored midsurface

## 6.5. Asymmetric force

To prove the success of the optimization not only for special cases like rotational symmetric forces or loaded beams, an asymmetric loading is chosen for the fourth example. It is given as

$$f(x, \varphi) = \begin{cases} 2 \left( x - \frac{2}{\pi} \varphi + 0.6 \right), & \text{if } x + \frac{2}{\pi} \in [1.2, 1.3] \text{ and } x - \frac{2}{\pi} \in [-0.6, 0.6] \\ 0, & \text{else.} \end{cases}$$

The load can be interpreted as some kind of bar with varying weight lying on the shell a bit off the diagonal. The load distribution is shown graphically in Figure 6.15.

The optimal thickness is shown in Figure 6.16. The left picture presents the thickness over the parameter domain on a grid with a step size of  $2^{-8}$  in  $x$ -direction and the right one contains the thickness distribution as a colored midsurface. In addition, the loaded region is enclosed with dashed lines. It can be seen that the

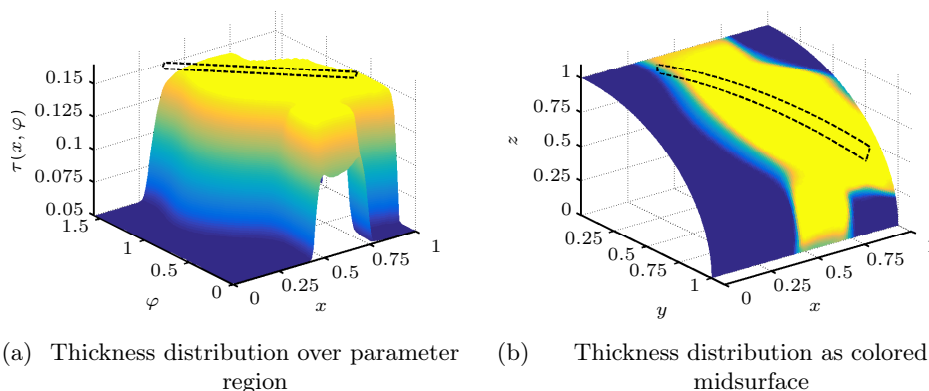


Figure 6.16.: Optimal thickness for Example 6.5, region of applied force (dashed-line)

thickness is distributed along the region of the applied force. The distribution is neither symmetric nor represented by any “regular” shape. There is also an interesting thick shaped domain near  $\varphi = 0$  which is a bit away from the actually loaded region. This could be caused by the strongest force applied at the bottom right end of the rectangular loaded area.

The optimal displacement and the comparison of the translational displacement norm is shown in Figure 6.17. On the upper left side, the displaced midsurface for the optimal thickness  $\tau_{\text{opt}}$  is shown. The upper right graphic contains the quantity  $u_{\text{diff}}(x, \varphi)$  from (6.2.1). It can be noticed that the optimal thickness results in an up to 50 % smaller displacement compared to the constant thickness at the region where the force is applied. Outside this region the graphic hardly shows differences between constant and optimized thickness. Those effects are better resolved in the bottom Figure 6.17(c): Here, the smaller displacement at the thickened region near the  $y = 1$ -boundary is visible as well as the bigger displacement near the  $x = 0$ -boundary. Certainly, both hardly influence the objective value. Altogether, the overall smaller displacement in particular around the loaded area is a satisfying result.

The rate of convergence between successively refined grids is shown in Table 6.3. It is not as satisfying as in Example 6.2 and 6.3 although some kind of convergence can be observed. The poorer rate could be explained by the additional thicker region off the loaded area that was mentioned above: On the coarser grids this region is not resolved as good as on the fine grid.

The solution of the linear program (5.3.11) leads to the optimality measure  $\varepsilon = 6.2 \cdot 10^{-7}$ , which suggests a good approximation of the solution from (5.3.4).

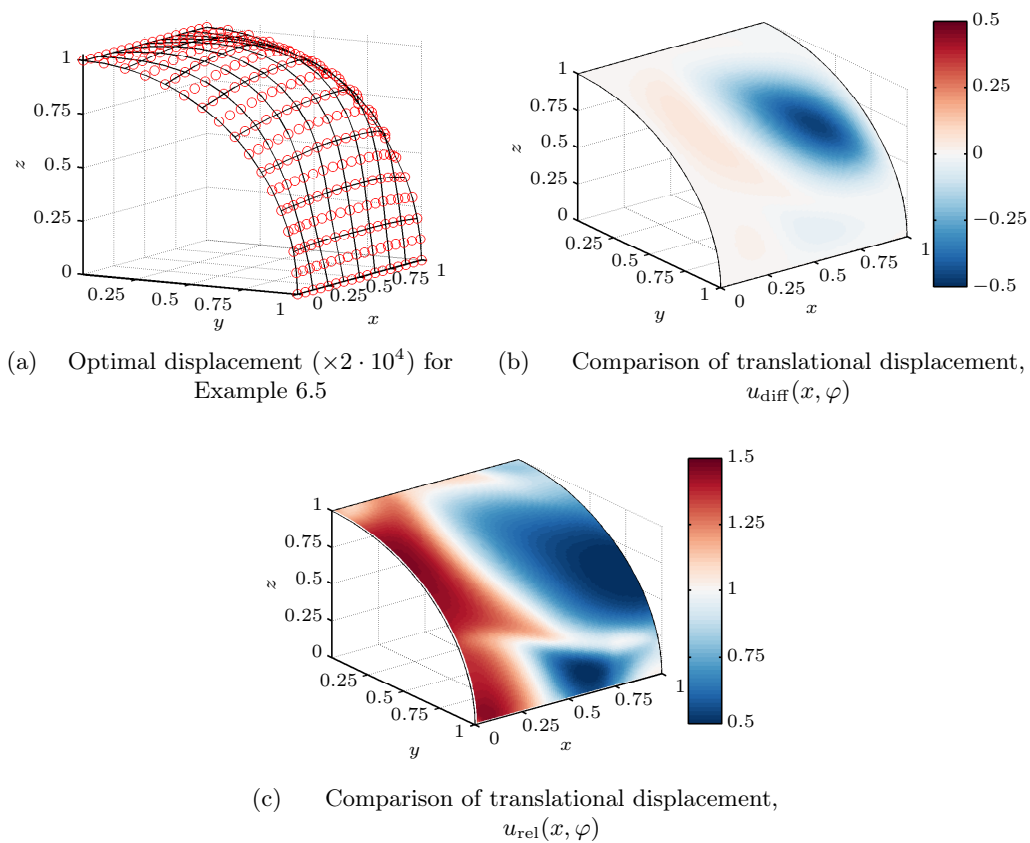


Figure 6.17.: Displacement for optimal thickness of Example 6.5

step size $h_x$	$\ \tau_h - \tau_{\text{opt}}\ _{\infty, h}$
$2^{-4}$	$4.53 \cdot 10^{-2}$
$2^{-5}$	$1.91 \cdot 10^{-2}$
$2^{-6}$	$1.23 \cdot 10^{-2}$
$2^{-7}$	$6.06 \cdot 10^{-3}$

Table 6.3.: Comparison of optimal thickness on different grids for Example 6.5

## 6.6. Summary

In this chapter numerical results for the steady-state optimization problem (3.2.1) were presented. The theoretical results were applied to shells with a cylindrical midsurface. Different boundary conditions were investigated, like softclamped over the whole boundary (Example 6.2 and 6.5) or partly hardclamped (Example 6.4). Moreover, different types of forces were applied, like rotational symmetric ones (Example 6.2) or discontinuous forces (Example 6.3). In addition, the model was tested for two limit cases: For a full tube where a numerical solution is available in [Nes12] and for a beam for which a comparable analytic solution is available (Example 6.4). For both cases the compliance with the available solutions is great.

In all examples a comparison between the displacement for the optimal thickness and the displacement for the initial (constant) thickness was presented. It was shown that the optimized displacement was significantly smaller than the initial one, especially in the region where the strongest force was applied. Additionally, a comparison between the optimal thickness on successively refined grids was investigated: In every example some sort of convergence could be observed. Moreover, the “optimality test” from (5.3.11) indicated a numerical solution which is near to an optimal solution of (5.3.4).

The overall numerical results show that the optimization workflow (including the use of the finite element gradient) and the implementation lead to satisfying optimization results.

## 7. Results for the dynamic problem

In this chapter the numerical results for the dynamic optimization problem (4.2.1) are presented. As in the steady-state case, different setups are chosen involving a shell with a cylindrical midsurface. Three different examples are presented, including not only the calculated optimal thickness but also convergence studies and the investigation of the “optimality measure” discussed in (5.3.11). In addition, the optimized displacement is compared to the initial displacement involving a constant thickness.

The geometry is the same as discussed in Section 6.1. The same is true for the general problem data, like upper and lower bounds for the thickness, the prescribed volume and the regularization parameter; the units are kept as well. The density is chosen as  $\rho = 7.8 \text{ g cm}^{-2}$ , still corresponding to steel. The time is measured in units of ms. The temporal step-size is chosen as  $5 \cdot 10^{-3}$  for all examples. The spatial grid is once less refined than in the steady-state case to keep a reasonable time and memory consumption for the optimization.

### 7.1. Force with time-dependent intensity

The first example involves a spatially stationary load, while its intensity changes over time. It is given as

$$f(x, \varphi; t) = \begin{cases} 2t \sin(2\pi x) \sin(2\varphi), & \text{if } x \leq 0.5 \\ -t \sin(2\pi x) \sin(2\varphi), & \text{if } x > 0.5. \end{cases}$$

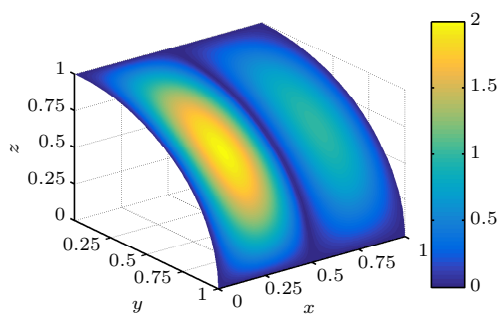


Figure 7.1.: Applied (maximal) force for Example 7.1 as colored midsurface

There is no initial momentum given, i.e.  $g = 0$ . The distribution of the maximal load (w.r.t. to  $t$ ) on the midsurface is shown in Figure 7.1.

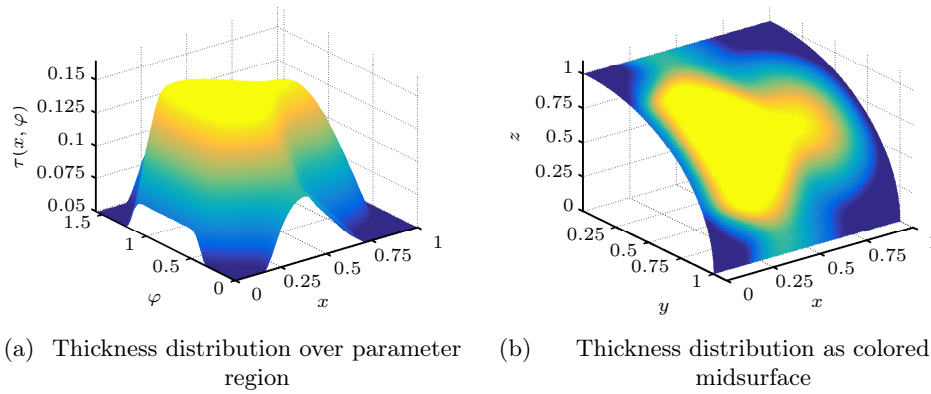


Figure 7.2.: Optimal thickness for Example 7.1

The optimal thickness is shown in Figure 7.2. The thickness distribution follows approximately the applied load. Clearly, the shell is thicker in the region  $x \leq 0.5$ , since the double amount of force is applied there. However, the hardly loaded area around  $x = 0.5$  is not reflected by the optimal thickness. The symmetry about the  $\varphi = \pi/4$ -axis is clearly visible.

The comparison of the optimal displacement and the initial one involving a constant thickness is shown in Figure 7.4. The time-integrated quantities  $u_{\text{diff}}$  derived from (6.2.1),

$$u_{\text{diff}}(x, \varphi) = \int_0^T \frac{\|u_{h,\text{opt}}(x, \varphi; t)\|_2 - \|u_{h,\text{const}}(x, \varphi; t)\|_2}{\sup \|u_{h,\text{const}}(x, \varphi; t)\|_2} dt \quad (7.1.1)$$

and  $u_{\text{rel}}$  derived from (6.2.2),

$$u_{\text{rel}}(x, \varphi) = \int_0^T \frac{\|u_{h,\text{opt}}(x, \varphi; t)\|_2}{\|u_{h,\text{const}}(x, \varphi; t)\|_2} dt \quad (7.1.2)$$

are presented there. It can be observed that the optimization leads to a decrease in displacement of over 30 % in the region  $x \leq 0.5$  where the strongest force is applied, and also a noticeable decrease in the area of  $x > 0.5$ . As in the steady-state case, there are also regions where the displacement increases – this can be seen best in the right subfigure of Figure 7.4 involving  $u_{\text{rel}}$ . Since the actual displacements (and hence the absolute differences) are small in those regions, the overall gain from thickness optimization is clearly visible.

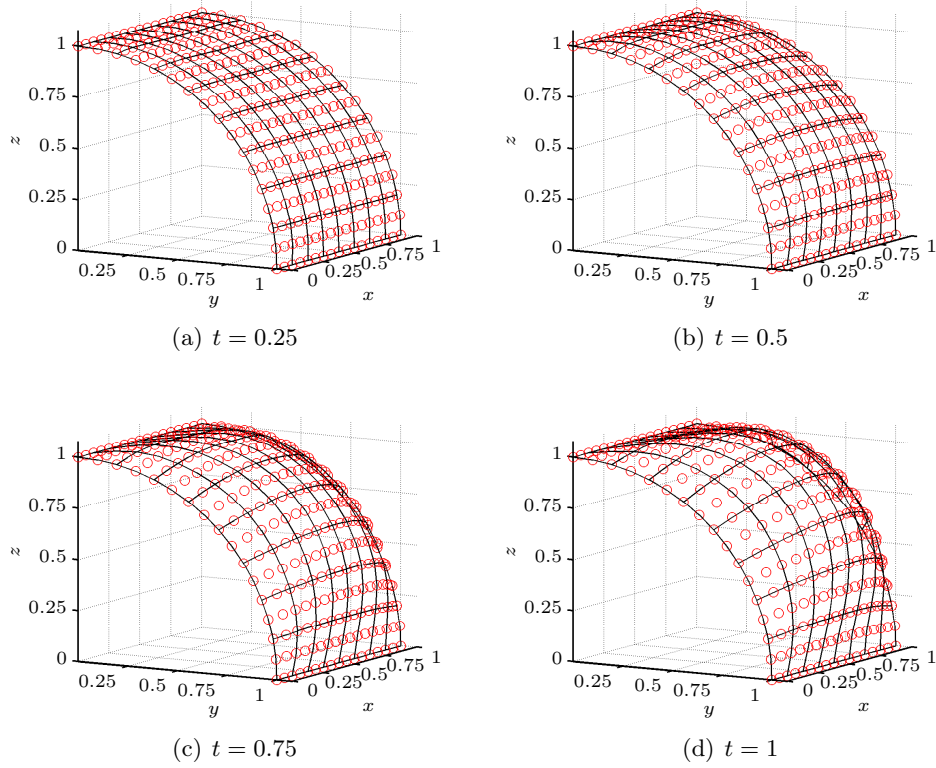


Figure 7.3.: Optimal displacement ( $\times 5 \cdot 10^3$ ) for Example 7.1

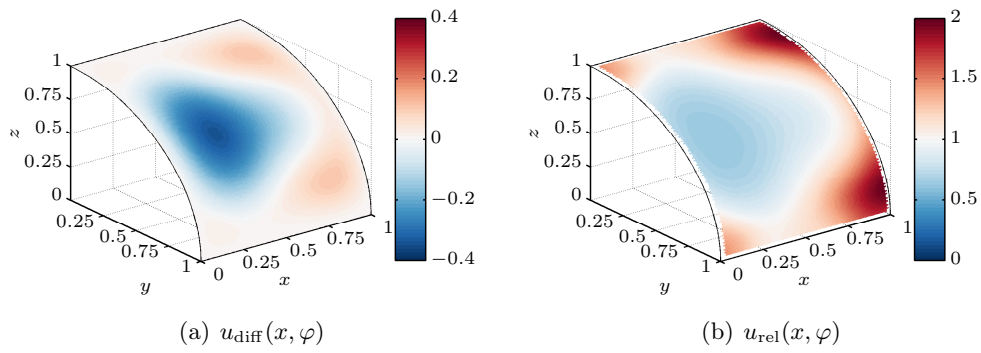
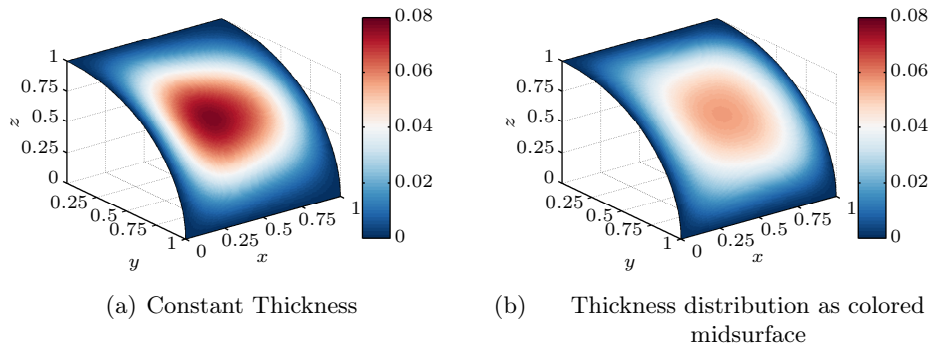


Figure 7.4.: Comparison of translational displacement for Example 7.1

Figure 7.5.: Maximal displacement (w.r.t. time)  $\times 10^3$  for Example 7.1

step size $h_x$	$\ \tau_h - \tau_{\text{opt}}\ _{\infty, h}$
$2^{-4}$	$3.76 \cdot 10^{-2}$
$2^{-5}$	$8.26 \cdot 10^{-3}$
$2^{-6}$	$1.62 \cdot 10^{-3}$

Table 7.1.: Comparison of optimal thickness on different grids for Example 7.1

It is also interesting to observe how the place of maximal displacement (w.r.t. to time) changes between constant and optimal thickness. In Figure 7.5 it can be seen that the constant thickness leads to the maximal displacement at the place of the maximal force. This is no more true for the optimal thickness: Here, the maximal displacement moves to the middle of the shell and also takes a symmetric shape. Hence, the thickness optimization provides a more even distribution of the displacement. Moreover, the maximal displacement for the optimized thickness is clearly smaller than for the constant one.

The difference between the solution on successively refined grids is shown in Table 7.1. The convergence towards the “exact” solution is noticeable and emphasizes the satisfying result for this example. Finally, the evaluation of the optimality measure (5.3.11) gives  $\varepsilon = 1.78 \cdot 10^{-5}$ .

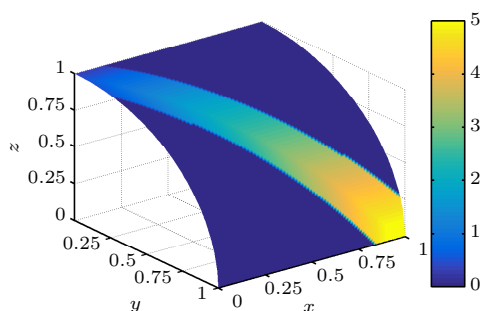


Figure 7.6.: Applied force for Example 7.2 as colored midsurface

## 7.2. Moving load

The second example consists of a moving load with time-varying intensity. It is given by

$$f(x, \varphi; t) = \begin{cases} 5(1-t), & \text{if } t \leq 1, x \in [0.9-t, 1.1-t], \varphi \in \left[\frac{\pi}{2}t - 0.1, \frac{\pi}{2}t + 0.1\right] \\ 0, & \text{else.} \end{cases}$$

The load moves along the diagonal of the shell with decreasing intensity. The path and the intensity is visualized as a colored midsurface in Figure 7.6. In this graphic, the maximal load (w.r.t. to  $t \in [0, T]$ ) at each point on the midsurface is presented. The end time is chosen as  $T = 1$ . There is no initial momentum given, i.e.  $g = 0$ .

The objective is taken as the  $L^2$ -norm of the translational displacement rather than the compliance functional, i.e.

$$J_{\|\cdot\|}(u, \theta; \tau) = \|u\|_{L^2(0,T;L^2(S)^2 \times L^2(S))}^2 + \frac{\lambda}{2} \|\tau\|_{H^1(S)}^2,$$

so the displacement is considered without further weighting. This objective is better suited than the compliance functional if an end time  $T > 1$  is considered (see the discussion in the following paragraph), because otherwise the displacement at time points  $t \geq 1$  would not contribute to the optimization at all.

The optimal thickness is shown in Figure 7.7. This is the result on a grid consisting of about 16,000 nodes. It can be observed that the shell is thickest along the path of the applied force. However, there are additional spikes near the  $\varphi = \pi/2$ -boundary. This is an effect of the specifically chosen end time  $T = 1$ . As can be seen in the displacement comparison in Figure 7.9, at  $t = 1$  the displacement norm takes maxima near those specific areas where the additional thickness is placed. A different choice of the end time would lead to a different or additional placement of those spikes, so they should not be over-pronounced in the interpretation of the result. This is visible in Figure 7.8, where the optimal thickness for  $T = 2$  (left picture) and  $T = 5$  (right picture) are shown as a colored midsurface. Note that also for  $T > 1$  the continuity of  $f$  with respect to the time is met, which is required in Theorem 4.4.

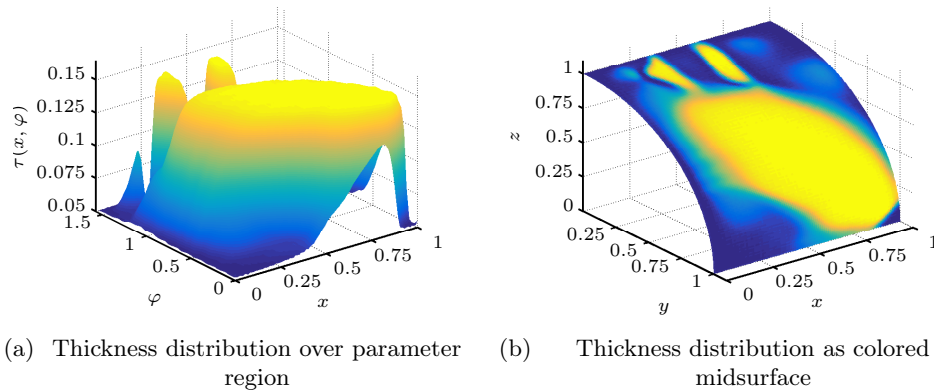
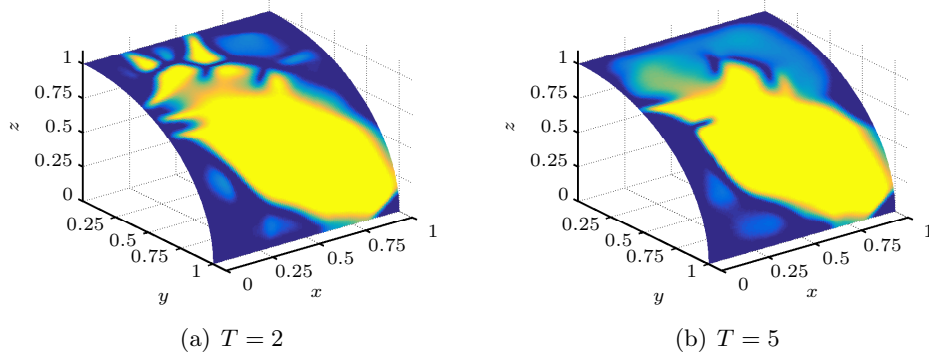


Figure 7.7.: Optimal thickness for Example 7.2

Figure 7.8.: Optimal thickness for different values of  $T$  for Example 7.2

Another interesting observation is the missing symmetry of the optimal thickness with respect to the diagonal: Although the spatial *path* of the load is symmetric, its intensity changes over time – so the displacement (and therefore the optimal thickness) cannot be expected to be symmetric either.

The comparison of the optimal displacement and the initial one, involving a constant thickness, is shown in Figure 7.10. The quantities  $u_{\text{diff}}$  from (7.1.1) and  $u_{\text{rel}}$  from (7.1.2) are presented there. It can be observed that the optimization leads to a decrease in displacement of up to 50 % in the region where the strongest force is applied. As in the steady-state case, there are also regions where the displacement increases – this can best be seen in the right figure involving  $u_{\text{rel}}$ . Since the actual displacements (and hence the absolute differences) are small in those regions, the overall gain from thickness optimization is clearly visible.

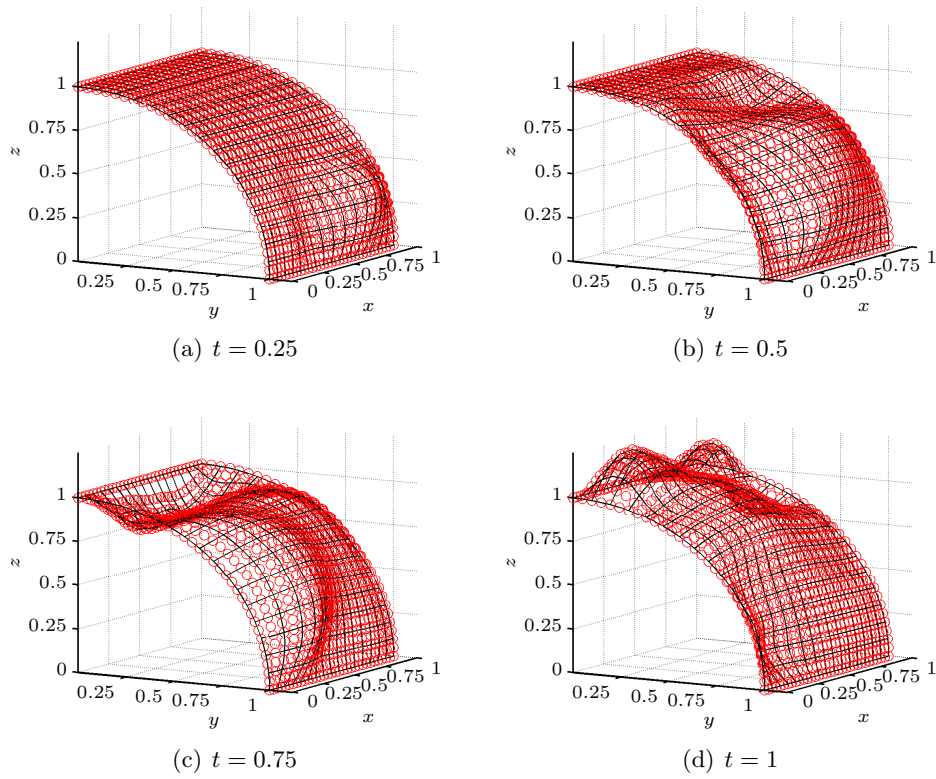
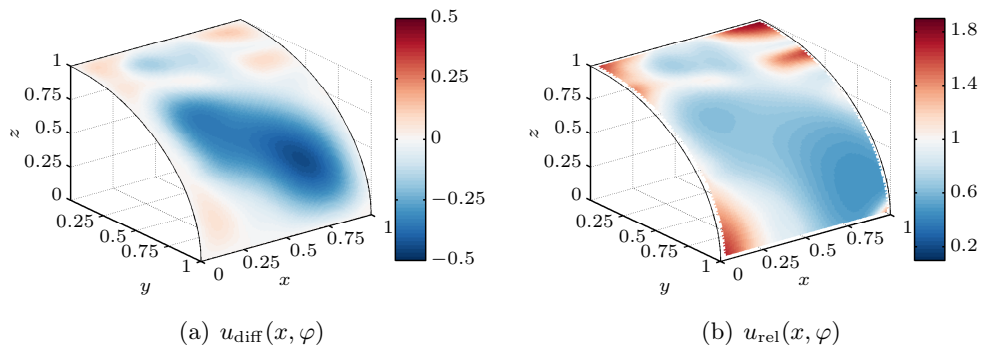
Figure 7.9.: Optimal displacement ( $\times 10^4$ ) for Example 7.2

Figure 7.10.: Comparison of translational displacement for Example 7.2

(a) $\infty$ -norm		(b) $L^2$ -norm	
step size $h_x$	$\ \tau_h - \tau_{\text{opt}}\ _{\infty, h}$	step size $h_x$	$\ \tau_h - \tau_{\text{opt}}\ _{L^2, h}$
$2^{-4}$	$7.66 \cdot 10^{-2}$	$2^{-4}$	$2.00 \cdot 10^{-2}$
$2^{-5}$	$5.77 \cdot 10^{-2}$	$2^{-5}$	$1.10 \cdot 10^{-2}$
$2^{-6}$	$6.05 \cdot 10^{-2}$	$2^{-6}$	$9.89 \cdot 10^{-3}$

Table 7.2.: Comparison of optimal thickness on different grids for Example 7.2

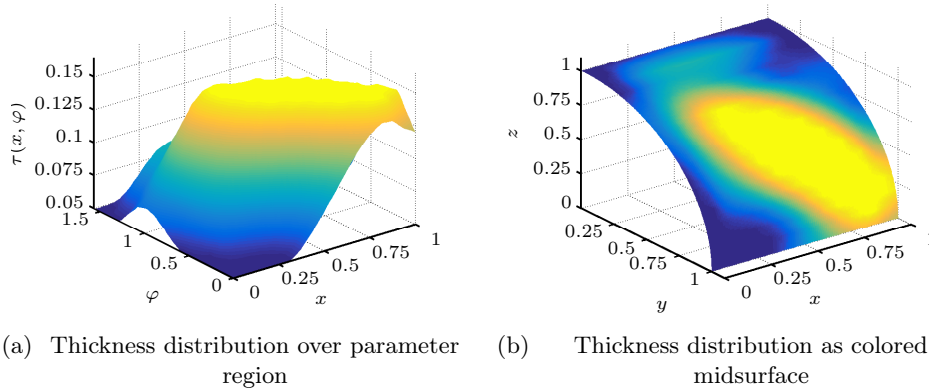


Figure 7.11.: Optimal thickness on a coarse grid for Example 7.2

It is also interesting to investigate the convergence properties of the solution on refined grids. In Table 7.2 two subtables are shown: The left one lists the difference between the “exact” solution (on the finest grid) and the solutions on coarser grids taken in the  $\infty$ -norm. It is obvious that the error does not decrease monotonically as in other examples. This is caused by the spikes placed off the loaded region. Those are not resolved at all on coarser grids, as can be seen in Figure 7.11, where the solution on a coarse grid consisting of only 289 nodes is shown. So for investigating the convergence properties, the  $L^2$ -norm is better suited: The “exact” solution is restricted to its nodal values on the coarse grid, which are then interpolated by the ansatz functions on that grid. This “restricted exact” solution is compared to the solution on the coarse grid with help of the  $L^2$ -norm. The results are shown in the right listing of Table 7.2. Here, the error decreases monotonically. The evaluation of the optimality measure (5.3.11) gives  $\varepsilon = 1.25 \cdot 10^{-6}$ .

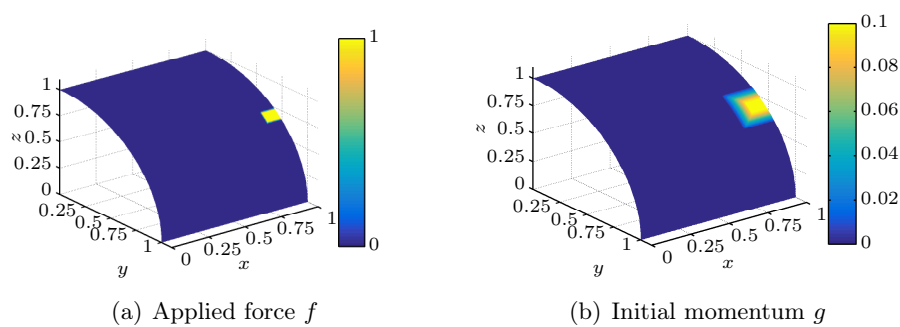


Figure 7.12.: Magnitude of force and momentum distribution as colored midsurface for Example 7.3

### 7.3. Impact on partially clamped shell

The third example of this chapter considers a partially hardclamped cylindrical shell. Given the usual parameter domain  $\omega = (0, 1) \times (0, \pi/2)$ , the shell is hardclamped at the edge  $x = 0$ , i.e. neither translational nor rotational displacements are allowed there. The other three edges are free. Note that these boundary conditions forbid rigid body motions, as required for the existence and uniqueness of the state equation solution (see discussion in Subsection 3.1.1). The setup is a body falling on the shell at the opposite of the clamped edge, giving an initial momentum to the shell body and staying at the place of the impact.

The load is given as

$$f(x, \varphi) = \begin{cases} -1, & \text{if } x \geq 0.9, \frac{\pi}{4} - 0.05 \leq \varphi \leq \frac{\pi}{4} + 0.05 \\ 0, & \text{else.} \end{cases}$$

The weak solution theory in particular allows for discontinuous forces and  $f \in L^2(\mathcal{S})$  meets the requirements of the differentiability Theorem 4.4. Certainly, the momentum cannot be chosen like  $f$ , because the corresponding operator  $\Phi_g = g/\tau$  induced by the function  $g$  must be  $L^\infty(\mathcal{S}) \rightarrow \mathcal{V}$ . In particular, there is no jump discontinuity allowed as in  $f$ . So the momentum is smoothened with help of linear interpolation,

$$g(x, \varphi) = \begin{cases} -0.1, & \text{if } (x, \varphi) \in (0.9, 1) \times (\frac{\pi}{4} - 0.05, \frac{\pi}{4} + 0.05) \\ 0, & \text{if } (x, \varphi) \notin (0.8, 1) \times (\frac{\pi}{4} - 0.15, \frac{\pi}{4} + 0.15) \\ -5 (|x + \varphi - (0.95 + \frac{\pi}{4})| + |x - \varphi - (0.95 - \frac{\pi}{4})|) + 1.5, & \text{else.} \end{cases}$$

The force distribution  $f$  and the momentum distribution  $g$  are shown in Figure 7.12.

The optimal thickness is shown in Figure 7.13. This is the result on a grid consisting of about 16,000 nodes. An interesting structure can be observed here: The region of impact (and the loaded region) is surrounded by a thick area. Evolving from this area, there are two supporting struts which lead along the  $\varphi = 0$  and

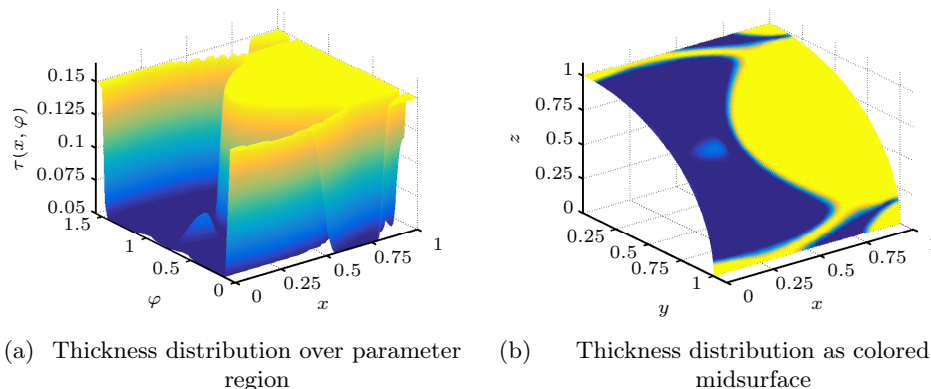


Figure 7.13.: Optimal thickness for Example 7.3

$\varphi = \pi/2$ -boundary to the clamped edge. In addition, the corners at the  $x = 1$ -boundary are thickened and there is a little thicker area around  $\varphi = \pi/4$  near the  $x = 0$ -edge. The symmetry with respect to the  $\varphi = \pi/4$  axis is clearly visible. This structure is an intriguing starting point for topology optimization: The thin areas would be erased completely from the shell, leading to a cut midsurface. For this purpose, the techniques from shape optimization discussed in Section 3.4 would come in place.

The comparison of the optimal displacement and the initial one involving a constant thickness is shown in Figure 7.15. Note the inverted  $x$ -axis in the plots of the optimal displacement for a better viewing angle. The wave caused by the initial momentum can well be tracked through the shell.

The quantities  $u_{\text{diff}}$  from (7.1.1) and  $u_{\text{rel}}$  from (7.1.2) are presented in Figure 7.15. The evaluation of  $u_{\text{diff}}$  shows an around 30 % lowered displacement near the impact region. In addition, the quantity  $u_{\text{rel}}$  indicates an evenly distributed overall decrease in displacement. Only near the hardclamped edge, where the thickness is chosen smaller, an increase can be observed. However, the displacement near the clamped edge is small in general. Altogether, the overall gain from thickness optimization is clearly visible.

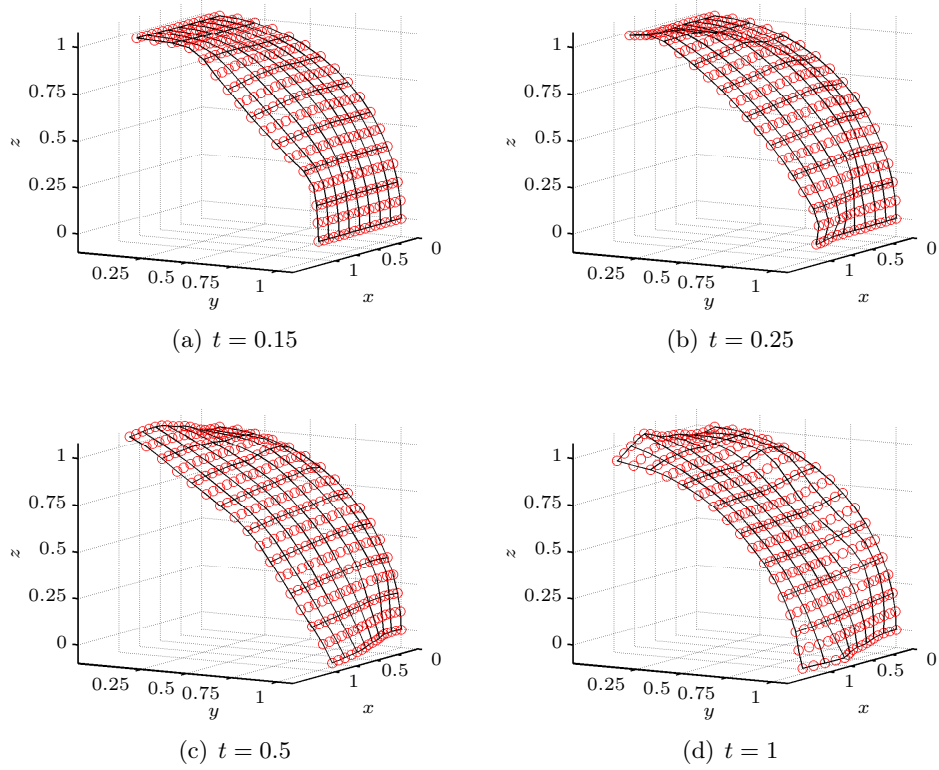


Figure 7.14.: Optimal displacement ( $\times 3 \cdot 10^3$ ) for Example 7.3

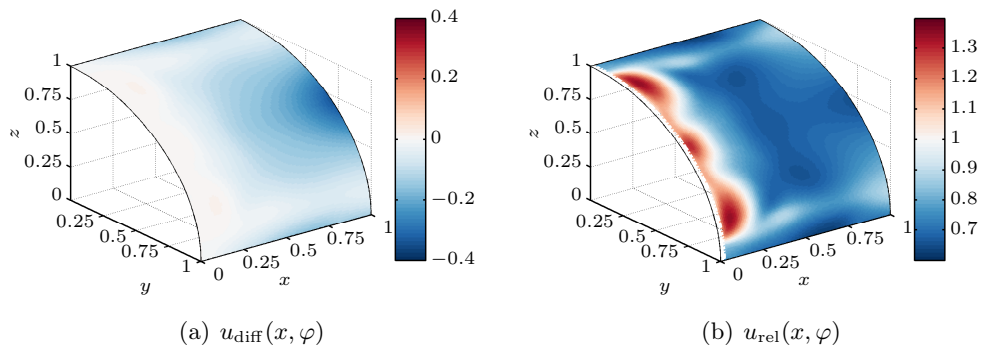


Figure 7.15.: Comparison of translational displacement for Example 7.3

(a) $\infty$ -norm		(b) $L^2$ -norm	
step size $h_x$	$\ \tau_h - \tau_{\text{opt}}\ _{\infty, h}$	step size $h_x$	$\ \tau_h - \tau_{\text{opt}}\ _{L^2, h}$
$2^{-4}$	$6.74 \cdot 10^{-2}$	$2^{-4}$	$2.76 \cdot 10^{-2}$
$2^{-5}$	$1.00 \cdot 10^{-1}$	$2^{-5}$	$1.92 \cdot 10^{-2}$
$2^{-6}$	$6.63 \cdot 10^{-2}$	$2^{-6}$	$6.78 \cdot 10^{-3}$

Table 7.3.: Comparison of optimal thickness on different grids for Example 7.3

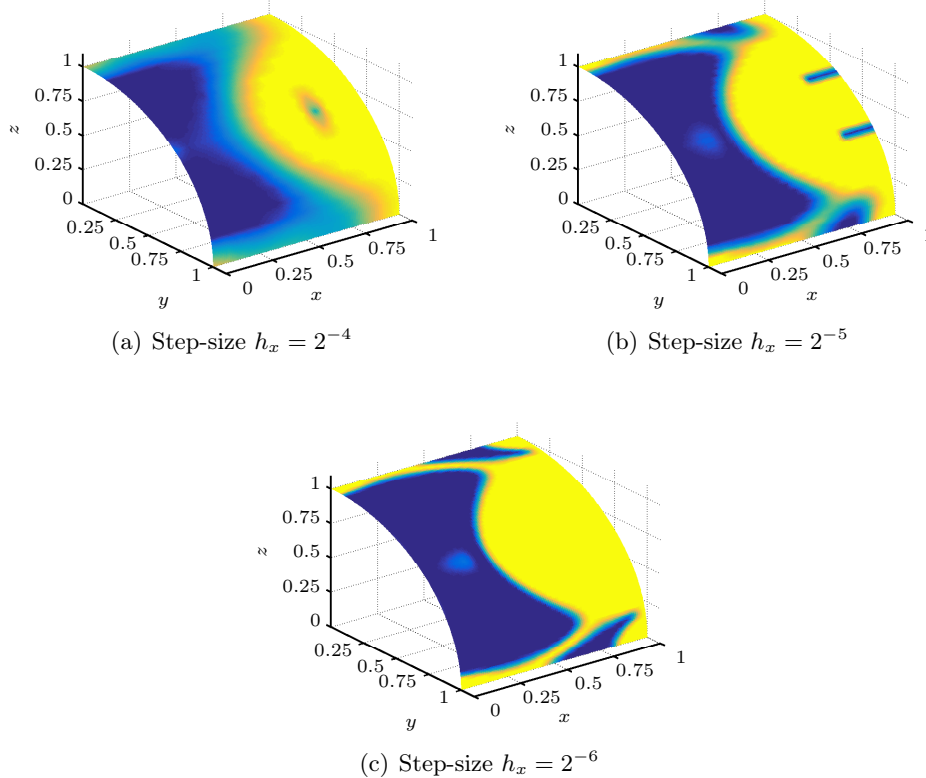


Figure 7.16.: Optimal thickness on coarse grids for Example 7.3

The convergence properties of the solution on refined grids is shown in Table 7.3, which consists of two subtables: The left one lists the difference between the “exact” solution (on the finest grid) and the solutions on coarser grids taken in the  $\infty$ -norm. It is obvious that the error does not decrease monotonically. This is caused by two possible reasons: The actual shape of the thick area is not resolved on the coarser grids, and since there is no smooth transition between the thick and thin area, a large error in the maximum norm can be expected if the thick and thin areas do not match. This is the case for the solution on the  $h_x = 2^{-6}$  step size grid: The thick corners at the  $x = 1$ -edge are still connected to the main thick area here, while this is

not true for the “exact” solution. The second explanation is an additional gap around the region of impact, which is only present on coarse grids (see Figure 7.16). Clearly, this cut causes errors in the maximum norm. For investigating the convergence properties, the  $L^2$ -norm is better suited, as explained in Example 7.2. The results are shown in the right listing of Table 7.3. Here, the error decreases monotonically. The evaluation of the optimality measure (5.3.11) gives  $\varepsilon = 2.55 \cdot 10^{-7}$ .

## 7.4. Summary

In this chapter numerical results for the dynamic optimization problem (4.2.1) were presented. The theoretical results were applied to shells with a cylindrical midsurface. Different boundary conditions were investigated, like softclamped over the whole boundary (Example 7.1 and 7.2) or partly hardclamped (Example 7.3). Moreover, different types of forces were applied: A spatially fixed force with time-varying intensity (Example 7.1), a discontinuous moving load (Example 7.2) and a force combined with an initial momentum (Example 7.3).

In all examples a comparison between the displacement for the optimal thickness and the displacement for the initial (constant) thickness was presented. It was shown that the optimized displacement was significantly smaller than the initial one, especially in the region where the strongest force was applied. This coincides with the results from the steady-state case. In addition, a more equal distribution of the (time-) maximal displacement was observed (Example 7.2).

Moreover, a comparison between the optimal thickness on successively refined grids was investigated: In every example some sort of convergence could be observed. It turned out that the  $\infty$ -norm is not always suited for this study, because some fine structures were not present on coarser grids. The  $L^2$ -norm is a better candidate, and indeed it decreased monotonically on refined grids. In addition, the “optimality test” from (5.3.11) indicated a numerical solution which is near to an optimal solution of (5.3.4).

Some differences evolved in comparison to the steady-state case: The optimal thickness depends on the end-time up to a certain degree (Example 7.2) and the comparison of the displacements at specific time-points was not suitable, so rather the time-average of the displacement differences was studied.

Altogether, the overall numerical results show that the optimization workflow (including the use of the finite element gradient) and the implementation lead to satisfying optimization results.



## 8. Summary and outlook

The last chapter gives a short review of the results and points out further questions that arise from the previous chapters.

In chapter two an overview of the equations of linear elasticity and different shell models was given. The decision for this thesis was made in favor of using the basic shell model from the book of Chapelle and Bathe ([CB03]). The corresponding equations were presented and specialized ones for the cylindrical test cases were calculated. In addition, the required smoothness for the thickness was discussed.

Clearly, optimization results involving different shell models are of further interest: A generalization to nonlinear models is possible and even full three-dimensional models serve as potential candidates. The computational effort would grow for such models, but processing power is widely available and efficient implementations would lead to reasonable calculation times. Comparisons to the results of different (simpler) models were presented in Section 6.2 and 6.4 and hardly showed discrepancies. This could be different when switching to three-dimensional equations, especially with thicker objects. Furthermore, the linearity assumptions on the strains is only valid for small deformations. For bigger loads or initial momenta, the use of nonlinear models becomes inevitable.

As further research, the optimization problem itself can be formulated for different objectives. As mentioned in Section 2.3, an interesting variation would be to look for a shell with as little material as possible (the most lightweight one), that does not deform more than a given value under a certain load. This introduces state constraints into the optimization problem which makes it more complex. Certainly, various other formulations are available from real-world applications. In addition, the use of robust optimization techniques brings an interesting approach: In Example 7.3 the area or the strength of the impact could be made variable, so that the optimization under this uncertainty yields a different thickness distribution.

Chapter three dealt with the steady-state optimization problem. Here, the investigation of the displacement sensitivity with respect to the thickness was of main interest. With the help of this result, also the sensitivity of the reduced objective functional and necessary conditions for an optimal solution could be given. From a mathematical perspective it would be interesting to view the role of the thickness  $\tau$  as integral boundaries actually as a mapping  $\tau \mapsto I_{(\tau)}$ , where  $I_{(\tau)}$  is an integral operator defining the innermost integral of the bilinear form  $a_{(\tau)}$ . This could simplify or shorten some of the proofs. Moreover, the Fréchet-differentiability of the control-to-state operator is worth looking at: The continuity of the Gâteaux-differential with respect to  $\tau$  is the missing piece here.

The existence of a solution to the optimization problem was answered for thickness

distributions lying in a certain restricted space  $U_{\text{ad}}^M$  (see Theorem 3.5). The existence of a solution in more general spaces like  $W^{1,\infty}(\mathcal{S})$  or even  $L^\infty(\mathcal{S})$  remains open.

The embedding of thickness optimization into the more general shape optimization theory was discussed in the second part of chapter three. Dropping the requirement of a fixed midsurface would lead to a whole new consideration of the problem and head away from the actual task of thickness optimization. However, the coincidence of shape derivative and Gâteaux-derivative with respect to the thickness was an enlightening result of Section 3.4.

The fourth chapter focused on the dynamic problem. Here, some effort was put onto the preparation of the sensitivity Theorem 4.4. As for the steady-state case, the objective sensitivity and necessary conditions for an optimal solution were provided. A further exploration of available results for hyperbolic partial differential equations could simplify some of the proofs. Moreover, the tighter regularity assumptions on the loading and initial momentum lead to restrictions in real-world applications. A reformulation of the theorems with the same assumptions as for the existence of a solution to the state equation (see Theorem 4.1) would be a huge profit for the theoretical results.

Looking at the optimization problem itself, the same variants as for the steady-state case are present. Moreover, the evaluation of the displacement at specific time points (e.g. the end time) creates more possibilities for objective functionals. This would introduce an observation operator into the objective which leads to additional considerations.

Methods for the numerical solution of the optimization problem were presented in chapter five. For the solution of problems from linear elasticity, a wide range of different finite elements is available. The rectangle elements used in this thesis rank among the simpler ones. The MITC elements mentioned in Section 5.4 are a suitable candidate to try for further investigation and to avoid the locking phenomenon discussed in Section 5.4. In addition, adaptive methods would serve as an interesting continuation.

Other available packages for the solution of linear equation systems or for nonlinear optimization are worth trying, and even a self-written solver tailored to thickness optimization problems is a possibility for further research. In addition, the availability of necessary conditions for optimal solutions enables the application of indirect optimization methods. Those methods try to solve the variational inequality (3.2.6) or (4.2.7) instead of the actual optimization problem and have gained much attention, in particular in optimal control of ordinary differential equations.

The results in chapter six and seven proof the functionality of the proposed models and the optimization workflow. While the theoretical results apply to general midsurfaces (except the overall restrictions for shell models (2.2.5) and the necessary conditions), the numerical calculations focus on cylindrical ones. The presented results look reasonable and the comparison of the optimized displacement to the non-optimized one shows the gain from thickness optimization. In addition, the convergence studies on successively refined grids propose a good quality of the results. This is as well underlined by the evaluation of the necessary conditions (3.2.6) and

(4.2.7), which are fulfilled up to a small error in every example.

In particular, the dynamic case gives rise for additional considerations: The influence of the end time is clearly visible in Example 7.2, so it should be discussed which time horizon is of importance. The result in Example 7.3 gives rise for studies in topology optimization where the thin parts of the shell would be cut out of the midsurface. This leads to a quite different geometry, in particular to a new midsurface and is settled nearer to the field of shape optimization. To stay in the context of cylindrical shells, the influence of the radius of the shell as well as its extension in angular direction is worth looking at. Moreover, the presence of cracks or parts with different material properties would lead to interesting investigations.

Regarding the focus on cylindrical shells, a more generalized implementation would be straightforward, however, general shell elements should be used then. This seems to be the most promising practical continuation of the work, since it enables a broad range of new problems to be solved. The midsurface would not have to be defined by a given parametrization, but rather by given nodes in Cartesian coordinates, e.g. from CAD systems. Moreover, the introduction of plastic behavior as well as the investigation of multi-layered materials following real-world problems would be a promising field for further research.



# Bibliography

- [AF03] R. A. Adams and J. J. Fournier. *Sobolev Spaces*. 2nd ed. Vol. 140. Pure and Applied Mathematics. Elsevier Science, 2003.
- [AFB12] S. Arnout, M. Firl, and K.-U. Bletzinger. “Parameter free shape and thickness optimisation considering stress response”. In: *Structural and Multidisciplinary Optimization* 45.6 (2012), pp. 801–814.
- [Alt12] H. Altenbach. *Kontinuumsmechanik: Einführung in die materialunabhängigen und materialabhängigen Gleichungen*. Springer, 2012.
- [AS04] V. Azhmyakov and W. Schmidt. “On the optimal design of elastic beams”. In: *Structural and Multidisciplinary Optimization* 27.1 (2004), pp. 80–88.
- [Bar+94] R. Barrett et al. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. Society for Industrial and Applied Mathematics, 1994.
- [Ben95] M. P. Bendsøe. *Optimization of Structural Topologie, Shape, and Material*. Springer, 1995.
- [Ber77] M. S. Berger. *Nonlinearity & Functional Analysis: Lectures on Non-linear Problems in Mathematical Analysis*. Vol. 74. Pure and Applied Mathematics. Academic Press, 1977.
- [BIC00] K. J. Bathe, A. Iosilevich, and D. Chapelle. “An evaluation of the MITC shell elements”. In: *Computers & Structures* 75.1 (2000), pp. 1–30.
- [BL01] A. Blouza and H. Le Dret. “Naghdi’s Shell Model: Existence, Uniqueness and Continuous Dependence on the Midsurface”. In: *J. Elasticity* 64 (2001), pp. 199–216.
- [BL99] A. Blouza and H. Le Dret. “Existence and uniqueness for the linear Koiter model for shells with little regularity”. In: *Quart. Appl. Math.* 57 (1999), pp. 317–337.
- [Ble09] J. R. Blevins. “A Generic Linked List Implementation in Fortran 95”. In: *ACM Fortran Forum* 28.3 (2009), pp. 2–7.
- [CB03] D. Chapelle and K.-J. Bathe. *The finite element analysis of shells - fundamentals*. 2nd ed. Engineering online library. Springer, 2003.
- [CBB04] N. Camprubí, M. Bischoff, and K.-U. Bletzinger. “Shape optimization of shells and locking”. In: *Computers & Structures* 82.29–30 (2004), pp. 2551–2561.

- [Cia00] P. G. Ciarlet. *Theory of Shells*. Mathematical Elasticity 3. Elsevier Science, 2000.
- [CJP08] B. Chapman, G. Jost, and R. van der Pas. *Using OpenMP: Portable Shared Memory Parallel Programming*. Scientific Computation Series 10. MIT Press, 2008.
- [DZ11] M. C. Delfour and J. P. Zolésio. *Shapes and geometries. Analysis, differential calculus, and optimization*. 2nd ed. Advances in design and control. Society for Industrial and Applied Mathematics, 2011.
- [EG15] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions*. Revised Edition. Textbooks in Mathematics. CRC Press, 2015.
- [Eva10] L. C. Evans. *Partial Differential Equations*. 2nd ed. Vol. 19. Graduate Studies in Mathematics. American Mathematical Society, 2010.
- [FGW02] A. Forsgren, P. E. Gill, and M. H. Wright. “Interior Methods for Nonlinear Optimization”. In: *SIAM Rev.* 44.4 (2002), pp. 525–597.
- [Ger04] J. M. Gere. *Mechanics of Materials*. 6th ed. Thomson Brooks/Cole, 2004.
- [Gou94] P. L. Gould. *Introduction to Linear Elasticity*. 2nd ed. Springer, 1994, p. 237.
- [GSH07] N. I. M. Gould, J. A. Scott, and Y. Hu. “A Numerical Evaluation of Sparse Direct Solvers for the Solution of Large Sparse Symmetric Linear Systems of Equations”. In: *ACM Trans. Math. Softw.* 33.2 (2007).
- [Heb96] E. Hebey. *Sobolev Spaces on Riemannian Manifolds*. Lecture Notes in Artificial Intelligence 1635. Springer, 1996.
- [Hei05] J. Heinonen. *Lectures on Lipschitz Analysis*. Lecture Notes, University of Jyväskylä. 2005.
- [Hel01] P. Helnwein. “Some remarks on the compressed matrix representation of symmetric second-order and fourth-order tensors”. In: *Computer Methods in Applied Mechanics and Engineering* 190.22-23 (2001), pp. 2753–2770.
- [HM03] J. Haslinger and R. A. E. Mäkinen. *Introduction to Shape Optimization: Theory, Approximation, and Computation*. Advances in Design and Control. Society for Industrial and Applied Mathematics, 2003.
- [Hug00] T. J. R. Hughes. *The finite element method: linear static and dynamic finite element analysis*. Mineola: Dover, 2000.
- [Hun14] J. K. Hunter. *Notes on Partial Differential Equations*. Lecture Notes, Department of Mathematics, University of California at Davis. 2014.
- [Ipo15] Ipopt. *Ipopt home page*. Nov. 2015. URL: <https://projects.coin-or.org/Ipopt>.

- [Jah07] J. Jahn. *Introduction to the Theory of Nonlinear Optimization*. 3rd ed. Springer, 2007.
- [KA82] L. Kantorovich and G. Akilov. *Functional analysis*. 2nd ed. Pergamon Press, 1982.
- [Kaw+15] Y. Kawajir et al. *Introduction to Ipopt: A tutorial for downloading, installing, and using Ipopt*. Revision: 2538. 2015.
- [KO82] C. Keno-Tung and N. Olhoff. “Regularized formulation for optimal design of axisymmetric plates”. In: *Internat. J. Solids Structures* 18.2 (1982), pp. 153–169.
- [Koi66] W. T. Koiter. “On the nonlinear theory of thin elastic shells”. In: *Koninklijke Nederlandse Akademie van Wetenschappen, Proceedings, Series B* 69.1 (1966), pp. 1–54.
- [Kuc11] H. Kuchling. *Taschenbuch der Physik: mit zahlreichen Tabellen*. Carl-Hanser-Verlag, 2011.
- [LB05] P. S. Lee and K. J. Bathe. “Insight into finite element shell discretizations by use of the “basic shell mathematical model””. In: *Computers & Structures* 83.1 (2005), pp. 69–90.
- [Lel+10] J. Lellep et al. “Optimization of stepped shells”. In: *WSEAS Transactions on Mathematics* 9.2 (2010), pp. 130–139.
- [Lel12] J. Lellep. *Shape Optimization of Cylindrical Shells*. Private Communication. Dec. 2012.
- [Lel15] J. Lellep. *Optimal beam*. Private Communication. June 2015.
- [Liu02] I.-S. Liu. *Continuum Mechanics*. Springer, 2002.
- [LP10] J. Lellep and A. Paltsepp. “Optimization of inelastic cylindrical shells with internal supports”. In: *Structural and Multidisciplinary Optimization* 41.6 (2010), pp. 841–852.
- [LP12] J. Lellep and E. Puman. “Optimization of conical shells of piece wise constant thickness”. In: *WSEAS Transactions on Mathematics* 11.3 (2012), pp. 242–251.
- [LS74] L. Lusternik and V. Sobolev. *Elements of Functional Analysis*. Vol. 55. Graduate texts in mathematics. Hindustan Publishing Corporation, 1974.
- [LSW99] T. Lepikult, W. Schmidt, and H. Werner. “Optimal design of rigid-plastic beams subjected to dynamical loading”. In: *Structural Optimization* 18.2 (1999), pp. 116–125.
- [Mat15a] The MathWorks, Inc. *Fortran Matrix Library API*. Dec. 2015. URL: <http://de.mathworks.com/help/matlab/fortran-mx-matrix-library.html>.

- [Mat15b] The MathWorks, Inc. *MAT-File API*. Dec. 2015. URL: <http://de.mathworks.com/help/matlab/read-and-write-matlab-mat-files-in-c-c-and-fortran.html>.
- [Mei07] J. Meiss. *Differential Dynamical Systems*. Mathematical Modeling and Computation. Society for Industrial and Applied Mathematics, 2007.
- [MKL12] MKL. *Intel Math Kernel Library, Reference Manual*. MKL 11.3. Intel, 2012.
- [MKL15] Intel Developer Zone. *Intel Math Kernel Library*. Oct. 2015. URL: <https://software.intel.com/en-us/intel-mkl>.
- [Nag63] P. M. Naghdi. “Foundations of elastic shell theory”. In: *Progress in solid mechanics*. Ed. by R. H. Ian Sneddon. Vol. 4. North Holland, 1963, pp. 1–90.
- [Nag72] P. M. Naghdi. “The theory of shells and plates”. In: *Mechanics of Solids II*. Ed. by C. Truesdell. Handbuch der Physik 6a/2. Springer, 1972, pp. 425–640.
- [Nes10a] P. Nestler. “Optimales Design einer Zylinderschale - eine Problemstellung der optimalen Steuerung in der Linearen Elastizitätstheorie”. PhD thesis. Greifswald: Mathematisch-Naturwissenschaftliche Fakultät der Universität Greifswald, 2010.
- [Nes10b] P. Nestler. “Optimales Design einer Zylinderschale - Existenz einer optimalen Lösung für diese Problemstellung”. Greifswalder Preprint-Reihe Mathematik. 2010.
- [Nes12] P. Nestler. “Optimal thickness of a cylindrical shell – an optimal control problem in linear elasticity theory”. In: *Ann. Acad. Romanian Sci., Ser. Math. Appl.* 4.2 (2012), pp. 183–208.
- [Nes13] P. Nestler. *pcg-methods working on element matrices only*. Private communication. Nov. 2013.
- [NST06] P. Neittaanmäki, J. Sprekels, and D. Tiba. *Optimization of elliptic systems: theory and applications*. 1st ed. Vol. 23. Springer Monographs in Mathematics. Springer, 2006.
- [Ola16] M. Ollagnon. *Orderpack 2.0*. May 2016. URL: <http://www.fortran-2000.com/rank/>.
- [OMP15] OpenMP. *OpenMP*. Oct. 2015. URL: <http://openmp.org/wp/>.
- [Par15] Intel Developer Zone. *Intel MKL PARDISO*. Oct. 2015. URL: <https://software.intel.com/en-us/node/470282>.
- [RHB06] K. Riley, P. Hobson, and S. Bence. *Mathematical Methods for Physics and Engineering: A Comprehensive Guide*. Cambridge University Press, 2006.

- [Saa03] Y. Saad. *Iterative Methods for Sparse Linear Systems: Second Edition*. Society for Industrial and Applied Mathematics, 2003.
- [ST09] J. Sprekels and D. Tiba. “Optimization Problems for Thin Elastic Structures”. In: *Optimal Control of Coupled Systems of Partial Differential Equations*. Ed. by K. Kunisch et al. Vol. 158. International Series of Numerical Mathematics. Birkhäuser, 2009, pp. 255–273.
- [SZ92] J. Sokolowski and J.-P. Zolésio. *Introduction to shape optimization*. Springer, 1992.
- [Tam14] J. Tambača. “A New Linear Shell Model for Shells with Little Regularity”. In: *J. Elasticity* 117.2 (2014), pp. 163–188.
- [TM05] R. Temam and A. Miranville. *Mathematical Modeling in Continuum Mechanics*. Cambridge University Press, 2005.
- [Trö10] F. Tröltzsch. *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*. Trans. by J. Sprekels. Vol. 112. Graduate studies in mathematics. American Mathematical Society, 2010.
- [Vin89] J. R. Vinson. *The Behavior of Thin Walled Structures: Beams, Plates and Shells*. 1st ed. Vol. 8. Mechanics of Surface Structure. Springer, 1989, p. 182.
- [Wal98] W. Walter. *Ordinary Differential Equations*. Graduate Texts in Mathematics. Springer, 1998.
- [WB06] A. Wächter and L. T. Biegler. “On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming”. In: *Mathematical Programming* 106 (2006), pp. 25–57.
- [Wlo87] J. Wloka. *Partial Differential Equations*. Cambridge University Press, 1987.
- [Yos80] K. Yosida. *Functional Analysis*. 6th ed. Vol. 123. Classics in Mathematics. Springer, 1980.
- [Zie13] P. Ziemann. “Optimal thickness of a cylindrical shell”. In: *International Conference on Optimization and Analysis of Structures*. Tartu, Estonia, 2013.
- [Zie14] P. Ziemann. “Optimal Thickness of a Cylindrical Shell”. In: *Ann. Acad. Rom. Sci. Ser. Math. Appl.* 6.2 (2014), pp. 214–234.
- [Zie15] P. Ziemann. “Optimal thickness of a cylindrical shell under dynamical loading”. In: *International Conference on Optimization and Analysis of Structures*. Tartu, Estonia, 2015.



# List of Figures

2.1. Example of a shell body with varying thickness based on a cylindrical midsurface . . . . .	15
5.1. 9-node reference element . . . . .	96
6.1. Initial geometry . . . . .	114
6.2. Applied force for Example 6.2 as colored midsurface . . . . .	115
6.3. Optimal thickness for Example 6.2 . . . . .	115
6.4. Displacement for optimal thickness of Example 6.2 . . . . .	116
6.5. Optimal thickness for Example 6.2 with a full tube . . . . .	118
6.6. Region (yellow) of applied force for Example 6.3 . . . . .	119
6.7. Optimal thickness for Example 6.3, region of applied force (dashed-line)	119
6.8. Displacement for optimal thickness of Example 6.3 . . . . .	120
6.9. Optimal thickness for Example 6.3 with unweighted objective . . . . .	121
6.10. Comparison of translational displacement for Example 6.3 with unweighted objective . . . . .	122
6.11. Cross section of the beam from Example 6.4 . . . . .	123
6.12. Optimal thickness for Example 6.4 . . . . .	124
6.13. Comparison of displacement component $u_3$ for Example 6.4 . . . . .	124
6.14. Comparison of analytical solution of (6.4.1) and numerical solution of Example 6.4 . . . . .	127
6.15. Applied force for Example 6.5 as colored midsurface . . . . .	127
6.16. Optimal thickness for Example 6.5, region of applied force (dashed-line)	128
6.17. Displacement for optimal thickness of Example 6.5 . . . . .	129
7.1. Applied (maximal) force for Example 7.1 as colored midsurface . . . . .	131
7.2. Optimal thickness for Example 7.1 . . . . .	132
7.3. Optimal displacement ( $\times 5 \cdot 10^3$ ) for Example 7.1 . . . . .	133
7.4. Comparison of translational displacement for Example 7.1 . . . . .	133
7.5. Maximal displacement (w.r.t. time) $\times 10^3$ for Example 7.1 . . . . .	134
7.6. Applied force for Example 7.2 as colored midsurface . . . . .	135
7.7. Optimal thickness for Example 7.2 . . . . .	136
7.8. Optimal thickness for different values of $T$ for Example 7.2 . . . . .	136
7.9. Optimal displacement ( $\times 10^4$ ) for Example 7.2 . . . . .	137
7.10. Comparison of translational displacement for Example 7.2 . . . . .	137
7.11. Optimal thickness on a coarse grid for Example 7.2 . . . . .	138
7.12. Magnitude of force and momentum distribution as colored midsurface for Example 7.3 . . . . .	139

7.13. Optimal thickness for Example 7.3 . . . . . 140  
7.14. Optimal displacement ( $\times 3 \cdot 10^3$ ) for Example 7.3 . . . . . 141  
7.15. Comparison of translational displacement for Example 7.3 . . . . . 141  
7.16. Optimal thickness on coarse grids for Example 7.3 . . . . . 142

## List of Tables

6.1. Comparison of optimal thickness on different grids for Example 6.2 .	117
6.2. Comparison of optimal thickness on different grids for Example 6.3 .	121
6.3. Comparison of optimal thickness on different grids for Example 6.5 .	129
7.1. Comparison of optimal thickness on different grids for Example 7.1 .	134
7.2. Comparison of optimal thickness on different grids for Example 7.2 .	138
7.3. Comparison of optimal thickness on different grids for Example 7.3 .	142
B.1. Source files for steady-state problem and referred implementation details	173
B.1. Source files for steady-state problem and referred implementation details	174
B.2. Source files for dynamic problem and referred implementation details	174
B.2. Source files for dynamic problem and referred implementation details	175

## List of Algorithms

5.1. Pseudo-code for CSR format (one degree of freedom per node) . . . .	103
5.2. Calculating the dynamic solution . . . . .	105
5.3. Calculating the adjoint solution . . . . .	110



# Publikationsliste

- Ziemann, P.: Optimal thickness of a cylindrical shell, International Conference on Optimization and Analysis of Structures. Tartu, Estonia (2013)
- Ziemann, P.: Optimal Thickness of a Cylindrical Shell. In: Ann. Acad. Rom. Sci., Ser. Math. Appl., 6.2, 214-234 (2015)
- Ziemann, P.: Optimal thickness of a cylindrical shell under dynamical loading. In: International Conference on Optimization and Analysis of Structures. Tartu, Estonia (2015)



# Danksagung

Ich möchte mich bei Prof. Dr. B. Kugelmann für die umfangreiche Betreuung während der Arbeit bedanken. Prof. Dr. W. Schmidt und Prof. Dr. J. Lellep danke ich für die Einführung in das Thema sowie für die aufschlussreichen Diskussionen bei meinen Besuchen an der Universität Tartu, Estland. Ein besonderer Dank geht an Dr. P. Nestler, dessen Arbeit eine wichtige Grundlage für meine Betrachtungen gegeben hat und dessen wertvolle Vorschläge, vor allem bei der numerischen Implementierung, zum Gelingen meiner Arbeit beigetragen haben. Auch an meinen ehemaligen Kollegen Dr. C. Lass geht ein großer Dank, da er stets für fruchtbare Diskussionen zur Verfügung stand. Dies gilt auch für Herrn J. Preuß, dem ich insbesondere für das gründliche Korrekturlesen der Arbeit danken möchte. Darüber hinaus bedanke ich mich bei allen meinen Kollegen, Freunden und Bekannten, deren Unterstützung ich sehr geschätzt habe.

Abschließend möchte ich mich bei meiner Familie  
und meiner Freundin Sarah bedanken,  
die während der gesamten Promotionszeit immer für mich da waren.



# A. Appendix

## A.1. Integral theorems

**Theorem A.1** (Gauss's theorem for second order tensors, [RHB06, p. 954]). Let  $T$  be a second order continuous differentiable tensor and  $V \subset \mathbb{R}^3$  a bounded subset with boundary  $\partial V = S$  and outward normal vector  $n$ . Then it holds

$$\int_V \sum_{j=1}^3 \frac{\partial}{\partial \xi^j} T^{ij} dV = \int_S \sum_{j=1}^3 T^{ij} n_j dS$$

or in short

$$\int_V \operatorname{div} T dV = \int_S T n dS.$$

**Theorem A.2.** Let  $\mathcal{B} \subset \mathbb{R}^3$  be a bounded region with Lipschitz-boundary  $\partial \mathcal{B}$  and outward normal vector  $n$  and  $v \in H^1(\mathcal{B})$ . Then it holds for any symmetric second-order tensor  $\sigma \in H^1(\mathcal{B})$  (e.g. the stress tensor, see Definition 2.8) and for the linearized strain tensor  $\varepsilon$  (see Definition 2.7)

$$\int_{\mathcal{B}} -(\operatorname{div} \sigma) \cdot v dV = \int_{\mathcal{B}} \sigma : \varepsilon(v) dV - \int_{\partial \mathcal{B}} (\sigma \cdot v) \cdot n dS.$$

*Proof.* Let  $\sigma^{ij}$  and  $v_j$  be the coordinates of  $\sigma$  and  $v$ , resp. It holds

$$\begin{aligned} \operatorname{div}(\sigma \cdot v) &= \sum_{i=1}^3 \frac{\partial}{\partial \xi^i} \sum_{j=1}^3 \sigma^{ij} v_j \\ &= \sum_{i=1}^3 \sum_{j=1}^3 \frac{\partial \sigma^{ij}}{\partial \xi^i} v_j + \sigma^{ij} \frac{\partial v_j}{\partial \xi^i} \\ &= (\operatorname{div} \sigma^T) \cdot v + \sigma^T : \nabla v \\ &= (\operatorname{div} \sigma) \cdot v + \sigma : \nabla v. \end{aligned}$$

Moreover, by using the property  $\sigma : \nabla v = \sigma^T : (\nabla v)^T$  it is

$$\sigma : \nabla v = \frac{1}{2} (\sigma + \sigma^T) : \nabla v = \sigma : \frac{1}{2} (\nabla v + (\nabla v)^T) = \sigma : \varepsilon(v).$$

Then it follows

$$\int_{\mathcal{B}} -(\operatorname{div} \sigma) \cdot v dV = \int_{\mathcal{B}} -\operatorname{div}(\sigma \cdot v) + \sigma : \varepsilon(v) dV = \int_{\mathcal{B}} \sigma : \varepsilon(v) dV - \int_{\partial \mathcal{B}} (\sigma \cdot v) \cdot n dS.$$

□

## A.2. Voigt notation

First- and second-order tensors can be represented with help of their components as vectors and matrices, resp. It is more difficult to find such representations for higher order tensors. Ideally, operations between higher order tensors like contraction or the  $\cdot$ -product are written as matrix-vector or matrix-matrix multiplications.

For symmetric tensors, which often arise in engineering applications, the *Voigt*- or *Mandel*-notation is used. It is based on the representation of symmetric tensors by reducing their order ([Hel01]). In this section the application to two-dimensional tensors will be discussed.

First, the representation of the tensors themselves: Let  $e$  be a symmetric second order tensor with covariant components  $e_{\alpha\beta}$ ,  $\alpha, \beta = 1, 2$ . Then the vector

$$e := \begin{pmatrix} e_{11} \\ e_{22} \\ \sqrt{2}e_{12} \end{pmatrix}$$

includes all information to represent  $e$ . The factor  $\sqrt{2}$  is actually due to Mandel, but the term *Voigt notation* is kept in further text. Let  $C$  be a symmetric fourth order tensor with contravariant components  $C^{\alpha\beta\lambda\mu}$ . With respect to fourth order tensors, *symmetric* means that the relations

$$C^{\alpha\beta\lambda\mu} = C^{\beta\alpha\lambda\mu} \quad C^{\alpha\beta\lambda\mu} = C^{\alpha\beta\mu\lambda}, \quad C^{\alpha\beta\lambda\mu} = C^{\lambda\mu\alpha\beta}$$

for  $\alpha, \beta, \lambda, \mu = 1, 2$  hold. Then the (symmetric) matrix

$$C := \begin{pmatrix} C^{1111} & C^{1122} & \sqrt{2}C^{1112} \\ C^{1122} & C^{2222} & \sqrt{2}C^{2212} \\ \sqrt{2}C^{1112} & \sqrt{2}C^{2212} & 2C^{1212} \end{pmatrix}$$

contains all information to represent  $C$ .

Second, operations between tensors can be rewritten in an elegant way: Let  $\sigma$  and  $e$  be two symmetric second-order tensors. Then for the double-dot product it holds

$$\sigma : e = \left( \begin{pmatrix} \sigma^{11} \\ \sigma^{22} \\ \sqrt{2}\sigma^{12} \end{pmatrix}, \begin{pmatrix} e_{11} \\ e_{22} \\ \sqrt{2}e_{12} \end{pmatrix} \right)$$

where  $(\cdot, \cdot)$  is the standard inner product for vectors. Let  $C$  and  $e$  be symmetric fourth- and second-order tensors, resp. Then the contraction  $\sigma := Ce$  with components

$$\sigma^{\alpha\beta} = (Ce)^{\alpha\beta} = \sum_{\lambda, \mu=1}^2 C^{\alpha\beta\lambda\mu} e_{\lambda\mu}$$

is a symmetric second order tensor which can be represented by the usual matrix-vector multiplication

$$\sigma = Ce.$$

### A.3. Functional Analysis

**Theorem A.3** (Sobolev embedding theorem, [AF03, p. 85f]). Let  $\omega \subset \mathbb{R}^n$  be a domain with Lipschitz-boundary and let  $j \geq 0$  and  $m \geq 1$  be integers and let  $1 \leq p < \infty$ . If  $mp > n$ , then there is a continuous embedding

$$W^{j+m,p}(\omega) \rightarrow C^j(\bar{\omega}).$$

**Theorem A.4** (Weakly sequentially relatively compact sets in Banach spaces, [Yos80, p. 126]). Let  $X$  be a reflexive Banach space and let  $x_n$  be a sequence in  $X$  which is bounded. Then there is a weakly convergent subsequence of  $x_n$ .

**Remark A.1.** *The above theorem can be generalized to: Every bounded subset of a reflexive Banach space is weakly sequentially relatively compact, see [Trö10, p. 46].*

### A.4. Vector-valued functions

This section is devoted to the concept of vector-valued functions. It is oriented on [Trö10, pp. 141ff.].

**Definition A.1** (Vector-valued function, [Trö10, p.141]). Any mapping from  $[a, b] \subset \mathbb{R}$  into a Banach space  $X$  is called a *vector-valued function*.

With help of the norm of the Banach space  $X$ , a concept of continuity can be introduced.

**Definition A.2** (Continuity of a vector-valued function, [Trö10, p. 142]). A vector-valued function  $y : [a, b] \rightarrow X$  is *continuous at the point*  $t \in [a, b]$  if

$$\lim_{s \rightarrow t} \|y(s) - y(t)\|_X = 0.$$

The space of vector-valued functions that are continuous at every  $t \in [a, b]$  is denoted by  $C(a, b; X)$ . It is a Banach space with respect to the norm

$$\|y\|_{C(a,b;X)} := \max_{t \in [a,b]} \|y(t)\|_X.$$

Moreover,  $L^p$ -spaces of vector-valued functions can be defined. For this, a little more preparation is necessary.

**Definition A.3** (Step function, [Trö10, p. 142]). A vector-valued function  $y : [a, b] \rightarrow X$  is called a *step function* if there are finitely many  $y_i \in X$  and Lebesgue measurable, pairwise disjoint sets  $M_i \subset [a, b]$ ,  $1 \leq i \leq m$ , such that  $[a, b] = \bigcup_{i=1}^m M_i$  and  $y(t) = y_i$  for every  $t \in M_i$ ,  $1 \leq i \leq m$ .

**Definition A.4** (Measurable vector-valued function, [Trö10, p. 142]). A vector-valued function  $y : [a, b] \rightarrow X$  is said to be *measurable* if there exists a sequence  $y_k$  of step functions  $y_k : [a, b] \rightarrow X$  such that  $y(t) = \lim_{k \rightarrow \infty} y_k(t)$  for almost every  $t \in [a, b]$ .

**Definition A.5** ( $L^p(a, b; X)$ -spaces, [Trö10, p. 143]). The space  $L^p(a, b; X)$ ,  $1 \leq p < \infty$  consists of all (equivalence classes of) measurable vector-valued functions  $y : [a, b] \rightarrow X$  having the property that

$$\int_a^b \|y(t)\|_X^p dt < \infty.$$

This space is a Banach space with respect to the norm

$$\|y\|_{L^p(a,b;X)} := \left( \int_a^b \|y(t)\|_X^p dt \right)^{\frac{1}{p}}.$$

It holds  $C(a, b; X) \subset L^p(a, b; X)$  for all  $p \geq 1$  ([Trö10, p. 143]).

For vector-valued functions in  $L^p$ -spaces, an integral can be defined. This is known as the Bochner integral.

**Definition A.6** (Bochner integral, [Trö10, p. 143]). For a step function  $y : [a, b] \rightarrow X$  with values  $y_i \in X$  on the sets  $M_i$  with Lebesgue-measure  $|M_i|$  the *Bochner integral* is defined as

$$\int_a^b y(t) dt := \sum_{i=1}^M y_i |M_i| \in X.$$

Since  $y \in L^p(a, b; X)$  is measurable there exists a sequence  $y_k$  of step functions converging almost everywhere on  $[a, b]$  to  $y$ . Then the Bochner integral of  $y$  is defined by

$$\int_a^b y(t) dt := \lim_{k \rightarrow \infty} \int_a^b y_k(t) dt.$$

The Bochner integral is indeed well-defined, because it does not depend on the choice of the sequence  $y_k$ , see [Trö10, p. 143].

In order to define derivatives of vector-valued functions, vector-valued distributions must be introduced first.

**Definition A.7** (Vector-valued distribution, [Trö10, p. 145]). Let  $V$  be a Banach space,  $y : [a, b] \rightarrow V$ . Then define  $\mathcal{T} : C_0^\infty(a, b) \rightarrow \mathcal{V}$  by

$$\mathcal{T}\varphi := \int_a^b y(t)\varphi(t) dt,$$

where the above integral is understood as the Bochner integral. The *vector-valued distribution*  $\mathcal{T}$  can be identified by the function  $y$ .

The derivative  $\mathcal{T}'$  of  $\mathcal{T}$  can be introduced as

$$\mathcal{T}'\varphi := - \int_a^b y(t)\varphi'(t) dt.$$

**Definition A.8** (Derivative of a vector-valued function, [Trö10, p. 146]). If for the above defined distribution  $\mathcal{T}$  with its derivative  $\mathcal{T}'$  there exists a vector-valued function  $w : [a, b] \rightarrow V$  with

$$\mathcal{T}'\varphi = \int_a^b w(t)\varphi(t) dt$$

for all  $\varphi \in C_0^\infty(a, b)$ , then  $\mathcal{T}'$  can be identified by  $w$  and  $w$  is called *derivative* of  $y$ ,

$$y'(t) := w(t).$$

In this sense it is  $y' \in L^1(a, b; V)$ .

This concept extends naturally for higher order derivatives using the formula of integration by parts

$$\int_a^b y(t)\varphi'(t) dt = y(t)\varphi(t)\Big|_a^b - \int_a^b y'(t)\varphi(t) dt = - \int_a^b y'(t)\varphi(t) dt$$

for all  $\varphi \in C_0^\infty(a, b)$  and continuously differentiable  $y : [a, b] \rightarrow V$ . Finally,  $H^k(a, b; X)$ -spaces can be introduced.

**Definition A.9** ( $H^k(a, b; X)$ -space). The space  $H^k(a, b; X)$ ,  $k \geq 0$ , consists of all vector-valued functions  $y \in L^2(a, b; X)$  which have all derivatives up to the order of  $k$  in  $L^2(a, b; X)$ .



## B. Source code overview

The following sections present an overview of the attached source files on the DVD. Please note that the DVD does not contain an executable program in order to avoid the redistribution of the utilized third-party libraries (Ipopt, Intel MKL, Matlab libraries, ...). However, the actual important part is the self-written code, which is the content of the DVD. There are two source subfolders contained in the folder `Source_Code`, one (`SteadyStateCase`) containing the source files for the steady-state optimization (Chapter 6), and the other one (`DynamicCase`) containing the source files for the dynamic optimization (Chapter 7).

The folder `mat-files` contains the files with the optimization results presented in Chapter 6 and 7. Those files can be opened with Matlab or Octave. The nodal values of optimal thickness are saved in the variable `x`, those for the corresponding displacement in `displ`, and those for the adjoint in `adj`. The coordinates of the nodes can be found in the matrix `FEMData.Coords`. The objective value is saved in `fval`. The several problem parameters (force, objective, ...) are stored in the structure `FEMData.opts`. The results can be plotted with the help of the Matlab routines contained in the folder `Plot_functions`.

### B.1. Steady-state case

The folder `SteadyStateCase` includes the self-written source code for the optimization of the steady-state problem. The below table contains selected files which implement the referred part of the discussion of the numerical solution in Chapter 5.

Table B.1.: Source files for steady-state problem and referred implementation details

file	implemented part	to find on page
<code>mod_BasisTransf</code>	provide basis transformation matrix (5.1.6) and their derivative (5.1.7)	99
<code>mod_calcLambda</code>	choose regularization parameter to ensure small influence of the regularization term on the objective value	113
<code>mod_calcVolume</code>	calculate volume of shell from nodal thickness values (5.3.3)	106
<code>mod_createStart</code>	interpolate solution on coarse grid for use as initial solution on refined grid	110
<code>mod_csrMat</code>	provide structure to manage sparse matrix CSR format	102

Table B.1.: Source files for steady-state problem and referred implementation details

file	implemented part	to find on page
<code>mod_erg2MatFile</code>	save results to mat-file	111
<code>mod_FEMData</code>	create 9-node rectangle elements, refine grid	96
<code>mod_FormFGP</code>	calculate values of shape functions from (5.1.2)	97
<code>mod_getElemSt</code>	calculate element stiffness matrix and load vector for state equation (5.1.1)	100, 103
<code>mod_getElemVals</code>	calculate objective value, objective gradient, and adjoint load vector on element	106, 109
<code>mod_main</code>	assemble element matrices (Algorithm 5.1), solve state and adjoint equation	103
<code>mod_pcg</code>	provide preconditioned conjugate gradient method	107
<code>mod_VerschVerzerrMat</code>	create strain-displacement matrices	99
<code>mod_zielfun</code>	calculate objective value and finite element gradient	106, 107
<code>mod_optimiere</code>	main program	

## B.2. Dynamic case

The folder `DynamicCase` includes the self-written source code for the optimization of the dynamic problem. The below table contains selected files which implement the referred part of the discussion of the numerical solution in Chapter 5.

Table B.2.: Source files for dynamic problem and referred implementation details

file	implemented part	to find on page
<code>mod_BasisTransf</code>	provide basis transformation matrix (5.1.6) and their derivative (5.1.7)	99
<code>mod_calcLambda</code>	choose regularization parameter to ensure small influence of the regularization term on the objective value	131
<code>mod_calcVolume</code>	calculate volume of shell from nodal thickness values (5.3.3)	106
<code>mod_createStart</code>	interpolate solution on coarse grid for use as initial solution on refined grid	110
<code>mod_csrMat</code>	provide structure to manage sparse matrix CSR format	102
<code>mod_erg2MatFile</code>	save results to mat-file	111

Table B.2.: Source files for dynamic problem and referred implementation details

file	implemented part	to find on page
mod_FEMData	create 9-node rectangle elements, refine grid	96
mod_FormFGP	calculate values of shape functions from (5.1.2)	97
mod_getAdjElemLoad	calculate element load vector for adjoint equation (5.3.5)	109
mod_getElemGrad	calculate objective gradient on element (5.3.7)	110
mod_getElemLoad	calculate element load vector for state equation (5.2.1)	104
mod_getElemMass	calculate element mass matrix for state equation (5.2.1)	104
mod_getElemSt	calculate element stiffness matrix	100
mod_getElemZielf	calculate objective value on element (5.3.1)	106
mod_InitDispl	create initial value $W_h^0$ from momentum $g$	104
mod_main	assemble element matrices (Algorithm 5.1), solve state and adjoint equation (Algorithm 5.2 and 5.3)	103, 105, 110
mod_pcg	provide preconditioned conjugate gradient method	107
mod_VerschVerzerrMat	create strain-displacement matrices	99
mod_zielfun	calculate objective value and finite element gradient	106, 107
mod_optimiere	main program	