Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

LARS LUBKOLL, ANTON SCHIELA & MARTIN WEISER

An optimal control problem in polyconvex hyperelasticity [1]

[1]Supported by the DFG Research Center MATHEON "Mathematics for key technologies", Berlin

Herausgegeben vom
Konrad-Zuse-Zentrum für Informationstechnik Berlin
Takustraße 7
D-14195 Berlin-Dahlem

Telefon: 030-84185-0
Telefax: 030-84185-125

e-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

An optimal control problem in polyconvex hyperelasticity [†]

Lars Lubkoll, Anton Schiela & Martin Weiser

February 27, 2012

Abstract

We consider a shape implant design problem that arises in the context of facial surgery. We introduce a reformulation as an optimal control problem, where the control acts as a boundary force. The state is modelled as a minimizer of a polyconvex hyperelastic energy functional. We show existence of optimal solutions and derive – on a formal level – first order optimality conditions. Finally, preliminary numerical results are presented.
AMS MSC 2000: 49J20, 74B20, 65N30

Keywords: polyconvex elasticity, implant design, optimal control

1 Introduction

As in many parts of modern medicine the design of implants is today more dependent on the experience of medical scientists than on technical tools. In most cases the determination of an implant's shape is done by visually comparing CT scans with implant models, choosing a model that seems to fit and possibly correct its shape during the insertion procedure. This approach is very sensitive to the surgeon's skills and the geometry of the implant. Especially in the case of heavy fractures or congenital deformations of the oral and maxillofacial bone structure it is often difficult to accurately predict the shape of the patients face after the medical treatment. Consequently it would be of advantage if one could delegate the determination of an implant's shape from a given desired shape of the skin to a computer-assisted tool. This would allow to give reliable assistance regarding the training, preparation and verification of implant insertions. In order to provide such a tool it is necessary to develop appropriate mathematical models and numerical schemes for the calculation of the implant's shape.

In Section 2 such a model will be derived, leading to an optimal control problem with constraints arising from elastostatics:

$$\min J(u, g) \quad \text{s.t.} \quad u \in \operatorname{argmin} \mathcal{E}(u, g) \tag{1}$$

[†]Supported by the DFG Research Center MATHEON "Mathematics for key technologies", Berlin

where g is the control and u the corresponding state (the material's displacement). The explicit description of the constraint $u \in \operatorname{argmin} \mathcal{E}(u, g)$ is determined by the chosen material law(s) for the soft tissue(s). As large strains in facial soft tissue should be allowed in the model, various nonlinear effects must be considered.

First of all there is the geometric nonlinearity, whose neglection leads to well-known overestimation of the displacements (for an illustration we refer to Fig. 4 in [34]). Then there are constitutive nonlinearities that possibly must be taken into account. The latter is to a great portion a consequence of the distribution of collagen in most types of human soft tissue. Being the main load carrying element and the most common protein in human soft tissue with particularly high concentration in the skin and, in contrast to other muscles, the facial muscle tissue, the high amount of collagen strongly determines the material behaviour [15, 17, 18]. On the one hand, the collagen distribution leads to a nonlinear stress-strain relationship, mainly dependent on the collagen fibre morphology corresponding to the current stress state. This observation is outlined in [17] and reflected by Fung-elastic material laws [15]. On the other hand, the distribution of the collagen fibres endows the material with directional properties, i.e. while the stiffness increases with muscle contraction in direction of the collagen fibres it remains constant in orthogonal directions [10, 17], thus leading to a strongly anisotropic behaviour. This is complemented by the observation that these fibre directions may change during the deformation. Thus the accurate modeling of anisotropic effects is not trivial and requires the knowledge of collagen fibre orientations and distributions in the considered materials. There also is a constitutive nonlinear inequality that is associated with limited compressibility and takes the form

$$\det(\nabla \Phi(x)) = \det((I + \nabla u)(x)) > 0 \tag{2}$$

where $\Phi = I + u$ is the deformation. In the case that $\Phi \in C^1$ this inequality serves as a local "orientation-preserving" condition that locally, not globally, prevents self-penetration of the considered material (see [11] and references therein).
Finally, normal pressure boundary conditions imposed on the deformed domain lead to nonlinear boundary conditions on the undeformed domain [5, 11].

The currently most general class of stationary material laws that can incorporate the mentioned nonlinearities and is accessible to mathematical analysis are hyperelastic constitutive laws given by polyconvex stored energy functions [5]. This class, that will be considered in this paper, includes popular material laws for large strains such as Neo-Hookean, Mooney-Rivlin [23, 29]), Ogden-type ([26, 27]) as well as anisotropic, Fung-elastic material laws as in [17]. Moreover polyconvexity is closely related to the Legendre-Hadamard condition [13, 24], which guarantees the ellipticity of the differential operator of linearized elasticity. For physical interpretations of the Legendre-Hadamard condition we refer to [11, 22, 28] and references therein. Eventually it is also related to (weak) lower semi-continuity [11, 24, 33], and thus admits John Ball's elegant proof of the existence problem in elasticity [5]. A generalization of these results by P.G. Ciarlet [11] will be used in Section 3 to

prove the existence of solutions of the corresponding optimal control problem in the subset of $W^{1,p},\ 2 \leq p < \infty$, that is associated with elastic, compressible deformations. Section 4 is concerned with first order optimality conditions. As in general hyperelastic theory it is not even clear if a local minimizer of the elastic energy functional satisfies the weak form of the corresponding Euler-Lagrange equation (see [7, Problems 5 & 6] and [11, 22]), the rigorous derivation of general first order optimality conditions currently appears to be out of reach. Despite of being related to the regularity of minimizers [6, 19], polyconvexity and coercivity are not sufficient for the determination of the desired regularity results. This will be illustrated and discussed for a compressible Mooney-Rivlin material in Section 4. Finally the formally derived optimality system will be solved numerically in Section 6.

The used notation follows the conventions of elasticity theory:

Notation.

1. *The (right) Green-St.Venant strain tensor is denoted by*
$$E(u) = \frac{1}{2}\left(\nabla u^T + \nabla u + \nabla u^T \nabla u\right)$$
2. *$\mathbb{M}^n$ denotes the set of $n \times n$ matrices and*
$$\mathbb{M}^n_+ := \{A \in \mathbb{M}^n \mid \det(A) > 0\}$$
3. *The scalar product on $\mathbb{M}^n$ is defined via*
$$F : G = \sum_{i,j} F_{ij} G_{ij} \quad \text{for } F, G \in \mathbb{M}^n$$
and induces the Frobenius norm $\|\cdot\|_M$.
4. *For derivatives subscripts will be used, i.e.*
$$\mathcal{E}_u = \frac{\partial}{\partial u}\,\mathcal{E}$$
5. *For invertible matrices $F \in \mathbb{M}^n$ the adjugate matrix is defined via*
$$\operatorname{adj}(F) = \det(F) F^{-T}$$

2 Modeling

In this section, we will derive from medical requirements a precise mathematical formulation of the implant shape design problem. We start with formulating the forward problem of finding the facial shape induced by a given implant shape as a contact problem in Section 2.1. In Section 2.2 we will see that the direct transcription of the forward problem into the inverse problem of finding an optimal implant shape such as to approximate a desired facial shape leads to quite difficult optimization problem. Surprisingly, a simple reformulation turns out to be a standard optimal control problem.

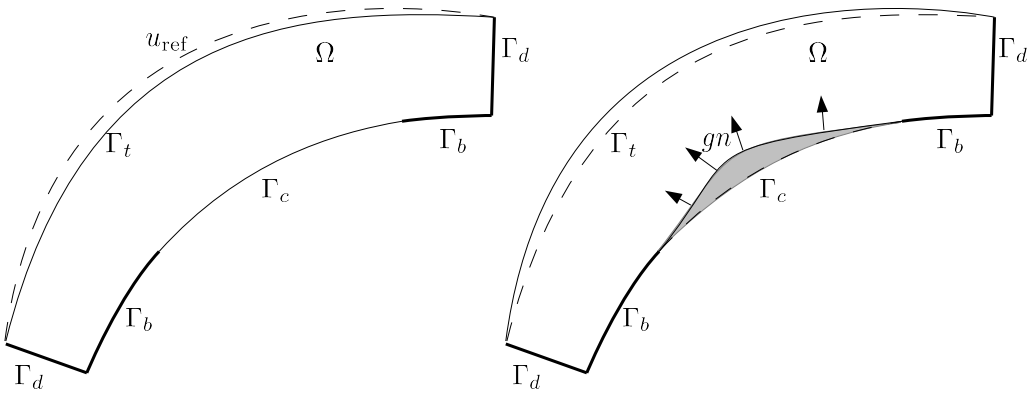


Figure 1: Cross-section of the reference configuration (left) and the deformed state due to the normal force gn defining the implant volume in gray (right).

2.1 Forward problem

The facial shape is determined by the elastic deformation of the soft tissue. In contrast, bone and implant are considered as rigid, such that only the soft tissue domain Ω is considered. We will restrict the attention to implants of limited geometric complexity and hence assume its manifold shape to be parametrized over Γ_c as a continuous normal displacement

$$y \mapsto y + s(y)n(y) \quad \text{for } y \in \Gamma_c,$$

where $n(y)$ is the unit outer normal of Ω at $y \in \Gamma_c$. Here, Γ_c is the part of the interior soft tissue boundary where it normally is in contact with bone, see Fig. 1. The implant displaces the soft tissue, which can freely glide over the implant surface but may not penetrate it. Hence, an obstacle condition has to be imposed on Γ_c. In a ring Γ_b around the implant region Γ_c we assume the soft tissue to be attached to the bone.

Due to the quickly vanishing Green's function of elastomechanics, the soft tissue domain may be restricted to a bounded region in the vicinity of the implant by introducing an artificial boundary Γ_d cutting the soft tissue. Here, transparent boundary conditions [20] might be imposed. For simplicity, we just assume the tissue to be fixed on Γ_d. On the skin surface Γ_t, natural boundary conditions hold.

The forward problem is therefore to find the minimizer of the stored strain energy

$$\mathcal{E}^S = \int_\Omega W(x, \nabla u(x))\,\mathrm{d}x, \tag{3}$$

where W is a stored energy function of hyperelasticity, subject to the constraints given by the boundary conditions:

$$\min_u \mathcal{E}^S(u) \tag{4a}$$

subject to

$$u = 0 \qquad \text{on } \Gamma_d \cup \Gamma_b, \tag{4b}$$

$$n(y)^T[x + u(x) - y] \geq s(y) \quad \text{for all } x, y \in \Gamma_c \text{ with } x + u(x) \in y + \mathbb{R}\, n(y) \tag{4c}$$

In particular the global non-penetration condition (4c) is difficult to address algorithmically, as a direct mapping from y to x in Γ_c depends on the solution, is potentially multi-valued, and usually not readily available. Note that W may also be heterogeneous, i.e. $W = W(x, \nabla u(x))$. Here we restricted the discussion to homogeneous stored energy functions,

This problem can also be written in strong form, if we introduce, the first Piola-Kirchhoff tensor $\sigma(u) = \tilde{\sigma}(\nabla u)$, using the definition of hyperelasticity, i.e. the point-wise relation

$$\tilde{\sigma}(F) = \frac{\partial W}{\partial F}(x, F) \quad x \in \Omega,\ F \in \mathbb{M}^3_+ \,. \tag{5}$$

Then we obtain as usual by formal partial integration

$$-\operatorname{div}(\sigma(u)) = 0 \qquad \text{in } \Omega \tag{6a}$$

$$u = 0 \qquad \text{on } \Gamma_b \cup \Gamma_d \tag{6b}$$

$$n(y)^T[x + u(x) - y] \geq s(y) \quad \text{for all } x, y \in \Gamma_c \text{ with } x + u(x) \in y + \mathbb{R}\, n(y) \tag{6c}$$

2.2 Inverse problem

Now the optimization problem consists of finding an implant shape given by the normal displacement $s(y)$, such that a desired facial shape is well approximated. Again for simplicity, we will consider the mismatch

$$J(u) = \frac{1}{2}\|u - u_{\text{ref}}\|^2_{L^2(\Gamma_t)}$$

of displacement u and a desired displacement u_{ref} on the facial surface, which is to be minimized subject to the obstacle problem (4). This formulation of the optimization problem as an MPEC (mathematical program with equilibrium constraints) has two mathematical drawbacks: it is algorithmically challenging, and the solutions are in general not unique (not even locally).

Moreover, additional medical requirements need to be satisfied. For instance no gaps should occur between soft tissue and implants since voids tend to be the source of infections. Such gaps can occur between soft tissue and implant surface wherever (4c) is a strict inequality. Consequently, the set of feasible implant shapes has to be restricted to those leading to contact everywhere in Γ_c, such that inequality (4c) can be replaced by the simpler corresponding equality. Taking into account that the soft tissue is not attached to the implant, thus being free to move in tangential directions at the contact surface, yields that the force, which is exerted by the

implant on the soft tissue, can only act in the surface's normal direction. In addition the implant cannot pull the soft tissue, eventually leading to pressure boundary conditions, i.e. the exerted force must be a negative multiple of the surface unit outer normal. In the reference configuration, the normal stress then assumes the form

$$\sigma(u)n = -g\,\mathrm{adj}(I+\nabla u)^T n, \quad g \geq 0 \quad \text{on } \Gamma_c. \tag{7}$$

Interpreting the nonnegative normal force g exerted by the implant as control variable instead of the implant shape itself leads to a significantly simplified optimization problem. In particular, the obstacle condition (4c) is replaced by (7).

Due to the change of control variable from normal displacement s to normal force g, an explicit mapping between different points in Γ_c as required in (4c) is no longer needed. Moreover, in applications large boundary forces are unwanted, thus motivating the expansion of the cost functional J by the penalization term

$$\frac{\alpha}{2}\|g\|^2_{L^2(\Gamma_c)}, \ \alpha > 0. \tag{8}$$

Note that the applied penalization coincides with the well-known Tikhonov regularization for inverse problems.

Then we end up with the control constrained optimal control problem subject to the following equations of elasticity in strong form:

$$\min_{u,g} J(u) + \frac{\alpha}{2}\|g\|^2_{L^2(\Gamma_c)} \tag{9a}$$

$$\begin{aligned}
\text{subject to} \qquad -\operatorname{div}(\sigma(u)) &= 0 && \text{in } \Omega && \text{(9b)}\\
u &= 0 && \text{on } \Gamma_b \cup \Gamma_d && \text{(9c)}\\
\sigma(u)n &= -g\,\mathrm{adj}(I+\nabla v)^T n, \ g \geq 0 && \text{on } \Gamma_c && \text{(9d)}
\end{aligned}$$

Eventually, from an optimal soft tissue displacement u solving (9), the implant shape can be reconstructed by filling the gap between reference and deformed inner soft tissue boundary. Again it is parametrized over Γ_c, but now in the form

$$x \mapsto x + u(x) \quad \text{for } x \in \Gamma_c.$$

By construction, no undesirable voids can occur, and no explicit a priori representation of the implant's shape is required. In the following, for sake of clarity, we will concentrate the parts of the boundary where homogeneous Dirichlet boundary conditions are imposed: $\Gamma_d = \Gamma_d \cup \Gamma_b$.

3 Existence of solutions

Our first step in the analysis of problem (9) is the study of existence of optimal controls g and corresponding deformations u. Our approach heavily relies on the

model assumption that for given g, the corresponding u is a minimizer of the hyperelastic energy functional. Thus, of necessity, there must be an energy functional that corresponds to the equilibrium of forces, imposed at the boundary, which reads in our case

$$\sigma n = -g \operatorname{adj}(I + \nabla u)^T n. \tag{10}$$

Unfortunately, except for the case of spatially constant control $g(x) = \text{const.}$ on Γ_c, a conservative formulation of these boundary conditions is in general not available ([5, 9]), leaving it as an open issue to model these conditions correctly.

For this reason, we will switch to a simplified setting, namely we will replace (10) by one of the following two dead load boundary conditions:

$$\begin{aligned} \sigma n &= -gn \quad g : \Gamma_c \to \mathbb{R} \\ \sigma n &= -g \quad g : \Gamma_c \to \mathbb{R}^3 \end{aligned}$$

Both conditions naturally enter linearly into the energy functional (see (14)) and can be augmented by a positivity constraint, such as $g \geq 0$ in the first case or $g^T n \geq 0$ in the second case. In the context of nonlinear elasticity even this simplified problem is already a delicate issue, since there is hardly more analytical structure available than weak lower semi-continuity of the energy functional. To render the discussion precise, we will now state a list of assumptions for the stored energy functional and the objective functional. Note that the assumptions on the stored energy functional are quite standard in non-linear elasticity (see [5, 11, 22, 28]).

Assumption 3.1.

1. *Ω is a bounded Lipschitz-domain and $\partial\Omega = \overline{\Gamma_d \cup \Gamma_t \cup \Gamma_c}$, $|\Gamma_c| > 0, |\Gamma_t| > 0$ a measurable partition of its boundary.*

2. *The space for of admissible deformations is contained in*

$$U := \{u \in \mathbf{W}^{1,p}(\Omega) : \operatorname{adj}(\nabla u) \in \mathbf{L}^q(\Omega), \det(I + \nabla u) \in L^r(\Omega)\},$$

   where $p \geq 2$, $q \geq p/(p-1)$, and $r > 1$.

3. *On Γ_d Dirichlet boundary conditions are imposed:*

$$u|_{\Gamma_d} = u_0 \in \mathbf{W}^{\frac{p-1}{p},p}(\Gamma_d)$$

4. *The stored energy function $W : \Omega \times \mathbb{M}^3 \to \mathbb{R} \cup \{+\infty\}$ exhibits the following properties:*

   Polyconvexity: *For almost all $x \in \Omega$ there exists a convex lower semi-continuous function*

$$\mathbb{W}(x, \cdot, \cdot, \cdot) : \mathbb{M}^3 \times \mathbb{M}^3 \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$$

such that

$$\mathbb{W}\left(x, F, \operatorname{adj}(F), \det(F)\right) = W\left(x, F\right) \quad \forall F \in \mathbb{M}^3_+$$

and

$$\mathbb{W}(\cdot, F, H, \delta) : \Omega \to \mathbb{R} \cup \{+\infty\}$$

is measurable for all $(F, H, \delta) \in \mathbb{M}^3 \times \mathbb{M}^3 \times]0, \infty[$.

Non-self penetration: *For almost all* $x \in \Omega$ *it holds that*

$$\lim_{\det(F) \to 0^+} W(x, F) = +\infty \tag{11}$$

and

$$W(x, F) = +\infty \text{ for all } F \in \mathbb{M}^3 \setminus \mathbb{M}^3_+ . \tag{12}$$

Coercivity: *There exist constants* $\alpha > 0,\ \beta \in \mathbb{R}^3$, *such that*

$$W(x, F) \geq \alpha \left(\|F\|_M^p + \|\operatorname{adj}(F)\|_M^q + |\det(F)|^r \right) + \beta \tag{13}$$

for all $F \in \mathbb{M}^3_+$ *and almost all* $x \in \Omega$.

The elastic strain energy is given by

$$\mathcal{E}^S(u) = \int_\Omega W\left(x, I + \nabla u(x)\right) \mathrm{d}x,$$

and there exists at least one admissible deformation $\overline{u}$ *such that* $\mathcal{E}^S(\overline{u}) < \infty$.

In view of Section 2 we will impose the following assumptions on the control and the objective functional:

Assumption 3.2.

1. *The control* g *is taken to be an element of* $G = \mathbf{L}^2(\Gamma_c)$ *and enters the total elastic energy functional via*

$$\mathcal{E}(u, g) = \mathcal{E}^S(u) - \mathcal{E}^{\Gamma_c}(u, g) \tag{14}$$

with $\mathcal{E}^{\Gamma_c}(u, g) = \int_{\Gamma_c} g(s) u(s) \, \mathrm{d}s$.

2. *The cost functional* $J(u, g) : U \times G \to \mathbb{R}$ *is weakly lower semicontinuous and there exist a constant* $\alpha_J > 0$ *such that*

$$J(u, g) \geq \alpha_J \|g\|_G^2 \tag{15}$$

Remark 3.1. *Extending* $\operatorname{dom}(W)$ *to* $\mathbb{M}^3$ *and* $\operatorname{ran}(W)$ *to* $\mathbb{R}_+ \cup \{\infty\}$, *compared with its classical definition in the context of elasticity theory, allows to reduce the length of the following proofs, as the orientation-preserving property* $\det(I + \nabla u) > 0$ *a.e. is a direct consequence of the assumption* $\mathcal{E}(u, g) < \infty$

Remark 3.2. *Note that the above assumptions include mixed* displacement-traction *as well as* pure traction problems*. With respect to the latter adequate choices of the cost functional may remove the "indeterminacy up to rigid translations" [11, 12].*

Theorem 3.1. *Suppose that Assumptions 3.1 and 3.2 hold. Then the optimal control problem*

$$\min_{(u,g)\in U\times G} J(u,g) \quad s.t. \quad u \in \operatorname*{argmin}_{v\in U} \mathcal{E}(v,g) \tag{16}$$

has at least one solution.

Before turning to the proof of this theorem we will first state two important lemmas that will be required therein. We start with a result on compensated compactness, which has been stated in [5, Section 6] and in a clearer version in [11, Chapter 7]. It can be viewed as the main step in the proof of existence of energy minimizers in nonlinear elasticity.

Lemma 3.2. *Let $\Phi \in \mathbf{W}^{1,p}(\Omega),\ p \geq 2$ and $r, q > 0$ such that $r^{-1} = p^{-1} + q^{-1} \leq 1$. Then the following implication holds:*

$$\left.\begin{aligned} \Phi^k &\rightharpoonup \Phi \text{ in } \mathbf{W}^{1,p}(\Omega),\ p \geq 2 \\ \operatorname{adj}(\nabla\Phi^k) &\rightharpoonup \rho \text{ in } \mathbf{L}^q(\Omega),\ \frac{1}{p}+\frac{1}{q} \leq 1 \\ \det(\nabla\Phi^k) &\rightharpoonup \delta \text{ in } L^r(\Omega), r \geq 1 \end{aligned}\right\} \Rightarrow \begin{cases} \rho = \operatorname{adj}(\nabla\Phi) \\ \delta = \det(\nabla\Phi) \end{cases}$$

Proof. See [11, Thm. 7.6-1]. □

Using the above result and the theorem of Mazur one can prove the sequential weak lower semi-continuity of $\mathcal{E}^S$ with respect to sequences u_k for which $\mathcal{E}^S$ remains bounded (see [11, Proof of Thm. 7.7-1]). This result can be extended in the following way:

Lemma 3.3. *Let Assumptions 3.1 and 3.2 hold. Consider a weakly converging sequence $(u_k, g_k) \rightharpoonup (\tilde{u}, \tilde{g})$ in $U \times G$ such that*

$$u_k \in \operatorname*{argmin}_{v\in U} \mathcal{E}(v, g_k)$$

and $\mathcal{E}(u_k, g_k)$ is bounded from above. Then

$$\lim_{k\to\infty} \mathcal{E}(u_k, g_k) = \mathcal{E}(\tilde{u}, \tilde{g}) = \min_{v\in U} \mathcal{E}(v, \tilde{g}) \tag{17}$$

Proof. First of all, we show the weak lower semi-continuity of $\mathcal{E}$ for sequences that leave the energy bounded from above.

Weak lower semi-continuity of the first part $\mathcal{E}^S$ with respect to u_k follows as in [11, Proof of Thm. 7.7-1] from Lemma 3.2 and convexity of the functional $\mathbb{W}$ with respect to its arguments. The second part

$$\mathcal{E}^{\Gamma_c}(u_k, g_k) = \int_{\Gamma_c} u_k g_k \, ds$$

is even *weakly continuous*. This follows via compactness of the trace mapping $\mathbf{W}^{1,p}(\Omega) \hookrightarrow \mathbf{L}^2(\Gamma_c)$, by strong convergence $u_k|_{\Gamma_c} \to \tilde{u}|_{\Gamma_c}$ in $\mathbf{L}^2(\Gamma_c)$ and weak convergence $g_k \rightharpoonup \tilde{g}$ in $\mathbf{L}^2(\Gamma_c)$. In summary, we can conclude weak lower semi-continuity of $\mathcal{E}$:

$$\mathcal{E}(\tilde{u}, \tilde{g}) \leq \liminf_{k\to\infty} \mathcal{E}(u_k, g_k),$$

and, if u is fixed,

$$\lim_{k\to\infty} \mathcal{E}(u, g_k) = \mathcal{E}(u, \tilde{g}).$$

Next, by the minimizing property of u_k, we obtain $\mathcal{E}(u_k, g_k) \leq \mathcal{E}(\tilde{u}, g_k)$ and

$$\limsup_{k\to\infty} \mathcal{E}(u_k, g_k) \leq \limsup_{k\to\infty} \mathcal{E}(\tilde{u}, g_k) = \lim_{k\to\infty} \mathcal{E}(\tilde{u}, g_k) = \mathcal{E}(\tilde{u}, \tilde{g}),$$

implying

$$\limsup_{k\to\infty} \mathcal{E}(u_k, g_k) \leq \mathcal{E}(\tilde{u}, \tilde{g}) \leq \liminf_{k\to\infty} \mathcal{E}(u_k, g_k)$$

and thus

$$\lim_{k\to\infty} \mathcal{E}(u_k, g_k) = \mathcal{E}(\tilde{u}, \tilde{g}).$$

The fact that $\tilde{u}$ is again an energy minimizer of $\mathcal{E}(\cdot, \tilde{g})$ follows from the minimizing property of u_k and the established convergence result. To this end let $\bar{u}$ be a minimizer of $\mathcal{E}(\cdot, \tilde{g})$. Then

$$\mathcal{E}(\bar{u}, \tilde{g}) \leq \mathcal{E}(\tilde{u}, \tilde{g}) = \lim_{k\to\infty} \mathcal{E}(u_k, g_k) \leq \lim_{k\to\infty} \mathcal{E}(\bar{u}, g_k) = \mathcal{E}(\bar{u}, \tilde{g}). \qquad \square$$

Observe the two structural properties that make this proof work. First, linearity of $\mathcal{E}^{\Gamma_c}$ with respect to g, second compactness of the trace mapping $W^{1,p}(\Omega) \hookrightarrow \mathbf{L}^2(\Gamma_c)$. Our proof extends to any $\mathcal{E}^{\Gamma_c}$ with the same abstract properties.

Proof of Theorem 3.1: First we show that we can apply Lemma 3.3. Then, using the weak lower semicontinuity of J, we will show that there exists an admissible minimizing sequence $(u_k, g_k)_{k\in\mathbb{N}}$ converging weakly in $U \times G$ to a minimizer $(\tilde{u}, \tilde{g})$ of the optimal control problem. Eventually exploiting the coerciveness of $\mathcal{E}$ will lead to the admissibility of the weak limit $(\tilde{u}, \tilde{g})$, i.e.

$$\operatorname{adj}(\nabla\tilde{u}) \in \mathbf{L}^q(\Omega) \quad \text{and} \quad \det(I + \nabla\tilde{u}) \in L^r(\Omega).$$

Existence of a weakly convergent subsequence:
As has been shown in [5, Thm. 7.3&7.6],[11, Thm. 7.7-1] for every $g_k \in G$ there exists a displacement $u_k \in U$ such that $u_k \in \operatorname{argmin}_{v\in U} \mathcal{E}(v, g_k)$. Thus, as the energy functional $J(u, g)$ is bounded from below, there exists a minimizing sequence $(u_k, g_k)_{k\in\mathbb{N}}$ of J with $g_k \in G$, $u_k \in U$ and u_k being a minimizer of $\mathcal{E}(\cdot, g_k)$. From (15) we deduce that the sequence $\{g_k\}_{k\in\mathbb{N}}$ is bounded in G by some constant C_g and by reflexivity of G there exists a weakly convergent subsequence which will again be denoted as $\{g_k\}_{k\in\mathbb{N}}$ with weak limit $\tilde{g} \in G$.
First, we have to show that the sequence $\{\mathcal{E}(u_k, g_k)\}_{k\in\mathbb{N}}$ is bounded from above.

Setting $\|\cdot\|_U := \|\cdot\|_{\mathbf{W}^{1,p}(\Omega)}$, $\|\cdot\|_G := \|\cdot\|_{\mathbf{L}^2(\Gamma_c)}$, using Hölder's inequality and the continuity of the trace operator we get an estimate for the sensitivity of the elastic energy functional with respect to changes in the Neumann boundary conditions:

$$\mathcal{E}(u, g_k) - \mathcal{E}(u, 0) = \int_{\Gamma_c} u(0 - g_k)\,\mathrm{d}s \leq \|u\|_U \|g_k\|_G \leq C_g \|u\|_U \tag{18}$$

Thus as u_k minimizes $\mathcal{E}(\cdot, g_k)$, the boundedness of $\{\mathcal{E}(u_k, g_k)\}_{k\in\mathbb{N}}$ is a consequence of (18), inserting $u = \overline{u}$ as defined at the end of Assumption 3.1:

$$\mathcal{E}(u_k, g_k) \leq \mathcal{E}(\overline{u}, g_k) \leq C_g \|\overline{u}\|_U + \mathcal{E}(\overline{u}, 0) < \infty.$$

Now the boundedness of $\{u_k\}_{k\in\mathbb{N}}$ follows from the coercivity of $\mathcal{E}$, i.e. there exist constants $\tilde{\gamma} > 0$, $\tilde{\beta} \in \mathbb{R}$ such that

$$\tilde{\gamma}\|u_k\|_U^p \leq \mathcal{E}(u_k, g_k) + \tilde{\beta} \leq C_g\|\overline{u}\| + \mathcal{E}(\overline{u}, 0) + \tilde{\beta}$$

Again reflexivity implies the existence of a subsequence $u_k \rightharpoonup \tilde{u}$ in U.
Admissibility of $(\tilde{u}, \tilde{g})$:
Now we can apply Lemma 3.3 to get

$$\lim_{k\to\infty} \mathcal{E}(u_k, g_k) = \mathcal{E}(\tilde{u}, \tilde{g}) = \min_{v\in U} \mathcal{E}(v, \tilde{g}).$$

Thus the pair $(\tilde{u}, \tilde{g})$ is an admissible candidate for a minimizer of J and a weak limit of the minimizing sequence (u_k, g_k) of J. As J is weakly lower semicontinuous $(\tilde{u}, \tilde{g})$ indeed minimizes J. Moreover the coercivity inequality (13) in combination with Lemma 3.2 guarantees that

$$\mathrm{adj}(I + \nabla\tilde{u}) \in \mathbf{L}^q(\Omega) \quad \text{and} \quad \det(I + \nabla\tilde{u}) \in L^r(\Omega)$$

and condition (12) assures that $\det(I + \nabla\tilde{u}) > 0$ a.e. in Ω. □

Remark 3.3.

1. *The argumentation in* [11, Thm. 7.9-1] *shows that the incorporation of the additional restriction* $\int_\Omega \det(\nabla\Phi)\,\mathrm{d}x \leq \mathrm{vol}(\Phi(\Omega))$ *allows to prove the weak injectivity condition* $\mathrm{card}\left(\Phi^{-1}(x)\right) = 1$ *for almost all* $x \in \Phi(\bar{\Omega})$ *if* $p > 3$.

2. *In general the admissible set* U *is not weakly closed. Here this is compensated by the coerciveness of* $\mathcal{E}$ *in combination with Lemma 3.2.*

4 Weak formulation

In the following we discuss weak formulations, corresponding to the energy minimization problem $\min_{u\in U} \mathcal{E}(u, g)$. This means that we derive first order necessary optimality conditions for the constraint of the optimal control problem under consideration. For sake of clarity we will from now on suppress the dependence on

x, i.e. we will assume that the material under consideration is homogeneous. The derived results also hold for heterogeneous materials.

As noted in the introduction it is, in general, not clear whether a local minimizer of the elastic energy functional satisfies the weak formulation (see [7, Problems 5 & 6])

$$\mathcal{E}'(u,g)h = 0 \quad \forall h \in C^\infty(\Omega)$$

In the context of compressible material laws the main difficulties are caused by condition (11). While being necessary in order to avoid local self-penetration and to model the observed material behaviour in a qualitatively correct way the introduced singularity leads to severe analytical difficulties.

In particular, it implies for the strain energy that

$$\mathcal{E}^S(u) = \int_\Omega W(u)\,\mathrm{d}x = \infty$$

on a dense subset of $W^{1,p}(\Omega)$ for any $p < \infty$ and thus also on a dense subset of U, i.e. for every $u \in U$ with $\mathcal{E}(u,g) < \infty$ one can construct a sequence $u_k \to u$ in U such that

$$\mathcal{E}(u_k, g) = \infty, \quad \forall k \in \mathbb{N}$$

Thus, we cannot expect differentiability in spaces weaker than $W^{1,\infty}(\Omega)$.
To make this discussion concrete, in the following we consider a compressible Mooney-Rivlin material law. This widely used constitutive relation is a special case of a compressible Ogden-type material. It is polyconvex, isotropic and may be written in terms of the left Cauchy-Green strain tensor E and the deformation gradient $\nabla\Phi = I + \nabla u$:

$$\tilde{W}(u) = a\,\mathrm{tr}(E) + b\,(\mathrm{tr}(E))^2 + c\,\mathrm{tr}(E^2) + \Gamma\left(\det(I + \nabla u)\right)$$

where

$$E = \frac{1}{2}\left(\nabla u^T + \nabla u + \nabla u^T \nabla u\right)$$

and $\lim_{s\to 0^+} \Gamma(s) = \infty$. Setting $\alpha = a - 2b,\ \beta = -c$, $\tilde{W}$ can be represented in the following way

$$\tilde{W}(u) = W(\nabla\Phi) = \frac{\alpha}{2}\|\nabla\Phi\|^2 + \frac{\beta}{2}\|\,\mathrm{adj}\,\nabla\Phi\|^2 + \Gamma\left(\det(\nabla\Phi)\right) + \mathrm{const.}$$

Popular choices for Γ take the forms (see [25, 26])

$$\Gamma(t) = \frac{1}{2}e_1 t^2 - e_2 \ln(t) \text{ or } \Gamma(t) = \frac{1}{2}e_1 t^2 + \frac{e_2}{k}t^{-k},\ k > 0 \tag{19}$$

In both cases the first summand t^2 guarantees, with $\alpha > 0, \beta > 0, e_1, e_2 > 0$, the validity of the coerciveness inequality (13) with $p = q = r = 2$. Moreover, for small strain, the material behaves like a St.Venant-Kirchhoff material. Thus, near $E = 0$ the stored energy function W should be a second order approximation

of the stored energy function of a St.Venant-Kirchhoff material. In the case of $\Gamma(t) = e_1 t^2 - e_2 \ln(t)$ it is always possible to determine $\alpha > 0, \beta > 0, e_1 > 0, e_2 > 0$ such that this is the case [11, Thm. 4.10-2]. This property comes at the expense of the model's quality, restricting its validity to rather academic questions. Thus we will focus on a non-logarithmic form as proposed in [25]. In this case the choice of parameters is dependent on the Poisson ratio $\nu = \frac{\lambda}{2(\lambda+\mu)}$. More precisely a lengthy computation shows that the following inequality

$$k < -1 + \frac{1}{1-2\nu} \text{ or equivalently } \nu > \frac{k}{2(k+1)} \tag{20}$$

restricts the possible range for k for given ν and vice versa, i.e. $k \geq 9$ requires $\nu > 0.45$, thus possibly implying the risk of constitutive locking (Poisson locking [8]). While being independent of Young's modulus, this inequality becomes less restrictive with growing ν.
With respect to the weak formulation we first focus on the energy minimization problem

$$\min_{u \in U} \mathcal{E}(u, g) \text{ for given, fixed } g \in G. \tag{21}$$

In the following we study the derivatives of $\mathcal{E}$ with respect to u, starting with a pointwise computation of the derivatives of W at non-singular F in direction δF:

$$W'(F)\delta F = \alpha F : \delta F + \beta \operatorname{adj} F : \operatorname{adj}'(F)\delta F + \Gamma'(\det F)\,(\operatorname{adj} F : \delta F)\,. \tag{22}$$

Here we used the differentiation rule $\det'(F)\delta F = \operatorname{adj} F : \delta F$. Further, we may also compute the second derivative:

$$\begin{aligned} &W''(F)(\delta F_1, \delta F_2) = \\ &\quad \alpha \delta F_1 : \delta F_2 + \beta \operatorname{adj}'(F)\delta F_1 : \operatorname{adj}'(F)\delta F_2 + \beta \operatorname{adj} F : \operatorname{adj}''(F)(\delta F_1, \delta F_2) \\ &\quad + \Gamma'(\det F)\left(\operatorname{adj}'(F)\delta F_1 : \delta F_2\right) + \Gamma''(\det F)\,(\operatorname{adj} F : \delta F_1)(\operatorname{adj} F : \delta F_2) \end{aligned} \tag{23}$$

The validity of the above pointwise formulae follows, for $F \in \mathbb{M}^3_+$, $\delta F_1, \delta F_2 \in \mathbb{M}^3$ directly from the definitions of det, adj and Γ. Having stated differentiability properties of W as a nonlinear function of the matrix $F \in \mathbb{M}^3$, we now turn to its study as superposition operators.

To this end, we consider the space $\mathbf{L}^p(\Omega)$ of p-integrable matrix valued functions $F : \Omega \to \mathbb{M}^3$, insert the matrix valued function $F \in \mathbf{L}^p(\Omega)$ pointwise into W and consider the result in another L^p-space. For this purpose we first need some properties of adj, Γ and an additional assumption on local minimizers of the energy functional $\mathcal{E}$.

Lemma 4.1. *Let $F \in \mathbf{L}^p(\Omega)$. Then the mapping*

$$\operatorname{adj}'(F) : \ \mathbf{L}^{p'}(\Omega) \to \mathbf{L}^1(\Omega) \tag{24}$$

is linear and continuous for $p^{-1} + (p')^{-1} \leq 1$. Moreover, the mapping

$$\operatorname{adj}''(F) : \ \mathbf{L}^{s_1}(\Omega) \times \mathbf{L}^{s_2}(\Omega) \to \mathbf{L}^1(\Omega) \tag{25}$$

is independent of F and bilinear and continuous for $s_1^{-1} + s_2^{-1} \leq 1$.
For $N > 2$ we have $\mathrm{adj}^{(N)} = 0$.

Proof. The assertion follows from the observation that adj is a second order polynomial in the entries of F and from Hölder's inequality. □

Definition 4.1. *Let $\Phi \in \mathbf{W}^{1,p}(\Omega)$ with $p \geq 1$. We call Φ* ***non-degenerate*** *if there exists a constant $\epsilon > 0$ such that*

$$\det(\nabla \Phi) \geq \epsilon \quad \text{a.e. in } \Omega. \tag{26}$$

Remark 4.1.

- *In the context of elasticity theory we will also call the displacement $u \in U$ non-degenerate if $\Phi = I + u$ is non-degenerate.*
- *Suppose there exists a local minimizer $u \in U$ of $\mathcal{E}_g$ that is degenerate , i.e. there exists a sequence*

  $$(x_k)_{k\in\mathbb{N}} \subset \Omega,\ x_k \to x \in \Omega \quad \text{such that} \quad \det(I + \nabla u(x_k)) \to 0$$

  Physically this corresponds to a deformation that becomes singular at $x \in \Omega$, thus being only reasonable in the modeling of cutting or piercing processes. In this cases other effects, like plasticity become dominant. In the context of applications like implant shape design the elastic behaviour is predominant, justifying the non-degeneracy assumption on minimizers of $\mathcal{E}$.
- *In the similar framework of barrier regularizations of optimal control problems examples can be given, where the violation of an analogue to non-degeneracy in the above sense yields minimizers that do not satisfy the formal optimality conditions [30].*

Lemma 4.2. *Assume that $F \in \mathbf{L}^p(\Omega)$ is non-degenerate,* $\mathrm{adj}\, F \in \mathbf{L}^q(\Omega)$*, and* $\det F \in L^r(\Omega)$*. Assume that the integrability indices $s_i \in [1, \infty], i = 1, \cdots, N$ satisfy*

$$\begin{aligned}
N = 1: &\qquad s_1^{-1} \leq 1 - (r^{-1} + q^{-1}) \\
N = 2: &\qquad s_1^{-1} + s_2^{-1} \leq 1 - \max(r^{-1} + p^{-1}, 2q^{-1}) \\
N = 3: &\qquad s_1^{-1} + s_2^{-1} + s_3^{-1} \leq 1 - \max(r^{-1}, p^{-1} + q^{-1}, 3q^{-1}),
\end{aligned}$$

which is only possible if the expressions on the corresponding right hand sides are non-negative.

Then, for the choice

$$\delta F_i \in \mathbf{L}^{s_i}(\Omega),\ s_i \in [1, \infty],\ i = 1, \cdots, N$$

we obtain

$$\frac{d^N}{dF^N} \Gamma(\det F)(\delta F_1, \ldots, \delta F_N) \in L^1(\Omega) \quad N = 1, 2, 3.$$

Proof. Differentiating Γ from (19) we get

$$\Gamma'(t) = e_1 t - e_2 t^{-(k+1)} \quad \Gamma''(t) = e_1 + e_2(k+1)t^{-(k+2)}$$
$$\Gamma'''(t) = -e_2(k+1)(k+2)t^{-(k+3)}.$$

Thus under our assumption of *non-degeneracy* follows that $\frac{d}{dF}\Gamma(\det F)$ grows linearly in $\det F$ and $\frac{d^2}{dF^2}\Gamma(\det F)$ is bounded independent of $\det F$. Then, using Hölder's inequality, inspection of the relevant terms in (22) and (23) yields our results for $N = 1$ and $N = 2$. For $N = 3$, we compute

$$\begin{aligned}\frac{d^3}{dF^3}\Gamma(\det F)(\delta F_1, \delta F_2, \delta F_3) &= \Gamma'(\det F)(\mathrm{adj}''(F)(\delta F_1, \delta F_2) : \delta F_3) \\ &\quad + 3\Gamma''(\det F)(\mathrm{adj}'(F)\delta F_1 : \delta F_2)(\mathrm{adj}\, F : \delta F_3) \\ &\quad + \Gamma'''(\det F)(\mathrm{adj}\, F : \delta F_1)(\mathrm{adj}\, F : \delta F_2)(\mathrm{adj}\, F : \delta F_3)\end{aligned}$$

and use again Hölder's inequality. □

Now we can turn to the study of the derivatives of W.

Proposition 4.3. *Assume that $F \in \mathbf{L}^p(\Omega)$ is non-degenerate, $\mathrm{adj}\, F \in \mathbf{L}^q(\Omega)$, and $\det F \in L^r(\Omega)$. (In the following, we take $s_i \in [1, \infty]$, and assume that the inequalities for the s_i are non void.)*

If $0 \le s_1^{-1} \le 1 - (q^{-1} + \max(r^{-1}, p^{-1}))$ then

$$W'(F)\delta F \in L^1(\Omega) \quad \textit{for all } \delta F \in \mathbf{L}^{s_1}(\Omega)$$

and $W'(F)$ is linear and continuous in δF.

If $0 \le s_1^{-1} + s_2^{-1} \le 1 - \max(2p^{-1}, r^{-1} + p^{-1}, 2q^{-1})$ then

$$W''(F)(\delta F_1, \delta F_2) \in L^1(\Omega) \quad \textit{for all } \delta F_i \in \mathbf{L}^{s_i}(\Omega),\ i = 1, 2$$

and $W''(F)$ is bilinear and continuous in $(\delta F_1, \delta F_2)$.

If $0 \le s_1^{-1} + s_2^{-1} + s_3^{-1} = 1 - \max(r^{-1}, p^{-1} + q^{-1}, 3q^{-1})$ then

$$W'''(F)(\delta F_1, \delta F_2, \delta F_3) \in L^1(\Omega) \quad \textit{for all } \delta F_i \in \mathbf{L}^{s_i}(\Omega),\ i = 1, 2, 3$$

and $W'''(F)$ is trilinear and continuous in $(\delta F_1, \delta F_2, \delta F_3)$.

Proof. The assertion follows from inspection of the particular terms. In (22) for W' and in (23) for W''. For W''' a similar term can be computed. Well definedness of the derivatives of Γ in suitable L_p spaces has been shown in Lemma 4.2, the remaining terms are second and fourth order polynomials in the coefficients of F. With this information, our result follows from repeated application of the Hölder inequality. □

Finally we can study conditions under which the formal directional derivatives of the strain energy

$$\mathcal{E}_u^S(u)v_1 := \int_\Omega W'(I+\nabla u)\nabla v_1 \,\mathrm{d}x \tag{27}$$

$$\mathcal{E}_{uu}^S(u)(v_1,v_2) := \int_\Omega W''(I+\nabla u)(\nabla v_1,\nabla v_2) \,\mathrm{d}x \tag{28}$$

$$\mathcal{E}_{uuu}^S(u)(v_1,v_2,v_3) := \int_\Omega W'''(I+\nabla u)(\nabla v_1,\nabla v_2,\nabla v_3) \,\mathrm{d}x \tag{29}$$

are well-defined. Moreover we have to verify if the remainder terms vanish, i.e. under which conditions the defined functionals really are the directional derivatives of the strain energy. For given $u \in W^{1,2}(\Omega)$ this is a delicate issue. Fortunately, the coerciveness inequality (13) implies that $\mathrm{adj}(I+\nabla u) \in \mathbf{L}^2(\Omega)$ and $\det(I+\nabla u) \in L^2(\Omega)$ if u is a minimizer of $\mathcal{E}$. Therefore

Corollary 4.4. *Assume that $u \in U$ is non-degenerate and $\mathcal{E}^S(u)$ is finite. Then $\mathcal{E}_u^S(u)$ and $\mathcal{E}_{uu}^S(u)$ are well defined in $W^{1,\infty}(\Omega)$, resp. $W^{1,\infty}(\Omega)\times W^{1,\infty}(\Omega)$.*
If further $u \in W^{1,\infty}(\Omega)$, then (27), (28), and (29) are well defined for $v_i \in W^{1,s_i}(\Omega)$ with $\sum s_i^{-1} = 1$, respectively.

Proof. By coercivity of $\mathcal{E}^S$ we conclude $p=2$, $q=2$, $r=2$. Thus, we can apply Proposition 4.3 for $s_i = \infty$ to obtain our first result for (27), (28)

Since adj and det are polynomials, it follows in the case of $u \in W^{1,\infty}$ and non-degeneracy that $p=q=r=\infty$, such that $\sum s_i^{-1} = 1$ can be chosen. □

Proposition 4.5. *If $u \in U$ is non-degenerate, then $\mathcal{E}^S$ is directionally differentiable for each $\delta u \in W^{1,\infty}(\Omega)$ with derivative given by (27). The corresponding remainder term is uniform in δu.*

If in addition $u \in W^{1,\infty}(\Omega)$, then $\mathcal{E}^S$ is twice directionally differentiable with second derivative given by (28). For sufficiently small $\|\delta u\|_{W^{1,\infty}}$ the corresponding remainder term can be estimated by

$$r_2(u,\delta u) \le c\|\delta u\|_{W^{1,\infty}}\|\delta u\|_{W^{1,2}}^2.$$

Proof. In order to prove the statement we consider for $\delta u \in \mathbf{W}^{1,\infty}(\Omega)$ the remainder term

$$\begin{aligned}|\mathcal{E}(u+\delta u,g) - \mathcal{E}(u,g) - \mathcal{E}_u(u,g)\delta u| &= \frac{1}{2}\left|\mathcal{E}_{uu}(u+\xi\delta u,g)(\delta u)^2\right| \\ &\le \frac{1}{2}\|\,\mathcal{E}_{uu}(u+\xi\delta u,g)\|\|\delta u\|_{W^{1,\infty}}^2\end{aligned}$$

By Corollary 4.4 we know that $\mathcal{E}_{uu}(u,g)(\delta u)^2$ is finite, and since $\mathcal{E}_{uu}$ is continuous at u in $W^{1,\infty}(\Omega)$, $\mathcal{E}_{uu}(u+\xi\delta u,g)(\delta u)^2$ is bounded.

The proof for the second derivative runs analogously, using the properties of W'''. □

Combination of these results allows us to prove the main theorem of this section:

Theorem 4.6. *Let $u \in U$ be a non-degenerate local minimizer of $\mathcal{E}$ with $\mathcal{E}(u) < \infty$. Then it satisfies the following weak formulation*

$$\mathcal{E}_u(u, g)\delta u = 0 \quad \text{for all } \delta u \in \mathbf{W}^{1,\infty}(\Omega). \tag{30}$$

If, in turn, $u \in W^{1,\infty}$ satisfies (30), and $\mathcal{E}_{uu}(u, g)v^2 \geq \delta \|v\|^2_{W^{1,2}}$ for all $v \in W^{1,\infty}$, then for sufficiently small $\delta u \in W^{1,\infty}(\Omega)$ and some $\varepsilon > 0$ we have the growth condition

$$\mathcal{E}(u + \delta u) \geq \mathcal{E}(u) + \varepsilon \|\delta u\|^2_{W^{1,2}}.$$

In particular, u is a $W^{1,\infty}$-local minimizer of $\mathcal{E}$.

Proof. The proof is standard: to show that $\mathcal{E}_u(u, g)\delta u = 0$, we compute

$$\mathcal{E}_u(u, g)(\pm\delta u) = \lim_{t\to 0} \frac{\mathcal{E}(u \pm t\delta u, g) - \mathcal{E}(u, g)}{t} \geq 0,$$

since u is a local minimizer of $\mathcal{E}$.

For our second assertion, we note that

$$\begin{aligned} \mathcal{E}(u + \delta u) - \mathcal{E}(u) &= \frac{1}{2}\,\mathcal{E}_{uu}(u, g)\delta u^2 + r(\delta u) \\ &\geq \frac{\delta}{2}\|\delta u\|^2_{W^{1,2}} + r(\delta u). \end{aligned}$$

Due to proposition 4.5

$$r(u, \delta u) \leq c\|\delta u\|_{W^{1,\infty}} \|\delta u\|^2_{W^{1,2}},$$

so that, for $\|\delta u\|_{W^{1,\infty}} \to 0$ we obtain

$$\mathcal{E}(u + \delta u) - \mathcal{E}(u) \geq \left(\frac{\delta}{2} - c\|\delta u\|_{W^{1,\infty}}\right) \|\delta u\|^2_{W^{1,2}} \geq \varepsilon \|\delta u\|^2_{W^{1,2}}. \qquad \square$$

5 Formal first order optimality conditions

Next we discuss first order optimality conditions of our optimal control problem. As we have seen above, differentiability of the equality constraints $\mathcal{E}_u(u, g) = 0$ requires the choice of $W^{1,\infty}(\Omega)$ (or stronger) as a topological framework. Thus we have to restrict our discussion to a formal level, as on one side we lack an existence result in this space, and on the other hand existing regularity results do not admit the application of the implicit function theorem in order to show that the set $\mathcal{E}_u(u, g) = 0$ is a smooth manifold. Its application requires continuous invertibility of the linearized weak formulation in suitable spaces. One possible framework would be to consider

$$\mathcal{E}_{uu} : W^{2,p} \to L_p$$

for $W^{2,p} \hookrightarrow W^{1,\infty}$ (cf. e.g. [11, Chapter 6]). However, the class of problems for which suitable regularity results hold is small.

Formally the first order optimality conditions of (16) can be derived via the Lagrangian function

$$L(u, g, p) = J(u, g) + p(\mathcal{E}_u(u, g)).$$

Computing the formal derivatives of L with respect to u, g, and p yields the system

$$J_u(u, g) + \mathcal{E}_{uu}(u, g)^* p = 0 \quad \text{in } U^* \tag{31a}$$
$$J_g(u, g) + \mathcal{E}_{ug}(u, g)^* p = 0 \quad \text{in } G^* \tag{31b}$$
$$\mathcal{E}_u(u, g) = 0 \quad \text{in } U^* \tag{31c}$$

If J is the sum of a measure of the error and a Tikhonov regularization term, i.e. if J is of the form $J(u, g) = J^{\text{err}}(u) + \frac{\alpha}{2}\|g\|^2_{L^2(\Gamma_c)}$, where α is the Tikhonov regularization parameter (as in (9a)), then these conditions can be written down explicitly:

$$J^{\text{err}}_u(u) + \mathcal{E}_{uu}(u, g)^* p = 0 \quad \text{in } U^* \tag{32a}$$
$$\alpha g(x) + \left(\text{adj}(I + \nabla u)^T n\right) p(x) = 0 \quad \text{a.e. on } \Gamma_c \tag{32b}$$
$$\mathcal{E}_u(u, g) = 0 \quad \text{in } U^* \tag{32c}$$

Elimination of g via (32b) reduces system (32) to

$$J^{\text{err}}_u(u) + \mathcal{E}_{uu}(u)^* p = 0 \tag{33a}$$
$$\mathcal{E}_u\left(u, -\frac{\left(\text{adj}(I + \nabla u)^T n\right) p}{\alpha}\right) = 0 \tag{33b}$$

6 Numerical Results

In order to perform first numerical experiments we consider the cost functional

$$J(u, g) = \frac{\beta}{2}\|u - u_{\text{ref}}\|^2_{\mathbf{L}^2(\Gamma_t)} + \frac{\alpha}{2}\|g\|^2_{L^2(\Gamma_c)}, \tag{34}$$

where the additional parameter $\beta \in]0, 1]$ is introduced in order to establish a numerical continuation scheme $\beta \to 1$. In a direct approach the occurring nonlinearities would lead to too small Newton steps in the solution of nonlinear problem for $\beta = 1$.

Then, setting $\tilde{n} = \text{adj}(I + \nabla u)^T n$ the reduced optimality system reads

$$\int_\Omega W''(\nabla u)\nabla p \nabla v \,\mathrm{d}x + \int_{\Gamma_t} \beta(u - u_{\text{ref}})v \,\mathrm{d}s = 0 \quad \forall v \in U \tag{35a}$$
$$\int_\Omega W'(\nabla u)\nabla w \,\mathrm{d}x + \int_{\Gamma_c} \frac{\tilde{n}p}{\alpha}\tilde{n}w \,\mathrm{d}s = 0 \quad \forall w \in U \tag{35b}$$

Further, in view of possibly large values for Young's modulus $\mathbb{E}$ we perform a rescaling of the problem via

$$W \mapsto \mathbb{E}^{-1} W \quad \text{and} \quad \alpha \mapsto \mathbb{E}^2 \alpha. \tag{36}$$

This is a problem formulation that is invariant with respect to Young's modulus, being of advantage as, in the presence of large $\mathbb{E}$, appropriate Tikhonov parameters satisfy $\alpha \sim \mathbb{E}^{-2}$ (see [21]) and thus may become very small. This in turn affects the condition number of the Newton matrix and thus the numerical accuracy of the Newton steps. As the coefficients of W depend linearly on Young's modulus, the application of the transformations (36) is equivalent to setting $\mathbb{E} = 1$.

In summary, we solve a sequence of problems

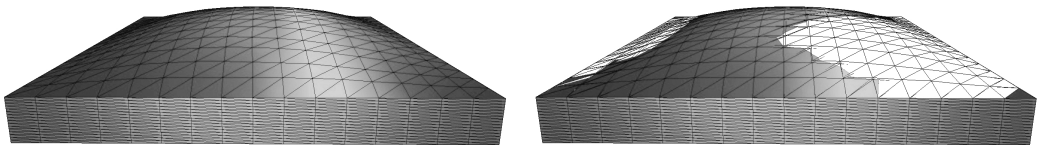
$$(P_k) \qquad \begin{cases} \displaystyle\int_\Omega W''(\nabla u)\nabla p \nabla v \,\mathrm{d}x + \int_{\Gamma_t} \beta_k (u - u_{\mathrm{ref}}) v \,\mathrm{d}x = 0 \\ \displaystyle\int_\Omega W'(\nabla u) h \,\mathrm{d}x + \int_{\Gamma_c} \frac{\tilde{n} p}{\alpha} \tilde{n} h \,\mathrm{d}s = 0 \end{cases} \tag{37}$$

with $0 < \beta_0 < \cdots < \beta_N = 1$, $N > 0$, $\mathbb{E} = 1$. The second material parameter of linearized elasticity, the Poisson ratio ν, is close to $\frac{1}{2}$ in order to correspond to a quasi-incompressible material, as encountered in soft tissue models. As constitutive locking is a commonly observed phenomenon for $\nu \to \frac{1}{2}$ [4, 8], that should be excluded in order to monitor the influences of the nonlinearities, we set $\nu = 0.45$. This choice keeps the risk of constitutive locking small while staying reasonable from a modelling point of view. In general, in order to allow the Poisson ratio to attain all values in the admissible range $[0, 0.5[$ mixed formulations and/or adjusted discretization schemes for the forward problem of elastostatics [3, 31, 32] must be used and adapted to the optimal control problem.
As noted in Section 4 a logarithmic dependence of Γ on the volume change is not sufficient to accurately model the soft tissue's behaviour. For sake of numerical simplicity we nevertheless choose the logarithmic penalty term $\Gamma(s) = s^2 + \log(s)$.

The systems (P_k) have been discretized on the cuboid $[-1, 1]\times[-1, 1]\times[-0.1, 0.1]$ with the finite element toolbox Kaskade7.1 [16] using linear elements. The resulting finite dimensional, nonlinear equations are solved with a covariant damped Newton-method as presented in [14, Chapter 3]. For the solution of the arising linear systems of equations we use the distributed multifrontal solver MUMPS [1, 2].

7 Conclusion

In this work, basic analytical and numerical results for the mathematical treatment of an implant design problem have been established. The design problem was formulated as an optimal control problem, and existence of optimal solutions was shown in the context of polyconvex hyperelastic materials. Optimality conditions were derived on a formal level and first numerical results were computed.



(a) computed deformation

(b) comparison between computed solution and desired deformation (white)

Figure 2: first numerical results for $\alpha = 0.1$

These results indicate that our problem is numerically very challenging, and refined algorithmic ideas are necessary to treat the nonlinear shape implant problem to full satisfaction. This includes on one hand globalization techniques for the nonlinear solver, and on the other hand adaptivity and iterative solution techniques for the linear systems.

References

[1] P.R. Amestoy, I.S. Duff, J. Koster, and J.-Y. L'Excellent. A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM J. Mat. Anal. and Appl.*, 23(1):15–41, 2001.

[2] P.R. Amestoy, A. Guermouche, J.-Y. L'Excellent, and S. Pralet. Hybrid scheduling for the parallel solution of linear systems. *Par. Comp.*, 32(2):136–156, 2006.

[3] O. Axelsson and A. Padiy. On a robust and scalable linear elasticity solver based on a saddle point formulation. *Int. J. Numer. Meth. Engn.*, 44:801–818, 1999.

[4] I. Babuška and M. Suri. Locking effects in the finite element approximation of elasticity problems. *Numer. Math.*, 62:439–463, 1992.

[5] J. M. Ball. Convexity conditions and existence theorems in nonlinear elasticity. *Arch. Rational Mesh. Anal.*, 63:337–403, 1977.

[6] J. M. Ball. Strict convexity, strong ellipticity, and regularity in the calculus of variations. *Math. Proc. Camb. Phil. Soc.*, 87:501–513, 1980.

[7] J. M. Ball. *Geometry, Mechanics and Dynamics*, chapter "Some open problems in elasticity", pages 3–59. Springer, 2002.

[8] D. Braess. *Finite Elemente: Theorie, schnelle Löser und Anwendungen in der Elastizitätstheorie.* Springer, 4^{th} edition, 1992.

[9] H. Bufler. Pressure loaded structures under large deformations. *Z. Angew. Math. u. Mech.*, 64(7):287–295, 1984.

[10] M. Chabanas, V. Luboz, and Y. Payan. Patient specific finite element model of the face soft tissues for computer-assisted maxillofacial surgery. *Medical Image Analysis*, 7(2):131–151, 2003.

[11] P. G. Ciarlet. *Mathematical Elasticity Vol. I: Three-dimensional Elasticity.* North-Holland, 1988.

[12] P.G. Ciarlet. *An Introduction to Differential Geometry with Applications to Elasticity.* Springer, 2005.

[13] B. Dacorogna. *Direct Methods in the Calculus of Variations.* Springer, 2^{nd} edition, 2008.

[14] P. Deuflhard. *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms.* Springer, 2004.

[15] Y. C. Fung. *Biomechanics: Mechanical Properties of Living Tissues.* Springer, 2^{nd} edition, 1993.

[16] S. Götschel, A. Schiela, and M. Weiser. Solving optimal control problems with the Kaskade 7 finite element toolbox. *ZIB-Report*, 10(25):1–13, 2010.

[17] G. A. Holzapfel. *Handbook of Material Behavior: Nonlinear Models and Properties*, chapter Biomechanics of Soft Tissue. Academic Press, 2001.

[18] G. A. Holzapfel and R. W. Ogden. *Mechanics of biological tissue.* Springer, 2006.

[19] J. K. Knowles and Eli Sternberg. On the failure of ellipticity of the equations for finite elastostatic plane strain. *Archive for Rational Mechanics and Analysis*, 63:321–336, 1976. 10.1007/BF00279991.

[20] S. Lee, R.E. Caflisch, and Y.-J. Lee. Exact artificial boundary conditions for continuum and discrete elasticity. *SIAM J. Appl. Math.*, 66(5):1749–1775, 2006.

[21] L. Lubkoll. Optimal control in implant shape design. Master's thesis, ZIB, TU Berlin, 2010.

[22] J. E. Marsden and J. R. Hughes. *Mathematical Foundations of Elasticity.* Prentice-Hall, 1983.

[23] M. Mooney. A theory of large elastic deformation. *Journ. App. Phys.*, 11(9):582–592, 1940.

[24] C. B. Morrey. *Multiple Integrals in the Calculus of Variations.* Springer, 2008 (reprint) edition, 1966.

[25] F. D. Murnaghan. *Finite deformation of an elastic solid.* John Wiley and Sons, 1951.

[26] R. W. Ogden. Large deformation isotropic elasticity: on the correlation of theory and experiment for compressible rubber-like solids. *Proc. Roy. Soc. London*, A(328):567–583, 1972.

[27] R. W. Ogden. Large deformation isotropic elasticity: on the correlation of theory and experiment for incompressible rubber-like solids. *Proc. Roy. Soc. London*, A(326):565–583, 1972.

[28] R. W. Ogden. *Non-linear elastic deformations.* Dover Publ., 1997.

[29] R. S. Rivlin. Large elastic deformations of isotropic materials. iv. further developments of the general theory. *Phil. Trans. Roy. Soc. London*, 241(835):379–397, 1948.

[30] A. Schiela. Barrier methods for optimal control problems with state constraints. *SIAM J. on Optimization*, 20(2):1002–1031, 2009.

[31] K. Shavan, B. P. Lamichhane, and Wohlmuth B. Locking-free finite element methods for linear and nonlinear elasticity in 2d and 3d. *Comp. Meth. App. Mech. Engn.*, 196(41-44):4075–4086, 2007.

[32] M. Vogelius. An analysis of the p-version of the finite element method for nearly incompressible materials. *Numer. Math.*, 41:39–53, 1983.

[33] V. Šverák. Rank-one convexity does not imply quasiconvexity. *Proc. Roy. Soc. Edinburgh*, A(120):185–189, 1992.

[34] M. Weiser, Deuflhard P., and B. Erdmann. Affine conjugate adaptive newton methods for nonlinear elastomechanics. *Opt. Meth. Softw.*, 22(3):414–431, 2007.