manuscript No.
(will be inserted by the editor)

Moreau-Yosida Regularization in Shape Optimization with Geometric Constraints

Moritz Keuthen · **Michael Ulbrich**

Received: date / Accepted: date

Abstract In the context of shape optimization with geometric constraints we employ the method of mappings (perturbation of identity) to obtain an optimal control problem with a nonlinear state equation on a fixed reference domain. The Lagrange multiplier associated with the geometric shape constraint has a low regularity (similar to state constrained problems), which we circumvent by penalization and a continuation scheme. We employ a Moreau-Yosida-type regularization and assume a second-order condition to hold. The regularized problems can then be solved with a semismooth Newton method and we study the properties of the regularized solutions and the rate of convergence towards a solution of the original problem. A model for the value function in the spirit of [17] is introduced and used in an update strategy for the regularization parameter. The theoretical findings are supported by numerical tests.

Keywords Shape Optimization · Moreau-Yosida Regularization · Method of Mappings · Semismooth Newton · Geometric Constraints

1 Introduction

Shape optimization is an very active research area with many important application, in particular in engineering. We refer to the monographs [43, 16, 9, 31] for an introduction to shape optimization and an overview of the various possible approaches. In this paper we will employ the method of mappings (also called perturbation of identity) which is applicable if the approximate optimal shape and in particular its topology

We gratefully acknowledge support from the International Research Training Group IGDK1754, funded by the German Science Foundation (DFG).

Moritz Keuthen · Michael Ulbrich
Chair of Mathematical Optimization, Department of Mathematics, Technische Universität München, Boltzmannstr. 3, 85748 Garching b. München, Germany
Tel.: +49-89-28917966
Fax: +49-89-28917932
E-mail: keuthen@ma.tum.de

is known a-priori. In many engineering applications this is the case. The method of mappings approach to shape optimization was originally proposed by Murat and Simon in [33] and [32] and has been successfully applied to various applications. See for example [4,5,3,27] for shape optimization with instationary Navier-Stokes flow. The idea is to describe varying domains $\Omega \subset \mathbb{R}^d$ as transformations $\tau = \mathrm{id} + V$ of a fixed reference domain Ω_{ref}, where $V : \mathbb{R}^d \to \mathbb{R}^d$ denotes the domain displacement. Since the state and objective should only depend on the shape of the domains Ω we are mainly interested in the transformation of the boundary $\partial\Omega_{ref}$. Let $\Gamma_B \subset \partial\Omega_{ref}$ denote the unknown part of the boundary, which is to be optimized, and $\Gamma_F \subset \partial\Omega_{ref}$ the fixed part of the boundary. We require $\Omega_{ref} \subset \mathbb{R}^d$ to be a Lipschitz domain and Γ_B to be a C^1-manifold. The displacement of the free boundary

$$u : \Gamma_B \to \mathbb{R}^d, u \in \mathscr{U},$$

acts as our control. Here $\mathscr{U}$ denotes some suitable space of displacements. For a given displacement $u \in \mathscr{U}$ we obtain the domain displacement V as the image of an extension operator $T : \mathscr{U} \to \mathbf{W}^{1,\infty}_{ref}$.

In this paper we will study the boundary shape optimization problem

$$\begin{aligned} \min_{u\in\mathscr{U}_{ad}, y\in Y(\Omega_{ref})} \quad & J(T(u),y) + \frac{\beta}{2}\|u\|^2_{\mathscr{U}} \\ \text{s.t. } & E(T(u),y) = 0 \qquad \text{in } Z(\Omega_{ref}), \end{aligned} \tag{1}$$

where $\mathscr{U}_{ad} \subset \mathscr{U}$ denotes an admissible set of displacements, and J some objective functional depending on $\tau(\Omega_{ref})$ and the state y. Furthermore we have added the regularization term $\frac{\beta}{2}\|u\|^2_{\mathscr{U}}$ with $\beta \geq 0$. We are specifically interested in geometric shape constraints of the following form

$$\mathscr{U}_{ad} = \{u \in \mathscr{U} \mid x + u(x) \in C \; \forall x \in \Gamma_B\} = \{u \in \mathscr{U} \mid \tau(\Gamma_B) \subset C\},$$

where $C \subset \mathbb{R}^d$ is some closed convex set. Note that $\mathscr{U} \hookrightarrow L^2(\Gamma_B, \mathbb{R}^d)$. Such design constraints appear in many applications and have been considered in various publications. Usually they are either only considered with regard to some particular parametrization (e.g. constraints on the control points of some Bézier curve), or discretization, see for example [2,4,5,3,27,22,35], or they are tacitly assumed to be inactive in the solution see for example [26,24]. To the best of our knowledge there is so far no satisfactory algorithmic treatment of geometric shape constraints in a function space setting available. The difficulty here is that the Lagrange multiplier associated with the constraint is a-priori only an element of $\mathscr{U}^*$, which in our numerical example will be $H^2(\Gamma_B, \mathbb{R}^2)^*$. This is a similar setting as in the case of state constraints in optimal control problems, where the multiplier associated to the state constraint is in general only a measure. Basically three approaches have been proposed in the literature on state constraints to deal with the associated difficulties. Moreau–Yosida regularization based inexact primal-dual path-following techniques were first investigated in [23,18,17], Lavrentiev regularization methods were proposed in [44,29,39] and barrier methods were studied in [41,42]. The Lavrentiev regularization which relaxes the state constraints to mixed control and state constraints is not applicable here and the theory of barrier methods is only available for convex optimal control problems. Since our shape optimization problem contains a highly non-linear state equation we follow the approach taken in [17] by introducing a Moreau-Yosida type penalty term and studying the properties of the solutions to the associated subproblems. Facing a non-linear problem we will assume some second-order condition to hold. We will show local Lipschitz continuity of the regularized solutions and prove some convergence rate estimates similar to [20]. The subproblems can be solved efficiently by a semismooth Newton method [45]. For the computation of higher order derivatives in shape optimization we refer to [15], some examples of second order optimization methods employed in shape optimization are [11,19,37,36].

This paper is organized as follows. In section 2 we describe in more detail how we obtain the shape optimization problem (1) and discuss existence of a solution as well as differentiability of the reduced objective. Section 3 is devoted to the functional analytic treatment of geometric shape constraints. We introduce the regularized problems employing a Moreau-Yosida type penalty term in section 3.1 and discuss the convergence of regularized solutions towards a solution of the original problem in section 3.2. We will state second order conditions in section 3.3 and show superliner convergence of a semismooth Newton method for the regularized problems. We exploit the second-order condition in section 3.4 to show local Lipschitz continuity of the regularized solutions. The value function, which maps the regularization parameter to the optimal objective value of the associated problem, will be studied in section 3.5 and a model function which is used for inexact path-following as in [17] is proposed. In section 3.6 we use the optimality conditions to show convergence of the approximate Lagrange multipliers to the Lagrange multiplier associated with the geometric constraint. Convergence rate estimates are derived in section 3.7. Finally we study as an example the potential flow through a channel in section 4. We check that the assumption made in sections 2 and 3 are satisfied and present numerical results which support our theoretic findings.

In the following, if not stated otherwise, we will denote with $c > 0$ some generic constant, which may change value in the computations. For a space $X(\Omega_{ref})$ we will use the shorthand notation X_{ref} and for X^d we write $\mathbf{X}$.

2 Shape optimization problem

2.1 The method of mappings

In this section we want to briefly present the method of mappings and the requirements it necessitates. For a concise treatment of this approach we refer to [4,5,27]. Furthermore we discuss assumptions which guarantee the existence of a solution and differentiability of the design-to-state mapping.

A general shape optimization problem with a state equation can be posed as

$$\min_{\Omega \in \mathscr{O}_{ad}, \tilde{y} \in Y(\Omega)} \bar{J}(\Omega, \tilde{y}) \text{ s.t. } \bar{E}(\Omega, \tilde{y}) = 0, \tag{2}$$

where $\mathscr{O}_{ad} \subset \mathscr{P}(\mathbb{R}^d)$ denotes an admissible set of domains. $Y(\Omega), Z(\Omega)$ are Banach spaces for all $\Omega \in \mathscr{O}_{ad}$ and

$$\begin{aligned} \bar{J} &: \{(\Omega,\tilde{y}) \mid \Omega \in \mathscr{O}_{ad}, \tilde{y} \in Y(\Omega)\} \to \mathbb{R}, \\ \bar{E} &: \{(\Omega,\tilde{y}) \mid \Omega \in \mathscr{O}_{ad}, \tilde{y} \in Y(\Omega)\} \to \{\tilde{z} \mid \Omega \in \mathscr{O}_{ad}, \tilde{z} \in Z(\Omega)\}, \end{aligned}$$

with $\bar{E}(\Omega,\tilde{y}) \in Z(\Omega)\ \forall \Omega \in \mathscr{O}_{ad},\ \forall \tilde{y} \in Y(\Omega)$. Let us now fix a bounded reference domain $\Omega_{ref} \in \mathscr{O}_{ad}$ and let $\mathscr{V} = \mathscr{V}(\Omega_{ref})$ be a suitable space of domain displacements. We present our choice of $\mathscr{V}$ below. We suppose

Assumption 1 *There exists a set $\mathscr{V}_{ad} \subset \mathscr{V}$ such that*

$$\mathscr{O}_{ad} = \{\tau(\Omega_{ref}) \mid \tau = id + V,\ V \in \mathscr{V}_{ad}\}, \tag{3}$$

and $\forall V \in \mathscr{V}_{ad}$

$$\begin{cases} Y(\Omega_{ref}) = \{\tilde{y} \circ \tau \mid \tilde{y} \in Y(\tau(\Omega_{ref}))\}, \\ Y(\tau(\Omega_{ref})) \ni \tilde{y} \mapsto y := \tilde{y} \circ \tau \in Y(\Omega_{ref}) \text{ is a homeomorphism.} \\ Z(\Omega_{ref}) = \{\tilde{z} \circ \tau \mid \tilde{z} \in Z(\tau(\Omega_{ref}))\}, \\ Z(\tau(\Omega_{ref})) \ni \tilde{z} \mapsto z := \tilde{z} \circ \tau \in Z(\Omega_{ref}) \text{ is a homeomorphism.} \end{cases} \tag{4}$$

Using this assumption we obtain the equivalent optimization problem

$$\min_{V \in \mathscr{V}_{ad}, y \in Y(\Omega_{ref})} J(V,y) \text{ s.t. } E(V,y) = 0, \tag{5}$$

where $E : \mathscr{V} \times Y(\Omega_{ref}) \to Z(\Omega_{ref})$ such that

$$\forall V \in \mathscr{V}_{ad},\ \tilde{y} \in Y(\tau(\Omega_{ref})): \quad E(V, \tilde{y} \circ \tau) = 0 \Leftrightarrow \bar{E}(\tau(\Omega_{ref}), \tilde{y}) = 0,$$

and J analogue.

Remark: (i) The operators E and J are usually obtained using the transformation rule for integrals. In particular E is highly nonlinear with respect to V.
(ii) Assumption (3) restricts the class of shape optimization problems we are able to handle with the method of mappings. Assumption (4) can be guaranteed for certain spaces Y, Z by appropriate choices of $\mathscr{V}$ and $\mathscr{V}_{ad}$ as we will now recall.

Murat and Simon showed in [33, Lemma 4.1] that (4) holds for $L^p, W^{1,p}$ and $W_0^{1,p}$ if

$$\mathscr{V}_{ad} \subset \mathscr{V}_{feas} := \{V \in \mathbf{W}^{1,\infty}(\mathbb{R}^d) \mid \tau = \mathrm{id} + V : \overline{\Omega_{ref}} \mapsto \tau(\overline{\Omega_{ref}}) \text{ is bi-Lipschitz}\}. \tag{6}$$

In [33, Lemma 2.4] they proved that $\mathscr{V}_{feas}$ is open in $\mathbf{W}^{1,\infty}(\mathbb{R}^d)$ and that small enough displacements are contained in it. If

$$\|V\|_{\mathbf{W}^{1,\infty}(\mathbb{R}^d)} \leq \frac{1}{2\sqrt{d^4}} \tag{7}$$

then $V \in \mathscr{V}_{feas}$ and it holds $\|\tau^{-1} - \mathrm{id}\|_{\mathbf{W}^{1,\infty}(\mathbb{R}^d)} \leq c\|V\|_{\mathbf{W}^{1,\infty}(\mathbb{R}^d)}$ for some fixed $c > 0$.

As already mentioned in section 1 we consider the boundary displacement $u \in \mathscr{U}$ as our control. For a given displacement $u \in \mathscr{U}$ we obtain the domain displacement V as the image of an extension operator $T : \mathscr{U} \to \mathbf{W}^{1,\infty}_{ref}$. Usually this will be the solution operator of some linear elliptic equation $B(u,V) = 0$ with Dirichlet boundary conditions, $V = u$ on Γ_B, and $V = 0$ on Γ_F.

$$T : \mathscr{U} \to \mathbf{W}^{1,\infty}_{ref},\ T(u) = V,\ \text{ where } V \text{ solves } \ B(u,V) = 0.$$

This will be described for the linear elasticity equation in more detail in section 4.2. Note that for a C^1-Manifold we can define the surface measure via the introduction of local coordinates and a smooth partition of unity (compare e.g. [43, section 2.2]). We assume that it is possible to choose the admissible set of boundary displacements $\mathscr{U}_{ad} \subset \mathscr{U}$ such that

$$u \in \mathscr{U}_{ad} \Leftrightarrow T(u) \in \mathscr{V}_{ad}.$$

We arrive at the shape optimization problem in terms of the boundary displacement

$$\min_{u\in\mathscr{U}_{ad}, y\in Y(\Omega_{ref})} \hat{J}(u,y) := J(T(u),y) + \beta \|u\|^2_{\mathscr{U}} \text{ s.t. } E(T(u),y) = 0, \tag{8}$$

where we have added a Tichonov regularization term $\beta\|u\|^2_{\mathscr{U}}$ with $\beta \geq 0$. If the state equation $E(V,y) = 0$ admits a unique solution $y(V) \in Y(\Omega_{ref})$ for every $V \in \mathscr{V}_{ad}$, we introduce the design-to-state operator

$$S : \mathscr{V}_{ad} \to Y(\Omega_{ref}),\quad V \mapsto S(V) := y(V).$$

This leads us to the reduced objectives

$$\begin{aligned}
&\mathfrak{j} : \mathscr{V}_{ad} \to \mathbb{R},\ \mathfrak{j}(V) := J(V,S(V)),\\
&j : \mathscr{U}_{ad} \to \mathbb{R},\ j(u) := \hat{J}(u,S(T(u))) = \mathfrak{j}(T(u)) + \beta\|u\|^2_{\mathscr{U}}.
\end{aligned}$$

In this paper we will study for certain admissible sets $\mathscr{U}_{ad}$ the reduced shape optimization problem

$$\min_{u\in\mathscr{U}} j(u) = \hat{J}(u,S(T(u))) \text{ s.t. } u \in \mathscr{U}_{ad}. \tag{9}$$

2.2 Existence of a solution and differentiability

It is a standard result that (9) possesses a solution if $\mathscr{U}$ is a reflexive Banach space, $\emptyset \neq U_{ad} \subset \mathscr{U}$ is closed and convex and j is coercive (i.e. $j(u) \to \infty$ if $\|u\|_{\mathscr{U}} \to \infty$) and weakly lower semicontinuous. We now want to present a setting which is appropriate for the shape optimization context and which implies those conditions. As was observed e.g. in [16, Remark 2.11] or [43] usually one requires $\mathscr{O}_{ad}$ to be compact and demands continuity of the design-to-state mapping and (weak) lower semicontinuity of the objective in a suitable setting. One possibility to apply this idea in the method of mapping context is to require

Assumption 2 *1. Ω_{ref} is a bounded Lipschitz domain and Γ_B is a C^1-manifold.*
2. $\mathscr{U}$ is a reflexive Banach space, $\mathscr{V}$, Y_{ref}, Z_{ref} are Banach spaces.
3. The boundary-to-domain operator $T : \mathscr{U} \to \mathscr{V}$ is completely continuous
4. There exists a design-to-state operator $S : \mathscr{V} \to Y_{ref}$ that is continuous.
5. The objective J is bounded from below by $\underline{J} > -\infty$ and is continuous.

6. *$\mathscr{U}_{ad} \subset \mathscr{U}_{feas}$ is nonempty, closed and convex. Here $\mathscr{U}_{feas} \subset \mathscr{U}$ is such that $T(u) \in \mathscr{V}_{feas}$ $\forall u \in \mathscr{U}_{feas}$.*

These conditions ensure that j is coercive and weakly lower semicontinuous. Indeed the coercivity follows from J being bounded from below and the coercivity of the norm. Furthermore $u \mapsto J(T(u), S(T(u)))$ is weakly continuous because T is completely continuous and S, J are continuous. It is well known that $u \mapsto \|u\|^2_{\mathscr{U}}$ is weakly lower semicontinuous. Hence we obtain the existence of a solution if Assumption 2 is satisfied.

We also give sufficient conditions for the twice continuous differentiability of the reduced objective functional $j : \mathscr{U}_{ad} \to \mathbb{R},\ j(u) = J(T(u), S(T(u))) + \beta\|u\|^2_{\mathscr{U}}$.

Assumption 3 *1. $T : \mathscr{U} \to \mathscr{V}$, $E : \mathscr{V}_{feas} \times Y_{ref} \to Z_{ref}$ and $J : \mathscr{V}_{feas} \times Y_{ref} \to \mathbb{R}$ are twice continuously Fréchet-differentiable.*
2. *For any bounded $\mathscr{V}_{cf} \subset \mathscr{V}_{feas}$, there exists a neighborhood $\hat{\mathscr{V}}_{cf} \subset \mathscr{V}_{feas}$, and a unique design-to-state operator $S : \hat{V}_{cf} \to Y_{ref}$. Furthermore*
3. *$E_y(V, S(V)) \in \mathscr{L}(Y_{ref}, Z_{ref})$ is continuously invertible for all $V \in \hat{\mathscr{V}}_{cf}$.*

The implicit function theorem yields that S is twice continuously Fréchet differentiable and hence the same holds for the reduced objective j.

Remark: Within the context of the method of mappings it is natural to employ the concept of Fréchet-differentiability. Other approaches to shape optimization (e.g. the speed method) use the shape calculus, where j' is called a material derivative. See [43] or [9] for details on shape calculus.

3 Pointwise geometric shape constraints

In this section we want to address the question of how to handle a restriction on the shape of the domains Ω by geometric constraints. We study the constraint

$$\mathscr{U}_{ad} = \{u \in \mathscr{U} \mid x + u(x) \in C\ \forall x \in \Gamma_B\} = \{u \in \mathscr{U} \mid \tau(\Gamma_B) \subset C\}, \tag{10}$$

where $C \subset \mathbb{R}^d$ is a closed, convex set.

Remark: Clearly $\mathscr{U}_{ad}$ is closed and convex. The same holds for its $\mathbf{L}^2$-relaxation $\mathscr{U}^{\mathbf{L}}_{ad} := \{u \in \mathbf{L}^2(\Gamma_B) \mid (\mathrm{id} + u)(x) \subset C \text{ for a. e. } x \in \Gamma_B\}$.

Lemma 1 *Let $\mathscr{U}$ be a Hilbert space and $\bar{u} \in \mathscr{U}$ be a local solution of* (9) *in which j is Gâteaux-differentiable. Furthermore suppose $T(\bar{u}) \in V_{feas}$. Then the following optimality conditions hold and are equivalent:*

$$\bar{u} \in \mathscr{U}_{ad}, \quad \langle j'(\bar{u}), u - \bar{u}\rangle_{\mathscr{U}^*,\mathscr{U}} \geq 0 \quad \forall u \in \mathscr{U}_{ad}. \tag{11}$$

$$\bar{u} = P_{geo}(\bar{u} - c\nabla j(\bar{u})). \tag{12}$$

Here $P_{geo} : \mathscr{U} \to \mathscr{U}$ denotes the projection onto $\mathscr{U}_{ad}$, $c > 0$ is arbitrary but fixed and $\nabla j(u) \in \mathscr{U}$ denotes the Riesz-representation of $j'(u) \in \mathscr{U}^*$.

Proof Compare [21, Corollary 1.2]. □

Lemma 1 demonstrates the difficulty of the problem (9). We can reformulate variational inequality (11) into equation (12), but there is no pointwise interpretation of the projection operator in the Hilbert space $\mathscr{U}$. In particular there is so far no semismoothness concept of this projection operator. Hence we can not use standard techniques like semismooth Newton methods to solve (12). One possible solution is to study a regularized problem.

3.1 The regularized problem

In this section we will introduce a regularization for the geometric constraint $\tau(\Gamma_B) \subset C$ and show existence of a solution to the original and the regularized problem. We introduce the superposition operator

$$P_C : P_C(v)(x) = \tilde{P}_C(v(x)), \tag{13}$$

where $\tilde{P}_C : \mathbb{R}^d \to \mathbb{R}^d$ is the projection onto $C \subset \mathbb{R}^d$.

We approximate (9) by the regularized problem

$$\min_{u\in\mathscr{U}} j_\gamma(u) := j(u) + \tfrac{\gamma}{2}\|\mathrm{id} + u - P_C(\mathrm{id}+u)\|^2_{\mathbf{L}^2(\Gamma_B)}. \tag{14}$$

The term $\frac{\gamma}{2}\|\mathrm{id} + u - P_C(\mathrm{id}+u)\|^2_{\mathbf{L}^2(\Gamma_B)} = \frac{\gamma}{2}\|\tau - P_C(\tau)\|^2_{\mathbf{L}^2(\Gamma_B)}$, is a Moreau-Yosida type regularization term with regularization parameter γ. We abbreviate $\delta_u := \mathrm{id} + u - P_C(\mathrm{id}+u)$ and $R_\gamma(u) := \frac{\gamma}{2}\|\mathrm{id}+u-P_C(\mathrm{id}+u)\|^2_{\mathbf{L}^2(\Gamma_B)}$. In Lemma 10 we will show that $P_C(\mathrm{id}+u) - \mathrm{id}$ describes the $\mathbf{L}^2$-projection onto $\mathscr{U}^{\mathbf{L}}_{ad}$.

We can use the Moreau-Yosida theory (c.f. [10, Section 6.4.1]) to deduce

Lemma 2 *Let $\Omega \subset \mathbb{R}^d$ be a measurable, bounded subset of $\mathbb{R}^d$ and $p > q \geq 1$. Further suppose $C \subset \mathbb{R}^d$ is nonempty, closed, and convex. Then the mapping $h : \mathbb{R}^d \to \mathbb{R}$, $h(x) := \frac{1}{2}\|x - \tilde{P}_C(x)\|_2^2$ is Lipschitz continuously differentiable and convex. Furthermore the associated superposition operator*

$$T_h : \mathbf{L}^p(\Omega) \to L^q(\Omega),\; T_h(v)(x) = \frac{1}{2}\|v(x) - \tilde{P}_C(v(x))\|_2^2,$$

is convex and continuously Fréchet-differentiable with derivative

$$T_h'(v) \in \mathscr{L}(\mathbf{L}^p(\Omega), L^q(\Omega)),\; [T_h'(v)w](x) = (v(x) - \tilde{P}_C(v(x)))^T w(x).$$

Proof Since C is closed, convex, and nonempty, the indicator function of $C \subset \mathbb{R}^d$,

$$\chi_C(x) = \begin{cases} 0 & x \in C, \\ \infty & \text{else.} \end{cases},$$

is a proper, convex, and lower semicontinuous function. The projection $\tilde{P}_C(x)$ onto C coincides with the proximal mapping $P_1\chi_C(x)$, which is defined by [40][Def. 1.22]

$$P_1\chi_C(x) = \operatorname*{argmin}_{y\in\mathbb{R}^d} \chi_C(y) + \frac{1}{2}\|y - x\|^2.$$

The Moreau envelope [40][Def. 1.22]

$$e_1\chi_C(x) = \min_{y\in\mathbb{R}^d} \chi_C(y) + \frac{1}{2}\|y-x\|^2 = \min_{y\in C}\frac{1}{2}\|y-x\|^2 = h(x),$$

is convex and continuously differentiable [40][Thm. 2.26] with derivative

$$\nabla h(x) = \nabla e_1\chi_C(x) = x - P_1\chi_C(x) = x - \tilde{P}_C(x).$$

Since the mapping $x \mapsto P_1\chi_C(x)$ is Lipschitz ($P_1\chi_C(x) = \tilde{P}_C(x)$ is a projection, but this is also true for general proximal mappings) we conclude that $e_1\chi_C(x) = h$ is Lipschitz continuously differentiable.

Finally h and ∇h satisfy the growth conditions for the differentiability of superposition operators from $\mathbf{L}^p(\Omega) \to L^q(\Omega)$ (note that $\|\tilde{P}_C(x)\|_2 \leq \|\tilde{P}_C(0)\|_2 + \|x\|_2$). Hence we have

$$[T_h'(v)w](x) = \nabla h(v(x))^T w(x) = (v(x) - \tilde{P}_C(v(x)))^T w(x),$$

as claimed. □

Remark: Noting that $\|x-\tilde{P}_C(x)\|^2 = d_C^2(x)$ where d_C is the distance function of the set C the result can also be obtained by a more geometric argumentation. In particular [9, Theorem 6.8.1] shows that d_C is convex if and only if $\overline{C}$ is convex and in this case it holds $d_C^2 \in C^{1,1}_{\mathrm{loc}}(\mathbb{R}^d)$.

Let us now compute $j_\gamma'(u)$. If j is differentiable the chain rule and Lemma 2 yield for all $v \in \mathscr{U}$

$$\begin{aligned}\langle j_\gamma'(u), v\rangle_{\mathscr{U}^*,\mathscr{U}} &= \langle j'(u), v\rangle_{\mathscr{U}^*,\mathscr{U}} + \gamma \int_{\Gamma_B} (x + u(x) - \tilde{P}_C(x+u(x)))^T v(x)\,\mathrm{d}x \\ &= \langle j'(u), v\rangle_{\mathscr{U}^*,\mathscr{U}} + \gamma(\mathrm{id} + u - P_C(\mathrm{id}+u), v)_{\mathbf{L}^2(\Gamma_B)}.\end{aligned}$$

Note that usually Assumption 2.7. will not be satisfied, i.e. $\mathscr{U}_{ad} \not\subset \mathscr{U}_{feas}$. To ensure (4) we introduce some closed, convex set $\mathscr{V}_{cf} \subset \mathscr{V}_{feas}$ and modify the objective functional by adding the indicator function $\chi_{\mathscr{V}_{cf}}$.

Corollary 1 *Let Assumptions 2.1.-4. hold. Further suppose $\tilde{J} : \mathscr{V} \times Y_{ref} \to \mathbb{R}$ is bounded from below and continuous. Define $J(V,y) := \tilde{J}(V,y) + \chi_{\mathscr{V}_{cf}}(V)$. Then the problem* (9) *admits a global solution. The same holds for the regularized problem* (14) *for any $\gamma \geq 0$.*

Proof Compare section 2.2. □

The method of mappings is only applicable if the optimal shape is already roughly known, i.e. if the reference domain can be chosen close enough to the optimal one. In many (e.g. engineering) applications this is the case. The adjacency translates into V being small. Hence the following simplification is reasonable:

Assumption 4 *Every solution $\bar{u}$ of* (9) *satisfies $T(\bar{u}) \in int\ \mathscr{V}_{cf}$. For every $\gamma \geq 0$ every solution u_γ of* (14) *satisfies*

$$T(u_\gamma) \in int\ \mathscr{V}_{cf}.$$

This assumption guarantees that a necessary optimality condition of (14) is

$$j'_\gamma(u) = 0.$$

We further specify the setting we are working with by making the following assumptions:

Assumption 5 *1.* $\mathscr{V} = \mathbf{W}^{1,\infty}_{ref}$ *and* $\mathscr{U}$ *is a Hilbert space with inner product* $(\cdot,\cdot)_{\mathscr{U}}$ *and associated norm* $\|\cdot\|_{\mathscr{U}} := \sqrt{(\cdot,\cdot)_{\mathscr{U}}}$. *For* $\alpha > 0$ *there holds the compact embedding*

$$\mathscr{U} \hookrightarrow\hookrightarrow \mathbf{C}^{1,\alpha}(\Gamma_B) \hookrightarrow \mathbf{W}^{1,\infty}(\Gamma_B).$$

2. If $\bar{u}$ *solves* (9) *it is* not *a solution of the unconstrained problem*

$$\min_{u\in\mathscr{U}} j(u).$$

3. $u = 0 \in \mathscr{U}_{ad}$ *and* $\beta > 0$.

We collect our working assumptions for ease of reference.

Assumption 6 *Assumptions 2.1.-5., 3, 4 , and 5 are satisfied.*

3.2 Properties of the regularized solutions

In this section we show that any strict local solution of (9) is a strong accumulation point of a sequence of local solutions of (14) for $\gamma \to \infty$. This convergence result and its proof are inspired by the ideas presented in [34]. The same ideas are used in [30] to obtain similar results. The result is easily extended to show that any weakly convergent sequence of global solutions of (14) converges strongly towards a global solution of (9). Let $(\gamma_n) \subset \mathbb{R}_{>0}$. We denote by $u_n \in \mathscr{U}_{ad}$ a local solution of (14) with $\gamma = \gamma_n$. We will show that any accumulation point of a sequence of local solutions u_γ, with $\gamma \to \hat{\gamma}$, is a local solution of (14) with $\gamma = \hat{\gamma}$.

We begin by observing that if j is locally Lipschitz continuous then $\|\delta_{u_\gamma}\|_{\mathbf{L}^2(\Gamma_B)} \to 0$ for $\gamma \to \infty$.

Lemma 3 *Let* j *be locally Lipschitz continuous. For all* $\bar{u} \notin \mathscr{U}_{ad}$ *there exists* $\gamma_0 > 0$ *such that for all* $\gamma > \gamma_0$ $\bar{u}$ *is* not *a local solution of* (14) *with that* γ.

Proof Consider an arbitrary $\bar{u} \notin \mathscr{U}_{ad}$ and an $0 < \varepsilon < 1$. By assumption there exists a local Lipschitz constant $L > 0$ of j on $B_{\mathscr{U}}(\bar{u},\varepsilon)$. Let $v = P_{\mathscr{U}_{ad}}(\bar{u})$ and define $u = \bar{u} + \frac{\varepsilon}{2}(v - \bar{u})$. Then $u \in B_{\mathscr{U}}(\bar{u},\varepsilon)$ and, by convexity (see Lemma 2),

$$\begin{aligned}\|\delta_u\|^2_{\mathbf{L}^2(\Gamma_B)} &= \|(1-\tfrac{\varepsilon}{2})(\mathrm{id}+\bar{u}) + \tfrac{\varepsilon}{2}(\mathrm{id}+v) - P_C\left((1-\tfrac{\varepsilon}{2})(\mathrm{id}+\bar{u}) + \tfrac{\varepsilon}{2}(\mathrm{id}+v)\right)\|^2_{\mathbf{L}^2(\Gamma_B)} \\ &\le (1-\tfrac{\varepsilon}{2})\|\delta_{\bar{u}}\|^2_{\mathbf{L}^2(\Gamma_B)} + \tfrac{\varepsilon}{2}\|\delta_v\|^2_{\mathbf{L}^2(\Gamma_B)} = (1-\tfrac{\varepsilon}{2})\|\delta_{\bar{u}}\|^2_{\mathbf{L}^2(\Gamma_B)},\end{aligned}$$

since $v \in \mathscr{U}_{ad}$. Now we choose $\gamma > \gamma_0 := \frac{4L}{\|\delta_{\bar{u}}\|^2_{\mathbf{L}^2(\Gamma_B)}}$ and compute

$$\begin{aligned} j_\gamma(u) - j_\gamma(\bar{u}) &= j(u) - j(\bar{u}) + \tfrac{\gamma}{2}(\|\delta_u\|^2_{\mathbf{L}^2(\Gamma_B)} - \|\delta_{\bar{u}}\|^2_{\mathbf{L}^2(\Gamma_B)}) \\ &< L\|u-\bar{u}\|_{\mathscr{U}} + \tfrac{2L}{\|\delta_{\bar{u}}\|^2_{\mathbf{L}^2(\Gamma_B)}}(-\tfrac{\varepsilon}{2}\|\delta_{\bar{u}}\|^2_{\mathbf{L}^2(\Gamma_B)}) \\ &\le L\varepsilon - L\varepsilon = 0. \end{aligned}$$

Hence for any $\bar{u} \notin \mathscr{U}_{ad}$ we find a γ_0 such that for all $\gamma > \gamma_0$, $\bar{u}$ is not a local minimum of (14). □

Adopting an idea from [34] we consider the following auxiliary problem. For a $\bar{u} \in \mathscr{U}_{ad}$ and $r > 0$ consider

$$\min_{u \in \mathscr{U}} j_\gamma(u) \text{ s.t. } u \in \overline{B_{\mathscr{U}}(\bar{u}, r)}. \tag{15}$$

Since $\overline{B_{\mathscr{U}}(\bar{u}, r)}$ is convex, closed, bounded and non-empty there exists a (global) solution $u^r_\gamma \in \mathscr{U}$ of (15). We begin by studying the properties of u^r_γ for $\gamma \to \infty$.

Lemma 4 *Let Assumption 6 hold and $(\gamma_n) \subset \mathbb{R}_{>0}$ tend to ∞. Let $\bar{u} \in \mathscr{U}_{ad}$ be a local solution of* (9) *on $B_{\mathscr{U}}(\bar{u}, \delta)$. Then there exists a weak accumulation point of the sequence of global solutions u^r_n of* (15) *for $\gamma = \gamma_n$ and $r = \frac{\delta}{2}$. Further, any weakly convergent subsequence $u^r_k \rightharpoonup u^* \in \mathscr{U}$ converges strongly and $u^* \in \mathscr{U}_{ad}$ is a local solution of* (9)*.*

Proof $u^r_n \subset \overline{B_{\mathscr{U}}(\bar{u}, r)}$ is bounded and hence there exists a weakly convergent subsequence. Now consider an arbitrary weakly convergent subsequence $u^r_k \rightharpoonup u^* \in \overline{B_{\mathscr{U}}(\bar{u}, r)}$. It holds $\|\delta_{u^r_k}\|_{\mathbf{L}^2(\Gamma_B)} \to 0$. Since $\mathscr{U} \hookrightarrow\hookrightarrow \mathbf{L}^2(\Gamma_B)$ we conclude $\|\delta_{u^*}\|_{\mathbf{L}^2(\Gamma_B)} = 0$ and hence $u^* \in \mathscr{U}_{ad}$. Obviously $\bar{u} \in \mathscr{U}_{ad}$ is feasible for (15). Since $u^r_k \rightharpoonup u^*$ in $\mathscr{U}$ and $j : \mathscr{U} \to \mathbb{R}$ is weakly lower semicontinuous we conclude

$$j(u^*) \le \liminf_{k\to\infty} j(u^r_k) \le \liminf_{k\to\infty} j_{\gamma_k}(u^r_k) \le \limsup_{k\to\infty} j_{\gamma_k}(u^r_k) \le \limsup_{k\to\infty} j_{\gamma_k}(\bar{u}) = j(\bar{u}). \tag{16}$$

We used the optimality of u^r_k in the last inequality. By $u^* \in \overline{B_{\mathscr{U}}(\bar{u}, \frac{\delta}{2})} \cap \mathscr{U}_{ad}$ we also have $j(\bar{u}) \le j(u^*)$, which implies $j(\bar{u}) = j(u^*)$. By assumption it holds $j(u) \ge j(\bar{u}) = j(u^*)$, $\forall u \in \mathscr{U}_{ad} \cap \overline{B_{\mathscr{U}}(\bar{u}, \delta)}$ and by construction $u^* \in \overline{B_{\mathscr{U}}(\bar{u}, \frac{\delta}{2})} \subset \overline{B_{\mathscr{U}}(\bar{u}, \delta)}$. We conclude that u^* is a local minimum of (9).

In particular $j(u^r_k) \le j(\bar{u}) = j(u^*)$ and $\liminf_{k\to\infty} j(u^r_k) = j(u^*)$ hence

$$\lim_{k\to\infty} j(u^r_k) = j(u^*) = J(T(u^*), S(T(u^*))) + \frac{\beta}{2}\|u^*\|^2_{\mathscr{U}}.$$

On the other hand T is completely continuous, S, J are continuous, hence $u^r_k \rightharpoonup u^*$ implies

$$J(T(u^r_k), S(T(u^r_k))) \to J(T(u^*), S(T(u^*))).$$

We conclude $\|u^r_k\|_{\mathscr{U}} \to \|u^*\|_{\mathscr{U}}$. Weak convergence plus convergence in the norm imply the strong convergence $u^r_k \to u^*$ in $\mathscr{U}$. □

Theorem 1 *Let Assumption 6 hold and $\bar{u} \in \mathscr{U}_{ad}$ be a strict local solution of* (9) *on $B_{\mathscr{U}}(\bar{u},\delta)$. Then for any $\gamma_n \to \infty$, any sequence of global solutions $u_n^r \subset \mathscr{U}$ of* (15) *with $r < \delta$ converges strongly in $\mathscr{U}$ to $\bar{u}$ and there exists $\hat{n} > 0$ such that for all $n \geq \hat{n}$ the u_n^r are local solutions of* (14).

Proof In Lemma 4 we showed that there exists a weakly convergent subsequence and that any such subsequence converges to a local solution u^* of (9). In particular we proved $j(u^*) = j(\bar{u})$. Since $\bar{u}$ is locally unique this implies $u^* = \bar{u}$. Hence $\bar{u}$ is the only weak accumulation point of the bounded sequence u_n^r which implies that the whole sequence converges weakly $u_n^r \rightharpoonup \bar{u}$ in $\mathscr{U}$. Lemma 4 implies that the convergence is strong. Finally since $u_n^r \to \bar{u}$ there exists $\hat{n} > 0$ such that for all $n \geq \hat{n}$ it holds $u_n^r \in B_{\mathscr{U}}(\bar{u}, r)$. Hence the u_n^r are local solutions of (14). □

Let us briefly consider global solutions $\bar{u}^g, u_\gamma^g$ of (9) and (14). Taking r large enough we can achieve $T(\overline{B_{\mathscr{U}}(\bar{u}^g, r)}) \supset \mathscr{V}_{feas}$. In particular $u_\gamma^r = u_\gamma^g$. The boundedness is achieved by j being coercive. Hence Lemma 4 implies

Corollary 2 *Let Assumption 6 hold and $(\gamma_n) \subset \mathbb{R}_{>0}$ tend to ∞. Then there exists a weakly convergent subsequence of (u_n^g). Furthermore any weakly convergent subsequence $u_k^g \rightharpoonup u^* \in \mathscr{U}$ converges strongly and u^* is a global solution of* (9).

We now turn towards the case $\gamma \to \hat{\gamma}$. We begin with the auxiliary problem

$$\min_{u \in \mathscr{U}} \quad j_\gamma(u), \qquad \text{s.t. } u \in A, \tag{17}$$

for some fixed, closed, convex and nonempty set $A \subset \mathscr{U}$.

Theorem 2 *Let Assumption 6 hold, $\hat{\gamma} > 0$ and $\gamma_n \to \hat{\gamma}$. Then there exists a weakly convergent subsequence (u_k) of the sequence of global solutions (u_n) of* (17). *Furthermore any weakly convergent subsequence $u_k \rightharpoonup u^* \in \mathscr{U}$ converges strongly and u^* solves* (17) *with $\gamma = \hat{\gamma}$.*

Proof (i) Since j is coercive the sequence (u_n) is bounded, hence there exists a weakly convergent subsequence.

(ii) Now let (u_k) be a weakly convergent subsequence $u_k \rightharpoonup u^*$ for some $u^* \in \mathscr{U}$. Since A is weakly closed it holds $u^* \in A$. We start by showing that u^* solves (17) with $\gamma = \hat{\gamma}$. Denote by $\hat{u} := u_{\hat{\gamma}}$ a global solution of (17) with $\gamma = \hat{\gamma}$. Since $u_k \rightharpoonup u^*$ in $\mathscr{U}$ we have by compact embedding $u_k \to u^*$ in $\mathbf{L}^2(\Gamma_B)$. Hence

$$j(u^*) \leq \liminf_{k\to\infty} j(u_k) \text{ and } R_{\gamma_k}(u_k) \to R_{\hat{\gamma}}(u^*).$$

Furthermore for all k it holds

$$j(u_k) + R_{\gamma_k}(u_k) = j_{\gamma_k}(u_k) \leq j_{\gamma_k}(\hat{u}) = j(\hat{u}) + R_{\gamma_k}(\hat{u}).$$

Thus we see that

$$j_{\hat{\gamma}}(u^*) = j(u^*) + R_{\hat{\gamma}}(u^*) \leq \liminf_{k\to\infty} j(u_k) + R_{\gamma_k}(u_k) \leq \liminf_{k\to\infty} j(\hat{u}) + R_{\gamma_k}(\hat{u}) = j_{\hat{\gamma}}(\hat{u}),$$

which implies that u^* solves (17) with $\gamma = \hat{\gamma}$.

(iii) The strong convergence $u_k \to u^*$ follows as in the proof of Lemma 4. □

Remark: Obviously $A = \mathscr{U}$ is possible and yields the result for global solutions of (14).

Now we consider a strict local solution $u_{\hat{\gamma}}$ of (14).

Corollary 3 *For $\hat{\gamma} > 0$ let Assumption 6 be satisfied. Denote by $u_{\hat{\gamma}}$ a strict local solution of* (14) *with $\gamma = \hat{\gamma}$ on $B_{\mathscr{U}}(\bar{u}, \delta)$. Set $0 < r < \delta$ and $A = \overline{B_{\mathscr{U}}(u_{\hat{\gamma}}, r)}$. Then for any $\gamma_n \to \hat{\gamma}$, any sequence of global solutions (u_n^r) of* (17) *converges strongly in $\mathscr{U}$ to $u_{\hat{\gamma}}$ and for γ_n close enough to $\hat{\gamma}$ the u_n^r are local solutions of* (14) *with $\gamma = \gamma_n$.*

Proof Using Theorem 2 we obtain a subsequence (γ_k) with $u_k^r \to u^*$ in $\mathscr{U}$, where u^* solves (17) for $\gamma = \hat{\gamma}$. Furthermore any weakly convergent subsequence converges towards such a solution. Since $u_{\hat{\gamma}}$ is a strict local solution of (14) and $r < \delta$ it is the unique solution of (17) for $\gamma = \hat{\gamma}$ which implies $u^* = u_{\hat{\gamma}}$. Hence $u_{\hat{\gamma}}$ is the only weak accumulation point of the bounded sequence u_n^r, therefore the whole sequence converges weakly: $u_n^r \rightharpoonup u_{\hat{\gamma}}$. Theorem 2 shows that the convergence is strong. Finally for γ_n close enough to $\hat{\gamma}$ we have $u_n^r \in B_{\mathscr{U}}(u_{\hat{\gamma}}, r)$, since $u_n^r \to u_{\hat{\gamma}}$. Hence the u_n^r are local solutions of (14) with $\gamma = \gamma_n$. □

3.3 Solving the regularized problem

In section 3.2 we showed that a solution of (9) can be found by solving a sequence of relaxed problems of the form (14) with regularization parameter γ_n tending to infinity. Of course this is only a practical strategy if the regularized problems can be solved efficiently. We will show that we can apply a semismooth Newton method to solve the first order optimality condition:

$$j'_{\gamma_n}(u_n) = 0 \text{ in } \mathscr{U}^*. \tag{18}$$

Remember $\langle j'_\gamma(u), v\rangle_{\mathscr{U}^*,\mathscr{U}} = \langle j'(u), v\rangle_{\mathscr{U}^*,\mathscr{U}} + \gamma(\mathrm{id} + u - P_C(\mathrm{id} + u), v)_{\mathbf{L}^2(\Gamma_B)}$. The superposition operator P_C is not differentiable, but as it turns out semismooth.

We start by defining the generalized differential of $P_C : \mathscr{U} \to \mathbf{L}^q(\Gamma_B)$, $q \geq 1$. We need

Assumption 7 *The projection $\tilde{P}_C : \mathbb{R}^d \to \mathbb{R}^d$ is $\partial\tilde{P}_C$-semismooth, where $\partial\tilde{P}_C$ denotes Clarke's generalized Jacobian [8].*

Remark: This is a condition on the convex, closed set $C \subset \mathbb{R}^d$.

Definition 1 We introduce

$$\begin{aligned} &\partial P_C : \mathscr{U} \rightrightarrows \mathscr{L}(\mathscr{U}, \mathbf{L}^q(\Gamma_B)), \\ &\partial P_C(w) := \{M \mid Mv = g^T v,\ g \in \mathbf{L}^\infty(\Gamma_B),\ g(x) \in \partial\tilde{P}_C(w(x)),\ x \in \Gamma_B\} \end{aligned}$$

This definition and the following theorem are based on [45]. For a compact overview see also [21, Chapter 2].

Theorem 3 *Let Assumption 7 hold. Then ∂P_C is well defined. Furthermore $P_C : \mathscr{U} \to \mathbf{L}^q(\Gamma_B)$, $q \geq 1$, is ∂P_C-semismooth.*

Proof The first statement is immediate. Let $q \geq 1$. It suffices to study the components $\tilde{P}_C^i$ and P_C^i (c.f. [45, Proposition 3.6]). The Lipschitz continuity of the projection $\tilde{P}_C : \mathbb{R}^d \to \mathbb{R}^d$ for $C \subset \mathbb{R}^d$ closed and convex is well known. Semismoothness of $\tilde{P}_C^i$ follows from Assumption 7. Furthermore for all $p > q$ the mapping $\mathscr{U} \ni u \mapsto \mathrm{id} + u \in \mathbf{L}^p(\Gamma_B)$ is continuously Fréchet-differentiable and Lipschitz continuous since $\mathscr{U} \hookrightarrow\hookrightarrow \mathbf{W}^{1,\infty}(\Gamma_B) \hookrightarrow \mathbf{L}^p(\Gamma_B)$. Hence we are able to employ [21, Theorem 2.13] and the claim follows. □

Hence if we want to solve (18) with a Newton-type method we need to employ semismooth calculus. We obtain j'_γ is $\partial j'_\gamma$-semismooth, where $\partial j'_\gamma : \mathscr{U} \rightrightarrows \mathscr{L}(\mathscr{U}, \mathscr{U}^*)$ and

$$\begin{aligned}&H_\gamma \in \partial j'_\gamma(u) \Leftrightarrow \exists M \in \partial P_C(\mathrm{id} + u) :\\ &\langle H_\gamma v, w\rangle_{\mathscr{U}^*,\mathscr{U}} = \langle j''(u)v, w\rangle_{\mathscr{U}^*,\mathscr{U}} + \gamma(v - Mv, w)_{\mathbf{L}^2(\Gamma_B)}, \ \forall v, w \in \mathscr{U}.\end{aligned}$$

For the superlinear convergence of the semismooth Newton method we need a regularity condition like $\exists C, \delta > 0$ such that

$$\|(H_\gamma)^{-1}\|_{\mathscr{L}(\mathscr{U},\mathscr{U}^*)} \leq C, \quad \forall H_\gamma \in \partial j'_\gamma(u), \quad \forall u \in B_{\mathscr{U}}(u_\gamma, \delta).$$

The following assumption assures this regularity condition for $u^* = u_\gamma$:

Assumption 8 *Let $\gamma > 0$ and $u^* \in \mathscr{U}$, furthermore suppose there exist $\alpha, \delta > 0$:*

$$\langle H_\gamma v, v\rangle_{\mathscr{U}^*,\mathscr{U}} \geq \alpha \|v\|^2_{\mathscr{U}}, \quad \forall v \in \mathscr{U}, \forall H_\gamma \in \partial j'_\gamma(u), \quad \forall u \in B_{\mathscr{U}}(u^*, \delta).$$

Theorem 4 *Let Assumptions 6, 7 hold, and Assumption 8 be satisfied for $\gamma > 0$ and a (local) solution u_γ. Then there exists $\delta > 0$ such that for all initial points*

$$u^0 \in \mathscr{U} \ \ \textit{with} \ \ \|u^0 - u_\gamma\|_{\mathscr{U}} < \delta,$$

the semismooth Newton method converges q-superlinearly to u_γ.

Proof Compare [21, Theorem 2.12]. □

Actually we can restrict ourselves to an assumption which solely depends on j

Assumption 9 *Let $u^* \in \mathscr{U}$ and suppose there exist $\alpha, \delta > 0$ such that*

$$\langle j''(u)v, v\rangle_{\mathscr{U}^*,\mathscr{U}} \geq \alpha \|v\|^2_{\mathscr{U}}, \quad \forall v \in \mathscr{U}, \forall u \in B_{\mathscr{U}}(u^*, \delta).$$

Theorem 5 *Let Assumptions 6, 7 hold and Assumption 9 be satisfied for some $u^* \in \mathscr{U}$. Then Assumption 8 holds as well for any $\gamma > 0$.*

Proof Let Assumption 9 be satisfied for some $u^* \in \mathscr{U}$ and let $\gamma > 0$ be arbitrary. Since the projection has Lipschitz constant one and the norm of Clarke's generalized Jacobian is bounded by the Lipschitz constant we have for all $u \in \mathscr{U}$ and $v \in \mathscr{U}$:

$$\forall M \in \partial P_C(u) : \quad \|Mv\|_{\mathbf{L}^2(\Gamma_B)} \leq \|v\|_{\mathbf{L}^2(\Gamma_B)}. \tag{19}$$

Thus for all $u \in B_{\mathscr{U}}(u^*, \delta)$ and any $H_\gamma \in \partial j'_\gamma(u)$:

$$\begin{aligned}\langle H_\gamma v, v\rangle_{\mathscr{U}^*,\mathscr{U}} &= \langle j''(u)v, v\rangle_{\mathscr{U}^*,\mathscr{U}} + \gamma(v - Mv, v)_{\mathbf{L}^2(\Gamma_B)}\\ &\geq \alpha\|v\|^2_{\mathscr{U}} + \gamma(\|v\|^2_{\mathbf{L}^2(\Gamma_B)} - \|v\|^2_{\mathbf{L}^2(\Gamma_B)}) = \alpha\|v\|^2_{\mathscr{U}}.\end{aligned}$$

Hence the claim follows. □

Assumption 9 seems to be restrictive since the coercivity of j'' is required in the strong norm $\|\cdot\|_{\mathscr{U}}$. In fact in the case of a Tikhonov-regularized objective this assumption is equivalent to the positivity assumption of j'' under certain assumptions (compare [7,24]). In particular we need that the mappings $v \mapsto T'(u)v$ and $v \mapsto T''(u)vv$ are completely continuous. For linear extension operators this follows immediately.

Assumption 10 *Let $u \in \mathscr{U}$ and suppose that*

$$\langle j''(u)v, v\rangle_{\mathscr{U}^*,\mathscr{U}} > 0 \quad \forall v \in \mathscr{U}\setminus\{0\}.$$

Lemma 5 *Let Assumption 6 hold. Furthermore suppose Assumption 10 is satisfied, and $T'(u) \in \mathscr{L}(\mathscr{U},\mathscr{V})$ as well as $v \mapsto (T''(u)v)v$ are completely continuous. Then there exists $\alpha > 0$ such that*

$$\langle j''(u)v, v\rangle_{\mathscr{U}^*,\mathscr{U}} \geq \alpha\|v\|^2_{\mathscr{U}}, \quad \forall v \in \mathscr{U}. \tag{20}$$

Proof In [7, Remark 2.7] this was shown if $j''(u)$ is a Legendre form i.e.,

(i) if $v_k \rightharpoonup v$ as $k \to \infty$, then $\langle j''(u)v, v\rangle_{\mathscr{U}^*,\mathscr{U}} \leq \liminf_{n\to\infty}\langle j''(u)v_n, v_n\rangle_{\mathscr{U}^*,\mathscr{U}}$

(ii) if additionally $\langle j''(u)v_k, v_k\rangle_{\mathscr{U}^*,\mathscr{U}} \to \langle j''(u)v, v\rangle_{\mathscr{U}^*,\mathscr{U}}$, then $\|v - v_k\|_{\mathscr{U}} \to 0$

are satisfied. Recall $j(u) = \mathrm{j}(T(u)) + \frac{\beta}{2}\|u\|^2_{\mathscr{U}}$. We have

$$\langle j''(u)v, v\rangle_{\mathscr{U}^*,\mathscr{U}} = \langle \frac{d^2}{du^2}\mathrm{j}(T(u))v, v\rangle_{\mathscr{U}^*,\mathscr{U}} + \beta\|v\|^2_{\mathscr{U}},$$

with

$$\langle \frac{d^2}{du^2}\mathrm{j}(T(u))v, v\rangle_{\mathscr{U}^*,\mathscr{U}} = \langle \mathrm{j}''(T(u))T'(u)v, T'(u)v\rangle_{\mathscr{V}^*,\mathscr{V}} + \langle \mathrm{j}'(T(u)), (T''(u)v)v\rangle_{\mathscr{V}^*,\mathscr{V}}.$$

Using the complete continuity of T' and T'' we see that

$$v_k \rightharpoonup v \Rightarrow \langle \frac{d^2}{du^2}\mathrm{j}(T(u))v_k, v_k\rangle_{\mathscr{U}^*,\mathscr{U}} \to \langle \frac{d^2}{du^2}\mathrm{j}(T(u))v, v\rangle_{\mathscr{U}^*,\mathscr{U}}.$$

Combined with the weak lower semicontinuity of the squared Hilbert space norm $\|\cdot\|^2_{\mathscr{U}}$ condition (i) follows. Furthermore $\langle j''(u)v_k, v_k\rangle_{\mathscr{U}^*,\mathscr{U}} \to \langle j''(u)v, v\rangle_{\mathscr{U}^*,\mathscr{U}}$ for $v_k \rightharpoonup v$ implies $\|v_k\|_{\mathscr{U}} \to \|v\|_{\mathscr{U}}$. Weak convergence plus convergence of the norm yields $\|v - v_k\|_{\mathscr{U}} \to 0$, hence (ii) is satisfied as well. □

3.4 Second-order conditions

In this section we study the second-order condition Assumption 8 more thoroughly. We show that a stationary point $u_{\hat{\gamma}}$ of (14) that fulfills this condition is a strict local minimum. Furthermore there exist neighborhoods of $\hat{\gamma}$ and $u_{\hat{\gamma}}$ such that the mapping $\gamma \mapsto u_\gamma$, restricted to those neighborhoods, is locally Lipschitz continuous.

We have the following quadratic growth property if Assumption 8 is satisfied:

Lemma 6 *For $\gamma > 0$ let Assumption 8 hold in a stationary point u_γ of* (14) *and let Assumptions 6, 7 be satisfied. Then there exist $\bar{\alpha}, \bar{\delta} > 0$ such that*

$$j_\gamma(u_\gamma) + \frac{\bar{\alpha}}{2} \|u_\gamma - u\|_{\mathscr{U}}^2 \leq j_\gamma(u) \text{ for all } u \in B_{\mathscr{U}}(u_\gamma, \bar{\delta}). \tag{21}$$

In particular u_γ is a strict local solution.

Proof Let $u \in B_{\mathscr{U}}(u_\gamma, \delta)$ with $\delta > 0$ as in Assumption 8. Since j_γ is continuously Fréchet-differentiable it holds

$$j_\gamma(u) - j_\gamma(u_\gamma) = \int_0^1 j'_\gamma(u_\gamma + t(u - u_\gamma))(u - u_\gamma) \text{ dt}.$$

Introducing $u_\gamma^t := (1-t)u_\gamma + tu$ and using the semismoothness of j'_γ we find

$$\begin{aligned}
&\int_0^1 \langle j'_\gamma(u_\gamma + t(u - u_\gamma)), u - u_\gamma \rangle_{\mathscr{U}^*,\mathscr{U}} \text{ dt} \\
&= \int_0^1 \langle j'_\gamma(u_\gamma), u - u_\gamma \rangle_{\mathscr{U}^*,\mathscr{U}} + \langle t H_\gamma(u_\gamma^t)(u - u_\gamma), u - u_\gamma \rangle_{\mathscr{U}^*,\mathscr{U}} + t o(\|u - u_\gamma\|_{\mathscr{U}}^2) \text{ dt} \\
&= \langle j'_\gamma(u_\gamma), u - u_\gamma \rangle_{\mathscr{U}^*,\mathscr{U}} + o(\|u - u_\gamma\|_{\mathscr{U}}^2) + \int_0^1 t \langle H_\gamma(u_\gamma^t)(u - u_\gamma), u - u_\gamma \rangle_{\mathscr{U}^*,\mathscr{U}} \text{ dt}.
\end{aligned}$$

Hence by Assumption 8 and stationarity we have

$$\begin{aligned}
j_\gamma(u) &\geq j_\gamma(u_\gamma) + \langle j'_\gamma(u_\gamma), u - u_\gamma \rangle_{\mathscr{U}^*,\mathscr{U}} + \int_0^1 t\alpha \|u_\gamma - u\|_{\mathscr{U}}^2 \text{ dt} + o(\|u - u_\gamma\|_{\mathscr{U}}^2) \\
&\geq j_\gamma(u_\gamma) + \tfrac{\bar{\alpha}}{2} \|u_\gamma - u\|_{\mathscr{U}}^2,
\end{aligned}$$

for some suitable $0 < \bar{\alpha} \leq \alpha$ and $\|u_\gamma - u\|_{\mathscr{U}}$ small enough. The last claim is clear. □

We now want to use Assumption 9 to show local Lipschitz continuity of the mapping $\gamma \mapsto u_\gamma$. First we state the following auxiliary lemma:

Lemma 7 *For any $u, v \in \mathscr{U}$ it holds $(\delta_u - \delta_v, u - v)_{\mathbf{L}^2(\Gamma_B)} \geq 0$.*

Proof Using the non-expansiveness of $\tilde{P}_C$ we compute

$$\begin{aligned}
&(\delta_u - \delta_v, u - v)_{\mathbf{L}^2(\Gamma_B)} \\
&= (\text{id} + u - P_C(\text{id} + u) - \text{id} - v + P_C(\text{id} + v), u - v)_{\mathbf{L}^2(\Gamma_B)} \\
&\geq \|u - v\|_{\mathbf{L}^2(\Gamma_B)}^2 - \|P_C(\text{id} + u) - P_C(\text{id} + v)\|_{\mathbf{L}^2(\Gamma_B)} \|u - v\|_{\mathbf{L}^2(\Gamma_B)} \\
&\geq \|u - v\|_{\mathbf{L}^2(\Gamma_B)}^2 - \|\text{id} + u - \text{id} - v\|_{\mathbf{L}^2(\Gamma_B)} \|u - v\|_{\mathbf{L}^2(\Gamma_B)} = 0.
\end{aligned}$$

□

Now we are able to show the local Lipschitz continuity of $\gamma \mapsto u_\gamma$.

Theorem 6 *For a $\hat{\gamma} > 0$ let Assumption 9 hold in a local solution $u_{\hat{\gamma}}$ of* (14) *with given $\alpha, \delta > 0$ and let Assumptions 6, 7 be satisfied. Then there exist neighborhoods $\mathscr{G}$ of $\hat{\gamma}$ and $\hat{\mathscr{U}}$ of $u_{\hat{\gamma}}$, such that for all $\gamma \in \mathscr{G}$ there exists a unique (strict) local solution u_γ of* (14) *in $\hat{\mathscr{U}}$. The mapping $\gamma \mapsto u_\gamma$ is Lipschitz continuous on $\mathscr{G}$.*

Proof By Corollary 3 we know that for γ close enough to $\hat{\gamma}$ there exists a local solution u_γ of (14) which lies close to $u_{\hat{\gamma}}$. Since Assumption 9 is satisfied in $u_{\hat{\gamma}}$ with $\delta > 0$, the same holds for all $u \in B_{\mathscr{U}}(u_{\hat{\gamma}}, \frac{\delta}{3})$ with radius $\frac{2\delta}{3}$. In particular any local solution u_γ of (14) in $B_{\mathscr{U}}(u_{\hat{\gamma}}, \frac{\delta}{3})$ is also strict and unique. This shows the first claim. The first order optimality conditions yield

$$j'_{\hat{\gamma}}(u_{\hat{\gamma}}) = 0 \text{ and } j'_\gamma(u_\gamma) = 0.$$

Testing with $u_\gamma - u_{\hat{\gamma}}$ and subtracting those two equations yields

$$\begin{aligned}0 &= \langle j'(u_\gamma) - j'(u_{\hat{\gamma}}), u_\gamma - u_{\hat{\gamma}}\rangle_{\mathscr{U}^*,\mathscr{U}} + \gamma(\delta_{u_\gamma}, u_\gamma - u_{\hat{\gamma}})_{\mathbf{L}^2(\Gamma_B)} - \hat{\gamma}(\delta_{u_{\hat{\gamma}}}, u_\gamma - u_{\hat{\gamma}})_{\mathbf{L}^2(\Gamma_B)} \\ &= \langle j'(u_\gamma) - j'(u_{\hat{\gamma}}), u_\gamma - u_{\hat{\gamma}}\rangle_{\mathscr{U}^*,\mathscr{U}} + \gamma(\delta_{u_\gamma} - \delta_{u_{\hat{\gamma}}}, u_\gamma - u_{\hat{\gamma}})_{\mathbf{L}^2(\Gamma_B)} \\ &\quad -(\gamma - \hat{\gamma})(\delta_{u_{\hat{\gamma}}}, u_\gamma - u_{\hat{\gamma}})_{\mathbf{L}^2(\Gamma_B)}.\end{aligned}$$

By Lemma 7 the second term can be bounded from below by zero. We will now use Assumption 9 to estimate the first term. Since j' is continuously Fréchet differentiable it holds

$$\begin{aligned}&\langle j'(u_\gamma) - j'(u_{\hat{\gamma}}), u_\gamma - u_{\hat{\gamma}}\rangle_{\mathscr{U}^*,\mathscr{U}} \\ &= \langle j''(u_{\hat{\gamma}})(u_\gamma - u_{\hat{\gamma}}), u_\gamma - u_{\hat{\gamma}}\rangle_{\mathscr{U}^*,\mathscr{U}} + o(\|u_\gamma - u_{\hat{\gamma}}\|^2_{\mathscr{U}}) \\ &\geq \alpha\|u_\gamma - u_{\hat{\gamma}}\|^2_{\mathscr{U}} + o(\|u_\gamma - u_{\hat{\gamma}}\|^2_{\mathscr{U}}).\end{aligned}$$

In the last step we used Assumption 9. Hence we have

$$\begin{aligned}\alpha\|u_\gamma - u_{\hat{\gamma}}\|^2_{\mathscr{U}} + o(\|u_\gamma - u_{\hat{\gamma}}\|^2_{\mathscr{U}}) &\leq (\gamma - \hat{\gamma})(\delta_{u_{\hat{\gamma}}}, u_\gamma - u_{\hat{\gamma}})_{\mathbf{L}^2(\Gamma_B)} \\ &\leq |\gamma - \hat{\gamma}|\|\delta_{u_{\hat{\gamma}}}\|_{\mathbf{L}^2(\Gamma_B)}\|u_\gamma - u_{\hat{\gamma}}\|_{\mathbf{L}^2(\Gamma_B)}.\end{aligned}$$

Since $u_\gamma \to u_{\hat{\gamma}}$ for $\gamma \to \hat{\gamma}$ we can choose a $\delta(\beta) > 0$ such that for all $|\gamma - \hat{\gamma}| < \delta$:

$$\bar{\alpha}\|u_\gamma - u_{\hat{\gamma}}\|^2_{\mathscr{U}} \leq |\gamma - \hat{\gamma}|\|\delta_{u_{\hat{\gamma}}}\|_{\mathbf{L}^2(\Gamma_B)}\|u_\gamma - u_{\hat{\gamma}}\|_{\mathbf{L}^2(\Gamma_B)} \leq |\gamma - \hat{\gamma}|\|\delta_{u_{\hat{\gamma}}}\|_{\mathbf{L}^2(\Gamma_B)}\|u_\gamma - u_{\hat{\gamma}}\|_{\mathscr{U}},$$

for $\bar{\alpha} > 0$. Boundedness of $u_{\hat{\gamma}}$ implies boundedness of $\delta_{u_{\hat{\gamma}}}$, which yields the claim. □

3.5 The value function and its model

Usually the value function is defined as $\gamma \mapsto \min_{u\in\mathscr{U}} j_\gamma(u)$. In [17] it is shown that the value function is differentiable in the linear-quadratic setting. In the nonlinear setting this is not necessarily true if the solutions of the regularized problems are not unique. Hence we have to differ between local and global solutions and restrict ourselves to a local analysis for the differentiability.

Lemma 8 *Let Assumption 6 hold. Denote the global solutions of* (14) *with* u_γ. *The mapping* $\gamma \mapsto V^g(\gamma) := j_\gamma(u_\gamma)$ *is globally Lipschitz continuous for all* $\gamma > 0$.

Proof It holds for any $\gamma_1, \gamma_2 > 0$

$$\begin{aligned} j_{\gamma_2}(u_{\gamma_2}) - j_{\gamma_1}(u_{\gamma_1}) &\le j_{\gamma_2}(u_{\gamma_1}) - j_{\gamma_1}(u_{\gamma_1}) = \tfrac{\gamma_2-\gamma_1}{2}\|\delta_{u_{\gamma_1}}\|^2_{\mathbf{L}^2(\Gamma_B)},\\ j_{\gamma_2}(u_{\gamma_2}) - j_{\gamma_1}(u_{\gamma_1}) &\ge j_{\gamma_2}(u_{\gamma_2}) - j_{\gamma_1}(u_{\gamma_2}) = \tfrac{\gamma_2-\gamma_1}{2}\|\delta_{u_{\gamma_2}}\|^2_{\mathbf{L}^2(\Gamma_B)}. \end{aligned} \tag{22}$$

u_γ is bounded in $\mathscr{U}$ for all $\gamma > 0$ and $\mathscr{U} \hookrightarrow \mathbf{L}^2(\Gamma_B)$, hence the same holds for $\|\delta_{u_\gamma}\|^2_{\mathbf{L}^2(\Gamma_B)}$ and the claim follows. □

Theorem 7 *Let Assumption 6 hold. Let $\hat{\gamma} > 0$ be arbitrary and $\hat{u}$ be a strict local solution of* (14). *Assume there exists a neighborhood $\hat{\mathscr{U}}$ of $\hat{u}$ such that for any sequence $\gamma \to \hat{\gamma}$ a sequence of local solution u_γ of* (14) *lies (for γ close enough to $\hat{\gamma}$) in $\hat{\mathscr{U}}$ and the u_γ are unique local solutions in $\hat{\mathscr{U}}$.*

Then the value function $V : \gamma \mapsto j_\gamma(u_\gamma)$ is differentiable at $\hat{\gamma}$ with

$$V'(\hat{\gamma}) = \frac{1}{2}\|\delta_{u_{\hat{\gamma}}}\|^2_{\mathbf{L}^2(\Gamma_B)}.$$

Proof (22) implies

$$\frac{1}{2}\|\delta_{u_{\hat{\gamma}}}\|^2_{\mathbf{L}^2(\Gamma_B)} \le \frac{j_{\hat{\gamma}}(\hat{u}) - j_\gamma(u_\gamma)}{\hat{\gamma}-\gamma} \le \frac{1}{2}\|\delta_{u_\gamma}\|^2_{\mathbf{L}^2(\Gamma_B)},$$

if $\hat{\gamma} \ge \gamma$ and the reverse inequality if $\hat{\gamma} < \gamma$. Using Corollary 3 we see that $u_\gamma \to u_{\hat{\gamma}}$ as $\gamma \to \hat{\gamma}$, in particular $\|\delta_{u_\gamma}\|^2_{\mathbf{L}^2(\Gamma_B)} \to \|\delta_{u_{\hat{\gamma}}}\|^2_{\mathbf{L}^2(\Gamma_B)}$. Thus we obtain $V'(\hat{\gamma}) = \frac{1}{2}\|\delta_{u_{\hat{\gamma}}}\|^2_{\mathbf{L}^2(\Gamma_B)}$. □

Remark: Note that we obtained this differentiability result without using the differentiability of j_γ. On the other hand usually we will be able to guarantee the requirements of Theorem 7 only if the second order condition Assumption 8 holds in $\hat{u}$, i.e. if j is twice continuously differentiable (see section 3.4).

It holds $V'(\gamma) \ge 0$. $V'(\gamma) = 0$ would imply $\delta_{u_\gamma} = 0$ which is only the case if $u_\gamma \in \mathscr{U}_{ad}$. This is a contradiction to Assumption 6.3. Thus we conclude that the value function V is strict monotonically increasing. Furthermore

$$\gamma_2 \ge \gamma_1 \;\Rightarrow\; \|\delta_{u_{\gamma_2}}\|^2_{\mathbf{L}^2(\Gamma_B)} \le \|\delta_{u_{\gamma_1}}\|^2_{\mathbf{L}^2(\Gamma_B)},$$

hence the mapping $\gamma \mapsto V'(\gamma)$ is monotonically decreasing.

We would like to use the value function V in an algorithmic setting to steer the γ-update. Since the function is not available explicitly we approximate it with a model function in the spirit of [17]. We define the model

$$m(\gamma) = C_1 - \frac{C_2}{(D+\gamma)^r},$$

with some constants $C_1 \in \mathbb{R}, C_2 > 0, D > 0$, and $r > 0$. Note that $m' > 0$, $m'' < 0$ corresponding to the properties of V. As proposed in [17] we will fix r whereas C_1, C_2, D will be updated iteratively. See section 4.4 for more details.

3.6 Optimality conditions and properties of the adjoint state

In this section we study the optimality conditions of (8) and its regularization. In particular we extend the results of the previous sections to the state, the adjoint state and the Lagrange multiplier associated with the geometric constraint. The results in this section were partly inspired by the ideas presented in [17] and [45]. We study

$$\min_{u\in\mathscr{U},y\in Y_{ref}} \hat{J}(u,y) \text{ s.t. } E(T(u),y)=0 \text{ in } Z_{ref},\ u\in\mathscr{U}_{ad}, \tag{23}$$

The admissible set $\mathscr{U}_{ad}$ is closed and convex, hence the tangent cone in $\bar{u}\in\mathscr{U}_{ad}$ is given by

$$T(\mathscr{U}_{ad},\bar{u}) := \mathrm{cl}\{d\in\mathscr{U} \mid d=\mu(v-\bar{u}),\ \text{where } \mu>0, v\in\mathscr{U}_{ad}\}.$$

Theorem 8 *Let Assumption 6 hold. Let $(\bar{u},\bar{y})\in\mathscr{U}\times Y_{ref}$ be a local solution of (23). Then there exists an unique adjoint state $\bar{p}\in Z^*_{ref}$ and a Lagrange multiplier $\bar{\lambda}\in\mathscr{U}^*$ such that the following optimality conditions hold*

$$\begin{aligned} \hat{J}_u(\bar{u},\bar{y})+(E_V(T(\bar{u}),\bar{y})T'(\bar{u}))^*\bar{p}+\bar{\lambda} &= 0 \textit{ in } \mathscr{U}^*,\\ \hat{J}_y(\bar{u},\bar{y})+E_y(T(\bar{u}),\bar{y})^*\bar{p} &= 0 \textit{ in } Y^*_{ref},\\ E(T(\bar{u}),\bar{y}) &= 0 \textit{ in } Z_{ref},\\ \bar{u}\in\mathscr{U}_{ad},\ \bar{\lambda} &\in T(\mathscr{U}_{ad},\bar{u})^\circ, \end{aligned} \tag{24}$$

where $T(\mathscr{U}_{ad},\bar{u})^\circ$ is the polar cone of $T(\mathscr{U}_{ad},\bar{u})$.

Proof Assumption 3.3 implies surjectivity of $E_y(T(u),S(T(u)))$ for all $u\in\mathscr{U}_{ad}$. Furthermore the tangent cone $T(\mathscr{U}_{ad},\bar{u})$ is equal to the linearization cone at $\bar{u}$. Hence Robinson's constraint qualification holds. The system can be derived in a standard way, compare for example [21, chapter 1]. The uniqueness of $\bar{p}$ follows also from Assumption 3.3. □

For the regularized problem

$$\begin{aligned} \min_{u\in\mathscr{U},y\in Y_{ref}} J_\gamma(u,y) &= \hat{J}(u,y)+\tfrac{\gamma}{2}\|\mathrm{id}+u-P_C(\mathrm{id}+u)\|^2_{\mathbf{L}^2(\Gamma_B)}\\ \text{s.t. } E(T(u),y) &= 0 \text{ in } Z_{ref}, \end{aligned} \tag{25}$$

we have under Assumption 6 for every $\gamma>0$ and a local solution $(u_\gamma,y_\gamma)\in\mathscr{U}\times Y_{ref}$ the existence of a unique adjoint state $p_\gamma\in Z^*_{ref}$ such that

$$\begin{aligned} \hat{J}_u(u_\gamma,y_\gamma)+(E_V(T(u_\gamma),y_\gamma)T'(u_\gamma))^*p_\gamma+\lambda_\gamma &= 0 \quad \text{in } \mathscr{U}^*,\\ \hat{J}_y(u_\gamma,y_\gamma)+E_y(T(u_\gamma),y_\gamma)^*p_\gamma &= 0 \quad \text{in } Y^*_{ref},\\ E(T(u_\gamma),y_\gamma) &= 0 \quad \text{in } Z_{ref},\\ \gamma(\mathrm{id}+u_\gamma-P_C(\mathrm{id}+u_\gamma)) &= \lambda_\gamma \quad \text{in } \mathbf{L}^2(\Gamma_B). \end{aligned} \tag{26}$$

Lemma 9 *Let Assumption 6 hold. If u_γ is uniformly bounded then $(y_\gamma,p_\gamma,\lambda_\gamma)$ is also bounded in $Y_{ref}\times Z_{ref}\times\mathscr{U}^*$.*

Proof If u_γ is uniformly bounded we can find a bounded, closed set $\tilde{\mathscr{U}} \subset \mathscr{U}$ with $(u_\gamma) \subset \tilde{\mathscr{U}}$. Hence the set $T(\tilde{\mathscr{U}})$ is relatively compact and $T(u_\gamma)$ is contained in the compact set $\tilde{\mathscr{V}} := cl(T(\tilde{\mathscr{U}}))$. Since S is continuous $S(\tilde{\mathscr{V}})$ is also compact. In particular we obtain $(y_\gamma) \subset S(\tilde{\mathscr{V}})$ uniformly bounded.

Since J is continuously differentiable $\hat{J}_y(u_\gamma, y_\gamma) = J_y(T(u_\gamma), y_\gamma) \subset J_y(\tilde{\mathscr{V}}, S(\tilde{\mathscr{V}}))$ is also bounded. The same holds for $E_y(T(u_\gamma), y_\gamma)$. Hence using the adjoint equation in (26) we can bound p_γ uniformly.

Finally the boundedness of λ_γ follows from the compact embedding $\mathscr{U} \hookrightarrow\hookrightarrow \mathbf{L}^2(\Gamma_B)$ and the last equation of (26). □

We will now show that $\lambda_\gamma \in \mathscr{U}^*$ approximates $\bar{\lambda}$. First note that $u \mapsto u^C := P_C(\mathrm{id} + u) - \mathrm{id}$ describes the $\mathbf{L}^2$-projection onto

$$\mathscr{U}_{ad}^{\mathbf{L}} = \{u \in \mathbf{L}^2(\Gamma_B) \mid (\mathrm{id} + u)(x) \subset C \text{ for a. e. } x \in \Gamma_B\}.$$

Lemma 10 *The $\mathbf{L}^2$-projection onto the closed convex set $\mathscr{U}_{ad}^{\mathbf{L}} \subset \mathbf{L}^2(\Gamma_B)$ is given by the mapping $u \mapsto u^C = P_C(id + u) - id$.*

Proof Recall $\tau = \mathrm{id} + u$. By construction we have $u^C = P_C(\tau) - \mathrm{id} \in \mathscr{U}_{ad}^{\mathbf{L}}$. It remains to check whether

$$(u - u^C, v - u^C)_{\mathbf{L}^2(\Gamma_B)} \leq 0 \quad \text{for all } v \in \mathscr{U}_{ad}^{\mathbf{L}}.$$

Let $v \in \mathscr{U}_{ad}^{\mathbf{L}}$ and $x \in \Gamma_B$. Since $\tilde{P}_C$ is the projection onto C in $\mathbb{R}^d$ and $v(x) + x \in C$ we have

$$\left(\tau(x) - \tilde{P}_C(\tau(x))\right)^T \left(v(x) + x - \tilde{P}_C(\tau(x))\right) \leq 0.$$

Thus it holds

$$\begin{aligned}(u - u^C, v - u^C)_{\mathbf{L}^2(\Gamma_B)} &= \int_{\Gamma_B} (u - u^C)^T (v - u^C)\, \mathrm{dx} \\ &= \int_{\Gamma_B} (\tau - P_C(\tau))^T (v + \mathrm{id} - P_C(\tau))\, \mathrm{dx} \leq 0.\end{aligned}$$

□

In the following we will need the next auxiliary Lemma which summarizes [45, Lemma 8.2] and [45, Lemma 8.20].

Lemma 11
(i) If $M \in \mathscr{L}(Z, X)$ is a surjective operator between Banach spaces, then there exists a constant $c > 0$ such that $\|x'\|_{X^*} \leq c\|M^* x'\|_{Z^*}$ for all $x' \in X^*$.
(ii) The linear operator

$$F : \mathscr{U} \times Y_{ref} \to \mathscr{U} \times Z_{ref},\ F(v, z) = \begin{pmatrix} v \\ E_V(T(u), y)T'(u)v + E_y(T(u), y)z \end{pmatrix}$$

is surjective if and only if $E_y(T(u), y) \in \mathscr{L}(Y_{ref}, Z_{ref})$ is surjective.

Recall that Assumption 3 implies the surjectivity of $E_y(u, S(T(u)))$ for all $u \in \mathscr{U}_{ad}$. We are now ready to prove the announced convergence result. For $\gamma_n \geq 0$ denote by $(u_n, y_n, p_n, \lambda_n) \in \mathscr{U} \times Y_{ref} \times Z_{ref}^* \times \mathscr{U}^*$ a solution of the optimality system (26).

Theorem 9 *Let Assumption 6 hold and $\gamma_n \to \infty$. Furthermore suppose that $\mathbf{C}^\infty(\Gamma_B)$ is dense in $\mathscr{U}$ and the mapping $u \mapsto T'(u) \in \mathscr{L}(\mathscr{U},\mathscr{V})$ is completely continuous. Then any weakly convergent subsequence $(u_k, y_k, p_k, \lambda_k) \rightharpoonup (\bar{u}, \bar{y}, \bar{p}, \bar{\lambda})$ in $\mathscr{U} \times Y_{ref} \times Z^*_{ref} \times \mathscr{U}^*$ converges strongly and $(\bar{u}, \bar{y}, \bar{p}, \bar{\lambda})$ solves the optimality system* (24)*.*

Proof 1. The strong convergence $y_\gamma = S(T(u_\gamma)) \to S(T(\bar{u}))$ follows from the complete continuity of T and the continuity of the design-to-state operator S (compare Assumption 2). Hence $\bar{y} = S(T(\bar{u}))$.

2. Using the complete continuity of T and $u \mapsto T'(u)$ and Assumption 3 we see that

$$\begin{aligned} \hat{J}_u(u_k, y_k) = J_V(T(u_k), y_k)T'(u_k) + \beta u_k \to J_V(T(\bar{u}), \bar{y})T'(\bar{u}) + \beta \bar{u} = \hat{J}_u(\bar{u}, \bar{y}), \\ \hat{J}_y(u_k, y_k) = J_y(T(u_k), y_k) \to J_y(T(\bar{u}, \bar{y}) = \hat{J}_y(\bar{u}, \bar{y}), \end{aligned} \tag{27}$$

and

$$\begin{aligned} E_V(T(u_k), y_k)T'(u_k) \to E_V(T(\bar{u}), \bar{y})T'(\bar{u}), \\ E_y(T(u_k), y_k) \to E_y(T(\bar{u}), \bar{y}). \end{aligned} \tag{28}$$

Combining this with the weak convergence of (p_k, λ_k) and with (26) we conclude that $(\bar{u}, \bar{y}, \bar{p}, \bar{\lambda})$ satisfies the first two equations in (24).

3. Next we show that $(p_k, \lambda_k) \to (\bar{p}, \bar{\lambda})$ in $Z^*_{ref} \times \mathscr{U}^*$. We set $F_{\bar{u}}$ to be the operator defined in Lemma 11 with $u = \bar{u}$ and similarly F_k with $u = u_{\gamma_k}$. The first two equations in (24) and (26) imply

$$F^*_{\bar{u}} \begin{pmatrix} \bar{\lambda} \\ \bar{p} \end{pmatrix} = - \begin{pmatrix} \hat{J}_u(\bar{u}, \bar{y}) \\ \hat{J}_y(\bar{u}, \bar{y}) \end{pmatrix} \text{ and } F^*_k \begin{pmatrix} \lambda_k \\ p_k \end{pmatrix} = - \begin{pmatrix} \hat{J}_u(u_k, y_k) \\ \hat{J}_y(u_k, y_k) \end{pmatrix}.$$

Hence we have

$$\begin{aligned} \begin{pmatrix} \hat{J}_u(u_k, y_k) - \hat{J}_u(\bar{u}, \bar{y}) \\ \hat{J}_y(u_k, y_k) - \hat{J}_y(\bar{u}, \bar{y}) \end{pmatrix} &= F^*_{\bar{u}} \begin{pmatrix} \bar{\lambda} \\ \bar{p} \end{pmatrix} - F^*_k \begin{pmatrix} \lambda_k \\ p_k \end{pmatrix} \\ &= F^*_{\bar{u}} \begin{pmatrix} \bar{\lambda} - \lambda_k \\ \bar{p} - p_k \end{pmatrix} + \left(F^*_{\bar{u}} - F^*_k\right) \begin{pmatrix} \lambda_k \\ p_k \end{pmatrix}. \end{aligned}$$

Using Lemma 11 we see that

$$\begin{aligned} &\left\| \begin{pmatrix} \bar{\lambda} - \lambda_k \\ \bar{p} - p_k \end{pmatrix} \right\|_{\mathscr{U}^* \times Z^*_{ref}} \le c \left\| F^*_{\bar{u}} \begin{pmatrix} \bar{\lambda} - \lambda_k \\ \bar{p} - p_k \end{pmatrix} \right\|_{\mathscr{U}^* \times Y^*_{ref}} \\ &\le c \left\| \begin{pmatrix} \hat{J}_u(u_k, y_k) - \hat{J}_u(\bar{u}, \bar{y}) \\ \hat{J}_y(u_k, y_k) - \hat{J}_y(\bar{u}, \bar{y}) \end{pmatrix} \right\|_{\mathscr{U}^* \times Y^*_{ref}} + c \left\| \left(F^*_k - F^*_{\bar{u}}\right) \begin{pmatrix} \lambda_k \\ p_k \end{pmatrix} \right\|_{\mathscr{U}^* \times Y^*_{ref}}. \end{aligned}$$

Let us adress these two terms separately. Because of (27) we have

$$\left\| \begin{pmatrix} \hat{J}_u(u_k, y_k) - \hat{J}_u(\bar{u}, \bar{y}) \\ \hat{J}_y(u_k, y_k) - \hat{J}_y(\bar{u}, \bar{y}) \end{pmatrix} \right\|_{\mathscr{U}^* \times Y^*_{ref}} \to 0.$$

Addressing the second term (28) implies

$$\|F_k - F_{\bar{u}}\|_{\mathscr{L}(\mathscr{U} \times Y_{ref}, \mathscr{U} \times Z_{ref})} \to 0.$$

Using the properties of the dual operator we obtain

$$\begin{aligned}\left\|\left(F_k^* - F_{\bar u}^*\right)\begin{pmatrix}\lambda_k\\ p_k\end{pmatrix}\right\|_{\mathscr{U}^*\times Y_{ref}^*} &\le \|F_k^* - F_{\bar u}^*\|_{\mathscr{L}(\mathscr{U}^*\times Z_{ref}^*,\mathscr{U}^*\times Y_{ref}^*)}\left\|\begin{pmatrix}\lambda_k\\ p_k\end{pmatrix}\right\|_{\mathscr{U}^*\times Y_{ref}^*}\\ &= \|F_k - F_{\bar u}\|_{\mathscr{L}(\mathscr{U}\times Y_{ref},\mathscr{U}\times Z_{ref})}\left\|\begin{pmatrix}\lambda_k\\ p_k\end{pmatrix}\right\|_{\mathscr{U}^*\times Y_{ref}^*}\end{aligned}$$

Since $\begin{pmatrix}\lambda_k\\ p_k\end{pmatrix}$ is uniformly bounded we conclude $\left\|\begin{pmatrix}\bar\lambda-\lambda_k\\ \bar p - p_k\end{pmatrix}\right\|_{\mathscr{U}^*\times Z_{ref}^*} \to 0.$

4. Testing the first equation in (24) and (26) with $\bar u - u_k$ and subtracting the two equations we see that

$$\begin{aligned}\beta\|\bar u - u_k\|_{\mathscr{U}}^2 = &\langle J_V(T(u_k),y_k)T'(u_k) - J_V(T(\bar u),\bar y)T'(\bar u), \bar u - u_k\rangle_{\mathscr{U}^*,\mathscr{U}}\\ &+\langle (E_V(T(u_k),y_k)T'(u_k))^* p_k - (E_V(T(\bar u),\bar y)T'(\bar u))^*\bar p, \bar u - u_k\rangle_{\mathscr{U}^*,\mathscr{U}}\\ &+\langle\lambda_k - \bar\lambda, \bar u - u_k\rangle_{\mathscr{U}^*,\mathscr{U}}.\end{aligned}$$

Combing now $u_k \rightharpoonup \bar u$, $p_k \to \bar p$, $\lambda_k \to \bar\lambda$, (27), and (28) shows that the right hand side tends to zero and hence $u_k \to \bar u$.

5. Using $\lambda_k \rightharpoonup \bar\lambda$ we have the boundedness of λ_k in $\mathscr{U}^*$, together with $\gamma_k \to \infty$ we see that $(\mathrm{id} + u_k - P_C(\mathrm{id} + u_k), v)_{\mathbf{L}^2(\Gamma_B)} \to 0$ for all $u \in \mathscr{U}$, hence $\mathrm{id} + u_k - P_C(\mathrm{id} + u_k) \to 0$ in $\mathbf{L}^2(\Gamma_B)$ (recall $\mathbf{C}^\infty(\Gamma_B)$ is dense in $\mathscr{U}$). Furthermore $u_k \to \bar u$ in $\mathscr{U}$ implies $u_k \to \bar u$ in $\mathbf{L}^2(\Gamma_B)$ and we conclude $\mathrm{id} + \bar u - P_C(\mathrm{id} + \bar u) = 0$ in $\mathbf{L}^2(\Gamma_B)$, i.e., $\bar u \in \mathscr{U}_{ad}$.

6. Let us now check if $\bar\lambda \in T(\mathscr{U}_{ad}, \bar u)^\circ$. Note that $\|u_k^C - \bar u\|_{\mathbf{L}^2(k_B)} \le \|u_k - \bar u\|_{\mathbf{L}^2(\Gamma_B)}$ by the projection property and $u_k \to \bar u$ in $\mathbf{L}^2(\Gamma_B)$, hence $u_k^C \to \bar u$ in $\mathbf{L}^2(\Gamma_B)$. Now let $d = \mu(v - \bar u) \in T(\mathscr{U}_{ad}, \bar u)$ for some $\mu > 0$ and $v \in \mathscr{U}_{ad}$. We have

$$\begin{aligned}\langle\bar\lambda, d\rangle_{\mathscr{U}^*,\mathscr{U}} &= \lim_{k\to\infty}\langle\lambda_k, d\rangle_{\mathscr{U}^*,\mathscr{U}} = \lim_{k\to\infty}\left(\gamma_k(\tau_k - P_C(\tau_k)), \mu(v - \bar u)\right)_{\mathbf{L}^2(\Gamma_B)}\\ &= \lim_{k\to\infty}\left(\gamma_k(\tau_k - P_C(\tau_k)), \mu(v - u_k^C)\right)_{\mathbf{L}^2(\Gamma_B)}\\ &= \lim_{k\to\infty}\gamma_k\mu\left(u_k - u_k^C, v - u_k^C\right)_{\mathbf{L}^2(\Gamma_B)} \le 0.\end{aligned}$$

Since the polar cone of a cone is the same as the polar cone of its closure we conclude that $\bar\lambda \in T(\mathscr{U}_{ad}, \bar u)^\circ$. □

Remark: Our numerical tests indicate that at least in a ”nice” setting the Lagrange multiplier $\bar\lambda$ enjoys a far better regularity than $\mathscr{U}^*$. If the same behavior is observed in other situations a more thorough investigation of the matter would be interesting. For the present work this is out of scope.

We can immediately carry Theorem 9 over to the case $\gamma_n \to \hat\gamma > 0$:

Corollary 4 *Let Assumption 6 hold, $\hat\gamma > 0$, and $\gamma_n \to \hat\gamma$. Furthermore suppose that the mapping $u \mapsto T'(u) \in \mathscr{L}(\mathscr{U}, \mathscr{V})$ is completely continuous. Then any weakly convergent subsequence $(u_k, y_k, p_k, \lambda_k) \rightharpoonup (u^*, y^*, p^*, \lambda^*)$ converges strongly and the limit $(u^*, y^*, p^*, \lambda^*)$ solves* (26) *with $\gamma = \hat\gamma$.*

Proof p^* solves the last equation of (26), since $u_k \rightharpoonup \bar{u}$ in $\mathscr{U}$ implies $u_k \to \bar{u}$ in $\mathbf{L}^2(\Gamma_B)$. Replacing $(\bar{u}, \bar{y}, \bar{p}, \bar{\lambda})$ by u^*, y^*, p^*, λ^* in the steps 1-4 of the proof of Theorem 9 we see that $(u^*, y^*, p^*, \lambda^*)$ solves also the other equations of (26) and the convergence is strong. □

Finally we extend Theorem 6.

Theorem 10 *Let Assumption 6 hold, $\hat{\gamma} > 0$ and $(u_\gamma, y_\gamma, p_\gamma, \lambda_\gamma)$ be a solution of* (26). *If the mapping $\gamma \mapsto u_\gamma$ is locally Lipschitz continuous for γ close enough to $\hat{\gamma}$, the mappings $\gamma \mapsto y_\gamma$, $\gamma \mapsto p_\gamma$ and $\gamma \mapsto \lambda_\gamma$ are also locally Lipschitz continuous.*

Proof By assumption the solution operators T and S are continuously differentiable, in particular locally Lipschitz continuous. Hence the mapping $\gamma \mapsto y_\gamma = S(T(u_\gamma))$ is locally Lipschitz continuous if $\gamma \mapsto u_\gamma$ is locally Lipschitz. Using again the operators $F_\gamma, F_{\hat{\gamma}}$ as defined in Lemma 11 with $u = u_\gamma$, respectively $u = u_{\hat{\gamma}}$ and copying the calculations in the proof of Theorem 9 we arrive at

$$\begin{aligned}
&\left\| \begin{pmatrix} \lambda_{\hat{\gamma}} - \lambda_\gamma \\ p_{\hat{\gamma}} - p_\gamma \end{pmatrix} \right\|_{\mathscr{U}^* \times Z^*_{ref}} \leq c \left\| F^*_{\hat{\gamma}} \begin{pmatrix} \lambda_{\hat{\gamma}} - \lambda_\gamma \\ p_{\hat{\gamma}} - p_\gamma \end{pmatrix} \right\|_{\mathscr{U}^* \times Y^*_{ref}} \\
&\leq c \left\| \begin{pmatrix} \hat{J}_u(u_{\hat{\gamma}}, y_{\hat{\gamma}}) - \hat{J}_u(u_\gamma, y_\gamma) \\ \hat{J}_y(u_{\hat{\gamma}}, y_{\hat{\gamma}}) - \hat{J}_y(u_\gamma, y_\gamma) \end{pmatrix} \right\|_{\mathscr{U}^* \times Y^*_{ref}} + c \left\| \left(F^*_\gamma - F^*_{\hat{\gamma}} \right) \begin{pmatrix} \lambda_\gamma \\ p_\gamma \end{pmatrix} \right\|_{\mathscr{U}^* \times Y^*_{ref}} .
\end{aligned}$$

Since $\hat{J}$ is continuously differentiable it holds for γ close to $\hat{\gamma}$

$$\left\| \begin{pmatrix} \hat{J}_u(u_{\hat{\gamma}}, y_{\hat{\gamma}}) - \hat{J}_u(u_\gamma, y_\gamma) \\ \hat{J}_y(u_{\hat{\gamma}}, y_{\hat{\gamma}}) - \hat{J}_y(u_\gamma, y_\gamma) \end{pmatrix} \right\|_{\mathscr{U}^* \times Y^*_{ref}} \leq c(\|u_{\hat{\gamma}} - u_\gamma\|_{\mathscr{U}} + \|y_{\hat{\gamma}} - y_\gamma\|_{Y_{ref}}),$$

and due to the continuous differentiability of E

$$\|F_\gamma - F_{\hat{\gamma}}\|_{\mathscr{L}(\mathscr{U} \times Y_{ref}, \mathscr{U} \times Z_{ref})} \leq c(\|u_{\hat{\gamma}} - u_\gamma\|_{\mathscr{U}} + \|y_{\hat{\gamma}} - y_\gamma\|_{Y_{ref}}),$$

for γ close to $\hat{\gamma}$. Copying the computations from the proof of Theorem 9 we see that

$$\begin{aligned}
\left\| \left(F^*_\gamma - F^*_{\hat{\gamma}} \right) \begin{pmatrix} \lambda_\gamma \\ p_\gamma \end{pmatrix} \right\|_{\mathscr{U}^* \times Y^*_{ref}} &\leq \|F_\gamma - F_{\hat{\gamma}}\|_{\mathscr{L}(\mathscr{U} \times Y_{ref}, \mathscr{U} \times Z_{ref})} \left\| \begin{pmatrix} \lambda_\gamma \\ p_\gamma \end{pmatrix} \right\|_{\mathscr{U}^* \times Y^*_{ref}} \\
&\leq c(\|u_{\hat{\gamma}} - u_\gamma\|_{\mathscr{U}} + \|y_{\hat{\gamma}} - y_\gamma\|_{Y_{ref}}) \left\| \begin{pmatrix} \lambda_\gamma \\ p_\gamma \end{pmatrix} \right\|_{\mathscr{U}^* \times Y^*_{ref}} .
\end{aligned}$$

Since $(\lambda_\gamma, p_\gamma)$ is (uniformly) bounded the local Lipschitz continuity of $\gamma \mapsto u_\gamma$ implies the local Lipschitz continuity of the mappings $\gamma \mapsto p_\gamma$ and $\gamma \mapsto \lambda_\gamma$. □

3.7 Convergence rate estimates

In this section we present some estimates on the rate of convergence towards feasibility as well as on the distance between a solution of (9) and a solution of (14). We first recall some facts on oriented distance functions of convex sets. For a detailed treatment we refer e.g. to [9, chapter 7]. The oriented distance function b_A of a set $A \subset \mathbb{R}^d$ is defined by $b_A(x) := d_A(x) - d_{A^C}(x)$ where d_A denotes the usual distance function and A^C the complement of A. b_A is differentiable almost everywhere with $\|\nabla b_A(x)\|_2 \leq 1$ almost everywhere. For all $x \in \mathbb{R}^d \backslash \overline{A}$ it holds $\nabla b_A(x) = \frac{x - \tilde{P}_{\partial A}(x)}{\|x - \tilde{P}_{\partial A}(x)\|_2}$ [9, Theorem 7.3.1]. Further if A is of class $C^{k,l}$ with $k \geq 2, 0 \leq l \leq 1$ or $k = 1, l = 1$, then [9, Theorem 7.8.2]

$$\forall x \in \partial A,\ \exists r > 0 \text{ such that } b_A \in C^{k,l}(\overline{B_r(x)}). \tag{29}$$

If A satisfies in addition $\partial A \cap \overline{\text{Sk}\ (\partial A)} = \emptyset$ the statement holds also for $k = 1$ and $0 \leq l < 1$. Here the skeleton $\text{Sk}(A)$ is defined as $\text{Sk}\ (A) := \{x \in \mathbb{R}^d \mid \tilde{P}_{\overline{A}}(x) \text{ is not unique}\}$. Moreover, in all cases, $\nabla b_A = n$ on ∂A, where n is the unit exterior normal to A. For a convex set A of class $C^{k,l}$ we see immediately that $b_A \in C^{k,l}(\mathbb{R}^d \backslash \overline{A})$.

Assumption 11 *There exists $c_1 > 0$, and a vector field v such that $v(z)^T \nabla b_C(z) \geq c_1$ $\forall z \in \mathbb{R}^d \backslash C$ and $v_\gamma := v \circ (id + u_\gamma) \in \mathscr{U}$, $\forall \gamma > 0$. If the local solutions u_γ of* (14) *are uniformly bounded for all $\gamma > 0$, then v_γ is bounded uniformly in $\mathscr{U}$.*

We are now ready to estimate the quantity δ_{u_γ} in the $\mathbf{L}^1$-norm. Compare [20] in the state-constrained setting.

Lemma 12 *Let Assumptions 6, 11 hold. If the local solutions u_γ of* (14) *are uniformly bounded for all $\gamma > 0$, then there exists a constant $c > 0$ such that*

$$\|\delta_{u_\gamma}\|_{\mathbf{L}^1(\Gamma_B)} \leq c\gamma^{-1}.$$

Proof Let $\gamma > 0$. A local solution u_γ satisfies the optimality system (26). Testing the first equation with $v_\gamma \in \mathscr{U}$ we obtain

$$\langle \hat{J}_u(u_\gamma, y_\gamma), v_\gamma \rangle_{\mathscr{U}^*, \mathscr{U}} + \langle p_\gamma, E_V(T(u_\gamma), y_\gamma) T'(u_\gamma) v_\gamma \rangle_{Z^*_{ref}, Z_{ref}} + \gamma (\delta_{u_\gamma}, v_\gamma)_{\mathbf{L}^2(\Gamma_B)} = 0.$$

Lemma 9 yields that the boundedness of u_γ implies the boundedness of y_γ and p_γ. Furthermore it was shown in the proof that $T(u_\gamma)$ is contained in a compact set $\tilde{\mathscr{V}}$ and y_γ is contained in the compact set $S(\tilde{\mathscr{V}})$. This implies the boundedness of

$$\hat{J}_u(u_\gamma, y_\gamma) = J_u(T(u_\gamma), y_\gamma) + \beta u_\gamma, \text{ and } E_V(T(u_\gamma), y_\gamma) T'(u_\gamma).$$

Hence

$$\begin{aligned}
\gamma (\delta_{u_\gamma}, v_\gamma)_{\mathbf{L}^2(\Gamma_B)} &\leq \|\hat{J}_u(u_\gamma, y_\gamma)\|_{\mathscr{U}^*} \|v_\gamma\|_{\mathscr{U}} \\
&\quad + \|p_\gamma\|_{Z^*_{ref}} \|E_V(T(u_\gamma), y_\gamma) T'(u_\gamma)\|_{\mathscr{L}(\mathscr{U}, Z_{ref})} \|v_\gamma\|_{\mathscr{U}} \\
&\leq c, \quad \text{for all } \gamma > 0.
\end{aligned}$$

Further for all $i = 1,\dots,d$: $\|(\delta_{u_\gamma})_i\|_{L^1(\Gamma_B)} \leq \int_{\Gamma_B} \|\delta_{u_\gamma}(x)\|_2 \,\mathrm{d}\Gamma$. Set $Z := \{x | \delta_{u_\gamma}(x) = 0\}$. We find

$$\begin{aligned}\gamma\|\delta_{u_\gamma}\|_{\mathbf{L}^1(\Gamma_B)} &\leq d\gamma \int_{\Gamma_B} \|\delta_{u_\gamma}(x)\|_2 \,\mathrm{d}\Gamma \leq \tfrac{d\gamma}{c_1} \int_{\Gamma_B\setminus Z} \|\delta_{u_\gamma}(x)\|_2 v_\gamma(x)^T \nabla b_C(x+u(x)) \,\mathrm{d}\Gamma \\ &= \tfrac{d\gamma}{c_1} \int_{\Gamma_B} v_\gamma(x)^T \delta_{u_\gamma}(x) \,\mathrm{d}\Gamma = \tfrac{d\gamma}{c_1}(v_\gamma, \delta_{u_\gamma})_{\mathbf{L}^2(\Gamma_B)} \leq \tfrac{dc}{c_1}.\end{aligned}$$

□

Now we want to have an estimate on the pointwise constraint violation which we measure with $\|\delta_{u_\gamma}\|_{\mathbf{L}^\infty}(\Gamma_B)$. A result from [20] which connects the L^∞, L^1 and $C^{1,\alpha}$ norms of a non-negative $C^{1,\alpha}$ function is employed.

Theorem 11 *Suppose Assumptions 6, 11 hold, $\gamma_n \to \infty$, $\alpha_1, \alpha_2, \alpha_3 > 0$, with $(u_\gamma) \subset \mathbf{C}^{1,\alpha_1}(\Gamma_B)$. Choose $x_\gamma \in \Gamma_B$ such that*

$$\|\delta_{u_\gamma}\|_{\mathbf{L}^\infty(\Gamma_B)} = \|\delta_{u_\gamma}(x_\gamma)\|_2.$$

If there exists $\hat\gamma$ such that $\forall \gamma \geq \hat\gamma$ there exists such a point $x_\gamma \notin \partial\Gamma_B$, Γ_B is a C^{1,α_2}-manifold, and C is of class C^{1,α_3}, satisfying $\partial C \cap \overline{Sk\,(\partial C)} = \emptyset$, then there holds

$$\|\delta_{u_\gamma}\|_{\mathbf{L}^\infty(\Gamma_B)} \leq c\gamma^{-\frac{\alpha+1}{\alpha+d}},$$

where $\alpha = \min\{\alpha_1, \alpha_2, \alpha_3\}$. Otherwise we have

$$\|\delta_{u_\gamma}\|_{\mathbf{L}^\infty(\Gamma_B)} \leq c\gamma^{-\frac{1}{d}}.$$

Proof Note that for any $x \in \Gamma_B$ it holds

$$\|\delta_{u_\gamma}(x)\|_2 = \|\tau_\gamma(x) - \tilde P_C(\tau_\gamma(x))\|_2 = (b_C \circ \tau_\gamma)(x).$$

Let us begin with the first claim. By (29) we have

$$\forall x \in \partial C,\ \exists r > 0 \text{ such that } b_C \in C^{1,\alpha_3}(\overline{B_r(x)}).$$

By assumption there exists a neighborhood $O \subset \mathbb{R}^d$ of x_γ, $r > 0$ and a $\mathbf{C}^{1,\alpha_2}$- diffeomorphism $h_\gamma : \mathbb{R}^{d-1} \supset B_r(0) \to O \cap \Gamma_B$. We define the composition

$$f_\gamma : B_r(0) \to \mathbb{R},\ f_\gamma(y) = (b_C \circ \tau_\gamma \circ h_\gamma)(y).$$

For all $y \in B_r(0)$ with $f_\gamma(y) > 0$ we have $f_\gamma \in C^{1,\alpha}$ by the chain rule ($b_C \in C^{1,\alpha_3},\ \tau_\gamma \in \mathbf{C}^{1,\alpha_1},\ h_\gamma \in \mathbf{C}^{1,\alpha_2}$). Hence we can employ [20, Proposition 2.4] to obtain

$$\|f_\gamma\|_{L^\infty(B_r(0))} \leq c\|f_\gamma\|_{\mathbf{C}^{1,\alpha}(B_r(0))}^{1-\frac{\alpha+1}{\alpha+d}} \|f_\gamma\|_{L^1(B_r(0))}^{\frac{\alpha+1}{\alpha+d}}.$$

We use $\|\delta_{u_\gamma}\|_{\mathbf{L}^\infty(\Gamma_B)} = \|f_\gamma\|_{L^\infty(B_r(0)}$, as well as

$$\begin{aligned}\|f_\gamma\|_{L^1(B_r(0))} &\leq c(h_\gamma)\|\delta_{u_\gamma}\|_{\mathbf{L}^1(\Gamma_B\cap O)} \leq c(h_\gamma)\|\delta_{u_\gamma}\|_{\mathbf{L}^1(\Gamma_B)}, \\ \|f_\gamma\|_{\mathbf{C}^{1,\alpha}(B_r(0)} &\leq c(h_\gamma)\|\delta_{u_\gamma}\|_{\mathbf{C}^{1,\alpha}(\Gamma_B\cap O)} \leq c(h_\gamma)\|\delta_{u_\gamma}\|_{\mathbf{C}^{1,\alpha}(\Gamma_B)},\end{aligned}$$

to see that

$$\|\delta_{u_\gamma}\|_{\mathbf{L}^\infty(\Gamma_B)} \le c(h_\gamma)\|\delta_{u_\gamma}\|_{\mathbf{C}^{1,\alpha}(\Gamma_B)}^{1-\frac{\alpha+1}{\alpha+d}}\|\delta_{u_\gamma}\|_{\mathbf{L}^1(\Gamma_B)}^{\frac{\alpha+1}{\alpha+d}}.$$

Since Γ_B is a fixed C^{1,α_2}-manifold we find an upper bound $c \ge c(h_\gamma)$ for all γ. Finally we invoke Lemma 12 and the uniform boundedness of u_γ in $\mathscr{U} \hookrightarrow \mathbf{C}^{1,\alpha_1}$ to obtain $\|\delta_{u_\gamma}\|_{\mathbf{L}^\infty(\Gamma_B)} \le c\gamma^{-\frac{\alpha+1}{\alpha+d}}$.
The second claim follows similarly from [20, Proposition 2.4] and the subsequent remark. Note that now $f_\gamma \in C^{0,1}$ since $b_C \in C^{0,1}$ and $h_\gamma \in \mathbf{C}^1$ by the general assumption that Γ_B is a C^1-manifold. □

Remark: In most cases we expect that this worst case estimate is not sharp. In our example we observe $\|\delta_{u_\gamma}\|_{\mathbf{L}^\infty(\Gamma_B)} \approx \gamma^{-1}$. See the discussion in [20] on convergence rates and their dependence on the structure of the Lagrange multiplier in the optimum in the state-constrained setting.

Finally we show an estimate on the distance between $\bar{u}$ and u_γ.

Theorem 12 *Let Assumptions 6, 7, 11 hold, and $\bar{u} \in \mathscr{U}_{ad}$ be a local solution of* (9) *in which Assumption 9 is satisfied with $\alpha, \delta > 0$. Further suppose there exists a sequence of local solutions (u_γ) of* (14) *with $u_\gamma \to \bar{u}$ in $\mathscr{U}$, and $\|\delta_{u_\gamma}\|_{\mathbf{L}^\infty(\Gamma_B)} \le c\gamma^{-s}$ for $s > 0$ and $\gamma \to \infty$. Finally denote $V : \gamma \mapsto j_\gamma(u_\gamma)$, then it holds*

$$0 \le j(\bar{u}) - V(\gamma) \le c\gamma^{-s}.$$

Furthermore there exists $\hat{\gamma} > 0$ such that for all $\forall \gamma \ge \hat{\gamma}$ we have

$$\|\bar{u} - u_\gamma\|_{\mathscr{U}} \le c\gamma^{-\frac{s}{2}}.$$

Proof We obtained in Theorem 7 that $V'(\gamma) = \frac{1}{2}\|\delta_{u_\gamma}\|^2_{\mathbf{L}^2(\Gamma_B)}$ for all γ such that $\|\bar{u} - u_\gamma\|_{\mathscr{U}} < \delta$. So using the estimate $\|v\|^2_{L^2} \le \|v\|_{L^1}\|v\|_{L^\infty}$, and Lemma 12 we obtain $V'(\gamma) \le c\gamma^{-1-s}$. Now let $\gamma_2 > \gamma_1$. Since $V(\cdot)$ is differentiable we obtain

$$V(\gamma_2) - V(\gamma_1) = \int_{\gamma_1}^{\gamma_2} V'(\gamma)\,\mathrm{d}\gamma \le \int_{\gamma_1}^{\gamma_2} c\gamma^{-1-s}\,\mathrm{d}\gamma = c(-\gamma_2^{-s} + \gamma_1^{-s}),$$

with c independant of γ. Hence passing to the limit $\gamma_2 \to \infty$ it holds (recall (16))

$$j(\bar{u}) - V(\gamma_1) = \lim_{\gamma_2 \to \infty} j_{\gamma_2}(u_{\gamma_2}) - V(\gamma_1) = \lim_{\gamma_2 \to \infty} V(\gamma_2) - V(\gamma_1) \le c\gamma_1^{-s}.$$

By Theorem 5 Assumption 8 is satisfied in $\bar{u}$ for any $\gamma > 0$. Choose $0 < r_0 < \frac{\delta}{2}$. For all $\gamma > 0$ such that $\|\bar{u} - u_\gamma\|_{\mathscr{U}} \le r_0$ it holds

$$\langle H_\gamma(u)v, v\rangle_{\mathscr{U}^*,\mathscr{U}} \ge \alpha\|v\|^2_{\mathscr{U}}, \quad \forall v \in \mathscr{U},\ \forall H_\gamma(u) \in \partial j'_\gamma(u), \quad \forall u \in B_{\mathscr{U}}(u_\gamma, r_0).$$

As we showed in Lemma 6 this implies

$$j_\gamma(u) \ge j_\gamma(u_\gamma) + \frac{\alpha}{2}\|u_\gamma - u\|^2_{\mathscr{U}} + R^u(\gamma),\ \forall u \in B_{\mathscr{U}}(u_\gamma, r_0),$$

where $R^u(\gamma) = o(\|u-u_\gamma\|^2_{\mathscr{U}})$. In particular it holds

$$\frac{\alpha}{2}\|\bar{u}-u_\gamma\|^2_{\mathscr{U}} + R^{\bar{u}}(\gamma) \le j_\gamma(\bar{u}) - j_\gamma(u_\gamma) = j(\bar{u}) - V(\gamma) \le c\gamma^{-s}.$$

Thus for γ big enough, i.e. $\|\bar{u}-u_\gamma\|_{\mathscr{U}}$ small enough, there exists a $c>0$ such that

$$\|\bar{u}-u_\gamma\|^2_{\mathscr{U}} \le c\gamma^{-s}.$$

This concludes our proof. □

4 Example: potential flow through a channel

We want to optimize a frictionless, irrotational and incompressible flow through a channel. A model for this situation is the potential flow, where the gradient of the potential corresponds to the velocity of the flow [25]. The application of shape optimization methods to potential flows in a channel were investigated by various authors, [1],[6],[38], see also [26] for a similar setting. We study the problem

$$\begin{aligned} \min_{\Omega\in\mathscr{O}_{ad},\hat{y}\in H^1(\Omega)} \quad & \int_Q \|\nabla\hat{y} - v_d\|_2^2 \, \mathrm{d}\tilde{x} \\ \text{s.t.} \quad & -\Delta\hat{y} = 0 \quad \text{in } \Omega, \\ & \hat{y} = y_\Gamma \quad \text{on } \Gamma_D, \\ & \tfrac{\partial}{\partial n}\hat{y} = 0 \quad \text{on } \Gamma_N, \end{aligned} \tag{30}$$

where $\mathscr{O}_{ad} \subset \{\Omega \subset \mathbb{R}^d | Q \subset \Omega\}$ denotes a suitable set of admissible domains, $\partial\Omega = \Gamma_D \dot\cup \Gamma_N$, v_d is the desired velocity of the flow in some pre-described region $Q \subset \mathbb{R}^d$, $v_d : Q \to \mathbb{R}^d$, $\hat{y}$ is the potential of the flow through the channel (depending on Ω), and $y_\Gamma \in H^1(\Gamma_D)$. We consider the two dimensional case $d=2$. We define Ω_{ref} to be the rectangle from Figure 1. Assuming a symmetric channel design we study only one half of the channel, where the left (Γ_1) and right (Γ_3) boundary are the entry and exit respectively, the lower boundary (Γ_4) corresponds to the middle axis of the channel (homogenous Neumann boundary) and the upper boundary (Γ_2) corresponds to the wall of the channel (also homogenous Neumann boundary). We fix the region Q at the exit of the channel. The fixed part of the boundary is denoted by $\Gamma_F = \Gamma_{22} \cup \Gamma_3 \cup \Gamma_4$. The free part of the boundary is thus $\Gamma_1 \cup \Gamma_{21}$ with $\Gamma_{21} = \Gamma_2 \backslash Q$. In order to reduce the number of freedoms and enhance numerical stability only horizontal displacements are allowed for the design boundary nodes. Furthermore the displacement of the nodes of the entrance Γ_1 is given by the scaled displacement of the corner node $x^* \in \Gamma_1 \cap \Gamma_2$. Hence we set $\Gamma_B = \Gamma_{21}$ and it suffices to study geometric constraints of the form $\mathscr{U}_{ad} = \{u = (0,u_2)^T \in \mathbf{H}^2(\Gamma_B) \mid a \le u_2 \le b\}$, for some bounds $a \le b$.

We set $\hat{y} = \tilde{y} + \tilde{y}_\Gamma$, where $\tilde{y}_\Gamma$ is the extension of y_Γ onto Ω and search for $\tilde{y} \in H^1_D(\Omega) := \{\tilde{w} \in H^1(\Omega) \mid \tilde{w} = 0 \text{ on } \Gamma_D\}$ such that the weak formulation of the state equation holds. Identifying $H^1_D(\Omega)^*$ with $H^1_D(\Omega)$ we have:

$$\begin{aligned} &\bar{E} : \{(\Omega,\tilde{y}) \mid \Omega \in \mathscr{O}_{ad}, \tilde{y} \in H^1_D(\Omega)\} \to \{\tilde{z} \mid \Omega \in \mathscr{O}_{ad}, \tilde{z} \in H^1_D(\Omega)\} \\ &\forall \tilde{h} \in H^1_D(\Omega) : \ (\bar{E}(\Omega,\tilde{y}),\tilde{h})_{H^1(\Omega)} = \int_\Omega \nabla_{\tilde{x}}\tilde{y}^T \nabla_{\tilde{x}}\tilde{h} \, \mathrm{d}\tilde{x} + \int_\Omega \nabla_{\tilde{x}}\tilde{y}_\Gamma^T \nabla_{\tilde{x}}\tilde{h} \, \mathrm{d}\tilde{x}. \end{aligned}$$

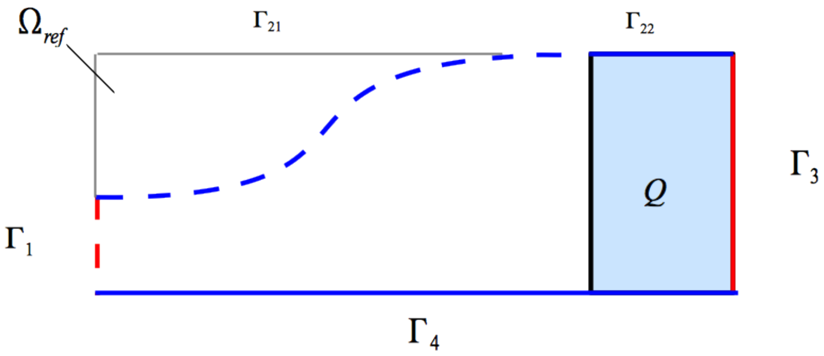

Fig. 1 The domain Ω

Further we set

$$\bar{J}: \{(\Omega,\tilde{y}) \mid \Omega \in \mathscr{O}_{ad}, \tilde{y} \in H^1_D(\Omega)\} \to \mathbb{R}$$
$$\bar{J}(\Omega,\tilde{y}) = \int_Q \|\nabla(\tilde{y}+\tilde{y}_\Gamma) - v_d\|_2^2 \, \mathrm{d}\tilde{x}.$$

In the following we will consider the weak form of the test problem:

$$\begin{aligned} \min_{\Omega \in \mathscr{O}_{ad}, \tilde{y} \in H^1_D(\Omega)} \quad & \bar{J}(\Omega,\tilde{y}) \\ \text{s.t. } & \bar{E}(\Omega,\tilde{y}) = 0 \text{ in } H^1_D(\Omega), \end{aligned} \tag{31}$$

with $\bar{J}: \mathscr{O}_{ad} \times H^1_D(\Omega) \to \mathbb{R}$.

4.1 Transformed shape optimization problem

Since $Q \subset \Omega \; \forall \Omega \in \mathscr{O}_{ad}$ we specify the extension operator T such that $T(u)(x) = 0$, for all $x \in Q$. Recall $\tau = \mathrm{id} + T(u) : \mathbb{R}^d \to \mathbb{R}^d$. Transforming the shape optimization problem to the reference domain Ω_{ref} yields

$$\begin{aligned} \min_{u \in \mathscr{U}, y \in H^1_{D,ref}} \quad & J(y) \\ \text{s.t. } & E(T(u),y) = 0 \text{ in } H^1_{D,ref}, \\ & u \in \mathscr{U}_{ad}. \end{aligned} \tag{32}$$

where $J : H^1_{D,ref} \to \mathbb{R}$, $J(y) = \int_Q \|\nabla(y+y_\Gamma) - v_d\|_2^2 \, \mathrm{dx}$ and $E : \mathscr{V} \times H^1_{D,ref} \to H^1_{D,ref}$,

$$\begin{aligned} (E(T(u),y),h)_{H^1_{ref}} = & \int_{\Omega_{ref}} \nabla y^T \tau'^{-1} \tau'^{-T} \nabla h |\det(\tau')| \, \mathrm{dx} \\ & + \int_{\Omega_{ref}} \nabla y_\Gamma^T \tau'^{-1} \tau'^{-T} \nabla h |\det(\tau')| \, \mathrm{dx}. \end{aligned}$$

As before we add the regularization term $\frac{\beta}{2}\|u\|_{\mathscr{U}}^2$ with $\beta > 0$ to the objective. We arrive at

$$\begin{aligned} \min_{u\in\mathscr{U},y\in H^1_{D,ref}} \quad & \hat{J}(u,y) = J(y) + \frac{\beta}{2}\|u\|_{\mathscr{U}}^2, \\ & \text{s.t. } E(T(u),y) = 0 \text{ in } H^1_{D,ref},\ u \in \mathscr{U}_{ad}. \end{aligned} \tag{33}$$

Let us now check the assumptions made in the previous sections. We employ the linear elasticity equation to map the boundary displacement u to a domain displacement V and choose $\mathscr{U} := \{u \in \mathbf{H}^2(\Gamma_{21}) \mid u(x^\dagger) = 0, x^\dagger = \Gamma_{21}\cap\Gamma_{22}\}$, $\mathscr{V} = \mathbf{W}^{1,\infty}_{ref}$.

4.2 The linear elasticity equation

In this section we want to demonstrate how the linear elasticity equation can be used to map the boundary displacement $u \in \mathscr{U}_{ad}$ to a domain displacement $V \in \mathbf{W}^{1,\infty}_{ref}$. In fact we first only compute V on $\Omega_B := \Omega_{ref}\backslash\overline{Q}$ so that $Tu(x) = 0$ for all $x \in Q$ is guaranteed. The linear elasticity equation on Ω_B (with volume forces set to zero) can be posed as

$$\begin{aligned} KV := (\lambda+\mu)(\nabla \mathrm{div} V)^T + \mu\Delta V &= 0 \ \text{ in } \Omega_B, \\ V &= u \ \text{ on } \Gamma_{21}, \\ V &= \hat{u} \ \text{ on } \Gamma_1, \\ V &= 0 \ \text{ on } \partial\Omega_B\backslash(\Gamma_1\cup\Gamma_{21}). \end{aligned} \tag{34}$$

Here $\lambda = \frac{\nu E}{(1+\nu)(1-2\nu)} > 0$, $\mu = \frac{E}{2(1+\nu)} > 0$ are the Lamé parameters, where E is Young's modulus and ν the Poisson ratio. The displacement at the entrance Γ_1 is obtained via $\hat{u}(x) = (0, \frac{x_2}{x_2^*}u_2(x^*))^T$, where $x^* \in \Gamma_1\cap\Gamma_2$ is the corner node.

Corollary 5 *Suppose $p > d$, $\mathscr{U} \hookrightarrow \mathbf{W}^{2-\frac{1}{p},p}(\Gamma_B)$, and $u \in \mathscr{U} \Rightarrow u(x^\dagger) = 0, x^\dagger = \Gamma_{21}\cap\Gamma_{22}$. Denote the solution of* (34) *by $V(u)$ and set $T : \mathscr{U} \to \mathbf{W}^{1,\infty}_{ref}$, $Tu|_{\Omega_B} = V(u)$, $Tu|_Q = 0$. Then T is completely continuous and twice continuously differentiable.*

Proof In [13, Theorem 6.1] it is shown, that for the solution of (34) with boundary data $u \in \mathbf{W}^{2-\frac{1}{p},p}(\Gamma_{21})$, $\hat{u} \in \mathbf{W}^{2-\frac{1}{p},p}(\Gamma_1)$ it holds $V(u) \in \mathbf{W}^{2,p}(\Omega_B)$ if the characteristic equation

$$\sin^2(z\frac{\pi}{2}) = \frac{(\lambda+\mu)^2}{(\lambda+3\mu)^2}z^2 \tag{35}$$

has no roots in the strip $0 < \mathrm{Re}(z) < \frac{2}{q}$, where $\frac{1}{p} + \frac{1}{q} = 1$. Here we depend heavily on our special case (we have only right angles). Indeed the proof of [14, Lemma 3.3.1] can be extended to show that (35) has no roots in the strip $0 < \mathrm{Re}(z) < 2$. Hence the solution operator $V : \mathbf{W}^{2-\frac{1}{p},p}(\Gamma_{21}) \to \mathbf{W}^{2,p}_{ref}$ exists. It is linear and, as we will now show using the closed graph theorem, also continuous. We want to show that the graph $(u, V(u)) \in \mathbf{W}^{2-\frac{1}{p},p}(\Gamma_{21}) \times \mathbf{W}^{2,p}_{ref}$ is closed. From this the continuity of V follows immediately. Let us consider a sequence $(u_n) \subset \mathscr{U}$ with $(u_n, V(u_n)) \to (u,y) \in \mathbf{W}^{2-\frac{1}{p},p}(\Gamma_{21}) \times \mathbf{W}^{2,p}_{ref}$. It is clear that $Ky = \lim_{n\to\infty} KV(u_n) = 0$. Denote by

$g:\mathbf{W}^{2,p}_{ref}\to\mathbf{W}^{2-\frac{1}{p},p}(\Gamma_{21})$ the trace operator. By [12, Theorem 1.5.2.1] g is continuous. It remains to show $\|g(y)-u\|_{\mathbf{W}^{2-\frac{1}{p},p}(\Gamma_{21})}=0$. Indeed we have

$$\|g(y)-u\|_{\mathbf{W}^{2-\frac{1}{p},p}(\Gamma_{21})}\le\|g(y)-g(V(u_n))\|_{\mathbf{W}^{2-\frac{1}{p},p}(\Gamma_{21})}+\|g(V(u_n))-u\|_{\mathbf{W}^{2-\frac{1}{p},p}(\Gamma_{21})}.$$

By the continuity of g the first term goes to zero and since $g(V(u_n))=u_n$ the second term also.

With the well known imbedding

$$\mathbf{W}^{2,p}(\Omega_B)\hookrightarrow\hookrightarrow\mathbf{C}^{1,\alpha}(\overline{\Omega_B})$$

for $p>d$ and $0<\alpha<1-\frac{d}{p}$ [12, section 1.4.4.] the mapping $u\mapsto V(u)\in\mathbf{C}^{1,\alpha}(\overline{\Omega_B})$ is completely continuous. Finally if we extend $V(u)$ by 0 onto Ω_{ref} we obtain $Tu\in\mathbf{W}^{1,\infty}_{ref}$. Hence the claim follows. □

4.3 Checking the assumptions

Theorem 13 *Let $\mathscr{V}=\mathbf{W}^{1,\infty}_{ref}$, $\mathscr{U}=\{u\in\mathbf{H}^2(\Gamma_{21})\mid u(x^\dagger)=0,x^\dagger=\Gamma_{21}\cap\Gamma_{22}\}$, then Assumption 2.1.-5. and 3 are satisfied.*

Proof Assumption 2.1, 2.2. are satisfied by construction. By the Sobolev imbedding theorem we have $\mathbf{H}^2(\Gamma_B)\hookrightarrow\mathbf{W}^{2-\frac{1}{p},p}(\Gamma_B)$ for $d=2<p\le4$, hence Corollary 5 yields the desired properties of T. Murat and Simon showed in [33, Lemma 4.2, 4.3] that $\forall V\in\mathscr{V}_{feas}$

$\mathbf{W}^{1,\infty}(\mathbb{R}^d)\ni V\mapsto|\det(\tau'))|\in L^\infty$ and
$\mathbf{W}^{1,\infty}(\mathbb{R}^d)\ni V\mapsto(\tau')^{-1}\in\mathbf{L}^\infty$ are twice continuously differentiable.

Hence $E:\mathbf{W}^{1,\infty}_{ref}\times H^1_{D,ref}\to H^1_{D,ref}$ and $J:H^1_{D,ref}\to\mathbb{R}$ are twice continuous differentiable. Furthermore the objective J is bounded from below by 0. This shows Assumption 2.3, 2.5 as well as Assumption 3.1.

We now choose a bounded set $\mathscr{V}_{cf}\subset\mathscr{V}_{feas}$. For fixed $V\in\mathscr{V}_{cf}$ and $\tau=\mathrm{id}+V$ we define the bilinear form

$$a:H^1_{D,ref}\times H^1_{D,ref}\to\mathbb{R},\quad a(y,h):=\int_{\Omega_{ref}}\nabla y^T\tau'^{-1}\tau'^{-T}\nabla h\det(\tau')\,\mathrm{d}x.$$

Since $\mathscr{V}_{cf}$ is bounded [33, Remark 4.1] yields that $a(\cdot,\cdot)$ is bounded and coercive. Hence the prerequisites of Lax-Milgram are fulfilled and we can conclude that there exists a unique bounded solution $y\in H^1_{D,ref}$ to the variational equation

$$a(y,h)=-a(y_\Gamma,h)\quad\forall h\in H^1_{D,ref}.$$

Thus there exists a bounded solution operator of the state equation

$$S:\mathscr{D}_{ad}\to H^1_{D,ref},\ V\mapsto y=S(V).$$

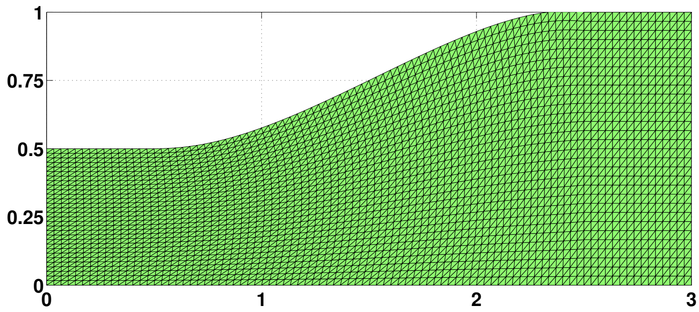

Fig. 2 The optimal domain

Since $\mathscr{V}_{cf} \subset \mathscr{V}_{feas}$ and $\mathscr{V}_{feas}$ is open (c.f. [33, Lemma 2.4]) we can extend the above calculations to an open neighborhood $\hat{\mathscr{V}}_{cf} \subset \mathscr{V}_{feas}$ of $\mathscr{V}_{cf}$ to obtain the bounded design-to-state operator $S : \hat{\mathscr{V}}_{cf} \to H^1_{D,ref}$.

We proceed to show that $E_y(V,S(V)) \in \mathscr{L}(H^1_{D,ref},(H^1_{D,ref})^*)$ is continuously invertible for every $V \in \hat{\mathscr{V}}_{cf}$. Let $V \in \hat{V}_{cf}$. We already proved that for any $f \in (H^1_{D,ref})^*$ there exists a unique solution $z \in H^1_{D,ref}$ to

$$\begin{aligned}&\left\langle \langle E_y(V,S(V)),z\rangle_{(H^1_{D,ref})^{**},(H^1_{D,ref})^*}, h\right\rangle_{(H^1_{ref})^*,H^1_{D,ref}}\\ &= a(z,h) \stackrel{!}{=} \langle f,h\rangle_{(H^1_{D,ref})^*,H^1_{D,ref}}, \quad \forall h \in H^1_{D,ref}.\end{aligned}$$

A direct computation shows that the mapping $f \mapsto z$ is linear. Furthermore Lax-Milgram provides boundedness of the mapping which implies its continuity. □

Assumption 4 can in general not be verified. Assumption 5.1. is satisfied by construction, Assumption 5.2.-3. are satisfied for suitable choices of C and y_γ. As already mentioned Assumption 7 can also be guaranteed to hold by a suitable choice of C.

4.4 Numerical results

γ	$j_\gamma(u_\gamma)$	$\lVert\tau - P_C(\tau)\rVert_{\mathbf{L}^\infty(\Gamma_B)}$	$\lVert u_{\gamma_{max}} - u_\gamma\rVert_{H^2(\Gamma_B)}$	iterations
0	2,8761113e-02	4,79e-02	1,86e-01	3
1,00e01	2,6600350e-02	5,11e-03	1,30e-02	3
2,00e02	2,6660391e-02	2,71e-04	7,79e-04	3
3,60e03	2,6663642e-02	1,54e-05	4,70e-05	1
5,83e04	2,6663823e-02	9,55e-07	2,75e-06	1
8,50e05	2,6663834e-02	6,56e-08	0	1

We implemented the proposed Moreau-Yosida penalty method for the introduced example in MATLAB [28]. The discretization uses piecewise linear finite elements.

Note that we do not use a conforming discretization of $\mathscr{U}$, but only an approximation of the discrete H^2 scalar product. As already mentioned only horizontal displacements are allowed for the design boundary nodes in order to reduce the number of degrees of freedoms and enhance numerical stability. The displacement of the nodes of the entrance Γ_1 is given by the scaled displacement of the corner node $x^* \in \Gamma_1 \cap \Gamma_2$. This ensures also a reasonably uniform discretization of the design boundary. The admissible set is given by

$$\mathscr{U}_{ad} := \{(0,u_2) \in \mathbf{H}^2(\Gamma_B) \mid u_2 \geq a\},$$

which enforces a minimum diameter of the channel. Note that the projection $\tilde{P}_C$ is given by the maximum function which is known to be semismooth, so Assumption 7 is satisfied. We use a dampened globalized semismooth Newton method to solve the regularized subproblems. The reduced gradient is computed with the adjoint approach. Recall that the gradient is the Riesz representative of j'_γ. We solve the Newton system using an iterative Krylov subspace method, where we solve two additional adjoint equations for each evaluation of $j''_\gamma(u)v$. Once the Newton direction is computed we check the angle condition. Either the Newton direction is used or the negative gradient. Finally we use the Armijo rule to choose the step size.

As already mentioned in section 3.5 we steer the γ-update with the help of the model

$$m_k(\gamma) = C_{1,k} - \frac{C_{2,k}}{(1+\gamma)^r},$$

with $C_{1,k} \in \mathbb{R}$ and $C_{2,k}, r > 0$. The constants $C_{1,k}$ and $C_{2,k}$ are updated in each iteration via the conditions $m_k(\gamma_k) = j_{\gamma_k}(u_k)$ and $m'_k(\gamma_k) = V'(\gamma_k)$. Following [17] we choose γ^+_{k+1} such that

$$|j_{\gamma_k}(u_k) + V'(\gamma_k)(\gamma^+_{k+1} - \gamma_k) - m_k(\gamma^+_{k+1})| \leq \alpha_1 |j_{\gamma_k}(u_k) - j_{\gamma_{k-1}}(u_{k-1})|,$$

and set $\gamma_{k+1} = \max\{\gamma^+_{k+1}, (1+\alpha_2)\gamma_k\}$, to ensure a minimum increase. Furthermore we solve the regularized problems for small γ only approximately and increase the accuracy along with γ. We use the stopping criterion

$$\|j'_{\gamma_k}(u_k)\|_{\mathscr{U}} \leq \frac{\alpha_3}{\gamma_k^r}.$$

Here $0 < \alpha_1, \alpha_2, \alpha_3 < 1$ are some constant parameters.

We present the results for the geometric shape constraint $u \geq -0.5$. The optimal domain is shown in Figure 4.4. Observe the good performance of the semismooth Newton method (last column in Table 4.4 shows the number of iterations until convergence). In particular for the last three loops of the penalty method only one step was needed. This demonstrates an efficient rule for the γ-update. As already announced in the theory sections we observe an order of convergence of $O(\gamma^{-1})$ for the constraint violation $\|\delta_{u_\gamma}\|_{\mathbf{L}^\infty(\Gamma_{21})} = \|\tau - P_C(\tau)\|_{\mathbf{L}^\infty(\Gamma_{21})}$ and the distance to the (approximate) optimum $\|u_{\gamma_{max}} - u_\gamma\|_{H^2(\Gamma_{21})}$, compare Figure 4.4. This can be explained by an increased regularity of the Lagrange multiplier $\bar{\lambda} \in \mathscr{U}^*$. If a similar behavior is observed in other applications this should be investigated further.

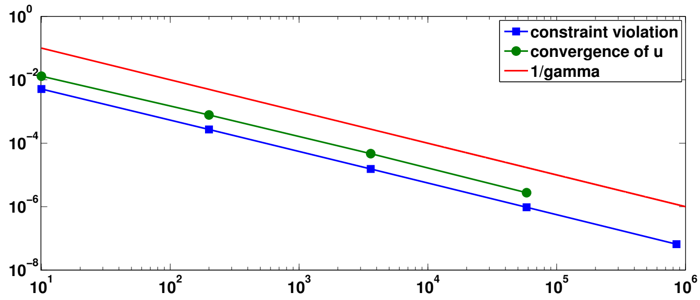

Fig. 3 $\|\tau - P_C(\tau)\}\|_{L^\infty(\Gamma_{21})}$ (blue) and $\|u_{\gamma_{max}} - u_\gamma\|_{H^2(\Gamma_{21})}$ (green)

References

1. Angrand, F.: Optimum design for potential flows. International Journal for Numerical Methods in Fluids **3**, 265–282 (1983)
2. Antonietti, P., Borzì, A., Verani, M.: Multigrid shape optimization governed by elliptic pdes. SIAM Journal on Control and Optimization **51**(2), 1417–1440 (2013)
3. Brandenburg, C.: Adjoint-based adaptive multilevel shape optimization based on goal-oriented error estimators for the instationary navier-stokes equations. Ph.D. thesis, TU Darmstadt (2011)
4. Brandenburg, C., Lindemann, F., Ulbrich, M., Ulbrich, S.: A continuous adjoint approach to shape optimization for navier-stokes flow. In: K. Kunisch, J. Sprekels, G. Leugering, F. Tröltzsch (eds.) Optimal Control of Coupled Systems of Partial Differential Equations, *International Series of Numerical Mathematics*, vol. 158, pp. 35–56. Birkhäuser Basel (2009)
5. Brandenburg, C., Lindemann, F., Ulbrich, M., Ulbrich, S.: Advanced numerical methods for pde constrained optimization with application to optimal design in navier stokes flow. In: G. Leugering, S. Engell, A. Griewank, M. Hinze, R. Rannacher, V. Schulz, M. Ulbrich, S. Ulbrich (eds.) Constrained Optimization and Optimal Control for Partial Differential Equations, *International Series of Numerical Mathematics*, vol. 160, pp. 257–275. Birkhäuser Basel (2011)
6. Butt, R.: Optimal shape design for a nozzle problem. The Journal of the Australian Mathematical Society, Series B pp. 71–86 (1993)
7. Casas, E., Tröltzsch, F.: A general theorem on error estimates with application to a quasilinear elliptic optimal control problem. Computational Optimization and Applications **53**(1), 173–206 (2012)
8. Clarke, F.H.: Optimization and Nonsmooth Analysis. John Wiley (1983)
9. Delfour, M.C., Zolésio, J.P.: Shapes and Geometries, second edn. Society for Industrial and Applied Mathematics (2011)
10. Geiger, C., Kanzow, C.: Theorie und Numerik restringierter Optimierungsaufgaben. Springer-Lehrbuch Masterclass. Springer Berlin Heidelberg (2002)
11. Goto, Y., Fujii, N.: Second-order numerical method for domain optimization problems. Journal on Optimization Theory and Applications **67**(3) (1990)
12. Grisvard, P.: Elliptic problems in nonsmooth domains. Pitman London (1985)
13. Grisvard, P.: Singularités en elasticité. Archive for Rational Mechanics and Analysis **107**(2), 157–180 (1989)
14. Grisvard, P.: Singularities in Boundary Value Problems. Masson, Paris (1992)
15. Guillaume, P., Masmoudi, M.: Computation of high order derivatives in optimal shape design. Numerische Mathematik **67**(2), 231–250 (1994)
16. Haslinger, J., Mäkinen, R.A.E.: Introduction to Shape Optimization: Theory, Approximation, and Computation. Series on Advances in Design and Control. SIAM (2003)
17. Hintermüller, M., Kunisch, K.: Feasible and noninterior path following in constrained minimization with low multiplier regularity. SIAM Journal on Control and Optimization **45**(4), 1198–1221 (2006)
18. Hintermüller, M., Kunisch, K.: Path-following methods for a class of constrained minimization problems in function space. SIAM Journal on Optimization **17**(1), 159–187 (2006)
19. Hintermüller, M., Ring, W.: A second order shape optimization approach for image segmentation. SIAM Journal on Applied Mathematics **64**(2), 442–467 (2004)

20. Hintermüller, M., Schiela, A., Wollner, W.: The length of the primal-dual path in moreau-yosida-based path-following for state constrained optimal control. Hamburger Beiträge zur Angewandten Mathematik **3** (2012)
21. Hinze, M., Pinnau, R., Ulbrich, M., Ulbrich, S.: Optimization with PDE Constraints. Springer (2009)
22. Hoppe, R.H.W., Linsenmann, C., Antil, H.: Adaptive path following primal dual interior point methods for shape optimization of linear and nonlinear stokes flow problems. In: I. Lirkov, S. Margenov, J. Waśniewski (eds.) Large-Scale Scientific Computing, *Lecture Notes in Computer Science*, vol. 4818, pp. 259–266. Springer Berlin Heidelberg (2008)
23. Ito, K., Kunisch, K.: Semi-smooth newton methods for state-constrained optimal control problems. Systems & Control Letters **50**(3), 221 – 228 (2003)
24. Kiniger, B., Vexler, B.: A priori error estimates for finite element discretization of a shape optimization problem. Mathematical Modelling and Numerical Analysis (subdmitted 2011)
25. Kluge, G., Stephani, H.: Theoretische Mechanik. Spektrum Akademischer Verlag (1995)
26. Laumen, M.: Newton's method for a class of optimal shape design problems. SIAM Journal on Optimization **10**(2), 503–533 (2000)
27. Lindemann, F.: Theoretical and numerical aspects of shape optimization with navier-stokes flows. Ph.D. thesis, Technische Universität München, München (2012)
28. The MathWorks, Inc., Natick, Massachusetts, United States: MATLAB R2012a
29. Meyer, C., Rösch, A., Tröltzsch, F.: Optimal control of pdes with regularized pointwise state constraints. Comput. Optim. Appl. **33**(2-3), 209–228 (2006)
30. Meyer, C., Yousept, I.: Regularization of state-constrained elliptic optimal control problems with nonlocal radiation interface conditions. Computational Optimization and Applications **44**(2), 183–212 (2009)
31. Mohammadi, B., Pironneau, O.: Applied Shape Optimization for Fluids. Oxford University Press (2001)
32. Murat, F., Simon, J.: Etudes de problems d'optimal design. Lecture Notes in Computer Science **41**, 54–62 (1976)
33. Murat, F., Simon, J.: Sur le contrôl par un domaine géométrique. Tech. rep., Universite P. et M. Curie (Paris IV) (1976)
34. Neitzel, I., Tröltzsch, F.: On convergence of regularization methods for nonlinear parabolic optimal control problems with control and state constraints. Control and Cybernetics **37**(4), 1013–1043 (2008)
35. Nemec, M., Zingg, D.W., Pulliam, T.H.: Multipoint and multiobjective aerodynamic shape optimization. AIAA Journal **42**(6), 1057–1065 (2004)
36. Novruzi, A., Roche, J.R.: Second order derivatives, newton method, application to shape optimization. Rapport de Recherche 2555, Institut National de Recherche en Informatique et en Automatique (1995)
37. Novruzi, A., Roche, J.R.: Newton's method in shape optimisation: A three-dimensional case. BIT Numerical Mathematics **40**(1), 102–120 (2000)
38. Pironneau, O.: Optimal shape design for elliptic systems. In: System Modeling and Optimization, Lecture Notes in Control and Information Sciences. Springer (1982)
39. Prüfert, U., Tröltzsch, F., Weiser, M.: The convergence of an interior point method for an elliptic control problem with mixed control-state constraints. Comput. Optim. Appl. **39**(2), 183–218 (2008)
40. Rockafellar, R.T., Wets, R.J.B.: Variational analysis, *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, vol. 317. Springer-Verlag, Berlin (1998)
41. Schiela, A.: Barrier methods for optimal control problems with state constraints. SIAM Journal on Optimization **20**(2), 1002–1031 (2009)
42. Schiela, A., Günther, A.: An interior point algorithm with inexact step computation in function space for state constrained optimal control. Numerische Mathematik **119**(2), 373–407 (2011)
43. Sokolowski, J., Zolésio, J.P.: Introduction to Shape Optimization. Series in Computational Mathematic. Springer (1992)
44. Tröltzsch, F.: Regular lagrange multipliers for control problems with mixed pointwise control-state constraints. SIAM Journal on Optimization **15**(2), 616–634 (2005)
45. Ulbrich, M.: Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces. MOS-SIAM series on optimization (2011)