Noname manuscript No.
(will be inserted by the editor)

A path-following inexact Newton method for PDE-constrained optimal control in BV

D. Hafemeyer · F. Mannel

Received: date / Accepted: date

Abstract We study a PDE-constrained optimal control problem that involves functions of bounded variation as controls and includes the TV seminorm of the control in the objective. We apply a path-following inexact Newton method to the problems that arise from smoothing the TV seminorm and adding an H^1 regularization. We prove in an infinite-dimensional setting that, first, the solutions of these auxiliary problems converge to the solution of the original problem and, second, that an inexact Newton method enjoys fast local convergence when applied to a reformulation of the optimality system in which the control appears as implicit function of the adjoint state. We show convergence of a Finite Element approximation, provide a globalized preconditioned inexact Newton method as solver for the discretized auxiliary problems, and embed it into an inexact path-following scheme. We construct a two-dimensional test problem with fully explicit solution and present numerical results to illustrate the accuracy and robustness of the approach.

Keywords optimal control · partial differential equations · TV seminorm · functions of bounded variation · path-following Newton method

Mathematics Subject Classification (2010) 35J70 · 49-04 · 49M05 · 49M15 · 49M25 · 49J20 · 49K20 · 49N60

Dominik Hafemeyer
TU München
Lehrstuhl für Optimalsteuerung, Department of Mathematics
Boltzmannstr. 3, 85748 Garching b. München, Germany
E-mail: dominik.hafemeyer@ma.tum.de

Florian Mannel
University of Graz
Institute of Mathematics and Scientific Computing
Heinrichstr. 36, 8010 Graz, Austria
E-mail: florian.mannel@uni-graz.at

Problem setting and introduction

This work is concerned with the optimal control problem

$$\min_{(y,u)\in H_0^1(\Omega)\times \mathrm{BV}(\Omega)} \underbrace{\frac{1}{2}||y-y_\Omega||^2_{L^2(\Omega)} + \beta|u|_{\mathrm{BV}(\Omega)}}_{=:J(y,u)} \qquad \text{s.t.} \qquad Ay=u, \quad \text{(OC)}$$

where throughout $\Omega \subset \mathbb{R}^N$ is a bounded $C^{1,1}$ domain and $N \in \{1,2,3\}$. The control u belongs to the space of functions of bounded variation $\mathrm{BV}(\Omega)$, the state y lives in $Y := H_0^1(\Omega)$, the parameter β is positive, and $Ay = u$ is a partial differential equation of the form

$$\begin{cases} \mathcal{A}y + c_0 y = u & \text{in } \Omega, \\ y = 0 & \text{on } \partial\Omega \end{cases}$$

with a non-negative function $c_0 \in L^\infty(\Omega)$ and a linear and uniformly elliptic operator of second order in divergence form $\mathcal{A} : H_0^1(\Omega) \to H^{-1}(\Omega)$, $\mathcal{A}y(\varphi) = \int_\Omega \sum_{i,j=1}^N a_{ij}\partial_i y \partial_j \varphi \,\mathrm{d}x$ whose coefficients satisfy $a_{ij} = a_{ji} \in C^{0,1}(\Omega)$ for all $i, j \in \{1, \ldots, N\}$. The specific feature of (OC) is the appearance of the BV seminorm $|u|_{\mathrm{BV}(\Omega)}$ in the cost functional, which favors piecewise constant controls and has recently attracted considerable interest in PDE-constrained optimal control, cf. [7,12,13,15,22,23,24,27,28,30] and the earlier works [16,17]. The majority of these contributions focuses on deriving optimality conditions and studying Finite Element approximations. In contrast, the main focus of this work is on a path-following method. Specifically,

– we propose to smooth the TV seminorm in J and add an H^1 regularization, and we show in an infinite-dimensional setting that the solutions of the resulting auxiliary problems converge to the solution of (OC);
– we present a non-standard reformulation of the optimality conditions of the auxiliary problems and show local convergence of an infinite-dimensional inexact Newton method when applied to this reformulation;
– we derive a practical path-following method that yields accurate solutions for (OC) and illustrate its capabilities in numerical examples for $\Omega \subset \mathbb{R}^2$.

To the best of our knowledge, these aspects have only been investigated partially for optimal control problems that involve the TV seminorm in the objective. In particular, there are few works that address the numerical solution when the measure ∇u is supported in a *two-dimensional* set. In fact, we are only aware of [22], where a doubly-regularized version of the Fenchel predual of (OC) is solved for fixed regularization parameters, but path-following is not applied. We stress that in our numerical experience the two-dimensional case is significantly more challenging than the one-dimensional case. A FeNiCs implementation of our path-following method is available at `https://imsc.uni-graz.at/mannel/publications.php`. It includes all the features that we discuss in section 6, e.g., a preconditioner for the Newton systems, a non-monotone line search globalization, and inexact path-following.

A further contribution of this work is that

– we provide an example of (OC) for $N = 2$ with fully explicit solution.

For the case that ∇u is defined in an interval ($N = 1$) such examples are available, e.g. [13,30], but for $N = 2$ this is new.

Let us briefly address three difficulties associated with (OC). First, the fact that (OC) is posed in the non-reflexive space $\mathrm{BV}(\Omega)$ complicates the proof of existence of optimal solutions. By now it is, however, well understood how to deal with this issue also in more complicated situations, cf. e.g. [13,15].

Second, we notice that $u \mapsto |u|_{\mathrm{BV}(\Omega)}$ is not differentiable. We will cope with this by replacing $|u|_{\mathrm{BV}(\Omega)}$ with a smoothed functional ψ_δ, $\delta \geq 0$, that satisfies $\psi_0(\cdot) = |\cdot|_{\mathrm{BV}(\Omega)}$. The functional ψ_δ that we use for this purpose is well-known, particularly in the imaging community, e.g. [1,20]. However, in most of the existing works the smoothing parameter $\delta > 0$ is fixed, whereas we are interested in driving δ to zero. We will also add the regularizer $\gamma\|u\|^2_{H^1(\Omega)}$, $\gamma \geq 0$, to J and drive γ to zero. For fixed $\gamma, \delta > 0$ the optimal control $\bar{u}_{\gamma,\delta}$ of the smoothed and regularized auxiliary problem turns out to be a $C^{1,\alpha}$ function for some $\alpha > 0$ and we will use this higher regularity in the convergence analysis of (an inexact) Newton's method, for instance to prove that the adjoint-to-control mapping has a locally Lipschitz continuous Fréchet derivative; cf. Theorem 7. In contrast, for $\gamma = 0$ only $\bar{u}_{0,\delta} \in \mathrm{BV}(\Omega)$ can be expected.

Third, numerical experiments show that for standard formulations of the optimality system, e.g. those that result from reduction to the control, path-following Newton methods are not able to sufficiently reduce the smoothing parameter δ. In fact, we have consistently encountered this phenomenon in our previous work [13,21,27,28,30] involving the TV seminorm. As a remedy we propose to eliminate the control from the optimality system by regarding it as an implicit function of the adjoint state, an approach that may be of interest in its own right. Since the control depends nonlinearly on the adjoint state, this increases the computational costs in comparison to standard formulations of the optimality system, e.g. reduction to the control or an all-at-once approach, where the dependencies are linear. On the other hand, the implicit approach enables us to reduce δ far below the levels that we achieved with standard formulations of the optimalty system. In addition, we provide measures that lower the computational burden of this approach.

Let us set our work in perspective with the available literature. We regard it as one of the main contributions that we show on the infinite-dimensional level that the solutions of the auxiliary problems converge to the solution of (OC), cf. section 2.5. The asymptotic convergence for vanishing H^1 seminorm regularization is analyzed in [15, Section 6] for a more general problem than (OC), but the fact that our setting is less general allows us to prove convergence in stronger norms than the corresponding [15, Theorem 10]. The asymptotic convergence for a doubly-regularized version of the predual of (OC) is established in [22, Appendix A], but one of the regularizations is left untouched, so convergence is towards the solution of a regularized problem, not towards the solution of (OC). Next, we demonstrate that an infinite-dimensional inexact Newton method, applied to the aforementioned non-standard reformulation of the optimality system, converges locally for the auxiliary problems. This is non-trivial to prove because the implicit control and the adjoint state are coupled by a quasilinear PDE. A related result is [22, Theorem 3.5], where local q-superlinear convergence of a semismooth Newton method is shown for the doubly-regularized Fenchel predual for fixed regularization parameters. Yet, since we work with a different optimality system, the overlap is rather small.

Turning to the discrete level we provide a Finite Element approximation and demonstrate that the Finite Element solutions of the auxiliary problems converge to the corresponding true solutions. Finite Element approximations for optimal control in BV involving the TV seminorm have also been studied in [7,12,13,15,23,24,27,28,30], but in our assessment the regularization of (OC) that we propose is not covered by these studies.

The BV-term in (OC) favors sparsity in the gradient of the control. Other sparsity promoting control terms that have been studied during recent years are measure norms and L^1–type functionals, e.g., [2,10,11,14,18,19,31,34,45].

TV-regularization is also of significant importance in imaging problems and its usefulness for, e.g., noise removal has long been known [41]. However, the character of imaging problems is substantially different from optimal control problems, for instance because the forward operator in imaging problems is usually cheap to evaluate and non-compact.

This paper is organized as follows. After some preliminaries in section 1, we consider existence, optimality conditions and convergence of solutions in section 2. In section 3 we establish differentiability of the adjoint-to-control mapping, which paves the way for proving local convergence of an inexact Newton method in section 4. Section 5 addresses the Finite Element approximation and its convergence, while section 6 provides the path-following method. Numerical experiments are presented in section 7, including for the test problem with explicit solution. Several technical results such as Hölder continuity of solutions to quasilinear PDEs are deferred to the appendix.

1 Preliminaries

We recall facts about the space $\mathrm{BV}(\Omega)$, introduce an index s, and collect properties of the solution operator of the PDE in (OC).

1.1 Functions of bounded variation

The following statements about $\mathrm{BV}(\Omega)$ can be found in [4, Chapter 3] unless stated otherwise. The space of functions of bounded variation is defined as

$$\mathrm{BV}(\Omega) := \left\{ u \in L^1(\Omega) : \sup_{v \in C_0^1(\Omega)^N, \| |v| \|_\infty \le 1} \int_\Omega u \operatorname{div} v \, \mathrm{d}x < \infty \right\}.$$

Here and throughout, $|\cdot|$ denotes the Euclidean norm, so we are using the isotropic total variation. It can be shown that $u \in \mathrm{BV}(\Omega)$ iff there exists a vector measure $(\partial_{x_1}u,\dots,\partial_{x_N}u)^T = \nabla u \in \mathcal{M}(\Omega)^N$ such that for all $i \in \{1,\dots,n\}$ there holds

$$\int_\Omega \partial_{x_i} uv \,\mathrm{d}x = -\int_\Omega u\partial_{x_i} v \,\mathrm{d}x \qquad \forall v \in C_0^\infty(\Omega),$$

where $\mathcal{M}(\Omega)$ denotes the linear space of regular Borel measures, e.g. [42, Chapter 2]. The BV seminorm (also called TV seminorm) is given by

$$|u|_{\mathrm{BV}(\Omega)} := \sup_{v\in C_0^1(\Omega)^N, \||v|\|_\infty \le 1} \int_\Omega u \operatorname{div} v \,\mathrm{d}x.$$

We endow $\mathrm{BV}(\Omega)$ with the norm $\|\cdot\|_{\mathrm{BV}(\Omega)} := \|\cdot\|_{L^1(\Omega)} + |\cdot|_{\mathrm{BV}(\Omega)}$ and recall from [5, Thm. 10.1.1] that this makes $\mathrm{BV}(\Omega)$ a Banach space. Obviously, we have the inclusion $W^{1,1}(\Omega) \subset \mathrm{BV}(\Omega)$. Moreover, $\mathrm{BV}(\Omega)$ embeds continuously (compactly) into $L^r(\Omega)$ for $r \in [1, \frac{N}{N-1}]$ ($r \in [1, \frac{N}{N-1})$), see, e.g., [4, Cor. 3.49 and Prop. 3.21]. We use the convention that $\frac{N}{N-1} = \infty$ for $N = 1$. Also important is strict convergence, e.g. [4,5].

Definition 1 For $r \in [1, \frac{N}{N-1}]$ the metric $d_{\mathrm{BV},r}$ is given by

$$\begin{aligned} d_{\mathrm{BV},r}\colon & \mathrm{BV}(\Omega) \times \mathrm{BV}(\Omega) \to \mathbb{R}, \\ & (u,v) \mapsto \|u-v\|_{L^r(\Omega)} + \left| |u|_{\mathrm{BV}(\Omega)} - |v|_{\mathrm{BV}(\Omega)} \right|. \end{aligned}$$

Convergence with respect to $d_{\mathrm{BV},1}$ is called *strict convergence*.

Remark 1 The embedding $\mathrm{BV}(\Omega) \hookrightarrow L^r(\Omega)$, for $r \in [1, \frac{N}{N-1}]$, implies that $d_{BV,r}$ is well-defined and continuous with respect to $\|\cdot\|_{\mathrm{BV}(\Omega)}$.

We will also use the following density property.

Lemma 1 *$C^\infty(\bar{\Omega})$ is dense in $(BV(\Omega) \cap L^r(\Omega), d_{\mathrm{BV},r})$ for $r \in [1, \frac{N}{N-1}]$.*

Proof By straightforward modifications the proof for the special case $r = 1$, [5, Thm. 10.1.2], can be extended, using that the sequence of mollifiers constructed in the proof converges in L^r, see [5, Prop. 2.2.4]. □

1.2 The smoothed BV seminorm

We employ the smoothed BV seminorm $\psi_\delta\colon \mathrm{BV}(\Omega) \to [0,\infty)$ given by

$$\psi_\delta(u) := \sup\left\{ \int_\Omega u \operatorname{div} v + \sqrt{\delta(1-|v|^2)} \,\mathrm{d}x :\ v \in C_0^1(\Omega)^N,\ \||v|\|_{L^\infty(\Omega)} \le 1 \right\}.$$

It has the following properties.

Lemma 2 *The following statements are true for all $\delta \ge 0$.*

1. *For any* $u \in BV(\Omega)$ *there holds*

$$|u|_{BV(\Omega)} = \psi_0(u) \leq \psi_\delta(u) \leq |u|_{BV(\Omega)} + \sqrt{\delta}|\Omega|.$$

2. ψ_δ *is lower semi-continuous with respect to the* $L^1(\Omega)$*-topology.*
3. ψ_δ *is convex.*
4. *For all* $u \in W^{1,1}(\Omega)$ *we have*

$$\psi_\delta(u) = \int_\Omega \sqrt{\delta + |\nabla u|^2}\, \mathrm{d}x.$$

5. *The function* $\psi_\delta|_{H^1(\Omega)}$ *is Lipschitz with respect to* $\|\cdot\|_{H^1(\Omega)}$.

Proof The first four statements are from [1, Section 2] and the last one follows from $H^1(\Omega) \hookrightarrow W^{1,1}(\Omega)$, 4. and and the Lipschitz continuity of $r \mapsto \sqrt{\delta + r^2}$. □

Remark 2 The smoothing function ψ_δ for the TV seminorm is frequently used in imaging problems, e.g. [1,20].

1.3 The index s

For the remainder of this work we fix a number $s = s(N) \in (1, \frac{N}{N-1})$ with

$$\boxed{\mathrm{BV}(\Omega) \hookrightarrow\hookrightarrow L^s(\Omega) \hookrightarrow H^{-1}(\Omega).}$$

Remark 3 Consider, for instance, $N = 2$ and any $r \in (1, 2)$. Then we have $\mathrm{BV}(\Omega) \hookrightarrow\hookrightarrow L^r(\Omega)$ and $H^1(\Omega) \hookrightarrow L^{\frac{r}{r-1}}(\Omega)$ so that any $s \in (1, 2)$ can be used.

1.4 The solution operator of the state equation

Lemma 3 *For every* $u \in H^{-1}(\Omega)$ *the operator equation* $Ay = u$ *in* (OC) *has a unique solution* $y = y(u) \in Y$*. The solution operator*

$$S\colon H^{-1}(\Omega) \to Y, \quad u \mapsto y(u)$$

is linear, continuous, and bijective. In particular, S *is* L^s-L^2 *continuous. Moreover, for given* $q \in (1, \infty)$ *there is a constant* $C > 0$ *such that*

$$\|Su\|_{W^{2,q}(\Omega)} \leq C\|u\|_{L^q(\Omega)}$$

is satisfied for all $u \in L^q(\Omega)$.

Proof Except for the estimate all statements follow from the Lax-Milgram theorem. The estimate is a consequence of [26, Lemma 2.4.2.1, Theorem 2.4.2.5]. □

Remark 4 From $\mathrm{BV}(\Omega) \hookrightarrow L^s(\Omega) \hookrightarrow H^{-1}(\Omega)$ and Lemma 3 we obtain that (OC) has a nonempty feasible set.

2 The solutions of original and regularized problems

In this section we prove existence of solutions for (OC) and the associated regularized problems, characterize the solutions by optimality conditions, and show their convergence in appropriate function spaces.

2.1 The original problem: Existence of solutions

To establish the existence of a solution for (OC) we use the *reduced problem*

$$\min_{u\in \mathrm{BV}(\Omega)} \underbrace{\frac{1}{2}\|Su-y_\Omega\|^2_{L^2(\Omega)}+\beta|u|_{\mathrm{BV}(\Omega)}}_{=:j(u)}. \tag{ROC}$$

Lemma 4 *The function $j: BV(\Omega)\to\mathbb{R}$ is well-defined, strictly convex, and continuous with respect to $d_{BV,s}$.*

Proof The term $\frac{1}{2}\|Su-y_\Omega\|^2_{L^2(\Omega)}$ is well-defined by Remark 4 and strictly convex in u due to the injectivity of S. Since $|\cdot|_{\mathrm{BV}(\Omega)}$ is convex, the strict convexity of j follows. The continuity holds because S is L^s-L^2 continuous. □

The strict convexity implies that j has at most one (local=global) minimizer.

Theorem 1 *The problem* (ROC) *has a unique solution $\bar u\in BV(\Omega)$.*

Proof The proof is included in the proof of Theorem 2. □

As usual, the *optimal state* $\bar y$ and the *optimal adjoint state* $\bar p$ are given by

$$\bar y := S\bar u \in Y\cap W^{2,r_N}(\Omega) \qquad \text{and} \qquad \bar p := S^*(\bar y - y_\Omega),$$

where, due to $\mathrm{BV}(\Omega)\hookrightarrow L^{\frac{N}{N-1}}(\Omega)$ and Lemma 3, we have $r_N=\frac{N}{N-1}$ for $N\in\{2,3\}$, respectively, $r_N\geq 1$ arbitrarily large for $N=1$. Moreover, S^* is the adjoint operator of S wrt. the L^2 inner product. Since $S^*=S$ and $\bar y - y_\Omega\in L^2(\Omega)$, Lemma 3 yields $\bar p\in P$ for

$$P := H^2(\Omega)\cap H^1_0(\Omega).$$

It is standard to show that $\bar p$ is the unique weak solution of

$$\begin{cases} \mathcal{A}p + c_0 p = \bar y - y_\Omega & \text{in } \Omega,\\ \qquad\quad p = 0 & \text{on } \partial\Omega. \end{cases}$$

2.2 The original problem: Optimality conditions

The optimality conditions for (ROC) are provided in appendix C. They are only needed for the construction of the test problem in appendix D.

2.3 The regularized problems: Existence of solutions

Smoothing the BV seminorm and adding an H^1 regularization to j yields

$$\min_{u\in \mathrm{BV}(\Omega)} \underbrace{\frac{1}{2}||Su-y_\Omega||^2_{L^2(\Omega)}+\beta\psi_\delta(u)+\frac{\gamma}{2}\left\|u\right\|^2_{H^1(\Omega)}}_{=:j_{\gamma,\delta}(u)}, \qquad (\mathrm{ROC}_{\gamma,\delta})$$

where we set $j_{\gamma,\delta}(u) := +\infty$ for $u \in \mathrm{BV}(\Omega) \setminus H^1(\Omega)$ if $\gamma > 0$.

Lemma 5 *For any $\gamma, \delta \geq 0$ the function $j_{\gamma,\delta} : BV(\Omega) \to \mathbb{R} \cup \{+\infty\}$ is well-defined and strictly convex, and the function $j_{\gamma,\delta}|_{H^1(\Omega)}$ is H^1 continuous.*

Proof The well-definition and strict convexity of $j_{\gamma,\delta}$ follow similarly as for j in Lemma 4. The continuity follows term by term. For the first term it is enough to recall from Lemma 3 the L^2-L^2 continuity of S. The second term is Lipschitz in H^1 by Lemma 2. The continuity of the third term is clear. □

To prove existence of solutions for $(\mathrm{ROC}_{\gamma,\delta})$ we use an auxiliary result.

Lemma 6 *Let $(u_k)_{k\in\mathbb{N}} \subset BV(\Omega)$ be such that $(j(u_k))_{k\in\mathbb{N}}$ is bounded. Then $\left(\|u_k\|_{BV(\Omega)}\right)_{k\in\mathbb{N}}$ is bounded.*

Proof We denote by $C > 0$ a generic constant. The sequence $\left(|u_k|_{\mathrm{BV}(\Omega)}\right)_{k\in\mathbb{N}}$ is bounded because for each $k \in \mathbb{N}$ we have

$$|u_k|_{\mathrm{BV}(\Omega)} \leq \frac{j(u_k)}{\beta} \leq C.$$

The Poincaré inequality holds in $\mathrm{BV}(\Omega)$, see [40, Theorem 4.10], hence

$$\|u_k - \hat{u}_k\|_{L^s(\Omega)} \leq C|u_k|_{\mathrm{BV}(\Omega)} \leq C \qquad \forall k \in \mathbb{N}, \tag{1}$$

where $\hat{u}_k := \frac{1}{|\Omega|}\int_\Omega u_k \,\mathrm{d}x$ denotes the integral mean of u_k. From

$$\|Su_k\|_{L^2(\Omega)} - \|y_\Omega\|_{L^2(\Omega)} \leq \|Su_k - y_\Omega\|_{L^2(\Omega)} \leq \sqrt{2j(u^k)} \leq C,$$

it follows that $(\|Su_k\|_{L^2(\Omega)})_{k\in\mathbb{N}}$ is bounded. Together with the L^s-L^2 continuity of S and (1) this gives

$$|\hat{u}_k|\ \|S1\|_{L^2(\Omega)} = \|S\hat{u}_k\|_{L^2(\Omega)} \leq \|Su_k - S\hat{u}_k\|_{L^2(\Omega)} + \|Su_k\|_{L^2(\Omega)} \leq C.$$

The injectivity of S yields $S1 \neq 0$, so $(|\hat{u}_k|)_{k\in\mathbb{N}}$ is bounded, which implies boundedness of $(\|u_k\|_{L^s(\Omega)})$ by (1) and thus also of $(\|u_k\|_{L^1(\Omega)})$. □

Theorem 2 *For any $\gamma, \delta \geq 0$, $(\mathrm{ROC}_{\gamma,\delta})$ has a unique solution $\bar{u}_{\gamma,\delta} \in BV(\Omega)$. For $\gamma > 0$ we have $\bar{u}_{\gamma,\delta} \in H^1(\Omega)$.*

Proof For $\gamma > 0$ the existence of $\bar{u}_{\gamma,\delta} \in H^1(\Omega)$ follows from standard arguments since $j_{\gamma,\delta}|_{H^1(\Omega)}$ is strongly convex and H^1 continuous by Lemma 5. It remains to argue for $\gamma = 0$. Let $\delta \geq 0$. There is a sequence $(u_k)_{k\in\mathbb{N}} \subset \mathrm{BV}(\Omega)$ such that

$$\lim_{k\to\infty} j_{\gamma,\delta}(u_k) = \inf_{u\in\mathrm{BV}(\Omega)} j_{\gamma,\delta}(u).$$

Moreover, there exists $C \in \mathbb{R}$ such that for all $k \in \mathbb{N}$

$$\begin{aligned} j(u_k) &= \frac{1}{2}\|Su_k - y_\Omega\|^2_{L^2(\Omega)} + \beta|u_k|_{\mathrm{BV}(\Omega)} \\ &\leq \frac{1}{2}\|Su_k - y_\Omega\|^2_{L^2(\Omega)} + \beta\psi_\delta(u_k) \leq j_{\gamma,\delta}(u_k) \leq C, \end{aligned}$$

where we used $|u_k|_{\mathrm{BV}(\Omega)} \leq \psi_\delta(u_k)$ from Lemma 2. By Lemma 6 we have that $(\|u_k\|_U)_{k\in\mathbb{N}}$ is bounded. Since $\mathrm{BV}(\Omega)$ is compactly embedded in $L^s(\Omega)$, there is a subsequence of $(u_k)_{k\in\mathbb{N}}$, denoted the same way, such that $\lim_{k\to\infty}\|u_k - \bar{u}_{\gamma,\delta}\|_{L^s(\Omega)} = 0$ for some $\bar{u}_{\gamma,\delta} \in L^s(\Omega)$. As S is L^s-L^2 continuous and ψ_δ is L^1 lower semi-continuous by Lemma 2, we obtain

$$j_{\gamma,\delta}(\bar{u}_{\gamma,\delta}) \leq \liminf_{k\to\infty} j_{\gamma,\delta}(u_k) = \inf_{u\in\mathrm{BV}(\Omega)} j_{\gamma,\delta}(u).$$

This implies $|\bar{u}_{\gamma,\delta}|_{\mathrm{BV}(\Omega)} \leq \psi_\delta(\bar{u}_{\gamma,\delta}) \leq j_{\gamma,\delta}(\bar{u}_{\gamma,\delta})/\beta < \infty$, so $\bar{u}_{\gamma,\delta} \in \mathrm{BV}(\Omega)$ is a minimizer of $(\mathrm{ROC}_{\gamma,\delta})$. As $j_{\gamma,\delta}$ is strictly convex, the minimizer is unique. □

Optimal state $\bar{y}_{\gamma,\delta}$ and *optimal adjoint state* $\bar{p}_{\gamma,\delta}$ for $(\mathrm{ROC}_{\gamma,\delta})$ are given by

$$\bar{y}_{\gamma,\delta} := S\bar{u}_{\gamma,\delta} \in Y \cap W^{2,r_N}(\Omega) \qquad \text{and} \qquad \bar{p}_{\gamma,\delta} := S^*\big(\bar{y}_{\gamma,\delta} - y_\Omega\big) \in P,$$

where $r_N = \frac{N}{N-1}$ for $N \in \{2,3\}$, respectively, $r_N \geq 1$ arbitrarily large for $N = 1$. In particular, $\bar{p}_{\gamma,\delta}$ is the unique weak solution of

$$\begin{cases} \mathcal{A}p + c_0 p = \bar{y}_{\gamma,\delta} - y_\Omega & \text{in } \Omega, \\ \qquad p = 0 & \text{on } \partial\Omega. \end{cases}$$

2.4 The regularized problems: Optimality conditions

The objective $j_{\gamma,\delta}$ has the following differentiability properties.

Lemma 7 *For $\gamma, \delta > 0$ the functional $j_{\gamma,\delta} : H^1(\Omega) \to \mathbb{R}$ is Lipschitz continuously Fréchet differentiable and twice Gâteaux differentiable. Its first derivative is*

$$j'_{\gamma,\delta}(u)v = (S^*(Su - y_\Omega), v)_{L^2(\Omega)} + \beta\psi'_\delta(u)v + \gamma(u,v)_{H^1(\Omega)} \qquad \forall v \in H^1(\Omega),$$

where

$$\psi'_\delta(u)v = \int_\Omega \frac{(\nabla u, \nabla v)}{\sqrt{\delta + |\nabla u|^2}}\,\mathrm{d}x \qquad \forall v \in H^1(\Omega).$$

Proof It suffices to establish the claim for ψ_δ, which is done in Lemma 17. □

For differentiable convex functions a vanishing derivative is both necessary and sufficient for a global minimizer. This yields the following result.

Theorem 3 *For $\gamma, \delta > 0$ the control $\bar{u}_{\gamma,\delta} \in H^1(\Omega)$ is the solution of* (ROC$_{\gamma,\delta}$) *iff*

$$j'_{\gamma,\delta}(\bar{u}_{\gamma,\delta})v = 0 \qquad \forall v \in H^1(\Omega),$$

which is the nonlinear Neumann problem

$$\gamma(\bar{u}_{\gamma,\delta}, v)_{H^1(\Omega)} + \beta \int_\Omega \frac{(\nabla \bar{u}_{\gamma,\delta}, \nabla v)}{\sqrt{\delta + |\nabla \bar{u}_{\gamma,\delta}|^2}} \, \mathrm{d}x = -(\bar{p}_{\gamma,\delta}, v)_{L^2(\Omega)} \qquad \forall v \in H^1(\Omega). \tag{2}$$

2.5 Convergence of the path of solutions

We prove that $(\bar{u}_{\gamma,\delta}, \bar{y}_{\gamma,\delta}, \bar{p}_{\gamma,\delta})$ converges to $(\bar{u}, \bar{y}, \bar{p})$ for $\gamma, \delta \to 0$. As a first step we show convergence of the objective values.

Lemma 8 *We have*

$$j_{\gamma,\delta}(\bar{u}_{\gamma,\delta}) \xrightarrow{\mathbb{R}^2_{\geq 0} \ni (\gamma,\delta) \to (0,0)} j(\bar{u}).$$

Proof Let $\epsilon > 0$ and let $((\gamma_k, \delta_k))_{k \in \mathbb{N}} \subset \mathbb{R}^2_{\geq 0}$ converge to $(0,0)$. There holds

$$0 \leq j_{\gamma_k,\delta_k}(\bar{u}_{\gamma_k,\delta_k}) - j(\bar{u}) = \left[j_{\gamma_k,\delta_k}(\bar{u}_{\gamma_k,\delta_k}) - j_{\gamma_k,0}(\bar{u}_{\gamma_k,0})\right] + \left[j_{\gamma_k,0}(\bar{u}_{\gamma_k,0}) - j(\bar{u})\right],$$

where we used $j(\bar{u}) \leq j(\bar{u}_{\gamma_k,\delta_k}) \leq j_{\gamma_k,\delta_k}(\bar{u}_{\gamma_k,\delta_k})$. The first term in brackets satisfies

$$\begin{aligned} j_{\gamma_k,\delta_k}(\bar{u}_{\gamma_k,\delta_k}) - j_{\gamma_k,0}(\bar{u}_{\gamma_k,0}) &\leq j_{\gamma_k,\delta_k}(\bar{u}_{\gamma_k,0}) - j_{\gamma_k,0}(\bar{u}_{\gamma_k,0}) \\ &= \beta\psi_{\delta_k}(\bar{u}_{\gamma_k,0}) - \beta|\bar{u}_{\gamma_k,0}|_{\mathrm{BV}(\Omega)} \leq \beta\sqrt{\delta_k}|\Omega|, \end{aligned}$$

where the last inequality follows from Lemma 2. For the second term in brackets we deduce from Lemma 1 and the $d_{BV,s}$ continuity of j established in Lemma 4 that there is $u_\epsilon \in C^\infty(\bar{\Omega})$ such that $|j(\bar{u}) - j(u_\epsilon)| < \epsilon$. This yields

$$\begin{aligned} j_{\gamma_k,0}(\bar{u}_{\gamma_k,0}) - j(\bar{u}) &\leq j_{\gamma_k,0}(u_\epsilon) - j(\bar{u}) \\ &= j(u_\epsilon) + \frac{\gamma_k}{2}||u_\epsilon||^2_{H^1(\Omega)} - j(\bar{u}) \leq \epsilon + \frac{\gamma_k}{2}||u_\epsilon||^2_{H^1(\Omega)}. \end{aligned}$$

Putting the estimates for the two terms together shows

$$|j_{\gamma_k,\delta_k}(\bar{u}_{\gamma_k,\delta_k}) - j(\bar{u})| \leq \beta\sqrt{\delta_k}|\Omega| + \epsilon + \frac{\gamma_k}{2}\|u_\epsilon\|^2_{H^1(\Omega)}.$$

For $k \to \infty$ this implies the claim since

$$0 \leq \liminf_{k\to\infty} |j_{\gamma_k,\delta_k}(\bar{u}_{\gamma_k,\delta_k}) - j(\bar{u})| \leq \limsup_{k\to\infty} |j_{\gamma_k,\delta_k}(\bar{u}_{\gamma_k,\delta_k}) - j(\bar{u})| \leq \epsilon.$$

□

We infer that the optimal controls $\bar u_{\gamma,\delta}$ converge to $\bar u$ in L^r for suitable r.

Lemma 9 *For any $r \in [1, \frac{N}{N-1})$ we have $||\bar u_{\gamma,\delta} - \bar u||_{L^r(\Omega)} \xrightarrow{(\gamma,\delta)\to(0,0)} 0$.*

Proof Let $((\gamma_k, \delta_k))_{k\in\mathbb{N}} \subset \mathbb{R}^2_{\geq 0}$ converge to $(0,0)$. Let C be so large that $\gamma_k, \delta_k \leq C$ for all k. The optimality of $\bar u_{\gamma_k,\delta_k}$ and Lemma 2 yield for each $k \in \mathbb{N}$

$$j(\bar u_{\gamma_k,\delta_k}) \leq j_{\gamma_k,\delta_k}(\bar u_{\gamma_k,\delta_k}) \leq j_{\gamma_k,\delta_k}(0) \leq j_{C,C}(0).$$

Lemma 6 and the compact embedding of $\mathrm{BV}(\Omega)$ into $L^r(\Omega)$, $r \in [s, \frac{N}{N-1})$, imply that there exists $\tilde u \in L^r(\Omega)$ such that a subsequence of $(\bar u_{\gamma_k,\delta_k})_{k\in\mathbb{N}}$, denoted in the same way, converges to $\tilde u$ in $L^r(\Omega)$. It is therefore enough to show $\tilde u = \bar u$. Since j is lower semi-continuous in the L^s topology, we have

$$j(\tilde u) \leq \liminf_{k\to\infty} j(\bar u_{\gamma_k,\delta_k}) \leq \liminf_{k\to\infty} j_{\gamma_k,\delta_k}(\bar u_{\gamma_k,\delta_k}) = j(\bar u),$$

where we used Lemma 8 to obtain the last equality. This shows $\tilde u \in \mathrm{BV}(\Omega)$, hence Theorem 1 implies $\tilde u = \bar u$. □

In fact, the convergence of $\bar u_{\gamma,\delta}$ to $\bar u$ is stronger.

Theorem 4 *For any $r \in [1, \frac{N}{N-1})$ we have $d_{BV,r}(\bar u_{\gamma,\delta}, \bar u) \xrightarrow{(\gamma,\delta)\to(0,0)} 0$.*

Proof For any $\gamma, \delta \geq 0$ we have $j(\bar u) \leq j(\bar u_{\gamma,\delta}) \leq j_{\gamma,\delta}(\bar u_{\gamma,\delta})$, so Lemma 8 yields $\lim_{(\gamma,\delta)\to(0,0)} j(\bar u_{\gamma,\delta}) = j(\bar u)$. Furthermore, there holds

$$\beta\Big||\bar u|_{\mathrm{BV}(\Omega)} - |\bar u_{\gamma,\delta}|_{\mathrm{BV}(\Omega)}\Big| \leq \big|j(\bar u) - j(\bar u_{\gamma,\delta})\big| + \frac{1}{2}\Big|\|S\bar u - y_\Omega\|^2_{L^2(\Omega)} - \|S\bar u_{\gamma,\delta} - y_\Omega\|^2_{L^2(\Omega)}\Big|.$$

By Lemma 9 and the continuity of S from $L^s(\Omega)$ to $L^2(\Omega)$ we thus find

$$|\bar u_{\gamma,\delta}|_{\mathrm{BV}(\Omega)} \xrightarrow{(\gamma,\delta)\to(0,0)} |\bar u|_{\mathrm{BV}(\Omega)}.$$

Together with Lemma 9 this proves the claim. □

We conclude this section with the convergence of $(\bar y_{\gamma,\delta}, \bar p_{\gamma,\delta})$ to $(\bar y, \bar p)$.

Theorem 5 *For any $r \in [1, \frac{N}{N-1})$ and any $r' \in [1, \infty)$ we have*

$$\lim_{(\gamma,\delta)\to(0,0)} \|\bar y_{\gamma,\delta} - \bar y\|_{W^{2,r}(\Omega)} = 0 \qquad \textit{and} \qquad \lim_{(\gamma,\delta)\to(0,0)} \|\bar p_{\gamma,\delta} - \bar p\|_{W^{2,r'}(\Omega)} = 0.$$

Proof The continuity of S from L^q to $W^{2,q}$ for any $q \in (1, \infty)$, see Lemma 3, implies with Lemma 9 that $\lim_{(\gamma,\delta)\to(0,0)} \|\bar y_{\gamma,\delta} - \bar y\|_{W^{2,r}(\Omega)} = 0$ for any $r \in [1, \frac{N}{N-1})$. Since for any $r' \in (1, \infty)$ there is $r \in [1, \frac{N}{N-1})$ such that $W^{2,r}(\Omega) \hookrightarrow L^{r'}(\Omega)$ is satisfied, we can use the $L^{r'}$-$W^{2,r'}$ continuity of $S^* = S$ to find $\lim_{(\gamma,\delta)\to(0,0)} \|\bar p_{\gamma,\delta} - \bar p\|_{W^{2,r'}(\Omega)} = \lim_{(\gamma,\delta)\to(0,0)} \|S^*(\bar y_{\gamma,\delta} - \bar y)\|_{W^{2,r'}(\Omega)} = 0$. □

Remark 5 The results of section 2 can also be established for nonsmooth domains Ω, but $\bar{y}, \bar{p}, \bar{y}_{\gamma,\delta}, \bar{p}_{\gamma,\delta}$ may be less regular since S may not provide the regularity stated in Lemma 3. A careful inspection reveals that only Theorem 5 has to be modified. If, for instance, $\Omega \subset \mathbb{R}^N$, $N \in \{2,3\}$, is a bounded Lipschitz domain, then [43, Theorem 3] implies that Theorem 5 holds if $W^{2,r}$ and $W^{2,r'}$ are both replaced by H^r, where $r \in [1, \frac{3}{2})$ is arbitrary. If Ω is convex, then [26, Theorem 3.2.1.2] further yields that $W^{2,r'}$ can be replaced by H^2.

3 Differentiability of the adjoint-to-control mapping

The main goal of this section is to show that the PDE

$$\begin{cases} -\operatorname{div}\left(\left[\gamma + \dfrac{\beta}{\sqrt{\delta + |\nabla u|^2}}\right]\nabla u\right) + \gamma u = p \quad \text{in } \Omega, \\ \left(\left[\gamma + \dfrac{\beta}{\sqrt{\delta + |\nabla u|^2}}\right]\nabla u, \nu\right) = 0 \quad \text{on } \partial\Omega \end{cases} \tag{3}$$

has a unique weak solution $u = u(p) \in C^{1,\alpha}(\Omega)$ for every right-hand side $p \in L^\infty(\Omega)$, and that $p \mapsto u(p)$ is Lipschitz continuously Fréchet differentiable in any open ball, having a Lipschitz constant that is independent of γ and δ, provided $\gamma > 0$ and $\delta > 0$ are bounded away from zero. This is accomplished in Theorem 7. Note that we suppress the dependency on γ, δ in $u = u(p; \gamma, \delta)$.

Assumption 6 *We are given constants* $0 < \gamma_0 \leq \gamma^0$, $0 < \delta_0 \leq \delta^0$ *and* $b^0 > 0$. *We denote* $I := [\gamma_0, \gamma^0] \times [\delta_0, \delta^0]$ *and write* $\mathbb{B} \subset L^\infty(\Omega)$ *for the open ball of radius* $b^0 > 0$ *centered at the origin in* $L^\infty(\Omega)$.

Let us first establish well-definition of $p \mapsto u(p)$ and a Lipschitz estimate.

Lemma 10 *Let Assumption 6 hold. Then there exist* $L > 0$ *and* $\alpha \in (0,1)$ *such that for each* $(\gamma, \delta) \in I$ *and all* $p_1, p_2 \in \mathbb{B}$ *the PDE* (3) *has unique weak solutions* $u_1 = u_1(p_1) \in C^{1,\alpha}(\Omega)$ *and* $u_2 = u_2(p_2) \in C^{1,\alpha}(\Omega)$ *that satisfy*

$$\|u_1 - u_2\|_{C^{1,\alpha}(\Omega)} \leq L \|p_1 - p_2\|_{L^\infty(\Omega)}.$$

In particular, we have the stability estimate

$$\|u_1\|_{C^{1,\alpha}(\Omega)} \leq L \|p_1\|_{L^\infty(\Omega)}.$$

Proof Unique existence and the first estimate are established in Theorem 13 in the appendix. The second estimate follows from the first for $p_2 = 0$. □

We introduce

$$f : \mathbb{R}^N \to \mathbb{R}^N, \qquad f(v) := \beta \frac{v}{\sqrt{\delta + |v|^2}},$$

so that (3) reads

$$-\operatorname{div}\Big(\gamma \nabla u + f(\nabla u)\Big) + \gamma u = p \quad \text{in } H^1(\Omega)^*. \tag{4}$$

We now show that the adjoint-to-control mapping is differentiable.

Lemma 11 *Let Assumption 6 hold and let $\alpha \in (0,1)$ be the constant from Lemma 10. For each $(\gamma,\delta) \in I$ the mapping $\mathbb{B} \ni p \mapsto u(p) \in C^{1,\alpha}(\Omega)$ is Fréchet differentiable. Its derivative $z = u'(p)d \in C^{1,\alpha}(\Omega)$ in direction $d \in L^\infty(\Omega)$ is the unique weak solution of the linear PDE*

$$\begin{cases} -\operatorname{div}\left(\Big[\gamma I + f'\big(\nabla u(p)\big)\Big]\nabla z\right) + \gamma z = d & \text{in } \Omega, \\ \left(\Big[\gamma I + f'\big(\nabla u(p)\big)\Big]\nabla z, \nu\right) = 0 & \text{on } \partial\Omega, \end{cases} \tag{5}$$

and there exists $C > 0$ such that for all $(\gamma,\delta) \in I$, all $p \in \mathbb{B}$, and all $d \in L^\infty(\Omega)$ we have

$$\|z\|_{C^{1,\alpha}(\Omega)} \leq C\, \|d\|_{L^\infty(\Omega)}.$$

Proof Let $p \in \mathbb{B}$ and $d \in L^\infty(\Omega)$ be such that $p+d \in \mathbb{B}$. From Lemma 10 we obtain $u(p) \in C^{1,\alpha}(\Omega)$ and $\|u(p)\|_{C^{1,\alpha}(\Omega)} \leq C\|p\|_{L^\infty(\Omega)}$, where C is independent of γ, δ, p. Combining this with Lemma 18 implies

$$f'(\nabla u(p)) = \frac{I}{\sqrt{\delta + |\nabla u(p)|^2}} - \frac{\nabla u(p)\nabla u(p)^T}{\big(\delta + |\nabla u(p)|^2\big)^{\frac{3}{2}}} \in C^{0,\alpha}(\Omega, \mathbb{R}^{N\times N}) \tag{6}$$

and the estimate $\|A\|_{C^{0,\alpha}(\Omega)} \leq a^0$ for $A := \gamma I + f'(\nabla u(p))$ with a constant a^0 that does not depend on γ, δ, p. This shows that Theorem 12 is applicable. Thus, it follows that the PDE (5) has a unique weak solution $z \in C^{1,\alpha}(\Omega)$ that satisfies the claimed estimate. Concerning the Fréchet differentiability we obtain for $r := u(p+d) - u(p) - z \in C^{1,\alpha}(\Omega)$

$$\begin{aligned} &-\operatorname{div}\left(\Big[\gamma I + f'\big(u(p)\big)\Big]\nabla r\right) + \gamma r \\ &\quad = -\operatorname{div}\Big(\gamma\nabla u(p+d)\Big) + \gamma u(p+d) + \operatorname{div}\Big(\gamma\nabla u(p)\Big) - \gamma u(p) \\ &\qquad + \operatorname{div}\left(\Big[\gamma I + f'\big(\nabla u(p)\big)\Big]\nabla z\right) - \gamma z - \operatorname{div}\Big(f'\big(\nabla u(p)\big)w\Big) \\ &\quad = \operatorname{div}\Big(f\big(\nabla u(p+d)\big) - f\big(\nabla u(p)\big) - f'\big(\nabla u(p)\big)w\Big), \end{aligned}$$

where we set $w := w(p,d) := \nabla u(p+d) - \nabla u(p)$. Theorem 12 implies that there is $C > 0$, independent of d, such that

$$\big\|r\big\|_{C^{1,\alpha}(\Omega)} \leq C\Big\|f\big(\nabla u(p+d)\big) - f\big(\nabla u(p)\big) - f'\big(u(p)\big)w\Big\|_{C^{0,\alpha}(\Omega)}.$$

The expression in the norm on the right-hand side satisfies the following pointwisely in Ω

$$\begin{aligned} &f\big(\nabla u(p+d)\big) - f\big(\nabla u(p)\big) - f'\big(\nabla u(p)\big)w \\ &\quad = \left(\int_0^1 f'\big(\nabla u(p) + tw\big) - f'\big(\nabla u(p)\big)\,\mathrm{d}t\right) w \\ &\quad = \left(\int_0^1 \int_0^1 f''\big(\nabla u(p) + \tau t w\big)\,\mathrm{d}\tau\, t\,\mathrm{d}t\right) [w, w]. \end{aligned}$$

Lemma 18 yields

$$\|r\|_{C^{1,\alpha}(\Omega)} \le C \int_0^1 \int_0^1 \Big\|f''\big(\nabla u(p)+\tau t w\big)\Big\|_{C^{0,\alpha}(\Omega)} \mathrm{d}\tau\,\mathrm{d}t\, \Big\|u(p+d)-u(p)\Big\|^2_{C^{1,\alpha}(\Omega)}.$$

As $f \in C^3(\mathbb{R}^N, \mathbb{R}^N)$ with bounded derivatives we have that f'' is Lipschitz continuous and bounded. We infer from Lemma 18 and Lemma 10 that

$$\|r\|_{C^{1,\alpha}(\Omega)} \le C\|d\|^2_{L^\infty(\Omega)},$$

which shows $\|r\|_{C^{1,\alpha}(\Omega)} = o\big(\|d\|_{L^\infty(\Omega)}\big)$ since C is independent of d. □

Theorem 7 *Let Assumption 6 hold and let $\alpha \in (0,1)$ be the constant from Lemma 10. Then the mapping $u' : \mathbb{B} \to \mathcal{L}(L^\infty(\Omega), C^{1,\alpha}(\Omega))$ is Lipschitz continuous and the Lipschitz constant does not depend on (γ, δ), but only on Ω, N, γ_0, γ^0, δ_0, δ^0 and b^0.*

Proof Let $p, q \in \mathbb{B}$ and $d \in L^\infty(\Omega)$. Set $z_p := \nabla\big(u'(p)d\big)$ and $z_q := \nabla\big(u'(q)d\big)$. Then

$$-\operatorname{div}\Big(\gamma\big[z_p - z_q\big] + f'\big(\nabla u(p)\big)z_p - f'\big(\nabla u(q)\big)z_q\Big) + \gamma\big[u'(p)d - u'(q)d\big] = 0$$

holds in $H^1(\Omega)^*$. Thus, the difference $r := u'(p)d - u'(q)d$ satisfies

$$\begin{aligned} -\operatorname{div}\big(\gamma\nabla r\big) + \gamma r &= \operatorname{div}\Big(f'\big(\nabla u(p)\big)z_p - f'\big(\nabla u(q)\big)z_q\Big) \\ &= \operatorname{div}\Big(f'\big(\nabla u(p)\big)\nabla r\Big) + \operatorname{div}\Big(\big[f'\big(\nabla u(p)\big) - f'\big(\nabla u(q)\big)\big]z_q\Big), \end{aligned}$$

from which we infer that

$$-\operatorname{div}\Big(\big[\gamma I + f'\big(\nabla u(p)\big)\big]\nabla r\Big) + \gamma r = \operatorname{div}\Big(\big[f'\big(\nabla u(p)\big) - f'\big(\nabla u(q)\big)\big]z_q\Big)$$

in $H^1(\Omega)^*$. By the same arguments as below (6), $A := \gamma I + f'\big(\nabla u(p)\big)$ satisfies $\|A\|_{C^{0,\alpha}(\Omega)} \le a^0$ with a constant a^0 that does not depend on γ, δ, p, q. Moreover, A is clearly elliptic with constant γ_0. By Theorem 12 this yields

$$\|r\|_{C^{1,\alpha}(\Omega)} \le C\Big\|\big[f'\big(\nabla u(p)\big) - f'\big(\nabla u(q)\big)\big]z_q\Big\|_{C^{0,\alpha}(\Omega)}.$$

Here, $C > 0$ does not depend on p, q, but only on the desired quantities. From Lemma 18 and Lemma 11 we infer that

$$\|r\|_{C^{1,\alpha}(\Omega)} \le C \Big\| f'\big(\nabla u(p)\big) - f'\big(\nabla u(q)\big) \Big\|_{C^{0,\alpha}(\Omega)} \|d\|_{L^\infty(\Omega)}.$$

Lemma 18 and Lemma 10 therefore imply

$$\begin{aligned}
&\|u'(p) - u'(q)\|_{\mathcal{L}(L^\infty(\Omega), C^{1,\alpha}(\Omega))} \\
&\le C \left\| \int_0^1 f''\Big(\nabla u(q) + t\big[\nabla u(p) - \nabla u(q)\big]\Big)\,\mathrm{d}t \right\|_{C^{0,\alpha}(\Omega)} \|\nabla u(p) - \nabla u(q)\|_{C^{0,\alpha}(\Omega)} \\
&\le C \int_0^1 \Big\| f''\Big(\nabla u(q) + t\big[\nabla u(p) - \nabla u(q)\big]\Big) \Big\|_{C^{0,\alpha}(\Omega)} \,\mathrm{d}t\, \|p - q\|_{L^\infty(\Omega)}.
\end{aligned}$$

The first factor is bounded since f'' is bounded and Lipschitz. This demonstrates the asserted Lipschitz continuity. □

Remark 6 Theorem 7 stays valid if Ω is of class $C^{1,\alpha'}$ for some $\alpha' > 0$.

4 An inexact Newton method for the regularized problems

In this section we introduce the formulation of the optimality system of $(\mathrm{ROC}_{\gamma,\delta})$ on which our numerical method is based, and we show that the application of an inexact Newton method to this formulation yields local convergence. We use the following assumption.

Assumption 8 *We are given constants* $0 < \gamma_0 \le \gamma^0$, $0 < \delta_0 \le \delta^0$ *and* $b^0 \ge 0$. *We denote* $I := [\gamma_0, \gamma^0] \times [\delta_0, \delta^0]$ *and fix* $(\gamma, \delta) \in I$.

Introducing

$$F : Y \times P \to Y^* \times L^2(\Omega), \qquad F(y,p) := \begin{pmatrix} Ay - u(-p) \\ y - y_\Omega - A^* p \end{pmatrix} \tag{7}$$

the optimality conditions from Theorem 3 are given by $F(\bar{y}_{\gamma,\delta}, \bar{p}_{\gamma,\delta}) = 0$, and the pair $(\bar{y}_{\gamma,\delta}, \bar{p}_{\gamma,\delta})$ is the unique root of F. We suppress the dependency of $u = u(p; \gamma, \delta)$ and $F = F(y, p; \gamma, \delta)$ on γ, δ. By standard Sobolev embeddings we have $P \subset H^2(\Omega) \hookrightarrow L^\infty(\Omega)$, hence $u(-p) \in C^{1,\alpha}(\Omega)$ for some $\alpha > 0$ by Lemma 10, so F is well-defined. We mention the work [44], where a Newton system with a somewhat similar structure is considered.

The next two lemmas yield convergence of an inexact Newton method.

Lemma 12 *Let Assumption 8 hold. Then F defined in* (7) *is locally Lipschitz continuously Fréchet differentiable. Its derivative at* $(y,p) \in Y \times P$ *is given by*

$$F'(y,p)\colon Y \times P \to Y^* \times L^2(\Omega), \qquad (\delta y, \delta p) \mapsto \begin{pmatrix} A & u'(-p) \\ I & -A^* \end{pmatrix} \begin{pmatrix} \delta y \\ \delta p \end{pmatrix}.$$

Proof Only $p \mapsto u(-p)$ is nonlinear, so the claims follow from Theorem 7. □

Lemma 13 *Let Assumption 8 hold. Then $F'(y,p)$ is invertible for all $(y,p) \in Y \times P$.*

Proof The proof consists of two parts. First we show that $F'(y,p)$ is injective and second that it is a Fredholm operator of index 0, see [32, Chapter IV, Section 5]. These two facts imply the bijectivity of $F'(y,p)$. For the injectivity let $(\delta y, \delta p) \in Y \times P$ with $F'(y,p)(\delta y, \delta p) = 0 \in Y^* \times L^2(\Omega)$, i.e.

$$0 = A\delta y + u'(-p)\delta p \in Y^* \qquad \text{and} \qquad 0 = \delta y - A^*\delta p \in L^2(\Omega), \tag{8}$$

and therefore

$$\|\delta y\|^2_{L^2(\Omega)} = (A^*\delta p, \delta y)_{L^2(\Omega)} = -(u'(-p)\delta p, \delta p)_{L^2(\Omega)}.$$

The representation of $z := u'(-p)\delta p$ from Lemma 11 yields

$$\begin{aligned} -\|\delta y\|^2_{L^2(\Omega)} &= \left(\Big[\gamma I + f'\big(\nabla u(-p)\big)\Big]\nabla z, \nabla z\right)_{L^2(\Omega)} + \gamma\big(z, z\big)_{L^2(\Omega)} \\ &\geq \Big(f'\big(\nabla u(-p)\big)\nabla z, \nabla z\Big)_{L^2(\Omega)}. \end{aligned} \tag{9}$$

Since f' is positive semi-definite, we find $\|\delta y\|^2_{L^2(\Omega)} \leq 0$. This shows $\delta y = 0$. By (8) this yields $A^*\delta p = 0$ in $L^2(\Omega)$, hence $\delta p = 0$, which proves the injectivity. To apply Fredholm theory we decompose $F'(y,p)$ into the two operators

$$F'(y,p) = \begin{pmatrix} A & 0 \\ 0 & -A^* \end{pmatrix} + \begin{pmatrix} 0 & u'(-p) \\ I & 0 \end{pmatrix}.$$

We want to use [32, Chapter IV, Theorem 5.26], which states: If the first operator is a Fredholm operator of index 0 and the second operator is compact with respect to the first operator (see [32, Chapter IV, Introduction to Section 3]), then their sum $F'(y,p)$ is also a Fredholm operator of index 0. By the injectivity of $F'(y,p)$ this implies its bijectivity.

The operators $A : Y \to Y^*$ and $A^* : P \to L^2(\Omega)$ are invertible by Lemma 3, and thus

$$Y \times P \to Y^* \times L^2(\Omega), \qquad (\delta y, \delta p) \mapsto \begin{pmatrix} A & 0 \\ 0 & -A^* \end{pmatrix} \begin{pmatrix} \delta y \\ \delta p \end{pmatrix}$$

is invertible and in particular a Fredholmoperator of index 0. It remains to show that

$$Y \times P \to Y^* \times L^2(\Omega), \qquad (\delta y, \delta p) \mapsto \begin{pmatrix} 0 & u'(-p) \\ I & 0 \end{pmatrix} \begin{pmatrix} \delta y \\ \delta p \end{pmatrix}$$

is compact with respect to the first operator. Thus, we have to establish that for any sequence $\big((\delta y_n, \delta p_n)\big)_{n\in\mathbb{N}} \subset Y \times P$ such that there exists a $C > 0$ with

$$\big(\|\delta y_n\|_Y + \|\delta p_n\|_P\big) + \big(\|A\delta y_n\|_{Y^*} + \|A^*\delta p_n\|_{L^2(\Omega)}\big) \leq C \qquad \forall n \in \mathbb{N}, \tag{10}$$

the sequence $\big((u'(-p)\delta p_n, \delta y_n)\big)_{n\in\mathbb{N}} \subset Y^* \times L^2(\Omega)$ contains a convergent subsequence. By (10) we have that $(\|\delta y_n\|_Y)_{n\in\mathbb{N}}$ is bounded. The compact embedding $Y \hookrightarrow\hookrightarrow L^2(\Omega)$ therefore implies the existence of a point $\hat{y} \in L^2(\Omega)$ and a subsequence, denoted in the same way, such that $\|\delta y_n - \hat{y}\|_{L^2(\Omega)} \to 0$. We also have that $(\|\delta p_n\|_P)_{n\in\mathbb{N}}$ is bounded. In particular $\|\delta p_n\|_{L^\infty(\Omega)} \leq b^0$ for all $n \in \mathbb{N}$ for some $b^0 > 0$. By Lemma 11 this implies that $(u'(-p)\delta p_n)_{n\in\mathbb{N}}$ is bounded in $C^{1,\alpha}(\Omega)$. Since $C^{1,\alpha}(\Omega) \hookrightarrow\hookrightarrow Y^*$, the proof is complete. ☐

We consider the following inexact Newton method to find the root of F given by (7). The norm that appears is that of $Y^* \times P^*$.

Algorithm 1: An inexact Newton method for $(\mathrm{ROC}_{\gamma,\delta})$

Input: $(y_0, p_0) \in Y \times P$, $(\gamma, \delta) \in \mathbb{R}^2_{>0}$, $\eta \in [0, \infty)$

```
1 for k = 0, 1, 2, ..., it_in do
2   if F(y_k, p_k) = 0 then set (y*, p*) := (y_k, p_k); stop
3   Compute (δy_k, δp_k) such that
      ‖F(y_k, p_k) + F'(y_k, p_k)(δy_k, δp_k)‖ ≤ η_k ‖F(y_k, p_k)‖, where η_k ∈ [0, η]
4   Set (y_{k+1}, p_{k+1}) = (y_k, p_k) + (δy_k, δp_k)
5 end
```

Output: (y^*, p^*)

It is well-known that the properties established in Lemma 12 and Lemma 13 are sufficient for local linear/q-superlinear/q-quadratic convergence of the inexact Newton method if the residual in iteration k is of appropriate order, e.g. [33, Theorem 6.1.4]. Thus, we obtain the following result.

Theorem 9 *Let Assumption 8 hold. If $(y_0, p_0) \in Y \times P$ is sufficiently close to $(\bar{y}_{\gamma,\delta}, \bar{p}_{\gamma,\delta})$, then Algorithm 1 either terminates after finitely many iterations with output $(y^*, p^*) = (\bar{y}_{\gamma,\delta}, \bar{p}_{\gamma,\delta})$ or it generates a sequence (y_k, p_k) that converges r-linearly [q-linearly/q-superlinearly/with q-order $1+\omega$] to $(\bar{y}_{\gamma,\delta}, \bar{p}_{\gamma,\delta})$, provided $\eta < 1$ [η is sufficiently small/$\eta_k \to 0$/$\eta_k = O(\|F(y_k, p_k)\|^\omega)$]. Here, $\omega \in (0, 1]$ is arbitrary; for $\omega = 1$ this means q-quadratic convergence.*

Remark 7 The same rates of convergence can also be established if an inexact Newton method is applied to standard formulations of the optimality system of $(\mathrm{ROC}_{\gamma,\delta})$. We focus on the implicit formulation (7) since in our numerical experiments this was the only approach that proved capable of sufficiently reducing γ, δ for $\Omega \subset \mathbb{R}^2$. For instance, with a control-based formulation and for a fixed coupling $\delta = 0.01\gamma$ we could not reduce γ below roughly 10^{-5}. Also, the algorithm was quite sensitive, e.g., a slight variation in the initial data could lead to a much higher number of total iterations. Both observations are well in line with our previous experience [13,21,27,28,30] on PDE-constrained optimal control problems involving the TV seminorm. In contrast, working with (7) is much more stable and enabled us to reduce γ to levels below 10^{-10}, as the numerical results in section 7 show. We point out that the homotopy path $(\gamma, \delta) \mapsto (\bar{u}_{\gamma,\delta}, \bar{y}_{\gamma,\delta}, \bar{p}_{\gamma,\delta})$ is not affected by the reformulation of the optimality system, so it is appropriate to compare the final values of γ.

5 Finite Element approximation

In this section we provide a discretization scheme for ($\mathrm{ROC}_{\gamma,\delta}$) and prove its convergence. Throughout, we work with a fixed pair $(\gamma, \delta) \in \mathbb{R}^2_{>0}$.

5.1 Discretization

We use Finite Elements for the discretization of ($\mathrm{ROC}_{\gamma,\delta}$). Control, state and adjoint state are discretized by piecewise linear and globally continuous elements on a triangular grid. We point out that discretizing the control by piecewise constant Finite Elements will not ensure convergence to the optimal control $\bar{u}_{\gamma,\delta}$, in general; cf. [6, Section 4].

For all $h \in (0, h_0]$ and a suitable $h_0 > 0$ let $\mathcal{T}_h$ denote a collection of open triangular cells $T \subset \Omega$ with $h = \max_{T\in\mathcal{T}_h} \operatorname{diam}(T)$. We write $\Omega_h := \operatorname{int}(\cup_{T\in\mathcal{T}_h} \bar{T})$. We assume that there are constants $C > 0$ and $c > \frac{1}{2}$ such that

$$\operatorname{dist}(\partial\Omega_h, \partial\Omega) \le Ch^c, \quad |\Omega \setminus \Omega_h| \xrightarrow{h\to 0} 0, \quad |\partial\Omega_h| \le C. \tag{11}$$

We further assume $(\mathcal{T}_h)_{h\in(0,h_0]}$ to be quasi-uniform and $\Omega_h \subset \Omega_{h'}$ for $h' \le h$. The assumptions in (11) are rather mild and in part implied if, for example, Ω and $(\Omega_h)_{h>0}$ are a family of uniform Lipschitz domains, cf. [29, Sections 4.1.2&4.1.3]. We also utilize the function spaces

$$V_h := \left\{ v_h \in C(\bar{\Omega}_h) : v_h|_T \text{ is affine linear } \forall\, T \in \mathcal{T}_h \right\}, \quad Y_h := V_h \cap H^1_0(\Omega_h).$$

Because $V_h \hookrightarrow H^1(\Omega_h)$ it follows that Y_h contains precisely those functions of V_h that vanish on $\partial\Omega_h$. We use the standard nodal basis $\varphi_1, \varphi_2, \dots, \varphi_{\dim(V_h)}$ in V_h and assume that it is ordered in such a way that $\varphi_1, \varphi_2, \dots, \varphi_{\dim(Y_h)}$ is a basis of Y_h. For every $u \in L^2(\Omega_h)$ there is a unique $y_h \in Y_h$ that satisfies

$$\int_{\Omega_h} \left(\sum_{i,j=1}^{N} a_{ij} \partial_i y_h \partial_j \varphi_h \right) + c_0 y_h \varphi_h \,\mathrm{d}x = \int_{\Omega_h} u \varphi_h \,\mathrm{d}x \qquad \forall \varphi_h \in Y_h$$

and by defining $S_h u := y_h$ we obtain the discrete solution operator $S_h : L^2(\Omega_h) \to Y_h$ to the PDE in (OC). The discretized version of ($\mathrm{ROC}_{\gamma,\delta}$) is given by

$$\min_{u\in V_h} \underbrace{\frac{1}{2} ||S_h u - y_{\Omega_h}||^2_{L^2(\Omega_h)} + \beta\psi_\delta(u) + \frac{\gamma}{2} \|u\|^2_{H^1(\Omega_h)}}_{=: j_{\gamma,\delta,h}(u)}, \tag{$\mathrm{ROC}_{\gamma,\delta,h}$}$$

where y_{Ω_h} represents the restriction of y_Ω to Ω_h. By standard arguments this problem has a unique optimal solution $\bar{u}_{\gamma,\delta,h}$. Based on $\bar{u}_{\gamma,\delta,h}$ we define $\bar{y}_{\gamma,\delta,h} := S_h \bar{u}_{\gamma,\delta,h}$ and $\bar{p}_{\gamma,\delta,h} := S_h^*(S_h \bar{u}_{\gamma,\delta,h} - y_{\Omega_h})$. For $h \to 0$ the triple $(\bar{u}_{\gamma,\delta,h}, \bar{y}_{\gamma,\delta,h}, \bar{p}_{\gamma,\delta,h})$ converges to the continuous optimal triple $(\bar{u}_{\gamma,\delta}, \bar{y}_{\gamma,\delta}, \bar{p}_{\gamma,\delta})$ in an appropriate sense, as we show next.

5.2 Convergence

In this section we prove convergence of the Finite Element approximation. We will tacitly use that extension-by-zero yields for each $v \in Y_h \subset H_0^1(\Omega_h)$ a function in $H_0^1(\Omega)$. Also, we need the following density result.

Lemma 14 *Let* (11) *hold. For each* $\varphi \in C^\infty(\bar{\Omega}) \cap H_0^1(\Omega)$ *there exists a sequence* $(\varphi_h) \subset Y_h$ *such that* $\lim_{h\to 0^+} \|\varphi_h - \varphi\|_{H^1(\Omega_h)} = 0$.

Proof Given φ and a sufficiently small h we define $\varphi_h \in Y_h$ on the inner nodes $x_1, x_2, \ldots, x_{\dim(Y_h)}$ of $\bar{\Omega}_h \setminus \partial\Omega_h$ as φ and as zero on the nodes $x_{\dim(Y_h)+1}, \ldots, x_{\dim(V_h)}$ on the boundary. That is, we set

$$\varphi_h(x_i) := \begin{cases} \varphi(x_i) & \text{if } x_i \in \Omega_h, \\ 0 & \text{if } x_i \in \partial\Omega_h. \end{cases}$$

Inserting the nodal interpolant $I_h\varphi \in V_h$ and utilizing an inverse inequality, e.g. [25, Corollary 1.141], we find

$$\|\varphi - \varphi_h\|_{H^1(\Omega_h)} \le \|\varphi - I_h\varphi\|_{H^1(\Omega_h)} + Ch^{-1}\|\varphi_h - I_h\varphi\|_{L^2(\Omega_h)},$$

where $C > 0$ is independent of h. An interpolation error estimate shows that the first term on the right-hand side converges to 0 for $h \to 0^+$, see for example [25, Theorem 1.103]. Owing to the definition of φ_h the second term only involves elements near the boundary $\partial\Omega_h$, hence

$$\begin{aligned} h^{-2}\|\varphi_h - I_h\varphi\|^2_{L^2(\Omega_h)} &= h^{-2} \sum_{\substack{T\in\mathcal{T}_h \\ \bar{T}\cap\partial\Omega_h\neq\emptyset}} \|\varphi_h - I_h\varphi\|^2_{L^2(T)} \\ &\le Ch^{N-2} \sum_{\substack{T\in\mathcal{T}_h \\ \bar{T}\cap\partial\Omega_h\neq\emptyset}} \|\varphi_h - I_h\varphi\|^2_{L^\infty(T)}. \end{aligned}$$

By (11) and the quasi-uniformity we find that the number of boundary triangles is proportional to $Ch^{-(N-1)}$, hence

$$h^{-2}\|\varphi_h - I_h\varphi\|^2_{L^2(\Omega_h)} \le Ch^{-1} \max_{x_i\in\partial\Omega_h} \left|\varphi_h(x_i) - I_h\varphi(x_i)\right|^2 = Ch^{-1} \max_{x_i\in\partial\Omega_h} \left|\varphi(x_i)\right|^2.$$

We find due to $\varphi = 0$ on $\partial\Omega$, $\operatorname{dist}(\partial\Omega_h, \partial\Omega) \le Ch^c$ and the boundedness of $\nabla\varphi$ that the term on the right-hand side is bounded by Ch^{-1+2c}. After taking square roots this concludes the proof as $c > \frac{1}{2}$. □

Theorem 10 *Let* (11) *hold. We have*

$$\lim_{h\to 0^+} \|(\bar{u}_{\gamma,\delta,h}, \bar{y}_{\gamma,\delta,h}, \bar{p}_{\gamma,\delta,h}) - (\bar{u}_{\gamma,\delta}, \bar{y}_{\gamma,\delta}, \bar{p}_{\gamma,\delta})\|_{L^2(\Omega)^3} = 0,$$

where $\bar{u}_{\gamma,\delta,h}$, $\bar{y}_{\gamma,\delta,h}$ *and* $\bar{p}_{\gamma,\delta,h}$ *are extended by zero to* Ω.

Proof By the optimality of the function values it is easy to see that there exists a constant $C > 0$, independent of h, such that $\|\bar{u}_{\gamma,\delta,h}\|_{H^1(\Omega_h)} \le C$. Using extension by zero we now find that $\|\bar{y}_{\gamma,\delta,h}\|_{H^1(\Omega)} \le C$.

Let $(h_n)_{n\in\mathbb{N}}$ be a zero sequence. After taking a subsequence, not relabeled, we may assume that it is monotonically decreasing. From the compact embedding of $H_0^1(\Omega)$ into $L^2(\Omega)$ and the reflexivity of $H_0^1(\Omega)$ we obtain a subsequence and a $\hat{y} \in H_0^1(\Omega)$ such that $\bar{y}_{\gamma,\delta,h_n} \xrightarrow{n\to\infty} \hat{y}$ strongly in $L^2(\Omega)$ and weakly in $H_0^1(\Omega)$. Extending $\bar{u}_{\gamma,\delta,h}$ by 0 to Ω and using the fact that $L^2(\Omega)$ is a Hilbert space we obtain on a subsequence, denoted the same way, that $\bar{u}_{\gamma,\delta,h_n} \cdot 1_{\Omega_{h_n}} \xrightarrow{n\to\infty} \hat{u}$ weakly in $L^2(\Omega)$ for some $\hat{u} \in L^2(\Omega)$. Let $\varphi \in C_c^\infty(\Omega)$ and φ_{h_n} be defined as in Lemma 14. We then have

$$0 = A(\bar{y}_{\gamma,\delta,h_n})\varphi_{h_n} - (\bar{u}_{\gamma,\delta,h_n}, \varphi_{h_n})_{L^2(\Omega_h)} \xrightarrow{n\to\infty} A(\hat{y})\varphi - (\hat{u}, \varphi)_{L^2(\Omega)}.$$

Thus $\hat{y} = S\hat{u}$ by the density of $C_c^\infty(\Omega)$ in $H_0^1(\Omega)$. The analogous arguments show that the adjoints converge in the same way to some $\hat{p} \in H_0^1(\Omega)$ with $\hat{p} = S^*(\hat{y} - y_\Omega)$. It therefore remains to show that $(\hat{u}, \hat{y}, \hat{p})$ is the unique optimal triple to ($\mathrm{ROC}_{\gamma,\delta}$). We will use Theorem 3 for that. Let $u \in H^1(\Omega) \cap C^\infty(\bar{\Omega})$ and $I_h u \in H^1(\Omega_h)$ denote the usual nodal interpolant. Then it is well-known, e.g. [25, Theorem 1.103], that $\|u - I_{h_n} u\|_{H^1(\Omega_{h_n})} \xrightarrow{n\to\infty} 0$. Moreover, it is straightforward to see that $\hat{u}$ and $\hat{p}$ satisfy (2) iff $\hat{u}$ minimizes

$$H^1(\Omega) \ni u \mapsto \frac{\gamma}{2}\|u\|^2_{H^1(\Omega)} + \beta \int_\Omega \sqrt{\delta + |\nabla u|^2}\,\mathrm{d}x + (\hat{p}, u)_{L^2(\Omega)} =: G(u).$$

By (13) we also have that $\bar{u}_{\gamma,\delta,h}$ minimizes G_h, which is defined analogously to G with Ω_h and $\bar{p}_{\gamma,\delta,h}$. Let $\hat{n} \in \mathbb{N}$ be arbitrary and $n \ge \hat{n}$. We therefore find by $h_{\hat{n}} \ge h_n$ that $\Omega_{h_{\hat{n}}} \subset \Omega_{h_n}$ and

$$G_{h_{\hat{n}}}(\bar{u}_{\gamma,\delta,h_n}) \le G_{h_n}(\bar{u}_{\gamma,\delta,h_n}) \le G_{h_n}(I_{h_n} u).$$

By $\|u - I_{h_n} u\|_{H^1(\Omega_{h_n})} \xrightarrow{n\to\infty} 0$, $|\Omega \setminus \Omega_{h_n}| \xrightarrow{n\to\infty} 0$ and $\bar{p}_{\gamma,\delta,h_n} \xrightarrow{n\to\infty} \hat{p}$ in $L^2(\Omega)$ we obtain

$$\limsup_{n\to\infty} G_{h_{\hat{n}}}(\bar{u}_{\gamma,\delta,h_n}) \le G(u).$$

As in previous arguments we have $1_{\Omega_{h_n}} \bar{u}_{\gamma,\delta,h_n} \xrightarrow{n\to\infty} 1_{\Omega_{h_n}} \hat{u}$ weakly in $H^1(\Omega_{h_{\hat{n}}})$ and strongly in $L^2(\Omega_{h_{\hat{n}}})$. Since $u \mapsto \int_{\Omega_{h_{\hat{n}}}} \sqrt{\delta + |\nabla u|^2}\,\mathrm{d}x$ is weakly lower semicontinuous with respect to the $L^2(\Omega_{h_{\hat{n}}})$ norm, cf. [1, Section 2], we obtain

$$G_{h_{\hat{n}}}(\hat{u}) \le \liminf_{n\to\infty} G_{h_{\hat{n}}}(\bar{u}_{\gamma,\delta,h_n}) \le \limsup_{n\to\infty} G_{h_{\hat{n}}}(\bar{u}_{\gamma,\delta,h_n}) \le G(u).$$

Sending $\hat{n} \to \infty$ shows that $\hat{u}$ is a minimizer of G, hence $G'(\hat{u}) = 0$, which implies the condition of Theorem 3. Together with $\hat{y} = S\hat{u}$ and $\hat{p} = S^*(\hat{y} - y_\Omega)$ this demonstrates $(\hat{u}, \hat{y}, \hat{p}) = (\bar{u}_{\gamma,\delta}, \bar{y}_{\gamma,\delta}, \bar{p}_{\gamma,\delta})$, thereby concluding the proof. □

Corollary 1 *Let* (11) *hold. We have*

$$\lim_{h\to 0^+} \|(\bar{y}_{\gamma,\delta,h}, \bar{p}_{\gamma,\delta,h}) - (\bar{y}_{\gamma,\delta}, \bar{p}_{\gamma,\delta})\|_{H^1(\Omega)^2} = 0,$$

where $\bar{y}_{\gamma,\delta,h}$ *and* $\bar{p}_{\gamma,\delta,h}$ *are extended by zero to* Ω.

Proof Let $R_h\bar{y} \in Y_h$ denote the Ritz projection with respect to A. Extending $\bar{y}_{\gamma,\delta,h} \in Y_h$ and $R_h\bar{y}$ by zero to Ω we clearly have

$$\|\bar{y}_{\gamma,\delta,h} - \bar{y}\|_{H^1(\Omega)} \le \|\bar{y}_{\gamma,\delta,h} - R_h\bar{y}\|_{H^1(\Omega_h)} + \|R_h\bar{y} - \bar{y}\|_{H^1(\Omega)}.$$

By definition, $\bar{y}_{\gamma,\delta,h} - R_h\bar{y}$ satisfies

$$A(\bar{y}_{\gamma,\delta,h} - R_h\bar{y})(\varphi_h) = (\bar{u}_{\gamma,\delta,h} - \bar{u}, \varphi_h)_{L^2(\Omega_h)} \qquad \forall \varphi_h \in Y_h.$$

Thus, choosing $\varphi_h = \bar{y}_{\gamma,\delta,h} - R_h\bar{y}$ and using the ellipticity of $\mathcal{A}$ and $c_0 \ge 0$ in Ω together with the Poincaré inequality in Ω yields a constant $C > 0$, independent of h, such that $\|\bar{y}_{\gamma,\delta,h} - R_h\bar{y}\|_{H^1(\Omega)} \le C\|\bar{u}_{\gamma,\delta,h} - \bar{u}\|_{L^2(\Omega)} \xrightarrow{h\to 0^+} 0$, where we also used extension by zero and Theorem 10. Since $R_h\bar{y} \xrightarrow{h\to 0^+} \bar{y}$ in Y, the $H^1(\Omega)$ convergence $\bar{y}_{\gamma,\delta,h} \xrightarrow{h\to 0^+} \bar{y}$ follows. The proof for $\bar{p}_{\gamma,\delta,h} - \bar{p}$ is analogue. □

6 Numerical solution

Based on the Finite Element approximation from section 5 we now study an inexact Newton method to compute the discrete solution $(\bar{y}_{\gamma,\delta,h}, \bar{p}_{\gamma,\delta,h}, \bar{u}_{\gamma,\delta,h})$ and we embed it into a practical path-following method.

6.1 A preconditioned inexact Newton method for the discrete problems

In this subsection we prove local convergence of an inexact Newton method when applied to a discretized version of (7) for fixed $(\gamma,\delta) \in \mathbb{R}^2_{>0}$. To this end, let us introduce the discrete adjoint-to-control mapping u_h. (We recall that the constant $h_0 > 0$ is introduced at the beginning of section 5.)

Lemma 15 *Let $h \in (0, h_0]$. For every $p \in L^2(\Omega_h)$ there exists a unique $u_h = u_h(p) \in V_h$ that satisfies the following discrete version of* (3)

$$\Big(\gamma\nabla u_h + f\big(\nabla u_h\big), \nabla\varphi_h\Big)_{L^2(\Omega_h)} + \gamma\big(u_h, \varphi_h\big)_{L^2(\Omega_h)} = (p, \varphi_h)_{L^2(\Omega_h)} \quad \forall \varphi \in V_h. \tag{12}$$

The associated solution operator $u_h : L^2(\Omega_h) \to V_h$ is Lipschitz continuously Fréchet differentiable. Its derivative $u_h'(p) \in \mathcal{L}(L^2(\Omega_h), V_h)$ at $p \in L^2(\Omega_h)$ in direction $d \in L^2(\Omega_h)$ is given by $z_h = u_h'(p)d \in V_h$, where z_h is the unique solution to

$$\left(\Big[\gamma I + f'\big(\nabla u_h(p)\big)\Big]\nabla z_h, \nabla\varphi_h\right)_{L^2(\Omega_h)} + \gamma\big(z_h, \varphi_h\big)_{L^2(\Omega_h)} = (d, \varphi_h)_{L^2(\Omega_h)} \quad \forall \varphi \in V_h. \tag{13}$$

Proof The proof is similar to the continuous case, but easier, so we omit it. □

With u_h at hand we can discretize (7) by

$$F_h : Y_h \times Y_h \to Y_h^* \times Y_h^*, \qquad F_h(y,p) := \begin{pmatrix} Ay - u_h(-p) \\ y - y_{\Omega_h} - A^* p \end{pmatrix}.$$

The same F_h is obtained if we consider the optimality conditions of ($\mathrm{ROC}_{\gamma,\delta,h}$) and express them in terms of (y,p). Moreover, $(\bar{y}_{\gamma,\delta,h}, \bar{p}_{\gamma,\delta,h})$ is the unique root of F_h and the properties of F from Lemma 12 and Lemma 13 carry over to F_h.

Lemma 16 *Let $h \in (0, h_0]$. The map $F_h : Y_h \times Y_h \to Y_h^* \times Y_h^*$ is Lipschitz continuously Fréchet differentiable. Its derivative at $(y,p) \in Y_h \times Y_h$ is given by*

$$F_h'(y,p) \colon Y_h \times Y_h \to Y_h^* \times Y_h^*, \qquad (\delta y, \delta p) \mapsto \begin{pmatrix} A & u_h'(-p) \\ I & -A^* \end{pmatrix} \begin{pmatrix} \delta y \\ \delta p \end{pmatrix}.$$

Moreover, $F_h'(y,p)$ is invertible for every $(y,p) \in Y_h$.

Proof The differentiability follows from Lemma 15. Since $\dim(Y_h \times Y_h) = \dim(Y_h^* \times Y_h^*)$, it is sufficient to show that $F_h'(y,p)$ is injective. This can be done exactly as in Lemma 13. □

Similar to Theorem 9 we have the following result.

Theorem 11 *Let $h \in (0, h_0]$ and $\eta \in [0, \infty)$. Then there is a neighborhood $N \subset Y_h \times Y_h$ of $(\bar{y}_{\gamma,\delta,h}, \bar{p}_{\gamma,\delta,h})$ such that for any $(y_0, p_0) \in N$ any sequence (y_k, p_k) that is generated according to $(y_{k+1}, p_{k+1}) = (y_k, p_k) + (\delta y_k, \delta p_k)$, where $(\delta y_k, \delta p_k) \in Y_h \times Y_h$ satisfies for all $k \geq 0$*

$$\|F_h(y_k, p_k) + F_h'(y_k, p_k)(\delta y_k, \delta p_k)\| \leq \eta_k \|F_h(y_k, p_k)\|$$

with $(\eta_k) \subset [0, \eta]$, converges r-linearly [q-linearly/q-superlinearly/with q-order $1+\omega$] to $(\bar{y}_{\gamma,\delta,h}, \bar{p}_{\gamma,\delta,h})$, provided $\eta < 1$ [η is sufficiently small/$\eta_k \to 0$/$\eta_k = O(\|F_h(y_k, p_k)\|^\omega)$]. Here, $\omega \in (0, 1]$ is arbitrary.

As a preconditioner for the fully discrete Newton system

$$F_h'(y,p) = \begin{pmatrix} \mathbf{A} & \mathbf{u_h'}(-\mathbf{p}) \\ \mathbf{M} & -\mathbf{A}^T \end{pmatrix} \quad \text{one can use} \quad \hat{\mathcal{P}} := \begin{pmatrix} \operatorname{diag}(\mathbf{A}) & 0 \\ \mathbf{M} & -\operatorname{diag}(\mathbf{A}^T) \end{pmatrix}. \tag{14}$$

It is sparse, cheaply invertible, and it does not change for fixed discretization. In [47] it is shown that diagonal preconditioning has a favorable effect on the distribution of the eigenvalues for Galerkin matrices. Our numerical experiments suggest that it can also be sensible to employ better approximations than $\operatorname{diag}(\mathbf{A})$ in $\hat{\mathcal{P}}$, e.g., a (modified incomplete) LU factorization of A.

Algorithm 2: Inexact path-following inexact Newton method

Input: $(\hat{y}_0, \hat{p}_0) \in Y_h \times Y_h$, $(\gamma_0, \delta_0) \in \mathbb{R}^2_{>0}$, $\kappa > 0$

1 **for** $i = 0, 1, 2, \dots$ **do**

2 set $(y_0, p_0) := (\hat{y}_i, \hat{p}_i)$

3 **for** $k = 0, 1, 2, \dots$ **do**

4 **if** $\|F_h(y_k, p_k)\| \leq \rho(\gamma_i, \delta_i)$ **then**

5 set $(\hat{y}_{i+1}, \hat{p}_{i+1}) := (y_k, p_k)$

6 **go to** line 11

7 **end**

8 choose $\eta_k > 0$ and use preconditioned GMRES to determine $(\delta y_k, \delta p_k)$ such that $\|r_k\| \leq \eta_k \|F_h(y_k, p_k)\|$

9 **call** Algorithm 3, input $w_k := (y_k, p_k)$, $\delta w_k := (\delta y_k, \delta p_k)$; output: λ_k

10 set $(y_{k+1}, p_{k+1}) := (y_k, p_k) + \lambda_k(\delta y_k, \delta p_k)$

11 **end**

12 select $\sigma_i \in (0, 1)$

13 **if** $\|(\hat{y}_{\iota+1}, \beta^{-1}\hat{p}_{\iota+1}) - (\hat{y}_\iota, \beta^{-1}\hat{p}_\iota)\|_{H^1} \leq (1 - \sigma_\iota)\kappa\|(\hat{y}_{i+1}, \beta^{-1}\hat{p}_{i+1})\|_{H^1}$ *for* $\iota = i, i-1$ **then** set $(y^*, p^*) := (\hat{y}_{i+1}, \hat{p}_{i+1})$; **stop**

14 set $(\gamma_{i+1}, \delta_{i+1}) := (\sigma_i \gamma_i, \sigma_i \delta_i)$

15 **end**

Output: (y^*, p^*)

6.2 A practical path-following method

The following Algorithm 2 is a practical path-following inexact Newton method to solve $(\text{ROC}_{\gamma,\delta,h})$. We use the residual $r_k := F_h(y_k, p_k) + F_h'(y_k, p_k)(\delta y_k, \delta p_k)$.

The function $\rho : \mathbb{R}^2_{>0} \to \mathbb{R}_{>0}$ prescribes how small the Newton residual should be for fixed γ, δ. In the implementation we use $\rho(\gamma, \delta) = \max\{10^{-6}, \gamma\}$, which may be viewed as inexact path-following. For the forcing term η_k we use the two choices $\eta_k = \bar{\eta}_k := 10^{-6}$ and $\eta_k = \hat{\eta}_k := \max\{10^{-6}, \min\{10^{-k-1}, \sqrt{\delta_i}\}\}$, where $k = k(i)$. For $\bar{\eta}_k$ we have $\bar{\eta}_k \leq \|F_h(y_k, p_k)\|$ since we terminate the inner loop if $\|F_h(y_k, p_k)\| < 10^{-6}$. Theorem 11 therefore suggests quadratic convergence for the choice $\eta_k = \bar{\eta}_k$ and this can indeed be observed. Similarly, $\eta_k = \hat{\eta}_k$ corresponds to superlinear convergence. For both choices, however, we found in the numerical experiments that it is more efficient to also terminate GMRES if the Euclidean norm of r_k drops below $\bar{\eta}_k$, respectively, $\hat{\eta}_k$ although this can prevent quadratic, respectively, superlinear convergence.

The control $u_h(-p_k)$ is computed with a globalized Newton method. The method terminates when the Newton residual falls below a threshold that decreases with (γ_i, δ_i). The linear systems are solved using SciPy's sparse direct solver `spsolve`. As an alternative we experimented with a preconditioned conjugate gradients method (PCG). The results were mixed: While the use of PCG diminished the total runtime of Algorithm 2 if all went well, we observed on several instances that it broke down for smaller values of (γ_i, δ_i).

We choose σ_i based on the number of Newton steps that are needed to compute the implicit controls $\{u_h(-p_k)\}_k$ in outer iteration i. If this number surpasses a predefined $m \in \mathbb{N}$, then we choose $\sigma_i > \sigma_{i-1}$. If it belongs to $[0, 0.75m]$, then we choose $\sigma_i < \sigma_{i-1}$. Otherwise, we let $\sigma_i = \sigma_{i-1}$. In addition, we respect the bound $\sigma_i \geq 0.25$ for all i, since we found in the numerical experiments, cf. Table 3 below, that choosing σ_i too small can prevent convergence in some cases. The weighing $1/\beta$ in the termination criterion is made since the amplitude of the adjoint state is roughly of order β in comparison to the state. In all experiments we use $\kappa = 10^{-3}$.

Algorithm 3 augments Algorithm 2 by a non-monotone line search globalization introduced in [35]. The non-monotonicity allows to always accept the inexact Newton step and yields potentially larger step sizes than descent-based strategies. The intention is to keep the number of trial step sizes low since every trial step size requires the evaluation of F_h and hence a recomputation of $u_h(-p_k)$. Assuming for simplicity that $u_h(-p_k)$ is determined exactly for each k, it is possible to show convergence of (y_k, p_k) from arbitrary starting points and to prove that eventually step size 1 will be accepted, which in turn ensures that the convergence rates of Theorem 11 are available for every fixed (γ_i, δ_i). In the numerical experiments we use $\tau = 10^{-4}$ and we observe that in the vast majority of iterations full steps are taken.

All norms without index in Algorithm 2 and 3 are $L^2(\Omega_h)$ norms.

Algorithm 3: Computation of step size

Input: $(w_k, \delta w_k)$, $\tau > 0$
for $l = 0, 1, 2, \ldots$ **do**
  if $\|F_h(w_k + 2^{-l}\delta w_k)\| \leq \left(1 + \frac{1}{(l+1)^2}\right)\|F_h(w_k)\| - \tau\|2^{-l}\delta w_k\|^2$ **then** set $\lambda_k := 2^{-l}$; **stop**
end
Output: λ_k

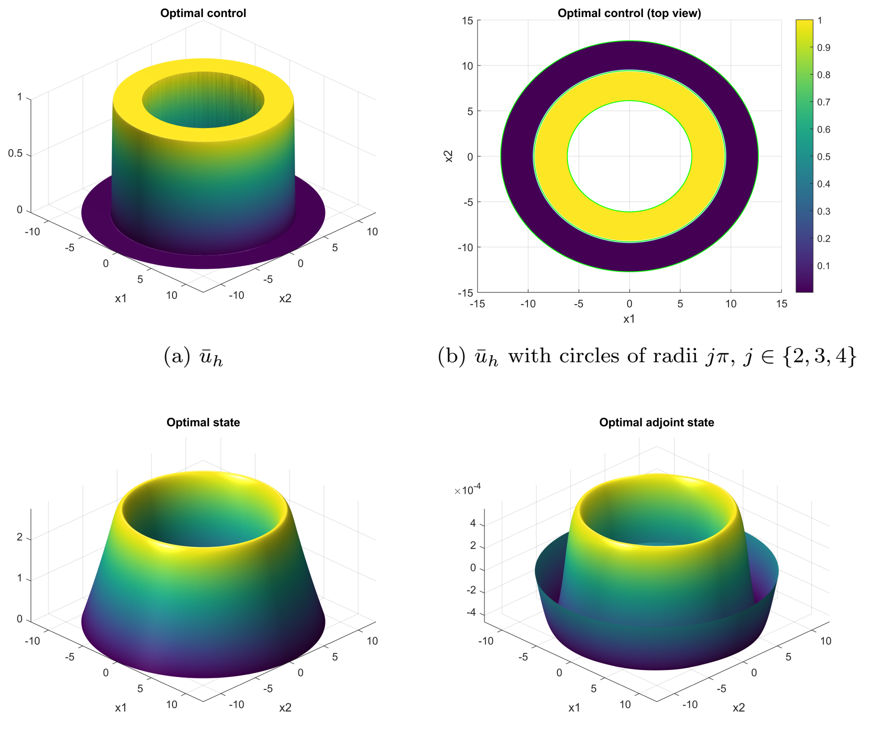
7 Numerical results

We provide numerical results for two examples. Our main goal is to illustrate that Algorithm 2 can robustly compute accurate solutions of (OC). The results are obtained from a Python implementation of Algorithm 2 using DOLFIN [38,39], which is part of FEniCS [3,37]. The code for the second example is available at `https://imsc.uni-graz.at/mannel/publications.php`.

7.1 Example 1: An example with explicit solution

The first example has an explicit solution and satisfies the assumptions used in this work. We consider (OC) for an arbitrary $\beta > 0$ with non-convex C^∞ domain $\Omega = B_{4\pi}(0) \setminus \overline{B_{2\pi}(0)}$ in $\mathbb{R}^2$, $\mathcal{A} = -\Delta$ and $c_0 \equiv 0$. The desired state is

$$y_\Omega(r) = \frac{\beta}{2r^3}\Big((1+r)\sin(r) - 1 - (2r^2-1)\cos(r)\Big) + \bar{y}$$



(a) $\bar{u}_h$ (b) $\bar{u}_h$ with circles of radii $j\pi$, $j \in \{2, 3, 4\}$

(c) $\bar{y}_h$ (d) $-\bar{p}_h$

Fig. 1: Numerically computed optimal solutions for Example 1

where $r(x, y) = \sqrt{x^2 + y^2}$, and the optimal state $\bar{y}$ is

$$\bar{y}(r) = \begin{cases} -\frac{r^2}{4} + A \ln(r/(4\pi)) + B & \text{if } r \in (2\pi, 3\pi), \\ C \ln(r/(4\pi)) & \text{if } r \in (3\pi, 4\pi) \end{cases}$$

with constants A, B, C whose values are contained in appendix D. The optimal control is

$$\bar{u}(r) = 1_{(2\pi,3\pi)}(r).$$

The optimal value is $j(\bar{u}) \approx 24.85\beta^2 + 59.22\beta$. In appendix D we provide details on the construction of this example and verify that $(\bar{y}, \bar{u})$ is indeed the optimal solution of (OC) If not stated otherwise, then $\beta = 10^{-3}$ is employed.

We use unstructured triangulations that approximate $\partial\Omega$ increasingly better as the meshes become finer, cf. (11). Figure 1 depicts the optimal control $\bar{u}_h$, optimal state $\bar{y}_h$ and negative optimal adjoint state $-\bar{p}_h$, which were computed by Algorithm 2 on a grid with 1553207 degrees of freedom (DOF).

We begin by studying convergence on several grids. We use the fixed ratio $(\gamma_i/\delta_i) \equiv 10^2$ and apply Algorithm 2 with $(\gamma_0, \delta_0) = (1, 0.01)$ and $(\hat{y}_0, \hat{p}_0) = (0, 0)$. We use the two forcing terms $\eta_k = 10^{-6} =: \bar{\eta}_k$ for all k (and all i) and

Table 1: Example 1: Number of Newton steps and errors for several meshes; the first value is for the forcing term $\bar{\eta}_k$, the second for $\hat{\eta}_k$ (only shown if different)

DOF	γ_{final}	#it	#it$_u$	$\mathcal{E}_j$	$\mathcal{E}_u$	$\mathcal{E}_y$	$\mathcal{E}_p$
1588	$1.2 \times 10^{-8}/9.3 \times 10^{-9}$	58/72	390/428	3.4×10^{-2}	18.7	2.3	6.0×10^{-2}
6251	$1.7/1.6 \times 10^{-10}$	78/91	597/608	4.0×10^{-3}	7.3	1.1	3.3×10^{-2}
24443	$2.1/1.5 \times 10^{-11}$	55/64	454/491	9.4×10^{-4}	5.1	0.50	1.3×10^{-2}
97643	$5.4/6.2 \times 10^{-11}$	46/48	407/380	3.3×10^{-4}	3.7	0.22	5.6×10^{-3}
389027	$4.6/4.3 \times 10^{-10}$	32/34	367/358	1.2×10^{-4}	2.8	0.09	2.9×10^{-3}

Table 2: Example 1: Course of Algorithm 2

i	γ_i	σ_i	(#it^i,#it$_u^i$)	$\mathcal{E}_j^i$	$\mathcal{E}_u^i$	$\mathcal{E}_y^i$	$\mathcal{E}_p^i$	τ^i	τ_u^i
0/1/2	1.0/0.45/0.18	0.45/0.41/0.37	$(0,0)$	575	155	38.5	9.8×10^{-3}	0	0
3	6.8×10^{-2}	0.33	$(1,1)$	5.7	44	1.8	1.4	1440	11.3
4	2.2×10^{-2}	0.30	$(1,2)$	2.0	38	1.0	0.48	959	0.83
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
12	8.7×10^{-6}	0.32	$(3,18)$	1.8×10^{-3}	5.1	0.23	8.8×10^{-3}	3.5	0.22
13	2.8×10^{-6}	0.31	$(3,20)$	8.7×10^{-4}	4.3	0.23	6.8×10^{-3}	2.0	0.15
14	8.7×10^{-7}	0.28	$(3,18)$	5.1×10^{-4}	3.9	0.22	6.0×10^{-3}	0.84	0.073
15	2.4×10^{-7}	0.26	$(5,20)$	3.8×10^{-4}	3.7	0.22	5.7×10^{-3}	0.32	0.027
16	6.4×10^{-8}	0.25	$(3,15)$	3.5×10^{-4}	3.7	0.22	5.6×10^{-3}	0.13	8.2×10^{-3}
17	1.6×10^{-8}	0.25	$(3,16)$	3.3×10^{-4}	3.7	0.22	5.6×10^{-3}	0.081	2.2×10^{-3}
18	4.0×10^{-9}	0.25	$(3,15)$	3.3×10^{-4}	3.7	0.22	5.6×10^{-3}	0.060	7.0×10^{-4}
19	9.9×10^{-10}	0.25	$(3,20)$	3.3×10^{-4}	3.7	0.22	5.6×10^{-3}	0.045	3.6×10^{-4}
20	2.5×10^{-10}	0.25	$(3,22)$	3.3×10^{-4}	3.7	0.22	5.6×10^{-3}	0.028	2.2×10^{-4}
21	6.2×10^{-11}	—	$(3,33)$	3.3×10^{-4}	3.7	0.22	5.6×10^{-3}	0.017	1.4×10^{-4}

$\eta_k = \max\{10^{-6}, \min\{10^{-k-1}, \sqrt{\delta_i}\}\} =: \hat{\eta}_k$. Table 1 shows #it, which represents the total number of inexact Newton steps for (y,p), and #it$_u$, which is the total number of Newton steps used to compute the implicit function u. Table 1 also contains the errors

$$\mathcal{E}_j := \left| j_{\gamma_{\text{final}},\delta_{\text{final}},h} - \bar{j} \right|, \qquad \mathcal{E}_u := \|\hat{u}_{\text{final}} - \bar{u}\|_{L^1(\Omega_*)},$$

as well as

$$\mathcal{E}_y := \|\hat{y}_{\text{final}} - \bar{y}\|_{H^1(\Omega_*)}, \qquad \mathcal{E}_p := \|\hat{p}_{\text{final}} - \bar{p}\|_{H^1(\Omega_*)}.$$

where Ω_* represents a reference grid with DOF = 1553207. To evaluate the errors, $\hat{u}_{\text{final}}$, $\hat{y}_{\text{final}}$ and $\hat{p}_{\text{final}}$ are extended to Ω_* using extrapolation. Table 2 provides details for the run from Table 1 with DOF = 97643 and $\eta_k = \hat{\eta}_k$. Table 2 includes $\tau^i := \|(\hat{y}_{i+1}, \beta^{-1}\hat{p}_{i+1}) - (\hat{y}_i, \beta^{-1}\hat{p}_i)\|_{H^1(\Omega_h)}$, which appears in the termination criterion of Algorithm 2, and also $\tau_u^i := \|u(\hat{p}_{i+1}) - u(\hat{p}_i)\|_{L^2(\Omega_h)}$.

Table 1 indicates convergence of the computed solutions $(\hat{u}_{\text{final}}, \hat{y}_{\text{final}}, \hat{p}_{\text{final}})$ to $(\bar{u}, \bar{y}, \bar{p})$ and of the objective value $j_{\gamma_{\text{final}},\delta_{\text{final}},h}$ to $\bar{j}$. It also suggests that convergence takes place at certain rates with respect to h. Moreover, the total number of Newton steps both for (y,p) and for u stays bounded as DOF increases, which may suggest mesh independence. The choice $\eta_k = \bar{\eta}_k$ frequently yields lower numbers of Newton steps for (y,p) and for u, yet the runtime (not depicted) is consistently higher than for $\eta_k = \hat{\eta}_k$ since more iterations

Table 3: Example 1: Results for fixed values $(\sigma_i) \equiv \sigma$; the first value is for the forcing term $\bar{\eta}_k$, the second for $\hat{\eta}_k$ (only shown if different)

σ	γ_{final}	#it	#it$_u$	$\mathcal{E}_j$	$\mathcal{E}_u$	$\mathcal{E}_y$	$\mathcal{E}_p$
0.2	6.6×10^{-12}	48/51	505/512	9.4×10^{-4}	5.1	0.50	1.3×10^{-2}
0.3	3.5×10^{-11}	51/57	408/405	9.4×10^{-4}	5.1	0.50	1.3×10^{-2}
0.5	2.3×10^{-10}	70/86	454/466	9.4×10^{-4}	5.1	0.50	1.3×10^{-2}
0.7	5.1×10^{-10}	97/130	522/552	9.4×10^{-4}	5.1	0.50	1.3×10^{-2}
0.9	$1.1/8.0 \times 10^{-9}$	276/474	1261/1329	$9.4/9.5 \times 10^{-4}$	5.1	0.50	1.3×10^{-2}

Table 4: Example 1: Results for a sequence of nested grids

DOF	γ_{final}	#it	#it$_u$	$\mathcal{E}_j$	$\mathcal{E}_u$	$\mathcal{E}_y$	$\mathcal{E}_p$
6655	6.7×10^{-3}	7	39	6.5×10^{-1}	30	1.2	1.8×10^{-1}
25596	4.8×10^{-5}	13	175	7.7×10^{-3}	8.3	0.51	1.7×10^{-2}
100336	8.6×10^{-7}	16	117	5.5×10^{-4}	4.1	0.23	5.6×10^{-3}
397248	4.4×10^{-11}	34	387	1.4×10^{-4}	2.8	0.09	3.2×10^{-3}

of GMRES are required to compute the step for (y, p). Specifically, using $\hat{\eta}_k$ saves between 5% and 36% of runtime, with 36% being the saving on the finest grid. In the vast majority of iterations, step size 1 is accepted for (y_k, p_k). For instance, all of the 52 iterations required for DOF $= 97643$ and $\eta_k = \hat{\eta}_k$ use full steps; for DOF $= 6251$ and $\eta_k = \bar{\eta}_k$, 86 of the 87 iterations use step size 1.

Table 3 displays the effect of fixing $(\sigma_i) \equiv \sigma$ in Algorithm 2. The mesh uses DOF $= 24443$ and is the same as in Table 1.

For both forcing terms, $\sigma = 0.3$ yields the lowest runtime. In comparison, the adaptive choice of σ_i that we employ requires about 6% more runtime. For $\sigma = 0.1$ the iterates failed to converge for both forcing terms once $\gamma_{12} = 10^{-12}$ is reached because $u_h(-p_k)$ could not be computed to sufficient accuracy within the 200 iterations that we allow for this process. Together with Table 3 this shows that small values of σ_i can increase the number of steps required for u and even prevent convergence. We therefore let $\sigma_i \geq 0.25$ for all i in all experiments, although this diminshes the efficacy of Algorithm 2 in some cases.

Table 4 shows results for $\eta_k = \hat{\eta}_k$ and a sequence of nested grids, where the grids are refined once $\gamma_i < 10^{-2}$, $\gamma_i < 10^{-4}$ and $\gamma_i < 10^{-6}$, respectively.

We note that the errors $\mathcal{E}_j$, $\mathcal{E}_u$, $\mathcal{E}_y$ and $\mathcal{E}_p$ in the last line of Table 4 are of similar size as their counterparts in the last line of Table 1. Since the iteration numbers in these lines are similar as well, the variant on the fixed grid somewhat surprisingly requires a lower runtime than the nested variant. The reason is that the computation of $u_h(-p_0)$ after the last grid refinement at $\gamma_i = 8.6 \times 10^{-7}$ requires 200 iterations. We leave the issue of reducing this large number (and correspondingly the runtime) as a future topic and mention that, in contrast, in example 2 the usage of nested grids is clearly advantageous.

Table 5: Example 1: Results for various values of β; the first line is for the choice $(\gamma_i/\delta_i) \equiv 10^2$, the second for $(\gamma_i/\delta_i) \equiv 1$

β	10^{-1}	10^{-2}	10^{-4}	10^{-5}
$(\#\text{it},\#\text{it}_u)/\mathcal{E}_u$	$(28, 373)/21$	$(37, 217)/7.2$	$(100, 914)/4.3$	$(153, 1230)/4.1$
$(\#\text{it},\#\text{it}_u)/\mathcal{E}_u$	$(40, 435)/21$	$(78, 795)/7.3$	$(126, 996)/4.3$	$(137, 1147)/4.1$

Table 6: Example 1: Iteration numbers and errors for several ratios γ_i/δ_i; the computations for $\gamma_i/\delta_i \in \{10^{-1}, 10^{-2}\}$ use a lower accuracy

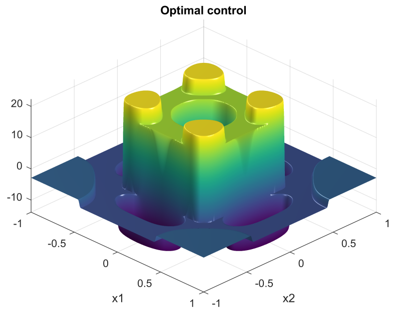
$\frac{\gamma_i}{\delta_i}$	δ_{final}	#it	$\#\text{it}_u$	$\mathcal{E}_j$	$\mathcal{E}_u$	$\mathcal{E}_y$	$\mathcal{E}_p$
10^{-2}	3.0×10^{-8}	71	288	9.9×10^{-4}	5.1	0.50	1.3×10^{-2}
10^{-1}	3.0×10^{-8}	58	246	9.9×10^{-4}	5.1	0.50	1.3×10^{-2}
1	1.8×10^{-12}	102	469	9.4×10^{-4}	5.1	0.50	1.3×10^{-2}
10^1	2.9×10^{-12}	80	396	9.4×10^{-4}	5.1	0.50	1.3×10^{-2}
10^2	2.3×10^{-12}	70	454	9.4×10^{-4}	5.1	0.50	1.3×10^{-2}
10^3	1.9×10^{-12}	59	475	9.4×10^{-4}	5.1	0.50	1.3×10^{-2}

We now turn to the robustness of Algorithm 2. We emphasize that in our numerical experience the robustness of algorithms for optimal control problems involving the TV seminorm in the objective is a delicate issue. Table 5 displays the iteration numbers required by Algorithm 2 for different values of β on the mesh with DOF = 24443 along with the error $\mathcal{E}_j$ for $\eta_k = \hat{\eta}_k$ for the two choices $(\gamma_i/\delta_i) \equiv 10^2$ and $(\gamma_i/\delta_i) \equiv 1$. The omitted values for $\beta = 10^{-3}$ and $(\gamma_i/\delta_i) \equiv 10^2$ are identical to those from Table 1 for DOF = 24443 and $\eta_k = \hat{\eta}_k$. Table 6 provides iteration numbers and errors for various fixed choices of (γ_i/δ_i) on the mesh with DOF = 24443 for $\beta = 10^{-3}$, $\eta_k = \bar{\eta}_k$ and $(\sigma_i) \equiv 0.5$. For the ratios 10^{-1} and 10^{-2} we increased κ from 10^{-3} to $5 \cdot 10^{-3}$ to obtain convergence. Since our goal is to demonstrate robustness, no further changes are made although this would lower the iteration numbers.

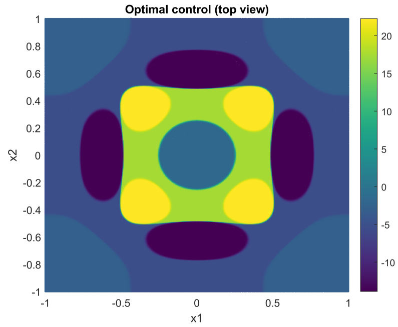
Table 5 and 6 suggest that Algorithm 2 is able to handle a range of parameter values without modification of its internal parameters.

7.2 Example 2

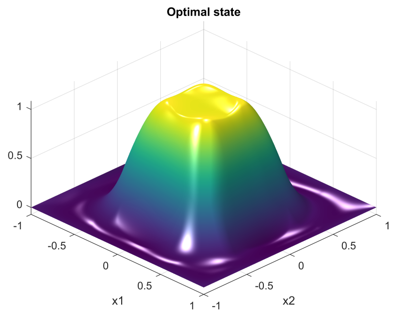
From section 3 onward we have required Ω to be of class $C^{1,1}$. To show that Algorithm 2 can still solve (OC) if Ω is only Lipschitz, we now consider an example from [22, section 4.2] on the square $\Omega = [-1, 1]^2$. We have $\mathcal{A} = -\Delta$, $c_0 \equiv 0$, $\beta = 10^{-4}$ and $y_\Omega = 1_D$, where $D = (-0.5, 0.5)^2$. We use uniform triangulations throughout this example and denote by $n + 1$ the number of nodes in coordinate direction. Figure 1 depicts the optimal control $\bar{u}_h$, optimal state $\bar{y}_h$ and negative optimal adjoint state $-\bar{p}_h$, which were computed with $n = 1024$. Apparently, $\bar{u}_h$ is piecewise constant.



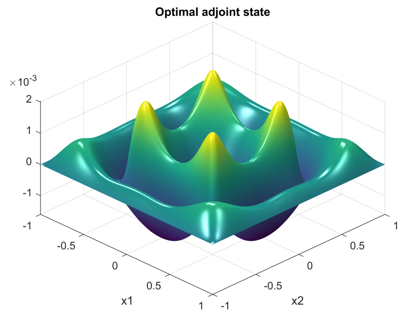
(a) $\bar{u}_h$



(b) $\bar{u}_h$ (top view)



(c) $\bar{y}_h$



(d) $-\bar{p}_h$

Fig. 2: Numerically computed optimal solutions for Example 2

Table 7: Example 2: Number of Newton steps and errors for several meshes

n	γ_{final}	#it	#it$_u$	$\mathcal{E}_j$	$\mathcal{E}_u$	$\mathcal{E}_y$	$\mathcal{E}_p$
32	1.7×10^{-11}	43	321	1.8×10^{-2}	8.6	0.75	9.0×10^{-3}
64	1.7×10^{-11}	48	551	9.9×10^{-3}	4.3	0.37	4.9×10^{-3}
128	3.2×10^{-11}	46	902	4.9×10^{-3}	2.3	0.19	2.4×10^{-3}
256	3.3×10^{-11}	50	1212	2.2×10^{-3}	1.1	0.081	1.1×10^{-3}
512	5.6×10^{-11}	58	2868	7.3×10^{-4}	0.42	0.031	4.2×10^{-4}

Throughout, we use the fixed ratio $(\gamma_i/\delta_i) \equiv 10^{-2}$ and apply Algorithm 2 with $(\gamma_0, \delta_0) = (0.01, 1)$ and $(\hat{y}_0, \hat{p}_0) = (0, 0)$. As in example 1, cf. Table 6, other ratios for γ_i/δ_i can be employed as well. We only provide results for $\bar{\eta}_k$ since the forcing term $\hat{\eta}_k$ does not yield lower runtimes in this example; both forcing terms produce the same errors, though. Table 7 displays iteration numbers and errors for different grids, while Table 8 shows details for $n = 256$.

Table 7 hints at possible mesh independence for (y, p), but suggests that the number of Newton steps for u increases with n. The depicted errors are

Table 8: Example 2: Course of Algorithm 2

i	γ_i	σ_i	$(\#\text{it}^i, \#\text{it}_u^i)$	$\mathcal{E}_j^i$	$\mathcal{E}_u^i$	$\mathcal{E}_y^i$	$\mathcal{E}_p^i$	τ^i	τ_u^i
0–4	0.01/...	0.45/...	(0, 0)	0.42	34	3.4	1.7×10^{-2}	0	0
5	6.7×10^{-5}	0.27	(2, 7)	6.0×10^{-2}	27	1.9	4.2×10^{-2}	498	7.3
6	1.8×10^{-5}	0.26	(2, 37)	3.4×10^{-2}	23	1.6	2.4×10^{-2}	212	3.4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
11	2.9×10^{-7}	0.55	(2, 59)	3.1×10^{-3}	12	0.50	4.6×10^{-3}	16.4	1.6
12	1.6×10^{-7}	0.55	(2, 59)	1.6×10^{-3}	9.6	0.39	3.4×10^{-3}	12.6	1.6
13	9.0×10^{-8}	0.53	(2, 56)	4.4×10^{-4}	7.4	0.29	2.5×10^{-3}	9.5	1.4
14	4.7×10^{-8}	0.53	(3, 70)	4.5×10^{-4}	5.3	0.21	1.9×10^{-3}	7.6	1.4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
21	9.5×10^{-10}	0.52	(3, 64)	2.1×10^{-3}	1.3	0.084	1.1×10^{-3}	0.62	0.32
22	5.0×10^{-10}	0.47	(2, 23)	2.1×10^{-3}	1.1	0.083	1.1×10^{-3}	0.38	0.24
23	2.4×10^{-10}	0.45	(2, 56)	2.1×10^{-3}	1.1	0.081	1.1×10^{-3}	0.28	0.27
24	1.1×10^{-10}	0.59	(2, 80)	2.1×10^{-3}	1.1	0.081	1.1×10^{-3}	0.15	0.17
25	6.2×10^{-11}	0.53	(2, 15)	2.2×10^{-3}	1.1	0.081	1.1×10^{-3}	0.062	0.052
26	3.3×10^{-11}	—	(2, 17)	2.2×10^{-3}	1.1	0.081	1.1×10^{-3}	0.042	0.037

Table 9: Example 2: Results for a sequence of nested grids

n	γ_{final}	#it	$\#\text{it}_u$	$\mathcal{E}_j$	$\mathcal{E}_u$	$\mathcal{E}_y$	$\mathcal{E}_p$
64	4.0×10^{-5}	5	26	3.9×10^{-2}	25	1.8	3.4×10^{-2}
128	4.8×10^{-7}	12	260	1.7×10^{-3}	14	0.64	6.3×10^{-3}
256	6.3×10^{-9}	19	481	1.7×10^{-3}	2.1	0.10	1.2×10^{-3}
512	5.6×10^{-11}	20	792	7.3×10^{-4}	0.42	0.031	4.2×10^{-4}

Table 10: Example 2: Results for various values of β. A sequence of nested grids is used and the displayed iteration numbers are for the finest grid only

β	10^{-3}	10^{-4}	10^{-5}	5×10^{-6}
$(\#\text{it}, \#\text{it}_u)/\mathcal{E}_j$	$(12, 117)/3.3 \times 10^{-3}$	$(22, 355)/2.8 \times 10^{-3}$	$(70, 958)/2.4 \times 10^{-3}$	$(104, 1569)/2.2 \times 10^{-3}$

computed by use of a reference solution that is obtained by Algorithm 2 with $\eta_k = \bar{\eta}_k$ on the mesh with $n = 1024$. As in the first example it seems that convergence with respect to h takes place at certain rates. The majority of iterations use full Newton steps for (y, p). For instance, all but one of the 50 iterations for $n = 256$ use step length one.

Table 9 shows the outcome of Algorithm 2 if a sequence of nested grids is used, where the grids are refined once $\gamma_i < 10^{-4}$, $\gamma_i < 10^{-6}$ and $\gamma_i < 10^{-8}$, respectively. This simple strategy reduces the runtime by about 57% while providing the same accuracy as a run for $n = 512$, cf. the last line of Table 7.

Table 10 adresses the robustness of Algorithm 2 with respect to β. The computations are carried out on nested grids and the displayed iteration numbers are those for the finest grid, which has $n = 128$. The reference solution is computed for $n = 256$. The final grid change happens once $\gamma_i < 10^{-8}$.

Table 10 indicates that Algorithm 2 is robust with respect to β. As in example 1 it is possible to achieve lower iteration numbers through manipulation of the algorithmic parameters. For instance, if the final grid change for $\beta = 10^{-5}$ happens once $\gamma_i < 10^{-9}$ instead of $\gamma_i < 10^{-8}$, then only $(41, 638)$ iterations are needed on the final grid instead of $(70, 958)$.

8 Summary

We have studied an optimal control problem with controls from BV in which the control costs are given by the TV seminorm. By smoothing the TV seminorm and adding an H^1 regularization term we obtained a family of auxiliary problems whose solutions converge to the optimal solution of the original problem in appropriate function spaces. For fixed smoothing and regularization parameter we showed local convergence of an infinite-dimensional inexact Newton method applied to a reformulation of the optimality system that involves the control as an implicit function of the adjoint state. Based on a convergent Finite Element approximation a practical algorithm was derived and it was demonstrated that the algorithm is able to robustly compute the optimal solution of the control problem with considerable accuracy. To verify this, a two-dimensional test problem with known solution was constructed.

A Differentiability of ψ_δ

Lemma 17 *Let $\delta > 0$, $N \in \mathbb{N}$ and let $\Omega \subset \mathbb{R}^N$ be open. The functional*

$$\psi_\delta : H^1(\Omega) \to \mathbb{R}, \qquad u \mapsto \int_\Omega \sqrt{\delta + |\nabla u|^2}\,\mathrm{d}x$$

is Lipschitz continuously Fréchet differentiable and twice Gâteaux differentiable. Its first derivative at u in direction v and its second derivative at u in directions v, w are given by

$$\psi_\delta'(u)v = \int_\Omega \frac{(\nabla u, \nabla v)}{\sqrt{\delta + |\nabla u|^2}}\,\mathrm{d}x \quad \textit{and} \quad \psi_\delta''(u)[v,w] = \int_\Omega \frac{(\nabla v, \nabla w)}{\sqrt{\delta + |\nabla u|^2}} - \frac{(\nabla u, \nabla v)\,(\nabla u, \nabla w)}{(\delta + |\nabla u|^2)^{\frac{3}{2}}}\,\mathrm{d}x.$$

Proof **First Gâteaux derivative**
Let $u, v \in H^1(\Omega)$. As $s \mapsto \sqrt{\delta + s}$ is Lipschitz on $[0, \infty)$ with constant $\frac{1}{2\sqrt{\delta}}$, we obtain for all $t \in [-1, 1]$, $t \neq 0$,

$$\left| \frac{\sqrt{\delta + |\nabla u + t\nabla v|^2} - \sqrt{\delta + |\nabla u|^2}}{t} \right| \le \frac{|\nabla v| \cdot (2\,|\nabla u| + |\nabla v|)}{2\sqrt{\delta}} \qquad \text{a.e. in } \Omega. \tag{15}$$

Thus, we can apply the theorem of dominated convergence, which yields

$$\lim_{t \to 0} \frac{\psi_\delta(u + tv) - \psi_\delta(u)}{t} = \int_\Omega \lim_{t \to 0} \frac{\sqrt{\delta + |\nabla u + t\nabla v|^2} - \sqrt{\delta + |\nabla u|^2}}{t}\,\mathrm{d}x = \int_\Omega \frac{(\nabla u, \nabla v)}{\sqrt{\delta + |\nabla u|^2}}\,\mathrm{d}x.$$

From

$$\left| \int_\Omega \frac{(\nabla u, \nabla v)}{\sqrt{\delta + |\nabla u|^2}}\,\mathrm{d}x \right| \le \int_\Omega \left| \frac{(\nabla u, \nabla v)}{\sqrt{\delta + |\nabla u|^2}} \right| \mathrm{d}x \le \frac{\|\nabla u\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)}}{\sqrt{\delta}} \le \frac{\|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)}}{\sqrt{\delta}}$$

we see that the functional $v \mapsto \psi_\delta'(u)v$ is linear and continuous.

Second Gâteaux derivative

Let $u, v, w \in H^1(\Omega)$. Since $g : \mathbb{R}^N \to \mathbb{R}$, $g(y) := \frac{(y,z)}{\sqrt{\delta+|y|^2}}$, with $z \in \mathbb{R}^N$ fixed, is Lipschitz continuous on $\mathbb{R}^N$ with constant $\frac{2}{\sqrt{\delta}}|z|$, we obtain for all $t \in \mathbb{R}$, $t \neq 0$,

$$\left|\frac{1}{t}\right| \left| \frac{(\nabla u + t\nabla w, \nabla v)}{\sqrt{\delta + |\nabla u + t\nabla w|^2}} - \frac{(\nabla u, \nabla v)}{\sqrt{\delta + |\nabla u|^2}} \right| \le \frac{2}{\sqrt{\delta}} |\nabla v||\nabla w| \qquad \text{a.e. in } \Omega. \tag{16}$$

Dominated convergence yields

$$\begin{aligned}\lim_{t\to 0} \frac{\psi_\delta'(u+tw)v - \psi_\delta'(u)v}{t} &= \int_\Omega \lim_{t\to 0} \frac{1}{t} \left(\frac{(\nabla u + t\nabla w, \nabla v)}{\sqrt{\delta + |\nabla u + t\nabla w|^2}} - \frac{(\nabla u, \nabla v)}{\sqrt{\delta + |\nabla u|^2}} \right) \mathrm{d}x \\ &= \int_\Omega \frac{(\nabla v, \nabla w)}{\sqrt{\delta + |\nabla u|^2}} - \frac{(\nabla u, \nabla v)\,(\nabla u, \nabla w)}{(\delta + |\nabla u|^2)^{\frac{3}{2}}} \,\mathrm{d}x,\end{aligned}$$

where we used the directional derivative of g to derive the last equality. From (16) we deduce the boundedness of the bilinear mapping $(v, w) \mapsto \psi_\delta''(u)[v, w]$ by

$$\left| \int_\Omega \frac{(\nabla v, \nabla w)}{\sqrt{\delta + |\nabla u|^2}} - \frac{(\nabla u, \nabla v)\,(\nabla u, \nabla w)}{(\delta + |\nabla u|^2)^{\frac{3}{2}}} \,\mathrm{d}x \right| \le \frac{2}{\sqrt{\delta}} \|v\|_{H^1(\Omega)} \|w\|_{H^1(\Omega)}. \tag{17}$$

Lipschitz continuous Fréchet differentiability

From (17) we infer that $\sup_{u\in H^1(\Omega)} \|\psi_\delta''(u)\|_{H^1(\Omega)^{**}} \le \frac{2}{\sqrt{\delta}}$, which implies that $u \mapsto \psi_\delta'(u)$ is Lipschitz with constant $\frac{2}{\sqrt{\delta}}$, hence $u \mapsto \psi_\delta(u)$ is Fréchet differentiable. □

B Hölder continuity for quasilinear partial differential equations

To prove results on the Hölder continuity of solutions to quasilinear elliptic PDEs, we first discuss linear elliptic PDEs.

Theorem 12 *Let $\alpha \in (0,1)$ and let Ω be a bounded $C^{1,\alpha}$ domain. Let $\gamma_0, \mu > 0$ be given. Let $A \in C^{0,\alpha}(\Omega, \mathbb{R}^{N\times N})$ be a uniformly elliptic matrix with ellipticity constant μ and let $\gamma \ge \gamma_0$. Let $a^0 > 0$ be such that $\gamma, \|A\|_{C^{0,\alpha}(\Omega)} \le a^0$. Then there is a constant $C > 0$ depending only on α, Ω, N, μ, a^0 and γ_0 such that for any $p \in L^\infty(\Omega)$ and any $f \in C^{0,\alpha}(\Omega, \mathbb{R}^N)$ the unique weak solution u to*

$$\begin{cases} -\operatorname{div}(A\nabla u) + \gamma u = p - \operatorname{div}(f) & \text{in } \Omega, \\ \partial_{A_\nu} u = 0 & \text{on } \Gamma, \end{cases} \tag{18}$$

satisfies $u \in C^{1,\alpha}(\Omega)$ and

$$\|u\|_{C^{1,\alpha}(\Omega)} \le C \left(\|p\|_{L^\infty(\Omega)} + \|f\|_{C^{0,\alpha}(\Omega)} \right).$$

Proof A standard ellipticity argument delivers unique existence and $\|u\|_{H^1(\Omega)} \le C\|p\|_{L^\infty(\Omega)}$, where C only depends on the claimed quantities. Moreover, by [46, Theorem 3.16(iii)]

$$\|u\|_{\mathcal{L}^{2,N+2\alpha}(\Omega)} + \|\nabla u\|_{\mathcal{L}^{2,N+2\alpha}(\Omega)} \le C\Big(\|p\|_{\mathcal{L}^{2,(N+2\alpha-2)^+}(\Omega)} + \|f\|_{\mathcal{L}^{2,N+2\alpha}(\Omega)} + \|u\|_{H^1(\Omega)} \Big).$$

Here, C depends on all of the claimed quantities except γ_0, and $\mathcal{L}^{2,\lambda}(\Omega)$ denotes a Campanato space; for details see [46, Chapter 1.4]. The definition of Campanato spaces implies $\|p\|_{\mathcal{L}^{2,(N+2\alpha-2)^+}(\Omega)} \le C\|p\|_{L^\infty(\Omega)}$. Using the isomorphism between $\mathcal{L}^{2,N+2\alpha}(\Omega)$ and $C^{0,\alpha}(\Omega)$ from [46, Theorem 1.17 (ii)] we obtain

$$\|u\|_{C^{1,\alpha}(\Omega)} \le C \left(\|p\|_{L^\infty(\Omega)} + \|f\|_{C^{0,\alpha}(\Omega)} + \|u\|_{H^1(\Omega)} \right).$$

The earlier ellipticity estimate concludes the proof. □

The next result follows directly from [36, Theorem 2] and requires no proof.

Theorem 13 *Let Ω be a bounded $C^{1,\alpha'}$ domain for some $\alpha' \in (0,1]$. Let $A : \Omega\times\mathbb{R}\times\mathbb{R}^N \to \mathbb{R}^N$, $B : \Omega \times \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}$, $M > 0$ and $0 < \lambda \le \Lambda$. Let $\kappa, m \ge 0$ and suppose that*

$$\sum_{i,j=1}^{N} \partial_{\eta_i} A_j(x,u,\eta)\xi_i\xi_j \ge \lambda\big(\kappa + |\eta|_2^2\big)^m |\xi|_2^2, \qquad \textit{(ellipticity)} \tag{19}$$

$$\sum_{i,j=1}^{N} |\partial_{\eta_i} A_j(x,u,\eta)| \le \Lambda\big(\kappa + |\eta|_2\big)^m, \qquad \textit{(boundedness of A)} \tag{20}$$

$$|B(x,u,\eta)| \le \Lambda\big(1 + |\eta|_2\big)^{m+2}, \qquad \textit{(boundedness of B)} \tag{21}$$

as well as the Hölder continuity property

$$|A(x_1,u_1,\eta) - A(x_2,u_2,\eta)| \le \Lambda\big(1+|\eta|_2\big)^{m+1}\big(|x_1-x_2|^{\alpha'} + |u_1-u_2|^{\alpha'}\big) \tag{22}$$

are satisfied for all $x, x_1, x_2 \in \Omega$, $u, u_1, u_2 \in [-M, M]$ and $\eta, \xi \in \mathbb{R}^N$. Then there exist constants $\alpha \in (0,1)$ and $C > 0$ such that each solution $u \in H^1(\Omega)$ of

$$\int_\Omega A(x,u,\nabla u)^T \nabla\varphi \,\mathrm{d}x = \int_\Omega B(x,u,\nabla u)\varphi \,\mathrm{d}x \qquad \forall \varphi \in H^1(\Omega)$$

satisfies

$$\|u\|_{C^{1,\alpha}(\Omega)} \le C.$$

Here, $C > 0$ only depends on α', Ω, N, Λ/λ, m, and M, while $\alpha \in (0,1)$ only depends on α', N, Λ/λ and m.

We collect elementary estimates for Hölder continuous functions.

Lemma 18 *Let $\Omega \subset \mathbb{R}^N$ be nonempty, let $\alpha > 0$, and let $f, g \in C^{0,\alpha}(\Omega)$. Then:*

- $\|fg\|_{C^{0,\alpha}(\Omega)} \le \|f\|_{C^{0,\alpha}(\Omega)}\|g\|_{C^{0,\alpha}(\Omega)}$.
- $\|\sqrt{\epsilon + f^2}\|_{C^{0,\alpha}(\Omega)} \le \sqrt{\epsilon} + \|f\|_{C^{0,\alpha}(\Omega)}$ *for all $\epsilon > 0$.*
- *If $|f| \ge \epsilon > 0$ on Ω for some constant $\epsilon > 0$, then there holds*
$$\|1/f\|_{C^{0,\alpha}(\Omega)} \le \epsilon^{-2}\|f\|_{C^{0,\alpha}(\Omega)} + \epsilon^{-1}.$$
- $\|\,|h|\,\|_{C^{0,\alpha}(\Omega)} \le \|h\|_{C^{0,\alpha}(\Omega,\mathbb{R}^N)}$ *for all $h \in C^{0,\alpha}(\Omega, \mathbb{R}^N)$.*
- *Let $N_i \in \mathbb{N}$ and let $U_i \subset \mathbb{R}^{N_i}$ be nonempty, $1 \le i \le 4$. For $\phi \in C^{0,1}(U_2, U_3)$, $h \in C^{0,\alpha}(U_1, U_2)$ and $H \in C^{0,\alpha}(U_3, U_4)$ there hold*
$$\|\phi \circ h\|_{C^{0,\alpha}(U_1,U_3)} \le |\phi|_{C^{0,1}(U_2,U_3)}|h|_{C^{0,\alpha}(U_1,U_2)} + \|\phi\|_{L^\infty(U_2,U_3)}$$
and
$$\|H \circ \phi\|_{C^{0,\alpha}(U_2,U_4)} \le |\phi|^{\alpha}_{C^{0,1}(U_2,U_3)}|H|_{C^{0,\alpha}(U_3,U_4)} + \|H\|_{L^\infty(U_3,U_4)}.$$

Proof **First claim:** Because of $|f(x)g(x) - f(y)g(y)| \le |f(x)|\,|g(x) - g(y)| + |g(y)|\,|f(x) - f(y)| \le \big(\|f\|_{L^\infty(\Omega)}|g|_{C^{0,\alpha}(\Omega)} + \|g\|_{L^\infty(\Omega)}|f|_{C^{0,\alpha}(\Omega)}\big)|x-y|^\alpha$ for all $x, y \in \Omega$, we infer $|fg|_{C^{0,\alpha}(\Omega)} \le \|f\|_{L^\infty(\Omega)}|g|_{C^{0,\alpha}(\Omega)} + \|g\|_{L^\infty(\Omega)}|f|_{C^{0,\alpha}(\Omega)}$. Together with $\|fg\|_{L^\infty(\Omega)} \le \|f\|_{L^\infty(\Omega)}\|g\|_{L^\infty(\Omega)}$ this implies the first claim.
Second claim: Since $\phi(t) := \sqrt{\epsilon + t^2}$ is Lipschitz continuous with constant 1 in $\mathbb{R}$, the assertion follows from the fifth claim by use of $\|\sqrt{\epsilon + f^2}\|_{L^\infty(\Omega)} \le \sqrt{\epsilon} + \|f\|_{L^\infty(\Omega)}$.
Third claim: Since $\phi(t) := |t|^{-1}$ is Lipschitz continuous with constant ϵ^{-2} in $\mathbb{R} \setminus (-\epsilon, \epsilon)$, the assertion follows from the fifth claim, applied with $U_2 := \{f(x) : x \in \Omega\}$, by use of $\|\phi\|_{L^\infty(U_2,U_3)} = \|\,|f|^{-1}\,\|_{L^\infty(\Omega)} \le \epsilon^{-1}$.
Fourth claim: The assertion follows from the fifth claim.
Fifth claim: For $x, y \in U_1$ we have

$$|\phi(h(x)) - \phi(h(y))| \le |\phi|_{C^{0,1}(U_2,U_3)}|h(x) - h(y)| \le |\phi|_{C^{0,1}(U_2,U_3)}|h|_{C^{0,\alpha}(U_1,U_2)}|x-y|^\alpha.$$

Together with $|\phi(h(x))| \le \sup_{y\in U_2}|\phi(y)| = \|\phi\|_{L^\infty(U_2,U_3)}$ for all $x \in U_1$ we obtain the assertion for $\phi \circ h$. The assertion for $H \circ \phi$ can be established analogously. ◻

We can now establish the desired regularity and continuity result for (3).

Theorem 14 *Let $\Omega \subset \mathbb{R}^N$ be a bounded $C^{1,\alpha'}$ domain for some $\alpha' \in (0,1]$. Let $\beta > 0$ and $\gamma^0 \geq \gamma \geq \gamma_0 > 0$ and $\delta^0 \geq \delta \geq \delta_0 > 0$. By $u = u(p) \in H^1(\Omega)$ we denote for each $p \in L^\infty(\Omega)$ the unique weak solution of*

$$\begin{cases} -\operatorname{div}\left(\left[\gamma + \frac{\beta}{\sqrt{\delta + |\nabla u|^2}}\right]\nabla u\right) + \gamma u = p & \text{in } \Omega, \\ \left(\left[\gamma + \frac{\beta}{\sqrt{\delta + |\nabla u|^2}}\right]\nabla u, \nu\right) = 0 & \text{on } \Gamma. \end{cases} \tag{23}$$

Then for every $b^0 > 0$ there exists $\alpha \in (0,1)$ such that $u : \mathbb{B}_{b^0} \to C^{1,\alpha}(\Omega)$ is well-defined and Lipschitz continuous, i.e. $\|u(p_1) - u(p_2)\|_{C^{1,\alpha}(\Omega)} \leq L\|p_1 - p_2\|_{L^\infty(\Omega)}$ for all $p_1, p_2 \in \mathbb{B}_{b^0} \subset L^\infty(\Omega)$ and some $L > 0$. The constants L and α are independent of γ and δ, but may depend on α', Ω, N, β, b^0, $\gamma_0, \gamma^0, \delta_0$ and δ^0.

Proof Let $b^0 > 0$ and let $p_1, p_2 \in L^\infty(\Omega)$ with $\|p_1\|_{L^\infty(\Omega)}, \|p_2\|_{L^\infty(\Omega)} < b^0$.

Part 1: Showing existence of $\mathbf{u_1}, \mathbf{u_2} \in \mathbf{H^1(\Omega)}$.
For $i = 1, 2$ we define

$$F_i \colon H^1(\Omega) \to \mathbb{R},$$
$$v \mapsto \gamma\|v\|^2_{H^1(\Omega)} + \beta \int_\Omega \sqrt{\delta + |\nabla v|^2}\,\mathrm{d}x - (p_i, v)_{L^2(\Omega)}.$$

Invoking the convexity of ψ_δ, cf. Lemma 2, we obtain that F_i is strongly convex, which implies the existence of a unique minimizer $u_i \in H^1(\Omega)$. Since F_i is Fréchet differentiable by Lemma 17, we have $F'(u_i) = 0$ in $H^1(\Omega)^*$, which is equivalent to (23).

Part 2: Showing $\mathbf{u_1}, \mathbf{u_2} \in \mathbf{L^\infty(\Omega)}$ and an estimate for $\|\mathbf{u_1}\|_{\mathbf{L^\infty(\Omega)}}$ and $\|\mathbf{u_2}\|_{\mathbf{L^\infty(\Omega)}}$
Fix $M > \gamma_0^{-1}b^0$ and let $u_{i,M} := \min(M, \max(-M, u_i))$, $i = 1, 2$. For any $\mathbb{N} \ni p \geq 1$ we have

$$\nabla\big(u_{i,M}^{2p-1}\big) = (2p-1)\, u_i^{2p-2}\, \nabla u_i \cdot 1_{\{-M<u_i<M\}} \in L^2(\Omega).$$

Testing (23) with $u_{i,M}^{2p-1}$ yields

$$\begin{aligned} &\gamma(u_i, u_{i,M}^{2p-1})_{L^2(\Omega)} \\ &\quad= (p, u_{i,M}^{2p-1})_{L^2(\Omega)} - (2p-1)\int_{\{-M<u_i<M\}} \gamma u_i^{2p-2}|\nabla u_i|^2 + \beta u_i^{2p-2}\frac{|\nabla u_i|^2}{\sqrt{\delta + |\nabla u_i|^2}}\,\mathrm{d}x \\ &\quad\leq \|p\|_{L^\infty(\Omega)}\|u_{i,M}^{2p-1}\|_{L^1(\Omega)} \leq b^0\|1\|_{L^{2p}(\Omega)}\|u_{i,M}^{2p-1}\|_{L^{\frac{2p}{2p-1}}(\Omega)} = b^0|\Omega|^{\frac{1}{2p}}\|u_{i,M}\|^{2p-1}_{L^{2p}(\Omega)}. \end{aligned}$$

In combination with

$$\begin{aligned} &\gamma(u_i, u_{i,M}^{2p-1})_{L^2(\Omega)} \\ &= \gamma\left(\int_{\{-M<u_i<M\}} u_i^{2p}\,\mathrm{d}x + \int_{\{u_i\leq -M\}} u_i(-M)^{2p-1}\,\mathrm{d}x + \int_{\{M\leq u_i\}} u_i M^{2p-1}\,\mathrm{d}x\right) \\ &\geq \gamma\left(\|u_i\|^{2p}_{L^{2p}(\{-M<u_i<M\})} + \int_{\{u_i\leq -M\}} (-M)^{2p}\,\mathrm{d}x + \int_{\{M\leq u_i\}} M^{2p}\,\mathrm{d}x\right) \geq \gamma_0\|u_{i,M}\|^{2p}_{L^{2p}(\Omega)} \end{aligned}$$

this yields $\gamma_0\|u_{i,M}\|_{L^{2p}(\Omega)} \leq b^0|\Omega|^{\frac{1}{2p}}$. Sending $p \to \infty$ gives $\|u_{i,M}\|_{L^\infty(\Omega)} \leq \gamma_0^{-1}b^0$. As $M > \gamma_0^{-1}b^0$ by assumption we conclude that

$$\|u_i\|_{L^\infty(\Omega)} \leq \gamma_0^{-1}b^0 \quad \text{for } i = 1, 2. \tag{24}$$

Part 3: Obtaining $\mathbf{C^{1,\alpha}}$ regularity of $\mathbf{u_1, u_2}$
We use Theorem 13 to establish the $C^{1,\alpha}(\Omega)$-regularity of u_1 and u_2. We apply it with $m = 0$, $A(x,u,\eta) = \gamma\eta + \beta\eta/\sqrt{\delta + |\eta|^2}$, $B(x,u,\eta) = p_i(x)$ for $i = 1,2$, $\kappa = 0$, identical values for α', $\lambda = \gamma_0$, $\Lambda = \max\{b^0, \gamma^0 N + \delta_0^{-1/2}\beta(N + N^2)\}$ and $M = \gamma_0^{-1} b^0$, cf. (24). Since A is independent of (x,u) and continuously differentiable, it is easy to see that the requirements of Theorem 13 are met. This shows $u_1, u_2 \in C^{1,\alpha}(\Omega)$ for some $\alpha > 0$ and yields

$$\|u_i\|_{C^{1,\alpha}(\Omega)} \le C, \tag{25}$$

where $C > 0$ and $\alpha \in (0,1)$ depend only on the quantities α', Ω, N, $\Lambda/\lambda = \gamma_0^{-1}\Lambda$ and $M = \gamma_0^{-1} b^0$.

Part 4: Lipschitz continuity of $\mathbf{p \mapsto u(p)}$
Taking the difference of the weak formulations supplies

$$\int_\Omega \nabla\varphi^T \left(\gamma\nabla\tilde{u} + \beta\frac{\nabla u_1}{\sqrt{\delta + |\nabla u_1|^2}} - \beta\frac{\nabla u_2}{\sqrt{\delta + |\nabla u_2|^2}} \right) + \gamma\varphi\tilde{u}\,\mathrm{d}x = \int_\Omega \varphi\tilde{p}\,\mathrm{d}x \quad \forall \varphi \in H^1(\Omega), \tag{26}$$

where we abbreviated $\tilde{u} := u_1 - u_2$ and $\tilde{p} := p_1 - p_2$. The function $H : \mathbb{R}^N \to \mathbb{R}$ given by $H(v) := \sqrt{\delta + |v|^2}$ is convex. Let $t \in [0,1]$ and denote by $u^\tau : \Omega \to \mathbb{R}$ the $C^{1,\alpha}(\Omega)$ function $u^\tau(x) := u_2(x) + \tau\tilde{u}(x)$. For every $x \in \Omega$ it holds that

$$\begin{aligned} \frac{\nabla u_1(x)}{\sqrt{\delta + |\nabla u_1(x)|^2}} - \frac{\nabla u_2(x)}{\sqrt{\delta + |\nabla u_2(x)|^2}} &= \nabla H\big(\nabla u_1(x)\big) - \nabla H\big(\nabla u_2(x)\big) \\ &= \int_0^1 \nabla^2 H(\nabla u^\tau(x))\,\mathrm{d}\tau\,\nabla\tilde{u}(x), \end{aligned}$$

where the integral is understood componentwise. Together with (26) we infer that $\tilde{u}$ satisfies

$$\begin{cases} -\operatorname{div}\Big(\tilde{A}\nabla\tilde{u}\Big) + \gamma\tilde{u} = \tilde{p} & \text{in } \Omega, \\ \partial_{\nu_{\tilde{A}}}\tilde{u} = 0 & \text{on } \Gamma, \end{cases}$$

where $\tilde{A} : \Omega \to \mathbb{R}^{N\times N}$ is given by

$$\tilde{A}(x) := \gamma I + \beta\int_0^1 \nabla^2 H(\nabla u^\tau(x))\,\mathrm{d}\tau.$$

In order to apply Theorem 12 to this PDE, we show $\tilde{A} \in C^{0,\alpha}(\Omega, \mathbb{R}^{N\times N})$. The convexity of H implies that $\nabla^2 H$ is positive semi-definite. Thus we find for any $v \in \mathbb{R}^N$ and any $x \in \Omega$

$$v^T\tilde{A}(x)v \ge \gamma|v|^2 \ge \gamma_0|v|^2.$$

For $x \in \Omega$ and $1 \le i,j \le N$ it holds that

$$|\tilde{A}_{ij}(x)| \le \gamma + \beta\int_0^1 \Big|\big[\nabla^2 H(\nabla u^\tau(x))\big]_{ij}\Big|\,\mathrm{d}\tau \le \gamma^0 + \beta\sup_{\tau\in[0,1]}\Big\|\big[\nabla^2 H(\nabla u^\tau)\big]_{ij}\Big\|_{L^\infty(\Omega)}.$$

We also have for all $x, y \in \Omega$

$$\begin{aligned} \Big|\tilde{A}_{ij}(x) - \tilde{A}_{ij}(y)\Big| &\le \beta\int_0^1 \Big|\Big[\nabla^2 H(\nabla u^\tau(x)) - \nabla^2 H(\nabla u^\tau(y))\Big]_{ij}\Big|\,\mathrm{d}\tau \\ &\le \beta\sup_{\tau\in[0,1]}\Big|\Big[\nabla^2 H(\nabla u^\tau(x)) - \nabla^2 H(\nabla u^\tau(y))\Big]_{ij}\Big| \\ &\le \beta\sup_{\tau\in[0,1]}\Big|\Big[\nabla^2 H(\nabla u^\tau)\Big]_{ij}\Big|_{C^{0,\alpha}(\Omega)}|x - y|^\alpha, \end{aligned}$$

which shows $|\tilde{A}_{ij}|_{C^{0,\alpha}(\Omega)} \le \beta \sup_{\tau\in[0,1]} |[\nabla^2 H(u^\tau)]_{ij}|_{C^{0,\alpha}(\Omega)}$. Together, we infer that

$$\|\tilde{A}_{ij}\|_{C^{0,\alpha}(\Omega)} \le \gamma^0 + 2\beta \sup_{\tau\in[0,1]} \left\| \left[\nabla^2 H(\nabla u^\tau)\right]_{ij} \right\|_{C^{0,\alpha}(\Omega)} \tag{27}$$

for all $1 \le i,j \le N$. From Lemma 18 we obtain for every fixed $1 \le i,j \le N$

$$\begin{aligned}\left\| \left[\nabla^2 H(\nabla u^\tau)\right]_{ij} \right\|_{C^{0,\alpha}(\Omega)} &\le \left\| \frac{1}{\sqrt{\delta + |\nabla u^\tau|^2}} \right\|_{C^{0,\alpha}(\Omega)} + \left\| \frac{\partial_{x_i} u^\tau \partial_{x_j} u^\tau}{\sqrt{\delta + |\nabla u^\tau|^2}^3} \right\|_{C^{0,\alpha}(\Omega)} \\ &\le C \left(1 + \left\| \sqrt{\delta + |\nabla u^\tau|^2} \right\|_{C^{0,\alpha}(\Omega)} + \|\nabla u^\tau\|^2_{C^{0,\alpha}(\Omega)} \left\| \left(\delta + |\nabla u^\tau|^2\right)^{-\frac{3}{2}} \right\|_{C^{0,\alpha}(\Omega)} \right),\end{aligned}$$

where C only depends on δ_0. Since $\|\nabla u_1\|_{C^{0,\alpha}(\Omega)}, \|\nabla u_2\|_{C^{0,\alpha}(\Omega)} \le C$ by (25), there holds $\|\nabla u^\tau\|_{C^{0,\alpha}(\Omega)} \le C$ with the same $C > 0$. This C only depends on α', Ω, N, β, b^0, γ_0, γ^0 and δ_0. This and Lemma 18 show

$$\begin{aligned}\left\| \left[\nabla^2 H(\nabla u^\tau)\right]_{ij} \right\|_{C^{0,\alpha}(\Omega)} &\le C \Big(1 + \left\| \sqrt{\delta + |\nabla u^\tau|^2} \right\|_{C^{0,\alpha}(\Omega)} + \left\| \sqrt{\delta + |\nabla u^\tau|^2} \right\|^3_{C^{0,\alpha}(\Omega)} \Big) \\ &\le C \Big(1 + \Big(\sqrt{\delta} + \|\nabla u^\tau\|_{C^{0,\alpha}(\Omega)} \Big)^3 \Big) \le C,\end{aligned}$$

where $C > 0$ is independent of τ and only depends on the quantities stated in the theorem. Hence, with the same C there holds

$$\sup_{\tau\in[0,1]} \left\| \left[\nabla^2 H(\nabla u^\tau)\right]_{ij} \right\|_{C^{0,\alpha}(\Omega)} \le C \qquad \forall\, 1 \le i,j \le N.$$

Inserting this into (27) yields $\tilde{A} \in C^{0,\alpha}(\Omega, \mathbb{R}^{N\times N})$ with $\|\tilde{A}\|_{C^{0,\alpha}(\Omega)} \le \gamma^0 + 2\beta C$, so Theorem 12 is applicable. We obtain $\|\tilde{u}\|_{C^{1,\alpha}(\Omega)} \le C\|\tilde{p}\|_{L^\infty(\Omega)}$, where C only depends on the claimed quantities. This proves the asserted Lipschitz continuity of $p \mapsto u(p)$. □

C The original problem: Optimality conditions

The first order optimality conditions of (ROC) can be obtained by use of [8]. The space $W_0^q(\mathrm{div};\Omega)$, $q \in [1,\infty)$, that appears in the following is defined in [8, Definition 10].

Theorem 15 *Let $\Omega \subset \mathbb{R}^N$, $N \in \{1,2,3\}$, be a bounded Lipschitz domain and let $r_N = \frac{N}{N-1}$ if $N > 1$, respectively, $r_N \in [1,\infty)$ if $N = 1$. Then we have: The function $\bar{u} \in BV(\Omega)$ is the solution of* (ROC) *iff there is*

$$\bar{h} \in L^\infty(\Omega, \mathbb{R}^N) \cap W_0^{r_N}(\mathrm{div};\Omega)$$

that satisfies $\||\bar{h}|\|_{L^\infty(\Omega)} \le \beta$ and $\operatorname{div} \bar{h} = \bar{p}$, where $\bar{p}$ is defined as in section 2.1, as well as

$$\begin{aligned} \bar{h} &= \beta \frac{\nabla \bar{u}_a}{|\nabla \bar{u}_a|} && \mathcal{L}^N\text{-a.e. in } \Omega \setminus \{x : \nabla \bar{u}_a(x) = 0\}, \\ T\bar{h} &= \beta \frac{\bar{u}^+(x) - \bar{u}^-(x)}{|\bar{u}^+(x) - \bar{u}^-(x)|} \nu_{\bar{u}} && \mathcal{H}^1\text{-a.e. in } J_{\bar{u}}, \\ T\bar{h} &= \beta \sigma_{C_{\bar{u}}} && |\nabla \bar{u}_c|\text{-a.e.} \end{aligned}$$

Here, the first, second and third equation correspond to the absolutely continuous part, the jump part, respectively, the Cantor part of the vector measure $\nabla\bar{u}$. Also, $\sigma_{C_{\bar{u}}}$ is the Radon-Nikodym density of $\nabla\bar{u}_c$ with respect to $|\nabla\bar{u}_c|$, cf. e.g. [9, Theorem 9.1]. Moreover, $\nu_{\bar{u}}$ is the jump direction of $\bar{u}$ and $J_{\bar{u}}$ denotes the discontinuity set of $\bar{u}$ in the sense of [4, Definition 3.63]. Further, $\mathcal{H}^1$ is the Hausdorff measure of $J_{\bar{u}}$. The operator $T\colon \mathrm{dom}(T) \subset W^{\mathrm{div},q}(\Omega) \cap L^\infty(\Omega,\mathbb{R}^N) \to L^1(\Omega,\mathbb{R}^N,|\nabla u|)$ is called the full trace operator *and is introduced in [8, Definition 12]. We emphasize that $\bar{h} \in \mathrm{dom}(T)$.*

Proof The well-known optimality condition $0 \in \partial j(\bar{u})$ from convex analysis can be expressed as $-\frac{\bar{p}}{\beta} \in \partial|\bar{u}|_{\mathrm{BV}(\Omega)}$, so the claim follows from [8, Proposition 8]. □

Remark 8 Theorem 15 implies the sparsity relation $\{x : \nabla\bar{u}_a(x) \neq 0\} \subset \{x : |\bar{h}(x)| = \beta\}$. Since $\{x : |\bar{h}(x)| = \beta\}$ typically has small Lebesgue measure (often: measure 0), $\bar{u}$ is usually constant a.e. in large parts (often: all) of Ω; cf. also the example in section D.

D An example with explicit solution

Using rotational symmetry we construct an example for (OC) for $N = 2$ with an explicit solution. We let $\mathcal{A} = -\Delta$ and $c_0 \equiv 0$ in the governing PDE. We define $\hat{h} : [0, \infty) \to \mathbb{R}$, $\hat{h}(r) := \frac{\beta}{2}(\cos(\frac{2\pi}{R}r) - 1)$ and $\Omega := B_{2R}(0) \setminus \overline{B_R(0)}$, where the parameters $R > 0$ and $\beta > 0$ are arbitrary. We introduce the functions

$$r(x, y) := \sqrt{x^2 + y^2}, \qquad \bar{h}(x, y) := \hat{h}(r(x, y))\nabla r(x, y) \quad \text{and} \quad \bar{u}(x, y) := 1_{(R, \frac{3R}{2})}(r(x, y)),$$

all of which are defined on Ω. The problem data is given by

$$\bar{p} := \operatorname{div} \bar{h}, \qquad \bar{y} := S\bar{u} \qquad \text{and} \qquad y_\Omega := \Delta\bar{p} + \bar{y}.$$

We now show that these quantities satisfy the properties of Theorem 15. By construction $\bar{y}$ and $\bar{p}$ are the state and adjoint state associated to $\bar{u}$ and we have $\bar{p} = \operatorname{div} \bar{h}$. We check the properties of $\bar{h}$. Since $|\nabla r| = 1$ for $(x, y) \in \Omega$, we obtain $|\bar{h}(x, y)| = |\hat{h}(r(x, y))| \le \frac{\beta}{2} 2 = \beta$. We also see that $\bar{h}$ is C^1 in $\bar{\Omega}$ and satisfies $\bar{h} = 0$ on $\partial\Omega$ so that $\bar{h} \in L^\infty(\Omega, \mathbb{R}^N) \cap W_0^q(\operatorname{div}; \Omega)$ for any $q \in [1, \infty)$. By [8, Proposition 6] we have $T\bar{h} = \bar{h}$. As $\nabla\bar{u}(x, y) = -\nabla r(x, y)\mathcal{H}^1_{\partial B_{\frac{3R}{2}}(0)}(x, y)$, we find that $\nabla\bar{u}$ has no Cantor part and no parts that are absolutely continuous with respect to the Lebesgue measure. Thus, the first and third condition on $\bar{h}$ in Theorem 15 are trivially satisfied. For $(x, y) \in \partial B_{\frac{3R}{2}}(0) = J_{\bar{u}}$ we have $\bar{h}(x, y) = -\beta\nabla r(x, y) = -\beta\nu_{\bar{u}}$ and $\bar{u}^+(x) = 0$, $\bar{u}^-(x) = 1$ for $x \in J_{\bar{u}}$, hence the second condition on $\bar{h}$ in Theorem 15 holds. Let us confirm that $\bar{p}$ satisfies the homogeneous Dirichlet boundary conditions. From $\Delta r = r^{-1}$ and $|\nabla r|^2 = 1$ we obtain

$$\bar{p} = \operatorname{div} \bar{h} = \nabla\hat{h}(r)^T\nabla r + \hat{h}(r)\Delta r = \hat{h}'(r)|\nabla r|^2 + r^{-1}\hat{h}(r) = \hat{h}'(r) + r^{-1}\hat{h}(r).$$

Thus, $\bar{p}$ satisfies the boundary conditions. Let us confirm that $\bar{y}$ satisfies the boundary conditions. The Ansatz $\bar{y}(x, y) = \hat{y}(r(x, y))$, with $\hat{y} : \Omega \to \mathbb{R}$ to be determined, yields

$$-1_{(R,3R/2)}(r) = -\bar{u}(x, y) = \Delta\bar{y}(x, y) = \operatorname{div}(\hat{y}'(r)\nabla r) = \hat{y}''(r) + r^{-1}\hat{y}'(r).$$

This leads to

$$\hat{y}(r) = \begin{cases} -\frac{r^2}{4} + A\ln(r/(2R)) + B & \text{if } r \in (R, 3R/2), \\ C\ln(r/(2R)) & \text{if } r \in (3R/2, 2R), \end{cases}$$

and it is straightforward to check that $\bar{y}$ satisfies the boundary conditions and is continuously differentiable for the parameters

$$A = \frac{R^2}{8}\cdot\frac{18\ln(3/4) - 5}{\ln(1/4)}, \qquad B = \frac{9R^2}{8}\left(\frac{1}{2} - \ln(3/4)\right) \qquad \text{and} \qquad C = \frac{R^2}{8}\cdot\frac{18\ln(3/2) - 5}{\ln(1/4)},$$

All in all, the optimality conditions of Theorem 15 are satisfied. Moreover, the optimal value in this example is given by

$$j(\bar{u}) = \frac{1}{2}\|\bar{y} - y_\Omega\|^2_{L^2(\Omega)} + \beta|\bar{u}|_{\mathrm{BV}(\Omega)} = \frac{1}{2}\|\Delta\bar{p}\|^2_{L^2(\Omega)} + \beta|\bar{u}|_{\mathrm{BV}(\Omega)},$$

which for $R = 2\pi$ results in

$$j(\bar{u}) = \frac{\beta^2\pi}{4}\left(3\pi^2 + \ln(8) + \frac{15}{4}\operatorname{Ci}(2\pi) - \frac{27}{4}\operatorname{Ci}(4\pi) + 3\operatorname{Ci}(8\pi)\right) + 6\pi^2\beta \approx 24.85\beta^2 + 59.22\beta$$

with $\operatorname{Ci}(t) := -\int_t^\infty \frac{\cos\tau}{\tau}\,\mathrm{d}\tau$.

Acknowledgements Dominik Hafemeyer acknowledges support from the graduate program TopMath of the Elite Network of Bavaria and the TopMath Graduate Center of TUM Graduate School at Technische Universität München. He received a scholarship from the Studienstiftung des deutschen Volkes. He receives support from the IGDK Munich-Graz. Funded by the Deutsche Forschungsgemeinschaft, grant no 188264188/GRK1754.

References

1. Acar, R., Vogel, C.R.: Analysis of bounded variation penalty methods for ill-posed problems. Inverse Probl. **10**(6), 1217–1229 (1994). DOI 10.1088/0266-5611/10/6/003
2. Allendes, A., Fuica, F., Otárola, E.: Adaptive finite element methods for sparse PDE-constrained optimization. IMA Journal of Numerical Analysis **40**(3), 2106–2142 (2019). DOI 10.1093/imanum/drz025
3. Alnæs, M.S., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M.E., Wells, G.N.: The FEniCS project version 1.5. Archive of Numerical Software **3**(100), 9–23 (2015). DOI 10.11588/ans.2015.100.20553
4. Ambrosio, L., Fusco, N., Pallara, D.: Functions of bounded variation and free discontinuity problems. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press (2000)
5. Attouch, H., Buttazzo, G., Michaille, G.: Variational analysis in Sobolev and BV spaces. Applications to PDEs and optimization. 2nd revised ed., *MPS/SIAM Series on Optimization*, vol. 6. SIAM (2014). DOI 10.1137/1.9781611973488
6. Bartels, S.: Total variation minimization with finite elements: convergence and iterative solution. SIAM J. Numer. Anal. **50**(3), 1162–1180 (2012). DOI 10.1137/11083277X
7. Bergounioux, M., Bonnefond, X., Haberkorn, T., Privat, Y.: An optimal control problem in photoacoustic tomography. Math. Models Methods Appl. Sci. **24**(12), 2525–2548 (2014). DOI 10.1142/S0218202514500286
8. Bredies, K., Holler, M.: A pointwise characterization of the subdifferential of the total variation functional (2012). Preprint, IGDK1754
9. Brokate, M., Kersting, G.: Measure and integral. Cham: Birkhäuser/Springer (2015). DOI 10.1007/978-3-319-15365-0
10. Casas, E., Clason, C., Kunisch, K.: Approximation of elliptic control problems in measure spaces with sparse solutions. SIAM J. Control Optim. **50**(4), 1735–1752 (2012). DOI 10.1137/110843216
11. Casas, E., Clason, C., Kunisch, K.: Parabolic control problems in measure spaces with sparse solutions. SIAM J. Control Optim. **51**(1), 28–63 (2013)
12. Casas, E., Kogut, P.I., Leugering, G.: Approximation of optimal control problems in the coefficient for the p-Laplace equation. I: Convergence result. SIAM J. Control Optim. **54**(3), 1406–1422 (2016). DOI 10.1137/15M1028108
13. Casas, E., Kruse, F., Kunisch, K.: Optimal control of semilinear parabolic equations by BV-functions. SIAM J. Control Optim. **55**(3), 1752–1788 (2017). DOI 10.1137/16M1056511
14. Casas, E., Kunisch, K.: Optimal control of semilinear elliptic equations in measure spaces. SIAM J. Control Optim. **52**(1), 339–364 (2014). DOI 10.1137/13092188X
15. Casas, E., Kunisch, K.: Analysis of optimal control problems of semilinear elliptic equations by BV-functions. Set-Valued Var. Anal. **27**(2), 355–379 (2019). DOI 10.1007/s11228-018-0482-7
16. Casas, E., Kunisch, K., Pola, C.: Some applications of BV functions in optimal control and calculus of variations. ESAIM, Proc. **4**, 83–96 (1998). DOI 10.1051/proc:1998022
17. Casas, E., Kunisch, K., Pola, C.: Regularization by functions of bounded variation and applications to image enhancement. Appl. Math. Optim. **40**(2), 229–257 (1999). DOI 10.1007/s002459900124
18. Casas, E., Ryll, C., Tröltzsch, F.: Sparse optimal control of the Schlögl and Fitzhugh-Nagumo systems. Comput. Methods Appl. Math. **13**(4), 415–442 (2013). DOI 10.1515/cmam-2013-0016
19. Casas, E., Vexler, B., Zuazua, E.: Sparse initial data identification for parabolic PDE and its finite element approximations. Math. Control Relat. Fields **5**(3), 377–399 (2015). DOI 10.3934/mcrf.2015.5.377
20. Chan, T.F., Zhou, H.M., Chan, R.H.: Continuation method for total variation denoising problems. In: F.T. Luk (ed.) Advanced Signal Processing Algorithms, vol. 2563, pp. 314–325. International Society for Optics and Photonics, SPIE (1995). DOI 10.1117/12.211408
21. Clason, C., Kruse, F., Kunisch, K.: Total variation regularization of multi-material topology optimization. ESAIM Math. Model. Numer. Anal. **52**(1), 275–303 (2018). DOI 10.1051/m2an/2017061

22. Clason, C., Kunisch, K.: A duality-based approach to elliptic control problems in non-reflexive Banach spaces. ESAIM Control Optim. Calc. Var. **17**(1), 243–266 (2011). DOI 10.1051/cocv/2010003
23. Engel, S., Kunisch, K.: Optimal control of the linear wave equation by time-depending BV-controls: A semi-smooth Newton approach. Math. Control Relat. Fields **10**(3), 591–622 (2020). DOI 10.3934/mcrf.2020012
24. Engel, S., Vexler, B., Trautmann, P.: Optimal finite element error estimates for an optimal control problem governed by the wave equation with controls of bounded variation. IMA Journal of Numerical Analysis (2020). DOI 10.1093/imanum/draa032
25. Ern, A., Guermond, J.L.: Theory and practice of finite elements, *Applied Mathematical Sciences*, vol. 159. Springer (2004). DOI 10.1007/978-1-4757-4355-5
26. Grisvard, P.: Elliptic problems in nonsmooth domains, vol. 69, reprint of the 1985 hardback edn. SIAM (2011). DOI 10.1137/1.9781611972030
27. Hafemeyer, D.: Optimale Steuerung von Differentialgleichungen mit BV-Funktionen. Bachelor's thesis (2016)
28. Hafemeyer, D.: Regularization and Discretization of a BV-controlled Elliptic Problem: A Completely Adaptive Approach. Master's thesis (2017)
29. Hafemeyer, D.: Optimal control of parabolic obstacle problems - optimality conditions and numerical analysis. Dissertation, Technische Universität München, München (2020)
30. Hafemeyer, D., Mannel, F., Neitzel, I., Vexler, B.: Finite element error estimates for one-dimensional elliptic optimal control by BV functions. Math. Control Relat. Fields **10**(2), 333–363 (2020). DOI 10.3934/mcrf.2019041
31. Herzog, R., Stadler, G., Wachsmuth, G.: Directional sparsity in optimal control of partial differential equations. SIAM J. Control Optim. **50**(2), 943–963 (2012). DOI 10.1137/100815037
32. Kato, T.: Perturbation theory for linear operators. Die Grundlehren der mathematischen Wissenschaften, Band 132. Springer (1966). DOI 10.1007/978-3-662-12678-3
33. Kelley, C.T.: Iterative methods for linear and nonlinear equations, vol. 16. SIAM (1995). DOI 10.1137/1.9781611970944
34. Li, C., Stadler, G.: Sparse solutions in optimal control of PDEs with uncertain parameters: the linear case. SIAM J. Control Optim. **57**(1), 633–658 (2019). DOI 10.1137/18M1181419
35. Li, D., Fukushima, M.: A derivative-free line search and global convergence of Broyden-like method for nonlinear equations. Optim. Methods Softw. **13**(3), 181–201 (2000). DOI 10.1080/10556780008805782
36. Lieberman, G.M.: Boundary regularity for solutions of degenerate elliptic equations. Nonlinear Anal. **12**(11), 1203–1219 (1988). DOI 10.1016/0362-546X(88)90053-3
37. Logg, A., Mardal, K.A., Wells, G.N., et al.: Automated Solution of Differential Equations by the Finite Element Method. Springer (2012). DOI 10.1007/978-3-642-23099-8
38. Logg, A., Wells, G.N.: Dolfin: Automated finite element computing. ACM Transactions on Mathematical Software **37**(2) (2010). DOI 10.1145/1731022.1731030
39. Logg, A., Wells, G.N., Hake, J.: DOLFIN: a C++/Python Finite Element Library, chap. 10. Springer (2012). DOI 10.1007/978-3-642-23099-8_10
40. Meyers, N.G., Ziemer, W.P.: Integral Inequalities of Poincaré and Wirtinger Type for BV Functions. American Journal of Mathematics **99**(6), 1345–1360 (1977). DOI 10.2307/2374028
41. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**, 259–268 (1992). DOI 10.1016/0167-2789(92)90242-F
42. Rudin, W.: Real and complex analysis, third edn. McGraw-Hill Book Co. (1987)
43. Savaré, G.: Regularity results for elliptic equations in Lipschitz domains. J. Funct. Anal. **152**(1), 176–201 (1998). DOI 10.1006/jfan.1997.3158
44. Schiela, A.: An interior point method in function space for the efficient solution of state constrained optimal control problems. Math. Program. **138**(1-2 (A)), 83–114 (2013). DOI 10.1007/s10107-012-0595-y
45. Stadler, G.: Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices. Comput. Optim. Appl. **44**(2), 159–181 (2009). DOI 10.1007/s10589-007-9150-9
46. Troianiello, G.M.: Elliptic differential equations and obstacle problems. The University Series in Mathematics. Plenum Press, New York (1987). DOI 10.1007/978-1-4899-3614-1
47. Wathen, A.J.: Realistic eigenvalue bounds for the Galerkin mass matrix. IMA J. Numer. Anal. **7**, 449–457 (1987). DOI 10.1093/imanum/7.4.449