

## TOTAL VARIATION REGULARIZATION OF MULTI-MATERIAL TOPOLOGY OPTIMIZATION

CHRISTIAN CLASON<sup>1,\*</sup>, FLORIAN KRUSE<sup>2</sup> AND KARL KUNISCH<sup>2,3</sup>

**Abstract.** This work is concerned with the determination of the diffusion coefficient from distributed data of the state. This problem is related to homogenization theory on the one hand and to regularization theory on the other hand. An approach is proposed which involves total variation regularization combined with a suitably chosen cost functional that promotes the diffusion coefficient assuming prespecified values at each point of the domain. The main difficulty lies in the delicate functional-analytic structure of the resulting nondifferentiable optimization problem with pointwise constraints for functions of bounded variation, which makes the derivation of useful pointwise optimality conditions challenging. To cope with this difficulty, a novel reparametrization technique is introduced. Numerical examples using a regularized semismooth Newton method illustrate the structure of the obtained diffusion coefficient.

**Mathematics Subject Classification.** 49Q10, 49K20, 49M15.

Received August 21, 2017. Accepted November 28, 2017.

### 1. INTRODUCTION

In this paper we revisit a challenging problem in the calculus of variations given by

$$\begin{cases} \min_{u \in \mathcal{U}} \frac{1}{2} \|y - z\|_{L^2(\Omega)}^2 + \mathcal{R}(u), \\ \text{s.t.} \quad -\operatorname{div}(u \nabla y) = f \quad \text{in } \Omega, \\ \quad \quad \quad y = 0 \quad \text{on } \partial\Omega, \end{cases} \quad (\text{PI})$$

where  $\mathcal{U}$  denotes the set of admissible controls and  $\mathcal{R}$  stands for a regularization term. This problem represents the optimization-theoretic formulation of the problem of determining the optimal distribution  $u$  of material in the domain  $\Omega$  from data  $z$ . If the data are only available in distributed part  $\omega \subsetneq \Omega$  of the domain, then the cost functional in (PI) can readily be adapted. Problem (PI) arises as the regularization of a coefficient inverse problem; if the focus is on the situation that  $u(x)$  is supposed to assume only preferred values  $u_i$  specific to different materials, it can also be considered as a topology optimization problem.

---

*Keywords and phrases:* Topology optimization, total variation, convex analysis, non-smooth optimization, semi-smooth Newton method.

<sup>1</sup> Faculty of Mathematics, University Duisburg-Essen, 45117 Essen, Germany.

<sup>2</sup> Institute of Mathematics and Scientific Computing, University of Graz, Heinrichstrasse 36, 8010 Graz, Austria.

<sup>3</sup> Radon Institute, Austrian Academy of Sciences, Linz, Austria.

\* Corresponding author: [christian.clason@uni-due.de](mailto:christian.clason@uni-due.de)

In the calculus of variation literature, different forms of (PI) have received a tremendous amount of attention. For the particular choice that  $\mathcal{R}$  is not present and

$$\mathcal{U} = \{u \in L^\infty(\Omega) : 0 < u_{\min} \leq u(x) \leq u_{\max}\}, \quad (1.1)$$

for constants  $u_{\min}$  and  $u_{\max}$ , it was shown in [33] that the problem may fail to have a solution. Historically, this goes along with the development of homogenization theory and deep analytical concepts such as H-convergence and compensated compactness; see, *e.g.*, [34, 39, 40]. Such concepts allow associating a solution to (PI) without the use of a regularization term  $\mathcal{R}$ .

Here we follow a different perspective and aim for a formulation that allows numerical realization; in such a context the use of regularization terms provides a powerful tool. The goal must be to choose a functional  $\mathcal{R}$  that guarantees existence to (PI) and at the same time does not affect the sought parameter  $u$  too much. The use of a regularization term involving semi-norms of Sobolev spaces would conflict with this second requirement, since such a choice would prevent jumps of  $u$  across hypersurfaces – a property that we want to retain here.

The choice for  $\mathcal{R}$  that we propose and investigate in this paper is

$$\mathcal{R}(u) = \alpha G(u) + \beta \text{TV}(u),$$

where  $G$  is a pointwise “multi-bang” penalty as in [20, 21] that promotes the attainment of the predefined states  $\{u_i\}_{i=1}^m$  almost everywhere, and TV denotes the total variation semi-norm. The use of TV will guarantee existence, while  $G$  models the desired structural properties. The usefulness of TV has been established in the calculus of variations and in image analysis for several decades now; see, *e.g.*, [4, 10, 19, 23, 36]. It has also been used in topology optimization in [8, 14], but the approaches in these contributions are different from our formulation and do not contain the multi-material concept (although the latter considers a three-phase formulation with two different non-material phases, “void” and “liquid”). Rather, this concept is an extension of our work from [21], where related topology optimization problems are considered in situations where well-posedness can be guaranteed without the need of employing TV-regularization. Concerning approaches for multi-material topology optimization, we refer to, *e.g.*, [5–7, 13, 26]; among these, our “multi-bang approach” is most closely related to the second. Finally, coefficient inverse problems have been studied in a wide variety of contexts.

The use of the TV functional entails an essential difficulty from an infinite dimensional optimization point of view. In fact, well-posedness of the PDE constraint in (PI) requires a strictly positive lower bound on  $u$  as in the definition (1.1) of  $\mathcal{U}$ . In the process of deriving optimality conditions, however, one is confronted with the problem of considering the subdifferential of  $\text{TV}(u) + I_{\mathcal{U}}$ , where  $I_{\mathcal{U}}$  denotes the indicator function of the set  $\mathcal{U}$ , *e.g.*, as extended real-valued functions on  $L^2(\Omega)$ . In this case, the sum rule cannot be used to compute this subdifferential since neither of the two functionals TV and  $I_{\mathcal{U}}$  is continuous at any point of its domain (which would be required to use a result as in [9] on the sum of subdifferentials of convex functions). The fact that the sum rule is not applicable constitutes a major obstacle for deriving useful optimality conditions. Thus, we propose a different approach to ensure the well-posedness of the PDE constraint in (PI): We introduce a reparametrization of the coefficient in the PDE constraint which allows us to drop the explicit pointwise bounds in the definition of  $\mathcal{U}$ . This novel approach could be of interest also in situations different from the one considered in this work.

For the numerical solution, we consider a finite element discretization of the problem that allows deriving optimality conditions in terms of the expansion coefficients that, after introducing a Moreau–Yosida regularization of the multi-bang and total variation penalties, can be solved by a semismooth Newton-type method with path-following.

The paper is organized as follows: Section 2 contains the problem statement, useful results on the state equation, and descriptions of the transformation announced above, as well as of the multi-bang penalty term. Sections 3 and 4 are devoted to the existence of minimizers and first-order optimality conditions, respectively. The discretization of the infinite dimensional problem as well as of the optimality conditions are provided

in Section 5. There we also provide a description of the semismooth Newton-type method, employing dual regularizations of the multi-bang penalty term and the TV term, which are needed for defining the Newton steps. Numerical examples are provided in Section 6 for two model problems motivated by the interpretation of (PI) as a topology optimization and a parameter identification problem, respectively. Finally, in Appendix A we prove that strongly Lipschitz domains are regular in the sense of Gröger, an elementary but not completely obvious result that is important in our analysis.

## 2. PROBLEM STATEMENT AND PRELIMINARY RESULTS

We consider for  $\alpha, \beta > 0$  the following problem:

$$\begin{cases} \min_{u \in BV(\Omega)} \frac{1}{2} \|y - z\|_{L^2(\Omega)}^2 + \alpha G(u) + \beta \text{TV}(u), \\ \text{s.t.} & -\text{div}(\Phi_\varepsilon(u)\nabla y) = f \quad \text{in } \Omega, \\ & y = 0 \quad \text{on } \partial\Omega. \end{cases} \quad (\text{P})$$

Here,  $\Omega \subset \mathbb{R}^d$ ,  $d \in \mathbb{N}$ , is a bounded strongly Lipschitz domain (see Def. A.1 for a rigorous definition),  $BV(\Omega)$  denotes the space of functions of bounded variation, and  $f \in L^2(\Omega)$  and  $z \in L^2(\Omega)$  are given. Furthermore, TV denotes the total variation,  $G$  is a multi-bang penalty, and  $\Phi_\varepsilon$  for  $\varepsilon \geq 0$  is a superposition operator defined by a (smoothed) pointwise projection onto the set  $[u_{\min}, u_{\max}] \subset (0, \infty)$ , each of which will be described in detail in the following subsections.

### 2.1. Functions of bounded variation

We recall, *e.g.*, from [4, 23, 44] that the space  $BV(\Omega)$  is given by those functions  $v \in L^1(\Omega)$  for which the distributional derivative  $Dv$  is a Radon measure, *i.e.*,

$$BV(\Omega) = \{v \in L^1(\Omega) : \|Dv\|_{\mathcal{M}(\Omega)} < \infty\}.$$

The *total variation* of a function  $v \in BV(\Omega)$  is then given by

$$\text{TV}(v) := \|Dv\|_{\mathcal{M}(\Omega)} = \int_{\Omega} d|Dv|_2,$$

*i.e.*, the total variation in the sense of measure theory of the vector measure  $Dv \in \mathcal{M}(\Omega; \mathbb{R}^d) = C(\overline{\Omega}; \mathbb{R}^d)^*$ . Here,  $|\cdot|_2$  denotes the Euclidean norm on  $\mathbb{R}^d$ ; we thus consider here the *isotropic* total variation. For  $v \in L^1(\Omega) \setminus BV(\Omega)$ , we set  $\text{TV}(v) = \infty$ .

The space  $BV(\Omega)$  is a Banach space if equipped with the norm

$$\|v\|_{BV(\Omega)} := \|v\|_{L^1(\Omega)} + \text{TV}(v),$$

see, *e.g.*, ([10], Thm. 10.1.1). Moreover, the space  $C^\infty(\overline{\Omega})$  is dense in  $BV(\Omega)$  with respect to *strict convergence*, *i.e.*, for any  $v \in BV(\Omega)$  there exists a sequence  $\{v_n\}_{n \in \mathbb{N}} \subset C^\infty(\overline{\Omega})$  such that

- (i)  $v_n \rightarrow v$  in  $L^1(\Omega)$  and
- (ii)  $\text{TV}(v_n) \rightarrow \text{TV}(v)$ ,

see, (*e.g.*, [10]), Thm. 10.1.2. In fact, a slight modification of the proof (which is based on approximation *via* mollification) shows that for  $v \in BV(\Omega) \cap L^p(\Omega)$  with  $1 < p < \infty$ , the convergence  $v_n \rightarrow v$  in (i) holds even strongly in  $L^p$  (since the constructed mollified sequence converges in  $L^p$  for any  $1 \leq p < \infty$ ; see, *e.g.*, ([10], Prop. 2.2.4)).

It follows that  $BV(\Omega)$  embeds into  $L^r(\Omega)$  continuously for every  $r \in [1, \frac{d}{d-1}]$  and compactly if  $r < \frac{d}{d-1}$ , see, e.g., ([4], Cor. 3.49 together with Prop. 3.21). Note that this requires  $\Omega$  to be a strongly Lipschitz domain. In addition, the total variation is lower semi-continuous with respect to strong convergence in  $L^1(\Omega)$ , i.e., if  $\{u_n\}_{n \in \mathbb{N}} \subset BV(\Omega)$  and  $u_n \rightarrow u$  in  $L^1(\Omega)$ , we have that

$$\text{TV}(u) \leq \liminf_{n \rightarrow \infty} \text{TV}(u_n), \quad (2.1)$$

see, e.g., ([44], Thm. 5.2.1). Note that this does not imply that  $\text{TV}(u) < \infty$  and hence that  $u \in BV(\Omega)$  unless  $\{\text{TV}(u_n)\}_{n \in \mathbb{N}}$  has a bounded subsequence. From (2.1), we also deduce that the convex extended real-valued functional  $\text{TV} : L^p(\Omega) \rightarrow \mathbb{R} \cup \{\infty\}$  is weakly lower semi-continuous for any  $p \in [1, \infty]$ .

## 2.2. Multibang penalty

Let  $u_1 < \dots < u_m$  be a given set of desired coefficient values. Here we assume that  $u_1 = 0$  and  $u_m = u_{\max} - u_{\min}$  such that for  $u(x) \in [u_1, u_m]$ , we have  $u(x) + u_{\min} \in [u_{\min}, u_{\max}]$ . The multi-bang penalty  $G$  is then defined similar to [21], where we have to replace the box constraints  $u(x) \in [u_1, u_m]$  by a linear growth to ensure that  $G$  is finite on  $L^r(\Omega)$ ,  $r < \infty$ . Specifically, we consider

$$G : L^1(\Omega) \rightarrow \mathbb{R}, \quad G(u) = \int_{\Omega} g(u(x)) \, dx,$$

where  $g : \mathbb{R} \rightarrow \mathbb{R}$  is given by

$$g(t) = \begin{cases} -u_m t & t \leq u_1, \\ \frac{1}{2}((u_i + u_{i+1})t - u_i u_{i+1}) & t \in [u_i, u_{i+1}], \quad 1 \leq i < m, \\ u_m t - \frac{1}{2}u_m^2 & t \geq u_m. \end{cases} \quad (2.2)$$

It can be verified easily that  $g$  is continuous (note that  $u_1 = 0$ ), convex, and linearly bounded from above and below, i.e.,

$$\frac{1}{2}u_2|t| \leq g(t) \leq u_m|t| \quad \text{for all } t \in \mathbb{R}.$$

**Remark 2.1.** The definition of  $g$  implies that  $g(t) > g(0) = 0$  for all  $t \neq 0$  and that  $g(t) > g(u_m)$  for all  $t > u_m = u_{\max} - u_{\min}$ . For the results of this section as well as of Sections 3 and 4, we only require these properties of  $g$  rather than the specific form of  $g$ . In particular, the results also hold for  $t \mapsto |t|$ , i.e., if  $G$  is replaced by the  $L^1$  norm.

Since  $g$  is finite (and hence proper), convex, and continuous, the corresponding integral operator  $G : L^r(\Omega) \rightarrow \mathbb{R}$  is finite, convex, and continuous (and hence *a fortiori* weakly lower semi-continuous) for any  $r \in [1, \infty]$ , see, e.g., ([11], Prop. 2.53). Also, the properties of  $g$  imply the following properties of  $G$ :

- (G1)  $G(v) > G(0) = 0$  for all  $v \in L^1(\Omega) \setminus \{0\}$ ,
- (G2)  $\frac{1}{2}u_2\|v\|_{L^1(\Omega)} \leq G(v) \leq u_m\|v\|_{L^1(\Omega)}$  for all  $v \in L^1(\Omega)$ .

Furthermore, for  $r < \infty$  and  $r' := \frac{r}{r-1}$  (with  $r' = \infty$  for  $r = 1$ ), the Fenchel conjugate

$$G^* : L^{r'}(\Omega) \rightarrow \mathbb{R} \cup \{\infty\}, \quad G^*(q) = \sup_{v \in L^r(\Omega)} \langle q, v \rangle_{L^{r'}(\Omega), L^r(\Omega)} - G(v),$$

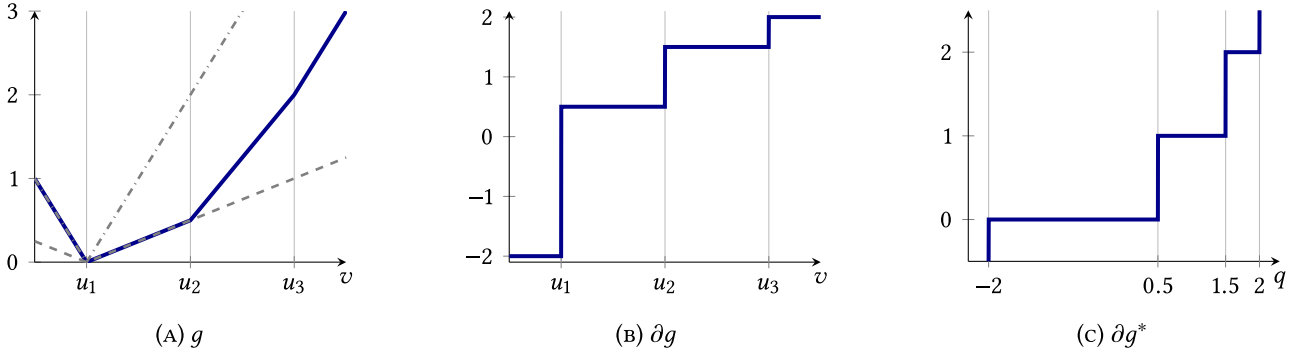


FIGURE 1. Pointwise multi-bang integrand  $g$ , subdifferential  $\partial g$ , and conjugate differential  $\partial g^*$  ( $u_1 = 0$ ,  $u_2 = 1$ ,  $u_3 = 2$ ).

as well as the convex subdifferential

$$\partial G(v) = \left\{ q \in L^r(\Omega) : \langle q, \tilde{v} - v \rangle_{L^r(\Omega), L^r(\Omega)} \leq G(\tilde{v}) - G(v) \quad \forall \tilde{v} \in L^r(\Omega) \right\},$$

can be computed pointwise, see, *e.g.*, ([22], Props. IV.1.2, IX.2.1) and ([11], Prop. 2.53), respectively. We point out that the pointwise representation of the subdifferential does *not* hold for  $r = \infty$ . From the definition of  $g$  we thus obtain that

$$[\partial G(v)](x) \in \begin{cases} \{-u_m\} & v(x) < u_1, \\ [-u_m, \frac{1}{2}(u_1 + u_2)] & v(x) = u_1, \\ \{\frac{1}{2}(u_i + u_{i+1})\} & v(x) \in (u_i, u_{i+1}), \quad 1 \leq i < m, \\ [\frac{1}{2}(u_{i-1} + u_i), \frac{1}{2}(u_i + u_{i+1})] & v(x) = u_i, \quad 1 < i < m, \\ [\frac{1}{2}(u_{m-1} + u_m), u_m] & v(x) = u_m, \\ \{u_m\} & v(x) > u_m, \end{cases} \quad (2.3)$$

where, by a slight abuse of notation,  $[\partial G(v)](x)$  stands for the evaluation of any  $q \in \partial G(v)$  at  $x \in \Omega$ . Using the fact that  $s \in \partial g(t)$  if and only if  $t \in \partial g^*(s)$  (see, *e.g.*, [38], Prop. 4.4.4), we deduce that

$$[\partial G^*(q)](x) \in \begin{cases} (-\infty, 0] & q(x) = -u_m, \\ \{0\} & q(x) \in (-u_m, \frac{1}{2}(u_1 + u_2)), \\ [u_i, u_{i+1}] & q(x) = \frac{1}{2}(u_i + u_{i+1}), \quad 1 \leq i < m, \\ \{u_i\} & q(x) \in (\frac{1}{2}(u_{i-1} + u_i), \frac{1}{2}(u_i + u_{i+1})), \quad 1 < i < m, \\ \{u_m\} & q(x) \in (\frac{1}{2}(u_{m-1} + u_m), u_m), \\ [u_m, \infty) & q(x) = u_m, \\ \emptyset & \text{else,} \end{cases} \quad (2.4)$$

almost everywhere; see Figure 1.

### 2.3. Superposition operator

To ensure well-posedness of the state equation, both coercivity of the differential operator and pointwise boundedness of the coefficients are required. This can be achieved by imposing pointwise bounds on the coefficients. Appending such bounds to the problem statement (P) would lead to difficulties when deriving pointwise

optimality conditions. As stated in the introduction, we therefore propose a reparametrization of the coefficient in the state equation. For this purpose we introduce the following family of (smoothed) pointwise projections onto the admissible set  $[u_{\min}, u_{\max}]$ . For fixed  $\varepsilon \geq 0$  we consider  $\varphi_\varepsilon : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$\varphi_\varepsilon(t) = u_{\min} + \begin{cases} -\varepsilon & \text{for } t \leq -\varepsilon, \\ -\frac{1}{\varepsilon^2}t^3 - \frac{1}{\varepsilon}t^2 + t & \text{for } t \in [-\varepsilon, 0], \\ t & \text{for } t \in [0, u_m], \\ -\frac{1}{\varepsilon^2}t^3 + \frac{3u_m + \varepsilon}{\varepsilon^2}t^2 + \frac{\varepsilon^2 - 2u_m\varepsilon - 3\varepsilon^2}{\varepsilon^2}t + \frac{u_m^3 + \varepsilon^2}{\varepsilon^2} & \text{for } t \in [u_m, u_m + \varepsilon], \\ u_m + \varepsilon & \text{for } t \geq u_m + \varepsilon, \end{cases} \quad (2.5)$$

where we have used that  $u_m = u_{\max} - u_{\min}$  from Section 2.2 (see Fig. 2A). For  $\varepsilon = 0$ , this coincides with the pointwise projection  $\text{proj}_{[u_{\min}, u_{\max}]}$ , while for  $\varepsilon > 0$  we have  $\varphi_\varepsilon \in C^{1,1}(\mathbb{R})$ . Clearly, there is a wide variety of choices which serves the purpose of making  $\varphi_\varepsilon$  continuously differentiable. It is appropriate to choose this exterior smoothing in such a manner that  $\varphi'_\varepsilon(t) \neq 0$  for  $t \in [0, u_m]$ . This will be further detailed in Remark 4.5. The reader will notice in the following that  $\varepsilon > 0$  is not used before deriving optimality conditions in Section 4.

Since  $\varphi_\varepsilon(t)$  is uniformly bounded and globally Lipschitz continuous, we deduce from [42], Lemma 4.11 that the corresponding superposition operator

$$\Phi_\varepsilon : L^r(\Omega) \rightarrow L^r(\Omega), \quad [\Phi_\varepsilon(v)](x) = \varphi_\varepsilon(v(x)) \quad \text{for a.e. } x \in \Omega,$$

is globally Lipschitz continuous for every  $r \in [1, \infty]$  and  $\varepsilon \geq 0$ .

Similarly, for any  $\varepsilon > 0$  it is easily verified that

$$\varphi'_\varepsilon(t) = \begin{cases} -\frac{3}{\varepsilon^2}t^2 - \frac{2}{\varepsilon}t + 1 & \text{for } t \in [-\varepsilon, 0] \\ 1 & \text{for } t \in [0, u_m], \\ -\frac{3}{\varepsilon^2}t^2 + \frac{6u_m + 2\varepsilon}{\varepsilon^2}t + \frac{\varepsilon^2 - 2u_m\varepsilon - 3u_m^2}{\varepsilon^2} & \text{for } t \in [u_m, u_m + \varepsilon], \\ 0 & \text{else,} \end{cases}$$

is locally Lipschitz continuous and uniformly bounded by 4/3 (see Fig. 2B). As a locally Lipschitz continuous function,  $\varphi'_\varepsilon$  is even globally Lipschitz on the compact set  $[-\varepsilon, u_m + \varepsilon]$ . Since  $\varphi'_\varepsilon(t) = 0$  for all  $t \in \mathbb{R} \setminus (-\varepsilon, u_m + \varepsilon)$ , we infer that  $\varphi'_\varepsilon$  is Lipschitz on all  $\mathbb{R}$ . Hence, it follows from ([42], Lem. 4.12, proof of Lem. 4.13) that  $\Phi_\varepsilon$  is Lipschitz continuously Fréchet differentiable from  $L^\infty(\Omega)$  to  $L^\infty(\Omega)$ , and that the Fréchet derivative  $\Phi'_\varepsilon(v) \in \mathcal{L}(L^\infty(\Omega), L^\infty(\Omega))$  at  $v \in L^\infty(\Omega)$  acting on  $h \in L^\infty(\Omega)$  is given by

$$[\Phi'_\varepsilon(v)h](x) = \varphi'_\varepsilon(v(x))h(x) \quad \text{for a.e. } x \in \Omega. \quad (2.6)$$

In particular,  $\Phi'_\varepsilon(v)$  can be represented pointwise almost everywhere by  $x \mapsto \varphi'_\varepsilon(v(x)) \in L^\infty(\Omega)$ . In the following, we will not distinguish the derivative and its representation.

## 2.4. State equation

It will be convenient to introduce for  $\varepsilon \geq 0$  the set

$$U_\varepsilon = \{v \in L^\infty(\Omega) : 0 < u_{\min} - \varepsilon \leq v \leq u_{\max} + \varepsilon \text{ a.e. in } \Omega\},$$

along with its open  $L^\infty(\Omega)$  neighborhood

$$\hat{U}_\varepsilon = \{v \in L^\infty(\Omega) : 0 < \frac{1}{2}u_{\min} - 2\varepsilon < v < 2u_{\max} + 2\varepsilon \text{ a.e. in } \Omega\}.$$

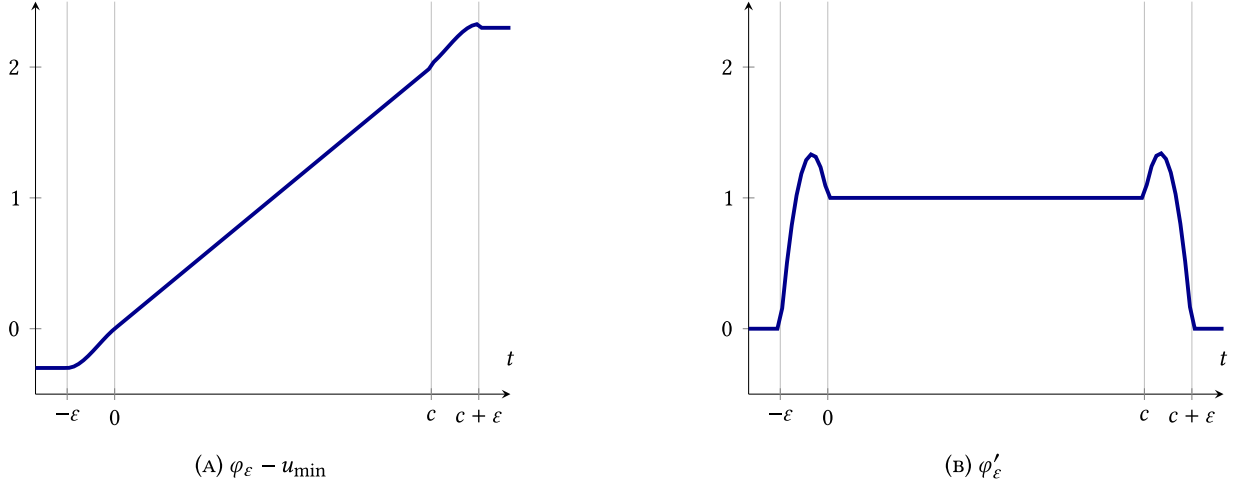


FIGURE 2. Smoothed projection  $\varphi_\varepsilon$  and derivative  $\varphi'_\varepsilon$  ( $c := u_m = 2$ ,  $\varepsilon = 0.3$ ).

Furthermore, we consider for  $w \in \hat{U}_\varepsilon$  and  $f \in L^2(\Omega)$  the elliptic partial differential equation

$$\begin{cases} -\operatorname{div}(w\nabla y) = f & \text{in } \Omega, \\ y = 0 & \text{on } \partial\Omega. \end{cases} \quad (2.7)$$

From standard arguments based on the Lax–Milgram lemma, we obtain the existence of a unique solution  $y \in H_0^1(\Omega)$  satisfying the uniform *a priori* estimate

$$\|y\|_{H_0^1(\Omega)} \leq K_2 \|f\|_{H^{-1}(\Omega)}, \quad (2.8)$$

for some  $K_2 > 0$  independent of  $w \in \hat{U}_\varepsilon$  (but depending on  $\hat{U}_\varepsilon$ ), where  $\|y\|_{H_0^1(\Omega)} = \|\nabla y\|_{L^2(\Omega)^d}$ . We also have the following global Lipschitz estimate for the solution mapping  $w \mapsto y =: y(w)$ .

**Lemma 2.2.** *For any  $\varepsilon \geq 0$  there exists a constant  $L > 0$  such that*

$$\|y(w_1) - y(w_2)\|_{H_0^1(\Omega)} \leq L \|w_1 - w_2\|_{L^\infty(\Omega)} \quad \text{for all } w_1, w_2 \in \hat{U}_\varepsilon.$$

*Proof.* Let  $y_1, y_2 \in H_0^1(\Omega)$  denote the solutions to (2.7) for  $w_1, w_2 \in \hat{U}_\varepsilon$ , respectively. Inserting  $y_1 - y_2 \in H_0^1(\Omega)$  as a test function in (2.7) for  $w = w_1$  and  $w = w_2$ , subtracting, inserting the productive zero, and rearranging yields

$$(w_1 \nabla(y_1 - y_2), \nabla(y_1 - y_2))_{L^2(\Omega)^d} = ((w_2 - w_1) \nabla y_2, \nabla(y_1 - y_2))_{L^2(\Omega)^d}.$$

Estimating the left-hand side using the uniform lower bound on  $w$  and the right-hand side using the Cauchy–Schwarz inequality and the *a priori* estimate (2.8), we obtain

$$\begin{aligned} \left(\frac{1}{2}u_{\min} - 2\varepsilon\right) \|\nabla(y_1 - y_2)\|_{L^2(\Omega)^d}^2 &\leq \|w_1 - w_2\|_{L^\infty(\Omega)} \|\nabla y_2\|_{L^2(\Omega)^d} \|\nabla(y_1 - y_2)\|_{L^2(\Omega)^d} \\ &\leq K_2 \|f\|_{H^{-1}(\Omega)} \|w_1 - w_2\|_{L^\infty(\Omega)} \|\nabla(y_1 - y_2)\|_{L^2(\Omega)^d}, \end{aligned}$$

from which the desired estimate follows with  $L := \frac{K_2}{\frac{1}{2}u_{\min} - 2\varepsilon} \|f\|_{H^{-1}(\Omega)}$ .  $\square$

Our next goal is to establish that there exists an  $s > 2$  such that the solution  $y$  of (2.7) belongs to  $W^{1,s}(\Omega)$ . This increase in regularity is crucial for obtaining pointwise optimality conditions. The proof relies on results from Gröger [24].

**Proposition 2.3.** *There exists an  $s > 2$  and a constant  $K_s > 0$  such that for all  $w \in \hat{U}_\varepsilon$  the solution  $y \in H_0^1(\Omega)$  of (2.7) satisfies*

$$\|y\|_{W^{1,s}(\Omega)} \leq K_s \|f\|_{W^{-1,s}(\Omega)}.$$

*Proof.* Fix  $w \in \hat{U}_\varepsilon$  and  $f \in L^2(\Omega)$  and denote by  $y \in H_0^1(\Omega)$  the solution to (2.7). By the Sobolev embedding theorem, there exists an  $\bar{s} > 2$  such that  $L^2(\Omega)$  is continuously embedded in  $W^{-1,s_1}(\Omega)$  for all  $s_1 \in (2, \bar{s}]$ . Furthermore, by Lemma A.3 the domain  $\Omega$  is regular in the sense of Gröger. Hence ([24], Thm. 3) implies that  $\Omega \in R_{s_2}$  for some  $s_2 > 2$  and thus by ([24], Lem. 1) for  $s := \min\{s_1, s_2\} > 2$  as well. We therefore obtain from ([24], Thm. 1) for any  $q \in W^{-1,s}(\Omega)$  that the unique solution  $\hat{y} \in H_0^1(\Omega)$  of

$$\begin{cases} -\operatorname{div}(w\nabla\hat{y}) + \hat{y} = q & \text{in } \Omega, \\ \hat{y} = 0 & \text{on } \partial\Omega, \end{cases} \quad (2.9)$$

satisfies  $\|\hat{y}\|_{W^{1,s}(\Omega)} \leq K\|q\|_{W^{-1,s}(\Omega)}$ , where  $K$  denotes a constant that depends on  $\hat{U}_\varepsilon$  but not on  $w$ ,  $\hat{y}$ , or  $q$ . For the choice  $q = y + f$  this yields  $\|\hat{y}\|_{W^{1,s}(\Omega)} \leq K(C\|y\|_{L^2(\Omega)} + \|f\|_{W^{-1,s}(\Omega)})$ , where  $C$  denotes the constant of the continuous embedding  $L^2(\Omega) \hookrightarrow W^{-1,s}(\Omega)$ . Using the continuous embedding  $H_0^1(\Omega) \hookrightarrow L^2(\Omega)$  with constant  $\hat{C}$ , the *a priori* estimate (2.8), and the continuous embedding  $W^{-1,s}(\Omega) \hookrightarrow H^{-1}(\Omega)$  with constant  $\bar{C}$ , we obtain  $\|\hat{y}\|_{W^{1,s}(\Omega)} \leq K(C\hat{C}K_2\bar{C}\|f\|_{W^{-1,s}(\Omega)} + \|f\|_{W^{-1,s}(\Omega)})$ . Since for fixed  $\hat{U}_\varepsilon$  all appearing constants are independent of  $w$ , the claim follows by noting that the choice of  $q$  implies that  $y$  solves (2.9), hence  $\hat{y} = y$ .  $\square$

### 3. EXISTENCE

To show existence of a solution to (P), we make use of the solution mapping  $w \mapsto y(w)$  to introduce the reduced functional

$$J : BV(\Omega) \rightarrow \mathbb{R}, \quad J(u) = \frac{1}{2} \|y(\Phi_\varepsilon(u)) - z\|_{L^2(\Omega)}^2 + \alpha G(u) + \beta \operatorname{TV}(u).$$

**Proposition 3.1.** *For every  $\varepsilon \geq 0$  there exists a global minimizer  $\bar{u} \in BV(\Omega)$  to (P).*

*Proof.* Since  $J$  is bounded from below due to (G1), there exists a minimizing sequence  $\{u_n\}_{n \in \mathbb{N}} \subset BV(\Omega)$ . Furthermore, by (G2), we may assume without loss of generality that there exists a  $C > 0$  such that

$$C(\|u_n\|_{L^1(\Omega)} + \operatorname{TV}(u_n)) \leq J(u_n) \leq J(0) \quad \text{for all } n \in \mathbb{N},$$

and hence that  $\{u_n\}_{n \in \mathbb{N}}$  is bounded in  $BV(\Omega)$ . By the compact embedding of  $BV(\Omega)$  into  $L^1(\Omega)$  for any  $d \in \mathbb{N}$ , we can thus extract a subsequence, denoted by the same symbol, converging strongly in  $L^1(\Omega)$  to some  $\bar{u} \in L^1(\Omega)$ . Lipschitz continuity of  $\Phi_\varepsilon$  from  $L^1(\Omega)$  to  $L^1(\Omega)$  now implies that  $\Phi_\varepsilon(u_n) \rightarrow \Phi_\varepsilon(\bar{u})$  in  $L^1(\Omega)$  as well. Furthermore, the corresponding sequence  $\{y(\Phi_\varepsilon(u_n))\}_{n \in \mathbb{N}}$  is uniformly bounded in  $H_0^1(\Omega)$  due to (2.8), and hence there exists a  $\bar{y} \in H_0^1(\Omega)$  such that, after passing to a further subsequence if necessary,  $y(\Phi_\varepsilon(u_n)) \rightharpoonup \bar{y}$  in  $H_0^1(\Omega)$ . Since  $\{\Phi_\varepsilon(u_n)\}_{n \in \mathbb{N}}$  is uniformly bounded in  $L^\infty(\Omega)$  by construction, we have that  $\Phi_\varepsilon(u_n) \rightarrow \Phi_\varepsilon(\bar{u})$  strongly in  $L^r(\Omega)$  for any  $r \in [1, \infty)$  and, in particular, for  $r = 2$ . We can thus pass to the limit in the distributional formulation of (2.7),

$$(\Phi_\varepsilon(u_n), \nabla y(\Phi_\varepsilon(u_n)) \cdot \nabla \psi)_{L^2(\Omega)} = (f, \psi)_{L^2(\Omega)} \quad \text{for all } \psi \in C_0^\infty(\Omega),$$



to obtain

$$(\Phi_\varepsilon(\bar{u}), \nabla \bar{y} \cdot \nabla \psi)_{L^2(\Omega)} = (f, \psi)_{L^2(\Omega)} \quad \text{for all } \psi \in C_0^\infty(\Omega).$$

By density, we obtain that  $\bar{y} = y(\Phi_\varepsilon(\bar{u}))$  and hence that  $y(\Phi_\varepsilon(u_n)) \rightarrow y(\Phi_\varepsilon(\bar{u}))$  strongly in  $L^2(\Omega)$ . Finally, lower semi-continuity of  $G$  and TV with respect to convergence in  $L^1(\Omega)$  and the strong convergence  $y(\Phi_\varepsilon(u_n)) \rightarrow y(\Phi_\varepsilon(\bar{u}))$  in  $L^2(\Omega)$  imply that

$$J(\bar{u}) \leq \liminf_{n \rightarrow \infty} J(u_n) \leq J(u) \quad \text{for all } u \in BV(\Omega),$$

and thus that  $\bar{u} \in BV(\Omega)$  is the desired minimizer.  $\square$

Due to the bilinear structure of the state equation the optimal control is not unique. Nonetheless, as a consequence of the reparametrization of the control by means of  $\Phi_\varepsilon$ , any solution to (P) automatically satisfies pointwise control constraints.

**Proposition 3.2.** *Let  $\varepsilon \geq 0$  and  $\bar{u} \in BV(\Omega)$  be a local solution to (P). Then,  $\bar{u} + u_{\min} \in U_\varepsilon \subset L^\infty(\Omega)$ .*

*Proof.* Let  $\varepsilon \geq 0$  and  $\bar{u} \in BV(\Omega)$  with  $\bar{u} + u_{\min} \notin U_\varepsilon$ . We will show that  $\bar{u}$  is not a local solution to (P). We start by comparing  $\bar{u}$  to  $\hat{u}$  defined pointwise almost everywhere by

$$\hat{u}(x) = \begin{cases} -\varepsilon & \bar{u}(x) < -\varepsilon, \\ \bar{u}(x) & \bar{u}(x) \in [-\varepsilon, u_m + \varepsilon], \\ u_m + \varepsilon & \bar{u}(x) > u_m + \varepsilon. \end{cases}$$

By definition of  $\varphi_\varepsilon$ , it follows that  $\Phi_\varepsilon(\hat{u}) = \Phi_\varepsilon(\bar{u})$  and thus that  $y(\Phi_\varepsilon(\hat{u})) = y(\Phi_\varepsilon(\bar{u}))$ .

Furthermore, from Stampacchia's lemma for BV functions ([37], Lem. 2.5) we obtain that  $\text{TV}(\hat{u}) \leq \text{TV}(\bar{u})$ . Using the pointwise definition of  $G$  together with the inequalities  $g(t) > g(-\varepsilon) > 0$  for all  $t < -\varepsilon$  and  $g(t) > g(u_m + \varepsilon)$  for all  $t > u_m + \varepsilon$ , we also deduce that  $G(\hat{u}) < G(\bar{u})$  since  $\bar{u} + u_{\min} \notin U_\varepsilon$ . Thus,  $J(\hat{u}) < J(\bar{u})$ . Similarly, we observe that  $y(\Phi_\varepsilon(u_t)) = y(\Phi_\varepsilon(\bar{u}))$  for all  $t \in [0, 1]$ , where we have denoted  $u_t := (1-t)\hat{u} + t\bar{u}$ . Using  $\text{TV}(\hat{u}) \leq \text{TV}(\bar{u})$  and  $G(\hat{u}) < G(\bar{u})$  together with the convexity of TV and  $G$  yields that  $\text{TV}(u_t) \leq \text{TV}(\bar{u})$  and  $G(u_t) < G(\bar{u})$  for all  $t \in [0, 1)$ . It follows that  $J(u_t) < J(\bar{u})$  for all  $t \in [0, 1)$  and hence that  $\bar{u}$  is not a local solution to (P).  $\square$

By Proposition 3.2, for any  $\varepsilon \geq 0$ , each locally optimal control to problem (P) is therefore also a local solution of

$$\min_{u \in BV(\Omega) \cap L^\infty(\Omega)} J(u),$$

and, moreover, the set of globally optimal controls is the same for both problems. In particular, the solutions  $\bar{u}$  to (P) for  $\varepsilon = 0$  coincide with the solutions to

$$\begin{cases} \min_{u \in BV(\Omega)} \frac{1}{2} \|y - z\|_{L^2(\Omega)}^2 + \alpha G(u) + \beta \text{TV}(u), \\ \text{s.t.} & u(x) + u_{\min} \in [u_{\min}, u_{\max}], \\ \text{and} & -\text{div}((u + u_{\min})\nabla y) = f \quad \text{in } \Omega, \\ & y = 0 \quad \text{on } \partial\Omega, \end{cases} \quad (\text{P}^*)$$

which is a particular case of the motivating problem (PI).

**Remark 3.3.** The same cut-off argument as in the proof of Proposition 3.2 can be applied to the minimizing sequence in the proof of Proposition 3.1 to construct a minimizing sequence that is bounded in  $L^\infty(\Omega)$  and hence in  $L^1(\Omega)$  even for  $\alpha = 0$ . We thus also obtain the existence of a solution  $\bar{u}$  to  $(P^*)$  with  $\alpha = 0$ . The results in the following Section 4 remain valid in this case, and the optimality conditions derived therein simplify in an obvious manner.

We close this section by briefly addressing the convergence of global solutions to  $(P)$  as  $\varepsilon \rightarrow 0^+$ . For this purpose we consider a family  $\{\bar{u}_\varepsilon\}_{\varepsilon>0}$  of solutions to  $(P)$ . From Proposition 3.2 and the fact that  $J(0)$  is independent of  $\varepsilon$ , we deduce that this family is bounded in  $L^\infty(\Omega) \cap BV(\Omega)$  as  $\varepsilon \rightarrow 0^+$ . Thus, there exists a sequence  $\{\bar{u}_{\varepsilon_k}\}_{k \in \mathbb{N}}$  converging strongly to some  $\bar{u}$  in  $L^r(\Omega)$  for every  $r \in [2, \infty)$  with  $\text{TV}(\bar{u}) \leq \liminf_{k \rightarrow \infty} \text{TV}(\bar{u}_{\varepsilon_k}) < \infty$ . With some modifications (in particular using that for every  $u \in BV(\Omega)$  there holds  $\Phi_{\varepsilon_k}(u) \rightarrow \Phi_0(u) = \text{proj}_{[u_{\min}, u_{\max}]}(u)$  strongly in  $L^1(\Omega)$  for  $k \rightarrow \infty$ ), the proof of Proposition 3.1 can now be used to verify that  $\bar{u}$  is a global solution to  $(P)$  for  $\varepsilon = 0$  and thus for  $(P^*)$ .

#### 4. OPTIMALITY CONDITIONS

In this section, we derive *pointwise* necessary optimality conditions for solutions to problem  $(P)$ . Since we will require differentiability of the control-to-state operator  $u \mapsto y(\Phi_\varepsilon(u))$ , we have to assume  $\varepsilon > 0$  from here on. To keep the presentation simple, we will from now omit the dependence on  $\varepsilon$ . The derivation rests crucially on the following two nontrivial properties:

- (i) By Proposition 3.2, we can work in the  $L^\infty(\Omega)$  topology rather than in the  $L^{\frac{d}{d-1}}(\Omega)$  topology induced by  $BV(\Omega)$ , which allows differentiability of the forward mapping.
- (ii) By Proposition 2.3, the derivative of the forward mapping is actually in  $L^r(\Omega)$  for some  $r > 1$ , which will yield multipliers in  $L^r(\Omega)$  instead of  $L^\infty(\Omega)^*$ .

We begin by showing differentiability of the reduced tracking term

$$F : \hat{U} \rightarrow \mathbb{R}, \quad F(w) = \frac{1}{2} \|y(w) - z\|_{L^2(\Omega)}^2. \quad (4.1)$$

This can be argued from differentiability of the forward mapping  $w \mapsto y(w)$  in  $L^\infty(\Omega)$  (see, *e.g.*, [1]) together with the chain rule. However, it actually holds under the weaker requirement of Lipschitz continuity of the forward mapping shown in Lemma 2.2. Since this argument may be of independent interest, we give a full proof here.

We first introduce for a given parameter  $w \in \hat{U} \subset L^\infty(\Omega)$  and  $y \in H_0^1(\Omega)$  the adjoint equation

$$\begin{cases} -\text{div}(w \nabla p) = -(y - z) & \text{in } \Omega, \\ p = 0 & \text{on } \partial\Omega. \end{cases} \quad (4.2)$$

By the same arguments as for the state equation (2.7) there exists a unique solution  $p = p(w, y) \in H_0^1(\Omega)$ , which depends continuously on  $y$  and for which the additional regularity  $p(w, y) \in W^{1,s}(\Omega)$  from Proposition 2.3 holds.

**Lemma 4.1.** *The mapping  $F$  defined in (4.1) is Lipschitz continuously Fréchet differentiable in every  $w \in \hat{U} \subset L^\infty(\Omega)$ . Furthermore, the Fréchet derivative of  $F$  in  $w \in \hat{U}$  is given by*

$$F'(w) = \nabla y(w) \cdot \nabla p(w) \in L^{\frac{s}{2}}(\Omega),$$

with  $s > 2$  from Proposition 2.3, where  $y(w) \in H_0^1(\Omega)$  is the solution to (2.7) and  $p(w) := p(w, y(w)) \in H_0^1(\Omega)$  is the corresponding solution to (4.2).

*Proof.* We first show directional differentiability in  $\hat{U} \subset L^\infty(\Omega)$ . Let  $w \in \hat{U}$  and  $h \in L^\infty(\Omega)$ . Then there exists a  $\rho_0 > 0$  sufficiently small such that  $w + \rho h \in \hat{U}$  for all  $\rho \in (0, \rho_0)$ . Consequently, for all such  $\rho$  there exists a solution  $y(w + \rho h) \in H_0^1(\Omega)$  to (2.7). We now insert the productive zero  $y(w) - y(w)$  in  $F(w + \rho h)$  and expand the square to obtain

$$\begin{aligned} F(w + \rho h) - F(w) &= \frac{1}{2} \|(y(w + \rho h) - y(w)) + (y(w) - z)\|_{L^2(\Omega)}^2 - \frac{1}{2} \|y(w) - z\|_{L^2(\Omega)}^2 \\ &= \frac{1}{2} \|y(w + \rho h) - y(w)\|_{L^2(\Omega)}^2 + (y(w + \rho h) - y(w), y(w) - z)_{L^2(\Omega)}. \end{aligned} \quad (4.3)$$

For the first term, we can use Lemma 2.2 to estimate

$$\frac{1}{2} \|y(w + \rho h) - y(w)\|_{L^2(\Omega)}^2 \leq \frac{L^2}{2} \rho^2 \|h\|_{L^\infty(\Omega)}^2. \quad (4.4)$$

For the second term, we introduce the adjoint state  $p(w)$ , integrate by parts, and use the state equation (2.7) for  $y = y(w)$  and  $y = y(w + \rho h)$  to obtain

$$\begin{aligned} (y(w + \rho h) - y(w), y(w) - z)_{L^2(\Omega)} &= (y(w + \rho h) - y(w), \operatorname{div}(w \nabla p))_{L^2(\Omega)} \\ &= (\operatorname{div}(w \nabla y(w + \rho h)), p)_{L^2(\Omega)} - (\operatorname{div}(w \nabla y(w)), p)_{L^2(\Omega)} \\ &= (-f, p)_{L^2(\Omega)} - (\operatorname{div}(\rho h \nabla y(w + \rho h)), p)_{L^2(\Omega)} - (-f, p)_{L^2(\Omega)} \\ &= \rho (h \nabla y(w + \rho h), \nabla p)_{L^2(\Omega)}. \end{aligned}$$

By Lemma 2.2 we have that  $y(w + \rho h) \rightarrow y(w)$  in  $H_0^1(\Omega)$  as  $\rho \rightarrow 0^+$ . Hence, dividing (4.3) by  $\rho > 0$  and passing to the limit implies in combination with (4.4) that

$$F'(w; h) := \lim_{\rho \rightarrow 0^+} \frac{1}{\rho} (F(w + \rho h) - F(w)) = \langle h, \nabla y \cdot \nabla p \rangle_{L^\infty(\Omega), L^1(\Omega)}.$$

Since the mapping  $h \mapsto F'(w; h)$  is linear and bounded,  $\nabla y \cdot \nabla p$  is the Gâteaux derivative of  $F$  at  $w \in \hat{U}$ . Thus,  $F$  is Gâteaux differentiable in  $\hat{U}$ . Due to Lemma 2.2 the mappings  $w \mapsto y(w)$  and  $w \mapsto p(w, y)$  are Lipschitz from  $L^\infty(\Omega)$  to  $H_0^1(\Omega)$  in  $\hat{U}$ . By using (2.8), we infer that the mapping  $y \mapsto p(w, y)$  is Lipschitz from  $H_0^1(\Omega)$  to  $H_0^1(\Omega)$  for any fixed  $w \in \hat{U}$ , with a Lipschitz constant independent of  $w$ . This shows that  $w \mapsto p(w) := p(w, y(w))$  is Lipschitz continuous from  $L^\infty(\Omega)$  to  $H_0^1(\Omega)$  in  $\hat{U}$ . Hence, the mapping  $w \mapsto \nabla y(w) \cdot \nabla p(w)$  is Lipschitz continuous from  $L^\infty(\Omega)$  to  $L^1(\Omega)$  in  $\hat{U}$ , and thus  $F$  is in fact Fréchet differentiable in  $\hat{U}$  with Lipschitz continuous derivative. The regularity  $\nabla y(w) \cdot \nabla p(w) \in L^{\frac{s}{2}}(\Omega)$  follows from Proposition 2.3.  $\square$

Together with the Fréchet differentiability of  $\Phi$  in  $L^\infty(\Omega)$ , this allows deriving abstract first-order necessary optimality conditions using classical tools from convex analysis. Here it is crucial that  $G$  does not incorporate pointwise constraints and is finite on  $L^p(\Omega)$  for  $p = \frac{s}{s-2} > 1$  instead of  $p = 1$  in order to apply the sum rule to its convex subdifferential (considered as a subset of  $L^q(\Omega)$  with  $q = \frac{s}{2} < \infty$ ), which requires the effective domain of  $G$  to have non-empty interior.

**Theorem 4.2.** *Any local minimizer  $\bar{u} \in BV(\Omega)$  to (P) satisfies*

$$-F'(\Phi(\bar{u}))\Phi'(\bar{u}) \in \alpha \partial G(\bar{u}) + \beta \partial \operatorname{TV}(\bar{u}) \subset L^{\frac{s}{2}}(\Omega), \quad (4.5)$$

where  $G$  and  $\operatorname{TV}$  are considered as extended real-valued convex functionals on  $L^{\frac{s}{s-2}}(\Omega)$ .

*Proof.* Let  $\bar{u} \in BV(\Omega)$  be a local minimizer to (P). Proposition 3.2 shows that  $\bar{u}$  is also a local minimizer in  $BV(\Omega) \cap L^\infty(\Omega)$ . Thus, for all  $u \in BV(\Omega) \cap L^\infty(\Omega)$  and  $t > 0$  sufficiently small, we have that

$$F(\Phi(\bar{u})) + \alpha G(\bar{u}) + \beta \text{TV}(\bar{u}) \leq F(\Phi(\bar{u} + t(u - \bar{u}))) + \alpha G(\bar{u} + t(u - \bar{u})) + \beta \text{TV}(\bar{u} + t(u - \bar{u})).$$

We now proceed as in the proof of Proposition 2.2 from [21], using the convexity of  $G$  and  $\text{TV}$  to obtain after rearranging that

$$\frac{1}{t} (F(\Phi(\bar{u} + t(u - \bar{u}))) - F(\Phi(\bar{u}))) + \alpha (G(u) - G(\bar{u})) + \beta (\text{TV}(u) - \text{TV}(\bar{u})) \geq 0.$$

By Lemma 4.1 and the chain rule,  $F \circ \Phi$  is Fréchet differentiable at  $\bar{u} \in L^\infty(\Omega)$ , and the Fréchet derivative is given by

$$(F \circ \Phi)'(\bar{u}) = F'(\Phi(\bar{u}))\Phi'(\bar{u}) \in L^\infty(\Omega)^*.$$

Since Lemma 4.1 further implies that  $F'(\Phi(\bar{u})) \in L^{\frac{s}{2}}(\Omega)$ , and since we have  $\Phi'(\bar{u}) \in L^\infty(\Omega)$  from the representation (2.6), we deduce that in fact  $(F \circ \Phi)'(\bar{u}) \in L^{\frac{s}{2}}(\Omega) \subset L^1(\Omega)$ . Hence, we can pass to the limit  $t \rightarrow 0^+$  to obtain

$$\langle F'(\Phi(\bar{u}))\Phi'(\bar{u}), u - \bar{u} \rangle_{L^1(\Omega), L^\infty(\Omega)} + \alpha (G(u) - G(\bar{u})) + \beta (\text{TV}(u) - \text{TV}(\bar{u})) \geq 0,$$

for all  $u \in BV(\Omega) \cap L^\infty(\Omega)$ .

By the density of  $C^\infty(\bar{\Omega})$  in  $L^{\frac{s}{s-2}}(\Omega) \cap BV(\Omega)$  with respect to strict convergence, there exists for any  $u \in L^{\frac{s}{s-2}}(\Omega) \cap BV(\Omega)$  a sequence  $\{u_n\}_{n \in \mathbb{N}} \subset C^\infty(\bar{\Omega})$  with  $u_n \rightarrow u$  strongly in  $L^{\frac{s}{s-2}}$ . Hence,  $G(u_n) \rightarrow G(u)$  by continuity of  $G$ ,  $\text{TV}(u_n) \rightarrow \text{TV}(u)$ , and

$$\langle F'(\Phi(\bar{u}))\Phi'(\bar{u}), u_n - \bar{u} \rangle_{L^1(\Omega), L^\infty(\Omega)} \rightarrow \langle F'(\Phi(\bar{u}))\Phi'(\bar{u}), u - \bar{u} \rangle_{L^{\frac{s}{2}}(\Omega), L^{\frac{s}{s-2}}(\Omega)}.$$

Taking  $\text{TV}(u) = \infty$  for  $u \in L^{\frac{s}{s-2}}(\Omega) \setminus BV(\Omega)$ , we deduce that

$$\langle F'(\Phi(\bar{u}))\Phi'(\bar{u}), u - \bar{u} \rangle_{L^{\frac{s}{2}}(\Omega), L^{\frac{s}{s-2}}(\Omega)} + \alpha (G(u) - G(\bar{u})) + \beta (\text{TV}(u) - \text{TV}(\bar{u})) \geq 0,$$

holds for all  $u \in L^{\frac{s}{s-2}}(\Omega)$ . But this implies by definition that

$$-F'(\Phi(\bar{u}))\Phi'(\bar{u}) \in \partial(\alpha G + \beta \text{TV})(\bar{u}) \subset L^{\frac{s}{2}}(\Omega),$$

where the subdifferentials are understood as those of the canonical restriction to  $L^{\frac{s}{s-2}}(\Omega)$ .

Finally, since  $\text{dom TV} = BV(\Omega) \cap L^{\frac{s}{s-2}}(\Omega) \subset L^{\frac{s}{s-2}}(\Omega) = \text{dom } G$  and  $G$  is continuous on  $L^{\frac{s}{s-2}}(\Omega)$ , we can apply the sum rule for convex subdifferentials (see, e.g., [38], Prop. 4.5.1) to obtain (4.5).  $\square$

Introducing explicit subgradients for the two subdifferentials, we obtain primal-dual optimality conditions.

**Corollary 4.3.** *For any local minimizer  $\bar{u} \in BV(\Omega)$  to (P), there exist  $\bar{q} \in L^{\frac{s}{2}}(\Omega)$  and  $\bar{\xi} \in L^{\frac{s}{2}}(\Omega)$  satisfying*

$$\begin{cases} 0 = F'(\Phi(\bar{u}))\Phi'(\bar{u}) + \alpha \bar{q} + \beta \bar{\xi}, \\ \bar{q} \in \partial G(\bar{u}), \\ \bar{\xi} \in \partial \text{TV}(\bar{u}). \end{cases} \tag{4.6}$$

From Corollary 4.3, we can further derive *pointwise* optimality conditions for optimal controls. For the Fréchet derivative of the tracking term and the subdifferential of the multi-bang penalty, we apply Lemma 4.1 together

with the representations (2.6) and (2.3), respectively. The characterization of  $\bar{\xi} \in \partial\text{TV}(\bar{u})$  is more involved. Formally, elements of the subdifferential  $\partial\text{TV}(u)$  have the form  $-\text{div}\left(\frac{\nabla u}{|\nabla u|_2}\right)$ , which is equal to the negative mean curvature of the level sets of  $u$ . This can be made rigorous using the *full trace* from [15], which requires some notation. First, we introduce for  $1 \leq q < \infty$  the space

$$W^{\text{div},q}(\Omega) := \left\{v \in L^q(\Omega; \mathbb{R}^d) : \text{div } v \in L^q(\Omega)\right\},$$

endowed with the graph norm. Furthermore, for any Radon measure  $\mu$ , let  $L^1_\mu(\Omega; \mathbb{R}^d)$  denote the space of  $\mu$ -measurable functions  $v : \Omega \rightarrow \mathbb{R}^d$  for which

$$\|v\|_{L^1_\mu(\Omega; \mathbb{R}^d)} := \int_\Omega |v(x)|_2 \, d\mu,$$

is finite. To any  $v \in W^{\text{div},q}(\Omega) \cap L^\infty(\Omega)$ , we can then assign a unique  $Tv \in L^1_{|Du|}(\Omega; \mathbb{R}^d)$ , called the *full trace* of  $v$ , using appropriate converging sequences; see ([15], Def. 12) for a precise definition. Finally, we recall the decomposition of the measure  $Du$  for  $u \in BV(\Omega)$  into an absolutely continuous part  $D^a u = \nabla u \, d\mathcal{L}^d$  with respect to the  $d$ -dimensional Lebesgue measure  $\mathcal{L}^d$ , a jump part

$$D^j u = (u^+ - u^-) \nu_u \, d\mathcal{H}^{d-1}|_{S_u},$$

where  $u^+ - u^-$  denotes the jump of  $u$  on the singularity set  $S_u$  with normal  $\nu_u$  and  $(d - 1)$ -dimensional Hausdorff measure  $\mathcal{H}^{d-1}$ , and the Cantor part  $D^c u$  with density  $\sigma_u$  with respect to  $|D^c u|$ . We can now state fully our pointwise optimality conditions.

**Theorem 4.4.** *For any local minimizer  $\bar{u} \in BV(\Omega)$  to (P), there exist  $\bar{y}, \bar{p} \in W^{1,s}(\Omega)$ ,  $\bar{q} \in L^{\frac{s}{2}}(\Omega)$ , and  $\bar{\psi} \in W^{\text{div}, \frac{s}{2}}(\Omega)$  satisfying*

$$\begin{cases} -\text{div}(\Phi(\bar{u})\nabla\bar{y}) = f & \text{in } \Omega, \\ \bar{y} = 0 & \text{on } \partial\Omega, \end{cases} \tag{4.7a}$$

$$\begin{cases} -\text{div}(\Phi(\bar{u})\nabla\bar{p}) = -(\bar{y} - z) & \text{in } \Omega, \\ \bar{p} = 0 & \text{on } \partial\Omega, \end{cases} \tag{4.7b}$$

$$(\nabla\bar{y} \cdot \nabla\bar{p})\Phi'(\bar{u}) + \alpha\bar{q} - \beta \text{div } \bar{\psi} = 0 \quad \text{in } L^{\frac{s}{2}}(\Omega), \tag{4.7c}$$

$$\bar{u}(x) \in \begin{cases} (-\infty, u_1] & \bar{q}(x) = -u_m, \\ \{u_1\} & \bar{q}(x) \in (-u_m, \frac{1}{2}(u_1 + u_2)), \\ [u_i, u_{i+1}] & \bar{q}(x) = \frac{1}{2}(u_i + u_{i+1}), \quad 1 \leq i < m, \\ \{u_i\} & \bar{q}(x) \in (\frac{1}{2}(u_{i-1} + u_i), \frac{1}{2}(u_i + u_{i+1})), \quad 1 < i < m, \\ \{u_m\} & \bar{q}(x) \in (\frac{1}{2}(u_{m-1} + u_m), u_m), \\ [u_m, \infty) & \bar{q}(x) = u_m, \\ \emptyset & \text{else,} \end{cases} \tag{4.7d}$$

$$\left\{ \begin{array}{ll} |\bar{\psi}(x)|_2 \leq 1 & \text{for a.e. } x \in \Omega, \\ \bar{\psi}(x) = \frac{\nabla \bar{u}(x)}{|\nabla \bar{u}(x)|_2} & \text{for a.e. } x \in \Omega \text{ with } \nabla \bar{u}(x) \neq 0, \\ (T\bar{\psi})(x) = \frac{\bar{u}^+(x) - \bar{u}^-(x)}{|\bar{u}^+(x) - \bar{u}^-(x)|} v_{\bar{u}}(x) & \text{for } \mathcal{H}^{d-1}\text{-a.e. } x \in S_{\bar{u}}, \\ (T\bar{\psi})(x) = \sigma_{\bar{u}}(x) & \text{for } |D^c \bar{u}|\text{-a.e. } x \in \Omega. \end{array} \right. \tag{4.7e}$$

*Proof.* We start with (4.7c), which is obtained from the first equation of (4.6) by using Lemma 4.1 to express  $F'(\Phi(\bar{u}))\Phi'(\bar{u})$  in terms of the solution  $\bar{y}$  to the state equation (4.7a) and the solution  $\bar{p}$  to the adjoint equation (4.7b). Furthermore, we have used ([15], Prop. 8), which states that any  $\bar{\xi} \in \partial \text{TV}(\bar{u}) \cap L^q(\Omega)$  can be expressed as  $\bar{\xi} = -\text{div } \bar{\psi}$  for a  $\bar{\psi} \in W^{\text{div},q}(\Omega)$  satisfying (4.7e).<sup>1</sup> We point out that the  $L^p(\Omega)$ ,  $p > 1$ , regularity of  $(F \circ \Phi)'(\bar{u})$  is crucial to allow applying this result, and that it holds for strongly Lipschitz domains. Finally, the second relation in (4.6) can be equivalently written as  $\bar{u} \in \partial G^*(\bar{q})$ , which by (2.4) admits the pointwise characterization (4.7d).  $\square$

Let us briefly comment on these optimality conditions. Clearly, (4.7d) implies that if  $\bar{q}$  does not have level sets of strictly positive measure,  $\bar{u}$  will be a pure multi-bang control, *i.e.*,  $\bar{u}(x) \in \{u_1, \dots, u_m\}$  almost everywhere. Moreover, from (4.7e) we can deduce that  $\nabla \bar{u}(x) = 0$  for almost every  $x \in \Omega$  with  $|\bar{\psi}(x)|_2 < 1$ . Further pointwise interpretations, in particular concerning the interaction between the multi-bang and the total variation penalty, is impeded by the fact that (4.7c) couples  $\bar{q}$  not with  $\bar{\psi}$  but with  $\text{div } \bar{\psi}$ , and the divergence operator does not act pointwise and has a nontrivial kernel.

**Remark 4.5.** As already mentioned, the regularization  $\varphi_\varepsilon$  of  $\text{proj}_{[u_{\min}, u_{\max}]}$  should be chosen in such a way that it does not become stationary in  $[0, u_m]$ . For example, if we define the function  $\varphi_\varepsilon$  of (2.5) in such a manner that it acts as an interior smoothing with  $\varphi'_\varepsilon(t) = 0$  for  $t \in (-\infty, 0] \cup [u_m, \infty)$ , then  $\bar{u} \equiv 0$  with  $\bar{q} \equiv 0$ ,  $\bar{\psi} \equiv 0$  and  $\bar{y}$ ,  $\bar{p}$  computed from (4.7a) and (4.7b) always provides a trivial solution to the optimality system. It could also be observed that this obstructs numerical algorithms.

Similarly,  $\varphi'_\varepsilon(u_m) = 0$  would restrict in an undesired manner the possibility that  $\varphi_\varepsilon(u(x)) = u_{\max}$ . In fact, if  $\bar{u}(x) = u_m$  on a ball  $B$  of radius  $\rho > 0$ , then  $\alpha \bar{q}(x) = \beta \text{div } \bar{\psi}(x)$  on  $B$ , where  $\bar{q}(x) \in (\frac{1}{2}(u_{m-1} + u_m), u_m]$  for almost every  $x \in B$  and  $|\bar{\psi}(x)|_2 \leq 1$  for almost every  $x \in \Omega$ . As a consequence, we have that

$$\frac{\alpha \pi^{\frac{d}{2}} \rho^d (u_{m-1} + u_m)}{2\Gamma(\frac{d}{2} + 1)} < \alpha \int_B \bar{q} \, dx = \beta \int_B \text{div } \bar{\psi} \, dx = \beta \int_{\partial B} \bar{\psi} \cdot n \, ds \leq \frac{2\beta \pi^{\frac{d}{2}} \rho^{d-1}}{\Gamma(\frac{d}{2})},$$

where  $n$  denotes the unit outer normal to  $B$ . Thus,  $\bar{u}(x) = u_m$  cannot occur on sets that contain a ball  $B$  of radius  $\rho \geq \frac{4\beta \Gamma(\frac{d}{2} + 1)}{\alpha (u_{m-1} + u_m) \Gamma(\frac{d}{2})} = \frac{2\beta d}{\alpha (u_{m-1} + u_m)}$ . Using the same argument for a general set  $B$  to which the divergence theorem applies, we infer that  $\bar{u} = u_m$  in  $B$  necessitates  $\frac{|B|}{|\partial B|} < \frac{2\beta}{\alpha (u_{m-1} + u_m)}$ .

## 5. NUMERICAL SOLUTION

This section is concerned with the numerical computation of solutions to (P). We proceed in several steps. First, we introduce in Section 5.1 a finite element discretization of (P), for which we derive in Section 5.2 necessary optimality conditions in terms of the coefficients with respect to the finite element basis functions. These can be solved by a semismooth Newton-type method with path-following that is described in Section 5.3.

<sup>1</sup>The result in [15] is stated for  $q = \frac{p}{p-1}$  for  $1 < p \leq \frac{d}{d-1}$ . However, the upper bound on  $p$  is not used in the proofs; it is merely the natural integrability of  $u \in BV(\Omega)$  through embedding and is assumed to avoid further restrictions. We can thus apply the result for arbitrary  $q > 1$ .

### 5.1. Discretization

We consider a finite element discretization of (P). Let  $\mathcal{T} = \{\mathcal{T}_h\}_{h>0}$  be a quasi-uniform triangulation of  $\Omega$ , which we assume in the following to be polyhedral for simplicity, consisting of triangular or tetrahedral elements  $T$  with volume  $|T|$ . For later use, let us also introduce the notation  $\mathcal{T}_h = \{T_j\}_{j=1}^{N_{\mathcal{T}_h}}$  for  $h > 0$ , *i.e.*,  $\mathcal{T}_h$  consists of  $N_{\mathcal{T}_h}$  elements that are denoted by  $T_j$ ,  $1 \leq j \leq N_{\mathcal{T}_h}$ .

For the state and adjoint equation, we choose a conforming piecewise linear discretization, *i.e.*, we set

$$Y_h := \{v_h \in C_0(\Omega) : v_h|_T \in \mathcal{P}_1 \text{ for all } T \in \mathcal{T}_h\}.$$

In  $Y_h$  we use the standard nodal basis  $\{\delta_i^{Y_h}\}_{i=1}^{N_{Y_h}}$  with respect to the vertices  $x_i \in \mathbb{R}^d$ ,  $1 \leq i \leq N_{Y_h}$ . For any  $v_h \in Y_h$ , we denote by  $\hat{v}_h \in \mathbb{R}^{N_{Y_h}}$  the coefficients of  $v_h$  with respect to this basis. Defining  $[v]_j$  to be the  $j$ th component of a vector  $v$ , we can express this for  $v_h \in Y_h$  as  $v_h = \sum_{i=1}^{N_{Y_h}} [\hat{v}_h]_i \delta_i^{Y_h}$ .

The control is also discretized as continuous and piecewise linear, *i.e.*, we set

$$U_h := \left\{ u_h \in C(\bar{\Omega}) : u_h|_T \in \mathcal{P}_1 \text{ for all } T \in \mathcal{T}_h \right\}.$$

This choice – as opposed to piecewise constants – yields a convergent (nonconforming) discretization even for the isotropic total variation, see [12, 17]. Again we use the standard nodal basis, denoted by  $\{\delta_i^{U_h}\}_{i=1}^{N_{U_h}}$ , and distinguish between  $u_h \in U_h$  and its coefficient vector  $\hat{u}_h \in \mathbb{R}^{N_{U_h}}$ .

For  $w_h \in U_h$ , the discrete state equation reads

$$(w_h \nabla y_h, \nabla v_h)_{L^2(\Omega)} = (f, v_h)_{L^2(\Omega)} \quad \text{for all } v_h \in Y_h,$$

and similarly for the discrete adjoint equation. We denote the corresponding (symmetric) stiffness matrix by  $A_h(w_h) \in \mathbb{R}^{N_{Y_h} \times N_{Y_h}}$  and the mass matrix by  $M_h \in \mathbb{R}^{N_{Y_h} \times N_{Y_h}}$ .

Since the discrete gradient of  $u_h \in U_h$  should be piecewise constant, we introduce the space

$$\Psi_h := \left\{ \psi_h \in L^2(\Omega)^d : \psi_h|_T \in \mathcal{P}_0^d \text{ for all } T \in \mathcal{T}_h \right\}.$$

In  $\Psi_h$  we work with the basis of characteristic functions of  $T \in \mathcal{T}_h$ , denoted by  $\{\chi_i\}_{i=1}^{N_{\Psi_h}}$ . For the coefficients of  $\psi_h \in \Psi_h$  associated to  $T \in \mathcal{T}_h$ , we write  $[\hat{\psi}_h]_T \in \mathbb{R}^d$  and assume that  $\hat{\psi}_h \in \mathbb{R}^{N_{\Psi_h}}$  is ordered in the way  $[\hat{\psi}_h]_{T_j} = ([\hat{\psi}_h]_{(j-1)d+1}, \dots, [\hat{\psi}_h]_{jd})^T \in \mathbb{R}^d$  for  $1 \leq j \leq N_{\mathcal{T}_h}$ . This allows us to infer that  $([\hat{\psi}_h]_{T_j})_{1 \leq j \leq N_{\mathcal{T}_h}} = \hat{\psi}_h$ . Moreover, let  $D_h \in \mathbb{R}^{N_{\Psi_h} \times N_{U_h}}$  denote the stiffness matrix arising from the bilinear form

$$(\nabla u_h, \psi_h)_{L^2(\Omega)} \quad \text{for all } (u_h, \psi_h) \in U_h \times \Psi_h.$$

We mention that  $-D_h^T \in \mathbb{R}^{N_{U_h} \times N_{\Psi_h}}$  corresponds to the discrete divergence. In the following, we assume that  $D_h$  is ordered in the way  $[D_h \hat{u}_h]_{T_j} = (D_{(j-1)d+1} \hat{u}_h, \dots, D_{jd} \hat{u}_h)^T \in \mathbb{R}^d$ , where  $D_i$  denotes for  $1 \leq i \leq N_{\Psi_h}$  the  $i$ th row of  $D_h$ . This allows us to infer that the Fréchet derivative of the mapping  $\hat{u}_h \mapsto ([D_h \hat{u}_h]_{T_j})_j \in \mathbb{R}^{N_{\Psi_h}}$ ,  $1 \leq j \leq N_{\mathcal{T}_h}$ , is given by  $D_h$ .

The multi-bang penalty is approximated *via* mass lumping, *i.e.*, we take

$$G_h(\hat{u}_h) := \sum_{i=1}^{N_{U_h}} d_i g([\hat{u}_h]_i),$$

where  $g : \mathbb{R} \rightarrow \mathbb{R}$  is given by (2.2) and  $d_i := \int_{\Omega} \delta_i^{U_h}(x) dx$ , see [18, 35, 41]. For later use, we also introduce the diagonal matrix  $M_h^\ell \in \mathbb{R}^{N_{U_h} \times N_{U_h}}$  with entries  $d_i$ , which corresponds to a lumped mass matrix in  $U_h$ . Similarly, the total variation is approximated by

$$\text{TV}_h(\hat{u}_h) := \sum_{T \in \mathcal{T}_h} |[D_h \hat{u}_h]_T|_2.$$

This is a correctly weighted discretization of the total variation since for all  $u_h \in U_h$  there holds

$$\text{TV}(u_h) = \sum_{T \in \mathcal{T}_h} |T| |\nabla u_h|_T|_2 = \sum_{i=1}^{N_{U_h}} |\nabla u_h, \chi_i|_2 = \sum_{T \in \mathcal{T}_h} |[D_h \hat{u}_h]_T|_2 = \text{TV}_h(\hat{u}_h).$$

Note that by these definitions,  $G_h$  and  $\text{TV}_h$  are defined on  $\mathbb{R}^{N_{U_h}}$ , allowing us to apply convex analysis in the standard Euclidean topology.

The discrete problem now reads

$$\begin{cases} \min_{\hat{u}_h \in \mathbb{R}^{N_{U_h}}} \frac{1}{2} \|y_h - z_h\|_{L^2}^2 + \alpha G_h(\hat{u}_h) + \beta \text{TV}_h(\hat{u}_h), \\ \text{s.t. } A_h(\Phi_h(u_h)) \hat{y}_h = M_h \hat{f}_h, \end{cases} \quad (5.1)$$

where  $z_h$  is the  $L^2(\Omega)$  projection of  $z$  onto  $Y_h$  and thus  $\frac{1}{2} \|y_h - z_h\|_{L^2}^2 = \frac{1}{2} (\hat{y}_h - \hat{z}_h)^T M_h (\hat{y}_h - \hat{z}_h)$ . Similarly,  $f_h$  denotes the  $L^2(\Omega)$  projection (or interpolation) of  $f$  onto  $Y_h$ . The existence of a solution  $\hat{u}_h^* \in \mathbb{R}^{N_{U_h}}$  to (5.1) then follows from standard arguments.

## 5.2. Discrete optimality system and regularization

We now derive numerically tractable optimality conditions for the discretized problem (5.1), exploiting the fact that functional-analytic difficulties that had to be circumvented to obtain (4.7) do not arise in the finite-dimensional setting. Specifically,

- (i) we can consider  $\varepsilon = 0$  or equivalently, by Proposition 3.2, the discrete analogue of (P\*), thus eliminating the need for  $\Phi_\varepsilon$ ;
- (ii) as in [20, 21], we can include the pointwise constraints in the definition of the multi-bang penalty  $G$ ;
- (iii) applying the chain rule to the convex subdifferential of the discrete total variation directly yields an explicit componentwise relation.

Hence, we replace (5.1) by

$$\begin{cases} \min_{\hat{u}_h \in \mathbb{R}^{N_{U_h}}} \frac{1}{2} \|y_h - z_h\|_{L^2}^2 + \alpha \hat{G}_h(\hat{u}_h) + \beta \text{TV}_h(\hat{u}_h), \\ \text{s.t. } A_h(u_h + u_{\min}) \hat{y}_h = M_h \hat{f}_h, \end{cases} \quad (\text{P}_h)$$

for

$$\hat{G}_h(\hat{u}_h) := \sum_{i=1}^{N_{U_h}} d_i \hat{g}([\hat{u}_h]_i), \quad \hat{g}(t) = \begin{cases} \infty & t < u_1, \\ \frac{1}{2} ((u_i + u_{i+1})t - u_i u_{i+1}) & t \in [u_i, u_{i+1}], \quad 1 \leq i < m, \\ \infty & t > u_m. \end{cases}$$



Proceeding as in the continuous case, we see that (4.7a) and (4.7b) are replaced by their finite element approximation. Introducing for  $y_h, p_h \in Y_h$  the vector

$$\hat{a}_h(y_h, p_h) := \nabla y_h \cdot \nabla p_h \in \mathbb{R}^{N_{U_h}},$$

we obtain analogously to (4.6) the primal-dual optimality conditions

$$\begin{cases} A_h(u_h^* + u_{\min})\hat{y}_h^* = M_h \hat{f}_h, \\ A_h(u_h^* + u_{\min})\hat{p}_h^* = M_h(\hat{z}_h - \hat{y}_h^*), \\ 0 = \hat{a}_h(y_h^*, p_h^*) + \alpha \hat{q}_h^* + \beta \hat{\xi}_h^*, \\ \hat{q}_h^* \in \partial \hat{G}_h(u_h^*), \\ \hat{\xi}_h^* \in \partial \text{TV}_h(u_h^*). \end{cases} \quad (5.2)$$

Let us remark that it is straightforward to derive a version of (5.2) in  $Y_h \times Y_h \times U_h \times U_h \times \Psi_h$  instead of  $\mathbb{R}^{N_{Y_h}} \times \mathbb{R}^{N_{Y_h}} \times \mathbb{R}^{N_{U_h}} \times \mathbb{R}^{N_{U_h}} \times \mathbb{R}^{N_{\Psi_h}}$ . It can then be observed that this version is exactly (4.6) but with  $(\bar{y}, \bar{p}, \bar{u}, \bar{q}, \bar{\xi}) \in Y \times Y \times U \times U \times \Psi$  replaced by their finite-dimensional counterparts  $(y_h^*, p_h^*, u_h^*, q_h^*, \xi_h^*) \in Y_h \times Y_h \times U_h \times U_h \times \Psi_h$ , and that (5.2) is its equivalent reformulation in  $\mathbb{R}^{N_{Y_h}} \times \mathbb{R}^{N_{Y_h}} \times \mathbb{R}^{N_{U_h}} \times \mathbb{R}^{N_{U_h}} \times \mathbb{R}^{N_{\Psi_h}}$ . In particular, the two approaches of *first discretize, then optimize* and *first optimize, then discretize* coincide.

The next step is to characterize these subgradients componentwise. For the first subdifferential, we can simply use the sum and chain rules and find that

$$[\hat{q}_h^*]_j \in d_j \partial \hat{g}([\hat{u}_h^*]_j), \quad 1 \leq j \leq N_{U_h},$$

or equivalently

$$[\hat{u}_h^*]_j \in \partial \hat{g}^*(d_j^{-1}[\hat{q}_h^*]_j), \quad 1 \leq j \leq N_{U_h},$$

with  $\partial \hat{g}^*$  given analogously to  $\partial g^*$  as

$$\partial \hat{g}^*(s) = \begin{cases} \{u_1\} & s \in (-\infty, \frac{1}{2}(u_1 + u_2)), \\ [u_i, u_{i+1}] & s = \frac{1}{2}(u_i + u_{i+1}), \quad 1 \leq i < m, \\ \{u_i\} & s \in (\frac{1}{2}(u_{i-1} + u_i), \frac{1}{2}(u_i + u_{i+1})), \quad 1 < i < m, \\ \{u_m\} & s \in (\frac{1}{2}(u_{m-1} + u_m), \infty), \end{cases}$$

see also ([20], Sect. 2.1). We will in the following replace the components  $[\hat{q}_h^*]_i$  of  $\hat{q}_h^*$  by their scaling  $d_i^{-1}[\hat{q}_h^*]_i$ ; using the definition of the lumped mass matrix, this means we have to replace  $\hat{q}_h^*$  in the third equation of (5.2) by  $M_h^\ell \hat{q}_h^*$ .

For the discrete total variation, we use the sum rule and the chain rule to deduce that there exists  $\hat{\psi}_h^* \in \mathbb{R}^{N_{\Psi_h}}$  such that

$$\hat{\xi}_h^* = D_h^T \hat{\psi}_h^* \quad \text{and} \quad [\hat{\psi}_h^*]_T \in \partial(|\cdot|_2)([D_h \hat{u}_h^*]_T) \quad \text{for all } T \in \mathcal{T}_h,$$

are satisfied. As before, we rewrite the subdifferential inclusion equivalently as

$$[D_h \hat{u}_h^*]_T \in \partial(|\cdot|_2^*)([\hat{\psi}_h^*]_T) \quad \text{for all } T \in \mathcal{T}_h.$$

Using

$$\hat{h} : \mathbb{R}^d \rightarrow \mathbb{R}, \quad \hat{h}(v) := |v|_2,$$

this reads

$$[D_h \hat{u}_h^*]_T \in \partial \hat{h}^*([\hat{\psi}_h^*]_T) \quad \text{for all } T \in \mathcal{T}_h.$$

To apply a Newton-type method, we replace the set-valued subdifferentials by their single-valued and Lipschitz-continuous Moreau–Yosida regularizations. Recall that the Moreau–Yosida regularization of  $\partial F$  for any proper, convex and lower semi-continuous functional  $F : X \rightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$  acting on a Hilbert space  $X$  is given by

$$(\partial F)_\gamma(v) = \frac{1}{\gamma}(v - \text{prox}_{\gamma F}(v)),$$

where  $\gamma > 0$  and

$$\text{prox}_{\gamma F}(v) := \arg \min_{w \in X} \frac{1}{2\gamma} \|w - v\|_X^2 + F(w) = (\text{Id} + \gamma \partial F)^{-1}(v).$$

For the regularized subdifferential  $(\partial \hat{g}^*)_\gamma$ , we have from Section 4.1 from [21] that for  $s \in \mathbb{R}$

$$(\partial \hat{g}^*)_\gamma(s) = \begin{cases} u_1 & s \in (-\infty, (\gamma + \frac{1}{2})u_1 + \frac{1}{2}u_2), \\ \frac{1}{\gamma}(s - \frac{u_i + u_{i+1}}{2}) & s \in [(\gamma + \frac{1}{2})u_i + \frac{1}{2}u_{i+1}, \frac{1}{2}u_i + (\gamma + \frac{1}{2})u_{i+1}], \quad 1 \leq i < m, \\ u_i & s \in (\frac{1}{2}u_{i-1} + (\gamma + \frac{1}{2})u_i, (\gamma + \frac{1}{2})u_i + u_{i+1}), \quad 1 < i < m, \\ u_m & s \in (\frac{1}{2}u_{m-1} + (\gamma + \frac{1}{2})u_m, \infty). \end{cases}$$

For  $\delta > 0$ , we denote the Moreau–Yosida regularization of  $\partial \hat{h}^*$  by  $(\partial \hat{h}^*)_\delta$ . To compute it, we recall that the Fenchel conjugate of a norm is the indicator function of the unit ball corresponding to the dual norm (which in this case is  $|\cdot|_2$  itself). Furthermore, the proximal mapping  $\text{prox}_{\delta F}$  of an indicator function to a convex set is for every  $\delta > 0$  the metric projection onto this set. This shows that for all  $v \in \mathbb{R}^d$  there holds

$$(\partial \hat{h}^*)_\delta(v) = \frac{1}{\delta}(v - \text{proj}_{\{|v|_2 \leq 1\}}(v)) = \begin{cases} 0 & |v|_2 \leq 1, \\ \frac{1}{\delta}(v - \frac{v}{|v|_2}) & |v|_2 > 1. \end{cases}$$

Combining the above, we obtain the regularized discrete optimality conditions

$$\begin{cases} A_h(u_h^* + u_{\min})\hat{y}_h^* = M_h \hat{f}_h, \\ A_h(u_h^* + u_{\min})\hat{p}_h^* = M_h(\hat{z}_h - \hat{y}_h^*), \\ 0 = \hat{a}_h(y_h^*, p_h^*) + \alpha M_h^\ell \hat{q}_h^* + \beta D_h^T \hat{\psi}_h^*, \\ [\hat{u}_h^*]_j = (\partial \hat{g}^*)_\gamma([\hat{q}_h^*]_j), \quad 1 \leq j \leq N_{U_h}, \\ [D_h \hat{u}_h^*]_T = (\partial \hat{h}^*)_\delta([\hat{\psi}_h^*]_T), \quad T \in \mathcal{T}_h. \end{cases} \quad (5.3)$$

Note that we have used the same notation  $\hat{y}_h^*$ ,  $\hat{u}_h^*$ , etc., as for solutions to the *unregularized* discrete optimality conditions (5.2) to avoid further complicating the notation. We point out that for the remainder of this work, this notation will always refer to solutions to (5.3).

Finally, we remark that since  $(\partial F^*)_Y = \nabla(F^*)_Y$  with  $((F^*)_Y)^* = F + \frac{\gamma}{2}\|\cdot\|_X^2$  holds for any proper, convex, and lower semi-continuous functional  $F : X \rightarrow \overline{\mathbb{R}}$ , the regularized optimality system coincides with the necessary optimality conditions of

$$\begin{cases} \min_{\hat{u}_h \in \mathbb{R}^{N_{U_h}}} \frac{1}{2}\|y_h - z_h\|_{L^2}^2 + \alpha \left( \hat{G}_h(\hat{u}_h) + \frac{\gamma}{2}\|\hat{u}_h\|_{M_h^\ell}^2 \right) + \beta \left( \text{TV}_h(\hat{u}_h) + \frac{\delta}{2}\|\hat{u}_h\|_{2,h}^2 \right), \\ \text{s.t. } A_h(u_h + u_{\min})\hat{y}_h = M_h\hat{f}_h, \end{cases}$$

where  $\|\hat{u}_h\|_{M_h^\ell} := (\hat{u}_h^T M_h^\ell \hat{u}_h)^{1/2}$  and  $\|\hat{u}_h\|_{2,h} := (\sum_{T \in \mathcal{T}_h} \|D_h \hat{u}_h|_T\|_2^2)^{1/2}$ . This can be interpreted as the mass-lumped approximation of an  $H^1$  regularization of (P). Note, however, that the problem is still nonsmooth since  $G_h$  and  $\text{TV}_h$  have not been modified; it has merely been made more strongly convex.

### 5.3. A semismooth Newton-type method

To apply a semismooth Newton method to the regularized optimality conditions (5.3), we reformulate them as a set of nonlinear implicit equations. Based on our numerical experience, it is preferable to consider the reduced system arising from (5.3) by eliminating the variables  $(\hat{u}_h, \hat{q}_h)$  rather than solving the full system (5.3) in the variables  $(\hat{y}_h, \hat{p}_h, \hat{u}_h, \hat{q}_h, \hat{\psi}_h)$ . In the following, we abbreviate  $\hat{\zeta}_h := (\hat{y}_h, \hat{p}_h, \hat{\psi}_h) \in \mathbb{R}^{N_{\hat{\zeta}_h}}$ , where  $N_{\hat{\zeta}_h} := 2N_{Y_h} + N_{\Psi_h}$ .

We begin the reformulation by noting that the third equation in (5.3) is equivalent to

$$\hat{q}_h^* = -\frac{1}{\alpha} M_h^{-\ell} \left( B_h(y_h)\hat{p}_h + \beta D_h^T \hat{\psi}_h^* \right), \quad (5.4)$$

where  $M_h^{-\ell}$  denotes the inverse of  $M_h^\ell$  and  $B_h(y_h) \in \mathbb{R}^{N_{U_h} \times N_{Y_h}}$  denotes the matrix induced by the bilinear form

$$((\nabla y_h \cdot \nabla v_h), w_h)_{L^2(\Omega)} \quad \text{for all } (w_h, v_h) \in U_h \times Y_h.$$

Defining

$$\hat{q}_h : \mathbb{R}^{N_{\hat{\zeta}_h}} \rightarrow \mathbb{R}^{N_{U_h}}, \quad \hat{q}_h(\hat{\zeta}_h) := -\frac{1}{\alpha} M_h^{-\ell} \left( B_h(y_h)\hat{p}_h + \beta D_h^T \hat{\psi}_h \right),$$

(5.4) becomes

$$\hat{q}_h^* = \hat{q}_h(\hat{\zeta}_h^*).$$

Inserting this into the fourth equation of (5.3) enables us to express  $\hat{u}_h^*$  by

$$\hat{u}_h^* = \hat{u}_h(\hat{\zeta}_h^*),$$

where

$$\hat{u}_h : \mathbb{R}^{N_{\hat{\zeta}_h}} \rightarrow \mathbb{R}^{N_{U_h}}, \quad \hat{u}_h(\hat{\zeta}_h) := \begin{pmatrix} (\partial \hat{g}^*)_Y([\hat{q}_h(\hat{\zeta}_h)]_1) \\ (\partial \hat{g}^*)_Y([\hat{q}_h(\hat{\zeta}_h)]_2) \\ \vdots \\ (\partial \hat{g}^*)_Y([\hat{q}_h(\hat{\zeta}_h)]_{N_{U_h}}) \end{pmatrix}.$$

We write  $u_h(\hat{\zeta}_h)$  for the function  $u_h \in U_h$  with coefficients  $\hat{u}_h(\hat{\zeta}_h)$ , *i.e.*,  $u_h(\hat{\zeta}_h) := \sum_{i=1}^{N_{U_h}} [\hat{u}_h(\hat{\zeta}_h)]_i \delta_i^{U_h}$ . Summarizing, (5.3) is equivalent to  $\mathcal{F}_{\gamma, \delta}(\hat{\zeta}_h^*) = 0$  for

$$\mathcal{F}_{\gamma, \delta} : \mathbb{R}^{N_{\hat{\zeta}_h}} \rightarrow \mathbb{R}^{N_{\hat{\zeta}_h}}, \quad \mathcal{F}_{\gamma, \delta}(\hat{\zeta}_h) := \begin{pmatrix} A_h(u_h(\hat{\zeta}_h) + u_{\min})\hat{p}_h + M_h(\hat{y}_h - \hat{z}_h) \\ A_h(u_h(\hat{\zeta}_h) + u_{\min})\hat{y}_h - M_h\hat{f}_h \\ \mathcal{H}(\hat{\zeta}_h) \end{pmatrix}, \quad (5.5)$$

where  $\mathcal{H} : \mathbb{R}^{N_{\hat{\zeta}_h}} \rightarrow \mathbb{R}^{N_{\Psi_h}}$ ,  $\mathcal{H} = (\mathcal{H}_1^T, \mathcal{H}_2^T, \dots, \mathcal{H}_{N_{\mathcal{T}_h}}^T)^T$  with

$$\mathcal{H}_j : \mathbb{R}^{N_{\hat{\zeta}_h}} \rightarrow \mathbb{R}^d, \quad \mathcal{H}_j(\hat{\zeta}_h) := [D_h \hat{u}_h(\hat{\zeta}_h)]_{T_j} - (\partial \hat{h}^*)_{\delta}([\hat{\psi}_h]_{T_j}) \quad \text{for } 1 \leq j \leq N_{\mathcal{T}_h}.$$

We recall that  $\mathcal{T}_h = \{T_j\}_{j=1}^{N_{\mathcal{T}_h}}$  and point out that  $N_{\Psi_h} = N_{\mathcal{T}_h} d$ .

Since all components of  $\mathcal{F}_{\gamma, \delta}$  are either continuously differentiable or continuous and piecewise continuously differentiable (PC<sup>1</sup>) in each variable,  $\mathcal{F}_{\gamma, \delta}$  is semismooth, see, *e.g.*, [27, 28, 32, 43]. To obtain Newton derivatives for the nonsmooth terms, we use the fact that for PC<sup>1</sup> functions we can take as Newton derivative any selection of the derivatives of the essentially active pieces; see ([43], Sect. 2.5.3). In the following, we denote Newton derivatives by  $D_N$ . For the partial Newton derivative of, say,  $\hat{u}_h(\cdot)$  with respect to the variable  $\hat{\psi}_h$  evaluated at  $\hat{\zeta}_h$ , we write  $D_{N_{\hat{\psi}}} \hat{u}_h(\hat{\zeta}_h)$ . Since the mapping  $\hat{u}_h(\cdot)$  is a composition of smooth mappings with  $(\partial \hat{g}^*)_{\gamma}$ , its Newton derivative is given by the chain rule in combination with our specific choice of

$$D_N(\partial \hat{g}^*)_{\gamma}(s) = \begin{cases} \frac{1}{\gamma} & s \in [(\gamma + \frac{1}{2})u_i + \frac{1}{2}u_{i+1}, \frac{1}{2}u_i + (\gamma + \frac{1}{2})u_{i+1}], \quad 1 \leq i < m, \\ 0 & \text{else.} \end{cases}$$

To determine  $D_N \mathcal{H}$ , it suffices to specify  $D_N(\partial \hat{h}^*)_{\delta}$ , where we make the choice

$$D_N(\partial \hat{h}^*)_{\delta}(v) = \begin{cases} 0 & |v|_2 \leq 1, \\ \frac{1}{\delta} \left( \text{Id} - \frac{1}{|v|_2} \text{Id} + \frac{1}{|v|_2^3} v v^T \right) & |v|_2 > 1. \end{cases}$$

Together, we obtain

$$D_N \mathcal{F}_{\gamma, \delta}(\hat{\zeta}_h) = \begin{pmatrix} C_p E_y + M_h & C_p E_p + C_{y/p} & C_p E_{\psi} \\ C_y E_y + C_{y/p} & C_y E_p & C_y E_{\psi} \\ D_h E_y & D_h E_p & D_h E_{\psi} - E_{\psi \psi} \end{pmatrix} \in \mathbb{R}^{N_{\hat{\zeta}_h} \times N_{\hat{\zeta}_h}},$$

where

$$\begin{aligned} C_p &:= B_h(p_h)^T, & C_y &:= B_h(y_h)^T, & C_{y/p} &:= A_h(u_h(\hat{\zeta}_h) + u_{\min}), \\ E_y &:= D_{N_y} \hat{u}_h(\hat{\zeta}_h), & E_p &:= D_{N_p} \hat{u}_h(\hat{\zeta}_h), & E_{\psi} &:= D_{N_{\hat{\psi}}} \hat{u}_h(\hat{\zeta}_h), \end{aligned}$$

and

$$E_{\psi \psi} := \begin{pmatrix} D_N(\partial \hat{h}^*)_{\delta}([\hat{\psi}_h]_{T_1}) & & & \\ & D_N(\partial \hat{h}^*)_{\delta}([\hat{\psi}_h]_{T_2}) & & \\ & & \ddots & \\ & & & D_N(\partial \hat{h}^*)_{\delta}([\hat{\psi}_h]_{T_{N_{\mathcal{T}_h}}}) \end{pmatrix} \in \mathbb{R}^{N_{\Psi_h} \times N_{\Psi_h}}.$$

Note that the Newton matrix can become singular. For instance, if  $\|[\hat{\psi}_h]_T\|_2 \leq 1$  for all  $T \in \mathcal{T}_h$ , then  $E_{\psi\psi} = 0$ . Hence,  $(0, 0, \hat{w}_h)^T \in \ker(D_N \mathcal{F}_{\gamma, \delta}(\hat{\zeta}_h))$  for every  $\hat{w}_h \in \ker(E_{\psi})$ . Clearly,  $\ker(E_{\psi})$  is nontrivial since this is true for  $\ker(D_h^T)$ . To cope with this singularity, we modify the (3,3) block of  $D_N \mathcal{F}_{\gamma, \delta}$  so that it reads  $D_h E_{\psi} - E_{\psi\psi} - \mu_{\gamma, \delta} \hat{M}_h$ , where  $\hat{M}_h \in \mathbb{R}^{N_{\Psi_h} \times N_{\Psi_h}}$  denotes the diagonal mass matrix in  $\Psi_h$ , and  $\mu_{\gamma, \delta} > 0$  is a weight that depends on  $\gamma$  and  $\delta$ ; in our numerical experiments we observed  $\mu_{\gamma, \delta} := \delta^{-1}$  to work well. In the following, we assume that this choice is made unless explicitly indicated otherwise. We denote this modified matrix by  $\widetilde{D_N \mathcal{F}_{\gamma, \delta}}$ . For later reference we notice that given  $(\gamma_j, \delta_j) \in \mathbb{R}_{>0} \times \mathbb{R}_{>0}$ , a semismooth Newton-type step  $\tilde{s}^j \in \mathbb{R}^{N_{\zeta_h}^{\tilde{\zeta}^j}}$  at  $\tilde{\zeta}^j \in \mathbb{R}^{N_{\zeta_h}^{\tilde{\zeta}^j}}$  is characterized by

$$\widetilde{D_N \mathcal{F}_{\gamma_j, \delta_j}}(\tilde{\zeta}^j) \tilde{s}^j = -\mathcal{F}_{\gamma_j, \delta_j}(\tilde{\zeta}^j). \quad (5.6)$$

This step is combined with a backtracking line search based on the residual norm as well as a path-following scheme for  $(\gamma_j, \delta_j)$ . The full procedure to compute an approximate solution to (P<sub>h</sub>) is given in Algorithm TVMB, where we have dropped the index  $h$  for better readability. We also write  $\|\hat{\zeta}\|_{L^2} := \|\zeta_h\|_{L^2(\Omega)^{d+2}}$  for  $\hat{\zeta} \in \mathbb{R}^{N_{\zeta}^{\hat{\zeta}}}$ , where  $\hat{\zeta}$  are the coefficients of the function  $\zeta_h \in Y_h \times Y_h \times \Psi_h$ , and  $\|(\hat{\zeta}, \hat{u}, \hat{q})\|_{L^2} := \|(\zeta_h, u_h, q_h)\|_{L^2(\Omega)^{d+4}}$ , where  $(\hat{\zeta}, \hat{u}, \hat{q})$  are the coefficients of  $(\zeta_h, u_h, q_h) \in Y_h \times Y_h \times \Psi_h \times U_h \times U_h$ .

Algorithm TVMB is structured as follows. Lines 4 to 14 constitute an inner iteration; in this inner iteration, a Newton-type method with line search is employed for fixed  $\gamma$  and  $\delta$  to find a root of  $\mathcal{F}_{\gamma, \delta}$ . The remaining lines form an outer iteration; in this outer iteration,  $\gamma$  and  $\delta$  are updated and the starting point for the next inner iteration is computed in line 19 or 21, respectively. Moreover, the  $L^2$  difference of subsequent outer iterates is stored in  $r_k$  and used in the termination criterion.

Let us comment on some important features of Algorithm TVMB. We start by pointing out that the line search in lines 7–12 of Algorithm TVMB is nonmonotone. That is, if backtracking does not yield a  $\sigma_j \in [\sigma_{\min}, 1]$  with  $\|\mathcal{F}_{\gamma_j, \delta_j}(\tilde{\zeta}^j + \sigma_j \tilde{s}^j)\|_{L^2} < \|\mathcal{F}_{\gamma_j, \delta_j}(\tilde{\zeta}^j)\|_{L^2}$ , then the step length  $\sigma_j = \sigma_{\text{nm}}$  is used regardless whether it satisfies  $\|\mathcal{F}_{\gamma_j, \delta_j}(\tilde{\zeta}^j + \sigma_{\text{nm}} \tilde{s}^j)\|_{L^2} < \|\mathcal{F}_{\gamma_j, \delta_j}(\tilde{\zeta}^j)\|_{L^2}$  or not.

Next we remark that the computation of  $\hat{\zeta}_{k+1}$  in line 19 is a predictor step: From the previous roots  $\hat{\zeta}_{\text{opt}}^k$  and  $\hat{\zeta}_{\text{opt}}^{k-1}$ , a prediction  $\hat{\zeta}_{\text{opt}}^{k+1}$  of  $\hat{\zeta}_{\text{opt}}^{k+1}$  is computed and used as the starting point for the next inner iteration (whose aim it is to find  $\hat{\zeta}_{\text{opt}}^{k+1}$ ). For  $k \geq 1$ , this prediction is taken to be the componentwise linear extrapolation

$$\hat{\zeta}_{\text{opt}}^{k+1} := \hat{\zeta}_{\text{opt}}^k + \frac{\gamma_k - \gamma_{k+1}}{\gamma_{k-1} - \gamma_k} \left( \hat{\zeta}_{\text{opt}}^k - \hat{\zeta}_{\text{opt}}^{k-1} \right) = (1 + \nu) \hat{\zeta}_{\text{opt}}^k - \nu \hat{\zeta}_{\text{opt}}^{k-1},$$

where we have used that  $\gamma_{k+1} = \nu \gamma_k = \nu^2 \gamma_{k-1}$ . Note that due to the coupling  $\delta_k / \gamma_k = \delta_0 / \gamma_0$  for all  $k$ , we obtain the same extrapolation step if  $\gamma$  is replaced by  $\delta$ . We thus perform a combined prediction for the continuation in  $\gamma$  as well as  $\delta$ . For  $k = 0$ , no predictor step is used as  $\hat{\zeta}_{\text{opt}}^{k-1} = \hat{\zeta}_{\text{opt}}^{-1}$  is not available; instead we set  $\hat{\zeta}^1 = \hat{\zeta}_{\text{opt}}^0$  in this case.

Finally, we embed Algorithm TVMB within a further continuation strategy for  $\nu$ : if a Newton iteration for a given pair  $(\gamma_k, \delta_k)$  does not terminate successfully, we increase  $\nu$  and restart Algorithm TVMB from the last successful solution; this outer continuation is terminated if  $\nu \approx 1$ .

We conclude this section with several practical remarks concerning Algorithm TVMB. First, we stress that while its numerical costs are negligible, the predictor step significantly increased the convergence speed in our numerical experiments. Also, due to the path-following strategy, it is not necessary to choose the initial guess  $\hat{\zeta}^0$  in a specific way. In fact, our numerical experiments indicate that arbitrary starting points can be used. In particular, the choice  $\hat{\zeta}^0 := 0$  was always sufficient to achieve convergence.

Furthermore, we found in our numerical experiments that for larger values of  $\gamma$  and  $\delta$  (e.g.,  $\gamma, \delta > 1$ ), the convergence of Algorithm TVMB can be accelerated if  $\mu_{\gamma, \delta} = \delta$  is used and  $\mathcal{F}_{\gamma, \delta}$  is modified such that its Newton derivative equals  $\widetilde{D_N \mathcal{F}_{\gamma, \delta}}$ . For small values of  $\gamma$  and  $\delta$ , however, this strategy did not work and we had to choose

**Algorithm TVMB:** Path-following method to solve  $(P_h)$ 


---

**Input:**  $\hat{\zeta}^0 \in \mathbb{R}^{N_\xi}$ ,  $\gamma_0 > 0$ ,  $\delta_0 > 0$ ,  $\nu \in (0, 1)$ ,  $TOL_r > 0$ ,  $TOL_{\mathcal{F}} > 0$ ,  
 $\sigma_{\min} \in (0, 1]$ ,  $\sigma_{\text{nm}} \in (0, 1]$

- 1 Set  $k = 0$  and  $r_{-1} = TOL_r + 1$
- 2 **repeat**
- 3   Set  $j = 0$  and  $\tilde{\zeta}^0 = \hat{\zeta}^k$
- 4   **while**  $\|\mathcal{F}_{\gamma_j, \delta_j}(\tilde{\zeta}^j)\|_{L^2} > TOL_{\mathcal{F}}$  **do**
- 5     Set  $\tilde{q}^j = \hat{q}(\tilde{\zeta}^j)$  and  $\tilde{u}^j = \hat{u}(\tilde{\zeta}^j)$
- 6     Compute the Newton-type step  $\tilde{s}^j$  at  $\tilde{\zeta}^j$  by solving (5.6) and set  $\sigma_j = 1$
- 7     **while**  $[\sigma_j \geq \sigma_{\min}$  and  $\|\mathcal{F}_{\gamma_j, \delta_j}(\tilde{\zeta}^j + \sigma_j \tilde{s}^j)\|_{L^2} \geq \|\mathcal{F}_{\gamma_j, \delta_j}(\tilde{\zeta}^j)\|_{L^2}]$  **do**
- 8       Set  $\sigma_j = \sigma_j/2$
- 9     **end**
- 10    **if**  $\sigma_j < \sigma_{\min}$  **then**
- 11     Set  $\sigma_j = \sigma_{\text{nm}}$
- 12    **end**
- 13    Set  $\tilde{\zeta}^{j+1} = \tilde{\zeta}^j + \sigma_j \tilde{s}^j$  and  $j = j + 1$
- 14   **end**
- 15   Set  $\hat{\zeta}_{\text{opt}}^k = \tilde{\zeta}^j$ ,  $\hat{u}_{\text{opt}}^k = \hat{u}(\hat{\zeta}_{\text{opt}}^k)$  and  $\hat{q}_{\text{opt}}^k = \hat{q}(\hat{\zeta}_{\text{opt}}^k)$
- 16   Set  $\gamma_{k+1} = \nu \gamma_k$  and  $\delta_{k+1} = \nu \delta_k$
- 17   **if**  $k \geq 1$  **then**
- 18     Set  $r_k = \|(\hat{\zeta}_{\text{opt}}^k, \hat{u}_{\text{opt}}^k, \hat{q}_{\text{opt}}^k) - (\hat{\zeta}_{\text{opt}}^{k-1}, \hat{u}_{\text{opt}}^{k-1}, \hat{q}_{\text{opt}}^{k-1})\|_{L^2}$
- 19     Set  $\hat{\zeta}^{k+1} = (1 + \nu)\hat{\zeta}_{\text{opt}}^k - \nu\hat{\zeta}_{\text{opt}}^{k-1}$
- 20   **else**
- 21     Set  $r_k = TOL_r + 1$  and  $\hat{\zeta}^{k+1} = \hat{\zeta}_{\text{opt}}^k$
- 22   **end**
- 23   Set  $k = k + 1$
- 24 **until**  $[r_{k-1} \leq TOL_r$  and  $r_{k-2} \leq TOL_r]$ ;

**Output:**  $\hat{\zeta}_{\text{opt}}^{k-1} \in \mathbb{R}^{N_\xi}$

---

$\mathcal{F}_{\gamma, \delta}$  as given in (5.5) and  $\mu_{\gamma, \delta} = \delta^{-1}$ . Note that for the choice  $\mu_{\gamma, \delta} = \delta^{-1}$  it is not sensible to modify  $\mathcal{F}_{\gamma, \delta}$  in such a way that its Newton derivative equals  $\overline{D_N \mathcal{F}_{\gamma, \delta}}$ . In fact, we can show that if  $\mathcal{F}_{\gamma, \delta}$  is modified in this way, then the sequence  $((\hat{\zeta}_{\text{opt}}^k, \hat{u}(\hat{\zeta}_{\text{opt}}^k), \hat{q}(\hat{\zeta}_{\text{opt}}^k)))_k$  can only converge to a solution to (5.2) with  $\beta = 0$ , *i.e.*, to a solution to the optimality conditions of the “pure multi-bang problem”.

## 6. NUMERICAL EXAMPLES

We illustrate the structure of optimal controls for  $(P_h)$  using two model problems. In particular, the goal is to show the difference between optimal controls of  $(P_h)$  for  $\beta > 0$  and for  $\beta = 0$ , *i.e.*, between solutions to a TV-regularized multi-bang problem and those to a “pure multi-bang” problem. We remark that  $\beta > 0$  is required in the infinite dimensional case but can be arbitrarily small, while taking  $\beta = 0$  is justified in the finite-dimensional setting only. More examples for the pure multi-bang approach can be found in [20, 21].

In all examples, we take  $\Omega = (-1, 1)^2 \subset \mathbb{R}^2$  and employ a uniform triangulation  $\mathcal{T}_h$  consisting of 8192 elements, *i.e.*,  $N_{U_h} = 64 \times 64$ . We use  $u_{\min} = 1.5$  and the algorithmic parameters  $\hat{\zeta}^0 = 0$ ,  $\gamma_0 = 10^5$ ,  $\delta_0 = 10^3$ ,  $\nu = 0.8$ ,  $\nu_{\max} = 0.9999$ ,  $TOL_r = 10^{-3}(u_{\max} - u_{\min})$ ,  $TOL_{\mathcal{F}} = 10^{-5}$ , as well as  $\sigma_{\min} = 10^{-6}$  and  $\sigma_{\text{nm}} = 10^{-2}$ . The remaining data and parameters are chosen individually for each example.

We implemented Algorithm [TVMB](#) in Python using DOLFIN [\[29, 31\]](#), which is part of the open-source computing platform FEniCS [\[2, 30\]](#). The linear system (5.6) arising from the Newton-type step is solved using the sparse direct solver `spsolve` from SciPy.

### 6.1. Example 1: Topology optimization

The first example is motivated by the possible application to topology optimization. The general idea is that we have a design  $\tilde{u} \in U$  making use of two materials characterized by their densities  $u_{\min} + \tilde{u}_1 = 1.5$  and  $u_{\min} + \tilde{u}_2 = 2.5$ ; we call this a *binary design*. Imagine that it has become possible to use also materials that have intermediate densities, *e.g.*, in total five materials with densities  $u_{\min} + u_j = 1.5 + 0.25(j - 1)$ ,  $1 \leq j \leq 5$ . The question is now whether it is possible to realize a similar state as arising from the (presumably optimal) binary design using the (presumably cheaper) intermediate materials.

Following this motivation, we start from the binary design

$$\tilde{u}(x) := \begin{cases} 1.5 & x \in \omega_1, \\ 2.5 & x \in \omega_2, \end{cases}$$

where

$$\omega_2 := \{x \in \Omega : 0.1 < |x_1| < 0.8 \text{ and } |x_2| < 0.8 \text{ and } [ |x_1| > 0.5 \text{ or } |x_2| > 0.5 ]\},$$

and  $\omega_1 := \Omega \setminus \omega_2$ . Denoting by  $\tilde{u}_h \in U_h$  the finite element function that interpolates  $\tilde{u}$  in all vertices of  $\mathcal{T}_h$ , we compute the target  $z_h \in Y_h$  as the state corresponding to  $\tilde{u}_h$  and  $f_h \equiv 10$ , *i.e.*, as the solution to  $-\operatorname{div}(\tilde{u}_h \nabla z_h) = f_h$  in  $\Omega$ ; see Figure [3a](#). We then compute a solution to  $(P_h)$  using the five desired coefficient values  $u_j = 0.25(j - 1)$ ,  $1 \leq j \leq 5$ , together with the parameters  $\alpha = 10^{-3}$  and  $\beta \in \{0, 10^{-6}, 5 \times 10^{-5}\}$ ; see Figure [3b-d](#) (with  $\gamma_{\text{final}} \approx 1 \times 10^{-4}$ ,  $\gamma_{\text{final}} \approx 1.8 \times 10^{-4}$ , and  $\gamma_{\text{final}} \approx 6.4 \times 10^{-2}$ , respectively).

Comparing the pure multi-bang design  $\bar{u}_h$  in Figure [3b](#) with the TV-multibang designs in Figure [3c](#) and [d](#), we clearly observe the well-known effect of TV regularization favoring level sets with smaller perimeter: While most jumps and the promotion of the desired parameter values are retained from the pure multi-bang design, the high-frequency “oscillations” between the level sets of  $\bar{u}_h = 1.0$  and  $\bar{u}_h = 0.75$  are removed. Similarly, the spurious “droplets” near  $x = (-1, 0)$  and  $x = (1, 0)$  are suppressed. (Here we recall that the multi-bang penalty acts purely pointwise and does not promote any spatial regularity.) The effect of the total variation penalty is also visible in Figure [3d](#), where the perimeters of the level sets for  $u_h^* = 0.5$  and  $u_h^* = 0.75$  have both been reduced, respectively, by closing the “slit” at  $x_1 = 0$  and by removing the highest-valued material. We point out that the simpler structure of the TV-regularized control may in itself be preferable in certain applications. (We also remark that if the admissible control values are restricted to  $(u_1, u_2) = (\tilde{u}_1, \tilde{u}_2) = (0, 1)$  and  $\alpha, \beta$  are chosen sufficiently small, the binary reference design is essentially recovered.)

### 6.2. Example 2: Parameter identification

The second example is motivated by a parameter identification related to electrical impedance tomography. Here, the goal is to reconstruct the spatially varying conductivity (which is a tissue-specific material parameter) from noisy observations of the electric field arising from external charges. It should be noted that in medical impedance tomography, external currents and observations are both taken on the boundary or a part thereof; for the sake of simplicity, however, we consider distributed charge density and observation.

We choose as true parameter

$$\tilde{u}(x) := \begin{cases} 1.5 & x \in \omega_1, \\ 1.6 & x \in \omega_2, \\ 1.7 & x \in \omega_3, \end{cases}$$

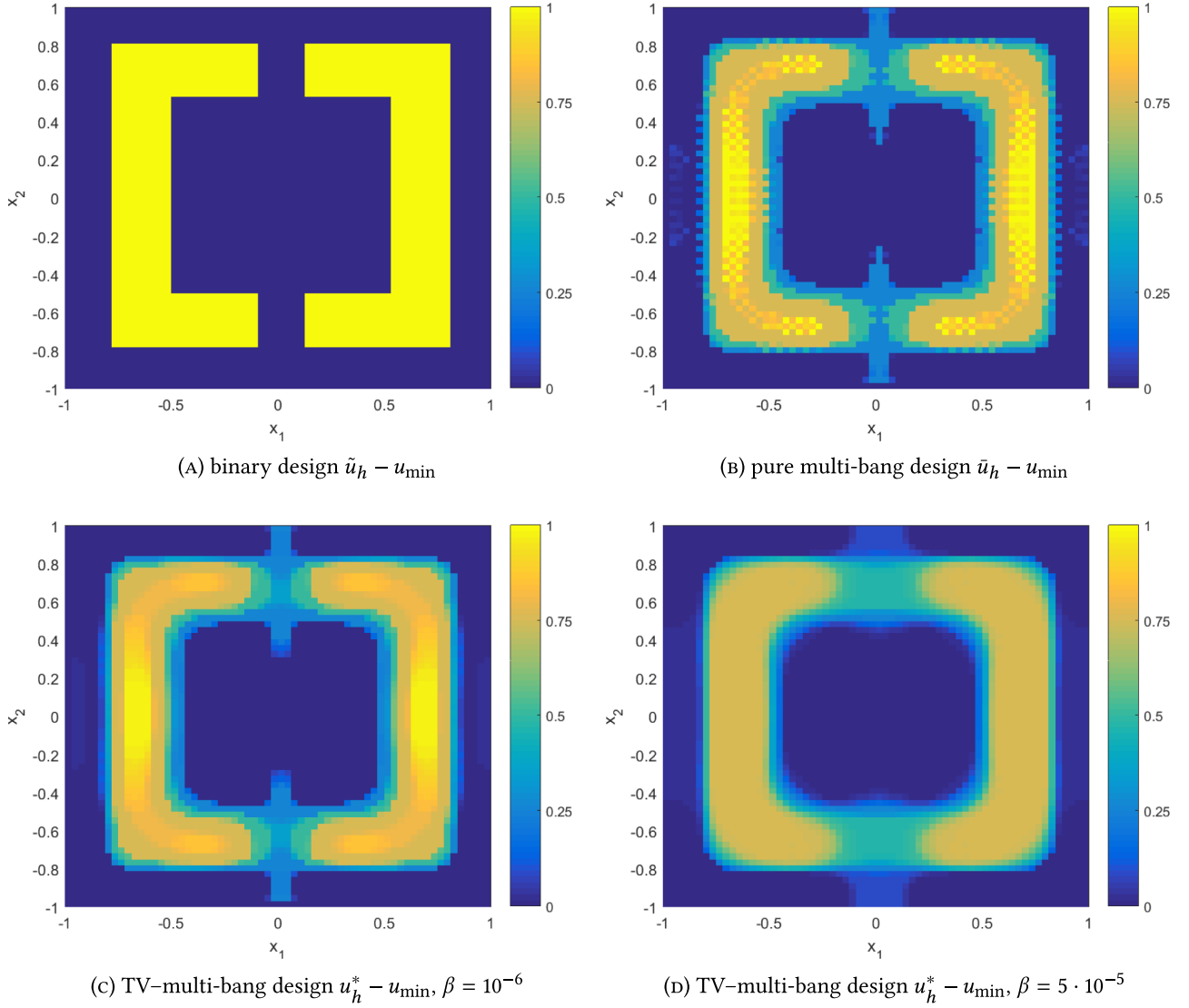


FIGURE 3. Comparison of binary, pure multi-bang, and total variation designs for Example 1.

where

$$\omega_1 := \{x \in \Omega : (x_1 + 0.1)^2 + (x_2 - 0.1)^2 \geq 0.4\}, \quad \omega_3 := \{x \in \Omega : (x_1 + 0.2)^2 + (x_2 - 0.2)^2 < 0.08\},$$

and  $\omega_2 := \Omega \setminus (\omega_1 \cup \omega_3)$  model background, tumor, and healthy tissue, respectively. Again,  $\tilde{u}_h \in U_h$  denotes the finite element function interpolating  $\tilde{u}$  in all vertices of  $\mathcal{T}_h$ ; see Figure 4a. For the target, we first compute a noise-free state  $\tilde{z}_h \in Y_h$  solving  $-\operatorname{div}(\tilde{u}_h \nabla \tilde{z}_h) = f_h$  in  $\Omega$ , where  $f_h \equiv 25$ . We now add noise to  $\tilde{z}_h$  to obtain  $z_h$ ; we use  $z_h := \tilde{z}_h + n_l \rho_h \max_{x \in \Omega} (|\tilde{z}_h(x)|)$ , where  $n_l := 10^{-3}$  and  $\rho_h \in Y_h$  is a finite element function whose coefficients  $\hat{\rho}_h \in \mathbb{R}^{N_{Y_h}}$  are sampled from a normal distribution with mean zero and standard deviation one. Corresponding to the assumption that strong *a priori* knowledge is available, we choose the desired coefficient values  $u_1 = 0$ ,  $u_2 = 0.1$  and  $u_3 = 0.2$ , together with the parameters  $\alpha = 5 \times 10^{-4}$  and  $\beta \in \{0, 10^{-5}, 10^{-6}\}$ ; see Figure 4b–d (with  $\gamma_{\text{final}} \approx 5.8 \times 10^{-6}$ ,  $\gamma_{\text{final}} \approx 2.9 \times 10^{-3}$ , and  $\gamma_{\text{final}} \approx 6.6 \times 10^{-3}$ , respectively).



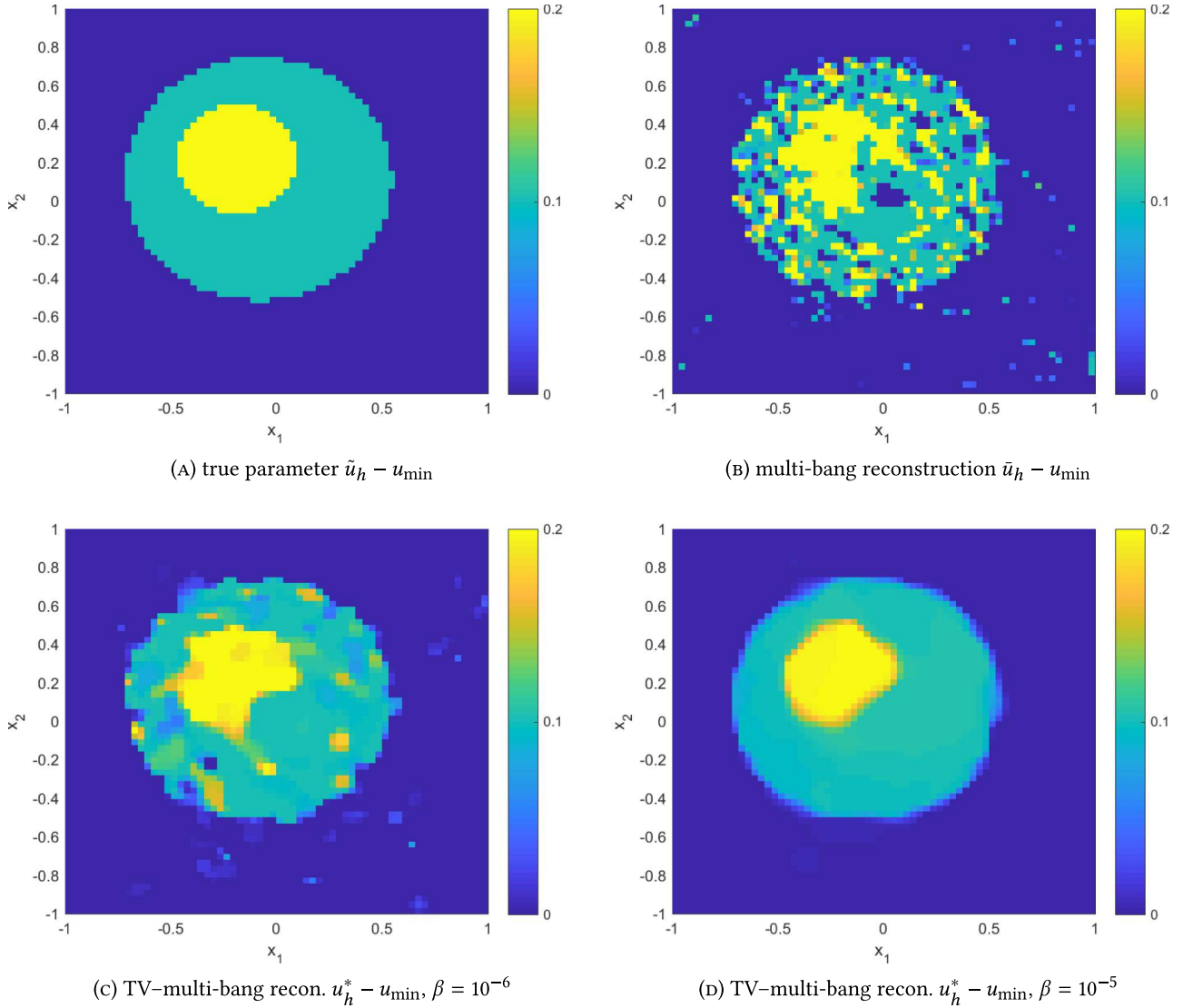


FIGURE 4. Comparison of true parameter, pure multi-bang, and total variation-regularized reconstructions for Example 2.

From Figure 4b, it is obvious that the pure multi-bang regularization fails for this challenging problem since the multi-bang penalty entails no spatial regularization. Specifically, noise remains in the homogeneous background, and many points in the healthy tissue region are misclassified as either tumor or background; the latter in particular in a large region near  $x = (0, 0)$  where  $\nabla \bar{y}_h \approx 0$  (compare (4.7)). The reconstruction is improved by adding the total variation regularization: with  $\beta = 10^{-6}$ , the “hole” near  $x = (0, 0)$  is gone, and the misclassified points are reduced; see Figure 4c. Increasing the total variation regularization parameter to  $\beta = 10^{-5}$  (Fig. 4d) again significantly improves the reconstruction by removing the small spurious inclusions while preserving the contrast and shape of the healthy tissue and tumor regions; merely the volume of the latter is slightly reduced. This indicates that regularization as understood in the context of inverse problems is predominantly provided by the total variation penalty, while the multi-bang penalty is responsible for maintaining the desired contrast

of the reconstruction. Hence, it suffices to investigate noise level-dependent parameter choice rules for  $\beta$  while keeping  $\alpha$  fixed, rather than having to consider – much more challenging – choice rules for multiple parameters.

## 7. CONCLUSION

Total variation regularization of topology optimization and parameter identification problems is challenging both analytically and numerically but is required in order to obtain existence of a solution without introducing additional smoothing. Furthermore, a pointwise multi-bang penalty can be used to promote optimal coefficients with desired (material) values. A reparametrization of the coefficient to be optimized allows proving existence as well as obtaining pointwise optimality conditions. The numerical solution is based on a finite element discretization and Moreau–Yosida regularization of reduced optimality conditions together with a semismooth Newton-type method combined with a predictive path-following strategy. Numerical examples indicate that in comparison to a pure multi-bang approach, the additional total variation regularization yields controls whose structure is much more regular.

## APPENDIX A. STRONGLY LIPSCHITZ DOMAINS ARE GRÖGER REGULAR

In this appendix, we address the relation between two different definitions of Lipschitz domains and the concept of Gröger regularity which are used in the literature. The first definition, sometimes referred to as a *strongly Lipschitz domain*, requires that, roughly speaking, the boundary can be represented locally as the graph of a Lipschitz function. A precise statement is the following from [3], A 8.2.

**Definition A.1** (Strongly Lipschitz domain). Let  $\Omega \subset \mathbb{R}^d$  be open and bounded. We say that  $\Omega$  has a *Lipschitz boundary* if there exists  $l \in \mathbb{N}$  such that  $\partial\Omega$  can be covered by open sets  $U^1, U^2, \dots, U^l$  and for  $j = 1, \dots, l$  there exist a Euclidean coordinate system  $e_1^j, e_2^j, \dots, e_d^j \in \mathbb{R}^d$ , a reference point  $y^j \in \mathbb{R}^{d-1}$ , numbers  $r^j > 0$  and  $h^j > 0$ , and a Lipschitz continuous function  $\eta^j : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$  that satisfy the following properties:

- (i)  $U^j = \left\{ x \in \mathbb{R}^d : |x_{-,d}^j - y^j|_2 < r^j \text{ and } |x_d^j - \eta^j(x_{-,d}^j)|_2 < h^j \right\}$ ;
- (ii) for all  $x \in U^j$ , if  $x_d^j = \eta^j(x_{-,d}^j)$  then  $x \in \partial\Omega$ ;
- (iii) for all  $x \in U^j$ , if  $0 < x_d^j - \eta^j(x_{-,d}^j) < h^j$  then  $x \in \Omega$ ;
- (iv) for all  $x \in U^j$ , if  $0 > x_d^j - \eta^j(x_{-,d}^j) > -h^j$  then  $x \notin \Omega$ .

Here, we have denoted  $x_{-,d}^j = (x_1^j, \dots, x_{d-1}^j)^T \in \mathbb{R}^{d-1}$  for  $x = x^j = (x_1^j, \dots, x_d^j)^T \in \mathbb{R}^d$ , and the coordinates of  $x^j$  are given in the local Euclidean coordinate system  $e_1^j, e_2^j, \dots, e_d^j$  in  $\mathbb{R}^d$ , i.e.,  $x^j = \sum_{i=1}^d x_i^j e_i^j$ .

A bounded domain with Lipschitz boundary is called a *strongly Lipschitz domain*.

Strongly Lipschitz domains are extension domains, which is required to obtain embeddings for Sobolev and BV functions into  $L^p$  spaces, and this definition is therefore used in [4, 15].

The second definition, sometimes referred to as a *weakly Lipschitz domain*, requires, roughly speaking, that the boundary can be locally flattened by a bi-Lipschitz transformation; a precise definition can be found in, e.g., ([16], Sect. 6). For our purposes, however, the following related concept from ([24], Def. 2) is more important.

**Definition A.2** (Gröger regularity). A set  $G \subset \mathbb{R}^d$  is called *regular (in the sense of Gröger)* if  $G$  is bounded and if for every  $y \in \partial G$  there exist subsets  $U$  and  $\tilde{U}$  of  $\mathbb{R}^d$  and a Lipschitz continuous bijection  $\Phi : U \rightarrow \tilde{U}$  with Lipschitz continuous inverse  $\Phi^{-1}$  such that  $U$  is an open neighborhood of  $y$  in  $\mathbb{R}^d$  and that  $\Phi(U \cap G)$  is one of the sets

$$E_1 := \left\{ x \in \mathbb{R}^d : |x| < 1, x_d < 0 \right\},$$

$$E_2 := \left\{ x \in \mathbb{R}^d : |x| < 1, x_d \leq 0 \right\},$$

$$E_3 := \{x \in E_2 : x_d < 0 \text{ or } x_1 > 0\},$$

where  $x = (x_1, x_2, \dots, x_d)^T$ .

The main result of [24] is that a second order elliptic mixed boundary value problem on a bounded domain  $\Omega$  admits higher regularity of the solution if  $G = \Omega \cup \Gamma_N$  is regular, where  $\Gamma_N \subset \partial\Omega$  denotes the Neumann boundary. For  $\Gamma_N = \emptyset$  (*i.e.*, pure Dirichlet conditions, where  $E_2$  and  $E_3$  are not needed), Definition A.2 reduces to that of  $\Omega$  being a weakly Lipschitz domain. Furthermore, [25], Section 5, shows also for mixed boundary conditions (under some assumptions on  $\Gamma_N$ ) that if  $\Omega \cup \Gamma_N$  is regular then  $\Omega$  is a weakly Lipschitz domain and, for  $d \in \{2, 3\}$ , *vice versa*.

In our analysis, we require the domain  $\Omega$  to satisfy both Definitions A.1 and A.2 since we use results from [4, 15] as well as from [24]. However, the notions of strongly and weakly Lipschitz domains are not equivalent; examples of weakly but not strongly Lipschitz domains can be found in, *e.g.*, ([16], Sect. 6). Although it is commonly accepted that strongly Lipschitz domains are regular (or, equivalently for domains, that they are weakly Lipschitz), despite our best efforts we could not find a proof of this fact in the literature. For the sake of completeness, we therefore provide one here.

**Lemma A.3.** *If a domain  $\Omega \subset \mathbb{R}^d$  satisfies Definition A.1, then it also satisfies Definition A.2.*

*Proof.* Let  $\Omega \subset \mathbb{R}^d$ ,  $d \in \mathbb{N}$ , denote the set in question and let  $\hat{x} \in \partial\Omega$ . Due to Definition A.1 there exist an open neighborhood  $\hat{V}$  of  $\hat{x}$  and a Lipschitz continuous function  $\eta : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$  such that  $\Omega \cap \hat{V} = \{x \in \hat{V} : \eta(x_1, \dots, x_{d-1}) < x_d\}$  and  $\eta(\hat{x}_1, \dots, \hat{x}_{d-1}) = \hat{x}_d$ . Defining

$$\hat{\Lambda} : \mathbb{R}^d \rightarrow \mathbb{R}^d, \quad \hat{\Lambda}(x) := (x_1, \dots, x_{d-1}, \eta(x_1, \dots, x_{d-1}) - x_d)^T,$$

we observe that  $\Omega \cap \hat{V} = \{x \in \hat{V} : \hat{\Lambda}_d(x) < 0\}$ . Clearly,  $\hat{\Lambda}$  is Lipschitz. Moreover, since  $\hat{\Lambda}(\hat{\Lambda}(x)) = x$  for all  $x \in \mathbb{R}^d$ , we infer that  $\hat{\Lambda}$  and its inverse mapping  $\hat{\Xi} := \hat{\Lambda}^{-1} = \hat{\Lambda}$  are bijective. (In the following, we nevertheless distinguish between  $\hat{\Lambda}$  and its inverse for the sake of transparency.) Since  $\hat{\Lambda}$  is Lipschitz continuous,  $\hat{\Xi}$  maps open sets to open sets. Defining  $\hat{y} := \hat{\Lambda}(\hat{x})$  we note that  $\hat{\Xi}$  maps  $B_\delta(\hat{y})$  for every  $\delta > 0$  bijectively to  $\hat{\Xi}(B_\delta(\hat{y}))$ , which is an open neighborhood of  $\hat{x}$ . In particular, there is  $\delta > 0$  such that  $\hat{\Xi}$  maps  $B_\delta(\hat{y})$  bijectively to  $V := \hat{\Xi}(B_\delta(\hat{y}))$  with  $V \subset \hat{V}$ . Consequently,  $\Xi(y) := \hat{\Xi}(\hat{y} + \delta y)$  maps  $B_1(0)$  bijectively to  $V$ , is Lipschitz continuous, and has the Lipschitz continuous inverse  $\Lambda(x) := (\hat{\Lambda}(x) - \hat{y})/\delta$ . It follows that

$$\Omega \cap V = \{x \in V : \hat{\Lambda}_d(x) < 0\} = \{\Xi(y) \in \mathbb{R}^d : y \in B_1(0), \hat{\Lambda}_d(\Xi(y)) < 0\}.$$

This implies that

$$\Lambda(\Omega \cap V) = \{y \in \mathbb{R}^d : y \in B_1(0), \hat{y}_d + \delta y_d < 0\} = \{y \in \mathbb{R}^d : y \in B_1(0), y_d < 0\},$$

where we have used  $\hat{y}_d = 0$ . Summarizing, we have established that for  $\hat{x} \in \partial\Omega$ , there is an open neighborhood  $V$  of  $\hat{x}$  and a Lipschitz continuous bijection  $\Lambda : V \rightarrow B_1(0)$  with Lipschitz continuous inverse such that  $\Lambda(\Omega \cap V) = \{y \in B_1(0) : y_d < 0\}$ . That is,  $G := \Omega$  satisfies Definition A.2.  $\square$

*Acknowledgements.* Support by the German Science Fund (DFG) under grant CL 487/1-1 for C.C. and by the ERC advanced grant 668998 (OCLOC) under the EU's H2020 research program for K.K. are gratefully acknowledged.

## REFERENCES

- [1] U. Aßmann and A. Röscher, Identification of an unknown parameter function in the main part of an elliptic partial differential equation. *Z. Anal. Anwend.* **32** (2013) 163–178.
- [2] M.S. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M.E. Rognes, and G.N. Wells, The FEniCS project version 1.5. *Arch. Numer. Softw.* **3** (2015) 9–23.

- [3] H.W. Alt, Linear Functional Analysis. An Application-Oriented Introduction (Universitext). Springer, London (2016).
- [4] L. Ambrosio, N. Fusco and D. Pallara, Functions of Bounded Variation and Free Discontinuity Problems. *Oxford Mathematical Monographs*. The Clarendon Press, Oxford University Press, New York (2000).
- [5] S. Amstutz, A semismooth Newton method for topology optimization. *Nonlinear Anal.* **73** (2010) 1585–1595.
- [6] S. Amstutz, Analysis of a level set method for topology optimization. *Optim. Meth. Softw.* **26** (2011) 555–573.
- [7] S. Amstutz and H. Andrä, A new algorithm for topology optimization using a level-set method. *J. Comput. Phys.* **216** (2006) 573–588.
- [8] S. Amstutz and N. Van Goethem, Topology optimization methods with gradient-free perimeter approximation. *Interfaces Free Bound.* **14** (2012) 401–430.
- [9] H. Attouch and H. Brezis, Duality for the sum of convex functions in general Banach spaces, in Aspects of Mathematics and Its Applications. Vol. 34 of *North-Holland Math. Library*. North-Holland, Amsterdam (1986) 125–133.
- [10] H. Attouch, G. Buttazzo and G. Michaille, Variational Analysis in Sobolev and BV Spaces. Applications to PDEs and Optimization. Vol. 6 of *MPS/SIAM Series on Optimization*, 2nd revised edn. SIAM, Philadelphia, PA (2014).
- [11] V. Barbu and T. Precupanu, Convexity and Optimization in Banach Spaces. *Springer Monographs in Mathematics*, 4th edn. Springer, Dordrecht (2012).
- [12] S. Bartels, Total variation minimization with finite elements: convergence and iterative solution. *SIAM J. Numer. Anal.* **50** (2012) 1162–1180.
- [13] L. Blank, M.H. Farshbaf-Shaker, H. Garcke, C. Rupprecht and V. Styles, Multi-material phase field approach to structural topology optimization, in Trends in PDE Constrained Optimization, edited by G. Leugering et al. Vol. 165 of *International Series of Numerical Mathematics*. Birkhäuser, Cham (2014) 231–246.
- [14] B. Bourdin and A. Chambolle, Design-dependent loads in topology optimization. *ESAIM: COCV.* **9** (2003) 19–48.
- [15] K. Bredies and M. Holler, A pointwise characterization of the subdifferential of the total variation functional. Preprint [arXiv:1609.08918](https://arxiv.org/abs/1609.08918) (2016).
- [16] K. Brewster and M. Mitrea, Boundary value problems in weighted Sobolev spaces on Lipschitz manifolds. *Mem. Differ. Eq. Math. Phys.* **60** (2013) 15–55.
- [17] E. Casas, K. Kunisch and C. Pola, Regularization by functions of bounded variation and applications to image enhancement. *Appl. Math. Optim.* **40** (1999) 229–257.
- [18] E. Casas, R. Herzog and G. Wachsmuth, Approximation of sparse controls in semilinear equations by piecewise linear functions. *Numer. Math.* **122** (2012) 645–669.
- [19] G. Chavent and K. Kunisch, Regularization of linear least squares problems by total bounded variation. *ESAIM: COCV.* **2** (1997) 359–376.
- [20] C. Clason and K. Kunisch, Multi-bang control of elliptic systems. *Annales de l'Institut Henri Poincaré (C) Analyse Non Linéaire* **31** (2014) 1109–1130.
- [21] C. Clason and K. Kunisch, A convex analysis approach to multi-material topology optimization. *ESAIM: M2AN* **50** (2016) 1917–1936.
- [22] I. Ekeland and R. Témam, Convex Analysis and Variational Problems. Vol. 28 of *Classics Appl. Math.* SIAM, Philadelphia (1999).
- [23] E. Giusti, Minimal Surfaces and Functions of Bounded Variation. Vol. 80 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel (1984).
- [24] K. Gröger, A  $W^{1,p}$ -estimate for solutions to mixed boundary value problems for second order elliptic differential equations. *Math. Ann.* **283** (1989) 679–687.
- [25] R. Haller-Dintelmann, C. Meyer, J. Rehberg and A. Schiela, Hölder continuity and optimal control for nonsmooth elliptic problems. *Appl. Math. Optim.* **60** (2009) 397–428.
- [26] J. Haslinger, M. Kočvara, G. Leugering and M. Stingl, Multidisciplinary free material optimization. *SIAM J. Appl. Math.* **70** (2010) 2709–2728.
- [27] K. Ito and K. Kunisch, Lagrange Multiplier Approach to Variational Problems and Applications. Vol. 15 of *Advances in Design and Control*. SIAM, Philadelphia, PA (2008).
- [28] B. Kummer, Newton's method for non-differentiable functions. *Math. Res.* **45** (1988) 114–125.
- [29] A. Logg and G.N. Wells, DOLFIN: automated finite element computing. *ACM Trans. Math. Softw.* **37** (2010) 20.
- [30] A. Logg, K.-A. Mardal, G.N. Wells et al., Automated Solution of Differential Equations by the Finite Element Method. Springer, New York (2012).
- [31] A. Logg, G.N. Wells and J. Hake, DOLFIN: A C++/Python Finite Element Library, Chapter 10. Springer, Berlin Heidelberg (2012).
- [32] R. Mifflin, Semismooth and semiconvex functions in constrained optimization. *SIAM J. Control Optim.* **15** (1977) 959–972.
- [33] F. Murat, Contre-exemples pour divers problèmes où le contrôle intervient dans les coefficients. *Ann. Mat. Pura Appl.* **112** (1977) 49–68.
- [34] F. Murat and L. Tartar,  $H$ -convergence, in Topics in the Mathematical Modelling of Composite Materials. Vol. 31 of *Progr. Nonlinear Differential Equations Appl.* Birkhäuser Boston, Boston, MA (1997) 21–43.
- [35] K. Pieper, Finite element discretization and efficient numerical solution of elliptic and parabolic sparse control problems. Dissertation, Technische Universität München, München (2015).
- [36] L.I. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms. *Physica D* **60** (1992) 259–268.

- [37] O. Scherzer, Explicit versus implicit relative error regularization on the space of functions of bounded variation, in *Inverse Problems, Image Analysis, and Medical Imaging*, edited by M.Z. Nashed and O. Scherzer, eds. Vol. 313 of *Contemp. Math.* American Mathematical Society, Providence, RI (2002) 171–198.
- [38] W. Schirotzek, *Nonsmooth Analysis (Universitext)*. Springer, Berlin (2007).
- [39] L. Tartar, The appearance of oscillations in optimization problems, in *Nonclassical Continuum Mechanics (Durham, 1986)*. Vol. 122 of *London Math. Soc. Lecture Note Ser.* Cambridge Univ. Press, Cambridge (1987) 129–150.
- [40] L. Tartar, The General Theory of Homogenization: A Personalized Introduction. Vol. 7 of *Lecture Notes of the Unione Matematica Italiana*. Springer, Berlin; UMI, Bologna (2009).
- [41] C.P. Trautmann, Sparse measure-valued optimal control problems governed by wave equations. Dissertation, Karl-Franzens-Universität Graz, Graz (2015).
- [42] F. Tröltzsch, *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*. Translated from the German by Jürgen Sprekels. American Mathematical Society, Providence, RI (2010).
- [43] M. Ulbrich, *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*, Vol. 11 of *MOS-SIAM Series on Optimization*. SIAM, Philadelphia, PA (2011).
- [44] W.P. Ziemer, *Weakly Differentiable Functions*. Vol. 120 of *Graduate Texts in Mathematics*. Springer, New York (1989).