Conditional gradients for total variation regularization with PDE constraints: a graph cuts approach

Giacomo Cristinelli*, José A. Iglesias*, Daniel Walter†

Abstract

Total variation regularization has proven to be a valuable tool in the context of optimal control of differential equations. This is particularly attributed to the observation that TV-penalties often favor piecewise constant minimizers with well-behaved jumpsets. On the downside, their intricate properties significantly complicate every aspect of their analysis, from the derivation of first-order optimality conditions to their discrete approximation and the choice of a suitable solution algorithm. In this paper, we investigate a general class of minimization problems with TV-regularization, comprising both continuous and discretized control spaces, from a convex geometry perspective. This leads to a variety of novel theoretical insights on minimization problems with total variation regularization as well as tools for their practical realization. First, by studying the extremal points of the respective total variation unit balls, we enable their efficient solution by geometry exploiting algorithms, e.g. fully-corrective generalized conditional gradient methods. We give a detailed account on the practical realization of such a method for piecewise constant finite element approximations of the control on triangulations of the spatial domain. Second, in the same setting and for suitable sequences of uniformly refined meshes, it is shown that minimizers to discretized PDE-constrained optimal control problems approximate solutions to a continuous limit problem involving an anisotropic total variation reflecting the fine-scale geometry of the mesh.

1 Introduction

In this paper, for $\Omega \subset \mathbb{R}^d$ a bounded domain with strongly Lipschitz boundary, we consider minimization problems of the form

$$\min_{u \in V} J(u) := [F(Ku) + \mathrm{TV}(u, \Omega)], \tag{$\mathcal{P}$}$$

where $V \subset L^q(\Omega)$ is a weakly closed linear subspace of $L^q(\Omega)$, $q = d/(d-1)$, and the objective functional consists of two parts: First, the composition of a smooth functional F with a linear continuous operator $K : V \to Y$ mapping to a Hilbert space of observations Y. Second, the total variation $\mathrm{TV}(u, \Omega)$ of a function $u \in V$. While this general setting encompasses denoising problems, [59], by setting $K = \mathrm{Id}$, our particular focus lies on instances in which the forward

---

*Department of Applied Mathematics, University of Twente, 7500AE Enschede, The Netherlands (`g.cristinelli@utwente.nl`, `jose.iglesias@utwente.nl`)

†Institut für Mathematik, Humboldt-Universität zu Berlin, 10117 Berlin, Germany (`daniel.walter@hu-berlin.de`)

mapping K is compact and potentially costly to evaluate. As a guiding example, we point out optimal control problems of the form

$$\min_{u,y} \left[\frac{1}{2} \|y - y_o\|^2_{L^2(\Omega)} + \alpha \|Du\|_{M(\Omega)} \right] \tag{$\mathcal{D}$}$$

where $\alpha > 0$, $\|Du\|_{M(\Omega)} = \mathrm{TV}(u, \Omega)$ denotes the norm of the distributional gradient Du seen as a vector-valued Radon measure and the control and state variables u and y are coupled by the PDE-constraint

$$-\mathrm{div}\,(a\nabla y) + cy = u \text{ in } \Omega, \quad y = 0 \text{ on } \partial\Omega. \tag{1.1}$$

Imposing suitable regularity assumptions on the coefficient functions a and c, respectively, Problem ($\mathcal{D}$) fits into the general setting ($\mathcal{P}$) by choosing $F(\cdot) = (1/2\alpha)\|\cdot - y_o\|^2_{L^2(\Omega)}$ and introducing the control-to-state operator $y = Ku$ which maps the right hand side of a Poisson-type problem to its unique solution. Problems of this or similar form have received considerable interest since the total variation regularization favors minimizers which only attain a finite number of values on Ω. We refer e.g. to [7, 34, 23, 24, 40] without pretence of completeness. However, this desirable structure of optimal solutions comes at the cost of additional difficulties such as the non-reflexivity of the space $\mathrm{BV}(\Omega)$, the nonsmoothness of the total variation as well as the possible lack of strong convexity in Problem ($\mathcal{P}$) due to the compactness of the observation operator K. These complicate both the theoretical analysis of the problem as well as its algorithmic solution.

A possible approach to alleviating some of these difficulties can be found by interpreting the TV-term as the Minkowski functional of the set

$$B_{\mathring{V}} = \left\{ u \in V \,\middle|\, \mathrm{TV}(u, \Omega) \leq 1, \ \int_\Omega u(x)\,\mathrm{d}x = 0 \right\}$$

and studying the convex geometry of the latter, in particular its set of extremal points $\mathrm{Ext}(B_{\mathring{V}})$. From this abstract perspective it turns out that Problem ($\mathcal{P}$) encourages minimizers that are given, up to a constant shift, by conic combinations of few elements in $\mathrm{Ext}(B_{\mathring{V}})$. Moreover, existence of such sparse solutions is ensured by convex representer theorems if the range of K is finite dimensional, [8, 13]. For $V = L^q(\Omega)$, an explicit characterization of $\mathrm{Ext}(B_{\mathring{V}})$ is available which identifies them as scaled characteristic functions of simple sets. This closes the circle to the aforementioned practical observation of piecewise constant solutions attaining few values. Moreover, this geometric interpretation also offers the opportunity for the application of efficient solution algorithms which can capitalize on these observations and produce fast-converging iterates comprising few extremal points. For example, the fully-corrective generalized conditional gradient method (FC-GCG) from [16] relies on alternating between the greedy update of finite sets $\mathcal{A}_k$ of extremal points by the solution of linear minimization problems over $\mathrm{Ext}(B_{\mathring{V}})$, and computing new iterates in the conic hull of $\mathcal{A}_k$ by the solution of convex finite dimensional minimization problems with an ℓ_1-type penalty term.

However, in many cases, e.g. for Problems of the form ($\mathcal{D}$), the action of the observation operator K on a function $u \in \mathrm{BV}$ cannot be computed analytically but requires the assistance of numerical methods and thus, ultimately, a discretization of the variable space. In the simplest case, this can be achieved by considering piecewise constant, $V = P_0(\mathcal{T}_h)$, or piecewise affine and continuous, $V = P_1(\mathcal{T}_h)$, finite element spaces subordinate to a triangulation $\mathcal{T}_h$ with grid width h of Ω for the discretization of u and by replacing K by a finite dimensional aproximation. While this yields a finite dimensional minimization problem, many of the original challenges, e.g. the nonsmoothness of the TV-term, still remain. Moreover, neglecting the infinite dimensional character of the continuous Problem ($\mathcal{P}$) and applying black box methods for the arising non-smooth minimization problems harbors the danger of mesh-dependence, i.e., the convergence

behavior of the chosen algorithm may quickly deteriorate with decreasing grid width. Finally, discretizing the problem also raises the question of the limiting behavior of its solutions and their convergence towards minimizers of the continuous problem as the grid width tends to zero. While the answer to this is affirmative for $V = P_1(\mathcal{T}_h)$, e.g. based on [20, Theorem 4.1] or [5], it is well known that the same does not hold true for $V = P_0(\mathcal{T}_h)$. More in detail, given a generic function $u \in \mathrm{BV}(\Omega)$ as well as a family of triangulations $\{\mathcal{T}_h\}_{h>0}$ of Ω, there may be no $\{u_h\}_{h>0}$ with $u_h \in P_0(\mathcal{T}_h)$ and

$$\lim_{h \to 0} \left[\|u - u_h\|_{L^1} + \left| \mathrm{TV}(u, \Omega) - \mathrm{TV}(u_h, \Omega) \right| \right] = 0,$$

see e.g. [5] or the counterexample [25, Example 3.6]. As a consequence, accumulation points of sequences of minimizers $\{\bar{u}_h\}_{h>0}$ to ($\mathcal{P}$) for $V = P_0(\mathcal{T}_h)$ might not constitute minimizers of the corresponding continuous problem.

1.1 Contribution

The present paper aims at contributing towards the efficient solution of discretized minimization problems with total variation regularization. Moreover, we further investigate the behavior of solutions to Problem ($\mathcal{P}$) for piecewise constant approximations, $V = P_0(\mathcal{T}_h)$, and vanishing grid width $h > 0$. In detail, the contributions of the present work are threefold:

Convex geometry of total variation balls for $P_0(\mathcal{T})$ and $P_1(\mathcal{T})$
Motivated by the recent interest in the interplay between nonsmooth minimization and convex geometry, we investigate the extremal points of the set $B_{\mathring{V}}$ for the particular choice of $V = P_0(\mathcal{T})$ and $V = P_1(\mathcal{T})$, respectively. This yields two key observations: On the one hand, in the piecewise constant case, we show that $\mathrm{Ext}(B_{\mathring{P}_0}) = \mathrm{Ext}(B_{\mathring{L}^q}) \cap P_0(\mathcal{T})$, i.e., piecewise constant discretizations are, loosely speaking, conforming with the convex geometry of the continuous ball $\mathrm{Ext}(B_{\mathring{L}^q})$. In particular, this implies that $B_{\mathring{P}_0}$ is a convex polytope. On the other hand, we also show that piecewise affine and globally continuous discretizations, $V = P_1(\mathcal{T}_h)$, fail to capture the geometry of the limiting problem, since one clearly has $\mathrm{Ext}(B_{\mathring{P}_1}) \cap \mathrm{Ext}(B_{\mathring{L}^q}) = \emptyset$, and, more surprisingly, the sublevel set $B_{\mathring{P}_1}$ is curved, thus admitting infinitely many extremal points.

A FC-GCG method for $V = P_0(\mathcal{T})$
We propose a fully-corrective generalized conditional gradient method for the solution of minimization problems with total variation regularization. For this purpose, an equivalent formulation of Problem ($\mathcal{P}$) is considered which introduces the mean value of u as a separable variable. The resulting problem fits into the abstract framework of [16] and thus allows for the immediate application of FC-GCG methods as well as of the presented worst-case convergence results therein. For the case of piecewise constant discretizations, we give a detailed description of its practical realization. Using the derived characterization of $\mathrm{Ext}(B_{\mathring{P}_0})$, this requires the repeated solution of problems of the form

$$\max_{v \in P_0(\mathcal{T}_h)} \frac{\int_\Omega pv \, \mathrm{d}x}{\mathrm{TV}(v, \Omega)} \quad \text{where} \quad p \in P_0(\mathcal{T}_h), \quad \int_\Omega p \, \mathrm{d}x = 0 \tag{1.2}$$

is a given dual variable. While finite, the size of its admissible set might grow exponentially with h, thus requiring iterative methods for its solution. Inspired by classical approaches from fractional minimization, in particular the so-called Dinkelbach method, [61], we argue that (1.2) is equivalent to solving a finite number of minimal cut problems on the dual graph of the mesh. The practical efficiency of the resulting algorithm is demonstrated on two instances of the PDE-constrained minimization problem ($\mathcal{D}$), each of them in two and three dimensions.

3

Discrete-to-continuum limits for $V = P_0(\mathcal{T}_h)$

We show that for a sequence of periodically refined triangulations $\{\mathcal{T}_k\}_{k \in \mathbb{N}}$ on the unit cube $\Omega = [0,1]^d$, see Definition 5.2, the restriction of the total variation onto $P_0(\mathcal{T}_k)$, denoted by $\mathrm{TV}_{\mathcal{T}_k}$, approximates a limiting energy of the form

$$\mathrm{TV}_{\varphi_{\mathcal{T}_1}}(u, \Omega) := \sup\left\{ \int_\Omega u \operatorname{div} \psi \, \mathrm{d}x \,\middle|\, \varphi_{\mathcal{T}_1}^*(\psi) \leq 1 \text{ for all } x \in \Omega \right\} = \int_\Omega \varphi_{\mathcal{T}_1}\left(\frac{Du}{|Du|} \right) \mathrm{d}|Du|.$$

in the sense of Γ-convergence w.r.t the topology on $L^1(\Omega)$. Here, the functional $\varphi_{\mathcal{T}_1}$ captures the anisotropy of the triangulation and is given implicitly by the optimal value function of a nonsmooth minimization problem. For the particular case of the usual "double-diagonal" triangulation, see Figure 4, we explicitly characterize $\varphi_{\mathcal{T}_1}$ by proving that the level set $\varphi_{\mathcal{T}_1}^{-1}(1)$ is given by a regular octagon. Finally, we point out that the convergence of the total variation immediately translates to minimizers of PDE-constrained problems of the form ($\mathcal{D}$). More in detail, we show that finite element discretizations of Problem ($\mathcal{D}$) with P_0 elements for the control and P_1 elements for the state variable asymptotically approximate

$$\min_{u,y} \left[\frac{1}{2} \|y - y_o\|_{L^2(\Omega)}^2 + \alpha \, \mathrm{TV}_{\varphi_{\mathcal{T}_1}}(u, \Omega) \right] \quad \text{s.t.} \quad (1.1).$$

1.2 Related work

Nonsmooth optimization & convex geometry

The empirical observation that the proper choice of a convex nonsmooth regularization function enhances desired structural features in the solutions of associated minimization problems has solidified their role as a cornerstone of modern approaches in optimal control, variational regularization in inverse problems as well as applications in data science and machine learning. In the case of one-homogeneous regularizers, this inherent property of the regularizer has recently, see e.g. [8, 13, 16], been linked to the geometry of its, in a suitable sense compact, sublevel sets, in particular the corresponding set of extremal points. More in detail, nonsmooth regularization tends to favor sparse solutions, i.e., elements that are given by a conic combination of finitely many extremals. This fundamental insight has led to an increased interest in the explicit characterizations of the extremals points for practically relevant regularizers as well as well as numerical solution methods which capitalize on this sparsity by using the precise structure of the extremal points in their formulation. For an incomplete list of examples, we refer, e.g., to [17, 14, 15, 51, 22, 2, 57, 65].

Solution algorithms for total variation regularized problems

Due to the significance of total variation regularization in the context of mathematical imaging, there is a wide variety of different numerical algorithms for the solution of (discretized) problems of the form ($\mathcal{P}$) with quadratic F. Without pretence of completeness, we mention dual and primal-dual approaches, e.g. semismooth Newton and primal-dual active set methods, [56], as well as split Bregman iteration, [43], and general splitting-type methods, e.g. [31]. In [6], the authors propose a conditional gradient method for the solution of the dual problem associated to ($\mathcal{P}$). For more details and further reference, we point to the literature reviews [26, 33]. While many of these approaches easily generalize to larger classes of functionals F and forward operators K, others, like minimization of graph cut energies to obtain level sets separately, [37, 28], are inherently tied to the structure of image denoising problems. Moreover, we again point out that, in the context of mathematical imaging, the operator K is usually either the identity or an integral operator with given kernel and thus relatively cheap to evaluate. This constitutes a conceptual difference to the current work in which K is thought to be connected to the solution of a differential equation. Hence, extensive evaluation of K or its adjoint K^*, as e.g. required in splitting type methods, can potentially create a computational bottleneck,

see also the discussion in [52]. Split Bregman methods for discretized PDE-constrained optimal control problems have been considered in [39], while, e.g. [35, 45], put the focus on the analysis of (generalized) Newton methods for infinite dimensional instances of Problem ($\mathcal{D}$) or the associated dual problem. Proving the well-posedness and local fast convergence convergence of the latter, however, requires further regularization and/or smoothing of the problem. This introduces new hyperparameters which require additional attention, both from a theoretical as well as from a practical perspective, due to the need of a suitable path-following strategy.

In contrast, the method presented in this manuscript is both simple and practically efficient, achieving satisfactory results in a moderate number of steps at the cost of two PDE solves per iteration. It is closest related to the modified Frank-Wolfe algorithm in the recent work [38] for off-the-grid image reconstruction on unbounded domains as well as the fully corrective conditional gradient in [47] for finite dimensional problems with TV-regularization. However, it differs from these earlier works in several key aspects, e.g. the specific form of the arising subproblems and/or their numerical treatment . We give a more detailed account of these differences in Remark 2.10 and 4.7, respectively. Finally, let us point out that FC-GCG methods for total variation regularization have already been considered in [62] for the one dimensional setting, i.e. $d = 1$ and $\Omega = (0, T)$. These results, however, rely on the identification of $\mathrm{BV}(\Omega) \simeq M(\Omega; \mathbb{R}^d) \times \mathbb{R}$ which does not hold for $d > 1$.

FE-approximation of optimal control problems & discrete-to-continuum limits
While there is a considerable amount of literature on finite element approximations of problems in imaging, see e.g. [5, 49, 50, 32] as well as their asymptotic analysis, previous works in the context of PDE-constrained optimal control with total variation regularization have mainly focused on the one-dimensional case, i.e. $d = 1$, [23, 41, 48, 46], or the finite element analysis of modified problems, e.g. by smoothing the total variation term, see e.g. [45] or the Master's thesis [44].

In the present work, rather than relying on approximation theory, the derivation of the limiting energy $\mathrm{TV}_{\varphi_{\mathcal{T}_1}}$ is inspired by results on discrete-to-continuum limits for variational problems on lattice systems, in particular the integral representation of [12, Thm. 2.4] as well as the cell formula of [30, Prop. 2.6]. Moreover, the explicit characterization of $\varphi_{\mathcal{T}_1}$ for the particular case of the double diagonal triangulation relies on a discrete analogue of slicing results used to fix boundary conditions in relaxation, lower semincontinuity and variational nonlinear homogenization results.

1.3 Structure of the paper

We collect notation and basic definitions in Subsection 1.4 below. In Section 2 we review the details of the existence of minimizers for ($\mathcal{P}$), optimality conditions, and introduce the abstract generalized conditional gradient strategy to be used. In Section 3, in view of Galerkin formulations of ($\mathcal{P}$), we focus on the convex geometry of the total variation balls of piecewise constant and piecewise affine functions on polygonal meshes. Having identified the extremal points in the piecewise constant case, in Section 4 we introduce our realization of the generalized conditional gradient for this problem, in which the insertion step is treated through a Dinkelbach method with inner iterations solved with graph cuts. Section 5 is dedicated to discrete to continuum convergence of the minimization problem for piecewise constant functions on periodic triangulations, and to the explicit computation of the resulting anisotropy for a simple 2D case. In Section 6, we numerically apply the proposed algorithm to a few model situations in 2D and 3D, and present some comparisons with other approaches to ($\mathcal{P}$). Appendix A contains additional proofs skipped in Section 2, in particular basic convergence results applicable to the method. Finally, in Appendix B we provide detailed proofs relating to the minimal path interpretation of the anisotropy used in Section 5.

1.4   Notation

We first recall basic facts (the proof of all of which can be found in [4, Ch. 3]) and fix notation for functions of bounded variation which feature repeatedly throughout later sections. For $u \in L^1(\Omega)$ one defines its total variation as

$$\mathrm{TV}(u, \Omega) = \sup \left\{ \int_\Omega u \operatorname{div} \psi \, dx \,\middle|\, \psi \in \mathcal{C}_c^1(\Omega; \mathbb{R}^d), \|\psi\|_\infty \leq 1 \right\}$$

where

$$\|\psi\|_\infty = \sup_{x \in \Omega} |\psi(x)| \quad \text{for all } \psi \in \mathcal{C}_c(\Omega; \mathbb{R}^d),$$

with $|\cdot|$ denoting the euclidean norm. We call $u \in L^1(\Omega)$ a function of bounded variation if $\mathrm{TV}(u, \Omega) < \infty$. Note that u is of bounded variation if and only if there is a regular Borel measure $Du = (\partial_{x_1} u, \ldots, \partial_{x_d} u) \in M(\Omega; \mathbb{R}^d)$ with

$$\int_\Omega \operatorname{div} \psi(x) u(x) \, dx = -\sum_{j=1}^d \int_\Omega \psi_j(x) \, d\partial_{x_j} u(x) \quad \text{for all } \psi \in \mathcal{C}_c^1(\Omega; \mathbb{R}^d),$$

and whenever $u \in W^{1,1}(\Omega)$, then we have that the weak gradient ∇u equals Du. The linear space of functions of bounded variation

$$\mathrm{BV}(\Omega) := \left\{ u \in L^1(\Omega) \,\middle|\, \mathrm{TV}(u, \Omega) < \infty \right\}$$

is a Banach space with respect to the norm

$$\|u\|_{\mathrm{BV}} = \|u\|_{L^1(\Omega)} + \mathrm{TV}(u, \Omega).$$

The space $\mathrm{BV}(\Omega)$ continuously embeds into $L^p(\Omega)$ for every $p \in [1, d/(d-1)]$, the embedding being compact for $p < d/(d-1)$. Every bounded sequence $\{u_k\}_k \subset \mathrm{BV}(\Omega)$ admits a subsequence, denoted by the same subscript, which satisfies

$$\lim_{k \to \infty} \|u_k - \bar{u}\|_{L^p(\Omega)} = 0 \quad \text{as well as} \quad Du_k \overset{*}{\rightharpoonup} D\bar{u}$$

for every $p \in [1, d/(d-1))$ and some $\bar{u} \in \mathrm{BV}(\Omega)$. By "$\overset{*}{\rightharpoonup}$" we refer to weak* convergence on the space $M(\Omega; \mathbb{R}^d)$. If we additionally have

$$\lim_{k \to \infty} [\mathrm{TV}(u_k, \Omega) - \mathrm{TV}(u, \Omega)] = 0,$$

the subsequence is called strictly convergent.

Throughout this paper, we interpret $\mathrm{BV}(\Omega)$ as a subspace of $L^q(\Omega)$, $q = d/(d-1)$, and canonically extend $\mathrm{TV}(u, \Omega)$ to $+\infty$ outside of $\mathrm{BV}(\Omega)$. Note that $\mathrm{TV}(u, \Omega)$ is convex and weakly lower semicontinuous on $L^q(\Omega)$. Weak convergence in $L^p(\Omega)$ is denoted by "$\rightharpoonup$".

Furthermore, denoting the Lebesgue measure on $\mathbb{R}^d$ as $\mathcal{L}^d$, we set

$$a_u = \frac{1}{\mathcal{L}^d(\Omega)} \int_\Omega u(x) \, dx \quad \text{as well as} \quad \mathring{u} = u - a_u$$

for every $u \in L^1(\Omega)$ as well as

$$\mathring{V} := \left\{ u \in V \,\middle|\, a_u = 0 \right\} = \left\{ \mathring{u} \,\middle|\, u \in V \right\}$$

and its particular instances

$$\mathring{L}^q(\Omega) := \{\, u \in L^q(\Omega) \,|\, a_u = 0 \,\} = \{\, \mathring{u} \,|\, u \in L^q(\Omega) \,\}\,,$$
$$\mathring{P}_0(\mathcal{T}) := \{\, u \in P_0(\mathcal{T}) \,|\, a_u = 0 \,\} = \{\, \mathring{u} \,|\, u \in P_0(\mathcal{T}) \,\}\,,$$
$$\mathring{P}_1(\mathcal{T}) := \{\, u \in P_1(\mathcal{T}) \,|\, a_u = 0 \,\} = \{\, \mathring{u} \,|\, u \in P_1(\mathcal{T}) \,\}\,.$$

Due to [4, Thm. 3.23] and the weak closedness of V, the unit ball

$$B_{\mathring{V}} = \left\{\, u \in V \,\middle|\, \mathrm{TV}(u, \Omega) \le 1, \;\; \int_\Omega u(x)\,\mathrm{d}x = 0 \,\right\}$$

is compact in the weak $L^q(\Omega)$ topology.

Finally, a subset $E \subset \Omega$ is called a set of finite perimeter if the associated characteristic function $\mathbb{1}_E$ is a function of bounded variation. In this case, we call $\mathrm{Per}(E, \Omega) := \mathrm{TV}(\mathbb{1}_E, \Omega)$ its perimeter. For such sets one can define the reduced boundary $\partial^* E$ as the set of points in $x \in \mathrm{supp}\, D\mathbb{1}_E$ for which

$$\nu_E(x) := \lim_{r \to 0^+} \frac{D\mathbb{1}_E(B(x, r))}{|D\mathbb{1}_E(B(x, r))|}$$

exists and satisfies $|\nu_E(x)| = 1$. Using this notion, the De Giorgi structure theorem [4, Thm. 3.59] provides the expression

$$D\mathbb{1}_E = \nu_E\, \mathcal{H}^{d-1} \mathbin{\llcorner} \partial^* E,$$

where the symbol $\llcorner$ denotes restriction of a measure, that is $(\mu \llcorner A)(B) = \mu(A \cap B)$, and $\mathcal{H}^{d-1}$ is the $(d-1)$-dimensional Hausdorff measure [4, Sec. 2.8]. A set of finite perimeter $E \subseteq \Omega$ is *decomposable* if there exist two nonempty subsets $E_1, E_2 \subseteq \Omega$ such that $E = E_1 \cup E_2$ and $\mathrm{Per}(E, \Omega) = \mathrm{Per}(E_1, \Omega) + \mathrm{Per}(E_2, \Omega)$, and *indecomposable* otherwise. Moreover, we say that $E \subseteq \Omega$ is *simple* if both E and $\Omega \setminus E$ are indecomposable. These notions play the role of connectedness and simple connectedness in the setting of sets of finite perimeter, for which the topological notions are not meaningful. In particular, one can [3, Thm. 1] express such sets as a union of their maximal indecomposable subsets, also referred to as *indecomposable components*.

2 Existence, optimality, and conditional gradient approach

The following properties of the misfit functional F and of the forward operator K are assumed throughout the paper.

Assumption A. *Let Y be a separable Hilbert space with inner product $(\cdot, \cdot)_Y$ and denote by $\|\cdot\|_Y$ the induced norm. We assume that:*

A1 *The operator $K : V \to Y$ is continuous from the weak topology of $L^q(\Omega)$ to the strong topology of Y.*

A2 *F is thrice continuously Fréchet differentiable as well as strictly convex. Moreover, we have $F(y) \ge 0$ for all $y \in Y$*

A3 *For every $y \in Y$ there is a unique $\bar{c}(y) \in \mathbb{R}$ with*

$$F(y + \bar{c}(y)K_{\mathbb{1}}) = \min_{c \in \mathbb{R}} F(y + c(y)K_{\mathbb{1}})$$

where we abbreviate $K_{\mathbb{1}} := K\mathbb{1}_\Omega$.

These requirements are, e.g., fulfilled for the PDE-constrained optimal control problem ($\mathcal{D}$), assuming $d \leq 4$ and setting

$$F(\cdot) = \frac{1}{2\alpha} \| \cdot -y_o \|_{L^2(\Omega)}^2 \quad \text{and} \quad Y = L^2(\Omega).$$

Moreover, under suitable assumptions on the coefficient functions a and b as well as the domain Ω, (1.1) admits a unique solution $y = Ku$ for $u \in L^2(\Omega)$ and the control-to-state operator K is weak-to-strong continuous (see for example [63], and the related considerations in Proposition 5.1 below).

In this section, we prove the existence of at least one minimizer to ($\mathcal{P}$) and state a particular set of necessary and sufficient optimality conditions which are tailored to the algorithmic approach considered in the remainder of the paper. For this purpose, similar to the one dimensional setting in [62], we rely on an equivalent reformulation of Problem ($\mathcal{P}$). More in detail, define

$$\mathcal{F} : Y \to \mathbb{R} \quad \text{where} \quad \mathcal{F}(y) = \min_{c \in \mathbb{R}} F(y + cK\mathbb{1}). \tag{2.1}$$

and consider the auxiliary problem

$$\inf_{v \in \mathring{V}} \mathcal{J}(v) := [\mathcal{F}(Kv) + \mathrm{TV}(v, \Omega)]. \tag{$\mathfrak{P}$}$$

The following proposition collects some of its properties. Its proof is relegated to Appendix A.

Proposition 2.1. *There holds:*

- *The restricted operator $K : \mathring{V} \to Y$ is weak-to-strong continuous.*

- *The mapping $\bar{c} : Y \to \mathbb{R}$ from Assumption A, **A**3, is two times continuously differentiable.*

- *The functional $\mathcal{F}$ is strictly convex and twice continuously Fréchet differentiable. Its gradient $\nabla \mathcal{F}$ is Lipschitz continuous on compact sets and there holds*

$$\nabla \mathcal{F}(y) = \nabla F(y + \bar{c}(y)K\mathbb{1}) \quad \text{for all} \quad y \in Y.$$

Remark 2.2. *Note that the minimization problem in (2.1) is convex. Calculating its first order necessary optimality conditions reveals*

$$\bar{c}(y) = \arg\min_{c \in \mathbb{R}} F(y + cK\mathbb{1}) \Leftrightarrow \int_\Omega K^* \nabla F(y + \bar{c}(y)K\mathbb{1}) = 0.$$

2.1 Existence of minimizers

In order to show the existence of a solution of ($\mathcal{P}$) we rely on the following equivalence result as well as Proposition 2.1.

Proposition 2.3. *The following equivalence holds:*

- *The function $\bar{u}$ is a minimizer of ($\mathcal{P}$).*

- *The function $\mathring{u}$ is a minimizer of ($\mathfrak{P}$) and*

$$a_{\bar{u}} = \arg\min_{c \in \mathbb{R}} F(K\mathring{u} + cK\mathbb{1}), \tag{2.2}$$

that is, $a_{\bar{u}} = \bar{c}(K\mathring{u})$.

Proof. Let $\bar{u} \in V$ and $\bar{v} \in \overset{\circ}{V}$ be arbitrary but fixed. On the one hand, note that we have

$$J(\bar{u}) = F(K\overset{\circ}{\bar{u}} + a_{\bar{u}}K_{\mathbb{1}}) + \text{TV}(\overset{\circ}{\bar{u}}) \geq \mathcal{F}(K\overset{\circ}{\bar{u}}) + \text{TV}(\overset{\circ}{\bar{u}}) = \mathcal{J}(\overset{\circ}{\bar{u}})$$

as well as

$$\mathcal{J}(\overset{\circ}{\bar{u}}) \geq \inf_{u \in V} \mathcal{J}(\overset{\circ}{u}) = \inf_{v \in \overset{\circ}{V}} \mathcal{J}(v).$$

On the other hand, there holds

$$\inf_{u \in V} J(u) \leq \inf_{c \in \mathbb{R}} J(\bar{v} + c\mathbb{1}_{\Omega}) = \mathcal{J}(\bar{v}) \quad \text{and thus} \quad \inf_{u \in V} J(u) \leq \inf_{v \in \overset{\circ}{V}} \mathcal{J}(v).$$

Now, if $\bar{u}$ is a minimizer of $(\mathcal{P})$, these observations imply

$$\min_{u \in V} J(u) = J(\bar{u}) = \mathcal{J}(\overset{\circ}{\bar{u}}) = \min_{v \in \overset{\circ}{V}} \mathcal{J}(v),$$

i.e., $\overset{\circ}{\bar{u}}$ is a minimizer of $(\mathfrak{P})$. Finally, from $J(\bar{u}) = \mathcal{J}(\overset{\circ}{\bar{u}})$ we conclude

$$F(K\bar{u}) = F(K\overset{\circ}{\bar{u}} + a_{\bar{u}}K_{\mathbb{1}}) = \min_{c \in \mathbb{R}} F(K\overset{\circ}{\bar{u}} + cK_{\mathbb{1}}).$$

Thus, we conclude (2.2) from Assumption **A**1. The converse direction follows by the same argument, finishing the proof. $\qquad\square$

Theorem 2.4. *Let Assumption A hold. Then Problem $(\mathcal{P})$ admits at least one minimizer $\bar{u} \in \text{BV}(\Omega)$.*

Proof. Using Proposition 2.1, Problem $(\mathfrak{P})$ can be embedded in the abstract setting of [16]. Thus, existence of at least one minimizer $\overset{\circ}{\bar{u}}$ is concluded from [16, Prop. 2.3]. Finally, applying Proposition 2.3 yields the desired statement. $\qquad\square$

2.2 First order optimality conditions

Next, we state a particular set of necessary and sufficient first order optimality conditions for Problem $(\mathcal{P})$ which will serve as a foundation for its algorithmic solution. In this context, an important role is played by the extremal points of $B_{\overset{\circ}{V}}$.

Definition 2.5. An element $u \in B_{\overset{\circ}{V}}$ is called an extremal point of $B_{\overset{\circ}{V}}$ (or extremal) if there is no $s \in (0,1)$ and $u_1, u_2 \in B_{\overset{\circ}{V}}$, $u_1 \neq u_2$ with $u = (1-s)u_1 + su_2$. The set of all extremal points of $B_{\overset{\circ}{V}}$ is denoted by $\text{Ext}(B_{\overset{\circ}{V}})$.

Since $B_{\overset{\circ}{V}} \neq \emptyset$ is weakly compact, the set $\text{Ext}(B_{\overset{\circ}{V}})$ is nonempty due to Krein-Milman Theorem [60, Thm. 3.23]. Utilizing this concept as well as the one-homogeneity of $\text{TV}(\cdot, \Omega)$, we arrive at the following sufficient first order optimality condition.

Theorem 2.6. *Let $\bar{u} \in V$ be given and define*

$$\bar{p} = -K^* \nabla F(K\bar{u}) \in V'.$$

The following are equivalent:

i) The function $\bar{u}$ is a solution to $(\mathcal{P})$.

ii) The functions $\bar{u} \in V$ and $\bar{p} \in V'$ satisfy

$$\int_{\Omega} \bar{p} \, \mathrm{d}x = 0, \quad \int_{\Omega} \bar{p}\bar{u} \, \mathrm{d}x = \text{TV}(\bar{u}, \Omega), \quad \int_{\Omega} \bar{p}u \, \mathrm{d}x \leq 1.$$

for all $u \in B_{\overset{\circ}{V}}$.

9

iii) The functions $\mathring{\bar{u}} \in \mathring{V}$ and $\bar{p} \in V'$ satisfy

$$\int_\Omega \bar{p} \, \mathrm{d}x = 0, \quad \int_\Omega \bar{p}\mathring{\bar{u}} \, \mathrm{d}x = \mathrm{TV}(\mathring{\bar{u}}, \Omega), \quad \int_\Omega \bar{p}u \, \mathrm{d}x \le 1.$$

for all $u \in \mathrm{Ext}(B_{\mathring{V}})$.

Proof. We proceed similarly to Theorem 2.4 and exploit the equivalence of Problem $(\mathcal{P})$ and Problem $(\mathfrak{P})$ in order to apply the abstract theory of [16]. According to Proposition 2.3, $\bar{u}$ is a solution to $(\mathcal{P})$ if and only if

$$\mathring{\bar{u}} \in \arg\min (\mathfrak{P}), \quad a_{\bar{u}} = \arg\min_{c \in \mathbb{R}} F(K\mathring{\bar{u}} + cK_{\mathbb{1}}).$$

Thus, it suffices to derive first-order necessary and sufficient optimality conditions for $\mathring{\bar{u}}$. Set $\bar{P} = -K^*\nabla\mathcal{F}(K\mathring{\bar{u}})$. Recalling Remark 2.2 and applying [16, Prop. 2.3] yields that $\mathring{\bar{u}}$ is a minimizer of Problem $(\mathfrak{P})$ if and only if

$$\int_\Omega \bar{P}\mathring{\bar{u}} \, \mathrm{d}x = \mathrm{TV}(\mathring{\bar{u}}, \Omega), \quad \max_{u \in \mathrm{Ext}(B_{\mathring{V}})} \int_\Omega \bar{P}\mathring{\bar{u}} \, \mathrm{d}x \le 1, \quad \int_\Omega \bar{P} \, \mathrm{d}x = 0.$$

Finally, from $a_{\bar{u}} = \bar{c}(K\mathring{\bar{u}})$, see Proposition 2.3, as well as the characterization of $\nabla\mathcal{F}$, see Proposition 2.1, we conclude

$$\nabla\mathcal{F}(K\mathring{\bar{u}}) = \nabla F(K\mathring{\bar{u}} + a_{\bar{u}}K_{\mathbb{1}}) = \nabla F(K\bar{u})$$

and thus $\bar{p} = \bar{P}$. Together with $\mathrm{TV}(\mathring{\bar{u}}, \Omega) = \mathrm{TV}(\bar{u}, \Omega)$, this yields $i) \Leftrightarrow ii)$. Finally, the equivalence between $ii)$ and $iii)$ follows from

$$\max_{v \in B_{\mathring{V}}} \int_\Omega \bar{p}v \, \mathrm{d}x = \max_{v \in \mathrm{Ext}(B_{\mathring{V}})} \int_\Omega \bar{p}v \, \mathrm{d}x \quad \text{as well as} \quad \int_\Omega \bar{p}\mathring{u} \, \mathrm{d}x = \int_\Omega \bar{p}u \, \mathrm{d}x$$

for all $u \in \mathrm{BV}(\Omega)$, see [16, Lem. A.1]. $\qquad\square$

2.3 Abstract fully-corrective generalized conditional gradient approach

For the algorithmic computation of a solution to $(\mathcal{P})$, we can again use the equivalence result in Proposition 2.3 and apply the fully-corrective generalized conditional gradient method (FC-GCG) from [16] to the surrogate Problem $(\mathfrak{P})$. In this section, we briefly summarize the abstract algorithm for the problem at hand. For this purpose, and given a finite, ordered set $\mathcal{A} = \{v_j\}_{j=1}^N \subset \mathrm{Ext}(B_{\mathring{V}})$ of extremal points, define the parametrization

$$\mathcal{U}_{\mathcal{A}} : \mathbb{R}^N \to \mathring{V}, \quad \mathcal{U}_{\mathcal{A}}(\gamma) = \sum_{j=1}^N \gamma_j v_j$$

as well as the finite-dimensional minimization problems

$$\bar{\gamma} \in \arg\min_{\gamma \in \mathbb{R}^N, \gamma \ge 0} \left[\mathcal{F}(K\mathcal{U}_{\mathcal{A}}(\gamma)) + \sum_{j=1}^N \gamma_j \right], \tag{2.3}$$

where the inequality constraint is understood componentwise. By definition of $\mathcal{F}$, the pair $\bar{\gamma}$ satisfies (2.3) if and only if

$$(\bar{\gamma}, \bar{c}) \in \arg\min_{\gamma \in \mathbb{R}^N, c \in \mathbb{R}, \gamma \ge 0} \left[F(K\mathcal{U}_{\mathcal{A}}(\gamma) + cK_{\mathbb{1}}) + \sum_{j=1}^N \gamma_j \right]. \tag{$\mathcal{P}(\mathcal{A})$}$$

where $\bar{c} = \bar{c}(K\mathcal{U}_\mathcal{A}(\bar\gamma))$. The FC-GCG method according to [16] relies on the alternating update of a finite active set $\mathcal{A}_k \subset \mathrm{Ext}(B_{\mathring{V}})$ as well as of iterates $\mathring{u}_k \in \mathring{V}$ and $u_k \in V$, respectively, given by

$$\mathring{u}_k = \mathcal{U}_{\mathcal{A}_k}(\gamma^k), \quad u_k = \mathcal{U}_{\mathcal{A}_k}(\gamma^k) + c^k$$

where (γ^k, c^k) is obtained from

$$(\gamma^k, c^k) \in \arg\min(\mathcal{P}(\mathcal{A}_k)).$$

More in detail, the algorithm proceeds as follows: Given the current active set $\mathcal{A}_k$ and iterates $\mathring{u}_k$ and u_k as defined above, we first calculate $p_k = -K^*\nabla\mathcal{F}(K\mathring{u}_k) \in V'$. By Proposition 2.1 and since from the definitions of c^k and u_k we have $c^k = \bar{c}(K\mathring{u}_k)$, we get

$$p_k = -K^*\nabla\mathcal{F}(K\mathring{u}_k) = -K^*\nabla F(K\mathring{u}_k + c^k K_{\mathbb{1}}) = -K^*\nabla F(Ku_k).$$

Moreover, from Remark 2.2, we also conclude that

$$-\int_\Omega K^*\nabla F(K\mathring{u}_k + c^k K_{\mathbb{1}}) \ \mathrm{d}x = \int_\Omega p_k \ \mathrm{d}x = 0.$$

Now, we compute a new candidate extremal point $\widehat{v}_k \in \mathrm{Ext}(B_{\mathring{V}})$ satisfying

$$\int_\Omega p_k \widehat{v}_k \ \mathrm{d}x = \max_{v \in \mathrm{Ext}(B_{\mathring{V}})} \int_\Omega p_k v \ \mathrm{d}x.$$

At this point we can check for convergence, since if we have

$$\int_\Omega p_k \widehat{v}_k \ \mathrm{d}x \le 1,$$

then there also holds

$$\int_\Omega p_k \mathring{u}_k \ \mathrm{d}x = \mathrm{TV}(\mathring{u}, \Omega)$$

see [16, Proposition 3.1], and the method stops, because in this case, by Theorem 2.6 we have that $\mathring{u}_k$ and u_k are minimizers of respectively $(\mathfrak{P})$ and $(\mathcal{P})$. Otherwise, we update the active set and iterates according to

$$\mathcal{A}_k^+ = \mathcal{A}_k \cup \{\widehat{v}_k\}, \quad \mathring{u}_{k+1} = \mathcal{U}_{\mathcal{A}_k^+}(\gamma^{k,+}) + c^{k+1}, \quad u_{k+1} = \mathring{u}_k + c^{k+1},$$

where

$$(\gamma^{k,+}, c^{k+1}) \in \arg\min(\mathcal{P}(\mathcal{A}_k^+)).$$

Finally, the new active set $\mathcal{A}_{k+1}$ is obtained from $\mathcal{A}_k^+$ is pruned by removing extremal points which are assigned a zero weight in $(\mathcal{P}(\mathcal{A}_k))$. Recalling Theorem 2.6, this procedure can be interpreted as a greedy method for the first order condition presented therein: In every iteration, it checks whether the current dual variable p_k satisfies

$$\int_\Omega p_k u \ \mathrm{d}x \le 1 \quad \text{for all} \quad u \in \mathrm{Ext}(B_{\mathring{V}}).$$

If this is not the case, an element $\widehat{v}_k \in \mathrm{Ext}(B_{\mathring{V}})$ which maximizes the violation of this constraint is added to $\mathcal{A}_k$ and the subproblem is solved anew. The abstract method is summarized in Algorithm 1. Since $(\mathfrak{P})$ is merely an auxiliary problem and our main interest lies in computing a solution to Problem $(\mathcal{P})$, both the pseudocode as well as the following theoretical results are only describing the update as well as the converge of the shifted sequence u_k.

We finish this section by summarizing some of its properties. Their proof is again relegated to Appendix A.

Algorithm 1: Abstract FC-GCG for Problem ($\mathcal{P}$)

Input: $u_0 = 0$, $\mathcal{A}_0 = \emptyset$, $N_0 = 0$

calculate $c^0 \in \arg\min_{c \in \mathbb{R}} F(cK\mathbb{1})$

$u_0 \leftarrow c^0 \mathbb{1}_\Omega$

for $k = 0, 1, 2, ...$ **do**

$\quad$ $p_k \leftarrow -K^* \nabla F(Ku_k)$

$\quad$ calculate:
$$\widehat{v}_k \in \arg\max_{v \in \mathrm{Ext}(B_{\mathring{V}})} \int_\Omega p_k v \, \mathrm{d}x$$

$\quad$ **if** $\int_\Omega p_k \widehat{v}_k \, \mathrm{d}x \leq 1$ **then**

$\quad\quad$ | terminate with a solution $\bar{u} = u_k$ to ($\mathcal{P}$)

$\quad$ **end**

$\quad$ $N_k^+ \leftarrow N_k + 1$, $u_{N_k^+}^k \leftarrow \widehat{v}_k$

$\quad$ $\mathcal{A}_k^+ \leftarrow \mathcal{A}_k \cup \{u_{N_k^+}^k\}$

$\quad$ calculate:
$$(\gamma^{k+1}, c^{k+1}) \in \arg\min_{(\gamma, c) \in \mathbb{R}_+^{N_k^+} \times \mathbb{R}} F\left(K\mathcal{U}_{\mathcal{A}_k^+}(\gamma) + cK\mathbb{1}\right) + \sum_{i=1}^{N_k^+} \gamma_i$$

$\quad$ $u_{k+1} \leftarrow \mathcal{U}_{\mathcal{A}_k^+}(\gamma^{k+1}) + c^{k+1}\mathbb{1}_\Omega = \sum_{i=1}^{N_k^+} \gamma_i^{k+1} u_i^k + c^{k+1}\mathbb{1}_\Omega$

$\quad$ $\mathcal{A}_{k+1} \leftarrow \mathcal{A}_k^+ \setminus \{u_i^k : \gamma_i^{k+1} = 0\}$

$\quad$ $N_{k+1} \leftarrow \#\mathcal{A}_{k+1}$

end

Proposition 2.7. *Let u_k, $k \in \mathbb{N}$, be generated by Algorithm 1. Then the method either stops after $\bar{k}$ steps with a minimizer $u_{\bar{k}}$ to Problem ($\mathcal{P}$) or there holds*

$$J(u_k) - \min_{u \in L^q(\Omega)} J(u) \leq \frac{c}{1+k}.$$

for some $c > 0$ and all $k \in \mathbb{N}$. Moreover, in this case, the sequence u_k admits at least one strictly convergent subsequence and every strict accumulation point is a minimizer of ($\mathcal{P}$).

Lemma 2.8. *Let $u_k = \mathcal{U}_{A_k}(\gamma^k) + c^k \mathbb{1}_\Omega$, $k \in \mathbb{N}$, be generated by Algorithm 1 and set*

$$M_k := \left(F\left(K\mathcal{U}_{\mathcal{A}_k}(\gamma^k) + c^k K\mathbb{1}\right) + \sum_{j=1}^{N_k} \gamma_j^k \right).$$

Then there holds

$$J(u_k) - \min_{u \in V} J(u) \leq \zeta_k := M_k \left(\int_\Omega p_k \widehat{v}_k \, \mathrm{d}x - 1 \right).$$

Remark 2.9. *By inspecting the respective proofs in [16], the results of Proposition 2.7 and Lemma 2.8 remain valid if $\widehat{v}_k$ is not an extremal point but satisfies $\widehat{v}_k \in B_{\mathring{V}}$ as well as*

$$\int_\Omega p_k \widehat{v}_k \, \mathrm{d}x = \max_{v \in \mathrm{Ext}(B_{\mathring{V}})} \int_\Omega p_k v \, \mathrm{d}x.$$

The incentive behind minimizing over $\mathrm{Ext}(B_{\mathring{V}})$ instead of the whole ball $B_{\mathring{V}}$ is that, once the extremal points are characterized, additional structural properties of this "simpler" set can be utilized to facilitate the algorithmic solution of the linear subproblems.

Remark 2.10. *It is worth pointing out that the convexity and one-homogeneity of* $\mathrm{TV}(\cdot, \Omega)$ *also imply*

$$J(\mathcal{U}_{\mathcal{A}_k}(\gamma)) \leq F(\mathcal{U}_{\mathcal{A}_k}(\gamma)) + \sum_{j=1}^{N_k} \gamma_j \quad \text{for all} \quad \gamma \in \mathbb{R}^{N_k}, \tag{2.4}$$

but the inequality can be strict for some γ. *In particular, Algorithm 1 is not a descent method, i.e., it does not ensure* $J(u_{k+1}) \leq J(u_k)$ *in general. In the recent paper [38], the authors consider a similar method based on the greedy insertion of extremal points for functions of bounded variation on unbounded domains, i.e.,* $V = \mathrm{BV}(\mathbb{R}^2)$. *Different to the current work, they rely on updating the iterate following*

$$u_{k+1} \in \underset{u \in \mathrm{span}(\mathcal{A}_k)}{\arg\min} \; J(u) = F(Ku) + \mathrm{TV}(u, \Omega).$$

This naturally ensures motonicity as well as a sublinear rate of convergence. However, ensuring the full resolution of the TV*-regularized subproblem is significantly more challenging than just using the* ℓ_1 *problem considered in the current work. Moreover, in our numerical examples presented in Section 6 the computed iterates always satisfied* (2.4) *with equality.*

3 Extremal points of total variation balls

From the discussion in the previous section, it becomes apparent that a practical realization of the abstract Algorithm 1, first and foremost, requires a characterization of the set of extremal points $\mathrm{Ext}(B_{\mathring{V}})$. While this can be directly concluded from known results for $\mathring{V} = \mathring{L}^q(\Omega)$, see Section 3.1, to the best of our knowledge, there are no explicitly stated results concerning the finite element spaces $\mathring{V} = \mathring{P}_0(\mathcal{T})$ and $\mathring{V} = \mathring{P}_1(\mathcal{T})$, respectively. Since many interesting instances of Problem ($\mathcal{P}$) inherently require a discretization of the variable space, we take a step at closing this gap in the following section. On the one hand, we give a precise characterization of $\mathrm{Ext}(B_{\mathring{P}_0})$ and show that piecewise constant discretizations are conforming with the convex geometry of $B_{\mathring{L}^q}$ in the sense that $\mathrm{Ext}(B_{\mathring{P}_0})$ is a finite set and $\mathrm{Ext}(B_{\mathring{P}_0}) \subset \mathrm{Ext}(B_{\mathring{L}^q})$. On the other hand, we provide a "negative" result on $\mathrm{Ext}(B_{\mathring{P}_1})$ showing that it contains infinitely many functions, albeit we do not provide of a full characterization of this set.

3.1 Continuous case with Neumann boundary

For $\mathring{V} = \mathring{L}^q(\Omega)$, the characterization of $\mathrm{Ext}(B_{\mathring{V}})$ follows from the results in [13] by exploiting the invariance of extremal points under embeddings.

Proposition 3.1. *There holds*

$$\mathrm{Ext}\left(B_{\mathring{L}^q}\right) = \left\{ \left. \frac{1}{\mathrm{Per}(E, \Omega)} \mathring{\mathbb{1}}_E \; \right| E \text{ simple} \right\}.$$

Proof. For $u \in \mathrm{BV}(\Omega)$, define the equivalence class

$$[u] = \{ u + c\mathbb{1}_\Omega \,|\, c \in \mathbb{R} \} \in \mathrm{BV}(\Omega)/\mathcal{C} \quad \text{where} \quad \mathcal{C} = \{ c\mathbb{1}_\Omega \,|\, c \in \mathbb{R} \}.$$

We equip the quotient space $\mathrm{BV}(\Omega)/\mathcal{C}$ with the canonical quotient norm

$$\|[u]\|_{\mathrm{BV}(\Omega)/\mathcal{C}} = \inf_{v \in \mathcal{C}} \|u + v\|_{\mathrm{BV}} = \inf_{v \in \mathcal{C}} \left[\|u + v\|_{L^1(\Omega)} + \mathrm{TV}(u, \Omega) \right].$$

Mutatis mutandis, we define the extremal points $\mathrm{Ext}(\mathcal{B})$ of the set

$$\mathcal{B} = \{ [u] \,|\, \mathrm{TV}(u, \Omega) \leq 1 \} \subset \mathrm{BV}(\Omega)/\mathcal{C}.$$

According to [13, Thm. 4.7], there holds

$$\text{Ext}(\mathcal{B}) = \left\{ \left[\frac{1}{\text{Per}(E,\Omega)} \mathbb{1}_E \right] \,\middle|\, E \text{ simple} \right\}.$$

Consider the mapping $L : \text{BV}(\Omega)/\mathcal{C} \to \mathring{L}^q(\Omega)$ with $L([u]) = \mathring{u}$. By definition, L is linear and there holds $L(\mathcal{B}) = L(B_{\mathring{L}^q})$. Moreover, L is continuous due to to

$$\|\mathring{u}\|_{L^q(\Omega)} = \|L([u])\|_{L^q(\Omega)} \leq C \, \text{TV}(u,\Omega) \leq C\|[u]\|_{\text{BV}(\Omega)/\mathcal{C}},$$

see, e.g., [4, Thm. 3.44], as well as injective. Summarizing these observations,[13, Lem. 3.2] yields

$$\text{Ext}\left(B_{\mathring{L}^q}\right) = \text{Ext}(L(\mathcal{B})) = L(\text{Ext}(\mathcal{B}))$$

and thus the desired statement. $\qquad\square$

3.2 Discretization with piecewise constant functions

In the following section, let $\Omega \subseteq \mathbb{R}^d$ be an open and bounded polyhedron and let $\mathcal{T}$ be a triangulation thereof, that is, a collection of open simplices

$$\mathcal{T} = \{T_1, \ldots, T_{n_\mathcal{T}}\} \quad \text{such that} \quad \overline{\Omega} = \bigcup_{i=1}^{n_\mathcal{T}} \overline{T_i}.$$

Definition 3.2. We define the class of triangulated subsets of Ω with respect to the triangulation $\mathcal{T}$, denoted as $\mathcal{S}_\mathcal{T}(\Omega)$, to be the open sets obtained as the interior of a union of closures of simplices belonging to $\mathcal{T}$. For such sets $E_1, E_2 \in \mathcal{S}_\mathcal{T}(\Omega)$, we implicitly redefine unions and set differences so that $E_1 \cup E_2, E_1 \setminus E_2, E_1 \triangle E_2 \in \mathcal{S}_\mathcal{T}(\Omega)$.

If not otherwise specified, any subset $E \subseteq \Omega$ considered in this section belongs to $\mathcal{S}_\mathcal{T}(\Omega)$. Moreover, for a subset $E \in \mathcal{S}_\mathcal{T}(\Omega)$ we have that $\text{Per}(E,\Omega) = \mathcal{H}^{d-1}(\partial E \cap \Omega)$. Let $P_0(\mathcal{T})$ be the set of piecewise constant functions on Ω with respect to the triangulation $\mathcal{T}$, and $\mathring{P}_0(\mathcal{T})$ its subspace of zero average. Inside $\mathring{P}_0(\mathcal{T})$, define the unit ball with respect to the total variation:

$$B_{\mathring{P}_0} := \left\{ u \in \mathring{P}_0(\mathcal{T}) : \text{TV}(u,\Omega) \leq 1 \right\}.$$

There is a convenient description of the total variation in this discrete setting.

Lemma 3.3. *For any $u \in P_0(\mathcal{T})$, the total variation corresponds to*

$$\text{TV}(u,\Omega) = \frac{1}{2} \sum_{i,j} |u(T_i) - u(T_j)| \mathcal{H}^{d-1}(\partial T_i \cap \partial T_j)$$

Proof. This result follows by standard facts about functions of bounded variation, see for example Section 4.2 and in particular Theorem 4.23 in [4]. It is also explicitly stated in [25, Thm. 3.1] and [5, Lem. 4.1]. $\qquad\square$

Proposition 3.4. *There holds*

$$\text{Ext}\left(B_{\mathring{P}_0}\right) = \left\{ \frac{1}{\text{Per}(E,\Omega)} \mathring{\mathbb{1}}_E \,\middle|\, E \in \mathcal{S}_\mathcal{T}(\Omega) \text{ simple} \right\}.$$

Proof. The proof of this lemma is structured in 3 cases. First, we prove that piecewise constant functions with at least three values in the image are not extremal. Then, in case 2, we do the same for characteristic functions on indecomposable but not simple domains. Finally, in case 3, we prove that, among the remaining functions, the extremal points of $B_{\mathring{P}_0}$ are characteristic functions on simple domains.

The proof of each case follows a standard procedure using perturbations that can be found, for example, in [2]. Namely, for some $u \in \partial B_{\mathring{P}_0}$, set

$$\delta_u := \min\{|u(T_i) - u(T_j)| : \mathcal{H}^{d-1}(\partial T_i \cap \partial T_j) > 0, u(T_i) \neq u(T_j)\},$$

which corresponds to the minimum non zero variation of u between two adjacent simplices. Consider a perturbation $h \in \mathring{P}_0(\mathcal{T})$ satisfying the following property: for every pair of adjacent simplices T_i, T_j, there exists $\lambda_{ij} \in \mathbb{R}$ such that

$$h(T_i) - h(T_j) = \lambda_{ij}(u(T_i) - u(T_j)). \tag{3.1}$$

This condition requires that if u has no variation between adjacent simplices, neither does h. Whenever $h(T_i) = h(T_j)$, we set $\lambda_{ij} = 0$. This condition limits the set of possible perturbations that we can choose from, but there are usually many that do satisfy this property, e.g. $h \in \text{span}(u)$. We will always exclude the trivial case $h = 0$, so that we can define

$$\Delta_h := \max\{|h(T_i) - h(T_j)| : \mathcal{H}^{d-1}(\partial T_i \cap \partial T_j) > 0\},$$

which is strictly positive since $h \notin \mathcal{C}$ by assumption. Choose an $\epsilon > 0$ smaller than δ_u/Δ_h, so that, by construction, $\epsilon|\lambda_{ij}| \leq 1$ for every i, j. Then, the crucial part of each step will be finding the right perturbation h for which $\text{TV}(u \pm \epsilon h, \Omega) \neq 0$. If that is the case, we then define

$$u^+ := \frac{u + \epsilon h}{\text{TV}(u + \epsilon h, \Omega)} \in B_{\mathring{P}_0} \qquad \text{and} \qquad u^- := \frac{u - \epsilon h}{\text{TV}(u - \epsilon h, \Omega)} \in B_{\mathring{P}_0},$$

and, by construction, it is possible to prove that u is a convex combination of u^+ and u^-. Indeed,

$$\text{TV}(u + \epsilon h, \Omega) + \text{TV}(u - \epsilon h, \Omega)$$
$$= \frac{1}{2}\sum_{i,j}\mathcal{H}^{d-1}(\partial T_i \cap \partial T_j)(|(1 + \epsilon\lambda_{ij})(u(T_i) - u(T_j))| + |(1 - \epsilon\lambda_{ij})(u(T_i) - u(T_j))|)$$
$$= \frac{1}{2}\sum_{i,j}\mathcal{H}^{d-1}(\partial T_i \cap \partial T_j)((1 + \epsilon\lambda_{ij})|u(T_i) - u(T_j)| + (1 - \epsilon\lambda_{ij})|u(T_i) - u(T_j)|)$$
$$= 2\,\text{TV}(u, \Omega) = 2.$$

Of course u^+ and u^- could coincide, but this is only true if $h \in \text{span}(u)$. So in conclusion, if there exists a zero average perturbation h that satisfies equation (3.1), such that $h \notin \text{span}(u)$ and $\text{TV}(u \pm \epsilon h, \Omega) \neq 0$ for a sufficiently small $\epsilon > 0$, then u is not extremal.

Case 1: Let $u \in \partial B_{\mathring{P}_0}$ be a piecewise constant function, whose image contains at least three values $\{\beta_1, \beta_2, \beta_3\}$. Without loss of generality, assume $\beta_1 \neq 0$, consider an indecomposable component $C_1 \in S_{\mathcal{T}}(\Omega)$ of the preimage $u^{-1}(\beta_1)$, and define

$$h := \beta_1 \mathring{\mathbb{1}}_{C_1}.$$

This perturbation satisfies equation 3.1, and for $0 < \epsilon < \delta_u/\Delta_h$, we get

$$\text{TV}(u \pm \epsilon h, \Omega) = \text{TV}(u, \Omega \setminus \overline{C_1}) + \sum_{T_i \subseteq C_1}\sum_{T_j \not\subseteq C_1}|(1 \pm \epsilon)\beta_1 - u(T_j)|\mathcal{H}^{d-1}(\partial T_i \cap \partial T_j)$$

Since the image of u has at least 3 values and C_1 is in the preimage of one of them, then the first term in the sum is always non zero. Therefore, u is not extremal.

Case 2: Let $u := \beta \mathring{\mathbb{1}}_E \in \partial B_{\mathring{P}_0}$ be a characteristic function over an indecomposable but not simple domain. Let $\{C_i\}_{i=1}^n$ be indecomposable components of $\Omega \setminus E$. By hypothesis, we know that $n \geq 2$, so define
$$h := \beta \mathring{\mathbb{1}}_{C_1}.$$
It clearly satisfies equation 3.1. Moreover, for any $0 < \epsilon \leq \delta_u / \Delta_h$, we get
$$\mathrm{TV}(u \pm \epsilon h, \Omega) = \mathrm{TV}(u, \Omega \setminus \overline{C_1}) + \sum_{T_i \subseteq C_1} \sum_{T_j \not\subseteq C_1} |(1 \pm \epsilon)\beta - u(T_j)| \mathcal{H}^{d-1}(\partial T_i \cap \partial T_j)$$
where the first term cannot be zero because u has variation at ∂C_2. Therefore u is not extremal.

Case 3: Let $u := \beta \mathring{\mathbb{1}}_E \in \partial B_{\mathring{P}_0}$ be the characteristic function of a simple domain E, and assume by contradiction that u is not an extremal point of $B_{\mathring{P}_0}$. Then, there exists a perturbation $h \in \mathring{P}_0(\mathcal{T})$ for which $\mathrm{TV}(u \pm \epsilon h) \neq 0$ for any $\epsilon > 0$ sufficiently small and satisfying equation (3.1), which implies that h has variation only in ∂E. This would mean that $h \in \mathrm{span}(u)$, which is a contradiction. Note also that
$$\mathrm{TV}(\beta \mathring{\mathbb{1}}_E, \Omega) = |\beta| \mathrm{Per}(E, \Omega) = 1 \qquad \Longrightarrow \qquad \beta = \pm \frac{1}{\mathrm{Per}(E, \Omega)}.$$

Since we are working with zero average characteristic functions on simple domains, we have
$$\frac{1}{\mathrm{Per}(E, \Omega)} \mathring{\mathbb{1}}_E = -\frac{1}{\mathrm{Per}(\Omega \setminus E, \Omega)} \mathring{\mathbb{1}}_{\Omega \setminus E},$$
which concludes the proof. $\qquad \square$

We remark that Proposition 3.4 above directly mimics the continuous case of extremal points of the TV ball in the whole of $\mathrm{BV}(\Omega)$, first proved in [42] (see [3] for a modern exposition) for $\Omega = \mathbb{R}^d$ and extended in [13] to bounded Ω.

3.3 Discretization with continuous piecewise affine functions

Let $\mathcal{T}$ be a triangulation in an open and bounded polyhedron $\Omega \subset \mathbb{R}^2$, and let $P_1(\mathcal{T})$ be the set of continuous piecewise affine functions on Ω. Analogously to the previous section, we aim to study the extremal points of
$$B_{\mathring{P}_1} = \left\{ f \in \mathring{P}_1(\mathcal{T}) \,|\, \mathrm{TV}(f, \Omega) \leq 1 \right\}.$$

We also have a convenient description of the total variation in this setting:

Lemma 3.5. *For any $u \in P_1(\mathcal{T})$, the total variation corresponds to*
$$\mathrm{TV}(u, \Omega) = \sum_{T \in \mathcal{T}} \mathcal{L}^2(T) |\nabla u|_T |,$$
where since $\nabla u|_T$ is constant for all T we have identified these with fixed vectors, and $|\cdot|$ is the Euclidean norm in $\mathbb{R}^d$.

Proof. A continuous piecewise affine function belongs in particular to $W^{1,p}(\Omega)$ for all $1 \leq p \leq \infty$. This implies [4, Sec. 3.1] that
$$\mathrm{TV}(u, \Omega) = \int_\Omega |\nabla u| \, \mathrm{d}x,$$
for the weak derivative $\nabla u \in L^\infty(\Omega)$, for which we have $\nabla u \in [P_0(\mathcal{T})]^d$. $\qquad \square$

An analytical study of the extremal points of $B_{\mathring{P}_1}$ needs to take into account the continuity constraints of the piecewise affine functions. Any variation of a function $u \in P_1(\mathcal{T})$ at a vertex of the triangulation changes the value of ∇u at all triangles that contains that vertex. For this reason, we simplify our study to the subset of 'double tent functions' as in Figure 1, from which we draw conclusions on the finiteness of the set of extremal points of $\mathring{P}_1(\mathcal{T})$. To do so, we need two preliminary Lemmas.

Lemma 3.6. *A function $u \in P_1(\mathcal{T})$ is completely described by its values at the nodes of the triangulation, i.e.*

$$P_1(\mathcal{T}) = \operatorname{span}\{g_z \mid z \in V(\mathcal{T})\},$$

where g_z is the continuous piecewise affine taking the value 1 function at the vertex z and vanishing at all other vertices, often called tent function.

Proof. See for example [18, Ch. 3]. □

Lemma 3.7. *Let Q be a polyhedron in $\mathbb{R}^n$. Then $Q \cap W$ is still a polyhedron for any k-dimensional plane $W \subset \mathbb{R}^n$.*

Proof. Q is defined as the intersection of a finite number of half-spaces in $\mathbb{R}^n$. Each half-space is defined as the set of points $x \in \mathbb{R}^n$ that satisfy a linear inequality of the form

$$\nu \cdot x \leq b,$$

where $\nu \in \mathbb{R}^n$ is a vector, and $b > 0$ is a constant. A posteriori, ν is the outward normal vector of a $(n-1)$-dimensional facet of Q, and b is the affine coordinate of the plane containing the facet. The set $Q \cap W$ consists of the points $x \in W$ that satisfy

$$\pi_W(\nu) \cdot x \leq b,$$

where ν and b are as above, and π_W is the orthogonal projection onto W. This set corresponds to a polyhedron in $\mathbb{R}^k$ through any linear isomorphism between W and $\mathbb{R}^k$. □

Let T_1 and T_2 be two adjacent triangles in $\mathcal{T}$, and z_1, z_2 be the two common vertices. Then, assume that, locally around z_1, z_2, the triangulation is regular and equilateral of area $\lambda > 0$, as depicted in Figure 1. We can classify the neighbor triangles based on which vertices they
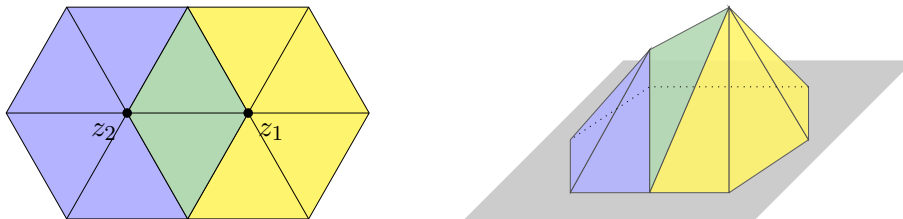

Figure 1: Support and graph of a $P_1(\mathcal{T})$ function with two contiguous nonzero vertex values.

depend on. Let C_2 be the set of the triangles that do not have z_1 as a vertex (blue in Figure 1), and C_1 be the set of triangles without z_2 as a vertex (yellow in Figure 1). Let's denote by $C_{1,2}$ the set of the remaining two triangles, i.e. $C_{1,2} = \{T_1, T_2\}$. Consider the set

$$W := \operatorname{span}\{g_{z_1}, g_{z_2}\},$$

which consists of the piecewise affine functions that have non-zero nodal values only at z_1 or z_2.
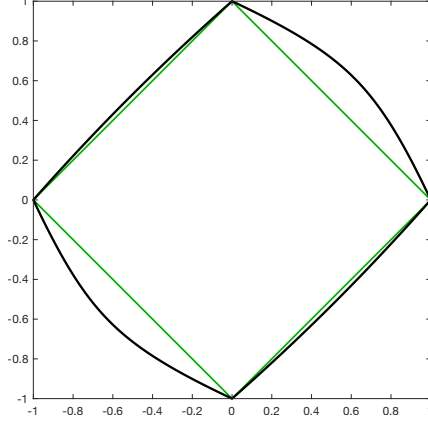
Figure 2: The set $\partial\mathfrak{B}_{1,2}$ for $\lambda = (4\sqrt{3} + \sqrt{11})^{-1}$ in black, and the ℓ^1 ball in $\mathbb{R}^2$ in green.

Lemma 3.8. *The total variation in W is*

$$\mathrm{TV}(c_1 g_{z_1} + c_2 g_{z_2}, \Omega) = 4\sqrt{3}\lambda|c|_1 + \lambda|c|_A, \quad for \quad |c|_A = \sqrt{c^\top A c} \quad with \quad A = \begin{pmatrix} 11 & -7 \\ -7 & 11 \end{pmatrix}$$

for every $c = (c_1, c_2) \in \mathbb{R}^2$.

Proof. Direct computation splitting the total variation into the components C_1, C_2 and $C_{1,2}$, i.e.

$$\mathrm{TV}(u, \Omega) = \lambda \sum_{T \in C_1} |\nabla u|_T| + \lambda \sum_{T \in C_2} |\nabla u|_T| + \lambda \sum_{T \in C_{1,2}} |\nabla u|_T|. \qquad \square$$

Based on Lemma 3.7, we conclude the following result.

Proposition 3.9. *The ball $B_{\mathring{P}_1}$ has infinitely many extremal points.*

Proof. First, note that for every $c = (c_1, c_2) \in \mathbb{R}^2$, using Lemma 3.8 we can write

$$h(c) := \mathrm{TV}(c_1 g_{z_1} + c_2 g_{z_2}, \Omega) = 4\sqrt{3}\lambda|c|_1 + \lambda|c|_A, \quad for \quad |c|_A = \sqrt{c^\top A c} \quad with \quad A = \begin{pmatrix} 11 & -7 \\ -7 & 11 \end{pmatrix}$$

a positive definite matrix. Define $\mathcal{B}_{1,2} := B_{\mathring{P}_1} \cap W$ as well as

$$\mathfrak{B}_{1,2} := \{\, c \in \mathbb{R}^2 \mid h(c) \le 1 \,\},$$

with the latter depicted in Figure 2. Let us show that $\mathrm{Ext}(\mathfrak{B}_{1,2}) = \partial\mathfrak{B}_{1,2}$. For this purpose, it suffices to show that every boundary point is extremal, since the converse inclusion is trivial. Indeed, assume that there is $c_s, c_a, c_b \in \mathbb{R}^2$ as well as $s \in (0,1)$ with

$$h(c_a) = h(c_a) = h(c_s) = 1, \quad c_s = (1-s)c_a + sc_b.$$

Then there also holds

$$|c_s|_A = (1-s)|c_a|_A + s|c_b|_A.$$

Since $|\cdot|_A$ is a weighted euclidean norm and $s \in (0,1)$, this implies that $c_a = tc_b$ for some $t > 0$. Due to $h(c_a) = h(c_b) = 1$ as well as the positive 1-homogeneity of h, we finally conclude $c_a = c_b = c_s$, hence, c_s is an extremal point of $\mathfrak{B}_{1,2}$. As a consequence, neither $\mathfrak{B}_{1,2}$ nor $\mathcal{B}_{1,2}$ are polytopes. In view of Lemma 3.7, this proves that $B_{\mathring{P}_1}$ has infinitely many extremal points. $\square$

18

4 Practical realization of Algorithm 1

Once the extremal points of $B_{\mathring{V}}$ are characterized, their structure can be exploited in order to devise efficient solution methods for the linear subproblems in Algorithm 1. We again point out [38] where the same philosophy is applied for $\Omega = \mathbb{R}^d$ and "simple" forward operators K induced by explicitly given convolution kernels. Since our main interest lies in more complicated examples such as the PDE-constrained optimization problem $(\mathcal{D})$, which inherently require a discretization of the space variable, and motivated by Section 3.2, we will further elaborate on the practical realization of Algorithm 1 for piecewise constant discretizations, i.e., $V = P_0(\mathcal{T})$. In this case, relying on Proposition 3.4, the linear subproblem is solved by

$$\widehat{v}_k = \mathbb{1}_{E_k} / \operatorname{Per}(E_k, \Omega) \quad \text{where} \quad E_k \in \underset{\substack{E \in \mathcal{S}_{\mathcal{T}}(\Omega) \\ E \text{ simple}}}{\arg\max} \frac{\int_E p_k \, \mathrm{d}x}{\operatorname{Per}(E, \Omega)}.$$

While the admissible set of the latter is finite, the restriction to simple sets significantly aggravates its efficient solution. For this reason, we drop this requirement and consider the relaxed problem

$$E_k \in \underset{E \in \mathcal{S}_{\mathcal{T}}(\Omega)}{\arg\max} \frac{\int_E p_k \, \mathrm{d}x}{\operatorname{Per}(E, \Omega)} \tag{4.1}$$

for the computation of $\widehat{v}_k$. Due to

$$\operatorname{Ext}(B_{\mathring{P}_0}) \subset \left\{ \, \mathring{\mathbb{1}}_E / \operatorname{Per}(E, \Omega) \mid E \in S_{\mathcal{T}}(\Omega) \, \right\} \subset B_{\mathring{P}_0}$$

and Remark 2.9, this relaxation does neither impact the convergence of the algorithm, Proposition 2.7, nor the upper bound in Lemma 2.8. In the following sections, we show that this, still challenging, fractional minimization can be replaced by a finite number of easier-to-solve subproblems which can be exactly solved by efficient graph-based algorithms.

4.1 Dinkelbach-Newton method for the insertion step

In the following, let u_k denote the k−th iterate generated by Algorithm 1. This implies that the associated dual variable $p_k = -K^* \nabla F(K u_k)$ satisfies

$$\bar{\lambda}_k := \left(\max_{E \subset \Omega} \frac{\int_E p_k \, \mathrm{d}x}{\operatorname{Per}(E, \Omega)} \right)^{-1} \geq 1. \tag{4.2}$$

Furthermore, for $\lambda \geq 0$, consider the regularized problem

$$\min_{E \in S_{\mathcal{T}}(\Omega)} j_\lambda(E) := \left[\operatorname{Per}(E, \Omega) - \lambda \int_E p_k \, \mathrm{d}x \right] \tag{P_λ}$$

as well as its associated value function

$$G : [0, \infty) \to \mathbb{R}, \ \lambda \mapsto \min (P_\lambda).$$

Note that (P_λ) is well-posed since the set $S_{\mathcal{T}}(\Omega)$ is finite. Denote by $E_\lambda \in S_{\mathcal{T}}(\Omega)$ an arbitrary solution. The following proposition formalizes the connection between (P_λ) and the fractional maximization problem (4.1).

Proposition 4.1. *We have that:*

- *There holds $G(\lambda) \leq 0$ for all $\lambda \geq 0$, with equality if and only if $\lambda \leq \bar{\lambda}_k$.*

- *For $\lambda < \bar{\lambda}_k$, the solutions of (P_λ) are given by $E_\lambda = \emptyset$ and $E_\lambda = \Omega$.*

19

- A set $\bar{E} \in \mathcal{S}_\mathcal{T}(\Omega)$, $\mathrm{Per}(\bar{E}) \neq 0$, is a solution of (P_λ) for $\lambda = \bar{\lambda}_k$ if and only if it is a maximizer of (4.1).

Proof. Since $\emptyset \in \mathcal{S}_\mathcal{T}(\Omega)$, we immediately conclude $G(\lambda) \leq 0$. Now, if $\lambda \leq \bar{\lambda}_k$, for all $E \in \mathcal{S}_\mathcal{T}(\Omega)$ with $\mathrm{Per}(E, \Omega) \neq 0$ there holds

$$\lambda^{-1} \geq \frac{\int_E p_k \, \mathrm{d}x}{\mathrm{Per}(E, \Omega)} \quad \text{and thus} \quad \mathrm{Per}(E, \Omega) - \lambda \int_E p_k \, \mathrm{d}x \geq 0.$$

Noting that since $\int_\Omega p_k \, \mathrm{d}x = 0$ we have $j_\lambda(\emptyset) = j_\lambda(\Omega) = 0$, we get $G(\lambda) = 0$ if $\lambda \leq \bar{\lambda}_k$. From this observation, we also conclude the second bullet point. Now assume that $\lambda > \bar{\lambda}_k$ and let $\bar{E} \in \mathcal{S}_\mathcal{T}(\Omega)$, $\mathrm{Per}(\bar{E}) \neq 0$, denote a maximizer of (4.1). Then there holds

$$0 > j_\lambda(\bar{E}) \geq G(\lambda)$$

which finishes the proof of the first bullet point. Now, fix $\lambda = \bar{\lambda}_k$ and let $\bar{E} \in \mathcal{S}_\mathcal{T}(\Omega)$, $\mathrm{Per}(\bar{E}) \neq 0$, denote a solution to (4.1). Then, by construction, there holds $j_\lambda(\bar{E}) = 0$. Thus, $\bar{E}$ is a minimizer of (4.1). Conversely, let $\bar{E} \in \mathcal{S}_\mathcal{T}(\Omega)$, $\mathrm{Per}(\bar{E}) \neq 0$, denote a solution of (P_λ) for $\lambda = \bar{\lambda}_k$. Then there holds

$$j_\lambda(\bar{E}) = -\lambda \int_{\bar{E}} p_k \, \mathrm{d}x + \mathrm{Per}(\bar{E}, \Omega) = 0 \quad \text{as well as} \quad \int_{\bar{E}} p_k \, \mathrm{d}x > 0$$

Reordering yields

$$\bar{\lambda}_k = \lambda = \frac{\mathrm{Per}(\bar{E}, \Omega)}{\int_{\bar{E}} p_k \, \mathrm{d}x} \quad \text{and thus} \quad \bar{E} \in \arg\max_{E \in \mathcal{S}_\mathcal{T}(\Omega)} \frac{\int_E p_k \, \mathrm{d}x}{\mathrm{Per}(E, \Omega)}. \qquad \square$$

As a consequence, solving the fractional maximization problem (4.1) is equivalent to computing a minimizer of (P_λ) with non-vanishing perimeter for $\lambda = \bar{\lambda}_k$. Since the latter is unknown, this requires two subroutines: first, a method to compute the largest zero of the value function G on $[0, \infty)$ and, second, an efficient algorithm for the solution of the discrete minimization problem (P_λ). We focus on the former in the following and derive a generalized Newton method which exhibits convergence in finitely many iterations. The efficient solution of (P_λ) is postponed to Section 4.1.1. First, we require the following additional properties of the value function.

Lemma 4.2. *The value function G is concave and locally Lipschitz continuous. For every $\lambda > 0$ there holds*

$$dG(\lambda, +) \leq - \int_{E_\lambda} p_k \, \mathrm{d}x \leq -dG(\lambda, -) \quad \text{for all} \quad E_\lambda \in \arg\min(P_\lambda). \tag{4.3}$$

where $dG(\lambda, \pm)$ denote the one-sided directional derivative of G at λ in the direction of ± 1, respectively.

Proof. Let $\lambda_1, \lambda_2 \geq 0$ and $\theta \in [0, 1]$ be arbitrary but fixed. Set $\lambda_\theta = (1 - \theta)\lambda_1 + \theta\lambda_2$. By definition of the value function G and that $G(\lambda_\theta)$ is attained at E_{λ_θ} we conclude

$$
\begin{aligned}
(1 - \theta)G(\lambda_1) + \theta G(\lambda_2) &= (1 - \theta)j_{\lambda_1}(E_{\lambda_1}) + \theta j_{\lambda_2}(E_{\lambda_2}) \\
&\leq (1 - \theta)j_{\lambda_1}(E_{\lambda_\theta}) + \theta j_{\lambda_2}(E_{\lambda_\theta}) \\
&= \mathrm{Per}(E_{\lambda_\theta}, \Omega) - \lambda_\theta \int_{E_{\lambda_\theta}} p_k \, \mathrm{d}x = j_{\lambda_\theta}(E_{\lambda_\theta}) = G(\lambda_\theta).
\end{aligned}
$$

Hence G is concave and thus locally Lipschitz continuous. It remains to prove (4.3). On the one hand, for $\delta > 0$ small enough, have

$$\frac{1}{\delta}(G(\lambda + \delta) - G(\lambda)) = \frac{1}{\delta}(j_{\lambda+\delta}(E_{\lambda+\delta}) - j_\lambda(E_\lambda)) \leq \frac{1}{\delta}(j_{\lambda+\delta}(E_\lambda) - j_\lambda(E_\lambda)) = -\int_{E_\lambda} p_k \, \mathrm{d}x.$$

On the other hand, we can estimate

$$-\int_{E_\lambda} p_k \, \mathrm{d}x = \frac{1}{\delta}(j_{\lambda+\delta}(E_\lambda) - j_\lambda(E_\lambda)) = \frac{1}{\delta}(j_\lambda(E_\lambda) - j_{\lambda-\delta}(E_\lambda)) \leq -\frac{1}{\delta}(G(\lambda - \delta) - G(\lambda))$$

Taking the limit for $\delta \to 0$ in both inequalities finally yields the desired statement by definition of the directional derivatives, which always exist by virtue of G being concave. $\qquad\square$

We draw several conclusions from this auxiliary result. First, the value function G is not differentiable at $\lambda = \bar{\lambda}_k$. Indeed, denoting by E_k a maximizer of (4.1) with $\mathrm{Per}(E_k, \Omega) \neq 0$, there holds

$$\mathrm{d}G(\bar{\lambda}_k, +) \leq -\int_{E_{\bar{\lambda}_k}} p_k \, \mathrm{d}x < 0 = -\int_\emptyset p_k \, \mathrm{d}x \leq -\mathrm{d}G(\bar{\lambda}_k, -)$$

due to $j_{\bar{\lambda}_k}(E_k) = 0$ and $\mathrm{Per}(E_k, \Omega) \neq 0$. However, secondly, we also deduce that G is semismooth at every $\lambda > 0$ and, due to (4.3), $-\int_{E_\lambda} p_k \, \mathrm{d}x$ provides a Clarke subgradient of G at λ. Hence, we can hope to compute $\bar{\lambda}_k$ by using a generalized (or semismooth) Newton method, see [64, Ch. 2] for an introduction to such methods and the definitions of the concepts mentioned. More in detail, starting from $\lambda_0 = 1 \geq \bar{\lambda}_k$, we iteratively update

$$\lambda_{\ell+1} = \lambda_\ell + \frac{G(\lambda_\ell)}{\int_{E_{\lambda_\ell}} p_k \, \mathrm{d}x} = \frac{\lambda_\ell \int_{E_{\lambda_\ell}} p_k \, \mathrm{d}x + G(\lambda_\ell)}{\int_{E_{\lambda_\ell}} p_k \, \mathrm{d}x} = \frac{\mathrm{Per}(E_{\lambda_\ell}, \Omega)}{\int_{E_{\lambda_\ell}} p_k \, \mathrm{d}x}$$

whenever $\int_{E_{\lambda_\ell}} p_k \, \mathrm{d}x \neq 0$. The resulting algorithm, also referred to as the Dinkelbach-Newton method [61], is given in Algorithm 2. Note that the method terminates in two scenarios. The first is if $\lambda_0 = 1$ satisfies $G(\lambda_0) = 0$, in which case we immediately conclude

$$\max_{v \in \mathrm{Ext}(B_{\tilde{V}})} \int_\Omega p_k y \, \mathrm{d}x = \max_{E \in \mathcal{S}_\mathcal{T}(\Omega)} \frac{\int_E p_k \, \mathrm{d}x}{\mathrm{Per}(E, \Omega)} = 1,$$

i.e., u_k is a minimizer of $(\mathcal{P})$. The second is if there holds $G(\lambda_{L+1}) = 0$ for some $L > 0$, implying that we have

$$\lambda_{L+1} = \bar{\lambda}_k, \quad \mathrm{Per}(E_{\lambda_L}, \Omega) > 0 \quad \text{as well as} \quad j_{\lambda_{L+1}}(E_{\lambda_L}) = 0$$

by construction. As a consequence, E_{λ_L} is also a solution of Problem $(\mathcal{P}_\lambda)$ for $\lambda = \lambda_{L+1}$ and thus, see Proposition 4.1, a maximizer of (4.1). The following proposition addresses the wellposedness of Algorithm 2 and summarizes some of its key properties.

Proposition 4.3. *Let λ_ℓ denote the current iterate in step $\ell \in \mathbb{N}$ of Algorithm 2. Then we either have $\lambda_\ell = \bar{\lambda}_k$ or there holds $\lambda_\ell > \bar{\lambda}_k$. In the second case, $\lambda_{\ell+1}$ is well-defined and there holds $\bar{\lambda}_k \leq \lambda_{\ell+1} < \lambda_\ell$.*

Proof. We only prove the statement for $\ell = 0$, the general case follows mutatis mutandis. If $\lambda_0 = 1 \neq \bar{\lambda}_k$, Proposition 2.3 yields $G(\lambda_0) < 0$ and thus $\int_{E_{\lambda_0}} p_k \, \mathrm{d}x > 0$. Consequently, λ_1 is well-defined and there holds

$$\lambda_1 = \lambda_0 + \frac{G(\lambda_0)}{\int_{E_{\lambda_0}} p_k \, \mathrm{d}x} < \lambda_0 \quad \text{as well as} \quad \lambda_1 = \frac{\mathrm{Per}(E_{\lambda_0}, \Omega)}{\int_{E_{\lambda_0}} p_k \, \mathrm{d}x} \geq \bar{\lambda}_k$$

due to $\int_{E_{\lambda_0}} p_k \, \mathrm{d}x > 0$, $G(\lambda_0) < 0$ as well as the definition of $\bar{\lambda}_k$. $\qquad\square$

Algorithm 2: Dinkelbach-Newton method for Problem (4.1)

Input: Initial parameter $\lambda_0 = 1$ and dual variable $p_k = -K^* \nabla F(K u_k)$.

1 Find $E_{\lambda_0} \in \arg\min_{E \in \mathcal{S}_{\mathcal{T}}(\Omega)} \mathrm{Per}(E, \Omega) - \lambda_0 \int_E p_k \, \mathrm{d}x$.

2 if $G(\lambda_0) = 0$ **then**

3 $\quad$ Set $\bar{\lambda}_k = \lambda_0$ as well as $E_k = \emptyset$ and stop.

4 end

5 for $G(\lambda_\ell) < 0$ **do**

6 $\quad$ $\lambda_{\ell+1} \leftarrow \mathrm{Per}(E_{\lambda_\ell}, \Omega) / \int_{E_{\lambda_\ell}} p_k \, \mathrm{d}x$

7 $\quad$ Find $E_{\lambda_{\ell+1}} \in \arg\min_{E \in \mathcal{S}_{\mathcal{T}}(\Omega)} \mathrm{Per}(E, \Omega) - \lambda_{\ell+1} \int_E p_k \, \mathrm{d}x$.

8 $\quad$ **if** $G(\lambda_{\ell+1}) = 0$ **then**

9 $\quad\quad$ Set $\bar{\lambda}_k = \lambda_\ell$ as well as $E_k = E_{\lambda_\ell}$ and stop.

10 $\quad$ **end**

11 $\quad$ $\ell \leftarrow \ell + 1$

12 end

Output: $\bar{\lambda}_k$ and E_k satisfying (4.1) and (4.2).

Following, e.g., [61] we readily obtain that the iterates λ_ℓ generated by Algorithm satisfy

$$\frac{1}{\bar{\lambda}_k} - \frac{1}{\lambda_{\ell+1}} \le c_\ell \left(\frac{1}{\bar{\lambda}_k} - \frac{1}{\lambda_\ell} \right) \quad \text{for some sequence} \quad c_\ell \ge 0 \quad \text{with} \quad c_\ell \searrow 0,$$

i.e., their reciprocals converge superlinearly towards the maximum value of Problem (4.1). Exploiting the finite cardinality of $\mathcal{S}_{\mathcal{T}}(\Omega)$ as well as the strict monotonicity of λ_ℓ, we further conclude the finite termination of Algorithm 2.

Proposition 4.4. *Let λ_ℓ be generated by Algorithm 2. Then there is $L \in \mathbb{N}$ such that $\lambda_L = \bar{\lambda}_k$, i.e., Algorithm 2 stops after finitely many steps.*

Proof. Assume that Algorithm 2 does not stop after finitely many iterations. Then there holds

$$\int_{E_{\lambda_\ell}} p_k \, \mathrm{d}x > 0 \quad \text{for all} \quad \ell \in \mathbb{N}.$$

By definition, we further get

$$\frac{\mathrm{Per}(E_{\lambda_{\ell+1}}, \Omega)}{\int_{E_{\lambda_{\ell+1}}} p_k \, \mathrm{d}x} = \lambda_{\ell+2} < \lambda_{\ell+1} = \frac{\mathrm{Per}(E_{\lambda_\ell}, \Omega)}{\int_{E_{\lambda_\ell}} p_k \, \mathrm{d}x} \quad \text{for all } \ell \in \mathbb{N}.$$

Consequently, we necessarily have $E_{\ell_1} \ne E_{\ell_2}$ for all $\ell_1, \ell_2 \in \mathbb{N}$, $\ell_1 \ne \ell_2$. Since $\mathcal{S}_{\mathcal{T}}(\Omega)$ is finite, we arrive at a contradiction. $\qquad\square$

4.1.1 Graph cuts for Problem (P_λ)

Finally, it remains to discuss the efficient solution of the perimeter minimization problem (P_λ) which is required in Algorithm 2. For this purpose, we briefly recap its equivalence to a suitable minimal graph cut problem, see also [58], which can be solved exactly and efficiently using, for example, maximum flow algorithms such as [9]. In the following, let $G = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ be the dual graph associated to $\mathcal{T}$, with a node for each $T_j \in \mathcal{T}$ and where the weight of each graph edge (i, j) is given by the length of the common face of the simplices T_i, T_j, i.e. $w_{i,j} = \mathcal{H}^{d-1}(\partial T_i \cap \partial T_j)$.

Thanks to this construction, the perimeter in Ω of any set $E \in \mathcal{S}_{\mathcal{T}}(\Omega)$ can be computed through a sum of weights of edges in $\mathcal{E}$. Indeed, let $E \in \mathcal{S}_{\mathcal{T}}(\Omega)$, and denote by $I \subset \mathcal{V}$ the collection of vertices that correspond to triangles in E. Then,

$$\text{Per}(E, \Omega) = \sum_{T_i \subset E} \sum_{T_j \not\subset E} \mathcal{H}^{d-1}(\partial T_i \cap \partial T_j) = \sum_{i \in I} \sum_{j \in V \setminus I} w_{i,j}.$$

Therefore, the dual graph encodes the perimeter information of the elements of $\mathcal{S}_{\mathcal{T}}(\Omega)$. In order to incorporate information regarding the dual variable p_k, an additional construction is required. For this purpose, let us introduce two new vertices, denoted as s (the source) and t (the sink), in the set of vertices $\mathcal{V}$. Next, for each node $i \in \mathcal{V}$, we create directed edges (s, i) from the source to the node i, and (i, t) from the node to the sink. Finally, noting that

$$\lambda \int_E p_k \, dx = \sum_{T_i \subset E} \lambda p_k(T_i) \mathcal{L}^d(T_i),$$

we can encode the integral $-\lambda \int_E p_k \, dx$ into the weights of the new edges by defining

$$w_{s,i} := \max\{0, -\lambda p_k(T_i)\mathcal{L}^d(T_i)\} \quad \text{as well as} \quad w_{i,t} := \max\{0, \lambda p_k(T_i)\mathcal{L}^d(T_i)\} \quad \text{for all} \quad i \in \mathcal{V}.$$

The resulting graph is commonly referred to as an (s, t)-graph. Visually, this construction adds an extra "dimension" to the graph, which can be illustrated graphically as shown in Figure 3.
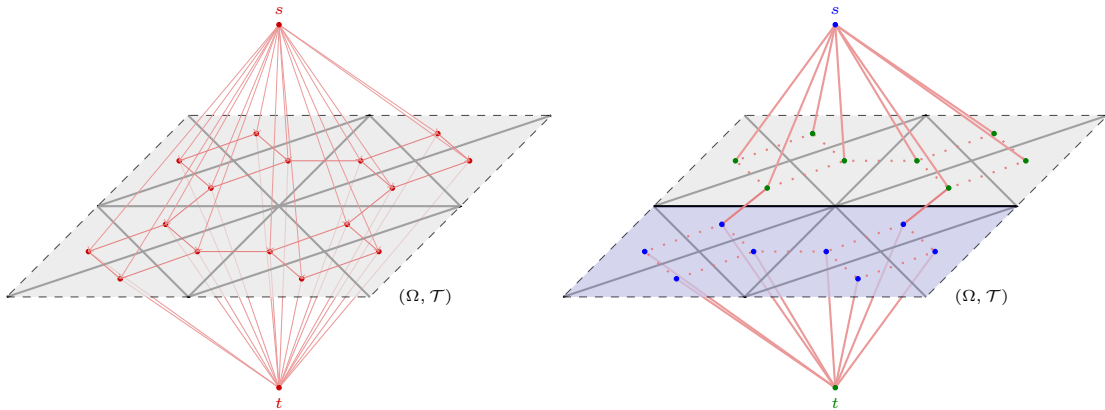

Figure 3: Our triangulated domain $(\Omega, \mathcal{T})$ in black, and the associated (s, t)-graph in red. On the right, an example of (s, t)-cut (blue vs green) for the unweighted (s, t)-graph, with the edges passing through the cut (i.e. connecting different color nodes) in red.

Given the (s, t)-graph as well as $I \subset \mathcal{V}$, we define the associated (s, t)-cut as the binary partition of the nodes of the graph into $I \cup \{t\}$ and $I^c \cup \{s\}$. The respective cut value is given by the flow through the cut, i.e.

$$\Theta_I := \sum_{i \notin I} \sum_{j \in I} w_{i,j} + \sum_{i \in I} w_{s,i} + \sum_{i \notin I} w_{i,t}.$$

Finally, a *minimal cut* is an (s, t) cut that minimizes the cut value. The following lemma addresses the equivalence between finding a minimal cut and solving the perimeter minimization problem (P_λ).

Lemma 4.5. *Let $I \subset \mathcal{V}$ and $E(I) = \text{int}\left(\bigcup_{i \in I} \overline{T_i}\right) \in \mathcal{S}_{\mathcal{T}}(\Omega)$ be the corresponding union of triangles. If $I \cup \{t\}$, $I^c \cup \{s\}$ is a minimal cut, then*

$$E(I) \in \underset{E \in \mathcal{S}_{\mathcal{T}}(\Omega)}{\arg \min} \ \text{Per}(E, \Omega) - \lambda \int_E p_k \, dx \tag{4.4}$$

23

Proof. By definition of the cut value, we have

$$\begin{aligned}
\Theta_I &= \operatorname{Per}(E(I), \Omega) + \sum_{i \in I} w_{s,i} + \sum_{j \notin I} w_{j,t} \\
&= \operatorname{Per}(E(I), \Omega) + \sum_{i \in I} \max(0, -\lambda p_k(T_i)\mathcal{L}^d(T_i)) + \sum_{i \in \mathcal{V} \setminus I} \max(0, \lambda p_k(T_i)\mathcal{L}^d(T_i)) \\
&= \operatorname{Per}(E(I), \Omega) - \sum_{i \in I} \lambda p_k(T_i)\mathcal{L}^d(T_i) + \sum_{i \in \mathcal{V}} \max(0, \lambda p_k(T_i)\mathcal{L}^d(T_i)) \\
&= \operatorname{Per}(E(I), \Omega) - \lambda \int_{E(I)} p_k \mathrm{d}x + \sum_{i \in \mathcal{V}} \max(0, \lambda p_k(T_i)\mathcal{L}^d(T_i)).
\end{aligned}$$

Since the last term is independent of $E(I)$, minimizing the cut value would minimize (4.4). $\square$

The full practical realization of FC-GCG for P_0 elements is given in Algorithm 3. Note that the scalars ζ_k can be computed as a by-product of Algorithm 3 and provide an upper bound on the suboptimality $J(u_k) - \min_{u \in P_0(\mathcal{T})} J(u)$, by Lemma 2.8.

Algorithm 3: Dinkelbach-FC-GCG for $V = P_0(\mathcal{T})$ discretization of Problem $(\mathcal{P})$

Input: $u_0 = 0$, $\mathcal{A}_0 = \emptyset$, $N_0 = 0$, $k = 0$, $\zeta_0 = 1$

1 calculate $c^0 \in \arg\min_{c \in \mathbb{R}} F(cK\mathbb{1})$
2 $u_0 \leftarrow c^0 \mathbb{1}_\Omega$
3 **while** $\zeta_k > 0$ **do**
4     $p_k \leftarrow -K_* \nabla F(Ku_k)$
5   calculate $(\overline{\lambda}_k, E_k)$ using Algorithm 2 with inputs $\lambda_0 = 1$ and p_k
6     $N_k^+ \leftarrow N_k + 1$
7     $u_{N_k^+}^k \leftarrow \mathbb{1}_{E_k} / \operatorname{Per}(E_k, \Omega)$
8     $\mathcal{A}_k^+ \leftarrow \mathcal{A}_k \cup \{u_{N_k^+}^k\}$
9   calculate (using SSN):

$$(\gamma^{k+1}, c^{k+1}) \in \arg\min_{(\gamma, c) \in \mathbb{R}_+^{N_k^+} \times \mathbb{R}} F\left(K\mathcal{U}_{\mathcal{A}_k^+}(\gamma) + cK\mathbb{1}\right) + \sum_{i=1}^{N_k^+} \gamma_i \quad \text{with minimum value } M_k$$

$$\zeta_{k+1} \leftarrow \left(-1/\overline{\lambda}_k - 1\right) M_k$$

10    $u_{k+1} \leftarrow \mathcal{U}_{\mathcal{A}_k^+}(\gamma^{k+1}) + c^{k+1}\mathbb{1}_\Omega = \sum_{i=1}^{N_k^+} \gamma_i^{k+1} u_i^k + c^{k+1}\mathbb{1}_\Omega$
11    $\mathcal{A}_{k+1} \leftarrow \mathcal{A}_k^+ \setminus \{u_i^k : \gamma_i^{k+1} = 0\}$
12    $N_{k+1} \leftarrow \#\mathcal{A}_{k+1}$
13    $k \leftarrow k + 1$
14 **end**

Output: $\overline{u} = u_k$

Finally, again exploiting the finite cardinality of $S_\mathcal{T}(\Omega)$, we conclude the convergence of Algorithm 3 in finitely many steps, assuming that the finite dimensional minimization problem for (γ^{k+1}, c^{k+1}) is solved exactly.

Proposition 4.6. *Let u_k be generated by Algorithm 3. Then there is $K \geq 0$ such that $\overline{u} = u_K$ is a minimizer of Problem $(\mathcal{P})$ for $V = P_0(\mathcal{T})$.*

Proof. Without loss of generality, assume that the iterate $u_k = $ in the k-th step of Algorithm 3, $j \geq 1$, is not optimal. Then, utilizing arguments from [16], there holds

$$F(Ku_{k+1}) + \sum_{i=1}^{N_k^+} \gamma_i^{k+1} < F(Ku_k) + \sum_{i=1}^{N_{k-1}^+} \gamma_i^k.$$

In particular, this implies $\mathcal{A}_{k+1} \neq \mathcal{A}_j$ for all $j \in \mathbb{N}$, $j \leq k$. Since the power set of $S_{\mathcal{T}}$ is also finite, we conclude the termination of Algorithm 3 in finitely many steps. $\square$

Remark 4.7. *The procedure summarized in Algorithm 3 bears some similarities to a version of the conditional gradient method for the deconvolution of zero-mean images presented in [47]. In this earlier work, the authors bypass the nonsmoothness of regularizer by introducing a dummy variable bounding the total variation of the image and considering the constrained minimization problem*

$$\min_{u \in \mathring{V}, \ r \in \mathbb{R}} [F(Ku) + r] \quad s.t. \quad \mathrm{TV}(u, \Omega) \leq r, \ 0 \leq r \leq D$$

where D is a given bound on the total variation of optimal reconstructed images. Subsequently, (accelerated) variants of the conditional gradient method are applied for the simultaneous update of the image u_k and the bound r_k. Similar to Algorithm 3, this requires the repeated solution of linear programs of the form

$$\max_{v \in \mathring{V}} \int_\Omega p_k v \ \mathrm{d}x \quad s.t. \quad \mathrm{TV}(v, \Omega) \leq 1. \quad (4.5)$$

which is realized by applying an interior point method to the dual of the equivalent problem

$$\min_{v \in \mathring{V}} \mathrm{TV}(v, \Omega) \quad s.t. \quad \int_\Omega p_k v \ \mathrm{d}x \geq 1. \quad (4.6)$$

The exact, geometry exploiting, solution approach considered in the present work, i.e. Algorithm 2, can also be motivated from this perspective but outperformed black-box linear programming approaches for (4.5) and (4.6) as well as their respective dual problems significantly in our numerical tests. Indeed, formulating the saddle-point problem for the Lagrangian of (4.6), we observe that

$$\max_{\lambda \geq 0} \min_{v \in \mathring{V}} \left[\mathrm{TV}(v, \Omega) - \lambda \int_\Omega p_k v \ \mathrm{d}x + \lambda \right] = \max_{\lambda \geq 0} \min_{E \in S_{\mathcal{T}}(\Omega)} \left[\mathrm{Per}(E, \Omega) - \lambda \int_E p_k \ \mathrm{d}x + \lambda \right]$$
$$= \max_{\lambda \geq 0} [G(\lambda) + \lambda] = \bar{\lambda}_k$$

due to $G(\lambda) \leq 0$ for all $\lambda \geq 0$ and since $\bar{\lambda}_k$ is the largest zero of G. Hence, Algorithm 2 computes a saddle-point of the Lagrangian in a finite number of iterations using the characterization of $\mathrm{Ext}(B_{\mathring{V}})$.

5 Discrete-to-continuum limits of TV problems on P0 functions

In this section, for sequences of triangulations $\mathcal{T}_h$ with $h > 0$ an upper bound on the diameter of the triangles, we investigate the behavior of minimizers of problems of the form $(\mathcal{D})$ formulated over functions $P_0(\mathcal{T}_h)$ as $h \to 0$. Specifically, we consider minimization problems

$$\min_{u_h \in P_0(\mathcal{T}_h)} F(K_h u_h) + \mathrm{TV}(u_h, \Omega)$$

where K_h is defined by $y_h = K_h u_h \in P_1(\mathcal{T}_h) \cap H_0^1(\Omega)$ being the unique solution to a Galerkin weak formulation of a linear elliptic PDE such as

$$\text{find } y \in P_1(\mathcal{T}_h) \cap H_0^1(\Omega) \quad \text{s.t.} \quad \int_\Omega \nabla y \cdot \nabla \theta_h + y\theta_h \, \mathrm{d}x = \int_\Omega u\theta_h \, \mathrm{d}x \tag{5.1}$$

for all $\theta_h \in P_1(\mathcal{T}_h) \cap H_0^1(\Omega)$, and the corresponding continuum K is the solution operator of the analogous weak formulation in $H_0^1(\Omega)$ without the $P_1(\mathcal{T}_h)$ constraint.

The challenge, as already indicated in the introduction (see the references therein), is that the total variation does not behave consistently with respect to approximation with piecewise constant functions. Therefore, we consider an approach based on Γ-convergence (see [10, Ch. 1] for the definition and basic properties) in which limits with respect to h of solutions over each mesh $\mathcal{T}_h$ are proved to be minimizers of a continuum variational problem, which in turn depends on which sequence of meshes $\mathcal{T}_h$ was used. Such an approach is based on formulating both the approximating and the limiting energies in the same functional space. To this end, we define the functional $\text{TV}_{\mathcal{T}_h} : L^1(\Omega) \to \mathbb{R}^+ \cup \{0, +\infty\}$ by

$$\text{TV}_{\mathcal{T}_h}(u) := \begin{cases} \text{TV}(u, \Omega) & \text{if } u \in P_0(\mathcal{T}_h), \\ +\infty & \text{otherwise,} \end{cases}$$

and the corresponding perimeter

$$\text{Per}_{\mathcal{T}_h}(E) := \begin{cases} \text{Per}(E, \Omega) & \text{if } E \in \mathcal{S}_{\mathcal{T}_h}(\Omega) \\ +\infty & \text{otherwise.} \end{cases}$$

In this framework and under suitable assumptions on $\mathcal{T}_h$, we will see that in the limiting energies the total variation term is replaced with an anisotropic version defined by

$$\text{TV}_\varphi(u) := \sup \left\{ \int_\Omega u \, \text{div} \, \psi \, \mathrm{d}x \, \bigg| \, \psi \in \mathcal{C}_c^1(\Omega; \mathbb{R}^d), \; \varphi^*(\psi(x)) \leq 1 \text{ for all } x \in \Omega \right\}$$
$$= \int_\Omega \varphi\left(\frac{Du}{|Du|}\right) \mathrm{d}|Du|,$$

for some convex positively one-homogeneous real valued function φ, and with $Du/|Du|$ denoting the Radon-Nikodym derivative of $Du \in M(\Omega; \mathbb{R}^d)$ with respect to the nonnegative measure $|Du|$.

A first step towards this goal is to notice that this type of convergence for the perimeters implies the same type of convergence for the PDE-constrained optimization problem with TV regularization, and convergence of the corresponding minimizers. To this end, we recall that a sequence of sets $E_n \subset \mathbb{R}^d$ is said to converge in L^1 to a limit set E whenever

$$\mathcal{L}^d(E_n \triangle E) = \mathcal{L}^d(E_n \setminus E) + \mathcal{L}^d(E \setminus E_n) \xrightarrow[n \to \infty]{} 0.$$

It is straightforward to check that the quantity $(E, F) \mapsto \mathcal{L}^d(E \triangle F)$ defines a metric on Lebesgue equivalence classes of subsets of $\mathbb{R}^d$, which in particular allows us to speak of Γ-convergence with respect to this topology.

Proposition 5.1. *Let $1 \leq d \leq 4$, $1 \leq p < d/(d-1)$, and $\mathcal{T}_h$ be triangulations of Ω (which we assume is polyhedral) satisfying $h \geq \max\{\text{diam}(T) \mid T \in \mathcal{T}_h\}$ and such that there is $C_s > 0$ with*

$$\frac{1}{h} \sup \left\{ r > 0 \mid B(x, r) \subset T \text{ for some } x \in T \right\} \geq C_s \quad \text{for all } h > 0 \text{ and } T \in \mathcal{T}_h.$$

Assume that F is lower semicontinuous with respect to the strong topology of $L^p(\Omega)$, and that the discrete perimeter $\text{Per}_{\mathcal{T}_h}$ Γ-converges with respect to the L^1 convergence of sets as $h \to 0$ to the anisotropic perimeter functional given by

$$A \mapsto \int_{\partial^* A} \varphi(\nu_A(x)) d\mathcal{H}^{d-1}(x),$$

where $\nu_A = D\mathbb{1}_A/|D\mathbb{1}_A|$ and $\partial^* A$ is the reduced boundary of A. Then we have

$$J_h := F \circ K_h + \mathrm{TV}_{\mathcal{T}_h} \xrightarrow{\Gamma - L^p} J := F \circ K + \mathrm{TV}_\varphi, \qquad (5.2)$$

and any sequence $\bar{u}_h$ of minimizers of J_h has a subsequence converging to a minimizer $\bar{u}$ of J strongly in $L^p(\Omega)$, for which $\mathrm{TV}(\bar{u}_h) \to \mathrm{TV}_\varphi(\bar{u})$.

Proof. By [29, Prop. 3.5] the Γ-convergence of $\mathrm{Per}_{\mathcal{T}_h}$ with respect to L^1 convergence of sets is equivalent to the convergence of the corresponding total variations, so we have

$$\mathrm{TV}_{\mathcal{T}_h} \xrightarrow{\Gamma - L^1(\Omega)} \mathrm{TV}_\varphi. \qquad (5.3)$$

Now assume $u_h \rightharpoonup u$ weakly in $L^{d/(d-1)}(\Omega)$, and let us consider the discrete and continuous weak formulations (5.1) with unique solution $y_h \in P_1(\mathcal{T}_h) \cap H_0^1(\Omega)$, and

$$\text{find } y \in H_0^1(\Omega) \text{ s.t. } \int_\Omega \nabla y \cdot \nabla \theta + y\theta \, \mathrm{d}x = \int_\Omega u\theta \, \mathrm{d}x \quad \text{for all } \theta \in H_0^1(\Omega), \qquad (5.4)$$

with unique solution $y \in H^1(\Omega)$. Choosing $\theta = y_h$ in (5.1), using the Hölder inequality and then Poincaré-Sobolev inequality in $H_0^1(\Omega)$ for y_h, we obtain the estimate

$$\|\nabla y_h\|_{L^2(\Omega)}^2 = \int_\Omega |\nabla y_h|^2 \, \mathrm{d}x \leq \int_\Omega |u_h y_h| \, \mathrm{d}x \leq \|u_h\|_{L^{d/(d-1)}(\Omega)} \|y_h\|_{L^d(\Omega)}$$
$$\leq C_\Omega \|u_h\|_{L^{d/(d-1)}(\Omega)} \|\nabla y_h\|_{L^2(\Omega)}, \qquad (5.5)$$

so that $\|\nabla y_h\|_{L^2(\Omega)} \leq C_\Omega \|u_h\|_{L^{d/(d-1)}(\Omega)}$. Notice that this is possible because Ω is bounded and

$$\left(\frac{d-1}{d}\right)' = d \leq \frac{2d}{d-2} \qquad (5.6)$$

is satisfied for $d \leq 4$, with equality if $d = 4$. Again using the Poincaré inequality, estimate (5.5) means that the sequence y_h is bounded in $H^1(\Omega)$. Therefore, using Banach-Alaoglu and Rellich-Kondrakov there is $y_\ell \in H_0^1(\Omega)$ such that (up to a subsequence) we have $\nabla y_h \rightharpoonup \nabla y_\ell$ weakly in $L^2(\Omega)$. For any $\theta \in C_c^\infty(\Omega)$ and by the assumptions on $\mathcal{T}_h$, we can find (see for example [18, Thm. 4.4.20]) a sequence $\theta_h \in P_1(\mathcal{T}_h) \cap H_0^1(\Omega)$ such that $\theta_h \to \theta$ strongly in $H^1(\Omega)$ and use it in (5.1). Using the Poincaré-Sobolev inequality in $H_0^1(\Omega)$ for θ_h and taking into account (5.6) we see that

$$\nabla y_h \rightharpoonup \nabla y \text{ in } L^2(\Omega), \quad \nabla \theta_h \to \nabla \theta \text{ in } L^2(\Omega),$$
$$u_h \rightharpoonup u \text{ in } L^{d/(d-1)}(\Omega), \quad \theta_h \to \theta \text{ in } L^d(\Omega),$$

so we can pass to the limit on both sides of (5.1) to obtain that y_ℓ satisfies

$$\int_\Omega \nabla y_\ell \nabla \theta \, \mathrm{d}x = \int_\Omega u\theta \, \mathrm{d}x \quad \text{for all } \theta \in C_c^\infty(\Omega).$$

Using the density of $C_c^\infty(\Omega)$ in the strong topology of $H_0^1(\Omega)$ we have that y_ℓ in fact satisfies (5.4), and uniqueness of solutions for this weak formulation implies that $y_\ell = y$.

Let us now prove (5.2). If we consider $u_h \to u$ strongly in $L^p(\Omega)$ for the Γ-$\liminf$ inequality, then since Ω is bounded also $u_h \to u$ strongly in $L^1(\Omega)$. This allows us to use (5.3) to obtain $\mathrm{TV}_\varphi(u) \leq \liminf_{h\to 0} \mathrm{TV}_{\mathcal{T}_h}(u_h)$. Moreover, if $\mathrm{TV}_\varphi(u) < +\infty$ this limit inferior can be approximated with (not relabelled) subsequences for which $\mathrm{TV}_{\mathcal{T}_h}(u_h) < +\infty$, so that $u_h \in P_0(\mathcal{T}_h)$ and $\mathrm{TV}_{\mathcal{T}_h}(u_h) = \mathrm{TV}(u_h) \leq C$. By weak compactness and the Poincaré-Sobolev inequality in $BV(\Omega)$ [4, Thm. 3.23,

Thm. 3.44] and since Ω is bounded we have (after possibly taking a further subsequence) that $u_h \rightharpoonup u$ in $L^{d/(d-1)}$, which combined with the compactness and continuity of y_h proved above yields

$$(F \circ K)(u) + \mathrm{TV}_\varphi(u) \leq \liminf_{h \to 0}(F \circ K_h)(u_h) + \mathrm{TV}_{\mathcal{T}_h}(u_h) \text{ for all } u_h \to u \text{ in } L^p(\Omega),$$

where we note that this inequality also holds if $\mathrm{TV}_\varphi(u) = +\infty$, since in that case by the use of (5.3) also the right hand side must equal $+\infty$. For the other half of the Γ-convergence, if we are given $u \in L^2(\Omega)$ we would like to find $u_h \to u$ strongly in $L^p(\Omega)$ for which

$$(F \circ K)(u) + \mathrm{TV}_\varphi(u) \geq \limsup_{h \to 0}(F \circ K_h)(u_h) + \mathrm{TV}_{\mathcal{T}_h}(u_h), \tag{5.7}$$

and may assume that $\mathrm{TV}_\varphi(u) < +\infty$, since otherwise the inequality is trivial. Using the Γ-convergence of the discrete perimeters (5.3) we can find a recovery sequence of functions u_h such that $u_h \to u$ in $L^1(\Omega)$, $\mathrm{TV}_{\mathcal{T}_h}(u_h) < \infty$ and $u_h \in P_0(\mathcal{T}_h)$ for all h, and

$$\limsup_{h \to 0} \mathrm{TV}_{\mathcal{T}_h}(u_h) = \limsup_{h \to 0} \mathrm{TV}(u_h) \leq \mathrm{TV}_\varphi(u).$$

Since $\mathrm{TV}_\varphi(u) < +\infty$ we have that $\mathrm{TV}(u_h)$ is bounded and we can proceed as above to see that, up to taking a not relabelled subsequence, also $u_h \to u$ strongly in $L^p(\Omega)$ and $y_h \to y$ strongly in $L^q(\Omega)$ for $q < d$, and from that conclude that for this particular sequence (5.7) holds.

To conclude the convergence of minimizers, we notice that $\mathrm{TV}_{\mathcal{T}_h}(u) \geq \mathrm{TV}(u)$ for every $u \in L^p(\Omega)$, which in combination with Assumption **A**3 and the Poincaré-Sobolev inequality in $\mathrm{BV}(\Omega)$ that the family J_h is equicoercive in the norm of this space. This allows us, for a sequence of minimizers $\bar{u}_h$ and by the same compact embedding arguments used before, to extract a subsequential limit $\bar{u}$ in the strong $L^p(\Omega)$ and weak $L^{d/(d-1)}$ topologies, and by the fundamental theorem of Γ-convergence [36, Cor. 7.20, Thm. 7.8] this limit must be a minimizer of J, and

$$J_h(\bar{u}_h) = \lim_{h \to 0} J(\bar{u}). \tag{5.8}$$

Moreover, since $\bar{u}_h \rightharpoonup \bar{u}$ in $L^{d/(d-1)}$, we can run the same continuity argument for $K(\bar{u}_h)$ to infer from (5.8) that $\lim_{h \to 0} \mathrm{TV}(\bar{u}_h) = \mathrm{TV}_\varphi(\bar{u})$. $\qquad \square$

The above result is in particular applicable for periodic triangulations of the cube $\Omega = (0,1)^d$:

Definition 5.2. Given an initial triangulation $\mathcal{T}_1 = \{T_1, ..., T_{n_\tau}\}$ of Ω, we say that $\{\mathcal{T}_k\}_k$ is a sequence of periodic triangulations on $\Omega = (0,1)^d$ if

$$\mathcal{T}_k = \bigcup_{i_1, ..., i_d = 0}^{k-1} \frac{1}{k}\big(\mathcal{T}_1 + (i_1, \ldots, i_d)\big).$$

In the previous equation, the two operations are conceptualized as shifting and rescaling of the elements of $\mathcal{T}_1$ to facilitate the division of Ω into k^d identical cubes of size $1/k$ with a rescaled copy of $\mathcal{T}_1$ inside each. A typical example of such triangulations is the 'double diagonal' depicted in Figure 4, which will be our main focus later on.

Proposition 5.3. *If $\{\mathcal{T}_k\}_k$ is a sequence of periodic triangulations on $\Omega = (0,1)^d$, the discrete perimeter $\mathrm{Per}_{\mathcal{T}_k}$ Γ-converges as $k \to \infty$ and with respect to the L^1 convergence of sets, to the anisotropic perimeter functional given by*

$$A \mapsto \int_{\partial^* A} \varphi_{\mathcal{T}_1}\big(\nu_A(x)\big) d\mathcal{H}^{d-1}(x), \tag{5.9}$$

where $\varphi_{\mathcal{T}_1}$ is Lipschitz continuous and given at rational directions ν by the single cell formula

$$\varphi_{\mathcal{T}_1}(\nu) = 3^{d-1} \inf \left\{ |Du| \left([1/3, 2/3]^d \right) \, \Big| \, u \in P_0(\mathcal{T}_3) \text{ s.t. } u(x) - \nu \cdot x \text{ is } 1/3\text{-periodic} \right\}, \quad (5.10)$$

in which the piecewise constant functions involved are real-valued, satisfy

$$u(x + 1/3z) - \nu \cdot (x + 1/3z) = u(x) - \nu \cdot x \quad \text{for all} \quad z \in \mathbb{Z}^d,$$

and the cost includes jumps at $\partial [1/3, 2/3]^d$.

Proof. We aim to use results from the literature on discrete-to-continuum convergence of variational problems on lattice systems, specifically the integral representation of [12, Thm. 2.4] and cell formula of [30, Prop. 2.6]. To this end, we define the lattice dual to $\mathcal{T}_k$ consisting of the centroids of the simplices in $\mathcal{T}_k$, that is

$$L_k = \bigcup_{i_1,\dots,i_d=0}^{k-1} \frac{1}{k} \left(\left\{ \mathrm{ce}(T_1^1), \dots, \mathrm{ce}(T_{n_\tau}^1) \right\} + (i_1, \dots, i_d) \right).$$

The obstacle to apply the mentioned results directly is that they are formulated for the embedding into continuum sets which are unions of Voronoi cells of the lattice, whereas the sets we consider are unions of triangles of $\mathcal{T}_k$. However, for any subset $A \in \mathcal{S}_{\mathcal{T}_k}(\Omega)$ we can define

$$L_k \supset \Xi_k(A) = \{ \mathrm{ce}(T_1), \dots, \mathrm{ce}(T_{N_A}) \} \text{ whenever } A = \bigcup_{i=1}^{N_A} T_i.$$

Noticing that the Ξ_k so defined is a bijection, we define the corresponding discrete perimeter on subsets of L_k given by

$$\mathrm{Per}_{L_k}(B) := \mathrm{Per}_{\mathcal{T}_k}\left(\Xi_k^{-1}(B) \right) \text{ for each } B \subset L_k.$$

Given this correspondence of the interaction energies, to conclude it is enough to check that for all sequences

$$\{A_k\} \subset \mathcal{S}_{\mathcal{T}_k}(\Omega) \text{ with } \mathrm{Per}_{\mathcal{T}_k}(A_k) \leq M < +\infty$$

we have

$$A_k \xrightarrow{L^1} A \subset (0,1)^d \quad \text{if and only if} \quad (V_k \circ \Xi_k)(A_k) \xrightarrow{L^1} A, \quad (5.11)$$

where V_k assigns to subsets of L_k the union of closures of cells of the Voronoi tesellation induced by L_k, which is the continuum embedding for lattice subsets used in [12] and [30]. To prove (5.11), we can denote the set of indices of simplices of A_k having one or more faces in ∂A_k as

$$I_k = \left\{ (i_1, \dots, i_d, \ell) \, \Big| \, \frac{1}{k}\left(T_\ell^1 + (i_1, \dots, i_d) \right) \subset A_k \text{ and } \frac{1}{k}\left(\partial T_\ell^1 + (i_1, \dots, i_d) \right) \cap \partial A_k \neq \emptyset \right\}.$$

Noticing that only these simplices and corresponding Voronoi cells can contribute to the symmetric difference between A_k and $(V_k \circ \Xi_k)(A_k)$, that their volume must be bounded by the size of a complete periodicity cube, and that $\mathrm{Per}_{\mathcal{T}_k}(A_k) = \mathcal{H}^{d-1}(\partial A_k \cap \Omega)$, we estimate

$$\mathcal{L}^d \left(A_k \triangle \left[(V_k \circ \Xi_k)(A_k) \right] \right) \leq k^{-d} \, \#\left(I_k \right)$$

$$\leq k^{-d} \left(\min_{\ell, F} \left\{ \mathcal{H}^{d-1}(F) \, \big| \, F \text{ a face of } \partial T_\ell^1 \right\} \right)^{-1} M k^{d-1} \xrightarrow[k \to \infty]{} 0.$$

Having proved (5.11), Γ-convergence of $\mathrm{Per}_{\mathcal{T}_k}$ is equivalent to that of the corresponding lattice system, which gives rise to a functional a functional of the form (5.9) as proved in [12, Thm. 2.4 and Sec. 2.1.1], and allows us to express φ using the cell formula (5.10) as a specialization of [30, Prop. 2.6]. For this formula, since the infimum needs to take into account interactions with the neighboring cubes, we have chosen to formulate it on the middle cube of $\mathcal{T}_3$, but clearly other analogous expressions are possible. $\qquad\square$
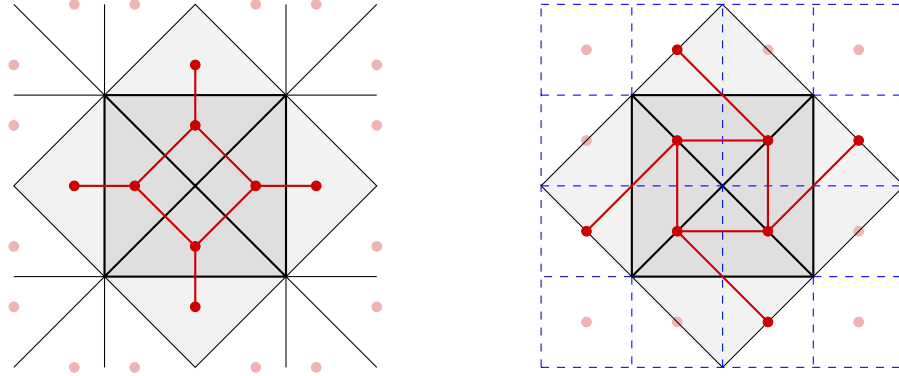
Figure 4: Representing the regular 'double diagonal' triangulation of the plane as interactions of periodic lattice systems, with the shaded central square being a complete period. The red segments represent the interactions corresponding to the thicker black edges of the triangulation. The assignment of Proposition 5.3 is depicted on the left, with the triangles themselves being the Voronoi cells. For the alternative construction on the right using the regular lattice $\frac{1}{2k}\mathbb{Z}^2$ each triangle is assigned counterclockwise to a red point, and the corresponding dashed blue square Voronoi cell has the same area.

Remark 5.4. *If $\mathcal{T}_k$ is the 'double-diagonal' triangulation in the plane with period $1/k$, we can also make an explicit correspondence to an interaction energy on $\frac{1}{2k}\mathbb{Z}^2$, as depicted in the right diagram of Figure 4. We remark as well that even though for this very symmetric triangulation the Voronoi cells corresponding to the centroids of the triangles (see the left diagram of Figure 4) are the triangles themselves, this is not true for more general triangulations, and in general Voronoi cells may also intersect the complement of their corresponding periodicity cube.*

Remark 5.5. *When using our algorithm on pseudo-random fine triangulations from a mesh generator (see Section 6 below), we empirically observe solutions resembling those corresponding to the isotropic case $\varphi \equiv 1$. Proving a result trying to formalize this behavior is beyond the scope of this article, but we point out that such results on isotropy of the limiting energy exist in the literature on stochastic homogenization of lattice systems, see for example [1].*

5.1 Explicit anisotropy for planar double-diagonal triangulations

We begin with a 'slicing' result, inspired by the continuous analogue (see for example [11, Lem. 3.2]), to modify sequences of triangulated sets converging to a half plane so that the intersection of their boundaries and $\partial(0,1)^2$ match the corresponding one for a half plane with rational normal: for $\nu = (\sigma, \tau) \in S^1$, we will denote by E_ν the subset of $(0,1)^2$ given by

$$E_\nu = \begin{cases} \{x \in (0,1)^2 : x \cdot \nu > 0\} & \text{if } \sigma\tau \leq 0, \\ \{x \in (0,1)^2 : |x \cdot \nu| > 1\} & \text{if } \sigma\tau > 0. \end{cases} \tag{5.12}$$

Definition 5.6. We say that a triangulation $\mathcal{T}$ on $\Omega = (0,1)^2$ is diagonally separable if there exist simple sets $A_1, A_2 \in S_\mathcal{T}(\Omega)$ such that

$$\partial A_i \cap \partial\Omega = \partial E_{\nu_i} \cap \partial\Omega,$$

where $\nu_1 = \frac{1}{\sqrt{2}}(1,1) \in S^1$, and $\nu_2 = \frac{1}{\sqrt{2}}(-1,1) \in S^1$.

Note that if $\mathcal{T}_1$ of a sequence of periodic triangulations $\{\mathcal{T}_k\}_k$ is diagonally separable, then so is any $\mathcal{T}_k$ of the sequence.

Lemma 5.7. *Let $\{\mathcal{T}_k\}_k$ be a sequence of periodic triangulations of $\Omega = (0,1)^2$, such that $\mathcal{T}_1$ is diagonally separable, and E_ν as in (5.12). For any rational direction $\nu \in S^1$, and any sequence $\{E_k \in S_{\mathcal{T}_k}(\Omega)\}_k$ converging in L^1 to E_ν, there exists a (sub)sequence $\{\widetilde{E}_{k_m} \in S_{\mathcal{T}_{k_m}}(\Omega)\}_m$ satisfying $\partial\widetilde{E}_{k_m} \cap \partial\Omega = \partial E_\nu \cap \partial\Omega$, and*

$$\lim_{m \to \infty} \mathcal{H}^1(\Omega \cap (\partial\widetilde{E}_{k_m} \setminus \partial E_{k_m})) = 0. \tag{5.13}$$

Proof. Let $\nu = \frac{1}{\sqrt{\sigma^2+\tau^2}}(\sigma,\tau) \in S^1$ be a rational direction for some coprime integers $\sigma,\tau \in \mathbb{Z}$ and let $n = \max\{|\sigma|,|\tau|\}$. We assume without loss of generality that neither σ nor τ is equal to zero. If this is not the case, then we can construct the new sequence as $\widetilde{E}_k := E_\nu$ which is admissible for every k and has zero perimeter inside Ω.

For $\delta > 0$, consider the neighbourhood of $\partial\Omega$ given by

$$\Omega_\delta = \{x \in \Omega : d(x,\partial\Omega) < \delta\},$$

where $d(x,\partial\Omega) = \inf_{y\in\partial\Omega} \|x-y\|$. It is clear that $\Omega_\delta \in S_{\mathcal{T}_k}(\Omega)$ only for specific choices of k and δ. More precisely, given $\delta = \frac{1}{m}$ for some fixed $m \in \mathbb{N}$, the previous statement holds if k is a multiple of m. For this reason, we will only consider elements of the sequence of triangulations $\mathcal{T}_k$ satisfying $k = \ell m n$ for some $\ell \in \mathbb{N}$. We set the index k to also be a multiple of n so that the boundary of E_ν intersects $\partial\Omega_{1/m}$ precisely at vertices of periodic squares of $\mathcal{T}_k$, as depicted in Figure 5. It will also be the case for any other neighborhood of $\partial\Omega$ of the form $\Omega_{\frac{nj}{k}}$ with $j \in \{1,...,\ell\}$, where in particular, $\frac{nj}{k} = \frac{1}{m}$ if $j = \ell$.

Let

$$C(\mathcal{T}_1) := \max_{T\in\mathcal{T}_1} \frac{\mathcal{H}^1(\partial T)}{\mathcal{L}^2(T)}.$$

be a bound on the shape of triangles in the initial triangulation. For every triangle $T \in \mathcal{T}_k$ and $j \in \{1,...,\ell\}$, since $d = 2$ and triangles in $\mathcal{T}_k$ are obtained by rescaling triangles of $\mathcal{T}_1$ by a factor $1/k$, we have

$$\mathcal{H}^1\left(\partial\Omega_{\frac{nj}{k}} \cap (E_k \triangle E_\nu) \cap \overline{T}\right) \leq C(\mathcal{T}_1)k\mathcal{L}^2(E_k \triangle E_\nu) \cap \overline{T}.$$

The previous inequality holds true even if we sum the lefthand side over $j \in \{1,...,\ell\}$. Indeed, by the assumption on the initial triangulation, each triangle in $\mathcal{T}_k$ intersects with at most one slice $\partial\Omega_{\frac{nj}{k}}$ (see Figure 5). Thus,

$$\sum_{j=1}^{\ell} \mathcal{H}^1\left(\partial\Omega_{\frac{nj}{k}} \cap (E_k \triangle E_\nu) \cap \overline{T}\right) \leq C(\mathcal{T}_1)k\mathcal{L}^2((E_k \triangle E_\nu) \cap \overline{T})$$

Summing the righthand side of the inequality over $T \in \mathcal{T}_k$ gives $C(\mathcal{T}_1)k\mathcal{L}^2(E_k \triangle E_\nu)$ because triangles are disjoint. On the other hand, for the lefthand side, we have

$$\sum_{T\in\mathcal{T}_k} \mathcal{H}^1\left(\partial\Omega_{\frac{nj}{k}} \cap (E_k \triangle E_\nu) \cap \overline{T}\right) = 2\mathcal{H}^1\left(\partial\Omega_{\frac{nj}{k}} \cap (E_k \triangle E_\nu)\right).$$

To see this, take a look at Figure 5, and note that, by periodicity, each piece of $\partial\Omega_{\frac{nj}{k}}$ (in blue) appears in two triangles of $\mathcal{T}_k$. All together,

$$\sum_{j=1}^{\ell} \mathcal{H}^1\left(\partial\Omega_{\frac{nj}{k}} \cap (E_k \triangle E_\nu)\right) \leq \frac{1}{2}C(\mathcal{T}_1)k\mathcal{L}^2(E_k \triangle E_\nu). \tag{5.14}$$

One of the terms of the sum in Equation (5.14) must be not larger than the average. Therefore, there exists $j(\ell, m) \in \{1, ..., \ell\}$ for which

$$\mathcal{H}^1\left(\partial\Omega_{\frac{nj(\ell,m)}{k}} \cap (E_k \triangle E_\nu)\right) \leq \frac{1}{\ell} \sum_{j=1}^{\ell} \mathcal{H}^1\left(\partial\Omega_{\frac{nj}{k}} \cap (E_k \triangle E_\nu)\right). \tag{5.15}$$

Together with Equation (5.14), we conclude that

$$\mathcal{H}^1\left(\partial\Omega_{\frac{nj(\ell,m)}{k}} \cap (E_k \triangle E_\nu)\right) \leq \frac{mn}{2} C(\mathcal{T}_1)\mathcal{L}^2(E_k \triangle E_\nu).$$

By the L^1 convergence of E_k to E_ν, we know that for every $m \in \mathbb{N}$, there exists $\ell(m) \in N$ such that for each $\ell \geq \ell(m)$,

$$\mathcal{L}^2(E_k \triangle E_\nu) < \frac{1}{m^2}.$$

Thus,

$$\mathcal{H}^1\left(\partial\Omega_{\frac{nj(\ell,m)}{k}} \cap (E_k \triangle E_\nu)\right) < \frac{n}{2m} C(\mathcal{T}_1) \xrightarrow[m\to\infty]{} 0.$$

In particular, if we consider the subsequence of triangulations $\mathcal{T}_{k_m}$ with $k_m := \ell(m)mn$, we get

$$\lim_{m\to\infty} \mathcal{H}^1\left(\partial\Omega_{\frac{nj_m}{k_m}} \cap (E_{k_m} \triangle E_\nu)\right) = 0, \tag{5.16}$$

where $j_m = j(\ell(m), m) \in \{1, ..., \ell(m)\}$ of Equation (5.15).

We construct a new sequence of triangulated sets $\widetilde{E}_{k_m} \in S_{\mathcal{T}_{k_m}}(\Omega)$ as

$$\widetilde{E}_{k_m} := \begin{cases} E_{k_m} & \text{on } \Omega \setminus \Omega_{\frac{nj_m}{k_m}} \\ A_m & \text{on } \Omega_{\frac{nj_m}{k_m}}, \end{cases}$$

where $A_m \in S_{\mathcal{T}_{k_m}}(\Omega)$ is a set satisfying the two attachment conditions:

$$\partial A_m \cap \partial\Omega = \partial E_\nu \cap \partial\Omega, \qquad \partial A_m \cap \partial\Omega_{\frac{nj_m}{m}} = \partial E_\nu \cap \partial\Omega_{\frac{nj_m}{m}},$$

and such that

$$\partial A_m \subset \bigcup \left\{\partial T \,\middle|\, T \in \mathcal{T}_{k_m} \text{ and } T \subset Q = \left[(0, 1/k_m)^2 + (i_1, i_2)\right] \text{ with } Q \cap \partial E_\nu \neq \emptyset\right\}. \tag{5.17}$$

The last condition enforces the boundary of A_m to be contained in periodic squares of $\mathcal{T}_{k_m}$ that intersect the boundary of the target E_ν. In this way, the maximum possible perimeter inside the squares containing $\partial E_\nu \cap \Omega_{\frac{nj_m}{m}}$ bounds the perimeter of A_m in $\Omega_{\frac{nj_m}{m}}$. The assumption that $\mathcal{T}_1$ is diagonally separable guarantees the existence of such an A_m. An example of this construction is depicted in Figure 5. Finally, to prove (5.13), we notice that

$$\Omega \cap (\partial\widetilde{E}_{k_m} \setminus \partial E_{k_m}) \subseteq \left(\Omega_{\frac{nj_m}{k_m}} \cap \partial A_m\right) \cup \left(\partial\Omega_{\frac{nj_m}{k_m}} \cap (E_{k_m} \triangle E_\nu)\right),$$

and by (5.17) we know that

$$\lim_{m\to\infty} \mathcal{H}^1\left(\Omega_{\frac{nj_m}{k_m}} \cap \partial A_m\right) = 0,$$

which combined with (5.16) proves (5.13). $\qquad\qquad\square$

We will now describe the anisotropy $\varphi_{\mathcal{T}_1}$ for a specific choice of periodic planar triangulations, the 'double diagonal'. The result implies, in particular, that sets with fixed mass which are optimal for $\text{Per}_{\varphi_{\mathcal{T}_1}}$ (i.e. the Wulff shape) are precisely the octagons which can be realized without error with piecewise constant functions on this type of mesh, as appearing for example in Figure 8(b). The strategy is to directly compute the limit energy of half-planes, for which we can modify the recovery sequence of Proposition 5.3 to match the boundary condition of the target set with the help of the previous lemma.
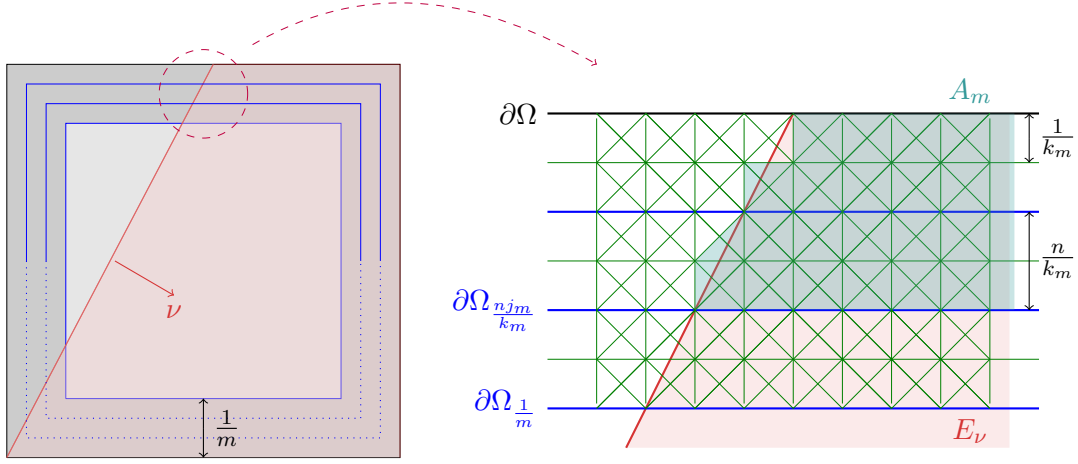
Figure 5: A sketch of the constructions in the proof of Lemma 5.7. In this case $\nu = (2, -1)/\sqrt{5}$.

Proposition 5.8 (Octagon). *Let $\{\mathcal{T}_k\}_k$ be a sequence of periodic triangulations of $\Omega = (0, 1)^2$, such that $\mathcal{T}_1$ is the 'double diagonal' triangulation of a square. Then the unit ball $\varphi_{\mathcal{T}_1}^{-1}(1) \subset \mathbb{R}^2$ corresponds to the regular octagon with a vertex at $(1, 0)$, and the Wulff shape is the rotation of $\varphi_{\mathcal{T}_1}^{-1}(1)$ by $\frac{\pi}{8}$.*

Proof. Given a rational direction $\nu \in S^1$, our strategy is to consider a recovery sequence $E_k \xrightarrow[k\to\infty]{} E_\nu$ coming from the Γ-convergence of Proposition 5.3, with E_ν defined as in Equation 5.12. By construction, it holds

$$\lim_k \operatorname{Per}_{\mathcal{T}_k}(E_k) = \operatorname{Per}_\varphi(E_\nu) = \varphi_{\mathcal{T}_1}(\nu) \mathcal{H}^1(\partial E_\nu \cap \Omega).$$

From Lemma 5.7 we can build a modified subsequence $\tilde{E}_{k_m}$ that matches the boundary of E_ν for which

$$\lim_k \operatorname{Per}_{\mathcal{T}_k}(E_k) = \lim_m \operatorname{Per}_{\mathcal{T}_{k_m}}(\tilde{E}_{k_m}).$$

Then, given a sequence of perimeter minimizers with fixed boundary constraint

$$G_{k_m} \in \operatorname*{arg\,min}_{\substack{A \in S_{\mathcal{T}_{k_m}}(\Omega), \\ \partial A \cap \partial\Omega = \partial E_\nu \cap \partial\Omega}} \operatorname{Per}_{\mathcal{T}_{k_m}}(A), \tag{5.18}$$

we obtain

$$\operatorname{Per}_{\mathcal{T}_{k_m}}(G_{k_m}) \leq \operatorname{Per}_{\mathcal{T}_{k_m}}(\tilde{E}_{k_m}).$$

If we have that $G_{k_m} \xrightarrow[m\to\infty]{} E_\nu$ in L^1, then the Γ-$\liminf$ inequality implies

$$\operatorname{Per}_\varphi(E_\nu) \leq \liminf_m \operatorname{Per}_{\mathcal{T}_{k_m}}(G_{k_m}),$$

and ultimately

$$\varphi_{\mathcal{T}_1}(\nu) \mathcal{H}^1(\partial E_\nu \cap \Omega) = \liminf_m \operatorname{Per}_{\mathcal{T}_{k_m}}(G_{k_m}). \tag{5.19}$$

Therefore, our goal now is to describe, for this particular triangulation, the minimal perimeter $\operatorname{Per}_{\mathcal{T}_{k_m}}(G_{k_m})$ and specific (since they are not unique) minimizers G_{k_m} with

$$\mathcal{L}^2(G_{k_m} \triangle E_\nu) \to 0 \tag{5.20}$$

Because the initial triangulation $\mathcal{T}_1$ is invariant under rotations by multiples of $\frac{\pi}{4}$, we can assume $\nu = \frac{1}{\sqrt{\sigma^2 + \tau^2}}(\sigma, -\tau) \in S^1$ for some coprime integers $\sigma, \tau \in \mathbb{N}$ with $0 \leq \sigma \leq \tau$, as in Figure 6(a).
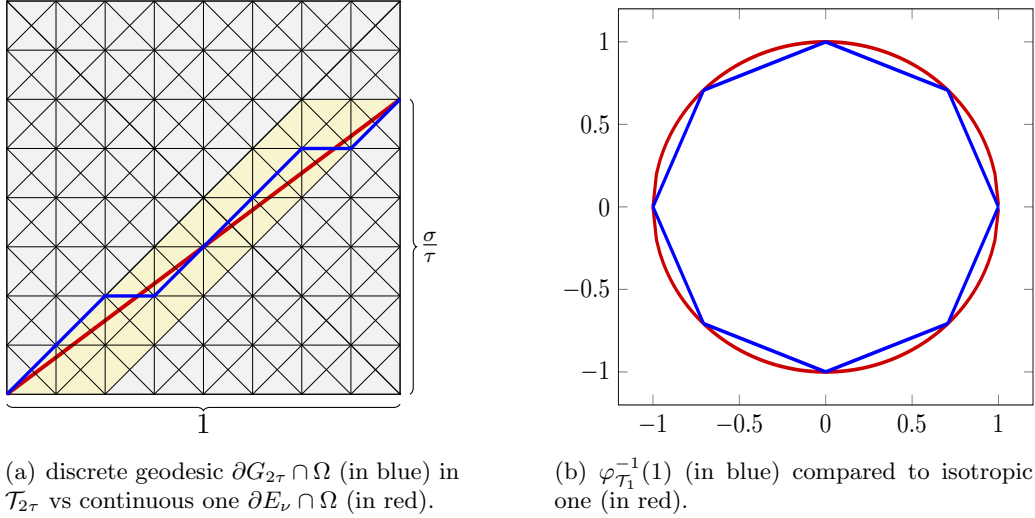
(a) discrete geodesic $\partial G_{2\tau} \cap \Omega$ (in blue) in $\mathcal{T}_{2\tau}$ vs continuous one $\partial E_\nu \cap \Omega$ (in red).

(b) $\varphi_{\mathcal{T}_1}^{-1}(1)$ (in blue) compared to isotropic one (in red).

Figure 6: Description of the anisotropy $\varphi_{\mathcal{T}_1}$ of the 'double diagonal' triangulation $\mathcal{T}_1$.

Note that E_ν is a triangle with vertices at $(0,0), (1,0)$ and $\left(1, \frac{\sigma}{\tau}\right)$, which, by the construction of the subsequence in Lemma 5.7, are all nodes of periodic squares in $\mathcal{T}_{k_m}$ for all m.

Consider the minimization problem (5.18) defining G_{k_m}. We claim that its minimal value can be found by minimizing over paths made up of edges none of which is contained in $\partial\Omega$. To see this, consider an arbitrary competitor A for (5.18). Then, since $\partial A \cap \partial\Omega = \partial E_\nu \cap \partial\Omega$ we have that all of the triangles of $\mathcal{T}_{k_m}$ with an edge contained in either $[0,1] \times \{0\}$ or $\{1\} \times [0, \sigma/\tau]$ must be included in A. Moreover, for each pair T_1, T_2 of those triangles such that $\overline{T_1} \cap \overline{T_2} = \{v_0\}$ for some $v_0 \in \partial\Omega$, we can assume that the two other triangles having v_0 as a vertex are also included in A, since adding them to it cannot increase the perimeter. Therefore, we have obtained that $\partial A \cap \partial\Omega$ must be part of the boundary of an indecomposable component of A, say A_0. Seeing A_0 as a positively oriented simplicial chain, we can consider the induced orientation for ∂A_0, in which each vertex of $\mathcal{T}_{k_m}$ must appear only in an even number of edges with by virtue of being a boundary, and in fact only twice since otherwise A_0 would not be indecomposable. Alternatively, one can use general results for two-dimensional sets of finite perimeter expressing their essential boundaries as unions of Jordan curves, see [3, Thm. 7, Cor. 1] or [21, Rem. II].

With this in mind, we only need to consider injective continuous paths $\gamma : [0,1] \to \overline{\Omega}$ connecting $(0,0)$ with $\left(1, \frac{\sigma}{\tau}\right)$ whose image is made up of (non-boundary) edges of triangles in $\mathcal{T}_{k_m}$. Furthermore, since we are only interested in $\mathcal{H}^1(\gamma([0,1]))$ and for injective curves this quantity equals (see [19, Thm. 2.6.2]) the length of γ defined as a supremum of sums of lengths of polygons with vertices in $\gamma([0,1])$ induced by partitions of $[0,1]$, we can assume γ to be piecewise linear with switching only at the preimages of vertices of triangles in $\mathcal{T}_{k_m}$. Let

$$D_\gamma^{0,1} := \{t_0, ..., t_{n+1}\} \subset [0,1]$$

be the ordered set of points where γ is non-differentiable. Then, $\gamma(D_\gamma^{0,1})$ must be vertices of the triangulation $\mathcal{T}_{k_m}$. It can be proven that if γ is length-minimizing among such curves, then $\gamma(t_{j+1}) - \gamma(t_j)$ must be either (up to positive rescaling) the vector $\nu_1 = (1,0)$ or $\nu_2 = (1,1)$, for all $t_j \in D_\gamma^{0,1}$, which in particular implies that $\gamma(D_\gamma^{0,1})$ consists of nodes of periodic squares inside the polygon with vertices

$$\left\{(0,0), \left(\frac{\sigma}{\tau}, \frac{\sigma}{\tau}\right), \left(1, \frac{\sigma}{\tau}\right), \left(1 - \frac{\sigma}{\tau}, 0\right)\right\},$$

illustrated in yellow in Figure 6(a). Moreover, the length of such an optimal path can be

computed from the amount of edges directed by ν_1 and ν_2 respectively, resulting in

$$\text{Per}_{\mathcal{T}_{k_m}}(G_{k_m}) = \frac{1}{\tau}\left(\sigma\sqrt{2} + (\tau - \sigma)\right) = 1 + (\sqrt{2} - 1)\frac{\sigma}{\tau}. \tag{5.21}$$

In the 'double-diagonal' geometry under consideration these facts are quite intuitive, so we relegate a detailed proof to Appendix B.

The only thing left to justify is that, as $m \to \infty$, one can choose optimal paths of the type we have just described, in such a way that (5.20) is satisfied. One adequate choice is to follow a diagonal path from the origin with $t \mapsto \frac{1}{k_m}t\nu_2$ until the first $t^* \in \mathbb{N}$ satisfying $t^*\nu_2 + \frac{1}{k_m}\nu_1 \notin \mathring{E}_\nu$, concatenating a horizontal segment $t \to \frac{1}{k_m}\nu_2$ for $t \in [t^*, t^* + 1]$, iterating this procedure until hitting the point $(1, \sigma/\tau)$, and rescaling the parametrization so that its domain is $[0,1]$. By construction, the image of a path γ constructed in this fashion satisfies $d(\gamma(t), \partial E_\nu) \leq 1/k_m$ for all t, implying in particular (5.20).

Finally, by combining Equation (5.19) and (5.21), we obtain that, for a rational direction $\nu = \frac{1}{\sqrt{\sigma^2+\tau^2}}(\sigma, -\tau) \in S^1$ with $0 \leq \sigma \leq \tau$,

$$\varphi_{\mathcal{T}_1}(\nu) = (\sqrt{2} - 1, -1) \cdot \nu.$$

By symmetry, we can extend the previous discussion to other rational directions $\nu \in S^1$, finding a description of $\varphi_{\mathcal{T}_1}$ on a dense subset of S^1. Then again, we can extend it to the whole S^1 by continuity and to $\mathbb{R}^2 \setminus \{0\}$ by homogeneity $\varphi_{\mathcal{T}_1}(v) = |v|\varphi_{\mathcal{T}_1}\left(\frac{v}{|v|}\right)$. The $1-$levelset of the resulting map $\varphi_{\mathcal{T}_1}$ is depicted in Figure 6(b), and it is precisely the expected regular octagon. The claim about the corresponding Wulff shape follows by the Wulff theorem [53, Thm. 20.8] and convex duality. $\qquad\square$

6 Numerical examples

Finally, we provide two numerical examples which illustrate the findings of the present work, in particular the practical efficiency of Algorithm 3 as well as the asymptotic behavior of the total variation as explained in Proposition 5.8. For this purpose, let $\mathcal{T}$ be a triangulation of $\Omega = (-1,1)^d$, $d = 2, 3$, either given by the "double-diagonal" triangulation from Figure 5.2 or pseudo-random ones obtained from a mesh generator. Associated to $\mathcal{T}$, we then consider discretized optimal control problems of the form

$$\min_{u_h \in P_0(\mathcal{T})} J_{\mathcal{T}}(u_h) := \left[\frac{1}{2}\|y_h - y_o\|_{L^2(\Omega)}^2 + \alpha\,\text{TV}(u_h, \Omega)\right]$$

where $y_h = K_h u_h \in P_1(\mathcal{T}) \cap H_0^1(\Omega)$ is the unique solution of

$$\text{find } y \in P_1(\mathcal{T}) \cap H_0^1(\Omega) \quad \text{s.t.} \quad \int_\Omega \nabla y \cdot \nabla\theta_h + cy\theta_h \,\mathrm{d}x = \int_\Omega u\theta_h \,\mathrm{d}x \tag{6.1}$$

for all $\theta_h \in P_1(\mathcal{T}) \cap H_0^1(\Omega)$, and $\alpha > 0$, $c \geq 0$ and $y_o \in L^2(\Omega)$ are example specific parameters.

6.1 Implementation of Algorithm 3

We give a brief overview on the implementation details concerning Algorithm 3. First, using standard adjoint calculus, we verify that the dual variable $p_k = (1/\alpha)K_h^*(y_o - K_h u_k) \in V' \simeq P_0(\mathcal{T})$ is given by $p_k = (1/\alpha)\Pi_0(z_k)$ where $\Pi_0 \colon P_1(\mathcal{T}) \to P_0(\mathcal{T})$ denotes the orthogonal projection onto $P_0(\mathcal{T})$ and $z_k \in P_1(\mathcal{T}) \cap H_0^1(\Omega)$ is the unique solution of the adjoint problem

$$\text{find } z \in P_1(\mathcal{T}) \cap H_0^1(\Omega) \quad \text{s.t.} \quad \int_\Omega \nabla z \cdot \nabla\theta_h + cz\theta_h \,\mathrm{d}x = \int_\Omega \theta_h \,\mathrm{d}x$$

for all $\theta_h \in P_1(\mathcal{T}) \cap H_0^1(\Omega)$. Second, alongside the current active set $\mathcal{A}_k = \{u_k^j\}_{j=1}^{N_k}$ and iterate $u_k = \sum_{j=1}^{N_k} \gamma_j^k u_k^j + c^k \mathbb{1}_\Omega$, we also update the associated set of states $\mathcal{S}_k := \{K_h u_k^j\}_{j=1}^{N_k} \cup \{K \mathbb{1}_\Omega\}$, accordingly. As a consequence, and due to the linearity of K_h, every iteration of Algorithm 3 requires two PDE-solves, one for the computation of the state associated to the new characteristic function as well as one solve of the adjoint equation to get p_k. The solution of the ℓ_1-regularized subproblems does not require further PDE-solves since a matrix representation of $K\mathcal{U}_{\mathcal{A}_k}$ is available from $\mathcal{S}_k$. The algorithm is implemented in Python, using FEniCS 2019 for the finite elements computations, the PyMaxflow library (https://github.com/pmneila/PyMaxflow) wrapping the Boykov-Kolmogorov implementation of maxflow [9] for the minimal cuts in the Dinkelbach iteration, as well as a semismooth Newton method based on the normal map approach for the finite dimensional subproblems of the form ($\mathcal{P}(\mathcal{A})$), see e.g. [54]. The latter is warmstarted using γ^k and c^k to construct a good starting point. In all examples, the progress of the algorithm is monitored by the upper bounds ζ_k, see Lemma 2.8, obtained as a by-product of Algorithm 2. We stop Algorithm 3 in iteration $\bar{k} \in \mathbb{N}$ if $\zeta_{\bar{k}} \leq 10^{-10}$ while each application of the semismooth Newton method is terminated once the corresponding residual is below 10^{-14}. The unstructured meshes we use are generated using CGAL through the mshr component of FEniCS. For abbreviation and with some abuse of notation, we set $\bar{u} = u_{\bar{k}}$ in the following. Note that these termination criteria imply

$$J_\mathcal{T}(\bar{u}) - \min_{u_h \in P_0(\mathcal{T})} J_\mathcal{T}(u_h) \leq 10^{-10}$$

in view of Lemma 2.8.

In order to assess the convergence behavior of Algorithm 3, we compute the suboptimality of u_k in terms of its objective functional value residual, i.e.,

$$J_\mathcal{T}(u_k) - J_\mathcal{T}(\bar{u}) \approx J_\mathcal{T}(u_k) - \min_{u_h \in P_0(\mathcal{T})} J_\mathcal{T}(u_h).$$

All calculations were performed on a 2021 Macbook Pro with 10-core M1 Max CPU. Python code for our implementation and configurations to reproduce some of the following examples are available at https://doi.org/10.5281/zenodo.10048384.

6.2   Two spheres

As a first simple example, we choose $c = 0.5$ and $y_o \in P_0(\mathcal{T})$ as the unique solution of (6.1) for the piecewise constant L^2 projection $u = \Pi_0(u_0)$ of

$$u_0 = \mathbb{1}_{D_1} + 2\mathbb{1}_{D_2} - \mathbb{1}_\Omega$$

where D_1 and D_2 are balls centered at $1/3\,\mathbf{1}$ and $-1/3\,\mathbf{1}$ with radius $\sqrt{0.15}$ as well $\sqrt{0.1}$, respectively. Here, $\mathbf{1} \in \mathbb{R}^d$ denotes the all-ones vector.

The regularization parameter is set to $\alpha = 10^{-5}$ for both $d = 2$ and $d = 3$. The obtained results from Algorithm 3 for $d = 2, 3$ on pseudo-random triangulations are depicted in Figure 7. Note that the computed approximations contain level sets which are not present in u_0. Indeed, even for low values of α, a loss of contrast is expected, so the input u_0 is never an exact minimizer of J; see [27] for a detailed analysis of this phenomenon in the planar denoising case with $K = \mathrm{Id}$. The presence of the catenoids in the outcomes of the algorithm reduces the value of the total variation while keeping the fidelity term close to zero. In particular, for the present example, we have

$$J_\mathcal{T}(\Pi_0(u_0)) = \alpha\,\mathrm{TV}(\Pi_0(u_0), \Omega) \approx \begin{cases} 8.29 \cdot 10^{-5} & d = 2 \\ 7.02 \cdot 10^{-5} & d = 3 \end{cases} \quad \text{and} \quad J_\mathcal{T}(u_{\bar{k}}) \approx \begin{cases} 6.15 \cdot 10^{-5} & d = 2 \\ 4.87 \cdot 10^{-5} & d = 3, \end{cases}$$
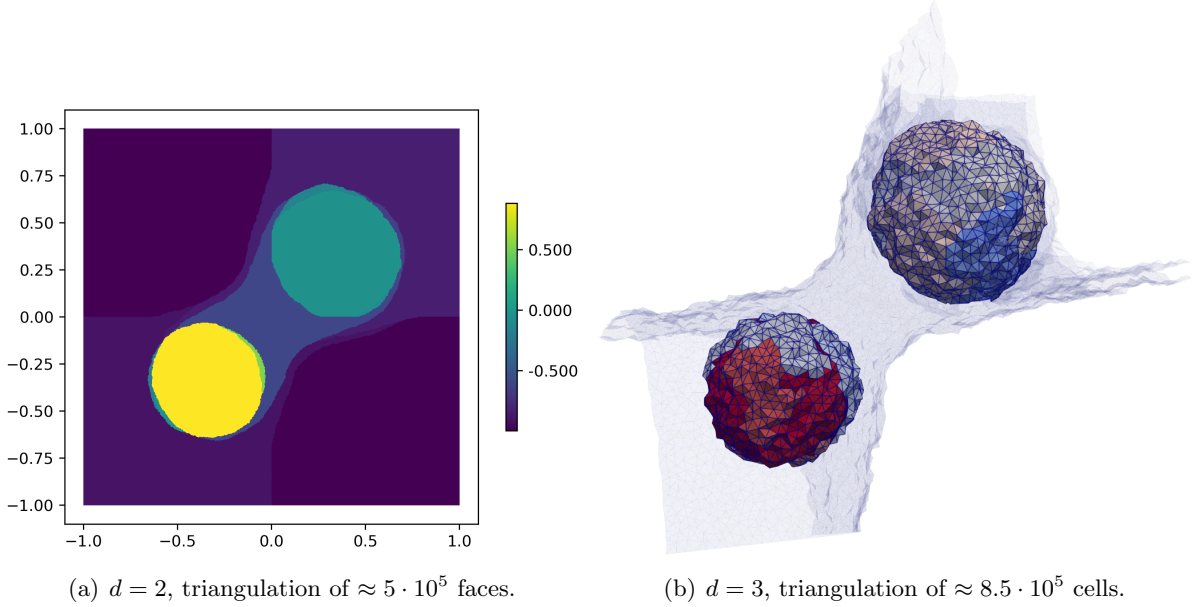
supporting this observation.

(a) $d = 2$, triangulation of $\approx 5 \cdot 10^5$ faces.

(b) $d = 3$, triangulation of $\approx 8.5 \cdot 10^5$ cells.

Figure 7: Output $\bar{u}$ of Algorithm 3 for $d = 2, 3$ on pseudo-random triangulations.

6.3 The castle

As a second, more, challenging setting, set $c = 0$ and $y_o = \mathbb{1}_{(-1/2,1/2)^d}$, see also [35, 45, 55] for the case of $d = 2$. In the absence of regularization, to perfectly match this observation one would need to have as a control the distributional Laplacian Δy_o, a distribution of order 2 which in particular does not belong to $\mathrm{BV}(\Omega)$. For this reason, the solution $\bar{u}$ of the regularized problem is expected to be more complex than the solution of the first example. The regularization parameter is set to $\alpha = 10^{-4}$.

(a) pseudo-random triangulation

(b) 'double diagonal' triangulation
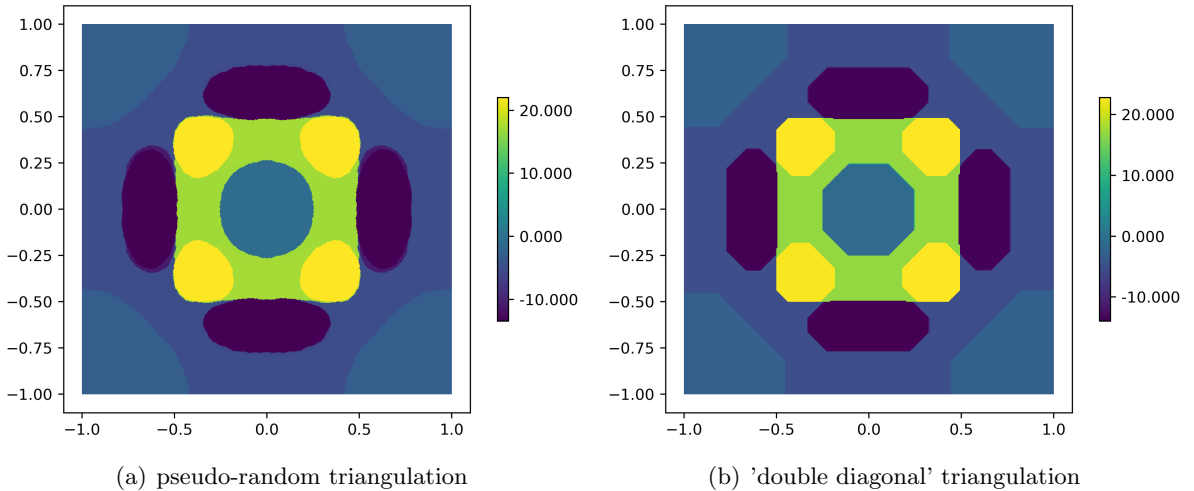
Figure 8: Top view of optimal control numerically computed through Algorithm 3 from different triangulations of $\approx 5 \cdot 10^5$ faces, $d = 2$ and $\alpha = 10^{-4}$.

For $d = 2$ and a pseudo-random triangulation $\mathcal{T}$ of Ω, the output of Algorithm 3 is depicted in Figure 8(a). This result is consistent with [45] where the authors computed the minimizer of problem $(\mathcal{P})$ for $V = P_1(\mathcal{T})$. When considering a regular 'double diagonal' triangulation $\mathcal{T}$ instead, the algorithm's outcome, see Figure 8(b), still shares some similarities to Figure 8(a) but clearly reflects the anisotropic behavior of the total variation in this setting, as we

37

previously discussed in Proposition 5.8. Finally, Figure 9 depicts the solution for $d = 3$ and a pseudo-random triangulation.
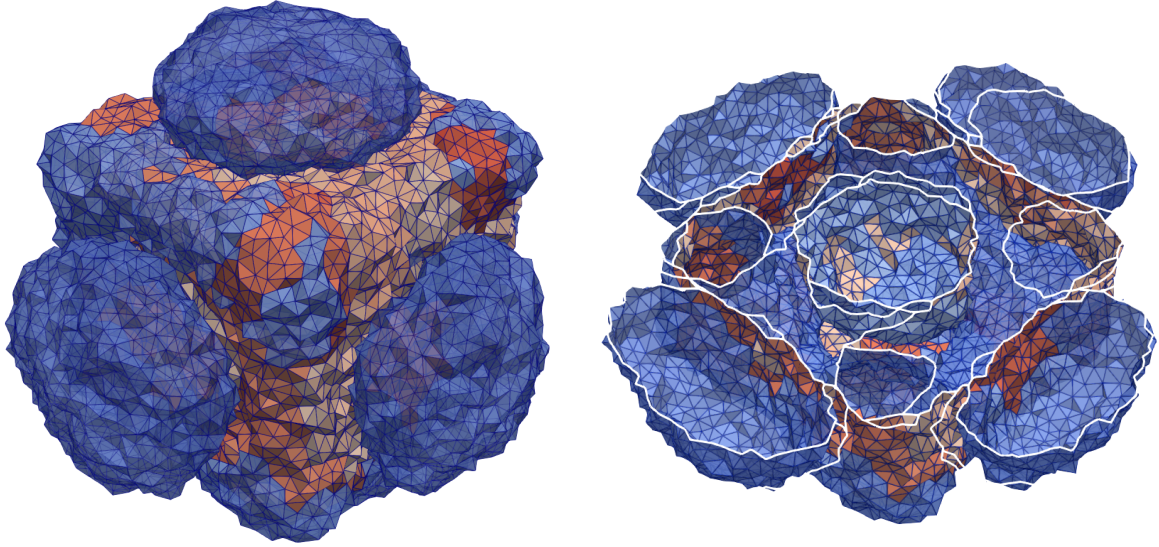

Figure 9: The plot shows where the variation of u_k between cells happens. The color gradient, ranging from blue to red, is an indicator of the magnitude of the variation. The setting is a pseudo-random triangulation of $\approx 8.5 \cdot 10^5$ cells, $d = 3$, and $\alpha = 10^{-4}$.

6.3.1 Practical performance & discussion

In order to investigate the practical performance of Algorithm 3, we plot the upper bounds ζ_k as well as the residuals in terms of the objective functional value associated to u_k for both examples and $d = 2$. We refer to Figure 10 where the convergence criterion was met after $\bar{k} = 66$ and $\bar{k} = 95$, respectively. Several observations can be made: First, as claimed by Lemma 2.8, ζ_k provides an upper bound on the residual. Furthermore, for the present examples, it solely deviates by a maximal factor of 10^2 from the latter. Thus, it represents an easily accessible and reliable convergence criterion. Second, as we can see from both plots, Algorithm 3 admits a global linear rate of convergence rather than a sublinear one as predicted by Proposition 4.4. While such improved convergence results can also be deduced from the abstract results in [16], their application to our setting is nontrivial and beyond the scope of the current work. Finally, we point out that, in both examples, we observed that there holds $\mathrm{TV}(u_k, \Omega) = \sum_{j=1}^{N_k} \gamma_j^k$ throughout every iteration. As a consequence, the residual exhibits a monotone convergence behavior.

We also want to point out that Algorithm 3 exhibits remarkable computational speed. We largely attribute this to the fast convergence of the Dinkelbach method, usually within 4/5 steps, as well as the efficiency of the employed minimal cut implementation [9]. Moreover, due to the mentioned warm starting strategy, the employed semismooth Newton method also converges in few steps. Coupled with the observed improved convergence behavior of Algorithm 3, this results in noteworthy runtime performance. In order to illustrate this observation, we again consider the second example on pseudo-random triangulations and plot the energy $J_{\mathcal{T}}(u_k)$ against the computational time. For comparison, we also consider the path-following based method from [45]. An implementation of the latter using $P_1(\mathcal{T})$ as discretized control space can be found at the source files of https://arxiv.org/abs/2010.11628. We particularly point out that Algorithm 3 provides near optimal iterates before the first subproblem in the path-following method is solved, see Figure 11. However, we also see that, while Algorithm 3 converges rapidly in the first iterations, its per-iteration decrease in the energy quickly deteriorates. For a closer investigation of this observation, we plot the relative change of the iterate u_k w.r.t to the L^1-norm for both
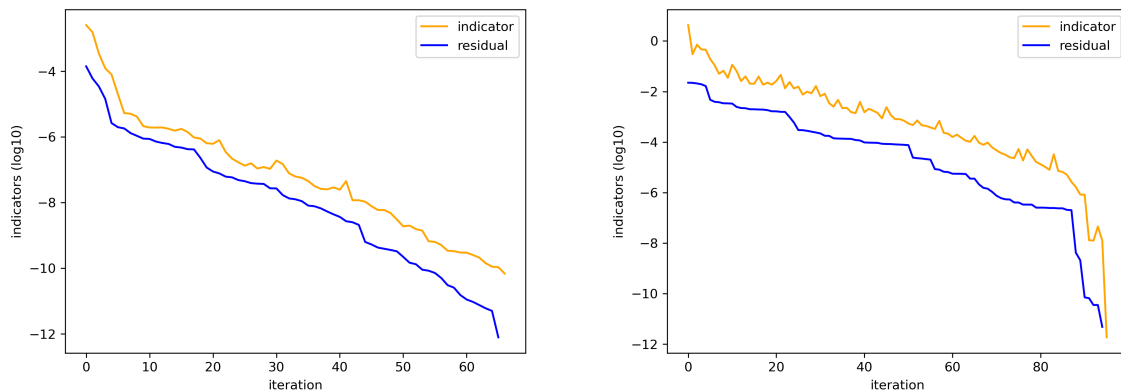
(a) Indicator and residual for the example of Fig. 7(b). (b) Indicator and residual for the example of Fig. 8(a).

Figure 10: Convergence plots for 3D spheres (left) and 2D castle (right)

examples in Figure 12.
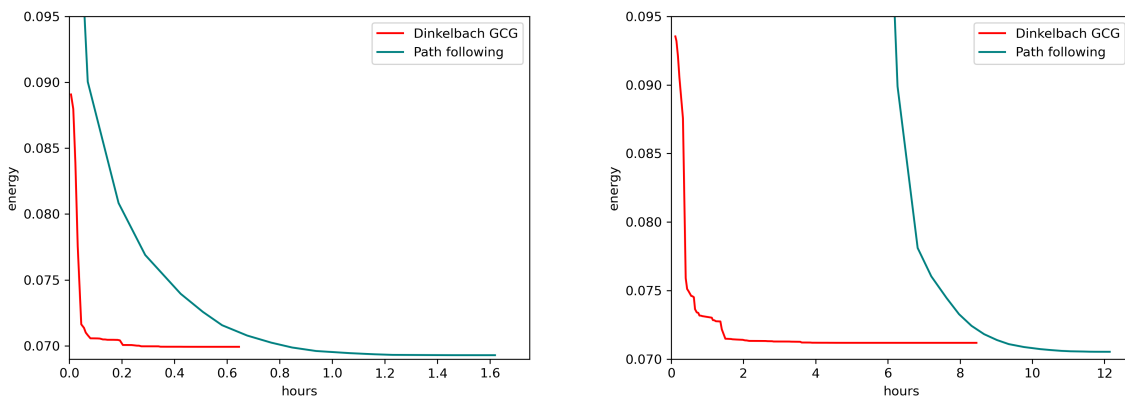
(a) triangulation of $\approx 1.25 \cdot 10^5$ faces   (b) triangulation of $\approx 5 \cdot 10^5$ faces

Figure 11: Comparison in terms of computational time

These results suggest that, after few iterations, u_k only changes slightly, which we suspect to be a consequence of clustering effects in the method, i.e. for k large enough, we have that $\min_{u \in \mathcal{A}_k} \|u^k_{n^+_k} - u\|_{L^1(\Omega)}$ is very small. In this case, faster convergence can be expected if additional "sliding" steps are included, i.e., instead of adding new scaled characteristic functions to the active set in very iteration, we also allow, in some sense, optimal deformations of those already included in $\mathcal{A}_k$. We again point out [38] for a potential practical realization of such a step for simpler deconvolution problems. However, similarly to sparse minimization problems, see, a straightforward inclusion of a sliding step might prove difficult due to the PDE-constraint. Moreover, additional speedup can potentially expected if the graph cut method can be warm-started using elements of the active set as initial guess. This is, however, not supported by the PyMaxflow library we used in our implementation. Since the present work is conceptual and puts the focus on the feasibility of FC-GCG methods for total variation regularization, both topics, sliding steps as well as warmstarting of the graph cut, are postponed to future work. Finally, it's important to acknowledge that the path-following seemingly yields smaller function values if run long enough. We attribute this behavior to the selection of different minimization spaces in both implementations, $V = P_0(\mathcal{T})$ for Algorithm 3 and $V = P_1(\mathcal{T})$ for [45], as well as

39

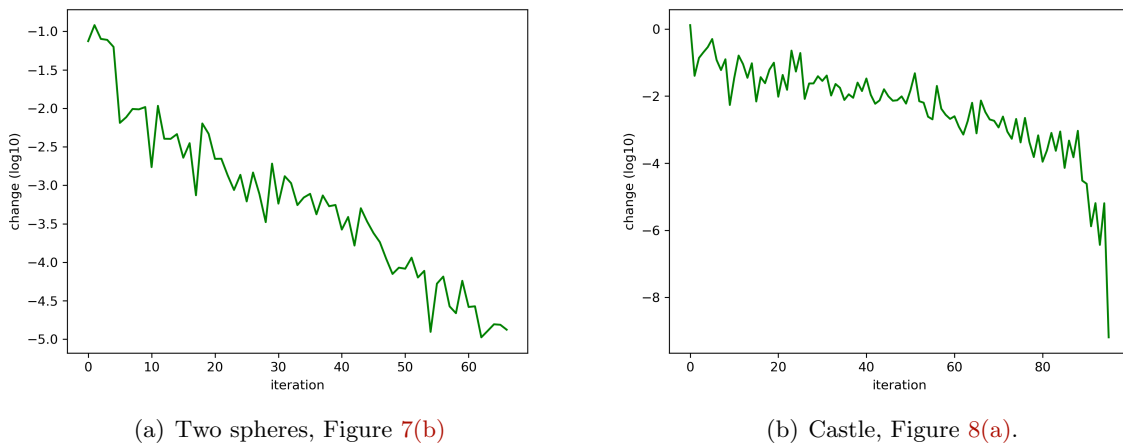(a) Two spheres, Figure 7(b) (b) Castle, Figure 8(a).

Figure 12: Relative change $\|u_{k+1} - u_k\|_{L^1(\Omega)} / \|u_k\|_{L^1(\Omega)}$.

the resulting anisotropic nature of the total variation, see Section 5, rather than inaccuracies in Algorithm 3.

Acknowledgements

A   Proofs related to Section 2

In this section, we provide the remaining technical proofs of Section 2 which were omitted until now for the sake of brevity.

Proof of Proposition 2.1. The weak-to-strong continuity of K follows immediately from Assumption A, **A**1. Now, in order to show that $\bar{c}$ is at least two times continuously differentiable, note that Assumption A, **A**3, implies $K_{\mathbb{1}} \neq 0$. Consider the nonlinear function

$$H(y, c) = \int_\Omega K^* \nabla F(y + cK_{\mathbb{1}}) \, \mathrm{d}x$$

Since F is thrice continuously differentiable, $H(\cdot, \cdot)$ is of class $\mathcal{C}^2$. According to Remark 2.2, there holds $H(y, \bar{c}(y)) = 0$ for all $y \in Y$. Moreover, the strict convexity of F implies

$$\partial_c H(y, c) = (K_{\mathbb{1}}, \nabla^2 F(y + \bar{c}(y)K_1)K_{\mathbb{1}})_Y > 0.$$

Hence, the differentiability of $\bar{c}$ is a consequence of the implicit function theorem as well as of the smoothness of F. Noting that

$$\mathcal{F}(y) = F(y + \bar{c}K_{\mathbb{1}})$$

we immediately conclude that $\mathcal{F}$ is two times continuously Fréchet differentiable and its gradient is Lipschitz continuous on compact sets. Furthermore, the chain rule together with Remark 2.2 implies

$$(\nabla \mathcal{F}(y), \delta y)_Y = (\nabla F(y + \bar{c}(y)K_{\mathbb{1}}), \delta y)_Y + (\nabla \bar{c}, \delta y)_Y \int_\Omega K^* \nabla F(y + \bar{c}(y)K_{\mathbb{1}}) \, \mathrm{d}x$$

$$= (\nabla F(y + \bar{c}(y)K_{\mathbb{1}}), \delta y)_Y$$

40

for all $\delta y \in Y$ which finally yields the characterization of $\nabla \mathcal{F}(y)$. $\qquad \square$

Proof of Proposition 2.7. If Algorithm 1 stops after $\bar{k}$ iterations, the function $\mathring{u}_{\bar{k}}$ is a minimizer of ($\mathfrak{P}$), see [16, Proposition 3.1]. Consequently, according to Proposition 2.3, the function $u_{\bar{k}} = \mathring{u}_{\bar{k}} + c^{\bar{k}}$ is a solution of ($\mathcal{P}$). Now assume that Algorithm 1 does not stop after finitely many iterations. Using [16, Theorem 3.3], the definition of $\mathcal{F}$ and c_k, respectively, as well as Proposition 2.3, we conclude

$$ J(u_k) - \min_{u \in L^q(\Omega)} J(u) = \mathcal{J}(\mathring{u}_k) - \min_{v \in \mathring{L}^q(\Omega)} \mathcal{J}(v) \leq \frac{c}{1+k}. $$

Moreover, the sequence $\mathring{u}_k$ is bounded in $\mathring{L}^q(\Omega)$ and note that u_k satisfies

$$ u_k = \mathring{u}_k + c^k = \mathring{u}_k + \bar{c}(K\mathring{u}_k). $$

Since the mapping $\bar{c}$ is continuous and $\{K\mathring{u}_k\}_k$ is strongly compact, we conclude that the sequence u_k is bounded and thus admits at least one subsequence converging weakly in $L^q(\Omega)$, denoted by the same subscript, and with limit $\bar{u} \in V$, since V is weakly closed. Since J is weakly* lower semicontinuous on $\mathrm{BV}(\Omega)$ and K is weak-to-strong continuous, we conclude

$$ 0 \leq J(\bar{u}) - \min_{u \in V} J(u) \leq \liminf_{k \to \infty}[J(u_k) - \min_{u \in V} J(u)] = 0, \quad F(Ku_k) \to F(K\bar{u}). $$

Hence, $\bar{u}$ is a minimizer of ($\mathcal{P}$) and $\mathrm{TV}(u_k, \Omega) \to \mathrm{TV}(\bar{u}, \Omega)$, i.e., u_k converges strictly. Finally, since the selected subsequence was arbitrary, we conclude the minimality of every strict accumulation point. $\qquad \square$

Proof of Lemma 2.8. Let $\bar{u}$ denote a minimizer of ($\mathcal{P}$). By one homogeneity and convexity of $\mathrm{TV}(\cdot, \Omega)$ as well as $\mathrm{TV}(u_j^k) = 1$, there holds

$$ \alpha \mathrm{TV}(u_k, \Omega) \leq \sum_{j=1}^{N_k} \gamma_j^k. $$

As a consequence, we conclude

$$ \alpha \mathrm{TV}(\bar{u}) \leq J(\bar{u}) \leq J(u_k) \leq F\left(K\mathcal{U}_{\mathcal{A}_k}(\gamma^k) + c^k K_{\mathbb{1}}\right) + \sum_{j=1}^{N_k} \gamma_j^k, $$

i.e., $\mathrm{TV}(\bar{u}) \leq M_k$. By convexity of F, there holds

$$ J(u_k) - \min_{u \in L^q(\Omega)} J(u) \leq \int_\Omega p_k(\bar{u} - u_k) \ \mathrm{d}x + \sum_{j=1}^{N_k} \gamma_j^k = \int_\Omega p_k\bar{u} \ \mathrm{d}x - \alpha \mathrm{TV}(\bar{u}, \Omega) $$

where the equality follows from the first-order optimality conditions of ($\mathcal{P}(\mathcal{A}_k)$). We further estimate

$$ \int_\Omega p_k\bar{u} \ \mathrm{d}x - \alpha \mathrm{TV}(\bar{u}, \Omega) \leq \alpha \mathrm{TV}(\bar{u}, \Omega)\left(\max_{v \in B} \int_\Omega p_k v \ \mathrm{d}x - 1\right) \leq M_k\left(\int_\Omega p_k\widehat{v}_k \ \mathrm{d}x - 1\right), $$

finishing the proof. $\qquad \square$

B   Discrete geodesics on regular triangulations

In this section, we give a formal proof for the claimed properties of length-minimizing piecewise linear and continuous curves connecting the points $(0,0)$ and $(1, \sigma/\tau)$ along interior edges of the triangulation $\mathcal{T}_{k_m}$. More in detail, with a slight abuse of notation, define $\mathcal{E} := \bigcup_{T \in \mathcal{T}_{k_m}} \partial T$ and denote by $\mathcal{N}$, the set of nodes of the mesh. Moreover, recalling the reasoning in the proof of Proposition 5.8, it is sufficient to consider piecewise linear, continuous and injective curves $\gamma : [0,1] \to \mathcal{E}$. Let the set of those curves be denoted by $\mathcal{C}(\mathcal{E})$. Note that each of these curves is differentiable everywhere apart from a finite ordered set

$$D_\gamma^{0,1} := \{t_0^\gamma, \cdots, t_{N_\gamma+1}^\gamma\} \quad \text{with} \quad 0 = t_0^\gamma < \cdots < t_{N_\gamma+1}^\gamma = 1,$$

as well as $\gamma(D_\gamma^{0,1}) \subset \mathcal{N}$. Its length, $L(\gamma) := \mathcal{H}^1(\gamma([0,1]))$, is given by

$$L(\gamma) = \sum_{j=0}^{N_\gamma} |\gamma(t_j^\gamma) - \gamma(t_{j+1}^\gamma)|.$$

With these prerequisites and for some coprime integers $\sigma, \tau \in \mathbb{Z}$ with $0 < \sigma < \tau$, consider the length minimization problem

$$\min_{\gamma \in \mathcal{C}(\mathcal{E})} L(\gamma) \quad \text{s.t.} \quad \gamma(0) = (0,0), \quad \gamma(1) = (1, \sigma/\tau) \tag{B.1}$$

as well as

$$\gamma(t) \in \mathcal{E} \setminus (((0,0), (1,0)] \cup ((1,0), (1, \sigma/\tau))) \quad \text{for all} \quad t \in [0,1]. \tag{B.2}$$

Since its admissible set is finite, the problem admits at least one solution. Indeed, the following characterization of its minimizers holds true.

Theorem B.1. *An admissible curve $\gamma \in \mathcal{C}(\mathcal{E})$ is a minimizer of* (B.1) *if and only if there holds*

$$\frac{\gamma(t_{j+1}^\gamma) - \gamma(t_j^\gamma)}{|\gamma(t_{j+1}^\gamma) - \gamma(t_j^\gamma)|} = \nu_j \quad \text{for all} \quad j = 0, \ldots, N_\gamma \tag{B.3}$$

where $\nu_j \in \{\nu^1, \nu^2\}$ and $\nu^1 = (1/\sqrt{2})(1,1), \ \nu^2 = (1,0)$.

Note that it suffices to show that a minimizer $\bar\gamma$ of (B.1) necessarily satisfies (B.3) since all admissible curves $\gamma \in \mathcal{C}(\mathcal{E})$ obeying (B.3) have the same length

$$L(\gamma) = 1 + \left(\sqrt{2} - 1\right) \frac{\sigma}{\tau}.$$

The former is a direct consequence of the following proposition:

Proposition B.2. *Let $\bar\gamma \in \mathcal{C}(\mathcal{E})$ be an optimal path and denote by $\bar\gamma_i, \ i = 1, 2$ its component functions. For every $j = 1, \ldots, N_{\bar\gamma}$ and all $t \geq t_j^\gamma$, we have*

$$\bar\gamma(t) \in \{\bar\gamma(t_j^\gamma)\} + \mathrm{cone}\{\nu_1, \nu_2\}.$$

Proof. In the following, we make frequent use of the fact that we can concatenate piecewise linear and continuous paths at nodes of triangulation, provided that both parts satisfy the correct boundary conditions. We start by proving that there holds

$$\bar\gamma_2(t) \leq \sigma/\tau, \quad \bar\gamma_1(t) \leq 1 - \sigma/\tau + \gamma_2(t) \quad \text{for all} \quad t \in [0,1].$$

Since $\bar{\gamma}$ is a piecewise linear and continuous path between nodes and satisfies the boundary condition $\bar{\gamma}(1) = (1, \sigma/\tau)$, it suffices to show that these estimates are satisfied whenever $\bar{\gamma}(\bar{t}) \in \mathcal{N}$ for some $\bar{t} \in [0, 1]$. Indeed, assume that there is $\bar{t} \in (0, 1)$ such that $\bar{\gamma}_2(\bar{t}) > \sigma/\tau$ and $\bar{\gamma}(\bar{t}) \in \mathcal{N}$. Then, by continuity of $\bar{\gamma}$ as well as due to the boundary conditions, there are $\bar{t}_1 < \bar{t} < \bar{t}_2$ with $\bar{\gamma}(\bar{t}_i) \in \mathcal{N}$ and $\bar{\gamma}_2(\bar{t}_i) = \sigma/\tau$, $i = 1, 2$. Define the horizontal augmented path

$$\tilde{\gamma}(t) = \begin{cases} \bar{\gamma}(t) & t \in [0, \bar{t}_1] \cup [\bar{t}_2, 1] \\ \frac{t - \bar{t}_2}{\bar{t}_1 - \bar{t}_2} \bar{\gamma}(\bar{t}_1) - \frac{t - \bar{t}_1}{\bar{t}_1 - \bar{t}_2} \bar{\gamma}(\bar{t}_2) & \text{else} \end{cases}.$$

Then $\tilde{\gamma}$ is admissible for (B.1) but there holds $L(\tilde{\gamma}) < L(\bar{\gamma})$, contradicting the minimality of $\bar{\gamma}$. For the estimate on $\bar{\gamma}_1$, we proceed analogously. Now, due to the constraint (B.2) as well as the derived estimates, we conclude

$$(1, \sigma/\tau) \in \{\bar{\gamma}(t_j^\gamma)\} + \operatorname{cone}\{\nu_1, \nu_2\}$$

for all $j = 1, \ldots, N_{\bar{\gamma}}$, noting that for every $x \in \Omega$ with $x_2 \leq \sigma/\tau$ and $x_1 \leq 1 - \sigma/\tau + x_2$, the set

$$(\{x\} + \operatorname{cone}\{\nu_1, \nu_2\}) \cap \Omega$$

is a triangle containing $(1, \sigma/\tau)$. In order to prove the statement claimed in the theorem, and w.l.o.g, assume that there is $t_j^\gamma \in D_{\bar{\gamma}}^{0,1}$ as well as $\bar{t} > t_j^\gamma$ with $\bar{\gamma}(\bar{t}) \in \mathcal{N}$ but

$$\bar{\gamma}(\bar{t}) \notin \{\bar{\gamma}(t_j^\gamma)\} + \operatorname{cone}\{\nu_1, \nu_2\}.$$

Then, due to continuity of $\bar{\gamma}$ as well as

$$(1, \sigma/\tau) \in \{\bar{\gamma}(t_j^\gamma)\} + \operatorname{cone}\{\nu_1, \nu_2\},$$

we have that $\bar{\gamma}$ necessarily intersects

$$\{\bar{\gamma}(t_j^\gamma)\} + \operatorname{cone}\{\nu_1\} \quad \text{or} \quad \{\bar{\gamma}(t_j^\gamma)\} + \operatorname{cone}\{\nu_2\}$$

for some $t > \bar{t}_2$. Repeating the previous construction, i.e. augmenting $\bar{\gamma}$ along ν_1 or ν_2, respectively, again yields a contradiction. This finishes the proof. $\square$

Proof of Theorem B.1. Assume that there is $t_j^\gamma \in D_{\bar{\gamma}}^{0,1}$ such that (B.3) does not hold. Then we also have

$$\bar{\gamma}(t_{j+1}^\gamma) \notin \{\bar{\gamma}(t_j^\gamma)\} + \operatorname{cone}\{\nu_1, \nu_2\}$$

yielding a contradiction to Proposition B.2. $\square$

References

[1] R. Alicandro, M. Cicalese, and M. Ruf. Domain formation in magnetic polymer composites: an approach via stochastic homogenization. *Arch. Ration. Mech. Anal.*, 218(2):945–984, 2015.

[2] L. Ambrosio, S. Aziznejad, C. Brena, and M. Unser. Linear inverse problems with Hessian-Schatten total variation. Preprint arXiv:2210.04077 [math.FA], 2022.

[3] L. Ambrosio, V. Caselles, S. Masnou, and J.-M. Morel. Connected components of sets of finite perimeter and applications to image processing. *J. Eur. Math. Soc. (JEMS)*, 3(1):39–92, 2001.

[4] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. Oxford University Press, New York, 2000.

[5] S. Bartels. Total variation minimization with finite elements: convergence and iterative solution. *SIAM J. Numer. Anal.*, 50(3):1162–1180, 2012.

[6] A. H. Bentbib, A. Bouhamadi, and K. Kreit. A conditional gradient method for primal-dual total variation-based image denoising. *ETNA, Electron. Trans. Numer. Anal.*, 48:310–328, 2018.

[7] M. Bergounioux, X. Bonnefond, T. Haberkorn, and Y. Privat. An optimal control problem in photoacoustic tomography. *Math. Models Methods Appl. Sci.*, 24(12):2525–2548, 2014.

[8] C. Boyer, A. Chambolle, Y. De Castro, V. Duval, F. De Gournay, and P. Weiss. On representer theorems and convex regularization. *SIAM J. Optim.*, 29(2):1260–1281, 2019.

[9] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26:1124–1137, 09 2004.

[10] A. Braides. Γ-*convergence for beginners*, volume 22 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2002.

[11] A. Braides, M. Maslennikov, and L. Sigalotti. Homogenization by blow-up. *Appl. Anal.*, 87(12):1341–1356, 2008.

[12] A. Braides and A. Piatnitski. Homogenization of surface and length energies for spin systems. *J. Funct. Anal.*, 264(6):1296–1328, 2013.

[13] K. Bredies and M. Carioni. Sparsity of solutions for variational inverse problems with finite-dimensional data. *Calc. Var. Partial Differential Equations*, 59(1):1–26, 2020.

[14] K. Bredies, M. Carioni, S. Fanzon, and F. Romero. On the extremal points of the ball of the Benamou-Brenier energy. *Bull. Lond. Math. Soc.*, 53(5):1436–1452, 2021.

[15] K. Bredies, M. Carioni, S. Fanzon, and F. Romero. A generalized conditional gradient method for dynamic inverse problems with optimal transport regularization. *Found. Comput. Math.*, 23(3):833–898, 2023.

[16] K. Bredies, M. Carioni, S. Fanzon, and D. Walter. Asymptotic linear convergence of fully-corrective generalized conditional gradient methods. *Math. Program.*, 0(0, Ser. A):000–000, 2023.

[17] K. Bredies and H. K. Pikkarainen. Inverse problems in spaces of measures. *ESAIM Control Optim. Calc. Var.*, 19(1):190–218, 2013.

[18] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[19] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[20] P. Bělík and M. Luskin. Approximation by piecewise constant functions in a BV metric. *Math. Models Methods Appl. Sci.*, 13(3):373–393, 2003.

[21] L. A. Caffarelli and N. M. Rivière. On the rectifiability of domains with finite perimeter. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 3(2):177–186, 1976.

[22] M. Carioni, J. A. Iglesias, and D. Walter. Extremal points and sparse optimization for generalized Kantorovich-Rubinstein norms. Preprint arXiv:2209.09167 [math.OC], 2022.

[23] E. Casas, F. Kruse, and K. Kunisch. Optimal control of semilinear parabolic equations by BV-functions. *SIAM J. Control Optim.*, 55(3):1752–1788, 2017.

[24] E. Casas, K. Kunisch, and C. Pola. Some applications of BV functions in optimal control and calculus of variations. In *Control and partial differential equations (Marseille-Luminy, 1997)*, volume 4 of *ESAIM Proc.*, pages 83–96. Soc. Math. Appl. Indust., Paris, 1998.

[25] E. Casas, K. Kunisch, and C. Pola. Regularization by functions of bounded variation and applications to image enhancement. *Appl. Math. Optim.*, 40(2):229–257, 1999.

[26] V. Caselles, A. Chambolle, and M. Novaga. Total variation in imaging. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, pages 1455–1499. Springer, New York, NY, 2015.

[27] V. Caselles, M. Novaga, and C. Pöschl. TV denoising of two balls in the plane. Preprint arXiv:1605.00247 [math.FA], 2016.

[28] A. Chambolle and J. Darbon. On total variation minimization and surface evolution using parametric maximum flows. *Int. J. Comput. Vis.*, 84(3):288–307, 2009.

[29] A. Chambolle, A. Giacomini, and L. Lussardi. Continuous limits of discrete perimeters. *M2AN Math. Model. Numer. Anal.*, 44(2):207–230, 2010.

[30] A. Chambolle and L. Kreutz. Crystallinity of the homogenized energy density of periodic lattice systems. *Multiscale Model. Simul.*, 21(1):34–79, 2023.

[31] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.*, 40(1):120–145, 2011.

[32] A. Chambolle and T. Pock. Approximating the total variation with finite differences or finite elements. In *Geometric partial differential equations. Part II*, volume 22 of *Handb. Numer. Anal.*, pages 383–417. Elsevier/North-Holland, Amsterdam, 2021.

[33] R. Chan, T. F. Chan, and A. Yip. Numerical methods and applications in total variation image restoration. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, pages 1501–1537. Springer New York, New York, NY, 2015.

[34] C. Clason, F. Kruse, and K. Kunisch. Total variation regularization of multi-material topology optimization. *ESAIM, Math. Model. Numer. Anal.*, 52(1):275–303, 2018.

[35] C. Clason and K. Kunisch. A duality-based approach to elliptic control problems in non-reflexive Banach spaces. *ESAIM Control Optim. Calc. Var.*, 17(1):243–266, 2011.

[36] G. Dal Maso. *An introduction to Γ-convergence*, volume 8 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser Boston, Inc., Boston, MA, 1993.

[37] J. Darbon and M. Sigelle. Image restoration with discrete constrained total variation. I. Fast and exact optimization. *J. Math. Imaging Vision*, 26(3):261–276, 2006.

[38] Y. De Castro, V. Duval, and R. Petit. Towards off-the-grid algorithms for total variation regularized inverse problems. *J. Math. Imaging Vision*, 65(1):53–81, 2023.

[39] O. L. Elvetun and B. F. Nielsen. The split Bregman algorithm applied to PDE-constrained optimization problems with total variation regularization. *Comput. Optim. Appl.*, 64(3):699–724, 2016.

[40] S. Engel and K. Kunisch. Optimal control of the linear wave equation by time-depending BV-controls: a semi-smooth Newton approach. *Math. Control Relat. Fields*, 10(3):591–622, 2020.

[41] S. Engel, B. Vexler, and P. Trautmann. Optimal finite element error estimates for an optimal control problem governed by the wave equation with controls of bounded variation. *IMA J. Numer. Anal.*, 41(4):2639–2667, 2021.

[42] W. H. Fleming. Functions with generalized gradient and generalized surfaces. *Ann. Mat. Pura Appl. (4)*, 44:92, 93–103, 1957.

[43] T. Goldstein and S. Osher. The split bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[44] D. Hafemeyer. Regularization and discretization of a BV-controlled elliptic problem: A completely adaptive approach. Master's thesis, Technische Universität München, 2017.

[45] D. Hafemeyer and F. Mannel. A path-following inexact Newton method for PDE-constrained optimal control in BV. *Comput. Optim. Appl.*, 82(3):753–794, 2022.

[46] D. Hafemeyer, F. Mannel, I. Neitzel, and B. Vexler. Finite element error estimates for one-dimensional elliptic optimal control by BV-functions. *Math. Control Relat. Fields*, 10(2):333–363, 2020.

[47] Z. Harchaoui, A. Juditsky, and A. Nemirovski. Conditional gradient algorithms for norm-regularized smooth convex optimization. *Math. Program.*, 152(1-2, Ser. A):75–112, 2015.

[48] E. Herberg and M. Hinze. Variational discretization of one-dimensional elliptic optimal control problems with BV functions based on the mixed formulation. *Math. Control Relat. Fields*, 13(2):695–720, 2023.

[49] M. Herrmann, R. Herzog, S. Schmidt, J. Vidal-Núñez, and G. Wachsmuth. Discrete total variation with finite elements and applications to imaging. *J. Math. Imaging Vis.*, 61(4):411–431, 2019.

[50] S. Hilb. *Total variation minimization via dual-based methods and its discretization aspects.* PhD thesis, Universität Stuttgart, 2023.

[51] J. A. Iglesias and D. Walter. Extremal points of total generalized variation balls in 1D: characterization and applications. *J. Convex Anal.*, 29(4):1251–1290, 2022.

[52] B. Jensen and T. Valkonen. A nonsmooth primal-dual method with interwoven PDE constraint solver. Preprint arXiv:2211.04807 [math.FA], 2022.

[53] F. Maggi. *Sets of finite perimeter and geometric variational problems*, volume 135 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2012.

[54] R. Milzarek and M. Ulbrich. A semismooth Newton method with multidimensional filter globalization for l_1-optimization. *SIAM J. Optim.*, 24(1):298–333, 2014.

[55] C. Natemeyer and D. Wachsmuth. A penalty scheme to solve constrained non-convex optimization problems in $BV(\Omega)$. *Pure Appl. Funct. Anal.*, 7(5):1857–1880, 2022.

[56] M. Ng, L. Qi, Y.-F. Yang, and Y.-M. Huang. On semismooth Newton's methods for total variation minimization. *J. Math. Imaging Vision*, 27(3):265–276, 2007.

[57] R. Parhi and R. D. Nowak. Banach space representer theorems for neural networks and ridge splines. *J. Mach. Learn. Res.*, 22:Paper No. 43, 40, 2021.

[58] J.-C. Picard and H. D. Ratliff. Minimum cuts and related problems. *Networks*, 5:357–370, 1975.

[59] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259–268, 1992.

[60] W. Rudin. *Functional analysis.* International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, second edition, 1991.

[61] S. Schaible. Fractional programming. II: On Dinkelbach's algorithm. *Manage. Sci.*, 22:868–873, 1976.

[62] P. Trautmann and D. Walter. A fast primal-dual-active-jump method for minimization in $BV((0,T);\mathbb{R}^d)$. *Optimization*, 0(0):1–45, 2023.

[63] F. Tröltzsch. *Optimal control of partial differential equations*, volume 112 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010.

[64] M. Ulbrich. *Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces*, volume 11 of *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011.

[65] M. Unser. A unifying representer theorem for inverse problems and machine learning. *Found. Comput. Math.*, 21(4):941–960, 2021.