SAND2025-14941J

DOMAIN DECOMPOSITION FOR INTEGER OPTIMAL CONTROL WITH TOTAL VARIATION REGULARIZATION*

ROBERT BARALDI† AND PAUL MANNS‡

Abstract. Total variation integer optimal control problems admit solutions and necessary optimality conditions via geometric variational analysis. In spite of the existence of said solutions, algorithms which solve the discretized objective suffer from high numerical cost associated with the combinatorial nature of integer programming. Hence, such methods are often limited to small- and medium-sized problems.

We propose a globally convergent, coordinate descent-inspired algorithm that allows tractable subproblem solutions restricted to a partition of the domain. Our decomposition method solves relatively small trust-region subproblems that modify the control variable on a subdomain only. Given nontrivial subdomain overlap, we prove that a global first-order necessary optimality condition is equivalent to a first-order necessary optimality condition per subdomain. We additionally show that sufficient decrease is achieved on a single subdomain by way of a trust-region subproblem solver using geometric measure theoretic arguments, which we integrate with a greedy patch selection to prove convergence of our algorithm. We demonstrate the practicality of our algorithm on a benchmark large-scale, PDE-constrained integer optimal control problem, and find that our method is faster than the state-of-the-art.

Key words. integer optimal control, total variation regularization, trust-region methods, domain decompositon.

AMS subject classifications. 49K30, 49Q15, 49Q20, 49M37

1. Introduction.

Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain, $d \in \mathbb{N}$, $\alpha > 0$, and $M \in \mathbb{N}$. We consider integer optimal control problems [14] of the form

$$\text{(P)} \qquad \begin{aligned} \min_{w \in L^1(\Omega)} \quad & J(w) := F(w) + \alpha \operatorname{TV}(w) \\ \text{s.t.} \quad & w(x) \in W := \{w_1, \dots, w_M\} \subset \mathbb{Z} \text{ for almost every (a.e.) } x \in \Omega, \end{aligned}$$

where $F : L^1(\Omega) \to \mathbb{R}$ is lower semicontinuous with respect to convergence in $L^p(\Omega)$ for some $p \geq 1$ and bounded below and $\operatorname{TV} : L^1(\Omega) \to [0, \infty]$ denotes the total variation seminorm. Typically, $F = j \circ S$ for j some tracking-type functional and S a solution operator of a PDE, ODE, or another integral operator. Such situations arise in control of switched systems [11, 19, 37], topology optimization [8, 9, 15, 29, 40], or optimal experimental design [36, 43].

Specifically, we consider the setting present in [32], where the authors proposed to solve a sequence of subproblems

$$\text{(TR)} \qquad \operatorname{TR}(\bar{w}, g, \Delta) := \begin{cases} \displaystyle\min_{w \in L^2(\Omega)} \; (g, w - \bar{w})_{L^2(\Omega)} + \alpha \operatorname{TV}(w) - \alpha \operatorname{TV}(\bar{w}) \\ \quad \text{s.t.} \quad \|w - \bar{w}\|_{L^1(\Omega)} \leq \Delta \text{ and } w(x) \in W \text{ for a.e. } x \in \Omega, \end{cases}$$

within a trust-region algorithm for globalization. Therein, the function g is (an approximation of) the Riesz representative $\nabla F(\bar{w}) \in L^\infty(\Omega)$ if F is Fréchet differentiable wrt. the L^1-norm so that the trust-region subproblem (TR) arises from (P) by solving a partially linearized model in an $L^1(\Omega)$-ball around a given point $\bar{w}$, that is, the current iterate of the algorithm. While we will assume said Fréchet differentiability of F to derive our results, we note that the presented algorithm and its iterations may still be well defined and meaningful if, e.g., subgradients are used if F is not differentiable. After discretization of the domain Ω and the introduction of a piecewise constant ansatz for w, the trust-region subproblems become integer linear programs [28]; these are often computationally expensive to solve in practice (see, e.g., [28, ?]). This difficulty stems from the large number of variables present in the integer linear programs, causing long running times or even subproblem computational infeasibility when particularly fine discretizations are chosen. While dynamic programming-based algorithms [33, 39] allow for efficient subproblem solvers on one-dimensional domains, no such approach is known for multi-dimensional domains. Instead, recent results [?] indicate that the problems are likely NP-hard, thereby requiring integer programming solver-based methods. These may moderately improve run times, but such structure-exploitation techniques are unlikely to decrease wallclock time by an appreciable amount. Moreover, when

*Submitted to the editors Oct. 21, 2024.

Funding: R. Baraldi acknowledges funding by the U.S. Air Force Office of Scientific Research no. FA9550-22-1-0248 and Sandia Laboratory Directed Research and Development John von Neumann Fellowship. P. Manns acknowledges funding by Deutsche Forschungsgemeinschaft (DFG) under project no. 515118017.

†Sandia National Laboratories, Optimization and Uncertainty Quantification. P.O. Box 5800, Albuquerque, New Mexico, 87125, USA, rjbaral@sandia.gov

‡Faculty of Mathematics, TU Dortmund University, 44227 Dortmund, Germany, paul.manns@tu-dortmund.de

considering a piecewise control function ansatz, as in [32, ?], the *exact* total variation is the sum of interface lengths between the level sets weighted by the jump heights across the sets; see, for example, [16]. Such settings are inevitable if the discretized controls are discrete-valued. Using uniform meshes and driving the mesh size to zero yields a gap between the discrete problem and the intended infinite-dimensional limit problem, because the restricted geometry of the discretization is reflected in the infinite-dimensional limit [10]. The difficulty is compounded in finite difference and finite element-based approximations, where the underlying meshes for the control and finite-element ansatz coincide [5], cannot be applied directly because interpolation and projection operators can violate integer feasibility. The authors of [?] overcome this problem via a two-level discretization combined with a cutting plane generation strategy, the latter of which successively enriches the integer subproblems, e.g. (TR), with linear inequalities. These allow isotropic approximations of the total variation seminorm despite the aforementioned geometric restriction. However, this process generates even larger problem formulations that are even less tractable; see the compute times reported in §6.4 of [38], in particular Table 6.11, where these problem formulations are used within the algorithm from [32]. Consequently, reducing the size of subproblems is a sensible starting point to scale the algorithm from [28, 32] to practical problem sizes, even given the potential cost incurred by computing more subproblem solutions.

Many large-scale problems, particularly in PDE numerics, are solved using domain decomposition approaches; see the monographs [6, 12, 21, 34]. In finite-dimensional optimization, coordinate-descent algorithms [41, 42] can be interpreted as decomposition-based methods and are particularly popular for settings with separable or block-separable objective functions. Coordinate-descent algorithms update only on a subset of the coordinates at a time, either cyclically or via random selection; in extreme cases, this may be only one coordinate. Hence, cost per iteration is reduced to an acceptable level although the number of iterations may be very large. Such methods have been explored for convex optimization with TV terms, primarily in the context of image denoising with an L^2-fidelity term. Common splitting methods, such as the Chambolle–Pock algorithm [4], perform well for small- and medium-scale problems and typically do not employ any domain decomposition. For larger TV-regularized problems, domain decomposition, or similarly utilized coordinate-descent methods, have been developed for both primal and (pre-)dual formulations; see the overview articles [27, 22]. While the setting therein concerns real-valued inputs and is in particular convex, these domain decomposition techniques also make concessions in of terms convergence guarantees or quality of the result. Specifically, naive coordinate-descent methods generally require a separability condition on the convex, nonsmooth term [41] to converge to global minimizers. This is, however, violated by the TV-term. Alternatively, one can consider the Euler–Lagrange equation for perturbed TV [7, 22], which, however, fails to preserve discontinuities and edges [22]. For predual TV formulations [17], convergence guarantees to a (global) minimizer have been proven for semismooth-Newton-type or accelerated splitting methods with overlapping and non-overlapping domain decomposition methods [22, 23]; additionally, patch subproblems can be solved in parallel [18, 24, 25, 26]. Discretizations considered include finite-differences [18] and finite elements [25]. The former can be solved with a variety of algorithms, e.g., Chambolle–Pock [4] or semismooth Newton [18], yielding convergence to a solution of the dual problem. The finite-element splitting approach [25] uses accelerated iterative soft-thresholding to achieve convergence to a solution of the dual problem for non-overlapping domains, allowing for direct parallelization. For primal decomposition methods to achieve such a result, a communication mechanism is required between or after solving the decomposed problems; see [23].

Our setting is inherently non-convex, hence we do not have strong duality and therefore follow a primal decomposition approach while striving for convergence to stationary points as in the case sans decomposition [28, ?].

1.1. Contribution. We transfer the idea of coordinate descent to (P) and the trust-region algorithm proposed in [32]. We decompose the domain Ω into smaller patches and solve the instances of (TR) on these patches. By prescribing a covering property and thus nontrivial overlap for the patches, we can localize the first-order necessary optimality condition on (P) from [32] to a first-order necessary optimality on all patch problems. This in turn permits construction of competitor sequences that allow us to prove sufficient decrease properties and determine stopping criteria for patch problem solution tabulation; the latter of which is based on the predicted reductions and trust-region radii. In turn, we are able to show convergence for a superordinate trust-region algorithm that tabulates solutions to patch subproblems and makes a greedy update of the iterate. Hence, there is no theoretical gap when compared to the asymptotics of the trust-region algorithm shown in [32]. We additionally note that this does not contradict the missing convergence guarantees of primal decomposition methods for the convex problems mentioned in the introduction; our optimality condition is localized on the interfaces of the level sets of w, and thus weaker than stationarity in a real-valued and differentiable setting.

We have executed the algorithm for a test problem on a one-dimensional domain and a test problem on a two-dimensional domain. On the one-dimensional domain, where we can use an efficient subproblem solver from [39], the SLIP algorithm scales very well and the subproblem solves are faster than the solution to state and adjoint equation. Consequently, the coordinate-based approach does not give a performance benefit in this case (except for extreme situations). This is in contrast to the two-dimensional problem, where a sufficiently large number of patches does not impair the quality of the resulting objective function values but gives high speedups for expensive instances. For the most expensive instance in our benchmark, our new algorithm returns a point of very similar objective value with a speedup of approximately 125 times the approach described [32]. Consequently, the proposed algorithm is an important step towards solving large-scale problems, as it can be computationally much cheaper without impairing the quality of the computed points.

While the proposed greedy update is clearly expensive, our current analysis deems it necessary to avoid situations like the one demonstrated in the famous example in [35], in which a coordinate descent-algorithm circumscribes the local optimum by traversing the surrounding level-sets. In such cases, coordinate descent fails to converge in $\mathbb{R}^n$ when using a fixed coordinate-selection scheme. Since, to our knowledge no other algorithmic approaches exist thus far and integer programming problems become otherwise quickly completely intractable, we believe that a greedy approach is justified. Additionally, we integrate a heuristic acceleration step into the algorithm that combines block updates sequentially (largest predicted decrease to smallest) until an *a posteriori* decrease condition fails. We also believe that the algorithm has a high potential for further improvements in terms of scalability, in particular by means of randomized patch selection. Such randomization is usually key to obtain good convergence properties for coordinate-descent algorithms without tabulation and greedy selection [42].

1.2. Structure of the remainder. We first introduce important notation in Section 2, and then the domain decomposition and localized/patch first-order necessary optimality (stationarity) in Section 3. Section 4 introduces the trust-region algorithm that includes the aforementioned tabulation and greedy update selection. In Section 5, we prove that instationary points lead to finite tabulation and acceptable patch updates and in turn convergence of the trust-region algorithm by means of the aforementioned competitor constructions. Section 6 provides preliminary computational results.

2. Notation. Let $[n] := \{1,\dots,n\}$ for $n \in \mathbb{N}$. The symmetric difference of sets $A, B \subset \mathbb{R}^d$ is $A \triangle B$. We denote the Lebesgue measure λ, and the Lebesgue space $L^p(\Omega)$ on Ω as $\|\cdot\|_{L^p(\Omega)}$ by $\|\cdot\|_{L^p}$ with inner product $(\cdot,\cdot)_{L^2}$. The restriction of a measure μ to set A is $\mu \llcorner A$. For a set A, the function χ_A is the $\{0,1\}$-valued characteristic function of A. We call a partition of a set into sets of finite perimeter a *Caccioppoli partition*. For a set $A \subset \Omega$, we denote its points of density 1 and 0 with respect to the Lebesgue measure by $A^{(1)}$ and $A^{(0)}$, see also [30, §5.3]. For measureable $E \subset \Omega$, the perimeter is defined as in [30, 32]:

$$P(E,\Omega) := \sup\left\{ \int_E \operatorname{div}\varphi(x)\,\mathrm{d}x \,\middle|\, \varphi \in C_c^1(\Omega,\mathbb{R}^d),\ \sup_{x\in\Omega}\|\varphi(x)\| \le 1 \right\}.$$

If $P(E,\Omega) < \infty$, it is a *Caccioppoli set*. A partition $\{E_i\}_{i\in I}$ of Ω is a Caccioppoli partition if $\sum_{i\in I} P(E_i,\Omega) < \infty$. The *topological boundary* is defined as ∂E and the *reduced boundary* is $\partial^* E$. The *essential boundary* $\partial^e A$ is the set of points with density of neither 1 nor 0 with respect to A. Note that unless noted otherwise, we consider $\partial^* E$, $\partial^e E$, $E^{(0)}$, $E^{(1)}$, ∂E with respect to Ω so that $P(E,\Omega) = \mathcal{H}^{d-1}(\partial^* E)$, where $\mathcal{H}^{d-1}$ denotes the $d-1$-dimensional Hausdorff measure. For a given Caccioppoli set E, we denote its unit outer normal vector on the reduced boundary by n_E. Moreover, if, in addition, a vector field $\phi \in C_c^\infty(\Omega,\mathbb{R}^d)$ is given, we recall that its boundary divergence $\operatorname{div}_{E_i}\phi : \partial^* E \to \mathbb{R}$ is defined by $\operatorname{div}_{E_i} := \operatorname{div}\phi - n_E \cdot \nabla\phi n_E$ [30, §17.3].

A function $u \in L^1(\Omega)$ is of bounded variation ($u \in \mathrm{BV}(\Omega)$) if its distributional derivative Du is a finite Radon measure over Ω, i.e., $\mathrm{TV}(u) := |Du|(\Omega) < \infty$ where $|u|$ is the total variation of measure μ. For a Borel set $E \subset \Omega$, $\mathrm{TV}(\chi_E) = P(E,\Omega)$. The fact that a sequence of functions $\{w^n\}_n \subset \mathrm{BV}(\Omega)$ converges *weakly-** to $w \in \mathrm{BV}(\Omega)$ is denoted by $w^n \overset{*}{\rightharpoonup} w$ when $w^n \to w$ in $L^1(\Omega)$ and $\limsup_{n\to\infty} \mathrm{TV}(w^n) < \infty$. A sequence $\{w^n\}_n$ converges *strictly* to $w \in \mathrm{BV}(w)$ if $w^n \to w$ in $L^1(\Omega)$ and $\mathrm{TV}(w^n) \to \mathrm{TV}(w) < \infty$.

Feasible points of (P) are functions in $\mathrm{BV}(\Omega)$ that attain values only in the finite set W; additionally, their distributional derivatives are absolutely continuous with respect to $\mathcal{H}^{d-1}$. The feasible set of (P) is $\mathrm{BV_W}(\Omega)$ defined by

$$\mathrm{BV_W}(\Omega) \coloneqq \{w \in \mathrm{BV}(\Omega) \,|\, w(x) \in W \text{ for a.e. } x \in \Omega\} \subset \mathrm{BV}(\Omega),$$

which is weak-* sequentially closed in $\mathrm{BV}(\Omega)$ [32]. Note that due to the discreteness restriction, $\mathrm{BV_W}(\Omega)$ is a bounded subset of $L^\infty(\Omega)$. This also implies that the assumed lower semicontinuity of F for some $p \geq 1$ gives lower semicontinuity of F for all $p \in [1, \infty)$ when restricting to $\mathrm{BV_W}(\Omega)$. In order to avoid cumbersome notation, we define for open sets $B \subset \Omega$ the restricted total variation to B as

$$\mathrm{TV_B}(w) \coloneqq \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} |w_i - w_j| \mathcal{H}^{d-1} \llcorner B(\partial^* E_i \cap \partial^* E_j)$$

for $w \in \mathrm{BV_W}(\Omega)$ with corresponding Caccioppoli partition $\{E_1, \ldots, E_M\}$ of Ω such that $w = \sum_{i=1}^M w_i \chi_{E_i}$. Clearly, $\mathrm{TV}(w) = \mathrm{TV}_\Omega(w)$.

Definition 2.1 (Definition 3.1 in [32]).

1. *A one parameter family of diffeomorphisms is the* $f \in C^\infty : (-\epsilon, \epsilon) \times \Omega \to \Omega$ *for some* $\epsilon > 0$ *such that for all* $t \in (-\epsilon, \epsilon)$*, the function* $f_t(\cdot) \coloneqq f(t, \cdot) : \Omega \to \Omega$ *is a diffeomorphism.*
2. *For open* $A \subset \Omega$*, the family* $(f_t)_{t \in (-\epsilon, \epsilon)}$ *is a local variation in* A *if, in addition to 1.,* $f_0(x) = x$ *for all* $x \in \Omega$ *and there is a compact set* $K \subset A$ *such that* $\{x \in \mathbb{R}^d | f_t(x) \neq x\} \subset K$ *for all* $t \in (-\epsilon, \epsilon)$.
3. *For a local variation, its initial velocity is defined as* $\phi(x) \coloneqq \frac{\partial f}{\partial t}(0, x)$ *for* $x \in \Omega$.

3. Relation of Stationarity for (P) to Patch Problems. We propose to solve subproblems on patches that only update the iterate in some part of the domain $D \subset \Omega$. We briefly recall stationarity for (P) from [32] and subsequently define a restricted *patch-stationarity* concept to patches D that decompose the domain Ω. Then we prove equivalence of stationarity and patch-stationarity. This section references several results from [32], which is written under the general assumption $d \geq 2$. Inspecting the arguments of [32], we observe that $d \geq 2$ is not required for the proofs of the referenced results (it is required in [32] for proving Theorem 5.2 therein and results building on it) so that the referenced results can all safely be used here in our unified setting $d \geq 1$.

Stationarity for (P). We provide the stationarity concept from Definition 4.4 in [32], which leans on stationarity for the prescribed mean curvature problem below.

Definition 3.1 (Stationarity, Definition 4.4 in [32]). *Let* $F : L^1(\Omega) \to \mathbb{R}$ *be continuously Fréchet differentiable. Let* $w \in \mathrm{BV_W}(\Omega)$*, that is* $w = \sum_{i=1}^M w_i \chi_{E_i}$ *for some Caccioppoli partition* $\{E_1, \ldots, E_M\}$ *of* Ω*. Let* $\nabla F(w) \in C(\bar{\Omega})$*. Then,* w *is* stationary *if*

(3.1)
$$\sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \int_{\partial^* E_i \cap \partial^* E_j} (w_j - w_i) \nabla F(w)(x) \phi(x) \cdot n_{E_i}(x) - \alpha |w_i - w_j| \operatorname{div}_{E_i} \phi(x) \, \mathrm{d}\mathcal{H}^{d-1}(x) = 0$$

holds for all $\phi \in C_c^\infty(\Omega, \mathbb{R}^d)$.

Stationarity as defined above is a first-order necessary optimality condition for (P) that arises from a first-order variation of the objective functional at a locally optimal point with respect to perturbations of a feasible point's level sets; see Proposition 3.2 below, which is shown in Theorem 4.6 in [32]. It is an extension of the variational first-order optimality for the *prescribed mean curvature problem* from geometric measure theory; see, e.g., [30, Equation 12.32 and Remark 17.11]. Moreover, [28, 32] prove the limit points of a trust-region algorithm in $\mathrm{BV_W}(\Omega)$ are stationary under regularity assumptions, the latter of which may be interpreted as *constraint qualifications* in analogy to classical nonlinear programming theory. Due to the non-convexity, (3.1) is a necessary and not a sufficient optimality condition.

The variational formulation (3.1) can be interpreted as follows: On $\partial^* E_i \cap \partial^* E_j$, the partition E_i has distributional mean curvature $-\nabla F(w) \frac{w_i - w_j}{\alpha |w_i - w_j|}$. We refer to [31, 32] for more details and present the following proposition:

Proposition 3.2 (Theorem 4.6 in [32], Proposition 2.4 in [31]). *Let* $F : L^1(\Omega) \to \mathbb{R}$ *and* $\bar{w} \in \mathrm{BV_W}(\Omega)$ *satisfy the assumptions of Definition* 3.1*. If there is* $r > 0$ *such that*

$$F(\bar{w}) + \alpha \, \mathrm{TV}(\bar{w}) \leq F(w) + \alpha \, \mathrm{TV}(w)$$

holds for all $w \in \mathrm{BV_W}(\Omega)$ *such that* $\|w - \bar{w}\|_{L^1} \leq r$*, then* $\bar{w}$ *is stationary.*

From the stationarity equality (3.1) and its derivation by means of local variations, it is clear that the above-defined stationarity concept is based on purely local information that is concentrated on the interfaces between the level sets of the feasible point $w \in \mathrm{BV_W}(\Omega)$ under consideration, see the analysis of stationarity and corresponding remarks in [28, 31, 32]. In particular, the variational *for all* character of the condition in Definition 3.1 implies that constant functions, which do not have level set boundaries inside Ω, are always stationary. We stress that this does not mean that the algorithm proposed in [32] and the algorithm analyzed in this work necessarily stop at constant functions. On the contrary, we initialize our algorithm with a constant function in our experiments in Section 6 and can observe that it produces a very different point. However, by choosing a very large value of $\alpha > 0$, it is possible to construct results where the algorithm cannot leave the stationary initial point; see also Example 3 in [22].

Patch-stationarity for (P)**.** We now assume a family of patches $\mathcal{D}$ that cover our computational domain Ω. We want to relate the concept of stationarity from the patches to the whole domain and vice versa. Therefore, we introduce and analyze a patch-based stationarity concept below.

ASSUMPTION 3.3. *Let $\mathcal{D} \subset 2^\Omega$ be a finite, open cover of Ω.*

DEFINITION 3.4 (Patch-stationarity with respect to $\mathcal{D}$). *Let $\mathcal{D} \subset 2^\Omega$, $F : L^1(\Omega) \to \mathbb{R}$, and $\bar{w} \in \mathrm{BV_W}(\Omega)$ satisfy the assumptions of Definition* 3.1 *and Assumption* 3.3*. Then, $\bar{w}$ is* patch-stationary with respect to $\mathcal{D}$ *if for all $D \in \mathcal{D}$ the identity* (3.1) *holds for all $\phi \in C_c^\infty(D, \mathbb{R}^d)$.*

We are ready to prove our main result for the localization of stationarity to the patch problems, namely that stationarity and local stationarity with respect to $\mathcal{D}$ are equivalent.

THEOREM 3.5. *Let $\mathcal{D} \subset 2^\Omega$, $F : L^1(\Omega) \to \mathbb{R}$, and $\bar{w} \in \mathrm{BV_W}(\Omega)$ satisfy the assumptions of Definition* 3.4*. Then $\bar{w} \in \mathrm{BV_W}(\Omega)$ is stationary if and only if it is patch-stationary with respect to $\mathcal{D}$.*

While the proof of the forward implication is straightforward, the reverse implication requires some preparatory work. Namely, for an arbitrary but fixed Caccioppoli partition of Ω and $\phi \in C_c^\infty(\Omega, \mathbb{R}^d)$, we assert the existence of a countable set of disjoint, closed balls $\{\overline{B_k} \,|\, k \in \mathbb{N}\}$ so that each $\overline{B_k}$ is contained in at least one of the patches. Moreover, the $\overline{B_k}$ exhaust $\operatorname{supp}\phi$ except for a set of $\mathcal{H}^{d-1}$-measure zero and their boundaries intersect the interfaces of the Caccioppoli partition only in a set of $\mathcal{H}^{d-1}$-measure zero. This auxiliary result is proven below in Lemma 3.6 as a direct consequence of the Vitali–Besicovitch covering theorem.

LEMMA 3.6. *Let $\mathcal{D}$ satisfy Assumption* 3.3*, $\{E_1, \dots, E_M\}$ be a Caccioppoli partition of Ω, and $\phi \in C_c^\infty(\Omega, \mathbb{R}^d)$. Then there is a countable set $\tilde{\mathcal{F}} = \{\overline{B_k} \,|\, k \in \mathbb{N}\}$ of pairwise disjoint and closed balls such that*

$$\overline{B_k} \subset\subset D \text{ for some } D \in \mathcal{D} \text{ for all } k \in \mathbb{N}, \tag{3.2}$$

$$\mathcal{H}^{d-1} \llcorner \left(\bigcup_{i=1}^M \partial^* E_i \right) \left(\operatorname{supp}\phi \setminus \bigcup_{k=1}^\infty \overline{B_k} \right) = 0, \text{ and} \tag{3.3}$$

$$\mathcal{H}^{d-1} \llcorner \left(\bigcup_{i=1}^M \partial^* E_i \right) (\partial \overline{B_k}) = 0. \tag{3.4}$$

Proof. The claim follows from the Vitali–Besicovitch covering theorem, see Theorem 2.19 in [1], if the (uncountable) set of closed balls

$$\mathcal{F} = \left\{ \overline{B_s(x)} \,\middle|\, x \in \operatorname{supp}\phi, 0 < s, \overline{B_s(x)} \subset D, D \in \mathcal{D}, \text{ and } \mathcal{H}^{d-1}\left(\partial \overline{B_s(x)} \cap \bigcup_{i=1}^M \partial^* E_i \right) = 0 \right\} \tag{3.5}$$

is a fine cover of $\operatorname{supp}\phi$. Specifically, we need to show (A) that for all $x \in \operatorname{supp}\phi$ there is a ball $B_{r_x}(x)$, $r_x > 0$, which is (compactly) contained in at least one patch $D \in \mathcal{D}$ and (B) that each of these balls contains infinitely many balls $\overline{B_{s_k}(x)}$ of radii $0 < s_k < r_x$ with $s_k \searrow 0$ that satisfy $\mathcal{H}^{d-1}\left(\partial \overline{B_{s_k}(x)} \cap \bigcup_{i=1}^M \partial^* E_i \right) = 0$.

Claim (A) follows directly from Assumption 3.3. Claim (B) follows from the fact that for balls $B_s(x)$ we have $\mathcal{H}^{d-1}\big(\partial \overline{B_s(x)} \cap \bigcup_{i=1}^M \partial^* E_i\big) = 0$ for λ-a.e. $s \in (0, r_x)$, which we briefly argue by way of contradiction. Assume that this assertion is false, then there exists a subset $A \subset (0, r_x)$ with positive Lebesgue measure, where $\mathcal{H}^{d-1}(\partial \overline{B_s} \cap \bigcup_{i=1}^M \partial^* E_i) > \varepsilon_0$ for some $\varepsilon_0 > 0$ and a.e. $s \in A$. This implies $\mathcal{H}^{d-1}(\overline{B_{\bar{s}}} \cap \bigcup_{i=1}^M \partial^* E_i) = \infty$ for some $\bar{s} \le r_x$ by virtue of the Fubini–Tonelli theorem, which contradicts that the E_i are sets of finite perimeter in Ω. ∎

Next, we briefly argue that when (3.1) is violated for some $\phi \in C_c^\infty(\Omega, \mathbb{R}^d)$, then there exists a closed ball in the countable set $\tilde{\mathcal{F}}$ from Lemma 3.6 so that the restriction of the integrand in (3.1) to this closed ball also implies a violation.

Lemma 3.7. *Let $F : L^1(\Omega) \to \mathbb{R}$ and $\bar{w} \in \mathrm{BV_W}(\Omega)$ satisfy the assumptions of Definition* 3.4*. Let $\mathcal{D} \subset 2^\Omega$ satisfy Assumption* 3.3*. Let $\{E_1, \ldots, E_M\}$ be the Caccioppoli partition of Ω associated with $\bar{w}$. Let* (3.1) *be violated, that is*

(3.6)
$$\sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \int_{\partial^* E_i \cap \partial^* E_j} \underbrace{(w_j - w_i)\nabla F(w)(x)\phi(x) \cdot n_{E_i}(x) - \alpha |w_i - w_j| \operatorname{div}_{E_i} \phi(x)}_{=:\psi_{ij}(x)} \,\mathrm{d}\mathcal{H}^{d-1}(x) > \eta$$

holds for some $\phi \in C_c^\infty(\Omega, \mathbb{R}^d)$ and $\eta > 0$. Let $\tilde{\mathcal{F}}$ be as in Lemma 3.6*.*

Then, there exist $\overline{B_k} \in \tilde{\mathcal{F}}$ with $\overline{B_k} \subset D$ for some $D \in \mathcal{D}$ and a positive scalar $\eta_D > 0$ such that

$$\sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \int_{\partial^* E_i \cap \partial^* E_j} \psi_{ij}(x)\chi_{\overline{B_k}}(x) \,\mathrm{d}\mathcal{H}^{d-1}(x) > \eta_D. \tag{3.7}$$

Proof. We first note that if the equality (3.1) is violated, then there is $\eta > 0$ such that the absolute value of the left hand side of (3.6) is greater than η. Thus by replacing ϕ with $-\phi$ if necessary, (3.6) holds. By virtue of the properties asserted in Lemma 3.6 and the countable additivity of the measure $\mathcal{H}^{d-1} \llcorner \left(\bigcup_{i=1}^M \partial^* E_i\right)$, we obtain

$$\sum_{k\in\mathbb{N}} \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \int_{\partial^* E_i \cap \partial^* E_j} \psi_{ij}(x)\chi_{\overline{B_k}}(x) \,\mathrm{d}\mathcal{H}^{d-1}(x) > \eta > 0$$

from the instationarity inequality (3.6). Consequently, there has to exist a strictly positive summand on the left hand side. We use (one of the) patches D asserted by (3.5). ∎

We now continue with the proof of Theorem 3.5.

Proof of Theorem 3.5. If $\bar{w} \in \mathrm{BV_W}(\Omega)$ is stationary and $D \in \mathcal{D}$, then (3.1) holds for all $\phi \in C_c^\infty(D, \mathbb{R}^d)$ because Assumption 3.3 ensures that D is an open subset of Ω.

Thus it remains to show that if $\bar{w} \in \mathrm{BV_W}(\Omega)$ is patch-stationary with respect to $\mathcal{D}$, then it is also stationary. We prove the claim by means of a contrapositive argument and assume that $\bar{w}$ is not stationary. Let $\{E_1, \ldots, E_M\}$ be the Caccioppoli partition of Ω associated with $\bar{w}$. Because $\bar{w}$ is not stationary, there are $\phi \in C_c^\infty(\Omega, \mathbb{R}^d)$, $0 \neq \phi$, and $\eta > 0$ such that the inequality (3.6) holds. Let $\tilde{\mathcal{F}}$ be as in Lemma 3.6. By virtue of Lemma 3.7 there exist $\overline{B_k} \in \tilde{\mathcal{F}}$ and $D \in \mathcal{D}$ such that $\overline{B_k} \subset D$ for some $D \in \mathcal{D}$ and a positive scalar $\eta_D > 0$ such that (3.7) holds. We close the proof by constructing a $C_c^\infty(D, \mathbb{R}^d)$-function that violates (3.1).

To this end, we use a mollification of the restriction of ϕ onto $\overline{B_k}$, that is we replace $\phi\chi_{\overline{B_k}}$ by a smooth function in $C_c^\infty(D, \mathbb{R}^d)$ in the integrand of (3.7). To this end, let $\phi_\delta := \eta_\delta * (\chi_{\overline{B_k}}\phi)$ for a family of positive standard mollifiers $(\eta_\delta)_{\delta>0}$. Then $\phi_\delta \in C_c^\infty(\Omega, \mathbb{R}^d)$. We show that there exists $\delta_0 > 0$ such that for all $\delta \in (0, \delta_0)$ it holds that $\operatorname{supp} \phi_\delta \subset D$ and

$$\lim_{\delta \searrow 0} \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \int_{\partial^* E_i \cap \partial^* E_j} \psi_{ij}^\delta(x) \,\mathrm{d}\mathcal{H}^{d-1}(x) > \eta_D, \tag{3.8}$$

where $\psi_{ij}^\delta := w_i(-\nabla F(w))(\phi_\delta \cdot n_{E_i}) - \alpha|w_i - w_j| \operatorname{div}_{E_i} \phi_\delta$ for all $i, j \in \{1, \ldots, M\}$.

For all $y \in \operatorname{int} \overline{B_k}$, we obtain $\operatorname{div} \phi_\delta(y) \to \operatorname{div} \phi(y)$ and $\nabla\phi_\delta(y) \to \nabla\phi(y)$ as $\delta \searrow 0$. Moreover, $\overline{B_k} \in \tilde{\mathcal{F}}$ gives (3.4) that $\mathcal{H}^{d-1}\big(\partial\overline{B_k} \cap \bigcup_{i=1}^M \partial^* E_i\big) = 0$. In combination with the properties of the mollification, this implies

$$\operatorname{div} \phi_\delta(y) \to \operatorname{div} \phi(y) \text{ and } \nabla\phi_\delta(y) \to \nabla\phi(y) \quad \text{as } \delta \searrow 0 \quad \text{for } \mathcal{H}^{d-1}\text{-a.e. } y \in \bigcup_{i=1}^M \partial^* E_i \cap \overline{B_k}$$

and

$$\operatorname{div} \phi_\delta(y) \to 0 \text{ and } \nabla\phi_\delta(y) \to 0 \quad \text{as } \delta \searrow 0 \quad \text{for all } y \in \Omega \setminus \overline{B_k}.$$

Because D is open, it holds that $\operatorname{dist}(\overline{B_k}, \partial D) > 0$ and we obtain that there exists $\delta_0 > 0$ such that for all $\delta \in (0, \delta_0)$ the inclusion $\operatorname{supp} \phi_\delta \subset\subset D$ is satisfied. These considerations imply

$$\psi_{ij}^\delta(y) \to \psi_{ij}(y)\chi_{\overline{B_k}}(y) \quad \text{as } \delta \searrow 0 \quad \text{for } \mathscr{H}^{d-1}\text{-a.e. } y \in \overline{B_k} \cap \bigcup_{i=1}^{M} \partial^* E_i.$$

Because of the L^∞-bounds on ϕ and $\chi_{\overline{B_k}}$, we can apply Lebesgue's dominated convergence theorem and obtain

$$\sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \int_{\partial^* E_i \cap \partial^* E_j} \psi_{ij}^\delta(x) \,\mathrm{d}\mathscr{H}^{d-1}(x) \underset{\delta \searrow 0}{\to}$$
$$\sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \int_{\partial^* E_i \cap \partial^* E_j} \psi_{ij}(x)\chi_{\overline{B_k}}(x) \,\mathrm{d}\mathscr{H}^{d-1}(x) > \eta_D > 0,$$

implying that the left hand side is eventually strictly greater than zero so that ϕ_δ eventually violates (3.1) and satisfies $\operatorname{supp} \phi_\delta \subset D$. ∎

4. Trust-region Patch Algorithm. We now introduce Algorithm 4.1, which is a modification of Algorithm 5.1 in [32], where we optimize subproblems over each patch in the collection $\mathscr{D}$ of patches determined *a priori*. Before stating it, we require a regularity assumption on the main part of the objective F.

Assumption 4.1 (see also Assumption 2.2 in [31], Assumption 4.3 in [32]).
1. *Let $F : L^1(\Omega) \to \mathbb{R}$ be continuously Fréchet differentiable.*
2. *Let $\nabla F : L^1(\Omega) \to L^\infty(\Omega)$ be Lipschitz continuous on the feasible set, that is*

$$\infty > L_{\nabla F} := \sup \left\{ \frac{\|\nabla F(w_1) - \nabla F(w_2)\|_{L^\infty}}{\|w_1 - w_2\|_{L^1}} \,\middle|\, w_1(x), w_2(x) \in \operatorname{conv} W \text{ for a.e. } x \in \Omega \right\}.$$

We note that Assumption 4.1 1 already implies the boundedness

$$c(b) := \sup \{ \|\nabla F(w)\|_{L^\infty} \,|\, w(x) \in W \text{ a.e. and } \operatorname{TV}(w) < b \} < \infty. \tag{4.1}$$

for every $b > 0$ because we can identify $L^1(\Omega)^*$ with $L^\infty(\Omega)$ and the bound $\operatorname{TV}(w) < b$ implies that the supremum is taken over a relatively compact subset of $L^1(\Omega)$ due to the compact embedding $\operatorname{BV}(\Omega) \hookrightarrow L^1(\Omega)$ [1, Corollary 3.49].

Specifically, for a given patch D, an iterate $\bar{w}$, a gradient (approximation) g of $\nabla F(\bar{w})$, and a trust-region radius $\Delta > 0$, the subproblem reads:

$$\text{(TRP)} \qquad \operatorname{TRP}(\bar{w}, g, D, \Delta) := \begin{cases} \min\limits_{w \in L^2(\Omega)} & (g, w - \bar{w})_{L^2(\Omega)} + \alpha \operatorname{TV}(w) - \alpha \operatorname{TV}(\bar{w}) \\ \text{s.t.} & \|w - \bar{w}\|_{L^1(\Omega)} \le \Delta \text{ and } w(x) \in W \text{ for a.e. } x \in D, \\ & w(x) = \bar{w}(x) \text{ for a.e. } x \in \Omega \setminus D. \end{cases}$$

We briefly recall that $\operatorname{TRP}(\bar{w}, g, D, \Delta)$ is well-defined, that is, it admits a solution.

Proposition 4.2. *Let $D \in \mathscr{D}$, $\bar{w} \in \operatorname{BV_W}(\Omega)$, $g \in L^2(\Omega)$, and $\Delta > 0$. Then $\operatorname{TRP}(\bar{w}, g, D, \Delta)$ has a minimizer.*

Proof. A proof can be found in [28, Proposition 3.2]. ∎

The regularity condition Assumption 4.1 as well as Proposition 4.2 allow for Algorithm 4.1 to converge to a stationarity point defined in (3.1). The method itself consists of an outer loop indexed by n, and an inner/block-update loop indexed by k.

At the beginning of each outer loop, we initialize two sets: the set of acceptable step candidates, $\mathscr{A} = \emptyset$; and the working set of patches, $\mathscr{W} = (0, D)$ for $D \in \mathscr{D}$. The inner/block-update loop starting at Line 3 solves a subproblem (TRP) every iteration k over all patches in the set $\mathscr{D}$ for which the pair (k, D) is contained in $\mathscr{W}$. These subproblem solves yield trial iterates $\tilde{w}^{n,k,D}$, with corresponding predicted, $\operatorname{pred}^{n,k,D}$, and actual, $\operatorname{ared}^{n,k,D}$, reductions. Based on their values, $\mathscr{W}$ is updated by means of the following case distinction. If $\operatorname{pred}^{n,k,D}$ is zero, this means that w^n is stationary in the patch D and (k, D) is just removed from $\mathscr{W}$. If this is not the case and the block update $\tilde{w}^{n,k,D}$ moreover sufficiently decreases the cost function by some fraction of the predicted reduction greater than zero, it is added to $\mathscr{A}$ as an "acceptable step" and (k, D) is also removed from $\mathscr{W}$. If $\operatorname{pred}^{n,k,D}$ is greater than zero but the sufficient decrease is not satisfied, it is checked

Algorithm 4.1 Sequential linear integer programming method with greedy patch updates

Input: F satisfying Assumption 4.1 with smoothness constant $L_{\nabla F}$, $\Delta^0 > 0$, $w^0 \in \mathrm{BV_W}(\Omega)$, $\sigma \in (0,1)$, set of patches $\mathscr{D}$.

for $n = 0, 1, 2, \dots$ **do**
  Set $\mathscr{A} \leftarrow \emptyset$, $\mathscr{W} \leftarrow \{(0, D) \,|\, D \in \mathscr{D}\}$.
  for $k = 0, 1, \dots$ **do**
    while $(k, D) \in \mathscr{W}$ **do**
      $\tilde{w}^{n,k,D} \leftarrow$ minimizer of $\mathrm{TRP}(w^n, \nabla F(w^n), D, \Delta^0 2^{-k})$.
      $\mathrm{pred}^{n,k,D} \leftarrow (\nabla F(w^n), w^n - \tilde{w}^{n,k,D})_{L^2} + \alpha\,\mathrm{TV}(w^n) - \alpha\,\mathrm{TV}(\tilde{w}^{k,n,D})$
      $\mathrm{ared}^{n,k,D} \leftarrow F(w^n) + \alpha\,\mathrm{TV}(w^n) - F(\tilde{w}^{n,k,D}) - \alpha\,\mathrm{TV}(\tilde{w}^{n,k,D})$
      if $\mathrm{ared}^{n,k,D} \geq \sigma\,\mathrm{pred}^{n,k,D}$ **and** $\mathrm{pred}^{n,k,D} > 0$ **then**
        $\mathscr{A} \leftarrow \mathscr{A} \cup \{(k, D)\}$.
      else if $\mathrm{pred}^{n,k,D} > 0$ **and** $\max_{(\tilde{k},\tilde{D})\in\mathscr{A}} \mathrm{ared}^{n,\tilde{k},\tilde{D}} < \mathrm{pred}^{n,k,D} + L_{\nabla F}\Delta^0 2^{-k}$ **then**
        $\mathscr{W} \leftarrow \mathscr{W} \cup \{(k+1, D)\}$
      end if
      $\mathscr{W} \leftarrow \mathscr{W} \setminus \{(k, D)\}$.
    end while
    if $\mathscr{W} = \emptyset$ **then**
      break
    end if
  end for
  if $\mathscr{A} = \emptyset$ **then**
    return (w^n is stationary).
  end if
  $\bar{w}^n \leftarrow w^n$.
  $\bar{j}^0 \leftarrow F(w^n) + \alpha\,\mathrm{TV}(w^n)$.
  while $\mathscr{A} \neq \emptyset$ **do**
    $\bar{k}, \bar{D} \leftarrow \arg\max\{\mathrm{ared}^{n,k,D} \,|\, (k, D) \in \mathscr{A}\}$
    $\tilde{w}^n \leftarrow \bar{w}^n \chi_{\Omega\setminus\bar{D}} + \chi_{\bar{D}} \tilde{w}^{n,\bar{k},\bar{D}}$
    $\bar{j} \leftarrow F(\tilde{w}^n) + \alpha\,\mathrm{TV}(\tilde{w}^n)$
    if $\bar{j} < \bar{j}^0$ **then**
      $\bar{w}^n \leftarrow \tilde{w}^n$, $\bar{j}^0 \leftarrow \bar{j}$, $\mathscr{A} \leftarrow \mathscr{A} \setminus \{(\bar{k}, \bar{D})\}$.
    else
      break
    end if
  end while
  $w^{n+1} \leftarrow \bar{w}^n$
end for

if $\max_{(\tilde{k},\tilde{D})\in\mathscr{A}} \mathrm{ared}^{n,\tilde{k},\tilde{D}}$, the maximum of already found acceptable reductions, is dominated from above by $\mathrm{pred}^{n,k,D} + L_{\nabla F}\Delta^0$, the predicted reduction plus the Lipschitz constant of the first part of the objective times the trust-region radius. We highlight that we use the convention $\max \emptyset = -\infty$ here. If this condition holds, a further reduction of the trust-region radius may give an acceptable step that then maximizes the predicted reductions over all steps that have been deemed acceptable so far. Thus (k, D) is replaced by $(k+1, D)$ in $\mathscr{W}$ so that the same patch but with reduced trust-region radius is now contained in the working set $\mathscr{W}$. If this condition does not hold, this implies that even if there is an acceptable step for this patch with a smaller trust-region radius, the resulting actual reduction cannot give the maximizer of $\max_{(\tilde{k},\tilde{D})\in\mathscr{A}} \mathrm{ared}^{n,\tilde{k},\tilde{D}}$ and, consequently, (k, D) is removed from $\mathscr{W}$ without replacement. This condition is derived from Assumption 4.1 and allows us to prove finite termination of the inner/block-update loop if w^n is not stationary because it can happen w^n is stationary on D but has positive predicted reduction for all positive trust-region radii, which would lead to infinite reduction of the trust-region radius without this safeguarding. Once $\mathscr{W} = \emptyset$, that is, if there are no more working patches we terminate Algorithm 4.1. An example run of this inner/block-update loop for four patches is given in Table 4.1. We note that a straightforward parallel version of Algorithm 4.1 arises by optimizing in parallel over the elements of $\mathscr{W}$. However, our computational results in subsection 6.2 indicate that the runtimes between the elements of $\mathscr{W}$ may vary dramatically and we find it more advisable to use parallelization on the subproblem solver level.

$\mathcal{D}$ \ k	D_1	D_2	D_3	D_4	$\mathcal{A}$	$\mathcal{W}$
0	$\tilde{w}^{n,0,1}$	$\tilde{w}^{n,0,2}$	$\tilde{w}^{n,0,3}$	$\tilde{w}^{n,0,4}$	$\emptyset$	$\{(1,1),(1,2),(1,4)\}$
1	$\tilde{w}^{n,1,1}$	$\tilde{w}^{n,1,2}$	–	$\tilde{w}^{n,1,4}$	$\{(1,4)\}$	$\{(2,1),(2,2)\}$
2	$\tilde{w}^{n,2,1}$	$\tilde{w}^{n,2,2}$	–	–	$\{(1,4)\}$	$\{(3,1)\}$
3	$\tilde{w}^{n,3,1}$	–	–	–	$\{(1,4)\}$	$\{(4,1)\}$
4	$\tilde{w}^{n,4,1}$	–	–	–	$\{(4,1),(1,4)\}$	$\emptyset$

Table 4.1: Example inner-loop iteration for $|\mathcal{D}| = 4$ at line 17. Orange signifies that $\text{pred}^{n,k,D} = 0$ so that w^n is already stationary on D and therefore (k, D) can safely be removed from $\mathcal{W}$. Green signifies that signifies Line 10 holds: $\tilde{w}^{n,k,D}$ is not yet acceptable and also not dominated by some previous iterate that was found acceptable. Blue signifies Line 8 holds: $\tilde{w}^{n,k,D}$ is acceptable and (k, D) can be transferred from $\mathcal{W}$ to $\mathcal{A}$. Violet signifies that $\text{pred}^{n,k,D} > 0$ but neither Line 10 nor Line 8 holds: the possible actual reduction for all further trust-region radii is dominated by some element of $\mathcal{A}$ and (k, D) can safely be removed from $\mathcal{W}$ too.

In the outer loop, once $\mathcal{A}$ is the nullset, then by construction w^n is stationary and we terminate the algorithm. If $\mathcal{A}$ is nonempty, then we define trial variable $\bar{w}^n = w^n$ and compute trial function value $\bar{j}^0$. We then enter a "greedy" update loop; this loop inserts the patch updates to the overall solution and measure decrease of the actual function. In effect, we iterate over stored maximum block actual reductions $(\bar{k}, \bar{D}) \leftarrow \arg\max\{\text{ared}^{n,k,D} \,|(k, D) \in \mathcal{A}\}$, insert the resulting solution $\tilde{w}^n$ associated with $(\bar{k}, \bar{D})$ into $\bar{w}^n$ and compute a new trial function value. If this function value is less than $\bar{j}^0$, then we eliminate this (k, D) from the set of acceptable steps, keep the $\bar{w}^n$ with the associated $\tilde{w}^n$, and repeat until $\mathcal{A} = \emptyset$.

By construction, the first (greedy) patch update will always satisfy the acceptance criterion; further updates are heuristic improvements that are accepted until the cost function increases.

5. Convergence Analysis of Algorithm 4.1. We provide a convergence analysis for a greedy patch selection approach under two assumptions on the set of patches.

Assumption 5.1 (Sufficient overlap and regularity of patches). *Let Ω be a bounded Lipschitz domain. Let the finite set of patches $\mathcal{D} \subset 2^\Omega$ satisfy the following conditions.*

(a) For all $x \in \Omega$, there exist $r > 0$ and $D \in \mathcal{D}$ such that $\overline{B_r(x)} \subset\subset D$. (Patch overlap)

(b) For all $D \in \mathcal{D}$, we assume that $D \in \mathcal{D}$ is a bounded Lipschitz domain. (Patch regularity)

Because $\mathcal{D}$ is finite, Assumption 5.1 (a) is equivalent to Assumption 3.3 and in particular, the analysis in Section 3 may be applied. Consequently, Theorem 3.5 gives that the instationarity condition (3.6) implies a localization to at least one patch, on which patch-instationarity (3.8) holds. Assumption 5.1 (b) implies that the patches themselves are sets of finite perimeter and the assumptions for the analysis of the inner loop of Algorithm 1 from [32] can be applied to trust-region subproblems $\text{TRP}(w^{n-1}, \nabla F(w^{n-1}), D^n, \Delta^0 2^{-k})$ for some $\Delta^0 > 0$ as are generated by Algorithm 4.1. While the analysis is not impaired by Assumption 5.1 (b), it facilitates the intuition greatly when considering patches that are convex and have boundaries that are finite unions of $d-1$-dimensional convex polyhedra.

In order to analyze Algorithm 4.1 and the trust-region subproblems, we briefly recall what we mean by ared and pred for a given sequence $\{w^\ell\}_\ell \subset \text{BV}_\text{W}(\Omega)$ that converges strictly to some limit $w \in \text{BV}_\text{W}(\Omega)$. Let $k \in \mathbb{N}$, $D \in \mathcal{D}$, and $\tilde{w}^{\ell,k,D}$ be a solution to $\text{TRP}(w^\ell, \nabla F(w^\ell), D, \Delta^0 2^{-k})$, which exists by means of Proposition 4.2. Then we define

$$\begin{aligned} \text{pred}^{\ell,k,D} &:= (\nabla F(w^\ell), w^\ell - \tilde{w}^{\ell,k,D})_{L^2(\Omega)} + \alpha \operatorname{TV}(w^\ell) - \alpha \operatorname{TV}(\tilde{w}^{\ell,n,D}) \text{ and} \\ \text{ared}^{\ell,k,D} &:= F(w^\ell) + \alpha \operatorname{TV}(w^\ell) - F(\tilde{w}^{n,k,D}) - \alpha \operatorname{TV}(\tilde{w}^{\ell,k,D}). \end{aligned}$$

We now aim to show that if the limit of such a sequence is instationary, then there is a patch such that the sufficient decrease condition that is also in Line 8 is eventually satisfied. Our argument uses so-called competitors that modify a weakly-* converging sequence $w^\ell \overset{*}{\rightharpoonup} w$ in $\text{BV}_\text{W}(\Omega)$ so that the resulting sequence coincides with $\{w^\ell\}_\ell$ on D^c but has properties of a different function $\hat{w}$ inside a patch D and does not affect the boundaries of the level sets close to ∂D. Their existence is asserted below.

Lemma 5.2. *Let Assumption 5.1 hold. Let $D \in \mathcal{D}$ and $D_r := \{x \in D \,|\, \text{dist}(x, D^c) > r\}$ for $r > 0$. Let $w^\ell \overset{*}{\rightharpoonup} w$ in $\text{BV}_\text{W}(\Omega)$. Let $\hat{w} \in \text{BV}_\text{W}(\Omega)$ and $K \subset\subset D$ with $\operatorname{supp} w - \hat{w} \subset\subset K$.*

Then there exists $s > 0$ only depending on K such that w, w^ℓ for $\ell \in \mathbb{N}$, and the functions defined by

$$\hat{w}^\ell(x) := \begin{cases} w^\ell(x) & \text{if } x \notin D_s, \\ \hat{w}(x) & \text{else} \end{cases} \tag{5.1}$$

for $\ell \in \mathbb{N}$ satisfy $\hat{w}^\ell \overset{*}{\rightharpoonup} \hat{w}$ in $\mathrm{BV_W}(\Omega)$ as $\ell \to \infty$ and

$$\begin{aligned} \mathrm{TV}(w) &= \mathrm{TV}_{\Omega\setminus\overline{D_s}}(w) + \mathrm{TV}_{D_s}(w) \\ \mathrm{TV}(\hat{w}^\ell) &= \mathrm{TV}_{\Omega\setminus\overline{D_s}}(w^\ell) + \mathrm{TV}_{D_s}(\hat{w}) + b^\ell \end{aligned}$$

for some $\{b^\ell\}_\ell \subset [0,\infty)$ with $\liminf_{\ell\to\infty} b^\ell = 0$. Moreover, if $w^\ell \to w$ also strictly in $\mathrm{BV_W}(\Omega)$, we have

$$\begin{aligned} \mathrm{TV}_{\Omega\setminus\overline{D_s}}(w^\ell) &\to \mathrm{TV}_{\Omega\setminus\overline{D_s}}(w) \text{ and} \\ \mathrm{TV}_{D_s}(w^\ell) &\to \mathrm{TV}_{D_s}(w). \end{aligned}$$

Proof. Let $\{E_1,\dots,E_M\}$, $\{E_1^\ell,\dots,E_M^\ell\}$, $\{\hat{E}_1,\dots,\hat{E}_M\}$, and $\{\hat{E}_1^\ell,\dots,\hat{E}_M^\ell\}$ denote the Caccioppoli partitions of Ω such that the identities $w = \sum_{i=1}^M w_i\chi_{E_i}$, $w^\ell = \sum_{i=1}^M w_i\chi_{E_i^\ell}$, $\ell \in \mathbb{N}$, $\hat{w} = \sum_{i=1}^M w_i\chi_{\hat{E}_i}$, and $\hat{w}^\ell = \sum_{i=1}^M w_i\chi_{\hat{E}_i^\ell}$, $\ell \in \mathbb{N}$, hold [32, Lemma 2.1]. Then we observe that there is a small enough $r_0 > 0$ such that $K \subset\subset D_{r_0} \subset\subset D$.

Next, we argue similar to [30, Theorem 21.14] to identify $0 < s < r_0$ such that modifying w^ℓ on D_s to obtain the competitor $\hat{w}^\ell$ implies the claimed properties. Specifically, we need that i) ∂D_s does not intersect with the reduced boundaries $\partial^* E_i$ and $\partial^* E_i^\ell$ (for $\ell \in \mathbb{N}$) on a set of strictly positive $\mathcal{H}^{d-1}$-measure, and ii) the $\mathcal{H}^{d-1}$-measure of the intersection of ∂D_s and the symmetric difference of the points of density one of E_i and E_i^ℓ vanishes as $\ell \to \infty$. This is used to ensure that ∂D_s does not affect the boundary of the level sets of the competitor and thus the value of its total variation.

To this end, let $r_{\max}$ be large enough such that $D_{r_{\max}} = \emptyset$. We obtain for all $r > 0$ that

$$\|w - w^\ell\|_{L^1} \ge \sum_{i=1}^M \left|D_r \cap \left(E_i^\ell \mathbin{\triangle} E_i\right)\right| = \int_r^{r_{\max}} \sum_{i=1}^M \mathcal{H}^{d-1}\left(\partial D_s \cap \left((E_i^\ell)^{(1)} \mathbin{\triangle} E_i^{(1)}\right)\right) \mathrm{d}s,$$

where the second identity holds due to the coarea formula, which can be applied using the Lipschitz continuity of the function $x \mapsto \operatorname{dist}(x, D^c)$.

From Proposition 2.16 in [30], we obtain for all $i \in \{1,\dots,M\}$ and a.e. $s > 0$ that

$$\mathcal{H}^{d-1}(\partial D_s \cap \partial^* E_i) = 0 \text{ and} \tag{5.2}$$

$$\mathcal{H}^{d-1}(\partial D_s \cap \partial^* E_i^\ell) = 0 \text{ for all } \ell \in \mathbb{N}. \tag{5.3}$$

We apply Fatou's lemma and obtain

$$\liminf_{\ell\to\infty} \sum_{i=1}^M \mathcal{H}^{d-1}\left(\partial D_s \cap \left((E_i^\ell)^{(1)} \mathbin{\triangle} E_i^{(1)}\right)\right) = 0 \tag{5.4}$$

for a.e. $s > 0$.

Let $s \in (0, r_0)$ satisfy (5.2), (5.3) for all $\ell \in \mathbb{N}$, and (5.4). Then $\hat{w}^\ell \overset{*}{\rightharpoonup} \hat{w}$ in $\mathrm{BV_W}(\Omega)$ as $\ell \to \infty$ as claimed.

Next, we use Lemma A.2 from the literature to characterize the distributional derivative of $\chi_{\hat{E}_i^\ell}$. To verify its assumptions, we first recall that $\mathcal{H}^{d-1}(\partial D_s \cap \partial^* E_i^\ell) = 0$ holds for all $\ell \in \mathbb{N}$ by virtue of (5.3). Second, $\hat{w}(x) \ne w(x)$ can only hold for $x \in K \subset\subset D_s$ so that $\mathcal{H}^{d-1}(\partial D_s \cap \partial^* \hat{E}_i) = \mathcal{H}^{d-1}(\partial^* D_s \cap \partial^* E_i) = 0$ holds by virtue of (5.2). Finally, we observe that $\hat{E}_i^\ell = (\hat{E}_i \cap D_s) \cup (E_i^\ell \setminus D_s)$ holds by definition in (5.1). We thus apply Lemma A.2 and obtain

$$\begin{aligned} D\chi_{\hat{E}_i^\ell} &= D\chi_{E_i^\ell} \llcorner (\Omega \setminus \overline{D_s}) + D\chi_{\hat{E}_i} \llcorner D_s \\ &\quad + D\chi_{D_s} \llcorner \Big((\hat{E}_i)^{(1)} \cap (E_i^\ell)^{(0)}\Big) - D\chi_{D_s} \llcorner \Big((\hat{E}_i)^{(0)} \cap (E_i^\ell)^{(1)}\Big). \end{aligned}$$

We employ the σ-additivity of $\mathcal{H}^{d-1}$ and the identity $(E_i)^{(b)} \cap \partial D_s = (\hat{E}_i)^{(b)} \cap \partial D_s$ for $b \in \{0,1\}$ due to $K \subset\subset D_s$ to obtain

$$\begin{aligned}\mathcal{H}^{d-1}\llcorner\partial^*\hat{E}_i^\ell &= \mathcal{H}^{d-1}\llcorner\big(\partial^* E_i^\ell \cap (\Omega\setminus\overline{D_s})\big)\\ &+ \mathcal{H}^{d-1}\llcorner\Big(\partial^*\hat{E}_i \cap D_s\Big)\\ &+ \mathcal{H}^{d-1}\llcorner\Big(\partial D_s \cap (E_i)^{(1)} \cap (E_i^\ell)^{(0)}\Big)\\ &+ \mathcal{H}^{d-1}\llcorner\Big(\partial D_s \cap (E_i)^{(0)} \cap (E_i^\ell)^{(1)}\Big)\end{aligned}$$

for the corresponding variation measures. We insert $\partial^*\hat{E}_j^\ell$, $j \neq i$, on both sides, use $\hat{E}_i^\ell \cap D_s = \hat{E}_i \cap D_s$ and $\hat{E}_i^\ell \cap (\Omega\setminus\overline{D_s}) = E_i^\ell \cap (\Omega\setminus\overline{D_s})$ due to (5.1), and the disjointness of the sets in the last two terms to obtain

$$\begin{aligned}\mathcal{H}^{d-1}\big(\partial^*\hat{E}_i^\ell \cap \partial^*\hat{E}_j^\ell\big) &= \mathcal{H}^{d-1}\left(\partial^* E_i^\ell \cap \partial^* E_j^\ell \cap (\Omega\setminus\overline{D_s})\right)\\ &+ \mathcal{H}^{d-1}\left(\partial^*\hat{E}_i \cap \partial^*\hat{E}_j \cap D_s\right)\\ &+ \mathcal{H}^{d-1}\left(\partial^*\hat{E}_j^\ell \cap \partial D_s \cap \Big(\big((E_i)^{(1)} \cap (E_i^\ell)^{(0)}\big) \cup \big((E_i)^{(0)} \cap (E_i^\ell)^{(1)}\big)\Big)\right).\end{aligned}$$

Using Lemma A.1 on the symmetric difference of sets of points of density 1, we obtain

$$\begin{aligned}\mathcal{H}^{d-1}\big(\partial^*\hat{E}_i^\ell \cap \partial^*\hat{E}_j^\ell \cap \Omega\big) &= \mathcal{H}^{d-1}\big(\partial^* E_i^\ell \cap \partial^* E_j^\ell \cap (\Omega\setminus\overline{D_s})\big)\\ &+ \mathcal{H}^{d-1}\big(\partial^*\hat{E}_i \cap \partial^*\hat{E}_j \cap D_s\big)\\ &+ \underbrace{\mathcal{H}^{d-1}\left(\partial^*\hat{E}_j^\ell \cap \partial D_s \cap \big((E_i)^{(1)} \bigtriangleup (E_i^\ell)^{(1)}\big)\right)}_{=:b_{ij}^\ell}.\end{aligned}$$

Multiplying by $|w_i - w_j|$ and summing over i and j yields

$$\mathrm{TV}(\hat{w}^\ell) = \mathrm{TV}_{\Omega\setminus\overline{D_s}}(w^\ell) + \mathrm{TV}_{D_s}(\hat{w}) + b^\ell.$$

with $b^\ell \coloneqq \sum_{i=1}^{M-1}\sum_{j=i+1}^{M} |w_i - w_j| b_{ij}^\ell$, where (5.4) gives $\liminf_{\ell\to\infty} b^\ell = 0$. Moreover, from (5.2) and $K \subset\subset D_s$, we obtain

$$\mathrm{TV}(w) = \mathrm{TV}_{\Omega\setminus\overline{D_s}}(w) + \mathrm{TV}_{D_s}(w).$$

In addition, (5.2), $\mathrm{TV}(w^\ell) \to \mathrm{TV}(w)$, and the lower semicontinuity of the total variation yield

$$\mathrm{TV}_{\Omega\setminus\overline{D_s}}(w^\ell) \to \mathrm{TV}_{\Omega\setminus\overline{D_s}}(w) \quad \text{and} \quad \mathrm{TV}_{D_s}(w^\ell) \to \mathrm{TV}_{D_s}(w). \qquad \square$$

Lemma 5.3. *Let Assumption 4.1 hold. Let $w \in \mathrm{BV_W}(\Omega)$ not be stationary. Let $\{w^\ell\}_\ell \subset \mathrm{BV_W}(\Omega)$ converge strictly to w. Let $\nabla F(w) \in C(\bar{\Omega})$. Then there exist $D \in \mathcal{D}$, $k_0 \in \mathbb{N}$, $\ell_0 \in \mathbb{N}$, and $\varepsilon > 0$ such that there are infinitely many $\ell \geq \ell_0$ that satisfy*

$$\mathrm{ared}^{\ell,k_0,D} \geq \sigma \mathrm{pred}^{\ell,k_0,D} \ \text{ and } \ \mathrm{pred}^{\ell,k_0,D} > \varepsilon.$$

Proof. Because w is not stationary, Theorem 3.5 gives that there exists a patch $D \in \mathcal{D}$ such that (3.1) is violated on D, that is there exist $\eta > 0$ and $\phi \in C_c^\infty(\Omega, \mathbb{R}^d)$ with $\operatorname{supp}\phi \subset\subset D$ such that

(5.5)
$$\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\int_{\partial^* E_i \cap \partial^* E_j} (w_j - w_i)\nabla F(w)(x)\phi(x)\cdot n_{E_i}(x) - \alpha|w_i - w_j| \operatorname{div}_{E_i}\phi(x)\,\mathrm{d}\mathcal{H}^{d-1}(x) > \eta.$$

Let $(f_t)_{t\in(-\varepsilon,\varepsilon)}$ be defined by $f_t(x) \coloneqq x + t\phi(x)$ for $x \in \Omega$ and $t \in (-\varepsilon,\varepsilon)$ and $f_t^\# w \coloneqq \sum_{i=1}^M w_i \chi_{f_t(E_i)}$ if $\{E_1,\ldots,E_M\}$ is the Caccioppoli partition of Ω such that $w = \sum_{i=1}^M w_i\chi_{E_i}$.

Lemma 3.8 in [32] gives that we can choose a sequence $t^k \searrow 0$ and $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$ we have $f_{t^k} \in C_c^\infty(D, \mathbb{R}^d)$ and

(5.6)
$$\|f_{t^k}^\# w - w\|_{L^1} = \left(1 - \frac{1}{k}\right)\Delta^0 2^{-k}.$$

Moreover, Lemmas 3.3 and 3.5 in [32] give that there exists some function $g : (-\varepsilon, \varepsilon) \to \mathbb{R}$ that satisfies

$$(5.7) \qquad g(t) \in o(t) \quad \text{and} \quad -(\nabla F(w), f_{t^k}^{\#} w - w)_{L^2} - (\alpha\, \mathrm{TV}(f_{t^k}^{\#} w) - \alpha\, \mathrm{TV}(w)) \geq t^k \eta + g(t^k).$$

Unfortunately, this violation cannot be used directly to obtain a bound from below on $\mathrm{pred}^{\ell,k,D}$ because $f_{t^k}^{\#} w$ is infeasible for $\mathrm{TRP}(w^\ell, \nabla F(w^\ell), D, \Delta^0 2^{-k_0})$ since it does not agree with w^ℓ on $\Omega \setminus D$. Therefore, in order to relate this reduction of the linearized objective back to $\mathrm{pred}^{\ell,k,D}$, we construct competitors $w^{\ell,k}$, which eventually become feasible for $\mathrm{TRP}(w^\ell, \nabla F(w^\ell), D, \Delta^0 2^{-k_0})$ for all $\ell \geq \ell_0$. The necessary interdependent choices for $k_0 \in \mathbb{N}$ and $\ell_0 \in \mathbb{N}$ will be determined after providing the competitor construction.

Competitor construction. To this end, we use that the local variation $(f_t)_{t\in(-\varepsilon,\varepsilon)}$ induced by ϕ has $\operatorname{supp} f_t - I \subset\subset D$ for all $t \in (-\varepsilon, \varepsilon)$ because of $\operatorname{supp}\phi \subset\subset D$. Specifically, we apply Lemma 5.2 with the choices $\hat{w} = f_{t^k} w$, $k \in \mathbb{N}$, and $K = \operatorname{supp}\phi$ and obtain that there exists $D_s \subset\subset D$ with $\operatorname{supp}\phi \subset\subset D_s$ such that for all $k \in \mathbb{N}$ there are functions $\{w^{\ell,k}\}_\ell$ such that $w^{\ell,k}(x) = w^\ell(x)$ for $x \in \Omega \setminus D_s$ and $w^{\ell,k}(x) = f_{t^k}^{\#} w(x)$ for $x \in D_s$ such that

$$\begin{aligned}
w^{\ell,k} &\overset{*}{\rightharpoonup} f_{t^k}^{\#} w \text{ in } \mathrm{BV_W}(\Omega) \text{ for } \ell \to \infty,\\
\mathrm{TV}(w^{\ell,k}) &= \mathrm{TV}_{\Omega\setminus\overline{D_s}}(w^\ell) + \mathrm{TV}_{D_s}(f_{t^k}^{\#} w) + b^\ell \text{ with } b^\ell \geq 0 \text{ and } \liminf_{\ell\to\infty} b^\ell = 0,\\
\mathrm{TV}(w) &= \mathrm{TV}_{\Omega\setminus\overline{D_s}}(w) + \mathrm{TV}_{D_s}(w),\\
\mathrm{TV}_{\Omega\setminus\overline{D_s}}(w^\ell) &\to \mathrm{TV}_{\Omega\setminus\overline{D_s}}(w),\\
\mathrm{TV}_{D_s}(w^\ell) &\to \mathrm{TV}_{D_s}(w).
\end{aligned}$$

Determination of $k_0 \in \mathbb{N}$ and $\ell_0 \in \mathbb{N}$. Let $k_0 \in \mathbb{N}$ be such that (5.6) and (5.7) hold for all $k \geq k_0$. Then $w^\ell \to w$ strictly, the competitor properties above, and Assumption 4.1 1 imply that we can find $\ell_0(k_0) \in \mathbb{N}$ such that

$$(5.8) \qquad \begin{aligned}
\|w^\ell - w\|_{L^1(D_s)} &\leq \frac{1}{k_0}\Delta^0 2^{-k_0},\\
\|\nabla F(w^\ell) - \nabla F(w)\|_{L^\infty(D_s)} &\leq \frac{1}{k_0},\\
|\mathrm{TV}_{D_s}(w^\ell) - \mathrm{TV}_{D_s}(w)| &\leq \frac{1}{k_0}\Delta^0 2^{-k_0}, \text{ and}\\
b^\ell &\leq \frac{1}{k_0}\Delta^0 2^{-k_0}
\end{aligned}$$

hold for infinitely many $\ell \geq \ell_0(k_0)$. Then the triangle equality gives that these w^{ℓ,k_0} are feasible for $\mathrm{TRP}(w^\ell, \nabla F(w^\ell), D, \Delta^0 2^{-k_0})$. For these ℓ, we obtain from Assumption 4.1 1 that

$$\mathrm{ared}^{\ell,k_0,D} = \mathrm{pred}^{\ell,k_0,D} + o(\Delta^0 2^{-k_0})$$

Consequently, the feasibility of w^{ℓ,k_0} and thus suboptimality for $\mathrm{TRP}(w^\ell, \nabla F(w^\ell), D, \Delta^0 2^{-k_0})$ gives for

$$p^{\ell,k_0} := (\nabla F(w^\ell), w^\ell - w^{\ell,k_0})_{L^2} + \alpha\, \mathrm{TV}(w^\ell) - \alpha\, \mathrm{TV}(w^{\ell,k_0})$$

that $\mathrm{pred}^{\ell,k_0,D} \geq p^{\ell,k_0}$ and in turn

$$\mathrm{ared}^{\ell,k_0,D} \geq \sigma \mathrm{pred}^{\ell,k_0,D} + (1-\sigma)p^{\ell,k_0} + o(\Delta^0 2^{-k_0})$$

hold.

We consider the first term in p^{ℓ,k_0}. After inserting suitable zeros, in particular using $w^\ell = w^{\ell,k_0}$

in $\Omega \setminus \overline{D_s}$ and $\operatorname{supp}\phi \subset D_s$, we obtain by means of (5.8) that

$$\begin{aligned}(\nabla F(w^\ell), w^\ell - w^{\ell,k_0})_{L^2(\Omega)} &= (\nabla F(w), w - w^{\ell,k_0})_{L^2(D_s)} + (\nabla F(w^\ell) - \nabla F(w), w - w^{\ell,k_0})_{L^2(D_s)} \\ &\quad + (\nabla F(w^\ell), w^\ell - w)_{L^2(D_s)} \\ &\geq (\nabla F(w), w - f^{\#}_{t^{k_0}} w)_{L^2(\Omega)} \\ &\quad - \|\nabla F(w^\ell) - \nabla F(w)\|_{L^\infty(D_s)} \|w - f^{\#}_{t^{k_0}} w\|_{L^1(D_s)} \\ &\quad - \|\nabla F(w^\ell)\|_{L^\infty(D_s)} \|w^\ell - w\|_{L^1(D_s)} \\ &\geq (\nabla F(w), w - f^{\#}_{t^{k_0}} w)_{L^2(\Omega)} \\ &\quad - \frac{1}{k_0} \|f^{\#}_{t^{k_0}} w - w\|_{L^1(D_s)} - c\|w^\ell - w\|_{L^1(D_s)} \\ &= (\nabla F(w), w - f^{\#}_{t^{k_0}} w)_{L^2(\Omega)} \underbrace{- \left(\frac{1}{1-k_0} + c\right) \frac{1}{k_0} \Delta^0 2^{-k_0}}_{=o(\Delta^0 2^{-k_0})}.\end{aligned}$$

with $c = \sup_{\tilde{\ell} \geq \ell_0(k_0))} \|\nabla F(w^{\tilde{\ell}})\|_{L^\infty(D_s)}$, which is bounded because of the continuous differentiability of $F : L^1(\Omega) \to \mathbb{R}$ and the set $\{w^\ell \mid \ell \in \mathbb{N}\} \cup \{w\}$ being a compact in $L^1(\Omega)$.

Next, we consider the difference $\mathrm{TV}(w^\ell) - \mathrm{TV}(w^{\ell,k_0})$ and obtain with a similar procedure

$$\begin{aligned}\mathrm{TV}(w^\ell) - \mathrm{TV}(w^{\ell,k_0}) &= \mathrm{TV}_{D_s}(w) - \mathrm{TV}_{D_s}(w^{\ell,k_0}) - b^\ell \\ &\quad + \mathrm{TV}_{D_s}(w^\ell) - \mathrm{TV}_{D_s}(w) \\ &\geq \mathrm{TV}(w) - \mathrm{TV}(f^{\#}_{t^{k_0}} w) \underbrace{- \frac{2}{k_0} \Delta^0 2^{-k_0}}_{=o(\Delta^0 2^{-k_0})}\end{aligned}$$

for all of the infinitely many suitable $\ell \geq \ell_0(k_0)$.

In combination, we can deduce

$$\begin{aligned}p^{\ell,k_0} &\geq (\nabla F(w), w - f^{\#}_{t^{k_0}} w)_{L^2} + \alpha\,\mathrm{TV}(w^\ell) - \alpha\,\mathrm{TV}(f^{\#}_{t^{k_0}} w) + o(\Delta^0 2^{-k_0}) \\ &\geq t^{k_0}\eta + o(t^{k_0}) + o(\Delta^0 2^{-k_0}),\end{aligned} \tag{5.9}$$

where the second inequality follows from (5.7). Using (5.6), we obtain

$$\Delta^0 2^{-k_0} = \frac{k_0}{k_0 - 1} \left\| f^{\#}_{t^{k_0}} w - w \right\|_{L^1} \leq \kappa t^{k_0}$$

for some $\kappa > 0$, where the last inequality follows from Lemma 3.8 in [32]. Consequently, we obtain

$$\mathrm{ared}^{\ell,k_0,D} \geq \sigma \mathrm{pred}^{\ell,k_0,D} + (1-\sigma)\kappa t^{k_0} + o(t^{k_0}).$$

Therefore, we can choose $k_0 \in \mathbb{N}$ large enough such that the sum of the second and third term is positive. We now set $\ell_0 \coloneqq \ell_0(k_0)$ and obtain for infinitely many $\ell \geq \ell_0$ due to (5.4) that $\mathrm{ared}^{\ell,k_0,D} \geq \sigma \mathrm{pred}^{\ell,k_0,D}$ By potentially choosing $k_0 \in \mathbb{N}$ larger and adjusting ℓ_0 accordingly, we also obtain $\mathrm{pred}^{\ell,k_0,D} \geq p^{\ell,k_0} > \varepsilon > 0$ for some $\varepsilon > 0$ from (5.9). ∎

Theorem 5.4. *Let Assumption* 4.1 *hold. Let* $\nabla F(w^n) \in C(\bar{\Omega})$. *Then the loop starting in Algorithm* 4.1 *Line 3 terminates after finitely many iterations and* $\mathcal{A}$ *is not empty if* w^n *is not stationary.*

Proof. We need to prove that eventually $\mathcal{W} = \emptyset$. We first observe that for every choice of k in the loop starting in Line 3, the set $\mathcal{W}$ contains at most $|\mathcal{D}|$ elements so that the loop starting in Line 4 always terminates finitely. In addition, elements are never removed but only added to the set $\mathcal{A}$.

We briefly argue that by Lemma 5.2, Algorithm 4.1 Line 8 will be executed at least once over the course of the inner loop of Line 3. Specifically, applying Lemma 5.3 with the choice $w^\ell \coloneqq w^n$ for all $\ell \in \mathbb{N}$ gives a tuple (k_1, D_1) such that the acceptance criterion in Line 8 holds. Consequently, there is some patch D such that for increasing k Line 11 is executed and $(k, D) \in \mathcal{W}$ holds until the acceptance criterion in Line 8 or $\max_{(\tilde{k},\tilde{D}) \in \mathcal{A}} \mathrm{ared}^{n,\tilde{k},\tilde{D}} \geq \mathrm{pred}^{n,k,D} + L_{\nabla F} \Delta^0 2^{-k}$ holds. In both cases, $\mathcal{A}$ is not empty after iteration k.

We also observe that the max in Line 10 is strictly greater than zero from this point on because pairs (k, D) are only inserted into $\mathcal{A}$ if $\mathrm{pred}^{n,k,D}$ and in turn also $\mathrm{ared}^{n,k,D}$ are strictly positive.

We proceed by contradiction and assume that the algorithm does not terminate finitely so that there is a patch D_2 such that for all $k \in \mathbb{N}$, the criterion in Line 10 is satisfied while the criterion in Line 8 is not satisfied. In combination with $\mathscr{A}$ eventually being non-empty and the left hand side of the max in Line 10 being strictly positive, this means that pred^{n,k,D_2} is uniformly bounded away from zero for all $k \in \mathbb{N}$.

Clearly, for all feasible points of $\text{TRP}(w^n, \nabla F(w^n), D_2, \Delta^0 2^{-k})$ and thus the minimizer $\tilde{w}^{n,k,D_2}$ it holds

$$
\begin{aligned}
J(w^n) - J(\tilde{w}^{n,k,D_2}) = (\nabla F(w^n), w^n - \tilde{w}^{n,k,D_2})_{L^2} + \alpha \operatorname{TV}(w^n) - \alpha \operatorname{TV}(\tilde{w}^{n,k,D_2}) \\
+ (\nabla F(\xi^{n,k,D_2}) - \nabla F(w^n), w^n - \tilde{w}^{n,k,D_2})_{L^2}
\end{aligned}
$$

for some ξ^{n,k,D_2} in the line segment between w^n and $\tilde{w}^{n,k,D_2}$. Moreover,

$$
|(\nabla F(\xi^{n,k,D_2}) - \nabla F(w^n), \tilde{w}^{n,k,D_2} - w^n)_{L^2}| \le L_{\nabla F} \Delta^0 2^{-k}
$$

holds by means of Assumption 4.1 2. In turn, we obtain

$$
\frac{\text{ared}^{n,k,D_2}}{\text{pred}^{n,k,D_2}} = \frac{\text{pred}^{n,k,D_2} + (\nabla F(\xi^{n,k,D_2}) - \nabla F(w^n), w^n - \tilde{w}^{n,k,D_2})_{L^2}}{\text{pred}^{n,k,D_2}} \to 1 > \sigma
$$

so that eventually Line 8 holds for D_2 and some large enough k too, thereby contradicting that the criterion in Line 8 is never satisfied for D_2. ∎

Remark 5.5. *A close inspection of the argument of Theorem* 5.4 *shows that it is possible to avoid Assumption* 4.1 *2 in Theorem* 5.4 *and thus in our overall arguments by replacing* $L_{\nabla F}$ *in Algorithm* 4.1 *by* $2c(b)$ *with* $c(b)$ *from* (4.1) *if a bound* $b > \operatorname{TV}(w^{n,k,D})$ *can be established uniformly for all iterates* $w^{n,k,D}$. *Such bounds exist due to the properties of* F *and the descent properties of the algorithm. If*

$$
c(\infty) = \sup\{\|\nabla F(w)\|_{L^\infty} \,|\, w(x) \in W \text{ a.e.}\} < \infty
$$

holds, this is of course sufficient too.

Lemma 5.6. *Let* $\{w^n\}_n$ *be the sequence of iterates produced by Algorithm* 4.1. *Then the sequence of objective values* $\{J(w^n)\}_n$ *is monotonously non-increasing and convergent. The sequence* $\{w^n\}_n$ *admits a feasible weak-** *accumulation point in* $\operatorname{BV_W}(\Omega)$.

Proof. The sequence of objective values is monotonically nonincreasing by construction and, because F and TV are both bounded below, also convergent. Because $\{\operatorname{TV}(w^n)\}_n$ is uniformly bounded above by $J(w^0) < \infty$, $\{w^n\}_n$ admits a weak-* accumulation point $\bar{w}$, which is feasible, that is in $\operatorname{BV_W}(\Omega)$. ∎

If $\bar{w}$ is also a strict accumulation point, that is $\operatorname{TV}(w^{n_k}) \to \operatorname{TV}(\bar{w})$ in addition to $w^{n_k} \overset{*}{\rightharpoonup} \bar{w}$ in $\operatorname{BV}(\Omega)$ holds for a subsequence $\{w^{n_k}\}_k$, then $J(w^{n_k}) \to J(\bar{w})$ under a continuity assumption on F. Consequently, we desire that every weak-* accumulation point is a strict accumulation point.

Lemma 5.7. *Let Assumptions* 4.1 *and* 5.1 *hold and* $\{w^n\}_n$ *be the sequence of iterates produced by Algorithm* 4.1. *Every weak-** *accumulation point* $\bar{w}$ *of* $\{w^n\}_n$ *is a strict accumulation point and* $J(w^n) \to J(\bar{w})$.

Proof. Let $\bar{w}$ be a weak-* accumulation point in $\operatorname{BV}(\Omega)$ with approximating subsequence $\{w^{n_\ell}\}_\ell$. Then $w^{n_\ell} \to \bar{w}$ in $L^1(\Omega)$ and in turn $F(w^{n_\ell}) \to F(\bar{w})$.

Outline. By way of contradiction, we assume that $\bar{w}$ is not strict, that is, after a possible restriction to a sub-subsequence,

$$
\lim_{\ell\to\infty} \operatorname{TV}(w^{n_\ell}) > \operatorname{TV}(\bar{w}) + \varepsilon \tag{5.10}
$$

for some $\varepsilon > 0$. We are going to localize this inequality, that is we show that there exists an open ball B such that $\overline{B} \subset D$ for some $D \in \mathscr{D}$, and $\varepsilon_B > 0$ such that

$$
\lim_{\ell\to\infty} \operatorname{TV_B}(w^{n_\ell}) > \operatorname{TV_B}(\bar{w}) + \varepsilon_B \tag{5.11}
$$

and whose boundary does not interfere with the interfaces of the level sets of the w^{n_ℓ} and $\bar{w}$. Specifically, we require

$$
\mathscr{H}^{d-1}\left(\partial B \cap \bigcup_{i=1}^{M} \partial^* E_i^{n_\ell}\right) = 0 \text{ for all } \ell \in \mathbb{N} \quad \text{and} \quad \mathscr{H}^{d-1}\left(\partial B \cap \bigcup_{i=1}^{M} \partial^* E_i\right) = 0, \tag{5.12}
$$

where $\{E_1, \ldots, E_M\}$ is a Caccioppoli partition of Ω such that $\bar{w} = \sum_{i=1}^{M} w_i \chi_{E_i}$ and similarly for w^{n_ℓ}, see [32, Lemma 2.1]. With this localization, we will then be able to construct competitors that are feasible for trust-region subproblems (TRP) on the patch D and guarantee a predicted and actual reduction that are bounded away from zero. In turn, we will obtain $J(w^n) \to -\infty$, which gives the desired contradiction.

Localization of (5.10) to (5.11). In order to obtain the existence of B and D, we employ a covering argument and consider the following set of closed balls:

$$\mathcal{F} = \left\{ \overline{B_s(x)} \;\middle|\; \begin{array}{c} x \in \Omega, 0 < s, D \in \mathcal{D}, \overline{B_s(x)} \subset D, \\ (5.12) \text{ holds for the choice } \overline{B} = \overline{B_s(x)} \end{array} \right\}. \tag{5.13}$$

Clearly, (5.13) would be a fine cover if (5.12) is not required for its elements. Because there are only countably many iterates w^n, there are only countably many Caccioppoli partitions $\{E_1^{n_\ell}, \ldots, E_M^{n_\ell}\}$. Consequently, we can always perturb s slightly (arbitrarily small if needed) to ensure both identities in (5.12) hold for a closed ball $\overline{B} = \overline{B_r(x)}$, see also [30, Proposition 2.16]. This means that for all $r_0 > 0$, we find some $r \in (0, r_0)$ such that $\overline{B_r(x)} \subset \overline{B_{r_0}(x)}$ and $\overline{B_r(x)} \subset D$.

The Vitali–Besicovitch covering theorem implies that there is a countable and pairwise disjoint subset $\tilde{\mathcal{F}} \subset \mathcal{F}$ such that $|Dw| = \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} |w_i - w_j| \mathcal{H}^{d-1} \llcorner (\partial^* E_i \cap \partial^* E_j \cap \Omega)$ satisfies $|Dw|(\Omega \setminus \bigcup_{\overline{B} \in \tilde{\mathcal{F}}} \overline{B}) = 0$ and $|Dw^{n_\ell}|(\Omega \setminus \bigcup_{\overline{B} \in \tilde{\mathcal{F}}} \overline{B}) = 0$ for all $\ell \in \mathbb{N}$.

In combination with the fact that (5.12) holds for all balls in $\mathcal{F}$ and thus also $\tilde{\mathcal{F}}$, we obtain (5.11) for some open ball B with $\overline{B} \subset \Omega$ and $\varepsilon_B > 0$.

Competitor construction. We seek competitors $\hat{w}^{n_\ell}$, which eventually become feasible for the problem $\mathrm{TRP}(w^{n_\ell}, \nabla F(w^{n_\ell}), D, \Delta^0 2^{-k})$ for large enough ℓ and suitable corresponding k. Let $r > 0$ and $\bar{x} \in D$ satisfy $B = B_r(\bar{x})$ for the above-asserted B and D. Moreover, let $r_1 > r$ be small enough such that $\overline{B_{r_1}(\bar{x})} \subset D$ holds. It is possible to find such r_1 due to the strictly positive distance of B to the boundary ∂D (note that D is open and $\overline{B}$ is closed). We apply Lemma 5.2 with the choices $\hat{w} = \bar{w}$ and $K = \overline{B_{r_1}(\bar{x})}$. We obtain that there exist $D_s \subset\subset D$ with $\overline{B_{r_1}(\bar{x})} \subset\subset D_s$ and functions $\{\hat{w}^{n_\ell}\}_\ell$ with $\hat{w}^{n_\ell}(x) = w^{n_\ell}(x)$ for $x \in \Omega \setminus D_s$ and $\hat{w}^{n_\ell}(x) = \bar{w}(x)$ for $x \in D_s$ such that

$$\begin{aligned}
\hat{w}^{n_\ell} &\overset{*}{\rightharpoonup} \bar{w} \text{ in } \mathrm{BV_W}(\Omega) \text{ for } \ell \to \infty, \\
\mathrm{TV}(\hat{w}^{n_\ell}) &= \mathrm{TV}_{\Omega \setminus \overline{D_s}}(w^{n_\ell}) + \mathrm{TV}_{D_s}(\bar{w}) + b^\ell \text{ with } b^\ell \geq 0 \text{ and } \liminf_{\ell \to \infty} b^\ell = 0, \\
\mathrm{TV}(\bar{w}) &= \mathrm{TV}_{\Omega \setminus \overline{D_s}}(\bar{w}) + \mathrm{TV}_{D_s}(\bar{w}).
\end{aligned}$$

Using (5.11) and (5.12), we obtain

$$\begin{aligned}
\mathrm{TV}(\hat{w}^{n_\ell}) &= \mathrm{TV}_{\Omega \setminus \overline{D_s}}(w^{n_\ell}) + \mathrm{TV}_{D_s}(\bar{w}) + b^\ell \\
&= \mathrm{TV}_{\Omega \setminus \overline{D_s}}(w^{n_\ell}) + \mathrm{TV}_{D_s \setminus B}(\bar{w}) + \mathrm{TV}_B(\bar{w}) + b^\ell \\
&< \mathrm{TV}_{\Omega \setminus \overline{D_s}}(w^{n_\ell}) + \lim_{\ell \to \infty} \mathrm{TV}_{D_s \setminus B}(w^{n_\ell}) + \mathrm{TV}_B(w^{n_\ell}) + b^\ell - \varepsilon_B.
\end{aligned}$$

Taking the lim inf over ℓ on both sides and using (5.12) again, we obtain

$$\mathrm{TV}(\bar{w}) \leq \liminf_{\ell \to \infty} \mathrm{TV}(\hat{w}^{n_\ell}) \leq \lim_{\ell \to \infty} \mathrm{TV}(w^{n_\ell}) - \varepsilon_B. \tag{5.14}$$

Contradiction $J(w^n) \to -\infty$. We follow the arguments of the proof of Theorem 6.4 in [32] (namely Outcome 3, part 2) in order to show that the competitors constructed above eventually become feasible and enforce a reduction of the objective that is bounded strictly away from zero infinitely often, which in turn contradicts that the objective is bounded below.

Let $\delta := \frac{\varepsilon_B}{2}$. Because of Assumption 4.1 and $\hat{w}^{n_\ell} \to \bar{w}$ in $L^2(\Omega)$ as well as $w^{n_\ell} \to \bar{w}$ in $L^2(\Omega)$, we obtain that there exist some large enough $k_0 \in \mathbb{N}$ and $\ell_0 \in \mathbb{N}$ and such that by virtue of (5.14) for infinitely many $\ell \geq \ell_0$ and all $\tilde{w} \in \mathrm{BV_W}(\Omega)$ with $\|\tilde{w} - w^{n_\ell}\|_{L^1} \leq \Delta^0 2^{-k_0}$ it holds that

$$|F(w^{n_\ell}) - F(\tilde{w})| \leq \frac{1-\sigma}{3-\sigma} \alpha \delta, \tag{5.15}$$

$$|(\nabla F(w^{n_\ell}), w^{n_\ell} - \tilde{w})_{L^2}| \leq \frac{1-\sigma}{3-\sigma} \alpha \delta, \tag{5.16}$$

$$\|\hat{w}^{n_\ell} - w^{n_\ell}\|_{L^1} \leq \Delta^0 2^{-k_0}, \text{ and} \tag{5.17}$$

$$\alpha \, \mathrm{TV}(w^{n_\ell}) - \alpha \, \mathrm{TV}(\hat{w}^{n_\ell}) \geq \alpha \delta, \tag{5.18}$$

where (5.17) gives that $\hat{w}^{n_\ell}$ is feasible for $\mathrm{TRP}(w^{n_\ell}, \nabla F(w^{n_\ell}), D, \Delta^0 2^{-k_0})$.

Because $\mathrm{pred}^{\,n_\ell,k_0,D}$ is computed in Algorithm 4.1 Line 6 as the negation of the objective value of a minimizer of $\mathrm{TRP}(w^{n_\ell}, \nabla F(w^{n_\ell}), D, \Delta^0 2^{-k_0})$, the feasibility (5.17) and the objective term estimates (5.16) and (5.18) yield

$$\mathrm{pred}^{\,n_\ell,k,D} \geq (\nabla F(w^{n_\ell}), w^{n_\ell} - \hat{w}^{n_\ell})_{L^2} + \alpha\,\mathrm{TV}(w^{n_\ell}) - \alpha\,\mathrm{TV}(\hat{w}^{n_\ell}) \geq -\frac{1-\sigma}{3-\sigma}\alpha\delta + \alpha\delta. \tag{5.19}$$

Moreover, (5.15) and (5.16) give

$$\mathrm{ared}^{\,n_\ell,k_0,D} \geq \mathrm{pred}^{\,n_\ell,k_0,D} - 2\frac{1-\sigma}{3-\sigma}\alpha\delta,$$

where the right-hand side is strictly positive because of (5.19) and $1 - \frac{1-\sigma}{3-\sigma} > 2\frac{1-\sigma}{3-\sigma}$. Consequently, for suitable ℓ it holds that

$$\frac{\mathrm{ared}^{\,n_\ell,k_0,D}}{\mathrm{pred}^{\,n_\ell,k_0,D}} \geq \frac{\mathrm{pred}^{\,n_\ell,k_0,D} - 2\frac{1-\sigma}{3-\sigma}\alpha\delta}{\mathrm{pred}^{\,n_\ell,k_0,D}} \geq \frac{1 - 3\frac{1-\sigma}{3-\sigma}}{1 - \frac{1-\sigma}{3-\sigma}} = \sigma,$$

where the second inequality follows from the monotonocity of the function $p \mapsto p^{-1}(p - 2\frac{1-\sigma}{3-\sigma})$.

Consequently, for infinitely many $\ell \geq \ell_0$, the loop starting at Algorithm 4.1 Line 3 terminates after finitely many iterations with $(\bar{k}, \bar{D}) \in \arg\max\{\mathrm{ared}^{n_\ell,\bar{k},\bar{D}} \,|\, (\bar{k}, \bar{D}) \in \mathscr{A}\}$ such that

$$\mathrm{ared}^{n_\ell,\bar{k},\bar{D}} \geq \mathrm{ared}^{n_\ell,k_0,D} \geq \sigma \mathrm{pred}^{n_\ell,k_0,D} \geq 2\frac{1-\sigma}{3-\sigma}\sigma\alpha\delta$$

Since the change induced by the maximizer $(\bar{k}, \bar{D})$ will definitely be applied in Line 28 and all further reductions decrease the objective even more, we obtain

$$F(w^{n_\ell+1}) + \alpha\,\mathrm{TV}(w^{n_\ell+1}) - F(w^{n_\ell}) - \alpha\,\mathrm{TV}(w^{n_\ell}) \geq \mathrm{ared}^{n_\ell,\bar{k},\bar{D}} \geq \sigma 2\alpha \frac{1-\sigma}{3-\sigma}\frac{\varepsilon_B}{2} > 0.$$

Because the sequence $\{J(w^n)\}_n$ decreases monotonically, we obtain the contradiction $J(w^n) \to -\infty$.

Objective value convergence. F is continuous and $w^{n_\ell} \stackrel{*}{\rightharpoonup} \bar{w}$ in $\mathrm{BV_W}(\Omega)$ implies $w^{n_\ell} \to \bar{w}$ in $L^1(\Omega)$ and in turn $w^{n_\ell} \to \bar{w}$ in $L^2(\Omega)$ because $\mathrm{BV_W}(\Omega)$ is bounded in $L^\infty(\Omega)$. In combination with the convergence of $J(w^n)_n$ and the strict convergence of $\{\mathrm{TV}(w^{n_\ell})\}_\ell\}$, we obtain $J(w^{n_\ell}) \to J(\bar{w}) = \inf_{n\in\mathbb{N}} J(w^n)$. ∎

Theorem 5.8. *Let Assumptions* 3.3, 4.1 *and* 5.1 *hold. Let $\{w^n\}_n$ be the sequence of iterates produced by Algorithm* 4.1*. Then one of the following mutually exclusive outcomes holds:*

1. *The sequence $\{w^n\}_n$ is finite and the final element w^N for some $N \in \mathbb{N}$ satisfies the following. For all $D \in \mathscr{D}$ there exists $k \in \mathbb{N}$ such that w^N solves* $\mathrm{TRP}(w^N, \nabla F(w^N), D, \Delta^0 2^{-k})$*. In particular, w^N is stationary if $\nabla F(w^N) \in C(\bar{\Omega})$.*
2. *The sequence $\{w^n\}_n$ is finite and the final element w^N for some $N \in \mathbb{N}$ satisfies the following. The loop over k that begins in Line 3 does not terminate finitely. In particular, w^N is stationary if $\nabla F(w^N) \in C(\bar{\Omega})$.*
3. *The sequence $\{w^n\}_n$ has a weak-* accumulation point in* $\mathrm{BV}(\Omega)$*. All weak-* accumulation points are in* $\mathrm{BV_W}(\Omega)$*. If a weak-* accumulation point $\bar{w}$ satisfies $\nabla F(\bar{w}) \in C(\bar{\Omega})$, it is stationary.*

Proof. We follow the basic proof strategy of Theorem 4.23 in [28] and Theorem 6.4 in [32] and extend it so that we can perform a localization to patches. We assume that Outcomes 1 and 2 do not hold and prove that Outcome 3 must hold in this case.

Lemma 5.6 gives the existence of a weak-* accumulation point $\bar{w}$ with approximating subsequence of iterates $\{w^{n_\ell}\}_\ell$. Because weak-* convergence implies pointwise a.e. convergence for a subsequence, we have $\bar{w} \in \mathrm{BV_W}(\Omega)$. Now let $\nabla F(\bar{w}) \in C(\bar{\Omega})$. Lemma 5.7 gives $\mathrm{TV}(w^{n_\ell}) \to \mathrm{TV}(\bar{w})$ and $F(w^{n_\ell}) \to F(\bar{w})$. We proceed by way of contradiction and assume that $\bar{w}$ is not stationary.

We apply Lemma 5.3 and obtain that there are $D \in \mathscr{D}$, $k_0 \in \mathbb{N}$, $\ell_0 \in \mathbb{N}$, and $\varepsilon > 0$ such that

$$\mathrm{ared}^{n_\ell,k_0,D} \geq \sigma \mathrm{pred}^{n_\ell,k_0,D} \;\text{ and }\; \mathrm{pred}^{n_\ell,k_0,D} > \varepsilon.$$

Consequently, for infinitely many $\ell \geq \ell_0$, the loop starting at Algorithm 4.1 Line 3 terminates after finitely many iterations with $(\bar{k}, \bar{D}) \in \arg\max\{\mathrm{ared}^{n_\ell,\bar{k},\bar{D}} \,|\, (\bar{k}, \bar{D}) \in \mathscr{A}\}$ such that

$$\mathrm{ared}^{n_\ell,\bar{k},\bar{D}} \geq \mathrm{ared}^{n_\ell,k_0,D} \geq \sigma \mathrm{pred}^{n_\ell,k_0,D} \geq \sigma\varepsilon.$$

Since the change induced by the maximizer $(\bar{k}, \bar{D})$ will definitely be applied in Line 28 and all further reductions decrease the objective even more, we obtain

$$F(w^{n_\ell+1}) + \alpha\,\mathrm{TV}(w^{n_\ell+1}) - F(w^{n_\ell}) - \alpha\,\mathrm{TV}(w^{n_\ell}) \geq \mathrm{ared}^{n_\ell,\bar{k},\bar{D}} \geq \sigma\varepsilon > 0.$$

Note that the while loop stops at the first instance we encounter an increase after the guaranteed first decrease or the set $\mathscr{A}$ becomes empty eventually so that we always get finite termination of this loop. Because the sequence $J(w^n)_n$ decreases monotonically, we obtain the contradiction $J(w^n) \to -\infty$. ∎

6. Numerical Experiments. This section performs numerical tests to compare the block-SLIP algorithm Algorithm 4.1 results to its primary competitor, SLIP [28, 32]. Note that block-SLIP is equivalent to SLIP if the number of patches is equal to one.

The goal of our numerical study is to assess if and when the block-SLIP algorithm can outperform SLIP and help to scale the problem sizes. To this end, we consider two test cases. First, we consider an integer control problem that is defined on a one-dimensional domain, where we can use the efficient subproblem solver from [39]. Second, we consider an integer control problem on a two-dimensional domain, where no efficient subproblem solver is known and we need to resort to an off-the-shelf integer programming solver.

Throughout this section, results of block-SLIP instances are designated by the subscript bs and results of SLIP instances are designated by the subscript s. We have carried out the experiments on a node of the Linux HPC cluster LiDO3 with two AMD EPYC 7542 32-Core CPUs and 64 GB RAM (computations were restricted to one CPU).

6.1. 1D Test Case. We solve the integer optimal control benchmark problem that is defined in (5.1) in [28] on the one-dimensional domain $\Omega = (-1, 1)$ with the discretization reported therein. Specifically, the problem reads

$$\min_{w} \frac{1}{2}\|Kw - f\|^2_{L^2(\Omega)} + \alpha\,\mathrm{TV}(w) \text{ s.t. } w(t) \in W \text{ for a.e. } t \in \Omega, \tag{6.1}$$

where we make the choices $W = \{-1, 0, 1\}$ and $f(t) = 0.2\cos(2\pi t - 0.25)\exp(t)$ for $t \in (-1, 1)$ and have $Kw = k * w$ with $(k * w)(t) = \int_{-1}^{t} k(t-\tau)w(\tau)\,\mathrm{d}\tau$ with k as in §5 in [28]. It is a deconvolution problem that stems from *Filtered Approximation* in electronics. It was analyzed as a finite-dimensional *convex quadratic integer program* in [3] and different variants have been used as benchmark problems for integer optimal control algorithms in [20, 28, 33, 39].

Subproblem Solution. We use the topological sorting-based algorithm from [39] as subproblem solver since (our latest implementation[1]) outperforms the A^*-based and integer programming-based solution approaches that are also discussed in [39] on our benchmark problems significantly.

Benchmark. For our experiments, we make the choices of $N \in \{2^{12}, 2^{14}\}$ for the number of discretization intervals for the piecewise constant ansatz for the control input function and $\alpha \in \{1.25 \cdot 10^{-4}, 5.0 \cdot 10^{-4}, 2.0 \cdot 10^{-3}\}$. We cover a spectrum of values for α because the the numbers of iterations that SLIP and block-SLIP require can differ substantially for different values α. We highlight that, in contrast to an integer programming solver as is used for the second test case, the performance of the topological sorting-based subproblem solver only depends on the current value of the trust-region radius and not on the data (current iterate), see also [39]. As initial value for the optimization, we choose the control $w^0 \equiv 0$ for all instances.

For all of these instances, we run SLIP, block-SLIP with $N_p = 4$ patches, and block-SLIP with $N_p = 9$ patches. The patches are uniform intervals, uniformly distributed over the domain, and always overlap by 0.2 with their left and right neighbors.

Algorithm Setup. Regarding the algorithm, we choose $\Delta_0 = 0.125$ and $\sigma = 10^{-4}$. We can determine contraction of the trust-region when Δ falls below the volume of one grid cell so that, due its discreteness, the subproblem solution coincides with the previous iterate and no progress is possible. Then we terminate the algorithm. We also prescribe a limit of 1000 outer iterations but note that all SLIP and block-SLIP runs on our instances terminate due to a contraction of the trust-region radius/no further progress being possible.

Results. SLIP and block-SLIP have generally returned points with similar objective values. Specifically, the objective values for all instances with $\alpha \in \{5.0 \cdot 10^{-4}, 2.0 \cdot 10^{-3}\}$ are very close; they differ by less than 1 %. Only for the instance with $\alpha = 1.25 \cdot 10^{-4}$, block-SLIP returns a better objective value than SLIP, which is approximately 8.7 % lower than the one produced by SLIP (consistently across discretizations and number of patches). This behavior is not unexpected

[1]https://github.com/paulmanns/trs4slip

since SLIP and block-SLIP are local optimization techniques that compute different steps that may converge to different (stationary) points. We highlight that due to the nonconvexity and the nontrivial nature of the optimality condition we do not know how many stationary points the problem has and how this information could straightforwardly be obtained. Regarding run times, SLIP outperforms block-SLIP by a wide margin. This is due to the run time complexity of the topological sorting-based subproblem solver, which is $O(\tilde{N}\tilde{K}M)$. Here, $\tilde{N}$ is equal to N for SLIP and is equal to the number of intervals of the patch problem for block-SLIP, which is larger than N/N_p because of the overlap of the patches; $\tilde{K} = \max\{\Delta N, \tilde{N}\}$, which is equal to ΔN on all of our instances for both SLIP and block-SLIP. As a consequence, if the same number of outer iterations is executed and the steps on the different patches become acceptable for the same trust-region radius, the run time spent in the subproblem solver for block-SLIP must be higher than for SLIP. We observe that block-SLIP generally requires more outer iterations on our benchmark, increasing the run time spent in the subproblem solver even more. Moreover, many additional evaluations of the control-to-state operator are necessary in block-SLIP, increasing the total run time of block-SLIP further. A detailed tabulation of the objective values and run times is given in Table 6.2.

In conclusion, our results for the first test case show that block-SLIP does not pay off for one-dimensional problems since the subproblem solver is extremely efficient and scales well irrespective of the specific data. The only reason that we can sensibly think of using block-SLIP in 1D are extreme cases, where very high values of N are required because the subproblem solver runs out of memory at some point. For our test case and compute environment with 64 GB RAM, this happens at $N = 2^{16}$.

Table 6.1: Objective values $J(x)$ and broken down to $f(x)$ and $\mathrm{TV}(x)$ and run times t_x for $x = x_s$ (solution returned by SLIP) and $x = x_{bs}$ (solution returned by block-SLIP) for the different instances of our one-dimensional benchmark problem. In each row, the winner(s) in terms of objective and run time up to the reported precision are highlighted with bold-faced text.

N	N_p	$\alpha \cdot 10^{-3}$	$J(x_\text{bs})$	$J(x_\text{s})$	$f(x_\text{bs})$	$\mathrm{TV}(x_\text{bs})$	$f(x_\text{s})$	$\mathrm{TV}(x_\text{s})$	t_bs	t_s
12	4	0.125	**0.002757**	0.003017	0.001257	12	0.001517	12	41	**21**
		0.500	**0.006074**	**0.006074**	0.002074	8	0.002074	8	27	**13**
		2.000	0.015788	**0.015787**	0.003788	6	0.003787	6	13	**6**
	9	0.125	**0.002759**	0.003017	0.001259	12	0.001517	12	135	**21**
		0.500	**0.006074**	**0.006074**	0.002074	8	0.002074	8	54	**13**
		2.000	**0.015787**	**0.015787**	0.003787	6	0.003787	6	22	**6**
14	4	0.125	**0.002743**	0.003010	0.001243	12	0.001510	12	1579	**615**
		0.500	**0.006072**	**0.006072**	0.002072	8	0.002072	8	636	**267**
		2.000	**0.015786**	**0.015786**	0.003786	6	0.003786	6	461	**119**
	9	0.125	**0.002744**	0.003010	0.001244	12	0.001510	12	4671	**615**
		0.500	**0.006072**	**0.006072**	0.002072	8	0.002072	8	1659	**267**
		2.000	**0.015786**	**0.015786**	0.003786	6	0.003786	6	432	**119**

6.2. 2D Test Case. As our 2D test case, we choose $W = \{0, 1\}$ and a convection-diffusion equation that is specified as follows. We consider the square domain $\Omega = (0, 1)^2$ and for a given control w with $w(x) \in W$ a.e., the state vector u is given by the solution to

$$
\begin{aligned}
-\varepsilon \Delta u + c_1 \cdot \nabla u + c_2 u w = f \quad & \text{in } \Omega \\
u = 0 \quad & \text{on } \{0,1\} \times (0,1) \cup ((0, 0.25) \cup (0.75, 1)) \times \{0\} \\
u = \sin(2\pi(x_1 - 0.25)) \quad & \text{on } (0.25, 0.75) \times \{0\} \\
\partial_n u = 0 \quad & \text{on } (0,1) \times \{1\},
\end{aligned}
\tag{6.2}
$$

where $c_2 = 2$, $c_1(x) = (\sin(\pi x_1) \quad \cos(2\pi x_2))^T$ for $x \in \Omega$, $f(x) = \sin(2\pi x_1 + 2\pi x_2) + 3$ for $x \in \Omega$, and $\varepsilon = 4 \cdot 10^{-2}$.

Let the solution operator to (6.2) be denoted by S. We choose the objective

$$F = j \circ S \quad \text{with} \quad j(u) = \frac{1}{2} \|u - u_d\|_{L^2}^2,$$

where u_d is computed by solving a variant of (6.2), where c_1 is replaced by $\tilde{c_1}(x) = (-x_2 \quad 2x_1)^T$, for $w = 2.5\chi_A - 4(x_1 - 0.35)^3\chi_A - 6(x_2 - 0.35)^3\chi_B$ with $A = (0, 0.35)^2$ and $B = \Omega \setminus (0, 0.35)^2$.

Note that we use a different PDE for computing u_d to ensure that u_d cannot be reached or almost be reached and thus avoid that $F(u)$ can have values very close to zero, which are more difficult to compare (with relative error computations which are more influenced by numerical errors when the denominator becomes close to zero).

Discretization and Subproblem Solution. In order to discretize the control and the PDE and to assemble the finite-element matrices for the PDE, we use the finite-element package `FEniCSx` [2], in which we choose a piecewise constant control ansatz on a uniform $N \times N$ of square grid cells and solve the PDE on the same grid with each grid of the cells being decomposed into 4 triangles, where we use continuous Lagrange elements of order one for u and u_d.

In order to compute $\nabla F(w)$ on the computer, we fix the discretization and then determine the adjoint using operator calculus and the finite-element system described above so that we follow a *first-discretize, then-optimize* principle. We evaluate the total variation directly on the control functions as described in [32], which may introduce an anisotropic effect on the resulting w, see the considerations in [?]. Since the convergence properties of the TV-discretization and accuracy of the geometry of the resulting functions are not the goal of our experiments, we find this reasonable to avoid much longer compute times that would otherwise be necessary when using the convergent discretization scheme from [?]. Since this discretization is applied to all instances over all algorithms, we still achieve a fair comparison. The resulting integer linear programs for the discretized trust-region subproblems are solved by means of `Gurobi` [13]; see [?] for a detailed MIP formulation and the accompanying repository[2] for a possible implementation using `Gurobi`'s python API.

Benchmark. For our experiments, we make the choices $N \in \{64, 96\}$ and $\alpha \in \{5 \cdot 10^{-4}, 7.5 \cdot 10^{-4}, 10^{-3}, 1.25 \cdot 10^{-3}, 1.5 \cdot 10^{-3}, 1.75 \cdot 10^{-3}, 2 \cdot 10^{-3}, 2.25 \cdot 10^{-3}\}$. We cover this spectrum of values for α because small and large values of α generally lead to inexpensive instances for integer programming solvers with a lot of chattering behavior in the resulting functions for small α and basically constant functions for large values of α. Consequently, we cover several values in between, where the run times are relatively high. Moreover, we solve one instance for $N = 128$ with $\alpha = 10^{-3}$. We restrict to one value of α here because the run time was very high. As initial value for the optimization, we choose the control $w^0 \equiv 0$ for all instances.

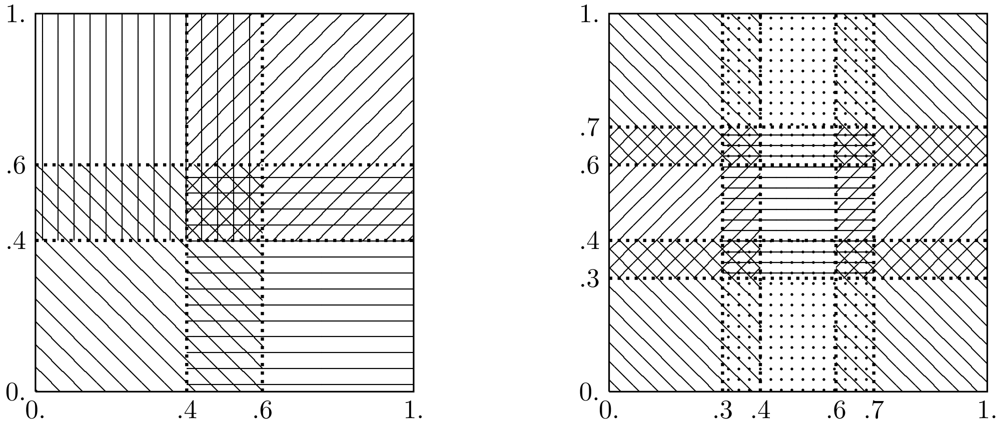
For all of these instances, we run SLIP, block-SLIP with $N_p = 4$ patches, and block-SLIP with $N_p = 9$ patches. The patches are of uniform size, uniformly distributed over the domain, and overlap by 0.1 in each axis with the neighboring patches. This is visualized in Figure 6.1.

Algorithm Setup. Regarding the algorithm, we choose $\Delta_0 = 0.125$ and $\sigma = 10^{-4}$. We can determine contraction of the trust-region when Δ falls below the volume of one grid cell so that, due its discreteness, the subproblem solution coincides with the previous iterate and no progress is possible. Then we terminate the algorithm. We prescribe a limit of 100 outer iterations, which has never been reached in our experiments.

Results. SLIP and block-SLIP have returned very similar objective values for all conducted experiments. The objective values equal to four digits of accuracy in almost all cases. In the three remaining cases, the point returned by SLIP is slightly better (relative improvement of objective value $< 1\%$). Figure 6.2 gives a visual impression for the computed solutions for $\alpha = 10^{-3}$. There are essentially two different points the algorithm variants produced.

For $N = 64$, the run times of SLIP and block-SLIP are similar but generally low (the most expensive instance for SLIP has a run time slightly less than 6 minutes). For $N = 96$, the comparison between SLIP and block-SLIP is different for the numbers of patches $N_p = 4$ and $N_p = 9$. For $N_p = 4$, there is no clear winner. While a run of block-SLIP $N_p = 4$ gives the highest absolute speedup with more than $16750\,s$ run time improvement from $66740\,s$ to $49990\,s$, there is also an instance, where block-SLIP takes $10807\,s$, which is almost twice than SLIP does on this instance with $5628\,s$. The reason is that for $N = 96$ and $N_p = 4$, the integer programs resulting from the discretized trust-region subproblems are already quite large on the different patches and therefore can have relatively long compute times. Moreover, more subproblems are solved in total due to the domain decomposition approach. This is different for the higher number of patches $N_p = 9$. Except for one inexpensive instance for the smallest value of α, the run time of block-SLIP is substantially lower compared to SLIP because the run time of the subproblems drops substantially now. For the two most expensive instances of SLIP with run times of $9151\,s$ and $66740\,s$, the achieved speedups are 25.28 and 7.4 and the run times for block-SLIP are $362\,s$ and $9017\,s$. For the instances with $N = 128$, the effects observed for $N = 96$ get amplified. The decomposition into $N_p = 4$ patches is counterproductive and the run time increases from already very expensive $134313\,s$ for SLIP to $393184\,s$ with block-SLIP for $N_p = 4$. Many trust-region

[2] https://github.com/INFORMSJoC/2024.0680



(a) $N_p = 4$ overlapping patches.

(b) $N_p = 9$ overlapping patches.

Fig. 6.1: Visualization of patch overlap for the domain $\Omega = (0, 1) \times (0, 1)$.

subproblems of block-SLIP for $N_p = 4$ are very expensive in this case. In contrast to this, the run time of block-SLIP drops to 1053 s for $N_p = 9$, which is a speedup of 127.55. A detailed tabulation of the objective values and compute times is given in Table 6.2.

The computational effort to solve the trust-region subproblems highly depends on the patch in our example. For both $N_\mathrm{p} = 4$ and $N_\mathrm{p} = 9$, the bottom left patch (see Figure 6.1) induces a much higher computational effort for the integer programming solver than all other patches. This computational effort is further concentrated to instances with comparatively large trust-region radii. We provide mean and median runtimes of the trust-region subproblems for the case $\alpha = 10^{-3}$ in Table 6.3 for $N_\mathrm{p} = 4$ and Table 6.4 for $N_\mathrm{p} = 9$. This observation is reflected in the properties of the linear programming relaxation of the trust-region subproblem. Specifically, the solution to the linear programming relaxation is already integer-valued and thus optimal for the other patches in almost all instances. We note that it can be expected but not guaranteed that the linear programming relaxation is integer-valued in large parts of the domain since Theorem 3 in [**?**] guarantees that there is at most one connected component of grid cells, where it can be fractional. In such cases, when the number of patches is relatively small and the runtimes vary greatly between the patches, we firmly believe that parallelization is advisable on the subproblem solver instead on the level of Algorithm 4.1.

In conclusion, our results show that block-SLIP pays off very well for large problem sizes and a sufficiently large number of patches so that the integer programs can be solved quickly. When compute times are already low, block-SLIP generally does not have a beneficial effect.

Acknowledgments. The authors are grateful to Francesco Maggi (UT Austin) for providing helpful guidance that lead to the competitor constructions in the proofs of Lemmas 5.3 and 5.7. The authors gratefully acknowledge computing time on the LiDO3 HPC cluster at TU Dortmund, partially funded in the Large-Scale Equipment 796 Initiative by the Deutsche Forschungsgemeinschaft (DFG) as project 271512359. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

REFERENCES

[1] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, vol. 254 of Oxford Mathematical Monographs, Clarendon Press Oxford, 2000.

[2] I. A. Baratta, J. P. Dean, J. S. Dokken, M. Habera, J. S. Hale, C. N. Richardson, M. E. Rognes, M. W. Scroggs, N. Sime, and G. N. Wells, *DOLFINx: the next generation FEniCS problem solving environment*. preprint, 2023, https://doi.org/10.5281/zenodo.10447666.

[3] C. Buchheim, A. Caprara, and A. Lodi, *An effective branch-and-bound algorithm for convex quadratic integer programming*, Mathematical programming, 135 (2012), pp. 369–395.

[4] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to*

Table 6.2: Objective values $J(x)$ and broken down to $f(x)$ and $\mathrm{TV}(x)$ and run times t_x for $x = x_s$ (solution returned by SLIP) and $x = x_{bs}$ (solution returned by block-SLIP) for the different instances of our two-dimensional benchmark problem. In each row, the winner(s) in terms of objective and run time up to the reported precision are highlighted with bold-faced text. The improvements in run time by block-SLIP with $N_p = 9$ for the three most expensive instances (for SLIP) are additionally highlighted with color boxes. Run times are reported in seconds.

N	N_p	$\alpha \cdot 10^{-3}$	$J(x_\mathrm{bs})$	$J(x_\mathrm{s})$	$f(x_\mathrm{bs})$	$\mathrm{TV}(x_\mathrm{bs})$	$f(x_\mathrm{s})$	$\mathrm{TV}(x_\mathrm{s})$	t_bs	t_s
64	4	0.50	**0.6749**	**0.6749**	0.6740	1.8593	0.6740	1.8593	**42**	55
		0.75	**0.6753**	**0.6753**	0.6740	1.7500	0.6740	1.7500	**51**	107
		1.00	**0.6758**	**0.6758**	0.6745	1.2812	0.6745	1.2812	**214**	353
		1.25	**0.6761**	**0.6761**	0.6745	1.2656	0.6745	1.2656	**222**	256
		1.50	**0.6764**	**0.6764**	0.6745	1.2656	0.6745	1.2656	172	**164**
		1.75	**0.6768**	**0.6768**	0.6745	1.2656	0.6745	1.2656	**108**	149
		2.00	**0.6771**	**0.6771**	0.6746	1.2500	0.6746	1.2500	**117**	123
		2.25	**0.6774**	**0.6774**	0.6746	1.2343	0.6746	1.2343	**115**	149
	9	0.50	**0.6749**	**0.6749**	0.6740	1.8593	0.6740	1.8593	114	**55**
		0.75	**0.6753**	**0.6753**	0.6740	1.7500	0.6740	1.7500	115	**107**
		1.00	**0.6758**	**0.6758**	0.6741	1.7187	0.6745	1.2812	**76**	353
		1.25	**0.6762**	**0.6761**	0.6741	1.6718	0.6745	1.2656	**95**	256
		1.50	0.6766	**0.6764**	0.6741	1.6718	0.6745	1.2656	**97**	164
		1.75	**0.6768**	**0.6768**	0.6745	1.2656	0.6745	1.2656	**138**	149
		2.00	**0.6771**	**0.6771**	0.6746	1.2500	0.6746	1.2500	**119**	123
		2.25	**0.6774**	**0.6774**	0.6746	1.2343	0.6746	1.2343	**148**	149
96	4	0.50	**0.6711**	**0.6711**	0.6702	1.8645	0.6702	1.8854	264	**234**
		0.75	**0.6716**	**0.6716**	0.6702	1.7604	0.6702	1.7604	**503**	980
		1.00	**0.6720**	**0.6720**	0.6703	1.7395	0.6707	1.2708	**612**	9151
		1.25	**0.6723**	**0.6723**	0.6707	1.2604	0.6707	1.2604	**49990**	66740
		1.50	**0.6726**	**0.6726**	0.6707	1.2604	0.6707	1.2604	10807	**5628**
		1.75	**0.6729**	**0.6729**	0.6707	1.2395	0.6707	1.2500	3894	**3733**
		2.00	**0.6732**	**0.6732**	0.6707	1.2395	0.6707	1.2395	1942	**1614**
		2.25	**0.6735**	**0.6735**	0.6707	1.2395	0.6707	1.2395	**1050**	1228
	9	0.50	**0.6711**	**0.6711**	0.6702	1.8645	0.6702	1.8854	323	**234**
		0.75	**0.6716**	**0.6716**	0.6702	1.7604	0.6702	1.7604	**434**	980
		1.00	**0.6720**	**0.6720**	0.6703	1.7395	0.6707	1.2708	**362**	9151
		1.25	**0.6723**	**0.6723**	0.6707	1.2604	0.6707	1.2604	**9017**	66740
		1.50	0.6728	**0.6726**	0.6703	1.6875	0.6707	1.2604	**489**	5628
		1.75	0.6733	**0.6729**	0.6704	1.6562	0.6707	1.2500	**434**	3733
		2.00	**0.6732**	**0.6732**	0.6707	1.2395	0.6707	1.2395	**820**	1614
		2.25	**0.6735**	**0.6735**	0.6707	1.2395	0.6707	1.2395	**586**	1228
128	4	1.00	**0.6738**	**0.6738**	0.6725	1.2734	0.6725	1.2734	393184	**134313**
	9	1.00	**0.6738**	**0.6738**	0.6721	1.7344	0.6725	1.2734	**1053**	134313

imaging, Journal of Mathematical Imaging and Vision, 40 (2010), pp. 120–145.

[5] A. Chambolle and T. Pock, *Approximating the total variation with finite differences or finite elements*, in Handbook of Numerical Analysis, vol. 22, Elsevier, 2021, pp. 383–417.

[6] T. F. Chan and T. P. Mathew, *Domain decomposition algorithms*, Acta Numerica, 3 (1994), pp. 61–143.

[7] K. Chen and X.-C. Tai, *A nonlinear multigrid method for total variation minimization from image restoration*, Journal of Scientific Computation, 33 (2007), pp. 115–138.

[8] C. Clason, F. Kruse, and K. Kunisch, *Total variation regularization of multi-material topology optimization*, ESAIM: Mathematical Modelling and Numerical Analysis, 52 (2018), pp. 275–303.

[9] C. Clason, C. Tameling, and B. Wirth, *Convex relaxation of discrete vector-valued optimization problems*, SIAM Review, 63 (2021), pp. 783–821.

[10] G. Cristinelli, J. A. Iglesias, and D. Walter, *Conditional gradients for total variation regularization with PDE constraints: A graph cuts approach*, arXiv preprint arXiv:2310.19777, (2023).

[11] A. De Marchi, *On the mixed-integer linear-quadratic optimal control with switching cost*, IEEE Control Systems Letters, 3 (2019), pp. 990–995.

[12] V. Dolean, P. Jolivet, and F. Nataf, *An introduction to domain decomposition methods: algorithms, theory, and parallel implementation*, SIAM, 2015.

[13] Gurobi Optimization, LLC, *Gurobi Optimizer Reference Manual*, 2023, https://www.gurobi.com.

[14] F. M. Hante and S. Sager, *Relaxation methods for mixed-integer optimal control of partial differential equations*, Computational Optimization and Applications, 55 (2013), pp. 197–225.

[15] J. Haslinger and R. A. Mäkinen, *On a topology optimization problem governed by two-dimensional Helmholtz equation*, Computational Optimization and Applications, 62 (2015), pp. 517–544.

Table 6.3: Mean and median solution times (seconds) of Gurobi for (TRP) instances per patch for $N_\mathrm{p} = 4$ and $\alpha = 1.00 \cdot 10^{-3}$ over the course of Algorithm 4.1.

N		1	2	3	4
64	mean	4.50	0.07	0.12	0.09
	median	3.82	0.06	0.08	0.10
96	mean	6.77	0.20	0.15	0.23
	median	4.86	0.19	0.14	0.22
128	mean	3477.81	0.40	0.30	0.31
	median	181.49	0.39	0.25	0.27

Table 6.4: Mean and median solution times (seconds) of Gurobi for (TRP) instances per patch for $N_\mathrm{p} = 9$ and $\alpha = 1.00 \cdot 10^{-3}$ over the course of Algorithm 4.1.

N		1	2	3	4	5	6	7	8	9
64	mean	0.39	0.05	0.04	0.06	0.03	0.04	0.10	0.04	0.05
	median	0.21	0.05	0.04	0.05	0.02	0.04	0.05	0.04	0.04
96	mean	1.35	0.10	0.04	0.12	0.06	0.08	0.16	0.25	0.07
	median	0.36	0.09	0.02	0.08	0.06	0.08	0.07	0.07	0.07
128	mean	5.40	0.21	0.07	0.24	0.19	0.15	0.13	0.03	0.10
	median	2.11	0.19	0.05	0.27	0.19	0.16	0.01	0.02	0.10

[16] M. Herrmann, R. Herzog, S. Schmidt, J. Vidal-Núñez, and G. Wachsmuth, *Discrete total variation with finite elements and applications to imaging*, Journal of Mathematical Imaging and Vision, 61 (2019), pp. 411–431.

[17] M. Hintermüller and K. Kunisch, *Total bounded variation regularization as a bilaterally constrained optimization problem*, SIAM Journal of Applied Mathematics, 64 (2004), pp. 1311–1333.

[18] M. Hintermüller and A. Langer, *Non-overlapping domain decomposition methods for dual total variation based image denoising*, Journal of Scientific Computing, 62 (2015), pp. 456 – 481.

[19] C. Y. Kaya, *Optimal control of the double integrator with minimum total variation*, Journal of Optimization Theory and Applications, 185 (2020), pp. 966–981.

[20] C. Kirches, P. Manns, and S. Ulbrich, *Compactness and convergence rates in the combinatorial integral approximation decomposition*, Mathematical Programming, 188 (2021), pp. 569–598.

[21] J. E. Lagnese and G. Leugering, *Domain decomposition methods in optimal control of partial differential equations*, vol. 148, Birkhäuser, 2012.

[22] A. Langer, *Domain decomposition for non-smooth (in particular TV) minimization*, in Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision, K. Chen, C.-B. Schönlieb, X.-C. Tai, and L. Younes, eds., Springer International Publishing, Cham, 2023, pp. 379–425.

[23] A. Langer and F. Gaspoz, *Overlapping domain decomposition methods for total variation denoising*, SIAM Journal of Numerical Analysis, 57 (2019), pp. 1411–1444.

[24] C.-O. Lee, C. Nam, and J. Park, *Domain decomposition methods using dual conversion for the total variation minimization with l^1 fidelity term*, Journal of Scientific Computing, 78 (2019), pp. 951 – 970.

[25] C.-O. Lee, E.-H. Park, and J. Park, *A finite element approach for the dual rudin–osher–fatemi model and its nonoverlapping domain decomposition methods*, SIAM Journal of Scientific Computation, 41 (2019), pp. B205 – B228.

[26] C.-O. Lee and J. Park, *Fast nonoverlapping block Jacobi method for the dual Rudin–Osher–Fatemi model*, SIAM Journal of Imaging Science, 12 (2019), pp. 2009–2034.

[27] C.-O. Lee and J. Park, *Recent advances in domain decomposition methods for total variation minimization*, Journal of the Korean Society for Industrial and Applied Mathematics, 24 (2020), pp. 161–197.

[28] S. Leyffer and P. Manns, *Sequential linear integer programming for integer optimal control with total variation regularization*, ESAIM: Control, Optimisation and Calculus of Variations, 28 (2022).

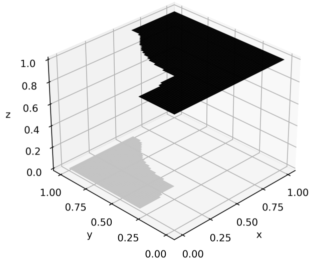
[29] S. Leyffer, P. Manns, and M. Winckler, *Convergence of sum-up rounding schemes for cloaking problems governed by the Helmholtz equation*, Computational Optimization and Applications, 79 (2021), pp. 193–221.

[30] F. Maggi, *Sets of Finite Perimeter and Geometric Variational Problems: an Introduction to Geometric Measure Theory*, no. 135, Cambridge University Press, 2012.

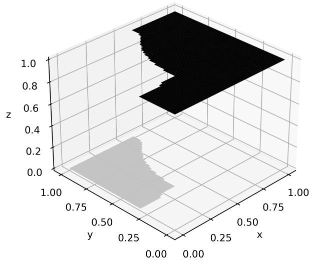
[31] P. Manns and V. Nikolić, *Homotopy trust-region method for phase-field approximations in perimeter-regularized binary optimal control*, ESAIM: Control, Optimisation and Calculus of Variations, 31 (2025).

[32] P. Manns and A. Schiemann, *On integer optimal control with total variation regularization on multidimensional domains*, SIAM Journal on Control and Optimization, 61 (2023), pp. 3415–3441.

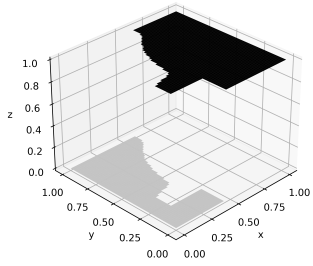
[33] J. Marko and G. Wachsmuth, *Integer optimal control problems with total variation regularization: Optimal-*



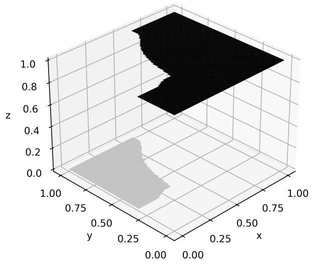
(a) $N = 64$, $N_{\mathrm{p}} = 1$



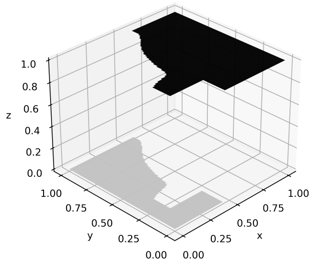
(b) $N = 64$, $N_{\mathrm{p}} = 4$



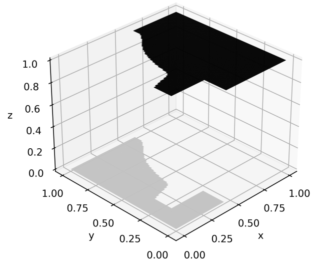
(c) $N = 64$, $N_{\mathrm{p}} = 9$



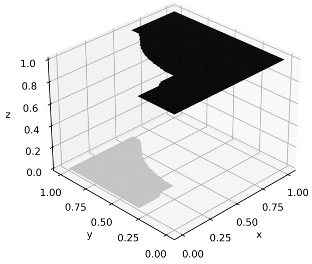
(d) $N = 96$, $N_{\mathrm{p}} = 1$



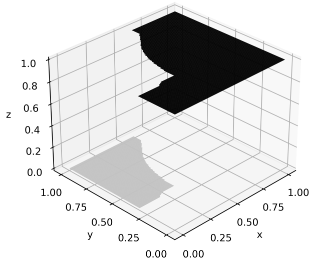
(e) $N = 96$, $N_{\mathrm{p}} = 4$



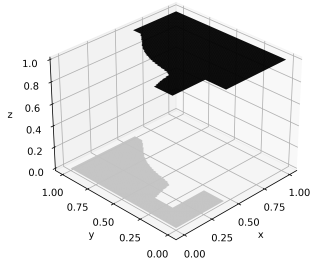
(f) $N = 96$, $N_{\mathrm{p}} = 9$



(g) $N = 128$, $N_{\mathrm{p}} = 1$



(h) $N = 128$, $N_{\mathrm{p}} = 4$



(i) $N = 128$, $N_{\mathrm{p}} = 9$

Fig. 6.2: Visualization of the controls produced by Algorithm 4.1 for $\alpha = 10^{-3}$, $N \in \{64, 96, 128\}$, and $N_{\mathrm{p}} \in \{1, 4, 9\}$.

ity conditions and fast solution of subproblems, ESAIM: Control, Optimisation and Calculus of Variations, 29 (2023).

[34] T. P. Mathew, *Domain decomposition methods for the numerical solution of partial differential equations*, Springer, 2008.

[35] M. J. Powell, *On search directions for minimization algorithms*, Mathematical Programming, 4 (1973), pp. 193–201.

[36] S. Sager, *Sampling decisions in optimum experimental design in the light of Pontryagin's maximum principle*, SIAM Journal on Control and Optimization, 51 (2013), pp. 3181–3207.

[37] S. Sager and C. Zeile, *On mixed-integer optimal control with constrained total variation of the integer control*, Computational Optimization and Applications, 78 (2021), pp. 575–623.

[38] A. Schiemann, *Integer optimization with total variation regularization*, PhD thesis, Technische Universität Dortmund, 2024.

[39] M. Severitt and P. Manns, *Efficient solution of discrete subproblems arising in integer optimal control with total variation regularization*, INFORMS Journal on Computing, 35 (2023), pp. 869–885.

[40] O. Sigmund and K. Maute, *Topology optimization approaches: A comparative review*, Structural and multidisciplinary optimization, 48 (2013), pp. 1031–1055.

[41] P. Tseng and S. Yun, *A coordinate gradient descent method for nonsmooth separable minimization*, Mathe-

matical Programming, 117 (2009), pp. 387–423.
[42] S. J. Wright, *Coordinate descent algorithms*, Mathematical Programming, 151 (2015), pp. 3–34.
[43] J. Yu and M. Anitescu, *Multidimensional sum-up rounding for integer programming in optimal experimental design*, Mathematical Programming, 185 (2021), pp. 37–76.

Appendix A. Auxiliary Results.

Lemma A.1. *Let A, B, C be sets of finite perimeter in Ω, $\mathcal{H}^{d-1}(C\cap\partial^* A)=0$, and $\mathcal{H}^{d-1}(C\cap\partial^* B)=0$. Then*

$$\mathcal{H}^{d-1}\big(C\cap((A^{(1)}\cap B^{(0)})\cup(A^{(0)}\cap B^{(1)}))\big)=\mathcal{H}^{d-1}\big(C\cap(A^{(1)}\vartriangle B^{(1)})\big).$$

Proof. Because A and B are sets of finite perimeter, we have

$$\mathcal{H}^{d-1}\left(\Omega\setminus\left(A^{(0)}\cup A^{(1)}\cup\partial^* A\right)\right)=0 \text{ and } \mathcal{H}^{d-1}\left(\Omega\setminus\left(B^{(0)}\cup B^{(1)}\cup\partial^* B\right)\right)=0.$$

Consequently, we obtain

$$\mathcal{H}^{d-1}\left(C\cap\left(\left(A^{(1)}\cap B^{(0)}\right)\cup\left(A^{(0)}\cap B^{(1)}\right)\right)\right)=\mathcal{H}^{d-1}\left(C\cap\left(\left(A^{(1)}\setminus B^{(1)}\right)\cup\left(B^{(1)}\setminus A^{(1)}\right)\right)\right).$$

∎

Lemma A.2. *Let A, B, C, F be sets of finite perimeter in Ω such that*

$$\begin{aligned}&\mathcal{H}^{d-1}(\partial^* A\cap\partial^* B)=0,\\&\mathcal{H}^{d-1}(\partial^* A\cap\partial^* C)=0, \text{ and}\\&F=(C\cap A)\cup(B\setminus A).\end{aligned}$$

Then

$$D\chi_F=D\chi_B\llcorner(\Omega\setminus\overline{A})+D\chi_C\llcorner A+D\chi_A\llcorner(C^{(1)}\cap B^{(0)})-D\chi_A\llcorner(C^{(0)}\cap B^{(1)}).$$

Proof. We first observe that sets of finite perimeter in Ω are also sets of finite perimeter in $\mathbb{R}^d$. To distinguish the distributional derivative of a set of finite perimeter G in $\mathbb{R}^d$ from the one when interpreting it as a set of finite perimeter in Ω, we denote the distributional derivative of the former by μ_G as in [30] and in contrast to $D\chi_G$ for the latter as introduced in Section 2.

We apply Theorem 16.16 from [30] and obtain from (16.35) in its proof that

$$\mu_F=\mu_B\llcorner(\mathbb{R}^d\setminus\overline{A})+\mu_C\llcorner A+\mu_A\llcorner(C^{(1)}\cap B^{(0)})-\mu_A\llcorner(C^{(0)}\cap B^{(1)}).$$

holds when interpreting $D\chi_G$ as the distributional derivative of a set of finite perimeter in $\mathbb{R}^d$. Restricting the measures on both sides to Ω, we obtain

$$\mu_F\llcorner\Omega=\mu_B\llcorner(\Omega\setminus\overline{A})+\mu_C\llcorner A+\mu_A\llcorner(C^{(1)}\cap B^{(0)})-\mu_A\llcorner(C^{(0)}\cap B^{(1)})$$

because A, $C^{(1)}\cap B^{(0)}$, $C^{(0)}\cap B^{(1)}\subset\Omega$ already. Since the distributional derivatives coincide in Ω, we obtain $D\chi_F=\mu_F\llcorner\Omega$ and similar for the other terms so that the claim follows. ∎