

# ON CONVERGENCE OF BINARY TRUST-REGION STEEPEST DESCENT

Paul Manns\*    Mirko Hahn†    Christian Kirches‡    Sven Leyffer§  
Sebastian Sager†

**Abstract** Binary trust-region steepest descent (BTR) and combinatorial integral approximation (CIA) are two recently investigated approaches for the solution of optimization problems with distributed binary-/discrete-valued variables (control functions). We show improved convergence results for BTR by imposing a compactness assumption that is similar to the convergence theory of CIA. As a corollary we conclude that BTR also constitutes a descent algorithm on the continuous relaxation and its iterates converge weakly-\* to stationary points of the latter. We provide computational results that validate our findings. In addition, we observe a regularizing effect of BTR, which we explore by means of a hybridization of CIA and BTR.

**Keywords:** mixed-integer optimal control, trust-region methods, relaxation-based methods

*MSC (2020):* 49J45, 49M05, 90C30

## 1 INTRODUCTION

For bounded domains  $\Omega \subseteq \mathbb{R}^d$  we are interested in optimization problems of the form

$$(P) \quad \inf_x J(x) \quad \text{s.t.} \quad x(s) \in \{0, 1\} \text{ for almost all (a.a.) } s \in \Omega \text{ and } x \in L^2(\Omega),$$

$J$  is a map from  $L^2(\Omega)$  to  $\mathbb{R}$ . For this problem class we study solutions of corresponding continuous relaxations of the form

$$(R) \quad \min_x J(x) \quad \text{s.t.} \quad x(s) \in [0, 1] \text{ for a.a. } s \in \Omega \text{ and } x \in L^2(\Omega)$$

and their relation to problem (P). The inf in the formulation of (P) and the min in the formulation of (R) are deliberately chosen to highlight that problem (R) (in contrast to (P)) admits a minimizer under mild assumptions [20, 28]. We restrict ourselves to this setting in the interest of a concise presentation. One can, however, extend our analysis in different directions, for example, to the case

---

This work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Scientific Discovery through the Advanced Computing (SciDAC) Program through the FASTMath Institute under Contract No. DE-AC02-06CH11357 and from the German Research Foundation under GRK 2297 MathCoRe (project No. 314838170) and SPP 1962 (projects No. SA 2016/1-2, KI 1839/1-2) and from the German Federal Ministry of Education and Research within the program “Mathematics for Innovations” under the project “Power to Chemicals.” Argonne National Laboratory Preprint ANL/MCS-P9652-0222.

\*Faculty of Mathematics, TU Dortmund University ([paul.manns@tu-dortmund.de](mailto:paul.manns@tu-dortmund.de)).

†Faculty of Mathematics, Otto von Guericke University Magdeburg.

‡Carl-Friedrich-Gauß-Faculty, Technical University of Braunschweig.

§Mathematics and Computer Science Division, Argonne National Laboratory.

where an  $L^1$ -regularization term is added to the objective function,  $J(x)$ , by following the arguments in [26].

A rich class of instances of (P) are mixed-integer PDE-constrained optimization problems, where  $J = j \circ S$ , where  $j$  is the objective of the optimization and  $S$  the control-to-state operator of an underlying partial differential equation (PDE). Such problems arise in many different areas such as topology optimization [23, 15], optimum experimental design [37], and gas network optimization [12, 13].

We build on recent work on two algorithmic solution approaches of (P). Specifically we use insights of the available analysis of combinatorial integral approximation (CIA) to improve the known convergence results for *binary trust-region steepest descent* (BTR). We note that there are more methods for (approximately) solving problems of the form (P). For example, an established method for topology optimization is the SIMP method [2], which employs a non-convex penalization and thus regularization of the controls (designs), and postprocessing of the solution of (R) by means of lumping or filtering techniques.

**Combinatorial integral approximation** [14, 20, 31, 32] The idea that underlies CIA is to split the solution process of (P) into solving the continuous relaxation (R) and then computing a  $\{0, 1\}$ -valued approximation of the relaxed solution. The approximation process can be analyzed in the weak-\* topology of  $L^\infty(\Omega)$  [27]. Relying on compactness properties of an underlying control-to-state operator (e.g.,  $J = j \circ S$ , where  $j \in C(L^2(\Omega), \mathbb{R})$  and  $S : L^2(\Omega) \rightarrow L^2(\Omega)$  is a compact operator), a tight approximation of the optimal objective value by the resulting approximants can be proved [20]. If stationary points are computed in the first step, the approximation properties generalize accordingly. We note that CIA can handle problem formulations, where the constraint  $x(s) \in \{0, 1\}$  is generalized to  $x(s) \in V$ , where  $V \subset \mathbb{R}$  is a finite set, by means of so-called special ordered set of type 1 (SOS1) reformulation of  $V$  so that its elements become the vertices of a unit simplex in  $\mathbb{R}^{|V|}$ .

**Binary trust-region steepest descent** [10, 33, 36] The BTR method solves trust-region subproblems in which the level sets of the control  $x$ , corresponding to the values 0 and 1, are manipulated to greedily improve the linearized objective. A trust-region constraint limits the volume of the level sets, which is the  $L^1$ -norm of the control function, and can change from one accepted iterate to the next. The analysis in [10] shows that BTR iterates eventually satisfy a condition called  $\varepsilon$ -stationarity under a regularity assumption used to obtain sufficient decrease of the aggregated volume of level set manipulations. Regarding the problem formulation (P), we note that generalizations of the BTR method to the constraint  $x(s) \in \{0, 1\}^k$  for  $k \in \mathbb{N}$  are conceivable but have not been considered in the literature so far.

The concept  $\varepsilon$ -stationarity as introduced in [10] measures the projected gradient of the objective and also provides, as we will show, a criticality measure for first-order necessary optimality conditions for (R). Moreover, similar to CIA, the structural assumptions on the quantities that appear in (P) are in general not able to prevent a fine microstructure from developing over the iterations. In fact, the weak-\* closure of the feasible set of (P) in  $L^\infty(\Omega)$  is the feasible set of its relaxation (R) [24, 25]. However, it is not known whether the iterates generated by BTR converge to a limit point that satisfies a first-order optimality condition of (R), in other words, if the termination tolerance of BTR is driven to zero. This lack of a convergence result is in contrast to CIA, which has stationary limits under a suitable compactness assumption.

**Standing assumptions** We provide our comparison of CIA and BTR and prove the new convergence result under the following set of assumptions, which will be discussed in detail in the remainder.

**Assumption 1.1.**

- (a) Let  $J : L^2(\Omega) \rightarrow \mathbb{R}$  be bounded from below.
- (b) Let  $J : L^1(\Omega) \rightarrow \mathbb{R}$  be Fréchet differentiable.
- (c) Let  $\nabla J : L^1(\Omega) \rightarrow L^2(\Omega)$  be Lipschitz continuous.
- (d) Let  $\nabla J : L^2(\Omega) \rightarrow L^2(\Omega)$  be weak-norm continuous (completely continuous).

**Remark 1.1.** We note that the domain of  $J$  can be restricted to  $L^\infty(\Omega)$  or the feasible set of (R) for all considerations in this work. However, we require the continuity (differentiability) properties with respect to the  $L^1(\Omega)$ -norm on the domain space in (b) and (d) in the remainder. We note that assuming continuity with respect to the codomain in  $L^2(\Omega)$  is well defined for our purpose because all feasible points and iterates are also  $L^\infty$ -functions.

**Contributions** We close the aforementioned theoretical gap between BTR and CIA. In particular, we use the compactness Assumption 1.1, (d) on the derivative of  $J$ , and we show that the BTR iterates produced by Algorithm 2 in [10] converge weakly- $*$  in  $L^\infty(\Omega)$  to a point that is feasible and satisfies a first-order optimality condition for the continuous relaxation (R). We perform several computational experiments on an example problem that is governed by an elliptic PDE to validate our theoretical findings: specifically, BTR validates the near-optimality of the solution produced by CIA.

We have observed that BTR tends to produce controls whose level sets have shorter interface lengths between them in practice when started from zero or a thresholded control. At the same time it is able to produce objective values of similar quality as CIA. While we cannot prove guarantees on this behavior, it motivates us to explore a hybrid method, where we apply CIA but use a coarser control mesh in order to compute the binary-valued approximation of the continuous relaxation. Then we start BTR from there, which allows us to combine the bounds and efficient running time behavior obtained with the CIA method while capitalizing on the regularization effect of BTR.

**Structure of the paper** In §2 we formally introduce the CIA method and show that its underlying approximation results hold under Assumption 1.1. In §3 we formally introduce and describe the BTR algorithm. In §4 we relate it to [10] and state our main convergence result. The proof is presented in §5. We provide a computational validation of our findings, demonstrate the aforementioned regularization effect, and investigate the observed regularization effects with a hybrid method in §6. We provide auxiliary results in §A and provide a brief discussion of Assumption 1.1 with respect to the assumptions imposed in the earlier work [10] in §B.

**Notation** Let  $d \in \mathbb{N}$  denote a dimension. For a measurable set  $A \subset \Omega$ ,  $\lambda(A)$  denotes the Borel–Lebesgue measure of  $A$  in  $\mathbb{R}^d$ . The function  $\chi_A$  denotes the  $\{0, 1\}$ -valued characteristic function of the set  $A$ . Let  $\mathcal{B}$  denote the Borel  $\sigma$ -algebra on  $\Omega$ . For a set  $A \subset \Omega$ , the set  $A^c$  denotes its complement in  $\Omega$ . For sets  $A, B \subset \Omega$ , the expression  $A \Delta B$  denotes the symmetric difference between  $A$  and  $B$ , that is,  $A \Delta B := (A \cup B) \setminus (A \cap B)$ . The inner product of the Hilbert space  $L^2(\Omega)$  is denoted by  $(\cdot, \cdot)_{L^2}$ . For a space  $X$  and its topological dual  $X^*$  we denote the pairing that puts  $X$  and  $X^*$  in duality by  $\langle \cdot, \cdot \rangle_{X^*, X}$ . We denote weak convergence with the arrow  $\rightharpoonup$  and weak- $*$  convergence with the arrow  $\rightharpoonup^*$ .

## 2 COMBINATORIAL INTEGRAL APPROXIMATION

CIA decomposes the solution process of (P) into two steps. First, the continuous relaxation (R) is solved (approximately) and then the result is used to compute a sequence of  $\{0, 1\}$ -valued functions that are feasible for (P) and converge to the computed solution (or stationary point) of (R) in the weak- $*$

topology of  $L^\infty(\Omega)$ . The CIA algorithm is given in [Algorithm 1](#). Its key ingredients and asymptotics are described below.

---

**Algorithm 1** CIA algorithm to optimize (P) and (R).

---

**Input:**  $J : L^2(\Omega) \rightarrow \mathbb{R}, \nabla J : L^2(\Omega) \rightarrow L^2(\Omega)$ .

**Input:** Order-conserving domain dissection  $(\mathcal{S}^n)_n \subset 2^{\mathcal{B}(\Omega)}$  (see [Definition A.1](#)).

```

1:  $y \leftarrow$  (Approximately) compute a stationary point of (R).
2: for  $n = 0, 1, 2, \dots$  do
3:    $x^n \leftarrow \text{Round}(y, \mathcal{S}^n)$ .
4: end for

```

---

**Inputs of Algorithm 1** [Algorithm 1](#) generally requires the objective function  $J$  as well as its gradient  $\nabla J$  as inputs in order to solve (R) in Line 1. We note that depending on the properties of (R) and the chosen algorithm, this may be relaxed or strengthened and subgradients of  $J$  may suffice (e.g., for projected subgradient methods) or Hessian evaluations (e.g., for a semi-smooth Newton’s method) may be desirable.

The second step of CIA, that is the for-loop in [Algorithm 1](#), requires a sequence of grids  $(\mathcal{S}^n)_n$  that decompose the domain  $\Omega$ . These grids need to abide a certain regularity that is defined in [Definition A.1](#) in order to obtain the aforementioned weak- $*$  convergence, which is explained in more detail below.

**The subroutine Round and order-conserving domain dissections** The subroutine Round in Line 3 takes an  $L^\infty(\Omega)$ -function  $y$  that is  $[0, 1]$ -valued and a partition  $\mathcal{S}^n$  of  $\Omega$  as inputs and computes a  $\{0, 1\}$ -valued function  $x^n$  from them.

In the literature on CIA, (P) is usually formulated with finite sets  $V \subset \mathbb{R}^m, m \in \mathbb{N}$ , instead of  $\{0, 1\}$  as the co-domain of the optimization variables in (P) and the output of Line 1 is a convex coefficient function  $\alpha : \Omega \rightarrow [0, 1]^M$  that satisfies  $\sum_{i=1}^m \alpha_i(s) = 1$  a.e. The rounding algorithm then transforms  $\alpha$  to a function  $\omega^n : \Omega \rightarrow \{0, 1\}^M$  such that exactly one entry of  $\omega^n(s)$  is one and all others are zero a.e. To relate this to our setting, we can simply choose  $m = 2, \alpha = (y, 1 - y)^T$  and recover  $x^n$  as  $x^n = \omega_1^n$ .

Under the assumption that the sequence of partitions is an order-conserving domain dissection and with a suitable implementation of Round that satisfies the prerequisites of [\[20, Proposition 3.5\]](#), one obtains

$$\alpha \rightharpoonup^* \omega \text{ in } L^\infty(\Omega),$$

which directly implies

$$(2.1) \quad x^n \rightharpoonup^* y \text{ in } L^\infty(\Omega).$$

Admissible choices for the subroutine Round are, for example, sum-up rounding (SUR) [\[30, 28\]](#), next-forced rounding [\[17\]](#), and the combinatorial optimization-based algorithms in [\[4, 18, 38\]](#). The key property of order-conserving domain dissections is that during the refinement of the grid from one iteration to the next, a spatial coherence property and a regular shrinkage property that allow to leverage Lebesgue’s differentiation theorem, see the analysis in [\[28\]](#). A formal definition is given in [Definition A.1](#) in [§A.1](#). The choices for the subroutine Round that are used in our computational experiments are described in [§A.1](#).

**Asymptotics of Algorithm 2 under Assumption 1.1** Which set of assumptions is necessary so that solution algorithms for (R) produce sequences with (weak) cluster points that are stationary for (R) in Line 1 depends on the properties of (R) and the desired algorithm. If (R) is convex, few assumptions

may suffice and Assumptions 1.1, (b), (c), (d) may be relaxed to a boundedness assumption of the  $\varepsilon$ -subdifferential on bounded sets and the projected (sub)gradient method will work [1]. If  $J$  is not convex, a projected gradient method with standard line search techniques, for example, Armijo linesearch, yields convergence to stationary points under Assumption 1.1 [9]. In this case, a metricization of the domain space in (b) and (c) with the  $L^2$ -norm instead of  $L^1$ -norm and the gradient needs not to be Lipschitz but only uniformly continuous. More regularity allows to employ second-order methods like semi-smooth Newton to solve for the first-order optimality conditions of (R).

While the assumption on the grids and the choice of the Round subroutine imply (2.1), the desired convergence of the objectives

$$J(x^n) \rightarrow J(y)$$

in CIA requires that  $J : L^2(\Omega) \rightarrow \mathbb{R}$  is weakly continuous, which can often be asserted by regularity (compactness) properties of an underlying differential equation in the context of optimal control [20, 28]. This is implied by Assumptions 1.1, (b), and (d) which is shown below.

**Proposition 2.1.** *Let Assumptions 1.1, (b), and (d) hold. Then  $J : L^2(\Omega) \rightarrow \mathbb{R}$  is weakly continuous.*

*Proof.* Let  $x^n, x \in L^2(\Omega)$  be such that  $x^n \rightharpoonup x$ , meaning  $x^n$  converging weakly to  $x$ , in  $L^2(\Omega)$ . We need to show  $J(x^n) \rightarrow J(x)$ . Assumption 1.1, (b) and the mean value theorem imply that  $J(x^n) - J(x) = (\nabla J(\xi^n), x^n - x)_{L^2}$  for some  $\xi^n \in L^2(\Omega)$  in the line segment between  $x^n$  and  $x$  for all  $n \in \mathbb{N}$ . Because  $(\xi^n)_n$  is bounded, there exists a weakly convergent subsequence  $\xi^{n_k} \rightharpoonup \xi$  for some  $\xi \in L^2(\Omega)$ . Assumption 1.1, (d) implies  $\nabla J(\xi^{n_k}) \rightarrow \nabla J(\xi)$  in  $L^2(\Omega)$ , and consequently  $J(x^{n_k}) - J(x) = (\nabla J(\xi^{n_k}), x^{n_k} - x)_{L^2} \rightarrow 0$ . Passing to subsequences proves the claim.  $\square$

**Remark 2.2.** We note that assumption of an order-conserving domain dissection and the weak continuity of  $J$  are sufficient to obtain the desired weak-\* convergence of the  $x^n$  to stationary points and the corresponding convergence objective values as well if one does not compute  $y$  first and then executes Round but instead executes Round on the iterates of produced by an optimization algorithm for (R), see Theorem 4.7 in [28].

### 3 BINARY TRUST-REGION STEEPEST DESCENT

The BTR algorithm operates on characteristic functions induced by measurable sets. We introduce the inputs and the trust-region subproblem. Then we describe the iterations of Algorithm 2 step by step. We relate the quantities in our variant of the algorithm to the one introduced as Algorithm 2 in [10], which purely takes the point of view of measurable sets.

The BTR algorithm is given as Algorithm 2 and is a special case of [10, Algorithm 2]. In particular, it corresponds to [10, Algorithm 2] with the choices  $\mathcal{J}(A) := J(\chi_A)$  for  $A \in \mathcal{B}$ . We also choose  $\varepsilon = 0$  because we aim to study the asymptotics of the algorithm when it is not stopped early.

---

**Algorithm 2** BTR algorithm to optimize (P) and (R).

---

**Input:**  $J : L^2(\Omega) \rightarrow \mathbb{R}, \nabla J : L^2(\Omega) \rightarrow L^2(\Omega), \Delta_{\max} \in (0, \lambda(\Omega)), 0 < \sigma_1 < \sigma_2 \leq 1, \omega \in (0, 1).$

**Input:**  $U^0 \in \mathcal{B}, \Delta^0 \in (0, \Delta_{\max})$

```

1: for  $n = 0, 1, 2, \dots$  do
2:    $g^n \leftarrow \nabla J(\chi_{U^n})(\chi_{(U^n)^c} - \chi_{U^n})$ 
3:    $D^n \leftarrow \text{FindStep}(g^n, \Delta^n, \min\{\omega, 0.5\} \|\min\{g^n, 0\}\|_{L^1} / \lambda(\Omega), \Delta^n)$  // e.g. [10, Proc. 1]
4:   if  $J(\chi_{U^n \Delta D^n}) - J(\chi_{U^n}) \leq \sigma_1 (\nabla J(\chi_{U^n}), \chi_{D^n \setminus U^n} - \chi_{U^n \cap D^n})_{L^2}$  then
5:      $U^{n+1} \leftarrow U^n \Delta D^n$ 
6:     if  $J(\chi_{U^n \Delta D^n}) - J(\chi_{U^n}) \leq \sigma_2 (\nabla J(\chi_{U^n}), \chi_{D^n \setminus U^n} - \chi_{U^n \cap D^n})_{L^2}$  then
7:        $\Delta^{n+1} \leftarrow \min\{2\Delta^n, \Delta_{\max}\}$ 
8:     else
9:        $\Delta^{n+1} \leftarrow \Delta^n$ 
10:    end if
11:  else
12:     $(U^{n+1}, \Delta^{n+1}) \leftarrow (U^n, 0.5\Delta^n)$ 
13:  end if
14: end for

```

---

**Inputs of Algorithm 2** The algorithm requires the objective function  $J$  as well as its gradient  $\nabla J$  as inputs. Using the latter requires assuming differentiability of  $J$  with respect to the  $L^2$ -norm. For the acceptance criterion of the computed descent step and the update of the trust-region radius, the algorithm requires a maximal trust-region radius  $\Delta_{\max}$  and control parameters  $\sigma_1$  and  $\sigma_2$  as inputs. To compute a descent step, the algorithm uses the subroutine FindStep (see below). The subroutine also requires the parameter  $\omega$ , which ensures that the volume of the returned set is always bounded from below by a fraction of  $\Delta^n$  that is smaller than 1.

**Trust-region subproblem and subroutine FindStep** The subroutine FindStep in Line 3 of Algorithm 2 approximately solves the subproblem

$$(3.1) \quad \min_D \int_D g(s) \, ds \quad \text{s.t.} \quad \begin{cases} D \subset g^{-1}((-\infty, 0]), \\ \lambda(D) \leq \Delta. \end{cases}$$

In Algorithm 2 FindStep is called with  $g = \nabla J(\chi_{U^n})(\chi_{(U^n)^c} - \chi_{U^n})$  and  $\Delta = \Delta^n$ . Changing from set optimization to function optimization, the minimization problem (3.1) is equivalent to the minimization problem

$$(TR(\Delta)) \quad \min_d (\nabla J(\chi_U), d)_{L^2} \quad \text{s.t.} \quad \begin{cases} \chi_U(s) + d(s) \in \{0, 1\} \text{ for a.a. } s \in \Omega, \\ \|d\|_{L^1} \leq \Delta \end{cases}$$

if  $g = \nabla J(\chi_U)(\chi_{U^c} - \chi_U)$ . We provide a proof of the equivalence in Proposition A.2. We note that using min in the definitions of (3.1) and (TR( $\Delta$ )) is justified because (TR( $\Delta$ )), and thus also (3.1), indeed admits a minimizer, which is shown in Proposition A.3.

The analysis in [10] employs that, in every iteration, FindStep produces a set  $D$ , or, equivalently, a corresponding function  $d = \chi_{U \Delta D} - \chi_U$  (see also Proposition A.2), such that

$$(\nabla J(\chi_U), d)_{L^2} \leq \min \{(\nabla J(\chi_U), s)_{L^2} \mid s \text{ feasible for } (TR(\Delta))\} + \delta \Delta \quad \text{and} \\ \|d\|_{L^1} \leq \Delta$$



hold, where  $\delta = \min \{ \omega, 0.5 \| \min \{ g, 0 \} \|_{L^1} / \lambda(\Omega), \Delta \}$  is the third input of FindStep in Algorithm 2. A bisection algorithm in function space that realizes this property is also provided in [10]. Other algorithmic approaches to (approximately) solving (TR( $\Delta$ )) and thus implementing FindStep are possible as well. In §6 we describe and take advantage of a variant that exploits uniform meshes for the control discretization in our computational experiments.

**Description of the steps of Algorithm 2** The for-loop starting in Line 1 of Algorithm 2 computes candidates for improvements of the objective function that are either accepted or rejected and then updates the trust-region radius accordingly.

Line 2 computes the function  $g^n$  such that it is equal to  $-\nabla J(\chi_{U^n})$  on  $U^n$ , where  $\chi_{U^n}$  may be decreased, and such that it is equal to  $\nabla J(\chi_{U^n})$  on  $(U^n)^c$ , where  $\chi_{U^n}$  may be increased. Thus  $\{s \in \Omega \mid g^n(s) \leq 0\}$  is the set on which  $U^n$  can be changed to obtain a first-order decrease of  $J$ . Because of the use of the  $L^1$ -norm, the trust-region radius  $\Delta^n$  limits the volume of the set that can be changed in the current iteration, see Proposition A.2.

The subroutine FindStep in Line 3 computes a set  $D^n \subset (g^n)^{-1}((-\infty, 0])$  that approximates the solution of (3.1).

The candidate for improving  $J$  is then the modification of the characteristic function of the set  $U^n$ , where the values on  $D^n$  are flipped, or formally  $\chi_{U^n \Delta D^n}$ . From Line 4 onward, the for-loop resembles common trust-region methods. Line 4 determines whether the reduction achieved by  $\chi_{U^n \Delta D^n}$  is at least a fraction of the reduction predicted by the linear model, in which case the step is accepted. A second (larger) ratio is used to determine whether the trust-region radius should be increased (doubled) or left unchanged after acceptance. The trust-region radius is reduced (halved) after rejection of a candidate step.

## 4 CONVERGENCE OF BTR TO FIRST-ORDER OPTIMAL POINTS

We introduce our main result. Algorithm 2 operates on iterates that are feasible for the integer problem (P). We prove that our algorithm generates a sequence of integer feasible points whose limits are first-order optimal points of the relaxation (R). Thus, we obtain a minimizing sequence for (P), which itself may not have a solution, if (R) is a convex problem. We introduce and a criticality measure for (R) and relate it to Algorithm 2 before introducing the main theorem.

### 4.1 CRITICALITY MEASURE FOR (R)

We define the criticality measure  $C : L^2(\Omega) \rightarrow [0, \infty)$  for  $x \in L^2(\Omega)$  as

$$(4.1) \quad C(x) := \max \left\{ \int_{\Omega} \nabla J(x)(x - f) \, ds \mid f \text{ feasible for (R)} \right\} = \int_{\Omega} \nabla J(x)x + \int_{\Omega} \max\{-\nabla J(x), 0\} \, ds,$$

where the identity follows from the structure of the feasible set of (R).  $C$  coincides with the function  $\Phi$  in [9] and is also known as *primal gap function* [21]. Local minimizers are zeros of  $C$ , which is well known and repeated here for convenience. In particular, this leads to the usual definition of stationary points below.

**Proposition 4.1.** *If  $x$  is a local minimizer of (R), then  $\nabla J(x)(x - f) \leq 0$  a.e. holds for all  $f$  that are feasible for (R). Moreover,  $C(x) = 0$ .*

*Proof.* The first claim follows from a Taylor expansion at  $x$ .  $C(x) = 0$  because  $x$  is feasible in the max. □

**Definition 4.2.** A function  $x$  that is feasible for (R) is called *stationary* if  $C(x) = 0$ .

The FindStep subroutine in Algorithm 2 operates with the quantity  $\|\min\{g, 0\}\|_{L^1}$  with the choice  $g = \nabla J(\chi_U)(\chi_{U^c} - \chi_U)$  for a set  $U \in \mathcal{B}$ . We show below that  $C(\chi_U) = \|\min\{g, 0\}\|_{L^1}$ . In this case, the first-order optimality condition  $C(\chi_U) = 0$  from Proposition 4.1 corresponds the first-order optimality condition of the set-based view point in [10], see Lemma 5 and Corollary 1 therein.

**Proposition 4.3.** *Let  $U \in \mathcal{B}$ . Then for  $g = \nabla J(\chi_U)(\chi_{U^c} - \chi_U)$  it holds that  $C(\chi_U) = \|\min\{g, 0\}\|_{L^1}$ .*

*Proof.* For a.a.  $s \in \Omega$ , we obtain

$$(4.2) \quad \min\{g(s), 0\} = \begin{cases} -\nabla J(\chi_U)(s) & \text{if } s \in U \text{ and } \nabla J(\chi_U)(s) \geq 0, \\ \nabla J(\chi_U)(s) & \text{if } s \in U^c \text{ and } \nabla J(\chi_U)(s) \leq 0, \\ 0 & \text{else.} \end{cases}$$

For the choice  $x = \chi_U$  in (4.1), the right hand side implies

$$C(x) = \int_{\Omega} \left\{ \begin{array}{l} |\nabla J(x)(s)| \text{ if } (s \in U \text{ and } \nabla J(x)(s) \geq 0) \text{ or } (s \in U^c \text{ and } \nabla J(x)(s) \leq 0), \\ 0 \text{ else} \end{array} \right\} ds.$$

The claim follows by a pointwise a.e. comparison of the integrands. □

An alternative criticality measure for (R) that uses the  $L^1$ -norm is the function  $\tilde{C}$  that is defined for  $x \in L^1(\Omega)$  as

$$\tilde{C}(x) := \|x - P_{[0,1]}(x - \nabla J(x))\|_{L^1}.$$

It is known that  $\tilde{C}(x) = 0$  implies that  $x$  satisfies a first-order necessary optimality condition for (R); see, for example, [16, Lemma 1.12]. Moreover, it can easily be verified that  $C(x) = 0$  if and only if  $\tilde{C}(x) = 0$ . We do not use  $\tilde{C}$  because it would complicate our analysis and lead to a less concise presentation.

#### 4.2 MAIN RESULT

Having introduced the necessary notation, concepts, and assumptions, we now state our main convergence results.

**Theorem 4.4.** *Let Assumptions 1.1, (a) and (b) hold. Let  $(U^n)_n \subset \mathcal{B}$ ,  $(D^n)_n \subset \mathcal{B}$ , and  $(\Delta^n)_n \subset (0, \Delta_{\max}]$  be the sequences of sets and trust-region radii produced by Algorithm 2. Then the sequence of objective values  $(J(\chi_{U^n}))_n$  is monotonically nonincreasing. Moreover, one of the following mutually exclusive outcomes holds:*

1. *There exists  $n_0 \in \mathbb{N}$  such that  $U^{n_0} = U^n$  a.e. holds for all  $n \geq n_0$ . Then  $\chi_{U^{n_0}}$  is stationary for (R).*
2. *For all  $n_0 \in \mathbb{N}$  there exists  $n_1 > n_0$  such that  $\lambda(U^{n_1} \Delta U^{n_0}) > 0$ . The sequence  $(\chi_{U^n})_n \subset L^\infty(\Omega)$  admits a weak-\* accumulation point. Every weak-\* accumulation point  $f$  of  $(\chi_{U^n})_n$  is feasible for (R).*

*If additionally Assumption 1.1, (c) holds, then*

$$\lim_{n \rightarrow \infty} C(\chi_{U^n}) = 0.$$

*If additionally Assumption 1.1, (d) holds and if a subsequence  $(\chi_{U^{n_k}})_k \subset (\chi_{U^n})_n$  satisfies  $C(\chi_{U^{n_k}}) \rightarrow 0$ , then every weak-\* accumulation point  $f$  of  $(\chi_{U^{n_k}})_k$  satisfies  $C(f) = 0$ , i.e.,  $f$  is stationary for (R).*

Theorem 4.4 is proven in §5. We obtain the following corollary that shows that Algorithm 2 produces a sequence of binary iterates that converge weakly-\* to stationary points of the continuous relaxation (R) of (P). Thus BTR yields results comparable to those produced by CIA because solution algorithms for (R) cannot be expected to perform better than producing a stationary point of (R) in practice.



**Corollary 4.5.** *Let Assumption 1.1 hold. Let  $(U^n)_n \subset \mathcal{B}$  be the sequences of sets produced by Algorithm 2. Then all weak- $*$  accumulation points of  $(\chi_{U^n})_n$  are feasible and stationary for (R).*

*Proof.* The claim follows by combining the two assertions in Outcome 2 of Theorem 4.4. □

## 5 PROOF OF THE MAIN THEOREMS

In this section we prove Theorem 4.4. We first prove preparatory results on the sufficient reduction condition (Algorithm 2, Line 4) for binary-valued control functions and trust-region steps in §5.1. Then we employ these results to analyze the asymptotics of Algorithm 2 in §5.2, finishing with the proof of Theorem 4.4.

### 5.1 SUFFICIENT DECREASE WITH A CHARACTERISTIC FUNCTION

The first step of the proof is to show that if  $\chi_{U^n}$  for a given iterate  $U^n \in \mathcal{B}$  of Algorithm 2 is not stationary, then there exists a set  $D^n$  such that  $\chi_{U^n \Delta D^n}$  satisfies a sufficient decrease condition with respect to  $\nabla J(\chi_{U^n})$  for sufficiently small trust-region radii. We briefly recap and the well-known result on existence of a descent direction in Lemma 5.1, which we adapt for our case of characteristic functions as Algorithm 2 operates on. This in turn implies acceptance of a new iterate after finitely many steps as is shown in Lemma 5.2.

**Lemma 5.1.** *Let Assumption 1.1 (b) hold. Let  $\chi_U$  for  $U \in \mathcal{B}$  not be stationary for (R). Then there exist  $\varepsilon > 0$  and  $\Delta_0 > 0$  such that for all  $0 < \Delta \leq \Delta_0$ , there exists  $d \in L^1(\Omega)$  that is feasible for (TR( $\Delta$ )) and satisfies  $(\nabla J(\chi_U), d) \leq -\varepsilon\Delta$ .*

*Proof.* Because  $\chi_U$  is not stationary, we have  $C(\chi_U) > 0$ , which implies that there is  $\varepsilon > 0$  and a set  $D \subset g^{-1}((-\infty, 0])$  with  $\lambda(D) > 0$  such that  $\nabla J(\chi_U)(s)(\chi_{U^c}(s) - \chi_U(s)) < -\varepsilon$  for a.a.  $s \in D$ . Let  $\Delta_0 := \lambda(D)$  and using the regularity properties of the Lebesgue measure, there is a subset  $D_\Delta \subset D$  with  $\lambda(D_\Delta) = \Delta$  for all  $0 < \Delta \leq \Delta_0$  (note that it is possible to use the greedy construction from Proposition A.3 here too), which implies

$$\int_{D_\Delta} \nabla J(\chi_U)(\chi_{U^c} - \chi_U) \, ds < -\Delta\varepsilon.$$

Using the equivalence asserted in Proposition A.2 and in particular setting  $d := \chi_{U \Delta D_\Delta} - \chi_U$  yields the claim. □

We employ this result to prove that Algorithm 2 accepts a step after finitely many iterations if the current iterate is not stationary for (R).

**Lemma 5.2.** *Let Assumptions 1.1, (b) and (c) be satisfied. Let  $(U^n)_n \subset \mathcal{B}$ ,  $(D^n)_n \subset \mathcal{B}$ , and  $(\Delta^n)_n \subset (0, \Delta_{\max}]$  be the sequences of sets and trust-region radii produced by Algorithm 2. Let  $\chi_{U^n}$  not be stationary for (R). Then the output of Algorithm 2, Line 3, is accepted after  $k \in \mathbb{N}$  steps: specifically,  $U^n = U^{n+j}$  for all  $0 \leq j < k$  and  $J(\chi_{U^n \Delta D^{n+k}}) - J(\chi_{U^n}) \leq \sigma_1(\nabla J(\chi_{U^n}), \chi_{D^{n+k} \setminus U^n} - \chi_{U^n \cap D^n})_{L^2}$ .*

*Proof.* For  $m \in \mathbb{N}$ , we define the optimal linear predicted reduction as

$$L^m := - \inf_{D' \in \mathcal{B}} \left\{ (\nabla J(\chi_{U^m}), d')_{L^2} \mid d' = \chi_{D' \setminus U^m} - \chi_{D' \cap U^m} \text{ and } \|d'\|_{L^1} \leq \Delta^m \right\}.$$

By design of Algorithm 2, we have  $\Delta^{n+j+1} = 0.5\Delta^{n+j}$  for all  $j \geq 0$  until the step  $D^{n+j}$  is accepted.

We prove the claim by contradiction and assume that the step  $D^{n+k}$  is not accepted for all  $k \in \mathbb{N}$ . Because  $U^n$  is not stationary for (R) and  $\Delta^{n+k} \rightarrow 0$  for  $k \rightarrow \infty$ , Lemma 5.1 implies that there exist  $\varepsilon > 0$  and  $k_0 \in \mathbb{N}$  such that for all  $k \geq k_0$  the estimate  $L^{n+k} \geq \Delta^{k+j}\varepsilon$  holds.

We apply Taylor’s theorem and obtain that

$$\begin{aligned}
 & J(\chi_{U^n}) - J(\chi_{U^n \Delta D^{n+k}}) \\
 &= -(\nabla J(\chi_{U^n}), \chi_{D^{n+k} \setminus U^n} - \chi_{U^n \cap D^{n+k}})_{L^2} + o\left(\lambda(D^{n+k})\right) \\
 (5.1) \quad &\geq L^{n+k} - (\Delta^{n+k})^2 + o(\Delta^{n+k}),
 \end{aligned}$$

where the inequality follows from the construction of  $D^{n+k}$  by means of the FindStep subroutine, specifically Lemma 9 in [10] with a choice  $\delta \leq \Delta^{n+k}$ . The inequality  $\delta \leq \Delta^{n+k}$  is satisfied in Algorithm 2, Line 3, because the parameter  $\delta$  of FindStep, the third argument of the subroutine, is given a value that is less than or equal to the trust-region radius in all iterations.

Because  $L^{n+k} \geq \varepsilon \Delta^{n+k}$  holds for all  $k \geq k_0$  and the two latter terms in (5.1) are  $o(\Delta^{n+k})$  there exists  $k_1 \in \mathbb{N}$  such that for all  $k \geq k_1$  it holds that

$$J(\chi_{U^n}) - J(\chi_{U^n \Delta D^{n+k}}) \geq \sigma_1 L^{n+k}.$$

By definition of  $L^{n+k}$  it follows that

$$J(\chi_{U^n}) - J(\chi_{U^n \Delta D^{n+k}}) \geq \sigma_1 L^{n+k} \geq -\sigma_1 (\nabla J(\chi_{U^n}), \chi_{D^{n+k} \setminus U^n} - \chi_{U^n \cap D^{n+k}})_{L^2},$$

and thus the step  $D^{n+k_1}$  is accepted in Algorithm 2. This contradicts the assumption that the step  $D^{n+k}$  is not accepted for all  $k \in \mathbb{N}$ . □

## 5.2 ASYMPTOTICS OF ALGORITHM 2

Before finalizing the proof of Theorem 4.4, we show three further preparatory lemmas. Lemma 5.3 states that the sequence of iterates produced by Algorithm 2 has a corresponding sequence of monotonically nonincreasing objective values. Lemma 5.4 shows that if the criticality measure  $C$  stays bounded away from zero over the iterations of Algorithm 2, then the trust-region radius contracts to zero.

**Lemma 5.3.** *Let Assumption 1.1, (b) hold. Let  $(U^n)_n, (D^n)_n \subset \mathcal{B}$ , and  $(\Delta^n)_n \subset (0, \Delta_{\max}]$  be the sequences of sets and trust-region radii produced by Algorithm 2. Then the sequence of objective values  $(J(\chi_{U^n}))_n$  is monotonically nonincreasing.*

*Proof.* By construction of  $D^n$  with FindStep, Procedure 1 of [10], it holds that  $D^n \subset \{s \in \Omega \mid g_{U^n}(s) < 0\}$ . A step that is accepted in Algorithm 2 Line 4 satisfies  $J(\chi_{U^n \Delta D^n}) < J(\chi_{U^n})$  because

$$(\nabla J(\chi_{U^n}), \chi_{D^n \setminus U^n} - \chi_{U^n \cap D^n})_{L^2} = \int_{D^n} g^n \, ds < 0,$$

while  $J(\chi_{U^n})$  remains unchanged for rejected steps. Thus, the sequence of objective values  $(J(\chi_{U^n}))_n$  is monotonically nonincreasing. □

**Lemma 5.4.** *Let Assumptions 1.1, (a) and (b) hold. Let  $(U^n)_n \subset \mathcal{B}, (D^n)_n \subset \mathcal{B}$ , and  $(\Delta^n)_n \subset (0, \Delta_{\max}]$  be the sequences of sets and trust-region radii produced by Algorithm 2. If there exists  $\varepsilon > 0$  and  $n_0 \in \mathbb{N}$  such that  $C(\chi_{U^n}) > \varepsilon$  for all  $n \geq n_0$ , then  $\Delta^n \rightarrow 0$ .*

*Proof.* We use the notation  $L^m$  for the optimal predicted reduction in iteration  $m \in \mathbb{N}$  as in the proof of Lemma 5.2. From Proposition A.3 and the definition of  $C$  it follows that  $L^m \geq C(\chi_{U^m}) \frac{\Delta^m}{\lambda(\Omega)}$  for all iterations  $m \in \mathbb{N}$ . This can be seen by using the greedily constructed set. From Algorithm 2 Line 3,

**Proposition 4.3**, and  $\Delta_{\max} \leq \lambda(\Omega)$  it follows that the third parameter  $\delta$  of the subroutine FindStep satisfies  $\delta \leq C(\chi_{U^n})/(2\Delta^n)$ . The analysis of FindStep, specifically [10, Lemma 9], implies that

$$\begin{aligned} -(J(\chi_{U^n}), \chi_{D^n \setminus U^n} - \chi_{U^n \cap D^n})_{L^2} &\geq L^n - \min\{\omega, C(\chi_{U^n})/(2\lambda(\Omega)), \Delta^n\} \Delta^n \\ &\geq C(\chi_{U^n}) \frac{\Delta^n}{\lambda(\Omega)} - \frac{1}{2} C(\chi_{U^n}) \frac{\Delta^n}{\lambda(\Omega)} \geq \frac{\varepsilon \Delta^n}{2\lambda(\Omega)} \end{aligned}$$

for all  $n \in \mathbb{N}$  and thus all  $n \geq n_0$ .

We close the proof with a contradictory argument and assume that  $\Delta^n \not\rightarrow 0$ . We deduce that there exists an infinite subsequence  $(\Delta^{n_k})_k$  of  $(\Delta^n)_n$  such that  $\liminf_{k \rightarrow \infty} \Delta^{n_k} > \underline{\Delta}$  for some  $\underline{\Delta} > 0$ . Consequently, there exists an infinite subsequence  $(n_\ell)_\ell$  of accepted iterates with trust-region radii  $\Delta^{n_\ell} \geq \underline{\Delta}$ .

Combining these insights on the accepted iterates with the lower bound on the linearly predicted reductions, we obtain that

$$J(\chi_{U^{n_\ell}}) - J(\chi_{U^{n_\ell} \Delta D^{n_\ell}}) \geq \sigma \frac{\varepsilon \underline{\Delta}}{4\lambda(\Omega)}.$$

Because the sequence of objective values is monotonically nonincreasing by virtue of Lemma 5.3, this implies  $J(\chi_{U^{n_\ell}}) \rightarrow -\infty$ , which contradicts Assumption 1.1, (a) and thus the assumption that  $\Delta^n \not\rightarrow 0$ . □

We are now ready to finish the proof of the two main results.

*Proof of Theorem 4.4.* Lemma 5.3 proves the claim that the sequence of objective values is monotonically nonincreasing.

We first analyze Outcome 1. Because there exists  $n_0 \in \mathbb{N}$  such that  $U^{n_0} = U^n$  holds a.e. for all  $n \geq n_0$ , the acceptance criterion in Algorithm 2, Line 4, is violated for all  $n \geq n_0$ . Then the claim of Outcome 1 follows from Lemma 5.2.

If there is no  $n_0 \in \mathbb{N}$  such that  $U^{n_0} = U^n$  holds a.e. for all  $n \geq n_0$ , then Outcome 1 does not hold true, and for all  $n_0 \in \mathbb{N}$  there exists  $n_1 > n_0$  such that  $\lambda(U^{n_1} \Delta U^{n_0}) > 0$ . It follows that Outcomes 1 and 2 are mutually exclusive. Moreover, the sequence  $(\chi_{U^n})_n \subset L^\infty(\Omega)$  is bounded and thus admits a weak-\* cluster point. By virtue of, for example, [34, Theorem 3], every weak-\* cluster point of  $(\chi_{U^n})_n$  is feasible for (R).

Next we assume that Assumptions 1.1, (a), (b), and (c) are satisfied and prove the claim  $\lim_{n \rightarrow \infty} C(\chi_{U^n}) = 0$  for Outcome 2. We do so in two steps. We first prove  $\liminf_{n \rightarrow \infty} C(\chi_{U^n}) = 0$  and then improve upon this finding to  $\lim_{n \rightarrow \infty} C(\chi_{U^n}) = 0$ .

*Step 1:* We prove

$$(5.2) \quad \liminf_{n \rightarrow \infty} C(\chi_{U^n}) = 0.$$

To this end, we consider the subsequence of accepted iterates (successful steps)  $(n_k)_k$  of Algorithm 2. Using the fundamental theorem of calculus and the notation  $d^n := \chi_{U^n \Delta D^n} - \chi_{U^n}$  for  $n \in \mathbb{N}$ , we may rewrite the decrease in the objective as

$$J(\chi_{U^n}) - J(\chi_{U^n \Delta D^n}) = -(\nabla J(\chi_{U^n}), d^n)_{L^2} - \int_0^1 (\nabla J(\chi_{U^n} + td^n) - \nabla J(\chi_{U^n}), d^n)_{L^2} dt.$$

We observe that  $d^n(s) \in \{-1, 0, 1\}$  for a.a.  $s \in \Omega$  implies  $\sqrt{\|d^n\|_{L^1}} = \|d^n\|_{L^2}$ . The Lipschitz continuity of  $\nabla J : L^1(\Omega) \rightarrow L^2(\Omega)$  (Assumption 1.1, (c)) with Lipschitz constant  $L > 0$  implies

$$J(\chi_{U^n}) - J(\chi_{U^n \Delta D^n}) \geq -(\nabla J(\chi_{U^n}), d^n)_{L^2} - \frac{L}{2} \|d^n\|_{L^1} \sqrt{\|d^n\|_{L^1}}.$$

As in the proof of Lemma 5.4, we observe that the estimate

$$-(\nabla J(\chi_{U^n}), d^n) \geq C(\chi_{U^n}) \frac{\Delta^n}{2\lambda(\Omega)}$$

holds for the steps  $d^n$ . Inserting this estimate yields

$$J(\chi_{U^n}) - J(\chi_{U^{n+\Delta D^n}}) \geq -\sigma_1(\nabla J(\chi_{U^n}), d^n)_{L^2} - \underbrace{(1 - \sigma_1) \frac{C(\chi_{U^n})\Delta^n}{2\lambda(\Omega)} - \frac{L}{2}(\Delta^n)^{\frac{3}{2}}}_{:=r^n}.$$

We show (5.2) by contradiction. If  $\liminf_{n \rightarrow \infty} C(\chi_{U^n}) > 0$ , then Lemma 5.4 implies that  $\Delta^n \rightarrow 0$  and  $r^n \geq 0$  holds for all  $n \geq n_2$  for some  $n_2 \in \mathbb{N}$ . But  $r^n \geq 0$  implies that the acceptance criterion in Algorithm 2, 4 is satisfied for all  $n \geq n_2$  and the trust-region radius is not decreased further from iteration  $n_2$  on. This contradicts  $\Delta^n \rightarrow 0$  and we thus obtain (5.2).

Step 2: We prove

$$(5.3) \quad \lim_{n \rightarrow \infty} C(\chi_{U^n}) = 0.$$

To this end, we follow the proof strategy of [35, Theorem 6]. We say that  $n \in \mathbb{N}$  is a successful iteration of Algorithm 2 if the acceptance test in Line 4 is successful. Let  $\mathcal{S} \subset \mathbb{N}$  denote the set of successful iterations of Algorithm 2. We observe that any successful iteration satisfies

$$J(\chi_{U^n}) - J(\chi_{U^{n+1}}) \geq \frac{1}{2\lambda(\Omega)} C(\chi_{U^n}) \Delta^n$$

because of the properties of the subroutine FindStep (see the proof of Lemma 5.4).

This implies

$$J(\chi_{U^0}) - J(\chi_{U^{n+1}}) \geq \frac{1}{2\lambda(\Omega)} \sum_{\ell=0, \ell \in \mathcal{S}}^n C(\chi_{U^\ell}) \Delta^\ell$$

for all  $n \in \mathbb{N}$ . We seek for a contradiction to the claim and assume that there exists a subsequence  $(n_k)_k \subset \mathcal{S}$  such that

$$(5.4) \quad C(\chi_{U^{n_k}}) \geq 2\varepsilon > 0$$

for some  $\varepsilon > 0$ . Let  $K := \{n \in \mathcal{S} \mid C(\chi_{U^n}) \geq \varepsilon\}$ . It follows that

$$J(\chi_{U^0}) - J(\chi_{U^{n+1}}) \geq \frac{1}{2\lambda(\Omega)} \sum_{\ell=0, \ell \in K}^n C(\chi_{U^\ell}) \Delta^\ell \geq \frac{1}{2\lambda(\Omega)} \varepsilon \sum_{\ell=0, \ell \in K}^n \Delta^\ell$$

for all  $n \in \mathbb{N}$ . Let  $n_0 \in \mathbb{N}$ . Then we obtain for all  $n \geq n_0$  that

$$\sum_{\ell=n_0, \ell \in K}^n \Delta^\ell \leq \frac{2\lambda(\Omega)}{\varepsilon} (J(\chi_{U^0}) - J(\chi_{U^{n+1}})) \leq \frac{2\lambda(\Omega)}{\varepsilon} (J(\chi_{U^0}) - \min(\mathbf{R})) < \infty,$$

which implies that  $\sum_{\ell=n_0, \ell \in K}^\infty \Delta^\ell < \kappa < \infty$  for some  $\kappa > 0$  for all  $n_0 \in \mathbb{N}$ . From (5.2) it follows for all  $k \in \mathbb{N}$  that there exists a smallest  $\ell(k) > n_k$  with  $\ell(k) \in \mathcal{S} \setminus K$ . We obtain

$$\|\chi_{U^{\ell(k)}} - \chi_{U^{n_k}}\|_{L^1} \leq \sum_{j=n_k, j \in \mathcal{S}}^{\ell(k)-1} \|\chi_{U^{j+1}} - \chi_{U^j}\|_{L^1} \leq \sum_{j=n_k, j \in \mathcal{S}}^{\ell(k)-1} \Delta^j = \sum_{j=n_k, j \in K}^{\ell(k)-1} \Delta^j \leq \sum_{j=n_k, j \in K}^\infty \Delta^j < \kappa.$$

This implies that for  $k \rightarrow \infty$  we obtain that

$$\|\chi_{U^{\ell(k)}} - \chi_{U^{n_k}}\|_{L^1} \rightarrow 0.$$

By virtue of Fatou’s lemma and the fact that every sequence that converges in  $L^1$  has a pointwise a.e. convergent subsequence, we obtain

$$\|\chi_{U^{\ell(k)}} - \chi_{U^{n_k}}\|_{L^2} \rightarrow 0$$

for a subsequence of  $(n_k)_k$ , which we denote with the same symbol for ease of notation. We conclude that

$$|C(\chi_{U^{\ell(k)}}) - C(\chi_{U^{n_k}})| \rightarrow 0 \text{ (for } k \rightarrow \infty),$$

which violates our assumption (5.4). Hence, (5.3) follows.

For the remainder of the proof, we restrict ourselves to a weakly convergent subsequence of  $(\chi_{U^n})_n$ , for ease of notation denoted by the same symbol, which satisfies  $C(\chi_{U^n}) \rightarrow 0$  and  $\chi_{U^n} \rightharpoonup f$  in  $L^2(\Omega)$ . It remains to show that  $C(f) = 0$ , if Assumptions 1.1, (b) and (d) hold. The criticality measure  $C$  is weakly lower semi-continuous under Assumption 1.1, (d), see [9, Lemma 4.1], so that

$$0 \leq C(f) \leq \liminf_{n \rightarrow \infty} C(\chi_{U^n}) = 0.$$

□

## 6 COMPUTATIONAL EXPERIMENTS

We carry out our experiments on an instance of (P) that satisfies Assumption 1.1. The instance is described in §6.1, and our computational setup is described in §6.2. Then validation experiments and their results for the presented theory are presented in §6.3. Motivated by observations in §6.3, we explore the effects of a hybridization of BTR and CIA in §6.4.

### 6.1 EXAMPLE PROBLEM

We consider the case  $d = 2$  and the domain  $\Omega = (0, 2)^2 \subset \mathbb{R}^2$ . For (P) we choose  $J(x) := j(S(x))$ , where  $j$  is a so-called tracking-type objective, specifically  $j(y) = .5\|y - y_d\|_{L^2}^2$  for a given

$$y_d(s) = \frac{1}{4} \sin(3(s_1 - 1)(s_2 - 1))^2 (|s_1 - 1| + |s_2 - 1|)$$

for  $s \in \Omega$ .  $S$  is the solution operator of the linear elliptic boundary value problem

$$(6.1) \quad -\varepsilon \Delta y + y = x, \quad y|_{\partial\Omega} = 0$$

for a given control input  $x \in L^1(\Omega)$  and the choice  $\varepsilon = 10^{-2}$ . This yields the following instance of (P):

$$(6.2) \quad \inf_x \frac{1}{2} \|y - y_d\|_{L^2}^2 \quad \text{s.t.} \quad y = S(x) \text{ and } x(s) \in \{0, 1\} \text{ for a.a. } s \in \Omega.$$

For a Poisson problem with right-hand side in  $L^1(\Omega)$ , the weak solution  $y$  is an element of the Sobolev space of  $q$ -integrable functions with a  $q$ -integrable distributional derivative that vanish at the boundary,  $W_0^{1,q}(\Omega)$ , for bounded Lipschitz domains  $\Omega$ , where we have the estimate  $\|u\|_{W_0^{1,q}(\Omega)} \leq c\|x\|_{L^1(\Omega)}$  for some  $c > 0$  if  $q < 2$ . A proof of this result (for more general elliptic operators) and more general right-hand sides can, for example, be found in [7, Theorem 1]; and for the case of mixed boundary

condition a proof can be found in [11] (note that the required Gröger regularity of the boundary therein reduces to requiring a strong Lipschitz condition if only a Dirichlet boundary condition is present).

Combining these considerations with the chain rule for Banach spaces and the Riesz representation theorem implies that Assumptions 1.1, (a) and (c) are satisfied for this example. To see that Assumption 1.1 (d) is also satisfied, we consider the compact embedding  $W_0^{1,q}(\Omega) \hookrightarrow W^{1,1}(\Omega) \hookrightarrow^c L^2(\Omega)$  (for  $d = 2$ ), where the compactness is due to the second embedding. Because  $S'(\bar{x})$  is linear and bounded for  $\bar{x} \in L^1(\Omega)$ , it maps weakly convergent sequences to weakly convergent sequences in  $W_0^{1,q}(\Omega)$  and, by compactness, to norm convergent sequences in  $L^2(\Omega)$ , which implies that  $S'$  and in turn  $\nabla J$  are weak-norm continuous.

## 6.2 SETUP

We solve the boundary value problem (6.1) numerically using a finite element method on a conforming uniform triangle mesh subdividing the domain  $\Omega$ . Specifically, the domain is partitioned into  $256 \times 256$  square cells that are split into 4 triangles each. In all experiments, the solution  $y = S(x)$  of (6.1) is computed in the space of cellwise affine and globally continuous functions on this mesh. All experiments are carried out on a laptop computer with Intel(R) Core(TM) i9-10885H CPU (2.40 GHz) and 32 GB RAM.

**Implementation of CIA (Algorithm 1)** We compute a solution of the continuous relaxation of (6.2) (the first step of CIA) by replacing the constraint  $x(s) \in \{0, 1\}$  by  $x(s) \in [0, 1]$  and optimizing the control function on the aforementioned triangle mesh in the space of cellwise constant discontinuous functions with a quasi-Newton method.

For the Round procedure, the second step of CIA, we consider three different choices: multidimensional sum-up rounding (SUR) [28], the combinatorial optimization-based rounding (COR) proposed in [18], see also [3, § 2.4.1], and a primal heuristic for switching cost aware rounding (SHG) [4, 5]. The choices for the Round procedure compute approximating controls. In our experiments, the computed controls are cellwise constant functions on cells of the grid of  $256 \times 256$  squares. The three choices for the Round procedure are explained in §A.1. In order to ensure the approximation property (2.1), we order the grid cells of the discretization along a Hilbert curve as in [28, 20], which yields an order-conserving domain dissection as mentioned in §2, see also Definition A.1.

While SUR and COR either minimize a certain approximation error or abide by an upper bound on that error, SHG minimizes the length of the interface between the level sets for the values zero and one in our setting while abiding by the approximation error bound. The integer programming formulation of SHG is computationally intractable in our setting (for our grid size), which is why we resort to a suboptimal heuristic, see also §A.1.

**Implementation of BTR (Algorithm 2)** We compute all iterates on the fixed uniform mesh of  $256 \times 256$  squares. The uniformity, namely, the fact that all cells have the same volume, has the advantage that a discretized variant of the FindStep method can be implemented efficiently. Specifically it is a Knapsack problem with all weights being equal to one, which is therefore not NP-hard. In particular, if  $x = \chi_A$  for some  $A \in \mathcal{B}$  is the current iterate, the discretized trust-region subproblem (TR( $\Delta$ )) can be solved as follows. We compute the average value of the function  $g = \nabla J(\chi_A)(\chi_{A^c} - \chi_A)$  on each grid cell. Then we sort the cell averages of  $g$  in ascending order and pick the cells with negative average values in a greedy fashion until the current trust region is filled. The picked cells constitute a difference set  $D$  so that the computed step  $d$  is  $d = \chi_{A \Delta D} - \chi_A$ . Our discretized implementation of BTR terminates when the trust-region radius contracts below the volume of one grid cell.

We note that while we have carried out our experiments on a fixed fine mesh, an alternative approach is to adaptively refine the mesh where required within the FindStep subroutine of BTR.



### 6.3 VALIDATION

We apply CIA to (6.2), where we initialize the solver for the continuous relaxation with the constant zero function,  $x = 0$ . As mentioned above, we use SUR, COR, and SHG for the Round subroutine in Algorithm 1 in order to compute the binary-valued approximation of the solution of the continuous relaxation on the uniform mesh of  $256 \times 256$  squares. COR and SUR produce the same resulting control. SHG also returns the same control if the approximation constraint in the problem formulation is used with  $\theta = 1$ , see Algorithm 5, indicating that the feasible set with this bound leaves (almost) no room for a reduction of the interface length. We therefore increase the feasible set by setting  $\theta = 10$ , which provides a trade-off that relaxes the approximation quality and allows to reduce the interface length within the prescribed approximation quality. The optimality gap is then higher but the interface length decreases and we obtain a qualitatively different solution.

Then we initialize BTR, which also operates on the uniform grid of  $256 \times 256$  squares, with the output control of SUR. Because of the near-optimality achieved by CIA and the fact that BTR does not produce a globally optimal solution of (6.2) for a fixed discretization, we expect that BTR can close a small part but not much of the remaining optimality gap between the upper bound given by the objective value for the output SUR and the lower bound given by the solution of the continuous relaxation.

This expectation is met by our computational results. Specifically, BTR is able to close a portion of the remaining optimality gap of CIA. We also start BTR from two further different initializations, specifically, from  $x = 0$  and from a cellwise rounding of the solution of the continuous relaxation to  $\{0, 1\}$ . All of the objective values are very close; the optimality gap is always around  $10^{-6}$  (at a magnitude of  $10^{-3}$ ) so that BTR also produces a near-optimal solution on this grid when it is initialized differently.

Moreover, we observe that the running times of BTR are longer than those of CIA, which we attribute to the facts that BTR is a pure first-order method, has restricted options for its feasible steps available (compared with usual gradient-based solvers for the continuous relaxation), and requires a sorting operation additionally to each adjoint solve of each accepted step. The higher running times are reflected by correspondingly high numbers of iterations of our implementation of Algorithm 2, as is typical for first-order methods. The running time of the continuous relaxation followed by an execution of BTR on cellwise rounding is moderately lower (about 20%) than that of BTR for initial control zero.

To give a qualitative and visual impression of the results, we provide the six computed controls in this experiment in Figure 1. A visual inspection of the results shows that starting BTR from a cellwise rounded solution or zero seems to have a regularizing effect on the resulting microstructure. We compute the length of the interface between the level sets for the values zero and the objective value one for all of the computed controls. We obtain that the none of the controls computed with CIA with SUR or COR as the Round procedure, CIA with SHG as the Round procedure, and BTR started from cellwise rounding dominates another in terms of low objective or low interface length value. Specifically, CIA with SUR or COR produces the lowest objective value but the highest interface length. The interface lengths for the controls with BTR started from zero or cellwise rounding are significantly lower (about 50 %) while the objective value increases slightly. The interface length decreases further (about 30 %) for CIA with SHG for  $\theta = 10$  but the increase in the objective is also higher.

While the behavior of the objective values follows our analysis, we cannot explain our consistent observation that some spatial coherence is maintained during the optimization with BTR. In order to do so, we need to better understand the possible trade-off between objective and interface length, which also motivated us to introduce SHG and report its results, which is impaired by its current computational intractability and us resorting to a heuristic solution. We leave further considerations in this direction to future research.

These findings also lead to the question whether one can obtain the regularization effect in a

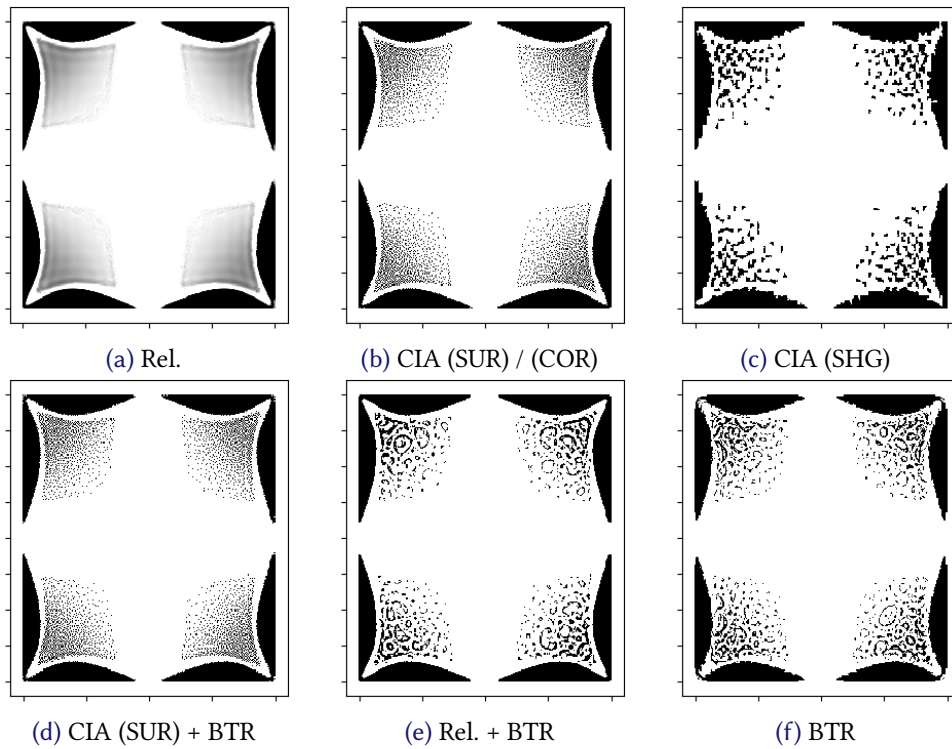


Figure 1: Visualization of the resulting control functions for *Rel.* (= continuous relaxation of (6.1)), *CIA (SUR) / (COR)* (= continuous relaxation and SUR / COR), *CIA (SHG)* (= continuous relaxation and SHG), *CIA (SUR) / (COR) + BTR*, *Rel. + BTR* (= continuous relaxation and BTR started from a cellwise rounding), and *BTR* (= BTR started from zero), where the value one is colored black and the value zero is colored white. For *Rel.*, the intermediate values in  $[0, 1]$  are depicted in grayscale.

**Table 1:** Objective values, remaining optimality gaps, BTR iteration numbers, running times, and interface lengths obtained for the cases *Rel.* (= continuous relaxation of (6.1)), *CIA (COR)* (= continuous relaxation and COR), *CIA (SUR)* (= continuous relaxation and SUR), *CIA (SHG)* (= continuous relaxation and SHG), *CIA + BTR* (= continuous relaxation and BTR started from SUR), *Rel. + BTR* (= continuous relaxation and BTR started from a cellwise rounding), and *BTR* (= BTR started from zero).

	Rel.	CIA (COR)	CIA (SHG)	CIA (SUR)	CIA + BTR	Rel. + BTR	BTR
Obj. [ $10^{-3}$ ]	4.0798	4.0808	4.1200	4.0808	4.0807	4.0837	4.0862
Opt. gap [ $10^{-6}$ ]	0	1.06	40.20	1.06	0.89	3.96	6.41
Time [s]	821	828	829	825	921	3928	4906
BTR iterations	n/a	n/a	n/a	n/a	51	1575	2481
Interface length	n/a	117.0	48.1	117.0	116.9	66.4	74.2

**Table 2:** Remaining optimality gaps, BTR iteration numbers, and interface lengths for BTR started from zero for different mesh sizes.

Mesh	Optimality gap [ $10^{-6}$ ]	BTR iterations	Interface length
$32 \times 32$	1200.79	116	21.8
$64 \times 64$	261.14	356	34.2
$128 \times 128$	38.02	938	50.6
$256 \times 256$	6.41	2481	74.2

hybridized method, where BTR is initialized with an approximation of SUR that is computed on a coarser control mesh. This is investigated in §6.4. We report the objective values, optimality gaps, running times, number of iterations required by our implementation of BTR, and interface lengths between the level sets in Table 1.

We also evaluate how our implementation of BTR behaves with respect to mesh refinement when initialized with zero and running it until the trust-region radius contracts. The optimality gap decreases for finer meshes as finer microstructures can be computed in order to more closely approximate a minimizer of the continuous relaxation. However, the number of iterations grows with a factor of approximately three when the mesh size is halved (the number of cells increases by a factor of four). In particular, our implementation of the algorithm is clearly not mesh-independent. The interface length also increases when finer meshes are chosen, which is consistent with the fact that this quantity tends to infinity if a function with values in  $(0, 1)$  on a set of strictly positive measure (in this case the solution of the relaxation) is approximated weakly- $*$  in  $L^\infty$  by binary functions. We provide the corresponding data in Table 2.

#### 6.4 HYBRIDIZATION OF SUR AND BTR

We explore the regularization effect that we have observed above by executing a hybridized method by initializing BTR with controls that are computed by CIA, where the second step is computed with SUR.

SUR operates on a mesh, and its approximation quality depends on the mesh size of this grid [20, 28]. We use a sequence of uniformly refined grids from  $8 \times 8$  grid cells to  $256 \times 256$  grid cells and start BTR, which itself operates on the  $256 \times 256$  grid of squares, with the resulting controls.

We assess the running times and iterations of BTR, the remaining optimality gaps, and the interface lengths between the level sets of the two control realizations one and zero.

The remaining optimality gaps are of an order of magnitude of  $10^{-6}$ , where the ones achieved for the

initializations with the SUR solutions for the  $128 \times 128$  and  $256 \times 256$  grids are slightly but noticeably smaller. This can be attributed to the fact that SUR provides already small optimality gaps, which are then further reduced by BTR.

Moreover, the running times and iterations of BTR do not differ much from the  $8 \times 8$  to  $64 \times 64$  initializations (all above 3000 s) and then drop in two large steps to 1074 s and 96 s for the two finest grids, again capitalizing on the fact that SUR already provides a near-optimal solution.

The interface lengths of the resulting level sets of the controls obtained for the  $8 \times 8$  to  $128 \times 128$  initializations are between 62.5 and 70.2, approximately 50 % smaller than the interface length obtained for the finest grid. This indicates that one may find a sensible trade-off, where the regularization effect of BTR is still pronounced and BTR can be meaningfully accelerated by means of CIA, in our case by executing SUR in CIA on the  $128 \times 128$  initialization.

We have recorded the obtained results in Table 3. To provide a visual impression again, we contrast the computed controls for SUR with the resulting ones after initializing BTR with them for the  $8 \times 8$  and  $128 \times 128$  grids in Figure 2.

**Table 3:** Remaining optimality gaps, running times, BTR iterations, and interface lengths for executing BTR on solutions of CIA, where the second step is computed by means of SUR, for refined grids used for SUR.

SUR Mesh	$8 \times 8$	$16 \times 16$	$32 \times 32$	$64 \times 64$	$128 \times 128$	$256 \times 256$
Optimality gap [ $10^{-6}$ ]	4.99	6.66	4.68	5.30	2.91	0.89
Time (only BTR) [s]	4376	3929	4049	3167	1074	96
BTR iterations	2242	2023	2069	1619	552	51
Interface length	70.2	67.9	67.4	62.5	67.0	116.9

### ACKNOWLEDGMENTS

We thank two anonymous referees for providing helpful feedback on the manuscript. We thank Peter Bella and Christian Meyer (both TU Dortmund University) for helpful discussions on the topic. This work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Scientific Discovery through the Advanced Computing (SciDAC) Program through the FASTMath Institute under Contract No. DE-AC02-06CH11357 and by the German Research Foundation under GRK 2297 MathCoRe (project No. 314838170), SPP 1962 (projects No. SA 2016/1-2, KI 1839/1-2), and SPP 2231 (project No. SA 2016/3-1, KI 417/9-1), and by the German Federal Ministry of Education and Research within the program “Mathematics for Innovations”, project “Power to Chemicals.” This work is also part of the Research Initiative “SmartProSys: Intelligent Process Systems for the Sustainable Production of Chemicals” funded by the Ministry for Science, Energy, Climate Protection and the Environment of the German State of Saxony-Anhalt.

### APPENDIX A AUXILIARY RESULTS

This appendix provides additional information on the Round subroutine in the CIA method, Algorithm 1, and establishes relationships between set-based and characteristic function. We also prove the equivalence between (3.1) and  $(TR(\Delta))$  and the existence of minimizers for them.

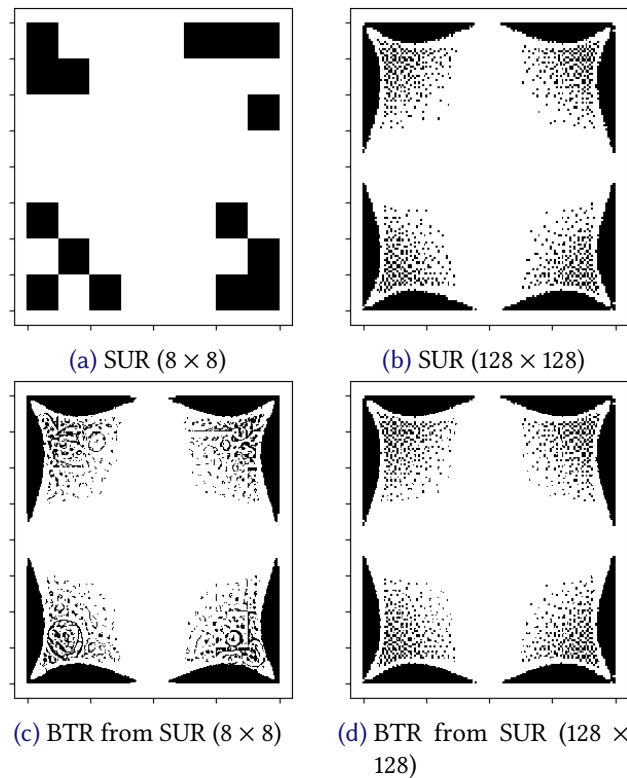


Figure 2: Visualization of the resulting control functions for SUR on uniform  $8 \times 8$  and  $128 \times 128$  square grids (top row) and BTR initialized with them (bottom row).

APPENDIX A.1 ROUND IN ALGORITHM 1

Algorithm 1 employs a Round subroutine, which takes grids as inputs. In order to establish the approximation property (2.1), see also [28, 20], it is assumed that the sequence of grids is a so-called *order-conserving domain dissection*, which is formally defined below.

Definition A.1 (Order-conserving domain dissection, Definition 4.3 in [28]). Let  $\Omega \subset \mathbb{R}^d$ . Then we call a sequence  $(\{S_1^n, \dots, S_{N^n}^n\})_n \subset 2^{\mathcal{B}(\Omega)}$  an *order-conserving domain dissection* if

1.  $\{S_1^n, \dots, S_{N^n}^n\}$  is a finite partition of  $\Omega$  for all  $n \in \mathbb{N}$ ,
2.  $\max\{\lambda(S_i^n) \mid i \in \{1, \dots, N^n\}\} \rightarrow 0$ ,
3. for all  $n \in \mathbb{N}$  for all  $i \in \{1, \dots, N^{n-1}\}$  there exists  $1 \leq j < k \leq N^n$  such that  $\bigcup_{\ell=j}^k S_\ell^n = S_i^{n-1}$ , and
4. the cells  $S_j^n$  shrink regularly (there exists  $C > 0$  such that for each  $S_j^n$  there exists a ball  $B_j^n$  with  $S_j^n \subset B_j^n$  and  $\lambda(S_j^n) \geq C\lambda(B_j^n)$ ).

We briefly introduce the three choices for Round in Algorithm 1, Line 3 that are used in our computational experiments in §6 below.

**Sum-up rounding (SUR)** The algorithm is stated as Algorithm 3 and works as follows. It starts from a  $[0, 1]^m$ -valued function  $\alpha$  such that  $\sum_{i=1}^m \alpha_i = 1$  a.e. and that is defined on an ordered sequence of grid cells that partition the domain  $\Omega$ . The algorithm iterates over the grid cells in the given order; identifies an entry  $i \in \{1, \dots, m\}$  such that the cumulative difference up to the current grid cell to a rounded function  $\omega$ , which is  $\{0, 1\}^m$ -valued, satisfies  $\sum_{i=1}^m \omega_i = 1$  a.e., and is defined on the same grid

---

**Algorithm 3** Sum-Up Rounding (multidimensional variant) [28]

---

**Input:** Ordered grid cells  $S_1, \dots, S_N \subset \Omega$  that partition  $\Omega$ .

**Input:** Function  $\alpha \in L^1(\Omega, \mathbb{R}^m)$  with averages  $a_{k,i}$  such that  $a_{k,i} = \frac{1}{\lambda(S_k)} \int_{S_k} \alpha_i(s) ds$  for all  $i \in \{1, \dots, m\}$ , and  $\sum_{i=1}^m a_{k,i} = 1$  for all  $k \in \{1, \dots, N\}$ .

```

1:  $\phi_0 := 0_{\mathbb{R}^m}$ 
2: for  $k = 1, \dots, N$  do
3:    $\gamma_k \leftarrow \phi_{k-1} + a_k \lambda(S_k)$  ( $a_k$  is short for the vector  $(a_{k,1}, \dots, a_{k,m})^T$ )
4:    $w_{k,i} \leftarrow \begin{cases} 1 & i \in \arg \max\{\gamma_{k,j} \mid j \in \{1, \dots, m\}\}, \\ 0 & \text{else} \end{cases}$  for all  $i \in \{1, \dots, m\}$ 
5:    $\phi_k \leftarrow \sum_{i=1}^k (a_k - w_k) \lambda(S_k)$  ( $w_k$  is short for the vector  $(w_{k,1}, \dots, w_{k,m})^T$ )
6: end for
7: return  $\omega := \sum_{k=1}^N w_k \chi_{S_k}$ 

```

---

as  $\alpha$ , is maximal. Then the algorithm sets  $\omega_i$  to one on the current grid cell and the other entries to zero on the respective grid cell.

**Combinatorial optimization-based rounding (COR)** The algorithm is stated as [Algorithm 4](#) and works as follows. It starts from a  $[0, 1]^m$ -valued function  $\alpha$  such that  $\sum_{i=1}^m \alpha_i = 1$  a.e. and that is defined on an

---

**Algorithm 4** Combinatorial optimization-based rounding (multidimensional variant)

---

**Input:** Ordered grid cells  $S_1, \dots, S_N \subset \Omega$  that partition  $\Omega$ .

**Input:** Function  $\alpha \in L^1(\Omega, \mathbb{R}^m)$  with averages  $a_{k,i}$  such that  $a_{k,i} = \frac{1}{\lambda(S_k)} \int_{S_k} \alpha_i(s) ds$  for all  $i \in \{1, \dots, m\}$ , and  $\sum_{i=1}^m a_{k,i} = 1$  for all  $k \in \{1, \dots, N\}$ .

```

1: compute  $w$  as minimizer of
   
$$\arg \min_{w, \eta} \eta \quad \text{s.t.} \quad \begin{cases} w \in \{0, 1\}^{N \times m} \\ \sum_{i=1}^m w_{k,i} = 1 \text{ for all } k \in \{1, \dots, N\} \\ \left| \sum_{j=1}^k (a_{j,i} - w_{j,i}) \lambda(S_j) \right| \leq \eta \text{ for all } k \in \{1, \dots, N\} \text{ and all } i \in \{1, \dots, m\} \end{cases}$$

2: return  $\omega := \sum_{k=1}^N w_k \chi_{S_k}$ 

```

---

ordered sequence of grid cells that partition the domain  $\Omega$ . The algorithm computes a  $\{0, 1\}^m$ -valued function that satisfies  $\sum_{i=1}^m \omega_i = 1$  a.e. that is piecewise constant on this grid. The function is computed such that it minimizes the maximum of the modulus of the cumulative difference (integration) to  $\alpha$  from the first to the  $k$ -th cell over  $k \in \{1, \dots, N\}$ . [Algorithm 4](#) can be implemented very efficiently using the shortest path approach described in [3, 5] on multidimensional domains because the costs are sequence independent in the sense of [3, § 2.6.1]. We use the open-source software `scarp_solver`<sup>1</sup> with the option `--sur_costs` in order to solve SHG in this work.

**Primal heuristic for switching cost aware rounding (SHG)** Switching cost aware rounding [5, 6] is stated as [Algorithm 5](#) and works as follows. It starts from a  $[0, 1]^m$ -valued function  $\alpha$  such that  $\sum_{i=1}^m \alpha_i = 1$  a.e. and that is defined on an ordered sequence of grid cells that partition the domain

---

<sup>1</sup> Accessed on <https://github.com/chrhansk/SCARP> on 02/15/2022.



**Algorithm 5** Switching cost aware rounding (multidimensional variant)

**Input:** Ordered grid cells  $S_1, \dots, S_N \subset \Omega$  that partition  $\Omega$ .

**Input:** Function  $\alpha \in L^1(\Omega, \mathbb{R}^m)$  with averages  $a_{k,i}$  such that  $a_{k,i} = \frac{1}{\lambda(S_k)} \int_{S_k} \alpha_i(s) ds$  for all  $i \in \{1, \dots, m\}$ , and  $\sum_{i=1}^m a_{k,i} = 1$  for all  $k \in \{1, \dots, N\}$ .

**Input:** Trade-off parameter  $\theta \geq 1$ .

1: compute  $w$  as minimizer of

$$\arg \min_w \frac{1}{m} \text{TV} \left( \sum_{k=1}^N w_k \chi_{S_k} \right) \quad \text{s.t.} \quad \begin{cases} w \in \{0, 1\}^{N \times m} \\ \sum_{i=1}^m w_{k,i} = 1 \text{ for all } k \in \{1, \dots, N\} \\ \left| \sum_{j=1}^k (a_{j,i} - w_{j,i}) \lambda(S_j) \right| \leq \theta \sum_{i=2}^m \frac{1}{i} \max_{\ell} \lambda(S_{\ell}) \\ \text{for all } k \in \{1, \dots, N\} \text{ and all } i \in \{1, \dots, m\} \end{cases}$$

2: **return**  $\omega := \sum_{k=1}^N w_k \chi_{S_k}$

$\Omega$ . The algorithm computes a  $\{0, 1\}^m$ -valued function that satisfies  $\sum_{i=1}^m \omega_i = 1$  a.e. that is piecewise constant on this grid. The function is computed such that it minimizes the total variation of  $w$  while constraining the modulus of the cumulative difference (integration) to  $\alpha$  from the first to the  $k$ -th cell over  $k \in \{1, \dots, N\}$  by  $\theta \sum_{i=2}^m \frac{1}{i} \max_{\ell} \lambda(S_{\ell})$ .  $\text{TV}(w)$  denotes the total variation of  $w$  in Algorithm 5. The constant  $\sum_{i=2}^m \frac{1}{i} \max_{\ell} \lambda(S_{\ell})$  is guaranteed by Algorithm 3, see the analysis in [19, 29], so that a feasible point always exists for  $\theta = 1$ . The feasible set may be increased by choosing  $\theta > 1$  in order to leave room for a better objective while allowing for a larger approximation error. After discretization, the optimization problem in Algorithm 5 becomes an integer linear program. While a shortest path reformulation and efficient solution algorithms exist for the case that  $\Omega \subset \mathbb{R}$ , it is not known if it can be solved efficiently for  $\Omega \subset \mathbb{R}^d, d \geq 2$ . By considering only the jumps between subsequent cells along the ordering of the grid cells in the minimization, we can obtain a suboptimal feasible point using the shortest path approach described in [3, 5]. This is introduced as the heuristic SCARP\_HG in [6]. We use the open-source software scarp\_solver<sup>1</sup> with the option `--scale  $\theta$`  in order to solve SCARP\_HG / (SHG).

**APPENDIX A.2 EQUIVALENCE OF (TR( $\Delta$ )) TO (3.1) AND EXISTENCE OF MINIMIZER**

**Proposition A.2.** Let  $g = \nabla J(\chi_U)(\chi_{U^c} - \chi_U)$ . A set  $D \in \mathcal{B}$  satisfies  $\lambda(D) \leq \Delta$  if and only if  $d = \chi_{U \Delta D} - \chi_U$  is feasible for (TR( $\Delta$ )) and the corresponding objective values for (3.1) and (TR( $\Delta$ )) coincide.

If  $D \in \mathcal{B}$  does not satisfy  $D \subset g^{-1}((-\infty, 0])$ , then its objective value is greater or equal than  $D \cap g^{-1}((-\infty, 0])$  so that the optimal objective for (TR( $\Delta$ )) is not altered by the additional feasible points.

*Proof.* For every  $D \in \mathcal{B}$ ,  $d$  can be computed with the formula above. On the other hand, the constraint  $\chi_U(s) + d(s) \in \{0, 1\}$  in (TR( $\Delta$ )) implies that  $\chi_U + d$  is a characteristic function of a measurable set  $A$ , which in turn can be represented as  $A = U \Delta D$  for the set  $D = U \Delta A$ . Moreover, we have  $\|d\|_{L^1} = \|\chi_{U \Delta D} - \chi_U\|_{L^1} = \lambda(D)$ , which shows the equivalence of the trust-region constraint.

Moreover, for any feasible  $D$  and  $d = \chi_{U \Delta D} - \chi_U$ , we distinguish the four cases whether  $s \in D$  and/or in  $s \in U$  holds and obtain

$$\int_D g ds = \int_D \nabla J(\chi_U)(\chi_{U^c} - \chi_U) ds = \int_{\Omega} \nabla J(\chi_U) d ds = (\nabla J(\chi_U), d)_{L^2},$$

which gives the coincidence of the objective values.

Finally, let  $D \in \mathcal{B}$  be given. Then

$$\int_D g \, ds = \int_{D \cap g^{-1}((-\infty, 0])} g \, ds + \int_{D \cap g^{-1}((0, \infty))} g \, ds \leq \int_{D \cap g^{-1}((-\infty, 0])} g \, ds.$$

□

**Proposition A.3.** *Let  $\Omega$  be a bounded domain. Let  $U \in \mathcal{B}$ . Let  $\Delta \in [0, \infty]$ . Let  $\nabla J \in L^1(\Omega)$ . Then  $(\text{TR}(\Delta))$  admits a minimizer  $\tilde{d}$  with  $(\nabla J(\chi_U), \tilde{d})_{L^2} \leq -\frac{\Delta}{\lambda(\Omega)} C(\chi_U)$ .*

*Proof.* Let  $g := \nabla J(\chi_U)(\chi_{U^c} - \chi_U)$ , and let  $D : \mathbb{R} \rightarrow \mathcal{B}$  be defined as  $D(x) := g_U^{-1}((-\infty, x))$  for  $x \in \mathbb{R}$ . Then  $D(x) \subset D(y)$  for all  $x \leq y$ , implying that  $\|\chi_{U \Delta D(x)} - \chi_U\|_{L^1} = \lambda(D(x))$  is monotone in  $x$ . Let  $d(x) := \chi_{U \Delta D(x)} - \chi_U$ . Then the  $d(x)$  are greedy solution candidates for  $(\text{TR}(\Delta))$  with  $\lim_{x \rightarrow -\infty} \|d(x)\|_{L^1} = 0$ . Specifically,  $d(0)$  minimizes  $(\text{TR}(\Delta))$  if  $\Delta = \infty$ .

Let  $\Delta < \infty$ . If  $\|d(0)\|_{L^1} \leq \Delta$ , then  $d(0)$  minimizes  $(\text{TR}(\Delta))$ . We restrict to  $\|d(0)\|_{L^1} > \Delta$ . Because of the greedy construction,  $d(\bar{x})$  is optimal if  $\|d(\bar{x})\|_{L^1} = \Delta$  for some  $\bar{x} < 0$ . We consider the case where there is no such  $\bar{x} \leq 0$ . We consider  $\bar{x} := \sup\{x \mid \|d(x)\|_{L^1} \leq \Delta\}$ . If  $d$  is continuous at  $\bar{x}$ , then  $\|d(\bar{x})\|_{L^1} = \Delta$ , and  $d(\bar{x})$  minimizes  $(\text{TR}(\Delta))$ . We distinguish two situations.

**Situation 1.** If  $d$  is only left continuous at  $\bar{x}$ , then we have  $\|d(\bar{x})\|_{L^1} \leq \Delta < \lim_{y \downarrow \bar{x}} \|d(y)\|_{L^1}$ . Thus there exists a set  $A \in \mathcal{B}$  satisfying  $A \cap D(\bar{x}) = \emptyset$ ,  $A \subset D(y)$  for all  $y > \bar{x}$ , and  $\lambda(A) = \lim_{y \downarrow \bar{x}} \|d(y)\|_{L^1} - \|d(\bar{x})\|_{L^1}$ . Such a set also exists if  $\|d(\bar{x})\|_{L^1} \leq \Delta$  and  $d$  is neither left nor right continuous at  $\bar{x}$ .

**Situation 2.** If  $d$  is only right continuous at  $\bar{x}$ , then we have  $\lim_{y \uparrow \bar{x}} \|d(y)\|_{L^1} \leq \Delta < \|d(\bar{x})\|_{L^1}$ . Thus there exists a set  $A \in \mathcal{B}$  satisfying  $A \subset D(\bar{x})$ ,  $A \cap D(y) = \emptyset$  for all  $y < \bar{x}$ , and  $\lambda(A) = \|d(\bar{x})\|_{L^1} - \lim_{y \downarrow \bar{x}} \|d(y)\|_{L^1}$ . Such a set also exists if  $\Delta < \|d(\bar{x})\|_{L^1}$  and  $d$  is neither left nor right continuous at  $\bar{x}$ .

Because of the monotony of  $\|d(\cdot)\|_{L^1}$  and the fact that the limits  $\lim_{y \uparrow \bar{x}} \|d(y)\|_{L^1}$  and  $\lim_{y \downarrow \bar{x}} \|d(y)\|_{L^1}$  always exist by virtue of continuity from below and above of the Lebesgue measure, this distinction is exhaustive. Because of the mean value property of the Lebesgue measure [8, Cor. 1.12.10], there exists  $\mathcal{B} \ni B \subset A$  with  $\lambda(B) = \Delta - \|d(\bar{x})\|_{L^1}$  (Situation 1) or  $\lambda(B) = \|d(\bar{x})\|_{L^1} - \Delta$  (Situation 2). In Situation 1 we set  $\tilde{D} := D(\bar{x}) \cup B$ , and in Situation 2 we set  $\tilde{D} := D(\bar{x}) \setminus B$ . In both situations,  $\tilde{d} := \chi_{U \Delta \tilde{D}} - \chi_U$  minimizes  $(\text{TR}(\Delta))$ .

The greedy construction of  $D(\bar{x})$  and thus  $\tilde{D}$  with respect to  $g$  imply

$$\begin{aligned} \frac{1}{\Delta} \int_{\tilde{D}} g \, ds &\leq \frac{1}{|\lambda(g^{-1}((-\infty, 0)))|} \int_{g^{-1}((-\infty, 0))} g \, ds \\ &= -\frac{1}{|\lambda(g^{-1}((-\infty, 0)))|} C(\chi_U) \leq -\frac{1}{|\lambda(\Omega)|} C(\chi_U), \end{aligned}$$

where we have used the definition of  $C$  for the equality and that the integrand is negative for the second inequality. Then the identity  $(\nabla J(\chi_U), \tilde{d})_{L^2} = \int_{\tilde{D}} g \, ds$  yields the claimed inequality. □

## APPENDIX B RELATIONSHIP BETWEEN SET-BASED AND CHARACTERISTIC FUNCTION POINTS OF VIEW

In this appendix, we discuss the relationship of Assumption 1.1 to [10] and the corresponding Taylor expansions.

APPENDIX B.1 RELATION OF THE SETTING OF ASSUMPTION 1.1 TO [10]

Our setting and the assumptions cannot be compared or embedded directly into the setting of [10] because [10] analyzes operations on sets in atomless measure spaces; see, for example, [8, Definition 1.12.7], while we restrict Algorithm 2 to functionals that operate on functions. We note, however, that our arguments do not hinge on the particular choice of the Lebesgue–Borel measure for  $L^1$  and  $L^2$  and it is still possible to find correspondences of the parts of Assumption 1.1 in [10], which we do below.

First, we note that the  $L^1$ -regularization that is used in the experiments in [10] does satisfy our assumptions because it is linear when restricted to the feasible set of (R).

Assumption 1.1, (a) is assumed in [10, Theorem 3], which essentially shows  $\liminf_{n \rightarrow \infty} C(\chi_{U^n}) = 0$ . We note that the assumption is not explicitly required if Assumption 1.1, (d) holds as well because the latter implies weak continuity of  $J$  by means of Proposition 2.1, the feasible set of (R) is weakly compact, and continuous functions assume their minimum on compact sets.

Assumption 1.1, (b) implies Assumption 1.3 in [10]. The local first-order Taylor expansion in [10, Theorem 1] for objective functions defined on measurable sets follows for the natural construction

$$J_s(A) := J(\chi_A), \text{ and } J'_s(A)D := \langle J'(\chi_A), \chi_{D \setminus A} - \chi_{D \cap A} \rangle_{L^\infty(\Omega), L^1(\Omega)}$$

for  $A, D \in \mathcal{B}$ . We give a short proof in Proposition B.1 below.

Assumption 1.1, (c) implies the assumption (5) in Lemma 3 and the (10) in Theorem 3 in [10]. They serve to obtain sufficient decrease in the algorithm, which is exactly the case, where it is needed the proof of Theorem 4.4 in this work. It is also similar to Assumption 4.1 in [22], where it serves the same purpose. It is no coincidence that assumptions of this type are required for the analysis of descent algorithms that manipulate binary control functions in  $L^p$ -norms for the following reason. All binary control functions  $v$  satisfy  $\|v\|_{L^1} = \|v\|_{L^2}^2$ . Therefore, bounding the error term of the Taylor expansion by the squared  $L^2$ -norm is not sufficient to obtain a sufficient decrease condition because the linear predicted reduction is bounded from below only by a fraction of the maximal  $L^1$ -norm of the trust-region step. Thus we cannot prove that the linear predicted reduction dominates the remainder terms for small trust-region radii without this further assumption. Because the trust-region subproblem does not allow fractional-valued control functions, a greedy strategy can always be used to approximate the infimal value of the trust-region subproblem regardless of the  $L^p$ -norm ( $p \in [0, \infty)$ ) that is used for the trust-region radius. Consequently, this assumption cannot be avoided by choosing a different  $L^p$ -norm for the trust-region radius.

Assumption 1.1, (d) is a compactness assumption on the derivative of the objective function, which allows us to infer the stationarity of weak-\* cluster points for (R). It is not assumed in [10], which does neither analyze the relationship to the continuous relaxation nor show such a result. It implies Assumption 1.4 for CIA in [20] by means of Proposition 2.1. The reason for this difference is that we need to pass to the limit in the derivative of the objective functional in the norm when certifying stationarity.

APPENDIX B.2 TAYLOR EXPANSION FOR SETS AND CHARACTERISTIC FUNCTIONS

Let  $J_s, J'_s$  be given as above. We say that  $J_s$  is Fréchet differentiable if

$$J_s(A \triangle D) = J_s(A) + J'_s(A)D + o(\lambda(D)),$$

which is the assertion of [10, Theorem 1]. Due to the assumed differentiability  $J : L^1(\Omega) \rightarrow \mathbb{R}$  in Assumption 1.1, (b), that is with respect to the  $L^1$ -norm on the domain, we obtain that  $J_s$  is Fréchet differentiable below.

**Proposition B.1.**  *$J_s$  is Fréchet differentiable.*

*Proof.* Let  $A, D \in B$ . We use the defined identity  $J_s(A \Delta D) = J(\chi_{A \Delta D})$ , Taylor's theorem for  $J$ , and the identities  $A \Delta D = (A \setminus D) \cup (D \setminus A)$ , and  $A = (A \setminus D) \cup (A \cap D)$ —where both unions are disjoint—to deduce

$$\begin{aligned} J_s(A \Delta D) &= J(\chi_A) + \langle J'(\chi_A), \chi_{A \Delta D} - \chi_A \rangle_{L^\infty, L^1} + o(\|\chi_{A \Delta D} - \chi_A\|_{L^1}) \\ &= J(\chi_A) + \langle J'(\chi_A), \chi_{D \setminus A} - \chi_{A \cap D} \rangle_{L^\infty, L^1} + o(\|\chi_{D \setminus A} - \chi_{A \cap D}\|_{L^1}). \end{aligned}$$

We observe that  $\|\chi_{D \setminus A} - \chi_{A \cap D}\|_{L^1} = \|\chi_D\|_{L^1}$ , and  $\|\chi_D\|_{L^1} = \lambda(D)$ . This implies

$$J_s(A \Delta D) = J(\chi_A) + \langle J'(\chi_A), \chi_{D \setminus A} - \chi_{A \cap D} \rangle_{L^\infty, L^1} + o(\lambda(D)).$$

Inserting the definitions of  $J_s(A)$  and  $J'_s(A)D$  yields the claim.  $\square$

## REFERENCES

- [1] Y. I. Alber, A. N. Iusem, and M. V. Solodov, On the projected subgradient method for nonsmooth convex optimization in a Hilbert space, *Mathematical Programming* 81 (1998), 23–35, [doi:10.1007/bf01584842](https://doi.org/10.1007/bf01584842).
- [2] M. P. Bendsøe and O. Sigmund, Extensions and applications, in *Topology Optimization*, Springer, 2004, 71–158, [doi:10.1007/978-3-662-05086-6](https://doi.org/10.1007/978-3-662-05086-6).
- [3] F. Bestehorn, *Combinatorial Algorithms and Complexity of Rounding Problems Arising in Mixed-Integer Optimal Control*, PhD thesis, Technical University of Braunschweig, 2022, [doi:10.24355/dbbs.084-202203101114-0](https://doi.org/10.24355/dbbs.084-202203101114-0).
- [4] F. Bestehorn, C. Hansknecht, C. Kirches, and P. Manns, A switching cost aware rounding method for relaxations of mixed-integer optimal control problems, in *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2019, 7134–7139, [doi:10.1109/cdc40024.2019.9030063](https://doi.org/10.1109/cdc40024.2019.9030063).
- [5] F. Bestehorn, C. Hansknecht, C. Kirches, and P. Manns, Mixed-integer optimal control problems with switching costs: a shortest path approach, *Mathematical Programming* 188 (2021), 621–652, [doi:10.1007/s10107-020-01581-3](https://doi.org/10.1007/s10107-020-01581-3).
- [6] F. Bestehorn, C. Hansknecht, C. Kirches, and P. Manns, Switching cost aware rounding for relaxations of mixed-integer optimal control problems: the 2-D case, *IEEE Control Systems Letters* 6 (2021), 548–553, [doi:10.1109/cdc40024.2019.9030063](https://doi.org/10.1109/cdc40024.2019.9030063).
- [7] L. Boccardo and T. Gallouët, Non-linear elliptic and parabolic equations involving measure data, *Journal of Functional Analysis* 87 (1989), 149–169, [doi:10.1016/0022-1236\(89\)90005-0](https://doi.org/10.1016/0022-1236(89)90005-0).
- [8] V. I. Bogachev, *Measure Theory*, volume 1, Springer, 2007, [doi:10.1007/978-3-540-34514-5](https://doi.org/10.1007/978-3-540-34514-5).
- [9] J. C. Dunn, Convergence rates for conditional gradient sequences generated by implicit step length rules, *SIAM Journal on Control and Optimization* 18 (1980), 473–487, [doi:10.1137/0318035](https://doi.org/10.1137/0318035).
- [10] M. Hahn, S. Leyffer, and S. Sager, Binary optimal control by trust-region steepest descent, *Mathematical Programming* 197 (2023), 147–190, [doi:10.1007/s10107-021-01733-z](https://doi.org/10.1007/s10107-021-01733-z).
- [11] R. Haller-Dintelmann, C. Meyer, J. Rehberg, and A. Schiela, Hölder continuity and optimal control for nonsmooth elliptic problems, *Applied Mathematics and Optimization* 60 (2009), 397–428, [doi:10.1007/s00245-009-9077-x](https://doi.org/10.1007/s00245-009-9077-x).

- [12] F. M. Hante, Mixed-integer optimal control for PDEs: relaxation via differential inclusions and applications to gas network optimization, in *Mathematical Modelling, Optimization, Analytic and Numerical Solutions*, Springer, 2020, 157–171, doi:[10.1007/978-981-15-0928-5\\_7](https://doi.org/10.1007/978-981-15-0928-5_7).
- [13] F. M. Hante, G. Leugering, A. Martin, L. Schewe, and M. Schmidt, Challenges in optimal control problems for gas and fluid flow in networks of pipes and canals: from modeling to industrial applications, in *Industrial Mathematics and Complex Systems*, Springer, 2017, 77–122, doi:[10.1007/978-981-10-3758-0\\_5](https://doi.org/10.1007/978-981-10-3758-0_5).
- [14] F. M. Hante and S. Sager, Relaxation methods for mixed-integer optimal control of partial differential equations, *Computational Optimization and Applications* 55 (2013), 197–225, doi:[10.1007/s10589-012-9518-3](https://doi.org/10.1007/s10589-012-9518-3).
- [15] J. Haslinger and R. A. Mäkinen, On a topology optimization problem governed by two-dimensional Helmholtz equation, *Computational Optimization and Applications* 62 (2015), 517–544, doi:[10.1007/s10589-015-9746-4](https://doi.org/10.1007/s10589-015-9746-4).
- [16] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE Constraints*, volume 23, Springer Science & Business Media, 2008, doi:[10.1007/978-1-4020-8839-1](https://doi.org/10.1007/978-1-4020-8839-1).
- [17] M. Jung, *Relaxations and Approximations for Mixed-integer Optimal Control*, PhD thesis, Heidelberg University, 2014, doi:[10.11588/heidok.00016036](https://doi.org/10.11588/heidok.00016036).
- [18] M. N. Jung, G. Reinelt, and S. Sager, The Lagrangian relaxation for the combinatorial integral approximation problem, *Optimization Methods and Software* 30 (2015), 54–80, doi:[10.1080/10556788.2014.890196](https://doi.org/10.1080/10556788.2014.890196).
- [19] C. Kirches, F. Lenders, and P. Manns, Approximation properties and tight bounds for constrained mixed-integer optimal control, *SIAM Journal on Control and Optimization* 58 (2020), 1371–1402, doi:[10.1137/18m1182917](https://doi.org/10.1137/18m1182917).
- [20] C. Kirches, P. Manns, and S. Ulbrich, Compactness and convergence rates in the combinatorial integral approximation decomposition, *Mathematical Programming* 188 (2021), 569–598, doi:[10.1007/s10107-020-01598-8](https://doi.org/10.1007/s10107-020-01598-8).
- [21] T. Larsson and M. Patriksson, A class of gap functions for variational inequalities, *Mathematical Programming* 64 (1994), 53–79, doi:[10.1007/bf01582565](https://doi.org/10.1007/bf01582565).
- [22] 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), 66, doi:[10.1051/cocv/2022059](https://doi.org/10.1051/cocv/2022059).
- [23] 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), 193–221, doi:[10.1007/s10589-020-00262-3](https://doi.org/10.1007/s10589-020-00262-3).
- [24] J. Lindenstrauss, A short proof of Liapounoff’s convexity theorem, *Journal of Mathematics and Mechanics* 15 (1966), 971–972.
- [25] A. A. Lyapunov, On completely additive vector functions, *Izv. Akad. Nauk SSSR* 4 (1940), 465–478.
- [26] P. Manns, Relaxed multibang regularization for the combinatorial integral approximation, *SIAM Journal on Control and Optimization* 59 (2021), 2645–2668, doi:[10.1137/20m1377187](https://doi.org/10.1137/20m1377187).

- [27] P. Manns and C. Kirches, Improved regularity assumptions for partial outer convexification of mixed-integer PDE-constrained optimization problems, *ESAIM: Control, Optimisation and Calculus of Variations* 26 (2020), 32, [doi:10.1051/cocv/2019016](https://doi.org/10.1051/cocv/2019016).
- [28] P. Manns and C. Kirches, Multidimensional sum-up rounding for elliptic control systems, *SIAM Journal on Numerical Analysis* 58 (2020), 3427–3447, [doi:10.1137/19m1260682](https://doi.org/10.1137/19m1260682).
- [29] P. Manns, C. Kirches, and F. Lenders, Approximation properties of sum-up rounding in the presence of vanishing constraints, *Mathematics of Computation* 90 (2021), 1263–1296.
- [30] S. Sager, *Numerical Methods for Mixed-integer Optimal Control Problems*, Der Andere Verlag Lübeck, 2005.
- [31] S. Sager, H. G. Bock, and M. Diehl, The integer approximation error in mixed-integer optimal control, *Mathematical Programming* 133 (2012), 1–23, [doi:10.1007/s10107-010-0405-3](https://doi.org/10.1007/s10107-010-0405-3).
- [32] S. Sager, M. Jung, and C. Kirches, Combinatorial integral approximation, *Mathematical Methods of Operations Research* 73 (2011), 363–380, [doi:10.1007/s00186-011-0355-4](https://doi.org/10.1007/s00186-011-0355-4).
- [33] M. Sharma, M. Hahn, S. Leyffer, L. Ruthotto, and B. van Bloemen Waanders, Inversion of convection–diffusion equation with discrete sources, *Optimization and Engineering* 22 (2021), 1419–1457, [doi:10.1007/s11081-020-09536-5](https://doi.org/10.1007/s11081-020-09536-5).
- [34] L. Tartar, Compensated compactness and applications to partial differential equations, in *Nonlinear Analysis and Mechanics: Heriot–Watt Symposium*, volume 4, 1979, 136–212.
- [35] P. L. Toint, Non-monotone trust-region algorithms for nonlinear optimization subject to convex constraints, *Mathematical Programming* 77 (1997), 69–94, [doi:10.1007/bf02614518](https://doi.org/10.1007/bf02614518).
- [36] R. H. Vogt, S. Leyffer, and T. Munson, A mixed-integer PDE-constrained optimization formulation for electromagnetic cloaking, *SIAM Journal on Scientific Computing* 44 (2022), B29–B50, [doi:10.1137/20m1315993](https://doi.org/10.1137/20m1315993).
- [37] J. Yu and M. Anitescu, Multidimensional sum-up rounding for integer programming in optimal experimental design, *Mathematical Programming* 185 (2021), 37–76, [doi:10.1007/s10107-019-01421-z](https://doi.org/10.1007/s10107-019-01421-z).
- [38] C. Zeile, N. Robuschi, and S. Sager, Mixed-integer optimal control under minimum dwell time constraints, *Mathematical Programming* 188 (2021), 653–694, [doi:10.1007/s10107-020-01533-x](https://doi.org/10.1007/s10107-020-01533-x).

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory (“Argonne”). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan <http://energy.gov/downloads/doe-public-access-plan>.