SIAM J. Optim.
Vol. 30, No. 3, pp. 2163–2196

A NONSMOOTH TRUST-REGION METHOD FOR LOCALLY LIPSCHITZ FUNCTIONS WITH APPLICATION TO OPTIMIZATION PROBLEMS CONSTRAINED BY VARIATIONAL INEQUALITIES[*]

CONSTANTIN CHRISTOF[†], JUAN CARLOS DE LOS REYES[‡], AND CHRISTIAN MEYER[§]

Abstract. We propose a general trust-region method for the minimization of nonsmooth and nonconvex, locally Lipschitz continuous functions that can be applied, e.g., to optimization problems constrained by elliptic variational inequalities. The convergence of the considered algorithm to C-stationary points is verified in an abstract setting and under suitable assumptions on the involved model functions. For a special instance of a variational inequality constrained problem, we are able to properly characterize the Bouligand subdifferential of the reduced cost function, and, based on this characterization result, we construct a computable trust-region model which satisfies all hypotheses of our general convergence analysis. The article concludes with numerical experiments that illustrate the main properties of the proposed algorithm.

Key words. trust-region methods, Bouligand subdifferential, optimal control, nonsmooth optimization, stationarity conditions, optimization with variational inequality constraints

AMS subject classifications. 35J87, 49J40, 49J52, 65K05, 90C26, 90C56

DOI. 10.1137/18M1164925

1. Introduction.

The aim of this paper is to study trust-region methods for the minimization of locally Lipschitz continuous, potentially nonsmooth and nonconvex functions $f : \mathbb{R}^n \to \mathbb{R}$ with a particular focus on optimization problems constrained by elliptic variational inequalities (VIs) of the second kind. We establish the convergence to C-stationary points of a general trust-region algorithm that couples an inaccurate but simple local model to a detailed but complicated nonlocal model of the objective function (see Algorithm 2.5), and we apply the resulting method to a discretized optimal control problem that is governed by an elliptic VI of the second kind involving the $\|\cdot\|_1$-norm. For the main results of this paper, see Theorem 2.12, Theorem 5.16, and Corollary 4.8, and see also the numerical experiments in section 6.

Let us put our work into perspective: Since VIs can be used to model processes and phenomena in various fields and application areas (see, e.g., [12, 13, 14, 16, 25, 27, 30, 38, 39, 42, 43, 46, 58, 59, 60] for examples ranging from machine learning and finance to contact mechanics and fluid dynamics), there is a natural interest in the analysis and numerical solution of optimization problems with VI-constraints, i.e., of problems whose reduced objective functions have the form $f(\cdot) = J(S(\cdot), \cdot)$ with a C^1-map $J : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$, $m, n \in \mathbb{N}$, and the solution operator $S : \mathbb{R}^n \to \mathbb{R}^m$ of a given elliptic VI of the first or the second kind. The study of this type of mathematical programming problem, however, turns out to be a challenging topic.

[*] Received by the editors January 12, 2018; accepted for publication (in revised form) May 14, 2020; published electronically August 12, 2020.
https://doi.org/10.1137/18M1164925
Funding: This research was supported by the German Research Foundation (DFG) under grant ME 3281/7-1 within the priority program "Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization" (SPP 1962).

[†] Technical University of Munich, Chair of Optimal Control, Center for Mathematical Sciences, M17, 85748 Garching, Germany (christof@ma.tum.de).

[‡] Escuela Politécnica Nacional, Research Center on Mathematical Modelling (MODEMAT), E11-253 Quito, Ecuador (juan.delosreyes@epn.edu.ec).

[§] Technical University of Dortmund, Faculty of Mathematics, Chair of Optimal Control, LSX, 44227 Dortmund, Germany (christian.meyer@math.tu-dortmund.de).

Due to the nonsmoothness induced by the inner, nondifferentiable map S, for example, classical optimality conditions become meaningless and various alternative concepts of stationarity arise (e.g., Clarke, Dini, Bouligand, and Mordukhovich stationarity), each with its own advantages and shortcomings. For an overview of the different types of stationarity in the nonsmooth setting and the related notions of subdifferential, see, for instance, [74] and the references therein.

Further (and even more severe) difficulties emerge due to the particular form of the cost function $f(\cdot) = J(S(\cdot), \cdot)$ in a VI-constrained optimization problem. Since the nonsmooth map S appears in the argument of the C^1-function J, and since S is itself only implicitly defined by the governing VI, the resulting objective f is in general hard to analyze and often has very poor structural properties. In particular, it can be shown (see section 3) that an f of the form $f(\cdot) = J(S(\cdot), \cdot)$ cannot be expected to be convex, regular in the sense of Clarke [19, Definition 2.3.4], or lower/upper-C^1 in the sense of Spingarn [72, Definition 10.29]. (Recall that, in the finite-dimensional setting, the concept of lower-C^1 regularity is also often referred to as approximate convexity; cf. [23].) Moreover, establishing properties like upper/lower semismoothness or semi-/quasi-differentiability for the reduced objective function of a VI-constrained optimization problem typically requires appropriate constraint qualifications which may be hard to check in practical applications. See, e.g., [66, Chapters 5 and 6], [8, section 3], and [5, section 2] for some results and more details on this topic. The potential lack of all these properties is a major issue since they are essential for the convergence analysis of various classical nonsmooth optimization algorithms. Compare, for instance, with the analysis of bundle methods in [8, 29, 34, 35, 36, 41, 45, 48, 52, 62, 63, 77, 78] in this context, which require convexity, lower-C^1 regularity, or, in the presence of a globalization by means of a line-search in the spirit of [47], semismoothness, or semidifferentiability, respectively, with the results for subgradient-oriented descent methods in [4, 5, 33, 54, 64], which are based on the approximability of the ε-subdifferential, upper-C^1 regularity, semismoothness, or the concept of quasi-differentiability, respectively, with the trust-region algorithms in [2, 20, 28, 56, 65, 69, 76], which rely, among other things, on the assumptions of upper-C^1 regularity, Clarke regularity, or B-differentiability, respectively, and with the alternative approaches in [31, 32, 49, 50, 51, 53, 57, 67, 73]. For an overview of nonsmooth optimization algorithms and their history, see also [3, 40, 55, 79, 81].

The easiest way to resolve the above difficulties with the poor structural properties of the objective function f in the presence of VI-constraints is to simply replace the inner, nonsmooth map S with a differentiable approximation, and to subsequently study the resulting smooth optimization problem instead of the original nonsmooth one. Such relaxation approaches have been considered, e.g., in [6, 7, 24, 37, 46, 82] with various different outcomes concerning the resulting stationarity conditions and solution algorithms. However, the main drawback of regularization techniques is that removing or relaxing the nonsmoothness of the map S necessarily alters the structure of the problem and may thus lead to unphysical or unsatisfactory solutions. This effect is in particular undesirable in situations where the nonsmoothness arises from an underlying physical model or is introduced to promote special effects such as, e.g., sparsity. Compare, for instance, with the applications in non-Newtonian fluid dynamics, mechanics, or image processing in [12, 27, 38, 42, 43, 46, 58] in this regard.

Nonsmooth optimization algorithms that do not rely on a regularization and can be proved to converge to stationary points under assumptions that are verifiable for an objective of the form $f(\cdot) = J(S(\cdot), \cdot)$ (e.g., Lipschitz continuity or directional differentiability) are comparatively rare in the literature and typically require an excessive amount of computational effort in each iteration. Compare, for instance, with the gradient sampling methods of [10, 11, 21, 22, 44] in this context, which converge with probability one to C-stationary points of functions, that are differentiable almost everywhere, but typically require the storage/calculation of $n+1$ gradients of the objective function in each iteration, or the trust-region method in [1], which needs a sufficiently accurate approximation of the ε-subdifferential around each iterate.

The aim of this paper is to construct a trust-region framework that is suitable for the solution of VI-constrained optimization problems and allows reduction of the number of iterations, in which an expensive optimization strategy comparable to the procedures in [1, 10, 11, 21, 22, 44] has to be used, to a minimum. The main idea of our approach is to couple a simple, local model of the objective function, which only requires a single subgradient evaluation, to a more detailed, nonlocal model, which is only computed when absolutely necessary. This structure makes it possible to exploit the frequently made observation that, in the trust-region context, simple models only fail in exceptionally rare situations, while still allowing for a rigorous convergence analysis (cf. the results of [26] and the numerical results in section 6). An essential and novel feature of our trust-region method is a special coupling of the two involved models of the objective function via an acceptance criterion that allows us to prove the C-stationarity of the accumulation points of the produced sequence of iterates by means of a Cauchy point argumentation. For further details on this topic, see the comments in sections 2 and 3. The structure of the paper is as follows.

After this introduction, we briefly comment on the notation that we use in our analysis. Note that the list of symbols collected in subsection 1.1 is not exhaustive. Additional notation is introduced wherever necessary in this paper and, for the sake of readability, defined where it first appears in the text.

In section 2, we present our trust-region algorithm in a general setting, i.e., in the context of minimization problems with locally Lipschitz continuous cost functions f. Here, we introduce in particular the already mentioned coupling of the two involved trust-region models and discuss which properties the "inner" detailed model of the objective has to have to ensure the C-stationarity of accumulation points of the produced sequence of iterates. The approach that we take in this section is an axiomatic one similar to that in [2, 62, 65]. We thus do not propose a particular model but collect conditions that have to be satisfied for our convergence results to hold (see Assumption 2.4). We would like to emphasize that the only assumption on f in section 2 is that of local Lipschitz continuity. This is a major difference from the analysis of, e.g., [5, 28, 35], where f is supposed to have further properties.

Section 3 contains comments on the existence and the construction of model functions with the properties in section 2 as well as additional remarks on the relationship of our approach to known results. Here, we prove in particular that every locally Lipschitz function possesses a model which satisfies the assumptions of our general convergence analysis, and that, for an upper-C^1 function f, such a model can be constructed with a single Bouligand subgradient. (Note that the latter observation has already been made in the context of strict models in [2, Theorem 2].) Section 3 concludes with two examples which demonstrate that simple, local models are not sufficient to construct a trust-region algorithm that is guaranteed to converge to C-stationarity when the objective is not upper-C^1, and that the objective of a VI-constrained problem typically lacks Clarke, upper-C^1, and lower-C^1 regularity.

In section 4, we turn our attention to composite objective functions of the form $f(\cdot) = J(S(\cdot), \cdot)$ with a C^1-map $J : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$, $m, n \in \mathbb{N}$, and a directionally differentiable and locally Lipschitz continuous operator $S : \mathbb{R}^n \to \mathbb{R}^m$. (Note that this setting covers in particular the VI-constrained case.) The main result of this section, Corollary 4.8, establishes that, for a function of the form $f(\cdot) = J(S(\cdot), \cdot)$, a model satisfying the assumptions of the general convergence analysis of section 2 can be constructed from an upper estimate of the ε-Bouligand differential of S at the current iterate; cf. the approaches in [1, 54].

In section 5, we apply the findings of section 4 to an optimal control problem governed by a VI of the second kind involving the $\|\cdot\|_1$-norm. The model function that we construct for this example is based on a precise characterization of the Bouligand differential of the control-to-state mapping (see Theorem 5.5) and can be evaluated numerically by solving a combinatorial auxiliary program that depends on the biactive and active set of the current iterate.

Section 6 finally contains numerical experiments that illustrate the main properties of the trust-region method developed in section 5. In this section, we also give some further remarks on implementation details and open questions.

1.1. Some notation. By λ^n, we denote the n-dimensional Lebesgue measure. Moreover, $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ denote the Euclidean norm and scalar product, and $\|\cdot\|_\infty$ and $\|\cdot\|_1$ stand for the maximum- and 1-norm, respectively. In addition, $\|\cdot\|_{\mathbb{R}^{n\times n}}$ denotes the spectral norm. We suppress the index $\mathbb{R}^{n\times n}$ if no ambiguity is possible. Given $x \in \mathbb{R}^n$ and $r > 0$, we denote by $B_r(x)$ the *closed* ball around x with radius r.

2. A nonsmooth trust-region algorithm.

In this section and the next, we discuss and analyze our trust-region approach in the context of general, nonlinear optimization problems of the form

$$\text{(P)} \qquad \min_{x\in\mathbb{R}^n} f(x)$$

with an objective function f satisfying the following assumption.

Assumption 2.1 (standing assumptions on the objective function). *The function $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz continuous (in the sense of* [19, section 1.2]*).*

The following definition collects some basic concepts of nonsmooth optimization that are going to be used throughout the paper.

Definition 2.2 (generalized differentials and subgradients). *Let $F : \mathbb{R}^n \to \mathbb{R}^m$, $m, n \in \mathbb{N}$, be locally Lipschitz continuous, and let $\mathcal{D}_F$ denote the set of points where F is differentiable. By Rademacher's theorem, it holds that $\lambda^n(\mathbb{R}^n \setminus \mathcal{D}_F) = 0$. For a given $x \in \mathbb{R}^n$, we define*

- *the* Bouligand (generalized) differential *by*

$$\partial_B F(x) := \{G \in \mathbb{R}^{m\times n} : \exists\,\{x_k\} \subset \mathcal{D}_F \text{ with } x_k \to x,\ F'(x_k) \to G\},$$

- *the* Clarke (generalized) *differential by $\partial F(x) := \mathrm{cl}\,(\mathrm{conv}(\partial_B F(x)))$. Here,* cl *is short for closure and* conv *denotes the convex hull.*

In the scalar case $m = 1$, we refer to $\partial_B F(x)$ and $\partial F(x)$ as the Bouligand and the Clarke subdifferential, respectively.

It is well known that, for a scalar, locally Lipschitz continuous $f : \mathbb{R}^n \to \mathbb{R}$, the Clarke subdifferential can equivalently be expressed as

$$\text{(2.1)} \qquad \partial f(x) = \{g \in \mathbb{R}^n : \langle g, v\rangle \le f^\circ(x; v) \quad \forall v \in \mathbb{R}^n\},$$

where f° denotes Clarke's generalized directional derivative; see [19, section 2.1]. For scalar functions, we moreover have the following notion of stationarity.

Definition 2.3 (Clarke stationarity). *Let $f : \mathbb{R}^n \to \mathbb{R}$, $n \in \mathbb{N}$, be locally Lipschitz continuous. Then a point $\bar{x} \in \mathbb{R}^n$ is called C(larke)-stationary if* $0 \in \partial f(\bar{x})$.

As already mentioned in the introduction, the main idea of our trust-region method is to use two models of the objective f, a simple one and a detailed one. To formulate the latter, we need a suitably chosen model function, whose existence is assumed as a start. In sections 3 to 5, we will see how to construct such a function for concrete problems. Our precise assumptions are as follows.

Assumption 2.4 (existence of an oracle and a model function).

1. *For every $x \in \mathbb{R}^n$, we can calculate a subgradient $g \in \partial f(x)$.*
2. *We are given a model function $\phi : \mathbb{R}^n \times \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}$ which satisfies the following:*
   (a) *For every $(x, \Delta) \in \mathbb{R}^n \times \mathbb{R}^+$, the mapping $\mathbb{R}^n \ni d \mapsto \phi(x, \Delta; d)$ is positively homogeneous and lower semicontinuous.*
   (b) Stationarity indicator property: *If $\{(x_k, \Delta_k)\} \subset \mathbb{R}^n \times \mathbb{R}^+$ is a sequence with $x_k \to x$, $\Delta_k \to 0$, and $\psi(x_k, \Delta_k) \to 0$, where ψ is defined by*

$$\psi(x, \Delta) := -\min_{\|d\| \le 1} \phi(x, \Delta; d) \ge 0, \tag{2.2}$$

   then the limit point x is Clarke-stationary, i.e., it holds that $0 \in \partial f(x)$.
   (c) Remainder term property: *For all $\{(x_k, \Delta_k)\} \subset \mathbb{R}^n \times \mathbb{R}^+$ satisfying*

$$x_k \to x, \quad \Delta_k \to 0, \quad \textit{and} \quad \lim_{k \to \infty} \psi(x_k, \Delta_k) > 0$$

   with ψ as defined in (2.2), *it holds that*

$$\limsup_{k \to \infty} \sup_{d \in B_{\Delta_k}(0)} \frac{f(x_k + d) - f(x_k) - \phi(x_k, \Delta_k; d)}{\Delta_k} \le 0. \tag{2.3}$$

Note that the non-negativity of the "stationarity measure" ψ in (2.2) follows immediately from the positive homogeneity and the lower semicontinuity of $\phi(x, \Delta; \cdot)$. It is moreover easy to see that the limit superior in (2.3) is actually a limit since the inner supremum is always greater than or equal to zero (just choose $d = 0 \in B_{\Delta_k}(0)$).

We would like to point out that the conditions on ϕ in Assumption 2.4 are closely related to the concept of "strict first-order model" used, e.g., in [2, 62, 64, 65]. To be more precise, it is easy to check that [2, Condition $(\widehat{M_2})$] is sufficient for the remainder term property in (2c) to hold. This property reflects that the model function typically has to incorporate certain information about the objective f in a neighborhood of the current iterate; cf. example (3.4). Note that, in contrast to [2, 62, 64, 65], we do not assume that ϕ is convex or that the subdifferential of the map $z \mapsto \phi(x, \Delta; z)$ is in a certain way related to the subdifferential of f (cf. [2, Definition 1]). Instead, we require positive homogeneity. A further difference from [2] is that, in Assumption 2.4, the model ϕ is allowed to depend explicitly on Δ. For more details on the relationship of our trust-region framework to the approaches of [2, 62, 64, 65], see section 3.

Given a model function ϕ satisfying the conditions in Assumption 2.4, our trust-region algorithm for the solution of (P) reads as follows.

Algorithm 2.5 (nonsmooth trust-region algorithm).

```
1: Choose constants $\Delta_{\min} > 0$, $0 < \eta_1 < \eta_2 < 1$, $0 < \beta_1 < 1 < \beta_2$, $0 < \mu \le 1$, an
   initial value $x_0 \in \mathbb{R}^n$, and an initial TR-radius $\Delta_0 > \Delta_{\min}$. Set $k := 0$.
2: for $k = 0, 1, 2, \dots$ do
```

```
Choose a subgradient $g_k \in \partial f(x_k)$ and a matrix $H_k \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$.
if $g_k = 0$ then
    STOP the iteration, $x_k$ is C-stationary, i.e., $0 \in \partial f(x_k)$.
else
    if $\Delta_k \geq \Delta_{\min}$ then
        Compute an inexact solution $d_k$ of the trust-region subproblem

        (Q_k)   $\begin{cases} \min\limits_{d\in\mathbb{R}^n} & q_k(d) := f(x_k) + \langle g_k, d\rangle + \frac{1}{2}\, d^\top H_k d \\ \text{s.t.} & \|d\| \leq \Delta_k \end{cases}$

        that satisfies the generalized Cauchy decrease condition

        (2.4)   $f(x_k) - q_k(d_k) \geq \frac{\mu}{2}\, \|g_k\| \, \min\Big\{\Delta_k, \frac{\|g_k\|}{\|H_k\|}\Big\}.$

        Compute the quality indicator

                $\rho_k := \frac{f(x_k) - f(x_k + d_k)}{f(x_k) - q_k(d_k)}.$

    else
        Compute an inexact solution $d_k$ of the modified trust-region subproblem

        ($\tilde{Q}_k$)   $\begin{cases} \min\limits_{d\in\mathbb{R}^n} & \tilde{q}_k(d) := f(x_k) + \phi(x_k, \Delta_k; d) + \frac{1}{2}\, d^\top H_k d \\ \text{s.t.} & \|d\| \leq \Delta_k \end{cases}$

        that satisfies the modified Cauchy decrease condition

        (2.5)   $f(x_k) - \tilde{q}_k(d_k) \geq \frac{\mu}{2}\, \psi(x_k, \Delta_k) \, \min\Big\{\Delta_k, \frac{\psi(x_k, \Delta_k)}{\|H_k\|}\Big\},$

        where $\psi(x_k, \Delta_k)$ is defined as in (2.2).
        Compute the modified quality indicator

        (2.6)   $\rho_k := \begin{cases} \frac{f(x_k) - f(x_k + d_k)}{f(x_k) - \tilde{q}_k(d_k)} & \text{if } \psi(x_k, \Delta_k) > \|g_k\|\,\Delta_k \\ 0 & \text{if } \psi(x_k, \Delta_k) \leq \|g_k\|\,\Delta_k. \end{cases}$

    end if
    Update: Set

        $x_{k+1} := \begin{cases} x_k & \text{if } \rho_k \leq \eta_1 \quad \text{(null step)}, \\ x_k + d_k & \text{otherwise} \quad \text{(successful step)}, \end{cases}$

        $\Delta_{k+1} := \begin{cases} \beta_1\, \Delta_k & \text{if } \rho_k \leq \eta_1, \\ \max\{\Delta_{\min}, \Delta_k\} & \text{if } \eta_1 < \rho_k \leq \eta_2, \\ \max\{\Delta_{\min}, \beta_2 \Delta_k\} & \text{if } \rho_k > \eta_2. \end{cases}$

    Set $k := k + 1$.
end if
```

```
16: end for
```

Remark 2.6 (Bouligand stationarity). If elements of the Bouligand subdifferential are chosen as g_k in Step 3 (as done, e.g., in section 6), then the termination criterion in Step 4 amounts to the stationarity condition $0 \in \partial_B f(x_k)$, which is stronger than Clarke stationarity; cf. the discussion in [17, Remark 4.8].

Remark 2.7 (comparison with other nonsmooth trust-region methods). The essential differences between Algorithm 2.5 and other nonsmooth trust-region methods such as, for instance, the ones presented in [2, 20, 28, 56, 65, 69, 76] are the following:

- By introducing the distinction of cases in Step 7, we allow for a classical trust-region subproblem, which is easy to solve by standard methods such as the dogleg method (see [68]) or Steihaug's CG-method (see [80]). As our numerical experiments in section 6 demonstrate, in the applications we are interested in, cases in which the complicated model in $(\tilde{\mathrm{Q}}_k)$ has to be used are exceptionally rare. Our algorithm therefore accounts for the fact that many nonsmooth problems can well be solved by classical trust-region methods and only switches to complicated model functions if it is absolutely necessary.
- Another essential feature of our algorithm, which ensures the convergence of the overall method, is the computation of the quality indicator in Step 12. It basically corresponds to a comparison of the "easy" and the complicated models weighted with the trust-region radius. Note in this context that both the simple and the complicated models in Algorithm 2.5 are, on their own, not able to reliably identify C-stationary points. Since the simple model only uses one subgradient at the current iterate x_k, it is typically unable to detect the case $0 \in \partial f(x_k)$, and since the model ϕ in (2.2) is only assumed to provide a meaningful stationarity measure for $\Delta \to 0$, it may become arbitrarily inaccurate if the trust-region radius does not tend to zero. In Algorithm 2.5, this issue is resolved by the comparison with the local "easy" model in Step 12. The coupling in Step 12 is thus indeed crucial for our trust-region framework.

Note that the modified trust-region subproblem $(\tilde{\mathrm{Q}}_k)$ admits an optimal solution due to the lower semicontinuity of ϕ w.r.t. the last argument by Assumption 2.4(2a). Moreover, it is always possible to find inexact solutions to (Q_k) and $(\tilde{\mathrm{Q}}_k)$ that fulfill the respective Cauchy decrease conditions.

LEMMA 2.8. *Global minimizers of* (Q_k) *and* $(\tilde{\mathrm{Q}}_k)$ *satisfy the respective Cauchy decrease conditions in* (2.4) *and* (2.5) *for every* $0 < \mu \leq 1$.

Proof. Since our model function ϕ is assumed to be positively homogeneous, we can argue as in [69, Lemma 3.2], which immediately gives the assertion. □

2.1. Convergence analysis. In what follows, we show that every accumulation point of the sequence of iterates generated by Algorithm 2.5 is C-stationary as defined in Definition 2.3. For this purpose, we need the following.

ASSUMPTION 2.9 (standing assumptions on H_k). *The matrices* $H_k \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$ *from Step* 3 *of Algorithm* 2.5 *satisfy* $\|H_k\| \leq C_H$ *for all* $k \in \mathbb{N}_0$ *with a constant* $C_H > 0$.

PROPOSITION 2.10. *Assume that Algorithm* 2.5 *does not terminate in finitely many steps, and suppose that* $\{x_{k_l}\}$ *is a subsequence of the sequence of iterates* $\{x_k\}$ *satisfying* $x_{k_l} \to \bar{x}$ *and* $\Delta_{k_l} \to 0$ *for* $l \to \infty$ *and some* $\bar{x} \in \mathbb{R}^n$. *Then it holds that* $0 \in \partial f(\bar{x})$.

Proof. Since $\Delta_{k_l} \to 0$, there exists an $L \in \mathbb{N}$ such that $\Delta_{k_l} < \beta_1 \Delta_{\min}$ for all $l \geq L$. This is only possible if the iterations $k_l - 1$, $l \geq L$, are all null steps, i.e., for all $l \geq L$, we have $x_{k_l - 1} = x_{k_l}$, $\Delta_{k_l} = \beta_1 \Delta_{k_l - 1} < \beta_1 \Delta_{\min}$, and $\rho_{k_l - 1} \leq \eta_1 < 1$. In particular, $x_{k_l - 1} \to \bar{x}$ and $\Delta_{k_l - 1} \to 0$. We next show $\psi(x_{k_l - 1}, \Delta_{k_l - 1}) \to 0$. Once this is established, the assertion follows immediately from Assumption 2.4(2b). To this end, we argue by contradiction and assume that there is an $\varepsilon > 0$ such that

$$\limsup_{l \to \infty} \psi(x_{k_l - 1}, \Delta_{k_l - 1}) \geq \varepsilon. \tag{2.7}$$

Let $\{(x_m, \Delta_m)\}_{m \in M}$ be a subsequence of $\{(x_{k_l - 1}, \Delta_{k_l - 1})\}$ which attains the limit superior in (2.7). Then, for $m \in M$ sufficiently large, we have $\psi(x_m, \Delta_m) \geq \varepsilon/2$. Since, in addition, the local Lipschitz continuity of f and $x_m \to \bar{x}$ for $M \ni m \to \infty$ imply that $\|g_m\| \leq L(\bar{x})$ holds for all large m, where $L(\bar{x})$ denotes the local Lipschitz constant of f at $\bar{x}$, the convergence $\Delta_m \to 0$ yields $\psi(x_m, \Delta_m) > \|g_m\| \, \Delta_m$ for $m \in M$ sufficiently large. Therefore, the first case in (2.6) applies in the computation of the quality indicator. Moreover, the modified Cauchy decrease condition in (2.5) and (2.2) imply $f(x_m) - \tilde{q}_m(d_m) > 0$. Thus, for all sufficiently large $m \in M$, we obtain

$$\begin{aligned}
\rho_m &= 1 - \frac{f(x_m + d_m) - f(x_m) - \phi(x_m, \Delta_m; d_m) - \frac{1}{2} d_m^\top H_m d_m}{f(x_m) - \tilde{q}_m(d_m)} \\
&\geq 1 - \frac{\sup_{d \in B_{\Delta_m}(0)} \big(f(x_m + d) - f(x_m) - \phi(x_m, \Delta_m; d) \big) + \frac{1}{2} \, C_H \, \Delta_m^2}{f(x_m) - \tilde{q}_m(d_m)} \\
&\geq 1 - \frac{\sup_{d \in B_{\Delta_m}(0)} \big(f(x_m + d) - f(x_m) - \phi(x_m, \Delta_m; d) \big) + \frac{1}{2} \, C_H \, \Delta_m^2}{\frac{\mu}{4} \, \varepsilon \, \min \left\{ \Delta_m, \frac{\varepsilon}{2 C_H} \right\}},
\end{aligned}$$

where we used the decrease condition (2.5) and $\psi(x_m, \Delta_m) \geq \varepsilon/2$ for the last estimate. From Assumption 2.4(2c), it now follows that

$$\begin{aligned}
\liminf_{M \ni m \to \infty} \rho_m \geq 1 &- \frac{2}{\mu \, \varepsilon} \, C_H \lim_{M \ni m \to \infty} \Delta_m \\
&- \frac{4}{\mu \, \varepsilon} \limsup_{M \ni m \to \infty} \sup_{d \in B_{\Delta_m}(0)} \frac{f(x_m + d) - f(x_m) - \phi(x_m, \Delta_m; d)}{\Delta_m} \geq 1,
\end{aligned}$$

which contradicts $\rho_{k_l - 1} \leq \eta_1 < 1$. Therefore, (2.7) is not true, which, together with the nonnegativity of ψ, results in

$$0 \leq \liminf_{l \to \infty} \psi(x_{k_l - 1}, \Delta_{k_l - 1}) \leq \limsup_{l \to \infty} \psi(x_{k_l - 1}, \Delta_{k_l - 1}) = 0.$$

This yields the desired convergence of $\psi(x_{k_l - 1}, \Delta_{k_l - 1})$ and proves the claim. ◻

LEMMA 2.11. *Assume that Algorithm* 2.5 *does not terminate in finitely many steps. If the sequence of iterates* $\{x_k\}$ *admits an accumulation point, then the sequence of function values* $\{f(x_k)\}$ *converges to some* $\bar{f} \in \mathbb{R}$.

Proof. By construction, the sequence $\{f(x_k)\}$ is monotonically decreasing so that $f(x_k) \to \bar{f} \in \mathbb{R} \cup \{-\infty\}$. If a subsequence $\{x_{k_l}\}$ converges to a point $\bar{x} \in \mathbb{R}^n$, then the continuity of f implies $\bar{f} = f(\bar{x}) > -\infty$, which yields the claim. ◻

THEOREM 2.12. *Assume that Algorithm* 2.5 *does not terminate in finitely many steps. Then every accumulation point of the sequence of iterates is C-stationary.*

Proof. If the number of successful steps is finite, then there is an $N \in \mathbb{N}$ such that all iterations $k \geq N$ are null steps. According to the update rules for null steps this implies $x_k \to x_N =: \bar{x}$, $\Delta_k \to 0$, and, ultimately, $0 \in \partial f(\bar{x})$ by Proposition 2.10.

We may thus focus on the case where there are infinitely many successful steps. Let $\bar{x}$ be an arbitrary accumulation point of the sequence of iterates and let $\{x_{k_l}\}$ be a subsequence with $x_{k_l} \to \bar{x}$. We assume w.l.o.g. that the iterations k_l are all successful (else, we shift the index forth to the next successful iteration, which does not change the sequence due to the update rule for null steps). Since the iterations k_l are all successful, the monotonicity of $\{f(x_k)\}$, (2.4), and (2.5) imply

$$f(x_{k_l}) - f(x_{k_{l+1}}) \geq f(x_{k_l}) - f(x_{k_l+1}) \geq \eta_1 \frac{\mu}{2}\, \nu(x_{k_l}, \Delta_{k_l}) \min\left\{\Delta_{k_l}, \frac{\nu(x_{k_l}, \Delta_{k_l})}{C_H}\right\} \tag{2.8}$$

with

$$\nu(x_{k_l}, \Delta_{k_l}) := \begin{cases} \|g_{k_l}\| & \text{if } \Delta_{k_l} \geq \Delta_{\min}, \\ \psi(x_{k_l}, \Delta_{k_l}) & \text{if } \Delta_{k_l} < \Delta_{\min}. \end{cases} \tag{2.9}$$

Due to the convergence of $\{f(x_k)\}$ and the nonnegativity of $\nu(x_{k_l}, \Delta_{k_l})$, the above can only be true if the right-hand side of (2.8) tends to zero. This, in turn, yields

$$\min\{\Delta_{k_l}, \nu(x_{k_l}, \Delta_{k_l})\} \to 0 \tag{2.10}$$

for $l \to \infty$. We now distinguish between three cases: If there exists a subsequence of $\{x_{k_l}\}$ (unrelabeled for simplicity) such that the associated trust-region radii Δ_{k_l} converge to zero, then the claim follows immediately from Proposition 2.10. In this case, there is nothing left to prove. If, on the other hand, there exists a subsequence of $\{x_{k_l}\}$ (again unrelabeled) with $\Delta_{k_l} \geq \Delta_{\min}$, then (2.9) and (2.10) imply $\|g_{k_l}\| \to 0$. In view of [19, Proposition 2.1.5(b)], this yields $0 \in \partial f(\bar{x})$ and the claim again follows. If, lastly, there exists a subsequence of $\{x_{k_l}\}$ (again unrelabeled) with $\varepsilon \leq \Delta_{k_l} < \Delta_{\min}$ for some $\varepsilon > 0$, then (2.10) gives $\nu(x_{k_l}, \Delta_{k_l}) \to 0$. We know, however, that the steps $\{x_{k_l}\}$ are all successful and, according to (2.6), this is only the case if

$$\nu(x_{k_l}, \Delta_{k_l}) = \psi(x_{k_l}, \Delta_{k_l}) \geq \|g_{k_l}\| \Delta_{k_l} \geq \|g_{k_l}\|\, \varepsilon \geq 0.$$

Accordingly, $\|g_{k_l}\| \to 0$ holds and we can argue as in the second case to deduce that $\bar{x}$ is C-stationary. This completes the proof. □

Remark 2.13. The proofs of Proposition 2.10 and Theorem 2.12 show not only that every accumulation point is C-stationary but also that, for every convergent subsequence $\{x_{k_l}\}$, the *stationarity indicator* $\min\{\|g_{k_l}\|, \psi(x_{k_l}, \Delta_{k_l})\}$ converges to zero. This lays the foundation for an implementable termination criterion of the form $\min\{\|g_{k_l}\|, \psi(x_{k_l}, \Delta_{k_l})\} \leq \texttt{tol}$ with a given tolerance $\texttt{tol} > 0$.

3. Comments on the model function ϕ. The aim of this section is to discuss in more detail why the inner model ϕ in Algorithm 2.5 is necessary, why ϕ typically has to incorporate information about the objective function f in a neighborhood of the current iterate, and how suitable functions ϕ can be constructed for particular problem classes. We begin by proving that every locally Lipschitz function f possesses a model which satisfies the conditions in Assumption 2.4. Note that this observation is already remarkable since a comparable existence result is, at least to the best of the authors' knowledge, currently not available for the concept of "strict model" employed, e.g., in [2, 62, 64, 65], which relies on axioms similar to those used in our framework.

Theorem 3.1 (existence of model functions). *Every locally Lipschitz continuous function $f : \mathbb{R}^n \to \mathbb{R}$ admits a model $\phi : \mathbb{R}^n \times \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}$ with the properties in Assumption 2.4. Further, for each locally Lipschitz continuous $f : \mathbb{R}^n \to \mathbb{R}$, a possible model function is given by*

$$\bar{\phi}(x, \Delta; d) := \begin{cases} \sup\limits_{t \in (0,\Delta]} \left(\dfrac{f\left(x + td/\|d\|\right) - f(x)}{t} \|d\| \right) & \text{if } d \neq 0, \\ 0 & \text{if } d = 0. \end{cases} \tag{3.1}$$

Proof. To prove the claim of the theorem, it suffices to check that the model $\bar{\phi}$ in (3.1) satisfies the conditions in Assumption 2.4. To this end, we first note that the function $\bar{\phi}$ in (3.1) is trivially real-valued due to the local Lipschitz continuity of f. It thus indeed makes sense to write $\bar{\phi} : \mathbb{R}^n \times \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}$. Further, $\bar{\phi}$ is clearly positively homogeneous and, as the pointwise supremum of continuous functions, lower semicontinuous in d, and from the definition of $\bar{\phi}$ it follows straightforwardly that

$$f(x + d) - f(x) - \phi(x, \Delta; d) \leq 0 \qquad \forall d \in B_\Delta(0), \quad \forall x \in \mathbb{R}^n, \quad \forall \Delta \in \mathbb{R}^+.$$

This shows that the conditions (2a) and (2c) in Assumption 2.4 are satisfied by $\bar{\phi}$. It remains to check (2b). So let us assume that a sequence $\{(x_k, \Delta_k)\} \subset \mathbb{R}^n \times \mathbb{R}^+$ and an $x \in \mathbb{R}^n$ with $x_k \to x$, $\Delta_k \to 0$, and

$$\bar{\psi}(x_k, \Delta_k) := - \min_{\|d\| \leq 1} \bar{\phi}(x_k, \Delta_k; d) \to 0$$

are given. Then the definitions of $\bar{\psi}$ and Clarke's generalized directional derivative (see [19, section 2.1]) imply that, for every arbitrary but fixed $d \in B_1(0) \setminus \{0\}$, we have

$$\begin{aligned} 0 &\leq \bar{\psi}(x_k, \Delta_k) + \sup_{t \in (0,\Delta_k]} \left(\frac{f\left(x_k + (t/\|d\|)d\right) - f(x_k)}{t/\|d\|} \right) \\ &\leq \bar{\psi}(x_k, \Delta_k) + \left(\sup_{y \in B_{\|x - x_k\|}(x),\ s \in (0, \Delta_k/\|d\|]} \frac{f(y + sd) - f(y)}{s} \right) \to 0 + f^\circ(x; d). \end{aligned}$$

Thus, $0 \in \partial f(x)$ by (2.1) and the stationarity property in (2b) is indeed satisfied. This completes the proof. ∎

Note that the possibility to vary the model ϕ with the trust-region radius is essential for the construction in (3.1), and that (3.1) typically does not define a strict model in the sense of [2, 62, 64, 65] since $\bar{\phi}$ is in general not convex in d and does not satisfy the condition on the subdifferential in point 1 of [2, Definition 1].

Due to its complicated form, the model (3.1) is, of course, typically not suitable for practical applications. Simpler models satisfying the conditions in Assumption 2.4 can be constructed if additional information about the objective function f is available. To discuss this topic properly, we recall the following concepts from [72, Definition 10.29] and [19, Definition 2.3.4].

Definition 3.2 (subsmoothness and regularity). *Suppose that a locally Lipschitz continuous function $f : \mathbb{R}^n \to \mathbb{R}$ is given. Then the following hold:*

- *f is called Clarke regular if f is directionally differentiable everywhere and if the directional derivative $f'(\cdot;\cdot)$ satisfies $f'(x; h) = f^\circ(x; h)$ for all $x, h \in \mathbb{R}^n$, where f° again denotes Clarke's generalized directional derivative.*

- *f is called lower-C^1 if for every $x \in \mathbb{R}^n$ there exist a neighborhood $U \subset \mathbb{R}^n$, a compact set $K \subset \mathbb{R}^m$, $m \in \mathbb{N}$, and a continuous function $\Theta : U \times K \to \mathbb{R}$ such that the derivative of Θ w.r.t. the first variable exists and is continuous as a map from $U \times K$ to $\mathbb{R}^n$, and such that f satisfies*

$$f(y) = \max_{t \in K} \Theta(y,t) \qquad \forall y \in U.$$

- *f is called upper-C^1 if $-f$ is lower-C^1.*

For an upper-C^1 function, we can prove the following theorem.

Theorem 3.3 (simple subgradient models for upper-C^1 functions). *Suppose that a locally Lipschitz continuous upper-C^1 function $f : \mathbb{R}^n \to \mathbb{R}$ is given, and that g_x is an arbitrary but fixed element of the Bouligand subdifferential $\partial_B f(x)$ for all $x \in \mathbb{R}^n$. Then the model function*

$$\phi : \mathbb{R}^n \times \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}, \qquad (x, \Delta, d) \mapsto \langle g_x, d \rangle \tag{3.2}$$

satisfies the conditions in Assumption 2.4, *and the convergence result in Theorem* 2.12 *holds true for Algorithm* 2.5 *when the inner model ϕ is chosen as in* (3.2).

Proof. Since the function in (3.2) is linear in d, it is clearly positively homogeneous and lower semicontinuous, and since the stationarity measure ψ defined in (2.2) is given by $\psi(x, \Delta) = \|g_x\|$ for the model in (3.2), condition (2b) in Assumption 2.4 follows immediately from [19, Proposition 2.1.5] in the situation of the theorem. It remains to verify condition (2c). To this end, let us assume that a sequence $\{(x_k, \Delta_k)\} \subset \mathbb{R}^n \times \mathbb{R}^+$ satisfying $x_k \to x$, $\Delta_k \to 0$, and $\lim_{k\to\infty} \psi(x_k, \Delta_k) > 0$ for some $x \in \mathbb{R}^n$ is given. Then the upper-C^1 regularity of f implies that we can find a neighborhood $U \subset \mathbb{R}^n$ of x, a compact set $K \subset \mathbb{R}^m$, $m \in \mathbb{N}$, and a map $\Theta : U \times K \to \mathbb{R}$ which is continuous and possesses a continuous first derivative w.r.t. the first variable, such that $f(y) = \min_{t \in K} \Theta(y, t)$ holds for all $y \in U$. Suppose now that $z \in U$ is an element of the set of points $\mathcal{D}_f$, where f is differentiable, and let t_z denote an arbitrary but fixed element of K with $f(z) = \Theta(z, t_z)$. (Note that such a t_z always exists due to the compactness of K and the continuity properties of Θ.) Then the differentiability of the functions f and $\Theta(\cdot, t_z)$ in z and the properties of Θ yield

$$\langle \nabla f(z), h \rangle = \lim_{s \to 0} \frac{f(z + sh) - f(z)}{s} \le \lim_{s \to 0} \frac{\Theta(z + sh, t_z) - \Theta(z, t_z)}{s} = \langle \nabla_1 \Theta(z, t_z), h \rangle$$

for all $h \in \mathbb{R}^n$, where ∇_1 denotes the gradient of Θ w.r.t. the first variable. This shows that, for all $z \in \mathcal{D}_f \cap U$, we have $\nabla f(z) = \nabla_1 \Theta(z, t_z)$. Consider now an arbitrary but fixed sequence $\{z_k\} \subset \mathcal{D}_f$ satisfying $\|x_k - z_k\| + \|\nabla f(z_k) - g_{x_k}\| \le \Delta_k^2$. (Such a sequence exists due to the definition of the Bouligand subdifferential $\partial_B f(x_k)$.) Then the above observations, the local Lipschitz continuity of f, the convergences $x_k \to x$

and $\Delta_k \to 0$, and the fundamental theorem of calculus imply that

$$\begin{aligned}
&\sup_{d \in B_{\Delta_k}(0)} \frac{f(x_k + d) - f(x_k) - \langle g_{x_k}, d\rangle}{\Delta_k} \\
&\qquad = \sup_{d \in B_{\Delta_k}(0)} \frac{f(z_k + d) - f(z_k) - \langle \nabla f(z_k), d\rangle}{\Delta_k} + o(1) \\
&\qquad \leq \sup_{d \in B_{\Delta_k}(0)} \frac{\Theta(z_k + d, t_{z_k}) - \Theta(z_k, t_{z_k}) - \langle \nabla_1 \Theta(z_k, t_{z_k}), d\rangle}{\Delta_k} + o(1) \\
&\qquad \leq \sup_{d \in B_{\Delta_k}(0)} \|\nabla_1 \Theta(z_k + d, t_{z_k}) - \nabla_1 \Theta(z_k, t_{z_k})\| + o(1)
\end{aligned}$$

holds for all large enough k, where the Landau symbol refers to the limit $k \to \infty$. Since z_k converges to x, since Δ_k converges to zero, and since $\nabla_1 \Theta$ is uniformly continuous in the second argument (due to the compactness of K), it now follows immediately that the model ϕ in (3.2) satisfies Assumption 2.4(2c). This proves the claim. □

As the last result shows, under the assumption of upper-C^1 regularity, a simple quadratic model of the objective function constructed from a single Bouligand subgradient is sufficient to guarantee the convergence of Algorithm 2.5. We remark that this observation has already been made, along different lines and in the context of strict models, in [2, Theorem 2].

An important (maybe the most important) point in the analysis of nonsmooth trust-region methods is to realize that, in the absence of upper-C^1 regularity, model functions that only use information about the objective f at the current iterate x_k can, in general, not be expected to yield a globally convergent algorithm. In fact, even the natural candidate for a trust-region model of a locally Lipschitz function, the so-called standard model

$$\tilde{\phi}(x, \Delta; d) := f^\circ(x, d) = \max_{g \in \partial f(x)} \langle g, d\rangle, \tag{3.3}$$

which has been proposed, e.g., in [69, section 4.1], may fail if the objective function is not well behaved or the parameters in the employed algorithm are chosen in a particular way. Consider, for example, the simple one-dimensional objective

$$f : \mathbb{R} \to \mathbb{R}, \quad x \mapsto \max\{-ax, -bx, x - (1 + b)\}, \tag{3.4}$$

where $0 < b < a < \infty$ are given constants. Then this function is trivially convex and piecewise affine with two kinks at $x = 0$ and $x = 1$, and the following holds true.

LEMMA 3.4. *Assume that the parameters in Step* 1 *of Algorithm* 2.5 *satisfy*

$$\begin{aligned}
&\beta_1 + \beta_1\beta_2 < 1, && \eta_1 \geq \Big(\frac{b}{a} - 1\Big)\frac{\beta_1}{\beta_1\beta_2 - 1} + \frac{b}{a}, \\
&x_0 \in \Big(\Big(-1 + \frac{\beta_1\beta_2 - 1}{\beta_1}\Big)^{-1}, 0\Big), && \Delta_0 := \frac{\beta_1\beta_2 - 1}{\beta_1}\, x_0.
\end{aligned} \tag{3.5}$$

Then the sequence of iterates generated by this trust-region algorithm applied to (3.4) *and performed with the model* $\tilde{\phi}$ *in* (3.3) *and* $H_k = 0$ *converges to* 0*, which is* not *stationary in any sense (in particular neither Clarke nor Bouligand stationary).*

Proof. The proof is not difficult, but it is rather technical and lengthy, and therefore we refer the reader to the preprint version of this article [18]. □

We would like to emphasize that the failure of the trust-region method in the situation of Lemma 3.4 is not caused by the distinction of cases contained in Algorithm 2.5. It is easy to see that, in both cases $\Delta_k \geq \Delta_{\min}$ and $\Delta_k < \Delta_{\min}$, the iteration is the same (unless the algorithm meets one of the kinks), and thus, Algorithm 2.5 turns into a standard (nonsmooth) trust-region iteration.

What is remarkable about the above one-dimensional counterexample is that it shows that a trust-region method based on a model which does not account for any neighborhood information may fail to converge even in the case of a convex and piecewise affine objective. Moreover, this behavior may occur even if the objective function is smooth at all of the produced iterates. Of course, this observation is not new, and we exemplarily refer the reader to [2, section 5.5], where a similar two-dimensional example is discussed. Note that, if the initial radius Δ_0 is chosen slightly differently from the setting in (3.5) or if numerical errors occur, then the trust-region algorithm with the model in (3.3) will converge to the global minimum at $x = 1$ in the situation of the objective (3.4). This indicates that cases in which simple quadratic models or the standard model in (3.3) fail are quite pathological. Our algorithmic framework accounts for this observation by means of the distinction of cases $\Delta_k \gtrless \Delta_{\min}$.

In the literature, the most popular strategy to overcome the shortcomings of purely local models is to use bundling approaches, i.e., to sample subgradients at trial iterates in the vicinity of the current iterate to progressively improve a working model of the objective function until it is sufficiently accurate and allows the calculation of the next, serious iterate. Compare, for instance, with the analyses in [8, 29, 34, 35, 36, 41, 45, 48, 52, 62, 63, 77, 78] in this regard. By using such a method, it is possible to construct trust-region algorithms that are guaranteed to converge to C-stationarity under, e.g., the assumption of lower-C^1 regularity; see [35, 62, 65]. We expect that, for lower-C^1 objectives, approaches similar to those in [35, 62, 65] can also be combined with the trust-region framework of section 2 so that the inner model ϕ in Algorithm 2.5 is generated successively. (Note that, even in this situation, the distinction of cases in Algorithm 2.5 is advantageous since it allows us to avoid the costly generation of a bundle until absolutely necessary.) We do not pursue this approach here since the rigorous analysis of bundling methods is cumbersome, and since, unfortunately, in the applications that we are interested in—the study of VI-constrained optimization problems—the objective function can typically not be expected to be upper-C^1, lower-C^1, or Clarke regular. To demonstrate the latter, we note the following.

Lemma 3.5. *Suppose that a locally Lipschitz continuous function $f : \mathbb{R}^n \to \mathbb{R}$ is given. Then the following hold true:*

- *If f is lower-C^1, then f is regular in the sense of Clarke.*
- *If f is Clarke regular, then, for all $x, h \in \mathbb{R}^n$, we have $-f'(x;-h) \leq f'(x;h)$.*

Proof. According to [23, Corollary 3], lower-C^1 regularity is equivalent to approximate convexity, and this property implies directional differentiability and regularity in the sense of Clarke by [61, Corollary 3.5, Theorem 3.6]. The first claim of the lemma thus follows immediately. To obtain the second one, it suffices to note that the Clarke regularity of f and the fact that Clarke's directional derivative is convex in the second argument yield

$$0 = f^\circ(x;0) \leq \frac{1}{2} f^\circ(x;h) + \frac{1}{2} f^\circ(x;-h) = \frac{1}{2} f'(x;h) + \frac{1}{2} f'(x;-h) \qquad \forall x, h \in \mathbb{R}^n.$$

Rearranging the above gives $-f'(x;-h) \leq f'(x;h)$ and completes the proof. ∎

Note that the last lemma implies that the restriction of a lower-C^1 function to an arbitrary line can only have locally convex kinks. This gives some intuition of what the concepts in Definition 3.2 actually mean.

Let us consider now the simple, one-dimensional, VI-constrained optimization problem

$$\begin{aligned} &\min_{x,y\in\mathbb{R}} \quad J(y,x) := (y-y_d)^2 + x^2 \\ &\quad \text{s.t.} \quad y(z-y) + |z| - |y| \geq x(z-y) \quad \forall z \in \mathbb{R}, \end{aligned} \tag{3.6}$$

where $y_d \in \mathbb{R}$ is an arbitrary but fixed real number (the desired state). Then it is easy to check that the solution operator $S : \mathbb{R} \to \mathbb{R}$, $x \mapsto y$, of the lower-level VI in (3.6) is precisely the function $y(x) = \min\{0, x+1\} + \max\{0, x-1\}$. If we plug in this formula on the upper level of (3.6), then we arrive at the reduced objective function

$$\begin{aligned} f(x) := J(y(x),x) = \min\{0,x+1\}^2 + \max\{0,x-1\}^2 + x^2 \\ - 2y_d\left(\min\{0,x+1\} + \max\{0,x-1\}\right) + y_d^2. \end{aligned}$$

The above f possesses precisely one convex and one concave kink for all $y_d \neq 0$. The reduced objective of (3.6) is thus indeed neither upper-C^1, nor lower-C^1, nor regular in the sense of Clarke by Lemma 3.5 for all y_d that are not equal to zero. Note that one of the fundamental problems here is that, in contrast to the functions studied, e.g., in [51, 73], in (3.6) we consider the composition of an outer smooth function with an inner nonsmooth function, and not the other way around. This complicates the analysis significantly since the outer C^1-map "distorts" the local behavior of the inner function and thus typically renders all a priori knowledge that might be available about the structure of the kinks of the inner map (e.g., information about their direction) useless. This effect is, of course, even more severe in the higher-dimensional setting.

As already explained in the introduction, the lack of upper- and lower-C^1 regularity of the objective function of a VI-constrained optimization problem makes it much more difficult and costly to construct a trust-region model that results in an algorithm that is guaranteed to converge to C-stationary points. (Recall that this high computational cost is one of the main reasons why we have introduced the distinction of cases in Algorithm 2.5.) To construct an inner model ϕ that can be evaluated numerically for a problem of the type (3.6) and satisfies the conditions in Assumption 2.4, we will consider functions of the form

$$\phi(x,\Delta;d) := \sup_{g\in\mathcal{G}(x,\Delta)} \langle g, d\rangle, \qquad \mathcal{G}(x,\Delta) \supset \bigcup_{\xi\in B_\Delta(x)} \partial f(\xi). \tag{3.7}$$

We remark that models similar to (3.7) have also been proposed in [33] and [1]. The approach in (3.7) is further related to those of [10, 11, 21, 22, 44] in that it explores the ε-Bouligand subdifferential around the current iterate to overestimate the objective.

4. Composite functions.

After the general discussion of the last two sections, we now turn our attention to (discretized) optimal control problems with nonsmooth constraints. In order to conform with the standard notation in optimal control, we denote the optimization variable by u from this point on. Although this causes a slight abuse of notation, we also tacitly replace x by u when referring to the results of section 2. Our general optimal control problem thus reads as follows:

$$\min_{u\in\mathbb{R}^n} f(u) := J(S(u),u). \tag{4.1}$$

Here, $J : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$, $m, n \in \mathbb{N}$, is continuously differentiable and $S : \mathbb{R}^n \to \mathbb{R}^m$ is assumed to be directionally differentiable and locally Lipschitz continuous. Note that these properties imply that S is Bouligand differentiable; see [75, Theorem 3.1.2]. In all of what follows, we will frequently abbreviate $y := S(u) \in \mathbb{R}^m$. To construct a model ϕ for the objective of (4.1), we suppose that, for every $u \in \mathbb{R}^n$ and $\Delta > 0$, we can determine an approximation $\mathcal{G}(u, \Delta)$ of the Bouligand differential of S such that the following holds true.

Assumption 4.1. *Given $u \in \mathbb{R}^n$ and $\Delta > 0$, the approximation $\mathcal{G}(u, \Delta) \subset \mathbb{R}^{m \times n}$ of the Bouligand differential satisfies the following:*

- *For all $u \in \mathbb{R}^n$ and all $\Delta > 0$, it holds that*

$$\bigcup_{\xi \in B_\Delta(u)} \partial_B S(\xi) \subset \mathcal{G}(u, \Delta). \tag{4.2}$$

- *If $(u_k, \Delta_k) \to (u, 0)$ with $0 \notin \partial f(u)$, then we have*

$$\operatorname{dist}(\mathcal{G}(u_k, \Delta_k), \partial_B S(u)) = \sup_{G \in \mathcal{G}(u_k, \Delta_k)} \inf_{W \in \partial_B S(u)} \|G - W\|_{\mathbb{R}^{m \times n}} \to 0. \tag{4.3}$$

With the approximation $\mathcal{G}$ at hand, we define (analogously to (3.7))

$$\phi(u, \Delta; d) := \sup_{G \in \mathcal{G}(u, \Delta)} \langle G^\top \nabla_y J(y, u) + \nabla_u J(y, u), d \rangle. \tag{4.4}$$

Note that the model (4.4) allows the following reformulation of the modified trust-region subproblem ($\tilde{\mathrm{Q}}_k$), which will be useful for the realization of the algorithm in case of the concrete optimization problem in section 5.

Lemma 4.2. *With the model function in* (4.4)*, the modified trust-region subproblem* ($\tilde{\mathrm{Q}}_k$) *from Step* 11 *of Algorithm* 2.5 *is equivalent to the following linear-quadratic problem in the sense that they admit the same (global) optima:*

$$(\mathfrak{Q}_k) \quad \begin{cases} \displaystyle\min_{\zeta \in \mathbb{R},\, d \in \mathbb{R}^n} \mathfrak{q}_k(d, \zeta) := J(y_k, u_k) + \zeta + \frac{1}{2}\, d^\top H_k d \\ \quad\text{s.t.} \quad \|d\| \le \Delta_k, \\ \qquad\qquad \langle g, d \rangle \le \zeta \;\; \forall g \in \{G^\top \nabla_y J(y_k, u_k) + \nabla_u J(y_k, u_k) : G \in \mathcal{G}(u_k, \Delta_k)\}. \end{cases}$$

If, further, $\bar{d}_k$ is a global minimizer of ($\tilde{\mathrm{Q}}_k$), then $(\bar{d}_k, \bar{\zeta}_k)$ with $\bar{\zeta}_k := \phi(u_k, \Delta_k; \bar{d}_k)$ solves ($\mathfrak{Q}_k$) with $\tilde{q}_k(\bar{d}_k) = \mathfrak{q}_k(\bar{d}_k, \bar{\zeta}_k)$, and if (d_k, ζ_k) is feasible for ($\mathfrak{Q}_k$) and satisfies

$$f(x_k) - \mathfrak{q}_k(d_k, \zeta_k) \ge \frac{\mu}{2}\, \psi(x_k, \Delta_k) \min\Big\{\Delta_k, \frac{\psi(x_k, \Delta_k)}{\|H_k\|}\Big\}, \tag{4.5}$$

then d_k fulfills the modified Cauchy decrease condition in (2.5).

Proof. The proof is straightforward and therefore omitted. □

Remark 4.3. Note that solutions $(\bar{d}_k, \bar{\zeta}_k)$ of ($\mathfrak{Q}_k$) satisfy the modified Cauchy decrease condition (4.5) (since the $\bar{d}_k$ do by Lemma 2.8 and $\tilde{q}_k(\bar{d}_k) = \mathfrak{q}_k(\bar{d}_k, \bar{\zeta}_k)$).

In what follows, our aim will be to show that the model function in (4.4) satisfies the conditions in Assumption 2.4. To this end, we need the following.

Assumption 4.4. *For every $u \in \mathbb{R}^n$ and every $h \in \mathbb{R}^n$, there exists a $G \in \partial_B S(u)$ such that $S'(u; h) = G\, h$.*

Assumption 4.4 is, e.g., satisfied for semismooth maps S. Indeed, we have the following lemma.

LEMMA 4.5. *If, in addition to our previous assumptions, the map $S : \mathbb{R}^n \to \mathbb{R}^m$ is semismooth, then Assumption* 4.4 *is fulfilled.*

Proof. Let $u, h \in \mathbb{R}^n$ be given and let $\{t_l\} \subset \mathbb{R}^+$ be some null sequence. Then, for every $l \in \mathbb{N}$, Rademacher's theorem implies the existence of h_l with $u + t_l h_l \in \mathcal{D}_S$ and $\|h_l - h\| = \mathcal{O}(t_l)$ for $l \to \infty$, and we obtain from the semismoothness of S that

$$\frac{S(u + t_l h_l) - S(u)}{t_l} - S'(u + t_l h_l)h_l \to 0. \tag{4.6}$$

Due to the local Lipschitz continuity of S, the sequence $\{S'(u+t_l h_l)\}$ is bounded and thus, at least after the transition to a subsequence, convergent to some $G \in \partial_B S(u)$. Since S is Bouligand differentiable, the properties of h_l further imply that the first addend in (4.6) converges to $S'(u; h)$. The claim now follows immediately. □

We would like to point out that semismoothness is not necessary for Assumption 4.4 to hold. The map $S : \mathbb{R} \to \mathbb{R}$, $u \mapsto u^2 \sin(1/u)$, for example, is not semismooth but still satisfies this condition. We can now prove the following.

LEMMA 4.6. *Let $\{(u_k, \Delta_k)\} \subset \mathbb{R}^n \times \mathbb{R}^+$ be a sequence such that $u_k \to \tilde{u}$ and $\Delta_k \to 0$ holds. Then the model* (4.4) *satisfies Assumption* 2.4*(*2c*), i.e., we have*

$$\limsup_{k\to\infty} \sup_{d \in B_{\Delta_k}(0)} \frac{J(S(u_k + d), u_k + d) - J(S(u_k), u_k) - \phi(u_k, \Delta_k; d)}{\Delta_k} \le 0.$$

Proof. Let $u \in \mathbb{R}^n$, $\Delta > 0$, and $d \in B_\Delta(0)$ be arbitrary. By [75, Proposition 3.1.1] and the chain rule for Bouligand differentiable functions, we have

$$\begin{aligned} &J(S(u + d), u + d) - J(S(u), u) \\ &= \int_0^1 \langle \nabla_y J(S(u + \theta d), u + \theta d), S'(u + \theta d; d)\rangle + \langle \nabla_u J(S(u + \theta d), u + \theta d), d\rangle \mathrm{d}\theta. \end{aligned}$$

By Assumption 4.4, for every $\theta \in [0, 1]$, there further exists a $G_\theta \in \partial_B S(u + \theta d)$ such that $G_\theta \, d = S'(u + \theta d; d)$. This, together with (4.2), (4.4), and $\|d\| \le \Delta$, yields

$$\begin{aligned} &J(S(u + d), u + d) - J(S(u), u) \\ &= \int_0^1 \langle G_\theta^\top \nabla_y J(S(u + \theta d), u + \theta d) + \nabla_u J(S(u + \theta d), u + \theta d), d\rangle \mathrm{d}\theta \\ &\le \phi(u, \Delta; d) \\ &\qquad + \sup_{G \in \cup_{\xi \in B_\Delta(u)} \partial_B S(\xi)} (\|G\| + 1) \int_0^1 \|J'(S(u + \theta d), u + \theta d) - J'(S(u), u)\| \mathrm{d}\theta \, \Delta. \end{aligned}$$

Now, let $\{(u_k, \Delta_k)\}$ be the sequence from the lemma and let $\tilde{L} > 0$ and $\widetilde{\mathcal{U}} \subset \mathbb{R}^n$ denote the local Lipschitz constant of S at $\tilde{u}$ and an associated (w.l.o.g. bounded) neighborhood of local Lipschitz continuity, respectively. Then, for $K \in \mathbb{N}$ sufficiently large, we have $B_{\Delta_k}(u_k) \subset \widetilde{\mathcal{U}}$ and therefore

$$\sup_{G \in \cup_{\xi \in B_{\Delta_k}(u_k)} \partial_B S(\xi)} \|G\| \le \tilde{L} \quad \forall k \ge K,$$

and the uniform continuity of $u \mapsto J'(S(u), u)$ on $\mathrm{cl}(\widetilde{\mathcal{U}})$ and $(u_k, \Delta_k) \to (\tilde{u}, 0)$ imply

$$\sup_{d \in B_{\Delta_k}(0)} \|J'(S(u_k + d), u_k + d) - J'(S(u_k), u_k)\| \to 0 \quad \text{as } k \to \infty.$$

By combining all of the above findings, we arrive at

$$\begin{aligned} &\sup_{d \in B_{\Delta_k}(0)} \frac{J(S(u_k + d), u_k + d) - J(S(u_k), u_k) - \phi(u_k, \Delta_k; d)}{\Delta_k} \\ &\qquad \leq \left(\tilde{L} + 1\right) \sup_{d \in B_{\Delta_k}(0)} \int_0^1 \|J'(S(u_k + \theta d), u_k + \theta d) - J'(S(u_k), u_k)\| \mathrm{d}\theta \to 0, \end{aligned}$$

which implies the assertion. □

LEMMA 4.7. *Let ϕ be the model function from* (4.4)*, and let $\psi : \mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}$ be defined as in* (2.2)*. Suppose that a sequence $\{(u_k, \Delta_k)\} \subset \mathbb{R}^n \times \mathbb{R}^+$ is given such that $u_k \to u$, $\Delta_k \to 0$, and $\psi(u_k, \Delta_k) \to 0$ hold. Then u is a C-stationary point of the reduced objective function $f(u) := J(S(u), u)$. The model function* (4.4) *thus satisfies Assumption* 2.4(2b)*.*

Proof. We argue by contradiction and assume that there is an $\varepsilon > 0$ such that

$$\mathrm{dist}(0, \partial f(u)) \geq \varepsilon. \tag{4.7}$$

Let us denote the set of points where S and f are differentiable by $\mathcal{D}_S$ and $\mathcal{D}_f$, respectively. Since J is continuously differentiable, the chain rule implies $\mathcal{D}_S \subset \mathcal{D}_f$ and, by Rademacher's theorem, we have $\lambda^n(\mathcal{D}_f \setminus \mathcal{D}_S) = 0$. Therefore, [19, Theorem 2.5.1] and the continuous differentiability of J imply

$$\begin{aligned} &\{G^\top \nabla_y J(y, u) + \nabla_u J(y, u) : G \in \partial_B S(u)\} \\ &\quad \subset \mathrm{cl}\Big(\mathrm{conv}\big(\{g \in \mathbb{R}^n : \exists \{u_l\} \subset \mathcal{D}_S : \\ &\qquad\qquad u_l \to u,\, S'(u_l)^\top \nabla_y J(y_l, u_l) + \nabla_u J(y_l, u_l) \to g\}\big)\Big) \\ &\quad = \partial f(u). \end{aligned} \tag{4.8}$$

Thus, due to (4.3), there exists $K \in \mathbb{N}$ such that, for all $k \geq K$, it holds that

$$\begin{aligned} &\{G^\top \nabla_y J(y_k, u_k) + \nabla_u J(y_k, u_k) : G \in \mathcal{G}(u_k, \Delta_k)\} \\ &\quad \subset \{G^\top \nabla_y J(y, u) + \nabla_u J(y, u) : G \in \partial_B S(u)\} + B_{\varepsilon/2}(0) \subset \partial f(u) + B_{\varepsilon/2}(0). \end{aligned}$$

Since $\partial f(u)$ is convex, this, in combination with (4.7), implies $\mathrm{dist}(\mathcal{C}_k, 0) \geq \varepsilon/2$, where we abbreviated

$$\mathcal{C}_k := \mathrm{cl}\left(\mathrm{conv}\left(\{G^\top \nabla_y J(y_k, u_k) + \nabla_u J(y_k, u_k) : G \in \mathcal{G}(u_k, \Delta_k)\}\right)\right). \tag{4.9}$$

Next, let us define $\bar{g}$ as the unique solution of the following VI:

$$\bar{g} \in \mathcal{C}_k, \quad \langle \bar{g}, g - \bar{g} \rangle \geq 0 \quad \forall g \in \mathcal{C}_k.$$

Using this VI in combination with $\operatorname{dist}(\mathcal{C}_k, 0) \geq \varepsilon/2$ results in

$$\begin{aligned}\psi(u_k, \Delta_k) &= \max_{\|h\|\leq 1} \Big(\inf_{G\in\mathcal{G}(u_k,\Delta_k)} \langle G^\top \nabla_y J(y_k, u_k) + \nabla_u J(y_k, u_k), -h\rangle \Big) \\ &\geq \inf_{G\in\mathcal{G}(u_k,\Delta_k)} \left\langle G^\top \nabla_y J(y_k, u_k) + \nabla_u J(y_k, u_k), \frac{\bar g}{\|\bar g\|} \right\rangle \\ &\geq \inf_{g\in\mathcal{C}_k} \frac{\langle g, \bar g\rangle}{\|\bar g\|} \geq \|\bar g\| \geq \frac{\varepsilon}{2} \qquad \forall k \geq K.\end{aligned}$$

This, however, contradicts $\psi(u_k, \Delta_k) \to 0$ so that the assertion follows. □

We may now collect the findings of this section as follows.

Corollary 4.8. *Under Assumptions* 4.1 *and* 4.4*, every accumulation point of the sequence of iterates generated by our nonsmooth trust-region algorithm applied to* (4.1) *with the model function in* (4.4) *is a C-stationary point of* (4.1)*.*

Proof. As shown in Lemmas 4.6 and 4.7, Assumption 2.4 is fulfilled, provided that Assumptions 4.1 and 4.4 hold true. Theorem 2.12 thus yields the claim. □

5. Optimization with variational inequalities of the second kind.

We now focus on the following class of nonsmooth optimization problems:

$$(\mathrm{P_{VI}}) \qquad \begin{cases} \min\limits_{y\in\mathbb{R}^m,\, u\in\mathbb{R}^n} & J(y,u) \\ \quad\text{s.t.} & \langle Ay, v-y\rangle + \|v\|_1 - \|y\|_1 \geq \langle Ru, v-y\rangle \quad \forall v \in \mathbb{R}^m, \end{cases}$$

where $J : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ is smooth, $A \in \mathbb{R}^{m\times m}$ is symmetric and positive definite, $R \in \mathbb{R}^{m\times n}$ is surjective, and $\|\cdot\|_1$ denotes the 1-norm, i.e., $\|v\|_1 = \sum_{i=1}^m |v_i|$. Note that the constraints have the form of a VI of the second kind here.

Remark 5.1. Problems of the type $(\mathrm{P_{VI}})$ arise, for instance, when a finite element method with mass-lumping is applied to certain infinite-dimensional optimal control problems governed by H_0^1-elliptic VIs of the second kind. Compare, e.g., with [26] and the example in section 6 in this context.

In the next proposition, we summarize some known results about $(\mathrm{P_{VI}})$. For more details on this topic, we refer the reader to [26].

Proposition 5.2. *Let $u \in \mathbb{R}^n$ be given. Then the following hold:*

- *There exists a unique solution $y \in \mathbb{R}^m$ of the VI in* $(\mathrm{P_{VI}})$*, i.e.,*

$$(\mathrm{VI}) \qquad \langle Ay, v-y\rangle + \|v\|_1 - \|y\|_1 \geq \langle Ru, v-y\rangle \quad \forall v \in \mathbb{R}^m.$$

- *A vector y solves* (VI) *if and only if there exists a $q \in \mathbb{R}^m$ such that*

$$(5.1) \qquad Ay + q = Ru, \quad y_i q_i = |y_i|, \quad |q_i| \leq 1, \quad i = 1, \ldots, m.$$

- *The solution map $S : \mathbb{R}^n \ni u \mapsto y \in \mathbb{R}^m$ of* (VI) *is globally Lipschitz continuous and directionally differentiable. Its directional derivative $\eta := S'(u; h)$ at u in direction $h \in \mathbb{R}^n$ is given by the unique solution of*

$$(5.2) \qquad \eta \in \mathcal{K}(y), \quad \langle A\eta, v-\eta\rangle \geq \langle Rh, v-\eta\rangle \quad \forall v \in \mathcal{K}(y),$$

where

$$(5.3) \quad \mathcal{K}(y) := \{v \in \mathbb{R}^m : v_i = 0, \ \text{if } |q_i| < 1, \ v_i\, q_i \geq 0, \ \text{if } y_i = 0 \wedge |q_i| = 1\}.$$

In what follows, we will frequently write $y(u)$ and $q(u)$ to indicate that $y = S(u)$ holds and that q is the associated slack variable from (5.1). With the solution operator S at hand, we may formulate problem $(\mathrm{P_{VI}})$ in reduced form as

$$\min_{u\in\mathbb{R}^n} \; f(u) := J(S(u), u),$$

so that a problem of the form (4.1) is obtained. Our aim in the following is to verify the hypotheses on the general problem (4.1), i.e., Assumptions 4.1 and 4.4. To this end, we first have to characterize the Bouligand differential of S, which is addressed in the next subsection.

5.1. Characterization of the Bouligand differential. Given $u \in \mathbb{R}^n$ with $y = S(u) \in \mathbb{R}^m$, we define the following sets:

$$\begin{aligned}
\mathcal{A} &:= \{i \in \{1, ..., m\} : y_i(u) = 0\} && \text{(active set)},\\
\mathcal{A}_s &:= \{i \in \{1, ..., m\} : |q_i(u)| < 1\} && \text{(strongly active set)},\\
\mathcal{I} &:= \{i \in \{1, ..., m\} : y_i(u) \neq 0\} && \text{(inactive set)},\\
\mathcal{B} &:= \{i \in \{1, ..., m\} : y_i(u) = 0 \wedge |q_i(u)| = 1\} && \text{(bi-active set)}.
\end{aligned}$$

Note that these sets depend on y and thus indirectly on u so that it would be more appropriate to write $\mathcal{A}(y)$ or $\mathcal{A}(u)$, etc., here. We suppress this dependency throughout this subsection for the sake of readability. This will be different in subsection 5.2, where we have to distinguish between the active sets in different points. Note that, because of (5.1), one has $\mathcal{A}_s \subset \mathcal{A}$, and, as a consequence, $\mathcal{A} = \mathcal{A}_s \cup \mathcal{B}$.

LEMMA 5.3. *S is differentiable at u iff $\mathcal{K}(y) = \{v \in \mathbb{R}^m : v_i = 0$ if $y_i(u) = 0\}$.*

Proof. It is clear that, if $\mathcal{K}(y)$ takes the form stated in the lemma, then $\mathcal{K}(y)$ is a linear subspace and, as a metric projection on a subspace, $S'(u; \cdot)$ is a linear mapping so that S is differentiable at u. To show the reverse implication, we first show that

$$S'(u; \mathbb{R}^n) = \mathcal{K}(y). \tag{5.4}$$

By (5.2), we already have $S'(u; \mathbb{R}^n) \subset \mathcal{K}(y)$. To see the reverse inclusion, let $z \in \mathcal{K}(y)$ be arbitrary. Since R is surjective by assumption, there is an $h \in \mathbb{R}^n$ such that $Rh = Az$. Then we trivially obtain $\langle Az, v - z\rangle = \langle Rh, v - z\rangle$ for all $v \in \mathcal{K}(y)$ so that $z = S'(u; h)$, which shows (5.4). Moreover, if S is differentiable so that $h \mapsto S'(u; h)$ is linear, then $S'(u; \mathbb{R}^n)$ becomes a linear subspace and, by (5.4), so does $\mathcal{K}(y)$. Therefore, $v \in \mathcal{K}(y)$ implies $-v \in \mathcal{K}(y)$, which, due to (5.3) yields

$$0 \leq v_i q_i(u) \leq 0 \quad \forall i \in \mathcal{B} = \{j \in \{1, \dots, m\} : y_j = 0 \wedge |q_j| = 1\}.$$

Since $q_i(u) \neq 0$ in $\mathcal{B}$, this gives $v_i = 0$ in $\mathcal{B}$, which, together with $v_i = 0$ in $\mathcal{A}_s$ (see (5.3)), finally results in $v_i = 0$ in $\mathcal{B} \cup \mathcal{A}_s = \mathcal{A}$ as claimed. ◻

We are now in the position to give a precise characterization of the Bouligand differential of S. To this end, we introduce the following definition.

DEFINITION 5.4. *Let $\mathcal{N} \subset \{1, \dots, m\}$ be an index set. Then we define the matrices $A(\mathcal{N}) \in \mathbb{R}^{m\times m}$ and $\chi(\mathcal{N}) \in \mathbb{R}^{m\times m}$ by*

$$A(\mathcal{N})_{ij} := \begin{cases} A_{ij} & \text{if } i, j \in \{1, \dots, m\} \setminus \mathcal{N},\\ 0 & \text{if } i \vee j \in \mathcal{N}, i \neq j,\\ 1 & \text{if } i = j \in \mathcal{N}, \end{cases} \qquad \chi(\mathcal{N})_{ij} := \begin{cases} 1 & \text{if } i = j \in \{1, \dots, m\} \setminus \mathcal{N},\\ 0 & \text{otherwise.} \end{cases}$$

THEOREM 5.5. *Let $u \in \mathbb{R}^n$ be arbitrary but fixed, and let $y = S(u)$. Then,*

$$\partial_B S(u) = \{A(\mathcal{A}_s \cup \mathcal{B}_0)^{-1}\, \chi(\mathcal{A}_s \cup \mathcal{B}_0)\, R : \mathcal{B}_0 \subset \mathcal{B}\}. \tag{5.5}$$

Remark 5.6. Note that $A(\mathcal{N})$ is indeed invertible for every set $\mathcal{N} \subset \{1, \dots, m\}$ since it is positive definite. Indeed, for an arbitrary $v \in \mathbb{R}^m$, we have

$$v^\top A(\mathcal{N}) v = [\chi(\mathcal{N})v]^\top A\, [\chi(\mathcal{N})v] + \sum_{i \in \mathcal{N}} v_i^2 \geq \min\{\lambda_{\min}, 1\} \|v\|^2,$$

where $\lambda_{\min} > 0$ denotes the minimal eigenvalue of A.

Remark 5.7. The last line in the definition of $\mathcal{A}(\mathcal{N})$ can be replaced by

$$A(\mathcal{N})_{ij} := c \quad \text{if } i = j \in \mathcal{N}$$

with some $c \neq 0$ since regardless of which value is chosen for $c \neq 0$, the matrix $A(\mathcal{N})^{-1}\chi(\mathcal{N})$ is always the same, as

$$A(\mathcal{N})\xi = \chi(\mathcal{N})\zeta \quad \Longleftrightarrow \quad \xi_i = 0 \ \forall i \in \mathcal{N}, \quad \sum_{j \notin \mathcal{N}} A_{ij}\xi_j = \zeta_i \ \forall i \in \{1, \dots, m\} \setminus \mathcal{N}, \tag{5.6}$$

and there is no more c appearing on the right-hand side of this equivalence.

Proof of Theorem 5.5. Recall that

$$\partial_B S(u) = \big\{B \in \mathbb{R}^{m \times n} : \exists\, \{u_l\} \subset \mathcal{D}_S \text{ with } u_l \to u,\ S'(u_l) \to B\big\}, \tag{5.7}$$

where $\mathcal{D}_S$ again denotes the (dense) set of points, where S is differentiable, and consider an arbitrary but fixed $B \in \partial_B S(u)$ and a sequence $\{u_l\}$ in $\mathcal{D}_S$ satisfying $u_l \to u$ and $S'(u_l) \to B$ for $l \to \infty$. Then the Lipschitz continuity of S implies

$$y_l := S(u_l) \to S(u) =: y \quad \text{and} \quad q_l := Ru_l - Ay_l \to Ru - Ay =: q. \tag{5.8}$$

Let us denote the active set associated with y_l by $\mathcal{A}^l$ and analogously for $\mathcal{I}^l$, etc. Then, from (5.8), we deduce the existence of an $L \in \mathbb{N}$ such that $\mathcal{I} \subset \mathcal{I}^l$ and $\mathcal{A}_s \subset \mathcal{A}_s^l$ hold for all $l \geq L$. Next, let $h \in \mathbb{R}^n$ be arbitrary but fixed. Since $u_l \in \mathcal{D}_S$, we know from Lemma 5.3 that $\eta_l := S'(u_l)h$ solves

$$\eta_i^l = 0 \quad \forall i \in \mathcal{A}^l, \quad \sum_{j=1}^m A_{ij}\eta_j^l = \sum_{k=1}^n R_{ik}h_k \quad \forall i \in \mathcal{I}^l. \tag{5.9}$$

Due to the convergence $S'(u_l) \to B$, we further know that $\eta_l \to Bh =: \tilde{\eta}$ for $l \to \infty$. Combining all of the above yields

$$\tilde{\eta}_i = 0 \quad \forall i \in \mathcal{A}_s, \quad \sum_{j=1}^m A_{ij}\tilde{\eta}_j = \sum_{k=1}^n R_{ik}h_k \quad \forall i \in \mathcal{I}.$$

It remains to study what happens on the set $\mathcal{B} = \mathcal{A} \setminus \mathcal{A}_s$. To this end, we introduce

$$\mathcal{B}_0 := \{i \in \mathcal{B} : \exists \text{ a subsequence } \{l_k\} \text{ such that } y_i(u)^{l_k} = 0\ \forall k \in \mathbb{N}\}$$

so that, for all $i \in \mathcal{B} \setminus \mathcal{B}_0$, it holds that $y_i^l \neq 0$ for all $l \in \mathbb{N}$ sufficiently large. Then we deduce from (5.9) that $\eta_i^{l_k} = 0$ for all $i \in \mathcal{B}_0$ and all $k \in \mathbb{N}$ and that

$\sum_{j=1}^m A_{ij}\eta_j^l = \sum_{k=1}^n R_{ik}h_k$ for all $i \in \mathcal{B} \setminus \mathcal{B}_0$, provided that $l \in \mathbb{N}$ is sufficiently large. Since $\eta_l \to \tilde{\eta}$, we obtain in this way

$$\tilde{\eta}_i = 0 \;\; \forall i \in \mathcal{A}_s \cup \mathcal{B}_0, \qquad \sum_{j \notin \mathcal{A}_s \cup \mathcal{B}_0} A_{ij}\tilde{\eta}_j = \sum_{k=1}^n R_{ik}h_k \;\; \forall i \in \{1, \dots, m\} \setminus (\mathcal{A}_s \cup \mathcal{B}_0).$$

Thus, in view of (5.6) and since h was arbitrary, we observe that B indeed has the form $B = A(\mathcal{A}_s \cup \mathcal{B}_0)^{-1}\,\chi(\mathcal{A}_s \cup \mathcal{B}_0)\,R$. This proves the inclusion "$\subset$" in (5.5).

It remains to prove that, for every set $\mathcal{B}_0 \subset \mathcal{B}$, the matrix B in (5.5) is an element of $\partial_B S(u)$. So let us fix a $\mathcal{B}_0 \subset \mathcal{B}$, denote the associated matrix with B, and write $\mathcal{B}_1 := \mathcal{B} \setminus \mathcal{B}_0$. In the following, we show that there exists a sequence $\{u_l\}$ satisfying

$$\begin{aligned} &u_l \in \mathcal{D}_S, \quad y_i^l = 0 \quad \forall i \in \mathcal{B}_0, \quad y_i^l \neq 0 \quad \forall i \in \mathcal{B}_1, \quad \forall l \in \mathbb{N}, \\ &\qquad\qquad \text{and} \quad u_l \to u, \quad S'(u_l) \to B \quad \text{as } l \to \infty, \end{aligned} \tag{5.10}$$

which, according to (5.7), implies $B \in \partial_B S(u)$. To verify the existence of such a sequence, let $\varepsilon > 0$ be arbitrary but fixed and define

$$y^\varepsilon := y + \sum_{k \in \mathcal{B}_1} \varepsilon \,\operatorname{sgn}(q_k)\,\mathrm{e}_k,$$

where e_i denotes the ith Euclidean unit vector. By construction, we obtain for the inactive and active set of y^ε that $\mathcal{I}^\varepsilon = \mathcal{I} \cup \mathcal{B}_1$ and $\mathcal{A}^\varepsilon = \mathcal{A} \setminus \mathcal{B}_1$. Let us further set

$$q^\varepsilon := q - \sum_{k \in \mathcal{B}_0} \varepsilon \,\operatorname{sgn}(q_k)\,\mathrm{e}_k.$$

Then, for $\varepsilon \in (0, 1]$, we obtain $|q_i^\varepsilon| \leq 1$ for all $i = 1, \dots, m$ and $\mathcal{A}_s^\varepsilon = \mathcal{A}_s \cup \mathcal{B}_0 = \mathcal{A} \setminus \mathcal{B}_1$, which, together with $\mathcal{A}^\varepsilon = \mathcal{A} \setminus \mathcal{B}_1$, shows that $\mathcal{B}^\varepsilon = \mathcal{A}^\varepsilon \setminus \mathcal{A}_s^\varepsilon = \emptyset$. The bi-active set associated with y^ε is thus empty. Furthermore, we define

$$u^\varepsilon := u + \varepsilon\, R^\dagger \Big(\sum_{k \in \mathcal{B}_1} \operatorname{sgn}(q_k)\, A\, \mathrm{e}_k - \sum_{k \in \mathcal{B}_0} \operatorname{sgn}(q_k)\, \mathrm{e}_k \Big), \tag{5.11}$$

where $R^\dagger \in \mathbb{R}^{n \times m}$ is the Moore–Penrose pseudoinverse. Since R is assumed to be surjective, it holds that $R^\dagger = R^\top(RR^\top)^{-1}$. Thus, we obtain by construction that

$$A\, y^\varepsilon + q^\varepsilon = R\, u^\varepsilon, \quad y_i^\varepsilon\, q_i^\varepsilon = |y_i^\varepsilon|, \quad |q_i^\varepsilon| \leq 1, \quad i = 1, \dots, m,$$

which, due to (5.1), implies $y^\varepsilon = S(u^\varepsilon)$. Because of $\mathcal{B}^\varepsilon = \emptyset$, we further have $\mathcal{K}(y^\varepsilon) = \{v \in \mathbb{R}^m : v_i = 0 \text{ if } y_i^\varepsilon = 0\}$, which, thanks to Lemma 5.3, implies $u^\varepsilon \in \mathcal{D}_S$. In addition, (5.11) immediately gives $u_\varepsilon \to u$ as $\varepsilon \to 0$. Because of $\mathcal{I}^\varepsilon = \mathcal{I} \cup \mathcal{B}_1$, we moreover have $y_i^\varepsilon \neq 0$ for all $i \in \mathcal{B}_1$ and, due to complementarity and $\mathcal{A}_s^\varepsilon = \mathcal{A} \setminus \mathcal{B}_1$, $y_i^\varepsilon = 0$ for all $i \in \mathcal{B}_0$. The sequence $\{u^\varepsilon\}_{\varepsilon > 0}$ thus satisfies all conditions in (5.10) except $S'(u^\varepsilon) \to B$. To establish this, let $\{\varepsilon_l\}_{l \in \mathbb{N}}$ be an arbitrary but fixed sequence tending to zero and write $u_l := u^{\varepsilon_l}$. Due to the global Lipschitz continuity of S, we have $\|S'(u_l)\|_{\mathbb{R}^{m \times n}} \leq L$ for all $l \in \mathbb{N}$ with some suitable constant $L > 0$. This allows us to pass over to a subsequence to achieve $S'(u_{l_k}) \to \tilde{B}$ for $k \to \infty$ with some matrix $\tilde{B}$. Since $y_i(u)^{l_k} = 0$ for all $i \in \mathcal{B}_0$ and $y_i(u)^{l_k} \neq 0$ for all $i \in \mathcal{B}_1$, we can argue completely analogously to the first part of the proof to obtain $\tilde{B} = A(\mathcal{A}_s \cup \mathcal{B}_0)^{-1}\,\chi(\mathcal{A}_s \cup \mathcal{B}_0)\,R$. It thus holds that $B = \tilde{B}$ and the proof is complete. $\square$

Note that, for finite-dimensional optimization problems subject to elliptic variational inequalities of the first kind, a result similar to that in Theorem 5.5 has already been obtained in [67, section 3]. For the construction of subgradients in infinite dimensions, see also [17, 70, 71]. To the best of the authors' knowledge, the question of how to characterize the generalized differential of the solution map of a VI of the second kind has not been addressed so far in the literature. From (5.5), we obtain the following lemma.

LEMMA 5.8. *For all $u, h \in \mathbb{R}^n$, there exists a $G \in \partial_B S(u)$ with $S'(u;h) = G\,h$. Hence, Assumption* 4.4 *is fulfilled by the control-to-state map of* ($\mathrm{P_{VI}}$).

Proof. Let $u, h \in \mathbb{R}^n$ be arbitrary, set $y = S(u)$, and denote the strongly active, inactive, and bi-active sets of y again with $\mathcal{A}_s$, $\mathcal{I}$, and $\mathcal{B}$, respectively. Then the cone $\mathcal{K}(y)$ in the VI (5.2) for the derivative $\eta := S'(u;h)$ can also be written as

$$\mathcal{K}(y) = \left\{ v \in \mathbb{R}^m : v_i = 0 \text{ if } |q_i(u)| < 1,\ v_i \begin{cases} \geq 0 & \text{if } y_i(u) = 0,\ q_i(u) = 1 \\ \leq 0 & \text{if } y_i(u) = 0,\ q_i(u) = -1 \end{cases} \right\}.$$

In combination with (5.2), the above implies that η satisfies

$$\eta_i = \begin{cases} \max\{0, (I-A)\eta + Rh\}_i & \text{if } y_i(u) = 0,\ q_i(u) = 1, \\ 0 & \text{if } |q_i(u)| < 1, \\ \min\{0, (I-A)\eta + Rh\}_i & \text{if } y_i(u) = 0,\ q_i(u) = -1, \\ ((I-A)\eta + Rh)_i & \text{else.} \end{cases} \tag{5.12}$$

So, if we define

$$\mathcal{B}_0 := \{i \in \{1, \dots, m\} : y_i(u) = 0,\ |q_i(u)| = 1,\ q_i(u)((I-A)\eta + Rh)_i < 0\} \subset \mathcal{B},$$

then a comparison of (5.12) with (5.6) gives $\eta = A(\mathcal{A}_s \cup \mathcal{B}_0)^{-1}\chi(\mathcal{A}_s \cup \mathcal{B}_0)R\,h$. The claim now follows immediately from Theorem 5.5. □

5.2. Approximation of the Bouligand differential. The aim of this section is to construct a *computable* approximation of $\partial_B S$ that satisfies Assumption 4.1. Again, we denote the state associated with a control u by $y(u)$ and, similarly, we write $q = q(u)$ for the slack variable in (5.1). Analogously, the strongly active, inactive, and bi-active sets at u are denoted by $\mathcal{A}_s(u)$, $\mathcal{I}(u)$, and $\mathcal{B}(u)$. (Note that these sets are determined by $y(u)$ and $q(u)$, which in turn uniquely depend on u.) We start with a sharpened Lipschitz continuity result for the solution map S associated with (VI).

LEMMA 5.9. *For all $u_1, u_2 \in \mathbb{R}^n$, it holds that*

$$\|y(u_1) - y(u_2)\|_\infty \leq L_y \|u_1 - u_2\| \qquad \text{with } L_y = \frac{\|R\|_{\mathbb{R}^{m\times n}}}{\lambda_{\min}}, \tag{5.13}$$

$$\|q(u_1) - q(u_2)\|_\infty \leq L_q \|u_1 - u_2\| \qquad \text{with } L_q = \Big(\frac{\lambda_{\max}}{\lambda_{\min}} + 1\Big)\|R\|_{\mathbb{R}^{m\times n}}. \tag{5.14}$$

Here, $\lambda_{\min}$ and $\lambda_{\max}$ denote the minimal and maximal eigenvalues of A, respectively.

Proof. To establish (5.13), it suffices to choose $y(u_2)$ as a test vector in the VI for $y(u_1)$ and $y(u_1)$ as a test vector in the VI for $y(u_2)$, to add the resulting inequalities, and to exploit the properties of A. The second estimate (5.14) follows immediately from (5.13) and the first equation in (5.1), which implies

$$\|q(u_1) - q(u_2)\| \leq \|A\|_{\mathbb{R}^{m\times m}}\|y(u_1) - y(u_2)\| + \|R\|_{\mathbb{R}^{m\times n}}\,\|u_1 - u_2\|.$$

This completes the proof. ◻

The construction of our model function is based on the following extended bi-active and sharpened strongly active set.

Definition 5.10. *Let $u \in \mathbb{R}^n$ and $\Delta > 0$ be given. Then we define the set of* possibly bi-active indices *and the* very active set*, respectively, by*

$$\mathcal{P}(u,\Delta) := \{i \in \{1,\dots,m\} : |y_i(u)| \le L_y\,\Delta \;\wedge\; |q_i(u)| \ge 1 - L_q\,\Delta\},$$
$$\mathcal{A}_v(u,\Delta) := \{i \in \{1,\dots,m\} : |q_i(u)| < 1 - L_q\,\Delta\}.$$

In view of Lemma 5.9, it is clear that

$$\bigcup_{\xi \in B_\Delta(u)} \mathcal{B}(\xi) \subset \mathcal{P}(u,\Delta) \quad \text{and} \quad \mathcal{A}_v(u,\Delta) \subset \bigcap_{\xi \in B_\Delta(u)} \mathcal{A}_s(\xi), \tag{5.15}$$

which is essential for the subsequent analysis. Given the set of possibly active indices, we construct our approximation of the Bouligand differential as follows:

$$\mathcal{G}(u,\Delta) := \{A(\mathcal{A}_v(u,\Delta) \cup \mathcal{B}_0)^{-1}\,\chi(\mathcal{A}_v(u,\Delta) \cup \mathcal{B}_0)\,R : \mathcal{B}_0 \subset \mathcal{P}(u,\Delta)\}. \tag{5.16}$$

As an immediate consequence of (5.15) and Theorem 5.5, we obtain the following lemma.

Lemma 5.11. *The approximation of $\partial_B S$ in* (5.16) *satisfies condition* (4.2).

Proof. Let $\xi \in B_\Delta(u)$ and $\mathcal{B}_0 \subset \mathcal{B}(\xi)$ be arbitrary and consider an arbitrary $i \in \mathcal{A}_s(\xi)$. Then either $|q_i(u)| < 1 - L_q\Delta$ so that $i \in \mathcal{A}_v(u,\Delta)$ or $|q_i(u)| \ge 1 - L_q\Delta$. In the latter case, we find $|y_i(u)| \le |y_i(\xi)| + |y_i(u) - y_i(\xi)| \le L_y\Delta$ (since $i \in \mathcal{A}_s(\xi)$) and, as a consequence, $i \in \mathcal{P}(u,\Delta)$. If we now define

$$\widetilde{\mathcal{B}}_0 := \{i \in \{1,\dots,m\} : i \in \mathcal{A}_s(\xi),\ |q_i(u)| \ge 1 - L_q\Delta\} \cup \mathcal{B}_0,$$

then it follows that $\mathcal{A}_s(\xi) \cup \mathcal{B}_0 = \mathcal{A}_v(u,\Delta) \cup \widetilde{\mathcal{B}}_0$, and, thanks to the first inclusion in (5.15), $\widetilde{\mathcal{B}}_0 \subset \mathcal{P}(u,\Delta)$, which, together with Theorem 5.5, yields the claim. ◻

On the other hand, we also have the following lemma.

Lemma 5.12. *Let $\{(u_k,\Delta_k)\} \subset \mathbb{R}^n \times \mathbb{R}^+$ be a sequence with $(u_k,\Delta_k) \to (u,0)$. Then there exists a $K \in \mathbb{N}$ (depending on u) such that $\mathcal{A}_s(u) \subset \mathcal{A}_v(u_k,\Delta_k)$ and $\mathcal{P}(u_k,\Delta_k) \subset \mathcal{B}(u)$ for all $k \ge K$. For all $k \ge K$, we thus have $\mathcal{G}(u_k,\Delta_k) \subset \partial_B S(u)$ and the approximation of $\partial_B S$ in* (5.16) *satisfies* (4.3).

Proof. We define

$$\delta_y := \min_{i \in \mathcal{I}(u)} |y_i(u)| > 0 \quad \text{and} \quad \delta_q := \min_{i \in \mathcal{A}_s(u)} \big(1 - |q_i(u)|\big) > 0.$$

Since $u_k \to u$ and S is globally Lipschitz, there exists $K_1 \in \mathbb{N}$ such that

$$\min_{i \in \mathcal{I}(u)} |y_i(u_k)| \ge \frac{\delta_y}{2} \quad \text{and} \quad \min_{i \in \mathcal{A}_s(u)} \big(1 - |q_i(u_k)|\big) \ge \frac{\delta_q}{2} \quad \forall k \ge K_1.$$

Moreover, as $\Delta_k \to 0$, we can find another index $K_2 \in \mathbb{N}$ such that

$$\Delta_k < \min\Big\{\frac{\delta_y}{2L_y}, \frac{\delta_q}{2L_q}\Big\} \quad \forall k \ge K_2.$$

Consequently, we obtain for all $i \in \mathcal{I}(u)$ that

$$|y_i(u_k)| \geq \frac{\delta_y}{2} > L_y\,\Delta_k \quad \Longrightarrow \quad i \notin \mathcal{P}(u_k,\Delta_k) \quad \forall k \geq K := \max\{K_1, K_2\}. \tag{5.17}$$

Analogously, for all $i \in \mathcal{A}_s(u)$, we have

$$|q_i(u_k)| < 1 - L_q\Delta_k \quad \Longrightarrow \quad i \in \mathcal{A}_v(u_k,\Delta_k) \text{ and } i \notin \mathcal{P}(u_k,\Delta_k) \quad \forall k \geq K.$$

Thus, $\mathcal{A}_s(u) \subset \mathcal{A}_v(u_k,\Delta_k)$ and, since $\mathcal{B}(u) = \{1,\dots,m\}\setminus(\mathcal{A}_s(u)\cup\mathcal{I}(u))$, it also follows that $\mathcal{P}(u_k,\Delta_k) \subset \mathcal{B}(u)$ as claimed. Furthermore, by (5.17) and complementarity, it holds that $|q_i(u_k)| = 1$ for all $i \in \mathcal{I}(u)$ and all $k \geq K$ and therefore $\mathcal{A}_v(u_k,\Delta_k) \subset \mathcal{A}_s(u)\cup\mathcal{B}(u)$. In summary, we thus find $\mathcal{A}_v(u_k,\Delta_k)\setminus\mathcal{A}_s(u) \subset \mathcal{B}(u)$ and $\mathcal{P}(u_k,\Delta_k) \subset \mathcal{B}(u)$. Therefore, for every $\mathcal{B}_0 \subset \mathcal{P}(u_k,\Delta_k)$, there is a set $\widetilde{\mathcal{B}}_0 \subset \mathcal{B}(u)$ with $\mathcal{A}_v(u_k,\Delta_k)\cup\mathcal{B}_0 = \mathcal{A}_s(u)\cup\widetilde{\mathcal{B}}_0$. The second assertion now follows immediately from Theorem 5.5 and (5.16). ∎

For convenience, we next state the precise algorithm that arises when we apply Algorithm 2.5 to $(\mathrm{P}_{\mathrm{VI}})$. Here, we again use the notation $f(\cdot) := J(S(\cdot),\cdot)$.

ALGORITHM 5.13 (trust-region algorithm for the solution of $(\mathrm{P}_{\mathrm{VI}})$).

```
Choose constants $\Delta_{\min} > 0$, $0 < \eta_1 < \eta_2 < 1$, $0 < \beta_1 < 1 < \beta_2$, $0 < \mu \leq 1$, an
initial value $u_0 \in \mathbb{R}^n$, and an initial TR-radius $\Delta_0 > \Delta_{\min}$. Set $k := 0$.
for $k = 0,1,2,\dots$ do
  Solve (VI) to compute the state $y_k$ associated with $u_k$.
  Choose a subset $\mathcal{B}_k \subset \mathcal{B}(u_k)$, solve the adjoint equation

    (5.18)   $A(\mathcal{A}_s(u_k)\cup\mathcal{B}_k)\,p_k = \chi(\mathcal{A}_s(u_k)\cup\mathcal{B}_k)\nabla_y J(y_k,u_k)$,

  and set $g_k = R^\top p_k + \nabla_u J(y_k,u_k)$.
  Choose a matrix $H_k \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$, e.g., via a BFGS-update using $g_k$.
  if $g_k = 0$ then
    STOP the iteration, $0 \in \partial_B f(u_k)$.
  else
    if $\Delta_k > \Delta_{\min}$ then
      Compute an inexact solution $d_k$ of the trust-region subproblem

        $(\mathrm{Q}_k)$   $\left\{ \begin{array}{ll} \min\limits_{d\in\mathbb{R}^n} & q_k(d) := f(u_k) + \langle g_k, d\rangle + \frac{1}{2}\,d^\top H_k d \\ \text{s.t.} & \|d\| \leq \Delta_k \end{array} \right.$

      that satisfies the generalized Cauchy decrease condition

        $f(u_k) - q_k(d_k) \geq \frac{\mu}{2}\,\|g_k\|\,\min\Big\{\Delta_k, \frac{\|g_k\|}{\|H_k\|}\Big\}$.

      Compute the quality indicator

        $\rho_k := \frac{f(u_k) - f(u_k + d_k)}{f(u_k) - q_k(d_k)}$.

    else
      Determine the index sets $\mathcal{A}_v(u_k,\Delta_k)$ and $\mathcal{P}(u_k,\Delta_k)$. Denote the elements
      of the powerset of $\mathcal{P}(u_k,\Delta_k)$ by $\mathcal{B}^k_1,\dots,\mathcal{B}^k_{m_k}$ with $m_k := 2^{|\mathcal{P}(u_k,\Delta_k)|}$.
```

```
            for $j = 1, \ldots, m_k$ do
              Solve the adjoint equation
                (5.19)  $A(\mathcal{A}_v(u_k, \Delta_k) \cup \mathcal{B}_j^k)\, p_j^k = \chi(\mathcal{A}_v(u_k, \Delta_k) \cup \mathcal{B}_j^k) \nabla_y J(y_k, u_k),$
              and set $g_j^k = R^\top p_j^k + \nabla_u J(y_k, u_k)$.
            end for
            Compute an inexact but feasible solution $(d_k, \zeta_k)$ of
              $(\mathfrak{Q}_k) \quad \begin{cases} \min\limits_{\zeta \in \mathbb{R}, d \in \mathbb{R}^n} & \mathfrak{q}_k(d, \zeta) := f(u_k) + \zeta + \frac{1}{2}\, d^\top H_k d \\ \text{s.t.} & \|d\| \le \Delta_k, \\ & \langle g_j^k, d \rangle \le \zeta \quad \forall j = 1, \ldots, m_k \end{cases}$
            that satisfies the modified Cauchy decrease condition
              (5.20)  $f(u_k) - \mathfrak{q}_k(d_k, \zeta_k) \ge \frac{\mu}{2}\, \psi(u_k, \Delta_k)\, \min\Big\{\Delta_k, \frac{\psi(u_k, \Delta_k)}{\|H_k\|}\Big\},$
            where the stationarity measure $\psi(u_k, \Delta_k)$ is determined by
            (5.21)
              $\psi(u_k, \Delta_k) = -\min\limits_{\xi \in \mathbb{R}, d \in \mathbb{R}^n} \{\xi : \|d\| \le 1,\ \langle g_j^k, d \rangle \le \xi \quad \forall j = 1, \ldots, m_k\}.$
            Compute the modified quality indicator
              $\rho_k := \begin{cases} \dfrac{f(u_k) - f(u_k + d_k)}{f(u_k) - \mathfrak{q}_k(d_k, \zeta_k)} & \text{if } \psi(u_k, \Delta_k) > \|g_k\|\, \Delta_k, \\ 0 & \text{if } \psi(u_k, \Delta_k) \le \|g_k\|\, \Delta_k. \end{cases}$
          end if
          Update: Set
            $u_{k+1} := \begin{cases} u_k & \text{if } \rho_k \le \eta_1 \quad \text{(null step)}, \\ u_k + d_k & \text{otherwise} \quad \text{(successful step)}, \end{cases}$
            $\Delta_{k+1} := \begin{cases} \beta_1\, \Delta_k & \text{if } \rho_k \le \eta_1, \\ \max\{\Delta_{\min}, \Delta_k\} & \text{if } \eta_1 < \rho_k \le \eta_2, \\ \max\{\Delta_{\min}, \beta_2 \Delta_k\} & \text{if } \rho_k > \eta_2. \end{cases}$
          Set $k := k + 1$.
        end if
      end for
```

Remark 5.14. The adjoint equation in (5.18) is derived as follows: According to Theorem 5.5 and the chain rule, a subgradient of the reduced objective reads

$$g_k = R^\top \chi(\mathcal{A}_s(u_k) \cup \mathcal{B}_k)^\top A(\mathcal{A}_s(u_k) \cup \mathcal{B}_k)^{-\top} \nabla_y J(y_k, u_k) + \nabla_u J(y_k, u_k).$$

From the symmetry of A and Definition 5.4, it follows that

$$\chi(\mathcal{A}_s(u_k) \cup \mathcal{B}_k)^\top A(\mathcal{A}_s(u_k) \cup \mathcal{B}_k)^{-\top} = A(\mathcal{A}_s(u_k) \cup \mathcal{B}_k)^{-1} \chi(\mathcal{A}_s(u_k) \cup \mathcal{B}_k),$$

which yields the adjoint equation in (5.18). The same argument applies to (5.19).

Remark 5.15. Completely analogously to Lemma 4.2, one can show that (5.21) is equal to the stationarity measure from Assumption 2.4(2b).

THEOREM 5.16. *Suppose that Algorithm* 5.13 *does not terminate in finitely many steps and that Assumption* 2.9 *holds. Then every accumulation point of the sequence of iterates is C-stationary.*

Proof. As seen in Lemma 4.2, since the inexact but feasible solution (d_k, ζ_k) of $(\mathfrak{Q}_k)$ satisfies (5.20), d_k also fulfills the modified decrease condition (2.5). Therefore, we can apply the results for our general Algorithm 2.5. As Lemmas 5.8, 5.11, and 5.12 show, the conditions in Assumptions 4.1 and 4.4, which guarantee the convergence of our trust-region algorithm for problems with composite functions of the form (4.1), are fulfilled in this concrete setting. Thus, Corollary 4.8 yields the claim. □

Remark 5.17. The main drawback of Algorithm 5.13 is, of course, the complexity of the complicated model in $(\mathfrak{Q}_k)$. This, however, is a deficit that is shared by all algorithmic approaches that ensure convergence without further assumptions like, e.g., approximate convexity. Compare, for instance, with gradient sampling methods in this context which require at least $n+1$ gradient evaluations in every iteration; cf. [21]. We remark that, in all of our numerical experiments, the complicated model was only needed to detect stationarity. The stationarity measure, however, can be efficiently approximated (see subsection 6.1 below) so that the complicated model in Algorithm 5.13 is—in practice—indeed only needed in pathological cases. Nonetheless, the reduction of the complexity of the complicated model without impairing the convergence analysis is a major challenge for future research.

6. Numerical experiments and validation.

To verify the theoretical results of the last sections numerically, in what follows, we will apply our trust-region method to the finite-dimensional minimization problem that arises when piecewise affine finite elements on a uniform Friedrichs–Keller triangulation and a standard mass-lumping scheme are used to discretize a tracking-type optimal control problem of the form

$$(\mathrm{P}_{\mathrm{ex}}^{\mathrm{c}})\quad \begin{cases} \min & \dfrac{1}{2}\displaystyle\int_\Omega (y-y_d)^2 \mathrm{d}x + \dfrac{\alpha}{2}\int_\Omega (u-u_d)^2\mathrm{d}x \\ \text{w.r.t.} & u \in L^2(\Omega),\ y \in H_0^1(\Omega), \\ \text{s.t.} & \displaystyle\int_\Omega \nabla y\cdot\nabla(v-y)\mathrm{d}x + \|v\|_{L^1(\Omega)} - \|y\|_{L^1(\Omega)} \geq \langle u, v-y\rangle \quad \forall v \in H_0^1(\Omega) \end{cases}$$

involving a Tikhonov parameter $\alpha > 0$, a desired state y_d, a desired control u_d, and a bounded, polyhedral domain $\Omega \subset \mathbb{R}^d$ (cf. [26]). Our model problem thus reads as follows:

$$(\mathrm{P}_{\mathrm{ex}})\quad \begin{cases} \displaystyle\min_{y\in\mathbb{R}^m, u\in\mathbb{R}^n} & J(y,u) := \dfrac{1}{2}\langle M_D y, y\rangle - \langle y, R y_d\rangle + \dfrac{\alpha}{2}\langle M(u-u_d), u-u_d\rangle \\ \text{s.t.} & \langle \nu^{-1}Ay, v-y\rangle + \|v\|_1 - \|y\|_1 \geq \langle \nu^{-1}Ru, v-y\rangle \quad \forall v\in\mathbb{R}^m. \end{cases}$$

Here, $M_D \in \mathbb{R}^{m\times m}$, $M \in \mathbb{R}^{n\times n}$, and $R \in \mathbb{R}^{m\times n}$ are the mass matrices that are obtained when the L^2-scalar product on Ω is discretized (taking into account the homogeneous Dirichlet boundary conditions of the states), A is the stiffness matrix associated with the finite element discretization, $\alpha > 0$ is the Tikhonov parameter, $\nu > 0$ is the weighting factor obtained from the mass-lumping discretization of the $L^1(\Omega)$-norm, and $u_d \in \mathbb{R}^n$ and $y_d \in \mathbb{R}^n$ are the function values of the desired control and the desired state at the nodes of the underlying triangulation of Ω, respectively.

(We denote these vectors with the same symbols as their continuous counterparts for the sake of simplicity.) For more details on how the matrices M_D, M, R, and A arise in the above context, we refer the reader to [9]. Note that (P_{ex}) is precisely of the form (P_{VI}). The analysis of the previous sections is thus indeed applicable here.

6.1. Algorithmic realization.

Before we present the results that we obtained by applying Algorithm 5.13 to (P_{ex}), some remarks are in order.

Mesh-independent norms. As (P_{ex}) arises from the finite element discretization of an infinite-dimensional optimal control problem, in order to be able to compare numerical results on different discretization levels (i.e., on finite element meshes with different widths), it is necessary to work with mesh-independent norms. Therefore, in our experiments, we exchanged the Euclidean norm and scalar product appearing in our algorithm with the norm $\|\cdot\|_M$ and scalar product $\langle\cdot,\cdot\rangle_M$ associated with the mass matrix M, respectively. Note that the Lipschitz constant L_y in Lemma 5.9 is not affected by this modification, whereas the Lipschitz constant of q is scaled by the inverse of the minimal eigenvalue of M.

Stopping criterion. As already mentioned in Remark 5.17, the major drawback of Algorithm 5.13 is that every evaluation of its inner, complicated model (and thus every iteration with $\Delta_k \leq \Delta_{\min}$) requires the solution of $m_k = 2^{|\mathcal{P}(u_k,\Delta_k)|}$ adjoint equations. As far as the computation of the stationarity measure is concerned, this issue can be resolved (at least heuristically) by running the following algorithm at the beginning of each iteration which makes use of the complicated model.

Algorithm 6.1 (successive approximation of the stationarity measure).

```
for $j = 1, \dots, m_k$ do
    Solve the adjoint equation
        $$A(\mathcal{A}_v(u_k,\Delta_k)\cup\mathcal{B}_j^k)\, p_j^k = \chi(\mathcal{A}_v(u_k,\Delta_k)\cup\mathcal{B}_j^k)\nabla_y J(y_k,u_k),$$
    and set $g_j^k = R^\top p_j^k + \nabla_u J(y_k,u_k)$.
    Compute the following approximation of the stationarity measure:
        (6.1)  $$\psi_j(u_k,\Delta_k) = -\min_{\xi\in\mathbb{R},d\in\mathbb{R}^n}\{\xi : \|d\|\leq 1,\ \langle g_\ell^k, d\rangle \leq \xi \quad \forall \ell = 1,\dots,j\}.$$
    if $\psi_j(u_k,\Delta_k) \leq \texttt{tol}$ then
        STOP the iteration
    end if
end for
Set $\psi(u_k,\Delta_k) := \psi_{m_k}(u_k,\Delta_k)$.
```

The stopping criterion in Step 4 of the above algorithm is motivated as follows: Analogously to the proof of Lemma 4.7, one shows that

$$\|\operatorname{proj}_{\mathcal{C}_k}(0)\| \leq \psi(u_k,\Delta_k) \leq \psi_j(u_k,\Delta_k) \leq \texttt{tol} \quad \Longrightarrow \quad 0 \in \mathcal{C}_k + B_{\texttt{tol}}(0),$$

where $\mathcal{C}_k$ is the set from (4.9) and $\operatorname{proj}_{\mathcal{C}_k}$ is the projection onto this convex and closed set. If now Δ_k is sufficiently small, then, by construction, $\mathcal{G}(u_k,\Delta_k) = \partial_B S(u_k)$ (see Definition 5.10), which in turn implies $\mathcal{C}_k \subset \partial f(u_k)$; cf. (4.8). Thus, in this case, the iterate is approximately C-stationary (up to the tolerance tol). Together with the stopping criterion in Step 6 of Algorithm 5.13 (in practice, of course, replaced by $\|g_k\| \leq \texttt{tol}$), Algorithm 6.1 therefore provides a termination criterion which is able to detect convergence even before the whole stationarity measure is computed. In the experiments below, we have always used the above algorithm.

The complicated subproblems and their solution. In contrast to the computation of the stationarity measure, it remains an open question if solutions of $(\mathfrak{Q}_k)$ can be approximated successively by iteratively considering more and more subgradients g_j^k without compromising the convergence behavior of the overall algorithm. This is subject to future research. However, in all of our numerical experiments presented below, the complicated model was *never* needed to prevent the algorithm from stalling at a nonstationary point. We only needed the complicated model to detect stationarity, which can be performed by Algorithm 6.1 as explained above. It moreover turned out that we *never* needed the whole number of m_k iterations to obtain a sufficiently small stationarity measure. Instead, the stationarity measure $\psi_j(u_k, \Delta_k)$ already fell below the tolerance for small values of j; see Table 2 below. For small values of j, the minimization problem in (6.1) can be solved by the MATLAB-inbuilt routine `fmincon`. In summary, we observed that the potential exponential complexity of the proposed algorithm did not play any role in our numerical computations. Nevertheless, an efficient successive approximation of the solution of $(\mathfrak{Q}_k)$ ensuring convergence of the method is an important issue that should be addressed in future research.

Further implementation details. Let us finally specify the remaining parts of the algorithm that we used for our numerical experiments: The solution of the VI in Step 3 was computed by means of an active set method applied to its dual problem. Globalization efforts were not necessary at this point; the active set iteration behaved robustly and converged quickly in all examples. The simple trust-region subproblems (Q_k) were approximately solved by the dogleg method. The matrices H_k were assembled via the inverse BFGS-update (w.r.t. the scalar product induced by the mass matrix) based on the subgradients computed in Step 4. In order to fulfill Assumption 2.9, we reset H_k to the identity (respectively, the mass matrix), when $\|H_k\|_\infty$ surpassed a given threshold (set to h^{-3}, where h denotes the width of the finite element mesh). Moreover, to avoid numerical instabilities, we also reset H_k in every 50th iteration. We remark that we have no theoretical evidence that such a BFGS-update accelerates the convergence. This goes beyond the scope of this work and has to be investigated in future research.

Chosen parameters. Unless stated otherwise, the parameters in Algorithm 5.13 were set as follows: For the update of the trust-region radius, we chose $\eta_1 = 0.1$, $\eta_2 = 0.9$, $\beta_1 = 0.5$, and $\beta_2 = 1.5$. The parameter in the Cauchy decrease condition was set to $\mu = 0.8$. The initial trust-region radius was $\Delta_0 = 10$. The Tikhonov parameter in the objective of $(\mathrm{P}_{\mathrm{ex}})$ was set to $\alpha = 10^{-4}$. The termination tolerance was $\mathtt{tol} = 10^{-5}$. For the threshold separating the simple from the complicated model, we chose $\Delta_{\min} = 10^{-6}$. (In some of the following numerical tests, $\Delta_{\min}$ was varied to investigate its influence on the performance of the algorithm.) As the underlying domain Ω, we considered the unit square $(0, 1)^2$. The piecewise affine finite element discretization was carried out on uniform Friedrichs–Keller triangulations with widths $h = 0.04$, $h = 0.02$, and $h = 0.01$. (Note that the matrices M_D, M, R, and A as well as the parameter ν are uniquely determined by these choices.)

Test cases. It is well known that constructing (locally) optimal, analytic solutions to infinite-dimensional VI-constrained optimal control problems is a nontrivial task. While the first-order necessary optimality conditions of such problems may become rather intricate (see [15, section 6.1.1]), second-order sufficient conditions ensuring (isolated) local optimality are, to the best of the authors' knowledge, currently unknown for this problem class (at least as far as VIs of the second kind are concerned). In order to construct test scenarios with known solutions for our algorithm, we therefore started with functions y_d, u_d, which are known to solve the variational inequality in ($\mathrm{P}_{\mathrm{ex}}^c$) and thus represent global minimizers of ($\mathrm{P}_{\mathrm{ex}}^c$) under the VI-constraint. For this purpose, it is beneficial to reformulate the VI in ($\mathrm{P}_{\mathrm{ex}}^c$) with the help of a slack variable $q \in L^\infty(\Omega)$ as follows:

$$\begin{aligned} -\Delta y + q = u \quad &\text{in } H^{-1}(\Omega), \\ |q(x)| \leq 1, \quad q(x)\, y(x) = |y(x)| \quad &\text{for almost all (f.a.a.) } x \in \Omega; \end{aligned} \tag{6.2}$$

see [24] and (5.1) for the finite-dimensional counterpart. Note that the solution map of the above problem is not Gâteaux differentiable in points u for which the bi-active set $\{x \in \Omega : y(x) = 0,\, |q(x)| = 1\}$ has positive Lebesgue measure. We will specifically consider examples below which provide this feature. The test cases that we applied our trust-region method to were as follows:

(a) *Generic test.* In our first test scenario, we chose y_d and u_d to be given by

$$y_d(x_1, x_2) = \sin(\pi\, x_1)\, \sin(\pi\, x_2) \qquad \text{and} \qquad u_d \equiv 0.$$

Note that the above y_d is positive everywhere in Ω. It is thus to be expected that the problems (P_{ex}) and ($\mathrm{P}_{\mathrm{ex}}^c$) are very well behaved for this choice of y_d and u_d.

(b) *Whole domain bi-active.* In our second test, the desired control, the desired state, and the associated multiplier in (6.2) were defined as

$$y_d \equiv 0, \qquad q_d \equiv 1, \qquad u_d = -\Delta y_d + q_d \equiv 1$$

so that the whole of Ω is bi-active. This can be seen as a worst-case example.

(c) *Half of the domain bi-active.* Third, we considered functions y_d and u_d for which the bi-active set equals $(0.5, 1) \times (0, 1)$:

$$y_d(x_1, x_2) = \begin{cases} \sin^2(2\pi\, x_1)\, \sin(\pi\, x_2), & x_1 < 0.5, \\ 0, & x_1 \geq 0.5, \end{cases} \qquad q_d \equiv 1,$$

$$u_d(x_1, x_2) = \begin{cases} \frac{\pi^2}{2}\big(1 - 17\cos(4\pi\, x_1)\big)\, \sin(\pi\, x_2) - 1, & x_1 < 0.5, \\ 1, & x_1 \geq 0.5. \end{cases}$$

As we will see below, even for the last two configurations, the complicated model was not needed to solve (P_{ex}) up to the desired tolerance of 10^{-5}. Actually, it was all but easy to find examples, where the complicated model occurred during the iteration, and, as already mentioned, we were only able to find scenarios where the complicated model was needed to detect stationarity. These read as follows:

(d) *Large desired control.*

$$y_d \equiv 0, \qquad u_d \equiv 50.$$

(e) *Discontinuous desired state.*

$$y_d(x_1, x_2) = \begin{cases} 0, & x_1 \leq 0.5, \\ 0.895, & x_1 > 0.5, \end{cases} \qquad u_d \equiv 0.$$

Note that, in (b) and (c), y_d solves the VI in (P_{ex}) with right-hand side u_d. For these cases, the solution of (P_{ex}) is thus known, as explained above.

6.2. Numerical results. We start with the first three scenarios (a), (b), and (c). In these tests, the simple model suffices to reach the tolerance `tol` $= 10^{-5}$. The results are shown in Table 1. Herein, the number in the brackets behind the total number of iterations refers to the number of successful iterations (in contrast to null steps). The fourth column shows the discrete L^2-norm of the subgradient calculated in Step 4 of Algorithm 5.13 in the last iteration. Furthermore, for the scenarios (b) and (c), the (discrete) L^2-distance to the known globally optimal solution (y_d, u_d) is shown in columns five and six (in relative norms, except for the difference in the state in case (b) since y_d vanishes identically there).

Table 1
Numerical results obtained by applying Algorithm 5.13 *to the test cases* (a), (b)*, and* (c).

Scenario	Mesh size h	#TR-iterations	$\|g\|_M$	$\|y - y_d\|_M$	$\|u - u_d\|_M$
(a)	0.04	7(7)	3.1848 e-6	–	–
(a)	0.02	7(7)	3.2928 e-6	–	–
(a)	0.01	7(7)	3.3231 e-6	–	–
(b)	0.04	30(30)	2.3893 e-6	8.4075 e-5	9.7765 e-3
(b)	0.02	32(32)	5.5489 e-6	5.5489 e-6	2.8709 e-2
(b)	0.01	46(29)	8.9626 e-6	9.1688 e-6	8.9616 e-2
(c)	0.04	33(27)	8.1245 e-6	8.7432 e-3	7.6691 e-3
(c)	0.02	33(33)	7.2360 e-6	1.3241 e-2	1.5274 e-2
(c)	0.01	24(24)	6.7408 e-6	5.3932 e-3	5.8847 e-3

We observe that, in test case (a), where there is no active set, our algorithm indeed performs as efficiently as a classical trust-region method. Turning to scenarios (b) and (c), where the solution map of the VI in ($\mathrm{P}^c_{\mathrm{ex}}$) is nondifferentiable in the global optimum (y_d, u_d), the iteration numbers increase but still remain moderate. In all of these examples, the algorithm behaved approximately mesh-independently.

Let us now turn to scenarios (d) and (e), where the complicated model was indeed necessary to fulfill the termination criterion. The results obtained for these cases can be seen in Table 2. Here, the mesh size was fixed ($h = 0.01$ for (d) and $h = 0.04$ for (e)), while $\Delta_{\min}$, i.e., the threshold between the simple and the complicated models, was varied. We again list the number of iterations and the number of successful steps in brackets. The discrete L^2-norm of the subgradient from Step 4 of Algorithm 5.13 computed in the last iteration performed with the simple model is shown in the fourth column of the table. Furthermore, "#bi-active" refers to the cardinality of $\mathcal{P}(u_k, \Delta_k)$ determining the complexity of the complicated model; see Step 12. Under the label "#subgrad" we list the number of iterations of Algorithm 6.1 needed to approximate the stationarity measure up to the desired tolerance `tol` $= 10^{-5}$. The corresponding result is shown in the last column.

One observes that in cases (d) and (e) the number of iterations is significantly higher than in the previous examples. Moreover, there is a substantial number of null steps in these two test cases. After performing a critical number of null steps, the trust-region radius reaches the threshold $\Delta_{\min}$, but the norm of the subgradient is still larger than the tolerance. Algorithm 5.13 then switches to the complicated model and computes the successive approximation of the stationarity measure according to Algorithm 6.1. Except for one test case, only two iterations were necessary to drive $\psi_j(u, \Delta)$ below the desired tolerance. In particular, in scenario (d), the number of elements in the set of possibly bi-active points determining the size of the complicated

TABLE 2
Numerical results obtained by applying Algorithm 5.13 *to the test cases* (d) *and* (e).

Scenario	$\Delta_{\min}$	#TR-it.	$\Vert g\Vert_M$	#bi-active	#subgrad.	$\psi_j(u,\Delta)$
(d)	10^{-4}	171(109)	9.3878 e-5	1549	2	8.2590 e-6
(d)	10^{-5}	176(110)	9.3853 e-5	201	2	8.2920 e-6
(d)	10^{-6}	179(110)	9.3855 e-5	61	2	8.2920 e-6
(e)	10^{-4}	144(94)	1.4461 e-5	23	4	7.6600 e-6
(e)	10^{-5}	152(99)	1.3716 e-5	13	2	8.2597 e-6
(e)	10^{-6}	196(128)	1.2147 e-5	2	2	9.0993 e-6

model is so large that an efficient solution of the complicated model in $(\mathfrak{Q}_k)$ would not be possible. This again calls for an efficient approximation of the solution of $(\mathfrak{Q}_k)$, which is subject to future research.

Acknowledgments. The authors thank Stephan Walther (TU Dortmund) for carrying out the proof of Lemma 3.4. We are moreover grateful to Maximilian Sperber and Sebastian Hillbrecht (both TU Dortmund) for their support with regard to the implementation and coding.

REFERENCES

[1] Z. Akbari, R. Yousefpour, and M. Reza Peyghami, *A new nonsmooth trust region algorithm for locally Lipschitz unconstrained optimization problems*, J. Optim. Theory Appl., 164 (2015), pp. 733–754.

[2] P. Apkarian, D. Noll, and L. Ravanbod, *Nonsmooth bundle trust-region algorithm with applications to robust stability*, Set-Valued Var. Anal., 24 (2016), pp. 115–148.

[3] A. Bagirov, N. Karmitsa, and M. M. Mäkelä, *Introduction to Nonsmooth Optimization*, Springer, New York, 2014.

[4] A. M. Bagirov, L. Jin, N. Karmitsa, A. Al Nuaimat, and N. Sultanova, *Subgradient method for nonconvex nonsmooth optimization*, J. Optim. Theory Appl., 157 (2013), pp. 416–435.

[5] A. M. Bagirov, B. Karasözen, and M. Sezer, *Discrete gradient method: Derivative-free method for nonsmooth optimization*, J. Optim. Theory Appl., 137 (2008), pp. 317–334.

[6] V. Barbu, *Necessary conditions for distributed control problems governed by parabolic variational inequalities*, SIAM J. Control Optim., 19 (1981), pp. 64–86, https://doi.org/10.1137/0319006.

[7] V. Barbu, *Optimal Control of Variational Inequalities*, Res. Notes Math. 100, Pitman, Boston, MA, 1984.

[8] A. Bihain, *Optimization of upper semidifferentiable functions*, J. Optim. Theory Appl., 44 (1984), pp. 545–568.

[9] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, 3rd ed., Springer, New York, 2008.

[10] J. V. Burke, A. Lewis, and M. L. Overton, *A robust gradient sampling algorithm for nonsmooth, nonconvex optimization*, SIAM J. Optim., 15 (2005), pp. 751–779, https://doi.org/10.1137/030601296.

[11] J. V. Burke, A. S. Lewis, and M. L. Overton, *A nonsmooth, nonconvex optimization approach to robust stabilization by static output feedback and low-order controllers*, IFAC Proc. Vol., 36 (2003), pp. 175–181.

[12] L. Calatroni, C. Chung, J. C. De los Reyes, C.-B. Schönlieb, and T. Valkonen, *Bilevel approaches for learning of variational imaging models*, in Variational Methods, Radon Ser. Comput. Appl. Math. 18, De Gruyter, Berlin, 2017, pp. 252–290.

[13] C. Carstensen, B. D. Reddy, and M. Schedensack, *A natural nonconforming FEM for the Bingham flow problem is quasi-optimal*, Numer. Math., 133 (2016), pp. 37–66.

[14] Y. J. Chen, T. Pock, and H. Bischof, *Learning l_1-based analysis and synthesis sparsity priors using bi-level optimization*, in Workshop on Analysis Operator Learning vs. Dictionary Learning, NIPS2012, 2012.

[15] C. Christof, *Sensitivity Analysis of Elliptic Variational Inequalities of the First and the Second Kind*, Ph.D. thesis, Technische Universität Dortmund, Dortmund, Germany, 2018.

[16] C. Christof, *Gradient-based solution algorithms for a class of bilevel optimization and optimal control problems with a nonsmooth lower level*, SIAM J. Optim., 30 (2020), pp. 290–318, https://doi.org/10.1137/18M1225707.

[17] C. Christof, C. Clason, C. Meyer, and S. Walther, *Optimal control of a non-smooth semilinear elliptic equation*, Math. Control Relat. Fields, 8 (2018), pp. 247–276.

[18] C. Christof, J. C. De los Reyes, and C. Meyer, *A Non-Smooth Trust-Region Method for B-Differentiable Functions with Application to Optimization Problems Constrained by Variational Inequalities*, preprint, https://arxiv.org/abs/1711.03208v1, 2017.

[19] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Classics Appl. Math. 5, SIAM, Philadelphia, 1990, https://doi.org/10.1137/1.9781611971309.

[20] B. Colson, P. Marcotte, and G. Savard, *A trust-region method for nonlinear bilevel programming: Algorithm and computational experience*, Comput. Optim. Appl., 30 (2005), pp. 211–227.

[21] F. E. Curtis, T. Mitchell, and M. Overton, *A BFGS-SQP method for nonsmooth, nonconvex, constrained optimization and its evaluation using relative minimization profiles*, Optim. Methods Softw., 32 (2017), pp. 148–181.

[22] F. E. Curtis and X. Que, *An adaptive gradient sampling algorithm for non-smooth optimization*, Optim. Methods Softw., 28 (2013), pp. 1302–1324.

[23] A. Daniilidis and P. Georgiev, *Approximate convexity and submonotonicity*, J. Math. Anal. Appl., 291 (2004), pp. 292–301.

[24] J. C. De los Reyes, *Optimal control of a class of variational inequalities of the second kind*, SIAM J. Control Optim., 49 (2011), pp. 1629–1658, https://doi.org/10.1137/090764438.

[25] J. C. De los Reyes, *Optimization of mixed variational inequalities arising in flow of viscoplastic materials*, Comput. Optim. Appl., 52 (2012), pp. 757–784.

[26] J. C. De los Reyes and C. Meyer, *Strong stationarity conditions for a class of optimization problems governed by variational inequalities of the second kind*, J. Optim. Theory Appl., 168 (2016), pp. 375–409.

[27] E. J. Dean, R. Glowinski, and G. Guidoboni, *On the numerical simulation of Bingham visco-plastic flow: Old and new results*, J. Non-Newton. Fluid Mech., 142 (2007), pp. 36–62.

[28] J. E. Dennis, S.-B. B. Li, and R. A. Tapia, *A unified approach to global convergence of trust region methods for nonsmooth optimization*, Math. Program., 68 (1995), pp. 319–346.

[29] G. Emiel and C. Sagastizábal, *Incremental-like bundle methods with application to energy planning*, Comput. Optim. Appl., 46 (2010), pp. 305–332.

[30] M. Fuchs and G. Seregin, *Variational Methods for Problems from Plasticity Theory and for Generalized Newtonian Fluids*, Springer, Berlin, 2000.

[31] A. Fuduli, M. Gaudioso, and G. Giallombardo, *A DC piecewise affine model and a bundling technique in nonconvex nonsmooth minimization*, Optim. Methods Softw., 19 (2004), pp. 89–102.

[32] G. Giallombardo and D. Ralph, *Multiplier convergence in trust-region methods with application to convergence of decomposition methods for MPECs*, Math. Program., 112 (2008), pp. 335–369.

[33] A. A. Goldstein, *Optimization of Lipschitz continuous functions*, Math. Program., 13 (1977), pp. 14–22.

[34] W. Hare and C. Sagastizábal, *A redistributed proximal bundle method for nonconvex optimization*, SIAM J. Optim., 20 (2010), pp. 2442–2473, https://doi.org/10.1137/090754595.

[35] L. Hertlein and M. Ulbrich, *An inexact bundle algorithm for nonconvex nonsmooth minimization in Hilbert space*, SIAM J. Control Optim., 57 (2019), pp. 3137–3165, https://doi.org/10.1137/18M1221849.

[36] M. Hintermüller, *A proximal bundle method based on approximate subgradients*, Comput. Optim. Appl., 20 (2001), pp. 245–266.

[37] M. Hintermüller and T. M. Surowiec, *A bundle-free implicit programming approach for a class of elliptic MPECs in function space*, Math. Program., 160 (2016), pp. 271–305.

[38] R. R. Huilgol and Z. You, *Application of the augmented Lagrangian method to steady pipe flows of Bingham, Casson and Herschel-Bulkley fluids*, J. Non-Newton. Fluid Mech., 128 (2005), pp. 126–143.

[39] K. Ito and K. Kunisch, *Optimal control of parabolic variational inequalities*, J. Math. Pures Appl., 93 (2010), pp. 329–360.

[40] N. Karmitsa, A. Bagirov, and M. M. Mäkelä, *Comparing different nonsmooth minimization methods and software*, Optim. Methods Softw., 27 (2012), pp. 131–153.

[41] N. Karmitsa and M. M. Mäkelä, *Limited memory bundle method for large bound constrained nonsmooth optimization: Convergence analysis*, Optim. Methods Softw., 25 (2010), pp. 895–916.

[42] A. M. Khludnev and J. Sokołowski, *Modelling and Control in Solid Mechanics*, Birkhäuser Boston, Boston, 1991.

[43] N. Kikuchi and J. T. Oden, *Contact Problems in Elasticity*, Stud. Appl. Math. 8, SIAM, Philadelphia, 1988, https://doi.org/10.1137/1.9781611970845.

[44] K. C. Kiwiel, *Convergence of the gradient sampling algorithm for nonsmooth nonconvex optimization*, SIAM J. Optim., 18 (2007), pp. 379–388, https://doi.org/10.1137/050639673.

[45] K. C. Kiwiel, *Improved convergence result for the discrete gradient and secant methods for nonsmooth optimization*, J. Optim. Theory Appl., 144 (2009), pp. 69–75.

[46] K. Kunisch and T. Pock, *A bilevel optimization approach for parameter learning in variational models*, SIAM J. Imaging Sci., 6 (2013), pp. 938–983, https://doi.org/10.1137/120882706.

[47] C. Lemaréchal, *A view of line-searches*, in Optimization and Optimal Control, A. Auslender, W. Oettli, and J. Stoer, eds., Springer, Berlin, 1981, pp. 59–78.

[48] C. Lemaréchal, J. J. Strodiot, and A. Bihain, *On a bundle algorithm for nonsmooth optimization*, in Nonlinear Programming 4, O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, eds., Academic Press, New York, 1981, pp. 245–282.

[49] A. S. Lewis, *Active sets, nonsmoothness, and sensitivity*, SIAM J. Optim., 13 (2002), pp. 702–725, https://doi.org/10.1137/S1052623401387623.

[50] A. S. Lewis and M. L. Overton, *Nonsmooth optimization via quasi-Newton methods*, Math. Program., 141 (2013), pp. 135–163.

[51] A. S. Lewis and S. J. Wright, *A proximal method for composite minimization*, Math. Program., 158 (2016), pp. 501–546.

[52] L. Lukšan and J. Vlček, *A bundle-Newton method for nonsmooth unconstrained minimization*, Math. Program., 83 (1998), pp. 373–391.

[53] Z.-Q. Luo, J.-S. Pang, and D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, UK, 1996.

[54] N. Mahdavi-Amiri and R. Yousefpour, *An effective nonsmooth optimization algorithm for locally Lipschitz functions*, J. Optim. Theory Appl., 155 (2012), pp. 180–195.

[55] M. M. Mäkelä, N. Karmitsa, and A. Bagirov, *Subgradient and bundle methods for nonsmooth optimization*, in Numerical Methods for Differential Equations, Optimization, and Technological Problems, Springer, Dordrecht, The Netherlands, 2013, pp. 275–304.

[56] P. Marcotte, G. Savard, and D. L. Zhu, *A trust region algorithm for nonlinear bilevel programming*, Oper. Res. Lett., 29 (2001), pp. 171–179.

[57] R. Mifflin, *An algorithm for constrained optimization with semismooth functions*, Math. Oper. Res., 2 (1977), pp. 191–207.

[58] P. P. Mosolov and V. P. Miasnikov, *Variational methods in the theory of the fluidity of a viscous-plastic medium*, J. Appl. Math. Mech., 29 (1965), pp. 545–577.

[59] P. P. Mosolov and V. P. Miasnikov, *On stagnant flow regions of a viscous-plastic medium in pipes*, PMM, 30 (1966), pp. 705–717.

[60] P. P. Mosolov and V. P. Miasnikov, *On qualitative singularities of the flow of a viscoplastic medium in pipes*, J. Appl. Math. Mech., 31 (1967), pp. 609–613.

[61] H. V. Ngai, D. T. Luc, and M. Thera, *Approximate convex functions*, J. Nonlinear Convex Anal., 1 (2011), pp. 155–176.

[62] D. Noll, *Cutting plane oracles to minimize non-smooth non-convex functions*, Set-Valued Var. Anal., 18 (2010), pp. 531–568.

[63] D. Noll, *Bundle method for non-convex minimization with inexact subgradients and function values*, in Computational and Analytical Mathematics, D. H. Bailey, H. H. Bauschke, P. Borwein, F. Garvan, M. Théra, J. D. Vanderwerff, and H. Wolkowicz, eds., Springer, New York, 2013, pp. 555–592.

[64] D. Noll, *Convergence of non-smooth descent methods using the Kurdyka–Lojasiewicz inequality*, J. Optim. Theory Appl., 160 (2014), pp. 553–572.

[65] D. Noll, *Cutting plane oracles for non-smooth trust-regions*, Pure Appl. Funct. Anal., Special Issue Dedicated to the Memory of Jon Borwein, to appear.

[66] J. Outrata, M. Kocvara, and J. Zowe, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications and Numerical Results*, Springer, Dordrecht, The Netherlands, 1998.

[67] J. Outrata and J. Zowe, *A numerical approach to optimization problems with variational inequality constraints*, Math. Program., 68 (1995), pp. 105–130.

[68] M. J. D. Powell, *A hybrid method for nonlinear equations*, in Numerical Methods for Nonlinear Algebraic Equations, P. Rabinowitz, ed., Gordon and Breach, London, 1970.

[69] L. Qi and J. Sun, *A trust region algorithm for minimization of locally Lipschitzian functions*, Math. Program., 66 (1994), pp. 25–43.

[70] A.-T. Rauls and S. Ulbrich, *Computation of a Bouligand generalized derivative for the solution operator of the obstacle problem*, SIAM J. Control Optim., 57 (2019), pp. 3223–3248, https://doi.org/10.1137/18M1187283.

[71] A.-T. Rauls and G. Wachsmuth, *Generalized derivatives for the solution operator of the obstacle problem*, Set-Valued Var. Anal., 28 (2020), pp. 259–285.

[72] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Grundlehren Math. Wiss. 317, Springer, Berlin, 1998.

[73] C. Sagastizábal, *Composite proximal bundle method*, Math. Program., 140 (2013), pp. 189–233.

[74] W. Schirotzek, *Nonsmooth Analysis*, Springer, Berlin, 2007.

[75] S. Scholtes, *Introduction to Piecewise Differentiable Equations*, Springer, New York, 2012.

[76] S. Scholtes and M. Stöhr, *Exact penalization of mathematical programs with equilibrium constraints*, SIAM J. Control Optim., 37 (1999), pp. 617–652, https://doi.org/10.1137/S0363012996306121.

[77] H. Schramm, *Eine Kombination von Bundle- und Trust-Region-Verfahren zur Lösung nichtdifferenzierbarer Optimierungsprobleme*, Ph.D. thesis, Universität Bayreuth, Bayreuth, Germany, 1989.

[78] H. Schramm and J. Zowe, *A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, convergence analysis, numerical results*, SIAM J. Optim., 2 (1992), pp. 121–152, https://doi.org/10.1137/0802008.

[79] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*, Springer Ser. Comput. Math. 3, Springer, New York, 1985.

[80] T. Steihaug, *The conjugate gradient method and trust regions in large scale optimization*, SIAM J. Numer. Anal., 20 (1983), pp. 626–637, https://doi.org/10.1137/0720042.

[81] T. M. Surowiec, *Numerical Optimization Methods for the Optimal Control of Elliptic Variational Inequalities*, Springer, New York, 2018, pp. 123–170.

[82] G. Wachsmuth, *Towards M-stationarity for optimal control of the obstacle problem with control constraints*, SIAM J. Control Optim., 54 (2016), pp. 964–986, https://doi.org/10.1137/140980582.