DFG Deutsche Forschungsgemeinschaft
Priority Programme 1962

A Non-smooth Trust-region Method for Locally Lipschitz Functions with Application to Optimization Problems Constrained by Variational Inequalities

Constantin Christof, Juan Carlos de los Reyes, Christian Meyer

SPP 1962 | Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization

Preprint Number SPP1962-051

received on February 15, 2018

Edited by
SPP1962 at Weierstrass Institute for Applied Analysis and Stochastics (WIAS)
Leibniz Institute in the Forschungsverbund Berlin e.V.
Mohrenstraße 39, 10117 Berlin, Germany
E-Mail: spp1962@wias-berlin.de

World Wide Web: http://spp1962.wias-berlin.de/

A NON-SMOOTH TRUST-REGION METHOD FOR LOCALLY LIPSCHITZ FUNCTIONS WITH APPLICATION TO OPTIMIZATION PROBLEMS CONSTRAINED BY VARIATIONAL INEQUALITIES

C. CHRISTOF[‡], J. C. DE LOS REYES[§], AND C. MEYER[‡]

Abstract. We propose a non-smooth trust–region method for solving optimization problems with locally Lipschitz continuous functions, with application to problems constrained by variational inequalities of the second kind. Under suitable assumptions on the model functions, convergence of the general algorithm to a C– stationary point is verified. For variational inequality constrained problems, we are able to properly characterize the Bouligand subdifferential of the reduced cost function and, based on that, we propose a computable trust–region model which fulfills the convergence hypotheses of the general algorithm. The article concludes with the experimental study of the main properties of the proposed method based on two different numerical instances.

Key words. Trust–region methods, Bouligand differentiability, stationarity conditions, optimization with variational inequality constraints.

1. Introduction. The study of optimization problems with variational inequality (VI) constraints is a challenging topic in mathematical programming, due to the intricate structure of the type of stationary points one aims to reach. Whereas for differentiable problems only one type of stationarity takes place, in the nonsmooth framework a family of concepts arise (e.g., Clarke, Dini, Bouligand or Mordukhovich stationarity), each one with advantages and shortcomings.

To overcome the difficulties related to the nonsmoothness of these types of problems, relaxation or penalization approaches have been frequently proposed, with different outcomes concerning optimality conditions and solution algorithms. The main criticism to these approaches, however, is that by removing/relaxing the nonsmoothness, the structure of the original problem is altered, possibly leading to undesired or unphysical solutions.

In the special case of *Mathematical Programs with Equilibrium Constraints (MPEC)* intensive efforts have been carried out and different methods proposed or extended (see [17] and the references therein). Typically, convergence to C-stationary points can be guaranteed for the proposed algorithms. The search for stronger (like B– or M–) stationary points remains a challenge.

In the case of optimization problems with variational inequality constraints of the second kind, much less work has been carried out. In [7] a semismooth Newton method based on a regularized version of the problem was proposed and superlinear convergence verified. The solution, however, corresponds to a regularized problem and not to the original one, although consistency is also proved there. A first attempt to find solutions without using regularization was tested in [10], where a trust–region algorithm was considered with promising results.

Trust-region methods have been investigated for nonsmooth optimization of locally Lipschitz continuous functions in, e.g., [2, 11, 20, 21, 26]. The underlying hypotheses, however, are difficult to verify for optimization problems with variational inequality

[‡]Faculty of Mathematics, Technische Universität Dortmund, Germany.
[§]Research Center on Mathematical Modelling (MODEMAT), Escuela Politécnica Nacional, Quito, Ecuador

constraints. This especially concerns the hypothesis that the cost function has to be *regular* (see [25] for a discussion on this matter), which happens to be very restrictive for problems as the ones considered in this manuscript. In combination with bundle methods, a convergence theory based on weaker assumptions, similar to ours, was recently studied in [2].

The purpose of this paper is twofold. First, we propose a general non-smooth trust-region method for locally Lipschitz continuous functions, and carry out the convergence analysis of the approach. Similarly to [1, 2], we prove convergence to C-stationary points without assuming regularity of the cost function. An essential and novel feature of our algorithm is the computation of the quality indicator, based on a comparison between an "easy" local model and a complicated one containing neighborhood information (see 12 below).

Second, and maybe more important, we consider the application of the proposed algorithm to optimal control problems governed by variational inequalities of the second kind. To that end, the Bouligand subdifferential of the control-to-state map is precisely characterized, which allows to construct a model function which satisfies the hypotheses of our general trust-region method. Differently from other contributions were the choice of a good candidate is assumed, we provide, for this special family of problems, a way to find them explicitely.

The paper is organized as follows. In Section 2 the general algorithm is presented, together with the assumptions on the model function and a general convergence result. In Section 3, we show how to construct a suitable model function for the case of composite functions with a non-smooth inner function as they appear in the implicit programming approach for optimal control of VIs. Afterwards, in Section 4, we apply these findings to an optimal control problem governed by a VI of the second kind. The construction of the model function associated with this example is based on a precise characterization of the Bouligand-subdifferential of the control-to-state mapping. In Section 5, numerical experiments are carried out to verify the main properties of the proposed algorithm. The paper ends with some concluding remarks.

1.1. Notation. By λ^n, we denote the n-dimensional Lebesgue-measure. Moreover, $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ denote the Euclidean norm and the associated scalar product, whereas $\|\cdot\|_\infty$ and $\|\cdot\|_1$ stand for the maximum and 1-norm, respectively. In addition, $\|\cdot\|_{\mathbb{R}^{n\times n}}$ denotes the spectral norm, and we sometimes 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 Non-Smooth Trust-Region Algorithm. We consider the general nonlinear optimization problem

$$\min_{x\in\mathbb{R}^n} f(x), \tag{P}$$

with an objective function satisfying the following conditions:

ASSUMPTION 2.1 (Objective function). *The function $f : \mathbb{R}^n \to \mathbb{R}$ is supposed to be* locally Lipschitz continuous*, i.e., for all $x \in \mathbb{R}^n$ there exist $\delta > 0$ and $L > 0$ so that*

$$|f(y) - f(z)| \le L\,\|y - z\| \quad \forall\, y, z \in B_\delta(x).$$

Next, we introduce some basic concepts of nonsmooth optimization that are going to be used along the paper.

DEFINITION 2.2 (Subdifferentials). *Let $F : \mathbb{R}^n \to \mathbb{R}^m$, $m, n \in \mathbb{N}$, be locally Lipschitz-continuous. We denote the set of points, where F is differentiable, by $\mathcal{D}_F$. By Rademacher's theorem, this set is dense in $\mathbb{R}^n$. For a given $x \in \mathbb{R}^n$ we define*

- *the* Bouligand-subdifferential *by*

$$\partial_B F(x) := \{G \in \mathbb{R}^{m\times n} : \exists\,(x_n) \subset \mathcal{D}_F \text{ with } x_n \to x,\ F'(x_n) \to G\}, \quad (2.1)$$

- *the* Clarke-subdifferential *by*

$$\partial F(x) = \mathrm{cl}\,(\mathrm{conv}(\partial_B F(x)))\,, \quad (2.2)$$

where conv *denotes the convex hull.*

It is well known that in case of a scalar-valued locally Lipschitz-continuous function $f : \mathbb{R}^n \to \mathbb{R}$ the Clarke-subdifferential can equivalently be expressed as

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

where f° denotes Clarke's generalized directional derivative. For scalar-valued functions we moreover define the following notion of stationarity:

DEFINITION 2.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$, $n \in \mathbb{N}$, be locally Lipschitz-continuous. We then call a point $\bar{x} \in \mathbb{R}^n$ C(larke)-stationary, if $0 \in \partial f(\bar{x})$.*

Our algorithm is based on a suitably chosen model function, whose existence is assumed as a start. In the later sections, we will see how to construct such a model for concrete problems. To be more precise, we require the following.

ASSUMPTION 2.4 (Model function).

1. *For every $x \in \mathbb{R}^n$, we can calculate a subgradient $g \in \partial f(x)$.*
2. *There is a model function $\phi : \mathbb{R}^n \times \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}$ satisfying the following conditions:*
   - *(a) For every $(x, \Delta) \in \mathbb{R}^n \times \mathbb{R}^+$, the mapping $\mathbb{R}^n \ni d \mapsto \phi(x, \Delta; \cdot)$ is positively homogeneous and lower semicontinuous.*
   - *(b)* Stationarity indicator property:
     The stationarity measure defined through

$$\psi(x, \Delta) := -\min_{\|d\|\le 1} \phi(x, \Delta; d) \ge 0 \quad (2.3)$$

     satisfies the following:
     If a sequence $\{x_k, \Delta_k\} \subset \mathbb{R}^n \times \mathbb{R}^+$ satisfies

$$x_k \to x, \quad \Delta_k \to 0, \quad \text{and} \quad \psi(x_k, \Delta_k) \to 0,$$

     then it follows that $0 \in \partial f(x)$.
   - *(c)* Remainder term property:
     For every sequence $\{x_k, \Delta_k\} \subset \mathbb{R}^n \times \mathbb{R}^+$ satisfying

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

     there holds

$$\limsup_{k\to\infty} \sup_{d\in B_{\Delta_k}(0)} \frac{f(x_k + d) - f(x_k) - \phi(x_k, \Delta_k; d)}{\Delta_k} \le 0. \quad (2.4)$$

Note that the inequality in (2.3) follows immediately from the positive homogeneity and the lower semicontinuity of $\phi(x, \Delta, \cdot)$. Note moreover that the limes superior in (2.4) is actually a limes, since the inner supremum is always greater or equal zero (just choose $d = 0 \in B_{\Delta_k}(0)$).

REMARK 2.5. *Assumption 2.4(2) is closely related to the hypotheses required in other contributions in the field of non-smooth trust-region methods such as in [21, Assumption A2] or [2, Theorem 1]. In particular, condition (2c) is similar to the assumption that the objective f admits a* strict first-order model, *cf. [2, Definition 1 and Axiom $(\widetilde{M}_2)$]. To be more precise, it is easy to see that [2, $(\widetilde{M}_2)$] is sufficient for the remainder term property in (2c). This property reflects the fact that the model function has to incorporate certain information about the objective function in a neighborhood of the current iterate.*

Given the model function, our algorithm reads as follows:

ALGORITHM 2.6 (Non-Smooth Trust-Region Algorithm).

Initialization:
Choose constants

$$\Delta_{\min} > 0, \quad 0 < \eta_1 < \eta_2 < 1, \quad 0 < \beta_1 < 1 < \beta_2, \quad 0 < \mu \leq 1,$$

an initial value $x_0 \in \mathbb{R}^n$, and an initial TR-radius $\Delta_0 > \Delta_{\min}$. Set $k = 0$.

for $k = 0, 1, 2, \ldots$ ***do***
  Choose a subgradient $g_k \in \partial f(x_k)$ and a matrix $H_k \in \mathbb{R}^{n\times n}_{\text{sym}}$.
  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

$$\left.\begin{aligned} \min_{d\in\mathbb{R}^n} \quad & q_k(d) := f(x_k) + \langle g_k, d\rangle + \frac{1}{2}\, d^\top H_k d \\ s.t. \quad & \|d\| \leq \Delta_k, \end{aligned}\right\} \tag{Q_k}$$

      that satisfies the generalized Cauchy-decrease condition

$$f(x_k) - q_k(d_k) \geq \frac{\mu}{2}\, \|g_k\| \, \min\left\{\Delta_k, \frac{\|g_k\|}{\|H_k\|}\right\}. \tag{2.5}$$

      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 following modified trust-region subproblem

$$\left.\begin{aligned} \min_{d\in\mathbb{R}^n} \quad & \tilde{q}_k(d) := f(x_k) + \phi(x_k, \Delta_k; d) + \frac{1}{2}\, d^\top H_k d \\ s.t. \quad & \|d\| \leq \Delta_k \end{aligned}\right\} \tag{$\tilde{\mathrm{Q}}_k$}$$

that satisfies the modified Cauchy-decrease condition

$$f(x_k) - \tilde{q}_k(d_k) \geq \frac{\mu}{2}\,\psi(x_k,\Delta_k)\,\min\Big\{\Delta_k, \frac{\psi(x_k,\Delta_k)}{\|H_k\|}\Big\}, \tag{2.6}$$

where $\psi(x_k,\Delta_k)$ *is as defined in* (2.3).

12: *Compute the modified quality indicator*

$$\rho_k := \begin{cases} \dfrac{f(x_k) - f(x_k+d_k)}{f(x_k) - \tilde{q}_k(d_k)}, & \text{if } \psi(x_k,\Delta_k) > \|g_k\|\,\Delta_k \\ 0, & \text{if } \psi(x_k,\Delta_k) \leq \|g_k\|\,\Delta_k. \end{cases} \tag{2.7}$$

13: ***end if***

14: *Update: Set*

$$x_{k+1} := \begin{cases} x_k, & \text{if } \rho_k \leq \eta_1 \quad \textit{(null step)}, \\ x_k + d_k, & \text{otherwise} \quad \textit{(successful step)}, \end{cases} \tag{2.8}$$

$$\Delta_{k+1} := \begin{cases} \beta_1\,\Delta_k, & \text{if } \rho_k \leq \eta_1, \\ \max\{\Delta_{\min}, \Delta_k\}, & \text{if } \eta_1 < \rho_k \leq \eta_2, \\ \max\{\Delta_{\min}, \beta_2\Delta_k\}, & \text{if } \rho_k > \eta_2. \end{cases} \tag{2.9}$$

Set $k = k+1$.

15: ***end if***

16: ***end for***

REMARK 2.7 (Bouligand-subgradients). *Our convergence analysis in Section 2.1 below does not require to choose a particular subgradient in Step 3, it works for every element of the Clarke-subdifferential. Therefore, one could well restrict to elements of the Bouligand-subdifferential in this step. In this case, the termination criterion in Step 4 would imply that* $0 \in \partial_B f(x_k)$*, i.e., a stationarity condition which is in general only meaningful in smooth points.*
In the applications we are interested in, we exactly proceed in this way and choose elements of the Bouligand-subdifferential in Step 3, cf. Algorithm 4.12 below.

REMARK 2.8 (Comparison to other non-smooth trust region algorithms). *The essential differences to other non-smooth trust-region algorithms such as for instance the ones presented in [2, 11, 21, 26] 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 or Steinhaug's CG method. As our numerical experiments in Section 5 show, in the applications we have in mind, the second case involving the complicated model in* ($\tilde{\mathrm{Q}}_k$) *occurs only if* Δ_{min} *is chosen comparatively large. Our algorithm therefore accounts for the fact that many non-smooth problems can well be solved by classical trust-region methods and only switches to complicated model functions if it is strictly necessary.*
- *Another essential feature of the algorithm, which ensures the convergence of the method, is the computation of the quality indicator in step 12. It basically corresponds to a comparison of the "easy" and the complicated model weighted with the trust-region radius. Since the complicated model contains neighborhood information of the objective function, it may become insufficiently accurate to measure stationarity. This issue is resolved by the comparison with the local "easy" model in step 12.*

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 2.9. *Global minimizers of* (Q_k) *and* ($\tilde{\mathrm{Q}}_k$) *satisfy the respective Cauchy-decrease conditions in* (2.5) *and* (2.6) *for every* $\mu \leq 1$.

Proof. Since our model function ϕ is assumed to be positively homogeneous, we can argue as in [21, Lemma 3.2], which immediately gives the assertion. ∎

2.1. Convergence Analysis. In what follows, we show that accumulation points of the sequence of iterates are C-stationary as defined in Definition 2.3. For this purpose, we need the following

ASSUMPTION 2.10 (Hessian approximation). *The matrices* $H_k \in \mathbb{R}^{d\times d}_{\mathrm{sym}}$ *from Step 3 of Algorithm 2.6 are supposed to satisfy*

$$\|H_k\| \leq C_H \quad \forall k \in \mathbb{N}$$

with a constant $C_H > 0$.

PROPOSITION 2.11. *Assume that Algorithm 2.6 does not terminate in finitely many iterations. Let* (x_k) *be the sequence of iterates generated by Algorithm 2.6, and suppose that* (x_{k_l}) *is a subsequence of* (x_k) *satisfying*

$$x_{k_l} \to \bar{x} \quad \textit{and} \quad \Delta_{k_l} \to 0 \quad \textit{as } l \to \infty.$$

Then $0 \in \partial f(\bar{x})$ *holds true.*

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

$$x_{k_l-1} = x_{k_l}, \qquad \Delta_{k_l} = \beta_1 \Delta_{k_l-1} < \beta_1 \Delta_{min}, \qquad \rho_{k_l-1} < \eta_1 < 1. \tag{2.10}$$

This shows

$$x_{k_l-1} \to \bar{x} \quad \text{and} \quad \Delta_{k_l-1} \to 0.$$

We next show $\psi(x_{k_l-1}, \Delta_{k_l-1}) \to 0$. Once this is established, the assertion immediately follows from Assumption 2.4(2b). For this end, we argue by contradiction and assume that there is an $\varepsilon > 0$ so that

$$\limsup_{l\to\infty} \psi(x_{k_l-1}, \Delta_{k_l-1}) \geq \varepsilon. \tag{2.11}$$

Consider now the subsequence of $(x_{k_l-1}, \Delta_{k_l-1})$, which attains the limes superior, denoted for simplicity by $(x_m, \Delta_m)_{m\in M}$. 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 $m \in M \to \infty$ imply that $\|g_m\| \leq L(\bar{x})$, where $L(\bar{x})$ denotes the local Lipschitz constant, the convergence of Δ_m to 0 implies that $\psi(x_m, \Delta_m) > \|g_m\| \Delta_m$ for $m \in M$ sufficiently large. Therefore, the first case in (2.7) applies in the computation of the

quality indicator. Moreover, the modified Cauchy-decrease condition in (2.6) and (2.3) imply $f(x_m) - \tilde{q}_m(d_m) > 0$. Thus, for all $m \in \mathbb{N}$ sufficiently large, we obtain

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

where we used (2.6) and (2.11) for the last estimate. Note that the supremum in the enumerator is always greater or equal zero, since $d = 0$ is feasible. Assumption 2.4(2c) then implies

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

which however contradicts the last inequality in (2.10). Therefore, (2.11) is not true, which, together with the non-negativity of ψ results in

$$0 \leq \liminf_{l \to \infty} \psi(x_{k_l - 1}, \Delta_{k_l - 1}) \leq \limsup_{l \to \infty} \psi(x_{k_l - 1}, \Delta_{k_l - 1}) = 0. \tag{2.12}$$

This finally yields the desired convergence of $\psi(x_{k_l - 1}, \Delta_{k_l - 1})$, which establishes the assertion. ∎

LEMMA 2.12. *Assume that Algorithm 2.6 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. The arguments are classical. 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 2.13. *Assume that Algorithm 2.6 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 iterations is finite, then there is an $N \in \mathbb{N}$ so that all iterations $k \geq N$ are null steps. According to the update rule for null steps, it follows that $x_k \to x_N =: \bar{x}$ and $\Delta_k \to 0$ and thus, Proposition 2.11 yields that $\bar{x}$ is C-stationary.

We can thus focus on the case with infinitely many successful iterations. Let $\bar{x}$ be an arbitrary accumulation point of the sequence of iterates and denote the corresponding convergent subsequence by (x_{k_l}). W.l.o.g. we may suppose that the iterations k_l are all successful (else, we just 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 are successful, the monotonicity of $(f(x_k))$ and the Cauchy-decrease

conditions in (2.5) and (2.6) imply

$$\begin{aligned} f(x_{k_l}) - f(x_{k_{l+1}}) &\geq f(x_{k_l}) - f(x_{k_l+1}) \\ &\geq \eta_1 \frac{\mu}{2} \nu(x_{k_l}, \Delta_{k_l}) \min\left\{\Delta_{k_l}, \frac{\nu(x_{k_l}, \Delta_{k_l})}{C_H}\right\} \geq 0 \end{aligned}$$

with

$$\nu(x_{k_l}, \Delta_{k_l}) := \begin{cases} \|g_{k_l}\|, & \text{if } \Delta_{k_l} \geq \Delta_{\min} \\ \psi(x_{k_l}, \Delta_{k_l}), & \text{if } \Delta_{k_l} < \Delta_{\min}. \end{cases} \tag{2.13}$$

Since the sequence $(f(x_k))$ converges by Lemma 2.12, it follows

$$\lim_{l\to\infty} \left(\nu(x_{k_l}, \Delta_{k_l}) \min\left\{\Delta_{k_l}, \frac{\nu(x_{k_l}, \Delta_{k_l})}{C_H}\right\}\right) = 0,$$

i.e., it has to hold

$$\min\{\Delta_{k_l}, \nu(x_{k_l}, \Delta_{k_l})\} \to 0 \tag{2.14}$$

as $l \to \infty$. We now distinguish between three cases:
(i) If there exists a subsequence of (x_{k_l}) (unrelabeled for simplicity) such that the associated Δ_{k_l} converge to zero, then the claim follows immediately from Proposition 2.11.
(ii) If there exists a subsequence of (x_{k_l}) (again unrelabeled) such that $\Delta_{k_l} \geq \Delta_{\min}$, then (2.13) and (2.14) imply $\|g_{k_l}\| \to 0$. In view of [3, Prop. 2.1.5(b)], this implies $0 \in \partial f(\bar{x})$ as claimed.
(iii) If there exists a subsequence of (x_{k_l}) (again unrelabeled) with $\varepsilon \leq \Delta_{k_l} < \Delta_{\min}$ for some $\varepsilon > 0$, then (2.14) 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 (2.7), 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 obtain the claim. ∎

REMARK 2.14. *The proofs of Proposition 2.11 and Theorem 2.13, in particular (2.12) and the distinction of cases after (2.14), do not only show 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, which is important for practical reasons, as it 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 $\texttt{TOL} > 0$.

2.2. A Pathological One-Dimensional Example. A crucial question in the context of Algorithm 2.6 of course concerns the choice of the model function. For a general non-smooth problem of the form (P), a naive choice would be

$$\widetilde{\phi}(x, \Delta; d) := \max_{g \in \partial f(x)} \langle g, d \rangle, \tag{2.15}$$

which is the model proposed in [21, Section 4.1]. However, it turns out that this model is not well suited for the minimization of non-smooth functions, as we will see by means of the one-dimensional counterexample below. The essential drawback of the model in (2.15) is that it does not account for neighboorhood information. Thus, it is rather natural to consider the following model function:

$$\phi(x,\Delta;d) := \max_{g\in\mathcal{U}(x,\Delta)} \langle g,d\rangle \quad \text{with} \quad \mathcal{U}(x,\Delta) := \bigcup_{\xi\in B_\Delta(x)} \partial f(\xi). \tag{2.16}$$

A similar model based on the ε-subdifferential as defined in [14] is used in [1] in the context of a non-smooth trust-region method. If f is Bouligand-differentiable and semi-smooth, then one can verify the conditions in Assumption 2.4(2) for the model function in (2.16) so that the above convergence analysis applies. The proof thereof is analogous to the ones of Lemma 3.6 and 3.7 below and therefore omitted. Of course, the model function ϕ is much more costly compared to $\widetilde{\phi}$, but the following counterexample shows that the simple model in (2.15) might not suffice. For this purpose, let us define

$$f:\mathbb{R}\to\mathbb{R}, \quad f: x\mapsto \max\{-ax,-bx,x-(1+b)\}, \tag{2.17}$$

where $0<b<a<\infty$ are given constants. This piecewise affine function is trivially convex and admits two kinks at $x=0$ and $x=1$. If one applies the trust-region method with the two different models to this function, the following lemma is obtained. Its proof is not difficult, but rather technical and therefore we refer to the preprint version of this article [6].

LEMMA 2.15. *Assume that the parameters and initial values in step 1 of the algorithm satisfy*

$$\beta_1+\beta_1\beta_2<1, \qquad \eta_1 \geq \Big(\frac{b}{a}-1\Big)\frac{\beta_1}{\beta_1\beta_2-1}+\frac{b}{a} \tag{2.18}$$

$$x_0 \in \Big(\Big(1+\frac{\beta_1\beta_2-1}{\beta_1}\Big)^{-1},0\Big), \quad \Delta_{\min}>\Delta_0 := \frac{\beta_1\beta_2-1}{\beta_1}\,x_0. \tag{2.19}$$

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

By contrast, if one uses the model function ϕ from (2.16) instead, then the iterates converges to the global minimum at $x=1$, no matter how the parameters and initial values are chosen.

REMARK 2.16. *We emphasize that the failure of the trust-region method in case of the model in (2.15) is not caused by the distinction of cases contained in Algorithm 2.6. 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 2.6 turns into a standard (non-smooth) trust-region iteration.*

In our opinion, it is remarkable that this one-dimensional counterexample shows that a method based on a local model, which does not account for any neighborhood information, fails to converge even in case of a convex and piecewise affine objective. Of course, this observation is not new and we exemplarily refer to [2, Section 5.7], where a similar two-dimensional example is discussed. However, if the initial radius Δ_0 is chosen slightly different from the setting in (2.19), then the trust-region algorithm

with the model in (2.15) will converge to the global minimum at $x = 1$. This indicates that, in frequent cases, it is not necessary to use an involved model of the form (2.16), while simpler local models will suffice. Our algorithmic approach accounts for this observation by incorporating the distinction of cases $\Delta_k \gtrless \Delta_{\min}$ into the algorithm.

3. Composite Functions. Our aim is to apply Algorithm 2.6 to (discretized) optimal control problems with non-smooth constraints. In order to conform with the standard notation in optimal control, we denote the optimization variable by u from now on. Although this causes a slight abuse of notation, we tacitly replace x by u, when referring to the results of Section 2. Our general optimal control problem reads as follows:

$$\min_{u\in\mathbb{R}^n} \; f(u) := J(S(u), u), \tag{3.1}$$

where $J : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$, $m, n \in \mathbb{N}$ is continuously differentiable and $S : \mathbb{R}^n \to \mathbb{R}^m$ is assumed to be directionally differentiable and locally Lipschitz continuous. Note that S is Bouligand-differentiable by [24, Thm. 3.1.2]. In all what follows, we will frequently abbreviate $y := S(u) \in \mathbb{R}^m$. Given $u \in \mathbb{R}^n$ and $\Delta > 0$, we suppose that we can construct an approximation of the Bouligand-subdifferential of S satisfying the following

ASSUMPTION 3.1. *Given $u \in \mathbb{R}^n$ and $\Delta > 0$, the approximation $\mathcal{G}(u, \Delta) \subset \mathbb{R}^{m\times n}$ of the Bouligand-subdifferential is supposed to fulfill the following conditions:*
For all $u \in \mathbb{R}^n$ and all $\Delta > 0$, there holds

$$\bigcup_{\xi\in B_\Delta(u)} \partial_B S(\xi) \subseteq \mathcal{G}(u, \Delta) \tag{3.2}$$

and, if $(u_k, \Delta_k) \to (u, 0)$ with $0 \notin \partial f(u)$, then

$$\operatorname{dist}(\mathcal{G}(u_k, \Delta_k), \partial_B S(u))) = \sup_{G\in\mathcal{G}(u_k,\Delta_k)} \inf_{W\in\partial_B S(u)} \|G - W\|_{\mathbb{R}^{m\times n}} \to 0 \tag{3.3}$$

is valid.

With this approximation at hand, we construct our model function as follows:

$$\phi(u, \Delta; d) := \sup_{G\in\mathcal{G}(u,\Delta)} \langle G^\top \nabla_y J(y, u) + \nabla_u J(y, u), d\rangle. \tag{3.4}$$

This model function allows the following reformulation of the modified trust-region subproblem, which will be useful for the realization of the algorithm in case of the concrete optimization problem in Section 4. Its proof is straightforward and therefore omitted.

LEMMA 3.2. *With the model function in (3.4), the modified trust-region subproblem ($\tilde{\mathrm{Q}}_k$) from step 11 is equivalent to the following linear quadratic problem in the sense that they admit the same (global) optima:*

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

In addition, if $\bar{d}_k$ is a global minimizer of $(\tilde{\mathfrak{Q}}_k)$, then $(\bar{d}_k, \bar{\zeta}_k)$ with $\bar{\zeta}_k = \phi(u_k, \Delta_k, ; \bar{d}_k)$ solves $(\mathfrak{Q}_k)$ so that $\tilde{q}_k(\bar{d}_k) = \mathfrak{q}_k(\bar{d}_k, \bar{\zeta}_k)$.

Moreover, if (d_k, ζ_k) is feasible for $(\mathfrak{Q}_k)$ and satisfies

$$f(x_k) - \mathfrak{q}_k(d_k, \zeta_k) \geq \frac{\mu}{2}\, \psi(x_k, \Delta_k)\, \min\Big\{\Delta_k, \frac{\psi(x_k, \Delta_k)}{\|H_k\|}\Big\}, \tag{3.5}$$

then d_k fulfills the modified Cauchy-decrease condition in (2.6), too.

REMARK 3.3. *Note that the optimal solution $(\bar{d}_k, \bar{\zeta}_k)$ of $(\mathfrak{Q}_k)$ satisfies the modified Cauchy-decrease condition (3.5), since $\bar{d}_k$ does so by Lemma 2.9 and $\tilde{q}_k(\bar{d}_k) = \mathfrak{q}_k(\bar{d}_k, \bar{\zeta}_k)$.*

Next, we show that the model function in (3.4) satisfies the conditions in Assumption 2.4. To this end, we require the following

ASSUMPTION 3.4. *For every $u \in \mathbb{R}^n$ and every $h \in \mathbb{R}^n$, there exists a $G \in \partial_B S(u)$ so that $S'(u; h) = G\, h$.*

There is a large class of functions satisfying this assumption such as for instance semi-smooth functions, as the next lemma shows.

LEMMA 3.5. *If $S : \mathbb{R}^n \to \mathbb{R}^m$ is Bouligand differentiable and semi-smooth, then Assumption 3.4 is fulfilled.*

Proof. Let u and h be given and (t_n) be an arbitrary null sequence. For every $n \in \mathbb{N}$, Rademacher's theorem implies the existence of h_n such that

$$u + t_n h_n \in \mathcal{D}_S \quad \text{and} \quad \|h_n - h\| = \mathcal{O}(t_n). \tag{3.6}$$

The semi-smoothness of S then implies

$$\frac{S(u + t_n h_n) - S(u)}{t_n} - S'(u + t_n h_n) h_n \to 0. \tag{3.7}$$

The local Lipschitz continuity of S moreover yields the boundedness of $\{S'(u+t_n h_n)\}$ so that there exists $G \in \mathbb{R}^{m \times n}$ with $S'(u + t_n h_n) \to G$. As S is Bouligand differentiable, (3.6) in turn implies the convergence of the first addend in (3.7) to $S'(u; h)$. This establishes the claim. ∎

LEMMA 3.6. *Let $(u_k, \Delta_k) \subset \mathbb{R}^n \times \mathbb{R}^+$ be a sequence such that $u_k \to \tilde{u}$ and $\Delta_k \to 0$. Then, the linearization error satisfies*

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

such that the model given by (3.4) fulfills Assumption 2.4(2c).

Proof. Let $u \in \mathbb{R}^n$, $\Delta > 0$, and $d \in B_\Delta(0)$ be arbitrary. By [24, Prop. 3.1.1] and the chain rule for Bouligand-differentiable functions, we have

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

By Assumption 3.4, for every $\theta \in [0,1]$, there exists a $G_\theta \in \partial_B S(u+\theta d)$ such that $G_\theta\, h = S'(u+\theta d; d)$. This, together with (3.2), the definition of our model ϕ, and $\|d\| \leq \Delta$, yields

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

Now, let (u_k, Δ_k) be the sequence from the statement of the lemma and denote by $\tilde{L} > 0$ and $\widetilde{\mathcal{U}} \subset \mathbb{R}^n$ the local Lipschitz constant of S at $\tilde{u}$ and the associated 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

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

Furthermore, 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

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

Collecting all findings, we arrive at

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

which implies the assertion. ∎

LEMMA 3.7. *Let ψ denote the stationarity measure from (2.3), i.e.,*

$$\psi(u, \Delta) := -\min_{\|h\| \leq 1} \phi(u, \Delta; h),$$

and $(u_k, \Delta_k) \subset \mathbb{R}^n \times \mathbb{R}^+$ so that

$$u_k \to u, \quad \Delta_k \to 0, \quad \text{and} \quad \psi(u_k, \Delta_k) \to 0. \tag{3.8}$$

Then $0 \in \partial f(u)$ holds true such that the model from (3.4) satisfies Assumption 2.4(2b). Herein f again denotes the reduced objective, i.e., $f(u) = J(S(u), u)$.

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

$$\mathrm{dist}(0, \partial f(u)) \geq \varepsilon. \tag{3.9}$$

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 $\mathcal{D}_S \subseteq \mathcal{D}_f$

and, by Rademacher's theorem, we have $\lambda^n(\mathcal{D}_f \setminus \mathcal{D}_S) = 0$. Therefore, [3, Thm. 2.5.1] and the continuous differentiability of J imply

$$\begin{aligned}
&\{G^\top \nabla_y J(y,u) + \nabla_u J(y,u) : G \in \partial_B S(u)\} \\
&\subset \operatorname{cl}\left(\operatorname{conv}\left(g \in \mathbb{R}^n : \exists\,(u_n) \subset \mathcal{D}_S : u_n \to u,\ S'(u_n)^\top \nabla_y J(y_n,u_n) + \nabla_u J(y_n,u_n) \to g\right)\right) \\
&= \partial f(u)
\end{aligned}$$

Therefore, due to Assumption (3.3), there exist $K \in \mathbb{N}$ such that, for all $k \geq K$, it holds

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

Since $\partial f(u)$ is convex, this in combination with (3.9) implies $\operatorname{dist}(\mathcal{C}_k, 0) \geq \varepsilon/2$, where we abbreviated

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

Next, let us define $\bar{g}$ as 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

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

This however contradicts the last assumptions in (3.8), which gives the claim. ∎

We collect the findings of this section in the following

COROLLARY 3.8. *Under Assumptions 3.1 and 3.4, every accumulation point of the sequence of iterates generated by our non-smooth trust-region algorithm applied to (3.1) with the model function in (3.4) is a C-stationary point of (3.1).*

Proof. As shown in the above lemmata, Assumption 2.4 is fulfilled, provided that Assumptions 3.1 and 3.4 hold true. Therefore, Theorem 2.13 gives the assertion. ∎

4. Optimization of Variational Inequalities of the Second Kind. We now focus on the following class of nonsmooth optimization problems:

$$\left.\begin{aligned}
\min_{y,u \in \mathbb{R}^n} \quad & J(y,u) \\
\text{s.t.} \quad & \langle Ay, v - y\rangle + \|v\|_1 - \|y\|_1 \geq \langle u, v - y\rangle \quad \forall\, v \in \mathbb{R}^n,
\end{aligned}\right\} \tag{$\mathrm{P_{VI}}$}$$

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

In the next proposition we summarize some known results about ($\mathrm{P_{VI}}$). For more details on this, we refer to [10].

PROPOSITION 4.1. *Let $u \in \mathbb{R}^n$ be given. Then there holds:*

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

$$\langle Ay, v - y\rangle + \|v\|_1 - \|y\|_1 \geq \langle u, v - y\rangle \quad \forall\, v \in \mathbb{R}^n. \tag{VI}$$

- *y is the solution of* (VI)*, iff there exists a $q \in \mathbb{R}^n$ such that*

$$Ay + q = u, \quad y_i\, q_i = |y_i|, \quad |q_i| \leq 1, \quad i = 1, ..., n \tag{4.1}$$

- *The solution mapping $S : \mathbb{R}^n \ni u \mapsto y \in \mathbb{R}^n$ 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*

$$\eta \in \mathcal{K}(y), \quad \langle A\eta, v - \eta\rangle \geq \langle h, v - \eta\rangle \quad \forall\, v \in \mathcal{K}(y), \tag{4.2}$$

where

$$\mathcal{K}(y) := \{v \in \mathbb{R}^n : v_i = 0,\ \textit{if } |q_i| < 1,\ v_i\, q_i \geq 0,\ \textit{if } y_i = 0 \wedge |q_i| = 1\}. \tag{4.3}$$

Thanks to these properties, we may formulate problem (P_{VI}) in reduced form as

$$\min_{u \in \mathbb{R}^n} \ f(u) := J(S(u), u), \tag{4.4}$$

so that a problem of the form (3.1) is obtained. Our aim in the following is to verify the hypotheses on the general problem (3.1), i.e., Assumptions 3.1 and 3.4. For this purpose, we first have to charactrize the Bouligand-subdifferential of S, which is addressed in the next subsection.

4.1. Characterization of the Bouligand-Subdifferential. Given $u \in \mathbb{R}^n$ with $y = S(u)$, we define the following sets

$$\mathcal{A} := \{i \in \{1, ..., n\} : y_i = 0\} \qquad \text{(active set)} \tag{4.5a}$$
$$\mathcal{A}_s := \{i \in \{1, ..., n\} : |q_i| < 1\} \qquad \text{(stongly active set)} \tag{4.5b}$$
$$\mathcal{I} := \{i \in \{1, ..., n\} : y_i \neq 0\} \qquad \text{(inactive set)} \tag{4.5c}$$
$$\mathcal{B} := \{i \in \{1, ..., n\} : y_i = 0 \wedge |q_i| = 1\} \qquad \text{(biactive set).} \tag{4.5d}$$

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., but, for the sake of readability, we suppress this dependency throughout this subsection. This will be different in Section 4.2, where we have to distinguish between the active sets in different points. Note that, because of the complementarity like relation in (4.1), one has $\mathcal{A}_s \subset \mathcal{A}$, and therefore $\mathcal{A} = \mathcal{A}_s \cup \mathcal{B}$.

LEMMA 4.2. *S is differentiable at u iff $\mathcal{K}(y) = \{v \in \mathbb{R}^n : v_i = 0,\ \textit{if } y_i = 0\}$.*

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 convex projection on a linear subspace, $S'(u, .)$ is a linear mapping so that S is differentiable at u.

To show the converse assertion, we first show that

$$S'(u; \mathbb{R}^n) = \mathcal{K}(y). \tag{4.6}$$

By (4.2), 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 and set $h = A\, z$. Then we trivially obtain $\langle Az, v - z\rangle = \langle h, v - z\rangle$ for all

$v \in \mathcal{K}(y)$ so that $z = S'(u;h)$, which shows (4.6). 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 (4.6), so does $\mathcal{K}(y)$. Therefore $v \in \mathcal{K}(y)$ implies $-v \in \mathcal{K}(y)$, which, due to (4.3) yields

$$0 \le v_i\, q_i \le 0 \quad \forall\, i \in \mathcal{B} = \{j \in \{1,...,n\} : y_j = 0 \wedge |q_j| = 1\}.$$

Since $q_i \neq 0$ in $\mathcal{B}$, this yields $v_i = 0$ in $\mathcal{B}$, which, together with $v_i = 0$ in $\mathcal{A}_s$, see (4.3), finally gives $v_i = 0$ in $\mathcal{B} \cup \mathcal{A}_s = \mathcal{A}$ as claimed. ∎

Now, we are in the position to give a precise characterization of the Bouligand-subdifferential of S. To this end, we introduce the following

DEFINITION 4.3. *Let $\mathcal{N} \subset \{1,...,n\}$ be an index set. Then we define the matrices $A(\mathcal{N}) \in \mathbb{R}^{n\times n}$ and $\chi(\mathcal{N}) \in \mathbb{R}^{n\times n}$ by*

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

THEOREM 4.4. *Let $u \in \mathbb{R}^n$ be fixed, but arbitrary, and let $y = S(u)$. Then there holds*

$$\partial_B S(u) = \{A(\mathcal{A}_s \cup \mathcal{B}_0)^{-1}\chi(\mathcal{A}_s \cup \mathcal{B}_0) : \mathcal{B}_0 \subseteq \mathcal{B}\}. \tag{4.7}$$

REMARK 4.5. *Note that $A(\mathcal{N})$ is indeed invertible for every index set $\mathcal{N} \subset \{1,...,n\}$, since it is positive definite: For an arbitrary $v \in \mathbb{R}^n$, we obtain*

$$\begin{aligned} v^\top A(\mathcal{N})v &= \sum_{i,j\notin\mathcal{N}} A_{ij}v_iv_j + \sum_{i\in\mathcal{N}} v_i^2 \\ &= [(I-\chi(\mathcal{N}))v]^\top A\,(I-\chi(\mathcal{N}))v + \sum_{i\in\mathcal{N}} v_i^2 \ge \min\{\lambda_{\min},1\}\|v\|^2, \end{aligned}$$

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

REMARK 4.6. *We could equivalently replace the last line in the definition of $\mathcal{A}(\mathcal{N})$ by*

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

with some $c \neq 0$, since no matter, which value is chosen for $c \neq 0$, the matrix $A(\mathcal{N})^{-1}\chi(\mathcal{N})$ is always the same, as

$$A(\mathcal{N})\tilde\eta = \chi(\mathcal{N})h \quad \Longleftrightarrow \quad \begin{cases} \tilde\eta_i = 0 & \forall\, i \in \mathcal{N}, \\ \displaystyle\sum_{j\notin\mathcal{N}} A_{ij}\tilde\eta_j = h_i & \forall\, i \in \{1,...,n\} \setminus \mathcal{N}, \end{cases} \tag{4.8}$$

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

Proof of Theorem 4.4. Recall that

$$\partial_B S(u) = \big\{B \in \mathbb{R}^{n\times n} : \exists\, \{u_n\} \subset \mathcal{D}_S \text{ with } u_n \to u,\ S'(u_n) \to B\big\}, \tag{4.9}$$

where $\mathcal{D}_S$ again denotes the (dense) set of points, where S is differentiable. Now consider an arbitrary $B \in \partial_B S(u)$ so that there is a sequence in $\mathcal{D}_S$ with

$$u_n \to u \quad \text{and} \quad S'(u_n) \to B. \tag{4.10}$$

The Lipschitz continuity of S implies

$$y_n := S(u_n) \to S(u) =: y \quad \Longrightarrow \quad q_n = u_n - Ay_n \to u - Ay = q. \tag{4.11}$$

Let us denote the active set associated with y_n by $\mathcal{A}^n$ and analogously for $\mathcal{I}^n$ etc. Then, from (4.11), we deduce the existence of an $N \in \mathbb{N}$ such that

$$\mathcal{I} \subset \mathcal{I}^n \quad \text{and} \quad \mathcal{A}_s \subset \mathcal{A}_s^n \qquad \forall\, n \geq N. \tag{4.12}$$

Next let $h \in \mathbb{R}^n$ be fixed, but arbitrary. Since $u_n \in \mathcal{D}_S$, we know from Lemma 4.2 that $\eta_n := S'(u_n)h$ solves

$$\eta_i^n = 0 \quad \forall\, i \in \mathcal{A}^n, \quad \sum_{j=1}^n A_{ij}\eta_j^n = h_i \quad \forall\, i \in \mathcal{I}^n. \tag{4.13}$$

By (4.10) we obtain that

$$\tilde{\eta} := B\,h = \lim_{n\to\infty} \eta_n. \tag{4.14}$$

Therefore, from (4.12)–(4.14) it follows that

$$\tilde{\eta}_i = 0 \quad \forall\, i \in \mathcal{A}_s, \quad \sum_{j=1}^n A_{ij}\tilde{\eta}_j = h_i \quad \forall\, i \in \mathcal{I}. \tag{4.15}$$

It remains to investigate what happens on the biactive set $\mathcal{B} = \mathcal{A} \setminus \mathcal{A}_s$. For this purpose we introduce

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

so that, for all $i \in \mathcal{B}\setminus\mathcal{B}_0$, it holds that $y_i^n \neq 0$ for all $n \in \mathbb{N}$ sufficiently large. Then we deduce from (4.13) that $\eta_i^{n_k} = 0$ for all $i \in \mathcal{B}_0$ and all $k \in \mathbb{N}$ and that $\sum_{j=1}^n A_{ij}\eta_j^n = h_i$ for all $i \in \mathcal{B}\setminus\mathcal{B}_0$, provided that $n \in \mathbb{N}$ is sufficiently large. Since $\eta_n \to \tilde{\eta}$, we obtain in this way

$$\tilde{\eta}_i = 0 \quad \forall\, i \in \mathcal{A}_s \cup \mathcal{B}_0, \qquad \sum_{j \notin \mathcal{A}_s \cup \mathcal{B}_0} A_{ij}\tilde{\eta}_j = h_i \quad \forall\, i \in \{1,\dots,n\} \setminus (\mathcal{A}_s \cup \mathcal{B}_0). \tag{4.16}$$

Thus, in view of (4.8) and since h was arbitrary, we observe that

$$B = A(\mathcal{A}_S \cup \mathcal{B}_0)^{-1}\chi(\mathcal{A}_s \cup \mathcal{B}_0). \tag{4.17}$$

Hence, B has indeed the form stated in the theorem.

To complete the proof, we need to show that, for every subset $\mathcal{B}_0 \subset \mathcal{B}$ the corresponding matrix B given by (4.7) is an element of $\partial_B S(u)$. To this end, let $\mathcal{B}_0 \subset \mathcal{B}$ be arbitrary, but fixed and let us abbreviate $\mathcal{B}_1 := \mathcal{B}\setminus\mathcal{B}_0$. In the following, we show that there exist a sequence $\{u_n\}$ satisfying

$$\begin{aligned} u_n \in D_S, \quad y_i^n = 0 \quad \forall\, i \in \mathcal{B}_0, \quad y_i^n \neq 0 \quad \forall\, i \in \mathcal{B}_1, \quad \forall\, n \in \mathbb{N}, \\ \text{and} \quad u_n \to u, \quad S'(u_n) \to B \quad \text{as } n \to \infty, \end{aligned} \tag{4.18}$$

which, according to (4.9) implies $B \in \partial_S S(u)$. To verify the existence of such a sequence, let $\varepsilon > 0$ and define

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

where e_i denotes the i-the Euclidian unit vector. By construction we obtain for the inactive and active set associated with y^ε that

$$\mathcal{I}^\varepsilon = \mathcal{I} \cup \mathcal{B}_1 \quad \text{and} \quad \mathcal{A}^\varepsilon = \mathcal{A} \setminus \mathcal{B}_1. \tag{4.19}$$

Moreover, we set

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

Thus, for $\varepsilon \in (0, 1]$, we obtain $|q_i^\varepsilon| \leq 1$ for all $i = 1, ..., n$. Moreover, the above construction leads to

$$\mathcal{A}_s^\varepsilon = \mathcal{A}_s \cup \mathcal{B}_0 = \mathcal{A} \setminus \mathcal{B}_1, \tag{4.20}$$

which, together with (4.19), shows that

$$\mathcal{B}^\varepsilon = \mathcal{A}^\varepsilon \setminus \mathcal{A}_s^\varepsilon = \emptyset, \tag{4.21}$$

i.e., the biactive set associated with y_ε is empty. Furthermore, if we define

$$u^\varepsilon := u + \varepsilon \sum_{k \in \mathcal{B}_0} \operatorname{sgn}(q_k) \, A \, \mathrm{e}_k + \varepsilon \sum_{k \in \mathcal{B}_1} \operatorname{sgn}(q_k) \, \mathrm{e}_k, \tag{4.22}$$

then we obtain by construction that

$$A \, y^\varepsilon + q^\varepsilon = u^\varepsilon, \quad y_i^\varepsilon \, q_i^\varepsilon = |y_i^\varepsilon|, \quad |q_i^\varepsilon| \leq 1, \quad i = 1, ..., n,$$

which, due to (4.1), implies $y^\varepsilon = S(u^\varepsilon)$. Because of (4.21), we further have $\mathcal{K}(y^\varepsilon) = \{v \in \mathbb{R}^n : v_i = 0, \text{ if } y_i^\varepsilon = 0\}$, which, thanks to Lemma 4.2, in turn implies $u^\varepsilon \in D_s$. In addition, (4.22) immediately gives $u_\varepsilon \to u$ as $\varepsilon \searrow 0$. Because of (4.19), we moreover have $y_i^\varepsilon \neq 0$ for all $i \in \mathcal{B}_1$ and, due to complementarity and (4.20), $y_i^\varepsilon = 0$ for all $i \in \mathcal{B}_0$. Therefore, the sequence $\{u^\varepsilon\}_{\varepsilon > 0}$ almost satisfies all conditions required in (4.18), except $S'(u^\varepsilon) \to B$. To establish this, let $\{\varepsilon_n\}_{n \in \mathbb{N}}$ be a sequence tending to zero and denote the associated u^{ε_n} simply by u_n. The global Lipschitz continuity of S yields that

$$\|S'(u_n)\|_{\mathbb{R}^{n \times n}} \leq L \quad \forall \, n \in \mathbb{N},$$

where $L > 0$ is the Lipschitz constant of S. Thus there exists a convergent subsequence, i.e.,

$$S'(u_{n_k}) \to \tilde{B} \quad \text{as } k \to \infty.$$

Since $y_i^{n_k} = 0$ for all $i \in \mathcal{B}_0$ and $y_i^{n_k} \neq 0$ for all $i \in \mathcal{B}_1$, we can argue completely analogously to the first part of the proof to show

$$\tilde{B} = A(\mathcal{A}_S \cup \mathcal{B}_0)^{-1} \chi(\mathcal{A}_s \cup \mathcal{B}_0) = B,$$

which finally establishes the claim. ∎

LEMMA 4.7. *For all $u, h \in \mathbb{R}^n$, there exists $G \in \partial_B S(u)$ such that $S'(u; h) = G \, h$. Hence Assumption 3.4 is fulfilled by the control-to-state map of* ($\mathrm{P_{VI}}$).

Proof. Let $u, h \in \mathbb{R}^n$ be arbitrary and again set $y = S(u)$. As above we denote by $\mathcal{A}_s$, $\mathcal{I}$, and $\mathcal{B}$ the active, inactive, and bi-active sets associated with y. Recall that the

directional derivative of the solution mapping in direction h is given by the unique solution of

$$\eta \in \mathcal{K}(y), \quad \langle A\eta, v-\eta\rangle \geq \langle h, v-\eta\rangle \quad \forall\, v \in \mathcal{K}(y), \tag{4.23}$$

with $\mathcal{K}(y)$ as defined in (4.3). This cone can equivalently be expressed as

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

which in turn leads to the following equivalent expression for η

$$\eta_i = \begin{cases} \max\{0, (I-A)\eta + h\}_i, & \text{if } y_i = 0,\ q_i = 1 \\ 0, & \text{if } |q_i| < 1 \\ \min\{0, (I-A)\eta + h\}_i, & \text{if } y_i = 0,\ q_i = -1 \\ ((I-A)\eta + h)_i, & \text{elsewhere.} \end{cases} \tag{4.24}$$

So, if we define

$$\mathcal{B}_0 := \{i \in \{1, ..., n\} : y_i = 0,\ |q_i| = 1,\ q_i((I-A)\eta + h)_i < 0\} \subseteq \mathcal{B},$$

then a comparison of (4.24) with (4.8) shows that

$$\eta = A(\mathcal{A}_s \cup \mathcal{B}_0)^{-1}\chi(\mathcal{A}_s \cup \mathcal{B}_0)h.$$

Since the matrix on the right hand side is an element of $\partial_B S(u)$, this establishes the assertion. ∎

4.2. Approximation of the Bouligand-Subdifferential. The aim of the upcoming section is to construct a *computable* approximation of $\partial_B S$, that satisfies the conditions in Assumption 3.1 so that the model function given by (3.4) fulfills Assumption 2.4 for the convergence result in Theorem 2.13. For this purpose, we need a sharpened Lipschitz continuity result for the solution operator S associated with (VI):

LEMMA 4.8. *For all $u_1, u_2 \in \mathbb{R}^n$, there holds*

$$\|y_1 - y_2\|_\infty \leq L_y \|u_1 - u_2\| \quad \text{with } L_y = \frac{1}{\lambda_{\min}},$$
$$\|q_1 - q_2\|_\infty \leq L_q \|u_1 - u_2\| \quad \text{with } L_q = \frac{\lambda_{\max}}{\lambda_{\min}} + 1,$$

where $y_i, q_i \in \mathbb{R}^n$, $i = 1, 2$, are the solutions of (4.1) associated with u_i, and $\lambda_{\min}$ and $\lambda_{\max}$ denote the minimal and maximal eigenvalue of A, respectively.

Proof. By testing the VI for y_1 with y_2 and vice versa and adding the arising inequalities, we obtain

$$\lambda_{\min}\|y_1 - y_2\| \leq \|u_1 - u_2\|, \tag{4.25}$$

which immediately gives the first assertion. The second directly follow from the first equation in (4.1), which yields

$$\|q_1 - q_2\| \leq \|A\|_{\mathbb{R}^{n\times n}}\|y_1 - y_2\| + \|u_1 - u_2\|.$$

Inserting (4.25) then implies the second estimate in the statement of the lemma. ∎

As the active, inactive, and biactive sets at multiple points will occur in what follows, we denote these sets for a given $u \in \mathbb{R}^n$ by $\mathcal{A}_s(u)$, $\mathcal{I}(u)$, and $\mathcal{B}(u)$. (Note that these sets are determined by y and q, which in turn uniquely depend on u.)

DEFINITION 4.9. *Let $u \in \mathbb{R}^n$ and $\Delta > 0$ be given and set $y = S(u)$. Then we define the set of* possibly biactive indices *by*

$$\mathcal{P}(u, \Delta) := \{i \in \{1, ..., n\} : |y_i| < L_y \, \Delta \ \wedge \ \big||q_i| - 1\big| < L_q \, \Delta\}.$$

In view of Lemma 4.8, it is clear that

$$\bigcup_{\xi \in B_\Delta(u)} \mathcal{B}(\xi) \subset \mathcal{P}(u, \Delta), \tag{4.26}$$

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

$$\mathcal{G}(u, \Delta) := \{A(\mathcal{A}_s(u) \cup \mathcal{B}_0)^{-1} \chi(\mathcal{A}_s(u) \cup \mathcal{B}_0) : \mathcal{B}_0 \subseteq \mathcal{P}(u, \Delta)\}. \tag{4.27}$$

As an immediate consequence of (4.26) and Theorem 4.4, we obtain

LEMMA 4.10. *The approximation of the Bouligand-subdifferential in (4.27) satisfies condition (3.2).*

On the other hand, we find the following:

LEMMA 4.11. *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 an index $K \in \mathbb{N}$ (depending on u) so that $\mathcal{P}(u_k, \Delta_k) \subseteq \mathcal{B}(u)$ for all $k \geq K$. Therefore, for all $k \geq K$, there holds $\mathcal{G}(u_k, \Delta_k) \subseteq \partial_B S(u)$ such that condition (3.3) is fulfilled, too.*

Proof. Again, we denote the state associated with the limit u by $y = S(u)$. Moreover, we define

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

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

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

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

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

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

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

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

$$\big||q_i^k| - 1\big| > L_q \, \Delta_k \quad \Longrightarrow \quad i \notin \mathcal{P}(u_k, \Delta_k) \quad \forall \, k \geq K := \max\{K_1, K_2\},$$

and therefore, since $\mathcal{B}(u) = \{1, ..., n\} \setminus (\mathcal{A}_s(u) \cup \mathcal{I}(u))$, it follows $\mathcal{P}(u_k, \Delta_k) \subset \mathcal{B}(u)$ as claimed. The second assertion of the lemma then immediately follows from Theorem 4.4 and the construction of our approximation in (4.27). ∎

For convenience of the reader, we next state the precise algorithm that arises, when applying Algorithm 2.6 to (P_{VI}). We again use the reduced objective function $f(\cdot) = J(S(\cdot), \cdot)$.

ALGORITHM 4.12 (Trust-Region Algorithm for the solution of (P_{VI})).

Initialization:
Choose constants

$$\Delta_{\min} > 0, \quad 0 < \eta_1 < \eta_2 < 1, \quad 0 < \beta_1 < 1 < \beta_2, \quad 0 < \mu \leq 1$$

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, ...* ***do***

Solve the variational inequality (VI) *to compute the state* y_k *associated with the control* u_k.

Choose a subset $\mathcal{B}_k \subseteq \mathcal{B}(u_k)$, *solve the* adjoint equation

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

and set $g_k = p_k + \nabla_u J(y_k, u_k)$.

Choose a matrix $H_k \in \mathbb{R}^{n\times n}_{\text{sym}}$, *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

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

that satisfies the generalized Cauchy-decrease condition

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

Compute the quality indicator

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

else

Identify the possibly biactive indices and denote the elements of the power-set of $\mathcal{P}(u_k, \Delta_k)$ *by* $\mathcal{B}^k_1, ..., \mathcal{B}^k_{m_k}$ *with* $m_k = 2^{|\mathcal{P}(u_k,\Delta_k)|}$.

for $j = 1, ..., m_k$ ***do***

Solve the adjoint equation

$$A(\mathcal{A}_s(u_k) \cup \mathcal{B}^k_j)p^k_j = \nabla_y J(y_k, u_k),$$

and set $g^k_j = p^k_j + \nabla_u J(y_k, u_k)$.

16: ***end for***

17: *Compute an inexact, but feasible solution* d_k *of the* modified trust-region subproblem

$$\left.\begin{aligned} \min_{\zeta\in\mathbb{R}, d\in\mathbb{R}^n} \quad & \mathfrak{q}_k(d,\zeta) := f(u_k) + \zeta + \frac{1}{2}\, d^\top H_k d \\ \text{s.t.} \quad & \|d\| \le \Delta_k, \\ & \langle g_j^k, d\rangle \le \zeta \quad \forall\, j = 1, ..., m_k. \end{aligned}\right\} \tag{$\mathfrak{Q}_k$}$$

that satisfies the modified Cauchy-decrease condition

$$f(x_k) - \mathfrak{q}_k(d_k, \zeta_k) \ge \frac{\mu}{2}\, \psi(u_k, \Delta_k)\, \min\left\{\Delta_k, \frac{\psi(u_k,\Delta_k)}{\|H_k\|}\right\}. \tag{4.28}$$

18: *Compute the stationarity measure* $\psi(u_k, \Delta_k)$ *as solution of*

$$\psi(u_k,\Delta_k) = -\min_{\xi\in\mathbb{R}, d\in\mathbb{R}^n} \{\xi : \|d\| \le 1,\ \langle g_j^k, d\rangle \le \xi \quad \forall\, j = 1, ..., m_k\}. \tag{4.29}$$

19: *Compute the modified quality indicator*

$$\rho_k := \begin{cases} \dfrac{f(u_k) - f(u_k + d_k)}{f(u_k) - \mathfrak{q}_k(d_k,\zeta_k)}, & \text{if } \psi(u_k,\Delta_k) > \|g_k\|\,\Delta_k \\ 0, & \text{if } \psi(u_k,\Delta_k) \le \|g_k\|\,\Delta_k. \end{cases}$$

20: ***end if***

21: *Update: Set*

$$u_{k+1} := \begin{cases} u_k, & \text{if } \rho_k \le \eta_1 \quad \textit{(null step)}, \\ u_k + d_k, & \text{otherwise} \quad \textit{(successful step)}, \end{cases}$$

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

Set $k = k + 1$.

22: ***end if***

23: ***end for***

REMARK 4.13. *Completely analogously to Lemma 3.2, one shows that the minimum on the right hand side of (4.29) equals* $-\min_{\|d\|\le 1} \phi(u_k, \Delta_k; d)$*, which is the stationarity measure from Assumption 2.4(2b).*

THEOREM 4.14. *Assume that Algorithm 4.12 does not terminate in finitely many steps and that Assumption 2.10 is satisfied for all* $k \in \mathbb{N}$*. Then every accumulation point of the sequence of iterates is C-stationary.*

Proof. As seen in Lemma 3.2, since the inexact, but feasible solution (d_k, ζ_k) of $(\mathfrak{Q}_k)$ satisfies (4.28), d_k also fulfills the modified Cauchy-decrease condition in (2.6). Therefore, we can apply the results for our general Algorithm 2.6. As Lemmata 4.7, 4.10, and 4.11 show, the conditions in Assumptions 3.1 and 3.4, that guarantee the convergence results for our trust-region algorithm applied to problems with composite functions as in (3.1), are fulfilled in this concrete setting. Thus, Corollary 3.8 yields the claim. ∎

5. Numerical results. In this section we verify the convergence properties of the proposed trust-region algorithm by means of two different examples. The first one is a toy problem in $\mathbb{R}^2$, for which the solution can be explicitely obtained and, moreover, the convergence hypotheses of the method analytically proved. Our purpose for this first experiment is to verify how the algorithm behaves in nondifferentiable points, i.e., biactive points, either when an optimal solution is biactive or when the algorithm has to move from such a point to further decrease the cost function.

The second example is concerned with the optimization of a variational inequality of the second kind arising from the discretization of an optimal control problem. The nondifferentiability in the variational inequality consists of the discretized L^1 norm of the state variable. The design variable (control) is the right hand side of the inequality, and represents a distributed control force on the whole geometric domain (see [10] for further details).

Unless otherwise specified, the used trust-region parameters are: $\eta_1 = 0.25$, $\eta_2 = 0.75, \beta_1 = 0.5, \beta_2 = 1.1$ and the initial radius for the algorithm was set to $\Delta_0 = 1$. We use a positive definite BFGS second order matrix H_k, which is updated in every successful trust-region step. For the fraction of Cauchy decrease, we consider the parameter $\mu = 0.8$. The algorithm stops whenever $\frac{|u_{k+1}-u_k|}{|u_0|}$ and the trust-region radius are smaller than a given tolerance. To accelerate the method we also compute the quasi Newton-step $-H_k^{-1}g$ and apply a dogleg strategy (see, e.g., [16]).

Experiment 1. We consider here the numerical solution of a toy example in $\mathbb{R}^2$ to illustrate the main problem and algorithm features. Let us consider the simplified variational inequality:

$$2y(v-y) + |v| - |y| \geq u(v-y), \quad \forall v \in \mathbb{R}, \tag{5.1}$$

whose solution can be obtained using the soft thresholding operator and is given in closed form by:

$$y = \begin{cases} 1/2(u-1) & \text{if } u \geq 1, \\ 0 & \text{if } u \in [-1,1], \\ 1/2(u+1) & \text{if } u \leq -1. \end{cases} \tag{5.2}$$

The solution mapping is clearly globally Lipschitz continuous and directionally differentiable. The directional derivative at u in direction h is given by $\eta \in \mathcal{K}(y)$ solution of

$$2\eta(v-\eta) \geq h(v-\eta), \quad \forall v \in \mathcal{K}(y), \tag{5.3}$$

where $\mathcal{K}(y)$ is the convex cone defined by

$$\mathcal{K}(y) := \{v \in \mathbb{R} : v = 0 \text{ if } |q| < 1; vq \geq 0 \text{ if } y = 0 \text{ and } |q| = 1\}. \tag{5.4}$$

Since for this simplified case the biactive set corresponds to the cases $u \in \{-1, 1\}$ and the set where $|q| < 1$ is the same as $u \in (-1, 1)$, the cone may also be written as

$$\mathcal{K}(y) = \{v \in \mathbb{R} : v = 0 \text{ if } u \in (-1,1); v \geq 0 \text{ if } u = 1, v \leq 0 \text{ if } u = -1\} \tag{5.5}$$

$$= \left\{ v \in \mathbb{R} : v \begin{cases} \geq 0 & \text{if } u = 1, \\ = 0 & \text{if } u \in (-1,1), \\ \leq 0 & \text{if } u = -1. \end{cases} \right\} \tag{5.6}$$

From the projection formula on convex sets it then follows that

$$\eta = \mathcal{P}_{\mathcal{K}(y)}(\eta - c(2\eta - h)) = \mathcal{P}_{\mathcal{K}(y)}((1-2c)\eta + ch)), \quad \forall c > 0. \tag{5.7}$$

Considering the quadratic cost function

$$J(y,u) = \frac{1}{2}|y - z_d|^2 + \frac{\alpha}{2}|u|^2, \tag{5.8}$$

we may define the reduced objective $f(u) := J(y(u), u)$, which, thanks to the Lipschitz continuity of the solution mapping, is a locally Lipschitz function. The directional derivative is then given by

$$f'(u)h = (y - z_d)^T \eta + \alpha u^T h \tag{5.9}$$

Taking the particular value $c = 1/2$ in the projection formula (5.7) we get that

$$\eta = \mathcal{P}_{\mathcal{K}(y)}\left(\frac{h}{2}\right) = \begin{cases} \frac{1}{2}\max(0,h) & \text{if } u = 1, \\ 0 & \text{if } u \in (-1,1), \\ \frac{1}{2}\min(0,h) & \text{if } u = -1, \\ \frac{1}{2}h & \text{elsewhere.} \end{cases} \tag{5.10}$$

Consequently,

$$f'(u)h = \begin{cases} \frac{1}{2}(y - z_d)\max(0,h) + \alpha h & \text{if } u = 1, \\ \alpha u h & \text{if } u \in (-1,1), \\ \frac{1}{2}(y - z_d)\min(0,h) - \alpha h & \text{if } u = -1, \\ \frac{1}{2}(y - z_d)h + \alpha u h & \text{elsewhere.} \end{cases} \tag{5.11}$$

For the solution of this particular instance, we consider the Algorithm 4.12 with the direction $g \in \partial_B f(u)$ given by $g = -(p + \alpha u)$, where the adjoint state p is computed through

$$p_i = \begin{cases} 0 & \text{if } i \notin \mathcal{I} \\ (y - z_d)_i & \text{if } i \in \mathcal{I}. \end{cases} \tag{5.12}$$

Since in this case we are working with a single real variable, the direction can be explicitely written as

$$g = \begin{cases} -\alpha u & \text{if } u \in [-1,1], \\ -\frac{1}{2}(y - z_d) - \alpha u & \text{elsewhere.} \end{cases} \tag{5.13}$$

Using (5.11) we may compute the directional derivative along the Bouligand direction g yielding

$$f'(u)g = \begin{cases} \frac{1}{2}(y - z_d)\max(0, -\alpha u) - \alpha^2 u^2 & \text{if } u = 1, \\ -\alpha^2 u^2 & \text{if } u \in (-1,1), \\ \frac{1}{2}(y - z_d)\min(0, -\alpha u) - \alpha^2 u^2 & \text{if } u = -1, \\ -\left[\frac{1}{2}(y - z_d) + \alpha u\right]^2 & \text{elsewhere.} \end{cases} \tag{5.14}$$

$$= \begin{cases} -\alpha^2 u^2 & \text{if } u \in [-1,1], \\ -\left[\frac{1}{2}(y - z_d) + \alpha u\right]^2 & \text{elsewhere.} \end{cases} \tag{5.15}$$

Consequently, $f'(u)g = -|g|^2$ for the direction considered, i.e., the element in Assumption 3.4 is explicitely given.

If the trust-region radius becomes smaller than $\Delta_{\min} = 1e-2$, problem $(\mathfrak{Q}_k)$ is solved for all possible variations of the possibly biactive set according to Definition 4.9. For the current toy problem the latter reduces to solve two auxiliary QP problems, and one extra for determining $\psi(u_k, \Delta_k)$. The auxiliary problems are solved using a Sequential Least Squares Programming (SLSQP) algorithm available in SciPy's Optimization library.

Since in this case the solution can be computed analytically, the different properties of the algorithm can also be easily verified. This involves in particular the attainability of minima, mainly when these points are biactive, or the escape from biactive points when they are reached along the iterative process.

In Figure 5.1 the solution operator and the composite cost function are plotted. As it can be oberved, the resulting objective function is piecewise differentiable with two local minima at $\bar{u}_1 = -1$ and $\bar{u}_2 = (4\alpha u_d + 2z_d + 1)/(4\alpha + 1)$. The points $u = -1$ and $u = 1$ are biactive, and only one of them corresponds to a local minimum for the problem.

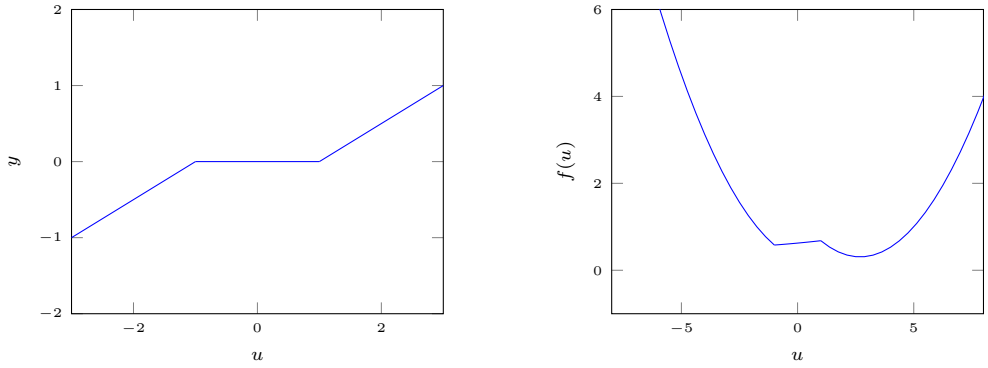

FIG. 5.1. *Solution operator (left) and cost function (right). Parameters:* $\alpha = 0,01$, $z_d = 1$, $u_d = -5$.

If the initial point of Algorithm 4.12 is chosen below or equal than one, then the method converges towards the local minimum $\bar{u}_1$, while it converges to $\bar{u}_2$ if the starting iterate is chosen greater than one. In Figure 5.2 the number of trust-region iterations for different initial points and α values is depicted. It can be easily observed that reaching the nonsmooth minimum requires significantly more iterations than reaching the smooth one, which is intuitively expected. Despite of that, the total number of trust-region iterations stays below 20 in all cases.

Experiment 2. Let us start this experiment by recalling problem $(\mathrm{P_{VI}})$:

$$\left.\begin{array}{ll} \min\limits_{y,u\in\mathbb{R}^n} & J(y,u) \\ \text{s.t.} & \langle \nu^{-1}Ay, v-y\rangle + |v|_1 - |y|_1 \geq \langle \nu^{-1}u, v-y\rangle, \quad \forall\, v\in\mathbb{R}^n, \end{array}\right\} \tag{5.16}$$

where $|.|_1$ denotes the 1-norm, i.e., $|v|_1 = \sum_{i=1}^n |v_i|$. We consider the matrix A as the homogeneous finite differences discretization matrix of the negative two-dimensional

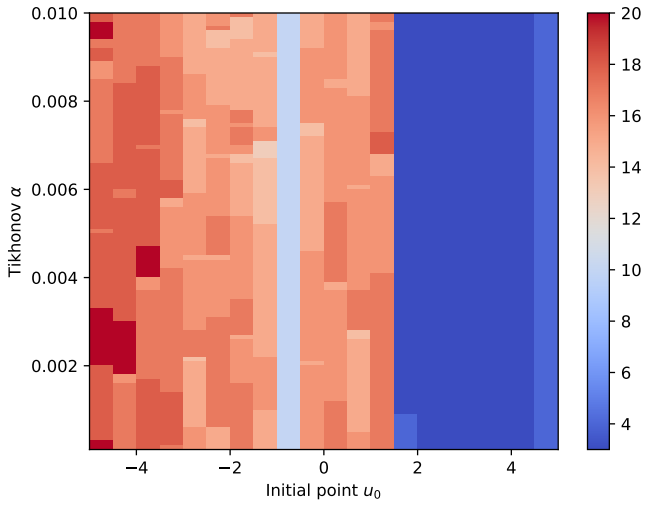

FIG. 5.2. *Number of trust-region iterations, depending on the initial guess* $u_0 \in [-5, 5]$ *and the Tikhonov parameter* $\alpha \in [1e-4, 1e-2]$.

Laplace operator with homogeneous Dirichlet boundary conditions on the domain $\Omega = (0,1)^2$. The 1-norm, together with the weight ν, is responsible for the level of sparsity in the computed state. When ν increases, the state becomes sparser up to a certain threshold where the state is identically zero.

As cost function we choose the classical tracking–type objective

$$J(y,u) := \frac{1}{2}\|y - z_d\|^2 + \frac{\alpha}{2}\|u\|^2,$$

where $\alpha > 0$ is the Tikhonov weight and $z_d = \chi_{x_1 \geq 1}$. In Figure 5 the controlled states for different values of ν are shown. As expected, it can be observed that as ν becomes larger, the region where the state has zero value also increases. This is something expected from the structure of the variational inequality constraint. Concerning the behaviour of the biactive sets, in Figure 5 these sets are plotted for different ν values. It can be realized that the biactive region is not dismissible in practice and should be carefully handled.

Implementation details. The computational domain Ω was discretized using homogeneous finite differences with step size $h = \frac{1}{n+1}$, where n is the number of inner discretization points. The resulting stiffness matrix A is therefore penta-diagonal, symmetric and positive definite. For the solution of the linear systems we therefore considered MATLAB's exact sparse solver. The trust-region radius lower bound was set to $\Delta_{\min} = 1e-3$

The main costly step in our algorithm is the solution of the variational inequality constraint. To improve the total computing time we therefore consider the inexact solution of the variational inequality constraint in the first trust-region iterations, and the exact solution of the problem in the last ones. Specifically, up to a tolerance of $1\mathrm{e}-2$ for the norm of the trust-region residuum, the variational inequality is solved by means of a semismooth Newton method (with a Huber regularization [15]). After

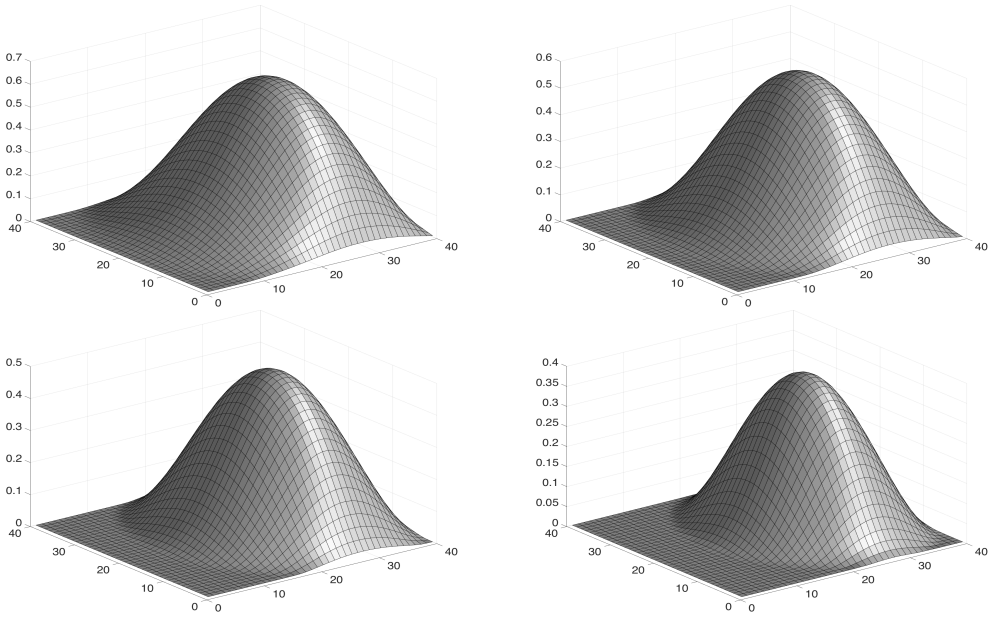

FIG. 5.3. *Optimal controlled state y for different parameters ν. From the left upper corner to the lower right corner: $\nu = 4$, $\nu = 8$, $\nu = 12$, $\nu = 18$. Tikhonov parameter: $\alpha = 0.001$; mesh size step $h = 1/41$.*

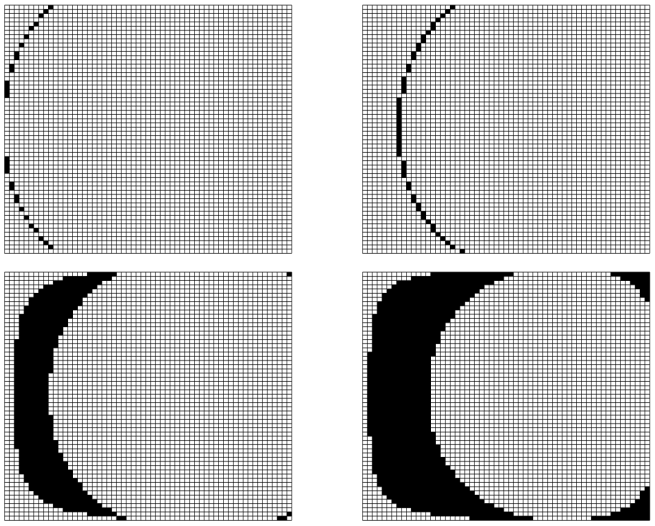

FIG. 5.4. *Biactive sets for different parameters ν. From the left upper corner to the lower right corner: $\nu = 4$, $\nu = 8$, $\nu = 12$, $\nu = 18$. Tikhonov parameter: $\alpha = 1\mathrm{e} - 3$; mesh size step $h = 1/61$.*

that, and until the required tolerance is reached (typically $1\mathrm{e} - 6$), the lower level problem is solved using a recently proposed orthantwise method [9]. The orthantwise method stops whenever the norm of the pseudo-gradient is smaller than $1\mathrm{e} - 7$.

Performance. The proposed inexact trust–region algorithm does not have significant difficulties for solving the optimization problem at hand, even when the solution

	$h = 1/20$				$h = 1/40$			
α \ ν	4	8	12	18	4	8	12	18
$1e-1$	46(20)	46(20)	46(20)	46(20)	53(25)	53(25)	53(25)	53(25)
$1e-2$	26(22)	69(20)	69(20)	69(20)	30(26)	76(24)	76(24)	76(24)
$1e-3$	35(31)	28(23)	26(21)	72(08)	46(41)	36(25)	29(25)	79(08)
$1e-4$	35(30)	38(33)	41(35)	72(08)	104(40)	105(37)	85(42)	79(08)

TABLE 5.1
Number of iterations in two different meshes for different values of α and ν.

has large biactive sets. The total iteration number of the trust–region method is provided in Table 5.1 for two medium size meshes and different values of the Tikhonov parameter α and the coefficient ν. The number of trust-region iterations where the semismooth Newton solver was used is registered in parenthesis.

6. Conclusions. We have presented a non-smooth trust–region algorithm for solving optimization problems with locally Lipschitz continuous functions and, under suitable assumptions, we prove convergence of the iterates to a C-stationary point. The structure of the trust-region subproblem is dependent on the size of the current trust-region radius and allows to escape from non-differentiable points which are not necessarily local minima.

As a particular instance we have considered optimization problems with a class of variational inequality constraints and proposed a computable model function to be used along the trust–region iterations. The construction of this model is based on a precise characterization of the Bouligand-subdifferential associated with the solution operator of the variational inequality. Moreover, thanks to the structure of these types of problems, the computation of the solution to the trust-region subproblem ($\mathfrak{Q}_k$) can be carried out just by adding a finite number of inequality constraints dependent on the size of the biactive set.

The proposed algorithm is general enough to deal with several classes of problems. We tested it for the particular case of optimization problems constrained by variational inequalities of the second kind and its performance was succesfully verified. The investigation of the behaviour of the algorithm for other types of problem is a matter of future research.

Acknowledgement. The authors thank Stephan Walther (TU Dortmund) for elaborating the proof of Lemma 2.15.

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, 733–754, 2015

[2] Apkarian, P., Noll, D. and Ravanbod, L. Nonsmooth bundle trust-region algorithm with applications to robust stability. *Set-Valued and Variational Analysis*, Vol. 24(1), pp. 115-148, 2016.

[3] Clarke, F.H. *Optimization and Nonsmooth Analysis*, SIAM, Philadelphia, 1990.

[4] Colson, B., Marcotte, P. and Savard, G. A trust-region method for nonlinear bilevel programming: algorithm and computational experience, *Computational Optimization and Applications*, Vol. 30(2), pp. 211–227, 2005.

[5] Conn, A. R., Gould, N. I. and Toint, P. L. *Trust region methods*. SIAM. 2000.
[6] Christof, C., De los Reyes, J.C. and Meyer, C. A non-smooth trust-region method for B-differentiable functions with application to optimization problems constrained by variational inequalities, *ArXiv:*1711.03208, 2018.
[7] De los Reyes, J.C. Optimal control of a class of variational inequalities of the second kind, *SIAM Journal on Control and Optimization*, Vol. 49, pp. 1629-1658, 2011.
[8] De los Reyes, J.C. *Numerical PDE-constrained optimization*. Springer Verlag. 2015.
[9] De los Reyes, J.C., Loayza, E. and Merino, P. Second-order orthant-based methods with enriched Hessian information for sparse ℓ_1-optimization, to appear in *Computational Optimization and Applications*.
[10] De los Reyes, J.C. and Meyer, C. Strong Stationarity Conditions for a Class of Optimization Problems Governed by Variational Inequalities of the Second Kind, *Journal of Optimization Theory and Applications*, Vol. 168(2), pp. 375–409, 2016.
[11] Dennis, J.E., Li, S.B.B. and Tapia, R.A. A unified approach to global convergence of trust region methods for nonsmooth optimization. *Mathematical Programming*, 68, 319-346. 1995.
[12] Evans, L.C. *Partial Differential Equations*, Graduate Studies in Mathematics, Vol. 19, AMS, Rhode Island, 1998.
[13] Giallombardo, G. and Ralph, D. Multiplier convergence in trust-region methods with application to convergence of decomposition methods for MPECs. *Mathematical Programming*, 112(2), 335-369. 2008.
[14] Goldstein, A., Optimization of Lipschitz continuous functions. *Math. Program.*, 13, 14-–22. 1977.
[15] Huber, Peter J. *Robust statistics*, Springer, 2011.
[16] Kelley, C.T. *Iterative methods for optimization*, SIAM, 1999.
[17] Luo, Z.-Q., Pang, J.-S. and Ralph, D. *Mathematical programs with equilibrium con- straints*, Cambridge University Press, Cambridge, 1996.
[18] Marcotte, P., Savard, G. and Zhu, D.L. A trust region algorithm for nonlinear bilevel programming. *Operations Research Letters*, Vol. 29(4), 171-179, 2001.
[19] Mignot, F. Contrôle dans les Inéquations Variationelles Elliptiques, *J. Func. Analysis*, Vol. 22, pp. 130–185, 1976.
[20] Outrata, J., and Zowe, J. A numerical approach to optimization problems with variational inequality constraints. *Mathematical Programming*, 68(1–3), 105–130, 1995.
[21] Qi, L. and Sun, J. A trust region algorithm for minimization of locally Lipschitzian functions, *Math. Prog.*, 66, pp. 25–43, 1994.
[22] Schramm, H., and Zowe, J. A Version of the Bundle Idea for Minimizing a Nonsmooth Function: Conceptual Idea, Convergence Analysis, Numerical Results. *SIAM Journal on Optimization*, 2(1), 121–152, 1992.
[23] Schirotzek, W. *Nonsmooth Analysis*, Springer Verlag. 2007.
[24] Scholtes, S. *Introduction to piecewise differentiable equations*, Springer Verlag. 2012.
[25] Scholtes, S. and Stöhr, M. Exact penalization of mathematical programs with equilibrium constraints. *SIAM Journal on Control and Optimization*, 37(2), 617-652. 1999.
[26] Sun, W. and Yuan, Y.-X. *Optimization theory and methods: nonlinear programming*, Springer, 2006.