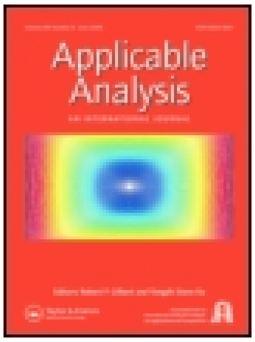

Taylor & Francis
Taylor & Francis Group

Applicable Analysis
An International Journal

ISSN: 0003-6811 (Print) 1563-504X (Online) Journal homepage: https://www.tandfonline.com/loi/gapa20

Nonsmooth optimization by successive abs-linearization in function spaces

Andrea Walther, Olga Weiß, Andreas Griewank & Stephan Schmidt

To cite this article: Andrea Walther, Olga Weiß, Andreas Griewank & Stephan Schmidt (2020): Nonsmooth optimization by successive abs-linearization in function spaces, Applicable Analysis, DOI: 10.1080/00036811.2020.1738397

To link to this article: https://doi.org/10.1080/00036811.2020.1738397

Published online: 12 Mar 2020.

Submit your article to this journal

View related articles

View Crossmark data

APPLICABLE ANALYSIS
https://doi.org/10.1080/00036811.2020.1738397

Nonsmooth optimization by successive abs-linearization in function spaces

Andrea Walther[a], Olga Weiß[b], Andreas Griewank[a] and Stephan Schmidt[b]

[a]Department of Mathematics, Humboldt-Universität zu Berlin, Berlin, Germany; [b]Department of Mathematics, Universität Paderborn, Paderborn, Germany

ABSTRACT
We present and analyze the solution of nonsmooth optimization problems by a quadratic overestimation method in a function space setting. Under certain assumptions on a suitable local model, we show convergence to first-order minimal points. Subsequently, we discuss an approach to generate such a local model using the so-called abs-linearization. Finally, we discuss a class of PDE-constrained optimization problems incorporating the L^1-penalty term that fits into the considered class of nonsmooth optimization problems.

ARTICLE HISTORY
Received 8 February 2019
Accepted 23 February 2020

COMMUNICATED BY
Jen-Chih Yao

KEYWORDS
Abs-linearization; quadratic overestimation method; first-order minimality

2010 MATHEMATICS SUBJECT CLASSIFICATIONS
49J52; 65K10; 47N10; 35Q93

1. Introduction and motivation

In a finite dimensional setting with $V = \mathbb{R}^n$, the minimization of a piecewise smooth function $\varphi : V \to \mathbb{R}, y = \varphi(x)$, based on successive abs-linearization has been analyzed extensively in a series of papers, see, e.g. [1–3]. An essential ingredient for the results obtained so far is the second order approximation property of the local abs-linear model, the proof of which relies essentially on the Lipschitz continuity of all quantities involved. For that purpose, the paper [2] studied a class of functions $\varphi : \mathbb{R}^n \to \mathbb{R}$ that are piecewise smooth in the sense of Scholtes [4], where the nonsmoothness is caused by the absolute value function exclusively, hence also covering max- and min-functions as well as complementary conditions formulated in an appropriate way. For a function of this class, one can define and number all arguments of absolute value evaluations successively as *switching variables* z_i for $i = 1 \dots s$, where it is assumed throughout that z_j can only influence z_i if $j < i$. Then, the paper [2] proposed a new approach to generate a local so-called abs-linear model $\Delta\varphi(x; \cdot) : \mathbb{R}^n \to \mathbb{R}$ of $\varphi(\cdot)$ using the technique of abs-linearization. One major result of that paper is the good approximation property of $\Delta\varphi$ in that

$$\varphi(x + \Delta x) - \varphi(x) = \Delta\varphi(x; \Delta x) + \mathcal{O}(\|\Delta x\|^2) \tag{1}$$

for $\Delta x \to 0$. We will show a similarly good approximation property also for a local model in the infinite dimensional setting using an appropriate extension of the abs-linearization.

Based on Equation (1), the following iterative optimization algorithm was proposed in [2]

$$x_{k+1} = x_k + \underset{\Delta x}{\arg\min} \left\{\varphi_{loc}(x_k; \Delta x) + q\|\Delta x\|^2\right\} \tag{2}$$

CONTACT Olga Weiß ✉ ebelo@math.upb.de

with $\varphi_{loc}(x_k; \Delta x) \equiv \varphi(x_k) + \Delta\varphi(x_k; \Delta x)$. We call this approach SALMIN for Successive Abs-Linear MINimization. The penalty factor q of the quadratic term is an estimated bound on the discrepancy between φ and its abs-linearization. This method was shown in [2] to generate a sequence of iterates $\{x_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ whose cluster points are first-order minimal. If the inner problem of minimizing the regularized piecewise linear model is not solved exactly, but increments Δx that are merely Clarke stationary for $\Delta\varphi$ are accepted also, then the cluster points are guaranteed to be also Clarke stationary as shown in [1]. However, when extending the results from the finite dimensional case to a function space setting, the Lipschitz continuity of the absolute value might be lost, see, e.g. [5]. Hence, it is not possible to transfer the results obtained for $V = \mathbb{R}^n$ directly to the infinite dimensional case. Nevertheless, it is interesting to analyze the more general infinite dimensional case in a Banach space setting. For example, this is needed, when appropriate discretizations of norms are required. Generally, infinite dimensional optimization problems, where nonsmoothness stems directly from the problem formulation, can arise in many applications. To name just one class of optimization tasks, we consider optimal control problems subject to a partial differential equation (PDE) as constraint where the objective functional contains the L^1-norm of the control and is therefore nondifferentiable. Problems of this type are of great interest for applications, in which one cannot put control devices all over the control domain. Since the L^1-term leads to sparsely supported optimal controls, the solution to such an optimization problem then provides information about where it is most efficient to put these control devices.

The SALMIN algorithm given by Equation (2) can be interpreted as a quadratic overestimation method, where the error between the model and the real function is bounded by the second power of the distance, see, e.g. [6, 7]. This approach is also related to proximal point methods as analyzed for the finite dimensional setting for instance in [8, 9], and for the infinite setting for example in [10–12] as well as in [13]. There the original function is still an additive component of the local subproblem, which is therefore not much easier to solve than the original problem. In contrast to the results presented in these papers, in Equation (2) the local model of the function to be minimized at the current iterate x_k is used instead of the original function. In any case, it is not possible to transfer the available results directly to the situation considered here. In infinite dimensions, optimization methods using a local model can be based on a bundle approach. Corresponding convergence results for convex target functions can be found in [14, 15]. An alternative bundle method covering also nonconvex target functions is presented in [16]. There, global convergence to approximate stationary points is shown.

The purpose of this paper is twofold. Assuming that one has a local model with similar approximation properties as in the finite dimensional case, we will first show that the SALMIN algorithm generates iterates that converge to a first-order minimal point also in the infinite dimensional case. To this end, we will consider two reflexive Banach spaces V and $\hat{V}$, where V is compactly embedded in $\hat{V}$. Then as a first result we will obtain a sequence that converges weakly to a limit point in V. Using the compact embedding, subsequently we will show that the limit point is Clarke stationary and under certain conditions even first-order minimal. Second, we propose an approach to generate a local model that provably has the required approximation properties. Finally, we will discuss an example from PDE-constrained optimization that fits into the considered setting.

This paper has the following structure. We define the function class that we want to consider in Section 2. This includes also the local model and its analysis. Here, it is important to note that the concepts of piecewise smoothness and piecewise linearity do not directly carry over to the infinite dimensional setting. However, extensions of these concepts are considered in [17]. In Section 3 we extend the proposed quadratic overestimation method given by Equation (2) to the infinite dimensional case and analyze its convergence behavior. Subsequently, we propose the generation of a local model in Section 4, where an approximation property similar to Equation (1) will be shown. In this section, we will also give an example for an optimization problem in function spaces that fits to our setting. Finally, we draw conclusions and provide an outlook in the final Section 5.

2. The considered function class and a suitable local model

From now on, we consider the function space $V = L^p(\Omega)$ with $1 < p < \infty$ for a given bounded domain $\Omega \subset \mathbb{R}^n$. Note, that due to the restriction on p, the resulting function space V is a reflexive Banach space. Furthermore, we consider a second reflexive Banach space $\hat{V}$ such that V is compactly embedded into $\hat{V}$. An example for this situation would be $V = L^2(\Omega)$ and $\hat{V} = H^{-1}(\Omega)$ as dual space of $H_0^1(\Omega)$.

To cover the kind of nonsmoothness studied in this paper we define

$$\begin{aligned} &\mathrm{abs} : V \to V, \\ &[\mathrm{abs}(v)](x) = |v(x)| \quad \text{for every } v \in V \text{ and for almost all } x \in \Omega. \end{aligned} \tag{3}$$

as the Nemytskii operator induced by the absolute value function. Such operators are also called superposition operators, see [18]. For better readability we will sometimes omit the local argument x and thus consider $\mathrm{abs}(\cdot)$ directly as an operator on the function space. The $\mathrm{abs}(\cdot)$ operator can enforce sparsity if included appropriately in the target function, see, e.g. [19, 20]. Furthermore, it can be used to describe a class of partial differential equations involving nonsmooth but Lipschitz continuous and directionally differentiable nonlinearities such as those appearing in the two-phase Stefan problem [21].

In general Banach spaces, it is not clear whether the absolute value operator is Lipschitz continuous, see, e.g. [5]. Therefore, we state the following result for the function spaces considered here:

Proposition 2.1 (Lipschitz continuity of abs(·)): *The absolute value operator* $\mathrm{abs} : V \to V, \mathrm{abs}(v) := |v|$, *is Lipschitz continuous.*

Proof: One has for almost every $x \in \Omega$ and $v, u \in V$ that

$$\big|\,|v(x)| - |u(x)|\,\big| \leqslant |v(x) - u(x)|$$

such that

$$\begin{aligned} \|\mathrm{abs}(v) - \mathrm{abs}(u)\|_V^p &= \int_\Omega \big||v(x)| - |u(x)|\big|^p \,\mathrm{d}x \\ &\leqslant \int_\Omega |v(x) - u(x)|^p \,\mathrm{d}x = \|v - u\|_V^p. \end{aligned} \qquad \blacksquare$$

It is worth mentioning that since the Lipschitz constant is 1, one obtains also that the absolute value operator is nonexpansive. Furthermore, one can conclude that from the kind of admissible nonsmoothness given by the absolute value operator, the directional derivative defined by

$$\varphi'(v; h) \equiv \lim_{t \to 0_+} \frac{1}{t}(\varphi(v + th) - \varphi(v)), \tag{4}$$

exists for all $v \in V$ and all directions $h \in V$.

Now, we define the class of operators considered here formally. In analogy to the class $C^d_{abs}(\mathbb{R}^n)$ in finite dimensions as defined in [3], we denote this class by $C^1_{\mathrm{abs}}(V)$.

Definition 2.2 (Operator Class $C^1_{\mathrm{abs}}(V)$): For a reflexive Banach space V, the class $C^1_{abs}(V)$ contains all operators $\varphi : V \to \mathbb{R}$ such that φ can be represented as a composition of continuously Fréchet differentiable operators $\psi_i : V_i \to \tilde{V}_i$ between some reflexive Banach spaces V_i and $\tilde{V}_i$ and the Nemitzkii operator induced by the absolute value operator abs as defined in Equation (3).

Depending on the specific situation, the Lipschitz-continuously Fréchet differentiable operators ψ_i are mappings between various Banach spaces, which preserve the Lipschitz continuity shown in Proposition 2.1. However, for our purpose, only the overall mapping from V to $\mathbb{R}$ is important. Using the well-known reformulations

$$\begin{aligned} \min(v,u) &= (v+u-\mathrm{abs}(v-u))/2 \quad \text{and} \\ \max(v,u) &= (v+u+\mathrm{abs}(v-u))/2, \end{aligned} \tag{5}$$

a large class of nonsmooth operators and functions is contained in $C^1_{\mathrm{abs}}(V)$.

Example 2.3: For a bounded domain $\Omega \subset \mathbb{R}^3$ and a given desired state $y_d \in H^1_0(\Omega)$, consider the optimization problem

$$\begin{aligned} \min_{(y,u)\in H^1_0(\Omega)\times L^2(\Omega)} \quad & \frac{1}{2}\|y-y_d\|^2_{L^2} + \frac{\alpha}{2}\|u\|^2_{L^2} + \beta\|u\|_{L^1} \\ \text{s.t.} \quad & Ay + l(y) = u + f \text{ in}\Omega, \end{aligned} \tag{6}$$

where y and u represent the state and the control, respectively, $f \in L^2(\Omega)$ some given function and $\alpha > 0$, $\beta > 0$ are parameters. Furthermore, $A : H^1_0(\Omega) \to H^{-1}(\Omega)$ is a linear elliptic differential operator of second order, e.g. the Laplace operator. Suppose, that the continuously differentiable operator $l : H^1_0(\Omega) \to L^2(\Omega)$ is such that there exists a Lipschitz-continuously Fréchet differentiable solution operator $S : L^2(\Omega) \to H^1_0(\Omega)$, which gives the solution $S(u) = y$ of the PDE in Equation (6) for any fixed control $u \in L^2(\Omega)$. As a very simple example one may consider $l(y) \equiv 0$ as in [20]. Then, the reduced problem formulation is given by the optimization problem

$$\min_{u\in L^2(\Omega)} \varphi(u) \quad \text{with} \quad \varphi(u) = \frac{1}{2}\|S(u)-y_d\|^2_{L^2} + \frac{\alpha}{2}\|u\|^2_{L^2} + \beta\|u\|_{L^1}. \tag{7}$$

As shown in Section 4, the assumed structure of the $C^1_{\mathrm{abs}}(V)$ class allows the construction of a suitable local model for the quadratic overestimation method discussed in the next section.

Since we want to minimize the objective functional φ, we restate first-order necessary conditions and introduce here Clarke's concept of generalized derivatives, see, e.g. [22, Section 1.2].

Definition 2.4 (Clarke Generalized Gradient): Suppose, $\varphi \in C^1_{\mathrm{abs}}(V)$, i.e. φ is also locally Lipschitz continuous. Let $\bar{v}, h \in V$ be given. Then the limit superior

$$\limsup_{\substack{v\to\bar{v}\\ \lambda\to 0_+}} \frac{1}{\lambda}\left(\varphi(v+\lambda h) - \varphi(v)\right) \equiv \varphi^C(\bar{v},h)$$

exists and is called *Clarke derivative* of φ at $\bar{v}$ in direction h. Since this limit superior exists for all $h \in V$, the function φ is called Clarke differentiable at $\bar{v}$. The set

$$\partial_C\varphi(\bar{v}) \equiv \{\xi \in V^* : \varphi^C(\bar{v},h) \geqslant \xi(h)\ \forall\, h \in V\} \subset V^*$$

denotes the Clarke *generalized gradient* or *subdifferential* of φ at $\bar{v}$, where V^* refers to the dual space of V.

Since one has for a function $\varphi : V \to \mathbb{R}$ that is Fréchet differentiable at $\bar{v}$ the inclusion $\varphi^C(\bar{v},\cdot) \in \partial_C\varphi(\bar{v})$ [22, Proposition 2.2.2], the concept of Clarke derivatives fits well for the nonsmooth case analyzed in this paper. As a necessary optimality condition, one has for φ being an element of the considered nonsmooth function class the following result: If v^* is a minimal point of φ then the functional 0_{V^*} is an element of $\partial_C\varphi(v^*)$, see e.g. [23, Proposition 6] and [24, Theorem 3.46].

However, in contrast to many other approaches, we aim at first-order minimality that is defined as follows

Definition 2.5 (First-order Minimality): Suppose, $\varphi \in C^1_{\text{abs}}(V)$. The operator φ is called *first-order minimal* at $v_* \in V$ if one has

$$0 \leqslant \varphi'(v_*; h) \quad \text{for all } h \in V. \tag{8}$$

Then, v_* is called *first-order minimal point.*

In the general nonconvex case this property is stronger than the frequently used concept of Clarke stationarity. For example, the function $-|x| : \mathbb{R} \to \mathbb{R}$ at the origin is Clarke stationary but not first-order minimal. However, both concepts coincide in the convex case, which will play an important role in the following chapter. Often, first-order minimality is also called criticality as defined in [25, 26], where $0 \in \mathbb{R}^n$ must be a Fréchet subgradient.

To prove convergence of the quadratic overestimation method proposed in the next section, we use the following properties.

Assumption 2.6 (Approximation Properties): Suppose, $\varphi \in C^1_{\text{abs}}(V)$ and $W \subset V$ is a closed convex subset. For all $\bar{v} \in W$, there exists a Lipschitz continuous local model $\varphi_{loc}(\bar{v}; \cdot) : V \to \mathbb{R}$ with $\varphi_{loc}(\bar{v}; \cdot) \in C^1_{\text{abs}}(V)$, given by a finite composition of linear functions and the absolute value operator. There exists a constant $\bar{q} > 0$ such that for all pairs $\bar{v}, v \in W$ one has

$$\varphi(\bar{v}) = \varphi_{loc}(\bar{v}; 0), \quad |\varphi(v) - \varphi_{loc}(\bar{v}; v - \bar{v})| \leqslant \bar{q}\|v - \bar{v}\|_V^2. \tag{9}$$

The quadratic model corresponding to such a local model is then defined by

$$\varphi_Q(\bar{v}; \cdot) \equiv \varphi_{loc}(\bar{v}; \cdot) + q\| \cdot \|_V^2, \tag{10}$$

with some $q \geqslant 0$.

For a local model satisfying Assumption 2.6, we can show the following result with respect to elements of the Clarke generalized gradients of φ at $\bar{v} \in V$ and the local model $\varphi_{loc}(\bar{v}; \cdot)$ at 0_V:

Lemma 2.7: *Let $\varphi \in C^1_{\text{abs}}(V)$. Furthermore, suppose that Assumption* 2.6 *holds. Then, one has for $\bar{v} \in V$ that*

$$\partial_C\varphi_{loc}(\bar{v}; 0) \subset \partial_C\varphi(\bar{v}) \tag{11}$$

Proof: Let $\xi \in \partial_C\varphi_{loc}(\bar{v}; 0)$ be arbitrary but fixed. We have to show that $\xi \in \partial_C\varphi(\bar{v})$ holds. Since ξ is an element of the Clarke generalized gradient of $\varphi_{loc}(\bar{v}; \cdot)$ at 0_V, the inequality

$$\varphi_{loc}^C(\bar{v}; 0)(h) \geqslant \xi(h) \quad \forall\, h \in V$$

is valid. Now assume that $\xi \notin \partial_C\varphi(\bar{v})$. Then, there exists a $\bar{h} \in V$ such that

$$\varphi^C(\bar{v}; \bar{h}) < \xi(\bar{h}),$$

where we can assume without loss of generality that $\|\bar{h}\|_V = 1$ due to the properties of the Clarke derivative. One obtains from the definition of the Clarke derivative and the properties of the local model that

$$\begin{aligned}\xi(\bar{h}) > \varphi^C(\bar{v}, h) &= \limsup_{\substack{v\to\bar{v}\\ \lambda\to 0_+}} \frac{1}{\lambda}\left(\varphi(v+\lambda\bar{h}) - \varphi(v)\right)\\ &= \limsup_{\substack{v\to\bar{v}\\ \lambda\to 0_+}} \frac{1}{\lambda}\left(\varphi_{loc}(v;\lambda\bar{h}) - \varphi_{loc}(v;0) + o\left(\|\lambda\bar{h}\|_V\right)\right)\\ &= \limsup_{\substack{v\to\bar{v}\\ \lambda\to 0_+}} \frac{1}{\lambda}\left(\varphi_{loc}(v;\lambda\bar{h}) - \varphi_{loc}(v;0) + o(\lambda)\right)\\ &= \limsup_{\substack{v\to\bar{v}\\ \lambda\to 0_+}} \frac{1}{\lambda}\left(\varphi_{loc}(v;\lambda\bar{h}) - \varphi_{loc}(v;0)\right) = \varphi^C_{loc}(\bar{v};0)(h) \geqslant \xi(\bar{h})\end{aligned}$$

and therefore a contradiction. Hence, $\xi \in \partial_C\varphi(\bar{v})$ must hold proving the assertion. ■

Note that Equation (11) may be a proper subset. The approximation quality stated in Assumption 2.6 also suffices to prove the following properties:

Proposition 2.8: *Suppose for $\varphi \in C^1_{\text{abs}}(V)$ and $v_* \in V$ that Assumption 2.6 holds for the local model $\varphi_{loc}(v_*, \cdot)$ in a neighborhood of v_*. Then one has:*

(1) *If the quadratic model $\varphi_Q(v_*;\cdot)$ is Clarke stationary at $\Delta v = 0$ for one $q \geqslant 0$, then φ is Clarke stationary at v_*.*
(2) *If φ is first-order minimal at v_*, then the quadratic model $\varphi_Q(v_*;\cdot)$ is first-order minimal at the argument $\Delta v = 0$ for all $q \in \mathbb{R}, q \geqslant 0$.*
(3) *If the quadratic model is first-order minimal at $\Delta v = 0$ for one $q \geqslant 0$, then φ is first-order minimal at v_*.*

Proof: To prove the first assertion, we define $\psi : V \to \mathbb{R}$ as $\psi(v) := \frac{q}{2}\|v\|^2_V$ which is a twice Fréchet differentiable function with the unique minimizer at $v = 0$ and $\partial_C\psi(0) = \{0\} \in V^*$. Then, the quadratic model in Equation (10) is given by $\varphi_{loc}(v^*;\Delta v) + \psi(\Delta v)$.

(1): If $\varphi_Q(v^*;\cdot)$ is Clarke stationary in $\Delta v = 0$, it implies that

$$\begin{aligned}0 \in \partial_C\varphi_Q(v^*;0) = \partial_C(\varphi_{loc}(v^*;0) + \psi(0)) &\subseteq \partial_C\varphi_{loc}(v^*;0) + \partial_C\psi(0)\\ &= \partial_C\varphi_{loc}(v^*;0) \subseteq \partial_C\varphi(v^*;0)\end{aligned}$$

using the inclusion of the Clarke generalized gradients shown in Lemma 2.7. Hence, φ is Clarke stationary in v^*.

To prove the first-order minimality statements 2. and 3. we consider the directional derivative defined in Equation (4). One has for all $h \in V$ and $q \geqslant 0$ that

$$\begin{aligned}\varphi'(v_*;h) &= \lim_{t\to 0_+} \frac{1}{t}(\varphi(v_*+th) - \varphi(v_*))\\ &= \lim_{t\to 0_+} \frac{1}{t}\left(\varphi_{loc}(v_*;th) - \varphi_{loc}(v_*;0) + o\left(\|th\|_V\right)\right)\\ &= \lim_{t\to 0_+} \frac{1}{t}\left(\varphi_{loc}(v_*;th) + q\|th\|^2_V - \varphi_{loc}(v_*;0) + o(t)\right)\\ &= \lim_{t\to 0_+} \frac{1}{t}\left(\varphi_Q(v_*;0+th) - \varphi_Q(v_*;0)\right) = \varphi'_Q(v_*;0)(h)\end{aligned}$$

yielding immediately the assertions (2) and (3). Here, $\varphi'_Q(v_*; 0)(h)$ denotes the directional derivative of $\varphi_Q(v_*; \cdot)$ at 0 in direction h. ∎

So far, we do not restrict the local model any further. However, the second order approximation given by Equation (9) justifies the term quadratic model in the last lemma.

3. The quadratic overestimation method

Assume for a given $v_0 \in V$ and a bounded subset $W \subset V$ that a local model is available for all $v \in W$. To update the factor in front of the quadratic penalty term appropriately we define the following function:

$$\hat{q}(v, \Delta v) \equiv \frac{|\varphi(v + \Delta v) - \varphi_{loc}(v; \Delta v)|}{\|\Delta v\|_V^2} \tag{12}$$

for all $v \in V, \Delta v \in V \setminus 0_V$. Before we devote ourselves to the quadratic overestimation method and the corresponding algorithm, we specify further necessary assumptions for the considered setting.

Assumption 3.1 (Considered Setting): Suppose $\varphi \in C^1_{\text{abs}}(V)$ has for a given $v_0 \in V$ a bounded level set

$$\mathcal{N}_0 \equiv \{v \in V : \varphi(v) \leqslant \varphi(v_0)\}.$$

and that $\varphi(\cdot)$ is bounded from below on $\mathcal{N}_0$.

Furthermore, let $\varphi_{loc}(v, \cdot)$ be a local model such that Assumption 2.6 holds. Assume that there exists a monotonic mapping $\bar{q} : [0, \infty) \to [0, \infty)$ such that for all $v \in \mathcal{N}_0$ and $\Delta v \in V$ with $v + \Delta v \in \mathcal{N}_0$

$$\hat{q}(v, \Delta v) \leqslant \bar{q}(\|\Delta v\|_V) \tag{13}$$

is valid. Furthermore, assume that the quadratic model $\varphi_Q(\bar{v}, \cdot)$ attains a minimum.

In the following chapter we will observe that for the problem class presented in Example 2.3, the resulting quadratic model actually has a minimal point. If the local model is such that Assumptions 2.6 and 3.1 hold then Proposition 2.8 yields the relationship between the minimizer of $\varphi_Q(v_k, \cdot)$ and φ.

The proposed SALMIN approach is stated in Algorithm 1. It builds substantially on the local model $\varphi_{loc}(v; \cdot)$.

To prepare the convergence analysis of the generated sequence, we discuss some intermediate results. If the local model is such that Assumption 2.6 and 3.1 hold then the arg min computation step of the algorithm for the quadratic model is well defined. Whenever the step is not successful in that $\varphi(v_k + \Delta v_k) \geqslant \varphi(v_k)$ and $v_{k+1} = v_k$, the new penalty factor q^{k+1} must be bigger than the current value q^k yielding a descent direction in finitely many steps. Hence, for the convergence analysis below, we can consider a subsequence of iterates, where we always have descent in the function value. All remaining iterates can be grouped together such that they form one update of the penalty factor in front of the quadratic term. For simplicity, we will denote the subsequence of iterates with strictly decreasing function values again with $\{v_k\}_{k\in\mathbb{N}}$.

For the sequence $\{q^k\}_{k\in\mathbb{N}}$, one obtains the following result:

Proposition 3.2: *Suppose $\varphi \in C^1_{\text{abs}}(V)$ satisfies Assumption 3.1. Then, the values $\{q^k\}_{k\in\mathbb{N}}$ generated by Algorithm 1 converge to a positive value $q^* \in (0, \infty)$.*

Algorithm 1 SALMIN

```
Require: Let $v_0 \in V$ be such that $\varphi(\cdot)$ is bounded on the bounded level set $\mathcal{N}_0$, $q^0 > 0, \tau > 0$.
  for $k = 0, 1, 2, \ldots$ do
      Compute
          $\Delta v_k = \arg\min_{\Delta v \in V} \varphi_{loc}(v_k; \Delta v) + \frac{1}{2}(1+\tau)q^k \|\Delta v\|_V^2$
      if $\Delta v_k = 0$ then
          STOP
      end if
      if $\varphi(v_k + \Delta v_k) < \varphi(v_k)$ then
          $v_{k+1} = v_k + \Delta v_k$
          Compute $q^{k+1} = \max\{q^k, \hat{q}(v_k, \Delta v_k)\}$
      else
          $v_{k+1} = v_k$
          Compute $q^{k+1} = \max\{(1+\tau)q^k, \hat{q}(v_k, \Delta v_k)\}$
      end if
  end for
```

Proof: Algorithm 1 ensures that all iterates v_k stay in the bounded level set $\mathcal{N}_0$. Hence, it follows from Equation (13) that there exists an upper bound $\check{q}$ with

$$\hat{q}(v_k, \Delta v_k) = \frac{|\varphi(v_k + \Delta v_k) - \varphi_{loc}(v_k; \Delta v_k)|}{\|\Delta v_k\|_V^2} \leqslant \bar{q}(\|\Delta v_k\|_V) \leqslant \check{q}.$$

Therefore, the sequence $\{q^k\}_{k\in\mathbb{N}}$ is increasing and bounded. Combining this with the fact that $q^0 > 0$ yields the assertion with $q^* \leqslant \check{q}$. ■

In finite dimensions, the existence of the monotone function $\bar{q}(\cdot)$ follows directly from the boundedness of the level set and the approximation property Assumption 2.6. However, this is not the case in the infinite dimensional case. Therefore, the existence of this function $\bar{q}(\cdot)$ has to be assumed. Obviously, the function $\bar{q}(\cdot)$ is usually not known. The quantities q^k in Algorithm 1 yield an approximation of $\bar{q}(\cdot)$ for the specific level set $\mathcal{N}_0$.

Now, everything is prepared to prove the main results of this paper:

Theorem 3.3: *Suppose $\varphi \in C^1_{\text{abs}}(V)$ satisfies Assumption* 3.1. *Then a subsequence of the sequence $\{v_k\}_{k\in\mathbb{N}}$ generated by Algorithm* 1 *converges weakly to an element $v_* \in V$ and the sequence $\{\Delta v_k\}_{k\in\mathbb{N}}$ converges strongly to 0_V in V.*

Proof: Algorithm 1 ensures that all iterates stay in the bounded level set $\mathcal{N}_0$. Furthermore, V is a reflexive Banach space. This ensures that a subsequence of $\{v_k\}_{k\in\mathbb{N}}$ converges weakly to a $v_* \in V$ proving the first assertion.

Second, we have to show that Δv_k converges strongly to 0 in V. For a given iterate v_k, the step Δv_k is generated by solving the overestimated quadratic problem

$$\Delta v_k = \arg\min_{\Delta v} \left\{ \varphi_{loc}(v_k; \Delta v) + \frac{1}{2}(1+\tau)q^k \|\Delta v\|_V^2 \right\}. \tag{14}$$

First assume that $\Delta v_\kappa = 0$ holds for one $\kappa \in \mathbb{N}$. Then, Algorithm 1 stops with $v_\kappa = v_{\kappa-1}$ and $\Delta v_\kappa = 0$ and the assertion is proven. Now assume that $\Delta v_k \neq 0$ for all $k \in \mathbb{N}$. Since Δv_k is defined by Equation (14), one has

$$\varphi_{loc}(v_k;\Delta v_k)+\tfrac{1}{2}(1+\tau)q^k\|\Delta v_k\|^2<\varphi_{loc}(v_k;0)=\varphi(v_k). \tag{15}$$

Algorithm 1 ensures that Δv_k is indeed a descent direction for $\varphi(\cdot)$ at the current iterate v_k. Due to the definition of $\hat{q}(v_k;\Delta v_k)$, one has

$$\varphi(v_k+\Delta v_k)-\varphi_{loc}(v_k,\Delta v_k)\leqslant\tfrac{1}{2}\hat{q}(v_k;\Delta v_k)\|\Delta v_k\|_V^2.$$

Combining this with Equation (15) yields for the descent directions Δv_k that

$$\begin{aligned}\varphi(v_k+\Delta v_k)&-\varphi(v_k)\\&=\varphi(v_k+\Delta v_k)-\varphi_{loc}(v_k,\Delta v_k)+\varphi_{loc}(v_k,\Delta v_k)-\varphi_{loc}(v_k)\\&\leqslant\tfrac{1}{2}\left[q^{k+1}-(1+\tau)q^k\right]\|\Delta v_k\|_V^2\end{aligned}\tag{16}$$

holds for all $k\in\mathbb{N}$. Here, we used $\hat{q}(v;\Delta v_k)\leqslant q^{k+1}$ which holds due to the update rule for q^{k+1} given in Algorithm 1. The fact that the sequence $\{q^k\}_{k\in\mathbb{N}}$ converges from below to q^* as shown in Proposition 3.2 implies

$$\varphi(v_k+\Delta v_k)-\varphi(v_k)\leqslant\tfrac{1}{2}\left[q^*-(1+\tau)q^k\right]\|\Delta v_k\|_V^2.$$

Exploiting once more that $\{q^k\}_{k\in\mathbb{N}}$ converges from below to q^*, it follows that for each $\tau>\epsilon>0$ there exists $\bar{k}\in\mathbb{N}$ such that for all $k\geqslant\bar{k}$ the inequality $0\leqslant q^*-q^k<\epsilon$ and therefore also $q^*-(1+\tau)q^k\leqslant c<0$ holds for a constant $c<0$. Since the objective function φ is bounded below on $\mathcal{N}_0$, infinitely many significant descent steps cannot be performed and thus $\varphi(v_k+\Delta v_k)-\varphi(v_k)$ has to converge to 0 as k goes towards infinity. As a consequence, the right hand side of the inequality (16) has to converge to 0 as well. This implies that the sequence $\{\Delta v_k\}_{k\in\mathbb{N}}$ converges strongly to 0. ■

The following theorem shows that Algorithm 1 generates a weakly convergent and bounded sequence in V with a strongly convergent subsequence in $\hat{V}$. Furthermore, strong convergence to a first-order minimal point in V can be proven, if $\Delta v_k=0$ holds for some $k>0$.

Theorem 3.4: *Suppose $\varphi\in C^1_{\text{abs}}(V)$ satisfies Assumption* 3.1. *Whenever there exists a Banach space $\hat{V}$ such that V is compactly embedded in $\hat{V}$, then a subsequence of the sequence $\{v_k\}_{k\in\mathbb{N}}$ generated by Algorithm* 1 *converges strongly to an element v_* in $\hat{V}$.*

Furthermore, if $\Delta v_k=0$ holds for some $k>0$, the cluster point v_* in $\hat{V}$ is first-order minimal, i.e.

$$0\leqslant\varphi'(v_*,h)\quad\textit{for all } h\in\hat{V}.$$

Proof: Theorem 3.3 ensures that a subsequence of $\{v_k\}_{k\in\mathbb{N}}$ converges weakly to an element $v_*\in V$. Taking into account that V is compactly embedded in $\hat{V}$, the weaker norm on $\hat{V}$ yields strong convergence of this subsequence to v_* in $\hat{V}$.

One obtains from Proposition 2.8 that an iterate v_κ is first-order minimal if $\Delta v_\kappa=0$ holds. Then, Algorithm 1 stops with the iterate $v_\kappa=v_{\kappa-1}$, which is first-order minimal and the assertion is proven. In this case one has immediately also strong convergence and uniqueness of the cluster point in V. Due to the compact embedding the same holds in $\hat{V}$. ■

In Algorithm 1 and in the corresponding proof of convergence only the update formula

$$q^{k+1} = \max\{q^k, \hat{q}(v_k, \Delta v_k)\}$$

and therefore a monotone increasing q^{k+1} is considered. Similar to the finite dimensional situation analyzed in [1, Theorem 4], one could also use the more general updating strategy

$$q^{k+1} = \max\{\hat{q}^{k+1}, \mu\, q^k + (1-\mu)\, \hat{q}^{k+1}, q^{lb}\}$$

with $\mu \in [0,1]$ and some fixed lower bound $q^{lb} \geqslant 0$. However, this approach complicates the convergence analysis considerably and is the subject of further research.

Additionally to the previous convergence results we will prove now that Algorithm 1 generates under certain conditions Clarke stationary and even first-order minimal solutions in the considered function space V. To this end, we first state and prove some useful results concerning the Clarke subdifferential defined in Definition 2.4.

Lemma 3.5: *Suppose $\varphi \in C^1_{\text{abs}}(V)$ and $v_* \in V$. If φ is convex and Clarke stationary at v_*, then v_* is already a first-order minimal point for φ.*

Proof: If φ is convex and Clarke stationary at v_*, then the Clarke derivative and the directional derivative of φ at v_* coincide, see e.g. [24, Theorem 3.42]. The Clarke subdifferential of φ at v_* is then given by

$$\partial\varphi(v_*) = \{\xi \in V^* \mid \varphi'(v_*; h) \geqslant \xi(h)\ \forall\, h \in V\}.$$

The first-order minimality follows directly from the requirement that 0_{V^*} has to be an element of the Clarke subdifferential. ■

Proposition 3.6: *Suppose $\varphi \in C^1_{\text{abs}}(V)$. Assume, there exists a sequence $\{w_k\}_{k\in\mathbb{N}} \subset V$ with $w_k \to w_* \in V$ as well as a sequence of functionals $\{\xi_k\}_{k\in\mathbb{N}} \subset V^*$ with a weak*-cluster point $\xi_* \in V^*$. In addition, assume that $\xi_k \in \partial_C\varphi(w_k)\ \forall k \in \mathbb{N}$. Then the limit ξ_* is an element of the Clarke subdifferential $\partial_C\varphi(w_*)$ of φ at the limit point w_*.*

Proof: For all elements $h \in V$ one has

$$\begin{aligned} \xi_*(h) &\leqslant \limsup_{k\to\infty} \xi_k(h) \\ &\leqslant \limsup_{k\to\infty} \varphi^C(w_k, h) \\ &\leqslant \varphi^C(w_*, h), \end{aligned}$$

where the last inequality follows from the fact that the Clarke derivative $\varphi^C(\cdot,\cdot)$ is upper semi-continuous, see e.g. [27]. Thus, one obtains: $\xi_* \in \partial_C\varphi(w_*)$. ■

It is worth mentioning that in the setting considered throughout this paper, weak and weak* convergence coincide since the underlying function spaces are reflexive Banach spaces. The above introduced property of the Clarke gradient and corresponding subdifferential, allows us to formulate the following theorem.

Theorem 3.7: *Let $\varphi \in C^1_{\text{abs}}(V)$. Suppose $\varphi \in C^1_{\text{abs}}(V)$ satisfies Assumption 3.1. Let $\{v_k\}_{k\in\mathbb{N}}$ be the sequence of iterates generated by Algorithm 1 and v_* be a weak cluster point of this sequence (which exists by Theorem 3.3). Further, assume that $\varphi_{loc}(v; \cdot)$ is convex for all $v \in \mathcal{N}_0$. Then φ is first-order minimal at v_*.*

Proof: Algorithm 1 requires the computation of the minimizer Δv_k of the quadratic model. For this purpose we will make use of the fact that the subdifferential in the sense of convex analysis of some Lipschitz continuous, proper and convex operator and the Clarke generalized gradient coincide, see e.g. [28]. This applies in particular to $\varphi_Q(v_k;\cdot)$ for a given $v_k \in \mathcal{N}_0$, which is convex due to the assumed convexity of $\varphi_{loc}(v_k;\cdot)$. Therefore, as a necessary optimality condition 0_{V^*} is an element of the subdifferential $\partial_C\varphi_Q(v_k;\Delta v_k)$ for all $k \in \mathbb{N}$ because each increment Δv_k in Algorithm 1 is a minimizer of the quadratic model. In this case in particular

$$0_{V^*} \in \partial_C\varphi_Q(v_*;\Delta v)$$

applies. Due to requirement of φ satisfying Assumption 3.1, Assumption 2.6 also applies and thus Theorem 3.3 provides that the sequence $\{\Delta v_k\}_{k\in\mathbb{N}}$ converges strongly to 0_V. Consequently, Proposition 3.6 applied to $\varphi_Q(v_*;\cdot)$ and $w_k = \Delta v_k, \xi_k \equiv 0$ together with the definition of the local and quadratic model yields

$$0_{V^*} \in \partial_C\varphi_Q(v_*;0) = \partial_C\varphi_{loc}(v_*;0) + \{0_V\} = \partial_C\varphi_{loc}(v_*;0).$$

Thus, $\varphi_{loc}(v_*;\cdot)$ is Clarke stationary at 0_{V^*}. By Lemma 3.5 the local model $\varphi_{loc}(v_*;\cdot)$ is already first-order minimal at 0_V due to the assumed convexity. With Proposition 2.8 we can conclude that in the considered case φ is first-order minimal at v_*. ■

Let us emphasize that the convexity assumption is by no means an excessive requirement since there is a large class of model problems for which this requirement is met, as discussed later.

It is also important to note that in proving the main convergence results and the quality of the minimizer in the previous theorems we essentially and only use the approximation properties of the local model. Thus, these results are independent of the way the local model is generated and characterized in detail.

4. Generating a suitable local model

After the convergence analysis for the quadratic overestimation method in the last section, we now present one possible approach to generate a suitable local model that fulfills the approximation requirements of Assumption 2.6.

Following the idea in the finite dimensional setting, we assume that the nonsmooth function $\varphi : V \to \mathbb{R}$ is an element of the considered function class C^1_{abs} and can be described as a composition of elemental operators that are either continuously Fréchet differentiable or the absolute value operator. Subsequently, consecutive continuously Fréchet differentiable elemental operators can be conceptually combined to obtain a representation, where all evaluations of the absolute value function can be clearly identified and exploited, see Table 1.

Under suitable conditions some of the elemental functions ψ_i, $i = 1,\dots,s$, may be linear differential operators, integral operators, and solution operators. This is also illustrated by the example given below. As hereinafter shown, the proposed type of reformulation proves to be extremely useful for creating a suitable algorithm for the considered class of nonsmooth problems.

Table 1. Structured Evaluation of $\varphi(v)$.

```
$v_0 = v$
for $i = 1,\dots,s$ do
    $z_i = \psi_i((v_j)_{j<i})$
    $\sigma_i = \mathrm{sign}(z_i)$
    $v_i = \sigma_i z_i = \mathrm{abs}(z_i)$
end for
$w = \psi_{s+1}(v_j)_{j<s+1} = \varphi(v)$
```

In the finite dimensional case $V = \mathbb{R}^n$, one has $z_i \in \mathbb{R}$ and therefore $\sigma_i \in \{-1, 0, 1\}$. For the function space scenario considered here, it follows that $z_i \in V$ and the functions σ_i are also Nemytskii operators defined by

$$\sigma_i : V \to V, \quad \sigma_i(z_i)(x) \cdot z_i(x) = \operatorname{sign}(z_i(x)) \cdot z_i(x) \quad \text{for almost all } x \in \Omega$$

as a function of z_i. This choice ensures that $v_i = \sigma_i z_i = \operatorname{abs}(z_i) \in V$ holds. Furthermore, it follows from the representation in Table 1 that φ is locally Lipschitz continuous. Hence, φ is also continuous due to the assumed smoothness of ψ_i, $i = 1, \dots, s$, [24, Theorem 3.15] and [29, Cha. 1].

Example 4.1: To exemplify the structured evaluation we will again consider the optimization problem as introduced in Example 2.3. The target function φ in Equation (7) can be written as structured evaluation with $s = 2$ using

$$\begin{aligned}
v_0 &= u \\
z_1 &= \psi_1(v_0) \equiv v_0, \quad \sigma_1 = \operatorname{sign}(z_1), \quad v_1 = \operatorname{sign}(z_1) z_1 \\
w &= \psi_2(v_0, v_1) \equiv \frac{1}{2}\|S(v_0) - y_d\|_{L^2}^2 + \frac{\alpha}{2}\|v_0\|_{L^2}^2 + \beta \int_\Omega v_1 \, \mathrm{d}x.
\end{aligned}$$

In this case, ψ_1 is a Lipschitz continuous operator mapping the Banach space $V = L^2(\Omega)$ into $L^2(\Omega)$ and ψ_2 is a Lipschitz continuous operator from the Banach space $L^2(\Omega) \times L^2(\Omega)$ into $\mathbb{R}$.

For the class of nonsmooth operators considered here, we can extend the propagation of derivative information in a suitable way to cover also the absolute value function. For given elements $v, u, \Delta v, \Delta u \in V$ and a continuously Fréchet differentiable ψ, we may use the linearizations

$$\Delta w = \Delta v \pm \Delta u \quad \text{for } w = v \pm u, \tag{17}$$

$$\Delta w = \psi'(v)(\Delta v) \quad \text{for } w = \psi(v) \neq \operatorname{abs}(v), \tag{18}$$

where $\psi'(v)$ denotes the Fréchet derivative of ψ. Here, we face a difference to the finite dimensional case since we do not have to introduce a linearization for a multiplication as this operation is not defined for two elements of the Banach space V. In the case of Banach algebras such a multiplication is defined but we will not consider this case here. For linear operators A, the linearizations are simply given by

$$\Delta w = A\,\Delta v \quad \text{for } w = A\,v. \tag{19}$$

If no absolute value evaluation occurs, the operator $w = \varphi(v)$ is indeed Fréchet differentiable and we obtain the relation

$$\Delta w = \varphi'(v)(\Delta v) \in \mathbb{R}$$

where $\varphi'(v) : V \to \mathbb{R}$ is the Fréchet derivative of φ. Thus we observe the fact that Fréchet differentiation is equivalent to linearizing all elemental operators. Now the question arises which linearization to take for the absolute value operator. Our method of choice is the so-called abs-linearization given by

$$\Delta w = \operatorname{abs}(v + \Delta v) - w \quad \text{for } w = \operatorname{abs}(v). \tag{20}$$

As can be seen, the linearized values Δw depend on both the argument v itself and the direction Δv. If required, we will denote this dependency by $\Delta w(v; \Delta v)$. However, most of the time we will drop these arguments v and Δv for notational simplicity. Similarly, the dependence of the intermediates v_i occurring during the evaluation of φ as described in Table 1 on the argument v is denoted by $v_i(v)$. The local model is constructed in the following way:

Definition 4.2 (Abs-Linearization): Suppose $\varphi : V \to \mathbb{R}$ is an element of the operator class $C^1_{\text{abs}}(V)$ as defined in Definition 2.2. For a fixed argument $v \in V$ and $w = \varphi(v)$ the abs-linearization $\Delta w(v; \cdot) : V \to \mathbb{R}$ based on the linearizations Equations (17)–(20) is constructed in the following way:

```
$v_0 = v, \Delta v_0 = \Delta v$
for $i = 1, \ldots, s$ do
    $z_i \;\;= \psi_i((v_j)_{j<i})$
    $\Delta z_i = \Delta\psi_i((v_j)_{j<i})((\Delta v_j)_{j<i})$
    $\sigma_i = \text{sign}(z_i)$
    $v_i = \sigma_i z_i = \text{abs}(z_i)$
    $\Delta v_i = \text{abs}(z_i + \Delta z_i) - \text{abs}(z_i)$
end for
$w = \psi_{s+1}(v_j)_{j<s+1} = \varphi(v), \Delta w = \Delta\psi_{s+1}((v_j)_{j<s+1})((\Delta v_j)_{j<s+1}).$
```

Once more, the σ_i are Nemytskii operators as defined already in Section 2.

Next, we will show below that the abs-linearization provides a local model that has the required approximation properties:

Proposition 4.3 (Approximation Properties of the Abs-linearization): *Suppose $\varphi \in C^1_{\text{abs}}(V)$. Then there exists a constant $\bar{q} > 0$, such that for all pairs $\bar{v}, v \in W \subset V$ with W some closed convex subset, one has for the local model defined by*

$$\varphi_{loc}(\bar{v}; \cdot) : V \to \mathbb{R}, \quad \varphi_{loc}(\bar{v}; \Delta v) = \varphi(\bar{v}) + \Delta\varphi(\bar{v}; \Delta v)$$

that

$$\varphi(\bar{v}) = \varphi_{loc}(\bar{v}; 0) \quad \textit{and} \quad |\varphi(v) - \varphi_{loc}(\bar{v}; \bar{v} - v)| \leqslant \bar{q}\|\bar{v} - v\|_V^2.$$

Proof: The first equality follows directly from the definition of the local model. The second equality is proven by induction on i. That is, we show that for all intermediates

$$v_i(v + \Delta v) - v_i(v) = \Delta v_i(v; \Delta v) + \mathcal{O}(\|\Delta v\|_V^2)$$

for $\Delta v = v - v$ in a neighborhood of v. For the first intermediate, i.e. v_0, this holds trivially since we set $\Delta v_0 = \Delta v$. For the arithmetic operations $+$ and $-$ as well as the continuously Fréchet differentiable elemental operators, the Taylor series theory in Banach spaces, see, e.g. [29, Section 4.5], ensures that the linearizations Equations (17) and (18) yield for the resulting Δv_i the asserted approximation property. For linear, continuous operators the approximation property holds trivially. Therefore, we only have to consider the case $w = \text{abs}(u)$. Equation (20) yields

$$\begin{aligned} & w(v) + \Delta w(v; \Delta v) - w(v + \Delta v) \\ & \quad = \text{abs}(u(v)) + [\text{abs}(u(v) + \Delta u(v; \Delta v)) - \text{abs}(u(v))] - \text{abs}(u(v + \Delta v)) \\ & \quad = \text{abs}(u(v) + \Delta u(v; \Delta v)) - \text{abs}(u(v + \Delta v)) = \mathcal{O}(\|\Delta v\|_V^2), \end{aligned}$$

where the last relation follows from the induction hypothesis and the Lipschitz continuity of all quantities involved. This yields for $w(v) = \varphi(v) \in \mathbb{R}$ and $w(v + \Delta v) = \varphi(v + \Delta v) \in \mathbb{R}$ that

$$w(v + \Delta v) - w(v) - \Delta w(v; \Delta v) = \mathcal{O}(\|\Delta v\|_V^2)$$

proving the assertion. ■

The next example illustrates that there is a whole class of PDE-constrained optimization problems that fulfills the requirements of the local model required in the previous section. Furthermore, one also has that the associated quadratic model always attains a minimum, which ensures that the arg min computation step in Algorithm 1 is well defined.

Example 4.4: Consider once more the optimization problem as introduced in Example 2.3

$$\min_{(y,u)\in H_0^1(\Omega)\times L^2(\Omega)} \quad \frac{1}{2}\|y-y_d\|_{L^2}^2+\frac{\alpha}{2}\|u\|_{L^2}^2+\beta\|u\|_{L^1}$$
$$\text{s.t.} \quad Ay+l(y)=u+f\text{in}\Omega,$$

with the operator $A: H_0^1(\Omega)\to H^{-1}(\Omega)$ representing a linear elliptic differential operator of second order. The operator $l: H_0^1(\Omega)\to L^2(\Omega)$ is such that there exists a continuously Fréchet differentiable solution operator $S: L^2(\Omega)\to H_0^1(\Omega)$. Defining the Fréchet differentiable operator

$$\varphi_1(u)=\frac{1}{2}\|S(u)-y_d\|_{L^2}^2+\frac{\alpha}{2}\|u\|_{L^2}^2,$$

and consequently the target function by

$$\varphi(u)=\varphi_1(u)+\beta\|u\|_{L^1}$$

and substituting the known quantities, the local model according to Definition 4.2 is given by

$$\begin{aligned}\varphi_{loc}(u;\Delta u)&=\varphi(u)+\Delta\varphi(u;\Delta u)\\&=\varphi(u)+\varphi_1'(u)(\Delta u)+\beta\int_\Omega \text{abs}(u+\Delta u)-\text{abs}(u)\,\mathrm{d}x\\&=\varphi_1(u)+\varphi_1'(u)(\Delta u)+\beta\|u+\Delta u\|_{L^1}.\end{aligned}$$

Note that the last relation follows from the definition of φ and φ_1 eliminating the term $\beta\|u\|_{L^1}$ in the last line.

As can be seen, the first term is constant with respect to Δu, the second term is linear in Δu and the third term convex in Δu. Therefore, the local model is even weakly lower semi-continuous in Δu. Once more a simple example for this scenario is given by $l(y)\equiv 0$ as in [20].

Due to the structure of the resulting reduced formulation for this optimization problem, Assumptions 2.6 and 3.1 are satisfied for the target function φ and the local model φ_{loc}. Hence, by Theorem 3.3, Algorithm 1 applied to this optimization problem generates a weakly convergent subsequence $v_k\rightharpoonup v_*\in L^2$, which converges even strongly in H^{-1} by Theorem 3.4, and a strongly convergent sequence $\Delta v_k\to 0_{L^2}$. In the case of S being a continuously Fréchet differentiable operator, $\varphi_{loc}(v_k;\cdot)$ is convex and hence, v_* is a first-order minimal point by Theorem 3.7.

5. Conclusion and outlook

We presented a new quadratic overestimation approach based on a local model with appropriate properties for the solution of nonsmooth optimization problems in function spaces. We proved convergence to first-order minimal points and hence a stronger stationarity concept than Clarke stationarity. Subsequently, we used the technique of abs-linearization to construct a local model that has the required approximation properties of second order. Finally, we illustrated the results on a class of model problems that fits into the considered setting, a PDE-constrained problem with an L^1 penalty term. Throughout the paper we assume $V=L^p(\Omega)$ with $1<p<\infty$. The presented theory can be extended easily to more general reflexive Banach spaces V where the absolute value function is Lipschitz continuous. It should be noted that the finite dimensional case represents a special case of the here presented setting, implied by $V=\hat{V}=\mathbb{R}^n$.

Future work will be dedicated to an extension of the theory for more general solution operators such that for example also nonsmooth PDE constraints can be handled.

Acknowledgements

We would like to thank Johannes Lankeit and Gerd Wachsmuth for the helpful comments and useful suggestions that greatly improved the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Furthermore, we acknowledge funding by the DFG priority program SPP 1962 within the project "Shape Optimization for Maxwell's Equations Including Hysteresis Effects in the Material Laws (HyLa)" (WA1607/12-1).

References

[1] Fiege S, Walther A, Griewank A. An algorithm for nonsmooth optimization by successive piecewise linearization. Math Program Ser A. 2018. DOI:10.1007/s10107-018-1273-5.

[2] Griewank A. On stable piecewise linearization and generalized algorithmic differentiation. Optim Methods Softw. 2013;28(6):1139–1178.

[3] Griewank A, Walther A. Relaxing kink qualifications and proving convergence rates in piecewise smooth optimization. SIAM J Optim. 2019;29(1):262–289.

[4] Scholtes S. Introduction to piecewise differentiable functions. Berlin: Springer; 2012.

[5] Dodds PG, Dodds TK, de Pagter B, et al. Lipschitz continuity of the absolute value and Riesz projections in symmetric operator spaces. J Funct Anal. 1997;148(1):28–69.

[6] Griewank A. The modification of Newton's method for unconstrained optimization by bounding cubic terms. University of Cambridge; 1981. (Technical report; NA/12).

[7] Griewank A, Fischer J, Bosse T. Cubic overestimation and secant updating for unconstrained optimization of $c^{2,1}$ functions. Optim Methods Softw. 2014;29(5):1075–1089.

[8] Knossalla M. Concepts on generalized ε-subdifferentials for minimizing locally lipschitz continuous functions. J Nonlinear Var Anal. 2017;1:265–279.

[9] Muu LD, Quy NV. Dc-gap function and proximal methods for solving nash-cournot oligopolistic equilibrium models involving concave cost. J Appl Numer Optim. 2019;1:13–24.

[10] Eckstein J, Bertsekas D. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math Program. 1992;55(3(A)):293–318.

[11] Guler O. On the convergence of the proximal point algorithm for convex minimization. SIAM J Control Optim. 1991;29(2):403–419.

[12] Rockafellar RT. Monotone operators and the proximal point algorithm. SIAM J Control Optim. 1976;14:877–898.

[13] Tung NL, Luu DV. Optimality conditions for nonsmooth multiobjective optimization problems with general inequality constraints. Nonlinear Funct Anal. 2018;2018:Article ID 2.

[14] Van Ackooij W, Bello Cruz JY, De Oliveira W. A strongly convergent proximal bundle method for convex minimization in hilbert spaces. Optimization. 2016;65:145–167.

[15] Correa R, Lemaréchal C. Convergence of some algorithms for convex minimization. Math Program. 1993;62:261–275.

[16] Hertlein L, Ulbrich M. An inexact bundle algorithm for nonconvex nondifferentiable functions in hilbert space. Technische Universität München; 2018. (Technical report; SPP1962-084).

[17] Clason C, Nhu VH, Rösch A. Optimal control of a non-smooth quasilinear elliptic equation. Preprint SPP1962-101, 12 2018.

[18] Ulbrich M. Nonsmooth Newton-like methods for variational inequalities and constrained optimization problems in function spaces [Habilitation]. Technische Universität München; 2002.

[19] Casas E, Herzog R, Wachsmuth G. Optimality conditions and error analysis of semilinear elliptic control problems with L^1 cost functional. SIAM J Optim. 2012;22(3):795–820.

[20] Stadler G. Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices. Comput Optim Appl. 2007 Nov;44(2):159.

[21] Christof C, Clason C, Meyer C, et al. Optimal control of a non-smooth semilinear elliptic equation. 2017. (Technical report; SPP1962-020, DFG SPP 1962).

[22] Clarke F. Optimization and nonsmooth analysis. Philadelphia (PA): SIAM; 1990.

[23] Clarke F. A new approach to lagrange multipliers. Math Oper Res. 1976;1:167–174.

[24] Jahn J. Introduction to the theory of nonlinear optimization. Berlin: Springer; 2007.

[25] Absil P-A, Mahony R, Andrews B. Convergence of the iterates of descent methods for analytic cost functions. SIAM J Optim. 2005;16(2):531–547.

[26] Attouch H, Bolte J. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. Math Program Ser B. 2009;116(1–2):5–16.
[27] Clarke F. Generalized gradients of lipschitz functionals. Adv Math (NY). 1981;40:52–67.
[28] Clarke F. Generalized gradients and applications. Trans Am Math Soc. 1975;205:247–262.
[29] Zeidler E. Applied functional analysis: applications to mathematical physics. Berlin: Springer; 1995.