Deutsche Forschungsgemeinschaft
Priority Programme 1962

Nonsmooth Optimization by Successive Abs-Linearisation in Function Spaces

Andrea Walther, Olga Ebel, Andreas Griewank, Stephan Schmidt

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

Preprint Number SPP1962-103

received on February 9, 2019

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/`

Nonsmooth Optimization by Successive Abs-Linearisation in Function Spaces

Andrea Walther[1], Olga Ebel[1],
Andreas Griewank[2], and Stephan Schmidt[3]

February 8, 2019

Abstract

We present and analyse the solution of nonsmooth optimization problems by a quadratic overestimation method in the 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-linearisation. Finally, we discuss a class of PDE-constrained optimisation problems incorporating the L^1-penalty term that fits into the considered class of non-smooth optimization problems.

Keywords: Abs-Linearisation, quadratic overestimation method, first-order minimality

1 Introduction and Motivation

In a finite dimensional setting with $V = \mathbb{R}^n$, the minimization of a piecewise smooth function $\varphi : V \mapsto \mathbb{R}, y = \varphi(x)$, based on successive abs-linearisation has been analysed already extensively in a series of papers, see, e.g., [11, 13, 15]. 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 [13] studies a class of functions $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$ that are piecewise smooth in the sense of Scholtes [22], 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 $i = 1 \ldots s$, where it is assumed throughout that z_j can only influence z_i if $j < i$. Then, the paper [13] proposes

[1]Institut für Mathematik, Universität Paderborn, Paderborn, Germany
[2]School of Mathematical and Computational Sciences, Yachay Tech, Urcuquí, Imbabura, Ecuador
[3]Institut für Mathematik, Universität Würzburg, Würzburg, Germany

a new approach to generate a local so-called abs-linear model $\Delta\varphi(x;.) : \mathbb{R}^n \mapsto \mathbb{R}$ of $\varphi(.)$ using the technique of abs-linearisation. 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 similar good approximation property also for a local model in the infinite dimensional setting using an appropriate extension of the abs-linearisation.

Based on Eq. (1), the following iterative optimization algorithm was proposed in [13]

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

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-linearisation. This method was shown in [13] 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 [11]. 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., [9]. 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.

The SALMIN algorithm given by Eq. (2) can be interpreted as a quadratic overestimation method, where the error between the model and the real function is bounded by a power of the distance, see, e.g., [12, 14]. This approach is also related to proximal point methods as analyzed for the infinite setting for example in [10, 16, 20]. There the original function is still an additive component of the local subproblem. In contrast to the results presented in these papers, in Eq. (2) the local model of the function to be minimized at the current iterate x_k is used instead of the original function. Hence, 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 [2, 8]. An alternative bundle method covering also nonconvex target functions is presented in [17]. 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 and epi-convergence, subsequently we will show that the limit point is 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 Sec. 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 transfer directly to the infinite dimensional setting. However, extensions of these concepts are considered in [7]. In Sec. 3 we extend the proposed quadratic overestimation method given by Eq. (2) to the infinite dimensional case and analyze its convergence behavior. Subsequently, we propose the generation of a local model in Sec. 4, where an approximation property similar to Eq. (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 Sec. 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 our choice of 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

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

as the Nemytskii operator induced by the absolute value function. Such operators are also called superposition operators, see [24]. For better readability we will sometimes omit the local argument x and thus consider abs(.) directly as an operator on the function space. The abs(.) operator can enforce sparsity if included appropriately in the target function, see, e.g., [5, 23]. 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 [6].

In general Banach spaces, it is not clear whether the absolute value operator is Lipschitz continuous, see, e.g., [9]. Therefore, we state the following result for the function spaces considered here:

Corollary 2.1 (Lipschitz continuity of abs(.))**.** *The absolute value operator* $\mathrm{abs} : V \mapsto V, \mathrm{abs}(v) := |v|$, *is Lipschitz continuous and nonexpansive.*

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

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

such that

$$\|\text{abs}(v) - \text{abs}(u)\|_V^p = \int_\Omega ||v(x)| - |u(x)||^p dx$$
$$\leqslant \int_\Omega |v(x) - u(x)|^p dx = \|v - u\|_V^p.$$

Since the Lipschitz constant is 1, one obtains also that the absolute value operator is nonexpansive. □

Following the idea in the finite dimensional setting, we assume that the nonsmooth function $\varphi : V \mapsto \mathbb{R}$ 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 Tab. 1. Under suitable conditions some of the elemental functions ψ_i, $i = 1, \ldots, s$, may be linear differential operators, integral operators, and solution operators. This is also illustrated by the example given below. As shown in the next sections, the proposed type of reformulation proves to be extremely useful for creating a suitable algorithm for the considered class of nonsmooth problems.

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

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

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

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

as a function of z_i. This choice ensures that $v_i = \sigma_i z_i = \text{abs}(z_i) \in V$ holds. Furthermore, it follows from the representation in Tab. 1 that φ is locally Lipschitz continuous. Hence, φ is also continuous due to the assumed smoothness of ψ_i, $i = 1, \ldots, s$, [19, Theo. 3.15] and [25, Cha. 1]. For the convergence result presented later, we also require that φ is continuous in $\hat{V}$. It follows from the kind of admissible nonsmoothness that the directional derivative defined by

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

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 [15], we denote this class by $C^1_{abs}(V)$.

Definition 2.2 (Operator Class $C^1_{abs}(V)$). *For a reflexive Banach space V, the class $C^1_{abs}(V)$ contains all operators $\varphi : V \mapsto \mathbb{R}$ such that φ can be represented by a structured evaluation as given in Tab. 1 with Fréchet differentiable mappings ψ_i, $i = 1, \ldots, s$. That is, $C^1_{abs}(V)$ is the set of all the operators $\varphi : V \mapsto \mathbb{R}$, for which there exist a natural number s, reflexive Banach spaces $V_1, \ldots, V_{s+1}$, $V_1 = V$, $V_{s+1} = \mathbb{R}$, and Fréchet differentiable mappings $\psi_i : V_i \mapsto V_{i+1}$, $1 \leqslant i \leqslant s+1$, with $(v_j)_{j<i} \in V_i$, such that the evaluation of $\varphi(v)$ can be described as a composition of continuously Fréchet differentiable operators and the absolute value operator.*

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

$$\begin{aligned} \min(v,u) &= (v+u-\operatorname{abs}(v-u))/2 \qquad \text{and} \\ \max(v,u) &= (v+u+\operatorname{abs}(v-u))/2\ , \end{aligned} \tag{4}$$

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

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*

$$\begin{aligned} \min_{(y,u)\in H^1_0(\Omega)\times L^2(\Omega)} \quad & \tfrac{1}{2}\|y-y_d\|^2_{L^2} + \tfrac{\alpha}{2}\|u\|^2_{L^2} + \beta\|u\|_{L^1} \\ \text{s.t.} \quad & Ay + l(y) = u + f \quad \text{in } \Omega\ , \end{aligned} \tag{5}$$

where y and u represent the state and the control, respectively, and $\alpha > 0$, $\beta > 0$ are parameters. Furthermore, $A : H^1_0(\Omega) \mapsto H^{-1}(\Omega)$ is a linear elliptic differential operator of second order. Suppose, that the operator l is such that there exists a continuously Fréchet differentiable solution operator $S : L^2(\Omega) \mapsto H^1_0(\Omega)$. As a very simple example one may consider $l(y) \equiv 0$ as in [23]. Then, the reduced problem formulation is given by the optimization problem

$$\min_{u\in L^2(\Omega)} \varphi(u) \quad \textit{with} \quad \varphi(u) = \tfrac{1}{2}\|S(u)-y_d\|^2_{L^2} + \tfrac{\alpha}{2}\|u\|^2_{L^2} + \beta\|u\|_{L^1}\ . \tag{6}$$

The target function $\varphi(u)$ can be written as structured evaluation with $s = 2$ using

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

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 $\mathbb{R}$.

As shown in Sec. 4, the assumed structure of the $C^1_{abs}(V)$ class allows the construction of a suitable local model for the quadratic overestimation method discussed in the next section. In contrast to many other approaches, we aim at first-order minimality that is defined as follows

Definition 2.4 (First-order Minimality). *Suppose, $\varphi \in C^1_{abs}(V)$. The operator φ is called* first-order minimal *at $v_* \in V$ if one has*

$$0 \leqslant \varphi'(v_*, h) \qquad \text{for all} \qquad h \in V \ . \tag{7}$$

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

It is important to note that this property is stronger than the frequently used concept of Clarke stationary. Often, first-order minimality is also called criticality as defined in [1] and [4], 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 2.5 (Approximation Properties). *Suppose, $\varphi \in C^1_{abs}(V)$ and $W \subset V$ is a closed convex subset. For all $\bar{v} \in W$, there exists a Lipschitz continuous local model $\varphi_{loc}(\bar{v}; .) : V \mapsto \mathbb{R}$ with $\varphi_{loc}(\bar{v}; .) \in C^1_{abs}(V)$ and constant $q > 0$ such that for all $v \in W$ one has*

$$\varphi(\bar{v}) = \varphi_{loc}(\bar{v}; 0), \qquad |\varphi(\bar{v}) - \varphi_{loc}(\bar{v}; v - \bar{v})| \leqslant q\|v - \bar{v}\|_V^2 \ . \tag{8}$$

Moreover, for any pair $\bar{v}, v \in W$ and $w \in V$ there exists a constant $\gamma > 0$ such that

$$\frac{|\varphi_{loc}(\bar{v}; w) - \varphi_{loc}(v; w)|}{1 + \|w\|_V} \leqslant \gamma\|\bar{v} - v\|_V \ . \tag{9}$$

For such a local model, we can show the following properties:

Lemma 2.6. *Suppose for $\varphi \in C^1_{abs}(V)$ and $v_* \in V$ that Assump. 2.5 holds for the local model $\varphi_{loc}(v_*, .)$ in a neigbourhood of v_*. Then one has:*

1. *If φ is first-order minimal at v_*, then the quadratic model*

$$\varphi_Q(v_*; .) \equiv \varphi_{loc}(v_*; .) + q\|.\|_V^2 \tag{10}$$

   is first-order minimal at the argument $\Delta v = 0$ for all $q \in \mathbb{R}, q \geqslant 0$.

2. *If the quadratic model given by Eq. (10) is first-order minimal at $\Delta v = 0$ for one $q \geqslant 0$, then φ is first-order minimal at v_*.*

Proof. One has for all $h \in V$ and $q \geqslant 0$ that

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

yielding immediately the assertions 1. and 2. Here, $\varphi'_Q(v_*; 0, h)$ denotes the directional derivative of $\varphi_Q(v_*; .)$ at 0 in direction h. □

So far, we do not restrict the local model any further. However, the second order approximation given by Eq. (8) justifies the term quadratic model in the last lemma. For the convergence proof of the overestimation method, we will exploit the second result of this lemma. For this purpose, we need in addition the following properties:

Lemma 2.7 (Convergence of Minimizers)**.** *Let $\varphi \in C^1_{abs}(V)$, $v_* \in V$ and a sequence $v_k \to v_*$ in V be given. Assume that Assump. 2.5 holds for the local models $\varphi_{loc}(v,.)$ at v in $W = \overline{B}_{\lambda c}(v_*) = \{v \in V \mid \|v - v_*\| \leqslant \lambda c\}$ for $c > 0$ and $\lambda > 1$. Then,*

1. *$\varphi_{loc}(v_k,.)$ epi-converges to $\varphi_{loc}(v_*,.)$ in $B_c(0_V) = \{v \in V \mid \|v\| < c\}$.*
2. *For any $k \geqslant 1$, let w_k denote a minimizer of $\varphi_{loc}(v_k,.)$ if it exists. If the sequence $(w_k)_{k\in\mathbb{N}}$ admits a subsequence converging to some $\bar{w} \in B_c(0_V)$, then $\bar{w}$ belongs to* $\arg\min \varphi_{loc}(v_*;.)$ *such that*

$$\limsup_{k\to\infty} \big(\operatorname*{arg\,min}_{w\in B_c(0_V)} \varphi_{loc}(v_k;w) \big) \subseteq \operatorname*{arg\,min}_{w\in B_c(0_V)} \varphi_{loc}(v_*;w) \ .$$

Proof. To prove the epi-convergence we need to show that for any $\bar{w} \in B_c(0_V)$ there exists a sequence $\tilde{w}_k \to \bar{w}$ such that

$$\limsup_{k\to\infty} \varphi_{loc}(v_k,\tilde{w}_k) \leqslant \varphi_{loc}(v_*,\bar{w}),$$

and for every sequence $w_k \to \bar{w}$ the inequality

$$\liminf_{k\to\infty} \varphi_{loc}(v_k,w_k) \geqslant \varphi_{loc}(v_*,\bar{w})$$

holds, see for instance [3, 18] and [21]. For the given sequence $v_k \to v_*$ in V consider an arbitrary sequence $\tilde{w}_k \to \bar{w}$ in $B_c(0_V)$. For k large enough one has that $v_k + w_k \in \overline{B}_{\lambda c}(v_*)$. With the triangle inequality and the approximation property Eq. (9) it follows that

$$\begin{aligned} &|\varphi_{loc}(v_k;w_k) - \varphi_{loc}(v_*;\bar{w})| \\ &\qquad \leqslant \gamma(\|v_k - v_*\|_V)(1 + \|w_k\|) + |\varphi_{loc}(v_*;w_k) - \varphi_{loc}(v_*;\bar{w})| \ . \end{aligned}$$

Since the last term on the right hand side converges to zero due to the continuity of $\varphi_{loc}(v;.)$, one obtains

$$\lim_{k\to\infty} |\varphi_{loc}(v_k;w_k) - \varphi_{loc}(v_*;\bar{w})| = 0$$

yielding the two inequalities required for the epi-convergence of the function sequence $\varphi_{loc}(v_k;.)$ to the function $\varphi_{loc}(v_*;.)$ and therefore the first assertion. The second assertion follows immediately from [21, Theorem 3.33]. □

Note that in the context of epi-convergence also the set of $\alpha-$approximate minimizers [18] is considered, which is always nonempty for $\alpha > 0$, in contrast to the set of exact minimizers which is obtained for $\alpha = 0$. In the present case however, the set of exact minimizers is usually also nonempty.

3 The Quadratic Overestimation Method

Assume for a given $v_0 \in V$ that the level set $\mathcal{N}_0 \equiv \{v \in V \ : \ \varphi(v) \leqslant \varphi(v_0)\}$ is bounded and that a local model is available for all $v \in \mathcal{N}_0$. To update the factor

Algorithm 1 SALMIN

Require: Let $v_0 \in V$ be such that $\varphi(.)$ is bounded on the bounded level set $\mathcal{N}_0$, $q^0 > 0, \tau > 0$.

for $k = 0, 1, 2, \ldots$ **do**

Compute

$$\Delta v_k = \underset{\Delta v \in V}{\arg\min}\, \varphi_{loc}(v_k; \Delta v) + \tfrac{1}{2}(1+\tau)q^k \|\Delta v\|_V^2$$

if $\varphi(v_k + \Delta v_k) < \varphi(v_k)$ **then**

$v_{k+1} = v_k + \Delta v_k$

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

$k = k + 1$

else

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

end if

end for

in front of the quadratic penalty term appropriately we define the following function:

$$\hat{q}(v, \Delta v) \equiv \frac{|\varphi(v + \Delta v) - \varphi_{loc}(v; \Delta v)|}{\|\Delta v\|_V^2} \tag{11}$$

for all $v, \Delta v \in V$. The proposed SALMIN approach is stated in Algo. 1. It builds substantially on the local model $\varphi_{loc}(v, .)$. No stopping criterion is given such that an infinite sequence of iterates $\{v_k\}_{k \in \mathbb{N}}$ is generated.

To prepare the convergence analysis of the generated sequence, we discuss some intermediate results. If the local model is such that Assump. 2.5 holds then Lemma 2.6 yields the existence of a minimizer Δv_k such that this step of the algorithm 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 finally many steps. Hence, for the convergence analysis below, we can consider a subsequence of iterates, where we 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:

Corollary 3.1. *Suppose $\varphi \in C^1_{abs}(V)$ has for a given $v_0 \in V$ a bounded level set*

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

Let $\varphi_{loc}(v, .)$ be a local model such that Assump. 2.5 holds. Assume that there exists a monotonic mapping $\bar{q} : [0, \infty) \mapsto [0, \infty)$ such that for all $v \in \mathcal{N}_0$ and $\Delta v \in V$ with $v + \Delta v \in \mathcal{N}_0$

$$\hat{q}(v, \Delta v) \leqslant \bar{q}(\|\Delta v\|_V) \tag{12}$$

is valid. Then, the values $\{q^k\}_{k \in \mathbb{N}}$ generated by Algo. 1 converge to a positive value $q^* \in (0, \infty)$.

Proof. Algo. 1 ensures that all iterates v_k stay in the bounded level set $\mathcal{N}_0$. Hence, it follows from Eq. (12) 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 $\bar{q}(.)$ follows directly from the boundedness of the level set and the approximation property Assump. 2.5. However, this is not the case in the infinite dimensional case. Therefore, the existence of this function $\bar{q}(.)$ has to be assumed. Obviously, the function $\bar{q}(.)$ is usually not known. The quantities q^k in Algo. 1 yield an approximation of $\bar{q}(.)$ for the specific level set $\mathcal{N}_0$.

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

Theorem 3.2. *Suppose $\varphi \in C^1_{abs}(V)$ has for a given $v_0 \in V$ a bounded level set $\mathcal{N}_0$ and that $\varphi(.)$ is bounded from below on $\mathcal{N}_0$. Suppose that there is a local model $\varphi_{loc}(v, .)$ for all $v \in \mathcal{N}_0$ fulfilling Assump. 2.5 and that there exists a monotonic mapping $\bar{q} : [0, \infty) \mapsto [0, \infty)$ such that for all $v \in \mathcal{N}_0$ and $\Delta v \in V$ with $v + \Delta v \in \mathcal{N}_0$*

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

is valid. Then a subsequence of the sequence $\{v_k\}_{k\in\mathbb{N}}$ generated by Algo. 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. Algo. 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

$$\Delta v_k = \arg\min_{\Delta v} \left\{\varphi_{loc}(v_k; \Delta v) + \tfrac{1}{2}(1 + \tau)q^k \|\Delta v\|_V^2\right\}. \tag{13}$$

First assume that $\Delta v^\kappa = 0$ holds for one $\kappa \in \mathbb{N}$. Then, Algo. 1 generates the subsequent iterates $v^k = v^\kappa, k \geqslant \kappa \in \mathbb{N}$, such that $\Delta v^k = 0$ for all $k \geqslant \kappa$ 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 Eq. (13), one has

$$\varphi_{loc}(v_k; \Delta v_k) + \tfrac{1}{2}(1 + \tau)q^k \|\Delta v_k\|^2 < \varphi_{loc}(v_k; 0) = \varphi(v_k)\ . \tag{14}$$

Algo. 1 ensures that Δv_k is indeed a descent direction for $\varphi(.)$ 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 Eq. (14) yields for the descent directions Δv_k that

$$\begin{aligned} &\varphi(v_k + \Delta v_k) - \varphi(v_k) \\ &\qquad = \varphi(v_k + \Delta v_k) - \varphi_{loc}(v_k, \Delta v_k) + \varphi_{loc}(v_k, \Delta v_k) - \varphi_{loc}(v_k) \\ &\qquad \leqslant \tfrac{1}{2}\left[q^{k+1} - (1 + \tau)q^k\right]\|\Delta v_k\|_V^2 \end{aligned} \tag{15}$$

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 Algo. 1. The fact that the sequence $\{q^k\}_{k\in\mathbb{N}}$ converges from below to q^* as shown in Cor. 3.1 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 can not 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 (15) has to converge to 0 as well. This implies that the sequence $\{\Delta v_k\}_{k\in\mathbb{N}}$ converges strongly to 0. □

Using the epi-convergence of the local models $\varphi_{loc}(v_k; .)$, we will now show that the cluster point v_* is first-order minimal.

Theorem 3.3. *Suppose $\varphi \in C^1_{abs}(V)$ has for a given $v_0 \in V$ a bounded level set $\mathcal{N}_0$ and that $\varphi(.)$ is bounded from below on $\mathcal{N}_0$. Suppose that there is a local model $\varphi_{loc}(v, .)$ for all $v \in \mathcal{N}_0$ fulfilling Assump. 2.5 and that there exists a monotonic mapping $\bar{q} : [0, \infty) \mapsto [0, \infty)$ such that for all $v \in \mathcal{N}_0$ and $\Delta v \in V$ with $v + \Delta v \in \mathcal{N}_0$*

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

is valid. Whenever there exists a Banach space $\hat{V}$ such that V is compactly embedded in $\hat{V}$ and φ is Lipschitz continuous in $\hat{V}$, then a subsequence of the sequence $\{v_k\}_{k\in\mathbb{N}}$ generated by Algo. 1 converges strongly to an element v_* in $\hat{V}$ that is first-order minimal, i.e.,

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

Furthermore, also all other cluster points of the sequence $\{v_k\}_{k\in\mathbb{N}}$ are first-order minimal.

Proof. Thm. 3.2 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 Prop. 2.6 that an iterate v_κ is first-order minimal if $\Delta v_\kappa = 0$ holds. Then, Algo. 1 generates the subsequent iterates $v_k = v_\kappa, k \geqslant \kappa \in \mathbb{N}$, that are all 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 $\hat{V}$.

Now assume that $\Delta v_k \neq 0$ for all $k \in \mathbb{N}$ and consider the cluster point v_*. For notational simplicity, we denote the strongly converging subsequence also by $\{v_k\}_{k\in\mathbb{N}}$. From Lemma 2.7 we know that $\varphi_{loc}(v_k; .)$ epi-converges to $\varphi_{loc}(v_*; .)$ in $B_c(0_V)$. Due to the assumed continuity of $\varphi(.)$ in $\hat{V}$, the epi-convergence

$$\varphi_Q^k(.) := \varphi_{loc}(v_k; .) + \tfrac{1}{2}(1+\tau)q^k\|.\|_V^2 \xrightarrow{\text{e}} \varphi_Q^*(.) := \varphi_{loc}(v_*; .) + q^*\|.\|_V^2$$

in $B_c(0_{\hat{V}})$ can be shown using similar arguments as in the proof of Lemma 2.7. Therefore, we can apply the second assertion of Lemma 2.7 to $\varphi_Q^k(.)$. That is, since $0 = \Delta v_*$ is a limit point of minimizers of $\varphi_Q^k(.)$, then it is also a minimizer of $\varphi_Q^*(.)$ yielding

$$\Delta v_* = 0 \in \limsup_{k\to\infty} \left(\operatorname*{arg\,min}_{\Delta v} \left\{ \varphi_{loc}(v_k; \Delta v) + \tfrac{1}{2}(1+\tau) q^k \|\Delta v\|_V^2 \right\} \right)$$
$$\subseteq \operatorname*{arg\,min}_{\Delta v} \left(\varphi_{loc}(v_*; \Delta v) - \varphi(v_*) + q^* \|\Delta v\|_V^2 \right) \ .$$

It follows from the necessary first-order conditions that $\Delta v_* = 0$ is first-order minimal for $\varphi_Q^*(.)$ in $\hat{V}$. Then, Lemma 2.6 ensures that v_* is first-order minimal for $\varphi(.)$ in $\hat{V}$.

The same argument can be used for any other cluster point of $\{v_k\}_{k\in\mathbb{N}}$ yielding the assertion. □

In Algo. 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 [11, Theo. 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 $\mu \in [0,1]$. However, this approach complicates the convergence analysis considerably and is the subject of further research.

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 Assump. 2.5.

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 linearisations

$$\Delta w = \Delta v \pm \Delta u \qquad \text{for} \quad w = v \pm u \ , \tag{16}$$
$$\Delta w = \psi'(v)(\Delta v) \qquad \text{for} \quad w = \psi(v) \neq \operatorname{abs}(v) \ , \tag{17}$$

where $\psi'(v)$ denotes the Fréchet derivative of ψ. Here, we face one difference to the finite dimensional case since we do not have to introduce a linearisation 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 linearisations are simply given by

$$\Delta w = A\,\Delta v \qquad \text{for} \quad w = A\,v \ . \tag{18}$$

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 $\varphi'(v) : V \mapsto \mathbb{R}$ 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 linearisation to take for the absolute value operator. Our method of choice is the so-called abs-linearisation given by

$$\Delta w = \mathrm{abs}(v + \Delta v) - w \qquad \text{for} \qquad w = \mathrm{abs}(v) \ . \tag{19}$$

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 Tab. 1 on the argument v is denoted by $v_i(v)$. The local model is constructed in the following way:

Definition 4.1 (Abs-Linearisation)**.** *Suppose $\varphi : V \mapsto \mathbb{R}$ is an element of the operator class $C^1_{abs}(V)$ as defined in Def. 2.2. For a fixed argument $v \in V$ and $w = \varphi(v)$ the abs-linearisation $\Delta w(v; .) : V \mapsto \mathbb{R}$ based on the linearisations Eqs.* (16)–(19) *is constructed in the following way:*

$v_0 = v$, $\Delta v_0 = \Delta v$
for $i = 1, \ldots, s$ *do*
$$\begin{aligned} z_i &= \psi_i((v_j)_{j<i}) \\ \Delta z_i &= \Delta\psi_i((v_j)_{j<i})((\Delta v_j)_{j<i}) \\ \sigma_i &= \mathrm{sign}(z_i + \Delta z_i) \\ v_i &= \sigma_i z_i = \mathrm{abs}(z_i) \\ \Delta v_i &= \mathrm{abs}(z_i + \Delta z_i) - \mathrm{abs}(z_i) \end{aligned}$$
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 Sec. 2.

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

Proposition 4.2 (Approximation Properties of the Abs-linearisation)**.** *Suppose $\varphi \in C^1_{abs}(V)$. Then there exists a constant $q > 0$, 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*

$$\varphi_{loc}(\bar{v}; .) : V \mapsto \mathbb{R}, \qquad \varphi_{loc}(\bar{v}; \Delta v) = \varphi(\bar{v}) + \Delta\varphi(\bar{v}; \Delta v) \tag{20}$$

that

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

Moreover, there exists a constant $\gamma > 0$ such that for any pair $\bar{v}, v \in \mathcal{W}$ and $w \in V$ one has

$$\frac{|\varphi_{loc}(\bar{v}; w) - \varphi_{loc}(v; w)|}{1 + \|w\|_V} \leqslant \gamma\|\bar{v} - v\|_V \ .$$

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 $\Delta v = \mathring{v} - v$ in a neigbourhood 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., [25, Sec. 4.5], ensures that the linearisations Eqs. (16) and (17) 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 $w = \text{abs}(u)$. Eq. (19) yields

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

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.

To show the second assertion we first note that one obtains from the Lipschitz continuity again by induction for all i

$$v_i(\bar{v}) - v_i(v) = \mathcal{O}(\|\bar{v} - v\|_V) \quad \text{and} \quad \|\Delta v_i(\bar{v}; \Delta v)\| \leqslant l_i \|\Delta v\|_V \tag{21}$$

hold for a suitable constants l_i, $1 \leqslant i \leqslant s + 1$. The actual assertion can now be derived by showing that

$$\frac{\|\Delta v_i(\bar{v}; w) - \Delta v_i(v; w)\|_{V_i}}{1 + \|w\|_V} = \mathcal{O}(\|\bar{v} - v\|_V)$$

holds for all $i = 1, \ldots, s+1$ This is true for the first intermediate v_0 whose increment $\Delta v_0 = \Delta v$ is chosen independently of v. Similar to the partial derivatives in finite dimension, one can consider $c_{ij}(v) := \partial_j \psi_i((v_j)_{j<i})$, i.e., the Fréchet derivative with respect to the intermediate value v_j obtained for the argument v. Then it follows for the Lipschitz continuously Fréchet differentiable elemental operators $v_i = \psi_i(v_j)_{j<i}$, $i = 1, \ldots, s + 1$ that

$$\|c_{ij}(\bar{v}) - c_{ij}(v)\|_{V_i} = \mathcal{O}(\|\bar{v} - v\|_V) \ .$$

Hence, one obtains for all Lipschitz continuously Fréchet differentiable elemental operators

$$\begin{aligned}
&\frac{\|\Delta v_i(\bar{v};w) - \Delta v_i(v;w)\|_{V_i}}{1+\|w\|_V}\\
\leqslant&\frac{\left\|\sum_{j<i}(c_{ij}(\bar{v}) - c_{ij}(v))\Delta v_j(\bar{v};w) + \sum_{j<i} c_{ij}(v)(\Delta v_j(\bar{v};w) - \Delta v_j(v;w))\right\|_{V_i}}{1+\|w\|_V}\\
\leqslant&\frac{\sum_{j<i}\mathcal{O}(\|\bar{v}-v\|_V)l_j\|w\|_V + \sum_{j<i}\|c_{ij}(\bar{v})\|_{V_i}\cdot\|\Delta v_j(\bar{v};w) - \Delta v_j(v;w)\|_{V_j}}{1+\|w\|_V}\\
\leqslant&\mathcal{O}(\|\bar{v}-v\|_V) + \sum_{j<i}\|c_{ij}(\bar{v})\|_{V_i}\mathcal{O}(\|\bar{v}-v\|_V) = \mathcal{O}(\|\bar{v}-v\|_V)
\end{aligned}$$

using the Lipschitz constants l_j, whose existence has been asserted in Eq. (21). Hence, once again, we only have to prove the assertion for the absolute value where

$$\begin{aligned}
&\|\Delta v_i(\bar{v};w) - \Delta v_i(v;w)\|_{V_i}\\
=&\|\mathrm{abs}(v_j(\bar{v}) + \Delta v_j(\bar{v};w)) - \mathrm{abs}(v_j(\bar{v})) - [\mathrm{abs}(v_j(v) + \Delta v_j(v;w)) - \mathrm{abs}(v_j(v))]\|_{V_i}\\
\leqslant&\|v_j(\bar{v}) + \Delta v_j(\bar{v};w) - [v_j(v) + \Delta v_j(v;w)]\|_{V_j} + \|v_j(\bar{v}) - v_j(v)\|_{V_j}\\
\leqslant&\|v_j(\bar{v}) - v_j(v)\|_{V_j} + \|\Delta v_j(\bar{v};w) - \Delta v_j(v;w)\|_{V_j} + \|v_j(\bar{v}) - v_j(v)\|_{V_j}\\
=&(1+\|w\|_V)\mathcal{O}(\|\bar{v}-v\|_V) + 2\mathcal{O}(\|\bar{v}-v\|_V) = (1+\|w\|_V)\mathcal{O}(\|\bar{v}-v\|_V)
\end{aligned}$$

yielding

$$\frac{|\Delta\varphi(\bar{v};w) - \Delta\varphi(v;w)|}{1+\|w\|_V} = \mathcal{O}(\|\bar{v}-v\|_V)\ .$$

Combining this with the definition Eq. (20) of the local model and the Lipschitz continuity this property follows also for the local model φ_{loc}, which completes the proof of the second assertion. □

The next example illustrates that there is a whole class of PDE-constrained optimization problems that fulfill the requirements of the local model required in the previous section.

Example 4.3. *Consider once more the optimization problem as introduced in Ex. 2.3*

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

Defining the Fréchet differentiable operator

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

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 Def. 4.1 is given by

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

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 [23]. Then the operator $\varphi_{loc}(.;.)$ is quasiconvex in both arguments.

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-linearisation to construct a local model that has the required approximation properties of second order. Finally, we discussed PDE-constrained problems that fit into the considered setting. This includes for example 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 dedicated to an extension of the theory for more general solution operators such that for example also nonsmooth PDE constraints can be handled.

Acknowledgment

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

References

[1] P.-A. Absil, R. Mahony, and B. Andrews. Convergence of the iterates of descent methods for analytic cost functions. *SIAM Journal on Optimization*, 16(2):531–547, 2005.

[2] W. Van Ackooij, J. Y. Bello Cruz, and W. De Oliveira. A strongly convergent proximal bundle method for convex minimization in hilbert spaces. *Optimization*, pages 145–167, 2016.

[3] H. Attouch. Epi-convergence and duality. Convergence of sequences of marginal and lagrangians functions. Applications to homogenization problems in mechanics. In R. Conti, E. De Giorgi, and F. Giannessi, editors, *Optimization and Related Fields*, pages 21–56, Berlin, Heidelberg, 1986. Springer Berlin Heidelberg.

[4] H. Attouch and J. Bolte. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Mathematical Programming. Series B*, 116(1-2):5–16, 2009.

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

[6] C. Christof, C. Clason, C. Meyer, and S. Walther. Optimal control of a non-smooth semilinear elliptic equation. Technical Report SPP1962-020, DFG SPP 1962, 2017.

[7] Christian Clason, Vu Huu Nhu, and Arnd Rösch. Optimal control of a non-smooth quasilinear elliptic equation. Preprint SPP1962-101, 12 2018.

[8] R. Correa and C. Lemaréchal. Convergence of some algorithms for convex minimization. *Math. Programming*, pages 261–275, 1993.

[9] P.G. Dodds, T.K. Dodds, B. de Pagter, and F.A. Sukochev. Lipschitz continuity of the absolute value and Riesz projections in symmetric operator spaces. *Journal of Functional Analysis*, 148(1):28–69, 1997.

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

[11] S. Fiege, A. Walther, and A. Griewank. An algorithm for nonsmooth optimization by successive piecewise linearization. *Mathematical Programming, Series A*, 2018. DOI: 10.1007/s10107-018-1273-5.

[12] A. Griewank. The modification of Newtons method for unconstrained optimization by bounding cubic terms. Technical Report NA/12, University of Cambridge, 1981.

[13] A. Griewank. On stable piecewise linearization and generalized algorithmic differentiation. *Optimization Methods and Software*, 28(6):1139–1178, 2013.

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

[15] A. Griewank and A. Walther. Relaxing kink qualifications and proving convergence rates in piecewise smooth optimization. *SIAM Journal on Optimization*, 2018. accepted for publication.

[16] O. Guler. On the convergence of the proximal point algorithm for convex minimization. *SIAM Journal on Control and Optimization*, 29(2):403–419, 1991.

[17] L. Hertlein and M. Ulbrich. An inexact bundle algorithm for nonconvex nondifferentiable functions in hilbert space. Technical Report SPP1962-084, Technische Universität München, 2018.

[18] C. Hess. Epi-convergence of sequences of normal integrands and strong consistency of the maximum likelihood estimator. *The Annals of Statistics*, 24(3):1298–1315, 1996.

[19] J. Jahn. *Introduction to the theory of nonlinear optimization.* Springer, 2007.

[20] R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14:877–898, 1976.

[21] R.T. Rockafellar and R.J.-B. Wets. *Variational analysis.* Springer, 1998.

[22] S. Scholtes. *Introduction to Piecewise Differentiable Functions.* Springer, 2012.

[23] G. Stadler. Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices. *Computational Optimization and Applications*, 44(2), Nov 2007.

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

[25] E. Zeidler. *Applied functional analysis. Applications to mathematical physics.* Springer, 1995.