LINE SEARCH GLOBALIZATION OF A SEMISMOOTH NEWTON METHOD FOR OPERATOR EQUATIONS IN HILBERT SPACES WITH APPLICATIONS IN OPTIMAL CONTROL

MATTHIAS GERDTS, STEFAN HORN AND SVEN-JOACHIM KIMMERLE

Institut für Mathematik und Rechneranwendung (LRT-1)
Universität der Bundeswehr München
Werner-Heisenberg-Weg 39
85577 Neubiberg/München, Germany

(Communicated by Liqun Qi)

Abstract. We consider the numerical solution of nonlinear and nonsmooth operator equations in Hilbert spaces. A semismooth Newton method is used for search direction generation. The operator equation is solved by a globalized semismooth Newton method that is equipped with an Armijo linesearch using a semismooth merit function. We prove that an accumulation point of the globalized algorithm is a solution and transition to fast local convergence under a directional Hadamard-like continuity assumption on the Newton matrix. In particular, no auxiliary descent directions or smoothing steps are required. Finally, we apply this method to a control-constrained and also to a regularized state-constrained optimal control problem subject to partial differential equations.

1. Introduction. In this article, we consider the problem of finding a \( \bar{z} \in Z \) with
\[
f(\bar{z}) = 0,
\]
where \( Z \) is a Hilbert space, \( Z^* \) denotes its topological dual, and \( f : Z \to Z^* \) is a locally Lipschitz continuous operator. In particular, we allow \( f \) to be nonlinear and nonsmooth. Problems of this type arise, e.g., from the reformulation of the necessary optimality conditions in control-constrained optimal control subject to partial differential equations (PDE), see Section 4.

Semismooth Newton methods for operator equations of the type of (1) have been examined in [4, 10, 15, 16, 18, 19, 21, 22, 23, 24]. A trust-region globalization of this method was presented in [22], global convergence is proved for a primal-dual active set method in finite dimensions in [15, 16]. A primal-dual active set strategy and semismooth Newton methods may lead to identical algorithms [8]. The application of semismooth Newton methods to optimal control of PDE was considered, e.g., in [9, 10, 12, 13, 20, 22, 23, 24]. For variationally discretized control-constrained optimal control problems subject to elliptic differential equations a globalization
strategy was developed by Hinze and Vierling [12, 13]. Discretized nonlinear optimal control problems subject to control and state constraints can be solved by globally convergent nonsmooth Newton methods, see for instance [1]. It has been demonstrated that a norm gap between range and domain of a projection operator is required for semismoothness, see e.g. [21 Th. 3.3], [22 Remark 3.34] or [24 Ch. 4]. In [21] the semismooth Newton method is revisited from a different point of view involving semismooth superposition operators and replacing Lipschitz continuity by weaker grow conditions.

We extend the local semismooth Newton method, see M. Ulbrich [22], by a globalization strategy in Hilbert spaces from [14]. We prove that if our global algorithm produces an accumulation point, then we have global convergence to a zero of $f$, see our main results Theorem 3.4 and Theorem 3.5 that, to our best knowledge, have not been proved so far. As in [22] we do not require a strict complementarity assumption for these results.

Our article is organized as follows. The concept of a semismooth Newton method is briefly presented in Section 2. In Section 3 we prove our main result. This globalization result is applied to optimal control problems with control constraints in Sect. 4 using an all-at-once approach. We close with numerical examples for control constraints and for state constraints in Sect. 5.

2. Semismooth Newton method. For applying a Newton method on (1) an appropriate substitute for the Fréchet-derivative $f'$ has to be found. Therefore we work with the following abstract semismoothness concept, cf. Def. 3.1 in [22]:

**Definition 2.1** ($\partial^* f$-Semismoothness). Let $f : Z \supset V \to Z^*$ be defined on an open subset $V$ of $Z$. Further, let be given a set-valued mapping $\partial^* f : V \rightrightarrows \mathcal{L}(Z, Z^*)$. $f$ is called

(a) $\partial^* f$-semismooth at $z \in V$ if $f$ is continuous near $z$ and

\[
\sup_{M \in \partial^* f(z+s)} \| f(z+s) - f(z) - M s \|_{Z^*} = o(\|s\|_Z) \quad \text{as } \|s\|_Z \to 0,
\]

i.e. $\forall \varepsilon > 0 \exists \delta > 0 : \|s\|_Z < \delta \Rightarrow \| f(z+s) - f(z) - M s \|_{Z^*} \leq \varepsilon \|s\|_Z$

\[\forall M \in \partial^* f(z+s).\]

(b) $\alpha$-order $\partial^* f$-semismooth at $z \in V$, $0 < \alpha \leq 1$, if $f$ is continuous near $z$ and

\[
\sup_{M \in \partial^* f(z+s)} \| f(z+s) - f(z) - M s \|_{Z^*} = \mathcal{O}(\|s\|_Z^{1+\alpha}) \quad \text{as } \|s\|_Z \to 0.
\]

i.e. $\exists \varepsilon > 0 \exists \delta > 0 : \|s\|_Z < \delta \Rightarrow \| f(z+s) - f(z) - M s \|_{Z^*} \leq \varepsilon \|s\|_Z^{1+\alpha}$

\[\forall M \in \partial^* f(z+s).\]

(c) uniformly $\partial^* f$-semismooth for all $z \in V$, if $f$ is continuous in $V$ and

\[\forall \varepsilon > 0 \exists \delta > 0 \forall z \in V : \|s\|_Z < \delta \Rightarrow \| f(z+s) - f(z) - M s \|_{Z^*} \leq \varepsilon \|s\|_Z\]

\[\forall M \in \partial^* f(z+s), \text{i.e. } \delta \text{ does not depend on } z.\]

The multifunction $\partial^* f$ is called generalized differential of $f$.

Please note that the uniform semismoothness of $f$ in part (c) of Definition 2.1 is only required later on in the particular case that a (sub)sequence of stepsizes in the globalized Newton method tends to zero, compare Part b) of Theorem 3.4. However, there exist projectors s.t. uniform semismoothness implies Fréchet differentiability. For this and a discussion of non-uniform semismoothness of projectors see [21 Sect. 3]. Note that a different definition of uniform semismoothness is used by
Correa and Joffre \[5, Remark 6.3\] in which the uniformity refers to a family of functions. For a uniform semismoothness within the context of mesh independence see \[24, Ch. 6\]. It is easy to prove that the direct product of (\(\alpha\)-order) semismooth operators is (\(\alpha\)-order) semismooth itself with respect to the direct product of the generalized differentials of its components. An equivalent relation holds for the composition of semismooth operators. For details see \[22\]. For \(Z = \mathbb{R}^n\) thanks to Rademacher’s theorem such a substitute for locally Lipschitz continuous functions is given by Clarke’s generalized Jacobian. In infinite dimensional spaces such a generalized differential is not given naturally and has to be defined separately, see e.g. \[6, 22\].

A semismooth Newton method for equation (1) is given by Algorithm 2.2 (Local semismooth Newton method).

(i) Choose a starting point \(z_0 \in Z\) and set \(k = 0\).
(ii) Is a stopping criterion satisfied: STOP.
(iii) Choose an arbitrary \(M(z_k) \in \partial^* f(z_k)\) and compute the search direction \(s_k \in Z\) by solving
\[
M(z_k) s_k = -f(z_k).
\]
(iv) Set \(z_{k+1} = z_k + s_k, k := k + 1\) and goto (ii).

Here we need some regularity condition for the operators \(M(z_k) \in \partial^* f(z_k)\) similar as in \[22, Assumption 3.11 (i)\].

**Assumption 2.3** (Uniform non-singularity of the Newton matrix). The operators \(M(z_k) \in \partial^* f(z_k)\) are continuously invertible elements of \(L(Z, Z^*)\) and there exists a constant \(C_{M^{-1}} > 0\) with
\[
\|M(z_k)^{-1}\|_{L(Z^*, Z)} \leq C_{M^{-1}} \quad \forall k \in \mathbb{N}.
\]

This assumption is a standard assumption for Newton methods and can be verified specifically for various examples e.g. \[9, 10, 17\]. In a finite dimensional setting, this assumption will be satisfied close to a zero \(\bar{z}\) of \(f\), if all elements \(M\) of \(\partial^* f(\bar{z})\) are non-singular, where \(\partial^* f\) denotes the compact-valued and upper semicontinuous subdifferentials of Bouligand, Clarke or Qi, compare \[23, Remark 2.8\].

For our globalization result we need Assumption 3.1, see below, for the subdifferentials. Furthermore we need a certain structure of the superposition operator arising within the context of the application to optimal control, see Assumption 4.2 below, that avoids a smoothing step in Algorithm 2.2 required in \[6\] within this context.

We recall the local convergence of Algorithm 2.2 that has been shown in \[6\] and \[22\].

**Theorem 2.4** (Local convergence of Algorithm 2.2). Let \(V \subset Z\) be open and \(f : V \to Z^*\). Let Assumption 2.3 hold. Assume \(\bar{z} \in V\) is a zero of \(f\), then the following assertions hold:

(a) If \(f\) is \(\partial^* f\)-semismooth in \(\bar{z}\), then there exists a constant \(\delta > 0\), such that for all \(z_0 \in \bar{z} + \delta B_Z\) Algorithm 2.2 generates a sequence \(\{z_k\} \subset V\) that converges superlinearly to \(\bar{z}\).

(b) If \(f\) in (a) is \(\alpha\)-order \(\partial^* f\)-semismooth in \(\bar{z}\), \(0 < \alpha \leq 1\), the rate of convergence is \(1 + \alpha\).
(c) If $f(z_k) \neq 0$ for all $k$, the sequence of residuals converges superlinearly, i.e.
\[
\lim_{k \to \infty} \frac{\|f(z_{k+1})\|_{Z^*}}{\|f(z_k)\|_{Z^*}} = 0.
\]

3. **Globalization.** To construct a globally convergent Newton method Algorithm 2.2 is extended by an Armijo line search. An appropriate merit function $\Theta : Z \to \mathbb{R}$ with $\Theta(z) \to 0$ for $\|f(z)\|_{Z^*} \to 0$ is given by
\[
\Theta(z) := \frac{1}{2} \|f(z)\|_{Z^*}^2.
\]
Due to the fact that $Z$ is a Hilbert space one can prove the semismoothness of (3).
To this end, let $\langle \cdot, \cdot \rangle_{Z^*,Z}$ denote the dual pairing between $Z^*$ and $Z$ while $\langle \cdot, \cdot \rangle_Z$ is the inner product on $Z$. We need

**Assumption 3.1** (Directional Hadamard-like continuity of the Newton matrix).

(a) For any $k \in \mathbb{N}$ there exist $M_{s} := M(z_k + ps_k) \in \partial^* f(z_k + ps_k)$ for every $p$ sufficiently small with a $\bar{\varepsilon} \in (0,1)$ such that
\[
\lim_{\rho \downarrow 0} \|(M_{\rho} - M(z_k))s_k\|_{Z^*} = \bar{\varepsilon}\|f(z_k)\|_{Z^*}.
\]
In particular, this assumption is fulfilled in the case $\bar{\varepsilon} = 0$, corresponding to a directional Hadamard-like continuity of $M$.

(b) For $k \to \infty$, $\alpha_k \to 0$ and a subsequence of $\{z_k\}$ converging to $\bar{z}$ there exists $\bar{M} \in \partial^* f(\bar{z})$ with
\[
\lim_{k \to \infty} \|(M_k - \bar{M})s_k\|_{Z^*} = 0
\]
for $M_k \in \partial^* f(z_k + \alpha_k s_k)$.

The conditions in Assumption 3.1 have some similarity with the definition of Hadamard differentiability [1, Def. 2.45] along curves. In our context, however, these conditions ensure the continuity or the smallness of jumps of certain elements of the subdifferentials at least in the directions $s_k$. Assumptions (a) and (b) are satisfied generically at all points of differentiability of $f$. At points of non-differentiability of $f$ the assumptions do impose restrictions on the choice of the elements from the subdifferentials and enforce a continuous selection. There might be cases where such a selection is not possible, because the search direction $s_k$ may be obtained with some $M(z_k)$ and points into a direction where $M_{s}$ is bounded away from $M(z_k)$. In the latter case we need at least bounds on the jump of $M$ along a search direction.

However, we are able to monitor this situation algorithmically since Assumption 3.1 (a) will be exploited in the subsequent lemma to prove that the search direction is a direction of descent. Hence, if the search direction turns out to be a direction of ascent, then Assumption 3.1 (a) will be violated. In our numerical experiments that, however, rely on a discretized version of Algorithm 3.3 we haven’t encountered this situation, though.

**Lemma 3.2** (Semismoothness of $\Theta$). Let $f : Z \to Z^*$ be $(\alpha$-order) $\partial^* f$-semismooth. Furthermore, let $H : Z^* \to \mathbb{R}$ be given by $H(\cdot) := \frac{1}{2} \|\cdot\|_{Z^*}^2 = \frac{1}{2} \langle \cdot, \cdot \rangle_{Z^*}$, where $\langle \cdot, \cdot \rangle_{Z^*}$ denotes the inner product in $Z^*$. Then the following assertions hold:

(a) $H$ is Fréchet-differentiable on $Z^*$ with $H'(f) h = (f, h)_{Z^*}$ for $f : Z \to Z^*, h \in Z^*$. 

(b) The merit function $\Theta : Z \to \mathbb{R}$ in (3) is $(\alpha$-order) $\partial^*\Theta$-semismooth, where the generalized differential $\partial^*\Theta := \{H' \circ \partial^* f : Z \rightrightarrows \mathcal{L}(Z, \mathbb{R})\}$ is defined by
\[
\partial^*\Theta(z) = \{H' \circ \partial^* f \}(z) = \{H'(f(z)) M(z) : M(z) \in \partial^* f(z)\}.
\]
(c) Let $\{z_k\} \subset Z$ be a sequence generated by Algorithm 2.2 and let Assumption 2.3 hold. For any $s_k$ being defined by $M(z_k)s_k = -f(z_k)$ and the associated $V(z_k) = H'(f(z_k))M(z_k) \in \partial^*\Theta(z_k)$, it holds that
\[
\langle V(z_k), s_k \rangle_{Z^*, Z} = -\|f(z_k)\|^2_{Z^*} = -2\Theta(z_k).
\]
Furthermore, if Assumption 3.1 (a) holds, then $s_k$ is a descent direction of $\Theta$ in $z_k$ unless $\Theta(z_k) = 0$.

Proof. (a) As common we identify the dual space $Z^{**}$ with $Z^*$ by means of the Riesz representation. We find
\[
\frac{H(f + h) - H(f) - (f, h)_{Z^*}}{\|h\|_{Z^*}} = \frac{1}{2} (f + h, f + h)_{Z^*} - \frac{1}{2} (f, f)_{Z^*} - (f, h)_{Z^*} = \frac{1}{2}\|h\|_{Z^*}.
\]
Taking the limit $\|h\|_{Z^*} \to 0$ yields the assertion.
(b) Due to the $(\alpha$-order) $\partial^* f$-semismoothness and local Lipschitz continuity of $f$, the validity of assertion (a) and the boundedness of $H'(f)$, all requirements of Proposition 3.7 in [22] are satisfied and $\Theta$ is $(\alpha$-order) $\partial^*\Theta$-semismooth. Thus assertion (b) holds.
(c) With (a) we find
\[
\langle V(z_k), s_k \rangle_{Z^*, Z} = H'(f(z_k))(M(z_k)s_k) = (f(z_k), M(z_k)s_k)_{Z^*} = -\|f(z_k)\|^2_{Z^*}.
\]
and thus $\langle V(z_k), s_k \rangle_{Z^*, Z} = -2\Theta(z_k)$.

The $\partial^*\Theta$-semismoothness of $\Theta$ according to (b) implies
\[
0 = \lim_{\rho \downarrow 0} \sup \frac{\|\Theta(z_k + \rho s_k) - \Theta(z_k)\|}{\rho} - \langle V(z_k + \rho s_k), s_k \rangle_{Z^*, Z}
\]
\[
\geq \lim_{\rho \downarrow 0} \sup \frac{\|\Theta(z_k + \rho s_k) - \Theta(z_k)\|}{\rho} - \langle V(z_k), s_k \rangle_{Z^*, Z}
\]
\[
- \langle V(z_k), s_k \rangle_{Z^*, Z} - \langle V(z_k + \rho s_k), s_k \rangle_{Z^*, Z}
\]
\[
= \lim_{\rho \downarrow 0} \sup \frac{\|\Theta(z_k + \rho s_k) - \Theta(z_k)\|}{\rho} + 2\Theta(z_k)
\]
\[
- \langle V(z_k), s_k \rangle_{Z^*, Z} - \langle V(z_k + \rho s_k), s_k \rangle_{Z^*, Z}.
\]
This relation holds for arbitrary choices of $M(z_k + \rho s_k) \in \partial^* f(z_k + \rho s_k)$ in $V(z_k + \rho s_k) = H'(f(z_k + \rho s_k))M(z_k + \rho s_k)$.

According to Assumption 3.1 (a) there exists $M_\rho := M(z_k + \rho s_k) \in \partial^* f(z_k + \rho s_k)$ such that
\[
\lim_{\rho \downarrow 0} \|\langle (M_\rho - M(z_k))s_k \rangle_{Z^*} = \varepsilon\|f(z_k)\|_{Z^*}.
\]
Then, together with the continuity of $H'$ we find
\[
\|V(z_k), s_k\|_{Z^*, Z} = \langle V(z_k + ps_k), s_k\rangle_{Z^*, Z}
\]
\[
= (H'(f(z_k)), M(z_k)s_k)_{Z^*, Z} - (H'(f(z_k + ps_k)), M(z_k + ps_k)s_k)_{Z^*, Z}
\]
\[
\leq (H'(f(z_k)), M(z_k)s_k)_{Z^*, Z} - (H'(f(z_k + ps_k)), M(z_k + ps_k)s_k)_{Z^*, Z}
\]
\[
+ \|H'(f(z_k + ps_k)), M(z_k + ps_k)s_k\|_{L(Z, Z)} \cdot \|f(z_k)\|_{Z^*}
\]
\[
+ \|H'(f(z_k + ps_k))\|_{L(Z, Z^*)} \cdot \|M(z_k + ps_k)s_k\|_{Z^*}
\]
\[
\rightarrow \|f(z_k)\|_{Z^*, Z}^2 \quad \text{as } \rho \downarrow 0.
\]

Hence, \[
\lim_{\rho \downarrow 0} \left| \frac{\Theta(z_k + ps_k) - \Theta(z_k)}{\rho} + 2\Theta(z_k) \right| = 2\epsilon \Theta(z_k)
\]
and since $\epsilon < 1$, $s_k$ is a direction of descent of $\Theta$ at $z_k$ unless $\Theta(z_k) = 0$.

Hence with $f(z_k) \neq 0$ the Armijo line search in the following algorithm is well defined. The descent property of $s_k$ w.r.t. $\Theta(z_k)$ follows particularly from the construction of the merit function $\Theta$.

**Algorithm 3.3** (Global semismooth Newton method).

(i) Choose $z_0 \in Z$, $\beta \in (0, 1)$, $\sigma \in (0, 1/2)$ and set $k = 0$.

(ii) Is a stopping criterion satisfied: STOP.

(iii) Choose an arbitrary $M(z_k) \in \partial^* f(z_k)$ and compute the search direction $s_k \in Z$ by solving

\[
M(z_k) s_k = -f(z_k).
\]

(iv) Calculate $V(z_k) = H'(f(z_k)) M(z_k) \in \partial^* \Theta(z_k)$, find the smallest $i_k \in \mathbb{N}_0$ with

\[
\Theta(z_k + \beta^{i_k} s_k) \leq \Theta(z_k) + \sigma \beta^{i_k} \langle V(z_k), s_k \rangle_{Z^*, Z}
\]

and set $\alpha_k = \beta^{i_k}$.

(v) Set $z_{k+1} = z_k + \alpha_k s_k$, $k := k + 1$, and goto (ii).

Using \[4\] the inequality in \[5\] can equivalently be written as

\[
\Theta(z_k + \beta^{i_k} s_k) \leq (1 - 2\sigma \beta^{i_k}) \Theta(z_k).
\]

Hence $\langle V_k, s_k \rangle_{Z^*, Z}$ has not to be computed explicitly. Global convergence of Algorithm 3.3 can be shown by adapting Theorem 4.2 in \[8\] that has been derived within the frame of optimal control of ordinary differential equations.

**Theorem 3.4** (Global convergence of Algorithm 3.3). Let $f$ be semismooth. Let $\bar{z} \in Z$ be an accumulation point of the sequence $\{z_k\}$ generated by Algorithm 3.3 and let Assumption 3.3 hold.

(a) If Assumption 3.1 (a) holds and

\[
\alpha := \liminf_{j \to \infty} \alpha_{kj} > 0,
\]

then $\bar{z}$ is a zero of $f$.

(b) If $\alpha = 0$, if Assumption 3.1 holds and if $f$ is locally uniformly semismooth in a neighbourhood of $\bar{z}$, then $\bar{z}$ is a zero of $f$. 

For clarification, we would like to mention already here that the locally uniform semismoothness and Assumption 3.1(b), are not fulfilled by the numerical examples in Sect. 5. However, it turns out that the pathological case (b) of the theorem, in which the Armijo step \( \alpha_k \) tends to zero, can be ruled out by the following Theorem 3.5 under Assumption 3.1(a) and is not met within both our examples.

**Proof.** We adapt the argumentation of Theorem 4.2 in [4]. Therefore let \( \{z_k\}_{j \in \mathbb{N}} \) be a subsequence with \( z_{k_j} \to \bar{z} \) and \( f(z_{k_j}) \neq 0 \). Then, \( (V(z_{k_j}), s_{k_j})_{Z^*Z} = -2\Theta(z_{k_j}) = -\|f(z_{k_j})\|_{Z^*}^2 < 0 \). According to Lemma 3.2(c), \( s_{k_j} \) is a descent direction of \( \Theta \) at \( z_{k_j} \) and the line search is well defined. There are two cases:

(a) Assume \( \alpha > 0 \). We have

\[
0 \leq \Theta(z_{k_j+1}) \leq \Theta(z_{k_j}) \leq \Theta(z_{k_j}) + \sigma \alpha_k (V(z_{k_j}), s_{k_j})_{Z^*Z} = (1 - 2\sigma \alpha_k)\Theta(z_{k_j}).
\]

With \( \sigma \in (0, 1/2) \) and \( \alpha \leq \alpha_k \leq 1 \) it follows that \( 0 < 1 - 2\sigma \alpha_k \leq 1 - 2\sigma \alpha < 1 \). Repeated application yields

\[
0 \leq \Theta(z_{k_j}) \leq (1 - 2\sigma \alpha)^j \Theta(z_{k_0}) \to 0.
\]

By the continuity of \( f, \bar{z} \) is a zero of \( f \).

(b) Assume that there is a subsequence \( \{z_k\}_{k \in J}, J \subseteq \{k_j : j \in \mathbb{N}\} \) with \( \alpha_k \to 0 \), \( k \in J \). The sequence \( \{s_k\} \) is bounded since \( \{M(z_k)^{-1}\} \) is bounded and

\[
0 \leq \|s_k\|_Z = \|M(z_k)^{-1}f(z_k)\|_Z \leq C_{M^{-1}} \|f(z_k)\|_{Z^*} \leq C_{M^{-1}} \|f(z_0)\|_{Z^*}.
\]

\( Z \) is a Hilbert space and thus reflexive. According to the Eberlein-Smulian theorem, there exists a weakly convergent subsequence \( \{s_k\}, k \in I \subseteq J \). Hence, there exists some \( \bar{s} \in Z \) such that for every \( V(\bar{z}) \in Z^* \) we have

\[
\langle V(\bar{z}), s_k \rangle_{Z^*Z} \to \langle V(\bar{z}), \bar{s} \rangle_{Z^*Z} \quad \text{as } k \to \infty.
\]  

(6)

We get

\[
\frac{\Theta(z_{k+1}) - \Theta(z_k)}{\alpha_k} \leq \frac{\Theta(z_{k+1}) - \Theta(z_k)}{\alpha_k} - \langle V(\bar{z}), \bar{s} \rangle_{Z^*Z} + \|V(z_{k+1}), s_k\|_{Z^*Z} + \|V(z_{k+1}), s_k\|_{Z^*Z} - \langle V(\bar{z}), \bar{s} \rangle_{Z^*Z}.
\]

The first term vanishes owing to the locally uniform semismoothness of \( \Theta \), the last term vanishes owing to (6) as \( k \to \infty \). For the remaining term we obtain
with $M$, as defined in Assumption 3.1(b), the following estimate

\[ |\langle V(z_{k+1}), s_k \rangle_{Z^*, Z} - \langle V(\bar{z}), s_k \rangle_{Z^*, Z}| \]

\[ = |H'(f(z_{k+1}))(M(z_{k+1})s_k) - H'(f(\bar{z}))(M s_k)| \]

\[ \leq |(f(z_{k+1}), M(z_{k+1})s_k)_{Z^*} - (f(\bar{z}), M s_k)_{Z^*}| \]

\[ \leq \|f(z_{k+1}) - f(\bar{z})\|_{Z^*} \cdot \|M(z_{k+1}) - M\|_{Z^*} \cdot \|s_k\|_{Z^*} + \|f(\bar{z})\|_{Z^*} \cdot \|M\|_{Z^*} \cdot \|s_k\|_{Z^*}. \]

The first term vanishes owing to the continuity of $f$ and the boundedness of $\{M(z_k)\}$ and $\{s_k\}$. The second term vanishes because of Assumption 3.1(b).

In summary we have shown that

\[ \lim_{k \to \infty, k \notin I} \frac{\Theta(z_{k+1}) - \Theta(z_k)}{\alpha_k} = \langle V(\bar{z}), s \rangle_{Z^*, Z}. \]

The line search in step (iv) of Algorithm 3.3 yields

\[ \frac{\Theta(z_k + \alpha_k s_k) - \Theta(z_k)}{\alpha_k} \leq \sigma \langle V(z_k), s_k \rangle_{Z^*, Z} < \frac{\Theta(z_k + \frac{\alpha_k}{\beta} s_k) - \Theta(z_k)}{\alpha_k}. \]

Passing to the limit and exploiting the previous considerations yields

\[ \sigma \langle V(\bar{z}), s \rangle_{Z^*, Z} = \langle V(\bar{z}), s \rangle_{Z^*, Z}. \]

Since $\sigma \in (0, 1/2)$ this only holds for $\langle V(\bar{z}), s \rangle_{Z^*, Z} = 0$. Thus, we have shown

\[ -\|f(z_k)\|_{Z^*} = \langle V(z_k), s_k \rangle_{Z^*, Z} \to \langle V(\bar{z}), s \rangle_{Z^*, Z} = 0. \]

By the continuity of $f$, $\bar{z}$ is a zero of $f$.

Now we can prove the transition of Algorithm 3.3 to fast local convergence, provided it converges at all. We present the proof here in details in order to show that we do not need Assumption 3.1(b) and the locally uniform semismoothness in Definition 2.1(c).

**Theorem 3.5** (Transition to fast local convergence). Let Assumptions 2.3 and 3.1(a) hold. Let $\{z_k\}$ be a sequence generated by Algorithm 3.3 with $f(z_k) \neq 0$ for all $k \in \mathbb{N}_0$. Furthermore, let $\bar{z}$ be an accumulation point of $\{z_k\}$ and a zero of $f$.

If $f$ is $\partial^* f$-semismooth the sequence $\{z_k\}$ converges superlinearly to $\bar{z}$ and $\alpha_k$ finally becomes 1. If $f$ is $\alpha$-order $\partial^* f$-semismooth the rate of convergence is $1 + \alpha$.

The idea of the proof is as follows. According to the local convergence result (Theorem 2.4), we know that the local semismooth Newton method (with $\alpha_k = 1$) converges superlinearly in some neighbourhood of a zero. Now, since $\bar{z}$ is an accumulation point of the sequence $\{z_k\}$ there is a subsequence converging to $\bar{z}$, which is assumed to be a zero. For $k$ sufficiently large, $z_k$ is in a neighbourhood of $\bar{z}$, where we have superlinear convergence of the local method. For one step of the local method we show that $\alpha_k = 1$ satisfies the Armijo condition. Hence, the globalized method at $z_k$ will accept $\alpha_k = 1$ and from that point on the sequences of the local method and the globalized method coincide.
Proof. Let \( \varepsilon > 0 \) be arbitrary. The superlinear convergence result of the local algorithm (Theorem 2.4) with \( f \) being semismooth implies the existence of a \( \delta > 0 \) such that for all \( \| z_k - \bar{z} \| Z \leq \delta \) we have

\[
\| z_k + s_k - \bar{z} \| Z \leq \varepsilon \| z_k - \bar{z} \| Z .
\]

(7)

Since \( f(\bar{z}) = 0 \) the Lipschitz continuity of \( f \) yields

\[
\| f(z_k + s_k) \| Z^* \leq L \| z_k + s_k - \bar{z} \| Z
\]

with a constant \( L > 0 \). Combining this estimate with (7) we have

\[
\| f(z_k + s_k) \| Z^* \leq L \varepsilon \| z_k - \bar{z} \| Z .
\]

(8)

On the other hand with Assumption 2.3

\[
\| s_k \| Z = \| M_k^{-1} \| L(\bar{z}, Z) \| f(z_k) \| Z^* \leq C M^{-1} \| f(z_k) \| Z^* ,
\]

and moreover

\[
\| z_k - \bar{z} \| Z \leq \| z_k - (z_k + s_k) \| Z + \| z_k + s_k - \bar{z} \| Z \leq C M^{-1} \| f(z_k) \| Z^* + \varepsilon \| z_k - \bar{z} \| Z .
\]

This yields

\[
\| z_k - \bar{z} \| Z \leq \frac{C M^{-1}}{1 - \varepsilon} \| f(z_k) \| Z^* .
\]

(9)

Together with (8) we have the estimate

\[
\| f(z_k + s_k) \| Z^* \leq L \varepsilon \frac{C M^{-1}}{1 - \varepsilon} \| f(z_k) \| Z^* .
\]

If we choose

\[
\varepsilon \leq \frac{\sqrt{1 - 2\sigma}}{L C M^{-1} + \sqrt{1 - 2\sigma}} < 1 ,
\]

then

\[
\| f(z_k + s_k) \| Z^* \leq \sqrt{1 - 2\sigma} \| f(z_k) \| Z^* ,
\]

i.e. \( \alpha_k = 1 \) is accepted. With (7) and \( \varepsilon < 1 \) we can repeat the argument and obtain convergence of the whole sequence. If \( f \) is \( \partial^* f \)-semismooth the convergence is superlinear according to Theorem 2.4 (a). If \( f \) is \( \alpha \)-order \( \partial^* f \)-semismooth the rate of convergence is \( 1 + \alpha \) according to Th. 2.4 (b).

Thereby we have proved the transition to fast local convergence of Algorithm 3.3 requiring only Assumption 3.1 (a) on the directional Hadamard-like continuity of the Newton matrix and a standard assumption on its uniform non-singularity.

4. Application to optimal control. In control-constrained optimal control subject to PDE constraints the reformulation of the necessary optimality conditions leads to problems of the type of (1). For Hilbert spaces \( Y, W, U \) and the feasible set \( U_{ad} \subset U \) such an optimal control problem, in general, is given as

\[
\min_{(y, u) \in Y \times U} J(y, u) \quad \text{s.t.} \quad E(y, u) = 0, \quad u \in U_{ad} \quad \text{(P1)}
\]

with objective functional \( J : Y \times U \to \mathbb{R} \) and state equation \( E : Y \times U \to W^* \). Here \( y \) denotes the state while \( u \) is the control. Semismooth Newton methods for optimal control problems in the general setting of (P1) are investigated in [22], for more special cases see also [9, 12, 13, 20, 24]. Globalization strategies for semismooth Newton methods applied to a reduced formulation of (P1) are investigated, e.g., in [12, 13, 22]. Although Algorithm 3.3 would be applicable to this reduced problem as well, we want to solve the full Karush-Kuhn-Tucker (KKT) system related to (P1) in [22] this is referred to as the all-at-once approach.
As usual the Lagrange function \( L : Y \times U \times W \to \mathbb{R} \) is defined as
\[
L(y, u, w) = J(y, u) + \langle E(y, u), w \rangle_{W^*, W}.
\] (10)

Then the KKT-system is given by

**Lemma 4.1** (cf. Corollary 1.3 in [1]). Let \((\bar{y}, \bar{u})\) be a solution of \((P1)\). Furthermore, let \(J : Y \times U \to \mathbb{R}\) and \(E : Y \times U \to W^*\) be continuously Fréchet-differentiable. Then there exists a Lagrange multiplier \(\bar{w} \in W\), such that the following optimality conditions hold:

\[
\begin{align*}
L_u(\bar{y}, \bar{u}, \bar{w}) &= E(\bar{y}, \bar{u}) = 0, \quad \tag{11} \\
L_y(\bar{y}, \bar{u}, \bar{w}) &= J_y(\bar{y}, \bar{u}) + E_y(\bar{y}, \bar{u})^* \bar{w} = 0, \quad \tag{12} \\
\bar{u} &\in U_{ad}, \quad \langle L_u(\bar{y}, \bar{u}, \bar{w}), u - \bar{u} \rangle_{U^*, U} \geq 0 \quad \forall u \in U_{ad}. \quad \tag{13}
\end{align*}
\]

We consider the Euclidean projection \(P_{U_{ad}}\) onto the set of admissible controls and introduce the continuous function \(\pi : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}\), \(\xi = (\xi_1, \xi_2, \xi_3) \mapsto \pi(\xi) = \xi_2 - P_{U_{ad}}(\xi_2 - \gamma L_u(\xi_1, \xi_2, \xi_3))\). With the superposition operator \(\Pi : Y \times U \times W \to U\), given as

\[
\Pi(y, u, w)(x) := \pi(y(x), u(x), w(x)) = u(x) - P_{U_{ad}}(u(x) - \gamma L_u(y(x), u(x), w(x))), \quad \gamma > 0,
\]
(14)

the variational inequality in [13] can be rewritten [22, Prop. 5.8] as

\[
\Pi(y, u, w) = 0.
\]

Thus, with \(z := (u, v, w)\) and

\[
f(z) := \begin{bmatrix} L_y(y, u, w) \\ \Pi(y, u, w) \\ E(y, u) \end{bmatrix}
\]
(15)

the solution of \((P1)\) is equivalent to finding a \(\bar{z} := (\bar{y}, \bar{u}, \bar{w}) \in Z := Y \times U \times W\) with

\[
f(\bar{z}) = 0.
\]

For further analysis let \(\Omega \subset \mathbb{R}^n\) be an open, bounded, measurable set with measure \(\mu(\Omega) > 0\). We define the set of control functions as \(U := L^2(\Omega)\). For simplicity we consider pointwise convex controls by using

\[
U_{ad} := \{ u \in L^2(\Omega) : a \leq u(x) \leq b \quad \text{a.e. in } \Omega; \quad a, b \in \mathbb{R} \}
\]
as the feasible set. The pointwise projection \(P_{U_{ad}} : \mathbb{R} \to \mathbb{R}\), \(\xi \mapsto \max\{a, \min\{\xi, b\}\}\) is (1-order) semismooth [22, Ex. 5.23]. Note that \(P_{U_{ad}}\) is not locally uniformly semismooth in any neighbourhood of \(a\) resp. \(b\). In order to establish semismoothness of the superposition operator \(\Pi\) we follow [22, Assumption 5.20] and assume the following:

**Assumption 4.2.** The following conditions hold:

(a) \(E : Y \times L^2(\Omega) \to W^*\) and \(J : Y \times L^2(\Omega) \to \mathbb{R}\) are twice continuously Fréchet-differentiable.

(b) \(L_u\) has the form \(L_u(y, u, w) = \lambda u + G(y, u, w)\) and there exist \(\lambda > 0\) and \(p > 2\),

such that

(i) \(G : Y \times L^2(\Omega) \times W \to L^2(\Omega)\) is continuously Fréchet-differentiable.

(ii) The operator \((y, u, w) \in Y \times L^2(\Omega) \times W \mapsto G(y, u, w) \in L^p(\Omega)\) is locally Lipschitz-continuous.
GLOBALIZATION OF A SEMISMOOTH NEWTON METHOD

| $k$ | $\alpha_k$ | $\|f(z_k)\|_{Z^*}$ | $\|s_k\|_Z$ |
|-----|-------------|----------------|------------|
| 0   | -           | 5.4311E-02     | -          |
| 1   | 9.76563E-04| 5.43015E-02    | 5.00085E+00|
| 2   | 3.12500E-02| 5.36304E-02    | 1.82556E+00|
| 3   | 5.00000E-01| 2.91839E-02    | 1.55585E+00|
| 4   | 6.25000E-02| 2.75202E-02    | 3.87423E-01|
| 16  | -           | 1.65715E-02    | 2.48095E-02|
| 17  | 0.50000E+00 | 1.38976E-02    | 1.28644E-02|
| 18  | 1.00000E+00 | 1.24060E-02    | 6.81858E-03|
| 19  | 1.00000E+00 | 9.44693E-03    | 1.63072E-03|
| 20  | 1.00000E+00 | 5.60965E-06    | 4.47294E-05|
| 21  | 1.00000E+00 | 2.27743E-15    | 1.57318E-11|

Table 1. Iteration history for the solution of problem (P2) for $h = 1/256$. Step size $\alpha_k$, norm $\|f(z_k)\|_{Z^*}$ and norm of the search direction $\|s_k\|_Z$ for the $k$-th iterate. These numerical results exhibit the superlinear convergence.

As already mentioned in Section 2, a generalized differential of (15) is not given naturally. Motivated by the sum and chain rule in finite dimensions in [22], the set-valued mapping $\partial C f : Y \times L^2(\Omega) \times W \rightarrow L(Y \times L^2(\Omega) \times W, Y^* \times L^2(\Omega) \times W^*)$ with

$$\partial C f := \left\{ M \in L(Y \times L^2(\Omega) \times W, Y^* \times L^2(\Omega) \times W^*) : M(y, u, w) = \begin{bmatrix} L_{yy}(y, u, w) & L_{yu}(y, u, w) & E_y(y, u)^* \\ \gamma D G_y(y, u, w) & I + \gamma D G_u(y, u, w) & \gamma D G_w(y, u, w) \\ E_u(y, u) & 0 & 0 \end{bmatrix} , D \in L^\infty(\Omega), D(x) \in \partial C P_{u,ad}(\gamma G(y, u, w)(x)) \text{ in } \Omega \right\} \tag{16}$$

is used as a generalized differential for [15]. Here the subscript “C” emphasizes the close relation to Qi’s C-subdifferential in finite dimensions.

**Theorem 4.3** (Semismoothness of $f$). Let Assumption 4.2 hold and choose $\gamma$ in (14) as $\gamma = 1/\lambda$. Then $f : Y \times L^2(\Omega) \times W \rightarrow Y^* \times L^2(\Omega) \times W^*$ in (15) is locally Lipschitz continuous and $\partial C f$-semismooth.

For a proof of Th. 4.3 the reader is referred to [22, Th. 5.21]. Therefore $f$ as in (15) meets the requirements needed for Algorithms 2.2 and 3.3. The merit function in (3) for problem (P1) is thus given as

$$\Theta(z) := \frac{1}{2} \|f(z)\|_{Z^*}^2 = \frac{1}{2} \|L_y(y, u, w)\|_{Y^*}^2 + \frac{1}{2} \|\Pi(y, u, w)\|_{L^2(\Omega)}^2 + \frac{1}{2} \|E(y, u)\|_{W^*}^2 . \tag{17}$$

5. **Numerical examples.** In this section we start with a numerical example for the application of Algorithm 3.3 to problems of the type of (P1).
5.1. Semilinear elliptic PDE with control constraints. For simplicity we restrict ourselves at first to the following control-constrained optimal control problem subject to a semilinear elliptic PDE that was taken from [10] Ch. 6 and is motivated by an application in superconductivity:

Minimize \( J(y, u) := \frac{1}{2} \int_{\Omega} (y(x) - y_d(x))^2 \, dx + \frac{\lambda}{2} \int_{\Omega} u(x)^2 \, dx \)

w. r. t. \((y, u) \in H^1_0(\Omega) \times L^2(\Omega)\)

subject to \( E(y, u) := -\Delta y + y^3 + y - u = 0 \) in \( \Omega := (0,1)^2 \),
\( u \in U_{ad} := \{ u \in L^2(\Omega) : -4 \leq u(x) \leq 0 \text{ a.e. in } \Omega \} \).

The desired state is \( y_d := \frac{1}{8} \sin(2\pi x_1) \sin(2\pi x_2) \exp(2x_1) \), the Tikhonov parameter is chosen as \( \lambda := 10^{-3} \). We convince ourselves that Assumption 4.2 is fulfilled for this example, where \( Y = W = H^1_0(\Omega), Y^* = W^* = H^{-1}(\Omega) \) and \( L_{\text{reg}}(y, u, w) = \lambda u + w \) yielding that \( G(y, u, w) = w \) has the required regularity. Hence \( Z^* = H^{-1}(\Omega) \times L^2(\Omega) \times H^{-1}(\Omega) \).

It is not clear, how to check a priori whether Assumption 2.3 and Assumption 3.1 (a) hold in general in an infinite dimensional function space.

However, for both our examples the well-posedness of the linear operator \( M \) and the optimal control problem in a function space setting can be shown by means of the Lax-Milgram theorem (see [10] Sect. 4 & 5 and [9] Sect. 3, respectively), exploiting a sufficient second-order condition [3] for the semilinear state equation. Otherwise, we can see within our numerics whether Assumption 2.3 holds.

We verify Assumption 3.1 (a) for our undiscretized example, assuming that the Tikhonov parameter \( \lambda \) is sufficiently large. In the matrix \( M \) we have only the non-differentiable component \( \gamma DG_w(y, u, w) \) (see [10]), in which in our example \( G(y, u, w) = w \). Thus, exploiting that \( E_y \) depends continuously on \( z \) in our example and \( \alpha_k \leq 1 \), we estimate

\[
\lim_{\rho \downarrow 0} \| (M_{\rho} - M(z_k)) s_k \| Z^* \\
\leq \gamma \lim_{\rho \downarrow 0} \| \partial_C P_{U_{ad}}(\gamma(w_k + \rho s_{k,3})) - \partial_C P_{U_{ad}}(\gamma w_k) \| L(Z, Z^*) s_k 3 \| Z \\
\leq \frac{1}{\lambda} C_{M^{-1}} \| f_k \| Z^*,
\]

where we use (2) in the last estimate. In case of \( \| \partial_C P_{U_{ad}}(\gamma(w_k + \rho s_{k,3})) - \partial_C P_{U_{ad}}(\gamma(w_k)) \| L(Z, Z^*) \in (0, 1) \) (or \( \in [0, 1] \) alternatively), it may happen that \( M_{\rho} \) can only be chosen such that \( \partial_C P_{U_{ad}}(\gamma w_k) = 0 \) (or \( = 1 \) otherwise) at a point of differentiability. In the latter situation \( \| \partial_C P_{U_{ad}}(\gamma w_k + \rho s_{k,3}) - \partial_C P_{U_{ad}}(\gamma w_k) \| L(Z, Z^*) \in (0, 1] \) and this does not vanish as \( \rho \) tends to zero from above. However the constant \( C_{M^{-1}} \) is uniform and for sufficiently large \( \lambda = 1/\gamma \), we may guarantee Assumption 3.1 (a). This estimate is also supported by numerical evidence, where we observe always descent directions, unless we reduce \( \lambda \) significantly.

For the approximation of the state as well as the control, piecewise linear finite elements are used. In this context, let \( T_h \) be a triangulation of \( \Omega \) for which the usual regularity assumptions hold. Furthermore, let \( h := \max_{T \in T_h} \text{diam}(T) \) denote the size of the mesh. For details we refer to [2]. Analogously for a finite element discretization the well-posedness of the Newton step and the discretized problem
can be derived for both examples \cite{10, 9}, together with a mesh independence result for the local algorithm under the further assumption that the set where the strict complementarity is violated has measure zero. As a starting point we use $z_0 = 0$.

Our stopping condition is $\|f(z_k)\|_Z \leq 10^{-10}$. Table 1 shows the iteration history for $h = 1/256$. In iterations 1 to 17 the step sizes are chosen with $\alpha_k < 1$ due to the Armijo line search. As predicted by Theorem 3.5 the acceptance of $\alpha_k = 1$ by the Armijo line search leads to the transition of Algorithm 3.3 into Algorithm 2.2. The expected superlinear convergence is confirmed by the evolution of the values $\|f(z_k)\|_Z$ and $\|s_k\|_Z$. We observe that $\alpha_k$ is uniformly bounded below and corresponding to Th. 3.4 (a) and Th. 3.5 the numerical accumulation point is a solution. Figure 1 shows the discrete solution of the optimal state $y^h$ and the optimal control $u^h$ for $h = 1/64$.

5.2. Semilinear elliptic PDE with state constraints. Problem (P2) has illustrated the performance of our algorithm. However, problems of the type of (P2) have been solved so far, see e.g. \cite{10, 12, 13}. Hence we close with an example for a semilinear elliptic PDE with state constraints \cite[Ch. 6]{9} where, to our knowledge, previous globalization strategies cannot be applied:

$$\text{Minimize } J(y, u) := \frac{1}{2} \|y - y_d\|^2_{L^2(\Omega)} + \frac{\lambda}{2} \|u\|^2_{L^2(\Omega)}$$

w. r. t. $(y, u) \in H^1_0(\Omega) \times L^2(\Omega)$

subject to $E(y, u) := -\Delta y + y^3 \exp(10y) + y - u = 0$ in $\Omega := (0, 1)^2$,
$$-10^{-2} \leq y(x) \leq 0 \text{ a.e. in } \Omega.$$

We would like to track the state $y_d := \frac{1}{2} \cos(\pi x_1) \cos(\pi x_2) \exp(x_1)$. As Tikhonov parameter we consider $\lambda := 10^{-4}$. We use a Lavrentiev regularization of the state-constraint in (P3)

$$-10^{-2} \leq \epsilon u + y \leq 0,$$

where $\epsilon > 0$ is a sufficiently small parameter, yielding a mixed control-state constraint.

Note that we consider the same objective functional as in (P2), but in addition to the regularized state constraints, we allow for a term $y^3 \exp(10y)$ instead of $y^3$.

**Figure 1.** Discrete solution of (P2) for $h = 1/64$. Left-hand side: Optimal state $y^h(x_1, x_2)$ on $x_3$ axis vs. $x_1$ and $x_2$. Right-hand side: Optimal control $u^h(x_1, x_2)$ on $x_3$ axis vs. $x_1$ and $x_2$. 
Table 2. Iteration history for the solution of problem (P3) for $h = 1/128$. Step size $\alpha_k$, norm $\|f(z_k)\|_{Z^*}$ and norm of the search direction $\|s_k\|_{Z^*}$ for the $k$-th iterate. We observe transition to local superlinear convergence.

| $k$ | $\alpha$ | $\|f(z_k)\|_{Z^*}$ | $\|s_k\|_{Z^*}$ |
|-----|--------|----------------|----------------|
| 0   | -      | 7.59736E+05     | -              |
| 1   | 1.00000E+00 | 1.14024E+05 | 1.93458E+03 |
| 2   | 1.00000E+00 | 3.61620E+04 | 7.83427E+02 |
| 3   | 1.00000E+00 | 1.59280E+04 | 1.62132E+03 |
| ... | ...    | ...            | ...            |
| 9   | 2.50000E-01 | 3.03640E-02 | 1.48894E-01 |
| 10  | 1.00000E+00 | 9.69843E-03 | 3.23249E-02 |
| 11  | 1.00000E+00 | 2.42234E-05 | 9.90030E-06 |
| 12  | 1.00000E+00 | 3.15754E-06 | 2.56947E-07 |
| 13  | 1.00000E+00 | 1.14583E-07 | 1.59876E-09 |
| 14  | 1.00000E+00 | 1.70426E-13 | 5.17916e-13 |

Figure 2. Discrete solution of (P3) for $h = 1/32$. Left-hand side: Optimal state $y^h(x_1, x_2)$ on $x_3$ axis vs. $x_1$ and $x_2$. Right-hand side: Optimal control $u^h(x_1, x_2)$ on $x_3$ axis vs. $x_1$ and $x_2$.

in the PDE. However, the mixed control-state constraint yields a slightly different setting as in Sect. 4. Concerning the validation of Assumption 2.3 and Assumption 3.1 (a) see the discussion for the first example. However, with respect to Assumption 3.1 (a) we have to keep in mind that the Lavrentiev regularization parameter $\epsilon$ and the Tikhonov parameter $\lambda$ should not be chosen independently.

Again, we work with the stopping condition $\|f(z_k)\|_{Z^*} \leq 10^{-10}$. For numerical results for $\epsilon = 10^{-3}$ as Lavrentiev regularization parameter, see Table 2 and Fig. 2.

Acknowledgments. The authors acknowledge gratefully the useful suggestions and comments of Prof. M. Ulbrich, Prof. A. Schiela, and an unknown referee.
REFERENCES

[1] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer Series in Operations Research, Springer, New York, 2000.

[2] E. Casas, ‘Using piecewise linear functions in the numerical approximation of semilinear elliptic control problems’, *Adv. Comput. Math.*, 26 (2007), 137–153.

[3] E. Casas and F. Tröltzsch, ‘Second order optimality conditions and their role in PDE control’, *Jahresber. Dtsh. Math.-Ver.*, 117 (2015), 3–44.

[4] X. Chen, Z. Nashed and L. Qi, ‘Smoothing methods and semismooth methods for nondifferentiable operator equations’, *SIAM J. Numer. Anal.*, 38 (2000), 1200–1216.

[5] R. Correa and A. Joffre, ‘Tangentially continuous directional derivatives in nonsmooth analysis’, *J. Optim. Theory Appl.*, 61 (1989), 1–21.

[6] M. Gerdts, ‘Global convergence of a nonsmooth Newton’s method for control-state constrained optimal control problems’, *SIAM J. Optim.*, 19 (2008), 326–350; M. Gerdts and B. Hüppling, ‘Erratum: Global convergence of a nonsmooth Newton’s method for control-state constrained optimal control problems’, Technical report, Universität der Bundeswehr München, Neubiberg (2011). Available online: http://www.unibw.de/lrt1/gerdts/publikationen/erratum-siam-19-1-2008-326-350-full.pdf.

[7] M. Gerdts and M. Kunkel, ‘A nonsmooth Newton’s method for discretized optimal control problems with state and control constraints’, *I. Ind. Manag. Opt.*, 4 (2008), 247–270.

[8] M. Hintermüller, K. Ito and K. Kunisch, ‘The primal–dual active set strategy as a semismooth Newton method’, *SIAM J. Optim.*, 13 (2003), 865–888.

[9] M. Hintermüller, F. Tröltzsch and I. Yousept, ‘Mesh-independence of semismooth Newton methods for Lavrentiev-regularized state constrained nonlinear optimal control problems’, *Numer. Math.*, 108 (2008), 571–603.

[10] M. Hinze and M. Ulbrich, ‘A mesh-independence result for semismooth Newton methods’, *Math. Program.*, Ser. B, 101 (2004), 151–184.

[11] M. Hinze, R. Pinnau, M. Ulbrich and S. Ulbrich, *Optimization with PDE Constraints*, Math. Modelling: Theory and Applications, Vol. 23, Springer, New York, 2009.

[12] M. Hinze and M. Vierling, ‘The semi-smooth Newton method for variationally discretized control constrained elliptic optimal control problems; implementation, convergence and globalization’, *Optim. Methods Softw.*, 27 (2012), 933–950.

[13] M. Hinze and M. Vierling, ‘A globalized semi-smooth Newton method for variational discretization of control constrained elliptic optimal control problems’, in *Constrained Optimization and Optimal Control for Partial Differential Equations* (eds. G. Leugering et al.), Int. Ser. Numer. Math., 160, Birkhäuser/Springer, Basel, 2012, 171–182.

[14] S. Horn, ‘Fixpunktkriterien für PDE-restringierte Optimierungsprobleme’, Master thesis, Universität der Bundeswehr München, Neubiberg, 2012.

[15] K. Ito and K. Kunisch, ‘Applications of semi-smooth Newton methods to variational inequalities in Control of Coupled Partial Differential Equations’ (eds. K. Kunisch, G. Leugering, J. Sprekels and F. Tröltzsch), Internat. Ser. Numer. Math., 155, Birkhäuser, Basel, 2007, 175–192.

[16] K. Ito and K. Kunisch, ‘On a semi-smooth Newton method and its globalization’, *Math. Program.*, Ser. A, 118 (2009), 347–370.

[17] A. Kröner, K. Kunisch and B. Vexler, ‘Semismooth Newton methods for optimal control of the wave equation with control constraints’, *SIAM J. Control Optim.*, 49 (2011), 830–858.

[18] B. Kummer, ‘Newton’s method for non-differentiable functions’, in *Advances in Mathematical Optimization* (eds. J. Guddat, et al.), Math. Res., 45, Akademie-Verlag, Berlin, 1988, 114–125.

[19] B. Kummer, ‘Newton’s method based on generalized derivatives for nonsmooth functions: Convergence analysis’, in *Advances in Optimization (Lambrecht 1991)* (eds. W. Oettli and D. Pallaschke), Lecture Notes in Econ. and Math. Systems, 382, Springer, Berlin, 1992, 171–194.

[20] A. Rösch and D. Wachsmuth, ‘Semi-smooth Newton’s method for an optimal control problem with control and mixed control-state constraints’, *Optim. Methods Softw.*, 26 (2011), 169–186.

[21] A. Schiela, ‘A simplified approach to semismooth Newton methods in function space’, *SIAM J. Optim.*, 19 (2008), 1417–1432.

[22] M. Ulbrich, *Nonsmooth Newton-like Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*, Habilitation thesis, Technical University Munich, Munich, 2001.
[23] M. Ulbrich, Semismooth Newton methods for operator equations in function spaces, *SIAM J. Optim.*, 13 (2003), 805–842.

[24] M. Ulbrich, *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*, MOS-SIAM Series on Optimization, 11, SIAM/MOS, Philadelphia, 2011.

Received May 2014; 1st revision April 2015; 2nd revision August 2015.

E-mail address: matthias.gerdts@unibw.de
E-mail address: stefanhorn87@aol.com
E-mail address: sven-joachim.kimmerle@unibw.de