Hamburger Beiträge
zur Angewandten Mathematik

Global convergence of a nonsmooth Newton's method for control-state constrained optimal control problems

Matthias Gerdts

Nr. 2007-04
February 2007

GLOBAL CONVERGENCE OF A NONSMOOTH NEWTON'S METHOD FOR CONTROL-STATE CONSTRAINED OPTIMAL CONTROL PROBLEMS

MATTHIAS GERDTS
Department of Mathematics
University of Hamburg
Bundesstraße 55
20146 Hamburg, Germany

Abstract. We investigate a nonsmooth Newton's method for the numerical solution of optimal control problems subject to mixed control-state constraints. The necessary conditions are stated in terms of a local minimum principle. By use of the Fischer-Burmeister function the local minimum principle is transformed into an equivalent nonlinear and nonsmooth equation in appropriate Banach spaces. This nonlinear and nonsmooth equation is solved by a nonsmooth Newton's method. We prove the global convergence and the locally quadratic convergence under certain regularity conditions. The globalized method is based on the minimization of the squared residual norm. Numerical examples for the Rayleigh problem conclude the article.

1. **Introduction.** We consider the following optimal control problem subject to mixed control-state constraints:

$$
(OCP) \qquad \begin{array}{ll}
\text{Minimize} & \displaystyle\int_0^1 f_0(x(t),u(t))dt \\
\text{w.r.t.} & x \in W^{1,\infty}([0,1],\mathbb{R}^{n_x}), u \in L^\infty([0,1],\mathbb{R}^{n_u}), \\
\text{s.t.} & x'(t) = f(x(t),u(t)) \text{ a.e. in } [0,1], \\
& \psi(x(0),x(1)) = 0, \\
& c(x(t),u(t)) \leq 0 \text{ a.e. in } [0,1].
\end{array}
$$

Without loss of generality the discussion is restricted to autonomous problems on the fixed time interval $[0,1]$. The functions $f_0 : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}$, $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$, $\psi : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_\psi}$, $c : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_c}$, are supposed to be at least twice continuously differentiable w.r.t. to all arguments. As usual, the Banach space $L^\infty([0,1],\mathbb{R}^n)$ consists of all measurable functions $h : [0,1] \to \mathbb{R}^n$ with

$$\|h\|_\infty := \operatorname*{ess\,sup}_{0\leq t\leq 1} \|h(t)\| < \infty,$$

where $\|\cdot\|$ denotes the Euclidian norm on $\mathbb{R}^n$. The Banach space $W^{1,\infty}([0,1],\mathbb{R}^n)$ consists of all absolutely continuous functions $h : [0,1] \to \mathbb{R}^n$ with

$$\|h\|_{1,\infty} := \max\{\|h\|_\infty, \|h'\|_\infty\} < \infty.$$

Several approaches towards the numerical solution of OCP have been investigated in the literature. The so-called direct discretization method is based on a discretization of the

2000 *Mathematics Subject Classification.* 49J15, 49J52, 49M15 .
Key words and phrases. optimal control, nonsmooth Newton's method, control-state constraints, global convergence.

.

infinite dimensional optimal control problem and leads to a finite dimensional nonlinear program, cf., e.g., Gerdts [5]. The latter can be solved numerically by suitable programming methods such as, e.g., sequential quadratic programming. The direct discretization method turns out to be very robust in practice. Nevertheless, the computational effort grows at a nonlinear rate with the number of grid points used for discretization.

The so-called indirect method for optimal control problems attempts to satisfy the necessary conditions that are provided by the well-known minimum principle numerically, cf., e.g. Oberle and Grimm [14]. The exploitation of the minimum principle leads to a nonlinear multi-point boundary value problem that has to be solved. Although the indirect method usually leads to the most accurate solutions, it suffers from the drawback that it requires a good initial guess in order to convergence. One crucial task is to estimate the sequence of active and inactive intervals of the control-state constraint.

Our intention is to analyze the local and global convergence properties of an alternative method – the nonsmooth Newton's method. The method is based on a nonsmooth reformulation of the necessary optimality conditions and it was introduced for the problem class OCP in Gerdts [7]. A brief outline of the essential ideas of the algorithm is as follows. The reformulation of the necessary conditions leads to the nonsmooth equation

$$F(z) = 0, \qquad F : Z \to Y,$$

where Z and Y are appropriate Banach spaces. Application of the globalized nonsmooth Newton's method generates sequences $\{z^k\}$, $\{d^k\}$ and $\{\alpha_k\}$ related by the iteration

$$z^{k+1} = z^k + \alpha_k d^k, \qquad k = 0, 1, 2, \ldots.$$

Herein, the search direction d^k is the solution of the linear operator equation $V_k(d^k) = -F(z^k)$ and the step length $\alpha_k > 0$ is determined by a line-search procedure of Armijo's type for a suitably defined merit function. The linear operator V_k is chosen from an appropriately defined generalized Jacobian $\partial_* F(z^k)$.

The nonsmooth Newton's method was investigated in finite dimensions amongst others by Qi [15] and Qi and Sun [16]. Extensions to infinite spaces can be found in Kummer [9, 10], Chen et al. [1], and Ulbrich [17, 18]. Our approach follows the general framework of Ulbrich [17, 18].

The paper is organized as follows. Section 2 introduces the nonsmooth Newton's method and establishes the locally superlinear resp. quadratic convergence under comparatively mild assumptions. In Section 3 details of the computation of the search direction are shown. It turns out that the search direction solves a linear boundary value problem with a differential-algebraic equation (DAE). If a certain operator is invertible, the so-called index of the DAE is one and the DAE can be transformed easily into an ordinary differential equation. A sufficient condition for the existence of the inverse operator is provided. Section 4 analyzes the global convergence properties of the nonsmooth Newton's method. Finally, numerical illustrations are presented in Section 5.

2. **Local Convergence of the Nonsmooth Newton's Method.** The (augmented) Hamilton function $H : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_c} \to \mathbb{R}$ is defined by

$$H(x, u, \lambda, \eta) := f_0(x, u) + \lambda^\top f(x, u) + \eta^\top c(x, u).$$

We summarize the well-known minimum principle for OCP, cf. Neustadt [13] and Gerdts [6], Th. 5.2, p. 19. Let (x_*, u_*) be a (weak) local minimum of OCP, let

$$\operatorname{rank}\left(c_u'(x_*(t), u_*(t))\right) = n_c$$

hold a.e. in $[0,1]$, and let the Mangasarian-Fromowitz constraint qualification hold, cf. Gerdts [6], Th. 6.3, p. 26.

Then there exist Lagrange multipliers $\lambda_* \in W^{1,\infty}([0,1],\mathbb{R}^{n_x})$, $\eta_* \in L^\infty([0,1],\mathbb{R}^{n_c})$, and $\sigma_* \in \mathbb{R}^{n_\psi}$ with

$$\begin{aligned} x_*'(t) - f(x_*(t),u_*(t)) &= 0, & (1)\\ \lambda_*'(t) + H_x'(x_*(t),u_*(t),\lambda_*(t),\eta_*(t))^\top &= 0, & (2)\\ \psi(x_*(0),x_*(1)) &= 0, & (3)\\ \lambda_*(0) + \psi_{x_0}'(x_*(0),x_*(1))^\top\sigma_* &= 0, & (4)\\ \lambda_*(1) - \psi_{x_1}'(x_*(0),x_*(1))^\top\sigma_* &= 0, & (5)\\ H_u'(x_*(t),u_*(t),\lambda_*(t),\eta_*(t))^\top &= 0. & (6)\end{aligned}$$

Furthermore, the complementarity conditions hold a.e. in $[0,1]$:

$$\eta_*(t) \geq 0, \qquad c(x_*(t),u_*(t)) \leq 0, \qquad \eta_*(t)^\top c(x_*(t),u_*(t)) = 0. \tag{7}$$

Unfortunately, these necessary conditions are not directly solvable for $(x_*,u_*,\lambda_*,\eta_*,\sigma_*)$ owing to the complementarity conditions. Therefore, the subsequent considerations aim at the reformulation of this set of equalities and inequalities as an equivalent system of equations, which will be solved by a generalized version of Newton's method. Throughout the rest of the paper for brevity we will use the notation $f[t]$ for $f(x(t),u(t))$.

The convex and locally Lipschitz continuous Fischer-Burmeister function $\varphi : \mathbb{R}^2 \to \mathbb{R}$ is defined by

$$\varphi(a,b) := \sqrt{a^2+b^2} - a - b, \tag{8}$$

cf. Fischer [3]. The Fischer-Burmeister function has the nice property that $\varphi(a,b) = 0$ holds if and only if $a,b \geq 0$ and $ab = 0$. Hence, the complementarity conditions (7) are equivalent with the equality

$$\varphi(-c_i(x_*(t),u_*(t)),\eta_{i,*}(t)) = 0, \qquad i = 1,\ldots,n_c,$$

that has to hold almost everywhere in $[0,1]$. Rather than working with the derivative of φ, which does not exist at the origin, we will work with Clarke's generalized Jacobian of φ:

$$\partial\varphi(a,b) = \begin{cases} \left\{\left(\dfrac{a}{\sqrt{a^2+b^2}} - 1, \dfrac{b}{\sqrt{a^2+b^2}} - 1\right)\right\}, & \text{if } (a,b) \neq (0,0),\\ \left\{(s,r) \;\middle|\; (s+1)^2 + (r+1)^2 \leq 1\right\}, & \text{if } (a,b) = (0,0).\end{cases}$$

Notice, that $\partial\varphi(a,b)$ is a nonempty, convex and compact set. Let the Banach spaces

$$\begin{aligned} Z &= W^{1,\infty}([0,1],\mathbb{R}^{n_x}) \times L^\infty([0,1],\mathbb{R}^{n_u}) \times L^\infty([0,1],\mathbb{R}^{n_x}) \times L^\infty([0,1],\mathbb{R}^{n_c}) \times \mathbb{R}^{n_\psi},\\ Y_1 &= L^\infty([0,1],\mathbb{R}^{n_x}) \times L^\infty([0,1],\mathbb{R}^{n_x}) \times \mathbb{R}^{n_\psi} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times L^\infty([0,1],\mathbb{R}^{n_u}),\\ Y_2 &= L^\infty([0,1],\mathbb{R}^{n_c})\end{aligned}$$

be equipped with the maximum norm for product spaces and $z_* = (x_*,u_*,\lambda_*,\eta_*,\sigma_*)$. Then, the necessary conditions (1)-(7) are equivalent with the nonlinear equation

$$F(z_*) = \begin{pmatrix} F_1(z_*)\\ F_2(z_*)\end{pmatrix} = 0, \tag{9}$$

where $F_1 : Z \to Y_1$ and $F_2 : Z \to Y_2$ denote the smooth and the nonsmooth part of $F : Z \to Y := Y_1 \times Y_2$, respectively:

$$F_1(z)(\cdot) := \begin{pmatrix} x'(\cdot) - f(x(\cdot), u(\cdot)) \\ \lambda'(\cdot) + H'_x(x(\cdot), u(\cdot), \lambda(\cdot), \eta(\cdot))^\top \\ \psi(x(0), x(1)) \\ \lambda(0) + \psi'_{x_0}(x(0), x(1))^\top \sigma \\ \lambda(1) - \psi'_{x_1}(x(0), x(1))^\top \sigma \\ H'_u(x(\cdot), u(\cdot), \lambda(\cdot), \eta(\cdot))^\top \end{pmatrix}, \quad F_2(z)(\cdot) := \omega(z(\cdot)), \tag{10}$$

where $\omega = (\omega_1, \ldots, \omega_{n_c})^\top : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_c} \times \mathbb{R}^{n_\psi} \to \mathbb{R}^{n_c}$ and

$$\omega_i(\bar{x}, \bar{u}, \bar{\lambda}, \bar{\eta}, \bar{\sigma}) := \varphi(-c_i(\bar{x}, \bar{u}), \bar{\eta}_i), \qquad i = 1, \ldots, n_c. \tag{11}$$

The standard approach to solve (9) numerically would be to apply the classical Newton's method. Unfortunately, the derivative $F'(z^k)$ does not exist since the component F_2 is not differentiable. Hence, we have to find a substitute for the derivative F' in the classical Newton's method. In finite dimensional spaces, such a substitute for locally Lipschitz continuous functions may be chosen from the generalized Jacobian of F defined by

$$\partial F(z) := \operatorname{co} \left\{ V \;\middle|\; V = \lim_{\substack{z_i \in D_F \\ z_i \to z}} F'(z_i) \right\},$$

where D_F denotes the set of points where F is differentiable, cf. Clarke [2]. However, in infinite dimensional spaces it is more difficult to define an appropriate generalized Jacobian since locally Lipschitz continuous functions in general are not differentiable almost everywhere. Motivated by the chain rule in finite dimensions we define the point to set mapping $\partial_* F : Z \rightrightarrows \mathcal{L}(Z, Y)$ according to

$$\partial_* F(z^k)(z) := \left\{ \begin{pmatrix} F_1'(z^k)(z) \\ -S\,(c'_x[\cdot]x + c'_u[\cdot]u) + R\eta \end{pmatrix} \;\middle|\; \begin{array}{l} S = diag(s_1, \ldots, s_{n_c}), \\ R = diag(r_1, \ldots, r_{n_c}), \\ (s_i, r_i) \in \partial\varphi[\cdot] \text{ a.e.}, \\ s_i(\cdot), r_i(\cdot) \text{ measurable} \end{array} \right\}$$

and use this set as a generalized Jacobian. The same idea was introduced earlier in Ulbrich [17], Def. 3.35, p. 47. Notice, that the first component F_1 of F in (10) is continuously Fréchet-differentiable with

$$F_1'(z^k)(z) = \begin{pmatrix} x'(\cdot) - f'_x[\cdot]x(\cdot) - f'_u[\cdot]u(\cdot) \\ \lambda'(\cdot) + H''_{xx}[\cdot]x(\cdot) + H''_{xu}[\cdot]u(\cdot) + H''_{x\lambda}[\cdot]\lambda(\cdot) + H''_{x\eta}[\cdot]\eta(\cdot) \\ \psi'_{x_0}x(0) + \psi'_{x_1}x(1) \\ \lambda(0) + \psi''_{x_0x_0}(\sigma, x(0)) + \psi''_{x_0x_1}(\sigma, x(1)) + \left(\psi'_{x_0}\right)^\top \sigma \\ \lambda(1) - \psi''_{x_1x_0}(\sigma, x(0)) - \psi''_{x_1x_1}(\sigma, x(1)) - \left(\psi'_{x_1}\right)^\top \sigma \\ H''_{ux}[\cdot]x(\cdot) + H''_{uu}[\cdot]u(\cdot) + H''_{u\lambda}[\cdot]\lambda(\cdot) + H''_{u\eta}[\cdot]\eta(\cdot) \end{pmatrix},$$

provided that the functions f_0, f, c, ψ are twice continuously differentiable w.r.t. all arguments. All functions are evaluated at $z^k = (x^k, u^k, \lambda^k, \eta^k, \sigma^k) \in Z$.

Replacing the non-existing Jacobian F' in the classical Newton's method by the generalized Jacobian $\partial_* F(z^k)$ leads to the following algorithm.

Algorithm 1 (Local Nonsmooth Newton's Method)**.**

(0) Choose z^0.
(1) If some stopping criterion is satisfied, stop.

(2) Choose an arbitrary $V_k \in \partial_* F(z^k)$ and compute the search direction d^k from the linear equation

$$V_k(d^k) = -F(z^k).$$

(3) Set $z^{k+1} = z^k + d^k$, $k = k+1$, and goto (1).

The assumptions needed to prove local convergence of the method are similar to those in Qi [15], Qi and Sun [16], Jiang [8], and Ulbrich [17]. $\partial_* F(z)$ is called nonsingular if for every $V \in \partial_* F(z)$ the inverse operator V^{-1} exists and if it is linear and bounded, i.e. $V^{-1} \in \mathcal{L}(Y, Z)$.

Theorem 1. *Let z_* be a zero of F. Suppose that there exist constants $\Delta > 0$ and $C > 0$ such that for every $z \in U_\Delta(z_*)$ the generalized Jacobian $\partial_* F(z)$ is nonsingular and $\|V^{-1}\|_{\mathcal{L}(Y,Z)} \leq C$ for every $V \in \partial_* F(z)$.*

(i) *Let*

$$\|F(z) - F(z_*) - V(z - z_*)\|_Y = o(\|z - z_*\|_Z) \qquad \forall V \in \partial_* F(z) \tag{12}$$

as $\|z - z_*\|_Z \to 0$. Then, for z^0 sufficiently close to z_* the nonsmooth Newton's method converges superlinearly to z_*.

(ii) *Let*

$$\|F(z) - F(z_*) - V(z - z_*)\|_Y = \mathcal{O}(\|z - z_*\|_Z^{1+p}) \qquad \forall V \in \partial_* F(z) \tag{13}$$

as $\|z - z_*\|_Z \to 0$. Then, for z^0 sufficiently close to z_* the nonsmooth Newton's method converges at order $1 + p$ to z_*.

Furthermore, if $F(z^k) \neq 0$ for all k, then the residual values converge superlinearly:

$$\lim_{k \to \infty} \frac{\|F(z^{k+1})\|_Y}{\|F(z^k)\|_Y} = 0.$$

Proof. Due to the first assumption, the algorithm is well-defined in some neighborhood of z_*. It holds

$$V_k(z^{k+1} - z_*) = V_k(z^k + d^k - z_*) = V_k(z^k - z_*) + V_k d^k = V_k(z^k - z_*) - F(z^k) + F(z_*).$$

The assertions in (i) and (ii) follow from

$$\begin{aligned}
\|z^{k+1} - z_*\|_Z &= \|V_k^{-1}\left(V_k(z^k - z_*) - F(z^k) + F(z_*)\right)\|_Y \\
&\leq \|V_k^{-1}\|_{\mathcal{L}(Y,Z)} \cdot \|F(z^k) - F(z_*) - V_k(z^k - z_*)\|_Y \\
&\leq C \cdot \|F(z^k) - F(z_*) - V_k(z^k - z_*)\|_Y \\
&= \begin{cases} o(\|z^k - z_*\|_Z), & \text{in case (i)}, \\ \mathcal{O}(\|z^k - z_*\|_Z^{1+p}), & \text{in case (ii)}. \end{cases}
\end{aligned} \tag{14}$$

Let $\varepsilon > 0$ be arbitrary. According to Equation (14) there exists $\delta > 0$ with

$$\|z^{k+1} - z_*\|_Z \leq \varepsilon \|z^k - z_*\|_Z \quad \text{whenever} \quad \|z^k - z_*\|_Z \leq \delta.$$

Notice, that for any $\delta > 0$ there exists some $k_0(\delta)$ such that $\|z^k - z_*\| \leq \delta$ for every $k \geq k_0(\delta)$ since z^k converges to z_*. By the local Lipschitz continuity of F we get

$$\|F(z^{k+1})\|_Y = \|F(z^{k+1}) - F(z_*)\|_Y \leq L\|z^{k+1} - z_*\|_Z \leq L\varepsilon \|z^k - z_*\|_Z$$

locally around z_* and the Newton iteration implies

$$\|z^{k+1} - z^k\|_Z \;\leq\; \|V_k^{-1}\|_{\mathcal{L}(Y,Z)} \cdot \|F(z^k)\|_Y \leq C\|F(z^k)\|_Y.$$

Thus,

$$\begin{aligned}\|z^k - z_*\|_Z &\leq \|z^{k+1} - z^k\|_Z + \|z^{k+1} - z_*\|_Z \\ &\leq C\|F(z^k)\|_Y + \|z^{k+1} - z_*\|_Z \\ &\leq C\|F(z^k)\|_Y + \varepsilon\|z^k - z_*\|_Z\end{aligned}$$

and

$$\|z^k - z_*\|_Z \leq \frac{C}{1-\varepsilon}\|F(z^k)\|_Y.$$

Finally,

$$\|F(z^{k+1})\|_Y \leq L\varepsilon\|z^k - z_*\|_Z \leq \frac{L\varepsilon C}{1-\varepsilon}\|F(z^k)\|_Y.$$

Since $F(z^k) \neq 0$ and ε may be arbitrarily small this shows the last assertion. $\square$

Remark 1.

- The properties (12) and (13) can be written as

$$\begin{aligned}\sup_{V\in\partial_* F(z)} \|F(z) - F(z_*) - V(z - z_*)\|_Y &= o(\|z - z_*\|_Z), \\ \sup_{V\in\partial_* F(z)} \|F(z) - F(z_*) - V(z - z_*)\|_Y &= \mathcal{O}(\|z - z_*\|_Z^{1+p})\end{aligned}$$

  as $\|z - z_*\|_Z \to 0$ and are referred to as semismoothness and p-order semismoothness of F at z_*, cf. Ulbrich [17], Def. 3.1, p. 34.
- It suffices if the assumptions are satisfied for certain elements of $\partial_* F$ provided that only these elements are used in the algorithm. For the upcoming computations we used the element corresponding to the choices

$$\begin{aligned}s_i(t) &= \begin{cases} -1, & \text{if } c_i[t] = 0,\ \eta_i(t) = 0, \\ \frac{-c_i[t]}{\sqrt{c_i[t]^2 + \eta_i(t)^2}} - 1, & \text{otherwise,} \end{cases} \\ r_i(t) &= \begin{cases} 0, & \text{if } c_i[t] = 0,\ \eta_i(t) = 0, \\ \frac{\eta(t)}{\sqrt{c_i[t]^2 + \eta_i(t)^2}} - 1, & \text{otherwise.} \end{cases}\end{aligned}$$

We will show that the conditions (12) and (13) with $p = 1$ hold for F in (9)-(10) under appropriate conditions.

The first component F_1 is continuously Fréchet-differentiable if f_0, f, c, ψ are twice continuously differentiable. The Fréchet-differentiability immediately yields (12) for the component F_1. If the second derivatives of f_0, f, c, ψ are even locally Lipschitz continuous, then F_1' also satisfies a local Lipschitz condition of type

$$\|F_1'(z + d) - F_1'(z)\|_{\mathcal{L}(Z,Y_1)} \leq L\|d\|_Z.$$

Using this property and the mean-value theorem we find

$$\begin{aligned}\|F_1(z + d) - F_1(z) - F_1'(z + d)(d)\|_{Y_1} &\leq \int_0^1 \|(F_1'(z + td) - F_1'(z + d))(d)\|_{Y_1} dt \\ &\leq \int_0^1 \|F_1'(z + td) - F_1'(z + d)\|_{\mathcal{L}(Z,Y_1)} dt \cdot \|d\|_Z \\ &\leq \frac{L}{2}\|d\|_Z^2\end{aligned}$$

and thus (13) with $p = 1$ holds for F_1.

The second component $F_2(z)(t) = \omega(z(t))$ of F in (10) is a superposition operator as in Ulbrich [17], Sec. 3.3, with the difference that F_2 maps from some subset of L^∞ to

L^∞. This allows us to consider the operator F_2 pointwise since $\|z - z_*\|_Z \to 0$ implies $\|z(t) - z_*(t)\| \to 0$ a.e. in $[0,1]$. Let us summarize some well-known results for finite dimensions. The Fischer-Burmeister function $\varphi : \mathbb{R}^2 \to \mathbb{R}$ is 1-order semismooth (and particularly semismooth) according to Fischer [4], Lemma 20. Furthermore, due to a result of Mifflin, the composition $g = g_1 \circ g_2$ of semismooth functions g_1, g_2 is again semismooth, cf. Fischer [4], page 527. Similarly, the composition of 1-order semismooth functions is again 1-order semismooth, cf. Fischer [4], Theorem 19. In particular, continuously differentiable functions are semismooth and functions having a locally Lipschitz continuous first derivative are 1-order semismooth. Consequently, the function ω in (11) is semismooth, if the function c is continuously differentiable. Moreover, ω is 1-order semismooth, if c' is locally Lipschitz continuous. With these remarks the semismoothness and the 1-order semismoothness of the superposition operator $F_2 : Z \to Y_2$ in (10) and (11) is established by the following lemma.

Lemma 1. *Consider the operator*

$$g : L^\infty([0,1], \mathbb{R}^n) \to L^\infty([0,1], \mathbb{R}^m), \qquad z \mapsto g(z)(t) = \omega(z(t)).$$

It holds:

(i) *g is semismooth at z (in the sense of Remark 1), if $\omega : \mathbb{R}^n \to \mathbb{R}^m$ is semismooth at $z(t) \in \mathbb{R}^n$ for a.e. $t \in [0,1]$.*

(ii) *g is p-order semismooth at z (in the sense of Remark 1), if ω is uniformly p-order semismooth at z, i.e. there exists $C_z > 0$ such that for almost every $\bar{z} \in \{z(t) \in \mathbb{R}^n \mid t \in [0,1]\}$ it holds*

$$\max_{V \in \partial\omega(\bar{z}+h)} \|\omega(\bar{z} + h) - \omega(\bar{z}) - Vh\| \le C_z \|h\|^{1+p} \qquad \text{as} \qquad \|h\| \to 0.$$

Proof. Define $\rho : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ by

$$\rho(x,h) := \max_{V \in \partial\omega(x+h)} \|\omega(x+h) - \omega(x) - Vh\|.$$

(i) Owing to the semismoothness of ω at $z(t)$ for a.e. $t \in [0,1]$ it holds

$$a(t) = \frac{\rho(z(t), d(t))}{\|d\|_\infty} = \frac{o(\|d(t)\|)}{\|d\|_\infty} \to 0$$

as $\|d(t)\| \to 0$ for a.e. $t \in [0,1]$. Since $\|d\|_\infty \to 0$ implies $\|d(t)\| \to 0$ a.e., it holds $\|\rho(z(\cdot), d(\cdot))\|_\infty = \|a\|_\infty \cdot \|d\|_\infty = o(\|d\|_\infty)$.

(ii) The uniform p-order semismoothness of ω at z yields

$$\rho(z(t), d(t)) \le C_z \|d(t)\|^{1+p} \le C_z \|d\|_\infty^{1+p}$$

a.e. in $[0,1]$, where C_z does not depend on t. The assertion follows from $\|\rho(z(\cdot), d(\cdot))\|_\infty \le C_z \|d\|_\infty^{1+p}$.

□

Application of the lemma and the previous considerations yield the following result.

Theorem 2. *Let z_* be a zero of F. Suppose that there exist constants $\Delta > 0$ and $C > 0$ such that for every $z \in U_\Delta(z_*)$ the generalized Jacobian $\partial_* F(z)$ is nonsingular and $\|V^{-1}\|_{\mathcal{L}(Y,Z)} \le C$ for every $V \in \partial_* F(z)$.*

The nonsmooth Newton's method converges locally at a superlinear rate, if f_0, f, c, ψ are twice continuously differentiable.

Moreover, the convergence is quadratic, if the second derivatives of f_0, f, c, ψ are locally Lipschitz continuous and if the uniform strict complementarity condition

$$\|\eta_*(t)\| + \|c(x_*(t), u_*(t))\| \ge \alpha \tag{15}$$

is satisfied a.e. in $[0,1]$ for some $\alpha > 0$.

Proof. The first assertion follows from Lemma 1 since the Fischer-Burmeister function is semismooth at every $(a,b)^\top \in \mathbb{R}^2$ and thus ω in (11) is semismooth everywhere provided that c is continuously differentiable.

Since the Fischer-Burmeister function is twice continuously differentiable at every $(a,b)^\top \neq (0,0)^\top$, it is uniformly 1-order semismooth on every compact set that does not contain the origin. Hence, the assumption in (ii) of Lemma 1 for $p = 1$ is satisfied at z_*, if c possesses a locally Lipschitz continuous first derivative and if the uniform strict complementarity condition (15) is satisfied. □

3. **Computation of the Search Direction.** For brevity we neglect the arguments whenever possible. The linear operator equation $V_k(d^k) = -F(z^k)$ in step (2) of Algorithm 1 reads as

$$\begin{pmatrix} x' \\ \lambda' \end{pmatrix} - \begin{pmatrix} f'_x & 0 \\ -H''_{xx} & -H''_{x\lambda} \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} - \begin{pmatrix} f'_u & 0 \\ -H''_{xu} & -H''_{x\eta} \end{pmatrix} \begin{pmatrix} u \\ \eta \end{pmatrix} = - \begin{pmatrix} (x^k)' - f \\ (\lambda^k)' + (H'_x)^\top \end{pmatrix} \tag{16}$$

and

$$\begin{pmatrix} \psi'_{x_0} & 0 & 0 \\ ({\psi'_{x_0}}^\top \sigma^k)'_{x_0} & I & {\psi'_{x_0}}^\top \\ -({\psi'_{x_1}}^\top \sigma^k)'_{x_0} & 0 & -{\psi'_{x_1}}^\top \end{pmatrix} \begin{pmatrix} x(0) \\ \lambda(0) \\ \sigma \end{pmatrix} + \begin{pmatrix} \psi'_{x_1} & 0 & 0 \\ ({\psi'_{x_0}}^\top \sigma^k)'_{x_1} & 0 & 0 \\ -({\psi'_{x_1}}^\top \sigma^k)'_{x_1} & I & 0 \end{pmatrix} \begin{pmatrix} x(1) \\ \lambda(1) \\ \sigma \end{pmatrix} = - \begin{pmatrix} \psi(x^k(0), x^k(1)) \\ \lambda^k(0) + {\psi'_{x_0}}^\top \sigma^k \\ \lambda^k(1) - {\psi'_{x_1}}^\top \sigma^k \end{pmatrix}, \tag{17}$$

and

$$\mathcal{A} \begin{pmatrix} u \\ \eta \end{pmatrix} + \begin{pmatrix} H''_{ux} & H''_{u\lambda} \\ -Sc'_x & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = - \begin{pmatrix} H'_u \\ \omega(z^k(\cdot)) \end{pmatrix}, \tag{18}$$

where

$$\mathcal{A} := \begin{pmatrix} H''_{uu} & (c'_u)^\top \\ -Sc'_u & R \end{pmatrix}. \tag{19}$$

Herein, every function is evaluated at the current iterate z^k. If the inverse operator $\mathcal{A}^{-1}$ exists, equation (18) can be solved for u and η according to

$$\begin{pmatrix} u \\ \eta \end{pmatrix} = -\mathcal{A}^{-1} \left[\begin{pmatrix} H''_{ux} & H''_{u\lambda} \\ -Sc'_x & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} + \begin{pmatrix} H'_u \\ \omega(z^k(\cdot)) \end{pmatrix} \right]. \tag{20}$$

Introducing this expression into the differential equation (16) yields

$$\begin{pmatrix} x' \\ \lambda' \end{pmatrix} - \left[\begin{pmatrix} f'_x & 0 \\ -H''_{xx} & -H''_{x\lambda} \end{pmatrix} - \begin{pmatrix} f'_u & 0 \\ -H''_{xu} & -H''_{x\eta} \end{pmatrix} \mathcal{A}^{-1} \begin{pmatrix} H''_{ux} & H''_{u\lambda} \\ -Sc'_x & 0 \end{pmatrix} \right] \begin{pmatrix} x \\ \lambda \end{pmatrix} = - \left[\begin{pmatrix} (x^k)' - f \\ (\lambda^k)' + {H'_x}^\top \end{pmatrix} + \begin{pmatrix} f'_u & 0 \\ -H''_{xu} & -H''_{x\eta} \end{pmatrix} \mathcal{A}^{-1} \begin{pmatrix} H'_u \\ \omega(z^k(\cdot)) \end{pmatrix} \right]. \tag{21}$$

Hence, in each iteration of Algorithm 1 we have to solve the linear boundary value problem given by the differential equation (21) and the boundary condition (17). Herein, the constant σ can be viewed as a solution of the differential equation $\sigma' = 0$. Under appropriate conditions that are very similar to those needed for the local minimum principle, the boundary value problem has a unique solution.

If the operator $\mathcal{A}$ is not invertible, the situation becomes more involved. In this case, equation (18) imposes algebraic constraints and the equations (16) and (18) form a differential algebraic equation (DAE) with an index of at least two. Actually, the case when $\mathcal{A}$ is invertible corresponds to the index one case. We will not go into detail here and remain this problem open for future research.

Finally, we state a sufficient condition for the existence of the inverse operator of $\mathcal{A}$ in (19).

Theorem 3. *Let there exist $\alpha > 0$, $\beta > 0$, $\gamma > 0$, $\delta > 0$, $\kappa > 0$ such that a.e. in $[0, 1]$ it holds*

$$\|H''_{uu}[t]\| \leq \gamma, \quad \|c'_u[t]\| \leq \delta, \quad \|c'_u[t]^\top\| \leq \kappa,$$

and

$$d^\top H''_{uu}[t] d \geq \alpha \|d\|^2 \qquad \forall d \in \mathbb{R}^{n_u}$$

and

$$\|c'_u[t]^\top \zeta\| \geq \beta \|\zeta\| \qquad \forall \zeta \in \mathbb{R}^{n_u}.$$

Then, the inverse operator $\mathcal{A}^{-1}$ exists and it is linear and bounded.

Proof. We will show that $\|\mathcal{A}w\| \geq C\|w\|$ holds for all w and some $C > 0$.

For brevity let $Q(t) := H''_{uu}[t]$ and recall that

$$(s_i, r_i) \in \partial\varphi[t] \subseteq \{(s, r) \in \mathbb{R}^2 \mid (s + 1)^2 + (r + 1)^2 \leq 1\}. \tag{22}$$

In particular, it holds $r_i, s_i \leq 0$. For an arbitrary but fixed $\varepsilon \in (0, 1)$ we define the index sets

$$J(t) := \{j \in \{1, \dots, n_c\} \mid r_j(t) \leq -\varepsilon\}, \quad J^c(t) := \{1, \dots, n_c\} \setminus J(t).$$

For an arbitrary but fixed $t \in [0, 1]$ define

$$\begin{aligned}
R_1 &:= \operatorname{diag}(r_j(t) \mid j \in J(t)), \\
R_2 &:= \operatorname{diag}(r_j(t) \mid j \in J^c(t)), \\
S_1 &:= \operatorname{diag}(s_j(t) \mid j \in J(t)), \\
S_2 &:= \operatorname{diag}(s_j(t) \mid j \in J^c(t)), \\
A_1 &:= \left(c'_{j,u}[t] \mid j \in J(t)\right), \\
A_2 &:= \left(c'_{j,u}[t] \mid j \in J^c(t)\right).
\end{aligned}$$

In the sequel we suppress the explicit dependence on t for brevity. Notice, that owing to (22) the matrices R_1 and S_2 are nonsingular. In particular, the following estimates hold (w.r.t. the spectral norm):

$$\varepsilon \leq \|R_1\| \leq 2, \quad \|R_2\| < \varepsilon, \quad \frac{1}{2} \leq \|R_1^{-1}\| \leq \frac{1}{\varepsilon},$$

$$0 \leq \|S_1\| \leq 2, \quad 1 - \sqrt{\varepsilon(2 - \varepsilon)} \leq \|S_2\| \leq 2, \quad \frac{1}{2} \leq \|S_2^{-1}\| \leq \frac{1}{1 - \sqrt{\varepsilon(2 - \varepsilon)}}.$$

Without loss of generality assume $J(t) = \{1, \dots, q\}$ and $J^c(t) = \{q + 1, \dots, n_c\}$ with $1 \leq q \leq n_c$. Consider the linear equation

$$\begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} = \begin{pmatrix} Q & A_1^\top & A_2^\top \\ -S_1 A_1 & R_1 & \Theta \\ -S_2 A_2 & \Theta & R_2 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}. \tag{23}$$

The second equation yields

$$w_2 = R_1^{-1} \left(e_2 + S_1 A_1 w_1\right). \tag{24}$$

Introducing this expression into the first equation yields

$$e_1 - A_1^\top R_1^{-1} e_2 = \left(Q + A_1^\top R_1^{-1} S_1 A_1\right) w_1 + A_2^\top w_3. \tag{25}$$

The symmetric operator $\hat{Q} := Q + A_1^\top R_1^{-1} S_1 A_1$ satisfies

$$d^\top \hat{Q} d \geq d^\top Q d \geq \alpha \|d\|^2 \qquad \forall d \in \mathbb{R}^{n_u}$$

since Q is supposed to be uniformly positive definite and $R_1^{-1} S_1$ is positive semidefinite according to (22). Furthermore, it holds $\|\hat{Q}\| \leq \|Q\| + \|A_1^\top\| \cdot \|R_1^{-1}\| \cdot \|S_1\| \cdot \|A_1\| \leq \gamma + \frac{2\delta\kappa}{\varepsilon} =: C_1$ and all eigenvalues of $\hat{Q}$ are located in $[\alpha, C_1]$. Consequently, all eigenvalues of $\hat{Q}^{-1}$ are located in $[\frac{1}{C_1}, \frac{1}{\alpha}]$ and

$$d^\top \hat{Q}^{-1} d \geq \frac{1}{C_1} \|d\|^2, \qquad \|\hat{Q}^{-1}\| \leq \frac{1}{\alpha}.$$

Solving (25) for w_1 leads to

$$w_1 = \hat{Q}^{-1} \left(e_1 - A_1^\top R_1^{-1} e_2 - A_2^\top w_3\right). \tag{26}$$

Introduction into the third equation of (23) yields

$$S_2^{-1} e_3 + A_2 \hat{Q}^{-1} \left(e_1 - A_1^\top R_1^{-1} e_2\right) = \left(A_2 \hat{Q}^{-1} A_2^\top + S_2^{-1} R_2\right) w_3.$$

The matrix

$$\tilde{Q} := A_2 \hat{Q}^{-1} A_2^\top + S_2^{-1} R_2$$

is again symmetric and uniformly positive definite because for all d it holds

$$d^\top \tilde{Q} d = \left(A_2^\top d\right)^\top \hat{Q}^{-1} A_2^\top d + d^\top S_2^{-1} R d \geq \frac{1}{C_1} \|A_2^\top d\|^2 \geq \frac{\beta^2}{C_1} \|d\|^2.$$

Hence, the estimate

$$\|w_3\| \leq \|\tilde{Q}^{-1}\| \cdot \left(\|S_2^{-1}\| + \|A_2\| \cdot \|\hat{Q}^{-1}\| \cdot \left(1 + \|A_1^\top\| \cdot \|R_1^{-1}\right) \|\right) \|e\|$$

holds. Since all norms are bounded we showed that $\|w_3\| \leq C_2 \|e\|$ for some $C_2 > 0$. Equations (24) and (26) yield

$$\|w_1\| \leq C_3 \|e\|, \qquad \|w_2\| \leq C_4 \|e\|$$

with appropriate constants $C_3, C_4 > 0$. Finally, we obtain $\|w\| \leq \|w_1\| + \|w_2\| + \|w_3\| \leq C\|e\| = C\|\mathcal{A}(t)w\|$ with $C := C_2 + C_3 + C_4$. Since C does not depend on t this estimate holds uniformly and the assertions follow with Ljusternik and Sobolew [11], Th. 1, p. 106. □

4. **Globalization.** One reason that makes the Fischer-Burmeister function appealing is the fact that its square

$$\phi(a, b) := \varphi(a, b)^2 = \left(\sqrt{a^2 + b^2} - a - b\right)^2$$

is continuously differentiable with $\phi'(a, b) = 2\varphi(a, b)v$, where $v \in \partial\varphi(a, b)$ is arbitrary. Hence, the mappings

$$(\bar{x}, \bar{u}, \bar{\eta}) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_c} \mapsto \phi(-c_i(\bar{x}, \bar{u}), \bar{\eta}_i), \quad i = 1, \ldots, n_c$$

are continuously differentiable by the chain rule. This allows to globalize the local nonsmooth Newton's method using the squared L^2-norm of F as a merit function:

$$\begin{aligned}
\Theta(z) &:= \frac{1}{2}\|F(z)\|_2^2 \\
&= \frac{1}{2}\int_0^1 \|x'(t) - f(x(t),u(t))\|^2\,dt \\
&\quad + \frac{1}{2}\int_0^1 \left\|\lambda'(t) + H'_x(x(t),u(t),\lambda(t),\eta(t))^\top\right\|^2 dt \\
&\quad + \frac{1}{2}\int_0^1 \|H'_u(x(t),u(t),\lambda(t),\eta(t))\|^2\,dt + \frac{1}{2}\sum_{i=1}^{n_c}\int_0^1 \phi(-c_i(x(t),u(t)),\eta_i(t))dt \\
&\quad + \frac{1}{2}\|\psi(x(0),x(1))\|^2 + \frac{1}{2}\|\lambda(0) + \psi'_{x_0}(x(0),x(1))^\top\sigma\|^2 \\
&\quad + \frac{1}{2}\|\lambda(1) - \psi'_{x_1}(x(0),x(1))^\top\sigma\|^2.
\end{aligned}$$

Θ is Fréchet-differentiable in Z if f_0, f, c, ψ are twice continuously differentiable. An analysis of the derivative of Θ reveals that for d^k with $V_k(d^k) = -F(z^k)$ it holds

$$\Theta'(z^k)(d^k) = -2\Theta(z^k) = -\|F(z^k)\|_2^2. \tag{27}$$

As a consequence, d^k is a direction of descent of Θ at z^k and the line-search in the following global version of the nonsmooth Newton's method is well-defined unless z^k is a zero of F.

Algorithm 2 (Global Nonsmooth Newton's Method)**.**

(0) Choose z^0, $\beta \in (0,1)$, $\sigma \in (0,1/2)$.
(1) If some stopping criterion is satisfied, stop.
(2) Chose an arbitrary $V_k \in \partial_* F(z^k)$ and compute the search direction d^k from
$$V_k(d^k) = -F(z^k).$$
(3) Find smallest $i_k \in \mathbb{N}_0$ with
$$\Theta(z^k + \beta^{i_k} d^k) \le \Theta(z^k) + \sigma\beta^{i_k}\Theta'(z^k)(d^k)$$
and set $\alpha_k = \beta^{i_k}$.
(4) Set $z^{k+1} = z^k + \alpha_k d^k$, $k = k+1$, and goto (1).

The upcoming global convergence proof to a large extend is similar to the proof presented in Jiang [8] for finite dimensions. However, some safeguards are necessary for infinite dimensions.

Theorem 4. *Let the inverse operators V_k^{-1} exist for all k and let $C > 0$ be a constant such that $\|V_k^{-1}\|_{\mathcal{L}(Y,Z)} \le C$ holds for all k. Let z_* be an accumulation point of the sequence $\{z^k\}$ generated by the global nonsmooth Newton's method.*

Then, z_* is a zero of F.

Proof. Let $\{z^k\}$ be a (sub)sequence with $z^k \to z_*$ and $F(z^k) \neq 0$. Then, $\Theta'(z^k)(d^k) = -2\Theta(z^k) = -\|F(z^k)\|_2^2 < 0$. The line-search is well-defined by the differentiability of Θ.

(i) Case 1: Assume
$$\alpha := \liminf_{k\to\infty} \alpha_k > 0.$$
Then
$$0 \le \Theta(z^{k+1}) \le \Theta(z^k) + \sigma\alpha_k\Theta'(z^k)(d^k) = \Theta(z^k)\left(1 - 2\sigma\alpha_k\right).$$

With $\sigma \in (0, 1/2)$ and $\alpha \le \alpha_k \le 1$ it follows $0 < 1 - 2\sigma\alpha_k \le 1 - 2\sigma\alpha < 1$ and repeated application yields

$$0 \le \Theta(z^k) \le \Theta(z^0)\,(1 - 2\sigma\alpha)^k \to 0.$$

By the continuity of F, z_* is a zero of F.

(ii) Case 2: Assume that there is a subsequence with $\alpha_k \to 0$, $k \in J \subset \mathbb{N}$.

The sequence $\{d^k\}$ is bounded since $\{V_k^{-1}\}$ is bounded and

$$0 \le \|d^k\|_Z = \|V_k^{-1}(F(z^k))\|_Z \le C\|F(z^k)\|_Y \le C\|F(z^0)\|_Y.$$

Unfortunately, the boundedness of $\{d^k\}$ in an infinite dimensional space does not imply that there exists a convergent subsequence. However, since d^k is bounded in $Z = W^{1,\infty}([0,1], \mathbb{R}^{n_x}) \times L^\infty([0,1], \mathbb{R}^{n_u}) \times L^\infty([0,1], \mathbb{R}^{n_x}) \times L^\infty([0,1], \mathbb{R}^{n_c}) \times \mathbb{R}^{n_\psi}$ it is also bounded in the space $\hat{Z} := W^{1,2}([0,1], \mathbb{R}^{n_x}) \times L^2([0,1], \mathbb{R}^{n_u}) \times L^2([0,1], \mathbb{R}^{n_x}) \times L^2([0,1], \mathbb{R}^{n_c}) \times \mathbb{R}^{n_\psi}$. $\hat{Z}$ is a Hilbert space and thus reflexive. According to Theorem III.3.7 in Werner [19] there exists a weakly convergent subsequence $\{d^k\}$, $k \in \hat{J} \subseteq J$. Hence, there exists some $d_* \in \hat{Z}$ such that for every element $g \in \hat{Z}^*$ it holds

$$g(d^k) \to g(d_*). \tag{28}$$

Herein, $\hat{Z}^*$ denotes the topological dual space of $\hat{Z}$. The derivative $\Theta'(z_*)(\cdot)$ is an element of Z^* and an investigation turns out that it is essentially made up of linear functionals of type

$$g_1(z) = \int_0^1 h_1(z_*(t))z(t)dt, \qquad g_2(z) = \int_0^1 h_2(z_*(t))z'(t)dt$$

with essentially bounded functions $h_1(z_*(\cdot))$ and $h_2(z_*(\cdot))$. Thus, by application of the Cauchy-Schwartz inequality, the functionals g_1 and g_2 are also linear continuous functionals on $\hat{Z}$ and thus g_1, g_2, and in particular $\Theta'(z_*)(\cdot)$ can be viewed as elements of $\hat{Z}^*$.

Hence, (28) holds for $g(\cdot) = \Theta'(z_*)(\cdot)$:

$$\Theta'(z_*)(d^k) \to \Theta'(z_*)(d_*).$$

Furthermore, due to the continuity of $\Theta'(\cdot)$ (in Z) for every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $\|z^k - z_*\|_Z \le \delta$ it holds

$$\begin{aligned} |\Theta'(z^k)(d^k) - \Theta'(z_*)(d^k)| &= \|d^k\|_Z \left| \Theta'(z^k)\left(\frac{d^k}{\|d^k\|_Z}\right) - \Theta'(z_*)\left(\frac{d^k}{\|d^k\|_Z}\right)\right| \\ &\le \|d^k\|_Z \cdot \sup_{\|d\|_Z = 1} |\Theta'(z^k)(d) - \Theta'(z_*)(d)| \\ &= \|d^k\|_Z \cdot \|\Theta'(z^k) - \Theta'(z_*)\|_{\mathcal{L}(Z,\mathbb{R})} \le \varepsilon\|d^k\|_Z. \end{aligned}$$

For arbitrary $\varepsilon > 0$ we find

$$\begin{aligned} |\Theta'(z^k)(d^k) - \Theta'(z_*)(d_*)| &\le |\Theta'(z^k)(d^k) - \Theta'(z_*)(d^k)| \\ &\quad + |\Theta'(z_*)(d^k) - \Theta'(z_*)(d_*)| \\ &\le \varepsilon\|d^k\|_Z + |\Theta'(z_*)(d^k) - \Theta'(z_*)(d_*)|. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary and since d^k is weakly convergent it holds

$$\Theta'(z^k)(d^k) \to \Theta'(z_*)(d_*) \qquad \text{as } k \to \infty.$$

In a similar way the Fréchet differentiability of Θ yields

$$\begin{aligned}
&\left|\frac{1}{\alpha_k}\left(\Theta(z^k+\alpha_k d^k)-\Theta(z^k)\right)-\Theta'(z_*)(d_*)\right| \\
&\le \left|\frac{1}{\alpha_k}\left(\Theta(z^k+\alpha_k d^k)-\Theta(z^k)\right)-\Theta'(z^k)(d^k)\right| + \left|\Theta'(z^k)(d^k)-\Theta'(z_*)(d_*)\right| \\
&\le \frac{1}{\alpha_k}o(\|\alpha_k d^k\|_Z) + \left|\Theta'(z^k)(d^k)-\Theta'(z_*)(d^k)\right| + \left|\Theta'(z_*)(d^k)-\Theta'(z_*)(d_*)\right| \\
&\le \|d^k\|_Z\frac{o(\alpha_k\|d^k\|_Z)}{\alpha_k\|d^k\|_Z} + \varepsilon\|d^k\|_Z + \left|\Theta'(z_*)(d^k)-\Theta'(z_*)(d_*)\right|.
\end{aligned}$$

Since d^k is weakly convergent it holds

$$\frac{1}{\alpha_k}\left(\Theta(z^k+\alpha_k d^k)-\Theta(z^k)\right) \to \Theta'(z_*)(d_*) \qquad \text{as } k\to\infty.$$

The line search in step (3) of the algorithm yields

$$\begin{aligned}
\frac{\Theta(z^k+\alpha_k d^k)-\Theta(z^k)}{\alpha_k} &\le \sigma\Theta'(z^k)(d^k), \\
\frac{\Theta(z^k+\frac{\alpha_k}{\beta} d^k)-\Theta(z^k)}{\frac{\alpha_k}{\beta}} &> \sigma\Theta'(z^k)(d^k).
\end{aligned}$$

Passing to the limit and exploiting the previous considerations yields

$$\sigma\Theta'(z_*)(d_*) = \Theta'(z_*)(d_*).$$

Since $\sigma\in(0,1/2)$ this only holds for $\Theta'(z_*)(d_*)=0$. Thus, we have shown

$$-\|F(z^k)\|_2^2 = \Theta'(z^k)(d^k) \to \Theta'(z_*)(d_*) = 0.$$

By the continuity of F, z_* is a zero of F.

□

The previous result only shows the convergence to a zero of F. It would be nice to have also the fast local convergence properties of the local method. The locally superlinear or even quadratic convergence would follow from the local convergence theorem 1 if we were able to show that $\alpha_k=1$ satisfies Armijo's rule for all sufficiently large k. But unfortunately, this leads to a two-norm discrepancy. The proof of the local convergence theorem 1 showed the superlinear convergence of the values $\|F(z^k)\|_Y$, i.e. for any $\varepsilon>0$ there exists $\delta>0$ such that for all $\|z-z_*\|\le\delta$ it holds

$$\|z+d-z_*\|_Z \le \varepsilon\|z-z_*\|_Z, \qquad \|F(z+d)\|_Y \le \varepsilon\|F(z)\|_Y,$$

where $d=-V^{-1}F(z)$, $V\in\partial_*F(z)$. In particular, with $z=z^k$ and $d=d^k$ there exists $\delta>0$ such that for all $\|z^k-z_*\|_Z\le\delta$ it holds

$$\|z^k+d^k-z_*\|_Z \le \frac{1}{2}\|z^k-z_*\|_Z, \qquad \|F(z^k+d^k)\|_Y \le \sqrt{1-2\sigma}\|F(z^k)\|_Y,$$

Unfortunately, we would need this property not for the norm $\|\cdot\|_Y$ but for the norm $\|\cdot\|_2$ since then

$$\Theta(z^k+d^k) = \frac{1}{2}\|F(z^k+d^k)\|_2^2 \le \frac{1-2\sigma}{2}\|F(z^k)\|_2^2 = (1-2\sigma)\Theta(z^k)$$

resp.

$$\Theta(z^k+d^k) \le \Theta(z^k) - 2\sigma\Theta(z^k) = \Theta(z^k) + \sigma\Theta'(z^k)(d^k),$$

i.e. Armijo's line-search would accept $\alpha_k = 1$ and $z^{k+1} = z^k + d^k$. Furthermore, $\|z^{k+1} - z_*\|_Z \leq \frac{1}{2}\|z^k - z_*\|_Z \leq \delta$ and we are in the same situation as above and the argument could be repeated.

Unfortunately, the superlinear convergence of the residual norms $\|F(z^k)\|_2$ could not be established by now. An additional assumption is needed to prove the fast local convergence.

Theorem 5. *Let the assumptions of theorem 4 be valid. If, in addition, there exists a constant $K > 0$ such that*

$$\|F(z^k)\|_Y \leq K\|F(z^k)\|_2$$

holds for $\{z^k\}$ with $z^k \to z_*$. Then, for sufficiently large k the step length $\alpha_k = 1$ is accepted and the global method turns into the local one.

Proof. Owing to the previous considerations it remains to show that

$$\lim_{k\to\infty} \frac{\|F(z^{k+1})\|_2}{\|F(z^k)\|_2} = 0.$$

Recall, that $\|\cdot\|_Y$ is essentially the L^∞-norm. Hence, there exits a constant $C_1 > 0$ with $\|y\|_2 \leq C_1\|y\|_Y$ for all $y \in Y$. Together with the superlinear convergence of the values $\|F(z^k)\|_Y$ in Theorem 1 for every $\varepsilon > 0$ and for sufficiently large k it holds

$$\|F(z^k + d^k)\|_2 \leq C\|F(z^k + d^k))\|_Y \leq \ C\varepsilon\|F(z^k)\|_Y \leq C \cdot K \cdot \varepsilon\|F(z^k)\|_2.$$

Since ε was arbitrary, this shows the superlinear convergence of the values $\|F(z^k)\|_2$. □

5. **Numerical Results.** All computations were performed on a PC with 3 GHz processing speed.

5.1. **Rayleigh Problem, Version 1.** We illustrate the method for the Rayleigh problem, cf. Maurer and Augustin [12], p. 39: Minimize

$$\int_0^{4.5} u(t)^2 + x_1(t)^2 dt \tag{29}$$

subject to

$$\begin{array}{rcll} x_1' &=& x_2, & x_1(0) = -5, \\ x_2' &=& -x_1 + x_2\,(1.4 - 0.14x_2^2) + 4u, & x_2(0) = -5 \end{array} \tag{30}$$

and

$$u + \frac{1}{6}x_1 \leq 0.$$

With $x = (x_1, x_2)^\top$, $\lambda = (\lambda_1, \lambda_2)^\top$, $\sigma = (\sigma_1, \sigma_2)^\top$ the Hamilton function reads as

$$H(x, u, \lambda, \eta) = u^2 + x_1^2 + \lambda_1 x_2 + \lambda_2\left(-x_1 + x_2\left(1.4 - 0.14x_2^2\right) + 4u\right) + \eta\left(u + \frac{1}{6}x_1\right).$$

With $z = (x, u, \lambda, \eta, \sigma)$ the function F in (9) is given by

$$F(z) = \begin{pmatrix} x_1' - x_2 \\ x_2' - (-x_1 + x_2\,(1.4 - 0.14x_2^2) + 4u) \\ \lambda_1' + 2x_1 - \lambda_2 + \frac{1}{6}\eta \\ \lambda_2' + \lambda_1 + \lambda_2\,(1.4 - 0.42x_2^2) \\ x_1(0) + 5 \\ x_2(0) + 5 \\ \lambda_1(0) + \sigma_1 \\ \lambda_2(0) + \sigma_2 \\ \lambda_1(4.5) \\ \lambda_2(4.5) \\ 2u + 4\lambda_2 + \eta \\ \varphi\left(-\left(u + \frac{1}{6}x_1\right), \eta\right) \end{pmatrix}.$$

In each iteration of the nonsmooth Newton's method we have to solve the linear boundary value problem (17), (21) for x, λ, σ. We leave the details of the boundary value problem (17),(21) and equation (20) to the reader. We note, that for all $(s+1)^2 + (r+1)^2 \leq 1$ it holds

$$\det \mathcal{A} = \det \begin{pmatrix} 2 & 1 \\ -s & r \end{pmatrix} = 2r + s \neq 0$$

and thus the operator $\mathcal{A}$ in (19) is invertible. The differential equations are discretized on $[0, 4.5]$ using forward Euler's method with N equidistant subintervals. The occurring derivatives $(x^k)'$ and $(\lambda^k)'$ are approximated by finite forward differences. The boundary value problem was solved by the single shooting method. Table 1 shows the output of the globalized nonsmooth Newton's method, i.e. step size α, residual norm $\|F\|$, and $\|d^k\|$ during iteration. The iterations show the rapid quadratic convergence at the end of the iteration sequence.

ITER	ALPHA	||F||	||dx||
0	0.000000E+00	0.245000E+04	0.173257E+04
1	0.531441E+00	0.173372E+04	0.316003E+04
2	0.717898E-01	0.170185E+04	0.897810E+03
3	0.185302E+00	0.155477E+04	0.653211E+03
...			
10	0.100000E+01	0.147905E-05	0.592231E-02
11	0.100000E+01	0.167034E-08	0.213155E-03
12	0.100000E+01	0.253557E-14	0.263768E-06
13	0.100000E+01	0.152582E-25	0.598877E-12

TABLE 1. Output of globalized nonsmooth Newton's method for the first version of Rayleigh's problem for $N = 100$ subintervals and Euler discretization: local quadratic convergence.

Figure 1 illustrates the iterates of the nonsmooth Newton's method. Notice the small inactive arc of the control-state constraint at the end of the time interval.

For comparison reasons the same optimal control problem was solved alternatively by a direct discretization method as in Gerdts [5] with Euler discretization and $N = 100$ subintervals. For this method the overall CPU time was 3.81 CPU seconds on the same processor. Furthermore, for the direct discretization method the CPU time grows nonlinearly with

N. Hence, if all regularity assumptions are fulfilled the nonsmooth Newton's method is an extremely efficient method.

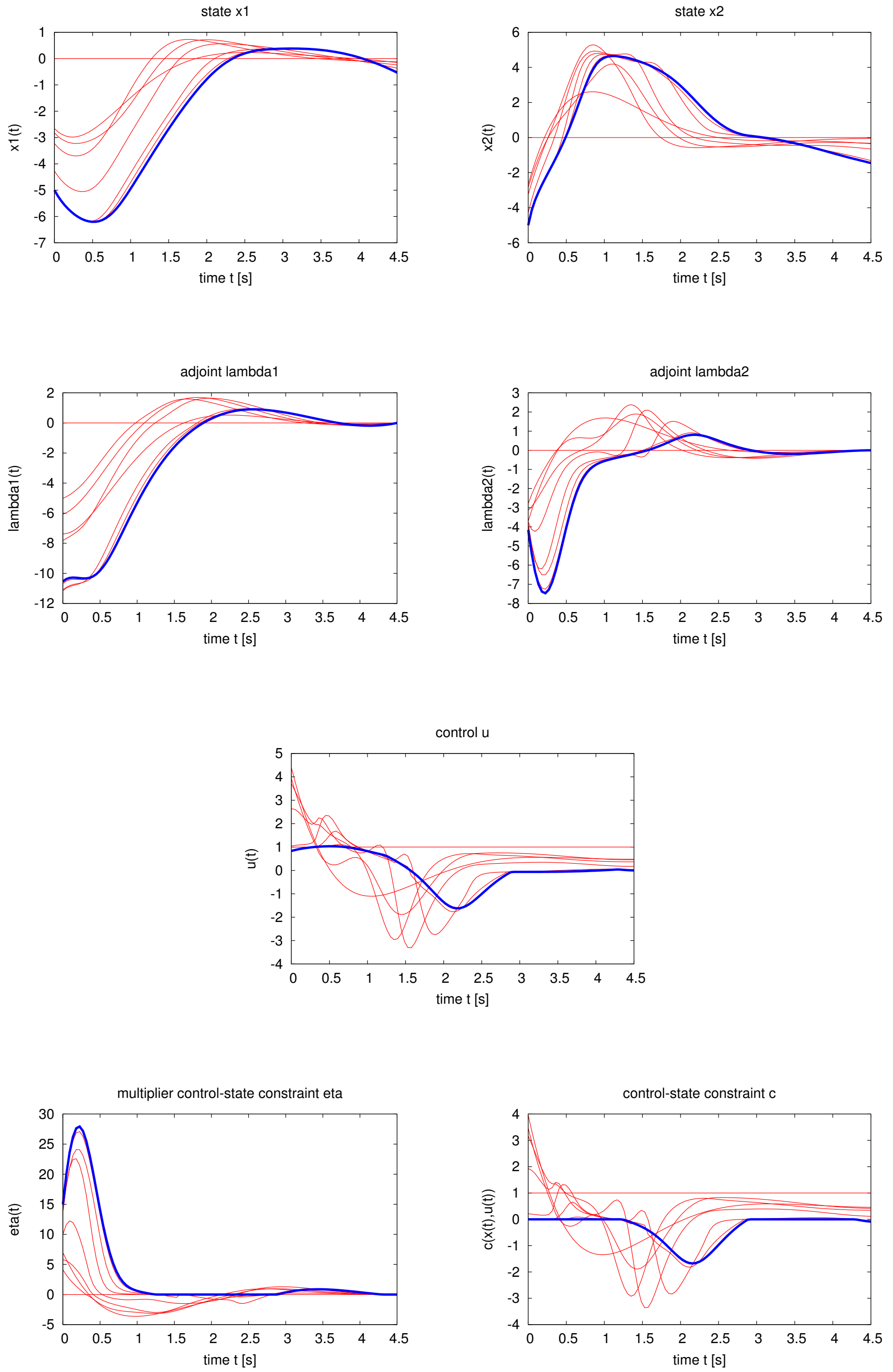

FIGURE 1. Numerical solution of the first version of Rayleigh's problem for $N = 100$ Euler steps: Intermediate iterates (thin lines) and converged solution (thick lines).

The following table summarizes results for different step sizes. The number of iterations differs only by one, which indicates – at least numerically – the mesh independence of the method. Furthermore, the CPU time grows at a linear rate with N.

N	CPU time [s]	Iterations
100	0.027	13
500	0.136	14
1000	0.271	14
2000	0.505	14
4000	1.083	14
8000	2.065	14

5.2. **Rayleigh Problem, Version 2.** We consider a slight variation of the Rayleigh problem where boundary conditions are added and the control-state constraint is replaced by box constraints for the control, cf. Maurer and Augustin [12], p. 39: Minimize (29) subject to (30) and $x_1(4.5) = 0$, $x_2(4.5) = 0$ and

$$-1 \le u \le 1.$$

With $x = (x_1, x_2)^\top$, $\lambda = (\lambda_1, \lambda_2)^\top$, $\sigma = (\sigma_1, \dots, \sigma_4)^\top$, $\eta = (\eta_1, \eta_2)^\top$ the Hamilton function reads as

$$\begin{aligned} H(x, u, \lambda, \eta) &= u^2 + x_1^2 + \lambda_1 x_2 + \lambda_2 \left(-x_1 + x_2 \left(1.4 - 0.14 x_2^2\right) + 4u\right) \\ &\quad + \eta_1(u-1) + \eta_2(-u-1). \end{aligned}$$

With $z = (x, u, \lambda, \eta, \sigma)$ the function F in (9) is given by

$$F(z) = \left(\begin{array}{l} x_1' - x_2 \\ x_2' - (-x_1 + x_2 (1.4 - 0.14 x_2^2) + 4u) \\ \lambda_1' + 2x_1 - \lambda_2 \\ \lambda_2' + \lambda_1 + \lambda_2 (1.4 - 0.42 x_2^2) \\ x_1(0) + 5 \\ x_2(0) + 5 \\ x_1(4.5) \\ x_2(4.5) \\ \lambda_1(0) + \sigma_1 \\ \lambda_2(0) + \sigma_2 \\ \lambda_1(4.5) - \sigma_3 \\ \lambda_2(4.5) - \sigma_4 \\ 2u + 4\lambda_2 + \eta_1 - \eta_2 \\ \varphi\left(-\left(u-1\right), \eta_1\right) \\ \varphi\left(-\left(-u-1\right), \eta_2\right) \end{array} \right).$$

Again, we leave the details of the linear boundary value problem (17), (21) and equation (20) to the reader. An investigation of the generalized differential of φ yields

$$\det \mathcal{A} = \det \begin{pmatrix} 2 & 1 & -1 \\ -s_1 & r_1 & 0 \\ s_2 & 0 & r_2 \end{pmatrix} = 2r_1 r_2 + r_1 s_2 + r_2 s_1 \neq 0$$

for any $(s_1, r_1) \in \partial\varphi(-(u-1), \eta_1)$ and $(s_2, r_2) \in \partial\varphi(-(-u-1), \eta_2)$.

Figure 2 illustrates the iterates of the nonsmooth Newton's method for $N = 100$.

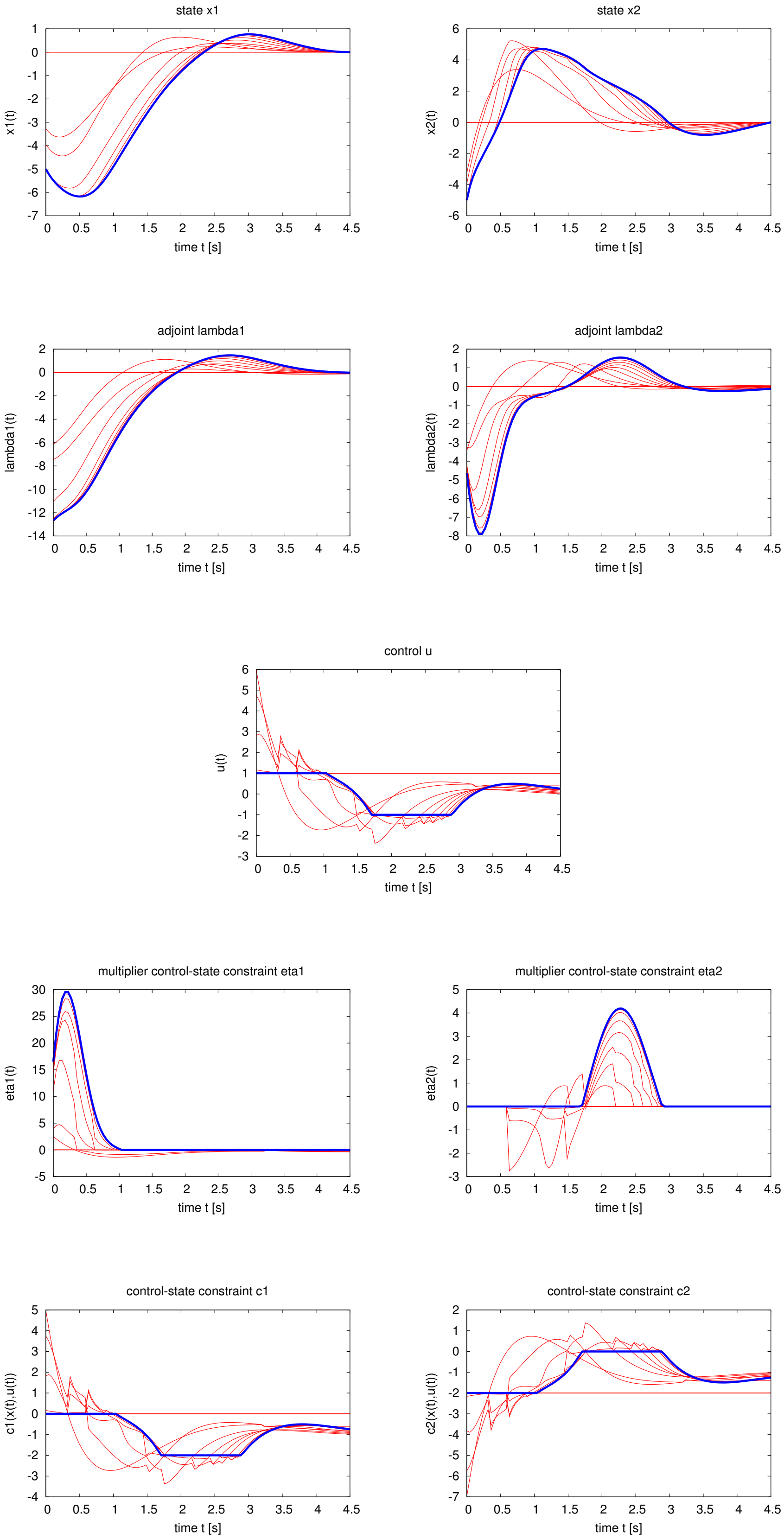

FIGURE 2. Numerical solution of the second version of Rayleigh's problem for $N = 100$ Euler steps: Intermediate iterates (thin lines) and converged solution (thick lines).

Table 2 shows more detailed information about the iterations, i.e. step size α, residual norm $\|F\|$, and $\|d^k\|$. The iterations show the rapid quadratic convergence at the end of the iteration sequence.

```
ITER        ALPHA              ||F||              ||dx||
-----------------------------------------------------------
   0     0.000000E+00      0.205000E+04      0.301353E+08
   1     0.898145E-07      0.205000E+04      0.772399E+06
   2     0.442969E-05      0.204999E+04      0.137827E+04
   3     0.656100E+00      0.137884E+04      0.533635E+03
...
  14     0.100000E+01      0.485899E-04      0.165212E+00
  15     0.100000E+01      0.710910E-07      0.678731E-02
  16     0.100000E+01      0.108957E-12      0.842304E-05
  17     0.100000E+01      0.271474E-24      0.130452E-10
```

TABLE 2. Output of globalized nonsmooth Newton's method for the second version of Rayleigh's problem for $N = 100$ subintervals and Euler discretization: local quadratic convergence.

The number of iterations remains nearly constant, which indicates – at least numerically – the mesh independence of the method. Furthermore, the CPU time grows at a linear rate with N.

N	CPU time [s]	Iterations
100	0.049	17
500	0.204	15
1000	0.502	18
2000	0.848	16
4000	1.785	17
8000	3.713	17

REFERENCES

[1] X. Chen, Z. Nashed, and L. Qi, *Smoothing methods and semismooth methods for nondifferentiable operator equations*, SIAM Journal on Numerical Analysis, 38 (2000), pp. 1200–1216.

[2] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, 1983.

[3] A. Fischer, *A special Newton-type optimization method*, Optimization, 24 (1992), pp. 269–284.

[4] ——, *Solution of monotone complementarity problems with locally Lipschitzian functions*, Mathematical Programming, 76 (1997), pp. 513–532.

[5] M. Gerdts, *Direct Shooting Method for the Numerical Solution of Higher Index DAE Optimal Control Problems*, Journal of Optimization Theory and Applications, 117 (2003), pp. 267–294.

[6] ——, *Local minimum principle for optimal control problems subject to index one differential-algebraic equations*, tech. report, Department of Mathematics, University of Hamburg, http://www.math.uni-hamburg.de/home/gerdts/Report_index1.pdf, 2005.

[7] ——, *A nonsmooth newton's method for control-state constrained optimal control problems*, vol. 30 of Proceedings 5th Vienna Symposium on Mathematical Modelling, ARGESIM Reports, ISBN 3-901608-30-3, Vienna, 2006.

[8] H. Jiang, *Global convergence analysis of the generalized Newton and Gauss-Newton methods of the Fischer-Burmeister equation for the complementarity problem*, Mathematics of Operations Research, 24 (1999), pp. 529–543.

[9] B. Kummer, *Newton's method for non-differentiable functions*, in Advances in Mathematical Optimization, J. Guddat et al., ed., Berlin, 1988, Akademie-Verlag, pp. 171–194.

[10] ———, *Newton's method based on generalized derivatives for nonsmooth functions: convergence analysis*, in Advances in Optimization, W. Oettli and D. Pallaschke, eds., Berlin, 1991, Springer, pp. 171–194.

[11] L. A. Ljusternik and W. I. Sobolew, *Elemente der Funktionalanalysis*, Verlag Harri Deutsch, Zürich-Frankfurt/Main-Thun, 1976.

[12] H. Maurer and D. Augustin, *Sensitivity Analysis and Real-Time Control of Parametric Optimal Control Problems Using Boundary Value Methods*, in Online Optimization of Large Scale Systems, M. Grötschel, S. O. Krumke, and J. Rambau, eds., Springer, 2001, pp. 17–55.

[13] L. W. Neustadt, *Optimization: A Theory of Necessary Conditions*, Princeton, New Jersey, 1976.

[14] H. J. Oberle and W. Grimm, *Bndsco – A Program for the Numerical Solution of Optimal Control Problems*, Tech. Report Reihe B, Bericht 36, Hamburger Beiträge zur Angewandten Mathematik, Department of Mathematics, University of Hamburg, http://www.math.uni-hamburg.de/home/oberle/software.html, 2001.

[15] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Mathematics of Operations Research, 18 (1993), pp. 227–244.

[16] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Mathematical Programming, 58 (1993), pp. 353–367.

[17] M. Ulbrich, *Nonsmooth Newton-like Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*, Habilitation, Technical University of Munich, Munich, 2002.

[18] M. Ulbrich, *Semismooth newton methods for operator equations in function spaces*, SIAM Journal on Optimization, 13 (2003), pp. 805–841.

[19] D. Werner, *Funktionalanalysis*, Springer, Berlin-Heidelberg-New York, 1995.

E-mail address: gerdts@math.uni-hamburg.de