T U M

FAKULTÄT FÜR MATHEMATIK

Non-Monotone Trust-Region Methods for Bound-Constrained Semismooth Equations with Applications to Nonlinear Mixed Complementarity Problems

Ulbrich, Michael

TUM-M9906

April 99

TECHNISCHE UNIVERSITÄT MÜNCHEN

Cataloging Data :

Ulbrich, Michael: Non-Monotone Trust-Region Methods for Bound-Constrained Semismooth Equations with Applications to Nonlinear Mixed Complementarity Problems; Techn. Univ. München, Fak. f. Math, Report TUM M9906 (99)

Mathematics Subject Classification : 90C30 90C33 49J40 65H10 65K05 49M37

Editor :

H. Wähling (waehling@mathematik.tu-muenchen.de),
Fakultät für Mathematik der Technischen Universität München, D-80290 München, Germany.

For the Electronic Version see : http://www-lit.mathematik.tu-muenchen.de/reports/

NON-MONOTONE TRUST-REGION METHODS FOR BOUND-CONSTRAINED SEMISMOOTH EQUATIONS WITH APPLICATIONS TO NONLINEAR MIXED COMPLEMENTARITY PROBLEMS

MICHAEL ULBRICH*

Abstract. We develop and analyze a class of trust-region methods for bound-constrained semismooth systems of equations. The algorithm is based on a simply constrained differentiable minimization reformulation. Our global convergence results are developed in a very general setting that allows for non-monotonicity of the function values at subsequent iterates. We propose a way of computing trial steps by a semismooth Newton-like method that is augmented by a projection onto the feasible set. Under a Dennis–Moré-type condition we prove that close to a BD-regular solution the trust-region algorithm turns into this projected Newton method, which is shown to converge locally q-superlinearly or quadratically, respectively, depending on the quality of the approximate BD-subdifferentials used.

As an important application we discuss in detail how the developed algorithm can be used to solve nonlinear mixed complementarity problems (MCPs). Hereby, the MCP is converted into a bound-constrained semismooth equation by means of an MCP-function. We propose and investigate a new class of MCP-functions that are motivated by affine-scaling techniques for nonlinear programming. These functions have attractive theoretical properties and prove to be efficient in practice. This is documented by our numerical results for a subset of the MCPLIB problem collection.

Key words. semismooth equation, non-monotone trust region method, nonlinear mixed complementarity problem, nonsmooth Newton method, global convergence, superlinear and quadratic convergence.

AMS subject classifications. 90C30, 90C33, 49J40, 65H10, 65K05, 49M37

1. Introduction.

In this paper we propose and analyze a class of trust-region methods for the solution of a simply constrained system of nonlinear nonsmooth equations

$$H(x) = 0, \quad x \in X. \tag{1.1}$$

Hereby, the function $H : \mathbb{R}^n \supset U \to \mathbb{R}^n$ is defined on the open set U containing the feasible set $X \stackrel{\text{def}}{=} [l, u] = \{x \in \mathbb{R}^n \,;\; l_i \le x_i \le u_i, 1 \le i \le n\}$. The bounds $l_i \in \mathbb{R} \cup \{-\infty\}$ and $u_i \in \mathbb{R} \cup \{+\infty\}$ are assumed to satisfy $l_i < u_i$, $1 \le i \le n$, (otherwise the variable $x_i = l_i = u_i$ could be eliminated).

We require that (with the definition of semismoothness to follow)

- (A1) The function H is semismooth or, stronger, p-order semismooth, $0 < p \le 1$.
- (A2) Each component function H_i of H is continuously differentiable on $U \setminus H_i^{-1}(0)$.

The locally q-superlinear/quadratic convergence of the algorithm to BD-regular solutions of (1.1) will be achieved by a Newton-type method that is augmented by a projection onto X to maintain feasibility. Local convergence results for Newton's method without projection were established in [41, 44, 46]. Similar to [41], our local convergence results hold under a Dennis-Moré-type condition, thus allowing for inexactness in the computation of B-subdifferentials and in the solution of linear systems. We safeguard this locally convergent iteration by a non-monotone trust-region globalization that is based on the minimization reformulation

$$\text{minimize } h(x) \quad \text{subject to} \quad x \in X, \tag{1.2}$$

* Lehrstuhl für Angewandte Mathematik und Mathematische Statistik, Technische Universität München, D-80290 München, Germany, E-mail: mulbrich@mathematik.tu-muenchen.de.

where $h : U \to \mathbb{R}$, $h(x) \stackrel{\text{def}}{=} \|H(x)\|^2/2$. Here and throughout the paper, $\|\cdot\|$ denotes the Euclidean norm. Obviously, (1.1) and (1.2) are equivalent if (1.1) possesses a solution. As will be shown in Lemma 4.2, our assumptions on H imply that h is continuously differentiable on U. This enables us to invoke smooth proof techniques for the trust-region algorithm. We stress that our analysis is not based directly on (A2) but on the – in connection with (A1) – weaker assumption

(A2') The function $h : U \to \mathbb{R}$, $h(x) = \|H(x)\|^2/2$ is continuously differentiable.

We believe, however, that (A2) is more concrete and easier to verify than (A2') and thus decide to choose (A1) and (A2) as our working assumptions. An assumption of the form (A2') was also used in [28].

The need for efficient algorithms for the solution of (1.1) comes from the fact that very general classes of problems can be converted to this form. Of particular importance are semismooth reformulations of mixed complementarity problems (MCPs), which will be discussed in the second part of this paper. They are defined as follows [13, 17]:

Nonlinear Mixed Complementarity Problem MCP($F, [l, u]$):

Find $x \in \mathbb{R}^n$ such that for all $1 \le i \le n$ holds

$$l_i \le x_i \le u_i \quad \text{and} \quad F_i(x) \begin{cases} = 0 & \text{if } l_i < x_i < u_i, \\ \ge 0 & \text{if } x_i = l_i, \\ \le 0 & \text{if } x_i = u_i. \end{cases} \tag{1.3}$$

Hereby, the bounds l, u and the sets $U \supset X = [l, u]$ have the same properties as in problem (1.1) and we assume $F \in S^p(U, \mathbb{R}^n)$, $0 \le p \le 1$. Further requirements on F will be stated below.

Mixed complementarity problems arise in a variety of areas, including computer sciences, economics, engineering, operations research, and mathematics. For a comprehensive discussion of applications see [17]. Further applications can be derived from the fact that the Karush-Kuhn-Tucker conditions of mathematical programs and, more general, of variational inequality problems are MCPs. Important special cases of MCP include

(a) Nonlinear systems of equations:

$$F(x) = 0 \qquad \text{(equivalent to: } x \text{ solves MCP}(F, \mathbb{R}^n))$$

(b) Nonlinear complementarity problems (NCPs):

$$x \ge 0,\ F(x) \ge 0,\ x^T F(x) = 0 \qquad \text{(equivalent to: } x \text{ solves MCP}(F, \mathbb{R}^n_+))$$

In order to apply our algorithm, we will reformulate MCP(F, X) equivalently in the form (1.1), where the function $H : U \to \mathbb{R}^n$ is defined by

$$H_i(x) \stackrel{\text{def}}{=} \psi_i(x_i, F_i(x)) \quad , \quad i = 1, \dots, n. \tag{1.4}$$

Hereby, the functions $\psi_i : \mathbb{R}^2 \to \mathbb{R}$ are MCP-functions, i.e., satisfy

$$\psi_i(a, b) = 0 \qquad \text{if and only if} \qquad l_i \le a \le u_i \ \text{ and } \ b \begin{cases} = 0 & \text{if } l_i < a < u_i, \\ \ge 0 & \text{if } a = l_i, \\ \le 0 & \text{if } a = u_i. \end{cases}$$

In the special case of NCPs, the functions ψ_i are NCP-functions, i.e., $\psi_i = \phi$ with

$$\phi(a, b) = 0 \qquad \text{if and only if} \qquad a, b \ge 0,\ ab = 0.$$

For details on the variety of available NCP-functions, the reader is referred to [5, 47]. See also section 7. In [1, 16] it is described how special NCP-functions can be used to construct MCP-functions. In this paper, however, we will not work with this approach. Rather, we prefer to introduce a new class of NCP- and MCP-functions that have strong theoretical properties and perform very well in practice, see the numerical results in section 8. These functions are closely related to affine-scaling techniques in nonlinear programming [7, 8, 24, 50, 51] and can be used to interpret affine-scaling Newton methods as special semismooth Newton methods. Given $\varepsilon > 0$, these new NCP-functions can be designed in such a way that $\phi(a, b) = \gamma ab$ holds with some constant $\gamma > 0$ for all $a, b > 0$ with $a + b > \varepsilon$. This is interesting also from the viewpoint of primal-dual interior-point methods, since there the perturbed complementarity condition is usually expressed in the form $a > 0$, $b > 0$, $ab = \mu$, with $\mu > 0$. Given the close relationship of the functions ϕ and $(a, b) \mapsto ab$ on the positive orthant, it might be attractive to try to get rid of the positivity requirements on a and b by choosing appropriate perturbations ϕ_μ of $\phi_0 = \phi$ and replacing the perturbed complementarity condition by $\phi_\mu(a, b) = 0$.

Most of the available literature on algorithms for problems of the form (1.1) focuses on special cases like nonsmooth reformulations of NCPs. Line search methods for problems (1.1) arising from reformulated NCPs and KKT-systems of variational inequality problems (VIPs) were analyzed in, e.g., [11, 16, 29]. Since x is a zero of the function H defined in (1.4) if and only if x solves $\mathrm{MCP}(F, X)$, it is possible to omit the box-constraint in (1.1). Below, we discuss this approach and mention some of its potential drawbacks. Line search methods for these unconstrained reformulations are investigated by several authors, see [11, 14, 27, 52, 53].

For unconstrained semismooth equations arising from reformulations of NCPs, trust-region algorithms were analyzed in [26, 30]. A trust-region method for a box-constrained reformulation without NCP-function was investigated in [37] under a strict complementarity condition. This assumption is not needed for our analysis. Moreover, we stress that the results in [26, 30, 37] are established for monotone trust-region methods only. Other approaches for the solution of the NCP, MCP, or VIP can be found in [3, 12, 39, 40]. For a survey, see [2, 15]. Algorithms for more general classes of nonsmooth equations are investigated in [21, 23, 41, 44, 45, 46, 52, 53].

Among the methods cited above, the trust-region algorithms in [26, 30] are probably the ones closest related to the class of methods proposed in this work. However, there are several important differences. In particular, we deal with a more general class of nonsmooth equations. Moreover, we allow for box-constraints and our algorithm generates feasible iterates with respect to these constraints. Numerical studies [5, 22, 49, 30] have shown that the performance of optimization methods for the solution of minimization problems can be significantly improved by using non-monotone line search- or trust-region techniques. Especially for problems with least-squares objective functions like (1.2), non-monotonicity helps to prevent from convergence to local-nonglobal solutions of (1.2). In this paper we introduce a new non-monotone trust-region technique and develop a global convergence theory that covers essentially all results that are known for monotone trust-region algorithms. These results appear to be new also for the special case of smooth problems (1.1). For other approaches to non-monotone line search and trust-region techniques we refer to [22, 49] and the references therein. Concerning literature on monotone trust-region methods for optimization problems with simple (or, more generally, convex) constraints we refer to [4, 8, 9, 20, 33, 48, 51].

Especially for problems (1.1) obtained from reformulating the MCP on the basis of (1.4) one might ask why we keep the box constraint although $H(x) \neq 0$ holds by definition for all $x \notin X$. More generally, if $\bar{x}$ is a solution to (1.1) then, obviously, it also solves the unconstrained semismooth equation

$$H(x) = 0. \tag{1.5}$$

The corresponding unconstrained counterpart to (1.2) is

$$\text{minimize } \ h(x). \tag{1.6}$$

We stress, however, that (1.5) and (1.6) contain the implicit constraint $x \in U$, which can be quite unstructured and complicated. There are good reasons to prefer the constrained formulations (1.1), (1.2) to the unconstrained ones (1.5) and (1.6). For instance, there might exist solutions to (1.5) that do not solve (1.1) since they are not feasible with respect to X. This cannot occur if $H(x) \neq 0$ for all $x \notin X$, cf. (1.4). Sometimes H is known to have nice properties on X (like, e.g., positive definiteness of the Jacobian $H'(x)$), but not outside of X. When working with (1.2) instead of (1.6) this can help to reduce the risk of finding a local, but non-global solution of the minimization problem only. Furthermore, in many applications the domain U of H is only implicitly available in the sense that H is numerically implemented as an oracle that returns the value $H(x)$ if $x \in U$ and an error message, otherwise. Therefore, algorithms for the solution of problem (1.6) can run into trouble by accidentally coming too close to the boundary of U. On the other hand, if a feasible point algorithm is applied to the reformulation (1.2), then it will never interfere with the boundary of U, since $X \subset U$ and U is open.

From the above discussion we conclude that algorithms that are based on the simply constrained problems (1.1) and (1.2) can, in fact, be more robust and efficient than those derived from the easier looking problems (1.5) and (1.6). This is confirmed by the numerical experiments in section 8.

The rest of this paper is organized in seven sections. In section 2 we collect important results of nonsmooth analysis needed for our investigations. The trust-region algorithm is developed in section 3. Hereby, we begin in section 3.1 with the underlying Newton-type method, which we augment by a projection to obtain feasible iterates. Then, in section 3.2 this iteration is embedded into a globally convergent non-monotone trust-region method. We proceed in section 4 by proving the global convergence of this algorithm. In section 5, the locally q-superlinear convergence and convergence of q-order $1+p$, respectively, are established. A concrete implementation of the decrease condition by means of a Cauchy step is discussed in section 6. Section 7 is devoted to the application of the developed method the the nonlinear mixed complementarity problem. In section 7.1 we derive important properties of the semismooth reformulation of the MCP. We use a new MCP-function, which we introduce and analyze in section 7.2. Numerical results for a subset of the MCPLIB test set [13] are reported in section 8. In the appendix we describe the computation of B-subdifferentials.

Notations. The ℓ_2-norm on $\mathbb{R}^n$ is denoted by $\|\cdot\|$. Given a set $M \subset \mathbb{R}^n$ and a point $x \in \mathbb{R}^n$, we set $M - x \stackrel{\text{def}}{=} \{y\,;\ x + y \in M\}$. $S^0(U, \mathbb{R}^n)$ denotes the set of all semismooth functions $f : \mathbb{R}^n \supset U \to \mathbb{R}^n$. $S^p(U, \mathbb{R}^n)$, $0 < p \leq 1$, is the set of all p-order semismooth functions. We write $f'(x, \cdot) : \mathbb{R}^n \to \mathbb{R}^m$ for the directional derivative, $f'(x) \in \mathbb{R}^{m\times n}$ for the Jacobian, $\partial_B f(x) \subset \mathbb{R}^{m\times n}$ for the B-subdifferential, and $\partial f(x) \in \mathbb{R}^{m\times n}$ for Clarke's generalized Jacobian of the function $f : \mathbb{R}^n \supset U \to \mathbb{R}^m$ at the point $x \in U$ (in case the respective objects exist). $\nabla f(x) = f'(x)^T$ denotes the gradient of the differentiable, real-valued function f at x.

2. Some Notions of Nonsmooth Analysis.

For convenience, we collect here all facts about nonsmooth analysis that are required for our investigations. Readers familiar with these concepts might want to skip this section.

Throughout, let $f : U \to \mathbb{R}^m$ be locally Lipschitz continuous on the nonempty open set $U \subset \mathbb{R}^n$. By D_f we denote the set of all $x \in U$ where f admits a (Fréchet-) derivative $f'(x) \in \mathbb{R}^{m\times n}$. According to Rademacher's Theorem [54], $U \setminus D_f$ has Lebesgue-measure zero. Hence, the following constructions make sense.

DEFINITION 2.1. [6, 41, 46] The set

$$\partial_B f(x) \stackrel{\text{def}}{=} \left\{V \in \mathbb{R}^{m\times n};\ \exists (x_k) \subset D_f :\ x_k \to x,\ f'(x_k) \to V\right\}$$

is called B-subdifferential of f at $x \in U$. Moreover, Clarke's generalized Jacobian of f at x is the set $\partial f(x) \stackrel{\text{def}}{=} \operatorname{conv}(\partial_B f(x))$. □

We collect some properties of $\partial_B f$ and ∂f.

PROPOSITION 2.2 ([6, Prop. 2.6.2]). *For all $x \in U$ holds:*

(a) *$\partial_B f(x)$ is nonempty and compact.*
(b) *$\partial f(x)$ is nonempty, compact, and convex.*
(c) *The setvalued mappings $\partial_B f$ and ∂f, respectively, are locally bounded and upper semicontinuous.*

Next, we recall the directional derivative and define semismoothness.

DEFINITION 2.3. [36, 41, 46]

(a) f is directionally differentiable at $x \in U$ if the directional derivative

$$f'(x,s) \stackrel{\text{def}}{=} \lim_{\tau\to 0^+} \frac{f(x+\tau s) - f(x)}{\tau}$$

exists for all $s \in \mathbb{R}^n$.

(b) f is semismooth at $x \in U$ if the following limit exists for all $s \in \mathbb{R}^n$:

$$\lim_{\substack{V\in\partial F(x+\tau d)\\ d\to s,\ \tau\to 0^+}} Vd.$$

(c) By $S^0(U,\mathbb{R}^m)$ we denote the set of all locally Lipschitz functions $f : U \to \mathbb{R}^m$ that are semismooth on U.

□

Note that $f'(x,\cdot)$ is positive homogeneous. The following Proposition gives an alternative definition of semismoothness.

PROPOSITION 2.4. [46, Thm. 2.3] *For $x \in U$ the following statements are equivalent:*

(a) *f is semismooth at x,*
(b) *$f'(x,\cdot)$ exists and*

$$\sup_{V\in\partial f(x+s)} \|Vs - f'(x,s)\| = o(\|s\|) \quad \text{as } s \to 0.$$

COROLLARY 2.5. *If f is semismooth at x then for all $s \in \mathbb{R}^n$*

$$f'(x,s) = \lim_{\substack{V\in\partial F(x+\tau s)\\ \tau\to 0^+}} Vs.$$

Based on Proposition 2.4 (b), a semismooth relaxation of Hölder-continuous differentiability can be established.

DEFINITION 2.6. [46] Let $0 < p \le 1$. The function f is called p-order semismooth at $x \in U$, if $f'(x,\cdot)$ exists and

$$\sup_{V \in \partial f(x+s)} \|Vs - f'(x,s)\| = O(\|s\|^{1+p}) \quad \text{as } s \to 0.$$

By $S^p(U,\mathbb{R}^m)$ we denote the set of all locally Lipschitz functions $f : U \to \mathbb{R}^m$ that are p-order semismooth on U. □

PROPOSITION 2.7. [46] *If f is semismooth at $x \in U$, then*

$$\|f(x+s) - f(x) - f'(x,s)\| = o(\|s\|) \quad \text{as } s \to 0.$$

If $0 < p \le 1$ and f is p-order semismooth at $x \in U$, then

$$\|f(x+s) - f(x) - f'(x,s)\| = O(\|s\|^{1+p}) \quad \text{as } s \to 0.$$

PROPOSITION 2.8. [19, Lem. 18 and Thm. 21] *Let $U_1 \subset \mathbb{R}^n$ and $U_2 \subset \mathbb{R}^l$ be open sets and $f_1 : U_1 \to U_2$, $f_2 : U_2 \to \mathbb{R}^m$ be locally Lipschitz mappings. Then, if f_1 is (p-order, $0 < p \le 1$) semismooth at $x \in U_1$ and f_2 is (p-order) semismooth at $f_1(x)$, the composite map $f \stackrel{\text{def}}{=} f_2 \circ f_1 : U_1 \to \mathbb{R}^m$ is (p-order) semismooth at x. Moreover,*

$$f'(x,\cdot) = f_2'(f_1(x), f_1'(x,\cdot)).$$

The following is obvious.

PROPOSITION 2.9. *If f is continuously differentiable in a neighborhood of $x \in U$ (with p-Hölder continuous derivative, $0 < p \le 1$), then f is (p-order) semismooth at x and $\partial f(x) = \partial_B f(x) = \{f'(x)\}$.*

Next, we describe how two C^1-functions can be glued together to obtain a new semismooth function.

LEMMA 2.10. *Let $j \in \{1,2\}$, $y \in \mathbb{R}$, and let the functions $f_i : U \to \mathbb{R}$, $i = 1,2$, be continuously differentiable on the open set $U \subset \mathbb{R}^2$ such that $f_1(x) = f_2(x)$ for all $x \in U$ with $x_j = y$. Define the function $f = \mathrm{pw}(j,y;f_1,f_2) : U \to \mathbb{R}$,*

$$f(x) = f_1(x), \quad \text{if } x_j \le y, \quad f(x) = f_2(x), \quad \text{otherwise.}$$

Then f is semismooth and

$$\partial_B f(x) = \begin{cases} \{\nabla f_1(x)^T\} & \text{if } x_j < y, \\ \{\nabla f_2(x)^T\} & \text{if } x_j > y, \\ \{\nabla f_1(x)^T, \nabla f_2(x)^T\} & \text{if } x_j = y. \end{cases}$$

If the derivatives of f_1 and f_2 are p-order Hölder continuous, $0 < p \le 1$, then f is p-order semismooth.

Proof. The local Lipschitz continuity of f is obvious. Moreover, the assertion on the form of the B-subdifferential is easily established.

For $x \in U$ with $x_j < y$ ($x_j > y$) we have $f = f_1$ ($f = f_2$) in a neighborhood. In this case, the assertions follow from Proposition 2.9.

Now let $x_j = y$. We fix an arbitrary $s \in \mathbb{R}^2$ with $z \stackrel{\text{def}}{=} x + s \in U$. If $s_j < 0$ then $\partial f(z) = \{\nabla f_1(z)^T\}$, $f'(x,s) = f_1'(x,s) = \nabla f_1(x)^T s$, and

$$\nabla f_1(z)^T s - f'(x,s) = (\nabla f_1(z) - \nabla f_1(x))^T s = o(\|s\|) \quad \text{as } s \to 0.$$

Similarly, for $s_j > 0$ we have $\partial f(z) = \{\nabla f_2(z)^T\}$ and

$$\nabla f_2(z)^T s - f'(x,s) = (\nabla f_2(z) - \nabla f_2(x))^T s = o(\|s\|) \quad \text{as } s \to 0.$$

For $s_j = 0$ holds $\partial f(z) = \{(\lambda \nabla f_1(z) + (1-\lambda)\nabla f_2(z))^T\,;\; 0 \le \lambda \le 1\}$ and $f'(x,s) = \nabla f_1(x)^T s = \nabla f_2(x)^T s$, and thus for arbitrary $v^T \in \partial f(z)$ and appropriate $0 \le \lambda \le 1$

$$v^T s - f'(x,s) = \lambda(\nabla f_1(z) - \nabla f_1(x))^T s + (1-\lambda)(\nabla f_2(z) - \nabla f_2(x))^T s = o(\|s\|)$$

as $s \to 0$. If the derivatives of f_1 and f_2 are p-order Hölder continuous, we can replace '$o(\|s\|)$' by '$O(\|s\|^{1+p})$' in the above estimates. This proves the (p-order) semismoothness of f at x. □

The following regularity property is essential for fast local convergence of Newton-like methods.

DEFINITION 2.11. [41] The point $x \in U$ is called BD-regular for f if all elements in $\partial_B f(x)$ are nonsingular. □

PROPOSITION 2.12. [41, Prop. 3] *Let $x \in U$ be a BD-regular for f. Then there exist $\varepsilon > 0$ and $C > 0$ such that all $V \in \partial_B f(y)$, $\|y - x\| \le \varepsilon$, are nonsingular with $\|V^{-1}\| \le C$. If, in addition, f is semismooth at x then there exist $\delta > 0$ and $\zeta > 0$ such that*

$$\|f(y) - f(x)\| \ge \zeta \|y - x\|$$

for all $y \in \mathbb{R}^n$, $\|y - x\| \le \delta$.

3. Development of the Algorithm.

As motivated above, our algorithm for the solution of (1.1) will be based on the reformulation

$$\text{minimize } h(x) \quad \text{subject to} \quad l \le x \le u.$$

Basically, the concept of trust-region methods is to make Newton's method globally convergent while maintaining its excellent local convergence behavior. Therefore, we begin the description of our algorithm with its core, the underlying Newton-like iteration.

3.1. A Newton-like Method with Projection.

In order to generate feasible iterates for (1.1) we introduce a Newton-like method that is augmented by the projection onto $X = [l, u]$. In the sequel, given a nonempty closed convex set $C \subset \mathbb{R}^n$, the mapping $P_C : \mathbb{R}^n \to C$

$$P_C(x) \stackrel{\text{def}}{=} \operatorname*{argmin}_{y \in C} \|y - x\|$$

denotes the projection onto C. Since $X = [l, u]$ is a box, P_X can be easily computed:

$$P_X(x) = \max\{l, \min\{x, u\}\} \quad \text{(componentwise)}.$$

We now can formulate the algorithm.

ALGORITHM 3.1 (NEWTON-LIKE ITERATION WITH PROJECTION).

1. Choose x_0 and set $k := 0$.
2. If $H(x_k) = 0$ then STOP.
3. Choose a nonsingular matrix $M_k \in \mathbb{R}^{n \times n}$ and compute the Newton-like step s_k^{N} by solving

$$M_k s_k^{\mathrm{N}} = -H(x_k),$$

4. Compute the projection of s_k^{N} onto $X - x_k$:

$$s_k^{\mathrm{PN}} := P_X(x_k + s_k^{\mathrm{N}}) - x_k.$$

5. Set $x_{k+1} := x_k + s_k^{\mathrm{PN}}$, $k := k + 1$, and go to Step 2.

Since $x_0 \in X$ and M_0 is invertible, the steps s_0^{N} and s_0^{PN} are well-defined. Moreover, $x_1 = P_X(x_0 + s_0^{\mathrm{N}}) \in X$. By iterating this argument we obtain the well-definedness of Algorithm 3.1 and the feasibility of the iterates x_k with respect to X.

Without the augmentation by a projection the local convergence properties of Newton-like iterations for semismooth equations were investigated in, e.g., [41, 44, 46]. Newton methods for nonsmooth equations are also discussed in [31, 32, 38, 39, 43]. The additional projection does not affect the convergence speed since it is Lipschitzian of rank 1 [25, p. 118]. We will give a full proof of the local convergence result for Algorithm 3.1 since it is instructive to see how semismoothness, BD-regularity and other concepts come into play, and, more importantly, because we will need all the estimates established in this proof for the more involved global-to-local analysis in Theorem 5.1.

THEOREM 3.2. *Assume that $H : U \to \mathbb{R}^n$ is locally Lipschitz continuous. Let $\bar{x} \in X$ be a BD-regular zero of H at which H is semismooth. Let $\|x_0 - \bar{x}\|$ and $\delta > 0$ be sufficiently small. Assume that Algorithm* 3.1 *generates infinitely many iterates, and that for all k holds*

$$\mu_k \stackrel{\mathrm{def}}{=} \min_{V \in \partial_B H(x_k)} \|(M_k - V)s_k^{\mathrm{N}}\| \leq \delta \|s_k^{\mathrm{N}}\|, \tag{3.1}$$

Then (x_k) converges to $\bar{x}$.

If in addition

$$\lim_{k \to \infty} \frac{\mu_k}{\|s_k^{\mathrm{N}}\|} = 0, \tag{3.2}$$

then the sequence (x_k) converges q-superlinearly to $\bar{x}$.

If H is p-order semismooth at $\bar{x}$, $0 < p \leq 1$, and

$$\limsup_{k \to \infty} \frac{\mu_k}{\|s_k^{\mathrm{N}}\|^{1+p}} < \infty, \tag{3.3}$$

then (x_k) converges with q-order $1 + p$ to $\bar{x}$.

Proof. By the local Lipschitz continuity, the BD-regularity of $\bar{x}$ and Proposition 2.12, there exist $\varepsilon > 0$, $L > 0$, and $C > 0$ such that H is Lipschitz continuous on $B_\varepsilon(\bar{x}) \stackrel{\mathrm{def}}{=} \{x \in \mathbb{R}^n\,;\ \|x - \bar{x}\| \leq \varepsilon\}$ of rank L, and

$$\|V^{-1}\| \leq C \quad \forall\, V \in \partial_B H(x),\ \forall\, x \in B_\varepsilon(\bar{x}). \tag{3.4}$$

Throughout the proof let $x_k \in B_\varepsilon(\bar{x})$ be arbitrary and let $V_k \in \partial_B H(x_k)$ be such that $\mu_k = \|(M_k - V_k)s_k^{\mathrm{N}}\|$. We define

$$d_k \stackrel{\mathrm{def}}{=} x_k - \bar{x}, \quad e_k \stackrel{\mathrm{def}}{=} x_k + s_k^{\mathrm{N}} - \bar{x}.$$

If $\delta \leq 1/(2C)$ then

$$\begin{aligned} \|s_k^{\mathrm{N}}\| &\leq \|V_k{}^{-1}\| \left(\|M_k s_k^{\mathrm{N}}\| + \|(V_k - M_k)s_k^{\mathrm{N}}\|\right) \leq C\|H(x_k)\| + C\mu_k \\ &\leq C\|H(x_k)\| + C\delta\|s_k^{\mathrm{N}}\| \leq C\|H(x_k)\| + \frac{1}{2}\|s_k^{\mathrm{N}}\|. \end{aligned} \tag{3.5}$$

In particular,

$$\|s_k^{\mathrm{N}}\| \le 2C\|H(x_k)\| \le 2CL\|d_k\|. \tag{3.6}$$

Further,

$$\begin{aligned} V_k e_k &= V_k s_k^{\mathrm{N}} + V_k d_k = M_k s_k^{\mathrm{N}} + (V_k - M_k)s_k^{\mathrm{N}} + V_k d_k \\ &= -H(x_k) + V_k d_k + (V_k - M_k)s_k^{\mathrm{N}} \\ &= (V_k d_k - H'(\bar{x}, d_k)) + (-H(x_k) + H(\bar{x}) + H'(\bar{x}, d_k)) \\ &\quad + (V_k - M_k)s_k^{\mathrm{N}}. \end{aligned} \tag{3.7}$$

If δ and ε are sufficiently small, we obtain from (3.4), Proposition 2.4(b), Proposition 2.7, (3.1), and (3.6) that

$$\|e_k\| \le \frac{1}{2}\|d_k\|.$$

Since $\bar{x} \in X$ and P_X is Lipschitz continuous of rank 1, we thus have

$$\begin{aligned} \|d_{k+1}\| &= \|P_X(x_k + s_k^{\mathrm{N}}) - \bar{x}\| = \|P_X(x_k + s_k^{\mathrm{N}}) - P(\bar{x})\| \\ &\le \|x_k + s_k^{\mathrm{N}} - \bar{x}\| = \|e_k\| \le \frac{1}{2}\|d_k\|. \end{aligned} \tag{3.8}$$

Therefore, it follows inductively that the sequence (x_k) converges to $\bar{x}$ if $\delta > 0$ and $\|x_0 - \bar{x}\|$ are sufficiently small.

If (3.2) holds, we see from (3.4), Proposition 2.4(b), and Proposition 2.7, that the right hand side of (3.7), and thus, by (3.4), also e_k is of the order $o(\|d_k\|)$. The q-superlinear convergence now follows from (3.8).

If H is p-order semismooth at $\bar{x}$ and if (3.3) holds, then the right hand side of (3.7) is obviously of the order $O(\|d_k\|^{1+p})$. The proof is completed as before. □

3.2. The Trust-Region Algorithm. We now wrap Algorithm 3.1 into a globally convergent trust-region method for problem (1.2). For the time being, let us take the continuous differentiability of the merit function h for granted. We return to this issue in Lemma 4.2. This amounts us to building the quadratic model

$$q_k(s) = g_k^T s + \frac{1}{2}\|M_k s\|^2,$$

around the current iterate x_k, where $g_k \stackrel{\text{def}}{=} \nabla h(x_k)$. q_k is an at least first-order accurate approximation of $h(x_k + s) - h(x_k)$. The matrices $M_k \in \mathbb{R}^{n\times n}$ are the same as in Algorithm 3.1. We stress, however, that the proposed trust-region method is globally convergent for much more general choices of M_k.

In each iteration of the trust-region algorithm, a trial step s_k is computed as approximate solution of the

Trust-Region Subproblem:

$$\text{minimize } q_k(s) \quad \text{subject to} \quad l \le x_k + s \le u, \ \|s\|_\infty \le \Delta_k. \tag{3.9}$$

This is a convex box-constrained QP with feasible set

$$X_k \stackrel{\text{def}}{=} [l - x_k, u - x_k] \cap [-\Delta_k, \Delta_k]^n.$$

We will assume that the trial steps meet the following two requirements:

Feasibility Condition:

$$l \le x_k + s_k \le u \quad \text{and} \quad \|s_k\|_\infty \le \beta_1 \Delta_k \tag{3.10}$$

Reduction Condition:

$$\text{pred}_k(s_k) \stackrel{\text{def}}{=} -q_k(s_k) \ge \beta_2 \chi(x_k) \min\{1, \Delta_k, \chi(x_k)\} \tag{3.11}$$

with constants $\beta_1 \ge 1$ and $\beta_2 > 0$ independent of k. Hereby, χ is a suitably chosen *criticality measure*:

$$\begin{aligned} &\chi : X \to \mathbb{R}_+ \text{ is continuous,} \\ &\chi(x) = 0 \text{ iff } x \text{ is a KKT-point of problem (1.2).} \end{aligned} \tag{3.12}$$

A well known criticality measure is the norm of the projected gradient

$$\chi(x) = \|x - P_X(x - \nabla h(x))\|.$$

We recall that $P_X(x) = \max\{l, \min\{x, u\}\}$ denotes the projection onto X. Usually, the update of the trust-region radius Δ_k is controlled by the ratio of *actual reduction*

$$\text{ared}_k(s) \stackrel{\text{def}}{=} h(x_k) - h(x_k + s)$$

and *predicted reduction* $\text{pred}_k \stackrel{\text{def}}{=} -q_k(s)$.

It has been observed [5, 22, 30, 49] that the performance of nonlinear programming algorithms can be significantly improved by using non-monotone line search- or trust-region techniques. Hereby, in contrast to the traditional approach, the monotonicity $h(x_{k+1}) \le h(x_k)$ of the function values is not enforced in every iteration. We introduce a new non-monotone trust-region technique for which all global convergence results for monotone methods remain valid. Hereby, the decrease requirement is significantly relaxed. Before we describe this approach and the corresponding reduction ratio $\rho_k(s)$ in detail, we first state the basic trust-region algorithm.

ALGORITHM 3.3 (TRUST-REGION ALGORITHM).

1. Initialization: Choose $\eta_1 \in (0,1)$, $\Delta_{\min} \ge 0$, and a criticality measure χ. Choose $x_0 \in X$, $\Delta_0 > \Delta_{\min}$, and a nonsingular matrix $M_0 \in \mathbb{R}^{n \times n}$. Choose an integer $m \ge 1$ and fix $\lambda \in (0, 1/m]$ for the computation of ρ_k. Set $k := 0$ and $i := -1$.
2. Compute $\chi_k := \chi(x_k)$. If $\chi_k = 0$ then STOP.
3. Compute a trial step s_k satisfying the conditions (3.10) and (3.11).
4. Compute the reduction ratio $\rho_k := \rho_k(s_k)$.
5. Compute the new trust-region radius Δ_{k+1} by invoking Algorithm 3.4.
6. If $\rho_k \le \eta_1$ then reject the step s_k, i.e., set $x_{k+1} := x_k$, $M_{k+1} := M_k$, increment k by 1, and go to Step 3.
7. Accept the step: Set $x_{k+1} := x_k + s_k$ and choose a nonsingular matrix $M_{k+1} \in \mathbb{R}^{n \times n}$. Set $j_{i+1} := k$, increment k and i by 1 and go to Step 2.

The increasing sequence $(j_i)_{i \ge 0}$, enumerates all indices of accepted steps. Moreover,

$$x_k = x_{j_i} \qquad \forall\, j_{i-1} < k \le j_i,\ \forall\, i \ge 1. \tag{3.13}$$

Conversely, if $k \ne j_i$ for all i then s_k was rejected. In the following we denote the set of all these 'successful' indices j_i by $\mathcal{S}$:

$$\mathcal{S} \stackrel{\text{def}}{=} \{j_i\,;\ i \ge 0\} = \{k\,;\ \text{trial step } s_k \text{ is accepted}\}\,. \tag{3.14}$$

Sometimes, accepted steps will also be called successful. We will repeatedly use the fact that

$$\{x_k\,;\ k \geq 0\} = \{x_k\,;\ k \in \mathcal{S}\}\,.$$

The trust-region updates are implemented as usual. We deal with two different flavors of update rules simultaneously by introducing a nonnegative parameter $\Delta_{\min}$. We require that after successful steps holds $\Delta_{k+1} \geq \Delta_{\min}$. If $\Delta_{\min} = 0$ is chosen, this holds automatically. For $\Delta_{\min} > 0$, however, it is an additional feature that allows for special proof techniques.

ALGORITHM 3.4 (UPDATE OF THE TRUST-REGION RADIUS).
Input: Δ_k, ρ_k. Output: Δ_{k+1}.
$\Delta_{\min} \geq 0$ and $\eta_1 \in (0,1)$ are the constants defined in Step 1 of Algorithm 3.3.
Let $\eta_1 < \eta_2 < 1$, and $0 < \gamma_0 < \gamma_1 < 1 < \gamma_2$ be fixed.

1. If $\rho_k \leq \eta_1$ then choose
   $\Delta_{k+1} \in (\gamma_0 \Delta_k, \gamma_1 \Delta_k]$.
2. If $\rho_k \in (\eta_1, \eta_2)$ then choose
   $\Delta_{k+1} \in [\gamma_1 \Delta_k, \max\{\Delta_{\min}, \Delta_k\}] \cap [\Delta_{\min}, \infty)$.
3. If $\rho_k \geq \eta_2$ then choose
   $\Delta_{k+1} \in (\Delta_k, \max\{\Delta_{\min}, \gamma_2 \Delta_k\}] \cap [\Delta_{\min}, \infty)$.

We still have to describe how the reduction ratios $\rho_k(s)$ are defined. Here is a detailed description of Step 4:

4.1. Compute $m_k := \min\{i+1, m\}$ and choose scalars

$$\lambda_{kr} \geq \lambda, \quad r = 0, \dots, m_k - 1, \qquad \sum_{r=0}^{m_k-1} \lambda_{kr} = 1.$$

4.2. Compute the relaxed actual reduction $\mathrm{rared}_k := \mathrm{rared}_k(s_k)$, where

$$\mathrm{rared}_k(s) \stackrel{\mathrm{def}}{=} \max\left\{ h(x_k), \sum_{r=0}^{m_k-1} \lambda_{kr} h(x_{j_{i-r}}) \right\} - h(x_k + s). \tag{3.15}$$

4.3. Compute the reduction ratio $\rho_k := \rho_k(s_k)$ according to

$$\rho_k(s) \stackrel{\mathrm{def}}{=} \frac{\mathrm{rared}_k(s)}{\mathrm{pred}_k(s)}. \tag{3.16}$$

REMARK 3.5. At the very beginning of the iteration, Step 4 is encountered with $i = -1$. In this case the sum in (3.15) is empty and thus

$$\mathrm{rared}_k(s) = \max\{h(x_k), 0\} - h(x_k + s) = h(x_k) - h(x_k + s) = \mathrm{pred}_k(s).$$

□

The idea behind the above update rule is the following: Instead of requiring that $h(x_k + s_k)$ be smaller than $h(x_k)$, it is only required that $h(x_k + s_k)$ is either less than $h(x_k)$ or less than the weighted mean of the function values at the last $m_k = \min\{i+1, m\}$ successful iterates. Of course, if $m = 1$, then $\mathrm{rared}_k(s) = \mathrm{ared}_k(s)$ and the usual reduction ratio is recovered. Our approach is a slightly stronger requirement than the straightforward idea to replace pred_k with

$$\mathrm{rared}_k^{\infty}(s) = \max_{0 \leq r < m_k} h(x_{j_{i-r}}) - h(x_k + s). \tag{3.17}$$

Unfortunately, for this latter choice it does not seem to be possible to establish all the global convergence results that are available for the monotone case. For our approach, however, this is possible without making the theory substantially more difficult. Moreover, we can approximate rared_k^∞ arbitrarily accurately by rared_k if we choose λ sufficiently small, in each iteration select $0 \le r_k < m_k$ satisfying $h(x_{j_{i-r_k}}) = \max_{0\le r<m_k} h(x_{j_{i-r}})$, and set

$$\lambda_{kr} = \lambda \ \text{ if } r \neq r_k, \quad \lambda_{kr_k} = 1 - (m_k - 1)\lambda. \tag{3.18}$$

To obtain a globally *and* locally fast convergent algorithm, we will embed the Newton-like algorithm 3.1 into the trust-region algorithm 3.3 by using it to compute trial steps s_k^{P}. Since we need trial steps that satisfy the feasibility condition (3.10), we compute s_k^{P} slightly different than s_k^{PN}. Whereas s_k^{PN} is the projection of $s_k^{\mathrm{N}} = -M_k^{-1}H(x_k)$ onto $X - x_k$, we obtain s_k^{P} by projection onto the feasible set X_k of the trust-region subproblem (3.9):

$$s_k^{\mathrm{P}} := P_{X_k}(s_k^{\mathrm{N}}).$$

Since we will show that finally holds $s_k^{\mathrm{P}} = s_k^{\mathrm{PN}}$, this modification does not change the local convergence behavior. If s_k^{P} satisfies the decrease condition (3.11), we choose $s_k = s_k^{\mathrm{P}}$ as trial step. Otherwise, a different trial step verifying (3.10) and (3.11) must be computed. In section 6 we propose a general way to do this. We thus arrive at the final version of the algorithm.

ALGORITHM 3.6 (TRUST-REGION PROJECTED-NEWTON ALGORITHM).
As Algorithm 3.3, but with Step 3 implemented as follows:

3.1. If coming from Step 6 then set $s_k^{\mathrm{N}} := s_{k-1}^{\mathrm{N}}$. Otherwise compute the Newton-like step s_k^{N} by solving $M_k s_k^{\mathrm{N}} = -H(x_k)$.
3.2. Compute the projected Newton step $s_k^{\mathrm{P}} := P_{X_k}(s_k^{\mathrm{N}})$.
3.3. If $s_k = s_k^{\mathrm{P}}$ satisfies the decrease condition (3.11), then set $s_k := s_k^{\mathrm{P}}$. Otherwise, compute a step s_k satisfying (3.10) and (3.11).

4. Global Convergence. We first establish the continuous differentiability of the merit function h. The proof is based on the following Lemma.

LEMMA 4.1. *Let $f : U \to \mathbb{R}$ be locally Lipschitz continuous on the nonempty open set $U \subset \mathbb{R}^n$. Assume that f is continuously differentiable on $U \setminus f^{-1}(0)$. Then the function f^2 is continuously differentiable on U. Moreover, $\nabla f^2(x) = 2f(x)v^T$ for all $v \in \partial f(x)$ and all $x \in U$.*

Proof. By assumption, f is continuously differentiable on the open set $U \setminus f^{-1}(0)$. Therefore, f^2 is C^1 on $U \setminus f^{-1}(0)$ with gradient $\nabla f^2(x) = 2f(x)\nabla f(x)$. Moreover, we have $\partial f(x) = \{\nabla f(x)^T\}$.

Now let $x_0 \in U$ be arbitrary with $f(x_0) = 0$. Since U is open and f is locally Lipschitz continuous, there exist $\varepsilon > 0$ and $L > 0$ such that $B_\varepsilon(x_0) \stackrel{\text{def}}{=} \{x\,;\ \|x - x_0\| \le \varepsilon\} \subset U$ and f is Lipschitz continuous on $B_\varepsilon(x_0)$ of rank L. Hence, for all $x \in B_\varepsilon(x_0)$ holds

$$|f^2(x) - f^2(x_0)| = |f^2(x)| = |f(x) - f(x_0)|^2 \le L^2\|x - x_0\|^2.$$

This shows that f^2 is differentiable at x_0 with gradient $\nabla f^2(x_0) = 0$. The Lipschitz continuity on $B_\varepsilon(x_0)$ implies $\|\nabla f(x)\| \le L$ for all $x \in B_\varepsilon(x_0)$ with $f(x) \neq 0$. Now let $x_k \in B_\varepsilon(x_0)$ tend to x_0. If $f(x_k) = 0$ then $\nabla f^2(x_k) = 0$. If $f(x_k) \neq 0$, we obtain

$$\|\nabla f^2(x_k)\| = 2\|\nabla f(x_k)\|\,|f(x_k)| \le 2L|f(x_k) - f(x_0)| \le 2L^2\|x_k - x_0\|.$$

Therefore, f^2 is continuously differentiable at x_0 with $\nabla f(x_0) = 0$. From $f(x_0) = 0$ we obtain $\nabla f^2(x_0) = 0 = 2f(x_0)v^T$ for all $v \in \partial f(x_0)$. The proof is finished. □

LEMMA 4.2. *Under the assumptions* (A1) *and* (A2) *on the mapping* H, *the function* $h(x) = \|H(x)\|^2/2$ *is continuously differentiable on* U *with gradient* $\nabla h(x) = V^T H(x)$, *where* $V \in \partial H(x)$ *is arbitrary. In particular,* (A1) *and* (A2) *imply* (A2').

Proof. For $1 \le i \le n$, the component function H_i of H is semismooth and thus locally Lipschitz continuous on U. Moreover, it is C^1 on $U \setminus H_i^{-1}(0)$. Therefore, H_i^2 is C^1 on U by Lemma 4.1. The same then holds true for $h(x) = \frac{1}{2}\sum H_i^2(x)$. Furthermore, for all $V = (v_1, \dots, v_n)^T \in \partial H(x)$ holds $v_i \in \partial H_i(x)$, $1 \le i \le n$, see [6, Prop. 2.6.2.(e)], and thus by Lemma 4.1

$$\nabla h(x) = \sum_{i=1}^{n} \nabla H_i^2(x) = \sum_{i=1}^{n} H_i(x) v_i = V^T H(x).$$

□

In the next Lemma an important decrease property of the function values $h(x_k)$ is established.

LEMMA 4.3. *Let x_k, s_k, Δ_k, j_i, etc., be generated by Algorithm* 3.3. *Then for all computed indices $i \ge 1$ holds*

$$h(x_{j_i}) < h(x_0) - \eta_1 \lambda \sum_{r=0}^{i-2} \mathrm{pred}_{j_r}(s_{j_r}) - \eta_1 \mathrm{pred}_{j_{i-1}}(s_{j_{i-1}}). \tag{4.1}$$

Proof. We will use the short notations $\mathrm{ared}_k = \mathrm{ared}_k(s_k)$, $\mathrm{rared}_k = \mathrm{rared}_k(s_k)$, and $\mathrm{pred}_k = \mathrm{pred}_k(s_k)$. First, let us note that (3.11) implies $\mathrm{pred}_k > 0$.

The rest of the proof is by induction. For $i = 1$ we have by (3.13) and using $\rho_{j_0}(s_{j_0}) > \eta_1$

$$h(x_{j_1}) = h(x_{j_0+1}) = h(x_{j_0}) - \mathrm{ared}_{j_0} < h(x_{j_0}) - \eta_1 \mathrm{pred}_{j_0} = h(x_0) - \eta_1 \mathrm{pred}_{j_0}.$$

Now assume that (4.1) holds for $1, \dots, i$.

If $\mathrm{rared}_{j_i} = \mathrm{ared}_{j_i}$ then, using (4.1) and $\lambda \le 1$,

$$\begin{aligned} h(x_{j_{i+1}}) = h(x_{j_i+1}) &= h(x_{j_i}) - \mathrm{ared}_{j_i} = h(x_{j_i}) - \mathrm{rared}_{j_i} \\ &< h(x_0) - \eta_1 \lambda \sum_{r=0}^{i-2} \mathrm{pred}_{j_r} - \eta_1 \mathrm{pred}_{j_{i-1}} - \eta_1 \mathrm{pred}_{j_i} \\ &\le h(x_0) - \eta_1 \lambda \sum_{r=0}^{i-1} \mathrm{pred}_{j_r} - \eta_1 \mathrm{pred}_{j_i}. \end{aligned}$$

If $\mathrm{rared}_{j_i} \ne \mathrm{ared}_{j_i}$ then $\mathrm{rared}_{j_i} > \mathrm{ared}_{j_i}$, and with $q = \min\{i, m-1\}$ we obtain

$$\begin{aligned} h(x_{j_{i+1}}) = h(x_{j_i+1}) &= \sum_{p=0}^{q} \lambda_{j_i p} h(x_{j_{i-p}}) - \mathrm{rared}_{j_i} \\ &< \sum_{p=0}^{q} \lambda_{j_i p} \left(h(x_0) - \eta_1 \lambda \sum_{r=0}^{i-p-2} \mathrm{pred}_{j_r} - \eta_1 \mathrm{pred}_{j_{i-p-1}} \right) - \eta_1 \mathrm{pred}_{j_i}. \end{aligned}$$

Using $\{0,\dots,q\}\times\{0,\dots,i-q-2\}\subset\{(p,r)\,;\ 0\le p\le q,\ 0\le r\le i-p-2\}$, $\lambda_{j_i0}+\cdots+\lambda_{j_iq}=1$, and $\lambda_{j_ip}\ge\lambda$, we can proceed

$$\begin{aligned}
h(x_{j_{i+1}}) &< h(x_0)-\eta_1\lambda\sum_{r=0}^{i-q-2}\left(\sum_{p=0}^{q}\lambda_{j_ip}\right)\mathrm{pred}_{j_r}-\eta_1\lambda\sum_{p=0}^{q}\mathrm{pred}_{j_{i-p-1}}-\eta_1\mathrm{pred}_{j_i}\\
&\le h(x_0)-\eta_1\lambda\sum_{r=0}^{i-q-2}\mathrm{pred}_{j_r}-\eta_1\lambda\sum_{r=i-q-1}^{i-1}\mathrm{pred}_{j_r}-\eta_1\mathrm{pred}_{j_i}\\
&= h(x_0)-\eta_1\lambda\sum_{r=0}^{i-1}\mathrm{pred}_{j_r}-\eta_1\mathrm{pred}_{j_i}.
\end{aligned}$$

□

LEMMA 4.4. *Let x_k, s_k, Δ_k, etc., be generated by Algorithm 3.3. Then for arbitrary $x\in X$ with $\chi(x)\ne 0$ and $0<\eta<1$ there exist $\Delta>0$ and $\delta>0$ such that*

$$\rho_k\ge\eta$$

holds whenever $\|x_k-x\|\le\delta$ and $\Delta_k\le\Delta$ is satisfied.

Proof. Since $\chi(x)\ne 0$, by continuity there exist $\delta>0$ and $\varepsilon>0$ such that $\chi(x_k)\ge\varepsilon$ for all k with $\|x_k-x\|\le\delta$. Now, for $0<\Delta\le\min\{1,\varepsilon\}$ and any k with $\|x_k-x\|\le\delta$ and $0<\Delta_k\le\Delta$, we obtain from the decrease condition (3.11):

$$\mathrm{pred}_k(s_k)=-q_k(s_k)\ge\beta_2\chi(x_k)\min\{1,\Delta_k,\chi(x_k)\}\ge\beta_2\varepsilon\Delta_k.$$

In particular, by (3.10)

$$\|s_k\|_\infty\le\beta_1\Delta_k\le\frac{\beta_1}{\beta_2\varepsilon}\mathrm{pred}_k(s_k).$$

Further, with appropriate $\tau_k\in[0,1]$

$$\begin{aligned}
\mathrm{ared}_k(s_k) &= h(x_k)-h(x_k+s_k)=-\nabla h(x_k+\tau_k s_k)^T s_k\\
&= -q_k(s)+(g_k-\nabla h(x_k+\tau_k s_k))^T s_k+\frac{1}{2}\|M_k s_k\|^2\\
&\ge \mathrm{pred}_k(s_k)+(g_k-\nabla h(x_k+\tau_k s_k))^T s_k.
\end{aligned}$$

By reducing Δ and δ, if necessary, we can bring x_k and $x_k+\tau_k s_k$ sufficiently close to x, and s_k sufficiently close to zero to achieve that

$$|(g_k-\nabla h(x_k+\tau_k s_k))^T s_k|\le(1-\eta)\frac{\beta_2\varepsilon}{\beta_1}\|s_k\|_\infty\le(1-\eta)\mathrm{pred}_k(s_k)$$

for all k with $\|x_k-x\|\le\delta$ and $0<\Delta_k\le\Delta$. Hence, for all these k holds

$$\mathrm{rared}_k(s_k)\ge\mathrm{ared}_k(s_k)\ge\mathrm{pred}_k(s_k)-|(g_k-\nabla h(x_k+\tau_k s_k))^T s_k|\ge\eta\,\mathrm{pred}_k(s_k).$$

The proof is complete. □

LEMMA 4.5. *Algorithm 3.3 either terminates after finitely many steps with a KKT-point x_k of (1.2) or generates an infinite sequence (s_{j_i}) of accepted steps.*

Proof. Assume that Algorithm 3.3 neither terminates nor generates an infinite sequence (s_{j_i}) of accepted steps. Then there exists a smallest index k_0 such that all steps s_k

are rejected for $k \geq k_0$. In particular, $x_k = x_{k_0}$, $k \geq k_0$, and the sequence of trust-region radii Δ_k tends to zero as $k \to \infty$, because

$$\Delta_{k_0+j} \leq \gamma_1^j \Delta_{k_0}.$$

Since the algorithm does not terminate, we know that $\chi(x_{k_0}) \neq 0$. But now Lemma 4.4 with $x = x_{k_0}$ yields that s_k is accepted as soon as Δ_k becomes sufficiently small. This contradicts our assumption. Therefore, the assertion of the Lemma is true. □

LEMMA 4.6. *Assume that Algorithm* 3.3 *generates infinitely many successful steps* s_{j_i} *and that there exists* $\mathcal{S}' \subset \mathcal{S}$ *with*

$$\sum_{k \in \mathcal{S}'} \Delta_k = \infty. \tag{4.2}$$

Then $\liminf_{\mathcal{S}' \ni k \to \infty} \chi(x_k) = 0$.

Proof. Let the assumptions of the lemma hold and assume that the assertion is wrong. Then there exists $\varepsilon > 0$ such that $\chi(x_k) \geq \varepsilon$ for all $k \in \mathcal{S}' \subset \mathcal{S}$. From (4.2) follows that $\mathcal{S}'$ is not finite. For all $k \in \mathcal{S}'$ holds by (3.11)

$$\text{pred}_k(s_k) \geq \beta_2 \chi(x_k) \min\{1, \Delta_k, \chi(x_k)\} \geq \beta_2 \varepsilon \min\{1, \Delta_k, \varepsilon\}.$$

From this estimate, the nonnegativity of h, and Lemma 4.3 we obtain for all $j \in \mathcal{S}'$, using $\lambda \leq 1$

$$\begin{aligned} h(x_0) \geq h(x_0) - h(x_j) &> \eta_1 \lambda \sum_{\substack{k \in \mathcal{S} \\ k<j}} \text{pred}_k(s_k) \geq \eta_1 \lambda \sum_{\substack{k \in \mathcal{S}' \\ k<j}} \text{pred}_k(s_k) \\ &\geq \eta_1 \lambda \beta_2 \varepsilon \sum_{\substack{k \in \mathcal{S}' \\ k<j}} \min\{1, \Delta_k, \varepsilon\} \to \infty \qquad (\text{as } j \to \infty). \end{aligned}$$

This is a contradiction. Therefore, the assumption was wrong and the lemma is proved. □

We now have everything at hand that we need to establish our first global convergence result. It is applicable in the case $\gamma_0 > 0$, $\Delta_{\min} > 0$ and says that accumulation points are KKT-points of (1.2).

THEOREM 4.7. *Let* $\gamma_0 > 0$ *and* $\Delta_{\min} > 0$. *Assume that Algorithm* 3.3 *does not terminate after finitely many steps with a KKT-point* x_k *of* (1.2). *Then the algorithm generates infinitely many accepted steps* (s_{j_i}). *Moreover, every accumulation point of* (x_k) *is a KKT-point of* (1.2).

Proof. Suppose that Algorithm 3.3 does not terminate after a finite number of steps. Then according to Lemma 4.5 infinitely many successful steps (s_{j_i}) are generated. Assume that $\bar{x}$ is an accumulation point of (x_k) that is not a KKT-point of (1.2). Since $\chi(\bar{x}) \neq 0$, invoking Lemma 4.4 with $x = \bar{x}$ yields $\Delta > 0$ and $\delta > 0$ such that $k \in \mathcal{S}$ holds for all k with $\|x_k - \bar{x}\| \leq \delta$ and $\Delta_k \leq \Delta$. Since $\bar{x}$ is an accumulation point, there exists an infinite increasing sequence $j_i' \in \mathcal{S}$, $i \geq 0$, of indices such $\|x_{j_i'} - \bar{x}\| \leq \delta$ and $x_{j_i'} \to \bar{x}$.

If $(j_i' - 1) \in \mathcal{S}$, then $\Delta_{j_i'} \geq \Delta_{\min}$. Otherwise, $s_{j_i'-1}$ was rejected, which, since then $x_{j_i'-1} = x_{j_i'}$, is only possible if $\Delta_{j_i'-1} > \Delta$, and therefore $\Delta_{j_i'} \geq \gamma_0 \Delta_{j_i'-1} > \gamma_0 \Delta$. We conclude that for all i holds $\Delta_{j_i'} \geq \min\{\Delta_{\min}, \gamma_0 \Delta\}$. Now Lemma 4.6 is applicable with $\mathcal{S}' = \{j_i' \,;\ i \geq 0\}$ and yields

$$0 \neq \chi(\bar{x}) = \lim_{i \to \infty} \chi(x_{j_i'}) = \liminf_{i \to \infty} \chi(x_{j_i'}) = 0,$$

where we have used the continuity of χ. This is a contradiction. Therefore, the assumption $\chi(\bar{x}) \neq 0$ was wrong. □

Next, we prove a result that holds also for $\Delta_{\min} = 0$. Moreover, the existence of accumulation points is not required.

THEOREM 4.8. *Let* $\gamma_0 > 0$ *or* $\Delta_{\min} = 0$ *hold. Assume that Algorithm* 3.3 *does not terminate after finitely many steps with a KKT-point* x_k *of* (1.2). *Then the algorithm generates infinitely many accepted steps* (s_{j_i}). *Moreover,*

$$\liminf_{k\to\infty} \chi(x_k) = 0. \tag{4.3}$$

In particular, if x_k *converges to* $\bar{x}$*, then* $\bar{x}$ *is a KKT-point of* (1.2).

Proof. By Lemma 4.5, infinitely many successful steps (s_{j_i}) are generated.

Now assume that (4.3) is wrong, i.e.,

$$\liminf_{k\to\infty} \chi(x_k) > 0. \tag{4.4}$$

Then we obtain from Lemma 4.6 that

$$\sum_{k\in\mathcal{S}} \Delta_k < \infty. \tag{4.5}$$

In particular, (x_{j_i}) is a Cauchy sequence by (3.10). Therefore, (x_k) converges to some limit $\bar{x}$, at which according to (4.4) and the continuity of χ holds $\chi(\bar{x}) \neq 0$.

Case 1: $\Delta_{\min} > 0$.

Then by assumption also $\gamma_0 > 0$, and Theorem 4.7 yields $\chi(\bar{x}) = 0$, which is a contradiction.

Case 2: $\Delta_{\min} = 0$.

Lemma 4.4 with $x = \bar{x}$ and $\eta = \eta_2$ yields $\Delta > 0$ and $\delta > 0$ such that $k \in \mathcal{S}$ and $\Delta_{k+1} \geq \Delta_k$ holds for all k with $\|x_k - \bar{x}\| \leq \delta$ and $\Delta_k \leq \Delta$. Since $x_k \to \bar{x}$, there exists $k' \geq 0$ with $\|x_k - \bar{x}\| \leq \delta$ for all $k \geq k'$.

Case 2.1: There exists $k'' \geq k'$ with $\Delta_k \leq \Delta$ for all $k \geq k''$.

Then $k \in \mathcal{S}$ and (inductively) $\Delta_k \geq \Delta_{k''}$ for all $k \geq k''$. This contradicts (4.5).

Case 2.2: For infinitely many k holds $\Delta_k > \Delta$.

By (4.5) there exists $k'' \geq k'$ with $\Delta_{j_i} \leq \Delta$ for all $j_i \geq k''$. Now, for each $j_i \geq k''$, there exists an index $k_i \geq j_i$ such that $\Delta_k \leq \Delta$, $j_i \leq k < k_i$, and $\Delta_{k_i} > \Delta$. If $k_i \in \mathcal{S}$, set $j_i' = k_i$, thus obtaining $j_i' \in \mathcal{S}$ with $\Delta_{j_i'} > \Delta$. If $k_i \notin \mathcal{S}$, we have $j_i' \stackrel{\text{def}}{=} k_i - 1 \geq j_i \geq k'$, and thus $j_i' \in \mathcal{S}$, since by construction $\Delta_{j_i'} \leq \Delta$. Moreover, $\Delta < \Delta_{k_i} \leq \gamma_2 \Delta_{j_i'}$ (here $\Delta_{\min} = 0$ is used) implies that $\Delta_{j_i'} > \Delta/\gamma_2$. By this construction, we obtain an infinitely increasing sequence $(j_i') \subset \mathcal{S}$ with $\Delta_{j_i'} > \Delta/\gamma_2$. Again, this yields a contradiction to (4.5).

Therefore, in all cases we obtain a contradiction. Thus, the assumption was wrong and the proof of (4.3) is complete.

Finally, if $x_k \to \bar{x}$, the continuity of χ and (4.3) imply $\chi(\bar{x}) = 0$. Therefore, $\bar{x}$ is a KKT-point of (1.2). □ The next result shows that under appropriate assumptions the lim inf in (4.3) can be replaced by lim.

THEOREM 4.9. *Let* $\gamma_0 > 0$ *or* $\Delta_{\min} = 0$ *hold. Assume that Algorithm* 3.3 *does not terminate after finitely many steps with a KKT-point* x_k *of* (1.2). *Then the algorithm generates infinitely many accepted steps* (s_{j_i}). *Moreover, if there exists a set* Ω *that contains* (x_k) *and on which* χ *is uniformly continuous and bounded, then*

$$\lim_{k\to\infty} \chi(x_k) = 0. \tag{4.6}$$

Proof. In view of Theorem 4.8 we only have to prove (4.6). Thus, let us assume that (4.6) is not true. Then there exists $\varepsilon > 0$ such that $\chi(x_k) \geq 2\varepsilon$ for infinitely many $k \in \mathcal{S}$. Since (4.3) holds, we thus can find increasing sequences $(j_i')_{i\geq 0}$ and $(k_i')_{i\geq 0}$ with $j_i' < k_i' < j_{i+1}'$ and

$$\chi(x_{j_i'}) \geq 2\varepsilon, \qquad \chi(x_k) > \varepsilon \quad \forall\, k \in \mathcal{S} \text{ with } j_i' < k < k_i', \qquad \chi(x_{k_i'}) \leq \varepsilon.$$

Setting $\mathcal{S}' = \bigcup_{i=0}^{\infty} \mathcal{S}_i'$ with $\mathcal{S}_i' = \{k \in \mathcal{S}\,;\; j_i' \leq k < k_i'\}$, we have

$$\liminf_{\mathcal{S}' \ni k \to \infty} \chi(x_k) \geq \varepsilon.$$

Therefore, with Lemma 4.6

$$\sum_{k \in \mathcal{S}'} \Delta_k < \infty.$$

In particular, $\sum_{k \in \mathcal{S}_i'} \Delta_k \to 0$ as $i \to \infty$, and thus, using (3.10),

$$\|x_{k_i'} - x_{j_i'}\|_\infty \leq \sum_{k \in \mathcal{S}_i'} \|s_k\|_\infty \leq \beta_1 \sum_{k \in \mathcal{S}_i'} \Delta_k \to 0 \quad (\text{as } i \to \infty).$$

This is a contradiction to the uniform continuity of χ, since

$$\lim_{i \to \infty} (x_{k_i'} - x_{j_i'}) = 0, \quad \text{but} \quad |\chi(x_{k_i'}) - \chi(x_{j_i'})| \geq \varepsilon \quad \forall\, i \geq 0.$$

Therefore, the assumption was wrong and the assertion is proved. □

The above results establish global convergence to a KKT-point $\bar{x}$ of the minimization reformulation (1.2). Of course, it may happen that $\bar{x}$ fails to be a global solution, i.e., fails to satisfy $H(\bar{x}) = 0$ if (1.1) possesses a solution. However, as can be seen from our numerical results in section 8, our non-monotone trust-region approach very successfully avoids convergence to local-nonglobal solutions. Moreover, in the context of semismooth reformulations of MCPs, conditions can be stated under which a local solution to (1.2) is a global solution. See [16, Thm. 2].

5. Local Convergence. In this section we will prove under relatively weak assumptions that Algorithm 3.6 turns into Algorithm 3.1 as soon as x_k comes sufficiently close to a BD-regular solution $\bar{x}$ of (1.1). Of course, we have to take care that close to $\bar{x}$ the projected Newton step is used as trial step which requires (3.11) to hold. Note that (3.10) is automatically satisfied since $x_k + s_k^{\mathrm{P}} \in X$. We will see that it suffices to require that the following implication holds:

$$\begin{aligned} &\|x_k - \bar{x}\| < \beta_3, \quad \operatorname{pred}_k(s_k^{\mathrm{P}}) \geq \beta_4 h(x_k) \\ &\Longrightarrow \quad (3.11) \text{ is satisfied for } s_k = s_k^{\mathrm{P}}, \end{aligned} \tag{5.1}$$

where $\beta_3 > 0$ and $0 < \beta_4 < 1$ are constants independent of k. Since $h(\bar{x}) = 0$, this means that for x_k close to $\bar{x}$ the step $s_k = s_k^{\mathrm{P}}$ satisfies (3.10) if the predicted decrease is at least a fraction of the maximum possible actual decrease $h(x_k) - h(\bar{x}) = h(x_k)$. Therefore, (5.1) is certainly a reasonable requirement, which in section 6 will be shown to hold for decrease conditions that are implemented by means of a Cauchy step.

THEOREM 5.1. *Let the assumptions* (A1) *and* (A2) *hold. Assume* $\Delta_{\min} > 0$ *and that Algorithm* 3.6 *generates infinitely many iterates. Let* $\bar{x} \in X$ *be a BD-regular zero of* H *and let* (5.1) *be satisfied. Then there exist* $\delta > 0$ *and* $\varepsilon > 0$ *such that the following holds:*

If the index k' *satisfies* $(k'-1) \in \mathcal{S}$, $\|x_{k'} - \bar{x}\| \le \varepsilon$, *and if* (3.1) *holds for all* $k \ge k'$, *then*

(a) *For all* $k \ge k'$ *the step* s_k^{P} *satisfies* $s_k^{\mathrm{P}} = s_k^{\mathrm{PN}} = P_X(x_k + s^{\mathrm{N}}) - x_k$, *is chosen as trial step, i.e.,* $s_k = s_k^{\mathrm{P}}$, *and is accepted, i.e.,* $k \in \mathcal{S}$.

(b) *The sequence* (x_k) *converges to* $\bar{x}$.

(c) *If* (3.2) *holds for all* $k \ge k'$ *then* (x_k) *converges q-superlinearly to* $\bar{x}$. *Moreover,*

$$\lim_{k\to\infty} \frac{q_k(s_k^{\mathrm{P}})}{h(x_k)} = \lim_{k\to\infty} \frac{q_k(s_k^{\mathrm{N}})}{h(x_k)} = -1. \tag{5.2}$$

(d) *If* $p > 0$ *and* (3.3) *holds for all* $k \ge k'$ *then* (x_k) *converges with q-order* $1+p$ *to* $\bar{x}$ *and* (5.2) *holds.*

Proof. Assume that

$$\delta > 0 \text{ and } \varepsilon > 0 \text{ are chosen sufficiently small} \tag{5.3}$$

with $\varepsilon \le \beta_4$, the constant in (5.1). Then Proposition 2.12 yields a constant $\zeta > 0$ such that

$$\|H(x)\| \ge \zeta \|x\| \qquad \forall\ B_\varepsilon(\bar{x}) \overset{\text{def}}{=} \{x\,;\ \|x - \bar{x}\| \le \varepsilon\}\,. \tag{5.4}$$

Now let k be arbitrary such that $(k-1) \in \mathcal{S}$, $x_k \in B_\varepsilon(\bar{x})$, and such that (3.1) is satisfied.

Let $d_k \overset{\text{def}}{=} x_k - \bar{x}$, $e_k \overset{\text{def}}{=} x_k + s_k^{\mathrm{N}} - \bar{x}$ and $V_k \in \partial_B H(x_k)$ such that $\mu_k = \|(M_k - V_k)s_k^{\mathrm{N}}\|$. We begin by copying the proof of Theorem 3.2 to derive the equations (3.4)–(3.7). From (3.6) and the fact that $(k-1) \in \mathcal{S}$ and $\Delta_{\min} > 0$ we conclude, using (5.3), that

$$\|s_k^{\mathrm{N}}\| \le 2C\|H(x_k)\| = 2C\|H(x_k) - H(\bar{x})\| \le 2CL\varepsilon \le \Delta_{\min} \le \Delta_k. \tag{5.5}$$

In particular,

$$s_k^{\mathrm{P}} = P_{X_k}(s_k^{\mathrm{N}}) = P_{X-x_k}(s_k^{\mathrm{N}}) = P_X(x_k + s_k^{\mathrm{N}}) - x_k = s_k^{\mathrm{PN}}. \tag{5.6}$$

Since $x_k \in X$ and P_X is Lipschitz continuous of rank 1 it follows that

$$\|s_k^{\mathrm{P}}\| = \|P_X(x_k + s_k^{\mathrm{N}}) - P_X(x_k)\| \le \|s_k^{\mathrm{N}}\|. \tag{5.7}$$

Now, using $g_k = V_k^T H(x_k)$, (3.1), (3.6), and (5.5)

$$\begin{aligned} |q_k(s_k^{\mathrm{N}}) + h(x_k)| &= \left| g_k^T s_k^{\mathrm{N}} + \frac{1}{2}\|M_k s_k^{\mathrm{N}}\|^2 + h(x_k) \right| = \left| H(x_k)^T V_k s_k^{\mathrm{N}} + 2h(x_k) \right| \\ &\le \left| H(x_k)^T M_k s_k^{\mathrm{N}} + 2h(x_k) \right| + \|H(x_k)\|\mu_k = \|H(x_k)\|\mu_k \\ &\le \delta \|H(x_k)\| \|s_k^{\mathrm{N}}\| \le 4C\delta h(x_k). \end{aligned} \tag{5.8}$$

This shows

$$-(1 + 4C\delta)h(x_k) \le q_k(s_k^{\mathrm{N}}) \le -(1 - 4C\delta)h(x_k). \tag{5.9}$$

Furthermore, since $\bar{x} \in X$, the properties of the projection P_X yield

$$\|s_k^{\mathrm{P}} - s_k^{\mathrm{N}}\| = \|P_X(x_k + s_k^{\mathrm{N}}) - (x_k + s_k^{\mathrm{N}})\| \le \|\bar{x} - (x_k + s_k^{\mathrm{N}})\| = \|e_k\|. \tag{5.10}$$

Thus by (5.7), (5.10), and (3.6)

$$
\begin{aligned}
|q_k(s_k^{\mathrm{P}}) - q_k(s_k^{\mathrm{N}})| &= \frac{1}{2}\left|g_k^T(s_k^{\mathrm{P}} - s_k^{\mathrm{N}}) + (s_k^{\mathrm{P}} + s_k^{\mathrm{N}})^T M_k^T M_k(s_k^{\mathrm{P}} - s_k^{\mathrm{N}})\right| \\
&\le \left(\|V_k\|\|H(x_k)\| + \|M_k\|^2\|s_k^{\mathrm{N}}\|\right)\|s_k^{\mathrm{P}} - s_k^{\mathrm{N}}\| \\
&\le \left(\|V_k\| + 2C\|M_k\|^2\right)\|H(x_k)\|\|e_k\|.
\end{aligned} \tag{5.11}
$$

From (3.7) we conclude as in the proof of Theorem 3.2 that

$$
\frac{\|e_k\|}{\|H(x_k)\|} \le \frac{1}{\zeta}\frac{\|e_k\|}{\|d_k\|} \to 0 \qquad (\text{as } (\delta, \varepsilon) \to 0). \tag{5.12}
$$

Now let $\theta \in (0,1)$ be arbitrary. Due to the upper semicontinuity of $\partial_B H$, cf. Proposition 2.2, the boundedness of $(\|M_k\|)$, and eq:ekdk, we achieve by invoking (5.3) that

$$
2\left(\|V_k\| + 2C\|M_k\|^2\right)\frac{\|e_k\|}{\|H(x_k)\|} + 4C\delta \le \theta.
$$

This in combination with (5.9) and (5.11) shows that

$$
-(1+\theta)h(x_k) \le q_k(s_k^{\mathrm{P}}) \le -(1-\theta)h(x_k). \tag{5.13}
$$

If θ was chosen $\le 1 - \beta_4$ then

$$
\mathrm{pred_k}(s_k^{\mathrm{P}}) = -q_k(s_k^{\mathrm{P}}) \ge (1-\theta)h(x_k) \ge \beta_4 h(x_k).
$$

Therefore, (5.1) implies that $s_k = s_k^{\mathrm{P}} = s_k^{\mathrm{PN}}$. As in (3.8) we obtain

$$
\|x_k + s_k^{\mathrm{P}} - \bar{x}\| \le \|e_k\|, \tag{5.14}
$$

and thus

$$
\|H(x_k + s_k)\| = \|H(x_k + s_k^{\mathrm{P}})\| \le L\|x_k + s_k^{\mathrm{P}} - \bar{x}\| \le L\|e_k\|, \tag{5.15}
$$

where L denotes the Lipschitz constant of H on $B_\varepsilon(\bar{x})$. By (5.3), (5.12) and (5.15) we thus may assume that

$$
\mathrm{ared}_k(s_k) \ge h(x_k) - h(x_k + s_k) \ge h(x_k) - \frac{L^2}{2}\|e_k\|^2 \ge (1-\theta)h(x_k). \tag{5.16}
$$

If θ was chosen $< (1-\eta_1)/(1+\eta_1)$, then by (5.13) and (5.16)

$$
\rho_k(s_k) = \frac{\mathrm{ared}_k(s_k)}{\mathrm{pred}_k(s_k)} \ge \frac{(1-\theta)h(x_k)}{(1+\theta)h(x_k)} > \eta_1.
$$

Consequently, $s_k = s_k^{\mathrm{P}}$ is accepted. Moreover, by (5.12), (5.14), and again invoking (5.3) we get $x_{k+1} \in B_\varepsilon(\bar{x})$.

We briefly resume what we have shown so far:

If $\delta > 0$ and $\varepsilon > 0$ are sufficiently small and k is such that $(k-1) \in \mathcal{S}$, $x_k \in B_\varepsilon(\bar{x})$, and (3.1) holds, then $s_k = s_k^{\mathrm{P}} = s_k^{\mathrm{PN}}$ is chosen as trial step and this step is accepted, i.e. $k \in \mathcal{S}$. Moreover, we have $x_{k+1} \in B_\varepsilon(\bar{x})$.

Consequently, $(k+1)$ satisfies again the requirements and we obtain inductively that assertion (a) holds.

Since by (a) we know that for $k \ge k'$ Algorithm 3.6 turns into Algorithm 3.1, the assertions (b)–(d), with the exception of (5.2), follow directly from Theorem 3.2. Finally, to prove (5.2) we observe that, if (3.2) holds, we can strengthen (5.8):

$$|q_k(s_k^{\mathrm{N}}) + h(x_k)| \le \|H(x_k)\|\mu_k = o(h(x_k)), \tag{5.17}$$

where we have used and (3.2) and (5.5). From (5.11) and (5.12) we conclude

$$|q_k(s_k^{\mathrm{P}}) - q_k(s_k^{\mathrm{N}})| = O(\|H(x_k)\|\|e_k\|) = o(h(x_k)).$$

This and (5.17) imply (5.2). □

6. An Implementable Decrease Condition. Our convergence analysis was carried out on the basis of the abstract condition (3.11) involving a criticality measure χ. For fast local convergence we also required (5.1). In this section we describe a concrete implementation of these condition by means of a Cauchy step.

We define the *Cauchy step* s_k^{C} as the solution of

$$\text{minimize } q_k(s) \quad \text{subject to} \quad s = -tD(x_k)^{2\gamma} g_k,\ t \ge 0,\ s \in X_k.$$

Hereby, $\gamma \ge 1$ is fixed and the diagonal affine-scaling matrix $D(x) \in \mathbb{R}^{n\times n}$ is defined by

$$D(x)_{ii} \overset{\text{def}}{=} \begin{cases} \min\{\kappa_D, x_i - l_i\} & \text{if } (\nabla h(x))_i > 0, \\ \min\{\kappa_D, u_i - x_i\} & \text{if } (\nabla h(x))_i < 0, \\ \min\{\kappa_D, x_i - l_i, u_i - x_i\} & \text{if } (\nabla h(x))_i = 0. \end{cases}$$

$\kappa_D > 0$ is a constant.

The following condition will replace the abstract condition (3.11).

Fraction of Cauchy decrease condition:

$$q_k(s_k) \le \alpha q_k(s_k^{\mathrm{C}}), \tag{6.1}$$

where $\alpha \in (0,1)$ is a constant.

Let χ denote any criticality measure such that

$$\beta_5 \chi(x) \le \chi_{\mathrm{AS}}(x) \overset{\text{def}}{=} \|D(x)^\gamma \nabla h(x)\| \tag{6.2}$$

holds on X for some $\beta_5 > 0$. We will show that there exists $\beta_2 > 0$ such that the validity of (6.1) implies (3.11). Certainly, a natural choice for χ satisfying (6.2) is $\chi = \chi_{\mathrm{AS}}/\beta_5$. Therefore, we first show that χ_{AS} is a criticality measure.

LEMMA 6.1. *The function χ_{AS} defined in* (6.2) *is a criticality measure, i.e., satisfies* (3.12).

Proof. It is easily seen that for $x \in X$ holds $D(x)^\gamma \nabla h(x)$ if and only if

$$(\nabla h(x))_i \begin{cases} \ge 0 & \text{if } x_i = l_i, \\ \le 0 & \text{if } x_i = u_i, \\ = 0 & \text{if } l_i < x_i < u_i, \end{cases}$$

which are the KKT-conditions of (1.2).

We still have to prove the continuity of χ_{AS}. Let $x \in X$ be arbitrary and set $I_+ = \{i\,;\ (\nabla h(x))_i > 0\}$, $I_- = \{i\,;\ (\nabla h(x))_i < 0\}$. Since ∇h is continuous, there exists $\delta > 0$ such that for all $y \in X$, $\|y - x\| \le \delta$, and all $i \in I_+ \cup I_-$ holds $(\nabla h(x))_i (\nabla h(y))_i > 0$.

Now let $y \in X$, $\|y - x\| \le \delta$, be arbitrary and set $r(x,y) = D(y)^\gamma \nabla h(y) - D(x)^\gamma \nabla h(x)$. For all $i \in I_+$ holds

$$\begin{aligned} |r(x,y)_i| &\le (\nabla h(x))_i |(D(y)^\gamma - D(x)^\gamma)_{ii}| + D(y)^\gamma_{ii} |(\nabla h(y) - \nabla h(x))_i| \\ &\le (\nabla h(x))_i |\min\{\kappa_D, y_i - l_i\}^\gamma - \min\{\kappa_D, x_i - l_i\}^\gamma| + \kappa_D^\gamma |(\nabla h(y) - \nabla h(x))_i|. \end{aligned}$$

The same calculation yields for $i \in I_-$

$$|r(x,y)_i| \le |(\nabla h(x))_i| |\min\{\kappa_D, u_i - y_i\}^\gamma - \min\{\kappa_D, u_i - x_i\}^\gamma| + \kappa_D^\gamma |(\nabla h(y) - \nabla h(x))_i|.$$

For all $i \notin I_+ \cup I_-$ holds $(\nabla h(x))_i = 0$ and thus

$$|r(x,y)_i| = D(y)^\gamma_{ii} |(\nabla h(y) - \nabla h(x))_i| \le \kappa_D^\gamma |(\nabla h(y) - \nabla h(x))_i|.$$

In all three cases we see that $|r(x,y)_i| \to 0$ as $y \to x$. Hence, χ_{AS} is continuous. ☐

The next Lemma gives sufficient conditions for χ_{AS} to be uniformly continuous. This is needed for Theorem 4.9.

LEMMA 6.2. *The criticality measure χ_{AS} defined in* (6.2) *is uniformly continuous on $\Omega \subset U$, if one of the following conditions holds true:*

(a) *∇h is bounded and uniformly continuous on Ω.*

(b) *H is bounded and uniformly continuous on Ω. There exists a constant $C > 0$ such that for all $\varepsilon > 0$ there exists $\delta > 0$ with*

$$\inf\left\{\|V_x - V_y\|\,;\ V_x \in \partial H(x),\ V_y \in \partial H(y),\ \|V_y\| \le C\right\} < \varepsilon \tag{6.3}$$

for all $x, y \in \Omega$ with $\|x - y\| < \delta$.

Proof. Under the assumptions (a) it follows easily from the estimates of $|r(x,y)_i|$ in the proof of Lemma 6.1 that χ_{AS} is uniformly continuous on Ω.

Now let (b) hold. Then $\|H(x)\| \le C_1$ for all $x \in \Omega$ with a suitable constant $C_1 > 0$. By assumption, for given $\varepsilon > 0$ there exists $\delta > 0$ such that (6.3) holds. By possibly reducing δ we achieve $\|H(x) - H(y)\| < \varepsilon$ for all $x, y \in \Omega$ with $\|x - y\| < \delta$. Condition (6.3) implies that the set over which the inf is taken is nonempty. In particular, there exists a selection $x \in \Omega \mapsto V(x) \in \partial H(x)$, with $\|V(x)\| \le C$ on Ω. Hence, for all $x \in \Omega$

$$\|\nabla h(x)\| = \|V(x)^T H(x)\| \le \|V(x)\| \|H(x)\| \le CC_1.$$

This proves the boundedness of ∇h on Ω. Now let $x, y \in \Omega$, $\|x - y\| < \delta$, be arbitrary. Since the set in (6.3) is nonempty and compact, there are $V_x \in \partial H(x)$ and $V_y \in \partial H(y)$, $\|V_y\| \le C$, for which the inf in (6.3) is attained. Now

$$\begin{aligned} \|\nabla h(x) - \nabla h(y)\| &\le \|V_x - V_y\| \|H(x)\| + \|V_y\| \|H(x) - H(y)\| \\ &\le C_1 \|V_x - V_y\| + C\|H(x) - H(y)\| < C_1\varepsilon + C\varepsilon. \end{aligned} \tag{6.4}$$

Thus, ∇h is uniformly continuous on Ω. Therefore, (a) applies, which proves that χ_{AS} is uniformly continuous. ☐

We have the following relation between (3.11) and (6.1).

LEMMA 6.3. *Let the criticality measure χ satisfy* (6.2). *Assume that the sequence $(\|M_k\|)$ is bounded above by a constant C_M. Then there exists a constant $\beta_2 > 0$ that only depends on α, β_5, γ, κ_D, and C_M such that the following holds: If $\chi(x_k) \neq 0$ and if the trial step s_k satisfies the fraction of Cauchy decrease condition* (6.1), *then also* (3.11) *holds.*

Proof. Set $D_k = D(x_k)$, $d_k = -D_k^{2\gamma} g_k$, and $\hat{g}_k = D_k^{\gamma} g_k$. We will derive an upper bound for $q_k(s_k^{\mathrm{C}}) = q_k(t^* d_k) = \min\{q_k(td_k)\,;\ t \geq 0,\ td_k \in X_k\}$, and then apply $q_k(s_k) \leq \alpha q_k(t^* d_k)$.

First, observe that d_k is a descent direction of q_k at 0, since by (6.2)

$$\nabla q_k(0)^T d_k = {g_k}^T d_k = -\chi_{\mathrm{AS}}(x_k)^2 \leq -\beta_5^2 \chi(x_k)^2 < 0.$$

The maximum stepsize allowed by the trust region constraint is

$$t_1 = \min\left\{ \frac{\Delta_k}{|(d_k)_i|}\,;\ (d_k)_i \neq 0 \right\} = \frac{\Delta_k}{(D_k)_{ii}^{\gamma} |(\hat{g}_k)_i|} \geq \frac{\Delta_k}{\kappa_D^{\gamma} \chi_{\mathrm{AS}}(x_k)}. \tag{6.5}$$

The maximum stepsize t_2 admitted by the lower bounds of the set $X - x_k$ is

$$\begin{aligned} t_2 &= \min\left\{ \frac{(x_k - l)_i}{|(d_k)_i|}\,;\ (d_k)_i < 0 \right\} = \min\left\{ \frac{(x_k - l)_i}{(D_k)_{ii}^{\gamma} (\hat{g}_k)_i}\,;\ (g_k)_i > 0,\ (x_k)_i > l_i \right\} \\ &\geq \min\left\{ \frac{(x_k - l)_i}{\min\{\kappa_D, (x_k - l)_i\}^{\gamma} \chi_{\mathrm{AS}}(x_k)}\,;\ (g_k)_i > 0,\ (x_k)_i > l_i \right\} \geq \frac{\kappa_D^{1-\gamma}}{\chi_{\mathrm{AS}}(x_k)}. \end{aligned} \tag{6.6}$$

In the same way, the stepsize t_3 admitted by the upper bounds of the set $X - x_k$ can be estimated:

$$t_3 = \min\left\{ \frac{(u - x_k)_i}{(d_k)_i}\,;\ (d_k)_i > 0 \right\} \geq \frac{\kappa_D^{1-\gamma}}{\chi_{\mathrm{AS}}(x_k)}.$$

In the case $M_k d_k = 0$ we set $t_4 = +\infty$. Otherwise, the function $q_k(td_k)$, $t \geq 0$, attains its global minimum at $t = t_4$, where

$$t_4 = \frac{-g_k^T d_k}{\|M_k d_k\|^2} = \frac{\|\hat{g}_k\|^2}{\|M_k d_k\|^2} \geq \frac{\|\hat{g}_k\|^2}{\|M_k\|^2 \|D_k^{\gamma}\|^2 \|\hat{g}_k\|^2} = \frac{1}{\|M_k\|^2 \|D_k^{\gamma}\|^2} \geq \frac{1}{C_M^2 \kappa_D^{2\gamma}}. \tag{6.7}$$

We have $t^* = \min\{t_1, t_2, t_3, t_4\}$. If $t^* < t_4$ then $\|\hat{g}_k\|^2 > t^* \|M_k d_k\|^2$ and

$$\begin{aligned} q_k(t^* d_k) &= -t^* \|\hat{g}_k\|^2 + \frac{1}{2}(t^*)^2 \|M_k d_k\|^2 < -\frac{t^*}{2} \chi_{\mathrm{AS}}(x_k)^2 \\ &= -\frac{\min\{t_1, t_2, t_3\}}{2} \chi_{\mathrm{AS}}(x_k)^2. \end{aligned} \tag{6.8}$$

If, on the other hand, $t^* = t_4$, then

$$q_k(t^* d_k) = -\frac{t_4}{2} \chi_{\mathrm{AS}}(x_k)^2. \tag{6.9}$$

The proof is completed by combining (6.2) and the estimates (6.5)–(6.9). ☐

Therefore, for any criticality measure satisfying (6.2), we can replace the decrease condition (3.11) by (6.1). Moreover, the Cauchy step s_k^{C}, which is easy to compute as could be seen in the proof of Lemma 6.3, is always an admissible trial step. Also, the optimal solution of the trust-region subproblem satisfies (6.1). Therefore, the global convergence theory of section 4 is applicable. Conditions that imply the uniform continuity of $\chi = \chi_{\mathrm{AS}}$ needed in Theorem 4.9 were established in Lemma 6.1.

On the basis of χ_{AS} and the fraction of Cauchy decrease condition (6.1) it is possible to state implementations for Step 3.3 of Algorithm 3.6 such that (a) the reduction condition (3.11) is easy to check, (b) condition (5.1) is satisfied.

There are several ways to do this.

(I) We begin with a universally applicable approach. Certainly, the function

$$\chi_{\mathrm{I}}(x) = \min\{\chi_{\mathrm{AS}}(x), (\beta_4 h(x))^{1/2}\} \tag{6.10}$$

with $\beta_4 \in (0,1)$ is a criticality measure. Now let χ be any criticality measure verifying

$$\beta_5 \chi(x) \le \chi_{\mathrm{I}}(x), \quad x \in X, \tag{6.11}$$

for some $\beta_5 > 0$. Then (6.2) holds. Let Step 3.3 of Algorithm 3.6 be implemented as follows:

3.3. If $s_k = s_k^{\mathrm{P}}$ satisfies

$$q_k(s_k) \le \max\{\alpha q_k(s_k^{\mathrm{C}}), -\beta_4 h(x_k)\} \tag{6.12}$$

Otherwise, compute a step s_k satisfying (3.10) and (6.12).

Note that condition (6.12) is a relaxation of (6.1). Obviously, (5.1) holds for arbitrary $\beta_3 > 0$. Furthermore, under the assumptions of Lemma 6.3 there is $\beta_2 > 0$ such that all computed steps s_k satisfy (3.11). In fact, if $q_k(s_k) \le -\beta_4 h(x_k)$, (3.11) immediately holds. On the other hand, if $q_k(s_k) \le -\alpha q_k(s_k^{\mathrm{C}})$, then Lemma 6.3 is applicable. Finally, it is easily seen that the particular choice $\chi = \chi_{\mathrm{I}}$ is uniformly continuous under condition (b) in Lemma 6.2.

(II) We now discuss a situation in which it suffices for χ to obey (6.2). Assume that $\min_{V\in\partial_B H(x)} \|V\| \le C_V < \infty$ on X. Then holds with $V(x) = \operatorname{argmin}_{V\in\partial_B H(x)} \|V\|$

$$\chi_{\mathrm{AS}}(x) = \|D(x)^\gamma \nabla h(x)\| \le \kappa_D^\gamma \|V(x)\| \|H(x)\| \le \kappa_D^\gamma C_V \|H(x)\|.$$

Therefore, if $\beta_4 \in (0,1)$ is fixed and the criticality measure χ satisfies (6.2) with β_5 replaced by β_5', then (6.11) holds for $0 < \beta_5 \le \beta_5' \min\{1, (\beta_4/2)^{1/2}/(\kappa_D^\gamma C_V)\}$. Therefore, (I) applies. We stress that in this scenario we can choose $\chi = \chi_{\mathrm{AS}}$.

(III) Finally, we state conditions under which we can implement Step 3.3 of Algorithm 3.6 by means of the fraction of Cauchy decrease condition:

3.3. If $s_k = s_k^{\mathrm{P}}$ satisfies (6.1) then set $s_k := s_k^{\mathrm{P}}$. Otherwise, compute a step s_k satisfying (3.10) and (6.1).

Let the criticality measure χ satisfy (6.2). Assume that the sequences $(\|M_k\|)$ and $(\|M_k^{-1}\|)$ are bounded, and that

$$\|M_k^T H(x_k) - g_k\| = o(\|H(x_k)\|) \quad (\text{as } k \to \infty).$$

Note that this holds true if, e.g., $M_k \in \partial H(x_k)$. By Lemma 6.3 there is $\beta_2 > 0$ such that (3.11) holds for all computed steps. Under the above assumptions we obtain

$$\begin{aligned} q_k(s_k^{\mathrm{C}}) &\ge q_k(-(M_k^T M_k)^{-1} g_k) = -\frac{1}{2} g_k^T (M_k^T M_k)^{-1} g_k \\ &= -h(x_k) - \frac{1}{2}(M_k^T H(x_k) - g_k)^T (M_k^T M_k)^{-1} (M_k^T H(x_k) + g_k) \\ &= -h(x_k) + \frac{1}{2}(M_k^T H(x_k) - g_k)^T (M_k^T M_k)^{-1} \left((M_k^T H(x_k) - g_k) - 2M_k^T H(x_k)\right) \\ &= -h(x_k) + o(\|H(x_k)\|^2) = -h(x_k) + o(h(x_k)). \end{aligned}$$

Therefore, for arbitrary $\beta_4 \in (\alpha, 1)$ we can find $\beta_3 > 0$ such that for all k satisfying the left hand side of (5.1) holds

$$q_k(s_k^{\mathrm{P}}) = -\mathrm{pred}_k(s_k^{\mathrm{P}}) \le -\beta_4 h(x_k) \le \alpha q_k(s_k^{\mathrm{C}}),$$

which implies that condition (5.1) is satisfied.

In all three scenarios the implementation of Step 3.3 yields a special case of Algorithm 3.6. Moreover, all global and local convergence results are applicable.

7. Application to the Mixed Complementarity Problem. In the introduction we defined the mixed complementarity problem MCP(F, X) and described how it can be equivalently reformulated in the form (1.1). Hereby, we used MCP-functions $\psi_i : \mathbb{R}^2 \to \mathbb{R}$, i.e., functions with the property

$$\psi_i(a,b) = 0 \quad \text{if and only if} \quad l_i \le a \le u_i \quad \text{and} \quad b \begin{cases} = 0 & \text{if } l_i < a < u_i, \\ \ge 0 & \text{if } a = l_i, \\ \le 0 & \text{if } = u_i \end{cases} \tag{7.1}$$

to define the function

$$H : U \to \mathbb{R}^n, \quad H_i(x) \stackrel{\text{def}}{=} \psi_i(x_i, F_i(x)). \tag{7.2}$$

7.1. Properties of the Semismooth Reformulation. We will assume that

(A3) The function $F : \mathbb{R}^n \supset U \to \mathbb{R}^n$ is Lipschitz continuously differentiable.

(A4) The functions ψ_i satisfy (7.1) and are continuously differentiable on $\mathbb{R}^2 \setminus \psi_i^{-1}(0)$. Moreover, $\psi_i \in S^1(\mathbb{R}^2, \mathbb{R})$. For all $a, b \in \mathbb{R}$ with $\psi_i(a,b) = 0$ and all $v \in \partial\psi_i(a,b)$ holds:

$$\begin{aligned} v \ge 0, \ & v \ne 0, \\ v_1 = 0 \ & \text{ if } l_i < a < u_i \text{ and } b = 0, \\ v_2 = 0 \ & \text{ if } a = l_i \text{ and } b > 0 \text{ or } a = u_i \text{ and } b < 0. \end{aligned}$$

LEMMA 7.1. *Let the assumptions* (A3) *and* (A4) *hold. Then the function H defined in* (7.2) *satisfies* (A1) *with $p = 1$ and* (A2). *Moreover, for all $x \in U$ and all $V \in \partial H(x)$ holds with appropriate positive semidefinite diagonal matrices $D_a, D_b \in \mathbb{R}^{n\times n}$:*

$$V = D_a + D_b F'(x).$$

If x solves MCP(F, X) then holds

$$\begin{aligned} D_a + D_b & \quad \textit{is positive definite}, \\ (D_a)_{ii} = 0 & \quad \textit{if } l_i < x_i < u_i \textit{ and } F_i(x) = 0, \\ (D_b)_{ii} = 0 & \quad \textit{if } x_i = l_i \textit{ and } F_i(x) > 0 \textit{ or } x_i = u_i \textit{ and } F_i(x) < 0. \end{aligned}$$

Proof. The functions ψ_i and F are 1-order semismooth by (A3) and (A4). Therefore, H is a composite of 1-order semismooth functions and thus 1-order semismooth itself. Thus, (A1) holds with $p = 1$.

Let $1 \le i \le n$ be fixed and $x \in U$ be arbitrary such that $H_i(x) \ne 0$. Then of course $\psi_i(x_i, F_i(x)) \ne 0$ and thus, by (A4), ψ_i is continuously differentiable in a neighborhood of $(x_i, F_i(x))$. Therefore, using (A4), H_i is C^1 in a neighborhood of x. This implies (A2).

Now let $x \in U$ be arbitrary. By [6, Prop. 2.6.2.(e)] holds

$$\partial H(x) \subset \partial H_1(x) \times \cdots \times \partial H_n(x).$$

Furthermore, setting $f_i(x) = (x_i, F_i(x))^T$, we have by [6, Thm. 2.6.6]

$$\begin{aligned}\partial H_i(x) &= \operatorname{conv}(\partial\psi_i(f_i(x))\partial f_i(x)) = \partial\psi_i(f_i(x))\nabla f_i(x)^T \\ &= \left\{v_1 e_i^T + v_2 \nabla F_i(x)^T\,;\ v \in \partial\psi_i(x_i, F_i(x))\right\}.\end{aligned}$$

If x is a solution of $\mathrm{MCP}(F,X)$, then $\psi_i(x_i, F_i(x)) = 0$ for all i. Now the remaining assertions follow from (A4). □

We now introduce the notion of strong regularity. It plays an important role in the stability analysis of solutions to the MCP. For a comprehensive discussion of strong regularity see [34, 42].

DEFINITION 7.2. [42] A solution $\bar{x}$ of $\mathrm{MCP}(F,X)$ is called *strongly regular* if there exist neighborhoods U_x of $\bar{x}$ and $U_y \subset \mathbb{R}^n$ of 0 such that for all $y \in U_y$ the perturbed linearized problem $\mathrm{MCP}(F^y, X)$, $F^y(x) = F(\bar{x}) + F'(\bar{x})(x - \bar{x}) + y$, admits exactly one solution $x(y)$ satisfying $x(y) \in U_x$, and, moreover, the function $x(y)$ is Lipschitz continuous on U_y. □

The following lemma gives an equivalent algebraic definition of strong regularity.

LEMMA 7.3. [14, Thm. 3.4] *The solution $\bar{x}$ of* $\mathrm{MCP}(F,X)$ *is strongly regular if and only if the submatrix* $F'(\bar{x})_{\bar{I}\bar{I}} = \left(\frac{\partial F_i}{\partial x_j}(\bar{x})\right)_{i,j\in\bar{I}}$ *is nonsingular and the Schur-complement*

$$F'(\bar{x})_{\bar{N}\bar{N}} - F'(\bar{x})_{\bar{N}\bar{I}} F'(\bar{x})_{\bar{I}\bar{I}}^{-1} F'(\bar{x})_{\bar{I}\bar{N}}$$

is a P-matrix, where $\bar{I} \overset{\text{def}}{=} \{i\,;\ l_i < \bar{x}_i < u_i\}$ *and* $\bar{N} \overset{\text{def}}{=} \{i\,;\ \bar{x}_i \in \{l_i, u_i\},\ F_i(\bar{x}) = 0\}$.

Using the Lemmas 7.1 and 7.3, the proof of [16, Thm. 1] can be easily modified to show the following result.

THEOREM 7.4. *If $\bar{x}$ is a strongly regular solution of* $\mathrm{MCP}(F,X)$ *then all elements of $\partial H(\bar{x})$ are nonsingular. In particular, $\bar{x}$ is BD-regular for H.*

7.2. A New MCP-Function. In this section we introduce a new MCP-function that is motivated by affine-scaling methods for nonlinear optimization. Throughout this section, let $1 \le i \le n$. In the affine-scaling approach [7, 8, 24] the condition on the right hand side of (7.1) is expressed equivalently as

$$d_i(a,b)b = 0, \quad l_i \le a \le u_i, \tag{7.3}$$

where the affine-scaling function $d_i : \mathbb{R}^2 \to \mathbb{R}$ is given by

$$d_i(a,b) \overset{\text{def}}{=} \begin{cases} a - l_i & \text{if } b \ge 0 \text{ and } l_i > -\infty, \\ u_i - a & \text{if } b < 0 \text{ and } u_i < +\infty, \\ 1 & \text{else.} \end{cases} \tag{7.4}$$

The ideas behind our construction are more conveniently explained by first discussing the special case $l_i = 0$, $u_i = +\infty$, in which ψ_i becomes an NCP-function. Using the same ideas, we then derive our MCP-function.

In the case $l_i = 0$ and $u_i = +\infty$, the right hand side of (7.1) can be written in the form

$$a, b \ge 0, \quad ab = 0, \tag{7.5}$$

and we have $d_i(a,b) = d(a,b) = a$, if $b \geq 0$, and $d_i(a,b) = d(a,b) = 1$, otherwise. Therefore, (7.3) becomes

$$d(a,b)b = ab_+ + b_- = 0, \quad a \geq 0. \tag{7.6}$$

where $t_+ \overset{\text{def}}{=} \max\{t,0\}$ and $t_- \overset{\text{def}}{=} \min\{t,0\}$. Since (7.5) is symmetric in a and b, we just as well can reverse the roles of a and b, yielding

$$d(b,a)a = a_+b + a_- = 0, \quad b \geq 0. \tag{7.7}$$

For all $a, b \in \mathbb{R}$ with $a \geq 0$ or $b \geq 0$, we now define the function

$$\varphi(a,b) \overset{\text{def}}{=} a_+b_+ + a_- + b_-$$

and note that (7.6) is the same as

$$\varphi(a,b) = 0, \quad a \geq 0,$$

and (7.7) is nothing else but

$$\varphi(a,b) = 0, \quad b \geq 0.$$

Since (7.5) holds if either of these conditions is satisfied, (7.5) is also equivalent to

$$\varphi(a,b) = 0, \quad a \geq 0 \text{ or } b \geq 0.$$

Therefore, if we extend the definition of φ to the negative orthant in such a way that $\varphi(a,b) \neq 0$ for all $a, b < 0$, φ becomes an NCP-function. Our particular choice of φ for $a, b < 0$ is

$$\varphi(a,b) \overset{\text{def}}{=} -\sqrt{a^2+b^2},$$

since this yields a C^1-transition across the negative coordinate axes. We thus have derived the following NCP-function:

$$\varphi(a,b) = a_+b_+ - \sqrt{a_-^2 + b_-^2},$$

which is obviously continuously differentiable at all $(a,b) \in \mathbb{R}^n$ with $a, b \neq 0$. Unfortunately, the resulting NCP-function does not satisfy (A4), since

$$\partial_B\varphi(0,0) \ni \lim_{a,b\to 0+} \nabla\varphi(a,b)^T = \lim_{a,b\to 0+} \nabla(ab)^T = \lim_{a,b\to 0+} (b,a) = 0.$$

Therefore, it would be necessary to require strict complementarity to establish Theorem 7.4 for $\mathrm{MCP}(F, \mathbb{R}^n_+)$ if we use $\psi_i = \varphi$.

REMARK 7.5. In passing we note that it is not possible to find a continuously differentiable NCP-function ϕ such that $\nabla\phi(0,0) \neq 0$. In fact, by the implicit function theorem, $\nabla\phi(0,0) \neq 0$ would imply that in a neighborhood of $(0,0)$, the zero set $\phi^{-1}(0)$ is a differentiable curve, which is not true since it has a kink at $(0,0)$. □

The difficulty with φ arises for $(a,b) > 0$ approaching zero. This problem can be resolved by replacing the term a_+b_+ in the definition of φ with $a_+b_+/\omega(|a|+|b|)$, where $\omega \in C^2([0,\infty), \mathbb{R})$ is any function satisfying

$$\omega(0) = 0, \quad \omega'(0) = 1, \qquad 0 < \omega(t) \leq \kappa, \quad \omega'(t) \geq 0, \quad \omega''(t) \leq 0 \quad \forall\, t > 0 \tag{7.8}$$

with some constant $\kappa > 0$. The reader should think about ω as a smooth variant of the cutoff-function $t \mapsto \min\{t, \kappa\}$. Typical examples are

$$\omega(t) = \kappa(1 - e^{-t/\kappa}) \quad \text{or} \quad \omega(t) = \min\{\kappa, t - t^2/(3\kappa) + t^3/(27\kappa^2)\}.$$

We thus obtain the NCP-function

$$\phi(a,b) \stackrel{\text{def}}{=} \frac{a_+ b_+}{\omega(|a| + |b|)} - \sqrt{a_-^2 + b_-^2} \tag{7.9}$$

with the interpretation $\phi(0,0) = 0$. As already φ, the function ϕ is continuously differentiable at (a,b) whenever $a, b \neq 0$.

For small $(a,b) > 0$ the function $\phi(a,b)$ essentially behaves like $ab/(a+b)$, a function that is linear along lines passing through the origin. On the other hand, for $(a,b) > 0$ sufficiently far away from the origin, we qualitatively have

$$\phi(a,b) \approx \frac{1}{\kappa} ab,$$

(if κ is chosen tight) showing that the term ab is recovered.

Before we proceed to the general case, we discuss the connection to other NCP-functions. For an overview see [47]. Probably the most popular NCP-function is the Fischer-Burmeister function [18]

$$\phi_{\text{FB}}(a,b) \stackrel{\text{def}}{=} a + b - \sqrt{a^2 + b^2}.$$

Hereby, we have reversed the sign of the original definition in [18] to achieve that the elements of the B-subdifferential are nonnegative. ϕ_{FB} is a smooth function except at the point $(0,0)$. It is 1-order semismooth [19] and for all $(a,b) \in \mathbb{R}^2$ and all $v \in \partial_B \phi_{\text{FB}}(a,b)$ holds $v \geq 0$, $v_1 + v_2 \geq 2 - \sqrt{2} \geq 1/2$. Recently, Chen, Chen and Kanzow [5] observed that the performance of methods that are based on Newton methods for semismooth reformulations (1.4) of NCPs can be increased by replacing the Fischer-Burmeister function by the following so-called penalized Fischer-Burmeister function

$$\phi_\lambda(a,b) \stackrel{\text{def}}{=} \lambda \phi_{\text{FB}}(a,b) + (1 - \lambda) a_+ b_+, \tag{7.10}$$

where $0 < \lambda < 1$. In [5], $\lambda = 0.95$ is used. The remarkable point is that the additional term $a_+ b_+$ is the same as in our NCP-function. Chen et al. [5] do not give a reason why they introduce the 'penalization' by $a_+ b_+$. In the derivation of our NCP-function ϕ it comes in completely naturally. In the numerical computations in section 8 we will see that, at least for our class of algorithms, our NCP-function ϕ is even superior to the penalized Fischer-Burmeister function. This seems to suggest the following procedure: On the positive orthant one should work with $a_+ b_+$ if the distance to the origin is not too small. Close to the origin, however, a different choice is necessary to achieve $0 \notin \partial_B \phi(0,0)$.

A possible explanation why the presence of the term $a_+ b_+$ leads to better performance compared to working with ϕ_{FB} may be that $a_+ b_+$ grows at least linearly along lines whereas $\phi_{\text{FB}}(a,b)$ asymptotically approaches the value b if b is fixed and a tends to $+\infty$.

Now let us return to the general case (7.1). It is clear that we can choose

$$\psi_i(a,b) \stackrel{\text{def}}{=} \begin{cases} \phi(a - l_i, b) & \text{if } -\infty < l_i < u_i = +\infty, \\ -\phi(u_i - a, -b) & \text{if } -\infty = l_i < u_i < +\infty, \\ b & \text{if } -\infty = l_i < u_i = +\infty. \end{cases} \tag{7.11}$$

It remains to discuss the case $-\infty < l_i < u_i < +\infty$. Then we use ϕ to define

$$\psi_i(a,b) \stackrel{\text{def}}{=} \begin{cases} \phi(a-l_i,b) & \text{if } a \le u_i \text{ and } b \ge 0, \\ -\phi(u_i-a,-b) & \text{if } a \ge l_i \text{ and } b \le 0. \end{cases} \tag{7.12}$$

Note that $\phi(a-l_i,0) = 0 = \phi(u_i-a,0)$ for $a \in [l_i,u_i]$. We still have to extend ψ_i to the whole of $\mathbb{R}^2$ by choosing $\psi_i(a,b) \ne 0$ for $a > u_i$ and $b > 0$ or $a < l_i$ and $b < 0$. We can do this as follows:

$$\psi_i(a,b) \stackrel{\text{def}}{=} \begin{cases} \sqrt{\phi(a-l_i,b)^2 + (u_i-a)^2} & \text{if } a > u_i \text{ and } b > 0, \\ -\sqrt{(a-l_i)^2 + \phi(u_i-a,-b)^2} & \text{if } a < l_i \text{ and } b < 0. \end{cases}$$

Analyzing signs, we see that the resulting MCP-function can be written in the closed form

$$\begin{aligned} &\psi_i(a,b) \stackrel{\text{def}}{=} \sqrt{\phi(a-l_i,b)_+^2 + (a-u_i)_+^2} - \sqrt{\phi(u_i-a,-b)_+^2 + (l_i-a)_+^2}, \\ &\text{if} \quad -\infty < l_i < u_i < +\infty, \end{aligned} \tag{7.13}$$

which might look complicated but simplifies the proof of smoothness results. Moreover, while verifying (7.13), one observes that always at least one of the two square roots vanishes.

LEMMA 7.6. *The function ϕ defined in* (7.9) *is* 1*-order semismooth on* $\mathbb{R}^2$ *and continuously differentiable on* $\mathbb{R}^2 \setminus \phi^{-1}(0)$*. For all $a,b \in \mathbb{R}$ and all $v \in \partial\phi(a,b)$ holds*

$$v \ge 0, \quad v_1 + v_2 \ge \frac{1}{2}, \tag{7.14}$$

$$v_1 = 0 \quad \textit{if } a > 0 \textit{ and } b = 0, \tag{7.15}$$

$$v_2 = 0 \quad \textit{if } a = 0 \textit{ and } b > 0. \tag{7.16}$$

Proof. Throughout, let $z = (a,b)^T$. We define

$$\phi_1(z) = \frac{ab}{\omega(a+b)}, \quad \phi_2(z) = b, \quad \phi_3(z) = -\|z\|, \quad \phi_4(z) = a. \tag{7.17}$$

On $\mathbb{R}^2 \setminus \phi^{-1}(0)$ the function ϕ is C^1, since $\phi = \phi_1$ on $\{z > 0\}$ and $\phi(z) = -\sqrt{a_-^2 + b_-^2}$ on $\mathbb{R}^2 \setminus \{z \ge 0\}$, respectively, both being C^1 on the respective sets.

On $\{z > 0\}$, $\phi = \phi_1$ even is C^2 and thus 1-order semismooth. The gradient is

$$\nabla\phi_1(z) = \frac{1}{\omega(a+b)}\begin{pmatrix} b \\ a \end{pmatrix} - \frac{ab\omega'(a+b)}{\omega(a+b)^2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{7.18}$$

With appropriate $c \in (0, a+b)$ holds $\omega(a+b) = \omega'(c)(a+b)$. Using $\omega'(c) \ge \omega'(a+b) \ge 0$, we derive

$$\begin{aligned} \omega(a+b)^2\frac{\partial\phi}{\partial a}(z) &= b\omega(a+b) - ab\omega'(a+b) = b\omega'(c)(a+b) - ab\omega'(a+b) \\ &\ge \omega'(c)(b(a+b) - ab) = b^2\omega'(c) = \frac{b^2\omega(a+b)}{a+b} > 0, \\ \omega(a+b)^2\frac{\partial\phi}{\partial b}(z) &\ge \frac{a^2\omega(a+b)}{a+b} > 0. \end{aligned}$$

Thus, for $v = \nabla\phi_1(z)$ we obtain

$$v_1 + v_2 \geq \frac{a^2+b^2}{(a+b)\omega(a+b)} \geq \frac{(a+b)^2}{2(a+b)\omega(a+b)} \geq \frac{1}{2}.$$

This establishes (7.14) on $\{z > 0\}$.

On $\{a > 0,\ b < 0\}$ the function $\phi = \phi_2$ is C^∞ with $\nabla\phi(z) = \nabla\phi_2(z) = (0,1)^T$, and thus 1-order semismooth. Here, (7.14) certainly holds.

On $\{z < 0\}$ the function $\phi = \phi_3$ is C^∞ with gradient

$$\nabla\phi_3(z) = -\frac{z}{\|z\|}.$$

In particular, ϕ is 1-order semismooth there and (7.14) is satisfied.

On $\{a < 0,\ b > 0\}$ we have $\phi = \phi_4$, being C^∞ and thus 1-order semismooth with $\nabla\phi(z) = \nabla\phi_4(z) = (1,0)^T$. Again, (7.14) holds trivially.

In a neighborhood of z with $a = 0$, $b < 0$ we can write $\phi = \mathrm{pw}(1,0;\phi_3,\phi_2)$, see Lemma 2.10, and $\phi_{2/3}$ are C^∞. Hence, ϕ is 1-order semismooth at z and (7.14) holds, since $\nabla\phi(z) = (0,1)^T$.

At z with $a < 0$, $b = 0$ locally holds $\phi = \mathrm{pw}(2,0;\phi_3,\phi_4)$ with $\phi_{3/4}$ being C^∞. Thus, ϕ is 1-order semismooth at z and (7.14) holds, since $\nabla\phi(z) = (1,0)^T$.

Around z with $a > 0$, $b = 0$ we locally have $\phi = \mathrm{pw}(2,0;\phi_2,\phi_1)$ with the C^2-functions $\phi_{1/2}$. Thus, ϕ is 1-order semismooth at z and (7.14), (7.15) are satisfied, since

$$\partial_B\phi(z) = \left\{\nabla\phi_1(z)^T, \nabla\phi_2(z)^T\right\} = \{(0, a/\omega(a)), (0,1)\}.$$

Similarly, in a neighborhood of z with $a = 0$, $b > 0$ we have $\phi = \mathrm{pw}(1,0;\phi_4,\phi_1)$ with the C^2-functions $\phi_{1/4}$. Now

$$\partial_B\phi(z) = \left\{\nabla\phi_1(z)^T, \nabla\phi_4(z)^T\right\} = \{(b/\omega(b), 0), (1,0)\}$$

implies (7.14) and (7.16).

It remains to discuss the case $z = 0$. First, we establish the local Lipschitz continuity of ϕ at 0. The continuity at 0 is obvious. From the properties of ω follows that there exists $\varepsilon > 0$ such that $\omega(t) \geq t/2$ on $[0,\varepsilon]$. This and the structure of the gradients of ϕ_i, $i = 1,\dots,4$, imply that there exists a constant L such that $\|\nabla\phi(a,b)\| \leq L$ at all differentiability points $(a,b) \in U_\varepsilon = \{(a,b)\,;\ |a|+|b| \leq \varepsilon\}$. Therefore, for all $v, w \in U_\varepsilon$ the function $\phi(v + t(w-v))$ is continuous on $[0,1]$ and is differentiable at all but finitely many points with $|\frac{d}{dt}\phi(v+t(w-v))| \leq L\|w-v\|$. Therefore, by the mean value theorem ϕ is Lipschitz continuous on U_ε of rank L.

Now we prove 1-order semismoothness. Let $s = (s_1,s_2)^T \in \mathbb{R}^2 \setminus \{0\}$ be arbitrary. For $s \in \phi^{-1}(0)$, i.e., $s \geq 0$, $s_1 s_2 = 0$, we have $[0,s] \subset \phi^{-1}(0)$ and see from (7.15) and (7.16) that for all $v^T \in \partial\phi(s)$ holds

$$v^T s - \phi'(0,s) = 0 - 0 = 0.$$

If $s_1 < 0$ or $s_2 < 0$ then with $s_- = ((s_1)_-, (s_2)_-)^T$ holds $\nabla\phi(s) = -s_-/\|s_-\|$ and

$$\phi'(0,s) = \lim_{t\to 0^+} -\frac{\|ts_-\|}{t} = -\|s_-\| = -\frac{s_-^T s}{\|s_-\|} = \nabla\phi(s)^T s.$$

Finally, if $s > 0$ we get with $y = s_1 + s_2$

$$\phi'(0,s) = \lim_{t\to 0^+} \frac{t^2 s_1 s_2}{t\omega(ty)} = \lim_{t\to 0^+} \frac{s_1 s_2}{\omega(ty)/t} = \frac{s_1 s_2}{\omega'(0)y} = \frac{s_1 s_2}{y}.$$

Since ω is C^2, we have, as $s \to 0$,

$$\omega(y) = \omega(0) + \omega'(0)y + O(y^2) = y + O(y^2), \quad \omega'(y) = \omega'(0) + O(y) = 1 + O(y).$$

Now

$$\begin{aligned}\nabla\phi(s)^T s - \phi'(0,s) &= \frac{2s_1 s_2}{\omega(y)} - \frac{s_1 s_2 \omega'(y) y}{\omega(y)^2} - \frac{s_1 s_2}{y} \\ &= \frac{s_1 s_2 (2y(y+O(y^2)) - (1+O(y))y^2 - (y+O(y^2))^2}{y(y+O(y^2))^2} \\ &= \frac{s_1 s_2 O(y^3)}{y^3 + O(y^4)} = O(s_1 s_2) = O(\|s\|^2).\end{aligned}$$

Therefore, we have shown that

$$\sup_{v\in\partial\phi(s)} \|v^T s - \phi'(0,s)\| = O(\|s\|^2) \qquad (\text{as } s \to 0),$$

which proves that ϕ is also 1-order semismooth at 0. Since (7.14) holds on $\mathbb{R}^2 \setminus \{0\}$, the definition of $\partial\phi$ implies that it is also satisfied at 0. □

We now discuss the properties of the MCP-function ψ_i.

LEMMA 7.7. *The MCP-function ψ_i defined by* (7.11) *and* (7.13) *is* 1*-order semismooth on $\mathbb{R}^2$ and continuously differentiable on $\mathbb{R}^2 \setminus \psi_i^{-1}(0)$. For all $a, b \in \mathbb{R}$ and all $v \in \partial\psi(a,b)$ holds*

$$v \geq 0, \quad v_1 + v_2 \geq \frac{1}{2}, \tag{7.19}$$

$$v_1 = 0 \quad \text{if } l < a < u \text{ and } b = 0, \tag{7.20}$$

$$v_2 = 0 \quad \text{if } a = l \text{ and } b > 0 \text{ or } a = u \text{ and } b < 0. \tag{7.21}$$

Proof. Again, we set $z = (a,b)^T \in \mathbb{R}^2$. In the cases $l_i = -\infty$ or $u_i = +\infty$ the assertion follows immediately from (7.11) and Lemma 7.6. Now let $-\infty < l_i < u_i < +\infty$. The continuous and piecewise linear function $t \mapsto t_+$ and the smooth function $t \mapsto t^2$ are 1-order semismooth, see Lemma 2.10 and Proposition 2.9. Moreover, $\sqrt{a^2+b^2} = a + b - \phi_{\text{FB}}(a,b)$ is 1-order semismooth [19, Lem. 20]. Since according to Proposition 2.9 the composite of 1-order semismooth functions is 1-order semismooth, we obtain that $\psi_i(a,b)$ as defined in (7.13) is 1-order semismooth.

On the set $\{z\,;\ a > u_i \vee (a > l_i \wedge b > 0)\}$ we have

$$\psi_i(z) = \left(\phi(a-l_i,b)_+^2 + (a-u_i)_+^2\right)^{1/2} > 0,$$

$$\nabla\psi_i(z) = \left(\nabla\phi(a-l_i,b), \begin{pmatrix}1\\0\end{pmatrix}\right) \frac{\begin{pmatrix}\phi(a-l_i,b)_+\\(a-u_i)_+\end{pmatrix}}{\sqrt{\phi(a-l_i,b)_+^2 + (a-u_i)_+^2}}.$$

Similarly, on $\{z\,;\ a < l_i \vee (a < u_i \wedge b < 0)\}$ holds

$$\psi_i(z) = -\left(\phi(u_i - a, -b)_+^2 + (l_i - a)_+^2\right)^{1/2} < 0,$$

$$\nabla\psi_i(z) = \left(\nabla\phi(u_i - a, -b), \begin{pmatrix}1\\0\end{pmatrix}\right) \frac{\begin{pmatrix}\phi(u_i - a, -b)_+\\(l_i - a)_+\end{pmatrix}}{\sqrt{\phi(u_i - a, -b)_+^2 + (l_i - a)_+^2}}.$$

In both cases the gradient assumes the form $v = \alpha w + \beta(1,0)^T$ with $\alpha, \beta \geq 0$, $\alpha^2 + \beta^2 = 1$, and $w \geq 0$, $w_1 + w_2 \geq 1/2$. Therefore, $v \geq 0$ and

$$v_1 + v_2 = \alpha(w_1 + w_2) + \beta \geq \frac{\alpha}{2} + \beta \geq \frac{\alpha + \beta}{2} \geq \frac{1}{2}\sqrt{\alpha^2 + \beta^2} = \frac{1}{2},$$

which proves (7.19) for all $z \notin \psi_i^{-1}(0)$. From the definition of ψ_i and the investigations on ϕ we know that for $a = l_i$, $b > 0$ or $a = u_i$, $b < 0$ holds

$$\partial_B\psi_i(z) = \partial_B\phi(0, |b|) = \{(|b|/\omega(|b|), 0), (1, 0)\},$$

which proves (7.19) for these z, and (7.20).

For $l_i < a < u_i$ and $b = 0$ we have in a neighborhood of z that $\psi_i = \mathrm{pw}(2, 0; f_1, f_2)$ with C^2-functions $f_1(z) = -\phi_1(u_i - a, -b)$ and $f_2(z) = \phi_1(a - l_i, b)$ and thus by Lemma 2.10

$$\partial_B\psi_i(z) = \left\{\nabla\phi_1(a - l_i, 0)^T, \nabla\phi_1(u_i - a, 0)^T\right\} = \left\{\left(0, \frac{a - l_i}{\omega(a - l_i)}\right), \left(0, \frac{u_i - a}{\omega(u_i - a)}\right)\right\},$$

thus yielding (7.19) for these z, and (7.21).

Finally, since $\partial_B\psi_i(0)$ consists of limits of derivatives and (7.19) holds on $\mathbb{R}^2 \setminus \{0\}$, it is also satisfied for all $v \in \partial\psi_i(0)$. □

8. Numerical Results. In this section we present numerical results for Algorithm 3.6. As test problems we use a subset of the MCPLIB [13] collection of mixed complementarity problems. The algorithm is applied to the reformulation (1.4) of the MCP. For the sake of comparison, test versions of four different algorithms were implemented in MATLAB. The difference between these methods consists in the choice of the MCP- (or NCP-) function and the choice of the reformulation (constrained/unconstrained):

ALG1: Algorithm 3.6 applied to (1.1) with H as in (1.4). For ψ_i we choose the new MCP-function as defined in (7.11), (7.13) with ϕ denoting the new affine-scaling NCP-function defined in (7.9). This is the new method we would like to emphasize. It is applicable to the general MCP.

ALG2: Here, the function H is the same as in ALG1. However, we apply Algorithm 3.6 to the *unconstrained* reformulation (1.5).

ALG3: As ALG1, but based on the penalized Fischer-Burmeister NCP-function $\phi = \phi_\lambda$, $\lambda = 0.95$, as defined in (7.10). Hereby, by using (7.11), we can handle any MCP for which every component x_i is subject to at most one bound, i.e., $u_i - l_i = \infty$, $i = 1, \ldots, n$.

ALG4: Here, the function H is the same as in ALG3. However, we apply Algorithm 3.6 to the *unconstrained* reformulation (1.5).

We applied ALG1 and ALG2 to all problems in MCPLIB of size $n <= 150$ that are accessible from within MATLAB (some of the problems are only available as GAMS files). ALG3 and ALG4 were tested for the subset of these problems with at most one bound per variable, i.e.,

$$u_i - l_i = \infty, \quad i = 1, \dots, n. \tag{8.1}$$

The MATLAB interface to MCPLIB provides an initialization routine that returns l, u, and an initial point x_0. Moreover, functions for the computation of F and its Jacobian F' are available. Therefore, as described in the Appendix, it is easy to compute exact (up to roundoff errors) elements of the B-subdifferential of H. Hence, we may assume that that matrices M_k are elements of $\partial_B H(x_k)$ up to machine accuracy. If M_k is too ill-conditioned, we add a small multiple of the identity to $M_k^T M_k$. We implemented Algorithm 3.6 on the basis of the fraction of Cauchy decrease condition (6.1) with Step 3.3 as described in (III) at the end of section 6. If $s_k = s_k^{\mathrm{P}}$ fails to satisfy (6.1), we compute s_k by solving the trust-region subproblem (3.9) exactly. Hereby, the QP-solver BPMPD [35] is used. The parameters in Algorithm 3.6 were chosen as follows:

• Termination criterion:
Successful termination if $\|\mathrm{mid}(x-l, x-u, F(x))\|_\infty \le 10^{-6}$, where mid denotes the componentwise median of the three arguments, i.e., the second largest entry.
Unsuccessful termination if $\Delta_k \le 10^{-10}$ or $k \ge 200$.

• Trust-region parameters: $\Delta_{\min} = 1$, $\Delta_0 = 100$, $\eta_1 = 10^{-4}$, $\eta_2 = 0.75$, $\gamma_1 = 1/2$, $\gamma_2 = 2$,

$$\Delta_{k+1} = \begin{cases} \gamma_1 \Delta_k & \text{if } \rho_k \le \eta_1, \\ \max\{\Delta_{\max}, \Delta_k\} & \text{if } \eta_1 < \rho_k < \eta_2, \\ \max\{\Delta_{\max}, \gamma_2 \Delta_k\} & \text{if } \rho_k \ge \eta_2. \end{cases}$$

• Non-monotonicity parameters: $m = 4$, $\lambda = 0.01$, and λ_{kr} as in (3.18).

• Affine-scaling parameter: $\gamma = 1$, $\kappa_D = 1$.

• Initial point: Denoting by $\hat{x}_0$ the initial point returned by the initialization routine, we computed x_0 by projecting $\hat{x}_0$ onto X and pushing it slightly away from the bounds into X.

The numerical results are shown in Table 1. For each of the four algorithms the number of major iterations (Maj. It), i.e., the value of $(i+1)$ at termination, the number of iterations (It), i.e., the value of k at successful termination, and the number of QP-subproblems that had to be solved (QPs) are reported. Note that the number of evaluations of F equals (It + 1) and that the Jacobian of F is evaluated once per major iteration. The entry '$*$' means that the algorithm (applies only to ALG3 and ALG4) is not applicable since (8.1) is violated. This is the case for the problems 'choi', 'gafni', and 'pies'. By '$-$' we mark that the algorithm terminated unsuccessfully. At the bottom we display the sum (Σ) over all column entries corresponding to problems that were successfully solved by all four methods. Moreover, we give an overall ranking of the algorithms for each of the three categories 'Maj. It', 'It', and 'QPs'. Hereby, in every category and for every problem to which all four algorithms can be applied (i.e., no '$*$' in this row), we compute for each algorithm a rank between 1 and 5 as follows:

Let *cat* denote the category, *prob* the problem, and i_k the entries for ALGk, $k = 1, \dots, 4$, where $i_k = +\infty$ for '$-$'-entries. We define the rank achieved by algorithm ALGk in category *cat* for problem *prob* by

$$\mathit{rank}(\mathrm{ALG}k,\mathit{cat},\mathit{prob}) \overset{\mathrm{def}}{=} \begin{cases} |\{j\,;\ i_j < i_k\}| + 1 & \text{if } i_k < +\infty, \\ 5 & \text{if } i_k = +\infty. \end{cases}$$

For example, in the category 'It' and for problem 'ehl_k40' we have $(i_1, \dots, i_4) = (9, 9, 24, +\infty)$ and thus $\mathit{rank}(\mathrm{ALG}k,\mathrm{It},\mathrm{ehl\_k40})_{1\le k\le 4} = (1, 1, 3, 5)$.

	Maj. It				It				QPs			
ALG:	1	2	3	4	1	2	3	4	1	2	3	4
badfree	1	1	4	5	1	1	4	5	0	0	0	0
bertsekas	8	7	20	20	10	9	40	39	0	0	21	23
billups	–	–	–	–	–	–	–	–	–	–	–	–
choi	4	4	∗	∗	4	4	∗	∗	0	0	∗	∗
colvdual	11	17	21	12	11	24	30	14	3	7	9	0
colvnlp	9	18	11	12	9	25	12	14	3	3	0	0
cycle	5	5	7	7	5	5	7	7	0	0	0	0
degen	3	3	4	5	3	3	4	5	1	0	0	0
duopoly	14	31	12	–	23	60	12	–	20	46	10	–
ehl_k40	9	9	24	–	9	9	45	–	0	0	5	–
ehl_k60	10	10	32	31	10	10	57	65	0	0	8	12
ehl_k80	11	11	25	–	11	11	46	–	0	0	4	–
ehl_kost	11	11	–	–	11	11	–	–	2	0	–	–
freebert	7	7	12	22	7	7	23	83	0	0	0	66
gafni	12	12	∗	∗	12	12	∗	∗	0	0	∗	∗
games	16	–	–	20	31	–	–	40	25	–	–	21
hanskoop	7	32	11	10	7	74	31	19	3	23	9	2
hydroc06	8	8	7	7	9	9	7	7	0	0	1	1
hydroc20	20	21	11	11	31	34	12	12	7	10	0	0
jel	4	4	7	7	4	4	7	7	0	0	0	0
josephy	6	6	6	7	14	15	6	7	6	7	1	0
kojshin	7	10	7	6	14	20	7	11	0	5	0	0
mathinum	4	4	8	4	4	4	18	4	0	0	1	0
mathisum	4	4	5	5	4	4	12	12	0	0	0	0
methan08	4	4	4	4	4	4	4	4	0	0	0	0
nash	6	6	8	8	6	6	8	8	0	0	0	0
ne-hard	–	–	–	–	–	–	–	–	–	–	–	–
pgvon105	21	–	–	–	50	–	–	–	27	–	–	–
pgvon106	–	–	–	–	–	–	–	–	–	–	–	–
pies	9	9	∗	∗	9	9	∗	∗	4	0	∗	∗
powell	6	13	7	9	14	21	7	16	1	1	0	0
powell_mcp	6	6	6	6	6	6	6	6	0	0	0	0
qp	4	4	3	5	4	4	3	5	0	0	0	0
scarfanum	5	7	9	11	12	7	15	19	9	0	1	1
scarfasum	5	6	7	7	5	6	18	18	1	0	0	0
scarfbnum	9	13	18	20	9	23	18	31	3	14	1	5
scarfbsum	8	12	14	14	8	23	14	24	4	14	4	7
shubik	7	7	25	32	9	9	69	82	3	3	59	70
simple-ex	–	–	–	–	–		–	–	–	–	–	–
simple-red	10	10	10	10	10	10	10	10	0	0	0	0
sppe	8	21	8	8	14	43	8	8	3	26	0	0
tinloi	6	7	8	7	6	7	27	7	0	0	3	0
tobin	15	16	11	14	31	30	15	17	12	12	3	0
Σ	213	289	311	326	281	447	499	566	59	125	121	187
Rank 1	29	18	10	8	25	18	12	7	21	22	20	23
Rank 2	4	6	3	5	5	3	3	7	6	0	3	1
Rank 3	3	3	16	10	5	6	14	8	6	5	7	3
Rank 4	0	7	4	8	1	7	4	9	3	7	3	4
Rank 5 (fails)	4	6	7	9	4	6	7	9	4	6	7	9

TABLE 1
Numerical Results

We make a number of observations.

- Apparently, the algorithms ALG1 and ALG3, which use the proposed *box-constrained* reformulations (1.1) and (1.2), are more robust than their counterparts ALG2 and ALG4, which are based on the *unconstrained* reformulations (1.5) and (1.6).
- Comparing ALG1 with ALG3 and ALG2 with ALG4 shows that the usage of the new affine-scaling NCP/MCP-function instead of the penalized Fischer-Burmeister function leads to an improvement in all three categories (major iterations, iterations, number of QPs). This makes the new MCP-function a good choice for numerical implementations.
- ALG1, which combines the proposed Algorithm 3.6, the new MCP-function and the constrained reformulation is the most robust and efficient method in the test. It only fails to solve 4 problems, followed by ALG2, ALG3, and ALG4 with 6, 7, and 9 fails. In addition, it needs substantially less iterations—and thus function evaluations—and only half as many calls of the QP solver than the other three competitors.
- To demonstrate the importance of non-monotone trust-region techniques in this context, we ran a monotone version of our algorithm on the testset. If in algorithm ALG1 we set $m = 1$, i.e., $\text{rared}_k(s) = \text{ared}_k(s)$, then the resulting algorithm fails for 12 problems compared to only 4 fails with $m = 4$. This shows that the proposed non-monotonicity technique increases the robustness of the algorithm considerably.

9. Conclusions. We have introduced a class of non-monotone trust region methods for box-constrained semismooth equations. For these algorithms a comprehensive global convergence theory was established. The method remains feasible with respect to the box-constraints and is based on a reformulation as a simply constrained smooth minimization problem. A Newton-like method with projection was proposed for the computation of trial steps that converges under a Dennis–Moré-type condition locally q-superlinearly/quadratically to a BD-regular solution. We showed that close to the solution the trust-region algorithm turns into this Newton-like method. The convergence analysis was carried out on the general basis of a criticality measure and a sufficient decrease condition. As a concrete implementation we discussed a fraction of Cauchy decrease condition in which the negative affinely scaled gradient is used as Cauchy direction.

The developed algorithm was applied to the solution of the nonlinear mixed complementarity problem. To this end, the MCP was converted into an equivalent bound-constrained semismooth equation. Hereby, a new MCP-function was used that is motivated by affine-scaling methods for nonlinear programming. The properties of this MCP-function were investigated in detail. In particular, we proved the 1-order semismoothness of the resulting equation and the BD-regularity of strongly regular solutions. Therefore, the global and local convergence results for the developed algorithm are applicable for this broad class of problems. The computation of elements of the BD-subdifferential is discussed in the Appendix. The numerical results presented for a subset of the MCPLIB testset show that—even in the case where H does not have a zero outside of X—incorporating the a priori knowledge $x \in X$ into the algorithm leads to more robustness and efficiency. This confirms the relevance of the problem class (1.1) and the need of algorithms for its solution together with the corresponding theory.

10. Appendix: Computation of B-subdifferentials. In our numerical computations we worked with exact B-subdifferentials, i.e., $M_k \in \partial_B H(x_k)$. Nevertheless, also in this case the general theory based on the different flavors (3.1)–(3.3) of Dennis-Moré-type conditions is of importance since it guarantees fast convergence if M_k and s_k^{N} are not exact (e.g., due to roundoff errors) but sufficiently accurate.

Let H be defined by (1.4) with $\psi_i = \psi$ as in (7.11) and (7.13), respectively. Let $x \in U$ be arbitrary and fixed. We know already from Lemma 7.1 that every $V \in \partial_B H(x)$ is of the form $V = D_a + D_b F'(x)$ with positive semidefinite diagonal matrices D_a and D_b. Our aim is to compute particular instances of D_a and D_b.

We stress that in our algorithm all iterates are feasible. Therefore, we only need to compute V for $x \in X$. Nevertheless, we treat the general case $x \in \mathbb{R}^n$, since this does not cause any additional difficulties. For $x \in \mathbb{R}^n$ we define

$$\begin{aligned} I_l &= \{i\,;\ x_i = l_i,\ F_i(x) \ge 0\}\,, \quad I_u = \{i\,;\ x_i = u_i,\ F_i(x) \le 0\}\,,\\ I_f &= \{i\,;\ x_i \notin \{l_i, u_i\}\}\,, \quad I_0 = \{i\,;\ l_i \le x_i \le u_i,\ F_i(x) = 0\}\,,\\ I_{00} &= \{i \in I_0\,;\ \nabla F_i(x) = 0\}\,, \quad I_{01} = \{i \in I_0\,;\ \nabla F_i(x) \ne 0\}\,. \end{aligned}$$

Now let $s \in \mathbb{R}^n$ be any vector satisfying

$$s_i > 0 \ \ (i \in I_l), \quad s_i < 0 \ \ (i \in I_u), \quad s_i \ne 0 \ \ (i \in I_{00}), \quad \nabla F_i(x)^T s \ne 0 \ \ (i \in I_{01}).$$

We do not describe how such an s can be obtained since almost all s that satisfy the first three conditions, which are easy to satisfy, also verify the fourth one. A constructive way to obtain a vector s with similar properties is described in [26]. If $\tau > 0$ is sufficiently small it follows that for all $t \in (0, \tau]$ holds

$$x_i + ts_i \notin \{l_i, u_i\} \ \ (i \in I_f \cup I_l \cup I_u), \quad F_i(x + ts)(\nabla F_i(x)^T s) > 0 \ \ (i \in I_{01}).$$

In particular, for all $i \notin I_{00}$ the functions H_i are C^1 in a neighborhood of $x + ts$, $t \in (0, \tau]$. Note also that $x_i + ts_i \in (l_i, u_i)$ for all $i \in I_l \cup I_u$ and $0 < t \le \tau$. Now let $(t_k) \downarrow 0$, $t_k \le \tau$ and $V_k \in \partial_B H(x + t_k s)$. Due to the upper semicontinuity of $\partial_B H$, see Proposition 2.2, we may assume (possibly after selecting a subsequence) that $V_k \to V \in \partial_B H(x)$. We decompose $V^T = (v_1, \dots, v_n)$.

1. $i \in I_{00}$: By Corollary 2.5 and Proposition 2.8

$$\begin{aligned} (D_a)_{ii} s_i = (Vs)_i &= (\lim_{k\to\infty} V_k s)_i = H_i'(x, s)\\ &= \psi_i'((x_i, F_i(x)), (s_i, \nabla F_i(x)^T s)) = \psi_i'((x_i, 0), (s_i, 0)) = 0. \end{aligned}$$

Therefore, $(D_a)_{ii} = 0$ and $(D_b)_{ii} > 0$ can be chosen arbitrary, since $\nabla F_i(x) = 0$.

2. $i \notin I_l \cup I_u \cup I_0$: ψ_i is locally C^1 around $(x_i, F_i(x))$, and thus by the chain rule

$$v_i = \nabla H_i(x) = \frac{\partial \psi_i}{\partial a}(x_i, F_i(x)) e_i + \frac{\partial \psi_i}{\partial b}(x_i, F_i(x)) \nabla F_i(x).$$

Hence, $(D_a)_{ii} = \frac{\partial \psi_i}{\partial a}(x_i, F_i(x))$, $(D_b)_{ii} = \frac{\partial \psi_i}{\partial b}(x_i, F_i(x))$.

3. $i \in (I_l \cup I_u \cup I_0) \setminus I_{00}$: By construction, ψ_i is locally C^1 around $(x_i + t_k s_i, F_i(x + t_k s))$. Therefore,

$$v_i = \lim_{k\to\infty} \left(\frac{\partial \psi_i}{\partial a}(x_i + t_k s_i, F_i(x + t_k s)) e_i + \frac{\partial \psi_i}{\partial b}(x_i + t_k s_i, F_i(x + t_k s)) \nabla F_i(x + t_k s) \right).$$

(i) $i \in I_{01} \setminus (I_l \cup I_u)$:
(a) If $\nabla F_i(x)^T s > 0$ and $l_i > -\infty$ then with ϕ_1 as in (7.17)

$$((D_a)_{ii}, (D_b)_{ii}) = \nabla\phi_1(x_i - l_i, 0)^T = \left(0, \frac{x_i - l_i}{\omega(x_i - l_i)}\right).$$

(b) If $\nabla F_i(x)^T s > 0$, and $-\infty = l_i < u_i < +\infty$ then with ϕ_2 as in (7.17)

$$((D_a)_{ii}, (D_b)_{ii}) = \nabla\phi_2(u_i - x_i, 0)^T = (0, 1).$$

(c) If $-\infty = l_i < u_i = +\infty$ then

$$((D_a)_{ii}, (D_b)_{ii}) = (0, 1).$$

(d) The case $\nabla F_i(x)^T s < 0$ can be treated analogously.
(ii) $i \in (I_l \cup I_u) \setminus I_0$: Then with ϕ_1 as in (7.17)

$$((D_a)_{ii}, (D_b)_{ii}) = \nabla\phi_1(0, |F_i(x)|)^T = \left(\frac{|F_i(x)|}{\omega(|F_i(x)|)}, 0\right).$$

(iii) $i \in (I_l \cup I_u) \cap I_{01}$:
(a) If $x_i = l_i$ and $\nabla F_i(x)^T s > 0$ then, using

$$F_i(x + t_k s) = t_k \nabla F_i(x)^T s + o(t_k),$$
$$\omega(x_i + t_k s_i - l_i + F_i(x_i + t_k s_i)) = t_k(s_i + \nabla F_i(x)^T s) + o(t_k),$$

and (7.18) yields

$$\begin{aligned}((D_a)_{ii}, (D_b)_{ii}) &= \lim_{k\to\infty} \nabla\phi_1(x_i + t_k s_i - l_i, F_i(x + t_k s))^T \\ &= \frac{(\nabla F_i(x)^T s, s_i)}{s_i + \nabla F_i(x)^T s} - \frac{s_i \nabla F_i(x)^T s}{(s_i + \nabla F_i(x)^T s)^2}(1,1) = \frac{((\nabla F_i(x)^T s)^2, s_i^2)}{(s_i + \nabla F_i(x)^T s)^2}.\end{aligned}$$

(b) If $x_i = l_i$, $\nabla F_i(x)^T s < 0$, and $u_i < +\infty$ then

$$((D_a)_{ii}, (D_b)_{ii}) = \nabla\phi_1(u_i - x_i, -F_i(x))^T = \nabla\phi_1(u_i - x_i, 0)^T = \left(0, \frac{u_i - x_i}{\omega(u_i - x_i)}\right).$$

(c) If $x_i = l_i$, $\nabla F_i(x)^T s < 0$, and $u_i = +\infty$ then with ϕ_2 as in (7.17)

$$((D_a)_{ii}, (D_b)_{ii}) = \nabla\phi_2(0, 0)^T = (0, 1).$$

(d) Due to symmetry, the case $x_i = u_i$ can be treated analogously.

Acknowledgments. I would like to thank my brother Stefan, Technische Universität München, for several helpful discussions. Moreover, I am grateful to Csaba Mészáros, Computer and Automation Research Center, Budapest, for making available his QP-code BPMPD.

REFERENCES

[1] S. C. Billups, *Algorithms for complementarity problems and generalized equations*, Ph.D. Thesis, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin, 1995.

[2] S. C. Billups, S. P. Dirkse, and M. C. Ferris, *A comparison of large scale mixed complementarity problem solvers*, Comput. Optim. Appl., 7 (1997), pp. 3–25.

[3] S. C. Billups and M. C. Ferris, *QPCOMP: A quadratic program based solver for mixed complementarity problems*, Math. Programming, 76 (1997), pp. 533–562.

[4] J. V. Burke, J. J. Moré, and G. Toraldo, *Convergence properties of trust region methods for linear and convex constraints*, Math. Programming, 47 (1990), pp. 305–336.

[5] B. Chen, X. Chen, and C. Kanzow, *A penalized Fischer-Burmeister NCP-function: Theoretical investigation and numerical results*, Preprint 126, Institute of Applied Mathematics, University of Hamburg, Hamburg, Germany, 1997.

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

[7] T. F. Coleman and Y. Li, *On the convergence of interior–reflective Newton methods for nonlinear minimization subject to bounds*, Math. Programming, 67 (1994), pp. 189–224.

[8] T. F. Coleman and Y. Li, *An interior trust region approach for nonlinear minimization subject to bounds*, SIAM J. Optimization, 6 (1996), pp. 418–445.

[9] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Global convergence of a class of trust region algorithms for optimization with simple bounds*, SIAM J. Numer. Anal., 25 (1988), pp. 433–460. See [10].

[10] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Correction to the paper on global convergence of a class of trust region algorithms for optimization with simple bounds*, SIAM J. Numer. Anal., 26 (1989), pp. 764–767.

[11] T. De Luca, F. Facchinei, and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Math. Programming, 75 (1996), pp. 407–439.

[12] S. P. Dirkse and M. C. Ferris, *The PATH solver: A non-monotone stabilization scheme for mixed comple-mentarity problems*, Optimization Methods and Software, 5 (1995), pp. 123–156.

[13] S. P. Dirkse and M. C. Ferris, *MCPLIB: A collection of nonlinear mixed complementarity problems*, Optimization Methods and Software, 5 (1995), pp. 319–345.

[14] F. Facchinei and C. Kanzow, *A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems*, Math. Programming, 76 (1997), pp. 493–512.

[15] M. C. Ferris and C. Kanzow, *Complementarity and related problems: A survey*, Mathematical Programming Technical Report 98-17, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin, 1998.

[16] M. C. Ferris, C. Kanzow, and T. S. Munson, *Feasible descent algorithms for mixed complementarity problems*, Math. Programming, forthcoming, 1998.

[17] M. C. Ferris and J. S. Pang, *Engineering and economic applications of complementarity problems*, SIAM Review, 39 (1997), pp. 669–713.

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

[19] A. Fischer, *Solution of monotone complementarity problems with locally Lipschitzian functions*, Math. Programming, 76 (1997), pp. 513–532.

[20] A. Friedlander, J. M. Martínez, and A. Santos, *A new trust region algorithm for bound constrained minimization*, Appl. Math. Optim., 30 (1994), pp. 235–266.

[21] S. A. Gabriel and J. S. Pang, *A trust region method for constrained nonsmooth equations*, in W. W. Hager, D. W. Hearn, and P. M. Pardalos, eds.; Large Scale Optimization: State of the Art, Kluwer Academic Publishers, Boston, MA, 1994, pp. 159–186.

[22] L. Grippo, F. Lampariello and S. Lucidi, *A nonmonotone line search technique for Newton's method*, SIAM J. Numer. Anal., 23 (1986), pp. 706–716.

[23] P. T. Harker and J. S. Pang, *Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications*, Math. Programming, 48 (1990), pp. 161–220.

[24] M. Heinkenschloss, M. Ulbrich, and S. Ulbrich, *Superlinear and quadratic convergence of affine-scaling interior-point Newton methods for problems with simple bounds without strict complementarity assumption*, Technical Report TR97–30, Department of Computational and Applied Mathematics, Rice University, Houston, Texas. To appear in Math. Programming.

[25] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms. Part I: Fundamentals*, Springer, Berlin, 1993.

[26] H. Jiang, M. Fukushima, L. Qi, and D. Sun, *A trust region method for solving generalized complementarity problems*, SIAM J. Optimization, 8 (1998), pp. 140–157.

[27] H. Jiang and L. Qi, *A new nonsmooth equations approach to nonlinear complementarity problems*, SIAM J. Control Optim., 35 (1997), pp. 178–193.

[28] H. Jiang and D. Ralph, *Global and local superlinear convergence analysis of Newton-type methods for semismooth equations with smooth least squares*, Manuscript, Department of Mathematics

& Statistics, University of Melbourne, 1997.
[29] C. Kanzow and H.-D. Qi, *A QP-free constrained Newton-type method for variational inequality problems*. To appear in Math. Programming, 1999.
[30] C. Kanzow and M. Zupke, *Inexact trust-region methods for nonlinear complementarity problems*, in M. Fukushima and L. Qi, eds.; Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods. Kluwer Academic Press, Dordrecht, The Netherlands, to appear.
[31] B. Kummer, *Newton's method for non-differentiable functions*, in J. Guddat et al., eds.; Mathematical Research, Advances in Mathematical Optimization. Akademie-Verlag, Berlin, 1988, pp. 114–125.
[32] B. Kummer, *Newton's method based on generalized derivatives for nonsmooth functions: Convergence analysis*, in W. Oettli and D. Pallaschke, eds.; Advances in Mathematical Optimization. Lecture Notes in Economics and Mathematical Systems, Vol. 382, Springer-Verlag, Heidelberg, 1992, pp. 171–194.
[33] C.-J. Lin and J. J. Moré, *Newton's method for large bound–constrained optimization problems*, Technical Report ANL/MCS-P724–0898, Mathematics and Computer Sciences Division, Argonne National Laboratories, 9700 South Cass Avenue, Argonne, Ill. 60439.
[34] J. Liu, *Strong stability in variational inequalities*, SIAM J. Control Optim., 15 (1995), pp. 725–749.
[35] Cs. Mészáros, *The BPMPD interior point solver for convex quadratic problems*, Working paper WP 98–8, Computer and Automation Research Institute, Hungarian Academy of Sciences, Budapest, Hungary, 1998.
[36] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim., 15 (1977), pp. 959–972.
[37] J. J. Moré, *Global methods for nonlinear complementarity problems*, Math. Oper. Res., 21 (1996), pp. 589–614.
[38] J. S. Pang, *Newton's method for B-differentiable equations*, Math. Oper. Res., 15 (1990), pp. 311–341.
[39] J. S. Pang, *A B-differentiable equation based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems*, Math. Programming, 51 (1991), pp. 101–132.
[40] J. S. Pang and S. A. Gabriel, *NE/SQP: A robust algorithm for the nonlinear complementarity problem*, Math. Programming, 60 (1993), pp. 295–337.
[41] J. S. Pang and L. Qi, *Nonsmooth equations: Motivation and algorithms*, SIAM J. Optimization, 3 (1993), pp. 443–465.
[42] S. M. Robinson, *Strongly regular generalized equations*, Math. Oper. Res., 5 (1980), pp. 43–62.
[43] S. M. Robinson, *Newton's method for a class of nonsmooth functions*, Set Valued Analysis, 2 (1994), pp. 291–305.
[44] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res., 18 (1993), pp. 227–244.
[45] L. Qi, *Trust region algorithms for solving nonsmooth equations*, SIAM J. Optimization, 5 (1995), pp. 219–230.
[46] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Programming, 58 (1993), pp. 353–367.
[47] D. Sun and L. Qi, *On NCP-functions*, to appear in Comput. Optim. Appl., 1998.
[48] Ph. L. Toint, *Global convergence of a class of trust-region methods for nonconvex minimization in Hilbert space*, IMA J. Numer. Anal., 8 (1988), pp. 231–252.
[49] Ph. L. Toint, *Non-monotone trust-region algorithms for nonlinear optimization subject to convex constraints*, Math. Programming, 77 (1997), pp. 69–94.
[50] M. Ulbrich and S. Ulbrich, *Superlinear convergence of affine-scaling interior-point Newton methods for infinite-dimensional nonlinear problems with pointwise bounds*, Tech. Report TR97–05, Department of Computational and Applied Mathematics, Rice University, Houston, Texas. Accepted subject to revisions for publication in SIAM J. Control Optim.
[51] M. Ulbrich, S. Ulbrich, and M. Heinkenschloss, *Global convergence of trust-region interior-point algorithms for infinite-dimensional nonconvex minimization subject to pointwise bounds*, SIAM J. Control Optim., 37 (1999), pp. 731–764.
[52] B. Xiao and P. T. Harker, *A nonsmooth Newton method for variational inequalities: I: Theory*, Math. Programming, 65 (1994), pp. 151–194.
[53] B. Xiao and P. T. Harker, *A nonsmooth Newton method for variational inequalities: II: Numerical results*, Math. Programming, 65 (1994), pp. 195–216.
[54] W. P. Ziemer, *Weakly differentiable functions. Sobolev spaces and functions of bounded variation*, Springer-Verlag, Berlin, 1989.