Math. Program., Ser. A 94: 167–187 (2002)

Digital Object Identifier (DOI) 10.1007/s10107-002-0305-2

Defeng Sun · Robert S. Womersley · Houduo Qi

A feasible semismooth asymptotically Newton method for mixed complementarity problems⋆

Received: December 1999 / Accepted: March 2002
Published online: September 5, 2002 –

Abstract. Semismooth Newton methods constitute a major research area for solving mixed complementarity problems (MCPs). Early research on semismooth Newton methods is mainly on infeasible methods. However, some MCPs are not well defined outside the feasible region or the equivalent unconstrained reformulations of other MCPs contain local minimizers outside the feasible region. As both these problems could make the corresponding infeasible methods fail, more recent attention is on feasible methods.

In this paper we propose a new feasible semismooth method for MCPs, in which the search direction asymptotically converges to the Newton direction. The new method overcomes the possible non-convergence of the projected semismooth Newton method, which is widely used in various numerical implementations, by minimizing a one-dimensional quadratic convex problem prior to doing (curved) line searches.

As with other semismooth Newton methods, the proposed method only solves one linear system of equations at each iteration.The sparsity of the Jacobian of the reformulated system can be exploited, often reducing the size of the system that must be solved. The reason for this is that the projection onto the feasible set increases the likelihood of components of iterates being active. The global and superlinear/quadratic convergence of the proposed method is proved under mild conditions. Numerical results are reported on all problems from the MCPLIB collection [8].

Key words. mixed complementarity problems – semismooth equations – projected Newton method – convergence

1. Introduction

In this paper we are concerned with finding a solution to the simply constrained system of nonlinear nonsmooth equations

$$H(x) = 0, \quad x \in X := \{x \in \mathbb{R}^n \mid l \le x \le u\}, \tag{1.1}$$

where the bounds $l_i \in \mathbb{R} \cup \{-\infty\}$, $u_i \in \mathbb{R} \cup \{+\infty\}$, $l_i < u_i$, $i = 1, \dots, n$, and the function $H : \mathbb{R}^n \to \mathbb{R}^n$ is assumed to have the following properties (with the definition of the semismoothness to follow):

D. Sun: Department of Mathematics, National University of Singapore, 2 Science Drive 2, Singapore 117543, Republic of Singapore. Fax: 0065-6779 5452. e-mail: matsundf@nus.edu.sg

R. S. Womersley: School of Mathematics, University of New South Wales, Sydney 2052, Australia.
e-mail: rsw@maths.unsw.edu.au

H. Qi: School of Mathematics, University of New South Wales, Sydney 2052, Australia.
e-mail: hdqi@maths.unsw.edu.au

⋆This work was supported in part by the Australian Research Council.

AMS subject classifications. 90C33, 90C30, 65H10

(H1) the function H is semismooth, and

(H2) the function $\theta(x) = \frac{1}{2}H^T(x)H(x)$ is continuously differentiable on $\mathbb{R}^n$.

An important application of (1.1) is to the mixed complementarity problem (MCP), which is to find a vector $x \in X$ such that

$$F(x)^T(y - x) \geq 0, \quad \forall\, y \in X. \tag{1.2}$$

Here, the function $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable. If $X = \mathbb{R}^n_+$, the MCP reduces to the nonlinear complementarity problem (NCP), which is to find $x \in \mathbb{R}^n$ such that

$$x \geq 0, \quad F(x) \geq 0, \quad x^T F(x) = 0. \tag{1.3}$$

See [16] and [11] for surveys on NCPs and MCPs. We will see in our numerical part Section 5 that through an NCP function, the MCP and NCP can be equivalently reformulated as (1.1) with H having the required properties. The proposed numerical method for (1.1) is then applied to solve the problems in the MCPLIB collection [8].

Problem (1.1) can be equivalently stated as a simply constrained minimization problem:

$$\begin{array}{ll} \min & \theta(x) \\ \text{s.t.} & x \in X. \end{array} \tag{1.4}$$

It is also noted that the complementarity property of NCP functions allows us to state that $x \in \mathbb{R}^n$ is a solution of (1.2) if and only if it satisfies $H(x) = 0$ (here constraints are not necessarily in place, in contrast to (1.1)). The reason that the box constraint set X is attached to (1.1) and (1.4), instead of dropping it, is based on several observations. Firstly, for some mixed complementarity problems in the MCPLIB collection [8], the function F or its Jacobian may not be well defined outside X, although for ease of discussion we assume that F is continuously differentiable on $\mathbb{R}^n$. Secondly, even if F is well defined on $\mathbb{R}^n$, some desirable properties of F like monotonicity which hold on X may not hold outside X. Thirdly, by forcing the iteration sequence to stay in X, one can avoid being trapped in a local minimizer of θ outside X. It is noted that adding bounds may increase the likelihood that the iteration sequence would be trapped in a constrained local minimizer. However, for NCPs and MCPs, conditions for assuring a stationary point of (1.4) to be a solution are not more restrictive than those for the unconstrained counterpart [13, 10].

Now we present a brief analysis which leads to what we call a feasible semismooth asymptotically Newton method for (1.1). At any $x \in \mathbb{R}^n$ such that $H(x) \neq 0$, let V be an element of the generalized Jacobian $\partial H(x)$ in the sense of Clarke [6]. Since θ is continuously differentiable, a solution d_N (if it exists) of the equation

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

is a descent direction of θ at x (note that $\nabla\theta(x) = V^T H(x)$ [7, 29]) and will be referred to as a (semismooth) Newton direction of θ at x. If the point at which the direction d_N is computed is obvious from the context, we will simply call d_N a (semismooth) Newton direction. This key feature enabled De Luca et al. to design a globally and locally superlinearly convergent (infeasible) semismooth Newton method [7]. The success of the semismooth Newton method [7] is heavily dependent on allowing the iteration sequence to stay outside X. The descent property of the Newton direction d_N further implies that if $x \in \text{int}X$, the interior part of X, and $\lambda > 0$ sufficiently small, one has

$$\theta(\Pi_X[x + \lambda d_N]) < \theta(x). \tag{1.6}$$

This has led several authors (e.g. [2, 3, 10, 17, 26]) to use (1.6) or a similar form as a rule to choose the next iterate. Methods based on (1.6) are referred to as projected (semismooth) Newton methods. However, (1.6) may not hold when $x \in X\backslash\text{int}X$, no matter how small λ is. In other words, the projected Newton direction $\bar{d}_N(\lambda)$ defined by

$$\bar{d}_N(\lambda) = \Pi_X[x + \lambda d_N] - x \tag{1.7}$$

may not be a descent direction for any small $\lambda > 0$. This, together with a couple of other features, is illustrated in Figure 1.

Example 1.1 Consider a two dimensional NCP where $X = \{x \in \mathbb{R}^2 | \ x \geq 0\}$ and $F_1(x) = -x_1 + x_2 - 1$ and $F_2(x) = 0$. Let the point considered be $x = (0, 0.5)$. The function H obtained with $\alpha = 1$ (cf. Section 5) is continuously differentiable around x and hence $V = H'(x)$. A particular instance of Newton directions at x is $d_N = (-2, -0.5)$. The problem has the following properties at x: (a) Newton direction d_N exists, (b) d_N is a descent direction for the unconstrained optimization problem (i.e., $\angle_2 > 90^0$ in Figure 1), and (c) the projected Newton direction is nonzero, but not a descent direction (i.e., $\angle_1 < 90^0$ in Figure 1).

Example 1.1 demonstrates that there is no theoretical guarantee of success for methods based on (1.6) despite the fact that it has been widely used in various projected Newton methods.

For any $x \in X$, define the gradient direction d_G by

$$d_G = -\gamma \nabla\theta(x)$$

for some $\gamma > 0$. Also for any $x \in X$ and $\lambda \in [0, 1]$ define the projected gradient direction $\bar{d}_G(\lambda)$ by

$$\bar{d}_G(\lambda) = \Pi_X[x + \lambda d_G] - x. \tag{1.8}$$

It is also well-known [1] that if $x \in X$ is not a stationary point of (1.4) then for any $\lambda > 0$ sufficiently small,

$$\theta(\Pi_X[x + \lambda d_G]) < \theta(x) \tag{1.9}$$

no matter whether $x \in \text{int}X$ or not (e.g., $\angle_3 > 90^0$ in Figure 1.) An immediate idea is to use (1.9) if (1.6) does not work (noticing again that the projected Newton direction is useless in Example 1.1 for all $\lambda > 0$.) However, it often happens that the projected Newton direction is not a descent direction for all $\lambda \in [\bar{\lambda}, 1]$ but is a descent direction for all $\lambda < \bar{\lambda}$ for some $\bar{\lambda} > 0$ (cf. Figure 2, $\angle = 90^0$). This illustrates a case that locally the projected Newton direction provides a descent direction and an Armijo-type search on the projected Newton direction may succeed. These arguments are illustrated in Figure 2.

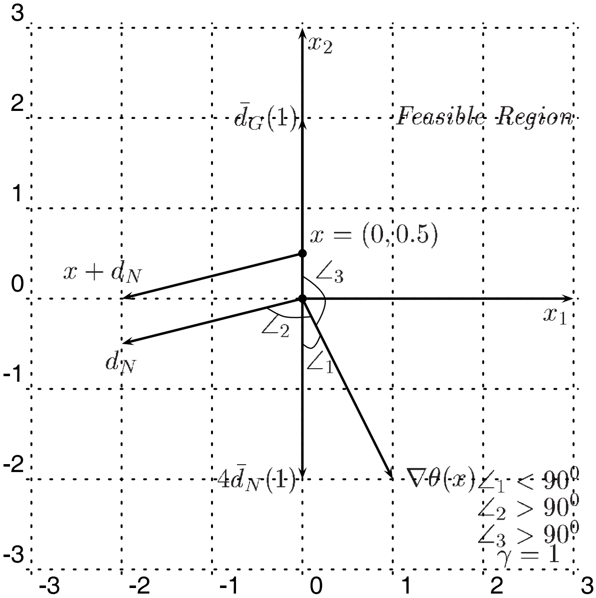

Fig. 1. Directions for Example 1.1 at $x = (0, 0.5)$ with $\alpha = 1$. All other projected Newton directions have the same orientation as of $\bar{d}_N(1)$

Example 1.2 Consider a two-dimensional MCP with $X = [0, \infty) \times [0, 1]$ and F given as in Example 1.1. Its unique solution is $(0, 1)$. Let the point considered be $x = (0, 0.5)$. The function H defined with $\alpha = 1$ (cf. Section 5) is continuously differentiable around x, and hence $V = H'(x)$. A particular instance of Newton directions at x is $d_N = (2, 1.5)$. Figure 2 illustrates the behavior of the projected Newton direction.

These examples leave us an intriguing question: how can one combine (1.9) with (1.6) into the design of algorithms so that both theoretical elegance and numerical excellence can be achieved simultaneously? In the circumstance of Example 1.2, $\bar{d}_N(\lambda)$ is a descent direction for $0 < \lambda < 0.5$, while in Example 1.1 $\bar{d}_N(\lambda)$ is not no matter how small λ is. Therefore, it is essential for good numerical performance to have a mechanism to judge if the projected Newton direction is suitable for a line search prior to trying it. To have such a mechanism or not is a major difference between our method and other feasible semismooth Newton methods [2, 10, 17, 18, 26].

In this paper, we shall address such a mechanism by introducing a projected (semismooth) asymptotically Newton method. For any $x \in X$ and $\lambda \in [0, 1]$, let

$$\bar{d}(\lambda) = t^*(\lambda)\bar{d}_G(\lambda) + [1 - t^*(\lambda)]\bar{d}_N(\lambda), \tag{1.10}$$

where for any fixed $\lambda \in [0, 1]$, $t^*(\lambda) \in [0, 1]$ is an optimal solution to the one-dimensional convex quadratic programming problem

$$\min_{t\in[0,1]} \frac{1}{2}\|H(x) + V[t\bar{d}_G(\lambda) + (1 - t)\bar{d}_N(\lambda)]\|^2. \tag{1.11}$$

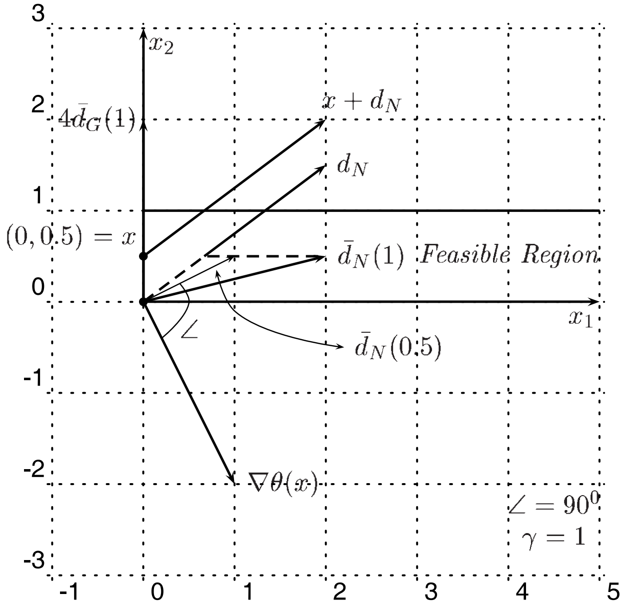

Fig. 2. Directions for Example 1.2. The dashed line is the trajectory of the projected Newton direction. $\bar{d}_N(0.5)$ forms an angle of 90^0 with $\nabla\theta(x)$, which indicates that the direction $\bar{d}_N(\lambda)$ with $\lambda \in (0, 0.5)$ is a descent direction, while $\bar{d}_N(\lambda)$ with $\lambda \in [0.5, 1]$ is not

We shall show that if $x \in X$ is not a stationary point of (1.4) and d_N exists,

$$\theta(x + \bar{d}(\lambda)) < \theta(x)$$

for all $\lambda > 0$ sufficiently small. More importantly, we will prove, under a nonsingularity assumption at a solution x^* of $H(x) = 0$, that

$$\lim_{x \to x^*} \sup_{0<\lambda\leq 1} \frac{\|\bar{d}(\lambda) - \lambda d_N\|}{\lambda\|d_N\|} = 0,$$

which means, locally, that $\bar{d}(\lambda)$ asymptotically converges to the Newton direction λd_N both in direction and magnitude for any $\lambda \in (0, 1]$. For this reason, we call $\bar{d}(\lambda)$, $\lambda \in [0, 1]$ a projected (semismooth) asymptotically Newton direction.

The organization of this paper is as follows. In the next section we review several results on semismoothness which will be used later on. In Section 3 we study properties of the search directions, which constitute the backbone of our algorithm. In Section 4 we state the algorithm and analyze its convergence. In Section 5 we first detail the reformulation of the MCP as (1.1) via an NCP function and then present numerical results on all problems in the MCPLIB collection [8]. Some concluding remarks are given in Section 6.

2. Properties on semismoothness

In this section we review some results on semismoothness which will be used later on. Semismoothness was originally introduced by Mifflin [19] for functionals. Convex functions, smooth functions, and piecewise smooth functions are examples of semismooth functions. The composition of semismooth functions is still a semismooth function (see [19]). In [24], Qi and Sun extended the definition of semismooth functions to $H : \mathbb{R}^n \to \mathbb{R}^n$. A locally Lipschitz continuous vector valued function $H : \mathbb{R}^n \to \mathbb{R}^n$ has a generalized Jacobian $\partial H(x)$ in the sense of Clarke [6]. H is said to be *semismooth* at $x \in \mathbb{R}^n$, if

$$\lim_{\substack{V \in \partial H(x+th') \\ h' \to h,\ t \downarrow 0}} \{Vh'\}$$

exists for any $h \in \mathbb{R}^n$. It has been proved in [24] that H is semismooth at x if and only if all its component functions are. Also $H'(x; h)$, the directional derivative of H at x in the direction h, exists for any $h \in \mathbb{R}^n$ and is equal to the above limit if H is semismooth at x.

Lemma 2.1 *[24, 22] Suppose that $H : \mathbb{R}^n \to \mathbb{R}^n$ is a locally Lipschitz function and semismooth at x. Then, for any $h \to 0$ and $V \in \partial H(x+h)$,*

$$H(x+h) - H(x) - Vh = o(\|h\|).$$

H is said to be *strongly semismooth* at x if H is semismooth at x and for any $V \in \partial H(x+h)$, $h \to 0$,

$$H(x+h) - H(x) - Vh = O(\|h\|^2).$$

A function H is said to be a (strongly) semismooth function if it is (strongly) semismooth everywhere on $\mathbb{R}^n$.

In [22], Qi defined the generalized Jacobian

$$\partial_B H(x) := \{V \in \mathbb{R}^{n\times n} |\ V = \lim_{x^k \to x} H'(x^k),\ H \text{ is differentiable at } x^k \text{ for all } k\}.$$

This concept will be used in the design of our algorithm. A locally Lipschitz function H is said to be BD-regular at $x \in \mathbb{R}^n$ if all $V \in \partial_B H(x)$ are nonsingular [22].

Lemma 2.2 *[22, Lemma 2.6] Suppose that $H : \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz continuous and H is BD-regular at $x \in \mathbb{R}^n$. Then there exist a neighborhood $\mathcal{N}(x)$ of x and a constant K such that for any $y \in \mathcal{N}(x)$ and $V \in \partial_B H(y)$, V is nonsingular and $\|V^{-1}\| \le K$.*

Lemma 2.3 *[21, Proposition 3] Suppose that $H : \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz continuous and H is BD-regular at a solution x^* of $H(x) = 0$. If H is semismooth at x^*, then there exist a neighborhood $\mathcal{N}(x^*)$ of x^* and a constant $\kappa > 0$ such that for any $x \in \mathcal{N}(x^*)$,*

$$\|H(x)\| \ge \kappa \|x - x^*\|.$$

The following two lemmas on properties of the projection operator $\Pi_X(\cdot)$ are useful in our analysis. Here the constraint set X can be any nonempty closed convex set.

Lemma 2.4 *[31] The projection operator $\Pi_X(\cdot)$ satisfies*

(i) for any $x \in X$, $[\Pi_X(z) - z]^T[\Pi_X(z) - x] \le 0$ for all $z \in \mathbb{R}^n$;
(ii) $\|\Pi_X(y) - \Pi_X(z)\| \le \|y - z\|$ for all $y, z \in \mathbb{R}^n$.

Lemma 2.5 *[14, 4] Given $x \in \mathbb{R}^n$ and $d \in \mathbb{R}^n$, the function ξ defined by*

$$\xi(\lambda) = \|\Pi_X(x + \lambda d) - x\|/\lambda, \quad \lambda > 0$$

is antitone (nonincreasing).

Lemma 2.5 actually implies that if $x \in X$ is a stationary point of (1.4), then

$$\bar{d}_G(\lambda) = \Pi_X[x + \lambda d_G] - x = 0 \quad \forall\, \lambda \geq 0.$$

3. Properties of search directions

In this section, we shall study some useful properties of $\bar{d}(\lambda)$, $\lambda \in [0, 1]$. We stress that the function H is always assumed to satisfy properties (H1) and (H2), which are stated at the beginning of Section 1.

Let $x \in X$ and $V \in \partial_B H(x)$ so that d_N exists. For $\lambda \in [0, 1]$, define $q_\lambda : \mathbb{R} \to \mathbb{R}_+$ by

$$q_\lambda(t) = \frac{1}{2}\|H(x) + V[t\bar{d}_G(\lambda) + (1-t)\bar{d}_N(\lambda)]\|^2, \quad t \in \mathbb{R}.$$

Then,

$$\begin{aligned} q_\lambda(t) = {} & \frac{1}{2}\|H(x) + V\bar{d}_N(\lambda)\|^2 + t[H(x) + V\bar{d}_N(\lambda)]^T V[\bar{d}_G(\lambda) - \bar{d}_N(\lambda)] \\ & + \frac{1}{2}t^2\|V[\bar{d}_G(\lambda) - \bar{d}_N(\lambda)]\|^2. \end{aligned}$$

For $\lambda \in [0, 1]$, let

$$t(\lambda) = \begin{cases} 0 & \text{if } V[\bar{d}_G(\lambda) - \bar{d}_N(\lambda)] = 0, \\ -\dfrac{[H(x) + V\bar{d}_N(\lambda)]^T V[\bar{d}_G(\lambda) - \bar{d}_N(\lambda)]}{\|V[\bar{d}_G(\lambda) - \bar{d}_N(\lambda)]\|^2} & \text{otherwise.} \end{cases} \tag{3.1}$$

Then $t(\lambda)$, $\lambda \in [0, 1]$ is a stationary point of $q_\lambda(\cdot)$, i.e., $\nabla q_\lambda(t(\lambda)) = 0$. For $\lambda \in [0, 1]$, define

$$t^*(\lambda) = \max\{0, \min\{1, t(\lambda)\}\}. \tag{3.2}$$

Lemma 3.1 *Suppose that $x \in X$ and $V \in \partial_B H(x)$ so that d_N exists. Then for any $\lambda \in [0, 1]$, $t^*(\lambda)$ is an optimal solution to (1.11), where $t^*(\lambda)$ is defined by (3.2).*

Proof. By the definition of (1.11) and $q_\lambda(\cdot)$ and the convexity of q_λ, $t^*(\lambda)$ is an optimal solution to

$$\min_{t \in [0,1]} q_\lambda(t).$$

□

The next proposition shows that $\bar{d}(\lambda)$ is a descent direction for all $\lambda > 0$ sufficiently small.

Theorem 3.1 *Suppose that $x \in X$ is not a stationary point of (1.4), $V \in \partial_B H(x)$ so that d_N exists and $\sigma \in (0, 1)$. Then there exists a constant $\lambda' \in (0, 1]$ such that for any $\lambda \in (0, \lambda']$, $\bar{d}(\lambda)$ is a descent direction of θ at x and*

$$\theta(x + \bar{d}(\lambda)) \leq \theta(x) + \sigma \nabla\theta(x)^T \bar{d}_G(\lambda). \tag{3.3}$$

Proof. Since θ is continuously differentiable from assumption (H2), for $\lambda \to 0$, we have

$$\theta(x + \bar{d}_G(\lambda)) = \theta(x) + \nabla\theta(x)^T \bar{d}_G(\lambda) + o(\lambda)$$

and

$$\theta(x + \bar{d}(\lambda)) = \theta(x) + \nabla\theta(x)^T \bar{d}(\lambda) + o(\lambda), \tag{3.4}$$

because from Lemma 2.4, $\|\bar{d}_G(\lambda)\| \leq \lambda\|d_G\|$, $\|\bar{d}_N(\lambda)\| \leq \lambda\|d_N\|$ and for some $t^*(\lambda) \in [0, 1]$,

$$\bar{d}(\lambda) = t^*(\lambda)\bar{d}_G(\lambda) + [1 - t^*(\lambda)]\bar{d}_N(\lambda).$$

On the other hand,

$$q_\lambda(1) = \frac{1}{2}\|H(x) + V\bar{d}_G(\lambda)\|^2 = \theta(x) + \nabla\theta(x)^T \bar{d}_G(\lambda) + O(\lambda^2) \tag{3.5}$$

and

$$q_\lambda(t^*(\lambda)) = \frac{1}{2}\|H(x) + V\bar{d}(\lambda)\|^2 = \theta(x) + \nabla\theta(x)^T \bar{d}(\lambda) + O(\lambda^2). \tag{3.6}$$

By using (3.5), (3.6) and Lemma 3.1, for any $\lambda \in [0, 1]$ we have

$$\nabla\theta(x)^T \bar{d}(\lambda) \leq \nabla\theta(x)^T \bar{d}_G(\lambda) + O(\lambda^2). \tag{3.7}$$

By Lemmas 2.4 and 2.5, for any $\lambda \in (0, 1]$,

$$\nabla\theta(x)^T \bar{d}_G(\lambda) \leq -\|\bar{d}_G(\lambda)\|^2/(\lambda\gamma) \leq -\lambda\|\bar{d}_G(1)\|^2/\gamma. \tag{3.8}$$

Hence, from (3.7) and (3.8), for all $\lambda > 0$ sufficiently small we have $\nabla\theta(x)^T \bar{d}(\lambda) < 0$. By (3.4) and (3.7) it holds that

$$\theta(x + \bar{d}(\lambda)) \leq \theta(x) + \nabla\theta(x)^T \bar{d}_G(\lambda) + o(\lambda),$$

which, together with (3.8), implies that there exists a constant $\lambda' \in (0, 1]$ such that for any $\lambda \in (0, \lambda']$, (3.3) holds. □

We now show that if for some $\lambda \in (0, 1]$, $x + \lambda d_N \in X$ and

$$(1 - \lambda)[2\lambda\theta(x) + \nabla\theta(x)^T \bar{d}_G(\lambda)] \geq 0,$$

then $\bar{d}(\lambda)$ always takes the Newton direction, i.e., $\bar{d}(\lambda) = \lambda d_N$. In particular, if $x + d_N \in X$, then it holds that $\bar{d}(1) = d_N$.

Proposition 3.1 *Suppose that $x \in X$ is not a stationary point of (1.4) and $V \in \partial_B H(x)$ so that d_N exists. If for some $\lambda \in (0, 1]$, $x + \lambda d_N \in X$, then*

$$\bar{d}(\lambda) = \begin{cases} d_N & \text{if } \lambda = 1, \\ \lambda d_N & \text{if } \lambda \in (0, 1) \;\&\; 0 < \gamma \leq 2\theta(x)/\|\nabla\theta(x)\|^2. \end{cases}$$

Proof. From the assumption $x + \lambda d_N \in X$, we have $\bar{d}_N(\lambda) = \lambda d_N$. This, together with (3.1), implies that

$$t(\lambda) = \begin{cases} 0 & \text{if } V[\bar{d}_G(\lambda) - \lambda d_N] = 0, \\ -\dfrac{(1-\lambda)[2\lambda\theta(x) + \nabla\theta(x)^T \bar{d}_G(\lambda)]}{\|V[\bar{d}_G(\lambda) - \lambda d_N]\|^2} & \text{otherwise.} \end{cases} \tag{3.9}$$

First, let us consider the case that $\lambda = 1$. Then from (3.2), we know that $t^*(1) = 0$, which implies that $\bar{d}(1) = d_N$.

Next, we discuss the situation for $\lambda \in (0, 1)$. For $0 < \gamma \le 2\theta(x)/\|\nabla\theta(x)\|^2$, we have

$$2\lambda\theta(x) + \nabla\theta(x)^T \bar{d}_G(\lambda) \ge 2\lambda\theta(x) - \lambda\gamma\|\nabla\theta(x)\|^2 \ge 0,$$

which, together with (3.9), implies that $t^*(\lambda) = 0$. This proves that $\bar{d}(\lambda) = \lambda d_N$. □

So far, we have considered some global properties of the search direction $\bar{d}(\cdot)$ for a nonstationary point $x \in X$. For the sake of superlinear (quadratic) convergence of our algorithm, we will next consider properties of the directions $\bar{d}_N(\cdot)$ and $\bar{d}(\cdot)$ around a solution point.

First, let us consider the direction $\bar{d}_N(\cdot)$. It has been shown in Section 1 that in general $\bar{d}_N(\lambda)$ is not a descent direction of θ for any $\lambda \in (0, 1]$. However, the next result shows that if x is sufficiently close to a BD-regular solution of $H(x) = 0$, $\bar{d}_N(\lambda)$ is indeed a descent direction of θ for all $\lambda \in (0, 1]$.

Proposition 3.2 *Suppose that H is BD-regular at a solution x^* of $H(x) = 0$. Then for any $\rho \in (0, 2)$, there exists a neighborhood $\mathcal{N}$ of x^* such that for any $\lambda \in (0, 1]$ and $x \in \mathcal{N} \cap X$, $\bar{d}_N(\lambda)$ is a descent direction of θ at x with*

$$\nabla\theta(x)^T \bar{d}_N(\lambda) \le -\rho\lambda\theta(x) \tag{3.10}$$

and

$$\bar{d}_N(\lambda) = -\lambda(x - x^*) + \lambda o(\theta(x)^{\frac{1}{2}}). \tag{3.11}$$

Furthermore, if H is strongly semismooth at x^*, then for any $\lambda \in (0, 1]$,

$$\bar{d}_N(\lambda) = -\lambda(x - x^*) + \lambda O(\theta(x)). \tag{3.12}$$

Proof. First, since H is BD-regular at a solution x^*, by Lemmas 2.2 and 2.3 there exist a neighborhood $\mathcal{N}$ of x^* and two positive numbers K and κ such that for any $x \in \mathcal{N}$ and $V \in \partial_B H(x)$,

$$\|V^{-1}\| \le K, \quad \text{and} \quad \|H(x)\| \ge \kappa\|x - x^*\|. \tag{3.13}$$

Next, for any $x \in \mathcal{N} \cap X$, define

$$R(x) := H(x) - H(x^*) - V(x - x^*).$$

By noting that $H(x^*) = 0$, we have for all $x \in \mathcal{N} \cap X$ that

$$\begin{aligned} x + \lambda d_N &= x - \lambda V^{-1} H(x) = x - \lambda V^{-1}[V(x - x^*) + R(x)] \\ &= (1-\lambda)x + \lambda x^* - \lambda V^{-1} R(x), \end{aligned}$$

which implies that

$$\begin{aligned}\bar{d}_N(\lambda) &= \Pi_X[x + \lambda d_N] - x \\ &= \Pi_X[(1-\lambda)x + \lambda x^*] - x \\ &\quad + \Pi_X[(1-\lambda)x + \lambda x^* - \lambda V^{-1} R(x)] - \Pi_X[(1-\lambda)x + \lambda x^*].\end{aligned}$$

Hence, since $(1-\lambda)x + \lambda x^* \in X$, we have

$$\bar{d}_N(\lambda) = (1-\lambda)x + \lambda x^* - x + \lambda \Delta_\lambda(x) = -\lambda(x - x^*) + \lambda \Delta_\lambda(x), \tag{3.14}$$

where

$$\Delta_\lambda(x) = \{\Pi_X[(1-\lambda)x + \lambda x^* - \lambda V^{-1} R(x)] - \Pi_X[(1-\lambda)x + \lambda x^*]\}/\lambda.$$

Therefore,

$$\begin{aligned}\nabla\theta(x)^T \bar{d}_N(\lambda) &= -\lambda \nabla\theta(x)^T (x - x^*) + \lambda \nabla\theta(x)^T \Delta_\lambda(x) \\ &= -\lambda H(x)^T V(x - x^*) + \lambda H(x)^T V \Delta_\lambda(x) \\ &= -2\lambda\theta(x) + \lambda H(x)^T R(x) + \lambda H(x)^T V \Delta_\lambda(x). \end{aligned}\tag{3.15}$$

By using (ii) of Lemma 2.4 we have

$$\|\Pi_X[(1-\lambda)x + \lambda x^* - \lambda V^{-1} R(x)] - \Pi_X[(1-\lambda)x + \lambda x^*]\| \le \lambda \|V^{-1} R(x)\|,$$

which implies that

$$\|\Delta_\lambda(x)\| \le \|V^{-1} R(x)\| \le K\|R(x)\|. \tag{3.16}$$

Hence, by (3.15), for all $x \in \mathcal{N} \cap X$,

$$\nabla\theta(x)^T \bar{d}_N(\lambda) \le -2\lambda\theta(x) + \sqrt{2}\lambda(1 + K\|V\|)\theta(x)^{\frac{1}{2}} \|R(x)\|. \tag{3.17}$$

Since H is semismooth at x^*, it follows from Lemma 2.1 that for any $V \in \partial_B H(x)$ and $x \to x^*$,

$$R(x) = o(\|x - x^*\|). \tag{3.18}$$

Then, because $\partial_B H(\cdot)$ is compact everywhere and upper semi-continuous [22], by shrinking $\mathcal{N}$ if necessary, we have for any $x \in \mathcal{N} \cap X$,

$$\|R(x)\| \le (2-\rho)\frac{\kappa\|x - x^*\|}{2(1 + K\|V\|)},$$

which, together with (3.13) and (3.17), implies that

$$\nabla\theta(x)^T \bar{d}_N(\lambda) \le -\rho\lambda\theta(x).$$

This proves (3.10).

Putting together (3.13), (3.14), (3.18) and (3.16), we have actually proved (3.11).

Now, we prove (3.12). Since H is strongly semismooth at x^*,

$$R(x) = O(\|x - x^*\|^2),$$

which, together with (3.13), (3.14) and (3.16), proves (3.12). □

The next lemma summarizes several results needed in proving the local properties of $\bar{d}(\cdot)$. Its proof can be obtained by using Lemmas 2.1–2.3 and the fact that $\partial_B H(\cdot)$ is compact on any compact set [22].

Lemma 3.2 *Suppose that H is BD-regular at a solution x^* of $H(x) = 0$. Then, there exists a neighborhood $\mathcal{N}$ of x^* such that for any $x \in \mathcal{N} \cap X$ and $V \in \partial_B H(x)$,*

$$H(x) - V(x - x^*) = o(\theta(x)^{\frac{1}{2}}),$$

and

$$\nabla\theta(x) = V^T H(x) = O(\theta(x)^{\frac{1}{2}}).$$

Moreover, if H is strongly semismooth at x^* then

$$H(x) - V(x - x^*) = O(\theta(x)).$$

Finally, we can characterize the properties of $\bar{d}(\cdot)$ locally.

Theorem 3.2 *Suppose that H is BD-regular at a solution x^* of $H(x) = 0$. Let η be a positive number in $(0, 1)$. Then for any $\lambda \in (0, 1]$,*

$$0 < \gamma \le \min\{1, \eta\theta(x)/\|\nabla\theta(x)\|^2\}$$

and $x \in X$ with $x \to x^*$, we have

$$\nabla\theta(x)^T \bar{d}(\lambda) = -2\lambda\theta(x) + \lambda o(\theta(x)) \tag{3.19}$$

and

$$\sup_{\lambda \in (0,1]} \frac{\|\bar{d}(\lambda) - \lambda d_N\|}{\lambda\|d_N\|} = o(1). \tag{3.20}$$

Moreover, if H is strongly semismooth at x^*, then for any $x \in X$ with $x \to x^*$, we have

$$\sup_{\lambda \in (0,1]} \frac{\|\bar{d}(\lambda) - \lambda d_N\|}{\lambda\|d_N\|} = O(\theta(x)^{\frac{1}{2}}). \tag{3.21}$$

Proof. By (3.11) in Proposition 3.2, for any $\lambda \in [0, 1]$ and $V \in \partial_B H(x)$,

$$V\bar{d}_N(\lambda) = -\lambda V(x - x^*) + \lambda o(\theta(x)^{\frac{1}{2}}),$$

which, together with Lemma 3.2, implies that

$$V\bar{d}_N(\lambda) = -\lambda H(x) + \lambda o(\theta(x)^{\frac{1}{2}}).$$

Hence, from the fact that for any $\lambda \in [0, 1]$,

$$\|\bar{d}_G(\lambda)\| \le \lambda\gamma\|\nabla\theta(x)\| \le \lambda\gamma\|V\|\|H(x)\|,$$

we have

$$\begin{aligned}
&[H(x) + V\bar{d}_N(\lambda)]^T V\bar{d}_N(\lambda) \\
&= [(1 - \lambda)H(x) + \lambda o(\theta(x)^{\frac{1}{2}})]^T[-\lambda H(x) + \lambda o(\theta(x)^{\frac{1}{2}})] \\
&= -2\lambda(1 - \lambda)\theta(x) + \lambda(1 - \lambda)o(\theta(x)) + \lambda^2 o(\theta(x)),
\end{aligned} \tag{3.22}$$

$$
\begin{aligned}
&-[H(x) + V\bar{d}_N(\lambda)]^T V\bar{d}_G(\lambda) \\
&= -[(1-\lambda)V^T H(x) + \lambda o(\theta(x)^{\frac{1}{2}})]^T \{\Pi_X[x - \lambda\gamma\nabla\theta(x)] - x\} \\
&\leq \|(1-\lambda)\nabla\theta(x) + \lambda o(\theta(x)^{\frac{1}{2}})\| \|\Pi_X[x - \lambda\gamma\nabla\theta(x)] - x\| \\
&\leq \gamma\lambda(1-\lambda)\|\nabla\theta(x)\|^2 + \lambda^2 o(\theta(x))
\end{aligned} \tag{3.23}
$$

and

$$
\begin{aligned}
&\|V[\bar{d}_N(\lambda) - \bar{d}_G(\lambda)]\|^2 \\
&= \|V\bar{d}_N(\lambda)\|^2 - 2[V\bar{d}_N(\lambda)]^T V\bar{d}_G(\lambda) + \|V\bar{d}_G(\lambda)\|^2 \\
&\geq \|V\bar{d}_N(\lambda)\|^2 - 2[V\bar{d}_N(\lambda)]^T V\bar{d}_G(\lambda) \\
&= 2\lambda^2\theta(x) + \lambda^2 o(\theta(x)) - 2[-\lambda H(x) + \lambda o(\theta(x)^{\frac{1}{2}})]^T V\bar{d}_G(\lambda) \\
&= 2\lambda^2\theta(x) + 2\lambda\nabla\theta(x)^T \bar{d}_G(\lambda) + \lambda^2 o(\theta(x)) \\
&\geq 2\lambda^2[\theta(x) - \gamma\|\nabla\theta(x)\|^2] + \lambda^2 o(\theta(x)).
\end{aligned} \tag{3.24}
$$

Now suppose that $0 < \gamma \leq \min\{1, \eta\theta(x)/\|\nabla\theta(x)\|^2\}$. Then, by using (3.22)–(3.24), for any $\lambda \in [0, 1]$ we have

$$
\begin{aligned}
Q(\lambda) &:= [H(x) + V\bar{d}_N(\lambda)]^T V[\bar{d}_N(\lambda) - \bar{d}_G(\lambda)] \\
&\leq -\lambda(1-\lambda)(2-\eta)\theta(x) + \lambda(1-\lambda)o(\theta(x)) + \lambda^2 o(\theta(x))
\end{aligned} \tag{3.25}
$$

and

$$
\|V[\bar{d}_N(\lambda) - \bar{d}_G(\lambda)]\|^2 \geq 2\lambda^2(1-\eta)\theta(x) + \lambda^2 o(\theta(x)). \tag{3.26}
$$

Hence, we can conclude from (3.25)-(3.26) and (3.1) that $t(\lambda) \leq 0$ if $Q(\lambda) \leq 0$ and

$$
t(\lambda) \leq \frac{|-\lambda(1-\lambda)(2-\eta) + \lambda(1-\lambda)o(1) + \lambda^2 o(1)|}{2\lambda^2(1-\eta) + \lambda^2 o(1)}
$$

if $Q(\lambda) > 0$. Therefore, from (3.2), we obtain

$$
t^*(\lambda) \leq o(1).
$$

Hence,

$$
\bar{d}(\lambda) = t^*(\lambda)\bar{d}_G(\lambda) + [1 - t^*(\lambda)]\bar{d}_N(\lambda) = \bar{d}_N(\lambda) + \lambda o(\theta(x)^{\frac{1}{2}}).
$$

Then, from Proposition 3.2, it holds that

$$
\bar{d}(\lambda) = -\lambda(x - x^*) + \lambda o(\theta(x)^{\frac{1}{2}}).
$$

This proves both (3.19) and (3.20) by using Lemma 3.2 and $d_N = -V^{-1}H(x)$.

Moreover, if H is strongly semismooth at x^*, then from Lemma 3.2, for $x \to x^*$ and $V \in \partial_B H(x)$,

$$
H(x) - V(x - x^*) = O(\theta(x)).
$$

Hence, by using this, Proposition 3.2, and the above argument, we get (3.21). We omit the details here. □

4. The algorithm and its convergence analysis

We first describe our algorithm and then discuss its convergence analysis. The algorithm parameters ρ, σ control an Armijo line search, the parameter η the scaling of the steepest descent direction, and p_1, p_2 when the Newton direction is acceptable.

Algorithm 4.1 *(A Projected Semismooth Asymptotically Newton Method)*

Step 0. *Choose constants* $\rho, \sigma, \eta \in (0, 1)$*,* $p_1 > 0$ *and* $p_2 > 2$*. Let* $x^0 \in X$ *and* $k := 0$*.*
Step 1. *Choose* $V_k \in \partial_B H(x^k)$ *and compute* $\nabla\theta(x^k) = V_k^T H(x^k)$*.*
Step 2. *If* x^k *is a stationary point, stop. Otherwise let*

$$d_G^k = -\gamma_k \nabla\theta(x^k),$$

where

$$\gamma_k = \min\{1, \eta\theta(x^k)/\|\nabla\theta(x^k)\|^2\},$$

and go to Step 3.
Step 3. *If the linear system*

$$H(x^k) + V_k d = 0, \tag{4.1}$$

has a solution d_N^k *and*

$$-\nabla\theta(x^k)^T d_N^k \geq p_1 \|d_N^k\|^{p_2}, \tag{4.2}$$

then use the direction d_N^k*. Otherwise, set*

$$d_N^k = d_G^k.$$

Step 4. *Let* m_k *be the smallest nonnegative integer* m *satisfying*

$$\theta(x^k + \bar{d}^k(\rho^m)) \leq \theta(x^k) + \sigma\nabla\theta(x^k)^T \bar{d}_G^k(\rho^m), \tag{4.3}$$

where for any $\lambda \in [0, 1]$*,*

$$\bar{d}^k(\lambda) = t_k^*(\lambda)\bar{d}_G^k(\lambda) + [1 - t_k^*(\lambda)]\bar{d}_N^k(\lambda),$$

$$\bar{d}_G^k(\lambda) = \Pi_X[x + \lambda d_G^k] - x^k, \quad \bar{d}_N^k(\lambda) = \Pi_X[x + \lambda d_N^k] - x^k,$$

and $t_k^*(\lambda)$ *is an optimal solution to*

$$\min_{t\in[0,1]} \frac{1}{2}\|H(x^k) + V_k[t\bar{d}_G^k(\lambda) + (1-t)\bar{d}_N^k(\lambda)]\|^2$$

and is computed by (3.2). Let $\lambda_k = \rho^{m_k}$ *and* $x^{k+1} = x^k + \bar{d}^k(\lambda_k)$*.*
Step 5. *Replace* k *by* $k + 1$ *and go to Step 1.*

Several comments on Algorithm 4.1 are in order. (a) The above algorithm is a feasible one because for any $\lambda \in [0, 1]$ and $x^k \in X$,

$$x^k + \bar{d}^k(\lambda) = t_k^*(\lambda)\Pi_X[x^k + \lambda d_G^k] + [1 - t_k^*(\lambda)]\Pi_X[x^k + \lambda d_N^k] \in X.$$

Hence, the whole iteration sequence generated by Algorithm 4.1 stays in the feasible region X. (b) In Step 3 of Algorithm 4.1, we need to solve the linear system (4.1) to get d_N^k. If (4.1) is unsolvable or if the matrix V_k is highly ill-conditioned, then d_N^k just takes the negative gradient direction, which then implies that $\bar{d}^k(\lambda) = \bar{d}_G^k(\lambda)$ for any $\lambda \in (0, 1]$. There is little extra cost to compute $\bar{d}^k(\cdot)$ once d_N^k is computed. It is also noted that in the application to the MCP, V_k not only keeps any sparse structure of $F'(x^k)$, but also has a special structure which could be exploited to reduce the cost of solving (4.1). For example, for the NCP, for any $1 \le i \le n$, $x_i^k = 0$ and $F_i(x^k) > 0$, we have $(V_k)_{ij} = 0$ for all $j \neq i$. This phenomenon was observed more often than in infeasible methods that allow x^k to stay outside X. (c) By Theorem 3.1, if x^k is not a stationary point of (1.4), then Step 4 of Algorithm 4.1 is well defined. Hence, our algorithm either stops at a stationary point or generates an infinite feasible sequence $\{x^k\} \in X$.

In our convergence analysis, we assume that our algorithm does not stop at a stationary point at any finite step.

Theorem 4.1 *Let $\{x^k\} \subset X$ be a sequence generated by Algorithm 4.1. Then any accumulation point of $\{x^k\}$ is a stationary point of (1.4).*

Proof. Let $\bar{x} \in X$ be an accumulation point of $\{x^k\}$. Suppose that $\bar{x}$ is not a stationary point of (1.4). By taking a subsequence if necessary, we assume that $\{x^k\} \to \bar{x}$. From the upper semicontinuity of the generalized Jacobian [6], there exists a constant $\kappa_2 > 0$ such that $\|V_k\| \le \kappa_2$ for all $k \ge 0$. It is also easy to see from Step 3 of Algorithm 4.1 that

$$\|d_N^k\| \le \max\left\{\gamma_k\|\nabla\theta(x^k)\|, (p_1^{-1}\|\nabla\theta(x^k)\|)^{\frac{1}{p_2-1}}\right\}.$$

It then follows from the continuity of $\nabla\theta(\cdot)$ that there exists a number $\kappa_3 > 0$ such that $\max\{\|d_G^k\|,\ \|d_N^k\|\} \le \kappa_3$ for all $k \ge 0$. By Lemma 2.4, for any $\lambda \in [0, 1]$ and $k \ge 0$ we have

$$\|\bar{d}_G^k(\lambda)\| \le \lambda\|d_G^k\| \le \lambda\kappa_3, \quad \|\bar{d}_N^k(\lambda)\| \le \lambda\|d_N^k\| \le \lambda\kappa_3,$$

and for some $t_k^*(\lambda)$ defined by (3.2),

$$\|\bar{d}^k(\lambda)\| = \|t_k^*(\lambda)\bar{d}_G^k(\lambda) + (1 - t_k^*(\lambda))\bar{d}_N^k(\lambda)\| \le \lambda\kappa_3.$$

Define

$$q_\lambda^k(t) := \frac{1}{2}\|H(x^k) + V_k[t\bar{d}_G^k(\lambda) + (1 - t)\bar{d}_N^k(\lambda)]\|^2, \quad t \in [0, 1].$$

Then we have

$$q_\lambda^k(1) = \frac{1}{2}\|H(x^k) + V_k\bar{d}_G^k(\lambda)\|^2 = \theta(x^k) + \nabla\theta(x^k)^T\bar{d}_G^k(\lambda) + \frac{1}{2}\|V_k\bar{d}_G^k(\lambda)\|^2$$

and

$$q_\lambda^k(t^*(\lambda)) = \frac{1}{2}\|H(x^k) + V_k\bar{d}^k(\lambda)\|^2 = \theta(x^k) + \nabla\theta(x^k)^T\bar{d}^k(\lambda) + \frac{1}{2}\|V_k\bar{d}^k(\lambda)\|^2.$$

Since $q_\lambda^k(t_k^*(\lambda)) \le q_\lambda^k(1)$ for any $\lambda \in [0, 1]$, we get

$$\begin{aligned}\nabla\theta(x^k)^T\bar{d}^k(\lambda) &\le \nabla\theta(x^k)^T\bar{d}_G^k(\lambda) + \frac{1}{2}\|V_k\bar{d}_G^k(\lambda)\|^2\\ &\le \nabla\theta(x^k)^T\bar{d}_G^k(\lambda) + \frac{1}{2}(\kappa_2\kappa_3)^2\lambda^2\\ &= \nabla\theta(x^k)^T\bar{d}_G^k(\lambda) + \kappa_1\lambda^2,\end{aligned} \tag{4.4}$$

where $\kappa_1 = \frac{1}{2}(\kappa_2\kappa_3)^2$.

For $\lambda \in [0, 1]$ and $k \ge 0$ we have

$$\begin{aligned}&\theta(x^k + \bar{d}^k(\lambda)) - \theta(x^k)\\ &\quad= \nabla\theta(x^k)^T\bar{d}^k(\lambda) + \int_0^1 [\nabla\theta(x^k + t\bar{d}^k(\lambda)) - \nabla\theta(x^k)]^T\bar{d}^k(\lambda)dt\\ &\quad\le \nabla\theta(x^k)^T\bar{d}^k(\lambda) + \left(\int_0^1 \|\nabla\theta(x^k + t\bar{d}^k(\lambda)) - \nabla\theta(x^k)\|dt\right)\|\bar{d}^k(\lambda)\|\\ &\quad\le \nabla\theta(x^k)^T\bar{d}^k(\lambda) + \lambda\kappa_3\int_0^1 \|\nabla\theta(x^k + t\bar{d}^k(\lambda)) - \nabla\theta(x^k)\|dt.\end{aligned} \tag{4.5}$$

Since both $\{x^k\}$ and $\{x^k + t\bar{d}^k(\lambda)\}$ with $t, \lambda \in [0, 1]$ are bounded and $\nabla\theta(\cdot)$ is uniformly continuous on any compact set, for any given $\varepsilon > 0$ there exists a number $\bar{\lambda} > 0$ (depending on ε) such that, for all $k \ge 0$ and $\lambda \in [0, \bar{\lambda}]$, it holds that

$$\int_0^1 \|\nabla\theta(x^k + t\bar{d}^k(\lambda)) - \nabla\theta(x^k)\|dt \le \varepsilon.$$

Hence, it follows from (4.5) that

$$\theta(x^k+\bar{d}^k(\lambda)) \le \theta(x^k)+\nabla\theta(x^k)^T\bar{d}^k(\lambda)+\lambda\varepsilon\kappa_3 \quad \text{for all } k \ge 0 \text{ and } \lambda \in [0, \bar{\lambda}]. \tag{4.6}$$

By Lemmas 2.4 and 2.5, for any x^k and any $\lambda \in (0, 1]$,

$$\nabla\theta(x^k)\bar{d}_G^k(\lambda) \le -\|\bar{d}_G^k(\lambda)\|^2/(\lambda\gamma_k) \le -\lambda\|\bar{d}_G^k(1)\|^2/\gamma_k. \tag{4.7}$$

Since $\bar{x}$ is not a stationary point of (1.4), there exists a number $\kappa_4 > 0$ such that

$$\|\bar{d}_G^k(1)\| = \|\Pi_X[x^k - \gamma_k\nabla\theta(x^k)] - x^k\| \ge \kappa_4.$$

Let

$$\varepsilon = \frac{1-\sigma}{2\kappa_3}\kappa_4^2, \quad \tilde{\lambda} = \frac{1-\sigma}{2\kappa_1}\kappa_4^2 \quad \text{and} \quad \lambda' = \min\{\bar{\lambda}, \tilde{\lambda}\}.$$

Then relations (4.4), (4.6), and (4.7) and the fact that $\gamma_k \leq 1$ imply that for all $k \geq 0$ and all $\lambda \in (0, \lambda']$ we have

$$\begin{aligned}\theta(x^k + \bar{d}^k(\lambda)) &\leq \theta(x^k) + \nabla\theta(x^k)^T \bar{d}_G^k(\lambda) + \kappa_1\lambda^2 + \lambda\varepsilon\kappa_3 \\ &\leq \theta(x^k) + \sigma\nabla\theta(x^k)^T \bar{d}_G^k(\lambda) - (1-\sigma)\lambda\|\bar{d}_G^k(1)\|^2/\gamma_k + \kappa_1\lambda^2 + \lambda\varepsilon\kappa_3 \\ &\leq \theta(x^k) + \sigma\nabla\theta(x^k)^T \bar{d}_G^k(\lambda). \end{aligned} \tag{4.8}$$

Hence, from (4.8), we know from the line search rule in Step 4 of Algorithm 4.1 that for all $k \geq 0$, $\lambda_k \geq \rho\lambda'$. From inequalities (4.3) and (4.7), we obtain that for $k \to \infty$,

$$\nabla\theta(x^k)\bar{d}_G^k(\lambda_k) \to 0 \ \text{ and } \lambda_k\|\bar{d}_G^k(1)\|^2/\gamma_k \to 0.$$

This is a contradiction because $\liminf_{k\to\infty} \lambda_k \geq \rho\lambda'$, γ_k is bounded, and $\|\bar{d}_G^k(1)\| \geq \kappa_4$. This contradiction shows that $\bar{x}$ is a stationary point of (1.4) and completes the proof. □

Next, we shall prove that Algorithm 4.1 converges superlinearly (quadratically) under the BD-regularity.

Theorem 4.2 *Suppose that $\{x^k\}$ is a sequence generated by Algorithm 4.1 and x^*, an accumulation point of $\{x^k\}$, is a solution of $H(x) = 0$. If H is BD-regular at x^*, then the whole sequence $\{x^k\}$ converges to x^* Q-superlinearly. Furthermore, if H is strongly semismooth at x^*, then the convergence rate is Q-quadratic.*

Proof. From Lemma 2.2, Theorem 3.2, and the choice of γ_k in Step 2 of Algorithm 4.1, for all x^k sufficiently close to x^*, (4.2) is satisfied, i.e., $d_N^k = -V_k^{-1}H(x^k)$ and

$$\begin{aligned}\|x^k + \bar{d}^k(1) - x^*\| &= \|x^k + d_N^k - x^* + o(\|d_N^k\|)\| \leq \|x^k + d_N^k - x^*\| + o(\|H(x^k)\|) \\ &\leq \|V_k^{-1}\|\|H(x^k) - H(x^*) - V_k(x^k - x^*)\| + o(\|H(x^k)\|),\end{aligned} \tag{4.9}$$

which, together with Lemmas 2.1-2.3, implies that for all x^k sufficiently close to x^*,

$$\|x^k + \bar{d}^k(1) - x^*\| = o(\|x^k - x^*\|) = o(\|H(x^k)\|). \tag{4.10}$$

Hence, from (4.10), for all x^k sufficiently close to x^*,

$$\begin{aligned}\theta(x^k + \bar{d}^k(1)) &= \frac{1}{2}\|H(x^k + \bar{d}^k(1))\|^2 = \frac{1}{2}\|H(x^k + \bar{d}^k(1)) - H(x^*)\|^2 \\ &= O(\|x^k + \bar{d}^k(1) - x^*\|^2) = o(\theta(x^k)).\end{aligned} \tag{4.11}$$

On the other hand, for all $k \geq 0$,

$$-\nabla\theta(x^k)^T \bar{d}_G^k(1) \leq \|\nabla\theta(x^k)\|\|\bar{d}_G^k(1)\| \leq \gamma_k\|\nabla\theta(x^k)\|^2 \leq \eta\theta(x^k). \tag{4.12}$$

Hence, we can conclude from relations (4.11) and (4.12) that for all x^k sufficiently close to x^*,

$$\theta(x^k + \bar{d}^k(1)) \leq \theta(x^k) + \sigma\nabla\theta(x^k)^T \bar{d}_G^k(1),$$

which further implies that,

$$x^{k+1} = x^k + \bar{d}^k(1).$$

Then from (4.10) we have proved that $\{x^k\}$ converges to x^* Q-superlinearly.

Finally, if H is strongly semismooth at x^*, we can easily modify the above arguments to get the Q-quadratic convergence of $\{x^k\}$ by invoking Theorem 3.2. □

5. Numerical experiments

In this section, we first outline a semismooth equation reformulation of the MCP based on a variant of a function discussed by Sun and Womersley [28]. We then report our numerical results on all problems in the MCPLIB collection [8].

The function $\psi_\alpha : \mathbb{R}^2 \to \mathbb{R}$ with $\alpha \in [0, 1]$ being prescribed in our reformulation is defined by

$$\psi_\alpha(a, b) := ([\phi_\alpha(a, b)]_+)^2 + ([-a]_+)^2,$$

where $[a]_+ := \max\{0, a\}$ for any $a \in \mathbb{R}$ and $\phi_\alpha : \mathbb{R}^2 \to \mathbb{R}$ is the penalized Fischer-Burmeister function introduced by Chen et al. [5] and has the form:

$$\phi_\alpha(a, b) := \alpha\phi_{\mathrm{FB}}(a, b) + (1 - \alpha)a_+b_+.$$

Here, $\phi_{\mathrm{FB}} : \mathbb{R}^2 \to \mathbb{R}$ is an NCP function, which is known as the Fischer-Burmeister function [12] and given by

$$\phi_{\mathrm{FB}}(a, b) := (a + b) - \sqrt{a^2 + b^2}.$$

Numerical tests indicate that the penalized Fischer-Burmeister function usually leads to better numerical performance than the Fischer-Burmeister function [5, 27]. For more discussions on advantages and/or disadvantages of NCP functions, see [23, 27, 28, 30] and references therein.

Let $N = \{1, \dots, n\}$ and

$$I_f := \{i \mid l_i = -\infty,\ u_i = \infty,\ i \in N\}, \quad I_l := \{i \mid l_i > -\infty,\ u_i = \infty,\ i \in N\},$$
$$I_u := \{i \mid l_i = -\infty,\ u_i < \infty,\ i \in N\}, \quad I_{lu} := N \backslash \{I_l \cup I_u \cup I_f\}.$$

Sun and Womersley [28] suggested reformulating the MCP as $H(x) = 0$ with

$$H_i(x) := \begin{cases} |F_i(x)| & \text{if } i \in I_f \\ |\phi_\alpha(x_i - l_i, F_i(x))| & \text{if } i \in I_l \\ |\phi_\alpha(u_i - x_i, -F_i(x))| & \text{if } i \in I_u \\ \sqrt{\psi_\alpha(x_i - l_i, F_i(x)) + \psi_\alpha(u_i - x_i, -F_i(x))} & \text{if } i \in I_{lu} \end{cases}, \quad i = 1, \dots, n. \tag{5.1}$$

A number of statements can be made about this reformulation. To name a few: (a) If F is continuously differentiable around $x \in \mathbb{R}^n$, then H is semismooth at x. Furthermore, if F' is locally Lipschitz continuous around x, then H is strongly semismooth at x; (b) If $x^* \in \mathbb{R}^n$ is a strongly regular solution of the MCP, then H satisfies the BD-regularity at x^*; and (c) Under mild conditions, any stationary point of (1.4) is already a solution of the MCP. These results and their proofs can be obtained from [28]. Before we go to the numerical part, we point out that the notion of the strong regularity introduced by Robinson [25] coincides the notion of R-regularity [9].

We report numerical results for the algorithm proposed in Section 4 using the whole set of test problems from the MCPLIB collection [8], which itself is being updated from time to time. The algorithm was implemented in Matlab and run on a SUN Solaris workstation. Instead of a monotone line search we used a nonmonotone version, which was based on an idea in [15] and can be stated as follows. Let $\ell \geq 1$ be a constant integer and calculate a steplength $\lambda_k > 0$ satisfying the nonmonotone Armijo-rule

$$\theta(x^k + \bar{d}^k(\lambda_k)) \leq \mathcal{W}_k + \sigma \nabla\theta(x^k)^T \bar{d}^k_G(\lambda_k), \tag{5.2}$$

where $\mathcal{W}_k := \max\{\theta(x^j) \mid j = k+1-\ell, \ldots, k\}$ denotes the maximal function value of θ over the last ℓ iterations. To choose an initial point, we follow a suggestion of Ulbrich [29] that interior starting points enable the constrained algorithm to identify the correct active constraints more efficiently than starting points close to the boundary. Let $\hat{x}^0$ be the initial point returned by the initialization routine. Then the initial point chosen is given by $x_i^0 = \Pi_{[l_i+0.1, u_i-0.1]}(\hat{x}^0)$, and if $u_i - l_i < 0.1$ for some i, we just let $x_i^0 = \max\{l_i, \min\{\hat{x}_i^0, u_i\}\}$, $i = 1, \ldots, n$. The parameters used in the algorithm were $\rho = 0.5$, $\alpha = 0.7$, $\ell = 4$, $\eta = 0.9$, $p_1 = 10^{-10}$, $p_2 = 2.1$ and $\sigma = 10^{-4}$. The iteration of the algorithm is stopped if either

$$\theta(x^k) \leq 10^{-12} \qquad \text{or} \qquad \|\nabla\theta(x^k)\| \leq 10^{-10}.$$

Our numerical results are summarized in Table 1, in which the first column gives the name of the problem, followed by `n`: the number of variables in the problem, `nl`: the number of lower bounds, `nlu`: the number of both lower and upper bounds, `nF`: the number of free variables (without bounds), `Nit`: the number of iterations, `Nf`: the number of evaluations of the function F, $\theta(x^f)$: the value of $\theta(\cdot)$ at the final iterate, $\|\nabla\theta(x^f)\|$: the value of $\|\nabla\theta(\cdot)\|$ at the final iterate, and $t^*(\lambda)_{ave}$: the average of all $t^*(\lambda_k)$. If $t^*(\lambda)_{ave}$ is close to zero then the projected Newton direction is used most of the time, while if it is close to one the projected gradient direction is used most of the time. `Nit` is equal to the number of evaluations of the Jacobian $F'(x)$ and the number of subproblems (4.1) or systems of linear equations solved.

The results presented in Table 1 show that the algorithm was able to solve most of the problems in the MCPLIB collection in a small number of iterations and are comparable to those results obtained with existing methods. For example, when restricted to problems of size under $n \leq 150$ and of at most one bound per variable, it failed to solve 6 problems, compared to 5 in [29] where a more costly quadratic program solver has to be invoked from time to time in order to solve subproblems. We also note that our algorithm appears to have more failures than the *infeasible* algorithm reported in [20] by Munson, et al. Although a direct comparison is not possible, partially because different initial points were used, we feel that with some fine tuning our *feasible* algorithm can be made just as reliable as the *infeasible* counterpart. However, such work is beyond the scope of the current paper. The problems we failed to solve are either ill-conditioned (with a large condition number) or badly scaled. It is also noted that almost all problems except `Billups` we failed to solve are recently added new problems to the MCPLIB, which are known to be very hard to solve. Some problems, for example, `ehl_k40`, which appeared in previous versions, contain quite different data and are actually new problems. To design more strategies including heuristic ones to efficiently solve these new but hard problems is left for our future research.

Table 1. Numerical results for all problems from MCPLIB

Problem	n	nl	nlu	nF	Nit	nf	$\theta(x^f)$	$\|\nabla\theta(x^f)\|$	$t^*(\lambda)_{ave}$
badfree	5	4	0	1	5	5	2.68e-13	8.87e-07	0.00e+00
bert_oc	5000	0	1000	4000	5	5	4.93e-24	7.03e-14	1.17e-03
bertsekas	15	15	0	0	13	21	9.85e-17	6.12e-07	2.37e-04
billups	1	1	0	0	–	–	–	–	–
bishop	1645	1645	0	0	–	–	–	–	–
bratu	5625	0	5625	0	14	14	3.31e-13	2.67e-06	3.41e-04
choi	13	0	13	0	5	5	1.31e-17	6.03e-10	1.45e-15
colvdual	20	20	0	0	11	12	1.03e-16	7.92e-08	3.21e-04
colvnlp	15	15	0	0	8	10	7.13e-14	1.02e-05	3.20e-16
cycle	1	1	0	0	6	6	2.18e-15	2.44e-07	1.90e-14
degen	2	2	0	0	5	5	7.72e-15	2.28e-07	7.40e-03
duopoly	63	63	0	0	–	–	–	–	–
ehl_k40	41	40	0	1	–	–	–	–	–
ehl_k60	61	60	0	1	13	14	1.73e-17	4.13e-05	1.43e-03
ehl_k80	81	80	0	1	14	15	4.60e-15	1.60e-03	3.62e-03
ehl_kost	101	100	0	1	13	13	4.72e-13	2.41e-02	5.24e-03
electric	158	48	98	12	51	98	5.74e-15	3.90e-01	0.00e+00
explcp	16	16	0	0	7	7	2.71e-21	5.15e-11	1.56e-02
forcebsm	184	118	0	66	–	–	–	–	–
forcedsa	186	116	0	70	–	–	–	–	–
freebert	15	10	0	5	12	13	1.66e-16	7.93e-07	5.38e-05
gafni	5	0	5	0	37	39	4.00e-18	2.59e-07	2.66e-16
games	16	12	0	4	9	11	5.82e-16	9.69e-07	4.34e-15
hanskoop	14	14	0	0	25	50	2.61e-17	1.73e-08	7.09e-02
hydroc06	29	11	0	18	8	8	7.32e-16	4.88e-05	9.27e-03
hydroc20	99	39	0	60	22	28	1.10e-17	5.78e-06	1.00e-02
jel	6	6	0	0	7	7	3.96e-14	5.01e-06	4.74e-14
josephy	4	4	0	0	5	5	3.36e-21	1.59e-10	8.81e-06
kojshin	4	4	0	0	5	5	1.25e-22	1.19e-10	3.21e-04
lincont	419	170	0	249	–	–	–	–	–
mathinum	3	3	0	0	6	6	8.78e-15	2.63e-07	2.93e-15
mathisum	4	4	0	0	10	10	3.92e-15	4.61e-07	3.25e-07
methan08	31	15	0	16	5	5	2.93e-16	1.83e-04	2.85e-13
nash	10	10	0	0	6	6	1.47e-17	2.87e-07	3.92e-16
ne-hard	3	0	0	3	–	–	–	–	–
obstacle	2500	0	2500	0	7	7	2.69e-13	2.36e-06	2.74e-03
opt_cont	288	0	144	144	5	5	7.96e-23	2.64e-11	0.00e+00
opt_cont127	4096	0	2048	2048	5	5	1.60e-18	5.05e-09	0.00e+00
opt_cont31	1024	0	512	512	5	5	1.27e-20	5.02e-10	0.00e+00
opt_cont255	8192	0	4096	4096	5	5	2.52e-18	7.12e-09	0.00e+00
opt_cont511	16384	0	8192	8192	5	5	7.62e-14	8.71e-07	2.05e-11
pgvon106	106	106	0	0	27	57	1.09e-14	1.88e-02	0.00e+00
pies	42	32	10	0	45	81	7.85e-20	1.02e-09	2.90e-01
powell	16	16	0	0	7	7	2.94e-14	1.51e-06	1.11e-05
powell_mcp	8	0	0	8	3	3	3.33e-13	7.38e-06	2.53e-16
qp	4	2	0	2	5	5	1.54e-22	2.28e-11	2.59e-19
scarfanum	13	13	0	0	11	11	8.24e-17	8.40e-08	1.74e-02
scarfasum	14	14	0	0	10	10	8.24e-17	8.82e-08	1.53e-02
scarfbsum	40	40	0	0	42	121	7.68e-17	1.48e-05	2.00e-01
shubik	45	45	0	0	–	–	–	–	–
simple-ex	17	13	0	4	–	–	–	–	–
simple-red	13	13	0	0	12	12	7.42e-22	1.78e-11	9.29e-13
sppe	27	27	0	0	5	5	1.54e-19	6.92e-10	0.00e+00
tinloi	146	146	0	0	16	22	2.85e-16	1.62e-04	3.55e-02
tobin	42	42	0	0	4	4	5.78e-22	5.09e-10	2.60e-10
trafelas	2904	2300	0	604	35	59	2.96e-18	2.45e-09	1.09e-01

6. Conclusions

In this paper, by introducing a projected asymptotically Newton direction and by doing a curved line search, we have proposed a new feasible projected Newton-type method for solving mixed complementarity problems. This new method achieves both theoretical and numerical excellence. We also feel that the idea introduced in this paper can be used in other projected Newton-type methods (e.g., [2, 3, 10, 17, 26]) to enhance those methods' theoretical results or numerical performance or even both.

Acknowledgement. The authors would like to thank the associate editor and three anonymous referees for their detailed comments which considerably improved the presentation of the paper.

References

[1] D.P. Bertsekas, "On the Goldstein-Levitin-Polyak gradient projection method", *IEEE Transactions on Automatic Control, 21* (1976), 174–184.
[2] S.C. Billups, "Improving the robustness of descent-based methods for semismooth equations using proximal perturbations", *Mathematical Programming, 87* (2000), 153–175.
[3] S.C. Billups, S.P. Dirkse and M.C. Ferris, "A comparison of algorithms for large-scale mixed complementarity problems", *Computational Optimization and Applications, 7* (1997), 3–25.
[4] P.H. Calamai and J.J. Moré, "Projected gradient methods for linearly constrained problems", *Mathematical Programming, 39* (1987), 93–116.
[5] B. Chen, X. Chen and C. Kanzow, "A penalized Fischer-Burmeister NCP-function", *Mathematical Programming, 88* (2000), 211–216.
[6] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
[7] T. De Luca, F. Facchinei and C. Kanzow, "A semismooth equation approach to the solution of nonlinear complementarity problems", *Mathematical Programming, 75* (1996), 407–439.
[8] S.P. Dirkse and M.C. Ferris, "MCPLIB: A collection of nonlinear mixed complementarity problems", *Optimization Methods and Software, 5* (1995), 319–345.
[9] F. Facchinei and J. Soares, "A new merit function for nonlinear complementarity problems and a related algorithm", *SIAM Journal on Optimization, 7* (1997), 225–247.
[10] M.C. Ferris, C. Kanzow and T.S. Munson, "Feasible descent algorithms for mixed complementarity problems", *Mathematical Programming, 86* (1999), 475–497.
[11] M.C. Ferris and J.-S. Pang, "Engineering and economic applications of complementarity problems", *SIAM Review, 39* (1997), 669–713.
[12] A. Fischer, "A special Newton-type optimization method", *Optimization, 24* (1992), 269–284.
[13] A. Fischer, "A new constrained optimization reformulation for complementarity problems", *Journal of Optimization Theory and Applications, 97* (1998), 105–117.
[14] E.H. Gafni and D.P. Bertsekas, "Two-metric projection methods for constrained optimization", *SIAM Journal on Control and Optimization, 22* (1984), 936–964.
[15] L. Grippo, F.L. Lampariello and S. Lucidi, "A nonmonotone linesearch technique for Newton's method", *SIAM Journal on Numerical Analysis*, 23 (1986), 707–716.
[16] P.T. Harker and J.-S. Pang, "Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications", *Mathematical Programming, 48* (1990), 161–220.
[17] C. Kanzow, "Strictly feasible equation-based methods for mixed complementarity problems", *Numerische Mathematik, 89* (2001), 135–160.
[18] C. Kanzow and H.-D. Qi, "A QP-free constrained Newton-type method for variational inequality problems", *Mathematical Programming, 85* (1999), 81–106.
[19] R. Mifflin, "Semismooth and semiconvex functions in constrained optimization", *SIAM Journal on Control and Optimization, 15* (1977), 957–972.
[20] T.S. Munson, F. Facchinei, M.C. Ferris, A. Fischer and C. Kanzow, "The semismooth algorithm for large scale complementarity problems", *INFORMS Journal on Computing, 13* (2001), 294–311.
[21] J.-S. Pang and L. Qi, "Nonsmooth equations: Motivation and algorithms", *SIAM Journal on Optimization, 3* (1993), 443–465.

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

[23] L. Qi, "Regular pseudo-smooth NCP and BVIP functions and globally and quadratically convergent generalized Newton methods for complementarity and variational inequality problems", *Mathematics of Operations Research, 24* (1999), 440–471.

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

[25] S.M. Robinson, *Generalized equations*, in Mathematical Programming: The State of the Art, A. Bachem, M. Grötschel and B. Korte, eds., Springer-Verlag, Berlin, 1983, 346–367.

[26] D. Sun, "Projected Newton-type methods for constrained nonsmooth equations", Talk presented at the "Workshop on Nonsmooth and Smoothing Methods" in Hong Kong City University, Hong Kong, China, December 1998.

[27] D. Sun and L. Qi, "On NCP-functions", *Computational Optimization and Applications, 13* (1999), 201–220.

[28] D. Sun and R.S. Womersley, "A new unconstrained differentiable merit function for box constrained variational inequality problems and a damped Gauss-Newton method", *SIAM Journal on Optimization, 9* (1999), 388–413.

[29] M. Ulbrich, "Nonmonotone trust-region methods for bound-constrained semismooth equations with applications to nonlinear mixed complementarity problems", *SIAM Journal on Optimization, 11* (2001), 889–917.

[30] N. Yamashita and M. Fukushima, "A new merit function and a descent method for semi-definite programming problems", in *Reformulations: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, Kluwer Academic Publishers B.V., 1999, 381–404.

[31] E.H. Zarantonello, "Projections on convex sets in Hilbert space and spectral theory", in E.H. Zarantonello, ed., Contributions to Nonlinear Functional Analysis, Academic Press, New York, 1971, 343–424.