Active-Set Projected Trust-Region Algorithm for Box-Constrained Nonsmooth Equations¹

L. Qi,² X. J. Tong,³ AND D. H. Li⁴

Communicated by L. Polak

Abstract. In this paper, by means of an active-set strategy, we present a trust-region method for solving box-constrained nonsmooth equations. Nice properties of the proposed method include: (a) all iterates remain feasible; (b) the search direction, as adequate combination of the projected gradient direction and the trust-region direction, is an asymptotic Newton direction under mild conditions; (c) the subproblem of the proposed method, possessing the form of an unconstrained trust-region subproblem, can be solved by existing methods; (d) the subproblem of the proposed method is of reduced dimension, which is potentially cheaper when applied to solve large-scale problems. Under appropriate conditions, we establish global and local superlinear/quadratic convergence of the method. Preliminary numerical results are given.

Key Words. Nonsmooth equations, active-set strategy, trust-region methods, global convergence, superlinear/quadratic convergence.

1. Introduction

Many practical problems, such as the nonlinear complementarity problem, the box constrained variational inequality problem, and the KKT system of a variational inequality problem or a constrained optimization problem, can be reformulated as a system of nonlinear equations $H(x) = 0$

¹This work was partially supported by the Research Grant Council of Hong Kong, the Australian Research Council, and a Hong Kong Polytechnic University Postdoctoral Fellowship, and Grant G.63.37.T494 of The Hong Kong Polytechnic University.
²Professor, Department of Applied Mathematics, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.
³Professor, Institute of Mathematics, Changsha University of Science and Technology, Changsha, China.
⁴Professor, Institute of Applied Mathematics, Hunan University, Changsha, China.
The function $H$ is generally not smooth, but semismooth (Ref. 3). In some cases, the squared norm function $\|H(x)\|^2$ is continuously differentiable (Ref. 6). Nonsmooth Newton methods, smoothing Newton methods, and quasi-Newton methods for solving this kind of nonsmooth equations have received much attention in the last decade. So far, most existing Newton-like methods are generally infeasible methods. In other words, the sequence generated by a nonsmooth or smoothing Newton method may not lie in the feasible set. However, there are some problems that are only defined in a feasible set or by some desirable properties held only in a subset; see Refs. 7 and 1. Kanzow (Ref. 8) and Bellavia, Macconi, and Morini (Ref. 9) considered feasible methods for solving smooth equations with bounded constraints. The former used an active-set strategy to reduce the dimension of subproblem, and the latter used an affine scaling trust-region approach so that the subproblem has an unconstrained form. Under some conditions, these two papers established global and local superlinear convergence of the proposed algorithms. On the other hand, some authors studied methods for solving constrained nonsmooth equations. In Ref. 10, Gabriel and Pang proposed a trust-region method for solving constrained nonsmooth equations. In Ref. 5, Ulbrich studied further the box-constrained system of nonsmooth equations

$$H(x) = 0, \quad x \in X,$$

(1)

where

$$X = \{x \in \mathbb{R}^n | l \leq x \leq u\}, \quad l \in \{\mathbb{R} \cup \{-\infty\}\}^n, \quad u \in \{\mathbb{R} \cup \{\infty\}\}^n,$$

and the function $H: \mathbb{R}^n \supseteq U \to \mathbb{R}^n$ is defined on an open set $U$ containing the feasible set $X$ and is locally Lipschitzian continuous.

In this paper, we consider also problem (1). Let

$$\theta(x) = (1/2)\|H(x)\|^2.$$

It is clear that problem (1) is equivalent to the following global optimization problem:

$$\min \theta(x), \quad \text{s.t. } x \in X,$$

(2)

in the sense that (1) has a solution if and only if the optimal objective function value of (2) is zero; in that case, $x$ is a solution of (1) if and only if it is a global optimal solution of (2).

We assume that the function $\theta$ is continuously differentiable in $U$.

In Ulbrich (Ref. 5), a nonmonotone trust-region method was presented for solving (1). At each iteration of the Ulbrich method, a projected Newton step was used if a Dennis-Moré type condition held. Otherwise, a trust-region problem with bounded constraints was solved. Global and superlinear/quadratic convergence of the method was established. More
generally, Gabriel and Pang (Ref. 10) studied trust-region methods for solving linear constrained nonsmooth equations where the function $H$ is locally Lipschitzian continuous, but not necessarily semismooth. By introducing an iteration function, Gabriel and Pang (Ref. 10) developed a wide class of trust-region methods which are globally convergent. Some of their methods are also superlinearly convergent. We note that the trust region subproblems of methods in Refs. 5 and 10 are constrained problems and are of full dimension. In this paper, we present a trust-region type method in a different way. First, the trust-region subproblem of our method is a reduced unconstrained trust-region problem. It can be solved by any existing effective method such as the truncated conjugate gradient method and the Lanczos method. Moreover, since the subproblem is of lower dimension, it has the potential of being numerically cheaper when applied to solve large-scale problems. In addition, motivated by recent works of Kanzow (Ref. 11) and of Sun, Womersley, and Qi (Ref. 12), at each step of our method, the search direction is chosen as an adequate combination of the projected gradient direction and the projected trust-region direction, which is an asymptotic Newton direction. Under appropriate conditions, the global and superlinear/quadratic convergence of the proposed method is established.

The remainder of the paper is organized as follows. In Section 2, we define an approximate active set and deduce the search direction used in the algorithm. In Section 3, we describe the steps of the proposed method. In Sections 4 and 5, we prove the global and superlinear/quadratic convergence of the proposed method. Finally, we give some preliminary numerical experiments in Section 6.

Some words about the notation. For a differentiable function $F: \mathbb{R}^n \rightarrow U \subset \mathbb{R}^n$, we use $F'(x)$ to denote the Jacobian of $F$ at $x \in \mathbb{R}^n$, whereas we use $\nabla F(x)$ to denote its transpose. We use $P_X(v)$ to denote the projection of vector $v$ onto $X$. Throughout the paper, without specification, $\| \cdot \|$ denotes the Euclidean norm of a vector. For a given positive-definite matrix $G$, $\| u \|_G$ is defined by $(u^T Gu)^{1/2}$. For a matrix $M = (m_{ij}) \in \mathbb{R}^{t \times t}$ and index sets $I, J \subseteq \{1, 2, \ldots, t \}$, $M_{IJ}$ stands for the submatrix of $M$ whose elements are $m_{ij}$, with $i \in I, j \in J$. If $I = \{1, 2, \ldots, t \}$ or $J = \{1, 2, \ldots, t \}$, then $M_{IJ}$ is simplified as $M_I$ or $M_J$, respectively.

Let $F: \mathbb{R}^n \supset U \rightarrow \mathbb{R}^n$ be locally Lipschitzian continuous. By the Rademacher theorem, $F$ is differentiable almost everywhere. Let $D_F$ denote the set of points where $F$ is differentiable. Then, the B-subdifferential of $F$ at $x \in X$ is defined to be

$$
\partial_B F(x) = \left\{ \lim_{x^k \rightarrow x} \nabla F(x^k)^T \right\}. 
$$

(3)
The Clarke generalized Jacobian (Ref. 13) of $F$ at $x$ is defined to be

$$\partial F(x) = \text{conv} \partial_B F(x).$$

(4)

$F$ is called semismooth at $x$ if $F$ is directionally differentiable at $x$ and, for all $V \in \partial F(x + h)$ and $h \to 0$, it holds that

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

(5)

$F$ is called strongly semismooth at $x$ if the term $o(\|h\|)$ in (5) is replaced by $O(\|h\|^2)$. $F$ is called BD-regular at $x$ if every element in $V \in \partial F(x)$ is nonsingular.

2. Active-Set Estimate and Search Direction

In this section, we define an estimate of the active set by using a way similar to Ref. 8 and describe how to construct a search direction.

Let

$$\xi_k := \min \{ \delta, c\sqrt{\|H(x^k)\|} \},$$

(6)

where $\delta$ and $c$ are positive constants such that

$$0 < \delta < (1/2) \min_{1 \leq i \leq n} (u_i - l_i).$$

Define the index sets

$$A_k := \{ i \in \{1, 2, \ldots, n \} | x^k_i - l_i \leq \xi_k \text{ or } u_i - x^k_i \leq \xi_k \},$$

$$I_k := \{1, 2, \ldots, n\} \setminus A_k = \{ i | l_i + \xi_k < x^k_i < u_i - \xi_k \}.$$  (7)

It $\xi_k$ is small, the set $A_k$ is an estimate of the active set. Moreover, as we shall see later, under reasonable conditions, $A_k$ coincides with the active set when $k$ is sufficiently large.

By using the active set estimate, we deduce the search direction used in our algorithm. Let $\Delta > 0$, $\Delta_{\text{max}} > 0$, and $\gamma_k > 0$ be given. Let $x^k$ be the current iterate. The search direction $d(\Delta)$ is obtained in three steps as follows.

Step 1. Projected Gradient Direction. Compute

$$d^k_G(\Delta) = - (\Delta/\Delta_{\text{max}}) \gamma_k \nabla \theta(x^k),$$  (8a)

$$\tilde{d}^k_G(\Delta) = P_k [x^k + d^k_G(\Delta)] - x^k.$$  (8b)

The direction $\tilde{d}^k_G(\Delta)$ is called the projected gradient of $\theta$. It possesses some nice properties (Ref. 14). In particular, it is a feasible and descent direction of $\theta$ at $x^k$ if $x^k$ is feasible. However, generally it does not guarantee superlinear convergence. To speed up the convergence, we compute the projected trust-region direction.
Step 2. Projected Trust-Region Direction. To this end, we determine first the $A_k$ part of $d_k^*(\Delta)$ by using an active-set strategy similar to that used in Refs. 14, 15. Define a subvector $v_k^{A}$ whose elements are given by

$$v_k^i := \begin{cases} x_k^i - l_i, & \text{if } x_k^i - l_i \leq \xi_k, \\ u_i - x_k^i, & \text{if } u_i - x_k^i \leq \xi_k. \end{cases}$$

Accordingly, we define the subvector $d_k^{*A}(\Delta)$ by

$$d_k^{*A}(\Delta) = \min \{1, \Delta/\|v_k^{A}\|\} v_k^{A}, \quad (9)$$

where we let $d_k^{*A}(\Delta) = 0$, if $v_k^{A} = 0$.

Next, we go to determine the subvector $d_k^{*I}(\Delta)$ of $d_k^*(\Delta)$ by solving a reduced trust-region subproblem. Let $V_k \in \partial H(x_k)$ be partitioned into

$$V_k = (V_k^{A}, V_k^{I}),$$

where

$$V_k^{A} \in \mathbb{R}^{n \times |A_k|} \quad \text{and} \quad V_k^{I} \in \mathbb{R}^{n \times |I_k|}.$$ 

Let $\tilde{d}_k^k(\Delta)$ be a solution of the following reduced trust-region subproblem:

$$\begin{align*}
\min & \quad ((V_k^{I})^T [H(x_k) + V_k^{A} \tilde{d}_k^k(\Delta)])^T d + (1/2)d^T (V_k^{I})^T V_k^{I} d, \\
\text{s.t.} & \quad \|d\| \leq \Delta. \quad (10a)
\end{align*}$$

Then, the trust-region direction is determined by

$$d_k^k(\Delta) = \begin{bmatrix} \tilde{d}_k^k(\Delta) \\ \tilde{d}_k^I(\Delta) \end{bmatrix}. \quad (11)$$

Accordingly, the projected trust-region direction is given by

$$\tilde{d}_k^k(\Delta) = P_X [x_k^k + d_k^k(\Delta)] - x_k^k. \quad (12)$$

Even though it speeds up the local convergence of the iterates, the projected trust-region direction may not be a descent direction of $\theta$ when $x_k^k$ is far from the solution. On the other hand, the projected gradient is always a descent direction of $\theta$. This motivates us to use a combination of these two directions as the search direction so that, as we hope, it is a descent direction and retains the fast local convergence property.

Step 3. Search Direction. Let

$$\tilde{d}_k^k(\Delta) = t_k^*(\Delta) \tilde{d}_k^k(\Delta) + (1 - t_k^*(\Delta)) d_k^k(\Delta), \quad (13)$$
where $t^*_k(\Delta) \in (0, 1)$ is a solution of the following one-dimensional quadratic minimization problem:

$$
\min_{t \in [0, 1]} \| H(x^k) + V^k [t \tilde{d}^k_G(\Delta) + (1 - t) \tilde{d}^k_{ip}(\Delta)] \|^2 \triangleq q^k(t).
$$

(14)

By letting

$$
\nabla q^k(\Delta) = 0,
$$

we get

$$
t_k(\Delta) = \begin{cases} 
- [H(x^k) + V^k \tilde{d}^k_{ip}(\Delta)]^T V^k [\tilde{d}^k_G(\Delta) - \tilde{d}^k_{ip}(\Delta)] / \| V^k [\tilde{d}^k_G(\Delta) - \tilde{d}^k_{ip}(\Delta)] \|^2, & \text{if } V^k \tilde{d}^k_G(\Delta) \neq V^k \tilde{d}^k_{ip}(\Delta), \\
\text{any number in } (-\infty, +\infty), & \text{if } V^k \tilde{d}^k_G(\Delta) = V^k \tilde{d}^k_{ip}(\Delta).
\end{cases}
$$

(15)

It is not difficult to prove the following lemma.

**Lemma 2.1.** Let $x^k \in X$. Then, the solution of (14) is

$$
t^*_k(\Delta) = \max \{0, \min \{1, t_k(\Delta)\}\},
$$

where $t_k(\Delta)$ is defined by (15).

**Remark 2.1.** We use a way similar to Ref. 8 to define the active-set estimate, but the search direction is constructed by using a different method.

3. Projected Trust-Region Algorithm

Section 2 has described the details of determining a feasible descent direction. We state now the steps of a projected trust-region algorithm as follows.

**Algorithm 3.1.** Projected Trust-Region Algorithm.

Step 0. Initialization. Given an $x^0 \in X$, constants satisfying $0 < \alpha_1 < 1 < \alpha_2$, $0 < \rho_1 < \rho_2 < 1$, $\sigma \in (0, 1)$, $\eta \in (0, 1)$, $\Delta_0 > 0$, $\Delta_{\max} > \Delta_{\min} > 0$, $c > 0$, $0 < \delta < (1/2)\min_{1 \leq i \leq n} (u_i - l_i)$. Let $k := 0$.

Step 1. Termination Criterion. Stop if $x^k$ is a stationary point of problem (2). Otherwise, let

$$
\Delta^k := \min \{\Delta_{\max}, \max \{\Delta_{\min}, \Delta_k\}\}, \quad \hat{\Delta} := \Delta^k.
$$

Choose $V^k \in \partial H(x^k)$.
Step 2. Active-Set Estimate. Determine the index sets $A_k$ and $I_k$ by (6) and (7).

Step 3. Trust-Region Subproblem. Let

$$d^k_{tr}(\hat{\Delta}) = \begin{pmatrix} \tilde{d}^k_{Ak}(\hat{\Delta}) \\ \tilde{d}^k_{Ik}(\hat{\Delta}) \end{pmatrix},$$

where $\tilde{d}^k_{Ak}(\hat{\Delta})$ and $\tilde{d}^k_{Ik}(\hat{\Delta})$ are determined by (9) and (10).

Step 4. Search Direction. Set

$$\gamma_k := \min \{ 1, \Delta_{\text{max}} / \| \nabla \theta(x^k) \|, \eta^\| H(x^k) \| / \| \nabla \theta(x^k) \|, \eta^\theta(x^k) / \| \nabla \theta(x^k) \|_2^2 \}.$$

(17)

Compute $\tilde{d}^k_{G}(\hat{\Delta})$, $\tilde{d}^k_{tr}(\hat{\Delta})$, $t^*_k(\hat{\Delta})$ by (8), (12), (16). Let

$$\tilde{d}^k(\hat{\Delta}) = t^*_k(\hat{\Delta}) \tilde{d}^k_{G}(\hat{\Delta}) + (1 - t^*_k(\hat{\Delta})) \tilde{d}^k_{tr}(\hat{\Delta}),$$

(18)

where $t^*_k(\hat{\Delta})$ is defined by (16).

Step 5. Test the Search Direction. Compute

$$\hat{r}_k := [\theta(x^k + \tilde{d}^k(\hat{\Delta})) - \theta(x^k)] / [(1/2) \| H(x^k) + V^k \tilde{d}^k(\hat{\Delta}) \|^2 - \theta(x^k)].$$

(19)

If the following two conditions hold:

$$\theta(x^k) - (1/2) \| H(x^k) + V^k \tilde{d}^k(\hat{\Delta}) \|^2 \geq - \sigma \nabla \theta(x^k)^T \tilde{d}^k_{G}(\hat{\Delta}),$$

(20)

$$\hat{r}_k \geq \rho_1,$$

(21)

let $s^k := \tilde{d}^k(\hat{\Delta}), x^{k+1} := x^k + s^k, \delta_k := \hat{\Delta}$, and

$$\Delta_{k+1} := \begin{cases} \hat{\Delta}, & \text{if } \rho_1 \leq \hat{r}_k < \rho_2, \\ \alpha_2 \hat{\Delta}, & \text{if } \hat{r}_k \geq \rho_2. \end{cases}$$

Let $k := k + 1; \text{return to Step 1. Otherwise, let } \hat{\Delta} := \alpha_1 \hat{\Delta} \text{ and return to Step 3.}$

Remark 3.1. The sequence $\{x^k\}$ generated by Algorithm 3.1 remains in $X$. The subproblem of Algorithm 3.1 is an unconstrained type problem. Moreover, it is of lower dimension. So, it saves computational cost when applied to solve large-scale problems.
Remark 3.2. The updating rule in Algorithm 3.1 was used by some authors (e.g. Ref. 17). At the beginning of iteration $k$, the trust-region radius $\Delta_k$ is set always to be no less than the fixed positive constant $\Delta_{\text{min}}$.

Remark 3.3. The scalar $\Delta_k$ stands for the initial trust-region radius of iteration $k$, whereas $\delta_k$ is the radius corresponding to the acceptable trial step $s^k$.

4. Global Convergence

In this section, we prove the global convergence of Algorithm 3.1. We show that, under appropriate conditions, the sequence generated by the algorithm converges to a stationary point of the equivalent minimization problem (2). A point $x$ is said to be a stationary point of problem (2) if it satisfies

\begin{align}
    x_i = l_i & \Rightarrow (\nabla \theta(x))_i \geq 0, \quad (22a) \\
    x_i = u_i & \Rightarrow (\nabla \theta(x))_i \leq 0, \quad (22b) \\
    x_i \in (l_i, u_i) & \Rightarrow (\nabla \theta(x))_i = 0, \quad (22c)
\end{align}

for $1 \leq i \leq n$.

The conditions that ensure a stationary point to be a solution of (1) can be found in Refs. 18, 19, etc.

The following two lemmas give some interesting properties of the projected operator $P_X(\cdot)$; see Ref. 14.

Lemma 4.1. The projection operator $P_X(\cdot)$ satisfies

(i) for any $x \in X$, $[P_X(z) - z]^T[P_X(z) - x] \leq 0$, for all $z \in R^n$;

(ii) $\|P_X(y) - P_X(x)\| \leq \|y - x\|$, for all $x, y \in R^n$.

Lemma 4.2. Given $x \in R^n$ and $d \in R^n$, the function $\zeta$ defined by

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

is nonincreasing.

Actually, Lemma 4.2 implies that, if $x \in X$ is a stationary point of problem (2), then

$\tilde{d}_G(\Delta) = P_X[x + d_G(\Delta)] - x = 0, \quad \forall \Delta \in (0, \Delta_{\text{max}}].$

From the above two lemmas, we can prove the descent property of the projected gradient direction.
Lemma 4.3. For all $\Delta \in (0, \Delta_{\text{max}}]$, we have
\[
\nabla \theta(x^k)^T \tilde{d}_G^k(\Delta) \leq - (\Delta/\Delta_{\text{max}} \gamma_k) \|\tilde{d}_G^k(\Delta_{\text{max}})\|^2. \tag{23}
\]

Proof. From Lemma 4.1, we have that, for any $\Delta \in (0, \Delta_{\text{max}}]$, \[
\nabla \theta(x^k)^T \tilde{d}_G^k(\Delta) = (\Delta_{\text{max}}/\Delta \gamma_k) \{x^k - [x^k - (\Delta/\Delta_{\text{max}}) \gamma_k \nabla \theta(x^k)]\}^T \{P_X[x^k - (\Delta/\Delta_{\text{max}}) \gamma_k \nabla \theta(x^k)] - x^k\} \\
\leq - (\Delta_{\text{max}}/\Delta \gamma_k) \|\tilde{d}_G^k(\Delta)\|^2. \tag{24}
\]

From Lemma 4.2, we have also \[
\|\tilde{d}_G^k(\Delta)\|/\Delta = \|P_X[x^k - (\Delta/\Delta_{\text{max}}) \gamma_k \nabla \theta(x^k)]x^k\|/\Delta \\
\geq \|P_X[x^k - \gamma_k \nabla \theta(x^k)] - x^k\|/\Delta_{\text{max}} \\
= \|\tilde{d}_G^k(\Delta_{\text{max}})\|/\Delta_{\text{max}},
\]
which combined with (24) yields (23). \hfill \square

The next lemma shows that Algorithm 3.1 is well-defined.

Lemma 4.4. Let \{x^k\} be generated by Algorithm 3.1. If $x^k$ is not a stationary point of problem (2), then the cycling between Step 3 and Step 5 terminates finitely.

Proof. The fact that $x^k$ is not a stationary point implies that $\gamma_k > 0$ and that there is a constant $b > 0$ such that \[
\|\tilde{d}_G^k(\Delta_{\text{max}})\| \geq b > 0.
\]
Since $\theta(x)$ is continuously differentiable, we have \[
\nabla \theta(x^k) = (V^k)^T H(x^k), \quad \text{for all } V^k \in \partial H(x^k).
\]
By (8), we deduce that there is a constant $b_1 > 0$ such that \[
\|V^k \tilde{d}_G^k(\Delta)\| \leq (\Delta/\Delta_{\text{max}}) \gamma_k \|V^k\| \|(V^k)^T H(x^k)\| \leq b_1 \Delta. \tag{25}
\]
Let \[
\hat{\Delta} = \min\{\Delta_{\text{max}}, (1 - \sigma)b^2/b_1^2\Delta_{\text{max}}\}.
\]
From the definition of \( \tilde{d}^k(\hat{\Delta}) \), we get that, for any \( \hat{\Delta} \in (0, \hat{\Delta}] \), it holds that

\[
\theta(x^k) - (1/2)\|H(x^k) + V^k \tilde{d}^k(\hat{\Delta})\|^2 \\
\geq \theta(x^k) - (1/2)\|H(x^k) + V^k \tilde{d}^k(\hat{\Delta})\|^2 \\
= -\nabla \theta(x^k)^T \tilde{d}^k(\hat{\Delta}) - (1/2)\|V^k \tilde{d}^k(\hat{\Delta})\|^2 \\
= -\sigma \nabla \theta(x^k)^T \tilde{d}^k(\hat{\Delta}) - (1 - \sigma)\nabla \theta(x^k)^T \tilde{d}^k(\hat{\Delta}) - (1/2)\|V^k \tilde{d}^k(\hat{\Delta})\|^2 \\
\geq -\sigma \nabla \theta(x^k)^T \tilde{d}^k(\hat{\Delta}) + [(1 - \sigma)\|\tilde{d}^k(\Delta_{\max})\|^2 + b^2/\gamma_k \Delta_{\max}]\hat{\Delta} - (1/2)b^2\hat{\Delta}^2 \\
\geq -\sigma \nabla \theta(x^k)^T \tilde{d}^k(\hat{\Delta}),
\]

(26)

where the second inequality follows from (23) and (25) and the last inequality holds because \( 0 < \gamma_k \leq 1 \) and \( \hat{\Delta} \leq \hat{\Delta} \). This means that condition (20) is satisfied for all \( \hat{\Delta} \) sufficiently small. To complete the proof, it suffices to show that condition (21) is satisfied for all \( \hat{\Delta} \) sufficiently small.

For any \( \hat{\Delta} \in (0, \hat{\Delta}] \), we deduce from Lemma 4.3 and (25) that

\[
(1/2)\|H(x^k) + V^k \tilde{d}^k(\hat{\Delta})\|^2 = (1/2)\|H(x^k)^2 + \nabla \theta(x^k)^T \tilde{d}^k(\hat{\Delta}) + (1/2)\|V^k \tilde{d}^k(\hat{\Delta})\|^2 \\
\leq \theta(x^k) - (\hat{\Delta}/\gamma_k \Delta_{\max})\|\tilde{d}^k(\Delta_{\max})\|^2 + (1/2)b^2\hat{\Delta}^2 \\
\leq \theta(x^k) - (\hat{\Delta}/2\gamma_k \Delta_{\max})\|\tilde{d}^k(\Delta_{\max})\|^2.
\]

(27)

Consequently, we have

\[
(1/2)\|H(x^k) + V^k \tilde{d}^k(\hat{\Delta})\|^2 - \theta(x^k) \\
\leq (1/2)\|H(x^k) + V^k \tilde{d}^k(\hat{\Delta})\|^2 - \theta(x^k) \\
\leq -\hat{\Delta}/2\gamma_k \Delta_{\max}\|\tilde{d}^k(\Delta_{\max})\|^2 < 0,
\]

(28)

where the first inequality follows from (13) and (14). Together with the fact that \( \|\tilde{d}^k(\Delta_{\max})\| \geq b \), inequality (28) implies that there is a constant \( \beta > 0 \) such that

\[
\theta(x^k) - (1/2)\|H(x^k) + V^k \tilde{d}^k(\hat{\Delta})\|^2 \geq \beta \hat{\Delta}.
\]

(29)

For the sake of contradiction, we suppose that the cycling between Steps 3 and 5 is infinite. This means that \( \hat{\Delta} \rightarrow 0 \).

Note that

\[
\|\tilde{d}^k(\hat{\Delta})\| \leq \hat{\Delta}, \quad \|\tilde{d}^k(\hat{\Delta})\| \leq \|\tilde{d}^k(\hat{\Delta})\| \\
\leq \|\tilde{d}^k(\hat{\Delta})\| + \|\tilde{d}^k(\hat{\Delta})\| \leq 2\hat{\Delta},
\]

(30)

where the first inequality follows from (8) and (17), the second inequality follows from (12) and the last inequality follows from (9) and (10), respectively. So, we get from (18)

\[
\|\tilde{d}^k(\hat{\Delta})\| \leq 2\hat{\Delta}.
\]

(31)
By the definition of $\hat{r}_k$, we have
\[
\hat{r}_k = \frac{\theta(x^k + \tilde{d}^k(\hat{\Delta})) - \theta(x^k)}{(1/2)\|H(x^k) + V^k\tilde{d}^k(\hat{\Delta})\|^2 - \theta(x^k)}
\]
\[
= 1 + \frac{\theta(x^k + \tilde{d}^k(\hat{\Delta})) - (1/2)\|H(x^k) + V^k\tilde{d}^k(\hat{\Delta})\|^2 - \theta(x^k)}{(1/2)\|H(x^k) + V^k\tilde{d}^k(\hat{\Delta})\|^2 - \theta(x^k)}
\]
\[
= 1 + \frac{\theta(x^k + \tilde{d}^k(\hat{\Delta})) - \nabla(\theta(x^k))^T\tilde{d}^k(\hat{\Delta}) - (1/2)\tilde{d}^k(\hat{\Delta})^T V^k V^k\tilde{d}^k(\hat{\Delta})}{(1/2)\|H(x^k) + V^k\tilde{d}^k(\hat{\Delta})\|^2 - \theta(x^k)}
\]
\[
= 1 + \frac{-(1/2)\tilde{d}^k(\hat{\Delta})^T(V^k)^T V^k\tilde{d}^k(\hat{\Delta}) + o(\|\tilde{d}^k(\hat{\Delta})\|)}{(1/2)\|H(x^k) + V^k\tilde{d}^k(\hat{\Delta})\|^2 - \theta(x^k)}
\]
\[
= 1 + \frac{-(1/2)\tilde{d}^k(\hat{\Delta})^T(V^k)^T V^k\tilde{d}^k(\hat{\Delta}) + o(\hat{\Delta})}{(1/2)\|H(x^k) + V^k\tilde{d}^k(\hat{\Delta})\|^2 - \theta(x^k)}
\]
\[
= 1 + o(\hat{\Delta})/O(\hat{\Delta}),
\]
where the last equality follows from (29). The last equality shows that, when $\hat{\Delta}$ is sufficiently small, it must hold that $\hat{r}_k \geq \rho_2$. By the updating rule for $\hat{\Delta}$ in Step 5 of Algorithm 3.1, this contradicts the assumption $\hat{\Delta} \to 0$. The contradiction shows that the cycling between Steps 3 and 5 is finite. \hfill \Box

**Proposition 4.1.** Suppose that $x^*$ is a limit point of a subsequence $\{x^k\}_{k \in K}$. If $x^*$ is not a stationary point, then there exist an index $\hat{k} > 0$ and a constant $\hat{\Delta} > 0$ such that, for all $k \geq \hat{k}$ with $k \in K$, the following inequalities hold for all $\hat{\Delta} \in (0, \hat{\Delta})$:
\[
\theta(x^k) - (1/2)\|H(x^k) + V^k\tilde{d}^k(\hat{\Delta})\|^2 \geq - \sigma \nabla(\theta(x^k))^T\tilde{d}^k_G(\hat{\Delta}),
\]
\[
\hat{r}_k = \frac{\theta(x^k + \tilde{d}^k(\hat{\Delta})) - \theta(x^k)}{(1/2)\|H(x^k) + V^k\tilde{d}^k(\hat{\Delta})\|^2 - \theta(x^k)} \geq \rho_1
\]

**Proof.** Denote
\[
\gamma^* = \min \{1_{\Delta_{\max}}/\|\nabla(\theta(x^*))\|, \eta\|H(x^*)\|/\|\nabla(\theta(x^*))\|, \eta\theta(x^*)/\|\nabla(\theta(x^*))\|^2 \} > 0.
\]
Since $x^*$ is not a stationary point of problem (2), we have that $\gamma^* > 0$ and
\[
\|\tilde{d}^k_G(\Delta_{\max})\| \to \|P_X[x^* - \gamma^*\nabla(\theta(x^*))] - x^*\| > 0, \quad \text{as } k \in K, k \to \infty.
\]
\[
(34)
\]
\[
(35)
\]
Therefore, there exist an integer \( \hat{k} > 0 \) and a constant \( \hat{b} > 0 \) such that
\[
\| \hat{d}_G^k(\Delta_{\max}) \| \geq \hat{b}, \quad \forall k \in K, k \geq \hat{k}.
\]

Similar to the proof of Lemma 4.4 [see (25)], there exists a constant \( b_2 > 0 \) such that
\[
\| V^k \hat{d}_G^k(\hat{\Delta}) \| \leq \hat{\Delta} \gamma_k \| V^k \| \| (V^k)^T H(x^k) \| \leq b_2 \hat{\Delta} / \Delta_{\max}.
\]

Let
\[
\Delta' = \min \{ \Delta_{\max}, (1 - \sigma) \hat{b}^2 / b_2^2 \Delta_{\max} \}.
\]

Following the proof of Lemma 4.4 [see (26)–(27)], we deduce that the following two inequalities hold for all \( \hat{\Delta} \in (0, \Delta') \):
\[
\begin{align*}
\theta(x^k) - (1/2) \| H(x^k) + V^k \hat{d}_G^k(\hat{\Delta}) \| ^2 & \geq - \sigma \nabla \theta(x^k)^T \hat{d}_G^k(\hat{\Delta}), \\
(1/2) \| H(x^k) + V^k \hat{d}_G^k(\hat{\Delta}) \| ^2 - \theta(x^k) & \leq - (\hat{\Delta}) / 2 \Delta_{\max} \gamma_k \| \hat{d}_G^k(\Delta_{\max}) \| ^2.
\end{align*}
\]

It is clear that inequality (36) yields (32).

Next, we prove (33). It follows from (34) and (35) that, as \( k \to \infty \) with \( k \in K \),
\[
- \| \hat{d}_G^k(\Delta_{\max}) \| ^2 / \gamma_k \to - \| P_{\Omega} [x^* - \gamma^* \nabla \theta(x^*)] - x^* \| ^2 / \gamma^* \triangleq - b_3 < 0.
\]
So, when \( k \in K \) is sufficiently large,
\[
- \| \hat{d}_G^k(\Delta_{\max}) \| ^2 / \gamma_k \leq - b_3 / 2 < 0.
\]

Inequality (37) together with (38) implies
\[
(1/2) \| H(x^k) + V^k \hat{d}_G^k(\hat{\Delta}) \| ^2 - \theta(x^k) \leq (1/2) \| H(x^k) + V^k \hat{d}_G^k(\hat{\Delta}) \| ^2 - \theta(x^k) \leq - \hat{\Delta} b_3 / 4 \Delta_{\max}.
\]

We rewrite \( r_k \) as
\[
\hat{r}_k = 1 + \frac{-(1/2) \hat{d}_G^k(\hat{\Delta})^T (V^k)^T V^k \hat{d}_G^k(\hat{\Delta}) + o(\| \hat{d}_G^k(\hat{\Delta}) \|)}{(1/2) \| H(x^k) + V^k \hat{d}_G^k(\hat{\Delta}) \| ^2 - \theta(x^k)}.
\]

Note that the inequalities (30) and (31) hold for all \( \hat{\Delta} \in [0, \Delta_{\max}] \). The numerator of the fraction in (40) is \( o(\hat{\Delta}) \). However, inequality (38) implies that the denominator of the fraction in (35) has a lower \( O(\hat{\Delta}) \). Therefore, inequality (33) holds for all \( k \in K \) sufficiently large and \( \hat{\Delta} \in (0, \Delta') \).
Proposition 4.2. Suppose that $x^*$ is a limit point of a subsequence $\{x^k\}_{k \in K}$. If $x^*$ is not a stationary point of problem (2), then
\[
\tilde{\delta} = \liminf_{k \in K, k \to \infty} \delta_k > 0,
\]
where $\delta_k$ is defined in Step 5 of Algorithm 3.1.

Proof. It follows from Proposition 4.1 that there exist a constant $\Delta > 0$ and an index $\hat{k} > 0$ such that inequality $r_{\hat{k}} \geq \rho_1$ holds for all $k \in K$, $k > \hat{k}$, whenever $\hat{\Delta} < \Delta$. By the updating rule of the trust-region radius, we get $\delta_k \geq \alpha_\Delta$. This implies (41). \qed

The following theorem establishes the global convergence of Algorithm 3.1.

Theorem 4.1. Let $\{x^k\}$ be generated by Algorithm 3.1. Then, every accumulation point of $\{x^k\}$ is a stationary point of problem (2).

Proof. Let $x^*$ be an accumulation point of $\{x^k\}$ and let
\[
\lim_{k \in K, k \to \infty} x^k = x^*.
\]
Suppose that $x^*$ is not a stationary point of problem (2). We are going to derive a contradiction. In a way similar to the proof of Proposition 4.1 [see (38)], there exists an integer $\hat{k} > 0$ such that, for all $k \geq \hat{k}$ with $k \in K$,
\[
\|\tilde{d}_G^k(\Delta_{\max})\|/\gamma_k \geq b_4,
\]
with a number $b_4 > 0$. From Step 5 of Algorithm 3.1, Lemma 4.3, and Proposition 4.2, we obtain that
\[
\theta(x^k) - (1/2)\|H(x^k) + V^k \tilde{d}^k(\delta_k)\|^2 \\
\geq -\sigma \nabla \theta(x^k)^T \tilde{d}_G(\delta_k) \\
\geq (\sigma \delta_k / \gamma_k \Delta_{\max}) \|\tilde{d}_G^k(\Delta_{\max})\| \\
\geq \sigma \delta_k b_4 / \Delta_{\max} \\
\geq \sigma \tilde{\delta} b_4 / \Delta_{\max} > 0.
\]
Together with (21), this implies
\[
\theta(x^k) - \theta(x^{k+1}) \geq \rho_1 \sigma \tilde{\delta} b_4 / \Delta_{\max} > 0, \quad \forall k \geq \hat{k}, \ k \in K.
\]
Consequently, we deduce that

\[
\theta(x^0) \geq \sum_{k=0}^{\infty} [\theta(x^k) - \theta(x^{k+1})] \\
\geq \sum_{k=0}^{\infty} \rho_1[\theta(x^k) - (1/2)\|H(x^k) + V^k\delta_k\|^2] \\
\geq \sum_{k > k, k \in K} \rho_1[\theta(x^k) - (1/2)\|H(x^k) + V^k\delta_k\|^2] \\
\geq \sum_{k > k, k \in K} \rho_1 \sigma \delta_k^5/\Delta_{\max} = \infty.
\]

Since \(\{\theta(x^k)\}\) is nonincreasing, the above inequality yields a contradiction, which completes the proof. 

\[\square\]

5. Superlinear Convergence

In this section, we prove the superlinear/quadratic convergence of Algorithm 3.1. Throughout this section, without specification, we assume always that there is an accumulation point \(x^*\) of \(\{x^k\}\) which is a BD-regular solution of (1) with \(H(x^*) = 0\).

Let \(\{x^k\}_{k \in K}\) be a subsequence of \(\{x^k\}\) that converges to \(x^*\). Denote

\[A_* = \{i \in \{1, 2, \ldots, n\} | x_i^* = l_i \text{ or } x_i^* = u_i\}, \quad I_* = \{1, 2, \ldots, n\} \setminus A_*\.
\]

The following lemma can be found in Refs. 3, 4.

**Lemma 5.1.** Let \(x^*\) be a BD-regular solution of \(H(x) = 0\). Then, the following statements hold:

(i) There exist positive constants \(\kappa_1\) and \(\epsilon_1\) such that, when \(\|x - x^*\| \leq \epsilon_1\), every \(V \in \partial H(x)\) is nonsingular and satisfies

\[\|V^{-1}\| \leq \kappa_1.\]

(ii) There exist positive constants \(\kappa_2\) and \(\epsilon_2\) such that

\[\|H(x)\| \geq \kappa_2\|x - x^*\|,

for all \(x\) satisfying \(\|x - x^*\| \leq \epsilon_2\).

The following lemma can be proved in a way similar to the proof of Lemma 7.1 in Ref. 16. In particular, it shows that the set \(A_k\) identifies actually the active constraints when \(k\) is sufficiently large.
**Lemma 5.2.** The following statements hold:

(i) We have $A_k = A_*$ and $I_k = I_*$ for $x^k \in K$ sufficiently large.

(ii) There is a constant $c_1 > 0$ such that the matrices $(V_{ik}^T V_{ik})^{-1}$ are nonsingular and

$$
\|(V_{ik}^T V_{ik})^{-1}\| \leq c_1,
$$

for all $x^k \in K$ sufficiently large.

**Lemma 5.3.** For $k \in K$ sufficiently large, we have

$$
\tilde{d}^k_{Ak}(\Delta^k) = x^* - x^k.
$$

Moreover, there exists a constant $\kappa_3 > 0$ such that

$$
\|\tilde{d}^k_{Ak}(\Delta^k)\| \leq \kappa_3 \|\nabla \theta(x^k)\|.
$$

**Proof.** Noting that $\Delta^k \geq \Delta_{\text{min}} > 0$ for each iteration, the first result (44) can be proved in a similar way to the proof of Corollary 9.5 in Ref. 8.

We turn to proving (45). First, by the continuous differentiability of $\theta$, we have

$$
\nabla \theta(x^k) = (V^k)H(x^k).
$$

So, we get from Lemma 5.1 that, when $k \in K$ is large enough,

$$
\|H(x^k)\| \leq \kappa_1 \|\nabla \theta(x^k)\|.
$$

Therefore, we deduce from (44) and Lemma 5.1 that, when $k \in K$ is sufficiently large,

$$
\|\tilde{d}^k_{Ak}(\Delta^k)\| = \|x^* - x_k^*\| \\
\leq \|x^k - x^*\| \\
\leq (1/\kappa_2) \|H(x^k)\| \\
\leq (\kappa_1/\kappa_2) \|\nabla \theta(x^k)\|.
$$

This proves (45) with $\kappa_3 = \kappa_1/\kappa_2$. \[Q.E.D.\]

The following lemma shows that the solution $\tilde{d}^k_{Ak}(\Delta^k)$ of the trust-region subproblem (10) reduces to the Newton direction when $k \in K$ is sufficiently large.
Lemma 5.4. Let $\tilde{d}_k^k(\Delta^k)$ be the solution of the trust-region subproblem (10) with $\Delta = \Delta^k$. Then, when $k$ is sufficiently large $k \in K$, it holds that

$$
\tilde{d}_k^k(\Delta^k) = -((V_{\lambda_k}^k)^T V_{\lambda_k}^k)^{-1}(V_{\lambda_k}^k)^T (H(x^k) + V_{\lambda_k}^k \tilde{d}_k^k(\Delta^k)).
$$

(47)

Proof. Denote

$$
s_k = -((V_{\lambda_k}^k)^T V_{\lambda_k}^k)^{-1}(V_{\lambda_k}^k)^T (H(x^k) + V_{\lambda_k}^k \tilde{d}_k^k(\Delta^k)).
$$

To show (47), it suffices to show that $s_k$ satisfies the constraint of (10), namely $\|s_k\| \leq \Delta_k$, because $s_k$ is the Newton direction of (10). It follows from Lemmas 5.2 and 5.3 that there exists a constant $b_3 > 0$ such that, for $k \in K$ sufficiently large,

$$
\|s_k\| \leq b_3 \|H(x^k)\| \leq \Delta_{\text{min}} \leq \Delta^k.
$$

This proves the lemma.

Lemma 5.5. We have that, for all $k \in K$ sufficiently large,

$$
x^k + d_{tr}^k(\Delta^k) = x^* + o(\theta(x^k)^{1/2}).
$$

(48)

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

$$
x^k + d_{tr}^k(\Delta^k) = x^* + O(\theta(x^k)).
$$

(49)

Proof. It follows from Lemmas 5.2–5.4 and the semismoothness of $H$ at $x^*$ that

$$
x^k_{\lambda_k} + \tilde{d}_k^k(\Delta^k)
= x^k_{\lambda_k} - ((V_{\lambda_k}^k)^T V_{\lambda_k}^k)^{-1}(V_{\lambda_k}^k)^T [H(x^k) + V_{\lambda_k}^k (x^k_{\lambda_k} - x^k_{\lambda_k})]
= x^k_{\lambda_k} - ((V_{\lambda_k}^k)^T V_{\lambda_k}^k)^{-1}((V_{\lambda_k}^k)^T [H(x^k) + V_{\lambda_k}^k (x^*_{\lambda_k} - x^k_{\lambda_k})] - (V_{\lambda_k}^k)^T V_{\lambda_k}^k (x^k_{\lambda_k} - x^k_{\lambda_k}))
= x^k_{\lambda_k} - ((V_{\lambda_k}^k)^T V_{\lambda_k}^k)^{-1}(V_{\lambda_k}^k)^T [H(x^k) - H(x^*) - V_{\lambda_k}^k (x^k - x^*)]
= x^k_{\lambda_k} + o(\|x^k - x^*\|)
= x^k_{\lambda_k} + o(\theta(x^k)^{1/2}).
$$

(50)

This proves that (48) holds for all $i \in I_k$. On the other hand, Lemma 5.3 shows that (48) holds for all $i \in A_k$. If $H$ is strongly semismooth at $x^*$, then the term $o(\theta(x^k)^{1/2})$ above can be replaced by $O(\theta(x^k))$. Consequently, (49) holds.
Lemma 5.6. We have that, for $k \in K$ large enough,
\[
\tilde{d}^k_r(\Delta^k) = -(x^k - x^*) + o(\theta(x^k)^{1/2}).
\] (51)
Moreover, if $H$ is strongly semismooth at $x^*$,
\[
\tilde{d}^k_r(\Delta^k) = -(x^k - x^*) + O(\theta(x^k)).
\] (52)

Proof. From Lemma 5.5 and the property of a projector, we obtain
\[
\tilde{d}^k_r(\Delta^k) = P_x[x^k + d^k_r(\Delta^k)] - x^k = P_x[x^* + o(\theta(x^k)^{1/2})] - x^k = P_x(x^*) - x^k + \{P_x[x^* + o(\theta(x^k)^{1/2})] - P_x(x^*)\} = -(x^k - x^*) + o(\theta(x^k)^{1/2}).
\] (53)
This proves (51). The equality (52) can be proved in a similar way. \hfill \Box

Now, we prove the asymptotic property of the search direction $\tilde{d}^k(\Delta^k)$.

Lemma 5.7. We have that, for $k \in K$ sufficiently large,
\[
\tilde{d}^k(\Delta^k) = -(x^k - x^*) + o(\theta(x^k)^{1/2})
\] (54)
and
\[
\|\tilde{d}^k(\Delta^k) - d^k_r(\Delta^k)\|/\|d^k_r(\Delta^k)\| = o(1).
\] (55)
Moreover, if $H$ is strongly semismooth at $x^*$, then
\[
\tilde{d}^k(\Delta^k) = -(x^k - x^*) + O(\theta(x^k)),
\] (56)
\[
\|\tilde{d}^k(\Delta^k) - d^k_r(\Delta^k)\|/\|d^k_r(\Delta^k)\| = O(\theta(x^k)^{1/2}).
\] (57)

Proof. By (15) and the nonsingularity of $V^k$, $t_k(\Delta^k)$ is expressed as
\[
t_k(\Delta^k) = \begin{cases} - [H(x^k) + V^k \tilde{d}^k_r(\Delta^k)]^T V^k [\tilde{d}^k_G(\Delta^k) - \tilde{d}^k_r(\Delta^k)]/\|V^k[\tilde{d}^k_G(\Delta^k) - \tilde{d}^k_r(\Delta^k)]\|^2, & \text{if } \tilde{d}^k_G(\Delta^k) \neq \tilde{d}^k_r(\Delta^k), \\ \text{any number in } (-\infty, \infty), & \text{if } \tilde{d}^k_G(\Delta^k) = \tilde{d}^k_r(\Delta^k). \end{cases}
\]
For the case $\tilde{d}^k(\Delta^k) \neq \tilde{d}^k_r(\Delta^k)$, we get from Lemma 5.6
\[
H(x^k) + V^k \tilde{d}^k_r(\Delta^k) = H(x^k) - V^k(x^k - x^*) + o(\theta(x^k)^{1/2}) = o(\theta(x^k)^{1/2}).
\] (58)
We have also
\[ V^k \bar{d}^k_{ir}(\Delta^k) = - V^k(x^k - x^*) + o(\theta(x^k)^{1/2}) = - H(x^k) + o(\theta(x^k)^{1/2}) = O(\theta(x^k)^{1/2}). \]
Together with (58), this gives
\[ [H(x^k) + V^k \bar{d}^k_{ir}(\Delta^k)]^T V^k \bar{d}^k_{ir}(\Delta^k) = o(\theta(x^k)). \] (59)
On the other hand, by the choice of \( \gamma_k \) in (17), we get
\[ \| \bar{d}^k_G(\Delta^k) \| = \| P_X [x^k - (\Delta^k/\Delta_{max}) \gamma_k \nabla \theta(x^k)] - x^k \| \leq (\Delta^k/\Delta_{max}) \gamma_k \| \nabla \theta(x^k) \| \leq \gamma_k \| \nabla \theta(x^k) \| \leq \eta \| H(x^k) \| = O(\theta(x^k)^{1/2}). \] (60)
Together with (58), this implies
\[ [H(x^k) + V^k \bar{d}^k_{ir}(\Delta^k)]^T V^k \bar{d}^k_{ir}(\Delta^k) = o(\theta(x^k)). \] (61)
Equations (58) and (61) show that the numerator of \( t_k(\Delta) \) is \( o(\theta(x^k)) \).
We estimate the denominator of \( t_k(\Delta^k) \) and get
\[
\begin{align*}
& \| V^k [\bar{d}^k_G(\Delta^k) - \bar{d}^k_{ir}(\Delta^k)] \|^2 \\
= & \| V^k \bar{d}^k_{ir}(\Delta^k) \|^2 - 2[V^k \bar{d}^k_{ir}(\Delta^k)]^T [V^k \bar{d}^k_G(\Delta^k)] + \| V^k \bar{d}^k_G(\Delta^k) \|^2 \\
\geq & - \| V^k (x^k - x^*) + o(\theta(x^k)^{1/2}) \|^2 - 2[ - V^k (x^k - x^*) + o(\theta(x^k)^{1/2})]^T [V^k \bar{d}^k_G(\Delta^k)] \\
= & \| - H(x^k) + o(\theta(x^k)^{1/2}) \|^2 - 2( - H(x^k) + o(\theta(x^k)^{1/2})]^T [V^k \bar{d}^k_G(\Delta^k)] \\
= & 2\theta(x^k) + o(\theta(x^k)) - 2[ - \nabla \theta(x^k) + o(\theta(x^k)^{1/2})]^T \bar{d}^k_G(\Delta^k) \\
\geq & 2\theta(x^k) + o(\theta(x^k)) - 2[\| \nabla \theta(x^k) \| \| \bar{d}^k_G(\Delta^k) \| + o(\theta(x^k)^{1/2})\| \bar{d}^k_G(\Delta^k) \|] \\
\geq & 2\theta(x^k) + o(\theta(x^k)) - 2(\Delta^k/\Delta_{max}) \gamma_k \| \nabla \theta(x^k) \|^2 + o(\theta(x^k)^{1/2})\Delta^k \gamma_k \| \nabla \theta(x^k) \| \\
\geq & 2(1 - \eta)\theta(x^k) + o(\theta(x^k)),
\end{align*}
\] (62)
where the second equality follows from (51), the third inequality follows from (60), and the last inequality follows from (17). The above discussion has shown that, if \( \bar{d}^k_G(\Delta^k) \neq \bar{d}^k_{ir}(\Delta^k) \), then
\[ t_k(\Delta^k) \leq o(\theta(x^k))/[2(1 - \eta)\theta(x^k) + o(\theta(x^k))] = o(1). \]
Consequently, from Lemma 2.1 we get that
\[ t_k^*(\Delta^k) \leq o(1) \]
and that
\[
\tilde{d}^k(\Delta^k) = t_k^*(\Delta^k)\tilde{d}_G^k(\Delta^k) + (1 - t_k^*(\Delta^k))\tilde{d}_r^k \\
= \tilde{d}_r^k(\Delta^k) + o(\theta(x^k)^{1/2}) \\
= -(x^k - x^*) + o(\theta(x^k)^{1/2}).
\]
For the case
\[ \tilde{d}_G^k(\Delta^k) = \tilde{d}_r^k(\Delta^k), \]
it is clear that
\[
\tilde{d}^k(\Delta^k) = \tilde{d}_r^k(\Delta^k) \\
= -(x^k - x^*) + o(\theta(x^k)^{1/2}).
\] (63)
Therefore, (54) holds. We turn to verifying (55). It follows from (54) and Lemma 5.5 that
\[
\|\tilde{d}^k(\Delta^k) - d_r^k(\Delta^k)\| = o(\theta(x^k)^{1/2}) \\
= o(\|x^k - x^*\|),
\]
\[
\|d_r^k(\Delta^k)\| = \|x^k - x^*\| + o(\theta(x^k)^{1/2}) \\
= \|x^k - x^*\| + o(\|x^k - x^*\|).
\]
Thus, we get (55). If \( H \) is strongly semismooth at \( x^* \), similar to above arguments, it is not difficult to deduce (56) and (57).

**Lemma 5.8.** The trial step \( \tilde{d}^k(\Delta^k) \) is accepted for all \( k \in K \) sufficiently large.

**Proof.** It follows from Lemmas 5.7 and 5.1 that
\[
\theta(x^k) - (1/2)\|H(x^k) + V^k \tilde{d}^k(\Delta^k)\|^2 \\
= \theta(x^k) - (1/2)\|H(x^k) - V^k(x^k - x^*) + o(\theta(x^k)^{1/2})\|^2 \\
= \theta(x^k) - o(\theta(x^k)).
\] (64)
On the other hand, we have
\[
- \nabla \theta(x^k)^T \tilde{d}_G^k(\Delta^k) \leq \|\nabla \theta(x^k)\| \|\tilde{d}_G^k(\Delta^k)\| \\
\leq \|\nabla \theta(x^k)\|^2 \gamma_k \leq \eta \theta(x^k) < \theta(x^k),
\] (65)
where the third inequality is due to the choice of $g_k$ in Algorithm 3.1. Inequalities (64) and (65) shows that the condition (20) with $\Delta^k$ holds for $k \in K$ sufficiently large.

Next, we prove that $\hat{r}_k \geq \rho_1$. We rewrite $\hat{r}_k$ as

$$
\hat{r}_k = 1 + \frac{\theta(x^k + \tilde{a}^k(\Delta^k)) - 1/2\|H(x^k) + V^k\tilde{a}^k(\Delta^k)\|^2}{(1/2)\|H(x^k) + V^k\tilde{a}^k(\Delta^k)\|^2 - \theta(x^k)}.
$$

From Lemma 5.7, it follows that

$$
(1/2)\|H(x^k) + V^k\tilde{a}^k(\Delta^k)\|^2 - \theta(x^k)
= (1/2)\|H(x^k) - V^k(x^k - x^*) + o(\theta(x^k)^{1/2})\|^2 - \theta(x^k)
= o(\theta(x^k)) - \theta(x^k)
$$

and that

$$
\theta(x^k + \tilde{a}^k(\Delta^k)) - (1/2)\|H(x^k) + V^k\tilde{a}^k(\Delta^k)\|^2
= (1/2)\|H(x^k + \tilde{a}^k(\Delta^k))\|^2 - (1/2)\|H(x^k) - V^k(x^k - x^*) + o(\theta(x^k)^{1/2})\|^2
= (1/2)\|H(x^* + o(\theta(x^k)^{1/2}) - H(x^*))\|^2 + o(\theta(x^k))
\leq (1/2)o(\theta(x^k)) + o(\theta(x^k)) = o(\theta(x^k)).
$$

Equality (66) and inequality (67) imply that, when $k \in K$ is large enough,

$$
\hat{r}_k \geq 1 + o(\theta(x^k))/[o(\theta(x^k)) - \theta(x^k)] \geq \rho_1.
$$

In other words, when $k \in K$ is sufficiently large, the trial step $\tilde{a}^k(\Delta^k)$ is accepted.

The following theorem establishes the superlinear/quadratic convergence of Algorithm 3.1

**Theorem 5.1.** Let $\{x^k\}$ be generated by Algorithm 3.1. Suppose that $x^*$ is an accumulation point of $\{x^k\}$ and a BD-regular solution of (1). Then, the whole sequence $\{x^k\}$ converges to $x^*$ superlinearly. In addition, if $H$ is strongly semismooth at $x^*$, the convergence rate is quadratic.

**Proof.** It follows from Lemmas 5.7 and 5.8 that, for $k \in K$ large enough,

$$
\|x^{k+1} - x^*\| = \|x^k + \tilde{a}^k(\Delta^k) - x^*\|
= o(\theta(x^k)^{1/2})
= o(\|x^k - x^*\|).
$$

(68)
This shows that \( \{x^k\}_{k \in K} \) converges to \( x^* \) superlinearly. On the other hand, the condition that \( H \) is BD-regular at \( x^* \) implies that \( x^* \) is an isolated solution of (1) and hence an isolated limit point of \( \{x^k\} \). We get also from equality (54) that

\[
\|x^{k+1} - x^k\| = \|\tilde{d}^k(\Delta_k)\| \to 0, \quad \text{as } k \to \infty \text{ with } k \in K.
\]

It is not difficult to prove that the whole sequence \( \{x^k\} \) converges to \( x^* \). Consequently, (68) holds for all \( k \) sufficiently large, which shows the superlinear convergence of \( \{x^k\} \). If \( H \) is strongly semismooth at \( x^* \), then the term \( o(\theta(x^k)^{1/2}) \) can be replaced by \( O(\theta(x^k)) \) and \( o(\|x^k - x^*\|) \) can be replaced by \( O(\|x^k - x^*\|^2) \), which implies the quadratic convergence of \( \{x^k\} \).

\[\square\]

6. Preliminary Numerical Experiments

In this section, we present some preliminary numerical results for Algorithm 3.1. The problems are extracted from semismooth reformulations of box-constrained variational inequality problems: find \( x \in X \) such that

\[ F(x)^T(y - x) \geq 0, \quad \forall y \in X. \]

The parameters used in Algorithm 3.1 are specified as follows:

\[ \alpha_1 = 0.5, \quad \alpha_2 = 2, \quad \rho_1 = 0.00001, \quad \rho_2 = 0.75, \quad \eta = 0.9, \quad \sigma = 0.5, \quad \Delta_0 = 5, \quad \Delta_{\text{min}} = 0.00001, \quad \Delta_{\text{max}} = 10, \quad \delta = 0.00001, \quad c = 1. \]

We use

\[ \min\{\theta(x^k), \|\nabla \theta(x^k)\|\} \leq 1.0^{-10} \]

as the stopping criterion. The trust-region subproblem (10) is solved by a truncated preconditioned conjugate gradient method (see Refs. 20, 21). The testing problems are listed as follows.

**Example 6.1.** This is linear complementarity problem from Geiger and Kanzow (Ref. 22) and Jiang and Qi (Ref. 23); see the first example in Ref. 23. Starting point: \( x^0 = 0 \).

**Example 6.2.** This is linear complementarity problem coming from Geiger and Kanzow (Ref. 22) and Jiang and Qi (Ref. 23). Starting point: \( x^0 = 0 \).
Example 6.3. This is the Kojima-Shindo problem; see Example 3 in Ref. 23. Starting point: \(a = (0, 0, 0, 0), \ b = (1, 1, 1, 1), \ c = (1, 2, 3, 4)\).

Example 6.4. This is nonlinear complementarity problem from Jiang and Qi (Ref. 23). Starting point: \(a = (0, 0, 0, 0), \ b = (1, 1, 1, 1), \ c = (2, 2, 2, 2)\).

Example 6.5. This is a modification of the Mathiesen problem from Jiang and Qi Ref. 23; see Example 5 in Ref. 23. Starting point: \(a = (1, 1, 1, 1), \ b = (2, 2, 2, 2), \ c = (0, 10, 1, 2)\).

Example 6.6. This is a Hock-Schittkowski problem (Ref. 24, Problem 35). Starting point: \(a = (0, 0, 0, 0), \ b = (1, 1, 1, 1), \ c = (100, 100, 100, 100)\).

Example 6.7. This is a Hock-Schittkowski problem (Ref. 24, Problem 76). Starting point: \(a = (0, 0, 0, 0, 0, 0, 0), \ b = (1, 1, 1, 1, 1, 1, 1), \ c = (0, 1, 2, 3, 4, 5, 6)\).

Example 6.8. This is a problem from Ralph and Wright (Ref. 25, Problem 3), which is defined by
\[
\begin{align*}
\min & \quad x_1^2 + x_1 x_2 + 2x_2^2 + x_1 + x_2, \\
\text{s.t} & \quad (1/2)(x_1 - 2)^2 + (1/2)(x_2 - 1)^2 \leq 5/2, \\
& \quad x_1 \geq 0, \quad x_2 \geq 0.
\end{align*}
\]
We reformulate its KKT system to a nonlinear complementarity problem with three variables. Starting point: \(a = (1, 1, 1), \ b = (1, 2, 3), \ c = (10, 10, 10)\).

Example 6.9. This is a problem from Chen and Ye (Ref. 26, Problem 3). Starting point: \(x^0 = 0\).

Example 6.10. This is a problem from Chen and Ye (Ref. 26, Problem 4), which is a linear complementarity problem involving the matrix
\[
M = \begin{pmatrix} P + D_1 & P + D_2 \\ -I & 0 \end{pmatrix}, \quad q = -\begin{pmatrix} e \\ 0 \end{pmatrix},
\]
where \(e\) is a vector whose elements are 1 and \(I\) is the unit matrix. Here, \(P = A^T A, \ A \in R^{(n/2) \times (n/2)}\), with \(0 < a_{ij} < 0\), \(D_1\) and \(D_2\) are diagonal matrices with \(1 \leq (D_1)_{ii}, (D_2)_{ii} \leq 3\). Starting point: \(x^0 = 0\).
The tests were done on a P-III 800 with MATLAB code. The numerical results are reported in Table 1. In the table, Dim denotes the number of the variables of the problem; Iter stands for the number of iterations and NF stands for the number of function evaluations. The number of trust-region subproblems is $NF - 1$. Error denotes the value of $\min\{\theta(x^k), \|\nabla \theta(x^k)\|\}$ at the final iteration, $r^*(\Delta_{ave})$ denotes the average of all $r^*(\Delta_k)$. If $r^*(\Delta_{ave})$ is close to zero, then the projected trust-region direction is used most of time, while if it is close to one then the projected gradient direction is used most of the time.

The results in Table 1 are promising. Moreover, for the tested optimization problems (Examples 6.6, 6.7, 6.8), the algorithm terminates at their optimal solutions. The average values of $r^*(\Delta_k)$ is generally not too large.

<table>
<thead>
<tr>
<th>Example</th>
<th>Dim.</th>
<th>Starting point</th>
<th>Iter</th>
<th>NF</th>
<th>Error</th>
<th>$r^*(\Delta_{ave})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>a</td>
<td>5</td>
<td>6</td>
<td>3.3534e-11</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>a</td>
<td>6</td>
<td>7</td>
<td>7.5511e-19</td>
<td>0.1667</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>a</td>
<td>6</td>
<td>7</td>
<td>7.5943e-19</td>
<td>0.1667</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>a</td>
<td>7</td>
<td>8</td>
<td>5.2864e-16</td>
<td>0.2613</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>a</td>
<td>14</td>
<td>15</td>
<td>6.3630e-13</td>
<td>0.4216</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>a</td>
<td>18</td>
<td>19</td>
<td>7.0740e-14</td>
<td>0.5564</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>a</td>
<td>14</td>
<td>22</td>
<td>2.7270e-14</td>
<td>0.2838</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>b</td>
<td>59</td>
<td>90</td>
<td>6.3627e-11</td>
<td>0.7400</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>c</td>
<td>28</td>
<td>33</td>
<td>5.9023e-11</td>
<td>0.6703</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>a</td>
<td>22</td>
<td>23</td>
<td>2.4551e-18</td>
<td>0.0455</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>b</td>
<td>16</td>
<td>17</td>
<td>2.4266e-18</td>
<td>7.2909e-13</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>c</td>
<td>19</td>
<td>20</td>
<td>2.3894e-14</td>
<td>7.2349e-11</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>a</td>
<td>2</td>
<td>3</td>
<td>0.0</td>
<td>1.1450e-17</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>b</td>
<td>2</td>
<td>3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>c</td>
<td>4</td>
<td>5</td>
<td>0.0</td>
<td>0.1755</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>a</td>
<td>6</td>
<td>7</td>
<td>8.3996e-12</td>
<td>0.1667</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>b</td>
<td>21</td>
<td>30</td>
<td>3.9565e-12</td>
<td>0.3966</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>c</td>
<td>57</td>
<td>58</td>
<td>6.4993e-12</td>
<td>0.6161</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>a</td>
<td>64</td>
<td>65</td>
<td>7.4113e-11</td>
<td>0.9805</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>b</td>
<td>84</td>
<td>121</td>
<td>9.9944e-11</td>
<td>0.9510</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>c</td>
<td>51</td>
<td>73</td>
<td>9.4331e-11</td>
<td>0.8817</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>a</td>
<td>2</td>
<td>3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>b</td>
<td>4</td>
<td>5</td>
<td>0.0</td>
<td>0.2884</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>c</td>
<td>21</td>
<td>26</td>
<td>0.0</td>
<td>0.5554</td>
</tr>
<tr>
<td>9</td>
<td>100</td>
<td>a</td>
<td>3</td>
<td>4</td>
<td>0.0</td>
<td>0.3851</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>a</td>
<td>3</td>
<td>4</td>
<td>7.7299e-13</td>
<td>0.3369</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>a</td>
<td>18</td>
<td>19</td>
<td>0.0</td>
<td>0.8333</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>a</td>
<td>8</td>
<td>9</td>
<td>1.6147e-18</td>
<td>0.4213</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>a</td>
<td>12</td>
<td>29</td>
<td>7.9478e-15</td>
<td>0.5007</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>a</td>
<td>24</td>
<td>44</td>
<td>4.7614e-15</td>
<td>0.8356</td>
</tr>
</tbody>
</table>
This shows that the trust-region direction was used very often and that the iterates converge fast. However, in the case that $t^*(\Delta_k)$ is close to 1 (e.g. in Example 6.7), the iterative numbers became large, because the projected gradient direction was used too many times. This feature shows the proposed method is computationally better than the projected gradient method.

References
