Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Computational and Applied Mathematics 187 (2006) 227–252
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
www.elsevier.com/locate/cam

Affine scaling inexact generalized Newton algorithm with interior backtracking technique for solving bound-constrained semismooth equations☆

Detong Zhu

Department of Mathematics, Shanghai Normal University, Shanghai 200234, P.R. China

Received 15 August 2004

Abstract

We develop and analyze an affine scaling inexact generalized Newton algorithm in association with nonmonotone interior backtracking line technique for solving systems of semismooth equations subject to bounds on variables. By combining inexact affine scaling generalized Newton with interior backtracking line search technique, each iterate switches to inexact generalized Newton backtracking step to strict interior point feasibility. The global convergence results are developed in a very general setting of computing trial steps by the affine scaling generalized Newton-like method that is augmented by an interior backtracking line search technique projection onto the feasible set. Under some reasonable conditions we establish that close to a regular solution the inexact generalized Newton method is shown to converge locally p-order q-superlinearly. We characterize the order of local convergence based on convergence behavior of the quality of the approximate subdifferentials and indicate how to choose an inexact forcing sequence which preserves the rapid convergence of the proposed algorithm. A nonmonotonic criterion should bring about speeding up the convergence progress in some ill-conditioned cases.

MSC: 90C30; 90C33; 49J40; 65H10; 65K05; 49M37

Keywords: Semismooth equation; Inexact generalized Newton method; Global convergence; Affine scaling interior; Superlinear convergence

☆ The author gratefully acknowledges the partial supports of the National Science Foundation Grant of China (10471094).
E-mail address: dtzhu@shtu.edu.cn.

0377-0427/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2005.03.045

1. Introduction

Consider an affine scaling inexact generalized Newton method for solving systems of the nonsmooth equations subject to the bound constraints on variable:

$$H(x)=0, \quad x \in \Omega \stackrel{\text{def}}{=} \{x \mid l \leqslant x \leqslant u\}. \tag{1.1}$$

Hereby, the function $H : \Re^n \supset \mathscr{X} \to \Re^n$ is defined on the open set $\mathscr{X}$ containing the n-dimensional feasible box constraint set $\Omega \stackrel{\text{def}}{=} [l, u] \stackrel{\text{def}}{=} \{x \in \Re^n; l_i \leqslant x_i \leqslant u_i, i=1, \ldots, n\}$. The vector $l \in (\Re \cup \{-\infty\})^n$ and $u \in (\Re \cup \{+\infty\})^n$ are specified lower and upper bounds on the variables such that $\operatorname{int}(\Omega) \stackrel{\text{def}}{=} \{x \mid l < x < u\}$ is nonempty. Nonsmooth systems (1.1) arise naturally in systems of equations modelling real-life problems when not all the solutions of the model have physical meaning. Various sources of nonlinear nonsmooth equations with the box constraint Ω drawn from mixed nonlinear complementarity problems, nonlinear optimization and several related problems have been described, where Ω maybe the positive orthant of $\Re^n$ or a closed box constraint. In the classic methods for solving the nonlinear smooth equations (1.1) when the function $H(x)$ is a continuously differentiable function, the Newton methods or quasi-Newton methods can be used. Computing the exact solution can be expensive if n is large and for any n, may be justified when x is far from a solution. Much analysis of many inexact and exact Newton algorithms have been done on smooth nonlinear equations but on nonsmooth equations based on convergent analysis.

These methods by using the Jacobain or version of generalized Newton's methods often solve the semismooth systems (1.1), which is known to have locally very rapid convergence. Both the approximate generalized Newton methods and inexact Newton methods used for semismooth system (1.1) did not ensure global convergence, that is, the convergence is only local. Therefore, the methods are available only when the initial start point is good enough. The locally q-superlinear/quadratic convergence of the algorithm to BD-regular (where "BD" stands for Bouligand differential) solutions of (1.1) will be achieved by a Newton-type method that is augmented by a projection onto Ω to maintain feasibility. Local convergence results for Newton's method without bound-constraints were established in [12–14]. One effective remedy when this occurs is to restrict the trial step to a region where the linear model can be trusted. Globally convergent methods for the unconstrained systems $H(x)=0$ may be unsuited for the purpose of solving (1.1), since a vector x^* satisfy $H(x)=0$, but does not belong to Ω. Brown and Saad [2] introduced the Euclidean norm, i.e., l_2 norm as the merit function to combine the line search to solving the unconstrained nonlinear smooth systems (1.1) and proved the global convergence of the proposed algorithms. Recently, Ulbrich in [18] presented a class of double trust-region approaches with a projection onto the feasible set for bound-constrained semismooth systems of equations (1.1) and further proved that close to a regular solution the trust-region algorithm turns into this projected Newton method, which is shown to converge locally q-superlinearly or quadratically, respectively, depending on the quality of the approximate subdifferentials used under a Dennis–Moré-type condition [7,8] and by allowing for inexactness in the computation of B-subdifferentials (where "B" stands for Bouligand). However, the search direction generated in trust region subproblem must satisfy strict interior feasibility, i.e., $l < x_k + s < u$, which results in computational difficulties. It is possible that the trust region subproblem with the strict feasibility constraint needs to be resolved many times before obtaining an acceptable step [11], and hence the total computational effort for completing one iteration might be expensive and difficult.

In this paper, we also introduce the following bound-constrained semismooth minimization as the merit function to reformulate the problem (1.1):

$$\min h(x) \stackrel{\text{def}}{=} \tfrac{1}{2}\|H(x)\|^2 \quad \text{s.t. } x \in \Omega. \tag{1.2}$$

Here throughout the paper, $\|\cdot\|$ denotes the Euclidean norm. Obviously, (1.1) and (1.2) are equivalent if problem (1.1) possesses a solution. In the optimization algorithms it is sometimes helpful to include an affine scaling matrix for the variables which is invertible in most cases, called an affine scaling projection. The idea of combining the inexact generalized Newton method and backtracking interior line search technique motivates to introduce the affine scaling matrix of interior point to generate the inexact generalized Newton method which switches to strict interior feasibility by line search backtracking technique. At each iteration, most modern global fit within determining an initial trial step and testing the trial step to determine whether it gives adequate progress toward a solution, the affine scaling interior point should switch to strict interior feasibility by line search backtracking interior technique. For most versions for solving smooth equations, these approaches only rather restrictive guarantees of global convergence have only been based on the line search procedure such as Armijo rule, damped Newton methods [9]. Trial steps are determined in a variety ways to enforce a monotone decrease of the merit function at each step. Nonmonotone technique is developed to combine with, respectively, line search technique and trust region strategy for solving unconstrained optimization in [10,6]. The nonmonotone idea also motivates the study of inexact generalized Newton methods in association with nonmonotone interior backtracking line search technique for approximating zeros of the semismooth equations (1.1). In this paper, we introduce the affine scaling matrix to generate affine scaling inexact generalized Newton method in association with two criterions of nonmonotone backtracking line search and strict interior feasibility accepting step for solving the bound-constrained semismooth equations (1.1).

In order to describe and design the algorithms for solving the semismooth equations (1.1), we first introduce the squared Euclidean norm as the merit function to quadratic model of the systems (1.1) and state the nonmonotone affine scaling inexact generalized Newton algorithm with backtracking interior point technique for solving the semismooth equations in the next section. In Section 3, we prove results of the weak and strong global convergence of the proposed algorithm. We characterize the order of local convergence of the affine scaling inexact generalized Newton method in terms of the rates of the relative residuals based on convergence behavior of affine scaling matrix and indicate how forcing sequence influences the rapid rate of convergence and details of this under some reasonable conditions are in Section 4. Finally, applications of the algorithm to various nonsmooth bound-constrained semismooth equations are discussed and the affine scaling inexact generalized Newton step is implemented by affine scaling trust region approach in Section 5. As will be shown in [18], we rewrite the following notations.

Notations: Given a set $\mathscr{X} \subset \Re^n$. $S^0(\mathscr{X}, \Re^n)$ denotes the set of all semismooth functions $f : \Re^n \supset \mathscr{X} \to \Re^n$.$S^p(\mathscr{X}, \Re^n)$, $0 < p \leqslant 1$, is the set of all p-order semismooth functions. We write $f'(x;\cdot) : \Re^n \to \Re^m$ for the directional derivative, $\nabla f(x) \in \Re^{m\times n}$ for the Jacobian, $\partial_{\mathrm{B}} f(x) \subset \Re^{m\times n}$ for the B-subdifferential, and $\partial f(x) \subset \Re^{m\times n}$ for Clarke's generalized Jacobian of the function $f : \Re^n \supset \mathscr{X} \to \Re^m$ at the point $x \in \mathscr{X}$ (in case the respective objects exist). $\nabla f(x)$ denotes the gradient of the differentiable, real-valued function f at x.

2. Algorithm

In this section, we describe and design our affine scaling inexact generalized Newton method in association with nonmonotone interior backtracking technique for approximating a solution of the bound-constrained nonlinear semismooth minimization (1.2) transformed by the bound-constrained semismooth equations (1.1) and present an interior point backtracking technique which enforces the variable generating strictly feasible interior point approximations to the solution. Nonsmooth analysis is an essential tool for the design of effective numerical methods for solving the bound-constrained semismooth equations (1.1) and for the development of the supporting convergence theory. In this section, for convenience, we collect first concepts about nonsmooth analysis and we first assume that the function H to be considered is locally Lipschizian. A function $H : \mathscr{X} \subset \Re^n \to \Re^n$ is said to be B-differentiable at a point x if it is directional differentiable at x and

$$\lim_{d\to 0} \frac{H(x+d) - H(x) - H'(x;d)}{\|d\|} = 0. \tag{2.1}$$

We may write the equation as

$$H(x+d) = H(x) + H'(x;d) + \mathrm{o}(\|d\|).$$

In a finite-dimensional Euclidean space $\Re^n$, Shapiro [15] showed that a locally Lipschizian function H is B-differentiable at x if and only if it is directional differentiable at x. For such function H is locally Lipschizian, Rademacher's theorem implies that H is almost everywhere F-differentiable. Let the set of points where H is F-differentiable be denoted D_H. Then for any $x \in \Re^n$ the generalized subdifferential of H at x in the sense of Clarke [3] is

$$\partial H(x) = \mathrm{conv}\ \{\lim \nabla H(x_j) : x_j \to x, x_j \in D_H\}, \tag{2.2}$$

which is a nonempty convex compact set. Considered as a set-valued mapping, ∂H is locally bounded and upper semicontinuous. We call $\partial_{\mathrm{B}} H(x)$ the B-subdifferential of H at x. This concept was introduced in [12], where an explanation was also given for its introduction.

For $x,\ d \in \Re^n$ with $d \neq 0$ we say that y tends to x in the direction d, denoted by $y \to_d x$, if $y \to x,\ y \neq x$, and $\frac{y-x}{\|y-x\|} \to \frac{d}{\|d\|}$. We say that H is semismooth at x if H is locally Lipschizian there and if any $d \in \Re^n$ with $d \neq 0$

$$\lim_{y\to_d x} \{Vd \mid V \in \partial H(d)\} \tag{2.3}$$

exists. If H is semismooth at x, then H must be directionally differentiable (B-differentiable) at x and $H'(x;d)$ is equal to the above limit for any $d \neq 0$. If H is semismooth at all points in a given set, we say that H is semismooth in this set. In [14], Qi and Sun gave the following two lemmas.

Lemma 2.1. *Suppose that $H : \mathscr{X} \subseteq \Re^n \to \Re^m$ is directionally differentiable at a neighborhood of x. Then*

(1) *$H'(x;\cdot)$ is Lipschizian*;
(2) *for any d, there exists a $V \in \partial H(x)$ such that $H'(x;d) = Vd$.*

Suppose further that H is B-differentiable at a neighborhood of x. We say that the directional derivative $H'(\cdot;\cdot)$ is semicontinuous at x if, for every $\varepsilon > 0$, there exists a neighborhood $\mathcal{N}$ of x such that for all $x + d \in \mathcal{N}$, $\|H'(x+d;d) - H'(x;d)\| \leqslant \varepsilon \|d\|$.

Lemma 2.2. *Suppose that $H : \Re^n \to \Re^n$ is directionally differentiable at a neighborhood of x. The following statements are equivalent*:

(1) *H is semismooth at x*;
(2) *$H'(\cdot;\cdot)$ is semicontinuous at x*;
(3) *for any $V \in \partial H(x+d)$, $d \to 0$, $Vd - H'(x;d) = \mathrm{o}(\|d\|)$*;
(4)

$$\lim_{x+d \in D_H, d \to 0} \frac{H'(x+d;d) - H'(x;d)}{\|d\|} = 0.$$

If for any $V \in \partial H(x+d)$, $d \to 0$,

$$Vd - H'(x;d) = \mathrm{O}(\|d\|^{1+p}), \tag{2.4}$$

where $0 < p \leqslant 1$, then we call H p-order semismooth at x. It is clear that p-order semismoothness $(0 < p \leqslant 1)$ implies semismoothness.

In [3], Clarke gave that for any $x, \ y \ \in \Re^n$,

$$H(y) - H(x) \in \operatorname{conv} \partial H([x,y])(y-x), \tag{2.5}$$

where the right-hand side denotes the convex hull of all points of form $V(y-x)$ with $V \in \partial H(u)$ for some point u in $[x, y]$. By Lemma 2.2(3) and (2.5), if H is semismooth at x, then for any $d \to 0$,

$$H(x+d) - H(x) - H'(x;d) = \mathrm{o}(\|d\|). \tag{2.6}$$

Based on the semismooth relaxation of Hölder-continuous differentiability can be established. Let $0 < p \leqslant 1$. The H is p-order semismooth at $x \in \mathcal{X}$ if H is locally Lipschitz at x, $H'(x;\cdot)$ exists, then for any $d \to 0$,

$$H(x+d) - H(x) - H'(x;d) = \mathrm{O}(\|d\|^{1+p}). \tag{2.7}$$

It is known (see [12, Proposition 1]) that semismoothness of H at x implies that

$$\sup_{V \in \partial H(x+d)} \{H(x+d) - H(x) - Vd\} = \mathrm{o}(\|d\|). \tag{2.8}$$

In the strongly p-order semismooth case at x, one has that

$$\sup_{V \in \partial H(x+d)} \{H(x+d) - H(x) - Vd\} = \mathrm{O}(\|d\|^{1+p}). \tag{2.9}$$

By $S^p(\mathcal{X}, \Re^m)$ we denote the set of all functions $H : \mathcal{X} \to \Re^m$ that are p-order semismooth on $\mathcal{X}$. The following is obvious.

Lemma 2.3. *If H is continuously differentiable in a neighborhood of $x \in \mathcal{X}$* (*with p-Hölder continuous derivative*, $0 < p \leqslant 1$), *then f is* (*p-order*) *semismooth at x and $\partial H(x) = \partial_{\mathrm{B}} H(x) = \{\nabla H(x)\}$.*

The following regularity property is essential for fast local convergence of generalized Newton-like methods. We say that H is BD-regular at $x \in \mathscr{X}$ if all the elements in $\partial_{\mathrm{B}} H(x)$ are $n \times n$ nonsingular matrices.

Lemma 2.4 (*see Pang and Qi [12, Proposition 3]*). *Let $x \in \mathscr{X}$ be a BD-regular for H. Then there exist $\varepsilon > 0$ and $\chi > 0$ such that all $V \in \partial_{\mathrm{B}} H(y)$, $\|y-x\| \leqslant \varepsilon$ are nonsingular with $\|V^{-1}\| \leqslant \chi$. If, in addition, H is semismooth at x, then there exist $\delta > 0$ and $\zeta > 0$ such that*

$$\|H(y) - H(x)\| \geqslant \zeta \|y - x\| \tag{2.10}$$

for all $y \in \Re^n$, $\|y - x\| \leqslant \delta$.

As motivated above, a classical algorithm for solving the problem (1.1) will be based on the reformulation (1.2). Optimality conditions refer to the continuous differentiability of the merit function h which was established by Ulbrich and found from Lemma 4.2 in [18].

Lemma 2.5. *Assume that the function $H : \Re^n \supset \mathscr{X} \to \Re^n$ is semismooth or, stronger, p-order semismooth, $0 < p \leqslant 1$, then the merit function $\frac{1}{2}\|H(x)\|^2$ is continuously differentiable on $\mathscr{X}$ with gradient $\nabla h(x) = V^{\mathrm{T}} H(x)$, where $V \in \partial H(x)$ is arbitrary.*

As motivated above, a classical algorithm for solving the problem (1.1) will be based on the reformulation (1.2). Ignoring primal and dual feasibility of the systems (1.2), the first-order necessary conditions for x^* to be a local minimizer and

$$\begin{aligned} &(g_*)_i = 0 \quad \text{if } l_i < (x^*)_i < u_i, \\ &(g_*)_i \geqslant 0 \quad \text{if } (x^*)_i = l_i, \\ &(g_*)_i \leqslant 0 \quad \text{if } (x^*)_i = u_i, \end{aligned}$$

where $g(x) \stackrel{\text{def}}{=} \nabla h(x) = V^{\mathrm{T}} H(x)$, with arbitrary $V \in \partial H(x)$, and $(g_*)_i$, $(x^*)_i$ are the ith component of g_* and x^*, respectively. The scaling matrix $D_k = D(x_k)$ arises naturally from examining the first-order necessary conditions for the bound-constrained semismooth minimization (1.2) transformed by the bound-constrained problem (1.1), where $D(x)$ is the diagonal scaling matrix such that

$$D(x) \stackrel{\text{def}}{=} \operatorname{diag}\{|v_1(x)|^{-1/2}, \ldots, |v_n(x)|^{-1/2}\} \tag{2.11}$$

and the ith component of vector $v(x)$ defined componentwise as follows

$$v_i(x) \stackrel{\text{def}}{=} \begin{cases} x_i - u_i & \text{if } g_i < 0 \quad \text{and} \quad u_i < +\infty, \\ x_i - l_i & \text{if } g_i \geqslant 0 \quad \text{and} \quad l_i > -\infty, \\ -1 & \text{if } g_i < 0 \quad \text{and} \quad u_i = +\infty, \\ 1 & \text{if } g_i \geqslant 0 \quad \text{and} \quad l_i = -\infty \end{cases} \tag{2.12}$$

here g_i is the ith component of vector $g(x)$. We remark that, even though $D(x)$ may be undefined on the boundary of Ω, $D(x)^{-1}$ can be extended continuously to it. We will denote this extension as a convention by $D(x)^{-1}$ for all $x \in \Omega$.

Definition 2.5 (*see Coleman and Li [4]*). A point $x \in \Omega$ is nondegenerate if, for each index i,

$$g_i(x) = 0 \Longrightarrow l_i < x_i < u_i. \tag{2.13}$$

The problem (1.2) is nondegenerate if (2.13) holds for every $x \in \Omega$.

A classical algorithm for solving the semismooth equations (1.1) is exact generalized Newton-like method when Ω is $\Re^n$, i.e., unconstrained semismooth equations as follows.

2.1. The exact generalized Newton method

Let x_0 be given. For $k=0$ step 1 until convergence do $(+\infty)$:
Find the step s_k which satisfies

$$M_k s_k = -H_k, \tag{2.14}$$

where M_k is an approximation to $V_k \in \partial H(x_k)$ or $M_k = V_k \in \partial H(x_k)$ (see [7,8]).
Set $x_{k+1} = x + s_k$.

The process is exact generalized Newton method if $M_k = V_k \in \partial H(x_k)$, and it represents an exact generalized Newton-like method if M_k is an approximation to $V_k \in \partial H(x_k)$. Computing the exact solution can be expensive if n is large and may not be justified when x_k is far from a solution. Thus, in [5] iterative processes of the following general form were led to the following algorithm, called inexact generalized Newton method:

2.2. The inexact generalized Newton method

Let x_0 be given. For $k=0$ step 1 until convergence do $(+\infty)$:
Find some step s_k which satisfies

$$\|H(x_k) + V_k s_k\| \leqslant \eta_k \|H(x_k)\|. \tag{2.15}$$

Set $x_{k+1} = x + s_k$. Here $V_k \in \partial H(x_k)$ and $\eta_k \in [0, 1)$ is a sequence of forcing terms.

Generally, the global convergence of the methods is obtained by augmenting the inexact generalized Newton condition with a sufficient monotone decrease condition on the merit function $h(x) \stackrel{\text{def}}{=} \frac{1}{2}\|H(x)\|_2^2$ introduced by Brown and Saad in [2] for solving the unconstrained smooth equations (1.1) wherein step length λ_k of s_k in (2.15) by using the Armijo–Goldstein rule, must satisfy

$$h(x_k + \lambda_k s_k) \leqslant h(x_k) + \beta \lambda_k \nabla h(x_k)^{\mathrm{T}} s_k, \tag{2.16}$$

where $\beta \in (0, 1)$ and $\nabla h(x_k) = V_k^{\mathrm{T}} H(x_k)$, $V_k = \nabla H(x_k)$ when H is continuously differentiable in a neighborhood of x_k.

Recently, Bellavia et al. in [1] presented the affine scaling double trust-region approach scheme for solving the constrained smooth nonlinear equations (1.1). The basic idea is based on the trust region subproblem

$$\begin{aligned} &\min \psi_k(d) \stackrel{\text{def}}{=} \tfrac{1}{2}\|\nabla H_k d + H_k\|^2 = \tfrac{1}{2}\|H_k\|^2 + H_k^{\mathrm{T}} \nabla H_k d + \tfrac{1}{2} d^{\mathrm{T}} (\nabla H_k^{\mathrm{T}} \nabla H_k) d \\ &\quad \text{s.t.} \|D_k d\| \leqslant \Delta_k, \end{aligned} \tag{2.17}$$

where Δ_k is the trust region radius, and $\psi_k(d)$ is trusted to be an adequate representation of the merit function $h(x) \stackrel{\text{def}}{=} \frac{1}{2}\|H(x)\|^2$.

In this paper, we set also the Euclidean norm as the merit function $h(x) \stackrel{\text{def}}{=} \frac{1}{2}\|H(x)\|_2^2$ with the scaling invertible matrix $D(x)$ for x where set $g(x) \stackrel{\text{def}}{=} \nabla h(x) = V^{\mathrm{T}} H(x)$, and $V \in \partial H(x)$ since Lemma 2.5 holds. In kth iteration of our algorithm, relaxing the acceptability conditions on the trial step s_k, we suggest to use the nonmonotone technique instead of monotone technique. The idea of affine scaling transformation motivates the study of inexact generalized Newton methods in association with nonmonotone interior backtracking line search technique for approximating zeros of the bounded-constrained semismooth equations (1.1).

Now, we describe an affine scaling inexact generalized Newton algorithm with nonmonotonic strick interior feasible backtracking line search technique for solving semismooth systems (1.1).

2.3. *Algorithm (nonmonotone backtracking affine scaling inexact generalized Newton algorithm)*

Given positive integer M, $m(0) = 0$ and the starting point $x_0 \in \mathrm{int}(\Omega)$.

For $k = 0$ step 1 until ∞ (if $\|D_k^{-1} g_k\| = \|D_k^{-1}(V_k^{\mathrm{T}} H_k)\| \leqslant \varepsilon$, where ε is a given suitable small quantity and $V_k \in \partial H(x_k)$) stop with the approximate solution x_k do:

Given D_k an invertible matrix as affine scaling transformation for each k. Given $\omega \in (0, 1)$ and $\beta \in (0, \frac{1}{2})$. Find some $\eta_k \in (0, 1)$, D_k an invertible matrix as affine scaling transformation for each k and $\widehat{s_k}$ that satisfy

$$\|H(x_k) + V_k D_k^{-1} \widehat{s_k}\|_2 \leqslant \eta_k \|H(x_k)\|_2 \tag{2.18}$$

and set

$$s_k = D_k^{-1} \widehat{s_k} \tag{2.19}$$

and then choose $\lambda_k = 1, \omega, \omega^2, \ldots$ until the following inequality is satisfied

$$h(x_k + \lambda_k s_k) \leqslant h(x_{l(k)}) + \lambda_k \beta \nabla h(x_k)^{\mathrm{T}} s_k, \tag{2.20}$$

$$\text{with } x_k + \lambda_k s_k \in \Omega, \tag{2.21}$$

where $\nabla h(x_k) = V_k^{\mathrm{T}} H(x_k)$ with $V_k \in \partial H(x_k)$, and $h(x_{l(k)}) = \max_{0 \leqslant j \leqslant m(k)} \{h(x_{k-j})\}$, with the nonmonotone index function $0 \leqslant m(k) \leqslant \min\{m(k-1)+1, M\}$, $k \geqslant 1$.

Further, set

$$p_k \stackrel{\text{def}}{=} \begin{cases} \lambda_k s_k & \text{if } x_k + \lambda_k s_k \in \mathrm{int}(\Omega), \\ \theta_k \lambda_k s_k & \text{otherwise}, \end{cases}$$

where $\theta_k \in (\theta_l, 1]$, for some $0 < \theta_l < 1$ and $\theta_k - 1 = \mathrm{O}(\|s_k\|)$ and then set

$$x_{k+1} = x_k + p_k. \tag{2.22}$$

Remark 1. V_k can be updated by updating formula for Clarke's generalized subdifferential and other updating generalized subdifferential approaching formula which can avoid to any computing of the generalized Jacobian of H at point x_k. When $D_k = I$ and $\Omega = \Re^n$, the affine scaling inexact generalized Newton-like algorithm is the usual inexact generalized Newton method for solving unconstrained smooth equations (1.1) if $H(x)$ is smooth. It is easy to see that the global affine scaling inexact Newton-like algorithm with the nonmonotone technique is the usual global affine scaling inexact quasi-Newton method when $M = 0$. It is clear to see that $k - m(k) \leqslant l(k) \leqslant k$.

Remark 2. The scalar λ_k given in (2.21), denotes the step size along the direction s_k to the boundary on the variables $l \leqslant x_k + \alpha_k s_k \leqslant u$, that is,

$$\Lambda_k \stackrel{\text{def}}{=} \min \left\{ \max \left\{ \frac{l_i - (x_k)_i}{(s_k)_i}, \frac{u_i - (x_k)_i}{(s_k)_i} \right\}, \ i = 1, 2, \ldots, n \right\}, \tag{2.23}$$

where l_i, u_i, $(x_k)_i$ and $(s_k)_i$ are the ith components of l, u, x_k and s_k, respectively.

3. Convergence analysis

Throughout this section we assume that $H : \mathscr{X} \subset \Re^n \to \Re^n$ is semismooth, or stronger, p-order semismooth. Given $x_0 \in \Omega \subset \Re^n$, the algorithm generates a sequence $\{x_k\} \subset \Omega \subset \Re^n$. In our analysis, we denote the level set of h by

$$\mathscr{L}(x_0) = \{x \in \Re^n \mid h(x) \leqslant h(x_0), l \leqslant x \leqslant u\}.$$

We first require the definition of semismoothness as follows.

Assumption A1. Sequence $\{x_k\} \subset \Omega$ generated by the algorithm is contained in a compact set $\mathscr{L}(x_0)$ on $\Re^n$.

Assumption A2. The function $H : \Re^n \supset \mathscr{X} \to \Re^n$ is semismooth or, stronger, p-order semismooth, $0 < p \leqslant 1$.

Assumption A3. Each component function H_i of H is continuously differentiable on $\mathscr{X}/H_i^{-1}(0)$.

In order to discuss the global convergence property of the proposed algorithm, we make the following assumptions in this section which are commonly used in convergence analysis of most methods for solving the bound-constrained semismooth minimization (1.2) reformulated by the bound-constrained systems (1.1). As will be shown in [18], the above assumptions are not based directly on Assumption A3 but on the—in connection with Assumption A3—weaker assumption.

Assumption A3′. The function $h : \mathscr{X} \to \Re$, $h(x) \stackrel{\text{def}}{=} \frac{1}{2}\|H(x)\|^2$ is continuously differentiable.

In particular, Assumption A3 is more concrete and easier to verify than Assumption A3′ and thus decide to choose Assumptions A1–A3.

Assumption A4. There exist some positive constants χ_V and χ_D such that

$$\|V\| \leqslant \chi_V, \quad \forall V \in \partial H(x), \quad \|D(x)^{-1}\| \leqslant \chi_D, \quad \text{for all } x \in \mathscr{L}(x_0).$$

In kth iteration, to guarantee that the current iterate will make progress towards the solution in one step of the proposed algorithm we must know how there is the affine scaling interior inexact generalized Newton step $\widehat{s}_k$ satisfying (2.18).

Lemma 3.1. *If there exists* $\widehat{s}_k$ *satisfying* (2.18) *when* $\|D_k^{-1} g_k\| = \|D_k^{-1}(V_k^{\mathrm{T}} H_k)\| = 0$ *where* $V_k \in \partial H(x_k)$, *then* $\|H(x_k)\|_2 = 0$. *Further, if* $V_k \in \partial H(x_k)$ *is nonsingular, then the proposed algorithm will generate the inexact generalized Newton step* $\widehat{s}_k$ *only if* $\widehat{s}_k = 0$ *satisfies* (2.18).

Proof. If $\widehat{s}_k$ satisfies the inequality (2.18) when $\|D_k^{-1} g(x_k)\| = \|D_k^{-1}(V_k^{\mathrm{T}} H_k)\| = 0$, squaring Euclidean norm to two sides of the inequality (2.18), we gave

$$\|H(x_k)\|_2^2 + 2[D_k^{-1} g(x_k)]^{\mathrm{T}} \widehat{s}_k + \|V_k D_k^{-1} \widehat{s}_k\|_2^2 = \|H(x_k) + V_k D_k^{-1} \widehat{s}_k\|_2^2 \leqslant \eta_k^2 \|H(x_k)\|_2^2, \tag{3.1}$$

which implies that noting $\eta_k \in (0, 1)$ and $D_k^{-1} g_k = 0$,

$$0 \leqslant \|V_k D_k^{-1} \widehat{s}_k\|_2^2 \leqslant -(1 - \eta_k^2)\|H(x_k)\|_2^2 \leqslant 0.$$

So, $\|H(x_k)\|_2 = 0$. Furthermore, that D_k^{-1} and $V_k \in \partial H(x_k)$ are nonsingular means $\widehat{s}_k = 0$. $\square$

Lemma 3.2. *Assume that there exists an* $\bar{s}_k$ *such that satisfies* $V_k \in \partial H(x_k)$*and* $\|H(x_k) + V_k D_k^{-1} \bar{s}_k\| < \|H(x_k)\|$. *Then there exists* $\eta_{\min} \in (0, 1)$ *such that, for any* $\eta_k \in (\eta_{\min}, 1)$, *there is an* $\widehat{s}_k$ *such that* (2.18) *holds.*

Proof. Clearly $H(x_k) \neq 0$ and hence $\bar{s}_k \neq 0$. Set

$$\eta_{\min} \stackrel{\text{def}}{=} \frac{\|H(x_k) + V_k D_k^{-1} \bar{s}_k\|}{\|H(x_k)\|}. \tag{3.2}$$

For any $\eta_k \in (\eta_{\min}, 1)$, let $\widehat{s}_k \equiv \frac{1-\eta_k}{1-\eta_{\min}} \bar{s}_k$. Since the norm function is convex, we have that

$$\begin{aligned}
\|H(x_k) + V_k D_k^{-1} \widehat{s}_k\| &\leqslant \frac{\eta_k - \eta_{\min}}{1 - \eta_{\min}} \|H(x_k)\| + \frac{1 - \eta_k}{1 - \eta_{\min}} \|H(x_k) + V_k D_k^{-1} \bar{s}_k\| \\
&= \frac{\eta_k - \eta_{\min}}{1 - \eta_{\min}} \|H(x_k)\| + \frac{1 - \eta_k}{1 - \eta_{\min}} \eta_{\min} \|H(x_k)\| \\
&= \eta_k \|H(x_k)\|.
\end{aligned} \tag{3.3}$$

So, Lemma 3.2 does. $\square$

The following lemma show the relation between the gradient $g_k = \nabla h_k = V_k^{\mathrm{T}} H_k$ of the objective function and the step s_k generated by the proposed algorithm. We can see from the lemma how the affine scaling interior inexact generalized Newton steps $\widehat{s}_k$ and s_k are a descent direction for h at the current approximation x_k.

Lemma 3.3. *Let $x_k \in \mathscr{X}$, D_k^{-1} be the invertible matrix given in* (2.11)–(2.12) *and $\widehat{s}_k$ satisfies* (2.18), *then s_k is descent direction for h at x_k, i.e.*,

$$-[D_k^{-1}\nabla h(x_k)]^{\mathrm{T}}\widehat{s}_k = -\nabla h(x_k)^{\mathrm{T}} s_k \geqslant (1-\eta_k)\|H(x_k)\|_2^2 \geqslant 0, \tag{3.4}$$

$$\frac{|[D_k^{-1}\nabla h(x_k)]^{\mathrm{T}}\widehat{s}_k|}{\|\widehat{s}_k\|_2} = \frac{|\nabla h(x_k)^{\mathrm{T}} s_k|}{\|\widehat{s}_k\|_2} \geqslant \frac{1-\eta_k}{(1+\eta_k)\kappa_k\tau_k}\|D_k^{-1}\nabla h(x_k)\| \geqslant 0, \tag{3.5}$$

where $\kappa_k \stackrel{\text{def}}{=} \mathrm{cond}_2(V_k) \stackrel{\text{def}}{=} \|V_k^{-1}\| \cdot \|V_k\|$ and $\tau_k \stackrel{\text{def}}{=} \mathrm{cond}_2(D_k) \stackrel{\text{def}}{=} \|D_k^{-1}\| \cdot \|D_k\|$ represent the Euclidean condition numbers of the matric V_k and D_k, respectively, with $V_k \in \partial H(x_k)$, and $\eta_k \in (0,1)$ given in (2.18).

Proof. Let r_k be the affine scaling residual associated with $\widehat{s}_k$ so that $H(x_k) + V_k D_k^{-1}\widehat{s}_k = r_k$, where $V_k \in \partial H(x_k)$. From

$$\nabla h(x_k)^{\mathrm{T}} s_k = H(x_k)^{\mathrm{T}} V_k D_k^{-1}\widehat{s}_k = H(x_k)^{\mathrm{T}}[r_k - H(x_k)], \tag{3.6}$$

hence, taking the norm in the right-hand side of (3.6), we have that

$$\nabla h(x_k)^{\mathrm{T}} s_k \leqslant \|H(x_k)\| \cdot \|r_k\| - \|H(x_k)\|_2^2 \leqslant -(1-\eta_k)\|H(x_k)\|_2^2. \tag{3.7}$$

So, noting $[D_k^{-1}\nabla h(x_k)]^{\mathrm{T}}\widehat{s}_k = \nabla h(x_k)^{\mathrm{T}} s_k$, for $\eta_k \in (0,1)$, the conclusion (3.4) of the lemma is true.

Clearly, $D_k^{-1}V_k^{\mathrm{T}}H(x_k) \neq 0$ and hence $\widehat{s}_k \neq 0$. Next, $V_k D_k^{-1}\widehat{s}_k = r_k - H(x_k)$. Thus, $\widehat{s}_k = D_k V_k^{-1}[r_k - H(x_k)]$, taking the norm, we have

$$\|\widehat{s}_k\| \leqslant \|D_k\|\|V_k^{-1}\|(\|r_k\| + \|H(x_k)\|) \leqslant (1+\eta_k)\|D_k\|\|V_k^{-1}\|_2 \cdot \|H(x_k)\|_2. \tag{3.8}$$

Combining (3.7) with (3.8), we have that

$$\frac{|\nabla h(x_k)^{\mathrm{T}} s_k|}{\|\widehat{s}_k\|_2} \geqslant \frac{(1-\eta_k)\|H(x_k)\|_2^2}{(1+\eta_k)\|D_k\|\|V_k^{-1}\|_2 \cdot \|H(x_k)\|_2} = \frac{(1-\eta_k)\|H(x_k)\|_2}{(1+\eta_k)\|D_k\|\|V_k^{-1}\|_2} \tag{3.9}$$

and hence as a result, using the fact that $\|D_k^{-1}\nabla h(x_k)\|_2 \leqslant \|D_k^{-1}\|\|H(x_k)\|_2\|V_k\|_2$, we get

$$\frac{|\nabla h(x_k)^{\mathrm{T}} s_k|}{\|D_k^{-1}\nabla h(x_k)\|_2\|\widehat{s}_k\|_2} \geqslant \frac{(1-\eta_k)\|H(x_k)\|_2}{(1+\eta_k)(\|D_k\|\|D_k^{-1}\|)\|H(x_k)\|_2(\|V_k\|_2\|V_k^{-1}\|_2)} \geqslant \frac{1-\eta_k}{(1+\eta_k)\tau_k\kappa_k},$$

where $\kappa_k = \mathrm{cond}_2(V_k)$ and $\tau_k = \mathrm{cond}_2(D_k)$. So, we have that the conclusions of the lemma is true. $\square$

Lemma 3.4. *Let $\beta \in (0,1)$ and s_k be proposed by* (2.18). *Assume that Assumptions* A1–A4 *hold, and there exist $\eta \in (0,1)$, τ and κ such that $\eta_k \leqslant \eta$, $\tau_k \leqslant \tau$ and $\kappa_k \leqslant \kappa$. If $\|D_k^{-1}g_k\| \neq 0$ then the proposed algorithm will produce an iterate $x_{k+1} = x_k + \lambda_k s_k$ in a finite number of backtracking steps in* (2.20)–(2.21).

Proof. Since $\|D_k^{-1}g(x_k)\| \neq 0$, by continuity there exist $\delta > 0$ and $\varepsilon > 0$ such that $\|D(x)^{-1}g(x)\| \geqslant \varepsilon$ for all x with $\|x_k - x\| \leqslant \delta$. It can be clearly seen that λ_k will satisfy $\lambda_k \leqslant \Lambda_k$ in a finite number of backtracking reductions where Λ_k given in (2.23). Using the mean value theorem, we have that with $0 \leqslant \vartheta_k \leqslant 1$, the following inequality holds

$$\begin{aligned} h(x_k + \lambda_k s_k) &= h(x_k) + \beta\lambda_k \nabla h(x_k)^{\mathrm{T}} s_k + (1-\beta)\lambda_k \nabla h(x_k)^{\mathrm{T}} s_k \\ &\quad + \lambda_k[\nabla h(x_k + \vartheta_k \lambda_k s_k)^{\mathrm{T}} s_k - \nabla h(x_k)^{\mathrm{T}} s_k] \\ &= h(x_k) + \beta\lambda_k \nabla h(x_k)^{\mathrm{T}} s_k + \lambda_k[(1-\beta)\nabla h(x_k)^{\mathrm{T}} s_k + \xi_k], \end{aligned} \tag{3.10}$$

where for convenience we have set $\xi_k \stackrel{\text{def}}{=} [\nabla h(x_k + \vartheta_k \lambda_k s_k) - \nabla h(x_k)]^{\mathrm{T}} s_k$. By the Assumption A4, there exists a constant $\chi_D > 0$ such that $\|D(x)^{-1}\| \leqslant \chi_D$, $\forall x \in \mathscr{L}(x_0)$ and there exist η and κ such that $\eta_k \leqslant \eta < 1$ and $\kappa_k \leqslant \kappa$, $\tau_k \leqslant \tau$, $\forall k$. Since $\nabla h(x)$ is continuous, there exists sufficiently small λ_k when $\|\vartheta_k \lambda_k s_k\| \leqslant \delta'$ such that

$$\|\nabla h(x_k + \vartheta_k \lambda_k s_k) - \nabla h(x_k)\| \leqslant (1-\beta)\,\frac{1-\eta}{(1+\eta)\tau\kappa\chi_D}\,\varepsilon.$$

Note that from the assumptions we have

$$|\xi_k| = |[\nabla h(x_k + \theta_k \lambda_k s_k) - \nabla h(x_k)]^{\mathrm{T}} s_k| \leqslant \frac{(1-\beta)(1-\eta)\varepsilon}{(1+\eta)\kappa\tau\chi_D}\,\|s_k\| \leqslant \frac{(1-\beta)(1-\eta)\varepsilon}{(1+\eta)\kappa\tau}\,\|\widehat{s}_k\|.$$

Since (3.5) means $\nabla h(x_k)^{\mathrm{T}} s_k \leqslant -\frac{1-\eta}{(1+\eta)\kappa\tau}\varepsilon\|\widehat{s}_k\|$, we have that after a finite number of reductions, the last term in brackets in the right-hand side of (3.10) will become negative and the corresponding λ_k will be acceptable, that is, we have that in a finite number of backtracking steps, λ_k must satisfy

$$h(x_k + \lambda_k s_k) \leqslant h(x_k) + \beta\lambda_k \nabla h(x_k)^{\mathrm{T}} s_k.$$

Since $h(x_k) \leqslant h(x_{l(k)})$, the conclusion of the lemma holds. □

In this section, we are now ready to state one of our main results of the proposed algorithm.

Theorem 3.5. *Let $\{x_k\} \subset \Omega \subset \Re^n$ be a sequence generated by the proposed algorithm. Assume that Assumptions* A1–A4 *hold and there exist η, τ and κ such that $\eta_k \leqslant \eta < 1$, $\tau_k \leqslant \tau$ and $\kappa_k \leqslant \kappa$. If λ_k given in* (2.21) *is bounded away from zero as $s_k \to 0$, then*

$$\liminf_{k\to\infty} \|D_k^{-1} g_k\| = \liminf_{k\to\infty} \|D_k^{-1} V_k^{\mathrm{T}} H_k\| = 0, \tag{3.11}$$

where $V_k \in \partial H(x_k)$.

Proof. According to the acceptance rule (2.20), we have that by s_k being a descent direction

$$h(x_{l(k)}) - h(x_k + \lambda_k s_k) \geqslant -\lambda_k \beta \nabla h(x_k)^{\mathrm{T}} s_k = \beta[D_k^{-1} V_k^{\mathrm{T}} H_k]^{\mathrm{T}}(D_k s_k), \tag{3.12}$$

where $V_k \in \partial H(x_k)$. Taking into account that $m(k+1) \leqslant m(k)+1$, and $h(x_{k+1}) \leqslant h(x_{l(k)})$, we have $h(x_{l(k+1)}) \leqslant \max\{h(x_{l(k)}), h(x_{k+1})\} = h(x_{l(k)})$. This means that the sequence $\{h(x_{l(k)})\}$ is nonincreasing for all k, and therefore $\{h(x_{l(k)})\}$ is convergent.

Since s_k is a descent direction $|\nabla h(x_k)^{\mathrm{T}} s_k| = -\nabla h(x_k)^{\mathrm{T}} s_k$, and $\tau_k \leqslant \tau$, $\kappa_k \leqslant \kappa$, $1-\eta_k \geqslant 1-\eta$ and hence $1+\eta_k \leqslant 1+\eta$ for all k. By (2.20) and (3.5), for all $k > M$,

$$\begin{aligned} h(x_{l(k)}) &= h(x_{l(k)-1} + \lambda_{l(k)-1} s_{l(k)-1}) \\ &\leqslant \max_{0 \leqslant j \leqslant m(l(k)-1)} \{h(x_{l(k)-j-1})\} + \lambda_{l(k)-1} \beta \nabla h(x_{l(k)-1})^{\mathrm{T}} s_{l(k)-1} \\ &\leqslant \max_{0 \leqslant j \leqslant m(l(k)-1)} \{h(x_{l(k)-j-1})\} \\ &\quad - \lambda_{l(k)-1} \beta \frac{1-\eta}{(1+\eta)\kappa\tau} \|D_{l(k)-1}^{-1} \nabla h(x_{l(k)-1})\|_2 \|\widehat{s}_{l(k)-1}\|_2. \end{aligned} \tag{3.13}$$

If the conclusion of the theorem is not true, then there exists some $\varepsilon > 0$ such that

$$\|D_k^{-1} \nabla h(x_k)\|_2 \geqslant \varepsilon, \quad k = 1, 2, \ldots . \tag{3.14}$$

Therefore, we have that

$$h(x_{l(k)}) \leqslant h(x_{l(l(k)-1)}) - \lambda_{l(k)-1} \beta \varepsilon \frac{1-\eta}{(1+\eta)\kappa\tau} \|\widehat{s}_{l(k)-1}\|_2. \tag{3.15}$$

As $\{h(x_{l(k)})\}$ is convergent, we obtain from (3.15) that

$$\lim_{k \to \infty} \lambda_{l(k)-1} \|\widehat{s}_{l(k)-1}\|_2 = 0.$$

This means that either

$$\liminf_{k \to \infty} \lambda_{l(k)-1} = 0, \tag{3.16}$$

or

$$\liminf_{k \to \infty} \|\widehat{s}_{l(k)-1}\| = 0. \tag{3.17}$$

If (3.17) holds, following by induction way used in [10], it can be derived that

$$\lim_{k \to \infty} h(x_{l(k)}) = \lim_{k \to \infty} h(x_k). \tag{3.18}$$

By the rule for accepting the step $\lambda_k s_k$,

$$h(x_{k+1}) - h(x_{l(k)}) \leqslant \beta \lambda_k \nabla h(x_k)^{\mathrm{T}} s_k \leqslant \beta \varepsilon \frac{1-\eta}{(1+\eta)\tau\kappa} \lambda_k \|\widehat{s}_k\|_2. \tag{3.19}$$

This means

$$\lim_{k \to \infty} \lambda_k \|\widehat{s}_k\|_2 = 0,$$

which implies that either

$$\lim_{k \to \infty} \lambda_k = 0, \tag{3.20}$$

or

$$\liminf_{k \to \infty} \|\widehat{s}_k\| = 0. \tag{3.21}$$

If above Eq. (3.21) holds, we have that from

$$\lim_{k\to\infty}\{\|H(x_k)+V_kD_k^{-1}\widehat{s}_k\|-\eta_k\|H(x_k)\|\}\leqslant 0.$$

Now, let x_* be any limit point of $\{x_k\}$, it means $(1-\eta)\|H(x_*)\|\leqslant 0$. Since $1-\eta>0$, we have $\|H(x_*)\|=0$ which implies $\|D_*^{-1}\nabla h(x_*)\|_2\leqslant\|D_*^{-1}\|_2\|V_*\|_2\|H(x_*)\|_2=0$, where $V_*\in\partial H(x_*)$. The conclusion of this theorem holds.

Furthermore, if (3.20) holds, then acceptance rule (2.20) means that, for large enough k,

$$h\left(x_k+\frac{\lambda_k}{\omega}s_k\right)-h(x_k)\geqslant h\left(x_k+\frac{\lambda_k}{\omega}s_k\right)-h(x_{l(k)})>\beta\frac{\lambda_k}{\omega}\nabla h(x_k)^{\mathrm{T}}s_k.$$

Since

$$h\left(x_k+\frac{\lambda_k}{\omega}s_k\right)-h(x_k)=\frac{\lambda_k}{\omega}\nabla h(x_k)^{\mathrm{T}}s_k+\mathrm{o}\left(\frac{\lambda_k}{\omega}\|s_k\|\right),$$

we have

$$(1-\beta)\frac{\lambda_k}{\omega}\nabla h(x_k)^{\mathrm{T}}s_k+\mathrm{o}\left(\frac{\lambda_k}{\omega}\|s_k\|\right)\geqslant 0. \tag{3.22}$$

Dividing (3.22) by $\frac{\lambda_k}{\omega}\|s_k\|$ and noting that $1-\beta>0$ and $\nabla h(x_k)^{\mathrm{T}}s_k\leqslant 0$, we obtain

$$\lim_{k\to\infty}\frac{\nabla h(x_k)^{\mathrm{T}}s_k}{\|s_k\|}=0. \tag{3.23}$$

From (3.5) and the assumptions of the theorem, we have that from $1-\eta>0$ and $\|s_k\|\leqslant\chi_D\|\widehat{s}_k\|$,

$$0\leqslant\lim_{k\to\infty}\frac{\nabla h(x_k)^{\mathrm{T}}s_k}{\|s_k\|}=\lim_{k\to\infty}\frac{\nabla h(x_k)^{\mathrm{T}}s_k}{\|\widehat{s}_k\|}\frac{\|\widehat{s}_k\|}{\|s_k\|}\leqslant-\frac{\beta\varepsilon(1-\eta)}{\tau\kappa\chi_D(1+\eta)}<0, \tag{3.24}$$

which also implies $s_k\to 0$ and hence λ_k given in (2.23) is bounded away from zero by the assumption of the theorem and $\lim_{s_k\to 0}\theta_k=1$. Thus acceptance rules (2.20) and (2.21) all mean that $\{\lambda_k\}$ cannot converges to zero, contradicting (3.20). So, the conclusion of the theorem is true. □

Theorem 3.5 indicates that at least one limit point of $\{x_k\}$ is a stationary point. Now, we shall extend this theorem to a stronger result.

Theorem 3.6. *Assume that the assumptions of Theorem 3.5 hold. Let $\{x_k\}$ be a sequence generated by the algorithm. Then*

$$\lim_{k\to\infty}\|D_k^{-1}g_k\|=\lim_{k\to\infty}\|D_k^{-1}(V_k^{\mathrm{T}}H_k)\|=0, \tag{3.25}$$

where $V_k\in\partial H(x_k)$.

Proof. If (3.25) is not true, then we assume that there are an $\varepsilon_1\in(0,1)$ and a subsequence $\{\|D_{m_i}^{-1}\nabla h(x_{m_i})\|\}$ of $\{\|D_k^{-1}\nabla h(x_k)\|\}$ such that $\|D_{m_i}^{-1}\nabla h(x_{m_i})\|\geqslant\varepsilon_1$ for all $m_i,\ i=1,2,\ldots$. Theorem 3.5 guarantees the existence of another subsequence $\{\|D_{l_i}^{-1}\nabla h(x_{l_i})\|\}$ such that

$$\|D_k^{-1}\nabla h(x_k)\| \geqslant \varepsilon_2, \quad \text{for } m_i \leqslant k < l_i \tag{3.26}$$

and

$$\|D_{l_i}^{-1}\nabla h(x_{l_i})\| \leqslant \varepsilon_2 \tag{3.27}$$

for an $\varepsilon_2 \in (0, \varepsilon_1)$.

By (3.5), we get

$$\nabla h(x_k)^{\mathrm{T}} s_k \leqslant -\frac{1-\eta}{(1+\eta)\tau\kappa}\|D_k^{-1}\nabla h(x_k)\|_2 \|\widehat{s}_k\|_2. \tag{3.28}$$

Set $\widehat{\omega}_1 \stackrel{\text{def}}{=} \frac{1-\eta}{(1+\eta)\tau\kappa}$. Eqs. (3.5) and (3.28) mean that

$$h(x_{l(k)}) \leqslant h(x_{l(l(k)-1)}) - \lambda_{l(k)-1}\beta\widehat{\omega}_1\|D_k^{-1}\nabla h(x_k)\|_2\|\widehat{s}_k\|_2. \tag{3.29}$$

Similar to the proof of Theorem 3.5, we can also obtain (3.18), i.e.,

$$\lim_{k\to\infty} h(x_{l(k)}) = \lim_{k\to\infty} h(x_k). \tag{3.30}$$

By the accepting rule of the step s_k,

$$h(x_{k+1}) - h(x_{l(k)}) \leqslant \beta\lambda_k \nabla h(x_k)^{\mathrm{T}} s_k \leqslant -\lambda_k\beta\widehat{\omega}_1\|D_k^{-1}\nabla h(x_k)\|_2\|\widehat{s}_k\|_2. \tag{3.31}$$

Eqs. (3.28), (3.30) and (3.31) imply that for $m_i \leqslant k < l_i$,

$$\lim_{k\to\infty, m_i\leqslant k<l_i} \lambda_k \ \|\widehat{s}_k\|_2 = 0. \tag{3.32}$$

By the accepting rule of the step $\lambda_k s_k$, for large enough i such that $m_i \leqslant k < l_i$, using the mean value theorem we have the following equality

$$h(x_k+\lambda_k s_k) - h(x_k) = \lambda_k[\nabla h(x_k+\xi_k\lambda_k s_k) - \nabla h(x_k)]^{\mathrm{T}} s_k + \lambda_k \nabla h(x_k)^{\mathrm{T}} s_k \tag{3.33}$$

with $\xi_k \in (0,1)$. Since $\nabla h(x)$ is continuous, there exists sufficiently small λ_k when $\|\lambda_k s_k\| \leqslant \delta'$ such that

$$|[\nabla h(x_k+\xi_k\lambda_k s_k) - \nabla h(x_k)]^{\mathrm{T}} s_k| \leqslant \frac{(1-\beta)(1-\eta)}{(1+\eta)\kappa\tau\chi_D}\varepsilon_2\|s_k\|.$$

So, (3.33) implies that by for large i, $m_i \leqslant k < l_i$,

$$\begin{aligned} &h_k - h(x_k+\lambda_k s_k) \\ &\quad\geqslant -\lambda_k\frac{(1-\beta)(1-\eta)}{(1+\eta)\kappa\tau\chi_D}\varepsilon_2\|s_k\| + \lambda_k\frac{1-\eta_k}{(1+\eta_k)\kappa_k\tau_k}\|D_k^{-1}\nabla h(x_k)\|\|\widehat{s}_k\| \\ &\quad\geqslant -\lambda_k\frac{(1-\beta)(1-\eta)}{(1+\eta)\kappa\tau\chi_D}\varepsilon_2\|s_k\| + \lambda_k\frac{1-\eta}{(1+\eta)\kappa\tau\chi_D}\varepsilon_2\|s_k\| \\ &\quad= \frac{\beta(1-\eta)}{(1+\eta)\kappa\tau\chi_D}\varepsilon_2(\lambda_k\|s_k\|) \stackrel{\text{def}}{=} \varpi\lambda_k\|s_k\|, \end{aligned} \tag{3.34}$$

where $\varpi \stackrel{\text{def}}{=} \frac{\beta(1-\eta)}{(1+\eta)\kappa\tau\chi_D}\varepsilon_2$. Similarly to prove (3.24), we can also obtain that $s_k \to 0$ and hence $\lim_{s_k\to 0}\theta_k = 1$. From $\|x_{k+1} - x_k\| \leqslant \lambda_k\|s_k\|$, we then deduce from this bound that for i sufficiently large,

$$\|x_{m_i} - x_{l_i}\| \leqslant \sum_{k=m_i}^{l_i-1} \|x_k - x_{k+1}\| \leqslant \frac{1}{\varpi} \sum_{k=m_i}^{l_i-1} [h_k - h(x_k + \lambda_k s_k)] = \frac{1}{\varpi}(h_{m_i} - h_{l_i}). \tag{3.35}$$

Therefore, (3.35) implies that $h_{m_i} - h_{l_i}$ tends to zero as i tends to infinity and hence $\|x_{m_i} - x_{l_i}\|$ tends to zero as i tends to infinity. Eq. (2.12) implies $|(v_{m_i})_j - (v_{l_i})_j| \leqslant |(x_{m_i})_j - (x_{l_i})_j| \to 0$, as i tends to infinity. Consequently,

$$\|D_{l_i}^{-1}\nabla h(x_{l_i}) - D_{m_i}^{-1}\nabla h(x_{m_i})\| \leqslant \|(D_{m_i}^{-1} - D_{l_i}^{-1})V_{l_i}^{\mathrm{T}}H_{l_i}\| + \|D_{m_i}^{-1}(V_{m_i}^{\mathrm{T}}H_{m_i} - V_{l_i}^{\mathrm{T}}H_{l_i})\|.$$

Finally, from the triangle inequality and continuity of the gradient $\nabla h(x) = V^{\mathrm{T}}(x)H(x)$ where $V(x) \in \partial H(x)$, we then deduce from this bound tending to zero for i sufficiently large. We thus deduce that $|\|D_{l_i}^{-1}\nabla h(x_{l_i})\| - \|D_{m_i}^{-1}\nabla h(x_{m_i})\||$ also tends to zero. However, this is impossible because of the definitions of $\{l_i\}$ and $\{m_i\}$, which imply that

$$|\|D_{l_i}^{-1}\nabla h(x_{l_i})\| - \|D_{m_i}^{-1}\nabla h(x_{m_i})\|| \geqslant \|D_{m_i}^{-1}\nabla h(x_{m_i})\| - \|D_{l_i}^{-1}\nabla h(x_{l_i})\| \geqslant \varepsilon_1 - \varepsilon_2 > 0.$$

Hence no subsequence satisfying (3.26) can exist, and the theorem is proved. □

Corollary 3.7. *Assume that the assumptions of Theorem* 3.5 *hold. If there exists a limit point x^* of the sequence $\{x_k\}$ generated by the proposed algorithm such that $x^* \in \mathrm{int}(\Omega)$ is a BD-regular point of H at which H is semismooth, then* $\lim_{k\to\infty}\|H_k\| = 0$, *and all the accumulation point solve the problem* (1.1).

Proof. For $x^* \in \mathrm{int}(\Omega)$, there exists sufficiently small $\delta \in (0, 2]$ such that the open ball $\mathcal{N}(x^*, \delta) \stackrel{\text{def}}{=} \{x \mid \|x - x^*\| < \delta\} \in \mathrm{int}(\Omega)$. Since x^* is a BD-regular point of H, we have that there exist, without loss of generality $\delta > 0$ and $\rho > 0$ such that all $V \in \partial_{\mathrm{B}}H(y)$, $\|y - x\| \leqslant \delta$ are nonsingular with $\|V^{-1}\| \leqslant \rho$. Also , let $\{x_{k_j}\}$ be subsequence such that $x_{k_j} \to x^*$ and j_0 be the index such that for $k > k_{j_0}$, the sequence $\{x_{k_j}\}$belongs to $\mathcal{N}(x^*, \delta/2)$. Assume $k_j > k_{j_0}$. Then $|l_i - (x_{k_j})_i| > \delta/2$ and $|u_i - (x_{k_j})_i| > \delta/2$ for $i = 1, \ldots, n$, where l_i, u_i and $(x_{k_j})_i$ are the ith components of l, u and x_{k_j}, respectively. Hence, $\|D_{k_j}\| \leqslant \sqrt{2n/\delta}$ where $\sqrt{2/\delta} \geqslant 1$. Further, (3.18) means that the sequence $\{h(x_k)\}$ is convergent. Then, from the following inequality and (3.25)

$$\frac{\|H_{k_j}\|}{\rho\sqrt{2n/\delta}} \leqslant \frac{\|V_{k_j}^{\mathrm{T}}H_{k_j}\|}{\|D_{k_j}\|} \leqslant \|D_{k_j}^{-1}V_{k_j}^{\mathrm{T}}H_{k_j}\| \to 0, \tag{3.36}$$

where $V_{k_j} \in \partial H(x_{k_j})$ which implies that the corollary is proved. □

Remark. Since $x^* \in \mathrm{int}(\Omega)$, we can obtain that there exists $\delta > 0$ such that $|l_i - (x_{k_j})_i| > \delta/2$ and $|u_i - (x_{k_j})_i| > \delta/2$ for $i = 1, \ldots, n$. So, the step size Λ_k given in (2.23) along the direction s_k is bounded away from zero, furthermore, $\Lambda_k \to +\infty$ as $s_k \to 0$.

4. The local convergence

We now discuss the convergence rate for the proposed algorithm. Defining the generalized Newton step s_k^{N} satisfies

$$V_k s_k^{\mathrm{N}} = -H(x_k), \tag{4.1}$$

that is, the inexact generalized Newton step $s_k \to s_k^{\mathrm{N}}$ if $\eta_k \to 0$.

Theorem 4.1. *Assume that Assumptions* A1–A4 *and the assumptions of Theorem* 3.5 *hold. Let x^* be any accumulation point of the sequence $\{x_k\}$ generated by the proposed algorithm and x^* be a BD-regular point of H at which H is semismooth. Furthermore, if $\eta_k \to 0$, $\tau = \mathrm{cond}(D_*) < +\infty$, $\kappa = \mathrm{cond}(V_*) < +\infty$ with $V_* \in \partial H(x_*)$ and Λ_k given in* (2.23) *is bounded away from* 1 *as $s_k^{\mathrm{N}} \to 0$, that is,*

$$\lim_{k\to\infty} \Gamma_k \overset{\mathrm{def}}{=} \min\{\Lambda_k, 1\} = 1, \tag{4.2}$$

$$\Lambda_k \overset{\mathrm{def}}{=} \min\left\{\max\left\{\frac{l_i - (x_k)_i}{(s_k)_i}, \frac{u_i - (x_k)_i}{(s_k)_i}\right\}, i = 1, 2, \ldots, n\right\}, \tag{4.3}$$

where $l_i, u_i, (x_k)_i$ and $(s_k)_i$ are the ith components of l, u, x_k and s_k, respectively. Then x^* is a BD-regular zero solution of H and $x_k \to x^*$, the step size $\lambda_k \equiv 1$ for large enough k.

Proof. If $H(x_k) = 0$ for some large k, then the residual condition implies $\widehat{s}_k = 0$. As a result, $s_k = D_k \widehat{s}_k \equiv 0$ for all large enough k and the step size $\lambda_k \equiv 1$. Therefore, without loss of generality we may assume that $D_k^{-1} V_k^{\mathrm{T}} H(x_k) \neq 0$. Similar to prove (3.24) in Theorem 3.5, we also have that

$$\lim_{k\to\infty} \frac{\nabla h(x_k)^{\mathrm{T}} s_k}{\|\widehat{s}_k\|} = 0. \tag{4.4}$$

By the continuities of $\mathrm{cond}(D_k)$ and $\mathrm{cond}(V_k)$, and $\tau_k = \mathrm{cond}(D_k) \to \mathrm{cond}(D_*) < +\infty$ and $\kappa_k = \mathrm{cond}(V_k) \to \mathrm{cond}(V_*) < +\infty$, we have that $\{\tau_k\}$ and $\{\kappa_k\}$ are uniformly bounded from above. Hence, (4.4) and (3.5) imply that as $\eta_k \to 0$

$$0 = \lim_{k\to\infty} \frac{\nabla h(x_k)^{\mathrm{T}} s_k}{\|\widehat{s}_k\|} \geqslant \lim_{k\to\infty} \frac{1-\eta_k}{\tau_k(1+\eta_k)\kappa_k} \|D_k^{-1}\nabla h(x_k)\|_2 = \frac{1}{\tau\kappa} \|D_*^{-1}\nabla h(x^*)\|_2, \tag{4.5}$$

which implies $D_*^{-1} V_*^{\mathrm{T}} H(x^*) = D_*^{-1}\nabla h(x^*) = 0$. Since $x^* \in \Omega$ is a BD-regular point of H, i.e., D_* and V_* nonsingular, this gives $H(x^*) = 0$ which means that $x^* \in \Omega$ is a BD-regular zero solution of H.

Further, x^* is a BD-regular zero of H which H is semismooth, then there exist $\delta > 0$ and $\zeta > 0$ such that

$$\|H(x_k)\| \geqslant \zeta \|x_k - x^*\| \quad \text{for all } \|x_k - x^*\| \leqslant \delta.$$

This gives $x_k \to x_* \in \Omega$.

Using Lemma 2.4 again, V_k is also nonsingular and uniformly bounded from above for all $V_k \in \partial_{\mathrm{B}} H(x_k)$ and all large enough k. By reducing δ, if necessary, such that for all k with $\|x_k + s_k - x^*\| < \delta'$

$$\|x_k + s_k - x^*\| < \delta', \quad \|x_k - x^*\| \leqslant \delta < \delta'.$$

For large enough k, by $V_k \in \partial_B H(x_k)$ and D_k by nonsingular, we have that by Lemma 2.2,

$$\begin{aligned} h(x_k+s_k)-h_k-\nabla h_k^T s_k &= \tfrac{1}{2}\|H_k+V_k s_k+o(\|s_k\|)\|^2-\tfrac{1}{2}\|H_k\|^2-\nabla h_k^T s_k \\ &= \tfrac{1}{2}\|V_k D_k^{-1}\widehat{s}_k\|^2+o(\|s_k\|^2) \\ &\geqslant \tfrac{1}{4}\|V_k D_k^{-1}\widehat{s}_k\|^2. \end{aligned} \tag{4.6}$$

This gives

$$\begin{aligned} h(x_k+s_k) \leqslant{}& h(x_{l(k)})+\beta\nabla h(x_k)^T s_k+(\tfrac{1}{2}-\beta)\nabla h(x_k)^T s_k \\ &+\tfrac{1}{2}(\nabla h(x_k)^T s_k+\|V_k D_k^{-1}\widehat{s}_k\|^2)+o(\|s_k\|^2). \end{aligned} \tag{4.7}$$

Next, by (3.7)–(3.8), we get

$$\nabla h(x_k)^T s_k \leqslant \|H(x_k)\|\cdot\|r_k\|-\|H(x_k)\|_2^2 \leqslant -(1-\eta_k)\|H(x_k)\|_2^2, \tag{4.8}$$

$$\|s_k\| \leqslant \|D_k^{-1}\|\|\widehat{s}_k\| \leqslant (1+\eta_k)\|D_k^{-1}\|\|V_k^{-1}\|\|H(x_k)\|_2 \tag{4.9}$$

and

$$\|V_k D_k^{-1}\widehat{s}_k\| \leqslant \|r_k\|_2+\|H(x_k)\|_2 \leqslant (1+\eta_k)\|H(x_k)\|_2.$$

So, (4.7) can be rewritten as follows

$$\begin{aligned} h(x_k+s_k) \leqslant{}& h(x_{l(k)})+\beta\nabla h(x_k)^T s_k+(\tfrac{1}{2}-\beta)\nabla h(x_k)^T s_k \\ &+\tfrac{1}{2}(\nabla h(x_k)^T s_k+\|V_k D_k^{-1}\widehat{s}_k\|^2)+o(\|s_k\|^2) \\ \leqslant{}& h(x_{l(k)})+\beta\nabla h(x_k)^T s_k \\ &-[(\tfrac{1}{2}-\beta)(1-\eta_k)+\tfrac{1}{2}(1-\eta_k)-\tfrac{1}{2}(1+\eta_k)^2]\|H(x_k)\|_2^2+o(\|s_k\|^2) \\ \leqslant{}& h(x_{l(k)})+\beta\nabla h(x_k)^T s_k \end{aligned} \tag{4.10}$$

for all large enough k, the last inequality is deduced because the third term in brackets in the right-hand side of (4.10) will become negative by

$$(\tfrac{1}{2}-\beta)(1-\eta_k)+\tfrac{1}{2}(1-\eta_k)-\tfrac{1}{2}(1+\eta_k)^2 \to (\tfrac{1}{2}-\beta), \quad \text{as } \eta_k \to 0,$$

and by (4.9), $\|s_k\|^2 = O(\|H(x_k)\|_2^2)$ for the last term. By the above inequality and λ_k given in (2.23) is bounded away from 1 as $s_k \to 0$ and hence $\lim_{s_k\to 0}\theta_k = 1$, we know that the acceptance rules (2.20)–(2.21) mean that for large k

$$x_{k+1} = x_k + s_k,$$

which implies that for large enough k, the step size $\lambda_k \equiv 1$ and hence the theorem is proved. □

Theorem 4.1 means that the local convergence rate for the proposed algorithm depends on the generalized subdifferential of the function at x^* and the local convergence rate of the step. If the nondegenerate property of the system (1.2) holds at point x^*, then at the kth iteration for large k, the active set is invariable. So, s_k becomes the projected generalized Newton step or the projected generalized Newton step, then the sequence $\{x_k\}$ generated by the algorithm converges to x_* quadratically or q-superlinearly.

We now present the characterizations of rate of convergence of the proposed algorithm. If the H is p-order semismooth at $x^* \in \Omega$, choosing $\delta > 0$ sufficiently small, then we define that

$$\gamma(x^*, \delta) \stackrel{\text{def}}{=} \sup_y \left\{ \sup_{V \in \partial H(y)} \frac{\|H(y) - H(x^*) - V(y - x^*)\|}{\|y - x^*\|} \,\middle|\, y \neq x^*;\ y \in \mathscr{N}(x^*; \delta) \right\}. \tag{4.11}$$

By p-order semismooth at $x^* \in \Omega$, it is clear that $\gamma(x^*, \delta)$ can be made as small as desired by making δ sufficiently small.

Theorem 4.2. *Assume that $\eta_k \to 0$. Suppose that H is locally Lipschizian, and $x^* \in \Omega$ be a BD-regular zero of H at which H is p-order semismooth. There exists $\varepsilon > 0$ such that $\|x_0 - x^*\| \leqslant \varepsilon$, then for each k, the sequence $\{x_k\}$ generated by*

$$\|H(x_k) + V_k D_k^{-1} \widehat{s}_k\| \leqslant \eta_k \|H(x_k)\|, \tag{4.12}$$

$$x_{k+1} = x_k + s_k = x_k + D_k^{-1} \widehat{s}_k, \tag{4.13}$$

converges to x^*. Moreover, the convergence is linear in the sense that there exists $\tau \in (0, 1)$ such that

$$\|x_{k+1} - x^*\| \leqslant \tau \|x_k - x^*\|. \tag{4.14}$$

Proof. Since H is BD-regular zero at x^*, there are a neighborhood $\mathscr{N}$ of x^* and constants $\bar{\mu} > 0$, $\mu > 0$ such that for any $y \in \mathscr{N}$ and $V \in \partial_{\mathrm{B}} H(y)$, V is nonsingular and $\max\{\|V\|\} \leqslant \mu$, $\max\{\|V^{-1}\|\} \leqslant \bar{\mu}$. By $\eta_k \to 0$, there exist sufficiently small $\gamma > 0$ and sufficiently small $\eta_{\max} \in (0, 1)$ such that $\eta_k \leqslant \eta_{\max}$ and

$$\bar{\mu}[\eta_{\max}(\mu + \gamma) + \gamma] \leqslant \tau. \tag{4.15}$$

Now, choose $\varepsilon > 0$ sufficiently small such that $\gamma(x^*, \varepsilon) \leqslant \gamma$ where $\gamma(x^*, \varepsilon)$ given in (4.11), and for any $y \in \mathscr{N}(x^*; \varepsilon)$ and $V \in \partial_{\mathrm{B}} H(y)$, that is,

$$\|H(y) - H(x^*) - V(y - x^*)\| \leqslant \gamma \|y - x^*\| \tag{4.16}$$

if $\|y - x^*\| \leqslant \varepsilon$. By reducing ε, if necessary, for all k, from (3) and (4) in Lemma 2.2, we have also that

$$\|H(y) - H(x^*) - V_*(y - x^*)\| \leqslant \gamma \|y - x^*\|$$

for all $V_* \in \partial_{\mathrm{B}} H(x^*)$, if $\|y - x^*\| \leqslant \varepsilon$.

Assume that $\|x_0 - x^*\| \leqslant \varepsilon$. We prove (4.14) by induction. Note that the induction hypothesis

$$\|x_k - x^*\| \leqslant \tau^k \|x_0 - x^*\| \leqslant \tau^k \varepsilon \leqslant \varepsilon,$$

so that (4.16) holds with $y = x_k$. Let $r_k = H(x_k) + V_k D_k^{-1} \widehat{s}_k$. Since

$$\begin{aligned} x_{k+1} - x^* &= x_k - x^* + V_k^{-1}[r_k - H(x_k)] \\ &= V_k^{-1} r_k - V_k^{-1}[H(x_k) - H(x^*) - V_k(x_k - x^*)] \end{aligned} \tag{4.17}$$

taking norms,

$$\begin{aligned} \|x_{k+1} - x^*\| &\leqslant \|V_k^{-1}\| \ [\|r_k\| + \|H(x_k) - H(x^*) - V_k(x_k - x^*)\|] \\ &\leqslant \bar{\mu} \ [\eta_k \|H(x_k)\| + \gamma \|x_k - x^*\|]. \end{aligned} \tag{4.18}$$

Since

$$H(x_k) = V_*(x_k - x^*) + [H(x_k) - H(x^*) - V_*(x_k - x^*)]$$

taking norms,

$$\|H(x_k)\| \leqslant \|V_*(x_k - x^*)\| + \|H(x_k) - H(x^*) - V_*(x_k - x^*)\| \leqslant \mu\|x_k - x^*\| + \gamma\|x_k - x^*\|,$$

using (4.16). Therefore,

$$\begin{aligned}\|x_{k+1} - x^*\| &\leqslant \bar{\mu}[\eta_k(\mu + \gamma)\|x_k - x^*\| + \gamma\|x_k - x^*\|] \\ &\leqslant \bar{\mu}[\eta_{\max}(\mu + \gamma) + \gamma]\|x_k - x^*\| \leqslant \tau\|x_k - x^*\|.\end{aligned} \tag{4.19}$$

The result now follows from the choice of γ. □

Theorem 4.3. *Suppose that H is locally Lipschizian, semismooth and BD-regular at x^* which is a solution of* (1.1) *and that the sequence $\{x_k\}$ generated by* (4.12)–(4.13) *converges to x^*. Then $x_k \to x^*$ superlinearly if and only if*

$$\|r_k\| = \mathrm{o}(\|H(x_k)\|) \quad \text{as } k \to \infty, \tag{4.20}$$

where $r_k = H(x_k) + V_k D_k^{-1}\widehat{s}_k$. Moreover, if H is p-order semismooth at x^*, then $x_k \to x^*$ with order at least $1 + p$ if and only if

$$\|r_k\| = \mathrm{O}(\|H(x_k)\|^{1+p}) \quad \text{as } k \to \infty. \tag{4.21}$$

Proof. Since H is BD-regular zero at x^*, there is a neighborhood $\mathcal{N}$ of x^* and a constant $\mu > 0$ such that for any $y \in \mathcal{N}$ and $V \in \partial_{\mathrm{B}} H(y)$, V is nonsingular and $\max\{\|V\|, \|V^{-1}\|\} \leqslant \mu$. By the equation

$$\begin{aligned} r_k &= H(x_k) + V_k D_k^{-1}\widehat{s}_k \\ &= [H(x_k) - H(x^*) - V_k(x_k - x^*)] + V_k(x_{k+1} - x^*), \end{aligned} \tag{4.22}$$

we have the first term in the right hand of (4.22) is $\mathrm{o}(\|x_k - x^*\|)$ since H is semismooth and BD-regular zero at x^*. Assume that $x_k \to x^*$ superlinearly, therefore, we have that any $V_k \in \partial_{\mathrm{B}} H(x_k)$ is nonsingular and uniformly bounded,

$$\|r_k\| = \mathrm{O}(\|x_{k+1} - x^*\|) = \mathrm{o}(\|x_k - x^*\|) = \mathrm{o}(\|H(x_k)\|).$$

On the other hand, since $V_k \in \partial_{\mathrm{B}} H(x_k)$ is nonsingular and $\max\{\|V_k\|, \|V_k^{-1}\|\} \leqslant \mu$, then we have that (4.22) means

$$\|x_{k+1} - x^*\| = \mathrm{o}(\|r_k\|) = \mathrm{o}(\|H(x_k)\|) = \mathrm{o}(\|x_k - x^*\|) \quad \text{as } k \to \infty.$$

If H is p-order semismooth at x^*, then the proof is essentially the same p-order semismooth and

$$H(x_k) - H(x^*) - V_k(x_k - x^*) = \mathrm{O}(\|x_k - x^*\|^{1+p})$$

instead of the semismooth of H. The conclusions of Theorem 4.2 hold. □

The above theorem characterizes the order of the convergence of the inexact generated Newton iterates in terms of the rate of convergence of the relative residuals. Let $x^* \in \Omega$ be a BD-regular zero of H at which H is p-order semismooth, then there exist $\delta > 0$ and $\zeta' > \zeta > 0$ such that

$$\zeta\|x_k - x^*\| \leqslant \|H(x_k)\| \leqslant \zeta'\|x_k - x^*\|$$

for all $\|x_k - x^*\| \leqslant \delta$. Therefore,

$$\begin{aligned}\frac{\zeta'}{1-\rho_k} &= \frac{\zeta'\|x_k - x^*\|}{\|x_k - x^*\| - \|x_{k+1} - x^*\|} \geqslant \frac{\|H(x_k)\|}{\|x_{k+1} - x_k\|}\\ &\geqslant \frac{\zeta\|x_k - x^*\|}{\|x_{k+1} - x^*\| + \|x_k - x^*\|} = \frac{\zeta}{1+\rho_k},\end{aligned}$$

where $\rho_k = \frac{\|x_{k+1}-x^*\|}{\|x_k - x^*\|}$. Then an equivalent result in terms of the steps $\{s_k\}$ for generalized-Newton methods is expressed in our notion that $x_k \to x^*$ superlinear if and only if

$$\|r_k\| = \mathrm{o}(\|s_k\|) \quad \text{as } k \to \infty, \tag{4.23}$$

where $r_k = H(x_k) + V_k D_k^{-1}\widehat{s}_k$, and $x_k \to x^*$ with order at least $1+p$ if and only if

$$\|r_k\| = \mathrm{O}(\|s_k\|^{1+p}) \quad \text{as } k \to \infty, \tag{4.24}$$

proved H is p-order semismooth at x^*. It should be noted that these conditions were not originally stated in terms of the residuals r_k and that, since they are not scale-invariant, they do not suggest a natural criterion for when to accept an approximate solution to the Newton equations.

If the condition

$$\|H(x_k) + V_k D_k^{-1}\widehat{s}_k\| \leqslant \eta_k \|H(x_k)\|$$

is replaced by the stronger condition which the forcing sequence $\{\eta_k\}$ is p-order for $\{\|H(x_k)\|\}$

$$\|H(x_k) + V_k D_k^{-1}\widehat{s}_k\| \leqslant \eta \|H(x_k)\|^{1+p}, \quad \forall k = 0, 1, \ldots, \tag{4.25}$$

that is, $\eta_k = \eta\|H(x_k)\|^p$, where η is any nonnegative constant, then one can show $1+p$ order superlinear convergence of the iterates $\{x_k\}$.

Theorem 4.4. *Suppose that H is locally Lipschizian, p-order semismooth and BD-regular at x^* which is a solution of* (1.1). *Assume also that in a neighborhood $\mathcal{N}$ of x^*, for any $y \in \mathcal{N}$ and $V \in \partial_B H(y) \cup \partial_B H(x^*)$, the following inequality holds*

$$\|H(y) - H(x^*) - V(y - x^*)\| \leqslant \gamma \|y - x^*\|^{1+p}, \tag{4.26}$$

where γ is called p-order semismooth constant at x^*. Then there exists an $\varepsilon \in (0, 1]$ such that $\|x_0 - x^*\| \leqslant \varepsilon$, the sequence $\{x_k\}$ generated by the forcing sequence (4.25) *superlinearly converges to x^* with order at least $1+p$ in the sense that*

$$\|x_{k+1} - x^*\| \leqslant \beta[\gamma + \eta(\beta' + \gamma\varepsilon^p)^{1+p}]\|x_k - x^*\|^{1+p}, \quad \forall k = 0, 1, \ldots, \tag{4.27}$$

where $\beta = \max\{\|V^{-1}\| \,|\, V \in \partial H(x^*)\}$ and $\beta' = \max\{\|V\| \mid V \in \partial H(x^*)\}$. Moreover,

$$\limsup_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^{1+p}} \leqslant \beta(\gamma + \eta(\beta')^{1+p}). \tag{4.28}$$

Proof. Since H is BD-regular zero solution of the problem (1.1) at x^* and (4.26) holds with p-order semismooth constant γ, we now choose $\varepsilon \in (0, 1]$ sufficiently small that for any $y \in \mathcal{N}(x^*, \varepsilon) \equiv \{y \mid \|y - x^*\| \leqslant \varepsilon\}$ and $V \in \partial_B H(y)$ and $V_* \in \partial_B H(x^*)$,

$$\|H(y) - H(x^*) - V(y - x^*)\| \leqslant \gamma \|y - x^*\|^{1+p},$$
$$\|H(y) - H(x^*) - V_*(y - x^*)\| \leqslant \gamma \|y - x^*\|^{1+p}.$$

Since H is BD-regular zero solution of the problem (1.1) at x^*, $\varepsilon > 0$ is chosen sufficiently small such that the semismooth constant γ and constants β, β' satisfy

$$\beta(\gamma + \eta(\beta' + \gamma\varepsilon^p)^{1+p})\varepsilon^p \leqslant \tfrac{1}{2}. \tag{4.29}$$

Assume that $\|x_0 - x^*\| \leqslant \varepsilon$. We prove by induction on k that (4.27) holds at each step and that

$$\|x_{k+1} - x^*\| \leqslant \tfrac{1}{2} \|x_k - x^*\|^{1+p}, \ \forall k = 0, 1, 2, \ldots,$$

giving $x_{k+1} \in \mathcal{N}(x^*, \varepsilon)$.

From $\|x_0 - x^*\| \leqslant \varepsilon$ and $\|H(x_0) + V_0 D_0^{-1}\widehat{s}_0\| \leqslant \eta \|H(x_0)\|^{1+p}$, it follows that, defining $r_k = H(x_k) + V_k D_k^{-1}\widehat{s}_k$,

$$\begin{aligned} x_1 - x^* &= x_0 - x^* + V_0^{-1}[r_0 - H(x_0)] \\ &= V_0^{-1} r_0 - V_0^{-1}[H(x_0) - H(x^*) - V_0(x_0 - x^*)] \end{aligned} \tag{4.30}$$

and taking norms,

$$\begin{aligned} \|x_1 - x^*\| &\leqslant \|V_0^{-1}\| \ [\|r_0\| + \|H(x_0) - H(x^*) - V_0(x_0 - x^*)\| \ ] \\ &\leqslant \beta \ [\eta \|H(x_0)\|^{1+p} + \gamma \|x_0 - x^*\|^{1+p}]. \end{aligned} \tag{4.31}$$

Taking $y = x_0$ in (4.26)

$$\|H(x_0) - H(x^*) - V_0(x_0 - x^*)\| \leqslant \gamma \|x_0 - x^*\|^{1+p}$$

and

$$H(x_0) = V_*(x_0 - x^*) + [H(x_0) - H(x^*) - V_*(x_0 - x^*)]$$

taking norms,

$$\begin{aligned} \|H(x_0)\| &\leqslant \|V_*(x_0 - x^*)\| + \|H(x_0) - H(x^*) - V_*(x_0 - x^*)\| \\ &\leqslant (\|V_*\| + \gamma \|x_0 - x^*\|^p) \|x_0 - x^*\|. \end{aligned}$$

Hence, by (4.31)

$$\begin{aligned} \|x_1 - x^*\| &\leqslant [\beta\gamma + \beta\eta(\beta' + \|x_0 - x^*\|^p \gamma)^{1+p}] \|x_0 - x^*\|^{1+p} \\ &\leqslant [\beta\gamma + \beta\eta(\beta' + \gamma\varepsilon^p)^{1+p}] \|x_0 - x^*\|^{1+p}. \end{aligned} \tag{4.32}$$

By (4.31), $\|x_1 - x^*\| \leqslant \frac{1}{2}\|x_0 - x^*\| \leqslant \varepsilon$, so that $x_1 \in \mathcal{N}(y^*, \varepsilon)$. This completes the case $k = 0$. The proof of the induction step is identical.

To (4.30), we have from (4.32) that

$$\|x_{k+1} - x^*\| \leqslant [\beta\gamma + \beta\eta(\beta' + \gamma\|x_k - x^*\|)^{1+p}]\|x_k - x^*\|^{1+p}, \text{ for } k = 0, 1, 2, \ldots.$$

Since $x_k \to x^*$ as $k \to \infty$, we immediately obtain (4.28). □

The above result gives that the steps $\{s_k\}$ for the inexact generalized Newton iterates can be accepted in which the order of the convergence of iterates depends on $\{\eta_k\}$.

5. Applications

The semismooth equations arise from the reformulation of the following mixed complementarity problems (MCP). In order to introduce the MCP, it is quite convenient to consider the variational inequality problem first. We only sketch the idea here and do not state any formal results. For a given function $F : \Re^n \to \Re^n$ and a nonempty, closed and convex set $\Omega \subseteq \Re^n$, this variational inequality problem consists in finding a point $y \in \Omega$ such that

$$F(y)^{\mathrm{T}}(x - y) \geqslant 0, \quad \forall x \in \Omega \stackrel{\text{def}}{=} \{x \mid l \leqslant x \leqslant u\}.$$

It can be seen that this is equivalent to $x \in \Re^n$ satisfying the following conditions: for every $i = 1, \ldots, n$,

$$\begin{aligned} &\text{if } F_i(y) > 0 \quad \text{then } x_i = l_i, \\ &\text{if } F_i(y) < 0 \quad \text{then } x_i = u_i, \\ &\text{if } F_i(y) = 0 \quad \text{then } l_i \leqslant x_i \leqslant u_i. \end{aligned}$$

It is well known and easy to see that this variational inequality problem is equivalent to the MCP, that is, the following nonlinear complementarity problems NCPs subject to bounded constraints on variables:

$$\begin{aligned} &x \geqslant 0, \quad F(x) \geqslant 0, \quad x^{\mathrm{T}} F(x) = 0, \quad \text{and} \\ &x \in \Omega \stackrel{\text{def}}{=} \{x \mid l \leqslant x \leqslant u\}, \end{aligned} \tag{5.1}$$

where $F : \Re^n \to \Re^n$ is assumed to be continuously differentiable and the n-dimensional feasible box constraint set Ω. It is easy to see that this mixed complementarity problem is the NCP when Ω is equal to the nonnegative orthant, i.e. if $\Omega = [0, \infty)$. The NCP is also a general framework for optimality conditions of mathematical programs. Hereby, the function $\phi : \Re^2 \to \Re$ is an NCP-function, i.e. it satisfies

$$\phi(a, b) = 0 \text{ if and only if } \quad a \geqslant 0, \ b \geqslant 0, \ ab = 0. \tag{5.2}$$

Probably the most popular NCP-function is the Fischer–Burmeister function which was used to study various methods for solving the NCPs and related problems (see [17] details)

$$\phi(a, b) \stackrel{\text{def}}{=} \phi_{\mathrm{FB}}(a, b) \stackrel{\text{def}}{=} \sqrt{a^2 + b^2} - a - b, \tag{5.3}$$

which is semismooth. In particular, these assumptions hold if F is Lipschitz continuously differentiable and if $\phi=\phi_{\mathrm{FB}}$ is chosen. For details on the variety of available NCP-functions, for example, other popular NCP-functions are min-function

$$\phi(a,b) \overset{\text{def}}{=} \phi_{\min}(a,b) \overset{\text{def}}{=} \min\{a,b\}$$

and max-function

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

where $z_{+} \overset{\text{def}}{=} \max\{0,z\}$ for $z \in \Re$ (also, referred to [17]). From this property, the NCP-function given in (5.2) can readily be recast as the system of semismooth equations

$$H(x) \overset{\text{def}}{=} \begin{pmatrix} H_1(x) \\ \vdots \\ H_n(x) \end{pmatrix} \overset{\text{def}}{=} \begin{pmatrix} \phi(x_1, F_1(x)) \\ \vdots \\ \phi(x_n, F_n(x)) \end{pmatrix} = 0. \tag{5.4}$$

In the sense that x solves (5.1) if and only if x solves (5.4) and $x \in \Omega$. Further, define a merit function $h : \Re^n \to \Re^1$ as

$$h(x) \overset{\text{def}}{=} \frac{1}{2}\sum_{i=1}^{n}[\phi(x_i, F_i(x))]^2 \overset{\text{def}}{=} \frac{1}{2}\|H(x)\|_2^2.$$

Finally, we discussed an affine scaling trust region approach to the affine scaling inexact generalized Newton step and hence, the step size Λ_k given in (2.23) along the direction s_k is bounded away from zero, moreover, $\Lambda_k \to +\infty$ as $s_k \to 0$. An augmented affine scaling trust region model is based on the local linear approximation of the squared Euclidean norm of the semismooth systems (1.1) at x_k

$$(S_k) \quad \min \phi_k(\widehat{d}) \overset{\text{def}}{=} \tfrac{1}{2}\|V_k D_k^{-1}\widehat{d} + H_k\|^2 + \tfrac{1}{2}\widehat{d}^{\mathrm{T}} C_k \widehat{d}$$
$$\text{s.t.} \|\widehat{d}\| \leqslant \Delta_k,$$

where Δ_k is the trust region radius, $V_k \in \partial H(x_k)$ and the diagonal matrix

$$C_k \overset{\text{def}}{=} \operatorname{diag}\{g_k\} J_k^{v} \tag{5.5}$$

suggested by Coleman and Li in [4], where $g(x) \overset{\text{def}}{=} \nabla h(x) = V^{\mathrm{T}} H(x)$ and $J^{v}(x) \in \Re^{n\times n}$ is the Jacobian matrix of $|v(x)|$ whenever $|v(x)|$ is differentiable. Each diagonal component of the diagonal matrix J^{v} equals zero or ± 1. Subproblem (S_k) is not solved exactly. A rather coarse solution is sufficient to guarantee basic global convergence.

It is well known (see [16]) that $d_k = D_k \widehat{d}_k$ is a solution to the subproblem (S_k) if and only if d_k is a solution to the following equations of the forms

$$[D_k^{-1}(V_k^{\mathrm{T}} V_k) D_k^{-1} + C_k + \mu_k I]\widehat{d}_k = -D_k^{-1} g_k, \tag{5.6}$$

$$\mu_k(\|\widehat{d}_k\| - \Delta_k) = 0, \quad \mu_k \geqslant 0. \tag{5.7}$$

Then there exist $\tau \in (0, \frac{1}{2})$ and $\kappa > 0$ such that

$$\phi_k(\widehat{d}) - \phi_k(0) \geqslant \tau \| D_k^{-1} g_k \| \min \left\{ \Lambda_k, \frac{\| D_k^{-1} g_k \|}{\kappa} \right\}. \tag{5.8}$$

As $\{d_k\}$ converges to zero, $[\mathrm{diag}\{g_k\} J_k^{\gamma} + \mu_k I]$ is a positive semidefinite diagonal matrix, and x^* is nondegenerate with $D_*^{-1} g_* = 0$, for any i with $(v_*)_i = 0$, $(d_k)_i$ and $(g_k)_i$ have the same sign for k sufficiently large. Hence, if Λ_k given in (2.23) is defined by some $(v_*)_j = 0$ and $(g_*)_j \neq 0$, then $\Lambda_k = \frac{|(v_k)_j|}{|(d_k)_j|}$ for k sufficiently large. Using (5.6),

$$\Lambda_k = \frac{|(g_k)_j| + \lambda_k}{|(g_k)_j + (V_k^{\mathrm{T}} V_k d_k)_j|} \geqslant \frac{|(g_k)_j| + \lambda_k}{\| g_k + V_k^{\mathrm{T}} V_k d_k \|_\infty}. \tag{5.9}$$

Assume that any limit point of $\{x_k\}$ is nondegenerate, similar to proof of the convergence in [4], then Λ_k given in (2.21) is bounded away from zero. Furthermore, Λ_k given in (2.23) is bounded away from 1 as $s_k^{\mathrm{N}} \to 0$.

Concluding remark. We have presented a new affine scaling inexact generalized Newton algorithm in association with nonmonotone interior backtracking line technique for solving systems of semismooth equations subject to bounds on variables. Important examples of the bounded semismooth equations are reformulations of NCPs. In particular, the augmented affine scaling trust region algorithm can be used to solve the inexact generalized Newton steps.

Acknowledgements

The author would like to thank the anonymous referee for their detailed comments and helpful suggestions.

References

[1] S. Bellavia, M. Macconi, B. Morini, An affine scaling trust-region approach to bound-constrained nonlinear systems, Appl. Numer. Math. 44 (2003) 257–280.
[2] P.N. Brown, Y. Saad, Convergence theory of nonlinear Newton–Kryloy algorithms, SIAM J. Optim. 4 (1994) 297–330.
[3] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.
[4] T.F. Coleman, Y. Li, An interior trust region approach for minimization subject to bounds, SIAM J. Optim. 6 (3) (1996) 418–445.
[5] R.S. Dembao, S.C. Eisenstat, T. Steinaug, Inexact Newton methods, SIAM J. Numer. Anal. 19 (1982) 400–408.
[6] N.Y. Deng, Y. Xiao, F.J. Zhou, A nonmonotonic trust region algorithm, J. Optim. Theory Appl. 76 (1993) 259–285.
[7] J.E. Dennis Jr., J.J. Moré, Quasi-Newton methods, motivation and theory, SIAM Rev. 19 (1977) 46–89.
[8] J.E. Dennis Jr., R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall Series in Computation; Mathematics, Englewood Cliffs, NJ, 1983.
[9] S.C. Eisenstat, H.F. Walker, Globally convergence inexact Newton methods, SIAM J. Optim. 4 (1994) 393–422.
[10] L. Gripp, F. Lampariello, S. Lucidi, A nonmonotone line search technique for Newton's methods, SIAM J. Numer. Anal. 23 (1986) 707–716.
[11] J. Nocedal, Y. Yuan, Combining trust region and line search techniques, in: Y. Yuan (Ed.), Advances in Nonlinear Programming, Kluwer, Dordrecht, 1998, pp. 153–175.

[12] J.S. Pang, L. Qi, Nonsmooth equation: Motivation and algorithms, SIAM J. Optim. 3 (1993) 443–465.
[13] L. Qi, Convergence analysis of some algorithms for solving nonsmooth equations, Math. Oper. Res. 18 (1993) 227–244.
[14] L. Qi, J. Sun, A nonsmooth version of Newton's method, Math. Programming 58 (1993) 353–367.
[15] A. Shapiro, On Concepts of directional Differentiability, J. Optim. Theory Appl. 66 (1990) 477–487.
[16] D.C. Sorensen, Newton's method with a model trust region modification, SIAM J. Numer. Anal. 19 (1982) 409–426.
[17] D. Sun, L. Qi, On NCP-function, Comput. Optim. Appl. 13 (1999) 201–220.
[18] M. Ulbrich, Nonmonotone trust-region methods for bound-constrained semismooth systems of equations with applications to nonlinear mixed complementarity problems, SIAM J. Optim. 11 (2001) 889–917.