Available online at www.sciencedirect.com
ScienceDirect

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

Journal of Computational and Applied Mathematics 219 (2008) 198–215

www.elsevier.com/locate/cam

# Affine scaling interior Levenberg–Marquardt method for bound-constrained semismooth equations under local error bound conditions

Detong Zhu

*Department of Mathematics, Shanghai Normal University, Shanghai 200234, PR China*

Received 12 July 2005; received in revised form 4 July 2007

**Abstract**

We develop and analyze a new affine scaling Levenberg–Marquardt method with nonmonotonic interior backtracking line search technique for solving bound-constrained semismooth equations under local error bound conditions. The affine scaling Levenberg–Marquardt equation is based on a minimization of the squared Euclidean norm of linear model adding a quadratic affine scaling matrix to find a solution that belongs to the bounded constraints on variable. The global convergence results are developed in a very general setting of computing trial directions by a semismooth Levenberg–Marquardt method where a backtracking line search technique projects trial steps onto the feasible interior set. We establish that close to the solution set the affine scaling interior Levenberg–Marquardt algorithm is shown to converge locally Q-superlinearly depending on the quality of the semismooth and Levenberg–Marquardt parameter under an error bound assumption that is much weaker than the standard nonsingularity condition, that is, BD-regular condition under nonsmooth case. A nonmonotonic criterion should bring about speed up the convergence progress in the contours of objective function with large curvature.


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

*Keywords:* Semismooth equation; Levenberg–Marquardt method; Affine scaling; Interior point; Superlinear convergence; Error bounds

## 1. Introduction

In this paper, we consider and analyze the problem of finding a solution of nonsmooth equation systems subjective 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 $H : \mathscr{X} \subseteq \Re^n \to \Re^m$ is defined on the open set $\mathscr{X}$ containing the $n$-dimensional feasible box constraint set $\Omega \stackrel{\text{def}}{=} [l, u] \stackrel{\text{def}}{=} \{x \in \Re^n \mid l_i \leqslant x_i \leqslant u_i,\ i=1,\ldots,n\}$. The vectors $l \in (\Re \cup \{-\infty\})^n$ 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. Note that the dimensions $n$ and $m$ do not necessarily coincide, that is, we consider systems of nonlinear equations (whose generalized Jacobian

*E-mail address:* dtzhu@shnu.edu.cn.

0377-0427/$ - see front matter 
doi:10.1016/j.cam.2007.07.039

not necessarily square) and want to find a solution that belongs to a certain feasible set. We denote by $X^*$ the set of solutions to the constrained systems (1.1).

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 $\Omega$ drawn from mixed nonlinear complementarity problems, nonlinear optimization and variational inequality problems have been described. In the classic methods for solving an unconstrained square system of nonlinear equations, when the function $H(x)$ is a continuously differentiable function, quasi-Newton methods, Levenberg–Marquardt methods, etc. can be used that have local fast convergence properties under a nonsingularity (BD-regular condition under nonsmooth case where "BD" stands for Bouligand differential) assumption at the solution. The nonsingularity (or BD-regular condition) assumption implies that the solution is locally unique. Much analysis of many well-known algorithms have been done on smooth nonlinear equations but on nonsmooth nonlinear equations based on convergent analysis. Generally, a basis Gauss–Newton- or Levenberg–Marquardt-type approach has been used in order to ensure global convergence toward local minima of nonlinear least squares reformulation of unconstrained nonsmooth equations (see [8]). Recently, An and Bai also proposed the globally convergent Newton–GMRES methods for solving large unconstrained (sparse) systems of nonlinear equations (see [1–3]). The possibility of dealing with constrained nonsmooth equations is very important (see [4]). However, globally convergent methods for the unconstrained semismooth systems may be unsuited for the purpose of solving (1.1), since a vector $x^*$ satisfies $H(x) = 0$, but does not belong to $\Omega$. Ulbrich in [15] presented a class of double trust-region approaches with a projection onto the feasible set for bound-constrained semismooth square systems of equations (1.1). Ulbrich further proved that close to a regular solution the 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 the BD-regular condition and by allowing for inexactness in the computation of B-subdifferentials (where "B" stands for Bouligand). Recently, Kanzow et al. in [8] presented Levenberg–Marquardt-type algorithms for solving a strictly convex minimization problem in which the smooth function $H$ is not required nonsingularity assumption, but satisfies an error bound condition. The main disadvantage of this method is that it has to solve relatively complicated quadratic programming subproblems at each iteration in the special case where the set $\Omega$ is polyhedral, and convex minimization problems in the general case. The search direction generated in the subproblem must satisfy strict interior feasibility, which results in computational difficulties and hence the total computational effort for completing one iteration might be expensive and difficult. Stimulated by the progress in these aspects, we present a variant of affine scaling Levenberg–Marquardt-type method that solves only a system of linear equations per iteration in order to avoid the drawback of the complicated quadratic programming subproblems. The new proposed algorithm is locally Q-superlinearly convergent under a weaker assumption that, in particular, allows the solution set to be (locally) nonunique. To this end, we replace the nonsingularity (BD-regular condition) assumption by an error bound condition. This is motivated by the recent paper [8] that deals with convex constrained equations. Another nonmonotone idea also motivates the study of affine scaling Levenberg–Marquardt method in association with nonmonotone interior backtracking line search technique for approximating zeros of the semismooth equations (1.1) which should bring about speeding up the convergence progress in some ill-conditioned cases.

The organization of the article is as follows: In Section 2, we introduce the squared Euclidean norm to quadratic model of the semismooth systems (1.1) and design the nonmonotone affine scaling Levenberg–Marquardt algorithm with backtracking interior point technique for solving (1.1). In Section 3, we prove the global convergence of the proposed algorithm. We discuss further the convergence property and characterize the order of local convergence of the Newton methods in terms of the rates of the relative residuals without the nonsingularity (BD-regular condition under nonsmooth case) assumption in Section 4. We write the following notations.

*Notations*: $(x)_i$ denotes the $i$th component of the vector $x$. The Euclidean norm is denoted by $\|\cdot\|$, $\mathscr{B}_\delta(x) \stackrel{\text{def}}{=} \{y \in \Re^n \,|\, \|y - x\| \leqslant \delta\}$ is the closed ball centered at $x$ with radius $\delta > 0$, $\operatorname{dist}(y, X^*) \stackrel{\text{def}}{=} \inf\{\|y - x\| \,|\, x \in X^*\}$ denotes the distance from a point $y$ to the solution set $X^*$, and $P_\Omega(x)$ is the projection of a point $x \in \Re^n$ onto the feasible set $\Omega$.

## 2. Algorithm

This section describes and investigates the affine scaling Levenberg–Marquardt method in association with nonmonotonic interior point backtracking technique for solving a bound-constrained semismooth minimization reformulated by the bound-constrained semismooth systems (1.1) under a weaker assumption that, in particular, allows the solution set to be (locally) nonunique.

For convenience, we collect first concepts about nonsmooth analysis and we first assume that the function $H$ to be considered is locally Lipschizian. We say that $H : \mathscr{X} \subset \Re^n \to \Re^m$ is directionally differentiable at $x \in \mathscr{X} \subseteq \Re^n$ if the direction derivative

$$H'(x; d) \stackrel{\text{def}}{=} \lim_{\tau \to 0^+} \frac{H(x + \tau d) - H(x)}{\tau}$$

exists for all $d \in \Re^n$ and hence is said to be B-differentiable at a point $x$ if it is directional differentiable at $x$ and

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

In a finite-dimensional Euclidean space $\Re^n$, Shapiro [14] 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. Then for any $x \in \Re^n$ the generalized subdifferential of $H$ at $x$ in the sense of Clarke [5] is

$$\partial H(x) = \text{conv } \{\lim \nabla H(x_j) : x_j \to x, H \text{ is F-differentiable at } x_j\} \tag{2.2}$$

which is a nonempty convex compact set. We call $\partial_B H(x)$ the B-subdifferential of $H$ at $x$ whose concept and explanation were introduced in [10,11]. 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 where $y \to_d x$ is said that $y$ tends to $x$ in the direction $d$. 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. Furthermore, If $H : \mathscr{X} \subseteq \Re^n \to \Re^m$ is directionally differentiable at a neighborhood of $x$, then $H'(x; \cdot)$ is Lipschizian and there exists a $V \in \partial H(x)$ such that $H'(x; d) = Vd$ for any $d$. In [12], Qi and Sun gave the following lemma.

**Lemma 2.1.** *Suppose that $H : \Re^n \to \Re^m$ 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 $x$, that is, for every $\varepsilon > 0$ there exists a neighborhood $\mathscr{N}$ of $x$ such that for all $x + d \in \mathscr{N}$, $\|H'(x + d; d) - H'(x; d)\| \leqslant \varepsilon \|d\|$*;
(3) *for any $V \in \partial H(x + d)$, $d \to 0$, $Vd - H'(x; d) = o(\|d\|)$*;
(4) *$H$ is F-differentiable at any $x + d$, $\lim_{d \to 0} H'(x + d; d) - H'(x; d)/\|d\| = 0$.*

In [5], Clarke gave that for any $x, y \in \Re^n$,

$$H(y) - H(x) \in \text{conv}\, \partial H([x, y])(y - x), \tag{2.4}$$

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 $[x, y]$. It is known (see [10, Proposition 1]) that semismoothness of $H$ at $x$ implies that

$$\sup_{V \in \partial H(x+d)} \{H(x + d) - H(x) - Vd\} = o(\|d\|). \tag{2.5}$$

It is obvious that if $H$ is continuously differentiable in a neighborhood of $x \in \mathscr{X}$, then $H$ is semismooth at $x$ and $\partial H(x) = \partial_B H(x) = \{\nabla H(x)\}$.

Typically, global extension of the semismooth Newton methods requires the additional assumption of the natural merit function $h : \Re^n \to \Re^1$ as follows

$$h(x) \stackrel{\text{def}}{=} \tfrac{1}{2} H(x)^{\mathrm{T}} H(x),$$

where $h$ is continuously differentiable when the function $H$ is semismooth. We safeguard this locally convergent iteration by a Levenberg–Marquardt-type globalization that is based on the minimization reformulation

$$\text{minimize}\, h(x) \stackrel{\text{def}}{=} \tfrac{1}{2}\,\|H(x)\|^2 \quad \text{subject to } x \in \Omega \stackrel{\text{def}}{=} \{x | l \leqslant x \leqslant u\}. \tag{2.6}$$

As motivated above, a classical algorithm for solving the constrained semismooth equations (1.1) will be based on the reformulated problem (2.6). Basically, the concept of nonsmooth (Gauss–Newton) Levenberg–Marquardt-type method is to make Newton-like method globally convergent while maintaining its excellent local convergence behavior. Now, we begin the description of the affine scaling interior Levenberg–Marquardt-type method with its core, the underlying Newton-like iteration.

Ignoring primal and dual feasibility of the reformulated problem (2.6), the first-order necessary conditions for $x^*$ to be a local minimizer and

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

where $g(x) \stackrel{\text{def}}{=} \nabla h(x)$. The scaling matrix $D_k = D(x_k)$ arises naturally from examining the first-order necessary conditions for the bound-constrained nonlinear minimization transformed by the bound-constrained problem (1.1), where $D(x)$ is the diagonal scaling matrix suggested in [5] such that

$$D(x) \stackrel{\text{def}}{=} \operatorname{diag}\{|\gamma_1(x)|^{-1/2}, \ldots, |\gamma_n(x)|^{-1/2}\} \tag{2.7}$$

and the $i$th component of vector $\gamma(x)$ defined componentwise as follows:

$$\gamma_i(x) \stackrel{\text{def}}{=} \begin{cases} (x)_i - u_i & \text{if } (g)_i < 0 \text{ and } u_i < +\infty, \\ (x)_i - l_i & \text{if } (g)_i \geqslant 0 \text{ and } l_i > -\infty, \\ -1 & \text{if } (g)_i < 0 \text{ and } u_i = +\infty, \\ 1 & \text{if } (g)_i \geqslant 0 \text{ and } l_i = -\infty. \end{cases} \tag{2.8}$$

**Definition 2.1** (*Coleman and Li [6]*). A point $x \in \Omega$ is nondegenerate if, for each index $i$,

$$g_i(x) = 0 \Longrightarrow l_i < (x)_i < u_i, \tag{2.9}$$

where $g_i(x)$ is the $i$th component of vector $g(x)$. A reformulated problem (2.6) is nondegenerate if (2.9) holds for every $x \in \Omega$.

The Levenberg–Marquardt-type equation and the affine scaling matrix $D_k$ arise naturally from examining the Kuhn–Tucker conditions for the reformulated problem (2.6),

$$D^{-2}(x)g(x) = D^{-2}(x)V(x)^{\mathrm{T}}H(x) = 0 \quad \text{for } V(x) \in \partial_B H(x). \tag{2.10}$$

We remark that, even though $D(x)$ may be undefined on the boundary of $\Omega$, $D(x)^{-1}$ can be extended continuously to it. We will denote this extension as a convention by $D(x)^{-1}$ for all $x \in \Omega$. The basic idea is based on the local linear approximation of the squared Euclidean norm of the semismooth systems (2.6) at $x_k$ and hence the affine scaling trust region subproblem is

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

where $\Delta_k$ is the trust region radius, $M_k$ is an approximation to $V_k \in \partial H(x_k)$ or $M_k = V_k \in \partial H(x_k)$. The Levenberg–Marquardt method is a modified trust region strategy that is designed to maintain advantages of trust region method. Examining the Kuhn–Tucker conditions (2.10) and considering the transformation $\widehat{d}_k = D_k d_k$, we take the continuous differentiability of the merit function $h$ for granted, and return to building the global minimum of the affine scaling quadratic model

$$(P_k)\ \min \psi_k(\widehat{d}) \stackrel{\text{def}}{=} \tfrac{1}{2}\,\|H_k\|^2 + (D_k^{-1}g_k)^{\mathrm{T}}\widehat{d} + \tfrac{1}{2}\,\widehat{d}^{\mathrm{T}} D_k^{-1}(M_k^{\mathrm{T}}M_k)D_k^{-1}\widehat{d} + \tfrac{1}{2}\,v_k\widehat{d}^{\mathrm{T}}\widehat{d},$$

where $v_k > 0$ is a positive parameter and quadratic affine scaling matrix model $\frac{1}{2}\,v_k\widehat{d}^{\mathrm{T}}\widehat{d}$ is added instead of the affine scaling trust region subproblem. We now state an affine scaling Levenberg–Marquardt-type method applied to the solution of the semismooth problem (1.1). Let $\widehat{d}_k$ be the solution of the subproblem $(P_k)$. Since $\psi_k(\widehat{d})$ is a strick convex function, $\widehat{d}_k$ is also the global minimum of the subproblem $(P_k)$ which is in fact equivalent to solving the following affine scaling Levenberg–Marquardt-type equation

$$(D_k^{-1}M_k^{\mathrm{T}}M_kD_k^{-1} + v_kI)\widehat{d}_k = -D_k^{-1}\nabla h(x_k) = -D_k^{-1}V_k^{\mathrm{T}}H_k, \tag{2.11}$$

where $M_k \cong V_k \in \partial_B H(x_k)$ and $v_k > 0$ is a positive parameter. The relevance of the used affine scaling matrix $D_k^{-1}$ and scaling matrix $v_kI$ depends on the fact that the affine scaled Levenberg–Marquardt trial step $d_k = D_k^{-1}\widehat{d}_k$ is angled away from the approaching bound. Consequently the bounds will not prevent a relatively large stepsize along $d_k$ from being taken. In order to maintain the strict interior feasibility, a step-back tracking along the solution $d_k$ of the Eq. (2.11) could be required by the strict interior feasibility and nonmonotonic line research technique.

Now, we describe an affine scaling Levenberg–Marquardt algorithm with nonmonotonic interior backtracking line search technique for approximating zeros of the bound-constrained semismooth Equations (1.1) under an error bound assumption.

**Algorithm**

**Initialization step**

Choose parameters $\beta \in (0, \frac{1}{2})$, $\omega \in (0, 1)$, $\varepsilon > 0$, $0 < \theta_l < 1$ and positive integer $M$ as nonmonotonic parameter. Let $m(0) = 0$. Give a starting point $x_0 \in \mathrm{int}(\Omega) \subset \Re^n$, select an initial matrix $M_0$ approximate to $V_0 \in \partial H(x_0)$. Set $k = 0$, go to the main step.

**Main step**

1. Evaluate $h_k = h(x_k) \stackrel{\text{def}}{=} \frac{1}{2}\,\|H(x_k)\|^2$ and $M_k \cong V_k \in \partial H(x_k)$. Calculate $D_k$ given in (2.7) and $g_k = \nabla h(x_k) \stackrel{\text{def}}{=} V_k^{\mathrm{T}}H_k$.
2. If $\|D_k^{-1}g_k\| \leqslant \varepsilon$, stop with the approximate solution $x_k$.
3. Solve a step $\widehat{d}_k$, based on the affine scaled Levenberg–Marquardt equation

$$(D_k^{-1}M_k^{\mathrm{T}}M_kD_k^{-1} + v_kI)\widehat{d}_k = -D_k^{-1}\nabla h(x_k) \tag{2.12}$$

and set

$$d_k = D_k^{-1}\widehat{d}_k. \tag{2.13}$$

4. Choose $\alpha_k = 1, \omega, \omega^2, \ldots, \omega^{l_k}$ with $l_k$ the smallest nonnegative integer such that

$$h(x_k + \alpha_k d_k) \leqslant h(x_{l(k)}) + \alpha_k\beta g_k^{\mathrm{T}} d_k \tag{2.14}$$

$$\text{with } x_k + \alpha_k d_k \in \Omega, \tag{2.15}$$

where $h(x_{l(k)}) = \max_{0\leqslant j\leqslant m(k)}\{h(x_{k-j})\}$.

5. Set

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

where $\theta_k \in (\theta_l, 1]$ and $\theta_k - 1 = \mathrm{O}(\|d_k\|)$ and then set

$$x_{k+1} = x_k + s_k. \tag{2.16}$$

6. Take the nonmonotone control parameter $m(k+1) = \min\{m(k) + 1, M\}$ and update $M_k$ to obtain $M_{k+1} \cong V_{k+1} \in \partial H(x_{k+1})$. Then set $k \leftarrow k + 1$ and go to step 1.

**Remark 1.** The scalar $\alpha_k$ given in (2.15) of step 4, denotes the step size along the direction $d_k$ to the boundary on the variables $l \leqslant x_k + \alpha_k d_k \leqslant u$, that is, $\alpha_k \in (0, \Gamma_k]$ and

$$\Gamma_k \stackrel{\text{def}}{=} \min\left\{\max\left\{\frac{l_i - (x_k)_i}{(d_k)_i}, \frac{u_i - (x_k)_i}{(d_k)_i}\right\}, i = 1, 2, \ldots, n\right\} \tag{2.17}$$

where $\Gamma_k = +\infty$ if $(d_k)_i = 0$ for all $i = 1, 2, \ldots, n$.

**Remark 2.** In order to investigate the convergence properties of our algorithm, we assume that the termination parameter $\varepsilon$ is equal to zero and $M_k = V_k \in \partial H(x_k)$. We further note that the proposed algorithm is well defined since $v_k > 0$ and the search direction $\widehat{d_k}$ is always a descent direction for the merit function $h$.

In order to obtain the global convergence result, for the sake of simplicity, we assume that from examining the Kuhn–Tucker conditions (2.10), $v_k$ is given in this paper by

$$v_k \stackrel{\text{def}}{=} \eta\|\nabla h(x_k)\| = \eta\|V_k^{\mathrm{T}} H_k\|, \quad V_k \in \partial H(x_k) \tag{2.18}$$

for a certain constant $\eta \geqslant 1$, although several other choices of $v_k$ yield the same result including the more realistic choices

$$v_k \stackrel{\text{def}}{=} \min\{\eta_1, \eta_2\|D_k^{-1}\nabla h(x_k)\|\}$$

for certain constants $\eta_1, \eta_2 > 0$. Note that these choices are consistent with the requirements for local superlinear/quadratic convergence in the following sections.

## 3. Convergence analysis

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

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

The following assumption is commonly used in convergence analysis of most methods for the constrained systems.

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

**Assumption (A2).** There exist some positive constants $\chi_D$ and $\chi_V$ such that

$$\|D(x)^{-1}\| \leqslant \chi_D, \quad \|V\| \leqslant \chi_V, \quad \forall V \in \partial H(x), \quad \forall x \in \mathscr{L}(x_0). \tag{3.1}$$

Ulbrich established the continuous differentiability of the merit function $h$ which can be found from Lemma 4.2 in [15].

**Lemma 3.1.** *Under the Assumptions* (A1) *and* (A2) *on the mapping* $H$, *the merit function* $h(x) \stackrel{\text{def}}{=} \frac{1}{2}\|H(x)\|^2$ *is continuously differentiable on* $\mathscr{X}$ *with gradient* $\nabla h(x) = V^{\mathrm{T}} H(x)$, *where* $V \in \partial H(x)$ *is arbitrary.*

**Lemma 3.2.** *If* $\|D_k^{-1} g_k\| \neq 0$, *then the proposed algorithm will produce an iterate* $x_{k+1} = x_k + \alpha_k d_k$ *in a finite number of backtracking steps in* (2.14)–(2.15).

**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 $\|x_k - x\| \leqslant \delta$. Hence, $v(x) \stackrel{\text{def}}{=} \eta\|g(x)\| \geqslant \eta\|D(x)^{-1} g(x)\| / \|D(x)^{-1}\| \geqslant \eta\varepsilon/\chi_D$ for all $x$ with $\|x_k - x\| \leqslant \delta$. It is clearly to see that $\alpha_k$ will satisfy $\alpha_k \leqslant \Gamma_k$ in a finite number of backtracking reductions where $\Gamma_k$ given in (2.17). Using the mean value theorem, we have that with $0 \leqslant \vartheta_k \leqslant 1$,

$$\begin{aligned} h(x_k + \alpha_k d_k) = h(x_k) &+ \beta\alpha_k \nabla h(x_k)^{\mathrm{T}} d_k + \{(1-\beta)\alpha_k \nabla h(x_k)^{\mathrm{T}} d_k \\ &+ \alpha_k[\nabla h(x_k + \vartheta_k \alpha_k d_k) - \nabla h(x_k)]^{\mathrm{T}} d_k\}. \end{aligned} \tag{3.2}$$

Since $\nabla h(x)$ is Lipschitzian continuous, there exists sufficiently small $\alpha_k$ when $\|\vartheta_k \alpha_k d_k\| \leqslant \delta$ such that

$$|[\nabla h(x_k + \theta_k \alpha_k d_k) - \nabla h(x_k)]^{\mathrm{T}} d_k| \leqslant \frac{\eta\varepsilon(1-\beta)}{2\chi_D}\|d_k\|^2.$$

By the affine scaling Levenberg–Marquardt equation (2.12), we have that noting the matrix $V_k^{\mathrm{T}} V_k$ is positive semidefinite

$$\begin{aligned}[\nabla h(x_k)]^{\mathrm{T}} d_k &= -\widehat{d}_k^{\mathrm{T}}(D_k^{-1} V_k^{\mathrm{T}} V_k D_k^{-1} + v_k I)\widehat{d}_k \\ &\leqslant -v_k \widehat{d}_k^{\mathrm{T}} \widehat{d}_k \leqslant -\eta\|\nabla h(x_k)\|\|d_k\|^2 \leqslant -\frac{\eta\varepsilon}{\chi_D}\|d_k\|^2,\end{aligned} \tag{3.3}$$

where the last inequality is deduced by $\|d_k\| \leqslant \|D_k^{-1}\widehat{d}_k\| \leqslant \chi_D\|\widehat{d}_k\|$. This gives that after a finite number of reductions, the last term in brackets in the right-hand side of (3.2) will become negative and the corresponding $\alpha_k$ will be acceptable. Since $h(x_k) \leqslant h(x_{l(k)})$, the conclusion of the lemma holds. □

**Theorem 3.3.** *Let $\{x_k\}$ be a sequence generated by the proposed algorithm. Assume that Assumptions* (A1)–(A2) *and the nondegenerate condition of the reformulated problem* (2.6) *hold, then*

$$\liminf_{k\to\infty} \|D_k^{-1}\nabla h_k\| = \liminf_{k\to\infty} \|D_k^{-1} V_k^{\mathrm{T}} H_k\| = 0, \tag{3.4}$$

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

**Proof.** According to the acceptance rule in step 4, we have

$$h(x_{l(k)}) - h(x_k + \alpha_k d_k) \geqslant -\beta\alpha_k g_k^{\mathrm{T}} d_k = -\beta\alpha_k (D_k^{-1} V_k^{\mathrm{T}} H_k)^{\mathrm{T}}(D_k d_k). \tag{3.5}$$

Taking into account that $m(k+1) \leqslant m(k)+1$, and $h(x_{k+1}) \leqslant h(x_{l(k)})$, we have $h(x_{l(k+1)}) \leqslant \max_{0\leqslant j\leqslant m(k)+1}\{h(x_{k+1-j})\} = h(x_{l(k)})$. This means that the sequence $\{h(x_{l(k)})\}$ is nonincreasing for all $k$ and hence $\{h(x_{l(k)})\}$ is convergent.

By (2.16) and (3.3), for all $k > M$, we get

$$\begin{aligned}h(x_{l(k)}) &\leqslant \max_{0\leqslant j\leqslant m(l(k)-1)}\{h(x_{l(k)-j-1})\} + \alpha_{l(k)-1}\beta\nabla h_{l(k)-1}^{\mathrm{T}} d_{l(k)-1} \\ &\leqslant \max_{0\leqslant j\leqslant m(l(k)-1)}\{h(x_{l(k)-j-1})\} - \alpha_{l(k)-1}\beta\eta\|V_{l(k)-1}^{\mathrm{T}} H_{l(k)-1}\|\|d_{l(k)-1}\|^2.\end{aligned} \tag{3.6}$$

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

$$\|D_k^{-1} V_k^{\mathrm{T}} H_k\| \geqslant \varepsilon, \quad k = 1, 2, \ldots . \tag{3.7}$$

Hence, $\|V_k^{\mathrm{T}} H_k\| \geqslant \|D_k^{-1} V_k^{\mathrm{T}} H_k\| / \|D_k^{-1}\| \geqslant \varepsilon/\chi_D$. As $\{h(x_{l(k)})\}$ is convergent, we obtain that from (3.7),

$$\lim_{k\to\infty} \alpha_{l(k)-1}\|d_{l(k)-1}\|^2 = 0.$$

Following the way of proof used in [7], we can also prove by induction that

$$\lim_{k\to\infty} \alpha_k\|d_k\|^2 = 0. \tag{3.8}$$

Therefore, we have that either Case (I)

$$\liminf_{k\to\infty} \alpha_k = 0, \tag{3.9}$$

or Case (II)

$$\lim_{k\to\infty} \|d_k\|^2 = 0 \tag{3.10}$$

holds.

Case (I), assume that $\alpha_k$ given in step 4 is the stepsize to the boundary of box constraints along $d_k$. From (2.17), we have

$$\Gamma_k \overset{\text{def}}{=} \min\left\{\max\left\{\frac{l_i-(x_k)_i}{(d_k)_i}, \frac{u_i-(x_k)_i}{(d_k)_i}\right\}, i=1,2,\ldots,n\right\}.$$

If (3.9) holds, we have that there exists a subset $\mathscr{K} \subset \{k\}$ such that

$$\lim_{k\to\infty,\ k\in\mathscr{K}} \alpha_k = 0$$

and hence, without loss of generality, assume $(x^*)_i = l_i$ for some $i$ where $x^*$ is any accumulation point of the sequence $\{x_k\}$ and without loss of generality, $\{x_k\}_{\mathscr{K}}$ a subsequence converging to $x^*$. Call (2.12), we can write

$$v_k d_k = -D_k^{-2}[\nabla h_k + (V_k^{\mathrm{T}} V_k) d_k]. \tag{3.11}$$

Since $v_k$ is a positive parameter, and $x^*$ is nondegenerate with $(v_*)_i = 0$ for any $i$, we have that from $x_i^* = l_i < u_i$ for some $i$, $(d_k)_i$ and $-(g_k)_i$ have the same sign for $k$ sufficiently large. Hence, if $\alpha_k$ is defined by some $(v_*)_j = 0$ and $(g_*)_j \neq 0$, then $\alpha_k = |(v_k)_j|/|(d_k)_j|$ for $k$ sufficiently large. Using (3.11), again, it is rewritten as follows:

$$\alpha_k = \frac{v_k}{|(g_k)_j + (V_k^{\mathrm{T}} V_k d_k)_j|} \geqslant \frac{v_k}{\|g_k + V_k^{\mathrm{T}} V_k d_k\|_\infty}. \tag{3.12}$$

It is clear that from (3.12) and $v_k = \eta\|g(x_k)\| = \eta\|V_k^{\mathrm{T}} H_k\| \geqslant \eta\varepsilon/\chi_D$, $\alpha_k$ given in step 4 is the stepsize to the boundary of box constraints along $d_k$,

$$\lim_{k\to\infty}\inf \alpha_k \geqslant \lim_{k\to\infty}\inf \frac{\eta\varepsilon}{\chi_D\|g_k + V_k^{\mathrm{T}} V_k d_k\|_\infty} > 0. \tag{3.13}$$

Furthermore, if (3.9) holds, the acceptance rule (2.14) means that, for large $k$,

$$\frac{\alpha_k}{\omega} g_k^{\mathrm{T}} d_k + \mathrm{o}\left(\frac{\alpha_k}{\omega}\|d_k\|\right) = h\left(x_k + \frac{\alpha_k}{\omega} d_k\right) - h_k \geqslant h\left(x_k + \frac{\alpha_k}{\omega} d_k\right) - h(x_{l(k)}) \geqslant \beta \frac{\alpha_k}{\omega} g_k^{\mathrm{T}} d_k.$$

Hence, we have

$$(1-\beta)\frac{\alpha_k}{\omega} g_k^{\mathrm{T}} d_k + \mathrm{o}\left(\frac{\alpha_k}{\omega}\|d_k\|\right) \geqslant 0. \tag{3.14}$$

Dividing (3.14) by $(\alpha_k/\omega)\|d_k\|$ and noting $g_k^{\mathrm{T}} d_k \leqslant 0$, we have

$$\lim_{k\to+\infty} \frac{g_k^{\mathrm{T}} d_k}{\|d_k\|} = 0. \tag{3.15}$$

From (3.3) and (3.7), we have that (3.15) means

$$0 = \lim_{k\to+\infty} \frac{g_k^{\mathrm{T}} d_k}{\|d_k\|} \leqslant \lim_{k\to+\infty} -\frac{\eta\varepsilon}{\chi_D} \frac{\|d_k\|^2}{\|d_k\|} \leqslant 0. \tag{3.16}$$

This also means that $\lim_{k\to+\infty, k\in\mathscr{K}} \|d_k\| = 0$ and hence (3.10) holds, that is, Case (II) holds.

On the other hand, i.e., Case (II), taking into account that $\{d_k\}_{\mathscr{K}} \to d_* = 0$ and $\{x_k\}_{\mathscr{K}} \to x^*$, we obtain from the affine scaling Levenberg–Marqwarddt equation (2.12) that

$$\begin{aligned} &-[D_k^{-1}\nabla h(x_k)]^{\mathrm{T}}[D_k^{-1} V_k^{\mathrm{T}} V_k D_k^{-1} + v_k I][D_k^{-1}\nabla h(x_k)] \\ &\quad = [D_k^{-1}\nabla h(x_k)]^{\mathrm{T}} \widehat{d}_k = [\nabla h(x_k)]^{\mathrm{T}} d_k \to 0. \end{aligned} \tag{3.17}$$

Since $\{x_k\}_{\mathscr{K}} \to x^*$, we get from the upper semicontinuity of the B-subdifferential that the sequence $\{V_k\}_{\mathscr{K}}$ is bounded. Without loss of generality, we therefore have $\{V_k\}_{\mathscr{K}} \to V_*$ for some matrix $V_* \in \partial_B H(x^*)$. Since $D(x)^{-1}\nabla h(x)$ is continuous, we also obtain $\{D_k^{-1}\nabla h(x_k)\}_{\mathscr{K}} \to D_*^{-1}\nabla h(x_*)$ and therefore $\{v_k\}_{\mathscr{K}} \to v_*$ with $v_* \stackrel{\text{def}}{=} \eta\|\nabla h(x_*)\|\geqslant 0$. We can obtain that

$$-[D_*^{-1}\nabla h(x^*)]^{\mathrm{T}}[D_*^{-1}V_*^{\mathrm{T}}V_*D_*^{-1}+v_*I]^{-1}[D_*^{-1}\nabla h(x^*)]=0.$$

This shows that $x^*$ is a stationary point of $D(x)^{-1}\nabla h(x)$. Hence the conclusion of the theorem is true. □

Theorem 3.3 indicates that at least one limit point of $\{x_k\}$ is a stationary point. Next, we shall extend this theorem to a stronger global convergent result.

**Theorem 3.4.** *Let $\{x_k\}$ be a sequence generated by the proposed algorithm. Assume that Assumptions* (A1)–(A2) *and the nondegenerate condition of the reformulated problem* (2.6) *hold*, *then*

$$\lim_{k\to+\infty}\|D_k^{-1}g_k\| = \lim_{k\to+\infty}\|D_k^{-1}V_k^{\mathrm{T}}H_k\|=0. \tag{3.18}$$

**Proof.** Assume that there are an $\varepsilon_1 \in (0,1)$ and a subsequence $\{D_{m_i}^{-1}g_{m_i}\}$ of $\{D_k^{-1}g_k\}$ such that for all $m_i$, $i=1,2,\ldots,$

$$\|D_{m_i}^{-1}g_{m_i}\|\geqslant\varepsilon_1. \tag{3.19}$$

Theorem 3.3 guarantees the existence of another subsequence $\{D_{n_i}^{-1}g_{n_i}\}$ such that

$$\|D_k^{-1}g_k\|\geqslant\varepsilon_2 \quad \text{for } m_i\leqslant k<n_i \tag{3.20}$$

and

$$\|D_{n_i}^{-1}g_{n_i}\|\leqslant\varepsilon_2 \tag{3.21}$$

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

Since the matrix $(D_k^{-1}V_k^{\mathrm{T}}V_kD_k^{-1}+v_kI)$ is nonsingular in the affine scaling Levenberg–Marqwardt equation (2.12), we have that

$$\begin{aligned}\nabla h(x_k)^{\mathrm{T}}d_k &= -[D_k^{-1}\nabla h(x_k)]^{\mathrm{T}}(D_k^{-1}V_k^{\mathrm{T}}V_kD_k^{-1}+v_kI)^{-1}[D_k^{-1}\nabla h(x_k)]\\ &\leqslant -\frac{\|D_k^{-1}\nabla h(x_k)\|^2}{\|D_k^{-1}V_k^{\mathrm{T}}V_kD_k^{-1}\|+v_k}\end{aligned} \tag{3.22}$$

and hence, by (3.3), i.e.,

$$\nabla h(x_k)^{\mathrm{T}}d_k\leqslant-\|D_k^{-1}\nabla h(x_k)\|\|d_k\|^2, \tag{3.23}$$

we have that

$$(\nabla h(x_k)^{\mathrm{T}}d_k)^2\geqslant\frac{\|D_k^{-1}\nabla h(x_k)\|^3}{\|(D_k^{-1}V_k^{\mathrm{T}}V_kD_k^{-1}\|+v_k}\|d_k\|^2. \tag{3.24}$$

This gives that from the matric $\|D_k^{-1}V_k^{\mathrm{T}}V_kD_k^{-1}\|$ and $v_k=\eta\|V_k^{\mathrm{T}}H_k\|$ being bounded

$$\nabla h(x_k)^{\mathrm{T}}d_k\leqslant-\frac{\|D_k^{-1}\nabla h(x_k)\|^{3/2}\|d_k\|}{\sqrt{\|D_k^{-1}V_k^{\mathrm{T}}V_kD_k^{-1}\|+v_k}}\leqslant-\frac{\sqrt{\varepsilon_2^3}\|d_k\|}{\chi_D\sqrt{\chi_V^2\chi_D^2+\eta\chi_V\chi_H}}. \tag{3.25}$$

Similar to the proof of Theorem 3.3, we have that the sequence $\{h(x_{l(k)})\}$ is nonincreasing for $m_i\leqslant k<n_i$, and hence $\{h(x_{l(k)})\}$ is convergent. Eqs. (3.20) and (3.6) mean that from setting $\upsilon\stackrel{\text{def}}{=}\sqrt{\varepsilon_2^3}/\chi_D\sqrt{\chi_V^2\chi_D^2+\eta\chi_V\chi_H}$,

$$h(x_{l(k)})\leqslant h(x_{l(l(k)-1)})-\beta\alpha_{l(k)-1}\upsilon\|d_{l(k)-1}\|. \tag{3.26}$$

That $\{h(x_{l(k)})\}$ is convergent means

$$\lim_{k\to\infty} \alpha_{l(k)-1}\|d_{l(k)-1}\| = 0.$$

Similar to the proof of (3.8) in Theorem 3.3, we have also that

$$\lim_{k\to\infty} \alpha_k\|d_k\| = 0. \tag{3.27}$$

Therefore, similar to the proof of (3.13), we can also get that there exists a subset $\mathscr{K} \subset \{k\}$ such that

$$\alpha_k \nrightarrow_{k\in\mathscr{K}} 0, \tag{3.28}$$

where $\alpha_k$ give in the step size to the boundary of box constraints along $d_k$, that is, the step size $\{\alpha_k\}$ cannot converge to zero.

Since $\nabla h(x)$ is continuous, and (3.27) holds, we have that for $\upsilon$ given in (3.26),

$$|[\nabla h(x_k + \xi_k\theta_k\alpha_k d_k) - \nabla h(x_k)]^{\mathrm{T}} d_k| \leqslant \tfrac{1}{2}(1-\beta)\upsilon\|d_k\|. \tag{3.29}$$

Similar to prove (3.2), using the mean value theorem, (3.25) and (3.29) mean that

$$\begin{aligned} h(x_k + \alpha_k\theta_k d_k) &= h(x_k) + \beta\alpha_k\theta_k\nabla h(x_k)^{\mathrm{T}} d_k + (1-\beta)\alpha_k\theta_k\nabla h(x_k)^{\mathrm{T}} d_k \\ &\quad + \alpha_k[\nabla h(x_k + \xi_k\theta_k\alpha_k d_k) - \nabla h(x_k)]^{\mathrm{T}} d_k \\ &\leqslant h(x_k) + \beta\alpha_k\theta_k\nabla h(x_k)^{\mathrm{T}} d_k, \end{aligned} \tag{3.30}$$

where $\xi_k \in [0, 1]$ and the last second inequality is deduced since the last term in brackets in the right-hand side of equality in (3.30) will become negative when $\alpha_k\theta_k\|d_k\|$ is small enough. And hence the corresponding $\theta_k \to 1$, as $\|d_k\| \to 0$. From (3.30) and (3.25), this means that for sufficiently large $i$, $m_i \leqslant k < n_i$,

$$h_k - h(x_k + s_k) \geqslant \alpha_k\beta\nabla h(x_k)^{\mathrm{T}} d_k \geqslant \beta\upsilon\theta_l\alpha_k\|d_k\|. \tag{3.31}$$

We then deduce from this bound that for $i$ sufficiently large,

$$\begin{aligned} \|x_{m_i} - x_{n_i}\| &\leqslant \sum_{k=m_i}^{n_i-1}\|x_k - x_{k+1}\| \leqslant \sum_{k=m_i}^{n_i-1}\alpha_k\|d_k\| \\ &\leqslant \frac{1}{\beta\upsilon\theta_l}\sum_{k=m_i}^{n_i-1}[h_k - h(x_k+s_k)] = \frac{1}{\beta\upsilon\theta_l}(h_{m_i} - h_{n_i}). \end{aligned} \tag{3.32}$$

Therefore, (3.24) implies that $h_{m_i} - h_{n_i}$ tends to zero as $i$ tends to infinity. (2.8) implies $|(v_{m_i})_j - (v_{n_i})_j| \leqslant |(x_{m_i})_j - (x_{n_i})_j| \to 0$, as $i$ tends to infinity. Finally, from (3.21)–(3.22) and triangle inequality, we get that from $\|V_{m_i}^{\mathrm{T}} H_{m_i} - V_{n_i}^{\mathrm{T}} H_{n_i}\| \leqslant \chi_V\|x_{m_i} - x_{n_i}\|$ and assuming $\|x_{m_i} - x_{n_i}\| \leqslant \varepsilon_2$,

$$\begin{aligned} \varepsilon_1 &\leqslant \|D_{m_i}^{-1} V_{m_i}^{\mathrm{T}} H_{m_i}\| \\ &\leqslant \|D_{m_i}^{-1}\|\|V_{m_i}^{\mathrm{T}} H_{m_i} - V_{n_i}^{\mathrm{T}} H_{n_i}\| + \|(D_{m_i}^{-1} - D_{n_i}^{-1})V_{n_i}^{\mathrm{T}} H_{n_i}\| + \|D_{n_i}^{-1} V_{n_i}^{\mathrm{T}} H_{n_i}\| \\ &\leqslant (\chi_D\varepsilon_2 + \chi_V\chi_H\varepsilon_2 + \varepsilon_2) \end{aligned}$$

which contradicts $\varepsilon_2 \in (0, \varepsilon_1)$, for arbitrarily small. From above, the conclusion of the theorem is true. □

## 4. The local convergence

Throughout this section we assume that the function $H$ is locally Lipschitz continuous in the region of interest. To establish the (local) convergence results for the proposed algorithm, we need the following assumptions in [16] under the basic properties of the B-differentiable function $H$ at any point $x^* \in X^*$.

**Assumption (A3).** The solution set $X^*$ of problem (1.1) is nonempty. For some solution $x^* \in X^*$, there exist constants $\delta > 0$, $\tau_1 > 0$, $\tau_2 > 0$ and local Lipschitz constant $L > 0$ such that the following inequalities hold:

$$\tau_1 \operatorname{dist}(x, X^*) \leqslant \|H(x)\|, \quad \forall x \in \mathscr{B}_\delta(x^*) \cap \Omega, \tag{4.1}$$

$$\|H(x) - H(x_k) - V_k(x - x_k)\| \leqslant \tau_2 \|x - x_k\|^2, \quad \forall V_k \in \partial H(x_k), \ \forall x, x_k \in \mathscr{B}_\delta(x^*) \cap \Omega, \tag{4.2}$$

$$\|H(x) - H(y)\| \leqslant L\|x - y\|, \quad \forall x, \ y \in \mathscr{B}_\delta(x^*) \cap \Omega. \tag{4.3}$$

Throughout this section, the constants $\delta$, $\tau_1$, $\tau_2$ and $L$ that appear in the subsequent analysis are always the constants from Assumption (A3).

Assumption (A3) only says that $H$ is locally Lipschitzian in a neighborhood of the solution $x^*$. Of course, this condition is automatically satisfied if $H$ is a semismooth function. Assumption (A3) is a local error bound condition and known to be much weaker than the more standard BD-regular zero of $H$ at which $H$ is semismooth in the case where the generalized subdifferential exists. Due to Hoffman's [13] famous error bound result, there exists $\tau > 0$ such that

$$\tau \operatorname{dist}(x, X^*) \leqslant \|H(x)\| + \|P_\Omega(x)\|. \tag{4.4}$$

If $x \in \mathscr{B}_\delta(x^*) \cap \Omega$ for some $x^* \in X^*$, then $P_\Omega(x) = 0$. So, (4.4) reduces to $\tau \operatorname{dist}(x, X^*) \leqslant \|H(x)\|$, which implies condition (4.1). Since the function $H$ is locally Lipschitz continuous in the region of interest, Pang [9] proved that if $H$ is B-differentiable on an open convex set $\mathscr{D}$, then for any $x, x + d, z \in \mathscr{D}$,

$$\|H(x + d) - H(x) - H'(z; d)\| \leqslant \sup_{0 \leqslant t \leqslant 1} \{\|H'(x + td; d) - H'(z; d)\|\}.$$

Therefore, if $H'(z; d)$ is Lipschitz continuous at $z \in \mathscr{D}$ with Lipschitz constant $L$, then

$$\|H(x + d) - H(x) - H'(z; d)\| \leqslant L \max\{\|x - z\|, \|x + d - z\|\}\|d\|,$$

which implies condition (4.2).

For this purpose of the locally convergent rate for the proposed algorithm, we need to prove following technical lemma.

**Lemma 4.1.** *Assume that Assumption* (A3) *holds. There exist constants* $\tau_3 > 0$ *and* $\tau_4 > 0$ *such that the following inequalities hold for each* $x_k \in \mathscr{B}_{\delta/2}(x^*) \cap \Omega$ *where* $x^* \in X^*$ *is some nondegenerate solution* (*here without loss of generality*, $\delta \leqslant \tau_1^2 / L\tau_2$)

(a) $\|d_k\| \leqslant \tau_3 \operatorname{dist}(x_k, X^*)$;
(b) $\|H(x_k) + V_k d_k\| \leqslant \tau_4 \operatorname{dist}(x_k, X^*)^{3/2}$.

**Proof.** (a) Let $\bar{x}_k \in X^*$. Denote the closest solution to $x_k$ so that

$$\|x_k - \bar{x}_k\| = \operatorname{dist}(x_k, X^*). \tag{4.5}$$

In fact the affine scaling Levenberg–Marquardt-type equation is equivalent that $\widehat{d_k}$ is the global minimum of the following subproblem:

$$(P_k) \min \psi_k(\widehat{d}) \stackrel{\text{def}}{=} \tfrac{1}{2}\|H_k\|^2 + (D_k^{-1} g_k)^{\mathrm{T}} \widehat{d} + \tfrac{1}{2} \widehat{d}^{\mathrm{T}} D_k^{-1} (M_k^{\mathrm{T}} M_k) D_k^{-1} \widehat{d} + \tfrac{1}{2} v_k \widehat{d}^{\mathrm{T}} \widehat{d}.$$

By the assumption $x_k \in B_{\delta/2}(x^*)$, we obtain

$$\|\bar{x}_k - x^*\| \leqslant \|\bar{x}_k - x_k\| + \|x_k - x^*\| \leqslant \|x^* - x_k\| + \|x_k - x^*\| \leqslant \delta$$

so that $\bar{x}_k \in \mathscr{B}_\delta(x^*) \cap \Omega$. Since

$$\|H(x_k)\|^2 = H(x_k)^{\mathrm{T}} [H(\bar{x}_k) + V_k(x_k - \bar{x}_k)] + H(x_k)^{\mathrm{T}} r_k,$$

where $r_k \stackrel{\text{def}}{=} H(x_k) - H(\bar{x}_k) + V_k(x_k - \bar{x}_k)$, we have that

$$H(x_k)^{\mathrm{T}} V_k(x_k - \bar{x}_k) = \|H(x_k)\|^2 - H(x_k)^{\mathrm{T}} r_k.$$

From (4.1)–(4.3), we can obtain that from $\|r_k\| \leqslant \tau_2 \|x_k - \bar{x}_k\|^2$,

$$\|H(x_k)^{\mathrm{T}} V_k\| \|x_k - \bar{x}_k\| \geqslant \tau_1^2 \|x_k - \bar{x}_k\|^2 - L\tau_2 \|x_k - \bar{x}_k\|^3 \geqslant (\tau_1^2 - L\tau_2\delta/2)\|x_k - \bar{x}_k\|^2.$$

As $\tau_1^2/L\tau_2 \geqslant \delta$, we have that

$$\|V_k^{\mathrm{T}} H(x_k)\| \geqslant \frac{\tau_1^2}{2} \|x_k - \bar{x}_k\|. \tag{4.6}$$

Moreover, the definition of $v_k$ in the proposed algorithm together with (4.1) and (4.5) gives

$$v_k = \eta \|V_k^{\mathrm{T}} H(x_k)\| \geqslant \frac{\eta\tau_1}{2} \|x_k - \bar{x}_k\| = \frac{\eta\tau_1}{2} \operatorname{dist}(x_k, X^*). \tag{4.7}$$

For the solution $x^* \in X^*$, the nondegenerate condition of the reformulated problem (2.6) implies that there exists sufficiently small $\rho \in (0, 2]$ such that $l_i + \rho \leqslant (x^*)_i \leqslant u_i - \rho$ for $i = 1, \ldots, n$. For each $x_k \in \mathscr{B}_{\rho/2}(x^*) \cap \Omega$, then $|l_i - (x_k)_i| > \rho/2$ and $|u_i - (x_k)_i| > \rho/2$ for $i = 1, \ldots, n$. Hence, $\|D_k\| \leqslant \sqrt{2n/\rho} \stackrel{\text{def}}{=} \varrho$ where $\sqrt{2/\rho} \geqslant 1$.

Using (4.5) and (4.2), we obtain from the definition of the function $\psi_k$ in the subproblem $(P_k)$ that for $\forall V_k \in \partial H(x_k)$,

$$\begin{aligned}
\|d_k\|^2 &\leqslant \frac{2}{v_k} \psi_k(\widehat{d_k}) \leqslant \frac{2}{v_k} \psi_k(D_k(\bar{x}_k - x_k)) \\
&= \frac{1}{v_k} (\|H(x_k) + V_k(\bar{x}_k - x_k)\|^2 + v_k \|D_k(\bar{x}_k - x_k)\|^2) \\
&\leqslant \frac{1}{v_k} \|H(x_k) - H(\bar{x}_k) - V_k(x_k - \bar{x}_k)\|^2 + \|D_k\|^2 \|\bar{x}_k - x_k\|^2 \\
&\leqslant \frac{2\tau_2^2}{\eta\tau_1} \|x_k - \bar{x}_k\|^2 + \varrho^2 \|x_k - \bar{x}_k\|^2 = \left(\frac{2\tau_2^2}{\eta\tau_1} + \varrho^2\right) \operatorname{dist}(x_k, X^*)^2.
\end{aligned} \tag{4.8}$$

Therefore, statement (a) holds with $\tau_3 \stackrel{\text{def}}{=} \sqrt{(2\tau_2^2/\eta\tau_1) + \varrho^2}$.

(b) Since (4.3) yields

$$v_k = \eta \|V_k^{\mathrm{T}} H(x_k)\| = \eta\chi_V \|H(x_k) - H(\bar{x}_k)\| \leqslant \eta L \chi_V \|x_k - \bar{x}_k\|,$$

we have that from the above inequality, (4.2) and the definition of $\psi_k$ in the subproblem $(P_k)$,

$$\begin{aligned}
\|H(x_k) + V_k d_k\|^2 &\leqslant \|H(x_k) - H(\bar{x}_k) - V_k(x_k - \bar{x}_k)\|^2 + v_k \varrho^2 \|\bar{x}_k - x_k\|^2 \\
&\leqslant \tau_2^2 \|x_k - \bar{x}_k\|^4 + \eta L \chi_V \varrho^2 \|x_k - \bar{x}_k\|^2.
\end{aligned} \tag{4.9}$$

Hence statement (b) holds with $\tau_4 \stackrel{\text{def}}{=} \sqrt{\tau_2^2 \delta + \varrho^2 \eta L \chi_V}$. $\square$

**Theorem 4.2.** *Assume that Assumptions* (A1)–(A3) *hold. If there exists a limit point $x^*$ of the sequence $\{x_k\}$ generated by the proposed algorithm such that $x^* \in$* int$(\Omega)$, *then* $\lim_{k\to\infty} \|H_k\| = 0$, *and all the accumulation point solve the semismooth systems* (1.1).

**Proof.** For $x^* \in \operatorname{int}(\Omega)$, there exists sufficiently small $\delta \in (0, 2]$ such that the open ball $\mathscr{B}(x^*, \delta) \stackrel{\text{def}}{=} \{x \mid \|x - x^*\| < \delta\} \subset \operatorname{int}(\Omega)$.

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 $\mathscr{B}(x^*, \delta/2)$. 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 $i$th components of $l$, $u$ and $x_{k_j}$, respectively. Hence, $\|D_{k_j}\| \leqslant \sqrt{2n/\delta}$ where $\sqrt{2/\delta} \geqslant 1$. Also, since

$$\|H(x_k)\|^2 = H(x_k)^{\mathrm{T}} H(x_k) = H(x_k)^{\mathrm{T}} [H(\bar{x}_k) + V_k(x_k - \bar{x}_k)] + H(x_k)^{\mathrm{T}} r_k,$$

we have that $H(x_k)^{\mathrm{T}} V_k(x_k-\bar{x}_k)=\|H(x_k)\|^2-H(x_k)^{\mathrm{T}} r_k$. From (4.1)–(4.3), we can obtain that from $\|r_k\| \leqslant \tau_2\|x_k-\bar{x}_k\|^2$,

$$\begin{aligned}\|H(x_k)^{\mathrm{T}} V_k\|\|x_k-\bar{x}_k\| &\geqslant \tau_1\|H(x_k)\|\|x_k-\bar{x}_k\|-\tau_2\|H(x_k)\|\|x_k-\bar{x}_k\|^2\\ &\geqslant(\tau_1-\tau_2\delta/2)\|H(x_k)\|\|x_k-\bar{x}_k\|.\end{aligned}$$

As $\tau_1/\tau_2 \geqslant \delta$, we have that

$$\|V_k^{\mathrm{T}} H(x_k)\| \geqslant \frac{\tau_1}{2}\|H(x_k)\|.$$

Further, similar to the proof of theorem in [7], (3.14) means that the sequence $\{h(x_k)\}$ is convergent. Then, from the following inequality and (3.18),

$$\frac{\|H_{k_j}\|}{\tau_1\sqrt{2n/\delta}} \leqslant \frac{\|V_{k_j}^{\mathrm{T}} H_{k_j}\|}{2\|D_{k_j}\|} \leqslant\|D_{k_j}^{-1} V_{k_j}^{\mathrm{T}} H_{k_j}\| \rightarrow 0,$$

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

**Theorem 4.3.** *Assume that Assumptions* (A1)–(A3) *hold and that $x^*$ is an accumulation point of $\{x_k\}$ such that $x^*$ is a nondegenerate zero of $H$ at which $H$ is semismooth. Then the full stepsize $\alpha_k=1$ and $\theta_k=1$ is always accepted for $k$ sufficiently large so that $x_{k+1}=x_k+d_k$ when $v_k$ given in* (2.18).

**Proof.** Since $x^*$ is an accumulation point of $\{x_k\}$ there exists $\varepsilon>0$ such that for sufficiently large $k$, $x_k \in \mathscr{B}_{\varepsilon}(x^*) \stackrel{\text{def}}{=}\{x \in \Re^n \mid \|x-x^*\| \leqslant \varepsilon\}$. From (2.11), we have that for $\forall V_k \in \partial H(x_k)$,

$$\begin{aligned}\nabla h(x_k)^{\mathrm{T}} d_k &= (V_k^{\mathrm{T}} H_k)^{\mathrm{T}} d_k\\ &= \frac{1}{2}\|H_k+V_k d_k\|^2-\frac{1}{2}\|H_k\|^2-\frac{1}{2} d_k^{\mathrm{T}} V_k^{\mathrm{T}} V_k d_k\\ &\leqslant \frac{1}{2}(\|H_k+V_k d_k\|^2+v_k\|D_k\|^2\|d_k\|^2)-\frac{1}{2}\|H_k\|^2\\ &\leqslant \frac{1}{2}(\|H(x_k)-H(\bar{x}_k)-V_k(x_k-\bar{x}_k)\|^2+v_k \varrho^2\|\bar{x}_k-x_k\|^2-\|H_k\|^2)\\ &\leqslant \frac{1}{2}(\tau_2^2\|x_k-\bar{x}_k\|^4+v_k \varrho^2\|x_k-\bar{x}_k\|^2-\tau_1^2\|x_k-\bar{x}_k\|^2)\\ &\leqslant-\frac{1}{2}[\tau_1^2-(\tau_2^2 \varepsilon^2+\eta L \chi_D \varrho^2 \varepsilon)] \operatorname{dist}(x_k, X^*)^2 \leqslant-\frac{\tau_1^2}{4} \operatorname{dist}(x_k, X^*)^2,\end{aligned} \tag{4.10}$$

where the last inequality is deduced by $\|x_k-\bar{x}_k\| \leqslant \varepsilon$ for sufficiently large $k$ and sufficiently small $\varepsilon$ such that $\varepsilon \leqslant \tau_1^2/2(\tau_2^2+\eta L \chi_D \varrho^2)$. According to the acceptance rule in step 4, we have

$$h(x_{l(k)})-h(x_k+\alpha_k d_k) \geqslant-\beta \alpha_k \nabla h(x_k)^{\mathrm{T}} d_k \geqslant \beta \frac{\tau_1^2}{4} \alpha_k \operatorname{dist}(x_k, X^*)^2. \tag{4.11}$$

Similar to the proof of Theorem 3.3, we can prove that

$$\lim_{k \rightarrow \infty} \alpha_{l(k)} \operatorname{dist}(x_{l(k)}, X^*)^2=0. \tag{4.12}$$

By (a) in Theorem 4.2, we can also have that

$$\lim_{k \rightarrow \infty} \alpha_{l(k)}\|d_{l(k)}\|^2=0. \tag{4.13}$$

Similar to the proof of theorem in [7], we can prove that $\lim_{k \rightarrow \infty} h(x_{l(k)})=\lim_{k \rightarrow \infty} h(x_k)$. From (4.11), we also have that

$$\lim_{k \rightarrow \infty} \alpha_k \operatorname{dist}(x_k, X^*)^2=0. \tag{4.14}$$

Hence,

$$\lim_{k\to\infty} \alpha_k \|d_k\|^2 = 0. \tag{4.15}$$

Let the step size scalar $\alpha_k$ be given in (2.17) along the direction $d_k$ to the boundary (2.15) of the box constraints. Since the nondegenerate of the reformulated problem (2.6) holds at every limit point of $\{x_k\}$, similar to the proof of Theorem 3.3, we can first obtain that $\liminf_{k\to+\infty} \alpha_k \neq 0$ when $d_k$ is given in (2.15) along $d_k$ to the boundary of the box constraints. Therefore, we assume that if $\liminf_{k\to+\infty} \alpha_k = 0$ when the acceptance rule (2.14) determines $\alpha_k$, similar to the proof of (3.15) in Theorem 3.3, we can also obtain that

$$\lim_{k\to+\infty} \frac{\nabla h_k^{\mathrm{T}} d_k}{\|d_k\|} = 0.$$

Hence, (4.15), (4.10) and (a) in Theorem 4.2 mean

$$0 = \lim_{k\to+\infty} \frac{\nabla h_k^{\mathrm{T}} d_k}{\|d_k\|} \leqslant \lim_{k\to+\infty} -\frac{\tau_1^2}{4} \frac{\operatorname{dist}(x_k, X^*)^2}{\|d_k\|} \leqslant \lim_{k\to+\infty} -\frac{\tau_1^2}{4\tau_3} \frac{\|d_k\|^2}{\|d_k\|} \leqslant 0. \tag{4.16}$$

This gives that if (4.15) holds, then

$$\lim_{k\to+\infty} \|d_k\| = 0. \tag{4.17}$$

We now prove that if (4.17) holds, then $\alpha_k = 1$ must satisfy the accepted condition (2.14) in step 4. For large enough $k$, there exists $v_k \geqslant 0$ such that from (2.11),

$$\begin{aligned} & h(x_k + d_k) - h(x_k) - \tfrac{1}{2}\nabla h_k^{\mathrm{T}} d_k \\ &\quad = \tfrac{1}{2}\|H_k + V_k d_k + \mathrm{o}(\|d_k\|)\|^2 - \tfrac{1}{2}\|H_k\|^2 - \tfrac{1}{2}(V_k^{\mathrm{T}} H_k)^{\mathrm{T}} d_k \\ &\quad = \tfrac{1}{2}(V_k^{\mathrm{T}} H_k)^{\mathrm{T}} d_k + \tfrac{1}{2} d_k^{\mathrm{T}} (V_k^{\mathrm{T}} V_k) d_k + \mathrm{o}(\|d_k\|^2) \\ &\quad = -\tfrac{1}{2} v_k \|d_k\|^2 + \mathrm{o}(\|d_k\|^2) \leqslant \mathrm{o}(\|d_k\|^2), \end{aligned} \tag{4.18}$$

where the last inequality is deduced by (4.8). Using the above equality, we have that from (4.10) and Lemma 4.1(a),

$$\begin{aligned} h(x_k + d_k) - h(x_k) - \beta \nabla h_k^{\mathrm{T}} d_k &= \left(\frac{1}{2} - \beta\right) \nabla h_k^{\mathrm{T}} d_k - \frac{1}{2} v_k \|d_k\|^2 + \mathrm{o}(\|d_k\|^2) \\ &\leqslant -\left(\frac{1}{2} - \beta\right) \frac{\tau_1^2}{4} \operatorname{dist}(x_k, X^*)^2 + \mathrm{o}(\|d_k\|^2) \leqslant 0, \end{aligned} \tag{4.19}$$

where inequality is deduced by (4.10) and (a) in Lemma 4.1. Therefore, the accepted condition (2.14) holds when $\alpha_k = 1$, since $h(x_k) \leqslant h(x_{l(k)})$.

Now, we prove that if (4.15) holds, when $\alpha_k = 1$ the accepted condition (2.15) given in step 4 also holds at the stepsize to the boundary of box constraints along $d_k$. Eq. (4.15) means that $(d_k)_i \to 0$, for all $i$. If $(g_*)_i = 0$ for any $i$, assume that $\alpha_k$ given in step 4 is the step size to the boundary of box constraints along $d_k$, the nondegenerate means that $l_i < (x_*)_i < u_i$, then

$$\lim_{k\to\infty} \Gamma_k \stackrel{\text{def}}{=} \min\left\{ \max\left\{ \frac{l_i - (x_k)_i}{(d_k)_i}, \frac{u_i - (x_k)_i}{(d_k)_i} \right\}, i = 1, 2, \ldots, n \right\} = +\infty.$$

If $(g_*)_i \neq 0$ for some $i$, without loss of generality, assume $(x_*)_i = l_i$ for some $i$. since $(V_k^{\mathrm{T}} V_k d_k)$ converges to zero and $v_k I$ is a positive semidefinite diagonal matrix in (3.11), the nondegenerate condition of reformulated problem (2.6) at the limit point implies that $(d_k)_i$ and $-(g_k)_i$ have the same sign for $k$ sufficiently large. Hence, if $\alpha_k$ is defined by some $(v_*)_j = 0$ and $(g_*)_j \neq 0$, then $\alpha_k = |(v_k)_j|/|(d_k)_j|$ for $k$ sufficiently large. Using (3.12), again, noting $\eta \geqslant 1$,

$$\begin{aligned} \min\{1, \Gamma_k\} &= \min\left\{1, \frac{v_k}{|(g_k)_j + (V_k^{\mathrm{T}} V_k d_k)_j|}\right\} \\ &\geqslant \min\left\{1, \eta - \frac{\eta \|V_k^{\mathrm{T}} V_k d_k\|}{\|g_k\| + \|V_k^{\mathrm{T}} V_k d_k\|}\right\} \longrightarrow 1 \quad \text{as } d_k \to 0. \end{aligned} \tag{4.20}$$

Further, by the condition on the strictly feasible stepsize $\theta_k \in (\theta_0, 1]$, for some $0 < \theta_0 < 1$ and $\theta_k - 1 = \mathrm{O}(\|d_k\|^2)$, $\lim_{k\to\infty} \theta_k = 1$, comes from $\lim_{k\to\infty} d_k = 0$.

So $\alpha_k = 1$, i.e., $s_k = d_k$ and hence $x_{k+1} = x_k + d_k$.

From above, when $\alpha_k$ given in step 4 is the step size to the boundary of box constraints along $d_k$, we always have that whether $(g_*)_i = 0$ for any $i$ or $(g_*)_i \neq 0$ for some $i$,

$$\lim_{k\to\infty} \min\{1, \alpha_k\} = 1$$

holds.

We have also obtained that as $\|d_k\| \to 0$ and $\theta_k \to 1$, the full step $\alpha_k \equiv 1$ is eventually accepted, for large enough $k$, if $\alpha_k$ is determined by (2.14)–(2.15), that is,

$$h(x_k + d_k) \leqslant h(x_{l(k)}) + \beta \nabla h(x_k)^{\mathrm{T}} d_k \quad \text{with } x_k + \alpha_k d_k \in \Omega.$$

The conclusion of Theorem holds. $\square$

Theorem 4.3 means that the local convergence rate for the proposed algorithm depends on the quality of the approximate subdifferentials, local error bound condition at $X^*$, and the local convergence rate of the step $d_k$.

We now show that the proposed algorithm is locally Q-superlinear convergent in the sense that the distance from the iterates $x_k$ to the solution set $X^*$ goes down to zero with a Q-superlinear convergent rate. In order to verify this result, we need to prove a couple of technical lemmas. These lemmas can be derived by suitable modifications of the corresponding constrained results in [8] by setting $v_k$ in (2.18). The next result is a major step in verifying local superlinear convergence of the distance function.

**Lemma 4.4.** *Assume that Assumption* (A3) *holds. If both $x_{k-1}$ and $x_k$ belong to the ball $\mathscr{B}_{\delta/2}(x^*)$ for each $k$ where the nondegenerate condition holds at the point $x^*$, then there is a constant $\tau_5 > 0$ such that*

$$\operatorname{dist}(x_k, X^*) \leqslant \tau_5 \operatorname{dist}(x_{k-1}, X^*)^{3/2}$$

*for each $k$.*

**Proof.** Since $x_k, x_{k-1} \in \mathscr{B}_{\delta/2}(x^*)$ and $x_k = x_{k-1} + d_{k-1}$, we obtain from (4.3) that

$$\begin{aligned} &\|H(x_{k-1} + d_{k-1})\| - \|H(x_{k-1}) + V_{k-1} d_{k-1}\| \\ &\quad \leqslant \|H(x_{k-1}) - H(x_{k-1} + d_{k-1}) + V_{k-1} d_{k-1}\| \leqslant \tau_2 \|d_{k-1}\|^2. \end{aligned}$$

Using the error bound assumption (4.2) and Lemma 4.1, we therefore obtain

$$\begin{aligned} \tau_1 \operatorname{dist}(x_k, X^*) &\leqslant \|H(x_{k-1} + d_{k-1})\| \\ &\leqslant \|H(x_{k-1}) + V_{k-1} d_{k-1}\| + \tau_2 \|d_{k-1}\|^2 \\ &\leqslant \tau_4 \operatorname{dist}(x_{k-1}, X^*)^2 + \tau_2 \tau_3^{3/2} \operatorname{dist}(x_{k-1}, X^*)^{3/2} \\ &= (\tau_4 \delta^{1/2} + \tau_2 \tau_3^{3/2}) \operatorname{dist}(x_{k-1}, X^*)^{3/2} \end{aligned}$$

and this completes the proof by setting $\tau_5 \stackrel{\text{def}}{=} (\tau_4 \delta^{1/2} + \tau_2 \tau_3^{3/2})/\tau_1$. $\square$

The next result shows that the assumption of Lemma 4.3 is satisfied if the starting point $x_0$ in the proposed algorithm is chosen sufficiently close to the solution set $X^*$. Let

$$r \stackrel{\text{def}}{=} \min\left\{\frac{\delta}{2(1+3\tau_3)}, \frac{2}{3\tau_5^2}\right\}. \tag{4.21}$$

**Lemma 4.5.** *Assume that Assumption* (A3) *holds. If the starting point* $x_0 \in \Omega$ *used in the proposed algorithm belongs to the ball* $\mathscr{B}_r(x^*)$, *where* $r$ *is defined by* (4.21) *and the nondegenerate condition holds at the point* $x^*$, *then all iterates* $x_k$ *generated by the proposed algorithm belong to the ball* $\mathscr{B}_{\delta/2}(x^*)$.

**Proof.** The proof is by induction on $k$. We start with $k=0$. By assumption, we have $x_0 \in \mathscr{B}_r(x^*)$. Since $r \leqslant \delta/2$, this implies $x_0 \in \mathscr{B}_{\delta/2}(x^*)$. Let $k \geqslant 0$ be arbitrarily given and assume that $x_j \in \mathscr{B}_{\delta/2}(x^*)$ for all $j=0,\ldots,k$, now we prove that $x_{k+1}$ also belongs to $\mathscr{B}_{\delta/2}(x^*)$.

From Lemma 4.4, we have that

$$\begin{aligned}
\operatorname{dist}(x_j, X^*) &\leqslant \tau_5 \operatorname{dist}(x_{j-1}, X^*)^{3/2} \leqslant \tau_5 \tau_5^{3/2}[\operatorname{dist}(x_{j-2}, X^*)^{3/2}]^{3/2} \leqslant \cdots \\
&\leqslant \tau_5 \tau_5^{3/2} \cdots \tau_5^{(3/2)^{(j-1)}} \operatorname{dist}(x_0, X^*)^{(3/2)^j} = (\tau_5^2)^{[(3/2)^j - 1]} \operatorname{dist}(x_0, X^*)^{(3/2)^j} \\
&\leqslant (\tau_5^2)^{[(3/2)^j-1]} \|x_0 - x^*\|^{(3/2)^j} \leqslant (\tau_5^2)^{[(3/2)^j-1]} r^{(3/2)^j}
\end{aligned}$$

for all $j=0,\ldots,k$. Using $r \leqslant 2/(3\tau_5^2)$, we therefore get that from Lemma 4.3

$$\begin{aligned}
\|x_{k+1} - x^*\| &= \|x_k + d_k - x^*\| \leqslant \|x_k - x^*\| + \|d_k\| \\
&\leqslant \|x_0 - x^*\| + \sum_{j=0}^{k} \|d_j\| \leqslant r + \tau_3 \sum_{j=0}^{k} \operatorname{dist}(x_j, X^*) \\
&\leqslant r + \tau_3 \sum_{j=0}^{k} (\tau_5^2)^{[(3/2)^j-1]} r^{(3/2)^j} \\
&\leqslant r + \tau_3 r \sum_{j=0}^{k} \left(\frac{2}{3}\right)^{(3/2)^j - 1} \leqslant r + \tau_3 r \sum_{j=0}^{\infty} \left(\frac{2}{3}\right)^{j} = (1+3\tau_3) r \leqslant \frac{\delta}{2}
\end{aligned}$$

where the last inequality follows from (4.21) of $r$. This completes the induction. □

We now obtain the following superlinear convergence result for the distance function as an immediate consequence of Lemmas 4.4 and 4.5.

**Theorem 4.6.** *Let Assumption* (A3) *be satisfied and* $\{x_k\}$ *be a sequence generated by the proposed Algorithm with starting point* $x_0 \in \mathscr{B}_r(x^*)$, *where* $r$ *is defined by* (4.21) *and the nondegenerate condition holds at the point* $x^*$. *Then the sequence* $\{\operatorname{dist}(x_k, X^*)\}$ 1.5-*order Q-superlinear converges to zero, i.e., the iterates* $x_k$ *approach the solution set* $X^*$ *at the* 1.5-*order rate of local convergence.*

Theorem 4.6 shows that the proposed affine scaling Levenberg–Marquardt-type algorithm is locally superlinearly convergent under fairly mild assumptions. In view of Theorem 4.6, we know that the distance $\operatorname{dist}(x_k, X^*)$ from the iterates $x_k$ to the solution set $X^*$ converges to zero locally superlinearly. We start by showing that the sequence $\{x_k\}$ proposed by the algorithm is convergent.

**Theorem 4.7.** *Let Assumption* (A3) *be satisfied and* $\{x_k\}$ *be a sequence generated by the proposed algorithm with starting point* $x_0 \in \mathscr{B}_r(x^*)$, *where* $r$ *is defined by* (4.21) *and the nondegenerate condition holds at the point* $x^*$. *Then the sequence* $\{x_k\}$ *converges to a solution* $\bar{x}$ *of* (1.1) *belonging to the ball* $\mathscr{B}_{\delta/2}(x^*)$.

**Proof.** Since the entire sequence $\{x_k\}$ remains in the closed ball $\mathscr{B}_{\delta/2}(x^*)$ by Lemma 4.5, every limit point of this sequence belongs to this set, too. As in the proof of Lemma 4.5, we have that for any positive integer $l$,

$$\|d_j\| \leqslant \tau_3 \operatorname{dist}(x_j, X^*) \leqslant \tau_3 (\tau_5^2)^{(3/2)^j - 1} r^{(3/2)^j} \leqslant \tau_3 r (\tfrac{2}{3})^{(3/2)^j - 1} \leqslant c_3 r (\tfrac{2}{3})^j,$$

where the first inequality follows from Lemma 4.3 and the third inequality follows from $r \leqslant 2/(3\tau_5)$. Therefore, for any positive integers $k$ and $m$ such that $k > t$, we have

$$\|x_k - x_t\| \leqslant \|x_{k-1} - x_t\| + \|d_{k-1}\| \leqslant \sum_{j=t}^{k-1} \|d_j\| \leqslant \tau_3 r \sum_{j=t}^{\infty} \left(\frac{2}{3}\right)^j \to 0 \quad \text{as } t \to \infty.$$

This means $\{x_k\}$ is a Cauchy sequence and hence convergent. $\square$

We now obtain our main local convergence result of this section.

**Theorem 4.8.** *Let Assumption* (A3) *be satisfied and* $\{x_k\}$ *be a sequence generated by the proposed algorithm with starting point* $x_0 \in \mathscr{B}_r(x^*)$ *where r is defined by* (4.21) *and the nondegenerate condition holds at the point* $x^*$ *and limit point* $\bar{x}$*. Then the sequence* $\{x_k\}$ *converges locally Q-superlinearly to* $\bar{x}$ *at* 1.5*-order rate of local convergence.*

**Proof.** Letting $\bar{x}_{k+1} \in X^*$ denote the closest solution to $x_{k+1}$, we then obtain that for all $k$ large enough

$$\begin{aligned}\|d_k\| &= \|x_k - x_{k+1}\| \geqslant \|x_k - \bar{x}_{k+1}\| - \|\bar{x}_{k+1} - x_{k+1}\| \\ &\geqslant \operatorname{dist}(x_k, X^*) - \operatorname{dist}(x_{k+1}, X^*) \geqslant \tfrac{1}{2}\operatorname{dist}(x_k, X^*),\end{aligned} \tag{4.22}$$

where last inequality follows from $\operatorname{dist}(x_{k+1}, X^*) \leqslant \frac{1}{2}\operatorname{dist}(x_k, X^*)$ for all $k$ sufficiently large in Theorem 4.6.

From Lemmas 4.3, 4.4, and (4.22), we have that setting $\tau_6 = \frac{1}{2}$,

$$\|d_{k+1}\| \leqslant \tau_3 \operatorname{dist}(x_{k+1}, X^*) \leqslant \tau_3 \tau_5 \operatorname{dist}(x_k, X^*)^{3/2} \leqslant \tau_3 \tau_5 \tau_6^{3/2} \|d_k\|^{3/2} \stackrel{\text{def}}{=} \tau_7 \|d_k\|^{3/2} \tag{4.23}$$

for all $k$ sufficiently large.

For sufficiently large $k$, without loss of generality, we assume that $\tau_7 \|d_k\|^{1/2} \leqslant \frac{1}{2}$ holds, and hence, (4.23) means that $\|d_{k+1}\| \leqslant \frac{1}{2}\|d_k\|$ holds. We can then apply (4.23) successively to obtain that for all $j = 0, 1, 2, \ldots,$

$$\|d_{k+j}\| \leqslant \tau_7 \|d_{k+j-1}\|^{3/2} \leqslant (\tfrac{1}{2})^{3/2} \tau_7 \|d_{k+j-2}\|^{3/2} \leqslant (\tfrac{1}{2})^2 \|d_{k+j-2}\| \leqslant (\tfrac{1}{2})^j \|d_k\|.$$

Let $\{x_k\}$ be a sequence generated by the proposed algorithm and converging to $\bar{x}$. Since

$$x_{k+l} = x_k + \sum_{j=0}^{t-1} d_{k+j} \quad \text{and} \quad \bar{x} = \lim_{t \to \infty} x_{k+t},$$

we therefore get

$$\begin{aligned}\|x_k - \bar{x}\| &= \|x_k - \lim_{t \to \infty} x_{k+t}\| = \|\lim_{t \to \infty} \sum_{j=0}^{t-1} d_{k+j}\| \\ &\leqslant \lim_{t \to \infty} \sum_{j=0}^{l-1} \|d_{k+j}\| = \sum_{j=0}^{\infty} \|d_{k+j}\| \leqslant \|d_k\| \sum_{j=0}^{\infty} \left(\frac{1}{2}\right)^j = 2\|d_k\|.\end{aligned} \tag{4.24}$$

Setting $\tau_9 \stackrel{\text{def}}{=} \tau_3$, Lemma 4.3 implies $\|d_k\| \leqslant \tau_3 \operatorname{dist}(x_k, X^*) \leqslant \|x_k - \bar{x}\|$ for all $k$. Setting, again, $\tau_8 \stackrel{\text{def}}{=} 1/2$ in (4.24), we have that for all $k$ sufficiently large,

$$\tau_8 \|x_k - \bar{x}\| \leqslant \|d_k\| \leqslant \tau_9 \|x_k - \bar{x}\|, \tag{4.25}$$

which implies that the length of the search direction $d_k$ is eventually in the same order as the distance from the current iterate $x_k$ to the limit point $\bar{x}$ of the sequence $\{x_k\}$. Using (4.25), we immediately obtain

$$\tau_8\|x_{k+1} - \bar{x}\| \leqslant \|d_{k+1}\| \leqslant \tau_7\|d_k\|^{3/2} \leqslant \tau_7\tau_9^{3/2}\|x_k - \bar{x}\|^{3/2}$$

for all $k$ sufficiently large. This shows that the sequence $\{x_k\}$ converges locally Q-superlinearly to $\bar{x}$ at 1.5-order rate of local convergence. □

In view of Theorem 4.6, we have known that the distance $\mathrm{dist}(x_k, X^*)$ from the iterates $x_k$ to the solution set $X^*$ converges to zero locally Q-superlinearly. However, in this section, we further see that about the behavior of the sequence $\{x_k\}$ itself, this sequence converges to a solution of (1.1), and that the rate of convergence is also locally Q-superlinear at 1.5-order rate of local convergence. We have presented the globalized and local version of the constrained affine scaling interior Levenberg–Marquardt method. We feel that the numerical test will be implemented in practice further.

## Acknowledgments

The author would like to thank the anonymous referee for their detailed comments and helpful suggestions. The author gratefully acknowledges the partial supports of the National Science Foundation Grant (10471094) of China, the Ph.D. Foundation Grant (0527003) of Chinese Education Ministry, the Shanghai Leading Academic Discipline Project (T0401), and the Science Foundation Grant (05DZ11) of Shanghai Education Committee.

## References

[1] H.B. An, Z.Z. Bai, NGLM: a globally convergent Newton–GMRES method, Math. Numer. Sinica 27 (2005) 151–174 (in Chinese).
[2] H.B. An, Z.Z. Bai, On efficient variants and global convergence of the Newton–GMRES method, J. Numer. Methods Comput. Appl. 26 (2005) 291–300 (in Chinese).
[3] H.B. An, Z.Z. Bai, A globally convergent Newton–GMRES method for large sparse systems of nonlinear equations, Appl. Numer. Math. 57 (2007) 235–252.
[4] S. Bellavia, M. Macconi, B. Morini, An affine scaling trust-region approach to bound-constrained nonlinear systems, Appl. Numer. Math. 44 (2003) 257–280.
[5] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.
[6] T.F. Coleman, Y. Li, An interior trust region approach for minimization subject to bounds, SIAM J. Optimi. 6 (3) (1996) 418–445.
[7] L. Gripp, F. Lampariello, L. Lucidi, A nonmonotone line search technique for Newton's methods, SIAM J. Numer. Anal. 23 (1986) 707–716.
[8] C. Kanzow, N. Yamashita, M. Fukushima, Levenberg–Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints, J. Comput. Appl. Math. 172 (2004) 375–397.
[9] J.S. Pang, Newton's method for B-differentiable equation, Math. Oper. Res. 15 (1990) 311–341.
[10] J.S. Pang, L. Qi, Nonsmooth equation: motivation and algorithms, SIAM J. Optim. 3 (1993) 443–465.
[11] L. Qi, Convergence analysis of some algorithms for solving nonsmooth equations, Math. Oper. Res. 18 (1993) 227–244.
[12] L. Qi, J. Sun, A nonsmooth version of Newton's method, Math. Programming 58 (1993) 353–367.
[13] S.M. Robinson, Some continuity properties of polyhedral multifunctions, Math. Programming Stud. 14 (1981) 206–214.
[14] A. Shapiro, On concepts of directional differentiability, J. Optim. Theory Appl. 66 (1990) 477–487.
[15] 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.
[16] N. Yamashita, M. Fukushima, On the rate of convergence of the Levenberg–Marquardt method, Computing 15 (Suppl.) (2001) 239–249.