2013年6月 运 筹 学 学 报 第17卷 第2期
June, 2013 Operations Research Transactions Vol.17 No.2

Affine scaling interior Levenberg-Marquardt method for KKT systems*

WANG Yunjuan[1,†] ZHU Detong[2]

Abstract We develop and analyze a new affine scaling Levenberg-Marquardt method with nonmonotonic interior backtracking line search technique for solving Karush-Kuhn-Tucker (KKT) system. By transforming the KKT system into an equivalent minimization problem with nonnegativity constraints on some of the variables, we establish the Levenberg-Marquardt equation based on this reformulation. Theoretical analysis are given which prove that the proposed algorithm is globally convergent and has a local superlinear convergent rate under some reasonable conditions. The results of numerical experiments are reported to show the effectiveness of the proposed algorithm.

Keywords KKT systems, Levenberg-Marquardt method, affine scaling, interior point, convergence

Chinese Library Classification O221.2

2010 Mathematics Subject Classification 90C30, 65K05

仿射变换内点Levenberg-Marquardt法解KKT系统*

王云娟[1,†] 朱德通[2]

摘要 提供了一类新的结合非单调内点回代线搜索技术的仿射变换Levenberg-Marquardt法解Karush-Kuhn-Tucker（KKT）系统。基于由KKT系统转化得到的等价的部分变量具有非负约束的最小化问题，建立了Levenberg-Marquardt方程。证明了算法不仅具有整体收敛性，而且在合理的假设条件下，算法具有超线性收敛速率。数值结果验证了算法的实际有效性。

关键词 KKT系统，Levenberg-Marquardt法，仿射变换，内点，收敛

中图分类号 O221.2

2010数学分类号 90C30, 65K05

收稿日期：2011年3月24日

* This work is supported by The National Natural Science Foundation of China (No. 10871130).

1. School of Mathematics and Information, Shanghai Lixin University of Commerce, Shanghai 201620, China; 上海立信会计学院数学与信息学院，上海 201620

2. Business School, Shanghai Normal University, Shanghai 200234, China; 上海师范大学商学院，上海 200234

† 通讯作者 Corresponding author, Email: yunjuanwang@163.com

0 Introduction

Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be once and $h : \mathbb{R}^n \to \mathbb{R}^p$, $c : \mathbb{R}^n \to \mathbb{R}^m$ be twice continuously differentiable functions. Define by

$$L(x, y, z) \stackrel{\text{def}}{=} F(x) + \nabla h(x)y - \nabla c(x)z$$

the Lagrangian function, then we consider the following Karush-Kuhn-Tucker (KKT) system:

$$\begin{aligned} L(x, y, z) &= 0, \\ h(x) &= 0, \\ c(x) \geqslant 0,\ z \geqslant 0,\ z^{\mathrm{T}} c(x) &= 0. \end{aligned} \tag{0.1}$$

The main purpose of this paper is to find a KKT point $\omega^* = (x^*, y^*, z^*) \in \mathbb{R}^{n+p+m}$ satisfying the KKT system (0.1). Due to its close relationship with the variational inequality problem and the nonlinear programming problems, many constructing efficient algorithms for (0.1) have been established.

The method we will describe in this paper is similar to the idea proposed in [1], where system (0.1) is transformed into a differentiable unconstrained minimization problem. Based on the Fischer-Burmeister function in [2]: $\varphi : \mathbb{R}^2 \to \Re$ defined by

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

which has a very interesting property that

$$\varphi(a, b) = 0 \Leftrightarrow a \geqslant 0,\ b \geqslant 0,\ ab = 0.$$

Hence we can reformulate system (0.1) as a nonlinear system of equations $\Phi(\omega) = 0$, where the nonsmooth mapping $\Phi : \mathbb{R}^{n+p+m} \to \mathbb{R}^{n+p+m}$ is defined by

$$\Phi(\omega) \stackrel{\text{def}}{=} \Phi(x, y, z) \stackrel{\text{def}}{=} \begin{pmatrix} L(x, y, z) \\ h(x) \\ \phi(c(x), z) \end{pmatrix}$$

and

$$\phi(c(x), z) \stackrel{\text{def}}{=} (\varphi(c_1(x), z_1), \cdots, \varphi(c_m(x), z_m))^T \in \mathbb{R}^m.$$

It is easy to see that solving (0.1) is equivalent to finding a global solution of the problem

$$\min\ \Psi(\omega) \stackrel{\text{def}}{=} \frac{1}{2} \Phi(\omega)^{\mathrm{T}} \Phi(\omega) = \frac{1}{2} \|\Phi(\omega)\|^2, \tag{0.2}$$

here $\Psi(\omega)$ denotes the natural merit function of the equation operator Φ.

This unconstrained optimization approach has been used in [1,3-4] to develop some Newton-type methods for the solution of (0.1). Despite their strong theoretical and numerical properties, these methods may fail to find the unique solution of (0.1) arising from strongly monotone variational inequalities because the variable z is not forced to be non-negative in [1, 3-4]. This, together with the fact that the variable z has to be nonnegative at a solution of (0.1), we are naturally led to consider the following variant of the problem:

$$\min\ \Psi(\omega) \quad \text{s.t.}\ \ z \geqslant 0. \tag{0.3}$$

Therefore, this paper will focus on the study of this reformulation of the KKT system (0.1).

It is well known that, it is rather difficult to solve the constrained optimization (0.3) directly. In order to avoid handling the constraints explicitly, we can use affine scaling strategies which contain an appropriate quadratic function and an scaling matrix to form a quadratic model similar to Coleman and Li in [5]. And finally, we will consider to apply an affine scaling Levenberg-Marquardt method to the quadratic model.

This paper is organized as follows. In the next section, we propose the nonmonotone affine scaling Levenberg-Marquardt algorithm with backtracking interior point technique for solving (0.1). In section 2, we prove the global convergence of the proposed algorithm. In section 3, we discuss the local convergence property. Finally, the results of numerical experiments of the proposed algorithm are reported in Section 4.

1 Algorithm

This section describes the affine scaling Levenberg-Marquardt method in association with nonmonotonic interior backtracking technique for solving a bound-constrained minimization reformulated by KKT system (0.1).

By the differentiability assumptions we made on the functions F, c and h, and by the convexity of φ, it is obvious that the mapping Φ is locally Lipschitzian and thus almost everywhere differentiable by Rademacher's theorem. Let us denote by D_Φ the set of points $\omega \in \mathbb{R}^{n+p+m}$ at which Φ is differentiable. Then, we can consider the B-subdifferential of Φ at ω,

$$\partial_B \Phi(\omega) \stackrel{\text{def}}{=} \{H \mid H = \lim_{\omega^k \to \omega,\ \omega^k \in D_\Phi} \nabla \Psi(\omega^k)^{\mathrm{T}}\}$$

which is a nonempty and compact set whose convex hull

$$\partial \Phi(\omega) \stackrel{\text{def}}{=} \mathrm{conv}(\partial_B \Phi(\omega))$$

is Clarke's[6] generalized Jocobian of Φ at ω.

Proposition 1.1[1] *Let $\omega = (x, y, z) \in \mathbb{R}^{n+p+m}$. Then, each element $H \in \partial\Phi(\omega)$ can be represented as follows:*

$$H = \begin{pmatrix} \nabla_x L(\omega) & \nabla c(x) & \nabla h(x) D_a(\omega) \\ \nabla c(x)^{\mathrm{T}} & 0 & 0 \\ -\nabla h(x)^{\mathrm{T}} & 0 & D_b(\omega) \end{pmatrix}^{\mathrm{T}},$$

where $D_a(\omega) = \mathrm{diag}(a_1(\omega), \cdots, a_m(\omega))$, $D_b(\omega) = \mathrm{diag}(b_1(\omega), \cdots, b_m(\omega)) \in \mathbb{R}^{m\times m}$ are diagonal matrices whose jth diagonal elements are given by

$$a_j(\omega) = \frac{h_j(x)}{\sqrt{h_j(x)^2 + z_j^2}} - 1,\ \ b_j(\omega) = \frac{z_j}{\sqrt{h_j(x)^2 + z_j^2}} - 1$$

if $(h_j(x), z_j) \neq (0,0)$, *and by*

$$a_j(\omega) = \xi_j - 1,\ b_j(\omega) = \zeta_j - 1 \ for\ any\ (\xi_j, \zeta_j)\ with\ \|(\xi_j, \zeta_j)\| \leqslant 1$$

if $(h_j(x), z_j) = (0,0)$.

Proposition 1.2[1] Ψ *is continuously differentiable, and* $\nabla\Psi = H^{\mathrm{T}}\Phi(\omega)$ *for every* H *in* $\partial\Phi(\omega)$.

Since the solving (0.1) is equivalent to finding a global solution of the problem (0.3), a classical algorithm for solving (0.1) will be based on the reformulated problem (0.3), i.e., $\Phi(\omega) = 0, z \geqslant 0$. As is known to all that the concept of nonsmooth Levenberg-Marquardt 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 method with its core, the underlying Newton-like iteration.

Ignoring primal and dual feasibility of the problem (0.3), the first-order necessary conditions for ω^* to be a local minimizer are

$$\begin{cases} (g^*)_i = 0, & \text{if } i \in I \cup P, \\ (g^*)_i = 0, & \text{if } i \in J \text{ and } (\omega^*_J)_i > 0, \\ (g^*)_i > 0, & \text{if } i \in J \text{ and } (\omega^*_J)_i = 0, \end{cases} \tag{1.1}$$

where $g(\omega) \stackrel{\text{def}}{=} \nabla\Psi(\omega)$, $(g^*)_i$ is the ith components of $g^* = g(\omega*)$, $I \stackrel{\text{def}}{=} \{1, \cdots, n\}$, $P \stackrel{\text{def}}{=} \{n+1, \cdots, n+p\}$ and $J \stackrel{\text{def}}{=} \{n+p+1, \cdots, n+p+m\}$.

In order to omit the constraints appeared in the problem (0.3), we introduce an affine scaling matrix similar to the idea in [5]. The scaling matrix $D_k = D(\omega^k)$ arises naturally from examining the first-order necessary conditions for nonlinear minimization transformed by KKT system (0.1), where

$$D(\omega) \stackrel{\text{def}}{=} \text{diag}\{|\gamma^1(\omega)|^{-\frac{1}{2}}, |\gamma^2(\omega)|^{-\frac{1}{2}}, \cdots, |\gamma^{n+p+m}(\omega)|^{-\frac{1}{2}}\}, \tag{1.2}$$

and the ith component of the vector function $\gamma(\omega)$ is defined as follows:

$$\gamma^i(\omega) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } (g^*)_i = 0 \text{ and } i \in I \cup P, \\ \omega_i, & \text{if } (g^*)_i \geqslant 0 \text{ and } i \in J. \end{cases} \tag{1.3}$$

Definition 1.1[5] *A point* ω *is nondegenerate if, for each index* $i \in J$,

$$g_i(\omega_J) = 0 \Rightarrow (\omega_J)_i > 0, \tag{1.4}$$

where $g_i(\omega_J)$ *is the* i*th component of vector* $g(\omega_J)$. *A reformulated problem (0.3) is nondegenerate if (1.4) holds for every* ω_J.

The Levenberg-Marquardt equation arises naturally from examining the Kuhn-Tucker conditions for the reformulate problem (0.3), i.e.,

$$D(\omega)^{-2} g(\omega) = 0. \tag{1.5}$$

For each k, given a positive parameter μ_k and an identity matrix I, we define $\psi_k : \mathbb{R}^{n+p+m} \to \Re$ by

$$\begin{aligned}\psi_k(\hat{d}) &\stackrel{\text{def}}{=} \frac{1}{2}\|\Phi(\omega^k) + H^k D_k^{-1}\hat{d}\|^2 + \frac{1}{2}\mu_k\|\hat{d}\|^2 \\ &= \frac{1}{2}\|\Phi(\omega^k)\|^2 + (D_k^{-1}g^k)^{\mathrm{T}}\hat{d} + \frac{1}{2}\hat{d}^{\mathrm{T}}D_k^{-1}(H^k)^{\mathrm{T}}H^kD_k^{-1}\hat{d} + \frac{1}{2}\mu_k\hat{d}^{\mathrm{T}}\hat{d}, \end{aligned} \tag{1.6}$$

and consider the minimization problem

$$\min\ \psi_k(\hat{d}). \tag{1.7}$$

We now state an affine scaling Levenberg-Marquardt method applied to the solution of the semismooth problem (0.1). Let $\hat{d}^k$ be the solution of the subproblem (1.6). Since $\psi_k(\hat{d})$ is a strict convex function, $\hat{d}^k$ is the global minimum of the subproblem (1.6) which is in fact equivalent to solving the following affine scaling Levenberg-Marquardt equation

$$[D_k^{-1}(H^k)^{\mathrm{T}}H^kD_k^{-1} + \mu_k I]\hat{d}^k = -D_k^{-1}g^k. \tag{1.8}$$

Zhu in [7] pointed out that the relevance of the used affine scaling matrix D_k^{-1} and matrix $\mu_k I$ depended on the fact that the affine scaled Levenberg-Marquardt trial step $d^k = D_k^{-1}\hat{d}^k$ was 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 equation (1.8) could be required by the strict interior feasibility and nonmonotonic line research technique.

Next we describe an affine scaling Levenberg-Marquardt algorithm which combines nonmonotonic interior backtracking technique for solving the KKT system (0.1).

Algorithm 1.1

Initialization step

Choose parameters $\beta \in (0, \frac{1}{2})$, $\tau \in (0, 1)$, $\varepsilon > 0$, $0 < \theta_l < 1$, $\mu \geqslant 1$, $0 < q \leqslant 1$ and positive integer M as nonmonotonic parameter. Let $m(0) = 0$. Give a starting point $\omega^0 = (x^0, y^0, z^0)$ with $z^0 > 0$, calculate $H^0 \in \partial\Phi(\omega^0)$. Set $k = 0$, go to the main step.

Main step

1. Evaluate $\Psi_k = \Psi(\omega^k) = \frac{1}{2}\|\Phi(\omega^k)\|^2$ and $H^k \in \partial\Phi(\omega^k)$. Calculate D_k, $g^k = \nabla\Psi(\omega^k) = (H^k)^{\mathrm{T}}\Phi(\omega^k)$ and $\mu_k = \mu\|D_k^{-1}g^k\|^{2q}$.

2. If $\|D_k^{-1}g^k\| \leqslant \varepsilon$, stop with the approximate solution ω^k.

3. Solve the affine scaling Levenberg-Marquardt equation (1.8) and obtain a step $\hat{d}^k$. Set $d^k = D_k^{-1}\hat{d}^k$.

4. Choose $\alpha_k = 1, \tau, \tau^2, \cdots$ until the following inequalities hold:

$$\Psi(\omega^k + \alpha_k d^k) \leqslant \Psi(\omega^{l(k)}) + \alpha_k\beta(g^k)^{\mathrm{T}}d^k, \tag{1.9}$$

$$\omega_J^k + \alpha_k d_J^k \geqslant 0. \tag{1.10}$$

where $\Psi(\omega^{l(k)}) = \max_{0\leqslant j\leqslant m(k)}\{\Psi(\omega^{k-j})\}$.

5. Set

$$s^k = \begin{cases} \alpha_k d^k, & \text{if } \omega_J^k + \alpha_k d_J^k > 0, \\ \theta_k \alpha_k d^k, & \text{otherwise}, \end{cases}$$

where $\theta_k \in (\theta_l, 1)$ and $\theta_k - 1 = o(\|d^k\|)$ and then set

$$\omega^{k+1} = \omega^k + s^k.$$

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

2 Global convergence

Throughout this section we assume that Φ is semismooth. Given $\omega^0 = (x^0, y^0, z^0) \in \mathbb{R}^{n+p+m}$ with $z^0 > 0$, the algorithm generates a sequence $\{\omega^k\}$. In our analysis, we denote the level set of Ψ by

$$\mathcal{L}(\omega^0) \stackrel{\text{def}}{=} \{\omega \in \mathbb{R}^{n+p+m} \mid \Psi(\omega) \leqslant \Psi(\omega^0), z \geqslant 0\}.$$

The following assumptions are commonly used in convergence analysis of most methods for the constrained systems.

Assumption A1 *Sequence $\{\omega^k\}$ generated by the algorithm is contained in a compact set $\mathcal{L}(\omega^0)$ on $\mathbb{R}^{n+p+m}$.*

Assumption A2 *There exist some positive constants χ_D, χ_Φ and χ_H such that*

$$\|D(\omega)^{-1}\| \leqslant \chi_D, \quad \|\Phi(\omega)\| \leqslant \chi_\Phi, \quad \|H\| \leqslant \chi_H, \quad \forall H \in \partial\Phi(\omega), \ \forall \omega \in \mathcal{L}(\omega^0). \tag{2.1}$$

The next lemma will show that the algorithm is well-defined.

Lemma 2.1 *At the kth iteration, let $\hat{d}^k$ be a solution of Levenberg-Marquardt equation. If $\|D_k^{-1} g^k\| \neq 0$, then we have*

$$(g^k)^{\mathrm{T}} d^k < 0. \tag{2.2}$$

Moreover, the proposed algorithm will produce an iterate $\omega^{k+1} = \omega^k + \alpha_k d^k$ in a finite number of backtracking steps in (1.9)-(1.10), i.e., a positive α_k can always be found in step 4.

Proof Suppose that $(g^k)^{\mathrm{T}} d^k = 0$. Since $\hat{d}^k$ is a solution of Levenberg-Marquardt equation, we have

$$\begin{aligned} (g^k)^{\mathrm{T}} d^k &= (D_k^{-1} g^k)^{\mathrm{T}} \hat{d}^k \\ &= -(\hat{d}^k)^{\mathrm{T}} [D_k^{-1} (H^k)^{\mathrm{T}} H^k D_k^{-1} + \mu_k I] \hat{d}^k. \end{aligned} \tag{2.3}$$

Noting that the matrix $D_k^{-1}(H^k)^{\mathrm{T}} H^k D_k^{-1} + \mu_k I$ is positive definite since $\mu > 0$, $(g^k)^{\mathrm{T}} d^k = 0$ can only imply that $\hat{d}^k = 0$. Using the Levenberg-Marquardt equation again, we obtain

$$\begin{aligned} \|D_k^{-1} g^k\| &= \| -[D_k^{-1}(H^k)^{\mathrm{T}} H^k D_k^{-1} + \mu_k I] \hat{d}^k \| \\ &\leqslant \|D_k^{-1}(H^k)^{\mathrm{T}} H^k D_k^{-1} + \mu_k I\| \|\hat{d}^k\| \\ &= 0, \end{aligned} \tag{2.4}$$

which contradicts $\|D_k^{-1}g^k\| \neq 0$. So we have $(g^k)^{\mathrm{T}}d^k < 0$, i.e., (2.2) holds.

Now let us prove the latter part of the lemma.

On the one hand, it is clear to see that, in a finite number of backtracking reductions, α_k will satisfy

$$\alpha_k \overset{\text{def}}{=} \min\{1, \Gamma_k\} \overset{\text{def}}{=} \min\left\{1, \max\left\{\frac{(\omega_J^k)_i}{(d_J^k)_i}, i = 1, 2, \cdots, m\right\}\right\}$$

with $\Gamma_k \overset{\text{def}}{=} +\infty$ if $\frac{(\omega_J^k)_i}{(d_J^k)_i} \leqslant 0$ for all $i \in \{1, 2, \cdots, m\}$.

On the other hand, assume that ω^k satisfies that

$$\Psi(\omega^k + \tau^n d^k) > \Psi(\omega^{l(k)}) + \beta\tau^n (g^k)^{\mathrm{T}} d^k \geqslant \Psi(\omega^k) + \beta\tau^n (g^k)^{\mathrm{T}} d^k \tag{2.5}$$

for all $n \geqslant 0$. Then

$$\frac{\Psi(\omega^k + \tau^n d^k) - \Psi(\omega^k)}{\tau^n} > \beta(g^k)^{\mathrm{T}} d^k \tag{2.6}$$

follows. Hence, for $n \to \infty$, we have $(1-\beta)(g^k)^{\mathrm{T}}d^k \geqslant 0$, i.e., $(g^k)^{\mathrm{T}}d^k \geqslant 0$ which contradicts $(g^k)^{\mathrm{T}}d^k < 0$. Therefore it is always possible to find a step length $\alpha_k > 0$ satisfying (1.9)-(1.10).

So, we can see that the latter part of the lemma is also true. The total conclusion of the lemma holds.

The main aim of the following two theorems is to prove global convergence results of the proposed algorithm. The former indicates that at least one limit point of $\{\omega^k\}$ is a stationary point. The latter extends this theorem to a stronger global convergent result.

Theorem 2.1 *Let $\{\omega^k\}$ be a sequence generated by the proposed algorithm. Assume that Assumptions A1-A2 hold and the nondegenerate condition of the reformulated problem* (0.3) *holds. Then*

$$\liminf_{k\to\infty} \|D_k^{-1} g^k\| = 0.$$

Proof According to the acceptance rule (1.9) in step 4, we have that

$$\Psi(\omega^{l(k)}) - \Psi(\omega^k + \alpha_k d^k) \geqslant -\beta\alpha_k (g^k)^{\mathrm{T}} d^k. \tag{2.7}$$

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

From (2.3), we have

$$\begin{aligned}(g^k)^{\mathrm{T}} d^k &= -(\hat{d}^k)^{\mathrm{T}}[D_k^{-1}(H^k)^{\mathrm{T}} H^k D_k^{-1} + \mu_k I]\hat{d}^k \\ &\leqslant -\mu_k \|\hat{d}^k\|^2 \\ &= -\mu \|D_k^{-1} g^k\|^{2q} \|\hat{d}^k\|^2 \\ &\leqslant -\frac{\mu}{\chi_D^2} \|D_k^{-1} g^k\|^{2q} \|d^k\|^2.\end{aligned} \tag{2.8}$$

Combing (2.7) and (2.8), for all $k > M$, we obtain

$$\begin{aligned}\Psi(\omega^{l(k)}) &\leqslant \max_{0\leqslant j\leqslant m(l(k)-1)}\{\Psi(\omega^{l(k)-j-1})\} + \beta\alpha_{l(k)-1}(g^{l(k)-1})^{\mathrm{T}}d^{l(k)-1} \\ &\leqslant \max_{0\leqslant j\leqslant m(l(k)-1)}\{\Psi(\omega^{l(k)-j-1})\} - \frac{\beta\mu}{\chi_D^2}\alpha_{l(k)-1}\|D_{l(k)-1}^{-1}g^{l(k)-1}\|^{2q}\|d^{l(k)-1}\|^2.\end{aligned} \tag{2.9}$$

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

$$\|D_k^{-1}g^k\| \geqslant \varepsilon,\ k = 1, 2, \cdots . \tag{2.10}$$

As $\{\Psi(\omega^{l(k)})\}$ is convergent, we obtain that from (2.9) and (2.10),

$$\lim_{k\to\infty} \alpha_{l(k)-1}\|d^{l(k)-1}\|^2 = 0. \tag{2.11}$$

Further, following by induction way used in [8], it can be derived that

$$\lim_{k\to\infty} \alpha_k\|d^k\|^2 = 0,$$

which implies that either

$$\liminf_{k\to\infty} \alpha_k = 0, \tag{2.12}$$

or

$$\lim_{k\to\infty} \|d^k\| = 0. \tag{2.13}$$

First, let us consider the case that (2.13) holds. Since $\mu_k{=}\mu\|D_k^{-1}g^k\|^{2q}{\leqslant}\ \mu\chi_D^{2q}\chi_H^{2q}\chi_\Phi^{2q}$, we can obtain from the Levenberg-Marquardt equation that

$$\begin{aligned}(g^k)^{\mathrm{T}}d^k &= -(D_k^{-1}g^k)^{\mathrm{T}}[D_k^{-1}(H^k)^{\mathrm{T}}H^kD_k^{-1} + \mu_kI]^{-1}(D_k^{-1}g^k) \\ &\leqslant -\frac{\|D_k^{-1}g^k\|^2}{\|D_k^{-1}(H^k)^{\mathrm{T}}H^kD_k^{-1}\| + \mu_k} \\ &\leqslant -\frac{\varepsilon^2}{\chi_D^2\chi_H^2 + \mu\chi_D^{2q}\chi_H^{2q}\chi_\Phi^{2q}} \not\to 0.\end{aligned} \tag{2.14}$$

So $\|d^k\| \not\to 0$, which means that (2.13) is not true. So (2.12) holds.

Next, we will consider the case (2.12) holds. (2.12) means that there exists a subset $\kappa \subseteq \{k\}$ such that

$$\lim_{k\to\infty,k\in\kappa} \alpha_k = 0. \tag{2.15}$$

Assume that α_k given in step 4 is the stepsize to the boundary to box constraints along d^k. From the definition 1.1, there must exist some i such that $(\omega_J^*)_i = 0$ where ω^* is any accumulation point of the sequence $\{\omega^k\}$ and without loss of generality, assume that $\{\omega^k\}_\kappa$ is a subsequence convergent to ω^*. Recall the Levenberg-Marquardt equation, we can rewrite it as

$$\mu_kd^k = -D_k^{-2}[g^k + (H^k)^{\mathrm{T}}H^kd^k]. \tag{2.16}$$

Since ω^* is nondegenerate with $\gamma^i(\omega_J^*) = 0$ for any i, we have that $(\omega_J^*)_i = 0$ and $(g_J^*)_i \neq 0$ for some i. Hence, if α_k is defined by some $\gamma^i(\omega_J^*) = 0$ and $(g^*)_i \neq 0$, then $\alpha_k = \frac{|\gamma^i(\omega_J^k)|}{|(d_J^k)_i|}$ for sufficiently large k. Using (2.16) again, we have

$$\alpha_k = \frac{\mu_k}{|(g^k)_i + [(H^k)^{\mathrm{T}} H^k d^k]_i|} \geqslant \frac{\mu_k}{\|g^k + (H^k)^{\mathrm{T}} H^k d^k\|_\infty}. \tag{2.17}$$

Combing the above inequality and $\mu_k = \mu\|D_k^{-1} g^k\|^{2q} \geqslant \mu\varepsilon^{2q} > 0$, if α_k given in step 4 is the stepsize to the boundary of box constraints along d^k, we have that

$$\liminf_{k\to\infty} \alpha_k \geqslant \liminf_{k\to\infty} \frac{\mu_k}{\|g_J^k + (H_J^k)^{\mathrm{T}} H_J^k d_J^k\|_\infty} > 0. \tag{2.18}$$

Furthermore, if (2.12) holds, the acceptance rule (1.9) means that, for large k,

$$\Psi(\omega^k + \frac{\alpha_k}{\tau} d^k) - \Psi(\omega^k) \geqslant \Psi(\omega^k + \frac{\alpha_k}{\tau} d^k) - \Psi(\omega^{l(k)}) \geqslant \beta \frac{\alpha_k}{\tau} (g^k)^{\mathrm{T}} d^k.$$

Noting that $\frac{\alpha_k}{\tau} > 0$, we can get from the above inequality that

$$\frac{\Psi(\omega^k + \frac{\alpha_k}{\tau} d^k) - \Psi(\omega^k)}{\frac{\alpha_k}{\tau}} \geqslant \beta (g^k)^{\mathrm{T}} d^k. \tag{2.19}$$

Taking limits to (2.19), we obtain

$$\lim_{k\to\infty} (g^k)^{\mathrm{T}} d^k \geqslant \lim_{k\to\infty} \beta (g^k)^{\mathrm{T}} d^k,$$

that is,

$$\lim_{k\to\infty} (1-\beta)(g^k)^{\mathrm{T}} d^k \geqslant 0. \tag{2.20}$$

Taking into account that $1 - \beta > 0$, we can obtain from (2.20) that

$$\lim_{k\to\infty} (g^k)^{\mathrm{T}} d^k \geqslant 0. \tag{2.21}$$

Noting that $(g^k)^{\mathrm{T}} d^k \leqslant 0$, we have

$$\lim_{k\to\infty} (g^k)^{\mathrm{T}} d^k = 0. \tag{2.22}$$

But from (2.8), we can see that (2.22) is not true.

(2.18) and (2.22) mean that (2.12) does also not hold. Hence (2.10) is not true and the conclusion of the theorem is true.

Theorem 2.2 *Let $\{\omega^k\}$ be a sequence generated by the proposed algorithm. Assume that Assumptions A1-A2 hold and the nondegenerate condition of the reformulated problem* (0.3) *holds, then*

$$\lim_{k\to\infty} \|D_k^{-1} g^k\| = 0.$$

Proof The proof is still by contradiction. Let $\varepsilon_1 \in (0,1)$ be given and assume that there is a subsequence $\{m_i\}$ such that

$$\|D_{m_i}^{-1} g^{m_i}\| \geqslant \varepsilon_1. \tag{2.23}$$

Theorem 2.1 guarantees that for any $\varepsilon_2 \in (0,\varepsilon_1)$ there is a subsequence of $\{m_i\}$(without loss of generality we assume that it is still the full sequence) and a sequence $\{l_i\}$ such that

$$\|D_k^{-1} g^k\| \geqslant \varepsilon_2, \quad \text{for } m_i \leqslant k < l_i \tag{2.24}$$

and

$$\|D_{l_i}^{-1} g^{l_i}\| < \varepsilon_2. \tag{2.25}$$

From the affine scaling Levenberg-Marquardt equation, we have that

$$\begin{aligned}
(g^k)^{\mathrm{T}} d^k &= -(\hat{d}^k)^{\mathrm{T}}[D_k^{-1}(H^k)^{\mathrm{T}} H^k D_k^{-1} + \mu_k I]\hat{d}^k \\
&\leqslant -\mu_k \|\hat{d}^k\|^2 \\
&= -\mu \|D_k^{-1} g^k\|^{2q} \|\hat{d}^k\|^2 \\
&\leqslant -\mu \varepsilon_2^{2q} \|\hat{d}^k\|^2 \\
&\leqslant -\frac{\mu \varepsilon_2^{2q}}{\chi_D^2} \|d^k\|^2.
\end{aligned} \tag{2.26}$$

Since the matrix $D_k^{-1}(H^k)^{\mathrm{T}} H^k D_k^{-1} + \mu_k I$ is nonsingular in the affine scaling Levenberg-Marquardt equation, we can get that

$$\begin{aligned}
(g^k)^{\mathrm{T}} d^k &= -(D_k^{-1} g^k)^{\mathrm{T}} [D_k^{-1}(H^k)^{\mathrm{T}} H^k D_k^{-1} + \mu_k I]^{-1} D_k^{-1} g^k \\
&\leqslant -\frac{\|D_k^{-1} g^k\|^2}{\|D_k^{-1}(H^k)^{\mathrm{T}} H^k D_k^{-1}\| + \mu_k} \\
&\leqslant -\frac{\varepsilon_2^2}{\chi_D^2 \chi_H^2 + \mu \chi_D^{2q} \chi_H^{2q} \chi_\Phi^{2q}}.
\end{aligned} \tag{2.27}$$

So, combing (2.26) and (2.27), we obtain

$$[(g^k)^{\mathrm{T}} d^k]^2 \geqslant \frac{\mu \varepsilon_2^{2q+2}}{\chi_D^2 (\chi_D^2 \chi_H^2 + \mu \chi_D^{2q} \chi_H^{2q} \chi_\Phi^{2q})} \|d^k\|^2.$$

Noting that $(g^k)^{\mathrm{T}} d^k \leqslant 0$, we can get

$$(g^k)^{\mathrm{T}} d^k \leqslant -\frac{\sqrt{\mu}\varepsilon_2^{q+1}}{\chi_D \sqrt{\chi_D^2 \chi_H^2 + \mu \chi_D^{2q} \chi_H^{2q} \chi_\Phi^{2q}}} \|d^k\| = -\sigma \|d^k\|, \tag{2.28}$$

where $\sigma \overset{\text{def}}{=} \frac{\sqrt{\mu}\varepsilon_2^{q+1}}{\chi_D \sqrt{\chi_D^2 \chi_H^2 + \mu \chi_D^{2q} \chi_H^{2q} \chi_\Phi^{2q}}}$.

(1.9) and (2.28) mean that

$$\begin{aligned}
\Psi(\omega^{l(k)}) &\leqslant \Psi(\omega^{l(l(k)-1)}) + \beta \alpha_{l(k)-1} (g^{l(k)-1})^{\mathrm{T}} d^{l(k)-1} \\
&\leqslant \Psi(\omega^{l(l(k)-1)}) - \beta \sigma \alpha_{l(k)-1} \|d^{l(k)-1}\|.
\end{aligned} \tag{2.29}$$

Similar to the proof of Theorem 2.1, we have that the sequence $\{\Psi(\omega^{l(k)})\}$ is nonincreasing for $m_i \leqslant k < l_i$, and hence $\{\Psi(\omega^{l(k)})\}$ is convergent. So

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

Similar to the proof in [8] , we have that

$$\lim_{k\to+\infty} \Psi(\omega^{l(k)}) = \lim_{k\to+\infty} \Psi(\omega^k). \tag{2.30}$$

So we can also obtain

$$\lim_{k\to\infty} \alpha_k\|d^k\| = 0. \tag{2.31}$$

Similar to the proof of (2.18), we can obtain that there exists a subset $\kappa \subseteq \{k\}$ such that

$$\lim_{k\in\kappa, k\to\infty} \alpha_k \not\to 0, \tag{2.32}$$

where α_k is given 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 $\nabla\Psi(\omega)$ is continuous, and (2.31) holds, we have that

$$|[\nabla\Psi(\omega^k + \xi_k\theta_k\alpha_k d^k) - \nabla\Psi(\omega^k)]^{\mathrm{T}} d^k| \leqslant \frac{1}{2}(1-\beta)\sigma\|d^k\|, \tag{2.33}$$

where σ is given in (2.28). Noting that Ψ is continuously differentiable and using the mean value theorem, we have the following result with $0 \leqslant \xi_k \leqslant 1$

$$\begin{aligned} \Psi(\omega^k + \alpha_k\theta_k d^k) &= \Psi(\omega^k) + \beta\alpha_k\theta_k\nabla\Psi(\omega^k)^{\mathrm{T}}d^k + (1-\beta)\alpha_k\theta_k\nabla\Psi(\omega^k)^{\mathrm{T}}d^k \\ &\quad +\alpha_k\theta_k[\nabla\Psi(\omega^k + \xi_k\alpha_k\theta_k d^k) - \nabla\Psi(\omega^k)]^{\mathrm{T}}d^k \\ &\leqslant \Psi(\omega^k) + \beta\alpha_k\theta_k\nabla\Psi(\omega^k)^{\mathrm{T}}d^k, \end{aligned} \tag{2.34}$$

here the last second inequality is deduced since the last term in brackets in the right-hand side of equality in (2.34) will become negative when $\alpha_k\theta_k\|d^k\|$ is small enough. And hence the corresponding $\theta_k \to 1$, as $\|d^k\| \to 0$.

We then deduce from

$$\Psi(\omega^k) - \Psi(\omega^k + \alpha_k d^k) \geqslant -\beta\alpha_k(g^k)^{\mathrm{T}}d^k \geqslant \beta\sigma\alpha_k\|d^k\|$$

that for i sufficiently large,

$$\begin{aligned} \|\omega^{m_i} - \omega^{l_i}\| &\leqslant \sum_{k=m_i}^{l_i-1} \|\omega^{k+1} - \omega^k\| \\ &= \sum_{k=m_i}^{l_i-1} \|\alpha_k d^k\| = \sum_{k=m_i}^{l_i-1} \alpha_k\|d^k\| \\ &\leqslant \frac{1}{\beta\sigma}\sum_{k=m_i}^{l_i-1} [\Psi(\omega^k) - \Psi(\omega^{k+1})]) \\ &= \frac{1}{\beta\sigma}[\Psi(\omega^{m_i}) - \Psi(\omega^{l_i})]. \end{aligned} \tag{2.35}$$

(2.30) and (2.35) mean that for large i, we have

$$\|\omega^{m_i} - \omega^{l_i}\| \leqslant \varepsilon_2.$$

(1.3) implies that

$$|(\gamma^{m_i})_j - (\gamma^{l_i})_j| \leqslant |(\omega^{m_i})_j - (\omega^{l_i})_j| \to 0,$$

as i tends to infinity. Finally, from (2.25), (2.27) and triangle inequality, we get that from

$$\|g^{m_i} - g^{l_i}\| = \|(H^{m_i})^{\mathrm{T}}\Phi^{m_i} - (H^{l_i})^{\mathrm{T}}\Phi^{l_i}\| \leqslant \chi_H\|\omega^{m_i} - \omega^{l_i}\| \leqslant \chi_H\varepsilon_2,$$

$$\begin{aligned}\varepsilon_1 &\leqslant \|D_{m_i}^{-1}g^{m_i}\| \\ &\leqslant \|D_{m_i}^{-1}g^{m_i} - D_{m_i}^{-1}g^{l_i}\| + \|D_{m_i}^{-1}g^{l_i} - D_{l_i}^{-1}g^{l_i}\| + \|D_{l_i}^{-1}g^{l_i}\| \\ &\leqslant \|D_{m_i}^{-1}\|\|g^{m_i} - g^{l_i}\| + \|(D_{m_i}^{-1} - D_{l_i}^{-1})g^{l_i}\| + \|D_{l_i}^{-1}g^{l_i}\| \\ &\leqslant \chi_D\varepsilon_2 + \chi_H\chi_\Phi\varepsilon_2 + \varepsilon_2,\end{aligned}$$

which contradicts $\varepsilon_2 \in (0, \varepsilon_1)$, for arbitrarily small. This implies that (2.23) is not true, and hence the conclusion of the theorem holds.

3 Local convergence

In this section we want to show that the Algorithm is locally fast convergent. The assumption in this section is that a KKT point ω^* of (0.1) is a BD-regualr solution of the system $\Phi(\omega) = 0$.

Definition 3.1 *The vector ω^* is called BD-regular for Φ if all elements $H \in \partial_B\Phi(\omega^*)$ are nonsingular.*

The next result follows from the fact that Φ is a (strongly) semismooth operator under certain smoothness assumptions for F, h and c (see [9-12]).

Proposition 3.1 *The following state holds:*

$$\|\Phi(\omega + h) - \Phi(\omega) - Hh\| = c_0\|h\|^{1+q} \quad \text{for } h \to 0 \ \text{ and } H \in \partial\Phi(\omega + h). \tag{3.1}$$

The following proposition refers to Robinson's strong regularity condition. Here, we will not restate its definition and interested reader may consult Robinson [13] and Liu [14] for several characterizations of a strongly regular KKT point.

Proposition 3.2 *A solution $\omega^* = (x^*, y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ of system (0.1) is strongly regular if and only if all matrices in Clarke's generalized Jacobian $\partial\Phi(\omega^*)$ are nonsingular. In particular, the strong regularity of ω^* is sufficient for ω^* to be a BD-regular solution of the system $\Phi(\omega) = 0$.*

The next property follows from the semicontinuity of the generalized Jacobian (see [15]) and the assumed BD-regularity. For the precise proofs, interested reader may refer to Lemma 2.6 in [9] and Proposition 3 in [11].

Proposition 3.3 *Let ω^* be a BD-regular solution of $\Phi(\omega) = 0$. Then the following statement hold:*

1. *There exist constants* $c_1 > 0$ *and* $\delta_1 > 0$ *such that the matrices* $H \in \partial_B \Phi(\omega)$ *are nonsingular and satisfy*

$$\|H^{-1}\| \leqslant c_1 \tag{3.2}$$

for all ω *with* $\|\omega - \omega^*\| \leqslant \delta_1$.

2. *There exist constants* $c_2 > 0$ *and* $\delta_2 > 0$ *such that*

$$\|\Phi(\omega)\| \geqslant c_2 \|\omega - \omega^*\| \tag{3.3}$$

for all ω *with* $\|\omega - \omega^*\| \leqslant \delta_2$.

In order to prove the local convergence theorem, we will need the following lemmas.

Lemma 3.1 *Assume that* $\omega^* = (x^*, y^*, z^*)$ *is a BD-regular solution of* $\Phi(\omega) = 0$. *Then we have that*

$$\|d^k\| \leqslant c_1 \|\Phi(\omega^k)\| \tag{3.4}$$

for all $\omega^k = (x^k, y^k, z^k)$ *sufficiently close to* ω^*, *where* c_1 *is given in Proposition 3.3 and* $d^k = D_k^{-1} \hat{d}^k$ *and* $\hat{d}^k$ *denotes a solution of Levenberg-Marquardt equation.*

Proof Since ω^* is a BD-regular KKT point, the matrix $H^k \in \partial_B \Phi(\omega^k)$ are uniformly nonsigular for all ω^k sufficiently close to ω^* by Proposition 3.3, i.e., there exists a constant $c_1 > 0$ such that

$$\|d^k\| \leqslant \|(H^k)^{-1}\| \|H^k d^k\| \leqslant c_1 \|H^k d^k\|. \tag{3.5}$$

On the other hand, using Levenberg-Marquardt equation, we have

$$\begin{aligned} 0 &= (D_k^{-1} g^k)^{\mathrm{T}} \hat{d}^k + (\hat{d}^k)^{\mathrm{T}} [D_k^{-1} (H^k)^{\mathrm{T}} H^k D_k^{-1} + \mu_k I] \hat{d}^k \\ &\geqslant (g^k)^{\mathrm{T}} d^k + (d^k)^{\mathrm{T}} (H^k)^{\mathrm{T}} H^k d^k \\ &= \Phi(\omega^k) H^k d^k + \|H^k d^k\|^2 \\ &\geqslant \|H^k d^k\| - \|\Phi(\omega^k)\| \cdot \|H^k d^k\|, \end{aligned}$$

so

$$\|H^k d^k\| \leqslant \|\Phi(\omega^k)\|. \tag{3.6}$$

Combining (3.5) with (3.6), we can easily obtain

$$\|d^k\| \leqslant c_1 \|\Phi(\omega^k)\|. \tag{3.7}$$

Theorem 3.1 *Assume that assumptions A1-A2 hold. Let* ω^* *be any accumulation point of the sequence* $\{\omega^k\}$ *generated by the proposed algorithm and* ω^* *be a BD-regular point of* Φ. *Then* ω^* *is a BD-regular zero solution of* Φ *and* $\omega^k \to \omega^*$, *the step size* $\alpha_k \equiv 1$ *for large enough* k.

Proof If $\Phi(\omega^k) = 0$ for some enough k, then from $\|d^k\| \leqslant c_1 \|\Phi(\omega^k)\|$ we have that $d^k = 0$ for all large enough k and the step size $\alpha_k \equiv 1$. Therefore, without loss of generality we may assume that $D_k^{-1} (H^k)^{\mathrm{T}} \Phi(\omega^k) \neq 0$.

From the Levenberg-Marquardt equation, we have

$$\begin{aligned}(g^k)^{\mathrm{T}}d^k &= -(\hat{d}^k)^{\mathrm{T}}[D_k^{-1}(H^k)^{\mathrm{T}}H^kD_k^{-1}+\mu_kI]\hat{d}^k\\ &\leqslant -(d^k)^{\mathrm{T}}(H^k)^{\mathrm{T}}H^kd^k\\ &\leqslant -\frac{\|d^k\|^2}{\|(H^k)^{-1}\|^2}\\ &\leqslant -\frac{\|d^k\|^2}{c_1^2}. \end{aligned} \tag{3.8}$$

Using the acceptance rule (1.9) and (3.8), we have

$$\begin{aligned}\Psi(\omega^k+\alpha_kd^k) &\leqslant \Psi(\omega^{l(k)})+\beta\alpha_k(g^k)^{\mathrm{T}}d^k\\ &\leqslant \Psi(\omega^k)-\frac{\beta}{c_1^2}\alpha_k\|d^k\|^2.\end{aligned} \tag{3.9}$$

Similar to the proof of theorem in [8], since $\{\Psi(\omega^k)\}$ is convergent, we have

$$\lim_{k\to\infty}\alpha_k\|d^k\|^2=0. \tag{3.10}$$

Assume that there exists a subsequence $\kappa\subseteq\{k\}$ such that

$$\lim_{k\to\infty,k\in\kappa}\|d^k\|>0. \tag{3.11}$$

Then, assumption (3.11) implies that

$$\lim_{k\to\infty,k\in\kappa}\alpha_k=0. \tag{3.12}$$

Similar to the proof of (2.18), we have that $\alpha_k\not\to 0$ if α_k is given by (1.10). And at the same time, the acceptance (1.9) means that for large enough k,

$$\Psi(\omega^k+\frac{\alpha_k}{\tau}d^k)-\Psi(\omega^k)\geqslant\beta\frac{\alpha_k}{\tau}(g^k)^{\mathrm{T}}d^k. \tag{3.13}$$

Since

$$\Psi(\omega^k+\frac{\alpha_k}{\tau}d^k)-\Psi(\omega^k)=\frac{\alpha_k}{\tau}(g^k)^{\mathrm{T}}d^k+o(\frac{\alpha_k}{\tau}\|d^k\|)$$

we have that

$$(1-\beta)\frac{\alpha_k}{\tau}(g^k)^{\mathrm{T}}d^k+o(\frac{\alpha_k}{\tau}\|d^k\|)\geqslant 0. \tag{3.14}$$

Dividing (3.14) by $\frac{\alpha_k}{\tau}\|d^k\|$ and noting that $1-\beta>0$, we have that from (3.8)

$$0\leqslant\lim_{k\to\infty,k\in\kappa}\frac{(g^k)^{\mathrm{T}}d^k}{\|d^k\|}\leqslant-\lim_{k\to\infty,k\in\kappa}\frac{\|d^k\|}{c_1^2}\leqslant 0 \tag{3.15}$$

From (3.15), we can get that

$$\lim_{k\to\infty,k\in\kappa}(g^k)^{\mathrm{T}}d^k=0\ \text{ and }\ \lim_{k\to\infty,k\in\kappa}\|d^k\|=0. \tag{3.16}$$

We now prove that if (3.16) holds, then $\alpha_k = 1$ must satisfy the accepted condition (1.9) in step 4.

For large enough k, we have that

$$\begin{aligned} & \Psi(\omega^k + d^k) - \Psi(\omega^k) - (g^k)^{\mathrm{T}} d^k \\ = & \frac{1}{2}\|\Phi(\omega^k) + (H^k)^{\mathrm{T}} d^k + o(\|d^k\|)\|^2 - \frac{1}{2}\|\Phi(\omega^k)\|^2 - (g^k)^{\mathrm{T}} d^k \\ = & \frac{1}{2}\|(H^k)^{\mathrm{T}} d^k\|^2 + o(\|d^k\|^2). \end{aligned} \tag{3.17}$$

This gives that

$$\begin{aligned} \Psi(\omega^k + d^k) &= \Psi(\omega^k) + (g^k)^{\mathrm{T}} d^k + \frac{1}{2}\|(H^k)^{\mathrm{T}} d^k\|^2 + o(\|d^k\|^2) \\ &\leqslant \Psi(\omega^{l(k)}) + \beta (g^k)^{\mathrm{T}} d^k + (\frac{1}{2} - \beta)(g^k)^{\mathrm{T}} d^k \\ &\quad + \frac{1}{2}[(g^k)^{\mathrm{T}} d^k + \|(H^k)^{\mathrm{T}} d^k\|^2] + o(\|d^k\|^2) \\ &\leqslant \Psi(\omega^{l(k)}) + \beta (g^k)^{\mathrm{T}} d^k - (\frac{1}{2} - \beta)\mu_k \|\hat{d}^k\|^2 - \frac{1}{2}\mu_k \|\hat{d}^k\|^2 + o(\|d^k\|^2) \\ &= \Psi(\omega^{l(k)}) + \beta (g^k)^{\mathrm{T}} d^k. \end{aligned} \tag{3.18}$$

Therefore, the accepted condition (1.9) holds when $\alpha_k = 1$.

Now, we prove that if (3.16) holds, when $\alpha_k = 1$, the accepted condition (1.10) given in step 4 also holds at the stepsize to the boundary of box constraints along d^k. (3.16) means that $(d^k)_i \to 0$, for all i. Further, $(d^k_J)_i \to 0$ for all $i = 1, 2, \cdots, m$. If $(g^*_J)_i = 0$ for any i, assume that α_k given in step 4 is the step size to the boundary of box constraints along d^k, the nondegenerate means that $(\omega^*_J)_i > 0$, then

$$\alpha_k = \min\left\{1, \max\left\{\frac{(\omega^k_J)_i}{(d^k_J)_i}, i = 1, 2, \cdots, m\right\}\right\} = \min\{1, +\infty\} = 1.$$

If $(g^*_J)_i \neq 0$ for some i, we have that $(\omega^*_J)_i = 0$. Since $(H^k)^{\mathrm{T}} H^k d^k$ converges to zero and $\mu_k I$ is a positive definite diagonal matrix in (1.8), the nondegenerate condition of reformulated problem (0.3) at the limit point implies that $(d^k_J)_i$ and $-(g^k_J)_i$ have the same sign for k sufficiently large. Hence, if α_k is defined by some $\gamma^i(\omega^*_J) = 0$ and $(g^*_J)_i \neq 0$, then $\alpha_k = |\gamma^i(\omega^k_J)|/|(g^k_J)_i|$ for k sufficiently large. Using (2.17), again, noting $\mu \geqslant 1$, we have

$$\begin{aligned} \alpha_k &= \min\left\{1, \frac{\mu_k}{|(g^k_J)_i + [(H^k_J)^{\mathrm{T}}(H^k_J) d^k_J]_i|}\right\} \\ &\geqslant \min\left\{1, \mu - \frac{\mu\|(H^k_J)^{\mathrm{T}}(H^k_J) d^k_J\|}{\|g^k_J\| + \|(H^k_J)^{\mathrm{T}}(H^k_J) d^k_J\|}\right\} \to 1 \quad \text{as } d^k \to 0. \end{aligned}$$

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

So $\alpha_k \equiv 1$, i.e., $s^k = d^k$ and hence $\omega^{k+1} = \omega^k + d^k$.

Similar to the proof of (2.28), we can get that

$$\begin{aligned}(g^k)^{\mathrm T}d^k &\leqslant -\frac{\sqrt{\mu}\|D_k^{-1}g^k\|^{1+q}\|d^k\|}{\chi_D\sqrt{\chi_H^2\chi_D^2+\mu\chi_D^{2q}\chi_H^{2q}\chi_\Phi^{2q}}}\\ &= -\varrho\|D_k^{-1}g^k\|^{1+q}\|d^k\|, \end{aligned}\tag{3.19}$$

where $\varrho \stackrel{\text{def}}{=} \frac{\sqrt{\mu}}{\chi_D\sqrt{\chi_H^2\chi_D^2+\mu\chi_D^{2q}\chi_H^{2q}\chi_\Phi^{2q}}}$.

So from (3.19), we have

$$0=\lim_{k\to\infty}-\frac{(g^k)^{\mathrm T}d^k}{\|d^k\|}\geqslant \lim_{k\to\infty}\varrho\|D_k^{-1}g^k\|^{1+q}=\varrho\|D_*^{-1}g^*\|^{1+q}, \tag{3.20}$$

which implies that $D_*^{-1}(H^*)^{\mathrm T}\Phi(\omega^*)=D_*^{-1}g^*=0$. Since ω^* is a BD-regular point of Φ, i.e., D_* and H^* nonsingular, this gives $\Phi(\omega^*)=0$ which means that ω^* is a BD-regular zero point of Φ.

Further, ω^* is a BD-regular zero of Φ, then there exist $\delta_2>0$ and $c_2>0$ such that

$$\|\Phi(\omega^k)\|\geqslant c_2\|\omega^k-\omega^*\| \quad \text{for all} \quad \|\omega^k-\omega^*\|\leqslant\delta_2.$$

All the above gives $\omega^k\to\omega^*$.

Theorem 3.2 *Suppose that ω^* is a BD-regular solution of $\Phi(\omega)=0$. Let $\{\omega^k\}$ denote any sequence that converges to ω^* for all k. For each ω^k let $\hat{d}^k$ denote a solution of Levenberg-Marquardt equation. Then*

$$\|\omega^k+d^k-\omega^*\|=O(\|\omega^k-\omega^*\|^{1+q}).$$

Proof By the BD-regularity of ω^* we have for ω^k sufficiently close to ω^* and $H^k\in\partial_B\Phi(\omega^k)$ that

$$\begin{aligned}\|\omega^k+d^k-\omega^*\| &\leqslant \|(H^k)^{-1}\|\|H^k(\omega^k+d^k-\omega^*)\|\\ &\leqslant c_1\|H^kd^k+H^k(\omega^k-\omega^*)\|\\ &\leqslant c_1(\|\Phi(\omega^k)+H^kd^k\|+\|\Phi(\omega^k)-\Phi(\omega^*)-H^k(\omega^k-\omega^*)\|)\\ &= c_1(\|\Phi(\omega^k)+H^kd^k\|+c_0\|\omega^k-\omega^*\|^{1+q}). \end{aligned}\tag{3.21}$$

Since $\hat{d}^k$ is a solution of Levenberg-Marquardt equation and $\bar{d}^k=D_k(\omega^*-\omega^k)$ is feasible for (1.7), we obtain that

$$\begin{aligned}\|\Phi(\omega^k)+H^kd^k\|^2 &= \|\Phi(\omega^k)+H^kD_k^{-1}\hat{d}^k\|^2\\ &\leqslant \|\Phi(\omega^k)+H^kD_k^{-1}\hat{d}^k\|^2+\mu_k\|\hat{d}^k\|^2\\ &\leqslant \|\Phi(\omega^k)+H^kD_k^{-1}\bar{d}^k\|^2+\mu_k\|\bar{d}^k\|^2\\ &\leqslant \|\Phi(\omega^k)-\Phi(\omega^*)+H^k(\omega^*-\omega^k)\|^2+\mu_k\|D_k\|^2\|\omega^*-\omega^k\|^2\\ &= \|\Phi(\omega^k)-\Phi(\omega^*)-H^k(\omega^k-\omega^*)\|^2+\mu_k\|D_k\|^2\|\omega^*-\omega^k\|^2. \end{aligned}\tag{3.22}$$

For ω^* with $\omega_J^*>0$, there exists sufficiently small $\delta\in(0,2]$ such that the open ball $\mathcal{B}(\omega^*,\delta)\stackrel{\text{def}}{=}\{\omega \mid \|\omega-\omega^*\|<\delta,\ \omega_J>0\}$.

Let $\{\omega^{k_j}\}$ be subsequence such that $\omega^{k_j} \to \omega^*$ and j_0 be the index such that for $k > k_{j_0}$, the sequence $\{\omega^{k_j}\}$ belongs to $\mathcal{B}(\omega^*, \frac{\delta}{2})$. Assume $k_j > k_{j_0}$. Then $|(\omega_J^{k_j})_i| > \frac{\delta}{2}$ for $i = 1, 2, \cdots, m$. Hence, $\|D_{k_j}\| \leqslant n + p + m\sqrt{\frac{2}{\delta}}$ where $\sqrt{\frac{2}{\delta}} \geqslant 1$. Taking into account that

$$\mu_k = \mu\|D_k^{-1}g^k\|^{2q} \leqslant \mu\chi_D^{2q}\chi_H^{2q}L^{2q}\|\omega^k - \omega^*\|^{2q},$$

we can obtain from (3.22) that

$$\|\Phi(\omega^k) + H^k d^k\|^2 \leqslant [c_0^2 + \mu\chi_D^{2q}\chi_H^{2q}L^{2q}(n + p + m\sqrt{\frac{2}{\delta}})^2]\|\omega^k - \omega^*\|^{2+2q}, \tag{3.23}$$

that is,

$$\|\Phi(\omega^k) + H^k d^k\| \leqslant c_3\|\omega^k - \omega^*\|^{1+q}, \tag{3.24}$$

where $c_3 \overset{\text{def}}{=} \sqrt{c_0^2 + \mu\chi_D^{2q}\chi_H^{2q}L^{2q}(n + p + m\sqrt{\frac{2}{\delta}})^2}$.

Together (3.21) with (3.24), we have that

$$\|\omega^k + d^k - \omega^*\| = O(\|\omega^k - \omega^*\|^{1+q}). \tag{3.25}$$

Theorem 3.2 shows that under the assumption that a KKT point ω^* of (0.1) is a BD-regualr solution of the system $\Phi(\omega) = 0$, the proposed algorithm has locally Q-superlinear at (1+q)-order of convergence rate.

4 Numerical experiments

Numerical experiments on a new affine scaling Levenberg-Marquardt method with non-monotonic interior backtracking line search technique given in this paper have been performed on computer. The experiments are carried out on 4 test problems which are quoted from [16]. Here, we first transform them into the Karush-Kuhn-Tucker system, and then start searching from the initial points given by [16] to get the numerical results. The computation terminates when one of the following stopping criterions is satisfied which is either $\|D_k^{-1}g^k\| \leqslant 10^{-8}$ or $\|\Psi_{k+1} - \Psi_k\| \leqslant 10^{-8}$. The selected parameter values are: $\varepsilon = 10^{-8}$, $\beta = 0.25$, $\mu = 1$, $q = 0.5$, $\tau = 0.5$, $\theta_l = 0.95$ and $M = 5$. NI and NF stand for the numbers of iterations and function evaluations of Ψ respectively. The numerical results of our Levenberg-Marquardt algorithm and the Modified BFGS given in [16] are presented in the following Table 4.1.

Table 4.1 Experimental results

Problem name	Levenberg-Marquardt algorithm	Modified BFGS in [16]
	NI/NF	NI/NF
Powell's function of four variables	24/24	45/51
Wood's function	35/35	54/66
A quartic function	18/18	55/61
A sine-valley function	35/36	39/54

References

[1] Facchinei F, Fischer A, Kanzow C. Regularity properties of a semismooth reformlation of variational inequalities [J]. *SIAM Journal on Optimization*, 1998, **8**(3): 850-869.

[2] Fischer A. A special Newton-type optimization method [J]. *Optimization*, 1992, **24**(3-4): 269-284.

[3] Facchinei F, Fischer A, Kanzow C. A semismooth Newton method for variational inequalities: The case of box constraints [C]//it Complementarity and Variational Problems: State of the Art, Philadelphia: [s.n.], 1997: 76-90.

[4] Qi L Q, Jiang H Y. Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton and Quasi-Newton methods for solving these equations [J]. *Mathematics of Operations Research*, 1997, **22**(2): 301-325.

[5] Coleman T F, Li Y Y. An interior trust region approach for nonlinear minimization subject to bounds [J]. *SIAM Journal on Optimization*, 1996, **6**(2): 418-445.

[6] Clarke F H. Optimization and nonsmooth analysis [M]. New York: John Wiley and Sons, 1983.

[7] Zhu D T. Affine scaling interior Levenberg-Marquardt method for bound-constrained semismooth equations under local error bound conditions [J]. *Journal of Compututional and Applied Mathematics*, 2008, **219**(1): 198-215.

[8] Gripp L, Lampariello F, Lucidi L. A nonmonotone line search technique for Newton's methods [J]. *SIAM Journal on Numerical Analysis*, 1986, **23**(4): 707-716.

[9] Qi L Q. Convergence analysis of some algorithms for solving nonsmooth equations [J]. *Mathematics of Operations Research*, 1993, **18**(1): 227-244.

[10] Qi L Q, Sun J. A nonsmooth version of Newton's method [J]. *Mathematical Programming*, 1993, **58**(1-3): 353-367.

[11] Pang J S, Qi L Q. Nonsmooth equations: motivation and algorithms [J]. *SIAM Journal on Optimization*, 1993, **3**(3): 443-465.

[12] Fischer A. Solution of monotone complementarity problems with locally Lipschitzian functions [J]. *Mathematical Programming*, 1997, **76**(3): 513-532.

[13] Robinson S M. Strongly regular generalized equations [J]. *Mathematics of Operations Research*, 1980, **5**(1): 43-62.

[14] Liu J M. Strong stability in variational inequalities [J]. *SIAM Journal on Control and Oprimization*, 1995, **33**(3): 725-749.

[15] Bertsekas D P. Constrained optimization and Lagrange multiplier methods [M]. New York: Academic Press, 1982.

[16] Yuan Y Y. A modified BFGS algorithm for unconstrained optimization [J]. *IMA Journal of Numerical Analysis*, 1991, **11**(3): 325-332.