Shou-qiang Du; Yan Gao

Global convergence property of modified Levenberg-Marquardt methods for nonsmooth equations

Applications of Mathematics, Vol. 56 (2011), No. 5, 481–498

Persistent URL: `http://dml.cz/dmlcz/141620`

GLOBAL CONVERGENCE PROPERTY OF MODIFIED LEVENBERG-MARQUARDT METHODS FOR NONSMOOTH EQUATIONS*

Shou-Qiang Du, Qingdao, Yan Gao, Shanghai

(Received May 21, 2009)

Abstract. In this paper, we discuss the globalization of some kind of modified Levenberg-Marquardt methods for nonsmooth equations and their applications to nonlinear complementarity problems. In these modified Levenberg-Marquardt methods, only an approximate solution of a linear system at each iteration is required. Under some mild assumptions, the global convergence is shown. Finally, numerical results show that the present methods are promising.

Keywords: nonsmooth equations, modified Levenberg-Marquardt method, global convergence, nonlinear complementarity problem

MSC 2010: 65H10, 90C30

1. Introduction

In the past few years, there has been a growing interest in the study of nonsmooth equations, which is a powerful tool to study the variational inequalities problem and the nonlinear complementarity problem, see for instance [2], [7], [8], [12], [13], [15]. The variational inequalities problem is to find $x \in C$ such that

$$f(x)^\top (y - x) \geqslant 0 \tag{1.1}$$

* Supported by National Science Foundation of China (10671126), Shanghai Leading Discipline Project (S30501), Innovation Program of Shanghai Education Commission (10YZ99), and Higher Educational Science and Technology Program of Shandong Province (J10LA05).

for all $y \in C$, where C is a closed convex set in $\mathbb{R}^n$ and $f\colon \mathbb{R}^n \to \mathbb{R}^n$. The problem (1.1) can be reformulated as the system of nonsmooth equations

$$x - \mathrm{Proj}_C(x - f(x)) = 0,$$

where $\mathrm{Proj}_C(z)$ is the projection of $z \in \mathbb{R}^n$ onto C.

The nonlinear complementarity problem (for short NCP) is to find a point in $\mathbb{R}^n$ satisfying

$$x \geqslant 0, \quad f(x) \geqslant 0, \quad x^\top f(x) = 0, \tag{1.2}$$

where $f(x) = (f_1(x), \ldots, f_n(x))^\top\colon \mathbb{R}^n \to \mathbb{R}^n$ is a continuously differentiable function. The problem (1.2) has many important applications in mathematical programming, economic equilibrium and mechanics, see [3], [7], [15]. Based on a nonlinear complementarity function, a nonlinear complementarity problem can be reformulated as a nonsmooth equation. Let us consider the nonlinear complementarity function, proposed by Fischer in [5]:

$$\varphi(a, b) = \sqrt{a^2 + b^2} - a - b. \tag{1.3}$$

The nonlinear complementarity function φ plays an important role in the area of numerical methods for complementarity problems, constrained optimization and variational inequality problems, see [6], [14]. It is easy to see that the function φ has the property:

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

Define $H\colon \mathbb{R}^n \to \mathbb{R}^n$ as

$$H(x) := (H_1(x), \ldots, H_n(x))^\top = \begin{pmatrix} \varphi(x_1, f_1(x)) \\ \vdots \\ \varphi(x_n, f_n(x)) \end{pmatrix}.$$

It is easy to see that the nonlinear complementarity problem (1.2) is equivalent to the nonsmooth system $H(x) = 0$.

Facchinei and Kanzow in [7] proposed an inexact Levenberg-Marquardt-type method for the solution of the nonlinear complementarity problem based on the nonsmooth equation method. Actually, nonsmooth equations are much more difficult than smooth ones. Many existing classical results for smooth equations cannot be extended to nonsmooth equations directly. This difficulty motivates us to invoke the classical tool for solving smooth equations to solve nonsmooth equations, for instance, the one based upon the generalized Jacobian and the inexact Levenberg-Marquardt-type method, see for instance [7], [15].

In this paper, we explore Levenberg-Marquardt methods for the solution of the general nonsmooth equation

$$H(x) = 0. \tag{1.4}$$

Then we study their applications to a nonlinear complementarity problem and nonsmooth equations of maximums of finitely many smooth functions. At each iteration, our methods require the approximate solution of a symmetric positive semidefinite and solvable linear system. Denote the natural merit function by

$$\Psi(x) = \frac{1}{2} H(x)^\top H(x).$$

In order to globalize the local method we perform line search to minimize the natural merit function

$$\psi(a,b) = \frac{1}{2}\varphi^2(a,b), \tag{1.5}$$
$$\Psi(x) = \sum_{i=1}^{n} \psi(x_i, f_i(x)).$$

This paper is organized as follows: In Section 2, we recall some results on the generalized Jacobian and semismoothness. Some important properties of the operators H and Ψ are also summarized in this section. In Section 3, the globalization of modified Levenberg-Marquardt methods for nonsmooth equations and convergence results are given. Numerical tests are reported in Section 4.

2. Preliminaries

We start with some notions and propositions, which can be found in [7], [8], [11], [15].

Let $F\colon \mathbb{R}^n \to \mathbb{R}^n$ be locally Lipschitzian. Then it is almost everywhere F-differentiable. Denote the set of points where F is F-differentiable by D_F. The B-differential of F at $x \in \mathbb{R}^n$ is defined as

$$\partial_B F(x) = \{V \in \mathbb{R}^{n\times n}\colon \exists \{x_k\} \in D_F,\ x_k \to x,\ \{F'(x_k)\} \to V\}.$$

The general Jacobian of $F\colon \mathbb{R}^n \to \mathbb{R}^n$ at x in the sense of Clark is defined by

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

Proposition 2.1. *The set $\partial_B F(x)$ is nonempty and compact for any x. The set-valued mapping $x \mapsto \partial_B F(x)$ is upper semicontinuous.*

Definition 2.1. $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is said to be semismooth at x if F is locally Lipschitz at x and

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

exists for any $h \in \mathbb{R}^n$.

Proposition 2.2. *Suppose that $x^\star$ is a solution of (1.4) and for any $V \in \partial_B H(x^\star)$ it is nonsingular. Then there exist a neighborhood $N(x)$ of $x^\star$ and a constant c such that*

$$\|V^{-1}\| \leqslant c, \quad \forall\, V \in \partial_B H(x),\ \forall\, x \in N(x).$$

Proposition 2.3. *If $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz continuous and semismooth at x, then we have*

$$\lim_{\substack{V \in \partial F(x+th) \\ h \to 0}} \frac{\|F(x+h) - F(x) - Vh\|}{\|h\|} = 0.$$

Furthermore, if $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz continuous, strongly semismooth at x and directionally differentiable in a neighborhood of x, then

$$\limsup_{\substack{V \in \partial F(x+th) \\ h \to 0}} \frac{\|F(x+h) - F(x) - Vh\|}{\|h\|^2} < \infty.$$

Proposition 2.4. *Let $f(x) = (f_1(x), \ldots, f_n(x))^\top$ be given in (1.2). We have:*

(I) *If f is continuously differentiable, then H is semismooth.*

(II) *If f is continuously differentiable and $f'(x)$ is locally Lipschitzian, then H is strongly semismooth.*

(III) *If f is continuously differentiable then Ψ is also continuously differentiable, and its gradient at a point $x \in \mathbb{R}^n$ is given by $\nabla\Psi(x) = V^\top H(x)$, where V can be an arbitrary element in $\partial_B H(x)$.*

Proposition 2.5. *Suppose f given in* (1.2) *is continuously differentiable. Then:*
(I) $\Psi(x)=0$ *if and only if x solves the NCP.*
(II) *The set of solutions of the NCP coincides with the set of global minima of Ψ if the NCP has a solution.*

Definition 2.2. $f\colon \mathbb{R}^n\to\mathbb{R}^n$ is said to be monotone if

$$(x-y)^\top(f(x)-f(y))\geqslant 0$$

for all $x,y\in\mathbb{R}^n$, and f is said to be strongly monotone with modulus $\mu>0$ if

$$(x-y)^\top(f(x)-f(y))\geqslant \mu\|x-y\|^2$$

for all $x,y\in\mathbb{R}^n$.

Proposition 2.6 (see [9])**.** *Suppose the function f is continuously differentiable. Then f is monotone if and only if $\nabla f(x)$ is positive semidefinite for all $x\in\mathbb{R}^n$.*

3. Modified Levenberg-Marquardt methods and their globalization

In this section, we describe some kind of modified Levenberg-Marquardt methods for nonsmooth equations (1.4) with applications to nonlinear complementarity systems and finitely many maximum functions systems and give global convergence results. Roughly speaking, the following algorithms can be taken as an attempt to solve the semismooth system of equations by using the inexact Levenberg-Marquardt type method. We now give a modified Levenberg-Marquardt method for (1.4).

Modified Levenberg-Marquardt Method (I)
Step 0. Given an initial point $x_0\in\mathbb{R}^n$ and parameters $\varrho>0$, $p>2$, $\beta\in(0,\frac12)$, $a<1$, $\varepsilon\geqslant 0$, $\lambda_i^k\in\mathbb{R}^n$ with $0<|\lambda_i^k|<\infty$.
Step 1. If $\Psi(x_k)\leqslant\varepsilon$, stop.
Step 2. Select an element $V_k\in\partial_B H(x_k)$, find an approximate solution $d_k\in\mathbb{R}^n$ of the system

$$((V_k)^\top V_k+\operatorname{diag}(\lambda_i^{(k)}H_i(x_k)))d=-(V_k)^\top H(x_k)+r_k,\quad V_k\in\partial_B H(x) \tag{3.1}$$

for $i=1,\dots,n$ which satisfies

$$\|r_k\|\leqslant\alpha_k\|(V_k)^\top H(x_k)\|, \tag{3.2}$$

where r_k is the vector of residuals and α_k is a sequence of positive numbers such that $\alpha_k \leqslant a < 1$ for every k. If the condition

$$\nabla\Psi(x_k)^\top d_k \leqslant -\varrho\|d_k\|^p \tag{3.3}$$

is not satisfied, set $d_k = -(V_k)^\top H(x_k)$.

Step 3. Find the smallest $i^k \in \{0, 1, 2, \ldots\}$ such that

$$\Psi(x_k + 2^{-i^k} d_k) \leqslant \Psi(x_k) + \beta 2^{-i^k} \nabla\Psi(x_k)^\top d_k. \tag{3.4}$$

Set $x_{k+1} = x_k + 2^{-i^k} d_k$, let $k := k+1$, and go to Step 1.

Notice that if V_k is nonsingular in (3.1), the choice of $\lambda_i^k = 0$, $\|r_k\| = 0$ at each step is allowed by the above algorithm. Then (3.1) is equivalent to the generalized Newton equation in [13]. In what follows, as usual in analyzing the behavior of algorithms, we shall assume that $\varepsilon = 0$. Then the algorithm produces an infinite sequence of points. Similarly to Theorem 12 in [7], we give the following global convergence result.

Theorem 3.1. *Suppose that there exist constants $M > 0$ such that*

$$\|\mathrm{diag}(\lambda_i^{(k)} H_i(x))\| \leqslant M < \infty.$$

Then each accumulation point of the sequence $\{x_k\}$ generated by the above Modified Levenberg-Marquardt Method (I) *is a stationary point of Ψ.*

P r o o f. Assume that $\{x_k\}_K \to x^\star$. If there are infinitely many $k \in K$ such that $d_k = -\nabla\Psi(x_k)$, then the assertion follows immediately from Proposition 1.16 in [1]. Without loss of generality, we assume that if $\{x_k\}_K$ is a convergent subsequence of $\{x_k\}$, then d_k is always given by the solution of (3.1). We show that for every convergent subsequence $\{x_k\}_K$ for which

$$\lim_{k\in K,\, k\to\infty} \nabla\Psi(x_k) \neq 0, \tag{3.5}$$

we have

$$\limsup_{k\in K,\, k\to\infty} \|d_k\| < \infty \tag{3.6}$$

and

$$\limsup_{k\in K,\, k\to\infty} |\nabla\Psi(x_k)^\top d_k| > 0. \tag{3.7}$$

In what follows, we assume that $x_k \to x^\star$. Suppose that $x^\star$ is not a stationary point of Ψ. From (3.1) we have

$$\begin{aligned}\|\nabla\Psi(x_k) - r_k\| &= \|((V_k)^\top V_k + \operatorname{diag}(\lambda_i^{(k)} H_i(x_k)))d_k\| \\ &\leqslant \|(V_k)^\top V_k + \operatorname{diag}(\lambda_i^{(k)} H_i(x_k))\|\|d_k\|,\end{aligned} \tag{3.8}$$

so

$$\|d_k\| \geqslant \frac{\|\nabla\Psi(x_k) - r_k\|}{\|(V_k)^\top V_k + \operatorname{diag}(\lambda_i^{(k)} H_i(x_k))\|}.$$

Note that the denominator in the above inequality is nonzero, otherwise we have $\nabla\Psi(x_k) - r_k = 0$ because of (3.8) together with (3.2), and we get $\|\nabla\Psi(x_k)\| = 0$. Then x_k is a stationary point and the algorithm has stopped. By assumption $\|\operatorname{diag}(\lambda_i^{(k)} H_i(x))\| \leqslant M < \infty$ and Proposition 2.1 there exists a constant $k_1 > 0$ such that

$$\|(V_k)^\top V_k + \operatorname{diag}(\lambda_i^{(k)} H_i(x_k))\| \leqslant k_1.$$

From the above inequality and (3.2), we obtain

$$\|d_k\| \geqslant \frac{1-\alpha_k}{k_1}\|\nabla\Psi(x_k)\| \geqslant \frac{1-a}{k_1}\|\nabla\Psi(x_k)\|. \tag{3.9}$$

Formula (3.6) now readily follows from the fact that we have assumed that the direction satisfies (3.3) with $p > 2$, while the gradient $\nabla\Psi(x_k)$ is bounded on the convergent sequence $\{x_k\}$. If (3.7) is not satisfied there exists a subsequence $\{x_k\}_{K'}$ of $\{x_k\}_K$ with

$$\lim_{k\in K',\, k\to\infty} |\nabla\Psi(x_k)^\top d_k| = 0.$$

This implies, by (3.3), that $\lim_{k\in K',\, k\to\infty} \|d_k\| = 0$. Together with (3.9) it implies

$$\lim_{k\in K',\, k\to\infty} \|\nabla\Psi(x_k)\| = 0,$$

which contradicts (3.5). The sequence $\{d_k\}$ is uniformly gradient related to $\{x_k\}$ according to the definition given in [1] and the assertion of the theorem also follows from Proposition 1.16 in [1]. □

Theorem 3.2. *Let the assumptions of Theorem* 3.1 *hold. Let* $\{x_k\}$ *be any sequence generated by the algorithm. If one of the accumulation points of* $\{x_k\}$*, say* $x^\star$*, is an isolated solution of NCP* (1.1)*, then the entire sequence* $\{x_k\}$ *converges to* $x^\star$*.*

Proof. The thesis follows by Theorem 3.1 and Proposition 2.5. Let K be a subset of $\{1, 2, \ldots\}$ such that $x_k \to x^\star$, $k \in K$. Since $\{\|\nabla\Psi(x_k)\|\}_K \to 0$, we get, either because $d_k = -\nabla\Psi(x_k)$ or by (3.3) with $p > 2$ that $\{d_k\}_K \to 0$. By Lemma 4.10 in [11], the entire sequence $\{x_k\}$ converges to $x^\star$. □

R e m a r k 3.1. In Modified Levenberg-Marquardt Method (I), we also can use the non-monotone line search stepsize

$$\Psi(x_k + 2^{-i^k} d_k) - \max_{0 \leqslant j \leqslant m(k)} \Psi(x_{k-j}) \leqslant \beta 2^{-i^k} \nabla \Psi(x_k)^\top d_k,$$

where $m(0) = 0$, $m(k) = \min\{m(k-1)+1, M_0\}$, M_0 is a nonnegative integer.

Finitely many maximum functions systems are also very useful in the study of nonlinear complementarity problems, variational inequality problems, Karush-Kuhn-Tucker systems of nonlinear programming problems and many problems in mechanics and engineering. The finitely many maximum functions system which have been proposed in [8] as

$$\begin{aligned} \max_{j \in J_1} H_{1j}(x) &= 0, \\ &\vdots \\ \max_{j \in J_n} H_{nj}(x) &= 0, \end{aligned}$$

where $H_{ij}\colon \mathbb{R}^n \to \mathbb{R}$ for $j \in J_i$, $i = 1, \ldots, n$ are continuously differentiable, J_i for $i = 1, \ldots, n$ are finite index sets. Denote

$$\begin{aligned} H_i(x) &= \max_{j \in J_i} H_{ij}(x), \quad x \in \mathbb{R}^n,\ i = 1, \ldots, n, \\ H(x) &= (H_1(x), \ldots, H_n(x))^\top, \quad x \in \mathbb{R}^n, \\ J_i(x) &= \{j_i \in N\colon\ H_{ij}(x) = H_i(x)\}, \quad x \in \mathbb{R}^n,\ i = 1, \ldots, n; \end{aligned}$$

the above finitely many maximum functions system can also be rewritten as (1.4).

R e m a r k 3.2. Modified Levenberg-Marquardt Method (I) can also be used for the above finitely many maximum functions system.

Now we give another Modified Levenberg-Marquardt method for nonsmooth equations (1.4) with applications to nonlinear complementarity systems (1.2).

Modified Levenberg-Marquardt Method (II)

Data. Given an initial $x_0 \in \mathbb{R}^n$ and parameters $\varrho > 0$, $p > 2$, $\beta \in (0,1)$, $\sigma \in (0,1)$, $a < 1$, $\varepsilon \geqslant 0$, $\lambda_i^k \in \mathbb{R}^n$ with $0 < |\lambda_i^k| < \infty$.
Step 1. If $\|\nabla \Psi(x_k)\| \leqslant \varepsilon$, stop.
Step 2. Select an element $V_k \in \partial_B H(x_k)$, find an approximate solution $d_k \in \mathbb{R}^n$ of the system

$$((V_k)^\top V_k + \operatorname{diag}(\lambda_i^{(k)} H_i(x_k)))d = -(V_k)^\top H(x_k) + r_k, \quad V_k \in \partial_B H(x),$$

for $i = 1, \ldots, n$, where r_k is the vector of residuals

$$\|r_k\| \leqslant \alpha_k \|(V_k)^\top H(x_k)\|,$$

where α_k is a sequence of positive numbers such that $\alpha_k \leqslant a < 1$ for every k. If the condition

$$\nabla \Psi(x_k)^\top d_k \leqslant -\varrho \|d_k\|^p$$

is not satisfied, set $d_k = -(\nabla_b \psi(x_1^k, f_1(x_k)), \ldots, \nabla_b \psi(x_n^k, f_n(x_k)))^\top$, where

$$\nabla_b \psi(0,0) = 0,\ (a,b) \neq (0,0),\ \nabla_b \psi(a,b) = (b/\sqrt{a^2+b^2} - 1)\varphi(a,b).$$

Step 3. Find the smallest nonnegative integer, say m^k, satisfying

$$\Psi(x_k + \beta^{m^k} d_k) - \Psi(x_k) \leqslant -\sigma(\beta^{m^k})^2 \Psi(x_k). \tag{3.10}$$

Set $x_{k+1} = x_k + \beta^{m^k} d_k$, let $k := k+1$, and go to Step 1.

Remark 3.3. In Step 3 of Modified Levenberg-Marquardt Method (II), a change is made for the line search rule. This line search rule uses only the function values of Ψ. This line search is motivated by the work [10].

Lemma 3.1. *Let φ and ψ be defined by* (1.3) *and* (1.5)*, respectively. Then*

(i) $\varphi(a,b) = 0$ *if and only if* $\psi(a,b) = 0$.

(ii) $\nabla_a \psi(a,b) \cdot \nabla_b \psi(a,b) \geqslant 0$ *for all* $(a,b) \in \mathbb{R}^n$. *The equality holds if and only if* $\varphi(a,b) = 0$.

Proof. (i) The desired result is satisfied by virtue of the definition (1.3) and (1.5).

(ii) By direct computation we obtain $\nabla_a \psi(0,0) = \nabla_b \psi(0,0) = 0$. For $(a,b) \neq (0,0)$,

$$\nabla_a \psi(a,b) = \Big(\frac{a}{\sqrt{a^2+b^2}} - 1\Big)\varphi(a,b),$$
$$\nabla_b \psi(a,b) = \Big(\frac{b}{\sqrt{a^2+b^2}} - 1\Big)\varphi(a,b).$$

Clearly,

$$\nabla_a \psi(a,b) \cdot \nabla_b \psi(a,b) = \Big(\frac{a}{\sqrt{a^2+b^2}} - 1\Big)\Big(\frac{b}{\sqrt{a^2+b^2}} - 1\Big)\varphi^2(a,b).$$

It follows immediately that $\nabla_a \psi(a,b) \cdot \nabla_b \psi(a,b) \geqslant 0$ for all $(a,b) \in \mathbb{R}^n$. The equality holds if and only if $\varphi(a,b) = 0$. □

Lemma 3.2. *Suppose f in* (1.2) *is continuously differentiable and monotone. Line search rule* (3.10) *is then well defined.*

P r o o f. Assume that there is no nonnegative integer satisfying the line search rule (3.10). It follows that for any integer $l \geqslant 0$ we have

$$\Psi(x_k + \beta^l d_k) - \Psi(x_k) > -\sigma(\beta^l)^2 \Psi(x_k).$$

Dividing the above inequality by β^l and letting $l \to \infty$, we get

$$\Psi'(x_k, d_k) \geqslant 0.$$

By the continuous differentiability of Ψ on $\mathbb{R}^n$, we find that

$$\nabla\Psi(x_k)^\top d_k = \Psi'(x_k, d_k) \geqslant 0.$$

On the other hand, if x_k is not a solution of $\Psi(x)$, from the gradient of $\nabla\Psi(x_k)$ we have

$$\nabla\Psi(x_k)^\top d_k = -\sum_{i=1}^n \nabla_a\psi(x_i^k, f_i(x_k))\nabla_b\psi(x_i^k, f_i(x_k)) - (d_k)^\top \nabla f(x_k)(d_k).$$

By Proposition 2.6, the second term of the above equation is nonnegative. By Lemma 3.1, the first term of the above equation is also nonnegative. So we get

$$\nabla\Psi(x_k)^\top d_k < 0.$$

This leads to a contradiction. Thus, the line search rule (3.10) is well defined. □

Lemma 3.3. *Suppose f in* (1.2) *is continuously differentiable and monotone. If $\nabla\Psi(x_k)^\top d_k = 0$, then we have $\Psi(x_k) = 0$.*

P r o o f. From $\nabla\Psi(x_k)^\top d_k = 0$ we get $\nabla_a\psi(x_i^k, f_i(x_k))\nabla_b\psi(x_i^k, f_i(x_k)) = 0$. By Lemma 3.1 we have $\varphi(x_k^i, f_i(x_k)) = 0$, $\Psi(x_k) = 0$. □

Lemma 3.4. *Assume f in* (1.2) *is continuously differentiable and strongly monotone. Then the level set*

$$L(\Psi, \gamma) = \{x \in \mathbb{R}^n \colon \Psi(x) \leqslant \gamma\}$$

is bounded for any $\gamma \in \mathbb{R}^n$.

P r o o f. Suppose there exists an unbounded sequence $\{x_k\}_{k\in K} \to \infty$ with $\{\|x_k\|\}_{k\in K} \subset L(\Psi,\gamma)$ for some $\gamma \geqslant 0$, where K is a subset of N. We define $J = \{i \in \{1,2,\ldots\}\}$ when $\{x_i^k\}$ is unbounded. Since $\{x_k\}$ is unbounded, $J \neq \emptyset$. Let $\{z_k\}$ denote a bounded sequence defined by $z_i^k = 0$ if $i \in J$, $z_i^k = x_i^k$ otherwise. Then from the definition of $\{z_k\}$ and the strong monotonicity of $f(x)$ we obtain

$$\begin{aligned}\mu\sum_{i\in J}(x_i^k)^2 &= \mu\|x_k - z_k\|^2 \leqslant (x_k - z_k)^\top(f(x_k) - f(z_k))\\ &= \sum_{i\in J} x_i^k(f_i(x_k) - f_i(z_k)) \leqslant \Bigl(\sum_{i\in J}(x_i^k)^2\Bigr)^{1/2}\sum_{i\in J}|f_i(x_k) - f_i(z_k)|.\end{aligned}$$

Since $\sum\limits_{i\in J}(x_i^k)^2 \neq 0$ for $k \in K$, hence dividing by $\sum\limits_{i\in J}(x_i^k)^2$ both sides of the above formula, we get

$$\mu\Bigl(\sum_{i\in J}(x_i^k)^2\Bigr)^{1/2} \leqslant \sum_{i\in J}|f_i(x_k) - f_i(z_k)|, \quad k \in K. \tag{3.11}$$

On the other hand, we know that $\{f_i(z_k)\}_{k\in K}$ is bounded $(i \in J)$, because $\{z_k\}_{k\in K}$ is bounded and $f(x)$ is continuous. From (3.11) we know that $\{|f_{i_0}(x_k)|\} \to \infty$ for some $i_0 \in J$. Also, $\{\|x_{i_0}^k\|\} \to \infty$ by the definition of the index set J. Thus, when $k \to \infty$ then

$$\varphi(x_{i_0}^k, f_{i_0}(x_k)) \to \infty.$$

This contradicts $\{x_k\} \subset L(\Psi,\gamma)$. □

Theorem 3.3. *Suppose f in* (1.2) *is continuously differentiable and monotone. Then each accumulation point of the sequence $\{x_k\}$ generated by the above procedure* (II) *is a stationary point of Ψ.*

P r o o f. Assume that $\{x_k\}_K \to x^\star$. If there are infinitely many $k \in K$ such that $d_k = -(\nabla_b\psi(x_1^k, f_1(x_k)),\ldots,\nabla_b\psi(x_n^k, f_n(x_k)))^\top$, assume that $x^\star$ is an accumulation point of $\{x_k\}$, say the limit of a subsequence of $\{x_k,\, k \in K\}$. Then $\{x_k,\, k \in K\}$ is bounded, which implies that $\{d_k,\, k \in K\}$ is also bounded by virtue of the continuous differentiability of ψ. Without loss of generality, we may assume $d_k \to d^\star$, $k \to \infty$, $k \in K$. Next, we discuss two cases. If $\{m^k,\, k \in K\}$ is bounded, then from (3.10)

$$\sum_{k\in K}\Psi(x_k) < \infty.$$

This shows that $\Psi(x^\star) = 0$, i.e., $x^\star$ is a solution of the NCP.

Assume that $\{m^k,\, k \in K\}$ is unbounded. Clearly,

$$\nabla\Psi(x^\star)^\top d^\star \leqslant 0. \tag{3.12}$$

On the other hand, the line search rule (3.10) yields

$$\Psi(x_k + \beta^{m^k} d_k) - \Psi(x_k) \leqslant -\sigma(\beta^{m^k})^2 \Psi(x_k), \quad k \in K$$

and

$$\Psi(x_k + \beta^{m^k-1} d_k) - \Psi(x_k) > -\sigma(\beta^{m^k-1})^2 \Psi(x_k), \quad k \in K.$$

Dividing the above inequality by β^{m^k-1}, taking the limit and using Lemma 3.4, we have

$$\nabla\Psi(x^\star)^\top d^\star \geqslant 0.$$

From (3.12) we get

$$\nabla\Psi(x^\star)^\top d^\star = 0.$$

By Lemma 3.3, $\Psi(x^\star) = 0$, i.e., $x^\star$ is a solution of the NCP. Hence, we can assume without loss of generality that if $\{x_k\}_K$ is a convergent subsequence of $\{x_k\}$, then d_k is always given by (3.1). The rest is similar to Theorem 3.1, so we omit it. We have completed the proof. □

Theorem 3.4. *Let the assumptions of Theorem* 3.3 *hold. Let* $\{x_k\}$ *be a sequence generated by the algorithm. If one of the accumulation points of* $\{x_k\}$*, say* $x^\star$*, is an isolated solution of NCP* (1.2)*, then the entire sequence* $\{x_k\}$ *converges to* $x^\star$*.*

Proof. By Theorem 3.3 and Proposition 2.5, we get the conclusion of the theorem immediately. □

Remark 3.4. In Modified Levenberg-Marquardt Method (II), we also can use the non-monotone line search; we have stepsize β^{m^k},

$$\Psi(x_k + \beta^{m^k} d_k) - \max_{0 \leqslant j \leqslant m(k)} \Psi(x_{k-j}) \leqslant -\sigma(\beta^{m^k})^2 \Psi(x_k),$$

where $m(0) = 0$, $m(k) = \min\{m(k-1) + 1, M_0\}$, and M_0 is a nonnegative integer.

4. Numerical test

In this section, we give the comparison of Modified Levenberg-Marquardt Method (I) with the algorithms in [4], [7]. We also present some numerical results of Modified Levenberg-Marquardt Method (II) for nonlinear complementarity.

Example 4.1. Consider the finitely many maximum functions systems

$$\max\{H_{11}(x_1,x_2),H_{12}(x_1,x_2)\}=0,$$
$$\max\{H_{21}(x_1,x_2),H_{22}(x_1,x_2)\}=0,$$

where

$$H_{11}=\frac{1}{2}x_1^2-x_2^2,\quad H_{12}=x_1^2,\quad H_{21}=\frac{4}{5}x_1^2,\quad H_{22}=x_1^2.$$

We get a nonsmooth equation

$$H(x)=(H_1(x),H_2(x))^\top,$$

where $H_1(x)=x_1^2$, $H_2(x)=x_1^2$, $x\in\mathbb{R}^2$. The natural merit function is

$$\Psi(x)=\frac{1}{2}H(x)^\top H(x).$$

Here we also use the differential of H proposed in [8]:

$$\partial_\star H(x)=\{(\nabla H_{1j_1},\dots,\nabla H_{nj_n})^\top:\ j_1\in J_1(x),\dots,j_n\in J_n(x)\},\quad x\in\mathbb{R}^n.$$

We use the Modified Levenberg-Marquardt Method (I) for the above finitely many maximum functions system, cf. Example 4.1. The comparison of Modified Levenberg-Marquardt Method (I) with algorithms in [7] are listed.

Results of the numbers of function evaluations and the CPU times for Example 4.1 with the initial point $x_0=(1000,0)^\top$, $\lambda_1=0.01$, $\lambda_2=1$, $\varrho=10$, $p=3$, $\beta=\frac{1}{10}$, $\varepsilon=1\mathrm{e}{-4}$ computed by the algorithm in paper [7] are listed in Tab. 4.1.

Results of the numbers of function evaluations and the CPU times for Example 4.1 with the initial point $x_0=(1000,0)^\top$, $\lambda_1=0.01$, $\lambda_2=1$, $\varrho=10$, $p=3$, $\beta=\frac{1}{10}$, $\varepsilon=1\mathrm{e}{-4}$ computed by method (I) are listed in Tab. 4.2.

Step	$H(x)$
1	$1.0e{+}005 * (2.500000, 2.500000)^\top$
2	$1.0e{+}004 * (6.250000, 6.250000)^\top$
3	$1.0e{+}004 * (1.562500, 1.562500)^\top$
4	$1.0e{+}003 * (3.906251, 3.906251)^\top$
5	$1.0e{+}002 * (9.765633, 9.765633)^\top$
6	$1.0e{+}002 * (2.441415, 2.441415)^\top$
7	$(61.035990, 61.035990)^\top$
8	$(15.259622, 15.259622)^\top$
9	$(3.815531, 3.815531)^\top$
10	$(0.954508, 0.954508)^\top$
11	$(0.239251, 0.239251)^\top$
12	$1.0e{-}003 * (0.442255, 0.442255)^\top$

Table 4.1. $x_0 = (1000, 0)^\top$. CPU time is 0.031000 seconds.

Step	$H(x)$
1	$1.0e{+}005 * (2.506246, 2.506246)^\top$
2	$1.0e{+}004 * (6.281269, 6.281269)^\top$
3	$1.0e{+}004 * (1.574241, 1.574241)^\top$
4	$1.0e{+}003 * (3.945434, 3.945434)^\top$
5	$1.0e{+}002 * (9.888230, 9.888230)^\top$
6	$1.0e{+}002 * (2.478233, 2.478233)^\top$
7	$(62.110638, 62.110638)^\top$
8	$(15.566454, 15.566454)^\top$
9	$(3.901336, 3.901336)^\top$
10	$(0.977770, 0.977770)^\top$
11	$(0.245053, 0.245053)^\top$
12	$1.0e{-}004 * (0.959363, 0.959363)^\top$

Table 4.2. $x_0 = (1000, 0)^\top$. CPU time is 0.078000 seconds.

Example 4.2. Consider the finitely many maximum functions systems

$$\begin{aligned}
\max\{H_{11}(x_1, x_2, x_3), H_{12}(x_1, x_2, x_3), H_{13}(x_1, x_2, x_3)\} = 0,\\
\max\{H_{21}(x_1, x_2, x_3), H_{22}(x_1, x_2, x_3), H_{23}(x_1, x_2, x_3)\} = 0,\\
\max\{H_{31}(x_1, x_2, x_3), H_{32}(x_1, x_2, x_3), H_{33}(x_1, x_2, x_3)\} = 0,
\end{aligned}$$

where

$$\begin{aligned}
&H_{11} = \frac{1}{2}x_1^2 - x_2^2 - 5, \quad &&H_{12} = x_1^2 - 3, \quad &&H_{13} = x_1^2 + x_2^2,\\
&H_{21} = x_1^2 + x_3^2, &&H_{22} = x_1^2, &&H_{23} = \frac{4}{5}x_1^2 - 8,\\
&H_{31} = \frac{1}{2}x_3^2, &&H_{32} = x_3^2, &&H_{33} = \frac{4}{5}x_3^2 - 8.
\end{aligned}$$

We get the nonsmooth equation

$$H(x) = (H_1(x), H_2(x), H_3(x))^\top,$$

where $H_1(x) = x_1^2 + x_2^2$, $H_2(x) = x_1^2 + x_3^2$, $H_3(x) = x_3^2$, $x \in \mathbb{R}^3$. The natural merit function is

$$\Psi(x) = \frac{1}{2} H(x)^\top H(x).$$

We give the comparison of Modified Levenberg-Marquardt Method (I) with algorithms in [4]. Results of the numbers of function evaluations for Example 4.2 with the initial point $x_0 = (1,1,1)^\top$, $\varrho = 10$, $p = 3$, $\beta = \frac{1}{10}$, $\varepsilon = 1\mathrm{e}{-4}$ computed by the algorithm in [4] and computed by Modified Levenberg-Marquardt Method (I) are listed in Tab. 4.3. We use $\|x_k - x_{k-1}\| \leqslant \varepsilon$ as the stop rule in Method (I) and the algorithm in paper [4]. The comparison of Modified Levenberg-Marquardt Method (I) with the algorithms in [4] are also in Tab. 4.3.

Algorithm	Step	$H(x)$
Algorithm in [4]	21	$(0.007933, 0.006719, 0.006716)^\top$
MLM(I)	3	$(0.003911, 0.007816, 0.003906)^\top$
$\|x_k - x_{k-1}\| \leqslant \varepsilon$ as stop rule		
Algorithm in [4]	1955	$1.0\mathrm{e}{-005} * (0.118430, 0.197182, 0.104059)^\top$
MLM(I)	143	$(0.000512, 0.001023, 0.000511)^\top$

Table 4.3. $x_0 = (1,1,1)^\top$.

Results of the numbers of function evaluations and the CPU times for Example 4.2 with the initial point $x_0 = (100000, 100000, 100000)^\top$, $\varrho = 10$, $p = 3$, $\beta = \frac{1}{10}$, $\varepsilon = 1\mathrm{e}{-4}$ computed by Modified Levenberg-Marquardt Method (I) by 28 steps gives $H(x) = (0.013596, 0.003396, 0.001831)^\top$, CPU time is 0.031000 seconds. Algorithm in [4] fails for Example 4.2.

E x a m p l e 4.3. Consider the finitely many maximum functions systems

$$\max\{H_{11}(x_1,\dots,x_8),\dots,H_{18}(x_1,\dots,x_8)\} = 0,$$
$$\max\{H_{21}(x_1,\dots,x_8),\dots,H_{28}(x_1,\dots,x_8)\} = 0,$$
$$\max\{H_{31}(x_1,\dots,x_8),\dots,H_{38}(x_1,\dots,x_8)\} = 0,$$
$$\max\{H_{41}(x_1,\dots,x_8),\dots,H_{48}(x_1,\dots,x_8)\} = 0,$$
$$\max\{H_{51}(x_1,\dots,x_8),\dots,H_{58}(x_1,\dots,x_8)\} = 0,$$
$$\max\{H_{61}(x_1,\dots,x_8),\dots,H_{66}(x_1,\dots,x_8)\} = 0,$$
$$\max\{H_{71}(x_1,\dots,x_8),\dots,H_{78}(x_1,\dots,x_8)\} = 0,$$
$$\max\{H_{81}(x_1,\dots,x_8),\dots,H_{88}(x_1,\dots,x_8)\} = 0,$$

where $H_{11} = \frac{1}{2}x_1^2 - x_2^2 - 5$, $H_{12} = x_1^2 - 3$, $H_{13} = x_1^2 - x_2^2$, $H_{14} = x_1^2$, $H_{15} = \frac{1}{2}x_1^2 - 5$, $H_{16} = x_1^2 - 9$, $H_{17} = x_1^2 - \frac{2}{3}x_2^2$, $H_{18} = x_1^2 - 6$, $H_{21} = \frac{1}{2}x_2^2 - x_7^2 - 5$, $H_{22} = x_2^2$, $H_{23} = x_2^2 - x_6^2$, $H_{24} = x_2^2 - 4$, $H_{25} = \frac{1}{2}x_2^2 - 5$, $H_{26} = x_2^2 - 9$, $H_{27} = x_2^2 - \frac{2}{3}x_8^2$, $H_{28} = x_2^2 - 6$, $H_{31} = \frac{1}{2}x_1^2 + x_3^2$, $H_{32} = x_1^2 + x_3^2$, $H_{33} = \frac{1}{2}x_1^2 + x_3^2 - 4$, $H_{34} = \frac{1}{8}x_1^2 + x_3^2$, $H_{35} = x_1^2 + \frac{1}{2}x_3^2$, $H_{36} = x_1^2 - 9$, $H_{37} = x_3^2 - \frac{2}{3}x_8^2$, $H_{38} = x_3^2 - 6$, $H_{41} = x_4^2$, $H_{42} = x_4^2 - 7$, $H_{43} = x_4^2 - x_6^2$, $H_{44} = x_4^2 - 4$, $H_{45} = \frac{1}{2}x_4^2 - 5$, $H_{46} = x_4^2 - 9$, $H_{47} = x_4^2 - \frac{2}{3}x_3^2$, $H_{48} = x_4^2 - 6$, $H_{51} = \frac{1}{2}x_1^2 + x_5^2$, $H_{52} = x_1^2 + x_5^2$, $H_{53} = \frac{1}{2}x_1^2 + x_5^2 - 4$, $H_{54} = \frac{1}{8}x_1^2 + x_5^2 - 89$, $H_{55} = x_1^2 + \frac{1}{2}x_5^2$, $H_{56} = x_5^2 - 9$, $H_{57} = x_1^2 - \frac{2}{3}x_8^2$, $H_{58} = x_5^2 - 6$, $H_{61} = x_6^2$, $H_{62} = x_6^2 - 7$, $H_{63} = x_6^2 - x_7^2$, $H_{64} = x_6^2 - 4$, $H_{65} = \frac{1}{2}x_6^2 - 5$, $H_{66} = x_6^2 - 9$, $H_{67} = x_6^2 - \frac{2}{3}x_7^2$, $H_{68} = x_6^2 - 6$, $H_{71} = x_2^2$, $H_{72} = x_2^2 - 7$, $H_{73} = x_2^2 - x_7^2$, $H_{74} = x_2^2 - 4$, $H_{75} = \frac{1}{9}x_2^2 - 5$, $H_{76} = x_2^2 - 5$, $H_{77} = x_2^2 - \frac{2}{3}x_7^2$, $H_{78} = x_2^2 - 6x_7^2$, $H_{81} = \frac{1}{6}x_8^2 + x_7^2$, $H_{82} = x_7^2 + x_8^2$, $H_{83} = \frac{1}{2}x_7^2 + x_8^2 - 4$, $H_{84} = \frac{1}{8}x_7^2 + x_8^2 - 9$, $H_{85} = x_7^2 + \frac{1}{2}x_8^2 - 3$, $H_{86} = x_7^2 - 9$, $H_{87} = x_8^2 - \frac{2}{3}x_3^2$, $H_{88} = x_8^2 - 1$. We get the nonsmooth equation

$$H(x) = (H_1(x), H_2(x), H_3(x), H_4(x), H_5(x), H_6(x), H_7(x), H_8(x))^\top,$$

where $H_1(x) = x_1^2$, $H_2(x) = x_2^2$, $H_3(x) = x_1^2 + x_3^2$, $H_4(x) = x_4^2$, $H_5(x) = x_1^2 + x_5^2$, $H_6(x) = x_6^2$, $H_7(x) = x_2^2$, $H_8(x) = x_7^2 + x_8^2$, $x \in \mathbb{R}^8$. The natural merit function is

$$\Psi(x) = \frac{1}{2}H(x)^\top H(x).$$

Results of the numbers of function evaluations and the CPU times for Example 4.3 with the initial point $x_0 = (10000, 10000, 10000, 10000, 10000, 10000, 10000, 10000)^\top$, $\varrho = 10$, $p = 3$, $\beta = \frac{1}{10}$, $\varepsilon = 1\text{e}{-4}$ computed by Modified Levenberg-Marquardt Method (I) are listed in Tab. 4.4. CPU time is 0.047000 seconds.

Step	$H(x)$
48	$1.0\text{e}{-004} * (1.4976, 59.1713, 37.0532, 68.6733, 37.0532, 68.6733, 59.1714, 9.5633)^\top$

Table 4.4. $x_0 = (10000, 10000, 10000, 10000, 10000, 10000, 10000, 10000)^\top$.

Results are shown for Example 4.3 with the initial point

$$x_0 = (100000, 100000, 100000, 100000, 100000, 100000, 100000, 100000)^\top.$$

We get by 54 steps that

$$H(x) \\ = (0.001448, 0.007901, 0.001774, 0.003845, 0.001774, 0.003845, 0.007901, 0.005789)^\top$$

and CPU time is 0.062000 seconds.

E x a m p l e 4.4. Similarly to [7], we consider the following nonlinear complementarity problem. The function in (1.2) is of the form

$$f(x) = (x_1^2 + 1 + x_3, x_1^2 + x_2 + 3, x_3 - 2)^\top.$$

Computed by Modified Levenberg-Marquardt Method (II), results of the numbers of iteration and the CPU times for Example 4.4 with different starting points are given in Tab. 4.5.

Starting points	Number of iterations	x_k	CPU times
$(0.1, 0.1, 1.5)^\top$	3	$(0.099922, 0.099949, 2.264911)^\top$	0.031000 seconds
$(0.1, 0.1, 1.8)^\top$	3	$(0.099937, 0.099949, 2.034386)^\top$	0.031000 seconds

Table 4.5. Numerical results for Example 4.4.

Conclusion

The numerical results of Modified Levenberg-Marquardt Method (I) and Modified Levenberg-Marquardt Method (II) for the above examples indicate that the algorithms work quite well in practice, which is a typical feature of Newton-type methods. And the algorithms are fairly robust and capable of finding a solution to the above examples with a limited amount of steps. Furthermore, in all cases the global convergence is observed. The assumptions necessary to establish the global convergence of the algorithms are usually met in practice. Supposing in Theorem 3.1 that there exist constants $M > 0$, $\|\mathrm{diag}(\lambda_i^{(k)} H_i(x))\| \leqslant M < \infty$, we can let $\lambda_1 = 0.01$, $\lambda_2 = 1$ in the computation of Example 4.1. We also can let $\lambda_1 = 0.1$, $\lambda_2 = 1$, by 18 steps $H(x) = (0.009616, 0.009616)^\top$, let $\lambda_1 = 1$, $\lambda_2 = 1$, by 14 steps $H(x) = (0.001385, 0.001385)^\top$, but if we let $\lambda_1 = 10$, $\lambda_2 = 1$, by 40 steps $H(x) = (0.009379, 0.009379)^\top$, if we let $\lambda_1 = 100$, $\lambda_2 = 1$, the algorithm fails for the example. The same situation occurs in the computation of Example 4.2. We also can let $\lambda_1 = 0.001$, $\lambda_2 = 0.001$, $\lambda_3 = 0.001$, by 29 steps $H(x) = (0.012876, 0.003265, 0.001740)^\top$, let $\lambda_1 = 0.01$, $\lambda_2 = 0.01$, $\lambda_3 = 0.01$, by 29 steps $H(x) = (0.012949, 0.003463, 0.001198)^\top$, but if we let $\lambda_1 = 10$, $\lambda_2 = 10$, $\lambda_3 = 10$, by 45 steps $H(x) = (0.009895, 0.009989, 0.000097)^\top$, if we let $\lambda_1 = 1000$, $\lambda_2 = 1000$, $\lambda_3 = 1000$, by 2174 steps $H(x) = (0.013207, 0.003102, 0.000001)^\top$, the algorithm almost fails for the example.

Acknowledgement. The authors are grateful to the anonymous referee for valuable suggestions and comments.

References

[1] *D. P. Bertsekas*: Constrained Optimization and Lagrange Multiplier Methods. Academic Press, New York, 1982.

[2] *X. Chen, L. Qi*: A parameterized Newton method and a quasi-Newton method for nonsmooth equations. Comput. Optim. Appl. *3* (1994), 157–179.

[3] *R. W. Cottle, J.-S. Pang, R. E. Stone*: The Linear Complementarity Problem. Academic Press, Boston, 1992.

[4] *S.-Q. Du, Y. Gao*: A modified Levenberg-Marquardt method for nonsmooth equations with finitely many maximum functions. Math. Prob. Eng. (2008).

[5] *A. Fischer*: A special Newton-type optimization method. Optim. *24* (1992), 269–284.

[6] *A. Fischer, V. Jeyakumar, D. T. Luc*: Solution point characterizations and convergence analysis of a descent algorithm for nonsmooth continuous complementarity problems. J. Optimization Theory Appl. *110* (2001), 493–513.

[7] *F. Facchinei, F. Kanzow*: A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems. Math. Program. *76* (1997), 493–512.

[8] *Y. Gao*: Newton methods for solving two classes of nonsmooth equations. Appl. Math. *46* (2001), 215–229.

[9] *C. Geiger, C. Kanzow*: On the resolution of monotone complementarity problems. Comput. Optim. Appl. *5* (1996), 155–173.

[10] *H. Jiang*: Unconstrained minimization approaches to nonlinear complementarity problems. J. Glob. Optim. *9* (1996), 169–181.

[11] *J. J. Moré, D. C. Sorensen*: Computing a trust region step. SIAM J. Sci. Stat. Comput. *4* (1983), 553–572.

[12] *L. Qi, P. Tseng*: On almost smooth functions and piecewise smooth functions. Nonlinear Anal. *67* (2007), 773–794.

[13] *L. Qi, J. Sun*: A nonsmooth version of Newton's method. Math. Program. *58* (1993), 353–367.

[14] *D. Sun, L. Qi*: On NCP-functions. Comput. Optim. Appl. *13* (1999), 201–220.

[15] *N. Yamashita, M. Fukushima*: Modified Newton methods for solving a semismooth reformulation of monotone complementarity problems. Math. Program. *76* (1997), 469–491.

Authors' addresses: *Shou-Qiang Du*, College of Mathematics, Qingdao University, Qingdao, 266071, P. R. China, e-mail: dsq8933@163.com; *Y. Gao*, School of Management, University of Shanghai for Science and Technology, Jungong Road 516, Shanghai, 200093, P. R. China, e-mail: gaoyan@usst.edu.cn.