Yan Gao

Newton methods for solving two classes of nonsmooth equations

Applications of Mathematics, Vol. 46 (2001), No. 3, 215--229

Persistent URL: http://dml.cz/dmlcz/134465

NEWTON METHODS FOR SOLVING TWO CLASSES OF NONSMOOTH EQUATIONS*

Yan Gao, Beijing

(Received July 27, 1999)

Abstract. The paper is devoted to two systems of nonsmooth equations. One is the system of equations of max-type functions and the other is the system of equations of smooth compositions of max-type functions. The Newton and approximate Newton methods for these two systems are proposed. The Q-superlinear convergence of the Newton methods and the Q-linear convergence of the approximate Newton methods are established. The present methods can be more easily implemented than the previous ones, since they do not require an element of Clarke generalized Jacobian, of B-differential, or of b-differential, at each iteration point.

Keywords: nonsmooth equations, Newton method, approximate Newton method, max-type function, composite function, convergence

MSC 2000: 65H10, 90C30

1. Introduction

In recent years, much attention has been devoted to various forms of Newton methods and approximate Newton methods for solving nonsmooth equations, see for instance [2], [8], [9], [10]. Newton methods for solving nonsmooth equations

$$H(x) = 0, \tag{1.1}$$

where $H\colon \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitzian and semismooth, are given as follows:

$$x^{k+1} = x^k - V_k^{-1} H(x^k), \tag{1.2}$$

* This paper has been supported by the Foundation of the Chinese Ministry of Machine-Building Industry.

where V_k was assumed to be an element of Clarke generalized Jacobian in [9], an element of B-differential in [8], and an element of b-differential in [10] of H at x^k, respectively. Under the assumption that all elements of $\partial_{\rm Cl} H(x^*)$, of $\partial_B H(x^*)$, or of $\partial_b H(x^*)$, respectively, where x^* is the solution of the equations (1.2), are nonsingular, the locally superlinear convergence properties were obtained. The definitions of the Clarke generalized Jacobian, B-differential and b-differential for a locally Lipschitzian function are as follows:

Let $H\colon \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitzian and let D_H denote the set where H is differentiable. Then

$$\partial_B H(x) = \{\lim_{x_n \to x} JH(x_n) \mid x_n \to x, x_n \in D_H\}, \quad x \in \mathbb{R}^n,$$

where $JH(x_n)$ denotes the Jacobian of H at x_n, is called the B-differential of H at x_n, see [8];

$$\partial_{\rm Cl} H(x) = \operatorname{conv} \partial_B H(x), \quad x \in \mathbb{R}^n,$$

where conv stands for the convex hull, is called the Clarke generalized Jacobian of H at x, in particular, in the case $m = 1$, $\partial_{\rm Cl} H(x)$ reduces to the Clarke generalized gradient of H at x, due to [3]; by the definition in [10],

$$\partial_b H(x) = \partial_B h_1(x) \times \ldots \times \partial_B h_m(x), \quad x \in \mathbb{R}^n,$$

where $h_i(x)$ is the i-th component of $H(x)$, is called the b-differential of H at x, in particular, if $m = 1$, $\partial_b H(x) = \partial_B H(x)$.

Any Newton method, given in (1.2), is implemented under the assumption that at least one element for the related differential of H, at each iteration point, can be calculated. Actually, the calculation of an element for each kind of differentials mentioned above is quite time consuming in some applications. In the present paper we consider two systems of nonsmooth equations, one is the system of equations of max-type functions, the other is the system of equations of smooth compositions of max-type functions. New Newton methods and approximate Newton methods for solving these two systems of nonsmooth equations are proposed. Since they do not require an element of the Clarke generalized Jacobian, of B-differential, or of b-differential at each iteration point, our methods can be implemented more easily. These two systems of nonsmooth equations are of concrete background, for instance, complementarity problems, variational inequality problems, Karush-Kuhn-Tucker (KKT) systems of nonlinear programs and many problems in mechanics and engineering lead to these two systems of nonsmooth equations.

The present paper is organized as follows: In the remainder of this section, some preliminaries on nonsmooth equations are presented. In the next section, the Newton method with Q-superlinear convergence and the approximate Newton method with Q-linear convergence for the equations of max-type functions are developed. In Section 3, the results of Section 2 are extended to the equations of smooth compositions of max-type functions.

By the definition in [9], a locally Lipschitzian function $H\colon \mathbb{R}^n \to \mathbb{R}^m$ is said to be semismooth at x provided that

$$\lim_{\substack{V\in\partial_{\mathrm{Cl}}H(x+th')\\ h'\to h,\, t\to 0^+}} Vh' \tag{1.3}$$

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

The semismoothness was originally introduced for functionals by Mifflin [6] and was immediately shown to be very important in nonsmooth optimization algorithms. Qi and Sun [9] extended the concept of semismoothness to vector-valued functions for the research of nonsmooth equations. As is well known, many existing methods for solving nonsmooth equations have been developed under the assumption of semismoothness. The semismooth functions lie between locally Lipschitzian functions and C^1 functions. Smooth functions, convex functions, maximums of a finite number of smooth functions and piecewise smooth functions are semismooth. Moreover, smooth compositions of semismooth functions are still semismooth. Semismooth functions have many important properties. Some of them are very useful in the convergence analysis of methods for solving nonsmooth equations. We now list two related properties of semismooth functions, which are borrowed from [9].

If H is semismooth, then the directional derivative

$$H'(x;h) = \lim_{t\to 0^+} [H(x+th) - H(x)]/t, \quad h \in \mathbb{R}^n$$

exists and equals to (1.3), i.e.,

$$H'(x;h) = \lim_{\substack{V\in\partial_{\mathrm{Cl}}H(x+th')\\ h'\to h,\, t\to 0^+}} Vh', \quad h \in \mathbb{R}^n.$$

Lemma 1.1. *Suppose that $H\colon \mathbb{R}^n \to \mathbb{R}^m$ is locally Lipschitzian and semismooth at x. Then*

(1) $Vh - H'(x;h) = o(\|h\|)\ \forall V \in \partial_{\mathrm{Cl}}H(x+h),\ h \in \mathbb{R}^n$;

(2) $H(x+h) - H(x) - H'(x;h) = o(\|h\|), \quad h \in \mathbb{R}^n$.

Actually, the relation (1) is equivalent to the semismoothness of H at x, see Theorem 2.3 of [9]. Of course, the relations (1) and (2) could be rewritten respectively as

$$(1.4)\quad V(x-x^*) - H'(x^*; x-x^*) = o(\|x-x^*\|) \ \forall V \in \partial_{\mathrm{Cl}} H(x),\ \ x, x^* \in \mathbb{R}^n$$

and

$$(1.5)\qquad H(x) - H(x^*) - H'(x^*; x-x^*) = o(\|x-x^*\|),\ \ x, x^* \in \mathbb{R}^n.$$

The combination of (1.4) and (1.5) leads to

$$(1.6)\qquad H(x) - H(x^*) - V(x-x^*) = o(\|x-x^*\|) \ \forall V \in \partial_{\mathrm{Cl}} H(x),\ \ x, x^* \in \mathbb{R}^n.$$

The terminology of the convergence rate, which is used in this paper, refers to the following: a sequence $\{x^k\}$ is said to be Q-linear convergent to x^* if there exist $k_0 > 0$ and $0 < \alpha < 1$ such that

$$\|x^{k+1} - x^*\| \leqslant \alpha \|x^k - x^*\| \ \ \forall k \geqslant k_0,$$

and Q-superlinear convergent to x^* if

$$\lim_{k\to\infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = 0.$$

2. Equations of max-type Functions

Consider equations of max-type functions

$$(2.1)\qquad \begin{aligned} \max_{j\in J_1} f_{1j}(x) &= 0,\\ &\ \vdots\\ \max_{j\in J_n} f_{nj}(x) &= 0 \end{aligned}$$

where $f_{ij}\colon \mathbb{R}^n \to \mathbb{R}$ for $j \in J_i$, $i = 1, \ldots, n$, are continuously differentiable, J_i for $i = 1, \ldots, n$ are finite index sets. Obviously, (2.1) is a system of semismooth equations. In this section, we will propose a Newton method and an approximate Newton method for solving this system of nonsmooth equations.

Throughout this section, we denote

$$(2.2)\qquad f_i(x) = \max_{j\in J_i} f_{ij}(x),\ \ x \in \mathbb{R}^n,\ \ i = 1, \ldots, n,$$

$$(2.3)\qquad F(x) = \big(f_1(x), \ldots, f_n(x)\big)^T,\ \ x \in \mathbb{R}^n,$$

$$(2.4)\qquad J_i(x) = \{j \in J_i \mid f_{ij}(x) = f_i(x)\},\quad x \in \mathbb{R}^n,\ \ i = 1, \ldots, n.$$

Thus, the equations (2.1) can be briefly rewritten as

$$F(x) = 0. \tag{2.5}$$

Now, we define a new kind of the differential for the function F denoted by $\partial_* F(x)$ as follows:

$$\partial_* F(x) = \{(\nabla f_{1j_1}(x), \dots, \nabla f_{nj_n}(x))^T \mid j_1 \in J_1(x), \dots, j_n \in J_n(x)\}, \quad x \in \mathbb{R}^n. \tag{2.6}$$

Evidently, $\partial_* F(x)$ is a finite set of points in $\mathbb{R}^{n\times n}$ and can be calculated by determining the index sets $J_i(x)$, $i = 1, \dots, n$ and evaluating the gradients $\nabla f_{ij_i}(x)$, $j_i \in J_i(x)$, $i = 1, \dots, n$. In what follows, we take the differential $\partial_* F(x)$ as a tool instead of the Clarke generalized Jacobian, B-differential and b-differential.

We now present the Newton method for solving the equations (2.1) as follows:

$$x^{k+1} = x^k - V_k^{-1} F(x^k), \quad V_k \in \partial_* F(x^k). \tag{2.7}$$

Compared with other Newton methods, for instance in [8], [9] or [10], for solving the nonsmooth equations (2.1), the present one can be executed more easily. Applying Newton methods in [8] or [9] to solving the equations (2.1), an element of Clarke generalized Jacobian, or of B-differential for F, at each iteration point, has to be obtained. Actually, in the case when the cardinalities $\operatorname{card} J_i(x)$ for $i = 1, \dots, n$ are quite large, the calculation of an element of Clarke generalized Jacobian, or of B-differential for F at a point, is expensive. Besides, executing the Newton method from [10] for solving the equations (2.1), the calculation of an element for $\partial_b F(x)$ at each iteration point, which can be transformed into the calculation of an element of B-differential for each f_i at an iteration point, is required. In [5], the formula for the B-differential of f_i at x is given as follows:

$$\partial_B f_i(x) = \{\nabla f_{ij}(x) \mid \nabla f_{ij}(x) \quad \text{is an extreme point of} \quad \partial_{\mathrm{Cl}} f_i(x),\ j \in J_i(x)\}, \quad i = 1, \dots, n. \tag{2.8}$$

By virtue of [3], the Clarke generalized gradient $\partial_{\mathrm{Cl}} f_i(x)$ of f_i at x has the form

$$\partial_{\mathrm{Cl}} f_i(x) = \operatorname{conv}\{\nabla f_{ij}(x) \mid j \in J_i(x)\}, \quad i = 1, \dots, n. \tag{2.9}$$

Given a fixed point $x \in \mathbb{R}^n$ as well as indices i and $j_i \in J_i(x)$, we construct the linear system as follows:

$$\nabla f_{ij_i}(x) = \sum_{j\in J_i(x)\setminus\{j_i\}} \lambda_j \nabla f_{ij}(x), \qquad \sum_{j\in J_i(x)\setminus\{j_i\}} \lambda_j = 1, \ \lambda_j \geqslant 0, \ j \in J_i(x) \setminus \{j_i\}. \tag{L_{ij_i}}$$

Formulas (2.8) and (2.9) as well as the system (L_{ij_i}) show us that determining whether the point $\nabla f_{ij_i}(x)$ is an element of $\partial_B f_i(x)$ can be transformed into checking the consistency of the system (L_{ij_i}). If (L_{ij_i}) is inconsistent, since only in that case $\nabla f_{ij_i}(x)$ is an extreme point of $\partial_{\mathrm{Cl}} f_i(x)$, then $\nabla f_{ij_i}(x) \in \partial_B f(x)$; otherwise, $\nabla f_{ij_i}(x) \notin \partial_B f(x)$.

Next we give the approximate Newton method for solving the nonsmooth equations (2.1) as follows:

$$x^{k+1} = x^k - U_k^{-1} F(x^k), \ \ U_k \in \mathbb{R}^{n\times n}, \tag{2.10}$$

where U_k is an approximation to V_k to be discussed later. Before the convergence theorem, we present a lemma.

Lemma 2.1. *Suppose that $F(x)$ and $\partial_* F(x)$ are defined by* (2.3) *and by* (2.6)*, respectively, and all $V \in \partial_* F(x)$ are nonsingular. Then there exists a scalar $\beta > 0$ such that*

$$\|V^{-1}\| \leqslant \beta \ \ \forall V \in \partial_* F(x). \tag{2.11}$$

Furthermore, there exists a neighbourhood $N(x)$ of x such that

$$\|V^{-1}\| \leqslant \frac{10}{9}\beta \ \ \forall V \in \partial_* F(y), \ \ y \in N(x). \tag{2.12}$$

P r o o f. From the fact that $\partial_* F(x)$ is a finite set of points, it follows that (2.11) holds. Since each f_{ij} is continuous, there exists a neighbourhood $N_1(x)$ of x such that $J_i(y) \subset J_i(x)$ $\forall y \in N_1(x)$ for $i = 1, \ldots, n$. Thus, the continuity of each ∇f_{ij} and (2.11) imply that there exists a neighbourhood $N(x)$ of x such that (2.12) holds. □

The next theorem gives both the Q-linear convergence of the approximate Newton method and the Q-superlinear convergence of the Newton method.

Theorem 2.1. *Suppose that x^* is a solution of the equations* (2.1) *and all $V \in \partial_* F(x^*)$ are nonsingular. Then there exist scalars $\varepsilon > 0$ and $\Delta > 0$ such that if $\|x^0 - x^*\| \leqslant \varepsilon$ and*

$$\|V_k - U_k\| \leqslant \Delta \tag{2.13}$$

for some $V_k \in \partial_* F(x^k)$, then the iteration method given by (2.10) *is well defined and the sequence $\{x^k\}$ converges to x^* Q-linearly. Moreover, if $\lim\limits_{k\to\infty} \|V_k - U_k\| = 0$, then $\{x^k\}$ converges to x^* Q-superlinearly.*

P r o o f. By virtue of Lemma 2.1, there exist a scalar $\beta > 0$ and a neighbourhood $N(x^*)$ of x^* such that

$$\|V^{-1}\| \leqslant \beta \;\; \forall V \in \partial_* F(x^*), \tag{2.14}$$

$$\|V^{-1}\| \leqslant \frac{10}{9}\beta \;\; \forall V \in \partial_* F(y), \;\; y \in N(x^*). \tag{2.15}$$

Choose a scalar $\Delta > 0$ satisfying

$$6\beta\Delta \leqslant 1. \tag{2.16}$$

Denote the i-th row of $V \in \partial_* F(x)$ by V^i. Note that f_i, $i = 1, \ldots, n$, given in (2.2) are semismooth. It follows from (1.6) that

$$f_i(x) - f_i(x^*) - V^i(x - x^*) = o(\|x - x^*\|), \quad i = 1, \ldots, n.$$

Therefore, one has

$$F(x) - F(x^*) - V(x - x^*) = o(\|x - x^*\|) \;\; \forall V \in \partial F(x).$$

Thus, we can choose a positive number ε small enough such that

$$\{x \in \mathbb{R}^n \mid \|x - x^*\| \leqslant \varepsilon\} \subset N(x^*)$$

and

$$\|F(x) - F(x^*) - V(x - x^*)\| \leqslant \Delta\|x - x^*\| \;\; \forall V \in \partial_* F(x), \;\; \text{if } \|x - x^*\| \leqslant \varepsilon. \tag{2.17}$$

It follows that, if $\|x - x^*\| \leqslant \varepsilon$, then all $V \in \partial_* F(x)$ are nonsingular, moreover,

$$\|V^{-1}\| \leqslant \frac{10}{9}\beta \;\; \forall V \in \partial_* F(x), \;\; x \in \{x \in \mathbb{R}^n \mid \|x - x^*\| \leqslant \varepsilon\}. \tag{2.18}$$

Recall Theorem 2.3.2 of [7], which asserts that if $A, B \in \mathbb{R}^{n\times n}$ and B is nonsingular, then

$$\|A^{-1}\| \leqslant \frac{\|B^{-1}\|}{1 - \|B^{-1}(B - A)\|}. \tag{2.19}$$

In what follows, we suppose that $\|x^k - x^*\| \leqslant \varepsilon$ for all k. Setting $A = U_k$ and $B = V_k$ in (2.19), one has

$$\|U_k^{-1}\| \leqslant \frac{\|V_k^{-1}\|}{1 - \|V_k^{-1}(V_k - U_k)\|} \leqslant \frac{\frac{10}{9}\beta}{1 - \frac{10}{9}\beta\Delta} \leqslant \frac{\frac{10}{9}\beta}{1 - \frac{5}{27}} \leqslant \frac{3}{2}\beta. \tag{2.20}$$

We deduce that

$$\begin{aligned}
\|x^{k+1} - x^*\| &= \|x^k - U_k^{-1}F(x^k) - x^*\| \\
&\leqslant \|U_k^{-1}\| \, \|F(x^k) - F(x^*) - U_k(x^k - x^*)\| \\
&\leqslant \|U_k^{-1}\| \, \big[\|F(x^k) - F(x^*) - V_k(x^k - x^*)\| \\
&\quad + \|V_k - U_k\| \, \|x^k - x^*\|\big].
\end{aligned} \tag{2.21}$$

Substituting (2.13), (2.17) and (2.20) into the above formula, one has

$$\begin{aligned}
\|x^{k+1} - x^*\| &\leqslant \frac{3}{2}\beta[\Delta\|x^k - x^*\| + \Delta\|x^k - x^*\|] \\
&= 3\beta\Delta\|x^k - x^*\| \\
&\leqslant \frac{1}{2}\|x^k - x^*\|.
\end{aligned} \tag{2.22}$$

By mathematical induction, the relation $\|x^0 - x^*\| \leqslant \varepsilon$ and (2.22) mean that $\|x^k - x^*\| \leqslant \varepsilon$ holds for all positive integers k. So, $\|x^{k+1} - x^*\| \leqslant \frac{1}{2}\|x^k - x^*\|$ is valid for all k under the assumption that $\|x^0 - x^*\| \leqslant \varepsilon$. We now conclude that the iterative method (2.10) is well defined for all k because of (2.20), and the sequence $\{x^k\}$ converges to x^* Q-linearly.

We next prove that the sequence $\{x^k\}$ converges to x^* Q-superlinearly under the hypothesis $\lim\limits_{k\to\infty} \|V_k - U_k\| = 0$. By virtue of (1.6), (2.20), (2.21) and $\lim\limits_{k\to\infty} \|V_k - U_k\| = 0$, we obtain that

$$\begin{aligned}
\|x^{k+1} - x^*\| &\leqslant \|U_k^{-1}\| \, [\|F(x^k) - F(x^*) - V_k(x^k - x^*)\| + \|V_k - U_k\| \, \|x^k - x^*\|] \\
&\leqslant \frac{3}{2}\beta[\|F(x^k) - F(x^*) - V_k(x^k - x^*)\| + o(\|x^k - x^*\|)] \\
&= \frac{3}{2}\beta[o(\|x^k - x^*\|) + o(\|x^k - x^*\|)] \\
&= o(\|x^k - x^*\|).
\end{aligned}$$

This proves the Q-superlinear convergence of $\{x^k\}$ to x^*. Thus, we have completed the proof of the theorem. $\square$

In this section we deal with the equations of smooth compositions of max-type functions, which have the form

$$(3.1)\qquad \begin{aligned} g_1(\max_{j\in J_1} f_{1j}(x),\dots,\max_{j\in J_m} f_{mj}(x)) &= 0,\\ &\vdots\\ g_n(\max_{j\in J_1} f_{1j}(x),\dots,\max_{j\in J_m} f_{mj}(x)) &= 0 \end{aligned}$$

where $f_{ij}\colon\ \mathbb{R}^n \to \mathbb{R}$ for $j \in J_i,\ i = 1,\dots,m$ and $g_s\colon\ \mathbb{R}^m \to \mathbb{R}$ for $s = 1,\dots,n$ are continuously differentiable, J_i for $i = 1,\dots,m$ are finite index sets. The results of the above section will be extended to the equations (3.1).

Throughout this section, we denote

$$(3.2)\qquad f_i(x) = \max_{j\in J_i} f_{ij}(x),\quad x\in\mathbb{R}^n,\ i=1,\dots,m,$$

$$(3.3)\qquad F(x) = (f_1(x),\dots,f_m(x))^T,\quad x\in\mathbb{R}^n,$$

$$(3.4)\qquad J_i(x) = \{j\in J_i \mid f_{ij}(x) = f_i(x)\},\quad x\in\mathbb{R}^n,\ i=1,\dots,m,$$

$$(3.5)\qquad G(y) = (g_1(y),\dots,g_n(y))^T,\quad y\in\mathbb{R}^m,$$

$$(3.6)\qquad \partial_* F(x) = \{(\nabla f_{1j_1}(x),\dots,\nabla f_{mj_m}(x))^T \mid j_1\in J_1(x),\dots,j_m\in J_m(x)\},\quad x\in\mathbb{R}^n.$$

Thus, the equations (3.1) can be rewritten as

$$(3.7)\qquad G(F(x)) = 0.$$

Since F is semismooth and G is smooth, $G(F(x))$ is semismooth.

Now we present the Newton method for solving the equations (3.1) as

$$(3.8)\qquad x^{k+1} = x^k - W_k^{-1} G(F(x^k)),\quad W_k \in JG(F)\,|_{F=F(x^k)}\,\partial_* F(x^k).$$

It should be mentioned that, in general, W_k is no longer guaranteed to be an element of Clarke generalized Jacobian of $G(F(x))$ at $x = x^k$, and the same holds for the set

$$\partial_{\mathrm{Cl}} g_1(F(x^k)) \times \ldots \times \partial_{\mathrm{Cl}} g_n(F(x^k)).$$

This follows from the fact that the relation

$$\partial_{\mathrm{Cl}} G(F(x)) \subset JG(F)\,|_{F=F(x)}\,\partial_{\mathrm{Cl}} F(x),\quad x\in\mathbb{R}^n$$

is only an inclusion and need not turn into an equality. On the other hand, it is not hard to verify that

$$\partial_B G(F(x)) \subset JG(F)|_{F=F(x)}\partial_* F(x), \quad x \in \mathbb{R}^n$$

and

$$\partial_b G(F(x)) \subset JG(F)|_{F=F(x)}\partial_* F(x), \quad x \in \mathbb{R}^n.$$

In view of the two above formulas, the procedure given in (3.8) could be regarded as an extension of those in [9] and [10] for solving the equations (3.1).

There exist various problems in mechanics and engineering whose mathematical models are to solve the system of the nonsmooth equations of the form

$$\max_{i\in I} F_i(x) - \max_{j\in J} G_j(x) + H(x) = 0, \tag{3.9}$$

where $F_i, G_j, H\colon \mathbb{R}^n \to \mathbb{R}^n$ are continuously differentiable, I and J are finite index sets, the max operator denotes the componentwise maximum, see [1], [4]. The left-hand side of (3.9) being a difference (not sum) of two vector-valued max-type functions, the calculation of an element of B-differential or of Clarke generalized gradient for each of its components can not be implemented easily, let alone its Clarke generalized Jacobian or B-differential. Hence, it is expensive to solve the system (3.9) by using the Newton methods proposed in [8], [9], or [10]. Indeed, it is convenient to apply our Newton method presented in (3.8) to solve the system (3.9).

Example 3.1. Consider the nonsmooth equations

$$\max_{j\in J_1} f_{1j}(x) - \max_{j\in J_2} f_{2j}(x) = 0, \tag{3.10a}$$

$$\max_{j\in J_3} f_{3j}(x) \max_{j\in J_4} f_{4j}(x) = 0 \tag{3.10b}$$

where $f_{ij}\colon \mathbb{R}^2 \to \mathbb{R}^2$ for $j \in J_i$, $i = 1,\dots,4$ are continuously differentiable, J_i for $i = 1,\dots,4$ are finite index sets. We next apply the Newton method given in (3.8) to the equations (3.10a), (3.10b). Let

$$f_i(x) = \max_{j\in J_i} f_{ij}(x), \quad x \in \mathbb{R}^2,\ i = 1,\dots,4,$$

$$F(x) = \big(f_1(x), f_2(x), f_3(x), f_4(x)\big)^T, \quad x \in \mathbb{R}^2,$$

$$g_1(f_1, f_2, f_3, f_4) = f_1 - f_2,$$

$$g_2(f_1, f_2, f_3, f_4) = f_3 f_4,$$

$$G = (g_1, g_2)^T,$$

$$J_i(x) = \{j \in J_i \mid f_{ij}(x) = f_i(x)\}, \quad x \in \mathbb{R}^2,\ i = 1,\dots,4.$$

It follows that

$$JG(F)\big|_{F=F(x)} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & f_4(x) & f_3(x) \end{pmatrix}.$$

Choosing

$$V_k = (\nabla f_{1j_1}(x^k), \nabla f_{2j_2}(x^k), \nabla f_{3j_3}(x^k), \nabla f_{4j_4}(x^k)^T) \in \partial_* F(x^k),$$

where $j_1 \in J_1(x^k),\ j_2 \in J_2(x^k),\ j_3 \in J_3(x^k),\ j_4 \in J_4(x^k)$, one has

$$\begin{aligned} W_k &= JG(F)\big|_{F=F(x^k)} V_k \\ &= \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & f_4(x^k) & f_3(x^k) \end{pmatrix} \begin{pmatrix} \nabla f_{1j_1}(x^k)^T \\ \nabla f_{2j_2}(x^k)^T \\ \nabla f_{3j_3}(x^k)^T \\ \nabla f_{4j_4}(x^k)^T \end{pmatrix} \\ &= \begin{pmatrix} \nabla f_{1j_1}(x^k)^T - \nabla f_{2j_2}(x^k)^T \\ f_4(x^k)\nabla f_{3j_3}(x^k)^T + f_3(x^k)\nabla f_{4j_4}(x^k)^T \end{pmatrix}. \end{aligned}$$

Thus, the iteration procedure of the Newton method for the equations (2.12a), (2.12b) has the form

$$x^{k+1} = x^k - \begin{pmatrix} f_{1j_1}(x^k)^T - \nabla f_{2j_2}(x^k)^T \\ f_4(x^k)\nabla f_{3j_3}(x^k)^T + f_3(x^k)\nabla f_{4j_4}(x^k)^T \end{pmatrix}^{-1} \begin{pmatrix} f_1(x^k) - f_2(x^k) \\ f_3(x^k) f_4(x^k) \end{pmatrix}.$$

We next consider the approximate Newton method for solving the nonsmooth equations (3.1) in the form

$$x^{k+1} = x^k - W_k^{-1} G(F(x^k)), W_k = JG(F)\big|_{F=F(x^k)} U_k, \quad U_k \in \mathbb{R}^{m\times n}, \tag{3.11}$$

where U_k is an approximation to V_k, see (3.14).

Lemma 3.1. *Suppose that $F(x)$ and $\partial_* F(x)$ are defined by (3.3) and (3.6), respectively, and all $V \in JG(F)\big|_{F=F(x)} \partial_* F(x)$, where $JG(F(x))$ denotes the Jacobian of $G(F)$ at $F = F(x)$, are nonsingular. Then there exists a scalar $\beta > 0$ such that*

$$\|W^{-1}\| \leqslant \beta \ \ \forall W \in JG(F)\big|_{F=F(x)} \partial_* F(x). \tag{3.12}$$

Furthermore, there exists a neighbourhood $N(x)$ of x such that

$$\|W^{-1}\| \leqslant \frac{10}{9}\beta \ \ \forall W \in JG(F)\big|_{F=F(y)} \partial_* F(y), \ \ y \in N(x). \tag{3.13}$$

P r o o f. Note that G is continuously differentiable. The proof of the lemma is similar to that of Lemma 2.1. □

Theorem 3.1. *Suppose that x^* is a solution of the equations* (3.1) *and all $W \in JG(F)\big|_{F=F(x^*)}\partial_* F(x^*)$ are nonsingular. Then there exist scalars $\varepsilon > 0$ and $\Delta > 0$ such that if $\|x^0 - x^*\| \leqslant \varepsilon$ and*

$$\|V_k - U_k\| \leqslant \Delta \tag{3.14}$$

for some $V_k \in \partial_* F(x^k)$, then the iteration method given by (3.11) *is well defined and the sequence $\{x^k\}$ converges to x^* Q-linearly. Moreover, if $\lim\limits_{k\to\infty} \|V_k - U_k\| = 0$, then $\{x^k\}$ converges to x^* Q-superlinearly.*

P r o o f. By virtue of Lemma 3.1 and the continuity of $JG(F)\big|_{F=F(x)}$, there exist scalars $\beta > 0$, $\gamma > 1$ and a neighbourhood $N(x^*)$ of x^* such that

$$\|W^{-1}\| \leqslant \beta \ \ \forall W \in JG(F)\big|_{F=F(x^*)}\partial_* F(x^*), \tag{3.15}$$

$$\|W^{-1}\| \leqslant \frac{10}{9}\beta \ \ \forall W \in JG(F)\big|_{F=F(y)}V, \quad V \in \partial_* F(y), \quad y \in N(x^*), \tag{3.16}$$

$$\|JG(F)\big|_{F=F(y)}\| \leqslant \gamma \ \ \forall y \in N(x^*). \tag{3.17}$$

Choose a scalar $\Delta > 0$ satisfying

$$6\beta\gamma\Delta \leqslant 1. \tag{3.18}$$

Denote the $s-$th row of $JG(F)\big|_{F=F(x)}V$, where $V \in \partial_* F(x)$, by

$$(JG(F)\big|_{F=F(x)}V)_s.$$

Then it has the form

$$(JG(F)\big|_{F=F(x)}V)_s = \sum_{i=1}^{m} \frac{\partial g_s(F)}{\partial f_i}\Big|_{F=F(x)} \nabla f_{ij_i}(x)^T, \tag{3.19}$$

where $j_i \in J_i(x)$, $i = 1\ldots, m$. Recall that a vector-valued function is semismooth if and only if each of its components is, so $G \circ F$ is semismooth. Hence, the composite function $g_s \circ F$ is semismooth. Note that f_i is semismooth, $\nabla f_{ij_i}(x) \in \partial_{\mathrm{Cl}} f_i(x)$, and $\frac{\partial g_s(F)}{\partial f_i}\big|_{F=F(x)}$ is continuous and bounded in the neighbourhood $N(x^*)$. According to Lemma 1.1 and (3.19), we deduce that

$$
\begin{aligned}
(3.20)\quad & \|g_s(F(x)) - g_s(F(x^*)) - (JG(F)\big|_{F=F(x)}V)_s(x-x^*)\| \\
&= \left\| g_s(F(x)) - g_s(F(x^*)) - \sum_{i=1}^{m} \frac{\partial g_s(F)}{\partial f_i}\Big|_{F=F(x)} \nabla f_{ij_i}(x)^T(x-x^*) \right\| \\
&\leqslant \|g_s(F(x)) - g_s(F(x^*)) - (g_s \circ F)'(x^*; x-x^*)\| \\
&\quad + \left\| (g_s \circ F)'(x^*; x-x^*) - \sum_{i=1}^{m} \frac{\partial g_s(F)}{\partial f_i}\Big|_{F=F(x)} \nabla f_{ij_i}(x)^T(x-x^*) \right\| \\
&= o(\|x-x^*\|) + \left\| \sum_{i=1}^{m} \frac{\partial g_s(F)}{\partial f_i}\Big|_{F=F(x^*)} f_i'(x^*; x-x^*) \right. \\
&\quad \left. - \sum_{j=1}^{m} \frac{\partial g_s(F)}{\partial f_i}\Big|_{F=F(x)} \nabla f_{ij_i}(x)^T(x-x^*) \right\| \\
&\leqslant o\|x-x^*\| + \left\| \sum_{i=1}^{m} \frac{\partial g_s(F)}{\partial f_i}\Big|_{F=F(x^*)} [f_i'(x^*; x-x^*) - \nabla f_{ij_i}(x)^T(x-x^*)] \right\| \\
&\quad + \left\| \sum_{i=1}^{m} \left[\frac{\partial g_s(F)}{\partial f_i}\Big|_{F=F(x)} - \frac{\partial g_s(F)}{\partial f_i}\Big|_{F=F(x^*)} \right] \nabla f_{ij_i}(x)^T(x-x^*) \right\| \\
&= o(\|x-x^*\|) + o(\|x-x^*\|) + o(\|x-x^*)\| \\
&= o(\|x-x^*\|), \quad s = 1, \dots, n.
\end{aligned}
$$

Therefore, for any $V \in \partial_* F(x)$, the following relation holds:

$$(3.21)\qquad \|G(F(x)) - G(F(x^*)) - JG(F)\big|_{F=F(x)}V(x-x^*)\| = o(\|x-x^*\|).$$

In view of (3.21), we can choose a positive number ε small enough such that the relation $\|x-x^*\| \leqslant \varepsilon$ implies that

$$(3.22)\qquad \|G(F(x)) - G(F(x^*)) - JG(F)\big|_{F=F(x)}V(x-x^*)\| \leqslant \Delta\|x-x^*\|$$

and

$$\{x \in \mathbb{R}^n \mid \|x-x^*\| \leqslant \varepsilon\} \subset N(x^*).$$

From (3.16) it follows that if $\|x-x^*\| \leqslant \varepsilon$, then for any $V \in \partial_* F(x)$, $JG(F)\big|_{F=F(x)}V$ is nonsingular and satisfies

$$(3.23)\qquad \|\big(JG(F)\big|_{F=F(x)}V\big)^{-1}\| \leqslant \frac{10}{9}\beta, \quad V \in \partial_* F(x).$$

We next suppose that $\|x^k - x^*\| \leqslant \varepsilon$ for all k. Setting $A = W_k = JG(F)\big|_{F=F(x^k)}U_k$ and $B = JG(F)\big|_{F=F(x^k)}V_k$ in (2.19), one has

$$\|W_k^{-1}\| \leqslant \frac{\|\big(JG(F)\big|_{F=F(x^k)}V_k\big)^{-1}\|}{1 - \|[JG(F)\big|_{F=F(x^k)}V_k]^{-1}[JG(F)\big|_{F=F(x^k)}(U_k - V_k)]\|} \leqslant \frac{\frac{10}{9}\beta}{1 - \frac{10}{9}\beta\gamma\Delta} \leqslant \frac{\frac{10}{9}\beta}{1 - \frac{5}{27}} \leqslant \frac{3}{2}\beta. \tag{3.24}$$

We deduce that

$$\begin{aligned} \|x^{k+1} - x^*\| &= \|x^k - W_k^{-1}G(F(x^k)) - x^*\| \\ &\leqslant \|W_k^{-1}\|\,\|G(F(x^k)) - G(F(x^*)) - W_k(x^k - x^*)\| \\ &\leqslant \|W_k^{-1}\|\,\big\{\|G(F(x^k)) - G(F(x^*)) - JG(F)\big|_{F=F(x^k)} \\ &\quad \times V_k(x^k - x^*)\| + \|JG(F)\big|_{F=F(x^k)}(V_k - U_k)\|\,\|x^k - x^*\|\big\}. \end{aligned} \tag{3.25}$$

Substituting (3.14), (3.22) and (3.24) into (3.25) we conclude

$$\begin{aligned} \|x^{k+1} - x^*\| &\leqslant \frac{3}{2}\beta(\Delta\|x^k - x^*\| + \gamma\Delta\|x^k - x^*\|) \\ &= \frac{3}{2}(\beta\Delta + \beta\gamma\Delta)\|x^k - x^*\| \\ &\leqslant \frac{1}{4}\Big(\frac{1}{\gamma} + 1\Big)\|x^k - x^*\| \leqslant \frac{1}{2}\|x^k - x^*\|. \end{aligned} \tag{3.26}$$

Similarly to the proof of Theorem 2.1, it follows that the iterative method (3.11) is well-defined and the sequence $\{x^k\}$ converges to x^* Q-linearly.

Next, we proceed to prove that the sequence $\{x^k\}$ generated by (3.11) converges to x^* Q-superlinearly under the assumption that $\lim_{k\to\infty} \|V_k - U_k\| = 0$. By virtue of (3.17), (3.21), (3.24) and (3.25), the relation $\|x^k - x^*\| \leqslant \varepsilon$ implies that

$$\begin{aligned} \|x^{k+1} - x^*\| &\leqslant \|W_k^{-1}\|\,\big[\|G(F(x^k)) - G(F(x^*)) - JG(F)\big|_{F=F(x^k)}V_k(x^k - x^*)\| \\ &\quad + \|JG(F)\big|_{F=F(x^k)}\|\,\|(V_k - U_k)\|\,\|x^k - x^*\|\big] \\ &\leqslant \frac{3}{2}\beta\big[\|G(F(x^k)) - G(F(x^*)) - JG(F)\big|_{F=F(x^k)}V_k(x^k - x^*)\| \\ &\quad + \gamma\|V_k - U_k\|\,\|x^k - x^*\|\big] \\ &= \frac{3}{2}\beta[o(\|x^k - x^*\|) + o\|x^k - x^*\|)] = o\|x^k - x^*\|. \end{aligned}$$

This means that the sequence $\{x^k\}$ converges to x^* Q-superlinearly. Thus we have completed the proof of the theorem. $\square$

R e m a r k 3.1. Theorem 3.1 shows that the Newton method for solving the equations (3.1) given by (3.8) has Q-superlinear convergence.

Acknowledgements. The author is grateful to the anonymous referee for valuable suggestions and comments. He is also indebted to Li-Wei Zhang for a fruitful discussion concerning the topic of the paper.

References

[1] *W. Chen, G. Chen and E. Feng*: Variational principle with nonlinear complementarity for three dimensional contact problems and its numerical method. Sci. China Ser. A *39* (1996), 528–539.

[2] *X. Chen*: A verification method for solutions of nonsmooth equations. Computing *58* (1997), 281–294.

[3] *F. H. Clarke*: Optimization and Nonsmooth Analysis. John Wiley and Sons, New York, 1983.

[4] *V. F. Demyanov, G. E. Stavroulakis, L. N. Polyakova, and P. D. Panagiotopoulous*: Quasidifferentiability and Nonsmooth Modelling in Mechanics, Engineering and Economic. Kluwer Academic Publishers, Dordercht, 1996.

[5] *F. Meng, Y. Gao and Z. Xia*: Second-order directional derivatives for max-type functions. J. Dalian Univ. Tech. *38* (1998), 621–624. (In Chinese.)

[6] *M. Mifflin*: Semismooth and semiconvex functions in constrained optimization. SIAM J. Control Optim. *15* (1977), 959–972.

[7] *J. M. Ortega, W. C. Rheinboldt*: Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York, 1970.

[8] *L. Qi*: Convergence analysis of some algorithms for solving nonsmooth equations. Math. Oper. Res. *18* (1993), 227–244.

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

[10] *D. Sun, J. Han*: Newton and quasi-Newton methods for a class of nonsmooth equations and related problems. SIAM J. Optim. *7* (1997), 463–480.

Author's address: *Yan Gao*, Department of Mathematics and Mechanics, China University of Mining and Technology, Beijing 100083, China, e-mail: gaoyan@mail.cumtb.edu.cn.