

A further result on an implicit function theorem for locally Lipschitz functions [☆]

Defeng Sun

Department of Mathematics, National University of Singapore Singapore 117543, Singapore

Received 21 August 2000; received in revised form 1 October 2000; accepted 5 April 2001

Abstract

Let $H : \mathfrak{R}^m \times \mathfrak{R}^n \rightarrow \mathfrak{R}^n$ be a locally Lipschitz function in a neighborhood of (\bar{y}, \bar{x}) and $H(\bar{y}, \bar{x}) = 0$ for some $\bar{y} \in \mathfrak{R}^m$ and $\bar{x} \in \mathfrak{R}^n$. The implicit function theorem in the sense of Clarke (Pacific J. Math. 64 (1976) 97; Optimization and Nonsmooth Analysis, Wiley, New York, 1983) says that if $\pi_x \partial H(\bar{y}, \bar{x})$ is of maximal rank, then there exist a neighborhood Y of \bar{y} and a Lipschitz function $G(\cdot) : Y \rightarrow \mathfrak{R}^n$ such that $G(\bar{y}) = \bar{x}$ and for every y in Y , $H(y, G(y)) = 0$. In this paper, we shall further show that if H has a superlinear (quadratic) approximate property at (\bar{y}, \bar{x}) , then G has a superlinear (quadratic) approximate property at \bar{y} . This result is useful in designing Newton's methods for nonsmooth equations. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Implicit function theorem; Locally Lipschitz function; Higher order approximation

1. Introduction

Let $F : \mathfrak{R}^n \rightarrow \mathfrak{R}^p$ be a locally Lipschitz function in a neighborhood of some $x \in \mathfrak{R}^n$. Then by Rademacher's theorem, F is almost everywhere differentiable near x . Let D_F be the set where F is differentiable. In order to study the convergence of generalized Newton's methods, Qi [8] defines

$$\partial_B F(x) := \left\{ V \in \mathfrak{R}^{p \times n} \mid V = \lim_{x^k \rightarrow x} F'(x^k), F \text{ is differentiable at } x^k \text{ for all } k \right\}.$$

Hence, the generalized Jacobian $\partial F(x)$ in the sense of Clarke [1,2] is the convex hull of $\partial_B F(x)$, i.e.,

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

[☆] This work was supported by the Australian Research Council while the author was working at School of Mathematics, the University of New South Wales, Australia.

E-mail address: matsundf@nus.edu.sg (D. Sun).

Suppose that $p = n$ and thus F maps \mathfrak{R}^n into itself. Both $\partial F(x)$ and $\partial_B F(x)$ are used in developing Newton’s methods for solving nonsmooth equations, e.g., Qi and Sun [11,8] and Pang and Qi [7]:

$$F(x) = 0. \tag{1.1}$$

For a given starting point $x^0 \in \mathfrak{R}^n$, (nonsmooth) Newton’s methods for solving (1.1) can be stated as follows:

$$x^{k+1} = x^k - V_k^{-1}F(x^k), \quad k = 0, 1, 2, \dots, \tag{1.2}$$

where $V_k \in \partial F(x^k)$ (or $\partial_B F(x^k)$ or a variant). Assume that $x \in \mathfrak{R}^n$ is a solution of (1.1). For the superlinear (quadratic) convergence of methods (1.2), apart from assuming that $\partial F(x)$ (or $\partial_B F(x)$ or a variant) is of maximal rank, one needs an additional sufficient condition (see Kummer [3] for an example on the necessity). One such sufficient condition proposed in [11,8] is called *semismoothness*.

Semismoothness was originally introduced by Mifflin [6] for functionals. Convex functions, smooth functions, and piecewise smooth functions are examples of semismooth functions. The composition of semismooth functions is still a semismooth function (see [6]). In [11], Qi and Sun extended the definition of semismooth functions to $F : \mathfrak{R}^n \rightarrow \mathfrak{R}^p$. A vector valued function $F : \mathfrak{R}^n \rightarrow \mathfrak{R}^p$, which is Lipschitz continuous in a neighborhood of $x \in \mathfrak{R}^n$, is said to be *semismooth* at x , if

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

exists for any $h \in \mathfrak{R}^n$. It has been proved in [11] that F is semismooth at x if and only if all its component functions are. Also $F'(x; h)$, the directional derivative of F at x in the direction h , exists for any $h \in \mathfrak{R}^n$ and is equal to the above limit if F is semismooth at x .

Lemma 1.1 (Qi and Sun [11], Qi [8]). *Suppose that $F : \mathfrak{R}^n \rightarrow \mathfrak{R}^p$ is a locally Lipschitz function in a neighborhood of $x \in \mathfrak{R}^n$. If F is semismooth at x , then for any $h \rightarrow 0$ and $V \in \partial F(x + h)$,*

$$F(x + h) - F(x) - Vh = o(\|h\|). \tag{1.3}$$

Condition (1.3) and the nonsingularity of any matrix in $\partial F(x)$ (or $\partial_B F(x)$ or a variant) are two key conditions for the superlinear convergence of nonsmooth Newton’s methods, e.g., Kummer [4] and [7,8,11].

In this paper, we say $F : \mathfrak{R}^n \rightarrow \mathfrak{R}^p$ has a *superlinear approximate property* at $x \in \mathfrak{R}^n$ if F is a locally Lipschitz function in a neighborhood of x and for any $h \rightarrow 0$ and $V \in \partial H(x + h)$, it holds that

$$F(x + h) - F(x) - Vh = o(\|h\|). \tag{1.4}$$

Apparently, according to Lemma 1.1, if F is a locally Lipschitz function in a neighborhood of x and semismooth at x , then it has a superlinear approximate property at x .

A stronger notion than semismoothness is strong semismoothness. F is said to be *strongly semismooth* at x [11,8] if F is semismooth at x and for any $V \in \partial F(x + h)$, $h \rightarrow 0$,

$$F(x + h) - F(x) - Vh = O(\|h\|^2). \tag{1.5}$$

Analogously, we say $F : \mathfrak{R}^n \rightarrow \mathfrak{R}^p$ has a *quadratic approximate property* at $x \in \mathfrak{R}^n$ if F is a locally Lipschitz function in a neighborhood of x and for any $h \rightarrow 0$ and $V \in \partial H(x + h)$, (1.5) holds. It is obvious that $F : \mathfrak{R}^n \rightarrow \mathfrak{R}^p$ has a quadratic approximate property at x if F is strongly semismooth at x .

In the study of smoothing Newton’s methods (see, e.g., Qi and Sun [9] and Qi et al. [10]) one often finds that x in (1.1) is an implicit function of $y \in \mathfrak{R}^m$ defined by

$$H(y, x) = 0, \tag{1.6}$$

where H is a locally Lipschitz function mapping $\mathfrak{R}^m \times \mathfrak{R}^n \rightarrow \mathfrak{R}^n$ and y lies in \mathfrak{R}^m and $x \in \mathfrak{R}^n$. Let $\pi_x \partial H(y, x)$ be the set of all $n \times n$ matrices M such that, for some $n \times m$ matrix N , the $n \times (n + m)$ matrix $[N, M]$ belongs to $\partial H(y, x)$. Let $\pi_y \partial H(y, x)$ be such that $[\pi_y \partial H(y, x), \pi_x \partial H(y, x)] = \partial H(y, x)$.

Next, we state an implicit function theorem due to Clarke [1,2].

Theorem 1.1 (Clarke [1,2]). *Suppose that $H : \mathfrak{R}^m \times \mathfrak{R}^n \rightarrow \mathfrak{R}^n$ is a locally Lipschitz function in a neighborhood of (\bar{y}, \bar{x}) , which is a solution of (1.6), i.e., $H(\bar{y}, \bar{x}) = 0$. If $\pi_x \partial H(\bar{y}, \bar{x})$ is of maximal rank, then there exist an open neighborhood Y of \bar{y} and a function $G(\cdot) : Y \rightarrow \mathfrak{R}^n$ such that G is locally Lipschitz in Y , $G(\bar{y}) = \bar{x}$ and for every y in Y ,*

$$H(y, G(y)) = 0. \tag{1.7}$$

As we have pointed out, the superlinear (quadratic) approximate property for locally Lipschitz functions is one of two key conditions for achieving the high order convergence for nonsmooth Newton’s methods. Hence, one natural question arises: if H has a superlinear (quadratic) approximate property at (\bar{y}, \bar{x}) , does G have the same property at \bar{y} ? We will address this question in an affirmative way in Section 2.

2. Main result

Throughout this section, we assume that $H : \mathfrak{R}^m \times \mathfrak{R}^n \rightarrow \mathfrak{R}^n$ is a locally Lipschitz function in a neighborhood of (\bar{y}, \bar{x}) , which is a solution of (1.6). We also assume that $\pi_x \partial H(\bar{y}, \bar{x})$ is of maximal rank. Then, by Theorem 1.1, there exist an open neighborhood Y of \bar{y} and a locally Lipschitz function $G(\cdot) : Y \rightarrow \mathfrak{R}^n$ such that $G(\bar{y}) = \bar{x}$ and for every y in Y , $H(y, G(y)) = 0$.

The next theorem is our main result.

Theorem 2.1. *Suppose that all conditions in Theorem 1.1 hold. If H has a superlinear (quadratic) approximate property at (\bar{y}, \bar{x}) , then G has a superlinear (quadratic) approximate property at \bar{y} .*

To prove the above theorem, we need two lemmas.

Lemma 2.1. *Let $F : \mathfrak{R}^n \rightarrow \mathfrak{R}^p$ be a locally Lipschitz function in a neighborhood of $x \in \mathfrak{R}^n$. Then the following two statements are equivalent:*

(i) *for any $V \in \partial F(x + h)$, $h \rightarrow 0$,*

$$F(x + h) - F(x) - Vh = o(\|h\|),$$

(ii) *for any $x + h \in D_F$, $h \rightarrow 0$,*

$$F(x + h) - F(x) - F'(x + h)h = o(\|h\|).$$

Proof. (i) \rightarrow (ii) is obvious.

Next we prove (ii) \rightarrow (i): Assume by contradiction that (ii) holds while (i) does not hold. Then, there exist a positive number c , a sequence $\{h^i\}_{i=1}^\infty$ ($h^i \neq 0$) converging to 0 and a corresponding generalized Jacobian sequence $V_i \in \partial F(x + h^i)$ such that

$$\limsup_{i \rightarrow \infty} \frac{\|F(x + h^i) - F(x) - V_i h^i\|}{\|h^i\|} \geq c > 0. \tag{2.1}$$

According to the Carathéodory theorem, any $V_i \in \partial F(x + h^i)$ can be expressed as

$$V_i = \sum_{j=1}^{np+1} \lambda_{ij} V_{ij}, \quad (2.2)$$

where $V_{ij} \in \partial_B F(x + h^i)$ and

$$\lambda_{ij} \in [0, 1], \quad \sum_{j=1}^{np+1} \lambda_{ij} = 1. \quad (2.3)$$

For each $V_{ij} \in \partial_B F(x + h^i)$, by the definition of $\partial_B F(x + h^i)$, there exists $y^{ij} \in D_F$ such that

$$\|y^{ij} - (x + h^i)\| \leq \|h^i\|^2 \quad (2.4)$$

and

$$\|V_{ij} - F'(y^{ij})\| \leq \|h^i\|. \quad (2.5)$$

By (2.2)–(2.5), we obtain

$$\begin{aligned} & \|F(x + h^i) - F(x) - V_i h^i\| \\ & \leq \sum_{j=1}^{np+1} \lambda_{ij} \|F(x + h^i) - F(x) - F'(y^{ij}) h^i\| + \sum_{j=1}^{np+1} \lambda_{ij} \|[V_i - F'(y^{ij})] h^i\| \\ & \leq \sum_{j=1}^{np+1} \lambda_{ij} \|F(x + h^i) - F(x) - F'(y^{ij}) h^i\| + O(\|h^i\|^2) \\ & = \sum_{j=1}^{np+1} \lambda_{ij} \|F(x + h^i) - F(x) - F'(y^{ij})[(y^{ij} - x) + (x + h^i - y^{ij})]\| + O(\|h^i\|^2) \\ & \leq \sum_{j=1}^{np+1} \lambda_{ij} \|F(x + h^i) - F(x) - F'(y^{ij})(y^{ij} - x)\| + \sum_{j=1}^{np+1} \lambda_{ij} \|F'(y^{ij})(x + h^i - y^{ij})\| + O(\|h^i\|^2), \end{aligned}$$

which, together with (2.4) and the fact that $\{F'(y^{ij})\}$ are uniformly bounded because of the local Lipschitz property of F [2, Proposition 2.6.2], implies that

$$\|F(x + h^i) - F(x) - V_i h^i\| \leq \sum_{j=1}^{np+1} \lambda_{ij} \|F(x + h^i) - F(x) - F'(y^{ij})(y^{ij} - x)\| + O(\|h^i\|^2). \quad (2.6)$$

Relations (2.6), (2.3) and (2.4), together with the Lipschitz continuity of F , imply that

$$\|F(x + h^i) - F(x) - V_i h^i\| \leq \sum_{j=1}^{np+1} \lambda_{ij} \|F(y^{ij}) - F(x) - F'(y^{ij})(y^{ij} - x)\| + O(\|h^i\|^2). \quad (2.7)$$

Thus, by (ii), (2.1) and (2.4), from (2.7) we obtain

$$\limsup_{i \rightarrow \infty} \frac{\|F(x + h^i) - F(x) - V_i h^i\|}{\|h^i\|} = 0,$$

which contradicts (2.1). This contradiction shows that (ii) \rightarrow (i). \square

The following lemma is included in the proof of Sun and Sun [12, Theorem 3.6].

Lemma 2.2. Let $F : \mathfrak{R}^n \rightarrow \mathfrak{R}^p$ be a locally Lipschitz function in a neighborhood of $x \in \mathfrak{R}^n$. Then the following two statements are equivalent:

(i) for any $V \in \partial F(x+h)$, $h \rightarrow 0$,

$$F(x+h) - F(x) - Vh = O(\|h\|^2),$$

(ii) for any $x+h \in D_F$, $h \rightarrow 0$,

$$F(x+h) - F(x) - F'(x+h)h = O(\|h\|^2).$$

Proof of Theorem 2.1. First, we prove that for any $y \in Y \cap D_G$, $y \rightarrow \bar{y}$,

$$G(y) - G(\bar{y}) - G'(y)(y - \bar{y}) = o(\|y - \bar{y}\|). \tag{2.8}$$

For $y \in Y \cap D_G$, let $h = y - \bar{y}$. Then for any $|t|$ sufficiently small,

$$H(y+th, G(y+th)) - H(y, G(y)) = 0 - 0 = 0. \tag{2.9}$$

According to the Mean Value Theorem [2, Proposition 2.6.5], for $y \in Y \cap D_G$,

$$H(y+th, G(y) + tG'(y)h) - H(y, G(y)) \in t\mathcal{M}, \tag{2.10}$$

where

$$\mathcal{M} = \text{conv} \left\{ \partial H[(y, G(y)), (y+th, G(y) + tG'(y)h)] \begin{pmatrix} h \\ G'(y)h \end{pmatrix} \right\}.$$

Hence, by [2, Proposition 2.6.2] there exists a $W \in \partial H(y, G(y))$ such that for $t \rightarrow 0$, $t \neq 0$,

$$[H(y+th, G(y) + tG'(y)h) - H(y, G(y))]/t \rightarrow W \begin{pmatrix} h \\ G'(y)h \end{pmatrix}. \tag{2.11}$$

Note that for any fixed $y \in Y \cap D_G$ and $t \rightarrow 0$, we have

$$G(y+th) = G(y) + tG'(y)h + o(t),$$

which, together with the Lipschitz continuity of H , (2.9) and (2.11), implies that

$$W \begin{pmatrix} h \\ G'(y)h \end{pmatrix} = 0.$$

Let $\pi_y W \in \pi_y \partial H(y, G(y))$ and $\pi_x W \in \pi_x \partial H(y, G(y))$ be such that

$$W = [\pi_y W, \pi_x W].$$

Hence, we have

$$(\pi_y W)(y - \bar{y}) + (\pi_x W)G'(y)(y - \bar{y}) = 0. \tag{2.12}$$

On the other hand, since H has a superlinear approximate property at (\bar{y}, \bar{x}) and G is Lipschitz continuous in Y , we have

$$\begin{aligned} & H(y, G(y)) - H(\bar{y}, G(\bar{y})) - (\pi_y W)(y - \bar{y}) - (\pi_x W)[G(y) - G(\bar{y})] \\ &= o(\|y - \bar{y}\|) + o(\|G(y) - G(\bar{y})\|) \\ &= o(\|y - \bar{y}\|). \end{aligned} \tag{2.13}$$

The fact that $H(y, G(y)) = H(\bar{y}, G(\bar{y})) = 0$, together with equation (2.13), implies that

$$(\pi_y W)(y - \bar{y}) + (\pi_x W)[G(y) - G(\bar{y})] = o(\|y - \bar{y}\|),$$

which, together with (2.12) and the fact that $\|(\pi_x W)^{-1}\|$ is uniformly bounded [11], proves (2.8). Hence, by Lemma 2.1, we conclude that G has a superlinear approximate property at \bar{y} .

If H has a quadratic approximate property at (\bar{y}, \bar{x}) , then by the above argument and Lemma 2.2 one can easily prove that G has a quadratic approximate property at \bar{y} . We omit the details here. \square

Corollary 2.1. *Suppose that all conditions in Theorem 1.1 hold. If H is (strongly) semismooth at (\bar{y}, \bar{x}) , then G has a superlinear (quadratic) approximate property at \bar{y} .*

The superlinear (quadratic) approximate property of the implicit function G at \bar{y} is very useful in analyzing the higher order convergence of smoothing Newton's methods for solving complementarity problems and variational inequalities [9,10]. We also believe it is useful in other subjects like bi-level programs, or generally, mathematical programs with equilibrium constraints, e.g., Luo et al. [5], where implicit functions are widely used.

References

- [1] F.H. Clarke, On the inverse function theorem, *Pacific J. Math.* 64 (1976) 97–102.
- [2] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
- [3] B. Kummer, Newton's method for nondifferentiable functions in: *Advances in Mathematical Optimization*, Math. Res., Vol. 45, Akademie-Verlag, Berlin, 1988, pp. 114–125.
- [4] B. Kummer, Newton's method based on generalized derivatives for nonsmooth functions: convergence analysis, in: *Advances in Optimization* Lambrecht, 1991, *Lecture Notes in Economics and Mathematical Systems*, Vol. 382, W. Oettli, D. Pallaschke (Eds.), Springer, Berlin, 1992, pp. 171–194.
- [5] Z.-Q. Luo, J.-S. Pang, D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, 1996.
- [6] R. Mifflin, Semismooth and semiconvex functions in constrained optimization, *SIAM J. Control Optim.* 15 (1977) 957–972.
- [7] J.-S. Pang, L. Qi, Nonsmooth equations: Motivation and algorithms, *SIAM J. Optim.* 3 (1993) 443–465.
- [8] L. Qi, Convergence analysis of some algorithms for solving nonsmooth equations, *Math. Oper. Res.* 18 (1993) 227–244.
- [9] L. Qi, D. Sun, Smoothing functions and a smoothing Newton method for complementarity and variational inequality problems. *J. Optim. Theory Appl.*, to appear.
- [10] L. Qi, D. Sun, G. Zhou, A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities, *Math. Programming* 87 (2000) 1–35.
- [11] L. Qi, J. Sun, A nonsmooth version of Newton's method, *Math. Programming* 58 (1993) 353–367.
- [12] D. Sun, J. Sun, Semismooth matrix valued functions, Manuscript, School of Mathematics, The University of New South Wales, Sydney 2052, Australia, November 1999 (revised, November 2000).