A Further Result on an Implicit Function Theorem for Locally Lipschitz Functions *

Defeng Sun
School of Mathematics
The University of New South Wales
Sydney 2052, Australia

August 7, 2000

Abstract

Let $H : \Re^m \times \Re^n \to \Re^n$ be a locally Lipschitz function in a neighborhood of $(\bar{y}, \bar{x})$ and $H(\bar{y}, \bar{x}) = 0$ for $\bar{y} \in \Re^m$ and $\bar{x} \in \Re^n$. The implicit function theorem in the sense of Clarke [1, 2] 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 $G(\cdot) : Y \to \Re^n$ 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.

Key words. Implicit function theorem, Locally Lipschitz function, Higher order approximation

1 Introduction

Let $F : \Re^n \to \Re^p$ be a locally Lipschitz function in neighborhood of $x \in \Re^n$. 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 convergence of generalized Newton's methods, Qi [8] defines

$$\partial_B F(x) := \{V \in \Re^{p \times n} |\ V = \lim_{x^k \to x} F'(x^k), F \text{ is differentiable at } x^k \text{ for all } k\}.$$

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 is supported by the Australian Research Council. E-mail: sun@maths.unsw.edu.au

Suppose that $p = n$ and thus F maps $\Re^n$ 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 $x^0 \in \Re^n$, (nonsmooth) Newton's methods for solving (1.1) can be stated as follows

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

where $V_k \in \partial F(x^k)$ (or $\partial_B F(x^k)$ or an variant). Assume that $x \in \Re^n$ is a solution of (1.1). For superlinear (quadratic) convergence of method (1.2), apart from assuming that $\partial F(x)$ (or $\partial_B F(x)$ or an 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 $F : \Re^n \to \Re^p$. A vector valued function $F : \Re^n \to \Re^p$, which is Lipschitz continuous in a neighborhood of $x \in \Re^n$, is said to be *semismooth* at x, if

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

exists for any $h \in \Re^n$. 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 $h \in \Re^n$ and is equal to the above limit if F is semismooth at x.

Lemma 1.1 *[11, 8] Suppose that $F : \Re^n \to \Re^p$ is a locally Lipschitz function in a neighborhood of $x \in \Re^n$. If F is semismooth at x, then for any $h \to 0$ 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 an variant) are two key conditions for superlinear convergence of nonsmooth Newton methods, e.g., Kummer [4] and [7, 8, 11].

In this paper, we say $F : \Re^n \to \Re^p$ to have a *superlinear approximate property* at $x \in \Re^n$ if F is a locally Lipschitz function in a neighborhood of x and for any $h \to 0$ 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 $V \in \partial F(x+h)$, $h \to 0$,

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

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

In the study of smoothing Newton 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 $y \in \Re^m$ defined by

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

where H is a locally Lipschitz function mapping $\Re^m \times \Re^n \to \Re^n$ and y lies in $\Re^m$ and $x \in \Re^n$. Let $\pi_x \partial H(y, x)$ be the set 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 $\partial H(y, x)$. Denote $\pi_y \partial H(y, x)$ 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 *[1, 2] Suppose that $H : \Re^m \times \Re^n \to \Re^n$ 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 $G(\cdot) : Y \to \Re^n$ 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 that the superlinear (quadratic) approximate property for locally Lipschitz functions is one of two key conditions for achieving 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 has the same property at $\bar{y}$? We will address this question in an affirmative way in next section.

2 Main Result

Throughout this section we assume that $H : \Re^m \times \Re^n \to \Re^n$ is a locally Lipschitz function in a neighborhood of $(\bar{y}, \bar{x})$, which is a solution of (1.6). We aslo 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 $G(\cdot) : Y \to \Re^n$ 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 $F : \Re^n \to \Re^p$ be a locally Lipschitz function in a neighborhood of $x \in \Re^n$. Then the following two statements are equaivalent:*

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

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

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

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

Proof: (i)→(ii) is obvious.

Next we prove (ii)→(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

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

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

$$V_i = \sum_{j=1}^{np+1} \lambda_{i_j} V_{i_j}, \tag{2.2}$$

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

$$\lambda_{i_j} \in [0,1], \quad \sum_{j=1}^{np+1} \lambda_{i_j} = 1 \tag{2.3}$$

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

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

and

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

By (2.2)–(2.5), we obtain

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

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

$$
\begin{aligned}
&\|F(x+h^i)-F(x)-V_ih^i\| \\
&\quad \leq \sum_{j=1}^{np+1}\lambda_{i_j}\|F(x+h^i)-F(x)-F'(y^{i_j})(y^{i_j}-x)\| + O(\|h^i\|^2).
\end{aligned} \tag{2.6}
$$

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

$$
\begin{aligned}
&\|F(x+h^i)-F(x)-V_ih^i\| \\
&\quad \leq \sum_{j=1}^{np+1}\lambda_{i_j}\|F(y^{i_j})-F(x)-F'(y^{i_j})(y^{i_j}-x)\| + O(\|h^i\|^2).
\end{aligned} \tag{2.7}
$$

Thus, by (ii), (2.3) and (2.4), from (2.7) we obtain

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

which contradicts (2.1). This contradiction shows that (ii)→(i). ∎

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

Lemma 2.2 *Let $F:\Re^n\to\Re^p$ be a locally Lipschitz function in a neighborhood of $x\in\Re^n$. Then the following two statements are equaivalent:*

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

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

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

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

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

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

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

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

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

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

Note that for any fixed $y \in Y \cap D_G$ and $t \to 0$, 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 (quadratic) approximate property at $(\bar{y}, \bar{x})$ and G is Lipschitz continuous in Y, we have

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

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 that (2.8) holds. 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. ∎

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 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 constrains, e.g., Luo et al [5], where implicit functions are widely used.

References

[1] F.H. Clarke, "On the inverse function theorem", *Pacific Journal of Mathematics, 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*, 114–125, Math. Res., 45, Akademie-Verlag, Berlin, 1988.

[4] B. Kummer, "Newton's method based on generalized derivatives for nonsmooth functions: convergence analysis", in: *Advances in Optimization* (Lambrecht, 1991), 171–194, Lecture Notes in Econom. and Math. Systems, 382, Springer, Berlin, 1992

[5] Z.-Q. Luo, J.-S. Pang and D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, 1996.

[6] R. Mifflin, "Semismooth and semiconvex functions in constrained optimization", *SIAM Journal on Control and Optimization, 15* (1977), 957–972.

[7] J.-S. Pang and L. Qi, "Nonsmooth equations: Motivation and algorithms", *SIAM Journal on Optimization, 3* (1993), 443–465.

[8] L. Qi, "Convergence analysis of some algorithms for solving nonsmooth equations", *Mathematics of Operations Research, 18* (1993), 227–244.

[9] L. Qi and D. Sun, "Smoothing functions and a smoothing Newton method for complementarity and variational inequality problems", AMR 98/24, Applied Mathematics Report, The University of New South Wales, Sydney 2052, Australia, 1998.

[10] L. Qi, D. Sun and G. Zhou, "A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities", *Mathematical Programming, 87* (2000), 1–35.

[11] L. Qi and J. Sun, "A nonsmooth version of Newton's method", *Mathematical Programming, 58* (1993), 353–367.

[12] D. Sun and J. Sun, "Semismooth matrix valued functions", Manuscript, School of Mathematics, The University of New South Wales, Sydney 2052, Australia, 1999.