Mathematical Programming 58 (1993) 353-367
North-Holland

A nonsmooth version of Newton's method

Liqun Qi*
School of Mathematics, The University of New South Wales, Kensington, NSW, Australia

Jie Sun**
Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, USA

Received 27 December 1989
Revised manuscript received 2 July 1991

Newton's method for solving a nonlinear equation of several variables is extended to a nonsmooth case by using the generalized Jacobian instead of the derivative. This extension includes the B-derivative version of Newton's method as a special case. Convergence theorems are proved under the condition of semismoothness. It is shown that the gradient function of the augmented Lagrangian for C^2-nonlinear programming is semismooth. Thus, the extended Newton's method can be used in the augmented Lagrangian method for solving nonlinear programs.

Key words: Newton's methods, generalized Jacobian, semismoothness.

1. Introduction

Newton's method

$$x^{k+1}=x^k-[F'(x^k)]^{-1}F(x^k) \tag{1.1}$$

is a classic method for solving the nonlinear equation

$$F(x)=0, \tag{1.2}$$

where $F:\mathbb{R}^n\to\mathbb{R}^n$ is a continuously differentiable function, i.e., a *smooth function.* Many other methods for solving (1.2) are related to this method [12].

Suppose now that F is not a smooth function, but a locally Lipschitzian function. Then the formula (1.1) cannot be used. Let $\partial F(x^k)$ be the *generalized Jacobian* of F at x^k, defined by Clarke [4]. In this case, instead of (1.1), one may use

$$x^{k+1}=x^k-V_k^{-1}F(x^k), \tag{1.3}$$

Correspondence to: Dr. Liqun Qi, School of Mathematics, The University of New South Wales, Kensington, NSW 2033, Australia.

* This author's work is supported in part by the Australian Research Council.
** This author's work is supported in part by the National Science Foundation under grant DDM-8721709.

where $V_k \in \partial F(x^k)$, to solve (1.2). In this paper, we show that local and global convergence results hold for (1.3) when F is semismooth. Semismoothness was originally introduced by Mifflin [10] for functionals. In Section 2, we extend this definition to function $F:\mathbb{R}^n \to \mathbb{R}^n$ and show that semismoothness is equivalent to the uniform convergence of directional derivatives in all directions. We also show that a locally Lipschitzian function is semismooth at a point if it has the strong Fréchet derivative at that point.

In Section 3 the local and global convergence results of (1.3) are proved under the semismooth condition and compared with another nonsmooth version of Newton's method developed by Robinson [17, 18, 19], Pang [13, 14], and Harker and Xiao [7], which uses the B-derivative instead of the Fréchet derivative in (1.1). The B-derivative was first introduced by Robinson [17]. Pang [13] proved convergence of this version of Newton's Method when F has strong and nonsingular Fréchet derivative at the solution point. We show that this B-derivative version of Newton's Method can be regarded as a special version of (1.3). Hence our convergence theorem is stronger than Pang's convergence theorem in a certain sense. We also compare our results with Kummer's results in [9].

In Section 4, we show that the augmented Lagrangian of a C^2-nonlinear programming problem has semismooth gradients. Thus, the extended Newton's method can be used in the augmented Lagrangian method for solving nonlinear programs. We also briefly mention other possible applications of this method.

2. Semismooth functions of several variables

Suppose $F:\mathbb{R}^n \to \mathbb{R}^m$ is a locally Lipschitzian function. According to Rademacher's Theorem, F is differentiable almost everywhere. Denote the set of points at which F is differentiable by D_F. We write $JF(x)$ for the usual $n \times m$ Jacobian matrix of partial derivatives whenever x is a point at which the necessary partial derivatives exist. Let $\partial F(x)$ be the generalized Jacobian defined by Clarke in 2.6 of [4]. Then

$$\partial F(x) = \mathrm{co}\left\{ \lim_{\substack{x_i \to x \\ x_i \in D_F}} JF(x_i) \right\}. \tag{2.1}$$

By 2.6.5 of [4], for any $x, y \in \mathbb{R}^n$,

$$F(y) - F(x) \in \mathrm{co}\, \partial F([x, y])(y - x), \tag{2.2}$$

where the right-hand side denotes the convex hull of all points of the form $V(y-x)$ with $V \in \partial F(u)$ for some point u in $[x, y]$. We will use (2.2) several times later.

Let x be a point in $\mathbb{R}^n$. Assume that for any $h \in \mathbb{R}^n$,

$$\lim_{\substack{V \in \partial F(x+th) \\ t \downarrow 0}} \{Vh\}, \tag{2.3}$$

exists.

Proposition 2.1. *Under this assumption, the classic directional derivative*

$$F'(x; h) = \lim_{t \downarrow 0} \frac{F(x+th) - F(x)}{t} \tag{2.4}$$

exists and is equal to the limit in (2.3); *i.e.*,

$$F'(x; h) = \lim_{\substack{V \in \partial F(x+th) \\ t \downarrow 0}} \{Vh\}. \tag{2.5}$$

Proof. Since F is locally Lipschitzian,

$$\frac{F(x+th) - F(x)}{t} \tag{2.6}$$

is bounded. Suppose that l is a limiting point of (2.6) as $t \downarrow 0$. Then there are $t_j \downarrow 0$ such that

$$l = \lim_{j \to \infty} \frac{F(x+t_j h) - F(x)}{t_j}.$$

It suffices to show that l is equal to the limit in (2.3). By (2.2),

$$\frac{F(x+t_j h) - F(x)}{t_j} \in \mathrm{co}\, \partial F([x, x+t_j h])h.$$

By the Carathéodory Theorem, there exist $t_j^{(k)} \in [0, t_j]$, $\lambda_j^{(k)} \in [0, 1]$, $V_j^{(k)} \in \partial F([x, x+t_j^{(x)} h])$, for $k = 0, \dots, m$, $\sum_{k=0}^{m} \lambda_j^{(k)} = 1$, such that

$$\frac{F(x+t_j h) - F(x)}{t_j} = \sum_{k=0}^{m} \lambda_j^{(k)} V_j^{(k)} h.$$

By passing to a subsequence, we can assume that $\lambda_j^{(k)} \to \lambda_j$ as $j \to \infty$. We have $\lambda_j \in [0, 1]$ for $k = 0, \dots, m$ and $\sum_{k=0}^{m} \lambda_j = 1$. Then,

$$l = \lim_{j \to \infty} \frac{F(x+t_j h) - F(x)}{t_j} = \lim_{j \to \infty} \left\{ \sum_{k=0}^{m} \lambda_j^{(k)} V_j^{(k)} h \right\}$$

$$= \sum_{k=0}^{m} \lim_{j \to \infty} \lambda_j^{(k)} \lim_{j \to \infty} \{V_j^{(k)} h\} = \sum_{k=0}^{m} \lambda_j \lim_{\substack{V \in \partial F(x+th) \\ t \downarrow 0}} \{Vh\} = \lim_{\substack{V \in \partial F(x+th) \\ t \downarrow 0}} \{Vh\}.$$

This completes the proof. $\square$

We say that F is *semismooth* at x if F is locally Lipschitzian at x and

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

exists for any $h \in \mathbb{R}^n$. Clearly, this assumption is stronger than (2.3) and it implies that the Hadamard directional derivative

$$\lim_{\substack{h' \to h \\ t \downarrow 0}} \frac{F(x+th') - F(x)}{t} = \lim_{\substack{V \in \partial F(x+th') \\ h' \to h, t \downarrow 0}} \{Vh'\}. \tag{2.7}$$

Semismoothness was originally introduced by Mifflin [10] for functionals. Convex functions, smooth functions and subsmooth functions are examples of semismooth functions. Scalar products and sums of semismooth functions are still semismooth functions, see [10]. For applications of semismoothness and its relationship with other nonsmooth properties, see [1, 2, 3, 5, 6, 10, 11, 15, 21, 26].

We need a lemma for our discussion.

Lemma 2.2. *Suppose that $F:\mathbb{R}^n \to \mathbb{R}^m$ is a locally Lipschitzian function and $F'(x; h)$ exists for any h at x. Then*

(i) *$F'(x; \cdot)$ is Lipschitzian;*

(ii) *for any h, there exists a $V \in \partial F(x)$ such that*

$$F'(x; h) = Vh.$$

Proof. Suppose L is the Lipschitzian constant near x. Let $\|\cdot\|$ stand for the Euclidean norm. For any $h, h' \in \mathbb{R}^n$,

$$\begin{aligned}\|F'(x; h) - F'(x; h')\| &= \left\| \lim_{t\downarrow 0} \frac{F(x+th) - F(x+th')}{t} \right\| \\ &\leq \lim_{t\downarrow 0} \frac{\|F(x+th) - F(x+th')\|}{t} \leq L\|h - h'\|.\end{aligned}$$

This proves (i). By (2.2) and (2.4), there is a sequence $\{t_k\}$ and a sequence $\{V_k\}$ such that $t_k \downarrow 0$, $V_k \in \operatorname{co} \partial F([x, x+t_k h])$,

$$F'(x; h) = \lim_{k\to\infty} \{V_k h\}.$$

Because of the local Lipschitzian property of F, $\{V_k\}$ is bounded. By passing to a subsequence, we may assume that $V_k \to V$. However, due to the definition of the generalized Jacobian, ∂F is closed. Thus, $V \in \partial F(x)$. This proves (ii). $\square$

Theorem 2.3. *Suppose that $F:\mathbb{R}^n \to \mathbb{R}^m$ is a locally Lipschitzian function. The following statements are equivalent:*

(i) *F is semismooth at x;*

(ii) *the right-hand side limit in (2.7) is uniformly convergent for all h with unit norm;*

(iii) *the right-hand side limit in (2.5) is uniformly convergent for all h with unit norm;*

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

$$Vh - F'(x; h) = o(\|h\|). \tag{2.8}$$

(v) $$\lim_{\substack{x+h \in D_F \\ h\to 0}} \frac{F'(x+h; h) - F'(x; h)}{\|h\|} = 0. \tag{2.9}$$

Proof. (i) $\to$ (ii). Suppose (ii) does not hold. Then there exist $\varepsilon > 0$, $\{h^k \in \mathbb{R}^n: \|h^k\| = 1, k = 1, 2, \ldots\}$, $\|\bar{h}^k - h^k\| \to 0$, $t_k \downarrow 0$, $V_k \in \partial F(x + t_k \bar{h}^k)$ such that

$$\|V_k \bar{h}^k - F'(x; h^k)\| \geq 2\varepsilon, \tag{2.10}$$

for $k=1,2,\ldots$. By passing to a subsequence, we may assume that $h^k \to h$. Thus, $\bar{h}^k \to h$ too. By Lemma 2.2(i), (2.10) implies that

$$\|V_k\bar{h}^k - F'(x;h)\| \geq \varepsilon,$$

for all sufficiently large k. This contradicts the semismoothness assumption.

(ii) → (iii) → (iv), obviously.

(iv) → (i). Suppose that F is not semismooth at x. Then there exist $h \in \mathbb{R}^n$, $h^k \to h$, $\varepsilon > 0$, $t_k \downarrow 0$, $V_k \in \partial F(x+t_k h^k)$ such that

$$\|V_k h^k - F'(x;h)\| \geq 2\varepsilon, \tag{2.11}$$

for $k=1,2,\ldots$. By Lemma 2.2(i), (2.11) implies that

$$\|V_k h^k - F'(x;h^k)\| \geq \varepsilon,$$

for all sufficiently large k. This contradicts (2.8).

(iv) → (v), according to Lemma 2.2(ii).

(v) → (iv). Given $\varepsilon > 0$, by (2.9), there exists $\delta > 0$ such that for any $h \in \mathbb{R}^n$ satisfying $\|h\| \leq \delta$ and $x+h \in D_F$,

$$\|F'(x+h;h) - F'(x;h)\| \leq \varepsilon\|h\|. \tag{2.12}$$

Assume now that $\|h\| \leq \frac{1}{2}\delta$ and $V \in \partial F(x+h)$. We will show that

$$\|Vh - F'(x;h)\| \leq 5\varepsilon\|h\|. \tag{2.13}$$

By (2.1),

$$Vh \in \operatorname{co}\left\{ \lim_{\substack{h^i \to h \\ x+h^i \in D_F}} F'(x+h^i;h) \right\}.$$

By the Carathéodory theorem, there exist $h^{(0)}, \ldots, h^{(m)}$ such that $\|h^{(k)} - h\| \leq \min\{\frac{1}{2}\delta, \|h\|, \varepsilon\|h\|/L\}$, $x+h^{(k)} \in D_F$, for $k=0,\ldots,m$, where L is the Lipschitzian constant of F near x, and

$$\left\| Vh - \sum_{k=0}^{m} \lambda_k F'(x+h^{(k)};h) \right\| \leq \varepsilon, \tag{2.14}$$

where $\lambda_k \geq 0$ for $k=0,\ldots,m$, $\sum_{k=0}^{m}\lambda_k = 1$. By (2.12) and Lemma 2.2(1),

$$\begin{aligned}
&\left\| \sum_{k=0}^{m} \lambda_k F'(x+h^{(k)};h) - F'(x;h) \right\| \\
&\leq \sum_{k=0}^{m} \lambda_k [\|F'(x+h^{(k)};h^{(k)}) - F'(x+h^{(k)};h)\| \\
&\qquad + \|F'(x+h^{(k)};h^{(k)}) - F'(x;h^{(k)})\| + \|F'(x;h) - F'(x;h^{(k)})\|] \\
&\leq \sum_{k=0}^{m} \lambda_k [L\|h^{(k)} - h\| + \varepsilon\|h^{(k)}\| + L\|h^{(k)} - h\|] \\
&\leq \sum_{k=0}^{m} \lambda_k \cdot 4\varepsilon\|h\| = 4\varepsilon\|h\|.
\end{aligned} \tag{2.15}$$

Now, (2.13) follows from (2.14) and (2.15). This proves (2.8). □

Corollary 2.4. *Suppose that $F:\mathbb{R}^n \to \mathbb{R}^m$ is a locally Lipschitzian function. If each component of F is semismooth at x, then F is semismooth at x.*

Proof. For $x+h \in D_F$, $h \to 0$,

$$\|F'(x+h; h) - F'(x; h)\| \leqslant \sum_{i=1}^{m} |F_i'(x+h; h) - F_i'(x; h)|.$$

Applying (2.9) to each F_i, we have

$$F_i'(x+h; h) - F_i'(x; h) = \mathrm{o}(\|h\|)$$

for each i. We thus get (2.9) for F. □

Recall the definition of *strong Fréchet derivative* [12]: The Fréchet derivative $F'(x)$ is said to be strong if

$$\lim_{\substack{y \to x \\ z \to x}} \frac{F(x) - F(y) - F'(x)(z-y)}{\|z-y\|} = 0. \tag{2.16}$$

Corollary 2.5. *If F has strong Fréchet derivative at x, then F is semismooth at x.*

Proof. In (2.16), let $y = x + h \in D_F$, $z = y + th$, and $t \downarrow 0$ first. Then we get (2.9). □

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

$$Vh - F'(x; h) = \mathrm{O}(\|h\|^{1+p}),$$

where $0 < p \leqslant 1$, then we call F is *p-order semismooth* at x. Note that p-order semismoothness $(0 < p \leqslant 1)$ implies semismoothness.

Remark. By (2.2), if F is semismooth at x, then for any $h \to 0$,

$$F(x+h) - F(x) - F'(x; h) = \mathrm{o}(\|h\|), \tag{2.17}$$

if F is p-order semismooth at x, then for any $h \to 0$,

$$F(x+h) - F(x) - F'(x; h) = \mathrm{O}(\|h\|^{1+p}).$$

3. A nonsmooth version of Newton's method

Suppose $F:\mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitzian. We wish to find the solution of the equation

$$F(x) = 0. \tag{3.1}$$

Let us consider the nonsmooth version of Newton's method

$$x^{k+1} = x^k - V_k^{-1} F(x^k), \tag{3.2}$$

where $V_k \in \partial F(x^k)$.

Proposition 3.1. *If all $V \in \partial F(x)$ are nonsingular, then there is a neighborhood $N(x)$ of x and a constant C such that for any $y \in N(x)$ and any $V \in \partial F(y)$, V is nonsingular and*

$$\|V^{-1}\| \leq C.$$

Proof. If the conclusion is not true, then there is a sequence $y^k \to x$, $V_k \in \partial F(y^k)$ such that either all V_k are singular or $\|V_k^{-1}\| \to \infty$. Since F is locally Lipschitzian, ∂F is bounded in a neighborhood of x. By passing to a subsequence, we may assume the $V_k \to V$. Then V must be singular, a contradiction to the assumption of this proposition. This completes the proof. □

Theorem 3.2 (local convergence). *Suppose that x^* is a solution of* (3.1), *F is locally Lipschitzian and smemismooth at x^*, and all $V \in \partial F(x^*)$ are nonsingular. Then the iteration method* (3.2) *is well-defined and convergent to x^* in a neighborhood of x^*. If in addition F is p-order semismooth at x^*, the the convergence of* (3.2) *is of order $1+p$.*

Proof. By Proposition 3.1, (3.2) is well-defined in a neighborhood of x^* for the first step $k=0$. Now

$$\begin{aligned}\|x^{k+1}-x^*\| &= \|x^k - x^* - V_k^{-1}F(x^k)\| \\ &\leq \|V_k^{-1}[F(x^k)-F(x^*)-F'(x^*; x^k-x^*)]\| \\ &\quad + \|V_k^{-1}[V_k(x^k-x^*)-F'(x^*; x^k-x^*)]\| \\ &= o(\|x^k-x^*\|). \end{aligned} \tag{3.3}$$

The last equality is due to Proposition 3.1, (2.8), and (2.17). The case that F is p-order semismooth at x is similar. □

Theorem 3.3 (global convergence). *Suppose that F is locally Lipschitzian and semismooth on $S=\{x \in \mathbb{R}^n\colon \|x-x^0\| \leq r\}$. Also suppose that for any $V \in \partial F(x)$, $x, y \in S$, V is nonsingular,*

$$\|V^{-1}\| \leq \beta,\ \|V(y-x)-F'(x; y-x)\| \leq \gamma\|y-x\|,$$

$$\|F(y)-F(x)-F'(x; y-x)\| \leq \delta\|y-x\|,$$

where $\alpha=\beta(\gamma+\delta)<1$ and $\beta\|F(x^0)\| \leq r(1-\alpha)$. Then the iterates (3.2) *remain in S and converge to the unique solution x^* of* (3.1) *in S. Moreover, the error estimate*

$$\|x^k-x^*\| \leq [\alpha/(1-\alpha)]\|x^k-x^{k-1}\| \tag{3.4}$$

holds for $k=1, 2, \ldots$.

Proof.

$$\|x^1-x^0\|=\|V_0^{-1}F(x^0)\|\leqslant\beta\|F(x^0)\|\leqslant r(1-\alpha).$$

So $x^1\in S$. Suppose now $x^1, x^2, \ldots, x^k\in S$. Then

$$\begin{aligned}\|x^{k+1}-x^k\|&=\|V_k^{-1}F(x^k)\|\leqslant\beta\|F(x^k)\|\\&\leqslant\beta\|F(x^k)-F(x^{k-1})-F'(x^{k-1};x^k-x^{k-1})\|\\&\quad+\beta\|V_{k-1}(x^k-x^{k-1})-F'(x^{k-1};x^k-x^{k-1})\|\\&\leqslant\beta(\delta+\gamma)\|x^k-x^{k-1}\|=\alpha\|x^k-x^{k-1}\|\leqslant\alpha^k\|x^1-x^0\|\\&\leqslant r\alpha^k(1-\alpha).\end{aligned}$$

Hence

$$\|x^{k+1}-x^0\|\leqslant\sum_{j=0}^{k}\|x^{j+1}-x^j\|\leqslant\sum_{j=0}^{k}r\alpha^j(1-\alpha)\leqslant r.$$

So $x^{k+1}\in S$, i.e., all the iterates (3.2) remain in S. For any k and p,

$$\|x^{k+p+1}-x^k\|\leqslant\sum_{j=k}^{k+p}\|x^{j+1}-x^j\|\leqslant\sum_{j=k}^{k+p}r\alpha^j(1-\alpha)\leqslant r\alpha^k.$$

So the iterates (3.2) converge to a point x^* in S. Since F is Lipschitzian in S, $\|V_k\|$ is uniformly bounded. Thus

$$\|F(x^*)\|=\lim_{k\to\infty}\|F(x^k)\|\leqslant\lim_{k\to\infty}\|V_k\|\,\|x^{k+1}-x^k\|=0,$$

i.e., $F(x^*)=0$. Suppose that $y^*\in S$ and $F(y^*)=0$. Let $V^*\in\partial F(x^*)$. Then

$$\begin{aligned}\|y^*-x^*\|&\leqslant\beta\|V^*(y^*-x^*)\|\\&\leqslant\beta\|V^*(y^*-x^*)-F'(x^*;y^*-x^*)\|\\&\quad+\beta\|F(y^*)-F(x^*)-F'(x^*;y^*-x^*)\|\\&\leqslant\beta(\delta+\gamma)\|y^*-x^*\|=\alpha\|y^*-x^*\|.\end{aligned}$$

This implies

$$\|y^*-x^*\|\leqslant 0,$$

i.e., $x^*=y^*$. This shows that x^* is the unique solution of (3.1) in S. Finally,

$$\begin{aligned}\|x^{k+p+1}-x^k\|&\leqslant\sum_{j=k}^{k+p}\|x^{j+1}-x^j\|\\&\leqslant\sum_{j=0}^{p}\alpha^{j+1}\|x^k-x^{k-1}\|\leqslant[\alpha/(1-\alpha)]\|x^k-x^{k-1}\|.\end{aligned}$$

Let $p\to\infty$, we have (3.4). This completes the proof. □

This theorem is an extension of the classical Newton–Kantorovich theorem [12]. In order that $\alpha < 1$, it needs that γ and δ are small. The smallness of γ and δ can be regarded as a global form of (2.8) and (2.17). The smallness of γ may be realized when the diameter of $\partial F(x)$ is small for all $x \in S$.

We now compare (3.2) with Newton's Method based upon B-derivatives.

According to Pang [13], a function $F: \mathbb{R}^n \to \mathbb{R}^n$ is said to be B-differentiable at a point x if there exists a function $BF(x): \mathbb{R}^n \to \mathbb{R}^n$, called the *B-derivative* of F at x, which is positively homogeneous of degree 1 (i.e. $BF(x)(th) = tBF(x)h$ for all $h \in \mathbb{R}^n$ and all $t \geqslant 0$), such that

$$\lim_{h \to 0} \frac{F(x+h) - F(x) - BF(x)h}{\|h\|} = 0. \tag{3.5}$$

In Robinson's original definition of B-differentiability [17] the positive homogeneity of the B-derivative was phrased in term of a cone property. In a finite-dimensional Euclidean space $\mathbb{R}^n$, Shapiro [25] showed that a locally Lipschitzian function F is B-differentiable at x if and only if it is directionally differentiable at x. In this case, the B-derivative and the directional derivative are identical, also see Harker and Xiao [7]. By (2.17), a semismooth function is B-differentiable.

The Newton Method using the B-derivative is defined by

$$x^{k+1} = x^k + d^k \tag{3.6}$$

and

$$F(x^k) + BF(x^k)d^k = 0. \tag{3.7}$$

In general, (3.7) is a system of nonlinear equations. It is a nontrivial task to solve it. Pang [13] suggested to solve it inexactly. There are other nonsmooth versions of Newton's Method. For example, see Kojima and Shindo [8] for Newton's Method for systems of piecewise continuously differentiable equations.

Proposition 3.4. *Suppose that F is locally Lipschitzian at x and any $V \in \partial F(x)$ is nonsingular. Then in a neighborhood of x, by suitable choices of $V_k \in \partial F(x^k)$, the iterates generated by* (3.6) *and* (3.7) *coincide with the iterates generated by* (3.2).

Proof. By above remarks (3.7) is identical with

$$F(x^k) + F'(x^k; d^k) = 0.$$

By Lemma 2.2, there exists $V_k \in \partial F(x^k)$ such that

$$F'(x^k; d^k) = V_k d^k.$$

According to Proposition 3.1, V_k is invertible when x^k is in an adequate neighborhood of x. This results in the conclusion of the proposition. $\square$

The B-derivative $BF(\cdot)$ is said Lipschitzian at x [13] if there exists a neighborhood $N(x)$ of x and a positive constant L such that for all $y \in N(x)$, $h \in \mathbb{R}^n$ with $\|h\| = 1$,

$$\|(BF(y) - BF(x))h\| \leq L\|y - x\|. \tag{3.8}$$

It is easy to see that, if $BF(\cdot)$ is Lipschitzian at x, then F is locally Lipschitzian at x.

Proposition 3.5. *If the B-derivative $BF(\cdot)$ is Lipschitzian at x, then F is 1-order semismooth at x.*

Proof. Suppose that (3.8) holds in $N(x)$. Let $y = x + h \in N(x)$. Let $V \in \partial F(x+h)$. By (2.1),

$$\partial F(x+h) \cdot h = \mathrm{co}\left\{ \lim_{\substack{x_i \to x \\ x_i + h \in D_F}} F'(x_i + h; h) \right\}.$$

Thus, there exists $x_i \to x$ such that $x_i + h \in D_F$ and

$$\lim_{i \to \infty} \|F'(x_i + h; h) - F'(x; h)\| \geq \|Vh - F'(x; h)\|. \tag{3.9}$$

By (3.8) and (3.9),

$$\|Vh - F'(x; h\| \leq L\|h\|^2.$$

This proves that F is 1-order semismooth at x. □

Pang [13] gave the following local convergence theorem for Newton's Method based upon B-derivatives:

Theorem 3.6 (Pang, 1989). *Let $F: \mathbb{R}^n \to \mathbb{R}^n$ be Lipschitzian and B-differentiable in a neighborhood of a solution x^* to* (3.1). *Suppose that the Fréchet derivative $F'(x^*)$ exists, and is strong and nonsingular. The, there exists a neighborhood $N(x^*)$ of x^* such that for any initial vector x^0 chosen from $N(x^*)$, the iterates x^k is well-defined by* (3.6) *and* (3.7), *remains in $N(x^*)$ and converges to x^*. Moreover, if $BF(\cdot)$ is Lipschitzian at x^*, then the rate of convergence is quadratic.* □

By Proposition 3.4, we see that Theorems 3.2 and 3.3 are still true for the iterates generated by (3.6) and (3.7) as long as (3.7) is solvable. By Corollary 2.5, the existence of the strong Fréchet derivative of F at x^* implies semismoothness of F at x^*. By Proposition 3.5, if $BF(\cdot)$ is Lipschitzian at x^* then F is 1-order semismooth at x^*. A semismooth function may have no Fréchet derivative at a point. Therefore, Theorem 3.2 is stronger than Pang's Theorem, i.e., Theorem 3.6 in a certain sense.

Kummer [9] discussed the convergence of (3.2) under the assumption that $\partial F(x^*)$ is single-valued, where x^* is the root. By the upper semicontinuity of ∂F (Proposition 2.6.2(c) of [4]), this assumption implies that F is semismooth at x^* but not vice versa. Therefore, our results are more general in a certain sense. On the other hand, Kummer gave a divergence example when F is Lipschitzian but $\partial F(x^*)$ is not single-valued. It is not difficult to check that the function in his example is not semismooth at the root. Thus, his pathological example does not contradict with our results.

4. The augmented Lagrangian

This example was suggested by Terry Rockafellar.

Consider the C^2-nonlinear program

$$\begin{aligned} \text{(NLP)} \quad & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x)=0, \quad i=1,2,\ldots,p, \\ &&& f_i(x)\leqslant 0, \quad i=p+1,p+2,\ldots,m, \end{aligned} \tag{4.1}$$

where $x\in\mathbb{R}^n$, $f_i\in C^2$ for $i=0,1,\ldots,m$. The augmented Lagrangian [20] of NLP is

$$L_r(x,y)=f_0(x)+\sum_{i=1}^{p} h(r_i,f_i(x),y_i)+\sum_{i=p+1}^{m}\phi(r,f_i(x),y_i),$$

where $r>0$,

$$h(r,f_i(x),y_i)=y_if_i(x)+\tfrac{1}{2}rf_i(x)^2,$$

$$\phi(r,f_i(x),y_i)=\begin{cases} y_if_i(x)+\frac{1}{2}rf_i(x)^2, & \text{if } y_i+rf_i(x)\geqslant 0,\\ -(1/2r)y_i^2, & \text{if } y_i+rf_i(x)\leqslant 0.\end{cases}$$

Here we prove the following result:

Theorem 4.1. *Under the above conditions, $L_r(\cdot,\cdot)\in C^1$, $\nabla L_r(\cdot,\cdot)$ is semismooth on "surfaces" where $y_i+rf_i(x)=0$ for $i=p+1,\ldots,m$, and smooth in other places.*

Proof. Let $i\in\{p+1,\ldots,m\}$ and $r>0$. Denote

$$\eta(x,s)\equiv\phi(r,f_i(x),s).$$

Then

$$\eta(x,s)=\begin{cases} sf_i(x)+\frac{1}{2}rf_i(x)^2, & \text{if } s+rf_i(x)\geqslant 0,\\ -(1/2r)s^2, & \text{if } s+rf_i(x)\leqslant 0.\end{cases}$$

It suffices to show that $\eta\in C^1$, $\nabla\eta$ is semismooth on the surface $s+rf_i(x)=0$ and smooth in other places. In fact,

$$\nabla\eta(x,s)=\begin{cases} \begin{pmatrix}(s+rf_i(x))\nabla f_i(x)\\ f_i(x)\end{pmatrix}, & \text{if } s+rf_i(x)\geqslant 0,\\ \begin{pmatrix}0\\ -s/r\end{pmatrix}, & \text{if } s+rf_i(x)\leqslant 0.\end{cases}$$

Thus, $\eta \in C^1$ and $\nabla \eta$ is locally Lipschitzian. Furthermore,

$$\nabla^2 \eta(x, s) = \begin{cases} \begin{pmatrix} r(\nabla f_i(x))(\nabla f_i(x))^{\mathrm{T}} + (s + rf_i(x))\nabla^2 f_i(x) & \nabla f_i(x) \\ (\nabla f_i(x))^{\mathrm{T}} & 0 \end{pmatrix}, & \text{if } s + rf_i(x) > 0, \\ \begin{pmatrix} 0 & 0 \\ 0 & -1/r \end{pmatrix}, & \text{if } s + rf_i(x) < 0. \end{cases} \tag{4.2}$$

This means that $\nabla \eta$ is smooth except on the surface $s + rf_i(x) = 0$.

Let $(\bar{x}, \bar{s})$ is on this surface, i.e., assume that

$$\bar{s} + rf_i(\bar{x}) = 0. \tag{4.3}$$

We now show that $\nabla \eta(\cdot, \cdot)$ is semismooth at $(\bar{x}, \bar{s})$. Let $(h, \alpha) \in \mathbb{R}^{n+1}$. It suffices to show that

$$\lim_{\substack{h' \to h, \alpha' \to \alpha \\ t \downarrow 0}} \left\{ V \begin{pmatrix} h' \\ \alpha' \end{pmatrix} : V \in \partial \nabla \eta(\bar{x} + th', \bar{s} + t\alpha') \right\} \tag{4.4}$$

exists. If

$$\bar{s} + t\alpha' + rf_i(\bar{x} + th') \tag{4.5}$$

keeps the same sign as $h' \to h$, $\alpha' \to \alpha$ and $t \downarrow 0$, then by (4.2) it is easy to see that (4.4) exists. Assume that (4.5) has different signs as $h' \to h$, $\alpha' \to \alpha$ and $t \downarrow 0$. Then there are $h_j \to h$, $\alpha_j \to \alpha$ and $t_j \downarrow 0$ such that

$$\bar{s} + t_j\alpha_j + rf_i(\bar{x} + t_jh_j) = 0. \tag{4.6}$$

By (4.3) and (4.6),

$$t_j\alpha_j + r[f_i(\bar{x} + t_jh_j) - f_i(\bar{x})] = 0,$$

i.e.,

$$\alpha_j + r[f_i(\bar{x} + t_jh_j) - f_i(\bar{x})]/t_j = 0.$$

Since $f_i \in C^2$, we have

$$\alpha + rf_i'(\bar{x}; h) = 0,$$

i.e.,

$$\alpha + r\nabla f_i(\bar{x})^{\mathrm{T}} h = 0. \tag{4.7}$$

Now consider (h', α', t). Let $V \in \partial \nabla \eta(\bar{x} + th', \bar{s} + t\alpha')$. If (4.5) is negative, by (4.2),

$$V = \begin{pmatrix} 0 & 0 \\ 0 & -1/r \end{pmatrix}.$$

Thus,

$$V \begin{pmatrix} h' \\ \alpha' \end{pmatrix} = \begin{pmatrix} 0 \\ -\alpha'/r \end{pmatrix} \to \begin{pmatrix} 0 \\ -\alpha/r \end{pmatrix}, \tag{4.8}$$

as $h'\to h$, $\alpha'\to\alpha$ and $t\downarrow 0$. If (4.5) is positive, by (4.2),

$$V=\begin{pmatrix} r(\nabla f_i(\bar{x}+th'))(\nabla f_i(\bar{x}+th'))^{\mathrm{T}}+(\bar{s}+t\alpha'+rf_i(\bar{x}+th'))\nabla^2 f_i(\bar{x}+th') & \nabla f_i(\bar{x}+th') \\ (\nabla f_i(\bar{x}+th'))^{\mathrm{T}} & 0 \end{pmatrix}.$$

Thus,

$$V\begin{pmatrix} h' \\ \alpha' \end{pmatrix}=\begin{pmatrix} r(\nabla f_i(\bar{x}+th'))(\nabla f_i(\bar{x}+th'))^{\mathrm{T}}h'+(\bar{s}+t\alpha'+rf_i(\bar{x}+th'))\nabla^2 f_i(\bar{x}+th')h'+\nabla f_i(\bar{x}+th')\alpha' \\ (\nabla f_i(\bar{x}+th'))^{\mathrm{T}}h' \end{pmatrix}$$

$$\to\begin{pmatrix} (\nabla f_i(\bar{x}))[r(\nabla f_i(\bar{x}))^{\mathrm{T}}h+\alpha]+(\bar{s}+rf_i(\bar{x}))\nabla^2 f_i(\bar{x})h \\ (\nabla f_i(\bar{x}))^{\mathrm{T}}h \end{pmatrix}=\begin{pmatrix} 0 \\ -\alpha/r \end{pmatrix}, \tag{4.9}$$

as $h'\to h$, $\alpha'\to\alpha$ and $t\downarrow 0$, where the last equality in (4.9) is due to (4.3) and (4.7). By (4.8) and (4.9), $V\binom{h'}{\alpha'}$ tends to the same limit in these two cases. When (4.5) is zero, by (2.1) and (4.2), V is a convex combination of the corresponding expressions in the above two cases. Hence, $V\binom{h'}{\alpha'}$ also tends to the same limit in this case. Thus, (4.4) exists. This shows that $\nabla\eta(\cdot,\cdot)$ is semismooth at $(\bar{x},\bar{s})$. The proof is completed. □

The basic step of the augmented Lagrangian method to solve (4.1) is to minimize $L_r(x, y)$. Thus, the extended Newton's method (3.2) and Theorem 3.2 can be applied here.

There are other examples of optimization problems where the objective functions are smooth with semismooth derivatives, but are not twice differentiable. For example, the extended linear-quadratic program, which arises from optimal control and stochastic programming, is such a problem. The primal and dual objective functions of the extended linear-quadratic programming problems are piecewise linear-quadratic and smooth, thus their derivatives are piecewise linear, hence semismooth. See [22–24].

For applications of nonsmooth versions of Newton's method in nonlinear complementarity, variational inequality and the Karush–Kuhn–Tucker system of nonlinear programming, see [13] and [14].

The extending Newton's method can also be used as a core step in interior point algorithms for convex programming. By using this step, the second-order continuity requirements commonly imposed by interior point methods on the objective function and the constraint are reduced to a semismooth condition on their gradients, while the total number of iterations remains the similar order to that of the smooth cases. See [16, 27] for such extensions.

Acknowledgments

We are grateful to Terry Rockafellar and Steve Robinson for their long-term help. In particular, Terry Rockafellar provided us the augmented Lagrangian example in Section 4 during his visit at Sydney; Steve Robinson provided us several references.

We also thank Peter Kall and Danny Ralph for references. In addition, the comments of Claude Lemarachel, the referees, and the associate editor were very helpful in revising this paper.

References

[1] J.V. Burke and L. Qi, "Weak directional closedness and generalized subdifferentials," *Journal of Mathematical Analysis and Applications* 159 (1991) 485-499.

[2] R.W. Chaney, "Second-order necessary conditions in constrained semismooth optimization," *SIAM Journal on Control and Optimization* 25 (1987) 1072-1081.

[3] R. W. Chaney, "Second-order necessary conditions in semismooth optimization," *Mathematical Programming* 40 (1988) 95-109.

[4] F.H. Clarke, *Optimization and Nonsmooth Analysis* (Wiley, New York, 1983).

[5] R. Correa and A. Jofre, "*Some properties of semismooth and regular functions in nonsmooth analysis*," in: L. Contesse, R. Correa and A. Weintraub, eds., *Proceedings of the IFIP Working Conference* (Springer, Berlin, 1987).

[6] R. Correa and A. Jofre, "Tangent continuous directional derivatives in nonsmooth analysis," *Journal of Optimization Theory and Applications* 6 (1989) 1-21.

[7] P.T. Harker and B. Xiao, "Newton's method for the nonlinear complementarity problem: A B-differentiable equation approach," *Mathematical Programming* 48 (1990) 339-357.

[8] M. Kojima and S. Shindo, "Extensions of Newton and quasi-Newton methods to systems of PC^1 Equations," *Journal of Operations Research Society of Japan* 29 (1986) 352-374.

[9] B. Kummer, "Newton's method for non-differentiable functions," in: J. Guddat, B. Bank, H. Hollatz, P. Kall, D. Karte, B. Kummer, K. Lommatzsch, L. Tammer, M. Vlach and K. Zimmermann, eds., *Advances in Mathematical Optimization* (Akademi-Verlag, Berlin, 1988) pp. 114-125.

[10] R. Mifflin, "Semismooth and semiconvex functions in constrained optimization," *SIAM J. Control and Optimization* 15 (1977) 957-972.

[11] R. Mifflin, "An algorithm for constrained optimization with semismooth functions," *Mathematics of Operations Research* 2 (1977) 191-207.

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

[13] J.S. Pang, "Newton's method for B-differentiable equations," *Mathematics of Operations Research* 15 (1990) 311-341.

[14] J.S. Pang, "A B-differentiable equation-based, globally, and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems," *Mathematical Programming* 51 (1991) 101-131.

[15] L. Qi, "Semismoothness and decomposition of maximal normal operators," *Journal of Mathematical Analysis and Applications* 146 (1990) 271-279.

[16] L. Qi and J. Sun, "A nonsmooth version of Newton's method and an interior point algorithm for convex programming," Applied Mathematics Preprint 89/33, School of Mathematics, The University of New South Wales (Kensington, NSW, 1989).

[17] S.M. Robinson, "Local structure of feasible sets in nonlinear programming, part III: stability and sensitivity," *Mathematical Programming Study* 30 (1987) 45-66.

[18] S.M. Robinson, "An implicit function theorem for B-differentiable functions," *Industrial Engineering Working Paper*, University of Wisconsin (Madison, WI, 1988).

[19] S.M. Robinson, "Newton's method for a class of nonsmooth functions," *Industrial Engineering Working Paper*, University of Wisconsin (Madison, WI, 1988).

[20] R.T. Rockafellar, "Proximal subgradients, marginal values, and augmented Lagrangians in nonconvex optimization," *Mathematics of Operations Research* 6 (1981) 427-437.

[21] R.T. Rockafellar, "Favorable classes of Lipschitz-continuous functions in subgradient optimization," in: E. Nurminski, ed., *Nondifferentiable Optimization* (Pergamon Press, New York, 1982) pp. 125-143.

[22] R.T. Rockafellar, "Linear-quadratic programming and optimal control," *SIAM Journal on Control and Optimization* 25 (1987) 781-814.

[23] R.T. Rockafellar, "Computational schemes for solving large-scale problems in extended linear-quadratic programming," *Mathematical Programming* 48 (1990) 447-474.

[24] R.T. Rockafellar and R.J-B. Wets, "Generalized linear-quadratic problems of deterministic and stochastic optimal control in discrete time," *SIAM Journal on Control and Optimization* 28 (1990) 810-822.

[25] A. Shapiro, "On concepts of directional differentiability," *Journal of Optimization Theory and Applications* 66 (1990) 477-487.

[26] J.E. Spingarn, "Submonotone subdifferentials of Lipschitz functions," *Transactions of the American Mathematical Society* 264 (1981) 77-89.

[27] J. Sun, "An affine-scaling method for linearly constrained convex programs," Preprint, Department of IEMS, Northwestern University (Evanston, IL, 1990).