Numer. Math. (1998) 80: 305–324

Numerische
Mathematik

Electronic Edition

Secant methods for semismooth equations

Florian A. Potra[1,⋆], **Liqun Qi**[2,⋆⋆], **Defeng Sun**[2,⋆⋆]

[1] Department of Mathematics, University of Iowa, Iowa City, IA 52242, USA; e-mail: potra@math.uiowa.edu
[2] School of Mathematics, The University of New South Wales, Sydney, New South Wales 2052, Australia; e-mail: (Liqun Qi) L.Qi@unsw.edu.au, (Defeng Sun) sun@alpha.maths.unsw.edu.au

Received October 16, 1996 / Revised version received July 25, 1997

Summary. Some generalizations of the secant method to semismooth equations are presented. In the one-dimensional case the superlinear convergence of the classical secant method for general semismooth equations is proved. Moreover a new quadratically convergent method is proposed that requires two function values per iteration. For the n-dimensional cases, we discuss secant methods for two classes of composite semismooth equations. Most often studied semismooth equations are of such form.

1. Introduction

The classical secant method is one of the most efficient algorithms for solving nonlinear equations. It has been used from the time of early Italian algebraists and has been extensively studied in the literature. It is well known that for smooth equations the classical secant method is superlinearly convergent with Q-order at least $(1+\sqrt{5})/2 = 1.618\ldots$ (cf. [27]). Since, with the exception of the first step, only one function value per step is used its efficiency index as defined by Ostrowski [13] is also $(1+\sqrt{5})/2$. The first generalization of the secant method for systems of two nonlinear equations goes back to Gauss (cf. Goldstine [7]). For different generalizations in the n-dimensional case see Ortega and Rheinboldt [12], Schwetlick [24], Dennis and Schnabel [3], Potra and Pták [18]. Newton-like methods based on finite difference approximations of the Jacobian can also be considered

⋆ The research of this author is supported by the Australian Research Council and the National Science Foundation under Grant DMS-9305760.
⋆⋆ The research of this author is supported by the Australian Research Council.

as generalized secant methods since they use only function values. Considering methods based only on function values is even more important in the non-smooth case since computation of generalized Jacobians [1] or B-differentials [19] may be very expensive. This paper represents an attempt to generalize the secant method to some important classes of semismooth equations.

In the third section of the present paper we make a complete analysis of the classical secant method for semismooth one-dimensional equations. We prove that the method retains superlinear convergence even in this case. More precisely, depending on the sign configuration of the lateral derivatives at the solution the secant method is either 2-step Q-quadratically convergent (if the lateral derivatives have different signs) or 3-step Q-quadratically convergent (if the lateral derivatives have the same sign). This implies that its R-order of convergence is either $\sqrt{2} = 1.4142\ldots$ or $\sqrt[3]{2} = 1.2599\ldots$ Thus its efficiency index in the sense of Ostrowski is at least $\sqrt[3]{2}$.

In section four we analyze a modification of the classical secant method that requires two function values per step and is Q-quadratically convergent both in the smooth and the semismooth case. The efficiency index of the method is at least $\sqrt{2}$ so that it is more efficient than the classical secant method in case the lateral derivatives at the solution are different but have the same sign. Moreover the distance between the iterates and the solution converges monotonically to zero (at least locally) which is not the case with the classical secant method where we only can guarantee that the distance between every third iterate and the solution converges monotonically to zero.

In section five we generalize the above mentioned method to the n-dimensional case for two classes of composite semismooth equations. The resulting method uses only function values to construct a special “finite difference approximation of the Jacobian” and is Q-quadratically convergent, the same as the generalizations of Newton’s method considered by [22,19, 15]. While these generalizations of Newton’s method require the computation of an element of the generalized Jacobian defined by Clarke [1] or of the B-differential considered by Qi [19] at each step, our method requires only computation of function values and therefore can be easily implemented.

Over the last couple of years, the superlinear convergence theory of the generalized Newton methods established in [22,19,15] has been extensively used in solving nonlinear complementarity problems, variational inequality problems, extended linear-quadratic programming, LC^1 optimization problems, etc. (see [21,2,4,5,14,28], and especially [10] for a recent survey). All these methods require computation of generalized Jacobian or B-differentials which is in general difficult. The secant methods presented in section three of the present paper can be extended to solve important subclasses of such problems as well. This subject will be treated in detail in a future paper.

2. Some properties of semismooth operators

In what follows we will review some results relevant to the concept of semismoothness. This concept was first introduced by Mifflin [11] for functionals. Convex functions, smooth functions, and piecewise linear functions are examples of semismooth functions. Products and sums of semismooth functions are still semismooth (see [11]). In [22], Qi and Sun extended the definition of semismooth functions to nonlinear operators of the form $F : \Re^n \to \Re^m$. We say that $F : \Re^n \to \Re^m$ is semismooth at x if F is locally Lipschitz at x and the following limit

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

exists for any $h \in \Re^n$, where ∂F is the generalized Jacobian defined by Clarke [1],

$$\partial F(x) = \mathrm{conv}\partial_B F(x),$$

where the B-differential $\partial_{\mathrm{B}} F(x)$ is defined as [19]:

$$\partial_B F(x) = \left\{ \lim F'(x^k) : \ x^k \to x, \ F \text{ is differentiable at } x^k \right\}.$$

Most nonsmooth equations arising in applications involve semismooth operators [15]. It was proved in [22] that if F is semismooth at x, then $F'(x;h)$, the directional derivative of F at x in direction h, exists and

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

Furthermore, the following lemma is proved:

Lemma 2.1 *[22] Suppose that $F : \Re^n \to \Re^m$ is semismooth at x. Then (i) if $h \to 0$ then for any $V \in \partial F(x+h)$ we have*

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

(ii) if $h \to 0$ then

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

We say that $F : \Re^n \to \Re^m$ is strongly semismooth at x if F is semismooth at x and for any $V \in \partial F(x+h)$, $h \to 0$,

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

Piecewise C^2 functions are examples of strongly semismooth functions.

Assume $n = m$. A generalized Newton method for solving the nonsmooth equation

$$F(x) = 0 \tag{2.1}$$

is defined by

$$x^{k+1} = x^k - V_k^{-1} F(x^k), \qquad V_k \in \partial F(x^k). \tag{2.2}$$

A particular case of (2.2) is

$$x^{k+1} = x^k - V_k^{-1} F(x^k), \qquad V_k \in \partial_{\mathrm{B}} F(x^k). \tag{2.3}$$

Suppose that x^* is a solution of (2.1). The generalized Newton method (2.2) ((2.3)) converges to x^* superlinearly in a neighborhood of x^* if F is semismooth at x^* and all $V \in \partial F(x^*)$ ($V \in \partial_{\mathrm{B}} F(x^*)$) are nonsingular; if the function F is strongly semismooth at x^*, then the convergence is quadratic. Although this superlinear convergence theory of the generalized Newton methods (2.2) and (2.3), first established in [22, 19, 15], is quite satisfactory, its practical implementation poses difficulties since the computation of generalized Jacobians may be quite time consuming in some applications. Therefore in the present paper we investigate iterative methods based only on function values. We will first consider the one-dimensional case, where the "generalized Jacobians" and the "B-differentials" have a simple form. Indeed, by using Lemma 2.1 we can easily prove the following result:

Lemma 2.2 *Suppose that $F : \Re \to \Re$ is semismooth at x. Then the lateral derivatives $F'(x+) = F'(x;1)$, $F'(x-) = F'(x;-1)$ exist and we have*

$$\partial_{\mathrm{B}} F(x) = \{F'(x+), F'(x-)\}, \partial F(x) = \mathrm{conv}\{F'(x+), F'(x-)\}.$$

Hence in order to apply a generalized Newton method we must be able to efficiently compute lateral derivatives. In what follows we will consider iterative methods where "generalized Jacobians" or "B-differentials" are replaced by divided differences of the form:

$$\delta F(x,y) = \frac{F(x) - F(y)}{x - y}. \tag{2.4}$$

If F is Lipschitz on an interval D containing x and y then F is differentiable almost everywhere according to Rademacher's Theorem, and by using the Lebesgue integral we obtain the following integral representation of the divided difference:

(2.5) $$\delta F(x,y) = \int_0^1 F'(sx + (1-s)y)ds.$$

The above representation will be used in the proof of the next lemma.

Lemma 2.3 *Suppose that F is semismooth at x^* and denote the lateral derivatives of F at x^* by*

(2.6) $$d^- = -F'(x^*;-1) \text{ and } d^+ = F'(x^*;1).$$

Then

(2.7) $$d^- - \delta F(u,v) = o(1) \quad \text{for all } u \uparrow x^*, v \uparrow x^*;$$

(2.8) $$d^+ - \delta F(u,v) = o(1) \quad \text{for all } u \downarrow x^*, v \downarrow x^*.$$

Moreover, if F is strongly semismooth at x^*, then

(2.9) $$d^- - \delta F(u,v) = O(|u-x^*| + |v-x^*|) \quad \text{for all } u,v < x^*;$$

(2.10) $$d^+ - \delta F(u,v) = O(|u-x^*| + |v-x^*|) \quad \text{for all } u,v > x^*.$$

Proof. From Lemma 2.1, for all $u \uparrow x^*$ and $v \uparrow x^*$ we have

$$d^- - \delta F(u,v) = -F'(x^*,-1) - \int_0^1 F'(tu + (1-t)v)dt$$

$$= \int_0^1 [F'(x^* + (tu + (1-t)v - x^*))(tu + (1-t)v - x^*)$$

$$-F'(x^*; tu + (1-t)v - x^*)]\tfrac{1}{x^* - tu - (1-t)v}dt$$

$$= o(1).$$

This proves (2.7). If F is strongly semismooth at x^*, then for all $u \uparrow x^*$ and $v \uparrow x^*$

$$d^- - \delta F(u,v) = -F'(x^*,-1) - \int_0^1 F'(tu + (1-t)v)dt$$

$$= \int_0^1 O(x^* - tu - (1-t)v)dt$$

$$= O(|u-x^*| + |v-x^*|).$$

This proves (2.9). Relations (2.8) and (2.10) are proved similarly. □

3. Secant methods for one-dimensional semismooth equations

With the divided difference defined in (2.4) the classical secant method can be written as

$$x^{k+1} = x^k - \delta F(x^k, x^{k-1})^{-1} F(x^k). \tag{3.1}$$

If $F : \Re \to \Re$ is smooth at a zero x^* and $F'(x^*) \neq 0$, then this method is superlinearly convergent in the sense that

$$|x^{k+1} - x^*| = o(|x^k - x^*|).$$

Moreover if F' is Lipschitz in a neighborhood of x^* then the Q-order of convergence of the classical secant method is $\frac{1+\sqrt{5}}{2}$ (see Traub [27]). However the convergence of this method is rather difficult to analyze in the nonsmooth case. In the first part of this section we will prove that the classical secant method is still R-superlinearly convergent under the semismoothness assumption but in general it is not Q-superlinearly convergent (for the notions of R-order and Q-order see [12,17]). In fact we will prove that the classical secant method is 3-step Q-superlinearly convergent for semismooth equations in the sense that

$$|x^{k+3} - x^*| = o(|x^k - x^*|).$$

This implies R-superlinear convergence of $\{x^k\}$ in the sense that

$$\sqrt[k]{|x^k - x^*|} = o(1).$$

The classical secant method (3.1) depends on two starting points. In order to simplify analysis we will consider a generic starting point and will take the other starting point of a special form. More precisely we are going to analyze the following iterative procedure depending on a generic starting point x^0.

Algorithm 3.1 *(Classical secant method)*

Step 1. Given $x^0 \in \Re^n$ *and* $\varepsilon \in (0, \infty)$*. Let* $x^{-1} = x^0 + \varepsilon |F(x^0)| F(x^0)$*.*
$k := 0$*.*
Step 2. Let

$$x^{k+1} = x^k - \delta F(x^k, x^{k-1})^{-1} F(x^k).$$

Step 3. $k := k + 1$*. Go to Step 2.*

Theorem 3.2 *Suppose that F is semismooth at a solution x^* of (2.1) and let d^- and d^+ be the lateral derivatives of F at x^*, as defined in (2.6). If d^- and d^+ are both positive (or negative), then there are two neighborhoods $\mathcal{U}$ and $\mathcal{V}$ of x^*, $\mathcal{U} \subseteq \mathcal{V}$, such that for each $x^0 \in \mathcal{U}$ the algorithm (3.1) is well defined and produces a sequence of iterates $\{x^k\}$ such that*

$$x^k \in \mathcal{V},\ k = -1, 0, 1, \dots$$

and $\{x^k\}$ converges to x^* 3-step Q-superlinearly. Furthermore, if

$$\alpha := \frac{|d^+ - d^-|}{\min\{|d^+|, |d^-|\}} < 1,$$

then $\{x^k\}$ is Q-linearly convergent with Q-factor α. If F is strongly semismooth at x^*, then $\{x^k\}$ converges to x^* 3-step Q-quadratically.

Proof. We only need to consider the case $0 < d^- \leq d^+$. The case $0 < d^+ \leq d^-$ may be discussed similarly. If both d^+ and d^- are negative, we may consider $G := -F$ instead of F. Since F is semismooth at x^* there is a convex neighbourhood $\mathcal{L}$ of x^* such that F is Lipschitz continuous on $\mathcal{L}$. We will construct the convex neighbourhoods $\mathcal{U}$ and $\mathcal{V}$ such that $x^* \in \mathcal{U} \subseteq \mathcal{V} \subseteq \mathcal{L}$. From the definition of x^{-1} and the Lipschitz continuity we deduce that $x^{-1} - x^* = x^0 - x^* + O(|x^0 - x^*|^2)$. Therefore for any $\mathcal{V} \subseteq \mathcal{L}$ we may choose $\mathcal{U}$ sufficiently small such that for all $x^0 \in \mathcal{U}$ we have

$$x^{-1} \in \mathcal{V} \text{ and } (x^0 - x^*)(x^{-1} - x^*) > 0. \tag{3.2}$$

According to Lemma 2.3, by choosing $\mathcal{V}$ small enough we can guarantee that

$$2\max\{d^+, d^-\} \geq |\delta F(x, y)| \geq 0.5 \min\{d^+, d^-\}, \tag{3.3}$$

whenever

$$x, y \in \mathcal{V} \text{ and } (x - x^*)(y - x^*) > 0.$$

Therefore according to (3.2) the algorithm is well defined for $k = 0$ for any $x^0 \in \mathcal{U}$.

Now we assume that $x^k, x^{k-1} \in \mathcal{V}$ and we consider the following three cases.

(i) $x^k, x^{k-1} > x^*$. Then according to (3.3) $\delta F(x^k, x^{k-1})$ is invertible and from Lemma 2.3 we obtain

$$\begin{aligned} x^{k+1} - x^* &= x^k - \delta F(x^k, x^{k-1})^{-1} F(x^k) - x^* \\ &= \delta F(x^k, x^{k-1})^{-1} [\delta F(x^k, x^{k-1}) - \delta F(x^k, x^*))](x^k - x^*) \\ &= (d^+ + o(1))^{-1} [d^+ + o(1) - (d^+ + o(1))](x^k - x^*) \\ &= o(x^k - x^*). \end{aligned}$$

(ii) $x^k, x^{k-1} < x^*$. Similarly to case (i), we can prove that

$$|x^{k+1} - x^*| = o(|x^k - x^*|).$$

(iii) $x^{k-1} < x^* < x^k$ or $x^k < x^* < x^{k-1}$. We will consider the first case. The latter may be treated similarly. As in case (i) we can write

$$x^{k+1} - x^* = \delta F(x^k, x^{k-1})^{-1}[\delta F(x^k, x^{k-1}) - \delta F(x^k, x^*))](x^k - x^*). \tag{3.4}$$

By applying Lemma 2.3 we obtain

$$\delta F(x^k, x^{k-1}) = \frac{F(x^k) - F(x^*) + F(x^*) - F(x^{k-1})}{x^k - x^{k-1}}$$

$$= \frac{d^+(x^k - x^*) + o(|x^k - x^*|) + d^-(x^* - x^{k-1}) + o(|x^{k-1} - x^*|)}{x^k - x^{k-1}}$$

$$= \lambda^k d^+ + (1 - \lambda^k)d^- + o(1),$$

where $\lambda^k = (x^k - x^*)/(x^k - x^{k-1})$. Hence for all k such that $x^{k-1} < x^* < x^k$ we have

$$d^- + o(1) \le \delta F(x^k, x^{k-1}) \le d^+ + o(1).$$

Such a relation also holds for all k such that $x^k < x^* < x^{k-1}$. Then from Lemma 2.3 and equation (3.4) we deduce that

$$|x^{k+1} - x^*| \le \frac{|d^+ - d^- + o(1)|}{\min\{d^-, d^+\} + o(1)}|x^k - x^*|, \tag{3.5}$$

and

$$\begin{aligned} &\delta F(x^k, x^{k-1})^{-1}[d^- - d^+ + o(1)](x^k - x^*) \\ &\le x^{k+1} - x^* \le \delta F(x^k, x^{k-1})^{-1}[o(1)](x^k - x^*). \end{aligned}$$

Therefore either $x^{k+1} < x^*$ or $x^{k+1} > x^*$ and in the latter case $|x^{k+1} - x^*| = o(|x^k - x^*|)$. We first consider the case $x^{k+1} < x^*$. In this case we have

$$\begin{aligned} &x^{k+2} - x^* \\ &= \delta F(x^{k+1}, x^k)^{-1}[\delta F(x^{k+1}, x^k) - \delta F(x^{k+1}, x^*))](x^{k+1} - x^*). \end{aligned} \tag{3.6}$$

Then from Lemma 2.3 and (3.6) we obtain in a similar manner that

$$|x^{k+2} - x^*| \le \frac{|d^+ - d^- + o(1)|}{\min\{d^-, d^+\} + o(1)}|x^{k+1} - x^*|, \tag{3.7}$$

and

$$\delta F(x^{k+1}, x^k)^{-1}[d^+ - d^- + o(1)](x^{k+1} - x^*)$$
$$\le x^{k+2} - x^* \le \delta F(x^{k+1}, x^k)^{-1}[o(1)](x^{k+1} - x^*).$$

Hence $x^{k+2} < x^*$ or $x^{k+2} > x^*$ and in the latter case $|x^{k+2} - x^*| = o(|x^{k+1} - x^*|)$. We first consider the case $x^{k+2} < x^*$. But now we have $x^{k+1}, x^{k+2} < x^*$. From (ii) and the above discussion we get

$$|x^{k+3} - x^*| = o(|x^{k+2} - x^*|) = o(|x^{k+1} - x^*|) = o(|x^k - x^*|).$$

If $x^{k+2}, x^{k+1} > x^*$, then from (i) and the above proof we obtain,

$$|x^{k+3} - x^*| = o(|x^{k+2} - x^*|) = o(|x^{k+1} - x^*|^2) = o(|x^k - x^*|^3).$$

If $x^{k+2} > x^*, x^{k+1} < x^*$ or $x^{k+2} < x^*, x^{k+1} > x^*$, then we deduce similarly that

$$|x^{k+3} - x^*| \le \frac{|d^+ - d^- + o(1)|}{\min\{d^+, d^-\} + o(1)} |x^{k+2} - x^*|.$$

So, if $x^{k+2} > x^*, x^{k+1} < x^*$, we have

$$|x^{k+3} - x^*| = O(|x^{k+2} - x^*|) = o(|x^{k+1} - x^*|) = o(|x^k - x^*|).$$

On the other hand, if $x^{k+2} < x^*, x^{k+1} > x^*$ then we deduce from (i) that

$$|x^{k+3} - x^*| = O(|x^{k+2} - x^*|) = o(|x^{k+1} - x^*|) = o(|x^k - x^*|^2).$$

Thus in case (iii), we have at least

(3.8) $$|x^{k+3} - x^*| = o(|x^k - x^*|).$$

If $\mathcal{U} = (x^* - \eta, x^* + \eta)$ and $\eta > 0$ is small enough then $x^0 \in \mathcal{U}$ and (3.8) implies

(3.9) $$x^{3k} \in \mathcal{U}, \quad k = 0, 1, 2, \ldots.$$

Also if η is small enough then from (3.5) it follows that

(3.10) $$|x^{k+1} - x^*| \le 2\alpha |x^k - x^*|, \quad k = 0, 1, 2, \ldots.$$

Finally from (3.9) and (3.10) it follows that

$$x^k \in \mathcal{V} = (x^* - \zeta, x^* + \zeta), \quad k = 0, 1, 2, \ldots,$$

where

$$\zeta = \max\{4\alpha^2, 1\}\eta.$$

Thus we have proved the 3-step Q-superlinear convergence of $\{x^k\}$.

Furthermore, if $\alpha < 1$ then we have

$$\limsup_k \frac{|x^{k+1} - x^*|}{|x^k - x^*|} \leq \alpha.$$

If F is strongly semismooth at x^*, we can prove that $\{x^k\}$ converges to x^* 3-step Q-quadratically by considering the above discussions and Lemma 2.3. □

In Theorem 3.2, we analyzed the convergence of the classical secant method under the assumption that d^-, d^+ have the same sign. In the next theorem we discuss the case where d^- and d^+ have different signs.

Theorem 3.3 *Suppose that F is semismooth at a solution x^* of (2.1). If d^- and d^+ do not vanish and have different signs, then there exist two neighborhoods $\mathcal{U}$ and $\mathcal{V}$ of x^*, $\mathcal{U} \subseteq \mathcal{V}$, such that for each $x^0 \in \mathcal{U}$ the algorithm (3.1) is well defined and produces a sequence $\{x^k\}$ such that*

$$x^k \in \mathcal{V},\ k = -1, 0, 1, \ldots$$

and $\{x^k\}$ converges to x^* 2-step Q-superlinearly. If F is strongly semismooth at x^*, then $\{x^k\}$ converges to x^* 2-step Q-quadratically.

Proof. We assume $d^- < 0$ and $d^+ > 0$. (The case $d^- > 0$ and $d^+ < 0$ can be treated similarly.) As in the proof of Theorem 3.2 we consider a convex neighbourhood $\mathcal{L}$ of x^* such that F is Lipschitz continuous on $\mathcal{L}$ and we will construct the convex neighbourhoods $\mathcal{U}$ and $\mathcal{V}$ such that $x^* \in \mathcal{U} \subseteq \mathcal{V} \subseteq \mathcal{L}$. By using the Lipschitz continuity of F it is easily seen that

$$\delta F(x, y) = \delta F(y, x^*) + o(1) \quad \text{whenever } x, y \in \mathcal{L} \text{ and } |x - x^*| = o(|y - x^*|). \tag{3.11}$$

Thus if $x^0 \in \mathcal{U}$ and if $\mathcal{U}$ is small enough then from Lemma 2.3 we deduce that $\delta F(x^0, x^{-1})$ does not vanish and therefore the iterate x^1 is well defined. Moreover for any $\mathcal{V} \subseteq \mathcal{L}$ we can choose $\mathcal{U}$ small enough such that $x^{-1} \in \mathcal{V}$.

Now let us assume that $x^k, x^{k-1} \in \mathcal{V}$ and let us consider the following three cases.

(i) $x^k, x^{k-1} > x^*$. From (3.11) and Lemma 2.3 we obtain

$$\delta F(x^k, x^{k-1}) = d^+ + o(1),$$

so that $\delta F(x^k, x^{k-1})^{-1}$ is well defined and we can write

$$\begin{aligned} x^{k+1} - x^* &= \delta F(x^k, x^{k-1})^{-1}[\delta F(x^k, x^{k-1}) - \delta F(x^k, x^*)](x_k - x^*) \\ &= (d^+ + o(1))^{-1}[d^+ + o(1) - (d^+ + o(1))](x^k - x^*) \\ &= o(x^k - x^*). \end{aligned}$$

(ii) $x^k, x^{k-1} < x^*$. As in case (i), we can prove that

$$x^{k+1} - x^* = o(x^k - x^*).$$

(iii) $x^{k-1} < x^* < x^k$ or $x^k < x^* < x^{k-1}$. We only consider the first case. The latter case may be treated similarly. According to the algorithm, we may choose x^0 such that $x^0, x^{-1} > x^*$ or $x^0, x^{-1} < x^*$. So (iii) does not occur for $k = 0$. Suppose that we are in case (iii) and that at the previous step we had either (i) or (ii). Then it is easily seen that

$$x^k - x^* = o(x^{k-1} - x^*)$$

and

$$\begin{aligned} x^{k+1} - x^* &= \delta F(x^k, x^{k-1})^{-1}[\delta F(x^k, x^{k-1}) - \delta F(x^k, x^*)](x^k - x^*) \\ &= [\delta F(x^{k-1}, x^*) + o(1)]^{-1} \\ &\qquad \times [\delta F(x^{k-1}, x^*) + o(1) - \delta F(x^k, x^*)](x^k - x^*) \\ &= (d^- + o(1))^{-1}[d^- - d^+ + o(1)](x^k - x^*). \end{aligned}$$

Therefore we have

$$x^{k+1} - x^* = O(x^k - x^*) \text{ and } x^{k+1} > x^*.$$

This shows that if case (iii) occurs at the k-th step and if at the previous step we had (i) or (ii) then case (iii) will not occur at the $(k+1)$-th step. Furthermore, since case (iii) does not occur for $k = 0$, we deduce that when case (iii) occurs at the k-th step then at the previous step we must have had case (i) or (ii). Hence according to the argument used in (i), we obtain

(3.12) $$x^{k+2} - x^* = o(x^{k+1} - x^*) = o(x^k - x^*).$$

If $\mathcal{U} = (x^* - \eta, x^* + \eta)$ and η is small enough then $x^0 \in \mathcal{U}$ and (3.12) implies

(3.13) $$x^{2k} \in \mathcal{U}, \quad k = 0, 1, 2, ...$$

Let

$$\beta := \frac{|d^- - d^+|}{\min\{|d^-|, |d^+|\}}.$$

Then if η is small enough from the discussion in (i)-(iii) it follows that

(3.14) $$|x^{k+1} - x^*| \le 2\beta |x^k - x^*|, \quad k = 0, 1, 2, ...$$

Finally from (3.13) and (3.14) it follows that

$$x^k \in \mathcal{V} = (x^* - \zeta, x^* + \zeta), \quad k = 0, 1, 2, ...,$$

Table 1. Iterates of secant method for (3.15)

i	x^i
1	-5.0761421$\times 10^{-5}$
2	-2.5126285$\times 10^{-5}$
3	1.2755427$\times 10^{-9}$
4	-1.2754773$\times 10^{-9}$
5	-4.2516638$\times 10^{-10}$
6	5.4229009$\times 10^{-19}$
7	-5.4229009$\times 10^{-19}$
8	-1.8076636$\times 10^{-19}$
9	9.8026179$\times 10^{-38}$
10	-9.8026179$\times 10^{-38}$
11	-3.2675393$\times 10^{-38}$
12	3.2 $\times 10^{-75}$

Table 2. Iterates of secant method for (3.16)

i	x^i
1	-5.0761421$\times 10^{-5}$
2	-7.6659448$\times 10^{-5}$
3	3.8918385$\times 10^{-9}$
4	1.1677298$\times 10^{-8}$
5	-4.5446157$\times 10^{-17}$
6	-6.8169236$\times 10^{-17}$
7	3.0980298$\times 10^{-39}$
8	9.2940895$\times 10^{-39}$
9	-2.8793367$\times 10^{-65}$

where

$$\zeta = \max\{2\beta, 1\}\eta.$$

Thus we have proved that the sequence $\{x^k\}$ is well defined and relation (3.12) shows that it converges to x^* 2-step Q-superlinearly.

If F is strongly semismooth at x^*, we may prove the 2-step Q-quadratic convergence by considering Lemma 2.3. □

We end this section by giving two numerical examples illustrating the order of convergence results obtained in Theorems 3.2 and 3.3. The computations were performed in *Mathematica* 3.0 with 80 digit precision, but only the first 8 significant digits are printed. For both examples we have

$$x^* = 0, x^0 = \frac{1}{200}, x^{-1} = \frac{1}{100}.$$

In Table 1 we list the iterates produced by the classical secant method for

$$F(x) = \begin{cases} x(x+1) & \text{if } x < 0 \\ -2x(x-1) & \text{if } x \geq 0 \end{cases} \tag{3.15}$$

In this case we have $d^- = 1$, $d^+ = 2$ and the numerical results show clearly that the sequence is 3-step Q-quadratically convergent.

Table 2 presents the iterates for

$$F(x) = \begin{cases} -x(x+1) & \text{if } x < 0 \\ -2x(x-1) & \text{if } x \geq 0 \end{cases} \tag{3.16}$$

Here $d^- = -1$, $d^+ = 1$, and the 2-step Q-quadratic convergence is evident.

4. A modified secant method for one-dimensional equations

Theorems 3.2 and 3.3 show that by using the classical secant method for strongly semismooth equations we obtain Q-quadratic convergence with at most 3 function evaluations per step. From the next lemma it follows that the classical secant method for strongly semismooth equations has R-order at least $\sqrt[3]{2}$ if $d^-d^+ > 0$, and $\sqrt{2}$ if $d^-d^+ < 0$.

Lemma 4.1 *If $\{x^k\}$ is a convergent sequence with limit x^* such that*

$$|x^{k+p} - x^*| = O(|x^k - x^*|^r),$$

then the R-order of convergence of $\{x^k\}$ is at least $\sqrt[p]{r}$.

Proof. The proof is straightforward by using the methods of Ortega and Rheinboldt [12] and Potra [17]. □

The next procedure uses only 2 function values per step and attains Q-superlinear (quadratic) convergence.

Algorithm 4.2 *(A modified secant method)*

Step 1. Given $x^0 \in \Re^n$ and $\varepsilon \in (0, \infty)$. Let $y^0 = x^0 + \varepsilon|F(x^0)|F(x^0)$. Compute $\delta F(x^0, y^0)$. $k := 0$.

Step 2. Set

$$x^{k+1} = x^k - \delta F(x^k, y^k)^{-1} F(x^k). \tag{4.1}$$

Step 3. $k := k + 1$. Set

$$y^k = x^k + \varepsilon|F(x^k)|F(x^k).$$

Go to step 2.

The above procedure is closely related to Steffensen's method [25]. An iterative procedure, very similar to Algorithm 4.2 was considered for general operator equations in Banach spaces in [16, formula (24)]. The above mentioned procedures are quadratically convergent in the smooth case. In the next theorem we prove that Algorithm 4.2 is quadratically convergent for strongly semismooth equations as well.

Theorem 4.3 *Suppose that x^* is a solution of (2.1) and F is semismooth at x^*. If d^- and d^+ are nonzero, then the iteration (4.1) is well defined in a neighborhood of x^* and converges to x^* Q-superlinearly. Furthermore, if F is strongly semismooth at x^*, the convergence is Q-quadratic.*

Proof. We may choose x^0 sufficiently close to x^* such that we have either $x^0, y^0 > x^*$ or $x^0, y^0 < x^*$. According to Lemma 2.3, (4.1) is well defined for $k = 0$. It is easy to see that

$$|y^k - x^k| = O(|F(x^k)|^2) = O(|x^k - x^*|^2).$$

Then from Lemma 2.3,

$$\delta F(x^k, y^k) = \delta F(x^k, x^*) + o(1) = d^+ + o(1) \text{ (or} = \ d^- + o(1)).$$

Therefore,

$$\begin{aligned} |x^{k+1} - x^*| &= |x^k - x^* - \delta F(x^k, y^k)^{-1} F(x^k)| \\ &\leq |\delta F(x^k, y^k)^{-1}||F(x^k) - F(x^*) - \delta F(x^k, y^k)(x^k - x^*)| \\ &\leq |\delta F(x^k, y^k)^{-1}||\delta F(x^k, x^*) - \delta F(x^k, y^k)||x^k - x^*| \\ &= o(|x^k - x^*|). \end{aligned} \tag{4.2}$$

This completes the proof of superlinear convergence of $\{x^k\}$. If F is strongly semismooth at x^*, we may prove similarly that $\{x^k\}$ converges to x^* Q-quadratically in a neighborhood of x^*. □

5. Composite semismooth equations

In the previous section we have discussed a modified secant method for one-dimensional semismooth equations. It appears difficult to extend it to general n-dimensional semismooth equations. However, most semismooth equations arising from concrete problems such as nonlinear complementarity problems and variational inequalities have a special structure which allows a generalization of the modified secant method considered above. In what follows we will show that such a generalization is possible when the operator $F : \Re^n \to \Re^n$ is of the form

$$F(x) = \Phi(G(x)) \tag{5.1}$$

or

$$F(x) = H(\Psi(x)), \tag{5.2}$$

where in (5.1) $\Phi : \Re^m \to \Re^n$ is nonsmooth but of special structure and $G : \Re^n \to \Re^m$ is continuously differentiable, while in (5.2) $H : \Re^p \to$

$\Re^n$ is continuously differentiable and $\Psi : \Re^n \to \Re^p$ is nonsmooth but of special structure. For example, given a continuously differentiable mapping $E : \Re^n \to \Re^n$, the nonlinear complementarity problem NCP(E) is to find $x \in \Re^n$ such that

$$x \geq 0, \quad E(x) \geq 0, \quad x^{\mathrm{T}} E(x) = 0.$$

Then NCP(E) is equivalent to finding a solution of the following (nonsmooth) equations:

$$F(x) = \min(x, E(x)) = 0 \tag{5.3}$$

or

$$F(x) = (\psi(x_1, E_1(x)), ..., \psi(x_n, E_n(x))) = 0, \tag{5.4}$$

where the operator $\min$ denotes the component-wise minimum of two vectors and $\psi : \Re^2 \to \Re$ is the Fischer-Burmeister function [6]:

$$\psi(a, b) = \sqrt{a^2 + b^2} - (a + b), \quad a, b \in \Re.$$

Both (5.3) and (5.4) are of the form (5.1). Solving NCP(E) is also equivalent to finding a solution of the nonsmooth equation

$$F(z) = E(z_+) + z - z_+ = 0, \tag{5.5}$$

in the sense that if x solves NCP(E) then $z := x - E(x)$ is a solution of (5.5), and, conversely, if z is a solution of (5.5) then $x := z_+$ solves NCP(E). The function F in (5.5) is clearly of the form (5.2). For more problems which can be written under the form (5.1) or (5.2), see [15] and [23].

Let us first give an analogue to Algorithm 4.2 for solving (2.1) with F of the form (5.1). Let $\varepsilon > 0$ be a given constant. Then for any x such that $F(x) \neq 0$ we define the matrix

$$V(x) = (V(x)_{ij}), \tag{5.6}$$

with

$$V(x)_{ij} = \frac{G_i(x + \varepsilon \|F(x)\| e_j) - G_i(x)}{\varepsilon \|F(x)\|}, \quad i = 1, ..., m, \ j = 1, ..., n, \tag{5.7}$$

where e_j is the jth unit vector of $\Re^n$.

With the above notation we define the following iterative procedure for solving equations of the form (5.1).

Algorithm 5.1 *(A secant method for $\Phi(G(x)) = 0$)*

Step 1. Given $x^0 \in \Re^n$. *Compute* $V_0 := V(x^0)$ *where* $V(x^0)$ *is defined as in (5.6) and (5.7) and set* $k := 0$.
Step 2. Compute an $U_k \in \partial_{\rm B}\Phi(y)|_{y=G(x^k)}$ *and set* $W_k = U_k V_k$.
Step 3. Set

$$x^{k+1} = x^k - W_k^{-1} F(x^k). \tag{5.8}$$

Step 4. $k := k+1$. *Compute* $V_k := V(x^k)$. *Go to Step 2.*

Before giving the convergence analysis of the above secant method, let us explain, by means of an example, how it works for operators F of the form (5.4). In this case $m = 2n$ and we have

$$G(x) = \begin{pmatrix} x \\ E(x) \end{pmatrix}, \quad V(x) = \begin{pmatrix} I \\ T(x) \end{pmatrix},$$

where $I \in \Re^{n\times n}$ is the identity matrix and

$$T(x)_{ij} = \frac{E_i(x + \varepsilon\|F(x)\|e_j) - E_i(x)}{\varepsilon\|F(x)\|},\ i = 1, ..., n, j = 1, ..., n.$$

Also

$$U(x) = (R(x), S(x)),$$

where $R(x)$, $S(x)$ are diagonal matrices. If $x_i^2 + E_i(x)^2 \neq 0$ then

$$R_{ii}(x) = \frac{x_i}{\sqrt{x_i^2 + E_i(x)^2}} - 1,\ S_{ii}(x) = \frac{E_i(x)}{\sqrt{x_i^2 + E_i(x)^2}} - 1.$$

If $x_i^2 + E_i(x)^2 = 0$ then we define

$$R_{ii}(x) = \alpha_i - 1,\ S_{ii}(x) = \beta_i - 1$$

for any $\alpha_i, \beta_i \in \Re$ such that $\alpha_i^2 + \beta_i^2 = 1$.

Note that we only need the value of $E(x)$ and we do not make any use of the derivative of $E(x)$.

Lemma 5.2 *Suppose that* x^* *is a solution of (2.1) and* F *is of the form (5.1) with* Φ *locally Lipschitz around* $G(x^*)$ *and* G *continuously differentiable.*

If all matrices $W \in \partial_{\rm B}\Phi(y^*)|_{y^*=G(x^*)}G'(x^*)$ *are nonsingular, then there exist a neighborhood* $N(x^*)$ *of* x^* *and a constant* C *such that all matrices* $W \in \partial\Phi(y)|_{y=G(x)}V(x)$, $x \in N(x^*)$ *are nonsingular and*

$$\|W^{-1}\| \le C.$$

Proof. The proof follows easily by noting that $\partial_{\rm B}\Phi(\cdot)$ is upper semi-continuous, the set $\partial_{\rm B}\Phi(y)|_{y=G(x)}$ is compact, and $V(x) \to G'(x^*)$ as $x \to x^*$. □

Theorem 5.3 *Suppose that x^* is a solution of (2.1) and F is of the form (5.1) with Φ semismooth at $G(x^*)$ and G continuously differentiable.*

If all matrices $W \in \partial_{\mathrm{B}}\Phi(y^*)|_{y^*=G(x^*)}G'(x^*)$ are nonsingular, then the iterative procedure (5.8) is well defined in a neighborhood of x^* and the sequence generated by it converges to x^* Q-superlinearly. Furthermore, if Φ is strongly semismooth at $G(x^*)$ and G' is Lipschitz continuous at x^*, then the convergence is Q-quadratic.

Proof. According to Lemma 5.2, formula (5.8) is well defined for $k := 0$. From Lemmas 2.1 and 5.2 we get

$$
\begin{aligned}
\|x^{k+1} - x^*\| &= \|x^k - W_k^{-1}F(x^k)\| \\
&\leq \|W_k^{-1}\|\|F(x^k) - F(x^*) - W_k(x^k - x^*)\| \\
&\leq \|W_k^{-1}\|(\|\Phi(G(x^k)) - \Phi(G(x^*)) - U_k(G(x^k) - G(x^*))\| \\
&\qquad + \|U_k(G(x^k) - G(x^*) - V_k(x^k - x^*))\|) \\
&= o(\|G(x^k) - G(x^*)\|) + o(\|x^k - x^*\|) \\
&= o(\|x^k - x^*\|, \qquad (5.9)
\end{aligned}
$$

which completes the proof of Q-superlinear convergence. In a similar manner we get Q-quadratic convergence under the strong semismoothness assumptions. □.

In order to construct a secant method for solving (2.1) when F is of the form (5.2), we consider the following matrix for any given constant $\varepsilon > 0$ and any x such that $F(x) \neq 0$:

$$U(x) = (U(x)_{ij}), \qquad (5.10)$$

with

$$U(x)_{ij} = \frac{H_i(\Psi(x) + \varepsilon\|F(x)\|e_j) - H_i(\Psi(x))}{\varepsilon\|F(x)\|}, \quad i = 1, ..., n, \ j = 1, ..., p, \qquad (5.11)$$

where e_j is the jth unit vector of $\Re^p$.

Algorithm 5.4 *(A secant method for $H(\Psi(x)) = 0$)*

Step 1. Given $x^0 \in \Re^n$. Compute $U_0 := U(x^0)$ as defined in (5.10) and (5.11). $k := 0$.
Step 2. Compute a $V_k \in \partial_{\mathrm{B}}\Psi(x^k)$ and Set $W_k = U_kV_k$.
Step 3. Set

$$x^{k+1} = x^k - W_k^{-1}F(x^k). \qquad (5.12)$$

Step 4. $k := k + 1$. Compute $U_k := U(x^k)$. Go to Step 2.

Similarly to Lemma 5.2 and Theorem 5.3, we have the following convergence results for the above secant method.

Lemma 5.5 *Suppose that x^* is a solution of (2.1) and F is of the form (5.2) with Ψ locally Lipschitz around x^* and H continuously differentiable.*

If all matrices $W \in H'(\Psi(x^*))\partial_{\mathrm{B}}\Psi(x^*)$ are nonsingular, then there exist a neighborhood $N(x^*)$ of x^* and a constant C such that all matrices $W \in U(x)\partial_{\mathrm{B}}\Psi(x)$, $x \in N(x^*)$ are nonsingular and

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

The proof of the above lemma is straightforward and therefore is omitted.

Theorem 5.6 *Suppose that x^* is a solution of (2.1) and F is of the form (5.2) with Ψ semismooth at x^* and H continuously differentiable.*

If all matrices $W \in H'(\Psi(x^*))\partial_{\mathrm{B}}\Psi(x^*)$ are nonsingular, then the iterative procedure (5.12) is well defined in a neighborhood of x^* and the sequence generated by it converges to x^* Q-superlinearly. Furthermore, if Ψ is strongly semismooth at x^* and H' is Lipschitz continuous at $\Psi(x^*)$, then the convergence is Q-quadratic.

Proof. According to Lemma 5.5, iteration (5.12) is well defined for $k = 0$. From Lemmas 2.1 and 5.5 we deduce that

$$\begin{aligned}
\|x^{k+1} - x^*\| &= \|x^k - x^* - W_k^{-1}F(x^k)\| \\
&\leq \|W_k^{-1}\|\|F(x^k) - F(x^*) - W_k(x^k - x^*)\| \\
&\leq \|W_k^{-1}\|(\|H(\Psi(x^k)) - H(\Psi(x^*)) - U_k(\Psi(x^k) - \Psi(x^*))\| \\
&\qquad + \|U_k(\Psi(x^k) - \Psi(x^*) - V_k(x^k - x^*))\|) \\
&= o(\|\Psi(x^k) - \Psi(x^*)\|) + o(\|x^k - x^*\|) \\
&= o(\|x^k - x^*\|, \qquad (5.13)
\end{aligned}$$

which completes the proof of Q-superlinear convergence. By a similar analysis, we can prove the Q-quadratic convergence when Ψ is strongly semismooth. □

6. Final remarks

In this paper we discuss several secant methods for semismooth equations. For one-dimensional equations, we prove that the classical secant method converges R-superlinearly under the standard assumptions used in the study of the generalized Newton method. A modification of the classical secant method requiring two function evaluations per step converges Q-superlinearly. For n-dimensional semismooth equations, we give secant methods for two classes of composite semismooth equations. At present, we cannot establish superlinear convergence for secant methods for solving general semismooth equations. Fortunately, in practice, most semismooth equations have the structure studied in this paper.

We note that for special composite semismooth equations, quasi-Newton methods have been discussed in [21, 26, 8]. By considering the discussion given in Sect. 5, we may generalize the quasi-Newton methods discussed in the above cited papers to composite semismooth equations of the form (5.1) or (5.2). For general nonsmooth equations, differentiability at the solution has to be assumed in order to ensure superlinear convergence of quasi-Newton methods as shown in [9, 10], [20].

References

1. Clarke, F.H. (1983) Optimization and Nonsmooth Analysis. Wiley, New York
2. De Luca, T., Facchinei, F., Kanzow, C. A semismooth equation approach to the solution of nonlinear complementarity problems. Mathematical Programming. to appear.
3. Dennis, J.E., Schnabel, R.B. (1983) Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice-hall, Englewood Cliffs, J., N.
4. Facchinei, F., Fischer, A., Kanzow, C. (1995) A semismooth Newton method for variational inequalities: The case of box constraints. Preprint 103, Institute of Applied Mathematics, University of Hamburg, Hamburg
5. Facchinei, F., Kanzow, C. A nonmonotone inexact Newton method for the solution of large-scale nonlinear complementarity problems. Mathematical Programming. to appear.
6. Fischer, A. (1992) A special Newton-type optimization method. Optimization, 24:269–284
7. Goldstine, H.H. (1997) A history of numerical analysis from the 16th through the 19th century. Springer, Berlin
8. Han, J., Sun, D. Newton and quasi-Newton methods for normal maps with polyhedral sets. Journal of Optimization Theory and Applications. to appear.
9. Ip, C.-M., Kyparysis, T. (1992) Local convergence of quasi-Newton methods for b-differentiable equations. Mathematical Programming, 56:71–89
10. Jiang, H., Qi, L., Chen, X., Sun, D. (1996) Semismoothness and superlinear convergence in nonsmooth optimization and nonsmooth equations. In: Di Pillo, G., Giannessi, F. (eds.), Nonlinear Optimization and Applications. Plenum Publishing Corporation, New York, pp. 197-212
11. Mifflin, R. (1977) Semismooth and semiconvex functions in constrained optimization. SIAM Journal on Control and Optimization, 15:959–972
12. Ortega, J.M., Rheinboldt, W.C. (1970) Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York
13. Ostrowski, A.M. (1973) Solution of Equations in Euclidean and Banach Spaces. Academic Press, New York
14. Outrata, J.V., Zowe, J. (1995) A Newton method for a class of quasi-variational inequalities. Computational Optimization and Applications, 4:5–21
15. Pang, J.-S., Qi, L. (1993) Nonsmooth equations: Motivation and algorithms. SIAM Journal on Optimization, 3:443–465

16. Potra, F.A. (1972) Processes of rough solution of operational equations (Romanian; French and English summaries). Buletin Ştiinţific Studenţesc, Timişoara, pages 123–132
17. Potra , F.A. (1989) On Q-order and R-order of convergence. Journal of Optimization Theory and Applications, 63:415–431
18. Potra, F.A., Pták, V. (1984) Nondiscrete induction and iterative processes. Number 103 in Research Notes in Mathematics. John Wiley & Sons, Boston–London–Melbourne
19. Qi, L. (1993) Convergence analysis of some algorithms for solving nonsmooth equations. Mathematics of Operations Research, 18:227–244
20. Qi, L. (1996) A note on superlinear convergence of quasi-Newton methods for nonsmooth equations. Applied Mathematics Report AMR 96/18, School of Mathematics, The University of New South Wales, Sydney 2052, Australia
21. Qi, L., Jiang, H. Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton methods and quasi-Newton methods for solving these equations. Mathematics of Operations Research. to appear.
22. Qi, L., Sun, J. (1993) A nonsmooth version of Newton's method. Mathematical Programming, 58:353–368
23. Ralph, D., Sholtes, S. Sensitivity analysis of composite piecewise smooth equations. Mathematical Programming. to appear.
24. Schwetlick, H. (1979) Numerische Lösung nichtlinearer Gleichungen. VEB, Berlin
25. Steffensen, J. (1933) Remarks on iteration. Skand. A**ktuarietidskr.** , 16:64–72
26. Sun, D., Han, J. Newton and quasi-Newton methods for a class of nonsmooth equations and related problems. SIAM Journal on Optimization. to appear.
27. Traub, J. (1964) Iterative methods for the solution of equations. Prentice-Hall, Englewood Cliffs,
28. Yamashita, N., Fukushima, M. Modified Newton methods for solving semismooth reformulations of monotone complementarity problems. Mathematical Programming. to appear.