



# **An approximate Newton method for solving nonsmooth equations with infinite max functions**

Marek J. Śmietański

## **► To cite this version:**

Marek J. Śmietański. An approximate Newton method for solving nonsmooth equations with infinite max functions. International Journal of Computer Mathematics, 2011, <10.1080/00207160.2010.541451>. <hal-00687735>

**HAL Id: hal-00687735**

**<https://hal.science/hal-00687735v1>**

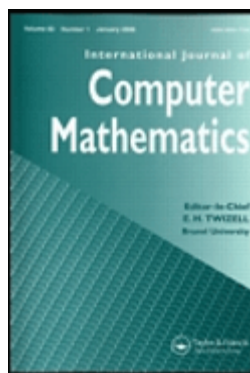
Submitted on 14 Apr 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



**An approximate Newton method for solving nonsmooth equations with infinite max functions**

Journal:	<i>International Journal of Computer Mathematics</i>
Manuscript ID:	GCOM-2008-0568-B.R2
Manuscript Type:	Original Article
Date Submitted by the Author:	02-Nov-2009
Complete List of Authors:	Śmietański, Marek; University of Lodz, Faculty of Mathematics and Computer Science
Keywords:	nonsmooth equation, semismooth function, weak consistently approximated Jacobian, difference approximation, superlinear convergence



An approximate Newton method for solving nonsmooth equations  
with infinite max functions

MAREK J. ŚMIETAŃSKI

Faculty of Mathematics and Computer Science,  
University of Lodz, Poland  
90-238 Lodz, ul.Banacha 22  
tel: (48)(42)635-58-77, fax: (48)(42)635-42-66  
e-mail: smietan@math.uni.lodz.pl

In this paper, a new version of approximate Newton method for solving nonsmooth equations with infinite max function is presented. This method uses a difference approximation of the generalized Jacobian based on a weak consistently approximated Jacobian. Numerical example is reported for the generalized Newton method using two versions of approximation.

*Keywords:* Nonsmooth equation; Semismooth function; Weak consistently approximated Jacobian; Difference approximation; Superlinear convergence.

*2000 Mathematics Subject classification codes:* 65H10, 90C33

*CCS:* G.1.5, G.1.2, F.2.1

## 1 Introduction

Many important practical problems of mathematical programming require the solving of nonlinear system of equations

$$F(x) = 0, \quad (1)$$

where  $F : R^n \rightarrow R^n$  is locally Lipschitz.

The generalized Jacobian method for solving such systems was proposed by Qi and Sun in [13] in the form

$$x^{(k+1)} = x^{(k)} - V_k^{-1} F(x^{(k)}), \quad V_k \in \partial F(x^{(k)}), \quad (2)$$

where  $\partial F(x^{(k)})$  is the generalized Jacobian of the function  $F$  at  $x^{(k)}$ , defined by Clarke in [5], and a matrix  $V_k$  is taken arbitrarily from  $\partial F(x^{(k)})$ . The iteration generated by (2) is locally superlinearly convergent under the assumption of semismoothness of the function  $F$ .

Xu and Chang in [16] expanded method (2), replacing an iteration matrix with an adequately defined consistently approximated Jacobian, to avoid the complicated evaluations of  $V_k$ . Xu and Chang proved the superlinear convergence of the method under the assumption of semismoothness. Other versions of the difference approximation were proposed also by Śmietański in [14].

In this paper we present the generalized Newton method for solving some new class of nonlinear equations. We construct the difference approximation method for solving problems with functions defined by the operator infinite maximum, i.e. the considered function  $F$  has the following form

$$F(x) = \max \{f_1(x), f_2(x), \dots\},$$

where  $f_p : R^n \rightarrow R^n$ ,  $p = 1, 2, \dots$  and 'max' denotes the componentwise maximum operator. We obtain again that the method is locally superlinearly convergent under the assumption of semismoothness of  $F$ . The iterations are based on some approximation of a weak consistently approximated Jacobian, which was introduced by Śmietański in [15]. Earlier, Kummer in [8] presented sufficient and necessary conditions for the convergence of Newton method based on

the generalized derivative (in particular the generalized Jacobian). But our approach permits to determine an approximation of generalized Jacobian in some special cases and involves the proof of superlinear convergence.

In Section 2 we give some needed good-known notions. In Section 3 we present an approximate Newton method for solving nonsmooth equations and describe a consistently approximated Jacobian, a weak consistently approximated Jacobian and their properties. In Section 4 we remind the finite difference approximation for the generalized Jacobian and propose a superlinearly convergent approximate method for solving nonsmooth equations with infinite max functions.

*Notation:* Throughout the paper, we use the following notation.  $R^n$  denotes the vector space of  $n$ -tuples with the Euclidean norm  $\|\cdot\|$  and  $L(R^n)$  the matrix space of  $n \times n$  real matrices with the induced norm  $\|\cdot\|$ .  $S(x, \delta)$  denotes an open ball in  $R^n$  with center  $x$  and radius  $\delta$ ,  $N(x, \delta)$  is a closed ball in  $R^l$  with center  $x$  and radius  $\delta$ . If  $a_j$ ,  $j = 1, \dots, n$  are vectors of  $R^n$ ,  $[a_1, \dots, a_n]$  will be a matrix whose  $j$ -th column is  $a_j$ , and if  $A_j$ ,  $j = 1, \dots, n$  are subsets of  $R^n$ , then  $[A_1, \dots, A_n]$  will be a set of matrices in the form  $[a_1, \dots, a_n]$ , where  $a_j \in A$  for all  $j = 1, \dots, n$ .

## 2 Preliminaries

Throughout this work, we assume that the functions  $f : R^n \rightarrow R$  and  $F : R^n \rightarrow R^n$  are locally Lipschitz.

Let  $S$  be some set with Lebesgue's measure 0 in  $R^n$ . The generalized gradient of  $f$  at  $x$  is

$$\partial f(x) = \text{co}\{\lim \nabla f(x^{(k)}) : x^{(k)} \rightarrow x, x^{(k)} \notin S, x^{(k)} \in D_f\},$$

where  $D_f$  denotes the set of points at which  $f$  is differentiable. Clarke in [4] and [5] showed that if  $f$  is  $C^1$  or differentiable at  $x$  and regular, then

$$\partial f(x) = \{\nabla f(x)\}. \quad (3)$$

Moreover, Clarke in [5] proved among the others that  $\partial f(x)$  is a compact subset of  $R^n$  and upper semicontinuous.

The following theorem describes the generalized gradient of a function defined by means of the infinite max operator:

**Theorem 1** (Clarke [4], Theorem 2.1) *Let  $U$  be a sequentially compact space, and let  $g : R^n \times U \rightarrow R$  has the following properties:*

- (a)  $g(x, u)$  is upper semicontinuous in  $(x, u)$ ;
  - (b)  $g$  is locally Lipschitz in  $x$ , uniformly for  $u \in U$ ;
  - (c)  $g_x^\circ(x, u; h) = g'_x(x, u; h)$  for each  $h \in R^n$ ;
  - (d)  $\partial_x g(x, u)$  is upper semicontinuous in  $(x, u)$ .
- Then, if we let  $f(x) = \max\{g(x, u) : u \in U\}$ ,*
- (i)  $f$  is locally Lipschitz;

- (ii)  $f'(x; h)$  exists;  
 (iii)  $f'(x; h) = f^\circ(x; h) = \max \{ \langle \xi, h \rangle : \xi \in \partial_x g(x, u), u \in M(x) \}$ , where  $M(x) = \{ u \in U : g(x, u) = f(x) \}$ ;  
 (iv)  $\partial f(x) = \text{co} \{ \partial_x g(x, u) : u \in M(x) \}$ .

**Remark:** (Clarke [4], Remark after Theorem 2.1) Hypotheses (b), (c), (d) from the above theorem will follow if  $g$  is convex in  $x$ . Moreover, the conditions on  $g$  also hold when  $\nabla_x g(x, u)$  exists and is continuous in  $(x, u)$ . Thereby, the above theorem is a generalization of a theorem originally due to Danskin [6].

According to Rademacher's Theorem, the local Lipschitz continuity of  $F$  implies that  $F$  is differentiable almost everywhere. Let  $D_F$  be the set of points at which  $F$  is differentiable. The generalized Jacobian of  $F : R^n \rightarrow R^n$  at  $x$  is

$$\partial F(x) = \text{co} \{ \lim JF(x^{(k)}) : x^{(k)} \rightarrow x, x^{(k)} \in D_F \},$$

where  $JF(x^{(k)})$  denotes the usual Jacobian of  $F$  at  $x^{(k)}$ . Clarke in [5] proved that

$$\partial F(x) \subset [\partial F^1(x), \dots, \partial F^n(x)]^T, \quad (4)$$

where the Cartesian product on the right side of the above inclusion denotes the set of all matrices, whose  $i$ -th row belongs to  $\partial F^i(x)$  for every  $i$ . If  $n = 1$ , then  $\partial F(x) = \partial F^1(x)$  (i.e. the generalized gradient and the generalized Jacobian coincide).

Following Chen [2], we will denote the set of matrices, whose rows are the generalized gradients of components  $F^i$  of the function  $F$  at  $x$  as  $\partial_C F(x)$ , i.e.  $\partial_C F(x) = [\partial F^1(x), \dots, \partial F^n(x)]^T$ . Hiriart-Urruty in [7] stated, that the equality in (4) holds at least in two particular cases. First, when all functions  $F^i$  except possibly one, e.g.  $F^1$ , are  $C^1$  at  $x$ , then  $\partial F(x) = [\partial F^1(x), \nabla F^2(x), \dots, \nabla F^n(x)]^T$ . The second case holds when the component functions are nondifferentiable with respect to nonrelated variables, e.g.  $F^i$  is nondifferentiable only with respect to variable  $x_i$  for every  $i = 1, \dots, n$ .

The generalized Jacobian of  $F$  with respect to the  $j$ -th variable (the partial generalized Jacobian) at  $x$  is

$$\partial_{x_j} F(x) = \text{co} \{ \lim \nabla_{x_j} F(x^{(k)}) : x^{(k)} \rightarrow x, x^{(k)} \in D_F \},$$

where  $\nabla_{x_j} F(x^{(k)})$  denotes the vector of the usual partial derivatives with respect to the  $j$ -th variable at  $x^{(k)}$ . From the definitions of the generalized Jacobian and the partial generalized Jacobian we may easily obtain the following inclusion:  $\partial F(x) \subset [\partial_{x_1} F(x), \dots, \partial_{x_n} F(x)]$ , where the Cartesian product on the right side denotes the set of all matrices, whose  $i$ -th column belongs to  $\partial_{x_i} F(x)$  for every  $i$  (see Hiriart-Urruty [7] - comment to the definition of the generalized partial Jacobian after Theorem 2.1).

We will say that  $\partial F(x)$  (or  $\partial_C F(x)$ ) is nonsingular if all  $V \in \partial F(x)$  ( $V \in \partial_C F(x)$ , respectively) are nonsingular.

**Lemma 2** If  $\partial_C F(x)$  is nonsingular, then there are constants  $\delta > 0$  and  $C_1 > 0$  such that for any  $y \in S(x, \delta)$ ,  $\partial_C F(y)$  is nonsingular and  $\|V^{-1}\| \leq C_1$  for any  $V \in \partial_C F(y)$ ,  $y \in S(x, \delta)$ .

**Proof.** Since  $\partial_C F(x)$  is nonsingular, then there exists  $\varepsilon > 0$  sufficiently small such that  $\partial_C F(x) + \varepsilon B$  is nonsingular, where  $B$  is unit ball in matrix space  $R^{n \times n}$ . By the upper semicontinuity of  $\partial_C F$ , for every  $\varepsilon > 0$ , there exists  $\delta > 0$  such that

$$\partial_C F(y) \subset \partial_C F(x) + \varepsilon B \text{ for all } y \in S(x, \delta).$$

■

The semismoothness was introduced for functionals by Mifflin in [9]. The following definition is presented after Qi and Sun [13].  $F$  is *semismooth* at  $x$  if  $F$  is locally Lipschitz at  $x$  and

$$\lim_{V \in \partial F(x+th'), h' \rightarrow h, t \downarrow 0} Vh'. \quad (5)$$

exists for any  $h \in R^n$ . Clearly, convex functions and smooth functions are semismooth. Scalar products and sums of semismooth functions are semismooth functions (see Mifflin [9]). The assumption (5) implies that the Hadamard directional derivative exists (see Qi and Sun [13]):

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

**Lemma 3** (Qi, Sun [13], Theorem 2.3) *The following statements are equivalent:*

- (i)  $F$  is semismooth at  $x$ ;
- (ii)  $Vh - F'(x; h) = o(\|h\|)$  for  $V \in \partial F(x + h)$ ,  $h \rightarrow 0$ ;
- (iii)  $\lim_{h \rightarrow 0} \frac{F'(x+h; h) - F'(x; h)}{\|h\|} = 0$ .

**Remark:** Chen in [2] suggested that the statement (ii) of Lemma 3 holds also for any  $V \in \partial_C F(x + h)$ , which is obvious because the semismoothness of  $F$  at  $x$  is equivalent to the semismoothness of each of its components  $F^i$  at  $x$ , what follows by Corollary 2.4 from [13] and Lemma 3(iii).

Additionally, Qi and Sun [13] remarked that if  $F$  is semismooth at  $x$ , then for any  $h \rightarrow 0$

$$F(x + h) - F(x) - F'(x; h) = o(\|h\|). \quad (6)$$

From (6) it follows that a semismooth function is B-differentiable.

### 3 Approximate Newton method

First, we introduce an approximate Newton iteration. Let us consider the version of Newton method in which the next point is computed by

$$x^{(k+1)} = x^{(k)} - J(x^{(k)}, s^{(k)})^{-1} F(x^{(k)}), \quad (7)$$

where  $J(x^{(k)}, s^{(k)}) \in L(R^n)$  is some approximation of the generalized Jacobian of  $F$  and  $s^{(k)}$  is an  $l$ -dimensional vector of parameters.

The iteration (7) is a generalization of the smooth approximate Newton method, which was established by Ortega and Rheinboldt in [11]. The construction of the matrix  $J(x, s)$  is here a principal difficulty. Two versions of possible constructions of  $J(x, s)$  were presented in [14].

Now, we present main results discussed in [16] and [14], which are necessary to prove the superlinear convergence. First, we have to remind a notion of CAJ introduced by Xu and Chang in [16]:

**Definition 4** Let  $J : D \times D_l \subset R^n \times R^l \rightarrow L(R^n)$ . If  $0 \in R^l$  is a limiting point of  $D_l$ , and if

$$\lim_{s \rightarrow 0} \text{dist}[J(x, s); \partial F(x)] = 0,$$

uniformly with respect to  $x \in D$ , then  $J(x, s)$  is called a consistently approximated Jacobian of  $F$  in  $D$  (CAJ).

Here,

$$\text{dist}[J(x, s); \partial F(x)] = \min_{V \in \partial F(x)} \|J(x, s) - V\|.$$

A simple approximate method for solving nonsmooth equations is based on a consistently approximated Jacobian. However, to allow the solving a nonsmooth equations with function defined by the operator maximum, the weak consistently approximated Jacobian has been introduced in [15].

**Definition 5** Let  $J : D \times D_l \subset R^n \times R^l \rightarrow L(R^n)$ . If  $0 \in R^l$  is a limiting point of  $D_l$ , and if

$$\lim_{s \rightarrow 0} \text{dist}[J(x, s); \partial_C F(x)] = 0, \quad (8)$$

uniformly with respect to  $x \in D$ , then  $J(x, s)$  is called a weak consistently approximated Jacobian of  $F$  in  $D$  (WCAJ).

Here, again

$$\text{dist}[J(x, s); \partial_C F(x)] = \min_{V \in \partial_C F(x)} \|J(x, s) - V\|.$$

**Remark:** (i) Let us notice that equality (8) in the above definition is similar to the Jacobian consistence property given by Chen, Qi and Sun in [3] in the form

$$\lim_{\varepsilon \downarrow 0} \text{dist}[(\nabla_x f(x, \varepsilon))^T; \partial_C F(x)] = 0 \text{ for any } x \in R^n,$$

where the B-subdifferential is approximated by the Jacobian of some  $C^1$  function  $f$  defined on  $R^n \times (0, +\infty)$ , called a smooth approximation of  $F$ .

(ii) If  $J(x, s)$  is a CAJ of  $F$  in  $D$ , then  $J(x, s)$  is a WCAJ of  $F$  in  $D$  (it follows obviously by Definitions 4, 5 and inclusion (4)).

The following lemma, theorem and corollary were presented in [15] to prove a convergence an approximate Newton method for nonsmooth equations with finite max function.



**Lemma 6** Let  $x^* \in R^n$ . Suppose that  $F$  is semismooth at  $x^*$  and that  $J(x, s)$  is a WCAJ of  $F(x)$  in a neighborhood of  $x^*$ . Then

$$\|F'(x^* + h, h) - J(x^* + h, s)h\| \leq w(h, s) \|h\|,$$

where  $\lim_{h \rightarrow 0, s \rightarrow 0} w(h, s) = 0$ .

**Theorem 7** Let  $x^* \in D$  be a solution of (1). Suppose that  $F$  is semismooth at  $x^*$  and that  $\partial_W F(x^*)$  is nonsingular. Let  $J : D \times D_l \subset R^n \times R^l \rightarrow L(R^n)$  be a WCAJ of  $F$  in  $D$  and assume that 0 is a limiting point of  $D_l$ . Then, there exist  $\delta > 0$  and  $\gamma > 0$  such that  $J(x, s)$  is nonsingular for  $x \in S(x^*, \delta)$ ,  $s \in D_l \cap N(0, \gamma)$ . Furthermore, the function

$$G(x, s) := x - J(x, s)^{-1}F(x) \quad (9)$$

is contractive in the sense that  $\|x^* - G(x, s)\| \leq u(x, s) \|x - x^*\|$  for all  $x \in D \cap S(x^*, \delta)$ ,  $s \in D_l \cap N(0, \gamma)$ , where  $\lim_{x \rightarrow x^*, s \rightarrow 0} u(x, s) = 0$ .

**Corollary 8** Suppose that the assumptions of Theorem 7 are satisfied. Let  $G(x, s)$  be defined by (9) and let  $g : R^n \rightarrow R^l$  be a continuous function with  $g(x^*) = 0$ . If  $s = g(x)$ , then there exists  $\delta > 0$  such that  $G(x, g(x))$  is contractive in  $S(x^*, \delta)$ . Thereby, the sequence produced by (7) converges to  $x^*$  superlinearly for a sufficiently good starting point.

## 4 Practical method for equations with max function

Now we will present some version of the difference approximation which offers a practical construction of a weak consistently approximated Jacobians for non-smooth equations with maximum function. Clearly, the existence of a WCAJ, which ensure convergence of the method, depends on the local properties of  $F$  (e.g. semismoothness of  $F$  and nonsingularity of the generalized Jacobian of  $F$ ). First, we remind a notation for the problem.

Assume that  $F$  in (1) has the following form:

$$F(x) = \max \{F_1(x), \dots, F_m(x)\}, \quad (10)$$

where  $F_p : R^n \rightarrow R^n$ ,  $p = 1, \dots, m$ , and "max" denotes the componentwise maximum operator.

Let  $I(x)$  denotes the active index set at  $x$ , i.e. the set  $\{i : F(x) = f_i(x)\}$ , and  $A_p^i$  - the effective region of an active smooth piece of function  $F_i$ , i.e. the set  $\{x : F_i(x) = f_{p,i}(x)\}$ .

**Remark:** If all components  $F_p$  are continuously differentiable, then  $F$  is a  $PC^1$  function. Some important useful property of such functions has been studied by Pang and Ralph in [12]: for  $PC^1$  function  $F$ , the convex hull of the set  $\{\nabla f(x) : i \in \tilde{I}(x)\}$  is equal to the generalized Jacobian of  $F$  at  $x$ , where  $\tilde{I}(x) = \{i : x \in \text{clint}\{x \in V : i \in I(z)\}\}$  and  $V$  is a neighborhood of  $x$ .

**Example:** An absolute value function is a simplest example of finite max function, which arises in practice. Really

$$|x| = \max \{x, -x\}.$$

For nonlinear equations with function  $F$  in the form (10) we may construct a weak consistently approximated Jacobian in the following way:

$$[J(x, s)]_{i,j} = \begin{cases} \frac{F_1^i(x+s_j e_j) - F_1^i(x)}{s_j} & \text{for } x \in A_1^i, \\ \dots \\ \frac{F_p^i(x+s_j e_j) - F_p^i(x)}{s_j} & \text{for } x \in A_p^i \text{ and } x \notin \bigcup_{k=1}^{p-1} A_k^i, \\ \dots \\ \frac{F_m^i(x+s_j e_j) - F_m^i(x)}{s_j} & \text{for } x \in A_m^i \text{ and } x \notin \bigcup_{k=1}^{m-1} A_k^i, \end{cases} \quad (11)$$

where  $i, j \in \{1, \dots, n\}$ ,  $s_i > 0$  for  $j = 1, \dots, n$ . Additionally, assume that  $x + s_j e_j \in A_p^i$  for suitable  $p$ . Because  $F$  is a  $PC^1$  function, then the considered points belong to the closure of the effective region of a considered active smooth piece. Otherwise one can always construct a  $PC^1$  function which is active at  $x$  but not active at any neighborhood of  $x$  (see Pang and Ralph [12]).

The following lemma permits to formulate the theorem about the superlinear convergence of the sequence (7) with a WCAJ defined by (11).

**Lemma 9** ([15], Lemma 17) Suppose that  $F_p$ ,  $p = 1, \dots, m$  are  $C^1$  or differentiable and regular. Then, for  $J(x, s)$  defined by (11), the following inclusion holds:

$$\lim_{s \downarrow 0} J(x, s) \in \partial_C F(x). \quad (12)$$

**Theorem 10** Let  $x^*$  be a solution of (1) with  $F$  in the form (10). Assume that  $F$  is semismooth in a neighborhood of  $x^*$  and that  $\partial F(x^*)$  is nonsingular. Suppose that, for  $J(x, s)$  defined by (11), the inclusion (12) holds uniformly with respect to  $x$  in a neighborhood of  $x^*$ . Let  $g: R^n \rightarrow R^l$  be a continuous function for which  $g(x^*) = 0$ . Then, for  $s = g(x)$ , there exists  $\delta > 0$  such that, if the starting point  $x_0$  belongs to  $S(x^*, \delta)$ , then the sequence  $\{x^{(k)}\}$  generated by (7) is well defined and converges to  $x^*$  superlinearly.

**Proof.** It follows from Lemma 9 that  $J(x, s)$  is a weak consistently approximated Jacobian of function  $F$  in some neighborhood of  $x^*$  under the assumption that (12) holds uniformly with respect to  $x$  in a neighborhood of  $x^*$  for  $s \in D_l \cap N(0, \gamma)$ , where  $\gamma$  is takeie, as in Theorem 7. The rest of proof follows directly from Corollary 8. ■

**Remark:** Theorem 10 states the convergence properties of a nonsmooth version of the classical approximate Newton method (the generalized Jacobian method). The standard difference approximations are only suitable for a restricted class of problems. Therefore, in the sequel we will introduce other techniques which allow to solve more complicated problems.

Many of the convex function, which arises in practical problems, are maximum value functions of the form

$$F(x) = \max \{ \Phi(x, t) : t \in \mathcal{T} \} \text{ for } x \in R^n,$$

where  $\mathcal{T}$  is a closed subset of  $R^m$  and  $\Phi(x, t)$  and its first- and second-order partial derivatives with respect to components of  $x$  are continuous on  $R^n \times \mathcal{T}$ . Such functions have been considered e.g. by Mifflin and Sagastizábal in [10]. Moreover, there is nonsmooth infinite max-operator in optimization programs which arise in automatic control applications for  $H_\infty$  controller synthesis (see e.g. Bompart, Noll and Apkarian, [1]). So, assume that the function  $F$  in (1) has the following form

$$F(x) = \max_{y \in Y} \{ \Phi(x, y) \}, \quad (13)$$

where  $\Phi : R^n \times Y \rightarrow R^m$ ,  $Y$  is some compact set in  $R^m$ , and "max", as previously, denotes the componentwise maximum operator.

For equation with infinite max it is possible to construct a difference approximation in the form

$$[J_1(x, s)]_{i,j} = \frac{\Phi^i(x + s_j e_j, y) - \Phi^i(x, y)}{s_j}, \quad (14)$$

where  $i, j \in \{1, \dots, n\}$ ,  $s_j > 0$  for  $j = 1, \dots, n$ , and  $y$  is an arbitrarily selected element of  $M^i(x) = \{y \in Y : F^i(x) = \Phi^i(x, y)\}$ . Now,  $A_y^i$  denotes the set  $\{x : F^i(x) = \Phi^i(x, y)\}$  for some  $y \in Y$ . Assume that  $x + s_j e_j \in A_y^i$  for suitable  $y \in Y$ .

**Lemma 11** Suppose that  $\nabla_x \Phi(x, y)$  exists and is continuous in  $(x, y)$ . Then, for  $J_1(x, s)$  defined by (14), the following inclusion holds:

$$\lim_{s \downarrow 0} J_1(x, s) \in \partial_C F(x).$$

**Proof.** The approximation (14) is well defined, because we suppose that  $x + s_j e_j \in A_y^i$  for suitable  $y \in Y$ . Since  $\Phi$  is directionally differentiable for each direction  $h \in R^n$ , we have, for  $y \in M^i(x)$ ,

$$\lim_{s \downarrow 0} [J_1(x, s)]_{i,j} = \lim_{s \downarrow 0} \frac{\Phi^i(x + s_j e_j, y) - \Phi^i(x, y)}{s_j} = (\Phi^i)'_x(x, y; e_j) = \langle \nabla_x \Phi^i(x, y); e_j \rangle.$$

Since  $\Phi^i$  are differentiable, by remark after Theorem 1 and (3), we obtain

$$\partial F^i(x) = \text{co} \{ \nabla_x \Phi^i(x, y) : y \in M^i(x) \}.$$

Then  $\lim_{s \downarrow 0} [[J_1(x, s)]_{i,1}, \dots, [J_1(x, s)]_{i,n}] = [\langle \nabla_x \Phi^i(x, y); e_1 \rangle, \dots, \langle \nabla_x \Phi^i(x, y); e_n \rangle] = \nabla_x \Phi^i(x, y)$  for  $x \in A_y^i$ ,  $y \in M^i(x)$ . All these facts give us that

$$\lim_{s \downarrow 0} [[J_1(x, s)]_{i,1}, \dots, [J_1(x, s)]_{i,n}] \in \partial F^i(x) \text{ for } i = 1, \dots, n,$$

and we obtain the desired conclusion. ■

Clearly, similarly as in finite maximum, the above lemma allows to formulate and to prove a suitable theorem about convergence of approximate Newton method (7) based on difference approximation.

Approximation (14) of the weak consistently approximated Jacobian uses the difference quotients based on two points (i.e. the one-sided approximation). In a similar way, it is possible to construct the approximation of a WCAJ based on three points (i.e. the central approximation), which gives better numerical results in some cases:

$$[J_2(x, s)]_{i,j} = \frac{\Phi^i(x + s_j e_j, y) - \Phi^i(x - s_j e_j, y)}{2s_j}, \quad (15)$$

where  $i, j \in \{1, \dots, n\}$ ,  $s_j > 0$  for  $j = 1, \dots, n$ , and  $y$  is an arbitrarily selected element of  $M^i(x) = \{y \in Y : F^i(x) = \Phi^i(x, y)\}$ . For the above approximations we have a lemma analogous to the previous one:

**Lemma 12** Suppose that  $\nabla_x \Phi(x, y)$  exists and is continuous in  $(x, y)$ . Then, for  $J_2(x, s)$  defined by (15), the following inclusion holds:

$$\lim_{s \downarrow 0} J_2(x, s) \in \partial_C F(x).$$

**Proof.** For every  $j \in \{1, \dots, n\}$  we have

$$\begin{aligned} \frac{\Phi^i(x + s_j e_j, y) - \Phi^i(x - s_j e_j, y)}{2s_j} &= \\ &= \frac{\Phi^i(x + s_j e_j, y) - \Phi^i(x, y) + \Phi^i(x, y) - \Phi^i(x - s_j e_j, y)}{2s_j} = \\ &= \frac{1}{2} \left[ \frac{\Phi^i(x + s_j e_j, y) - \Phi^i(x, y)}{s_j} + \frac{\Phi^i(x, y) - \Phi^i(x - s_j e_j, y)}{s_j} \right]. \end{aligned}$$

Since  $\Phi$  is directionally differentiable for each direction  $h \in R^n$ , we have, for  $y \in M^i(x)$

$$\lim_{s_j \downarrow 0} \frac{\Phi^i(x + s_j e_j, y) - \Phi^i(x, y)}{s_j} = (\Phi^i)'_x(x, y; e_j), \quad j \in \{1, \dots, n\}.$$

and

$$\lim_{s_j \downarrow 0} \frac{\Phi^i(x, y) - \Phi^i(x - s_j e_j, y)}{s_j} = -(\Phi^i)'_x(x, y; -e_j), \quad j \in \{1, \dots, n\},$$

Therefore, similarly as in the previous lemma, the differentiability of  $\Phi^i$ , remark after Theorem 1, (3) and a convexity of the generalized Jacobian (Proposition 2.6.2(i) in [5]) give the conclusion. ■

## 5 Examples

The approximate Newton method presented in this paper has been implemented in DevC++ to test the numerical behavior of algorithm. All computations **have been** performed in double precision, which has the absolute range  $[5.0 \cdot 10^{-324}, 1.7 \cdot 10^{308}]$  and gives the accuracy  $15 \div 16$  significant digits. The termination criterion was

$$\|x^{(k+1)} - x^{(k)}\| \leq \varepsilon_1 \text{ or } \|F(x^{(k)})\| \leq \varepsilon_2,$$

where  $\varepsilon_1 = 1.0E - 08$  and  $\varepsilon_2 = 1.0E - 12$ . In the tables containing the results,  $x^{(0)}$  denotes the starting point,  $N_{it}$  - the number of iterations needed to reach the desired precision,  $\Delta_{\tilde{x}}$  - an absolute error of last approximation of solution and  $\tilde{x}$  is the exact solution, to which the iteration sequence was convergent.

### Example 1.

Consider the equation (1) with the following infinite max function

$$F(x) = \max_{n \in N} \left\{ \frac{-n \cdot \sin(x/n)}{x} \right\}.$$

The equation  $F(x) = 0$  has four solutions  $x = k\pi, k \in \{-2, -1, 1, 2\}$ . The problem was solved by the generalized Jacobian method for infinite max functions with a WCAJ approximated by (14) and (15). We have tested the method with  $s = 1.0E - 5, 1.0E - 10$ ,  $|F(x)|$  (in the latter case  $s$  decreases to 0, similarly as in the classical Steffensen method). The Tables 1 and 2 contain the results of the computations. Depending on the starting point, we notice the convergence to different solutions. For all initial points we have obtained the given solution with the precision at least  $1.0E - 09$ . Usage (15) allows us to start the computations far from a solution. Unfortunately, parameter  $s$  can not be too large.

If a similar equation with cosine has been solved then the results were very similar despite the fact that equation has only one solution  $x = \pi$ .

### Example 2.

Consider the system of equations with the infinite max functions

$$F(x) = \max_{n \in N} \{\Phi(x, n)\}$$

where

$$\Phi(x, n) = \begin{bmatrix} \sqrt{\frac{n}{1+n}} x_1^2 + \frac{2\sqrt{n}}{1+n} x_2 \\ -n|x_1| + |x_2| \end{bmatrix}.$$

This equation has only one solution  $x^* = (0, 0)$ . We have tested the method with  $s = 1.0E - 10$ ,  $|F(x)|$ . The Tables 3 and 4 contain the results of the computations. If we used the approximation (14), we obtained the convergence to the solution almost for all tested starting points. However, a degenerate behavior appeared for some cases, especially for too large  $s$ . Using (15) we notice troubles with choice of a starting points to ensure the convergence to the solution.

## 6 Conclusions

Newton-like methods are important tools for solving nonsmooth equations. They are useful also when the evaluation of the Jacobian matrix is difficult or impossible (at least analytically) for smooth equations. As stated above, we have developed some version of approximate Newton method for solving nonsmooth equations with infinite max function. We have proved that under mild assumptions the sequence generated by the method is locally and superlinearly convergent for semismooth equations.

The performance of the method is evaluated in the terms of the number of iterations required. The analysis of the numerical results seems indicate that the method is promising for solving semismooth systems because the tested equations were solved satisfactorily. The experimental results also indicate that the method has at least superlinear convergence. The detailed analysis of the results for Example 2 lets state that the convergence of method was even almost quadratic for some initial points. So, the important conclusion is that the algorithm allows us to find various solution of equation, however if there are many solutions of the problem then the relationship between the starting point and the obtained solution is unpredictable. It is easy to observe in Tables 1 and 4 that choice of parameter  $s$  has not relevant effect on speed of convergence in terms of the number of iterations. In Table 3 some degenerate behavior of the method can be noticed for some initial points, if we use not sufficiently small constant parameter  $s$  in an approximation of WCAJ. Moreover, the method with the approximation based on central formula (14) is sensitive because to the choice of the parameter and the starting point.

## References

- [1] Bompart, V., Noll, D., Apkarian, P.: Second-order nonsmooth optimization for  $H_\infty$  synthesis. 5th IFAC Symposium on Robust Control Design, Toulouse (2006)
- [2] Chen, X.: Superlinear convergence of smoothing quasi-Newton methods for nonsmooth equations. J. Comput. Appl. Math. 80, 105-126 (1997)
- [3] Chen, X., Qi, L., Sun, D.: Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities. Math. Comp. 67, 519-540 (1998)
- [4] Clarke, F.H.: Generalized gradients and applications. Trans. Amer. Math. Soc. 205, 247-262 (1975)
- [5] Clarke, F.H.: Optimization and nonsmooth analysis, New York, Wiley 1983
- [6] Danskin Jr, J.M.: The theory of max-min and its application to weapons allocation problems, New York, Springer Verlag (1967)

[7] Hiriart-Urruty, J.-B.: Refinements of necessary optimality conditions in nondifferentiable programming I. *Appl. Math. Optim.* 5, 63-82 (1979)

[8] Kummer B.: Newton's method based on generalized derivatives for nonsmooth functions: convergence analysis. In: Oettli, W. et al. (eds.): *Advances in Optimization*, Berlin, Springer-Verlag 1992, 171-194

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

[10] Mifflin, R., Sagastizábal, C.: On  $\mathcal{WU}$ -theory for functions with primal-dual gradient structure. *SIAM J. Optim.* 11, 547-571 (2000)

[11] Ortega, J.M., Rheinboldt, W.C.: *Iterative solution of nonlinear equations in several variables*, New York, Academic Press (1970)

[12] Pang, J.-S., Ralph, D.: Piecewise smoothness, local invertibility and parametric analysis of normal maps. *Math. Oper. Res.* 21, 401-426 (1996)

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

[14] Śmietański, M.J.: New versions of approximate Newton method for solving nonsmooth equations, Ph.D. Thesis, Faculty of Mathematics, University of Lodz 1999.

[15] Śmietański, M.J.: An approximate Newton method for nonsmooth equations with finite max functions. *Numer. Algorithms* 41, 219-238 (2006)

[16] Xu, H., Chang, W.: Approximate Newton methods for nonsmooth equations. *J. Optim. Theory Appl.* 93, 373-394 (1997)

Table 1. Results obtained for Example 1 using (14).

	$s = 10^{-5}$		$s = 10^{-10}$			$s =  F(x) $		
$x^{(0)}$	$N_{it}$	$\Delta_{\tilde{x}}$	$N_{it}$	$\Delta_{\tilde{x}}$	$\tilde{x}$	$N_{it}$	$\Delta_{\tilde{x}}$	$\tilde{x}$
-15.0	7	$10^{-13}$	7	$10^{-14}$	$-2\pi$	6	$10^{-14}$	$-\pi$
-10.0	9	$10^{-14}$	8	$10^{-14}$	$-\pi$	7	$10^{-10}$	$\pi$
-5.0	5	$10^{-13}$	5	$10^{-14}$	$-2\pi$	5	$10^{-12}$	$-2\pi$
-2.0	4	$10^{-12}$	4	$10^{-13}$	$-\pi$	4	$10^{-13}$	$-\pi$
-1.0	5	$10^{-9}$	5	$10^{-9}$	$-\pi$	5	$10^{-16}$	$-2\pi$
-0.5	4	$10^{-13}$	4	$10^{-14}$	$-2\pi$	—	—	—
-0.1	6	$10^{-12}$	6	$10^{-13}$	$\pi$	6	$10^{-14}$	$2\pi$
0.1	6	$10^{-13}$	6	$10^{-14}$	$-\pi$	5	$10^{-16}$	$2\pi$
0.5	7	$10^{-11}$	4	$10^{-14}$	$2\pi$	5	$10^{-9}$	$\pi$
1.0	5	$10^{-9}$	5	$10^{-9}$	$\pi$	4	$10^{-13}$	$\pi$
2.0	4	$10^{-12}$	4	$10^{-13}$	$\pi$	4	$10^{-14}$	$\pi$
5.0	8	$10^{-13}$	5	$10^{-14}$	$2\pi$	5	$10^{-16}$	$2\pi$
10.0	7	$10^{-10}$	—	—	$2\pi$	7	$10^{-12}$	$-\pi$
15.0	20	$10^{-13}$	7	$10^{-15}$	$2\pi$	6	$10^{-15}$	$2\pi$

Table 2. Results obtained for Example 1 using (15).

	$s = 10^{-5}$		$s = 10^{-10}$			$s =  F(x) $		
$x^{(0)}$	$N_{it}$	$\Delta_{\tilde{x}}$	$N_{it}$	$\Delta_{\tilde{x}}$	$\tilde{x}$	$N_{it}$	$\Delta_{\tilde{x}}$	$\tilde{x}$
-30.0	5	0.0	5	$10^{-15}$	$\pi$	5	$10^{-16}$	$\pi$
-20.0	—	—	—	—	—	6	$10^{-13}$	$2\pi$
-15.0	looped	—	7	$10^{-14}$	$-2\pi$	7	$10^{-12}$	$-2\pi$
-10.0	—	—	11	$10^{-12}$	$2\pi$	7	$10^{-9}$	$2\pi$
-5.0	—	—	5	$10^{-13}$	$-2\pi$	5	$10^{-14}$	$-2\pi$
-2.0	4	$10^{-12}$	4	$10^{-12}$	$-\pi$	4	$10^{-13}$	$-\pi$
-1.0	5	$10^{-9}$	5	$10^{-9}$	$-\pi$	6	$10^{-13}$	$-\pi$
-0.5	looped	—	4	$10^{-13}$	$-2\pi$	5	$10^{-14}$	$-2\pi$
-0.1	6	$10^{-13}$	6	$10^{-13}$	$\pi$	—	—	—
0.1	6	$10^{-13}$	6	$10^{-13}$	$-\pi$	—	—	—
0.5	looped	—	4	$10^{-13}$	$2\pi$	5	$10^{-14}$	$2\pi$
1.0	5	$10^{-9}$	5	$10^{-9}$	$\pi$	6	$10^{-13}$	$\pi$
2.0	4	$10^{-12}$	4	$10^{-12}$	$\pi$	4	$10^{-13}$	$\pi$
5.0	looped	—	5	$10^{-13}$	$2\pi$	5	$10^{-14}$	$2\pi$
10.0	—	—	11	$10^{-12}$	$-2\pi$	7	$10^{-9}$	$-2\pi$
15.0	looped	—	7	$10^{-14}$	$2\pi$	7	$10^{-12}$	$2\pi$
20.0	—	—	—	—	—	6	$10^{-13}$	$-2\pi$
30.0	5	0.0	5	$10^{-15}$	$-\pi$	5	$10^{-16}$	$-\pi$



Table 3. Results obtained for Example 2 using (14).

$x^{(0)}$	$s = 10^{-5}$		$s = 10^{-10}$		$s =  F(x) $	
	$N_{it}$	$\Delta_{\tilde{x}}$	$N_{it}$	$\Delta_{\tilde{x}}$	$N_{it}$	$\Delta_{\tilde{x}}$
$(-5.0, -5.0)$	193	$10^{-09}$	looped	—	31	$10^{-16}$
$(-2.0, -2.0)$	51	$10^{-16}$	looped	—	47	$10^{-19}$
$(-0.75, -0.75)$	3459	$10^{-09}$	looped	—	15	$10^{-21}$
$(-0.5, -0.5)$	807	$10^{-10}$	looped	—	47	$10^{-17}$
$(-0.25, -0.25)$	5	$10^{-10}$	5	$10^{-17}$	5	$10^{-18}$
$(-0.1, -0.1)$	4	$10^{-14}$	4	$10^{-17}$	5	$10^{-16}$
$(0.1, 0.1)$	4	$10^{-11}$	4	$10^{-12}$	6	$10^{-24}$
$(0.25, 0.25)$	2976	$10^{-09}$	looped	—	124	$10^{-18}$
$(0.5, 0.5)$	6	$10^{-14}$	6	$10^{-17}$	7	$10^{-20}$
$(0.75, 0.75)$	9	$10^{-15}$	8	$10^{-18}$	8	$10^{-18}$
$(1.0, 1.0)$	4586	$10^{-15}$	looped	—	37	$10^{-16}$
$(2.0, 2.0)$	2436	$10^{-13}$	looped	—	—	—
$(5.0, 5.0)$	3552	$10^{-10}$	looped	—	39	$10^{-17}$
$(-0.5, 0.5)$	7	$10^{-10}$	7	$10^{-10}$	7	$10^{-16}$
$(-1.0, 1.0)$	669	$10^{-10}$	looped	—	67	$10^{-21}$

Table 4. Results obtained for Example 2 using (15).

$x^{(0)}$	$s = 10^{-10}$		$s =  F(x) $	
	$N_{it}$	$\Delta_{\tilde{x}}$	$N_{it}$	$\Delta_{\tilde{x}}$
$(-0.75, -0.75)$	looped	—	looped	—
$(-0.5, -0.5)$	looped	—	looped	—
$(-0.25, -0.25)$	5	$10^{-17}$	5	$10^{-16}$
$(-0.1, -0.1)$	4	$10^{-17}$	5	$10^{-17}$
$(0.1, 0.1)$	4	$10^{-10}$	6	$10^{-17}$
$(0.25, 0.25)$	looped	—	looped	—
$(0.5, 0.5)$	6	$10^{-17}$	7	$10^{-17}$
$(0.75, 0.75)$	8	$10^{-10}$	8	$10^{-17}$
$(1.0, 1.0)$	looped	—	looped	—
$(2.0, 2.0)$	looped	—	looped	—
$(-0.5, 0.5)$	7	$10^{-10}$	7	$10^{-17}$