Applications of Mathematics

Yan Gao; Xuewen Li

Nonsmooth equations approach to a constrained minimax problem

Applications of Mathematics, Vol. 50 (2005), No. 2, 115–130

Persistent URL: http://dml.cz/dmlcz/134596

NONSMOOTH EQUATIONS APPROACH TO A CONSTRAINED MINIMAX PROBLEM*

Yan Gao, Shanghai, Xuewen Li, Beijing

(Received May 9, 2003, in revised version November 23, 2003)

Abstract. An equivalent model of nonsmooth equations for a constrained minimax problem is derived by using a KKT optimality condition. The Newton method is applied to solving this system of nonsmooth equations. To perform the Newton method, the computation of an element of the b-differential for the corresponding function is developed.

Keywords: nonsmooth optimization, nonsmooth equations, minimax problems, Newton methods, KKT systems, quasidifferential calculus

MSC 2000: 90C47, 65H10

1. Introduction

Consider the nonlinear programming problem of the form

$$(\mathrm{P}_1)\qquad \begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && g_i(x) \leqslant 0, \quad i = 1, \ldots, m, \\ & && h_j(x) = 0, \quad j = 1, \ldots, l \end{aligned}$$

where $f, g_i, h_j\colon \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, $j = 1, \ldots, l$, are continuously differentiable. Under some constraint conditions, the problem (P_1) is equivalent to the following Karush-Kuhn-Tucker (KKT) system:

$$(1.1)\qquad \begin{cases} \nabla f(x) + \sum\limits_{i=1}^{m} u_i \nabla g_i(x) + \sum\limits_{j=1}^{l} v_j \nabla h_j(x) = 0, \\ u_i g_i(x) = 0, \quad i = 1, \ldots, m, \quad h_j(x) = 0, \quad j = 1, \ldots, l, \\ u_i \geqslant 0, \quad g_i(x) \leqslant 0, \quad i = 1, \ldots, m. \end{cases}$$

* This work has been supported by Shanghai Education Committee (04EA01).

The system (1.1) could be transformed into nonsmooth equations of the form

$$(1.2)\qquad \begin{cases} \nabla f(x) + \sum\limits_{i=1}^{m} u_i \nabla g_i(x) + \sum\limits_{j=1}^{l} v_j \nabla h_j(x) = 0, \\ \min\{u_i, -g_i(x)\} = 0, \quad i = 1, \ldots, m, \\ h_j(x) = 0, \quad j = 1, \ldots, l. \end{cases}$$

Many publications dealt with the system (1.2) to solve the problem (P_1), see for instance [4], [11]–[15].

Nevertheless, no papers which would apply directly the nonsmooth equations methods to solving minimax problems have appeared. In this paper, we intend to explore a nonsmooth equations method for solving the constrained minimax problem, see the problem (P_2) below. The paper is organized as follows: In the remainder of this section, some preliminaries are reviewed. In the next section, the nonsmooth equations model of the minimax problem is derived. In Section 3, the method for solving our system of nonsmooth equations is developed. In Section 4, numerical examples are listed.

Let $F\colon \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitzian and let D_F denote the set where F is differentiable. By the definition in [12],

$$\partial_B F(x) = \big\{ \lim_{x_n \to x} JF(x_n) \mid x_n \to x, \ \ x_n \in D_F \big\},$$

where "J" denotes the Jacobian, is called the B-differential of F at x. According to [1],

$$\partial_{\mathrm{Cl}} F(x) = \operatorname{conv} \partial_B F(x)$$

is called the Clarke generalized Jacobian of F at x; in particular, in the case $m = 1$, $\partial_{\mathrm{Cl}} F(x)$ reduces to the Clarke generalized gradient. Following [15],

$$\partial_b F(x) = \partial_B f_1(x) \times \ldots \times \partial_B f_m(x),$$

where $f_i(x)$ is the ith component of $F(x)$, is called the b-differential of F at x; particularly, if $m = 1$, $\partial_b F(x) = \partial_B F(x)$.

Following the definition in [14], a locally Lipschitzian function $F\colon \mathbb{R}^n \to \mathbb{R}^m$ is said to be semismooth at x provided that

$$\lim_{\substack{V \in \partial_{\mathrm{Cl}} F(x+th') \\ h' \to h,\, t \to 0^+}} Vh'$$

exists for any $h \in \mathbb{R}^n$. Convex functions, smooth functions, maximums of smooth functions and piecewise C^1 functions are semismooth. It was shown that a function F from $\mathbb{R}^n$ to $\mathbb{R}^m$ is semismooth if and only if all its components are semismooth.

According to the definition in [3], $f\colon \mathbb{R}^n \to \mathbb{R}$ is said to be quasidifferentiable at a point $x \in \mathbb{R}^n$, in the sense of Demyanov and Rubinov, if it is directionally differentiable at x and its directional derivative $f'(x;\cdot)$ is representable as the difference of two sublinear functions. In other words, there exists a pair of convex compact sets $\underline{\partial} f(x), \overline{\partial} f(x) \subset \mathbb{R}^n$ such that

$$f'(x;d) = \max_{v \in \underline{\partial} f(x)} v^T d + \min_{w \in \overline{\partial} f(x)} w^T d, \quad \forall\, d \in \mathbb{R}^n.$$

The pair of sets $Df(x) = [\underline{\partial} f(x), \overline{\partial} f(x)]$ is called the quasidifferential of f at x, $\underline{\partial} f(x)$ and $\overline{\partial} f(x)$ are called the subdifferential and the superdifferential, respectively.

Let U be a convex compact set in $\mathbb{R}^n$. The support function P_U of U is defined by $P_U(x) = \max\limits_{u \in U} u^T x$.

Given $\psi_i\colon \mathbb{R}^n \to \mathbb{R}$, $i \in I$, where I is a finite index set, we denote

$$g(x) = \max_{i \in I} \psi_i(x), \tag{1.3}$$

$$h(x) = \min_{i \in I} \psi_i(x), \tag{1.4}$$

$$I(x) = \{i \in I \mid \psi_i(x) = g(x)\}, \tag{1.5}$$

$$\overline{I}(x) = \{i \in I \mid \psi_i(x) = h(x)\}. \tag{1.6}$$

If ψ_i, $i \in I$ are continuously differentiable, then the Clarke generalized gradients of g and of h, given in (1.3) and (1.4), are formulated as follows:

$$\partial_{\mathrm{Cl}} g(x) = \operatorname{conv}\{\nabla \psi_i(x) \mid i \in I(x)\}, \tag{1.7}$$

$$\partial_{\mathrm{Cl}} h(x) = \operatorname{conv}\{\nabla \psi_i(x) \mid i \in \overline{I}(x)\}. \tag{1.8}$$

Suppose that each ψ_i is quasidifferentiable with the quasidifferential $[\underline{\partial}\psi_i(x), \overline{\partial}\psi_i(x)]$. Then both g and h, given in (1.3) and (1.4), respectively, are again quasidifferentiable and their quasidifferentials $[\underline{\partial} g(x), \overline{\partial} g(x)]$ and $[\underline{\partial} h(x), \overline{\partial} h(x)]$ have the form

$$\underline{\partial} g(x) = \operatorname{conv} \bigcup_{k \in I(x)} \Bigg(\underline{\partial}\psi_k(x) - \sum_{i \in I(x),\, i \neq k} \overline{\partial}\psi_i(x) \Bigg), \tag{1.9}$$

$$\overline{\partial} g(x) = \sum_{k \in I(x)} \overline{\partial}\psi_k(x), \tag{1.10}$$

$$\underline{\partial} h(x) = \sum_{k \in \overline{I}(x)} \underline{\partial}\psi_k(x), \tag{1.11}$$

$$\overline{\partial} h(x) = \operatorname{conv} \bigcup_{k \in \overline{I}(x)} \Bigg(\overline{\partial}\psi_k(x) - \sum_{i \in \overline{I}(x),\, i \neq k} \underline{\partial}\psi_i(x) \Bigg). \tag{1.12}$$

2. The model of nonsmooth equations

Consider the constrained minimax problem of the form

$$
(\mathrm{P}_2) \qquad \begin{aligned} &\text{minimize} && \max_{1\leqslant i\leqslant m} f_i(x) \\ &\text{subject to} && g_j(x)\leqslant 0, \quad j=1,\dots,p, \\ & && h_k(x)=0, \quad k=1,\dots,q \end{aligned}
$$

where $f_i, g_j, h_k\colon \mathbb{R}^n \to \mathbb{R}$, $i=1,\dots,m$, $j=1,\dots,p$, $k=1,\dots,q$ are continuously differentiable.

We know that the Karush-Kuhn-Tucker (KKT) optimality condition for (P_2) has the form [1], [2]

$$
\begin{cases} 0\in \partial_{\mathrm{Cl}}\bigl(\max\limits_{1\leqslant i\leqslant m} f_i(x)\bigr)+\sum\limits_{j=1}^{p} v_j\nabla g_j(x)+\sum\limits_{k=1}^{q} w_k\nabla h_k(x), \\ v_jg_j(x)=0, \quad j=1,\dots,p, \\ h_k(x)=0, \quad k=1,\dots,q, \\ v_j\geqslant 0, \quad g_j(x)\leqslant 0, \quad j=1,\dots,p. \end{cases} \tag{2.1}
$$

Under some constraint qualifications, for instance f_i, $i=1,\dots,m$ and g_j, $j=1,\dots,p$ are convex, h_k, $k=1,\dots,q$ are affine [1], [7], the problem (P_2) is equivalent to the system (2.1). Actually, most existing algorithms for solving the problem (P_2) generate a KKT point, i.e., a point satisfying (2.1). They are not ensured to generate a minimizer. According to [1], [3], we have

$$
\begin{aligned} &\partial_{\mathrm{Cl}}\bigl(\max_{1\leqslant i\leqslant m} f_i(x)\bigr)=\mathrm{conv}\{\nabla f_i(x)\mid f_i(x)=\max_{1\leqslant s\leqslant m} f_s(x)\} \\ &=\Bigl\{\sum_{i=1}^{m} u_i\nabla f_i(x)\mid u_i\bigl(\max_{1\leqslant s\leqslant m} f_s(x)-f_i(x)\bigr)=0,\ \sum_{i=1}^{m} u_i=1,\ u_i\geqslant 0\Bigr\}. \end{aligned}
$$

Thus (2.1) is equivalent to the system

$$
\begin{cases} \sum\limits_{i=1}^{m} u_i\nabla f_i(x)+\sum\limits_{j=1}^{p} v_j\nabla g_j(x)+\sum\limits_{k=1}^{q} w_k\nabla h_k(x)=0, \\ u_i\bigl(\max\limits_{1\leqslant s\leqslant m} f_s(x)-f_i(x)\bigr)=0, \quad i=1,\dots,m, \\ \sum\limits_{i=1}^{m} u_i=1, \quad u_i\geqslant 0, \quad i=1,\dots,m, \\ v_jg_j(x)=0, \quad j=1,\dots,p, \\ h_k(x)=0, \quad k=1,\dots,q, \\ v_j\geqslant 0, \quad g_j(x)\leqslant 0, \quad j=1,\dots,p. \end{cases} \tag{2.2}
$$

We now consider a constraint qualification at a point x denoted by (CQ): The following nonlinear system has no non-zero solutions:

$$\begin{cases} \sum_{j=1}^{p} v_j \nabla g_j(x) + \sum_{k=1}^{q} w_k \nabla h_k(x) = 0, \\ v_j g_j(x) = 0, \quad j = 1, \dots, p, \\ h_k(x) = 0, \quad k = 1, \dots, q, \\ v_j \geqslant 0, \quad g_j(x) \geqslant 0, \quad m, j = 1, \dots, p. \end{cases}$$

Under the constraint qualification (CQ), it is easy to see that (2.2) is equivalent to the system

$$\begin{cases} \sum_{i=1}^{m} u_i \nabla f_i(x) + \sum_{j=1}^{p} v_j \nabla g_j(x) + \sum_{k=1}^{q} w_k \nabla h_k(x) = 0, \\ u_i\big(\max\limits_{1 \leqslant s \leqslant m} f_s(x) - f_i(x)\big) = 0, \quad i = 1, \dots, m, \\ v_j g_j(x) = 0, \quad j = 1, \dots, p, \\ h_k(x) = 0, \quad k = 1, \dots, q, \\ u_i \geqslant 0, \quad i = 1, \dots, m, \quad v_j \geqslant 0, \quad g_j(x) \leqslant 0, \quad m, j = 1, \dots, p. \end{cases} \tag{2.3}$$

On the other hand, (2.3) can be rewritten by

$$\begin{cases} \sum_{i=1}^{m} u_i \nabla f_i(x) + \sum_{j=1}^{p} v_j \nabla g_j(x) + \sum_{k=1}^{q} w_k \nabla h_k(x) = 0, \\ \min\{u_i, \max\limits_{1 \leqslant s \leqslant m} f_s(x) - f_i(x)\} = 0, \quad i = 1, \dots, m, \\ \min\{v_j, -g_j(x)\} = 0, \quad j = 1, \dots, p, \\ h_k(x) = 0, \quad k = 1, \dots, q, \end{cases} \tag{2.4}$$

because $0 \leqslant \max\limits_{1 \leqslant s \leqslant m} f_s(x) - f_i(x)$ for $i = 1, \dots, m$.

Evidently, the corresponding function of the system (2.4) is Lipschitzian and semismooth. Therefore, many existing methods for solving nonsmooth equations can be applied to this system. In this paper, we will apply a Newton method to solving the system (2.4). Of course, the problem (P_2) can be transformed into the following equivalent nonlinear programming problem by means of introducing an auxiliary variable:

$$(\mathrm{P}_3) \qquad \begin{array}{lll} \text{minimize} & z & \\ \text{subject to} & f_i(x) - z \leqslant 0, & i = 1, \dots, m, \\ & g_j(x) \leqslant 0, & j = 1, \dots, p, \\ & h_k(x) = 0, & k = 1, \dots, q. \end{array}$$

Some nonlinear optimization methods (for instance the SQP method) can be applied to (P_3). In this paper we are interested in nonsmooth equations methods for the minimax problem. It should be said that the system of nonsmooth equations, transformed from the Karush-Kuhn-Tucker system of the problem (P_3), has $m+n+p+q+1$ equations and variables, whereas the system (2.4) has $m+n+p+q$ equations and variables.

3. Solving the nonsmooth equations

This section is devoted to solving the system (2.4) by using the Newton method proposed by Sun and Han [15]. We start with considering the nonsmooth equations

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

where $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitzian. Newton methods for solving the nonsmooth equations (3.1) are given as follows:

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

where V_k can be taken as an element of various subdifferentials of F at x^k [6], [12], [14], [15]. In [15], V_k is an element of the b-differential. Locally superlinear convergence of these Newton methods was shown when F is semismooth and all elements of the corresponding subdifferentials are nonsingular at the solution x^*. Of course, all Newton methods, given in (3.2), work on the assumption that at least one element for their corresponding subdifferentials of F at each iteration point can be calculated.

To perform the Newton method in [15] for solving the nonsmooth equations (2.4), an element of the b-differential for the function on the left-hand side of (2.4), in other words, an element of the B-differential for each component of this function is required. On the left-hand side of (2.4) we have to do with two nonsmooth functions. For the computation of the b-differential of the first one, the quasidifferential calculus will be used. We know that an element of the B-differential for the other function can be computed in a direct way.

Let

$$H(x) = \min\bigl\{\varphi(x), \max_{i\in I} \psi_i(x)\bigr\}, \tag{3.3}$$

where $\varphi, \psi_i\colon \mathbb{R}^n \to \mathbb{R}$, $i \in I$ are continuously differentiable and I is a finite index set. We employ the notation $g(x)$ and $I(x)$ introduced by (1.3) and (1.5), respectively.

Evidently, H is quasidifferentiable. Based on the quasidifferential calculus, we obtain that

$$\begin{gathered}\underline{\partial} g(x) = \operatorname{conv}\{\nabla\psi_i(x) \mid i \in I(x)\}, \quad \overline{\partial} g(x) = \{0\},\\ \underline{\partial}\varphi(x) = \{0\}, \quad \overline{\partial}\varphi(x) = \{\nabla\varphi(x)\}.\end{gathered}$$

Using formulas (1.11) and (1.12), one has

$$\begin{aligned}(3.4)\qquad \underline{\partial} H(x) &= \underline{\partial} g(x) + \underline{\partial}\varphi(x)\\ &= \operatorname{conv}\{\nabla\psi_i(x) \mid i \in I(x)\} \quad \text{if } \varphi(x) = \max_{i\in I}\psi_i(x),\\ (3.5)\qquad \overline{\partial} H(x) &= \operatorname{conv}[(\overline{\partial} g(x) - \underline{\partial}\varphi(x)) \cup (\overline{\partial}\varphi(x) - \underline{\partial} g(x))]\\ &= \operatorname{conv}[\{0\} \cup \operatorname{conv}\{\nabla\varphi(x) - \nabla\psi_i(x) \mid i \in I(x)\}]\\ &= \operatorname{conv}\{0, \nabla\varphi(x) - \nabla\psi_i(x) \mid i \in I(x)\} \quad \text{if } \varphi(x) = \max_{i\in I}\psi_i(x).\end{aligned}$$

Early results on the relation between the Clarke generalized gradient and the quasidifferential were obtained by [2], [9]. In this paper we apply the results of [5]. In what follows, we review the notion of the Demyanov difference of convex compact sets, due to [3], which can be used to calculate the Clarke generalized gradient via the quasidifferential for a certain class of functions.

A set $T \subset \mathbb{R}^n$ is said to be of full measure (with respect to $\mathbb{R}^n$), if the Lebesgue measure of the set $\mathbb{R}^n \setminus T$ is zero. Let $U, V \subset \mathbb{R}^n$ be convex compact sets and $T \subset \mathbb{R}^n$ a full measure set such that their support functions P_U and P_V are differentiable at every point $x \in T$. The set $U \dot{-} V$ is defined by

$$U \dot{-} V = \operatorname{cl}\operatorname{conv}\{\nabla P_U(x) - \nabla P_V(x) \mid x \in T\}. \tag{3.6}$$

We call $U \dot{-} V$ the Demyanov difference of U and V. It has been shown that $U \dot{-} V$ does not depend on the specific choice of the set T, so it is well-defined.

Let f be quasidifferentiable. Given a fixed point $x \in \mathbb{R}^n$, the function $f'(x;\cdot)$ is locally Lipschitzian. The following relation holds:

$$\partial_{\mathrm{Cl}} f'(x;y)|_{y=0} = \underline{\partial} f(x) \dot{-} (-\overline{\partial} f(x)) \tag{3.7}$$

and if both $[U_1, V_1]$ and $[U_2, V_2]$ are quasidifferentials of f at x, then $U_1 \dot{-} (-V_1) = U_2 \dot{-} (-V_2)$. That is to say, the set $\underline{\partial} f(x) \dot{-} (-\overline{\partial} f(x))$ is independent of the specific choice of the quasidifferential.

By the definition in [10], $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is said to be piecewise C^k on an open set $S \subset \mathbb{R}^n$, where k is a positive integer, if there exists a finite family of C^k functions

$f_i\colon S \to \mathbb{R}^m$ for $i = 1, \dots, l$, called the C^k pieces of f, such that f is continuous on S and for every $x \in S$, $f(x) = f_i(x)$ for at least one index $i \in \{1, \dots, l\}$. The family of piecewise C^k functions is denoted by PC^k. By [10] the following relation holds:

$$\partial_B f'(x; y)|_{y=0} \subset \partial_B f(x), \quad \forall f \in PC^1. \tag{3.8}$$

Combining (3.7) and (3.8) with the relation between the B-differential and the Clarke generalized gradient yields

$$\underline{\partial} f(x) \dot{-} (-\overline{\partial} f(x)) \subset \partial_{Cl} f(x), \quad \forall f \in PC^1. \tag{3.9}$$

(3.9) enables us to compute some elements of the Clarke generalized gradient via the quasidifferential for a PC^1 function.

Obviously, H defined by (3.3) is a PC^1 function.

The next proposition gives us a formula of the Demyanov difference for a pair of polyhedrons.

Proposition 3.1 [5]. *Let* $U = \operatorname{conv}\{u_i \mid i \in I\}$ *and* $V = \operatorname{conv}\{v_j \mid j \in J\}$, *where* $u_i, v_j \in \mathbb{R}^n$, I *and* J *are finite index sets. Without loss of generality, suppose that* $u_s \neq u_t, \forall\, s, t \in I,\ s \neq t$ *and* $v_s \neq v_t, \forall\, s, t \in J,\ s \neq t$. *Given a pair of indices* $i \in I$ *and* $j \in J$, *define a system of linear inequalities* (L_{ij}) *as follows:*

$$(\mathrm{L}_{ij},) \qquad \begin{aligned} (u_s - u_i)^T y < 0, &\quad \forall\, s \in I \setminus \{i\} \\ (v_t - v_j)^T y < 0, &\quad \forall\, t \in J \setminus \{j\} \end{aligned}$$

where $y \in \mathbb{R}^n$. *We then have*

$$U \dot{-} V = \operatorname{conv}\{u_i - v_j \mid (\mathrm{L}_{ij}) \text{ is consistent}, \ i \in I, \ j \in J\}.$$

The system (L_{ij}) has $\operatorname{card}(I) + \operatorname{card}(J) - 2$ linear inequalities, where "card" denotes cardinality.

In the light of Proposition 3.1, we investigate the Clarke generalized gradient of the function H at a point $\overline{x}$ satisfying $\varphi(\overline{x}) = \max_{i \in I} \psi_i(\overline{x})$ in the next theorem.

Theorem 3.1. *Let H be defined by* (3.3) *and* $\overline{x} \in \mathbb{R}^n$ *with* $\varphi(\overline{x}) = \max_{i \in I} \psi_i(\overline{x})$. *Given an index* $i \in I(\overline{x})$, *we construct two systems of linear inequalities as follows:*

$$\begin{cases} (\nabla\psi_s(\overline{x}) - \nabla\psi_i(\overline{x}))^T y < 0, & \forall\, s \in I(\overline{x}) \setminus \{i\}, \\ (\nabla\varphi(\overline{x}) - \nabla\psi_i(\overline{x}))^T y < 0 \end{cases} \tag{3.10}$$

and

$$\begin{cases} (\nabla\psi_s(\overline{x}) - \nabla\psi_i(\overline{x}))^T y < 0, & \forall\, s \in I(\overline{x}) \setminus \{i\}, \\ (\nabla\psi_i(\overline{x}) - \nabla\varphi(\overline{x}))^T y < 0 \end{cases} \tag{3.11}$$

where $y \in \mathbb{R}^n$. *Then* $\nabla\varphi(\overline{x}) \in \partial_{\mathrm{Cl}} H(\overline{x})$ *if* (3.10) *is consistent and* $\nabla\psi_i(\overline{x}) \in \partial_{Cl} H(\overline{x})$ *if* (3.11) *is consistent.*

P r o o f. Suppose (3.10) is consistent. According to (3.5), one has

$$-\overline{\partial} H(\overline{x}) = \mathrm{conv}\{0, \nabla\psi_i(\overline{x}) - \nabla\varphi(\overline{x}) \mid i \in I(\overline{x})\}.$$

Obviously, (3.10) can be rewritten as

$$\begin{cases} (\nabla\psi_s(\overline{x}) - \nabla\psi_i(\overline{x}))^T y < 0, \quad \forall\, s \in I(\overline{x}) \setminus \{i\}, \\ [(\nabla\psi_s(\overline{x}) - \nabla\varphi(\overline{x})) - (\nabla\psi_i(\overline{x}) - \nabla\varphi(\overline{x}))]^T y < 0, \quad \forall\, s \in I(\overline{x}) \setminus \{i\}, \\ [0 - (\nabla\psi_i(\overline{x}) - \nabla\varphi(\overline{x}))]^T y < 0. \end{cases} \tag{3.12}$$

Notice that both $\underline{\partial} H(\overline{x})$ and $-\overline{\partial} H(\overline{x})$ are convex hulls of finitely many points. By virtue of Proposition 3.1, the consistency of (3.12) implies

$$\nabla\psi_i(\overline{x}) - (\nabla\psi_i(\overline{x}) - \nabla\varphi(\overline{x})) = \nabla\varphi(\overline{x}) \in \underline{\partial} H(\overline{x}) \dot{-} (-\overline{\partial} H(\overline{x})).$$

From (3.9) and $H \in PC^1$, it follows that $\nabla\varphi(\overline{x}) \in \partial_{\mathrm{Cl}} H(\overline{x})$.

Suppose that (3.11) is consistent. It is not hard to see that (3.11) is equivalent to the system

$$\begin{cases} (\nabla\psi_s(\overline{x}) - \nabla\psi_i(\overline{x}))^T y < 0, & \forall\, s \in I(\overline{x}) \setminus \{i\}, \\ (\nabla\psi_s(\overline{x}) - \nabla\varphi(\overline{x}))^T y < 0, & \forall\, s \in I(\overline{x}). \end{cases} \tag{3.13}$$

Moreover, (3.13) can be rewritten as

$$\begin{cases} (\nabla\psi_s(\overline{x}) - \nabla\psi_i(\overline{x}))^T y < 0, & \forall\, s \in I(\overline{x}) \setminus \{i\}, \\ [(\nabla\psi_s(\overline{x})) - \nabla\varphi(\overline{x})) - 0]^T y < 0, & \forall\, s \in I(\overline{x}). \end{cases}$$

Similarly to the above argument, it can be obtained that $\nabla\psi_i(\overline{x}) = \nabla\psi_i(\overline{x}) - \nabla 0 \in \partial_{\mathrm{Cl}} H(\overline{x})$. This completes the proof of the theorem. $\square$

Let S be a closed convex set in $\mathbb{R}^n$. By the definition in [7], we say $x \in S$ is an extreme point, if there exists no convex combination $x = \sum_{i=1}^{k} \lambda_i x_i$, where $x_i \in S$, $\sum_{i=1}^{k} \lambda_i = 1$, $\lambda_i \geqslant 0$, other than $x_1 = \ldots = x_k = x$.

Proposition 3.2. *Let $S = \operatorname{conv}\{a_i \mid i \in I\}$, $a_i \in \mathbb{R}^n$, and let I be a finite index set. Without loss of generality, suppose $a_s \neq a_t$, $\forall\, s,t \in I, s \neq t$. Given an index $i \in I$, a_i is an extreme point of S if and only if the following system is consistent:*

$$(a_k - a_i)^T y < 0, \quad y \in \mathbb{R}^n, \quad \forall\, k \in I \setminus \{i\}. \tag{3.14}$$

P r o o f. Suppose that the system (3.14) is consistent. Let $\overline{y} \in \mathbb{R}^n$ be a solution of (3.14). Any $x \in S$, $x \neq a_i$, can be expressed as $x = \sum_{k \in I} \lambda_k a_k$, $\sum_{k \in I} \lambda_k = 1$, $\lambda_k \geqslant 0$, with at least one index $k \in I \setminus \{i\}$ satisfying $\lambda_k \neq 0$. This leads to

$$x^T \overline{y} = \sum_{k \in I} \lambda_k a_k^T \overline{y} < \sum_{k \in I} \lambda_k a_i^T \overline{y} = a_i^T \overline{y}. \tag{3.15}$$

Suppose a_i is expressed as a convex combination, i.e., $a_i = \sum_{s=1}^{m} \mu_s x_s$, where $x_s \in S$, $\sum_{s=1}^{m} \mu_s = 1$, $\mu_s \geqslant 0$. It follows that $a_i^T \overline{y} = \sum_{s=1}^{m} \mu_s x_s^T \overline{y}$. This yields $x_s = a_i$ since (3.15) implies $x_s^T \overline{y} < a_i^T \overline{y}$ whenever $x_s \neq a_i$. By definition, a_i is an extreme point of the set S.

Conversely, suppose a_i is an extreme point of S. Then $a_i \notin \operatorname{conv}\{a_k \mid k \in I \setminus \{i\}\}$, otherwise $a_i = \sum_{k \in I \setminus \{i\}} \lambda_k a_k$, $\sum_{k \in I \setminus \{i\}} \lambda_k = 1$, $\lambda_k \geqslant 0$, which contradicts the fact that a_i is an extreme point. Applying the separation theorem (see [7]) to sets $\{a_i\}$ and $\operatorname{conv}\{a_s \mid s \in I \setminus \{i\}\}$, we obtain that there exists $y_1 \in \mathbb{R}^n$ such that

$$\max\{u^T y_1 \mid u \in \operatorname{conv}\{a_s \mid s \in I \setminus \{i\}\}\} < a_i^T y_1.$$

Moreover,

$$(a_s - a_i)^T y_1 < 0, \quad \forall\, s \in I \setminus \{i\},$$

i.e., (3.14) is consistent. We have thus completed the proof of the proposition. $\square$

Corollary 3.1. *Let S be given as in Proposition 3.2. Suppose that an index $i \in I$ satisfies*

$$\|a_i\| = \max_{s\in I} \|a_s\|,$$

where $\|\cdot\|$ refers to the Euclidean norm. Then a_i is an extreme point of the set S.

Proof. It is easy to see that

$$(a_s - a_i)^T a_i < 0, \quad \forall\, s \in I \setminus \{i\}, \tag{3.16}$$

which is to say the system (3.14) is consistent. By virtue of Proposition 3.2, a_i is an extreme point of the set S. □

Proposition 3.3. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitzian. If u is an extreme point of $\partial_{\mathrm{Cl}} f(x)$, then $u \in \partial_B f(x)$.*

Proof. From the relation between the Clarke generalized gradient and the B-differential, it follows that u can be expressed as a convex combination $u = \sum\limits_{i=1}^{k} \lambda_i u_i$, where $u_i \in \partial_B f(x) \subset \partial_{\mathrm{Cl}} f(x)$, $\sum\limits_{i=1}^{k} \lambda_i = 1$, $\lambda_i \geqslant 0$. The definition of extreme point implies $u_1 = \ldots = u_k = u$. Therefore, $u \in \partial_B f(x)$. □

The following theorem is devoted to the B-differential of the function

$$\min\bigl\{u_i, \max_{1\leqslant s\leqslant m} f_s(x) - f_i(x)\bigr\},$$

which appears on the left-hand side of (2.4). We hereafter take $(u, v, w, x) \in \mathbb{R}^{m+n+p+q}$ as the variable, denote by e_i the unit vector in $\mathbb{R}^m$ whose ith component is 1, and by 0_s null vector in $\mathbb{R}^s$.

Theorem 3.2. *Let x be a point in $\mathbb{R}^n$ and u_i a point in $\mathbb{R}$ satisfying*

$$u_i = \max_{1\leqslant s\leqslant m} f_s(x) - f_i(x).$$

Then $(e_i^T, 0_{n+p+q}^T)^T$ belongs to the B-differential of the function

$$\min\bigl\{u_i, \max_{1\leqslant s\leqslant m} f_s(x) - f_i(x)\bigr\}.$$

Let

$$I_1(x) = \bigl\{k \in \{1, \ldots, m\} \mid f_k(x) = \max_{1\leqslant s\leqslant m} f_s(x)\bigr\},$$

and let the index $i_0 \in I_1(x)$ be such that

$$\|\nabla f_{i_0}(x)\| = \max_{1\leqslant s\leqslant m} \|\nabla f_s(x)\|. \tag{3.17}$$

Then $(0_{m+p+q}^T, \nabla f_{i_0}(x)^T - \nabla f_i(x)^T)^T$ is contained in the B-differential of the function $\min\bigl\{u_i, \max\limits_{1\leqslant s\leqslant m} f_s(x) - f_i(x)\bigr\}$.

P r o o f. Let

$$F_i(u,v,w,x) = \min\bigl\{u_i + f_i(x), \max_{1\leqslant s\leqslant m} f_s(x)\bigr\}.$$

Noting that

$$\min\bigl\{u_i, \max_{1\leqslant s\leqslant m} f_s(x) - f_i(x)\bigr\} = F_i(u,v,w,x) - f_i(x),$$

one has

$$\partial_B\bigl(\min\bigl\{u_i, \max_{1\leqslant s\leqslant m} f_s(x) - f_i(x)\bigr\}\bigr) = \partial_B F_i(u,v,w,x) - (0_{m+p+q}^T, \nabla f_i(x)^T)^T.$$

Hence, it is sufficient to show that $(e_i^T, 0_{p+q}^T, \nabla f_i(x)^T)^T$ and $(0_{m+p+q}^T, \nabla f_{i_0}(x)^T)^T$ are contained in the B-differential of the function F_i at (u,v,w,x). Let $y_1 = (0_{m+p+q}^T, \nabla f_{i_0}(x)^T)^T$. We know that $\nabla(u_i + f_i(x)) = (e_i^T, 0_{p+q}^T, \nabla f_i(x)^T)^T$ and $\nabla f_s(x) = (0_{m+p+q}^T, \nabla f_s(x)^T)^T$. By virtue of (3.16) and (3.17), we have

$$\begin{cases} [(0_{m+p+q}^T, \nabla f_s(x)^T) - (0_{m+p+q}^T, \nabla f_{i_0}(x)^T)]y_1 \\ \quad = (\nabla f_s(x) - \nabla f_{i_0}(x))^T \nabla f_{i_0}(x) < 0, \quad \forall\, s \in I_1(x) \setminus \{i_0\} \\ [(e_i^T, 0_{p+q}^T, \nabla f_i(x)^T) - (0_{m+p+q}^T, \nabla f_{i_0}(x)^T)]y_1 \\ \quad = (\nabla f_i(x) - \nabla f_{i_0}(x))^T \nabla f_{i_0}(x) < 0. \end{cases} \tag{3.18}$$

From Theorem 3.1, in particular (3.10), it follows that $(e_i^T, 0_{p+q}^T, \nabla f_i(x)^T)^T = \nabla(u_i + f_i(x)) \in \partial_{\mathrm{Cl}} F_i(u,v,w,x)$. On the other hand, notice that $y = (e_i^T, 0_{p+q+n}^T)^T$ is a solution of the system

$$[(0_{m+p+q}^T, \nabla f_s(x)^T) - (e_i^T, 0_{p+q}^T, \nabla f_i(x)^T)]^T y < 0, \quad \forall\, s \in I_1(x) \setminus \{i\}.$$

Based on Proposition 3.2, this yields the assertion that $(e_i^T, 0_{p+q}^T, \nabla f_i(x)^T)^T$ is an extreme point of the set $\mathrm{conv}\{(e_i^T, 0_{p+q}^T, \nabla f_i(x)^T)^T, (0_{m+p+q}^T, \nabla f_s(x)^T)^T \mid s \in I_1(x)\}$. By the calculus of the Clarke generalized gradient, we obtain

$$\begin{aligned} &\partial_{\mathrm{Cl}} F_i(u,v,w,x) \\ &\qquad \subset \mathrm{conv}\{(e_i^T, 0_{p+q}^T, \nabla f_i(x)^T), (0_{m+p+q}^T, \nabla f_s(x)^T)^T \mid s \in I_1(x)\}. \end{aligned} \tag{3.19}$$

Hence, $(e_i^T, 0_{p+q}^T, \nabla f_i(x)^T)^T$ is also an extreme point of the set $\partial_{\mathrm{Cl}} F_i(u,v,w,x)$. By virtue of Proposition 3.3, $(e_i^T, 0_{p+q}^T, \nabla f_i(x))^T \in \partial_B F_i(u,v,w,x)$.

Let $y_2 = (Me_i^T, 0_{p+q}^T, \nabla f_{i_0}(x)^T)^T$ and let M be a large number such that

$$\begin{cases} [(0_{m+p+q}^T, \nabla f_s(x)^T) - (0_{m+p+q}^T, \nabla f_{i_0}(x)^T)]^T y_2 < 0, \quad \forall\, s \in I_1(x) \setminus \{i_0\} \\ [(0_{m+p+q}^T, \nabla f_{i_0}(x)^T) - (e_i^T, 0_{p+q}^T, \nabla f_i(x)^T]^T y_2 < 0. \end{cases} \tag{3.20}$$

According to Theorem 3.1, in particular (3.11), one has that $(0_{m+p+q}^T, \nabla f_{i_0}(x)^T)^T \in \partial_{\mathrm{Cl}} F_i(u,v,w,x)$. Notice that $y = (0_{m+p+q}^T, \nabla f_{i_0}(x)^T)^T$ is a solution of the system

$$\begin{cases} [(0_{m+p+q}^T, \nabla f_s(x)^T) - (0_{m+p+q}^T, \nabla f_{i_0}(x)^T)]^T y < 0, \quad \forall\, s \in I_1(x) \setminus \{i_0\} \\ [(e_i^T, 0_{p+q}^T, \nabla f_i(x)^T) - (0_{m+p+q}^T, \nabla f_{i_0}(x)^T)] y < 0. \end{cases} \tag{3.21}$$

From (3.19), (3.21), Proposition 3.3 and $(0_{m+p+q}^T, \nabla f_{i_0}(x)^T)^T \in \partial_{\mathrm{Cl}} F_i(u,v,w,x)$, it follows that $(0_{m+p+q}^T, \nabla f_{i_0}(x)^T)^T \in \partial_B F_i(u,v,w,x)$. This completes the proof of the theorem. □

Theorem 3.2 enables us to obtain two elements of the B-differential $(e_i^T, 0_{n+p+q}^T)^T$ and $(0_{m+p+q}^T, \nabla f_{i_0}(x)^T - \nabla f_i(x)^T)^T$, where i_0 satisfies (3.17) for the function $\min\{u_i, \max\limits_{1\leqslant s\leqslant m} f_s(x) - f_i(x)\}$ at the point x when $u_i = \max\limits_{1\leqslant s\leqslant m} f_s(x) - f_i(x)$. Evidently, these two elements can be calculated. In particular, it should be mentioned that $(0_{m+p+q}^T, \nabla f_{i_0}(x)^T - \nabla f_i(x)^T)^T$ can be calculated by determining the maximum of $\|\nabla f_i(x)\|$ for $i = 1, \dots, m$.

Evidently, $\min\{u_i, \max\limits_{1\leqslant s\leqslant m} f_s(x) - f_i(x)\}$ is differentiable and $(e_i^T, 0_{n+p+q}^T)^T$ is its gradient in the case $u_i < \max\limits_{1\leqslant s\leqslant m} f_s(x) - f_i(x)$. If $u_i > \max\limits_{1\leqslant s\leqslant m} f_s(x) - f_i(x)$, then

$$\partial_B\big(\min\{u_i, \max_{1\leqslant s\leqslant m} f_s(x) - f_i(x)\}\big) = \partial_B\big(\max_{1\leqslant s\leqslant m} f_s(x) - f_i(x)\big). \tag{3.22}$$

It is easy to see that $(0_{m+p+q}^T, \nabla f_{i_0}(x)^T - \nabla f_i(x)^T)^T$, where i_0 satisfies (3.17), is an element of the set on the right-hand side of (3.22), so it is an element of the B-differential of the function $\min\{u_i, \max\limits_{1\leqslant s\leqslant m} f_s(x) - f_i(x)\}$ if $u_i > \max\limits_{1\leqslant s\leqslant m} f_s(x) - f_i(x)$. To sum up, we give a formula for the element of the B-differential for the function

$\min\{u_i, \max_{1\leqslant s\leqslant m} f_s(x) - f_i(x)\}$ as follows:

$$
(3.23) \qquad \left.\begin{array}{l}
(e_i^T, 0_{n+p+q}^T)^T, \\
(0_{m+p+q}^T, \nabla f_{i_0}(x)^T - \nabla f_i(x)^T)^T \\
\quad (\text{if } u_i = \max_{1\leqslant s\leqslant m} f_s(x) - f_i(x)) \\
(e_i^T, 0_{n+p+q}^T)^T \\
\quad (\text{if } u_i < \max_{1\leqslant s\leqslant m} f_s(x) - f_i(x)) \\
(0_{m+p+q}^T, \nabla f_{i_0}(x)^T - \nabla f_i(x)^T)^T \\
\quad (\text{if } u_i > \max_{1\leqslant s\leqslant m} f_s(x) - f_i(x))
\end{array}\right\} \in \partial_B \min\{u_i, \max_{1\leqslant s\leqslant m} f_s(x) - f_i(x)\},
$$

where i_0 satisfies (3.17). From (3.23), it is seen that computing an element of the B-differential for the function $\min\{u_i, \max_{1\leqslant s\leqslant m} f_s(x) - f_i(x)\}$ can be performed by finding an index i_0 in (3.17). This work is very cheap. By [15], elements of the B-differential of the function $\min\{v_j, -g_j(x)\}$ are determined by

$$
\left.\begin{array}{l}
(0_m^T, e_j^T, 0_{q+n}^T)^T \\
\quad (\text{if } v_j < -g_j(x)) \\
(0_{m+p+q}^T, -\nabla g_j(x)^T)^T \\
\quad (\text{if } v_j > -g_j(x)) \\
(0_m^T, e_j^T, 0_{q+n}^T)^T, (0_{m+p+q}^T, -\nabla g_j(x)^T)^T \\
\quad (\text{if } v_j = -g_j(x))
\end{array}\right\} \in \partial_B\big(\min\{v_j, -g_j(x)\}\big).
$$

Finally, each function h_k and the first component of the function on the left-hand side of (2.4) are continuously differentiable. So, their B-differentials are the gradients.

Up to now, we can calculate an element of the B-differential for each component of the function on the left-hand side of (2.4) at a point. That is, an element of the b-differential of this function at a point can be calculated. Hence, the Newton method from [15] for solving the nonsmooth equations (2.4) can be applied.

4. Numerical examples

In this section, we present the result of a numerical test for four examples which come from [16]. A comparison between our nonsmooth equations (NE) method and the SQP method is given. They are implemented in Fortran with single precision on a microcomputer. The termination accuracy is fixed at $\varepsilon < 10^{-20}$.

E x a m p l e 1. Let $f_1(x) = -2x_1 + 3x_2^2$, $f_2(x) = x_1$, $f_3(x) = x_1 - x_2^2$, $h_1(x) = x_1 - x_2^2$. The solution is $x^* = (0,0)$, $f^* = 0$. The initial points for the SQP method are (a) $x_0 = (-1,3)$, (b) $x_0 = (5,-2)$, (c) $x_0 = (10,-50)$. Applying our nonsmooth equations method, Example 1 is transformed to a system of nonsmooth equations with 7 variables. The solution is $(x^*, u^*, v^*) = (0,0,0,1,0,0,2)$. The initial points are given as (a) $(x_0, u_0, v_0) = (-1,3,9,\frac{1}{3},\frac{1}{3},\frac{1}{3},0)$, (b) $(x_0, u_0, v_0) = (5,-2,12,\frac{1}{3},\frac{1}{3},\frac{1}{3},0)$, (c) $(x_0, u_0, v_0) = (10,-50,7000,\frac{1}{3},\frac{1}{3},\frac{1}{3},0)$.

E x a m p l e 2. Let $f_1(x) = -x_1 + 10(x_1^2 + x_2^2 - 1)$, $f_2(x) = -x_1^2 - 1$, $f_3(x) = x_2^2 - 2$, $h_1(x) = x_1^2 + x_2^2 - 1$. The solution is $x^* = (1,0)$, $f^* = -1$. The initial points for the SQP method are given by (a) $x_0 = (-1,1)$, (b) $x_0 = (5,-2)$, (c) $x_0 = (10,-50)$. Using the nonsmooth equations method, Example 2 is transformed to a system of nonsmooth equations with 7 variables, the solution is $(x^*, u^*, v^*)^* = (1,0,-1,1,0,0,9.5)$. The initial points are given as (a) $(x_0, u_0, v_0) = (-1,1,11,\frac{1}{3},\frac{1}{3},\frac{1}{3},9.5)$, (b) $(x_0, u_0, v_0) = (0.8,0.6,-1,\frac{1}{3},\frac{1}{3},\frac{1}{3},0)$, (c) $(x_0, u_0, v_0) = (8,6,80,\frac{1}{3},\frac{1}{3},\frac{1}{3},0)$.

E x a m p l e 3. Let $f_1(x) = 10(x_2 - 3x_1^2)$, $f_2(x) = -f_1(x)$, $f_3(x) = 1 - x_1$, $f_4(x) = -f_3(x)$, $g_1(x) = 100(x_1^2 + x_2 - 101)$, $g_2(x) = 80(x_1^2 - x_2^2 - 79)$. The solution is $x^* = (1.0,1.0)$, $f^* = 0$. The initial points for the SQP method are given by (a) $x_0 = (-1.2,1.0)$, (b) $x_0 = (2,-20)$. Applying the nonsmooth equations method, Example 3 is reformulated as a system of nonsmooth equations with 9 variables, whose solution is $(x^*, u^*, v^*) = (1.0,1.0,0,0,0,0.5,0.5,0,0)$. The initial points are given as (a) $(x_0, u_0, v_0) = (-1.2,1.0,4.4,\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4},0,0)$, (b) $(x_0, u_0, v_0) = (2,-20,0,0,0,1,0,0,0)$.

E x a m p l e 4. Let $f_1(x) = (x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4)$, $f_2(x) = f_1(x) + 10g_1(x)$, $f_3(x) = f_1(x) + 10g_2(x)$, $f_4(x) = f_1(x) + 10g_3(x)$, $g_1(x) = (x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8)$, $g_2(x) = (x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10)$, $g_3(x) = (x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 - 5)$. The solution is $x^* = (0,1,2,-1)$, $f^* = -44$. The initial points are (a) $x_0 = (0,1,1,0)$. Applying the proposed nonsmooth equations method, Example 4 can be changed to a system of nonsmooth equations with 12 variables, the solution is $x^* = (0,1,2,-1,-44,0.7,0.1,0,0.2,0,0,0)$. The initial points are given by (a) $x_0 = (0,1,1,0,-27,\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4},0,0,0)$.

Tab. 1 is a comparison of the iterations of the examples applying the SQP method from [16] and the nonsmooth equations (NE) method proposed in this paper.

iterations	E1(a)	E1(b)	E1(c)	E2(a)	E2(b)	E2(c)	E3(a)	E3(b)	E4(a)
SQP	13	10	15	11	8	–	6	–	10
NE	3	3	6	16	8	8	3	2	10

Table 1. Results for Examples 1–4.

From Tab. 1 we observe that the NE method practically converges after much less iterations than the SQP method. However, its shortcoming is that we cannot ensure convergence theoretically. Now proving its convergence is underway.

References

[1] *F. H. Clarke, Yu. S. Ledyaev, R. J. Stern and P. R. Wolenski*: Nonsmooth Analysis and Control Theory. Springer-Verlag, New York, 1998.

[2] *V. F. Demyanov*: On a relation between the Clarke subdifferential and the quasi-differential. Vestn. Leningr. Univ., Math. *13* (1981), 183–189.

[3] *V. F. Demyanov, A. M. Rubinov*: Constructive Nonsmooth Analysis. Peter Lang, Frankfurt am Main, 1995.

[4] *F. Facchinei, A. Fischer, C. Kanzow, and J. Peng*: A simple constrained optimization reformulation of KKT systems arising from variational inequalities. Appl. Math. Optimization *40* (1999), 19–37.

[5] *Y. Gao*: Demyanov difference of two sets and optimality conditions of Lagrange multipliers type for constrained the quasidifferential optimization. J. Optimization Theory Appl. *104* (2000), 377–394.

[6] *Y. Gao*: Newton methods for solving nonsmooth equations via a new subdifferential. Math. Methods Oper. Res. *54* (2001), 239–257.

[7] *J. B. Hiriart-Urruty, C. Lemaréchal*: Convex Analysis and Minimization. Springer-Verlag, Berlin, 1993.

[8] *D. Li, N. Yamashita, and M. Fukushima*: Nonsmooth equations based BFGS method for solving KKT system in mathematical programming. J. Optimization Theory Appl. *109* (2001), 123–167.

[9] *J. V. Outrata*: Minimization of nonsmooth nonregular functions: Applications to discrete-time optimal control problems. Probl. Control Inf. Theory *13* (1984), 413–424.

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

[11] *F. A. Potra, L. Qi, and D. Sun*: Secant methods for semismooth equations. Numer. Math. *80* (1998), 305–324.

[12] *L. Qi*: Convergence analysis of some algorithms for solving nonsmooth equations. Math. Oper. Res. *18* (1993), 227–244.

[13] *L. Qi, H. Jiang*: Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton and quasi-Newton methods for solving these equations. Math. Oper. Res. *22* (1997), 301–325.

[14] *L. Qi, J. Sun*: A nonsmooth version of Newton's method. Mathematical Program., Ser. A *58* (1993), 353–367.

[15] *D. Sun, J. Han*: Newton and quasi-Newton methods for a class of nonsmooth equations and related problems. SIAM J. Optim. *7* (1997), 463–480.

[16] *Y. H. Yu, L. Gao*: Nonmonotone line search algorithm for constrained minimax problems. J. Optimization Theory Appl. *115* (2002), 419–446.

Authors' addresses: *Y. Gao*, School of Management, University of Shanghai for Science and Technology, Jungong Road 516, Shanghai 200093, P.R. China, e-mail: `gaoyan1962@263.net`; *X. Li*, Department of Applied Mathematics, School of Science, Beijing Institute of Technology, Zhongguanchun Road 5, Beijing 100081, P.R. China, e-mail: `bj-lixuewen@hotmail.com`.