Control and Cybernetics

Vol. 19 (1990) No. 1-2

A method of feasible directions for certain quasi-differentiable inequality constrained minimization problems

by

KRZYSZTOF C. KIWIEL

Systems Research Institute
Polish Academy of Sciences
Newelska 6
01-447 Warsaw, Poland

An algorithm is given for inequality constrained minimization problems in which objectives and constraints are smooth compositions of max-type functions. At each iteration several search directions are found by solving several quadratic programming subproblems. Then an Armijo-type search is performed simultaneously along these directions to produce the next approximation to a solution. The algorithm is readily implementable and globally convergent to inf-stationary points.

AMS 1980 subject classification. Primary: 65KO5. Secondary 9OC25

KEY WORDS: Nondifferentiable optimization, Quasidifferential calculus, Constrained optimization, Descent methods. Feasible directions methods.

1. Introduction

We are concerned with methods for solving a nondifferentiable and nonconvex problem of the form

$$\text{minimize } f(x), \text{ subject to } g(x) \leqslant 0, \tag{1.1}$$

where

$$f(x) = \tilde{f}(x, h(x)), \tag{1.2a}$$

$$g(x) = \tilde{g}(x, h(x)), \tag{1.2b}$$

$$h(x) = (h_1(x), \dots, h_M(x)),$$

$$h_i(x) = \max\{h_{ji}(x) : j \in J_i\} \quad \text{for} \quad i \in I,$$

the functions $\tilde{f} : R^N \times R^M \to R$, $\tilde{g} : R^N \times R^M \to R$ and $h_{ji} : R^N \to R$ are continuously differentiable, $I: = \{1,...,M\}$ and J_i, $i \in I$, are nonempty finite sets of indices. We shall also consider problems with objectives and constraints given by pointwise maxima of functions of the form (1.2). Such problems arise in many applications (e.g. minimax problems, l_1 and l_∞ approximation problems, exact penalty methods) and have been treated in several papers; see, e.g. [1], [2], [3], [4], [5], [6], [8], [11].

Several algorithms for problem (1.1) are known. The methods of [3], [11], which treat the original problem indirectly by solving an infinite sequence of unconstrained differentiable problems, converge under additional assumptions which may be difficult to check a priori. The algorithm of [1], as well as general purpose non smooth optimization methods (e.g. [7], [10]), can be used when the functions $\tilde{f}(x, y_1,...,y_M)$ and $\tilde{g}(x, y_1,...,y_M)$ are nondecreasing with respect to each y_i, $i \in I$ (see also [8]). If this assumption fails, e.g.

$$g(x) = -\max\{h_{j1}(x) : j \in J_1\} = \min\{-h_{j1}(x) : j \in J_1\},$$

only the method of [4] may solve (approximately) problem (1.1).

This paper presents a method of feasible directions that is tailored to the structure of (1.1). The algorithm generalizes one given in [9] for the unconstrained case. At each iteration several search directions are found by solving several quadratic programming subproblems. Then an Armijo-type search is performed simultaneously along all the search directions to produce the next improved estimate of a solution. The algorithm is "globally" convergent in the sense that each of its accumulation points is inf-stationary for problem (1.1) (see Section 2 for the definition).

Our algorithm differs from its predecessor of [4] in two aspects. First, its line search procedure needs only a finite number of function evaluations, whereas [4] requires exact directional minimizations. Secondly, we have modified the direction finding subproblems of [4] which ensure convergence to only approximately inf-stationary points. In effect, our algorithm seems to be the first readily implementable and globally convergent method for the problem in question.

The method is derived and stated in Section 2. Its global cnvergence is established in Section 3. Extensions to more general problems are discussed in Section 4. Finally, we have a conclusion section.

R^N denotes the N-dimensional Euclidean space with the usual inner product $< \cdot, \cdot >$ and the associated norm $|\cdot|$. Superscripts are used to denote different vectors, eg. x^1 and x^2. All vectors are row vectors.

2. Derivation of the method

We start by reviewing well-known properties of problem (1.1) (see [1], [2], [4] for details). Let $S = \{x \in R^N : g(x) \leqslant 0\}$ denote the feasible set. For any fixed $x \in R^N$, let

$$H(y, x) = \max\{f(y) - f(x), g(y)\} \quad \text{for all } y \in R^N$$

denote the improvement function. To justify this name, suppose one can find y such that $H(y, x) < H(x, x)$. Then y is better than x, since if $H(x, x) = g(x) > 0$ then $g(y) < g(x)$, whereas if $H(x, x) = g(x)_+ = 0$ then $f(y) < f(x)$ and $y \in S$, where $g(x)_+ = \max\{g(x), 0\}$. It follows that any local solution $\bar{x} \in S$ of (1.1) is a local unconstrained minimum point of $H(\cdot, \bar{x})$. In particular, we have

$$H'(\bar{x}, \bar{x}; d) \geqslant 0 \quad \text{for all} \quad d \in R^N, \tag{2.1}$$

where

$$H'(x, \bar{x}; d) = \lim_{t \downarrow 0} [H(x + td, \bar{x}) - H(x, \bar{x})]/t$$

denotes the derivative of $H(\cdot, \bar{x})$ at x in the direction d.
Note that

$$H'(x, \bar{x}; d) = \begin{cases} f'(x; d) & \text{if} \quad g(x) < 0, \\ \max\{f'(x; d), g'(x; d)\} & \text{if} \quad g(x) = 0, \\ g'(x; d) & \text{if} \quad g(x) > 0. \end{cases} \tag{2.2}$$

Points $\bar{x}$ satisfying (2.1) are called inf-stationary for problem (1.1).

We shall need the following notation. For $z = (x, y) \in R^N \times R^M$ we denote by $\nabla\tilde{f}(x, y)$ the N-vector $(\frac{\partial \tilde{f}}{\partial z_1}(z), \ldots, \frac{\partial \tilde{f}}{\partial z_N}(z))$, while $\frac{\partial \tilde{f}}{\partial y_i}(x, y)$ denotes $\frac{\partial \tilde{f}}{\partial z_{i+N}}(z)$, $i \in I$. For $x \in R^N$ and $i \in I$, let

$$a_i^f(x) = \frac{\partial \tilde{f}}{\partial y_i}(x, h(x)),$$

$$b^f(x) = \nabla\tilde{f}(x, h(x)).$$

Then

$$f'(x; d) = \langle b^f(x), d \rangle + \sum_{i \in I} a_i^f(x) h_i'(x; d) =$$

$$= \langle b^f(x), d \rangle + \sum_{i \in I} a_i^f(x) \max_{j \in J_i(x)} \langle \nabla h_{ji}(x), d \rangle,$$

so that

$$f'(x;d) = \langle b^f(x), d\rangle + \Sigma_{i\in I_+^f(x)} \max_{j\in J_i(x)} \langle a_i^f(x)\nabla h_{ji}(x), d\rangle +$$

$$+ \Sigma_{i\in I_-^f(x)} \min_{j\in J_i(x)} \langle a_i^f(x)\nabla h_{ji}(x), d\rangle,$$

where

$$J_i(x) = \{j\in J_i \ : \ h_{ji}(x) = h_i(x)\}, \qquad i\in I,$$

$$I_+^f(x) = \{i\in I \ : \ a_i^f(x) > 0\},$$

$$I_-^f(x) = \{i\in I \ : \ a_i^f(x) < 0\}$$

and the summation over an empty index set yields zero. Observe that $f'(\cdot;d)$ may be discontinuous if so are $J_i(\cdot)$. Therefore a better model of $f(x+d)-f(x)$ is given by the family of functions

$$\hat{f}(d;x,w^f,\delta) = \langle b^f(x),d\rangle + \Sigma_{i\in I_+^f(x)} a_i^f(x) \max_{j\in J_i(x,\delta)} [h_{ji}(x) - h_i(x) +$$

$$+ \langle \nabla h_{ji}(x),d\rangle] + \langle w^f, d\rangle$$

parametrized by w^f in

$$B^f(x,\delta) = \{w^f \in R^N : w^f = \Sigma_{i\in I_-^f(x)} a_i^f(x)\nabla h_{ji}(x)$$

$$\text{for some} \qquad j\in J_i(x,\delta)\}, \tag{2.3}$$

where the use of

$$J_i(x,\delta) = \{j\in J_i \ : \ h_{ji}(x) \geqslant h_i(x) - \delta\}$$

with a fixed "anticipation" tolerance $\delta > 0$ may take into account possible changes in $J_i(\cdot)$ around x, as will be explained later. Note that

$$f'(x;d) = \min\{\hat{f}(d;x,w^f,0) \ : \ w^f\in B^f(x,0)\}.$$

Replacing "f" by "g" in the preceding paragraph, we may approximate

$g(x+d)-g(x)$ by functions $\hat{g}(d;x,w^g,\delta)$ with w^g in $B^g(x,\delta)$. Hence for $w=(w^f,w^g)\in B^f(x,\delta)\times B^g(x,\delta)=:B(x,\delta)$ functions

$$\hat{H}(d;x,w,\delta)=\max\{\hat{f}(d;x,w^f,\delta),g(x)+\hat{g}(d;x,w^g,\delta)\}-g(x)_+$$

may serve as models of $H(x+d,x)-H(x,x)$. Since we want $x+d$ to improve on x, we may choose $d(w)$ to

$$\text{minimize } \hat{H}(d;x,w,\delta)+\frac{1}{2}|d|^2 \quad \text{over all} \quad d\in R^N \tag{2.4}$$

where the term $|d|^2/2$ ensures that $d(w)$ stays in the region where $\hat{H}(\cdot;x,w,\delta)$ is a close approximation to $H(x+\cdot,x)-H(x,x)$. In view of the strong convexity of the objective function of (2.4), the unique solution $d(w)$ of (2.4) exists and satisfies (see [6], [8])

$$d(w)=\mu^f(w)d^f(w)+\mu^g(w)d^g(w) \tag{2.5a}$$

for some

$$\mu^f(w)\geqslant 0,\quad \mu^g(w)\geqslant 0,\quad \mu^f(w)+\mu^g(w)=1, \tag{2.5b}$$

where

$$d^f(w)=-[b^f(x)+\Sigma_{i\in I_+^f(x)}a_i^f(x)\Sigma_{j\in J_i(x,\delta)}\lambda_{ji}^f(w)\nabla h_{ji}(x)+w^f], \tag{2.5c}$$

for some

$$\begin{aligned}&\lambda_{ji}^f(w)\geqslant 0 \quad \text{for} \quad j\in J_i(x,\delta),\\ &\Sigma_{j\in J_i(x,\delta)}\lambda_{ji}^f(w)=1, \quad \text{for} \quad i\in I_+^f(x)\end{aligned} \tag{2.5d}$$

and $d^g(w)$ satisfies a similar relation. Moreover, since

$$\hat{H}(d(w);x,w,\delta)+\frac{1}{2}|d(w)|^2\leqslant \hat{H}(0;x,w,\delta)+\frac{1}{2}|0|^2=0,$$

we have

$$\hat{H}(d(w);x,w,\delta)\leqslant -\frac{1}{2}|d(w)|^2 \tag{2.6}$$

and, since the preceding definitions imply

$$H'(x,x;d)\leqslant \hat{H}(d;x,w,\delta) \quad \text{if} \quad w\in B^f(x,0)\times B^g(x,0), \tag{2.7}$$

we see that $d(w)$ is a descent direction for $H(\cdot,x)$ at x if $d(w)\neq 0$ and $w\in B(x,0)$.

We may now state the method in detail.

ALGORITHM 2.1.

Step 0 (Initialization). Select a starting point $x^1\in R^N$, a final accuracy tolerance $\varepsilon_f\geqslant 0$, an anticipation tolerance $\delta>0$ and a line search parameter $m>0$. Set $k=1$.

Step 1 (Direction finding). For each $w\in B^f(x,\delta)\times B^g(x,\delta)$, find $d(w)$ from the solution $(d(w),u(w);u_i^f(w),i\in I_+^f(x);u_i^g(w),i\in I_+^g(w))$ to the quadratic programming subproblem

$$\min_{d,u,u_i^f,u_i^g}\frac{1}{2}|d|^2+u,$$

$$\text{s.t.}\quad \langle b^{fk},d\rangle+\sum_{i\in I_+^{fk}}a_i^{fk}u_i^f+\langle w^f,d\rangle\leqslant u,$$

$$g^k+\langle b^{gk},d\rangle+\sum_{i\in I_+^{gk}}a_i^{gk}u_i^g+\langle w^g,d\rangle\leqslant u,$$

$$h_{ji}^k-h_i^k+\langle\nabla h_{ji}^k,d\rangle\leqslant u_i^f\quad\text{for}\quad j\in J_i^k,\ \ i\in I_+^{fk},$$

$$h_{ji}^k-h_i^k+\langle\nabla h_{ji}^k,d\rangle\leqslant u_i^g\quad\text{for}\quad j\in J_i^k,\ \ i\in I_+^{gk},$$

where $x=x^k$, $b^{fk}=b^f(x^k)$, $I_+^{fk}=I_+^f(x^k)$, $I_i^k=I_i^k(x^k,\delta)$ etc.

Step 2 (Stopping criterion). If $|d(w)|\leqslant\varepsilon_f$ for all $w\in B^f(x,0)\times B^g(x,0)$, terminate. Otherwise, set

$$u^k=-\max\{|d(w)|^2: w\in B(x^k,\delta)\}\tag{2.8}$$

and continue.

Step 3 (Stepsize selection).

(i) Set $t=1$.

(ii) Find $\widetilde{w}$ in $B(x^k,\delta)$ such that

$$H(x^k+td(\widetilde{w}),x^k)=\min\{H(x^k+td(w),x^k): w\in B(x^k,\delta)\}.$$

(iii) If

$$H(x^k + td(\tilde{w}), x^k) \leqslant H(x^k, x^k) + m(t)^2 u^k,$$

set $t^k = t$, $d^k = d(\tilde{w})$, $x^{k+1} = x^k + t^k \tilde{d}^k$ and go to Step 4; otherwise replace t by $t/2$ and go to Step 3 (ii).

Step 4 Increase k by 1 and go to Step 1.

Observe that the algorithm cannot cycle infinitely at Step 3, since Step 3 is always entered with $d(\hat{w}) \neq 0$ for some $\hat{w} \in B(x^k, 0)$, so that $t \downarrow 0$ and

$$H'(x^k, x^k; d(\hat{w})) \geqslant \liminf_{t \downarrow 0} \; [\min_{w \in B(x^k, \delta)} H(x^k + td(w), x^k) -$$

$$- H(x^k, x^k)]/t \geqslant \lim_{t \downarrow 0} \; mtu^k = 0$$

would contradict (2.6) and (2.7). Moreover, $H(x^{k+1}, x^k) < H(x^k, x^k)$. This means that if $g(x^1) > 0$ then the algorithm decreases constraint violation (without unduly increasing objective values) until a feasible x^k is found; then the successive points stay feasible and $f(x^k)$ decreases monotonously.

In order to see why the algorithm has to use a positve anticipation tolerance δ, suppose that $f(x) = x$ and $g(x) = -|x| = \min\{x, -x\}$ for $x \in R^1$, $x^1 = 1$, $\varepsilon_f = 0$ and $m = 0.1$. For $\delta = 0$ we would get $x^k = 1/2^{k-1}$ for all k, whereas for $\delta = 1$ we have $x^2 = 0.5$, $B^g(x^2, 1) = \{1, -1\}$, $d(0, 1) = -1$, $d(0, -1) = -0.25$ and the algorithm "jumps" over the nonstationary point $\bar{x} = 0$ to $x^3 = -0.5$, continuing with $x^k \to -\infty$ and $f(x^k) \downarrow -\infty$.

3. Convergence

In this section we shall establish global convergence of the method. In the absence of convexity, we will content ourselves with finding an inf-stationary point of problem (1.1). Naturally, we assume that the final accuracy tolerance ε_f is set to zero.

We start by analyzing properties of search directions generated around nonstationary points.

LEMMA 3.1. *Suppose that $\bar{x} \in R^N$, $\bar{w} \in B(\bar{x}, 0)$ and $\bar{d} \in R^N$ are such that $\hat{H}(\bar{d}; \bar{x}, \bar{w}, 0) < 0$. Then there exist $\bar{\varepsilon} > 0$ and neighborhoods $S(\bar{x})$ and $S(\bar{w})$ of $\bar{x}$ and $\bar{w}$, respectively, such that*

$$H'(\bar{x}, \bar{x}; d(x, w)) \leqslant -\bar{\varepsilon} \quad \text{for all } (x, w) \in S(\bar{x}) \times S(\bar{w}), \tag{3.1}$$

$$|d(x, w)| \geqslant \bar{\varepsilon} \quad \text{for all } (x, w) \in S(\bar{x}) \times S(\bar{w}), \tag{3.2}$$

where $d(x, w)$ denotes the solution of (2.4).

P r o o f. By assumption, for some $\varepsilon > 0$ we have

$$-g(\bar{x})_+ + \langle b^f(\bar{x}), \bar{d}\rangle + \sum_{i \in I_+^f(\bar{x})} a_i^f(\bar{x}) \max_{j \in J_i(\bar{x})} \langle \nabla h_{ji}(\bar{x}), \bar{d}\rangle +$$

$$+ \langle \bar{w}^f, \bar{d}\rangle < -\varepsilon \tag{3.3a}$$

$$g(\bar{x}) - g(\bar{x})_+ + \langle b^g(\bar{x}), \bar{d}\rangle + \sum_{i \in I_+^g(\bar{x})} a_i^g(\bar{x}) \max_{j \in J_i(\bar{x})} \langle \nabla h_{ji}(\bar{x}), \bar{d}\rangle +$$

$$+ \langle \bar{w}^g, \bar{d}\rangle < -\varepsilon \tag{3.3b}$$

and $\bar{d} \neq 0$ with $\bar{w} = (\bar{w}^f, \bar{w}^g)$. Hence, using the continuity of the functions involved in (3.3a), we may choose $S(\bar{x}) \times S(\bar{w})$ such that

$$I_+^f(\bar{x}) \subset I_+^f(x) \quad \text{and} \quad I_+^f(x) \setminus I_+^f(\bar{x}) \subset \{ i \in I : a_i^f(\bar{x}) = 0 \},$$

$$\langle b^f(x) + w^f, \bar{d}\rangle + \sum_{i \in I_+^f(\bar{x})} a_i^f(\bar{x}) \max_{j \in J_i(\bar{x})} \langle \nabla h_{ji}(x), \bar{d}\rangle \leqslant -\varepsilon/2 + g(x)_+ \tag{3.4}$$

for all $x \in S(\bar{x})$, $w = (w^f, w^g) \in S(\bar{w})$. Next, since h_i and h_{ji} are continuous, while $\delta > 0$ is fixed, we have $J_i(\bar{x}) \subset J_i(x, \delta)$ and $h_{ji}(x) - h_i(x) < -\tilde{\varepsilon}$ for some fixed $\tilde{\varepsilon} > 0$ if x is close to $\bar{x}$ and $j \in J_i(x, \delta) \setminus J_i(\bar{x})$, so we may shrink $S(\bar{x})$ and choose small $\bar{t} \in (0, 1)$ such that

$$\max_{j \in J_i(\bar{x})} \langle \nabla h_{ji}(x), \bar{t}\bar{d}\rangle \geqslant \max_{j \in J_i(\bar{x})} -\bar{t}|\nabla h_{ji}(x)|\,|\bar{d}| > -\tilde{\varepsilon}/2 >$$

$$> h_{ji}(x) - h_i(x) + \langle \nabla h_{ji}(x), \bar{t}\bar{d}\rangle$$

for any $j \in J_i(x, \delta) \setminus J_i(\bar{x})$, $x \in S(\bar{x})$. Hence we may multiply (3.4) by $\bar{t}$, use the preceding inequality and subtract $g(x)_+$ to obtain

$$-g(x)_+ + \langle b^f(x), \hat{d}\rangle + \sum_{i \in I_+^f(x)} a_i^f(x) \max_{j \in J_i(x, \delta)} [h_{ji}(x) - h_i(x) +$$

$$+ \langle \nabla h_{ji}(x), \hat{d}\rangle] + \langle w^f, \hat{d}\rangle \leqslant -\hat{\varepsilon} - (1-\bar{t})g(x)_+$$

for $\hat{d} = \bar{t}\bar{d}$, $\hat{\varepsilon} = \bar{t}\varepsilon/2 > 0$ and all $x \in S(\bar{x})$, $w \in S(\bar{w})$. But $(1 - \bar{t})g(x)_+ \geqslant 0$, so we obtain

$$\hat{f}(\hat{d};x,w^f,\delta)-g(x)_+\leqslant-\hat{\varepsilon}$$

and, since the preceding arguments apply also to (3.3b) (with $g(x)_+$ replaced by $g(x)_+-g(x)\geqslant 0$), we deduce that

$$g(x)+\hat{g}(\hat{d};x,w^g,\delta)-g(x)_+\leqslant-\hat{\varepsilon}.$$

Thus $H(\hat{d};x,w,\delta)\leqslant-\hat{\varepsilon}$ and, since $\hat{H}(\cdot;x,w,\delta)$ is convex and $\hat{H}(0;x,w,\delta)=0$, $\hat{H}(t\hat{d};x,w,\delta)\leqslant-t\hat{\varepsilon}$ for all $t\in[0,1]$, so, since $d(x,w)$ solves (2.4) and $\hat{\varepsilon}$ may be decreased,

$$\hat{H}(d(x,w);x,w,\delta)\leqslant\min_{t\in[0,1]}\{\hat{H}(t\hat{d};x,w,\delta)+\frac{1}{2}|t\hat{d}|^2\}\leqslant-\hat{\varepsilon}/2|\hat{d}|^2$$

and therefore

$$\hat{f}(d(x,w);x,w^f,\delta)\leqslant-\hat{\varepsilon}/2|\hat{d}|^2+g(x)_+,$$

$$\hat{g}(d(x,w);x,w^g,\delta)\leqslant-\hat{\varepsilon}/2|\hat{d}|^2+g(x)_+-g(x) \tag{3.5}$$

for all $x\in S(\bar{x})$, $w\in S(\bar{w})$. By (2.5), $d(\cdot,\cdot)$ is locally bounded, so we may use (3.5) and the continuity relations between f, f', g and g' established in the proof of Lemma 2 in [9] together with the continuity of g and $g(\cdot)_+$ for shrinking $S(\bar{x})\times S(\bar{w})$ to obtain

$$f'(\bar{x};d(x,w))\leqslant-\bar{\varepsilon}+g(\bar{x})_+,$$

$$g'(\bar{x};d(x,w))\leqslant-\bar{\varepsilon}+g(\bar{x})_+-g(\bar{x}) \tag{3.6}$$

for $\bar{\varepsilon}=\hat{\varepsilon}/4|\hat{d}|^2$ and all $x\in S(\bar{x})$, $w\in S(\bar{w})$. (3.6) and (2.2) yield (3.1). Since $H'(\bar{x},\bar{x};\cdot)$ is continuous and $H'(\bar{x},\bar{x};0)=0$, (3.1) implies (3.2) for small $\bar{\varepsilon}>0$.

Our stopping criterion is justified below.

LEMMA 3.2. *If Algorithm* 2.1 *terminates at the* k*-th iteration, then* $\bar{x}=x^k$ *is inf-stationary for problem* (1.1).

P r o o f. For contradiction purposes, suppose that $\bar{x}$ is nonstationary, but $d(\bar{x},w)=0$ for all $w\in B(\bar{x},0)$. Then there are $\bar{w}\in B(\bar{x},0)$ and $\bar{d}\in R^N$ such that $\hat{H}(\bar{d};\bar{x},\bar{w},0)=H'(\bar{x},\bar{x};\bar{d})<0$, so Lemma 3.1 yields $d(\bar{x},\bar{w})\neq\emptyset$, a contradiction.

Our main result is

THEOREM 3.3. *Every accumulation point of an infinite sequence* $\{x\xrightarrow{k}\}$ *generated by Algorithm* 2.1 *is inf-stationary for problem* (1.1).

P r o o f. Suppose, for contradiction purposes, that there exist a nonstationary point $\bar{x}$ and an infinite set $K \subset \{1,2,...\}$ such that $x^k \xrightarrow{K} \bar{x}$. By (2.2) and Lemma 3.1, there are $\bar{w} \in B(\bar{x},0)$ and $\bar{\varepsilon} > 0$ such that (3.1) and (3.2) hold for some $S(\bar{x}) \times S(\bar{w})$. Since $x^k \xrightarrow{K} \bar{x}$ and $\delta > 0$ is fixed, it is easy to see from (2.3) that $B(x^k,\delta) \cap S(\bar{w}) \neq \varnothing$ for large $k \in K$, so there exist $w^k \in B(x^k,\delta)$ and $d^k = d(x^k,w^k)$ such that, by (3.1),

$$H'(\bar{x},\bar{x};d^k) \leqslant -\bar{\varepsilon} \quad \text{for large} \quad k \in K, \tag{3.7}$$

$$|d^k| \geqslant \bar{\varepsilon} \quad \text{for large} \quad k \in K. \tag{3.8}$$

Since $x^k \xrightarrow{K} \bar{x}$, (2.5) and (2.8) imply the existence of $\bar{u} < 0$ such that $\bar{u} \leqslant u^k \leqslant 0$ for all k; in particular, $\{d^k\}_{k \in K}$ is bounded. Hence one may use Taylor's expansion to show that

$$H(\bar{x} + t d^k, \bar{x}) \leqslant H(\bar{x}, \bar{x}) + t H'(\bar{x},\bar{x};d^k) + o(t,k),$$

where $o(t,k)/t \to 0$ as $t \downarrow 0$ uniformly with respect to $k \in K$. Combining this with (3.7), we get

$$\max\{f(\bar{x} + t d^k) - f(\bar{x}), g(\bar{x} + t d^k)\} \leqslant g(\bar{x})_+ - \bar{\varepsilon} t + o(t,k)$$

for large $k \in K$. Therefore, using the continuity of f and g, the boundedness of $\{d^k\}_{k \in K}$ and the fact that $x^k \xrightarrow{K} \bar{x}$, for any $\varepsilon > 0$ and $\tilde{\varepsilon} \in (0,\bar{\varepsilon})$ we may choose $t(\tilde{\varepsilon})$ such that

$$\max\{f(x^k + t d^k) - f(x^k), g(x^k + t d^k)\} - g(x^k)_+ \leqslant$$

$$\leqslant \max\{f(\bar{x} + t d^k) - f(\bar{x}), g(\bar{x} + t d^k)\} - g(\bar{x})_+ + \varepsilon \leqslant$$

$$\leqslant -\bar{\varepsilon} t + o(t,k) + \varepsilon \leqslant -\bar{\varepsilon} t + \tilde{\varepsilon} t + \varepsilon$$

to obtain for $\hat{\varepsilon} = \bar{\varepsilon} - \tilde{\varepsilon} > 0$

$$H(x^k + t d^k, x^k) \leqslant H(x^k, x^k) + \varepsilon - \hat{\varepsilon} t \tag{3.9}$$

for all $t \in [0, t(\tilde{\varepsilon})]$ and large $k \in K$. Let us choose ε such that the interval $[\underline{t}(\varepsilon), \bar{t}(\varepsilon)]$ of solutions to the inequality

$$\varepsilon - \hat{\varepsilon} t \leqslant m(t)^2 \bar{u} \tag{3.10}$$

contains $1/2^i \leqslant t(\tilde{\varepsilon})$ for some $i \rangle 0$. This is possible, since $[\underline{t}(\varepsilon), \bar{t}(\varepsilon)] \to [0, -\hat{\varepsilon}/m\bar{u}]$ as $\varepsilon \downarrow 0$. Then $t = 1/2^i$ satisfies, by (3.9) – (3.10) and the fact that $m\bar{u} \leqslant mu^k$ for $k \in K$,

$$H(x^k + td^k, x^k) \leqslant H(x^k, x^k) + m(t)^2 u^k \quad \text{for large} \quad k \in K.$$

Therefore, by construction, $t^k \geqslant t$ and

$$\max\{f(x^{k+1}) - f(x^k), g(x^{k+1})\} \leqslant g(x^k)_+ + m(t)^2 u^k \tag{3.11}$$

for large $k \in K$. First, suppose that $g(x^k) \leqslant 0$ for some k. Then $g(x^k)_+ = 0$ and $f(x^{k+1}) < f(x^k)$ for large k by construction, so that $f(x^{k+1}) - f(x^k) \to 0$ because $x^k \xrightarrow{K} \bar{x}$ and f is continuous. Hence (3.11) and the negativity of u^k yield $u^k \xrightarrow{K} 0$. On the other hand, if $g(x^k)_+ = g(x^k) > 0$ for all k, then $g(x^k) \downarrow g(\bar{x})$ and (3.11) again implies $u^k \xrightarrow{K} 0$. But $-u^k \geqslant |d^k|^2 \geqslant \bar{\varepsilon}^2$ for large $k \in K$ from (2.8) and (3.8). This contradiction completes the proof.

REMARK 3.4. The algorithm's accumulation points need not be feasible, e.g. when problem (1.1) has no feasible points. This will not occur if the following constraint qualification is satisfied: for each x such that $0 < g(x) \leqslant g(x^1)$ one has $g'(x;d) < 0$ for some d. (It holds trivially if $g(x^1) \leqslant 0$.)

4. Extensions

It is easy to extend the preceding results to the case when

$$f(x) = \max\{\tilde{f}_l(x, h(x)) : l \in L^f\},$$

$$g(x) = \max\{\tilde{g}_l(x, h(x)) : l \in L^g\}.$$

with continuously differentiable $\tilde{f}_l$, $\tilde{g}_l$ and finite L^f, L^g.
To this end, define $\hat{f}_l(d; x, w^{fl}, \delta)$ for each $\tilde{f}_l(x, h(x))$ as we did in Section 2 for $\tilde{f}(x, h(x))$, and use

$$\hat{f}(d; x, w^f, \delta) = \max_{l \in L^f(x, \delta)} [\tilde{f}_l(x, h(x)) - f(x) + \hat{f}_l(d; x, w^{fl}, \delta)]$$

with $w^f = \bigcap_{l \in L^f(x,\delta)} w^{fl}$ and $L^f(x, \delta) = \{l \in L^f : \tilde{f}_l(x, h(x)) \geqslant f(x) - \delta\}$ as

a model of $f(x+d) - f(x)$; the model $\hat{g}(d; x, w^g, \delta)$ of $g(x+d) - g(x)$ is defined analogously.

The resulting algorithm is essentially the same, although it uses more complicated quadratic programming subproblems (see also [8].) Straightforward extensions of all the preceding convergence results are left to the interested reader.

5. Conclusions

We have extended the method of [9] to the case of inequality constrained minimization problems with certain quasi differentiable functions. The algorithm compares favorably with its predecessor of [4], since it does not require exact minimizations at line searches and converges to inf-stationary points.

References

[1] AUSLENDER A. Minimisation de fonctions localement Lipschitziennes: applications a la programmation mi-convexe, midifferentiable, in "Nonlinear Programming 4" (O.L.Mangasarian, R.R. Meyer, and S.M. Robinson, Eds.), pp. 429-460, Academic Press, New York, 1981.

[2] BEN-TAL A., ZOWE J. Necessary and sufficient optimality conditions for a class of nonsmooth minimization problems. *Math. Programming* **24** (1982), 70-91.

[3] BERTSEKAS D. Approximation procedures based on the method of multipliers. *J. Optim. Theory Appl.* **23** (1977), 487-510.

[4] DEMYANOV V.F., GAMIDOV S., SIVELINA T.I. An Algorithm for Minimizing a Certain Class of Quasidifferentiable Functions, Mathematical Programming Study, **29** (1986) 74-84.

[5] FLETCHER R. Practical Methods of Optimization, Vol. II, Constrained Optimization. Wiley, New York, 1981.

[6] KIWIEL K.C. A phase I - phase II method for inequality constrained minimax problems. *Control Cyb.* **12** (1983), 55-75.

[7] KIWIEL K.C. A linearization algorithm for constrained non-smooth minimization, in "System Modelling and Optimization" (P. Thoft-Christensen, ED.), pp. 311-320, Lecture Notes in Control and Information Sciences 59, Springer, Berlin, 1984.

[8] KIWIEL K.C. A quadratic approximation method for minimizing a class of quasidifferentiable functions. *Numer. Math.* **45** (1984), 411-430.

[9] KIWIEL K.C. A method of linearizations for minimizing certain quasidifferentiable functions, in "Quasidifferential Calculus" (V.F. Demyanov, and L.C.W. Dixon, Eds.), pp. 85-94, Mathematical Programming Study 29, North-Holland, Amsterdam 1986.

[10] MIFFLIN R. A modification and an extension of Lemarechal's algorithm for nonsmooth minimization, in "Nondifferential and Variational Techniques in Optimization" (D.C. Sorensen, and R.J.-B. Wets, Eds.), pp. 77-90, Mathematical Programming Study 17, North-Holland, Amsterdam, 1982

[11] PAPAVASSILOPOULOS G. Algorithms for a class of nondifferentiable problems *J. Optim. Theory Appl.* **34** (1981), 31-82.

Received, January 1986.

Metoda kierunków dopuszczalnych dla pewnych zadań optymalizacji quasiróżniczkowalnej z ograniczeniami nierównościowymi

W pracy przedstawiono algorytm dla zadań minimalizacji z ograniczeniami nierównościowymi, w których funkcje celu i ograniczeń są gładkimi superpozycjami funkcji typu maksimum. W każdej iteracji metoda znajduje kilka kierunków poszukiwań, będących rozwiązaniami pomocniczych zadań programowania kwadratowego. Jednoczesne przeszukanie tych kierunków zgodnie z regułą Armijo daje kolejne przybliżenie rozwiązania. Algorytm ten jest łatwo implementowalny i globalnie zbieżny do punktów inf-stacjonarnych.

Метод допустимых направлений для некоторых задач квази-дифференцируемой оптимизации с ограничениями в виде неравенств

В работе представлен алгоритм для задач минимизации с ограничениями в виде неравенств, в которых функции цели и ограничений являются гладкими супперпозициями функций типа максимума. В каждой итерации метод находит несколько направлений поиска, являющихся решениями вспомогательных задач квадратного программирования. Одновременный поиск в этих направлениях согласно правилу Армие дает очередное приближение решения. Этот алгоритм легко применяем и глобально сходим к инф-стационарным точкам.