NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

ON THE STEEPEST DESCENT METHOD FOR A CLASS OF QUASI-DIFFERENTIABLE OPTIMIZATION PROBLEMS

Diethard Pallaschke*
Peter Recht*

December 1984
CP-84-57

*Institute of Statistics and Mathematical Economy,
University of Karlsruhe, Karlsruhe, F.R.G..

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

PREFACE

In a recent paper, V.F. Demyanov, S. Gamidov and I. Sivelina developed an algorithm for solving optimization problems, given by smooth compositions of max-type functions.

In this paper the authors apply this algorithm to a larger class of quasidifferentiable functions.

This paper is a contribution to research on nondifferentiable optimization currently underway with the System and Decision Sciences Program.

A.B. Kurzhanskii

Chairman
System and Decision
Sciences Program

On the steepest descent method for a class of quasi-differentiable optimization problems

D.Pallaschke, P.Recht

0. Introduction

In a recent paper V.F.Demyanov, S.Gamidov and T.J.Sivelina presented an algorithm for solving a certain type of quasidifferentiable optimization problems [3].
More precisely, they considered the class $\mathcal{F}$ of all functions given by

$$\mathcal{F} = \{f:\mathbb{R}^n \longrightarrow \mathbb{R} \mid f(x) = F(x, y_1(x), \dots, y_m(x))\} ,$$

where

$$y_i : \mathbb{R}^n \longrightarrow \mathbb{R} \quad \text{is defined by}$$

$$y_i(x) = \max_{j \in I_i} \phi_{ij}(x) \qquad I_i = 1,\dots,N_i ;\ i=1,\dots,m$$

and

$$\phi_{ij} : \mathbb{R}^n \longrightarrow \mathbb{R} \quad \text{for all } i \in \{1,\dots,m\} \text{ and all } j \in I_i .$$

The functions F and ϕ_{ij} under consideratiens are assumed to belong to the classes $C_1(\mathbb{R}^{n+m})$ and $C_1(\mathbb{R}^n)$ respectively.
The optimization problem consists in minimizing a function $f \in \mathcal{F}$ under constraints.

In this paper we will apply the minimization algorithm of [3] to another class of quasidifferentiable functions.
We are able to prove for this type of optimization problems a convergence theorem similar to that in [3].

1. Steepest descent method

We will shortly recall the steepest descent algorithm for minimizing a quasidifferentiable function in the unconstrained case.

Let $f:\mathbb{R}^n \longrightarrow \mathbb{R}$ be a quasidifferentiable function.

Then for every $\tilde{x}\in\mathbb{R}^n$ there exist two compact, convex sets $\overline{\partial} f|_{\tilde{x}}$ and $\underline{\partial} f|_{\tilde{x}}$, such, that for every $g\in\mathbb{R}^n$, $\|g\|_2 = 1$, the directional derivative is given by:

$$\left.\frac{df}{dg}\right|_{\tilde{x}} = \max_{v\in\underline{\partial} f|_{\tilde{x}}} \langle v,g\rangle + \min_{w\in\overline{\partial} f|_{\tilde{x}}} \langle w,g\rangle \quad .$$

Here $\langle,\rangle$ denotes the canonical inner product in $\mathbb{R}^n$.

In terms of these two sets, a steepest descent direction for f at $\tilde{x}$ is given by

$$g|_{\tilde{x}} := g(\tilde{x}) \quad = - \frac{v_o+w_o}{\|v_o+w_o\|_2}$$

with

$$\|v_o+w_o\|_2 = \max_{w\in\overline{\partial} f|_{\tilde{x}}} \left(\min_{v\in\underline{\partial} f|_{\tilde{x}}} \|v+w\|_2\right).$$

Now, in the steepest descent algorithm, we start with an arbitrary point $x_o\in\mathbb{R}^n$.

Let us assume that for $k\geq 0$ the point $x_k\in\mathbb{R}^n$ has already been defined, then define

$$x_{k+1} := x_k + \alpha_k\cdot g(x_k) \quad ,$$

where $g(x_k)$ is a steepest descent direction of f at x_k and the real number $\alpha_k \geq 0$ is choosen in such a way, that

$$\min_{\alpha\geq 0} f(x_k+\alpha g(x_k)) = f(x_k+\alpha_k g(x_k)) .$$

Obviously, the sequence $(x_k)_{k\in\mathbb{N}}$ induces a monotonously decreasing sequence $(f(x_k))_{k\in\mathbb{N}}$ of values of the function f.

A modification of the steepest descent algorithm is proposed in [3]. Therefore we define:

<u>Definition:</u> Let ε, μ be positive real numbers and $f: \mathbb{R}^n \to \mathbb{R}$ be quasidifferentiable. Let N be a neighbourhood of all points $x_o \in \mathbb{R}^n$, where f is not differentiable. Then for $x_o \in N$ we define:

$$\underline{\partial}_\varepsilon f|_{x_o} := \operatorname{conv}\left(\bigcup_{\substack{s \in \mathbb{R}^n \\ \|s\|_2 \leq \varepsilon}} \underline{\partial} f|_{x_o+s}\right)$$

$$\overline{\partial}_\mu f|_{x_o} := \operatorname{conv}\left(\bigcup_{\substack{s \in \mathbb{R}^n \\ \|s\|_2 \leq \mu}} \overline{\partial} f|_{x_o+s}\right) .$$

If $x_o \notin N$, then $\underline{\partial}_\varepsilon f|_{x_o} := \underline{\partial} f|_{x_o}$ and $\overline{\partial}_\mu f|_{x_o} := \overline{\partial} f|_{x_o}$.

If $\underline{\partial}_\varepsilon f|_{x_o}$ and $\overline{\partial}_\mu f|_{x_o}$ are compact sets, then f is called (ε,μ)- quasidifferentiable in x_o.

With the introduction of these two sets, we now give a modified steepest descent algorithm to find an ε-inf-stationary point x^* of f.

Let us assume that $f: \mathbb{R}^n \to \mathbb{R}$ is quasidifferentiable and moreover that, for given $\varepsilon, \mu > 0$, it is (ε,μ) - quasidifferentiable. Then choose an arbitrary $x_o \in \mathbb{R}^n$. Suppose, that x_k has already be defined.

If $-\overline{\partial} f|_{x_k} \subset \underline{\partial}_\varepsilon f(x_k)$ then x_k is an ε-inf stationary point and the algorithm stops.

Otherwise, if $-\overline{\partial} f|_{x_k} \not\subset \underline{\partial}_\varepsilon f(x_k)$, then compute

$$g(x_k) := -\frac{v_o+w_o}{\|v_o+w_o\|_2}$$

with

$$\|v_o+w_o\|_2 = \max_{w \in \overline{\partial}_\mu f|_{x_k}} \left(\min_{v \in \underline{\partial}_\varepsilon f|_{x_k}} \|v+w\|_2\right)$$

and define $x_{k+1} := x_k + \alpha_k g(x_k)$, where $\alpha_k \geq 0$ is chosen in such a way, that

$$\min_{\alpha \geq 0} f(x_k + \alpha\, g(x_k)) = f(x_k + \alpha_k g(x_k)) .$$

In this paper we want to apply this modification for finding an ε-inf stationary point for a class of quasidifferentiable functions.

2. A motivating example

Let $F,G:\mathbb{R}^n \longrightarrow \mathbb{R}$ be two arbitrary functions with $F,G \in C_1(\mathbb{R})$. Then define the following, quasidifferentiable function $f:\mathbb{R}^n \longrightarrow \mathbb{R}$ by

$$f := \max(|G|, -F-|G|) - ||G| - 2|F|| .$$

This type of function is considered in [1] and does obviously not belong to the class $\mathcal{F}$, defined in the introduction. For illustration in figure 1 the graph of a function f of such a type is given for

$$F:\mathbb{R}^2 \longrightarrow \mathbb{R}, \quad F(x_1,x_2) = x_1^2 - x_2$$

$$G:\mathbb{R}^2 \longrightarrow \mathbb{R}, \quad G(x_1,x_2) = -x_1^2 - x_2^2 + 1.2$$

in the set $\Omega = [-1,1.4] \times [-2,1.25]$.

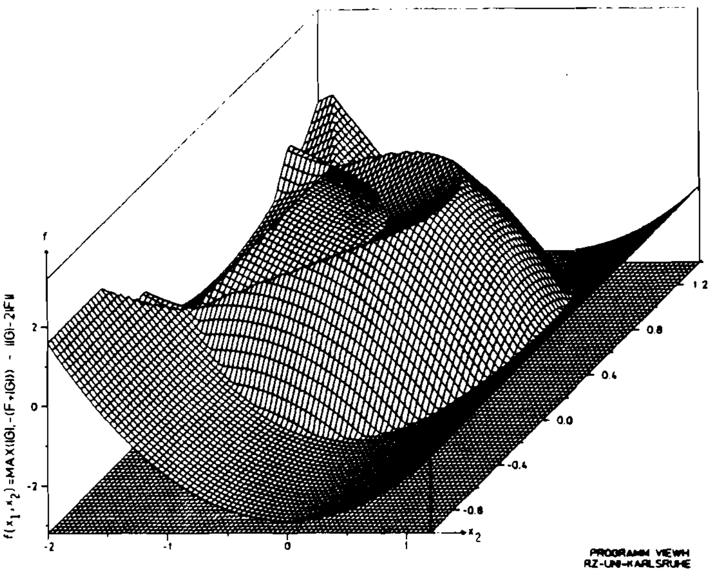

Figure 1

For functions of that type, as well as for the class $\mathcal{F}$, the following properties are valid, as observed in [3].

I. If for all $x \in \mathbb{R}^n$, the convex, compact sets $\underline{\partial} f|_x$ and $\overline{\partial} f|_x$ are computed as in [3] the two mappings

$$x \longmapsto \underline{\partial} f|_x \quad \text{and} \quad x \longmapsto \overline{\partial} f|_x$$

are upper-semi-continuous. Moreover for suitable $\varepsilon,\mu > 0$ the function $\underline{\partial}_\varepsilon f$, $\overline{\partial}_\mu f$ are also upper-semi-continous.

II. If $x \in \mathbb{R}^n$ is not a stationary point, then there exist a real number $M > 0$ and a neighbourhood U_o of $0 \in \mathbb{R}^n$, such that for all $y \in U_o$

$$\left| \frac{df}{dg}\Big|_x - \frac{df}{d(g+y)}\Big|_x \right| \leq M \cdot \|y\|_2 .$$

3. A convergence Theorem

Theorem:

Let $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ be a quasidifferentiable function with the following properties:

(i) There exist real numbers $\varepsilon > 0$, $\mu > 0$ such that for all $x \in \mathbb{R}^n$ f is (ε, μ)-quasidifferentiable and the mappings

$$x \longmapsto \underline{\partial}_{\varepsilon} f\big|_x \quad , \quad x \longmapsto \overline{\partial}_{\mu} f\big|_x$$

and

$$x \longmapsto \underline{\partial} f\big|_x \quad , \quad x \longmapsto \overline{\partial} f\big|_x$$

are upper semi-continuous (u.s.c.)

(ii) If $x \in \mathbb{R}^n$ is not an ε-inf stationary point, then there exist an $M>0$ and a neighbourhood U_o of $0 \in \mathbb{R}^n$ such that for all $y \in U$, $g \in \mathbb{R}^n$

$$\left| \frac{df}{dg}\Big|_x - \frac{df}{d(g+y)}\Big|_x \right| \leq M \cdot \|y\|_2 .$$

Then: Every limit point of the sequence $(x_n)_{n \in \mathbb{N}}$, constructed by the modified steepest descent algorithm, is an ε-inf stationary point of f.

Proof:

Let x^* be a limit point of $(x_n)_{n \in \mathbb{N}}$ and let us assume, that x^* is not ε-inf stationary.

Hence there exist a $v_o \in \underline{\partial}_{\varepsilon} f|_{x^*}$ and a $w_o \in \overline{\partial} f|_{x^*}$ such that

$$\|v_o + w_o\|_2 = \sup_{w \in \overline{\partial} f|_{x^*}} \left(\inf_{v \in \underline{\partial}_{\varepsilon} f|_{x^*}} \|v+w\|_2 \right) = a > 0.$$

Thus $g:= -\frac{v_o+w_o}{\|v_o+w_o\|_2}$ is a normalized descent direction in x^*.

Observe that $w_o \in \bar{\partial}_\mu f|_x$.

Since $x \longmapsto \underline{\partial}_\varepsilon f|_x$ is u.s.c.,there exist a neighbourhood $\tilde{U}$ of $\underline{\partial}_\varepsilon f|_{x^*}$ and a neighbourhood U of x^* such that for all $x \in U$

$$\underline{\partial}_\varepsilon f|_x \subset \tilde{U} .$$

Moreover,to $\bar{\partial}_\mu f|_{x^*}$ there exist a neighbourhood $\tilde{V}$ of $\bar{\partial}_\mu f|_{x^*}$ and a neighbourhood V of x^* such that for all $x \in V$

$$\bar{\partial}_\mu f|_x \subset \tilde{V} .$$

Choose U_o according to assumption (ii) of the theorem. To $W:=U \cap V \cap (U_o+x^*)$ there exists a $k_o \in \mathbb{N}$ such that for all $k \geq k_o$, $x_k \in W$. (Here k isthe index of the convergent subsequence .)
Let us denote by $w_k^* \in \bar{\partial}_\mu f|_{x_k}$ the point, which is nearest to w_o.

From the upper semicontinuity of $\bar{\partial}_\mu f$ we have

$$\lim w_k^* = w_o \quad .$$

Now, let $v_k \in \underline{\partial}_\varepsilon f|_{x_k}$ be a point of minimal distance to $-w_k^*$.

Then $\lim_k (\mathrm{dist}(v_k, \underline{\partial}_\varepsilon f|_{x^*})=0$.

This follows from the fact, that for every k (k large enough)

$$\mathrm{dist}(v_k, \underline{\partial}_\varepsilon f|_{x^*}) \leq \delta_{\mathrm{Hausd.}}(\underline{\partial}_\varepsilon f|_{x_k}, \underline{\partial}_\varepsilon f|_{x^*}) \quad ,$$

which tends to 0 by the choice of $\tilde{U}$.

The neighbourhoods of $\underline{\partial}_\varepsilon f|_{x^*}$ can be assumed to be bounded, since $\underline{\partial}_\varepsilon f|_{x^*}$ is compact.

Hence, there exists a subsequence $(v_k)_{k \in \mathbb{N}}$, also indexed by k, which converges to $\tilde{v} \in \underline{\partial}_\varepsilon f|_{x^*}$.

Thus, for a suitable subsequence and an index K we have:

$$\lim_{k\to\infty} \|w_k+v_k\|_2 = \|w_o+\tilde{v}\|_2 \geq \mathrm{dist}(w_o, \underline{\partial}_\varepsilon f|_{x^*}) = a \quad .$$

We see that $\tilde{v}=v_o$ since the Euclidian norm is strict. Therefore, for all $k \geq K$

$$\|w_k + v_k\|_2 \geq \frac{a}{2} \quad .$$

Now, we want to show, that for k large enough

$$\hat{g}_k = - \frac{v_k+w_k}{\|v_k+w_k\|_2}$$

is a descent direction in x^*.
For this, let $\alpha > 0$. Then:

$$f(x_k+\alpha\hat{g}_k)=f(x^*) + \left.\frac{df}{d[(x_k-x^*)+\alpha\hat{g}_k]}\right|_{x^*} +o(\|x_k-x^*+\alpha\hat{g}_k\|) \quad .$$

From assumption (ii) follows

$$\left.\frac{df}{d[(x_k-x^*)+\alpha\hat{g}_k]}\right|_{x^*} = \alpha \left.\frac{df}{d\hat{g}_k}\right|_{x^*} + O\ (\|x_k-x^*\|_2)$$

and therefore

$$f(x_k+\alpha\hat{g}_k)=f(x^*)+\alpha\left.\frac{df}{d\hat{g}_k}\right|_{x^*} +o(\|x_k-x^*+\alpha\hat{g}_k\|_2)+O(\|x_k-x^*\|_2)$$

$$=f(x^*)+\alpha\left.\frac{df}{d\hat{g}_k}\right|_{x^*} +O(\|x_k-x^*\|_2)+o(\alpha) \quad .$$

By definition of quasidifferentiability we have:

$$\left.\frac{\partial f}{\partial\hat{g}_k}\right|_{x_k} = \min_{w\in\overline{\partial}_\mu f|_{x_k}} (\max_{v\in\underline{\partial}_\varepsilon f|_{x_k}} \langle w+v,\hat{g}_k\rangle)$$

and therefore,by definition of v_k:

$$\left.\frac{\partial f}{\partial\hat{g}_k}\right|_{x_k} \leq \max_{v\in\underline{\partial}_\varepsilon f|_{x_k}} \langle w_k+v,\hat{g}_k\rangle$$

$$\leq \max_{v\in\underline{\partial}_\varepsilon f|_{x_k}} \left(- \langle w_k+v,w_k+v_k\rangle\cdot\|w_k+v_k\|_2^{-1}\right)$$

$$= -\|w_k+v_k\|_2^2 \cdot \|w_k+v_k\|_2^{-1} = -\|v_k+w_k\|_2 \le -\frac{a}{2}\ .$$

Since
$$\left.\frac{df}{d\hat{g}_k}\right|_{x^*} \le \max_{\substack{v\in\tilde{U} \\ \underline{\partial}_\varepsilon f|_{x^*}\subset\tilde{U}}} \langle v,\hat{g}_k\rangle + \min_{w\in\overline{\partial} f|_{x^*}} \langle w,\hat{g}_k\rangle\ ,$$

$$w_o\in\overline{\partial} f|_{x^*} \text{ and } \lim_{k\to\infty} w_k=w_o$$

we find for a given $\delta > o$ an index K_1 such that for all $k \ge K_1$

$$\left.\frac{df}{d\hat{g}_k}\right|_{x^*} \le \max_{\substack{v\in\tilde{U} \\ \underline{\partial}_\varepsilon f|_{x^*}\subset\tilde{U}}} \langle v,\hat{g}_k\rangle + \min_{w\in\overline{\partial} f|_{x^*}} \langle w,\hat{g}_k\rangle$$

$$\le (\max_{v\in\underline{\partial}_\varepsilon f|_{x_k}} \langle v,\hat{g}_k\rangle+\delta) + \langle w_o,\hat{g}_k\rangle$$

$$\le (\max_{v\in\underline{\partial}_\varepsilon f|_{x_k}} \langle v,\hat{g}_k\rangle+\delta) + \langle w_k,\hat{g}_k\rangle + \|w_k-w_o\|_2$$

$$= \left.\frac{df}{d\hat{g}_k}\right|_{x_k} + 2\,\delta \le -\frac{a}{2} + 2\,\delta\ .$$

Thus, for all $k \ge K_1$, we see that $\hat{g}_k$ is a descent direction in x^*.
Hence, there is $\tau_o > 0$ such that

$$f(x_k+\tau_o\hat{g}_k) < f(x^*)\ .$$

Now, by the definition of the sequence $(x_k)_{k\in\mathbb{N}}$ via the modified steepest descent algorithm we have:

$$f(x_{k+1}) = f(x_k+\alpha_k g(x_k)) = \min_{\alpha>o} f(x_k+\alpha g(x_k))$$

$$\le \min_{\alpha>o} f(x_k+\alpha\hat{g}_k) = f(x_k+\hat{\alpha}_k\hat{g}_k)$$

$$\le f(x_k+\tau_o\hat{g}_k) < f(x^*)\ .$$

This contradicts the facts that $(f(x_k))_{k\in\mathbb{N}}$ is monotonously decreasing and $\lim f(x_k)=f(x^*)$.

QED.

Remark: The proof also remains valid for $\varepsilon=o$, i.e. replacing "ε-inf-stationary" by "inf-stationary".

4. Numerical experiences

The above mentioned modification of the steepest descent method was implemented on the Siemens 7780 at the Computer Center of the University of Karlsruhe.
Applying this procedure to the motivating example of section 2, ε-inf stationary points ,also for problems under constraints ,(cf. [2]) could easily be found.
Let us now discuss a further

example

let $\quad f:\mathbb{R}^3 \longrightarrow \mathbb{R}$ be given by

$$f_1(x_1,x_2,x_3) = ((x_1+x_2) + \sqrt{(x_1-x_2)^2+4x_3^2}\,)\,/\,2$$

and

$$f_2(x_1,x_2,x_3) = ((x_1+x_2) - \sqrt{(x_1-x_2)^2+4x_3^2}\,)\,/\,2$$

with:

$$f(x_1,x_2,x_3) = |f_1(x_1,x_2,x_3)| - |f_2(x_1,x_2,x_3)|$$

Obviously $f_1, f_2 \notin C_1(\mathbb{R}^3)$.

This function naturally occurs in the investigation of condition of matrices, namely if we assign to any symmetric $(n\times n)$-matrix $A=(a_{ij})_{1\leq i,j\leq n}$ the difference of moduli of the maximal and minimal eigenvalue $|\lambda_{max}|$ and $|\lambda_{min}|$ respectively, i.e.

$$\varphi : L(\mathbb{R}^n,\mathbb{R}^n) \longrightarrow \mathbb{R}$$

$$\varphi_n(A) := |\lambda_{max}| - |\lambda_{min}| \quad .$$

This function is quasidifferentiable, since $\lambda_{max} = \sup\limits_{\|x\|=1} \langle Ax,x\rangle$ is a convex function and $\lambda_{min} = \inf\limits_{\|x\|=1} \langle Ax,x\rangle$ is a concave function.

For $n = 2$, ϕ_n coincides with the above defined function $f: \mathbb{R}^3 \longrightarrow \mathbb{R}$. Morover, the properties i) and ii) of the theorem are valid for the sets $\underline{\partial}_\varepsilon f$ and $\overline{\partial}_\mu f$ for suitable ε and μ. Figure 2 below gives an illustration of the graph of the function f for 4 different values of x_3, i.e. $x_3 = 0.3$; $x_3 = 0.2$; $x_3 = 0.1$; $x_3 = 0.0$.

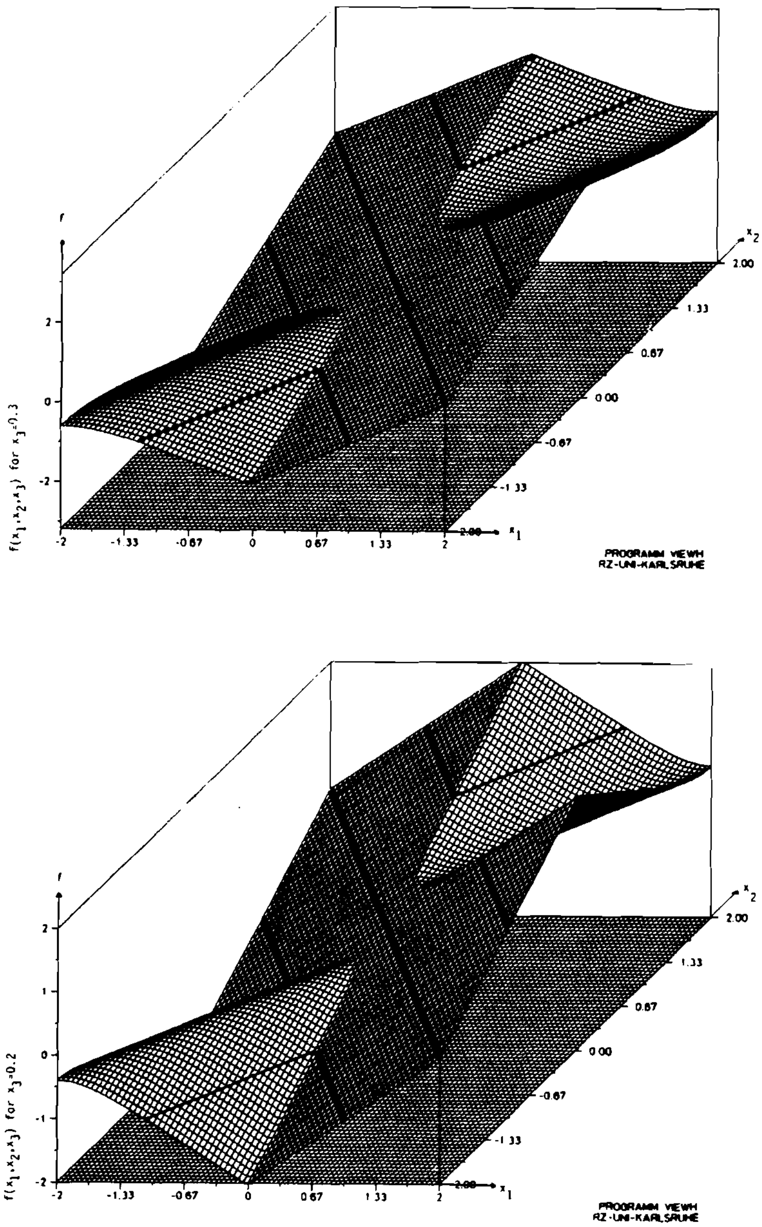

Figure 2

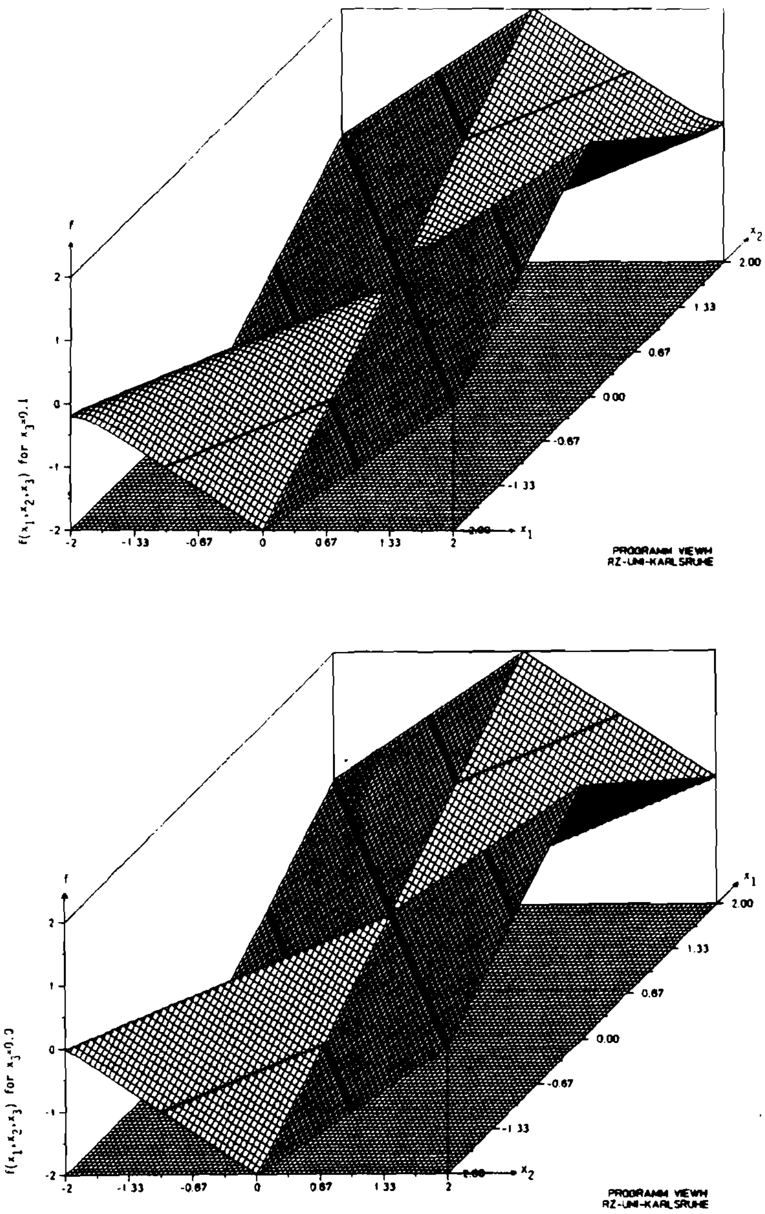

Figure 2

The behaviour of this function $x_3 = 0$ is similar to that, given in example 2.1 of [4].

In Clarke's sense, the point (0,0,0) is stationary, but is neither minimum or maximum, nor a saddle-point, it is a monkey-saddle point. Moreover, $0 \in \text{int } (\partial_{cl} f|_0)$, i.e., 0 is an inner point of the Clarke subdifferential. Of course, using quasidifferentials, the algorithm could find a descent direction (0,0,0).

Impressively, the "cummulative character" of Clarke's subdifferential can be observed in the pictures of Figure 2.

References:

[1] V.F.Demyanov, A.M.Rubinov
On Quasidifferentiable Mappings.
Math.Operationsforschung und Statistik, Ser. Optimization 14 (1983) pp. 3-21

[2] V.F.Demyanov
Quasidifferentiable functions: Necessary Conditions and Descent Directions.
IIASA-Working Papers, WP-83-64 (June 1983)

[3] V.F.Demyanov, S.Gamidov and T.I.Sivelina
An Algorithm for Minimizing a Certain Class of Quasidifferentiable Functions.
IIASA-Working Papers, WP-83-122 (Dec. 1983)

[4] V.F.Demyanov, L.N.Polykova, A.M.Rubinov
Nonsmoothness and Quasidifferentiability.
IIASA-Working Papers, WP-84-22 (March 1984)