WORKING PAPER

THE INVERSE OF A LIPSCHITZ FUNCTION IN R^n: COMPLETE CHARACTERIZATION BY DIRECTIONAL DERIVATIVES

Bernd Kummer

November 1989
WP-89-084

IIASA
International Institute for Applied Systems Analysis

THE INVERSE OF A LIPSCHITZ FUNCTION IN R^n: COMPLETE CHARACTERIZATION BY DIRECTIONAL DERIVATIVES

Bernd Kummer

November 1989
WP-89-084

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

Foreword

The new concept of directional derivative introduced for Lipschitzian vector valued functions helps to formulate a necessary and sufficient condition for the existence of locally Lipschitz inverse and to characterize its directional derivatives. For $C^{1,1}$ optimization, this allows to establish a necessary and sufficient condition for a critical point to be stable. Fundamentals of the calculus are developed, too.

The results were completed within the frame of the IIASA Contracted Study "The Development of Parametric Optimization and its Applications."

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

Lipschitzian inverse functions, directional derivatives and application in $C^{1,1}$ - optimization

Bernd Kummer [1)]

Key words

Inverse Lipschitz function, implicit function, directional derivative, mean-value theorem, chain-rules, strongly stable critical points

Abstract

The paper shows that L. Thibault's [22] limit sets allow an iff-characterization of local Lipschitzian invertibility in finite dimension. We consider these sets as directional derivatives and extend the calculus in a way that it can be used to clarify whether critical points are strongly stable in $C^{1,1}$- optimization problems.

1.1 Introduction

During the last fifteen years, various concepts of generalized derivatives have been developed to derive optimality conditions for nonsmooth problems or to describe implicit functions. As a selection of the rich literature to this field we refer to the basic work [2] and to [1,3,...,7,12,13,15,16,19,...,23]. The present paper aims at the existence of a locally Lipschitz inverse f^{-1} of a function f from R^n into itself. It turns out that, for this purpose, the limit sets $D_f(x;u)$ of L. Thibault [21,22] play an important role. In his papers, they are used in order to extend Clarke's calculus to functions taking values in topological vector spaces. There, our basic properties of § 2.2 (except connectness) are shown to hold more general. Concerning these historical facts the author is in debt to Prof. L. Thibault for sending the correspondent informations and papers.

In what follows, reserving the index for partial derivatives, we denote $D_f(x;u)$ by $\Delta f(x;u)$ and call these sets directional derivatives. They are related to F.H. Clarke's generalized

1) Humboldt-University Berlin, Section Mathematics, PSF 1297, GDR, Berlin 1086

Jacobian $\partial f(x)$ and its extremal points $ex\,\partial f(x)$ by

(1.1) $\quad (ex\,\partial f(x))\, u \subset \quad \Delta f(x;u) \subset \quad \partial f(x)\, u$

whenever $f\colon R^n \to R^m$ is locally Lipschitz (for similar relations in abstract spaces see [21]). Our main result will show that L. Thibault's sets are crucial in view of the inverse function

Theorem 1. Let $f\colon R^n \to R^n$ be locally Lipschitz. Then, f is Lipschitzian invertible at x (Def. 1) if and only if $0 \notin \Delta f(x;u)$ for all $u \in R^n \setminus \{0\}$. In this case, the relations $u \in \Delta f^{-1}(f(x);v)$ and $v \in \Delta f(x;u)$ are equivalent.

Example 2 will demonstrate that, even if $\partial f(x)$ is singular, some Lipschitz function can have a Lipschitzian inverse. This way, the consideration of $\Delta f(x;u)$ is motivated by new reasons which form the content of the next section. In section 3 we show the relation between "strong" and "weak" stability by an implicit function theorem and verify that known (from Clarke's calculus) mean-value theorems including Taylor expansion for $C^{1,1}$-functions [7] remain valid (and can be directly shown) with the present derivatives. In order to apply the calculus and to preserve the iff-condition of Theorem 1 some chain-rules of equation-type (section 4) and simple functions will be of interest. Finally, we are able (in section 5) to derive a complete characterization of so-called strongly stable critical points of optimization problems involving $C^{1,1}$-functions. It should be noted that this task (which is "almost" completely solved for the C^2-case by [11, 17, 8]) has essentially stimulated the following investigations and seems to be unsolvable without using the sets $\Delta f(x;u)$ (cp. Theorem 4).

1.2. Notations, basic definitions

Given a bounded subset X of the Euclidean space R^n we denote the linear space of all Lipschitz functions f from X into R^m by $C^{0,1}(X,R^m)$. The number Lip (f/X) is the smallest Lipschitz module of f on X, and by the equation

$$|f|^{0,1}_X = \max\left\{ \sup_{x\in X} \|f(x)\| \;,\; \mathrm{Lip}\,(f/X) \right\}$$

the so-called Lip-norm is defined. It can be regarded as a seminorm for the space $C^{0,1}(R^n,R^m)$ of all locally Lipschitz functions from R^n into R^m. Similarily, the space $C^{1,1}(R^n,R^m)$ consists of all continuously differentiable functions from R^n into R^m having locally Lipschitz Jacobians, and

$$|f|^{1,1}_X = \max \{ \sup_{x \in X} \|f(x)\| \, , \, |Df|^{0,1}_X \}.$$

Let $B(x, \varepsilon)$ denote the closed ball with center x and radius ε in the underlying space. For a function f from R^n into R^m we call

$$\operatorname{Lip} f(x) = \inf_{\varepsilon > 0} \operatorname{Lip}(f/ B(x,\varepsilon)) \in R_+ \cup \{\infty\}$$

the Lipschitz module of f at x.

Def. 1. A continuous function $f: R^n \to R^n$ is said to be Lipschitzian invertible at x if there are positive ε and δ such that

(i) the equation $f(y) = z$, $y \in B(x,\varepsilon)$ has a unique solution $y = f^{-1}(z)$ whenever $z \in B(f(x), \delta)$, and

(ii) the function f^{-1} is Lipschitz on $B(f(x),\delta)$.

Def. 2. A continuous function $f: R^n \to R^n$ is said to be weakly stable at x (with respect to some subset G of $C^{0,1}(R^n,R^n)$) if there are positive ε and δ such that the equation $f(y) + g(y) = f(x)$, $y \in B(x,\varepsilon)$ has a unique solution $y = y(g)$ whenever $g \in G$ and $|g|^{0,1}_{B(x,\varepsilon)} < \delta$.

If, additionally, the mapping $g \mapsto y(g)$ is Lipschitz on its domain with the norm $|\cdot|^{0,1}_{B(x,\varepsilon)}$, then f is called strongly stable at x (with respect to G).

Now, let $f: R^n \to R^m$ be a continuous function and suppose $x, u \in R^n$ to be fixed.

Def. 3. The set $\Delta f(x)$ consists of all points $z \in R^m$ being a limit of points

$$z^k = (f(y^k) - f(x^k)) \cdot \|y^k - x^k\|^{-1} \qquad k= 1,2,\ldots$$

where $x^k \to x$, $y^k \to x$ and $x^k \neq y^k$.

Def. 4. The set $\Delta f(x;u)$ consists of all points $z \in R^m$ being a limit of points

$$z^k = \left(f(x^k + \lambda_k u) - f(x^k) \right) \cdot \lambda_k^{-1} \qquad k = 1, 2, \ldots$$
where $x^k \to x$ and $\lambda_k \downarrow 0$.

We call $\Delta f(x;u)$ <u>directional derivative</u> of f at x.

2. Motivation and basic properties of the derivatives

2.1. Motivation

The main motivation of considering the set $\Delta f(x)$ is given by

<u>Lemma 1.</u> A continuous function $f: R^n \to R^n$ is Lipschitzian invertible at x if and only if $0 \notin \Delta f(x)$.

<u>Proof:</u> Indeed, if $0 \in \Delta f(x)$ then f^{-1} cannot be Lipschitz near $(x, f(x))$. This directly follows from Def. 3. Conversely, if $0 \notin \Delta f(x)$, then there is some positive ε such that

$\| f(x'') - f(x') \| \geq \varepsilon \| x'' - x' \|$ for all $x'', x' \in B(x, \varepsilon)$.

By the invariance of domain theorem, the set $f(B(x,\varepsilon))$ contains some ball $B(f(x), \delta)$ $(\delta > 0)$. [1)]

Thus, the requirements of Def. 1. are satisfied. □

Because of the chaotic structure of the sequences included in the definition of $\Delta f(x)$, it is hard to apply Lemma 1 to concrete functions. Therefore, a representation of $\Delta f(x)$ by means of the "better" sets $\Delta f(x;u)$ is desirable.

<u>Lemma 2.</u> If $f \in C^{0,1}(R^n, R^m)$ then $\Delta f(x) = \bigcup_{\|u\| = 1} \Delta f(x;u)$.

<u>Proof:</u> The inclusion " $\supset$ " is trivial. Let $z \in \Delta f(x)$, and consider the sequences x^k, y^k, z^k as in Def. 3. Setting

$\lambda_k = \| y^k - x^k \|, \quad u^k = (y^k - x^k) / \lambda_k$

some (infinite) subsequence of u^k converges to u, $\|u\| = 1$. We may assume that the original sequence already shows this property. Defining $v^k = (f(x^k + \lambda_k u) - f(x^k)) / \lambda_k$

we can estimate, for large k:

$$\| z^k - v^k \| \leq \| f(y^k) - f(x^k + \lambda_k u) \| / \lambda_k \leq$$
$$\leq (1 + \mathrm{Lip}\, f(x)) \| \lambda_k (u^k - u) \| / \lambda_k$$

Hence, we obtain $z = \lim z^k = \lim v^k \in \Delta f(x;u)$. □

---

1) This fact was already mentioned in [2] , Remark 2.

2.2. Basic properties of the derivatives

Most of the following statements are immediate consequences of the definitions and the Lipschitz property and need only elementary proofs which will be omitted here.
Further, we suppose throughout this section that the functions under consideration map R^n into R^m and are locally Lipschitz near the points of interest.

(P1) $\Delta f(x) \subset B(0, \mathrm{Lip}\, f(x))$

$\emptyset \neq \Delta f(x;u), \quad \Delta f(x;tu) = t\Delta f(x;u) \quad , \quad t \in R$

$\Delta f(x; u+v) \subset \Delta f(x;u) + \Delta f(x;v)$

$\Delta(f+g)(x;u) \subset \Delta f(x;u) + \Delta g(x;u)$.

(P2) The multifunctions $\Delta f(.)$ and $\Delta f(.;.)$ are closed and locally bounded.

(P3) If f is a functional (m=1) then $\Delta f(x;u)$ is the interval $[-(-f)^o(x;u)\ ,\ f^o(x;u)]$ where f^o denotes F.H. Clarke's directional derivative of f.

(P4) $\Delta f(x;u) \subset \partial f(x)\, u := \{ Au : A \in \partial f(x) \}$ where $\partial f(x)$ is F.H. Clarke's [2] generalized Jacobian.

Indeed, we may apply the mean-value theorem [3], proposition 2.6.5., to the points $x^k + \lambda_k u$ and x^k in Def. 4. This yields the existence of some matrices

$A^k \in \mathrm{conv}\left(\bigcup_{0 \le \tau \le 1} \partial f(\tau x^k + (1-\tau)(x^k + \lambda_k u)) \right)$

such that

$f(x^k + \lambda_k u) - f(x^k) = \lambda_k A^k u.$

Since the multifunction $\partial f(.)$ is closed, we obtain $Au = z$ for each accumulation "point" A of the bounded sequence A^k. □

Note that property (P4) implicitly makes use of Rademacher's theorem. However, (P4) has illustrative character and will not be used in what follows. The same is true for

(P5) If A is extremal in $\partial f(x)$ then $Au \in \Delta f(x;u)$.
This statement follows from the fact that there is a sequence $x^k \to x$ such that $Df(x^k)$ exist and converge to A.

(P6) The sets $\Delta f(x;u)$ are connected.

To verify (P6) we introduce the sets

$$\Delta^{\varepsilon} f(x;u) = \{(f(y+\lambda u) - f(y))/\lambda : 0 < \lambda \le \varepsilon , \ y \in B(x, \varepsilon) \}$$

and note that

$$\Delta f(x;u) = \limsup_{\varepsilon \downarrow 0} \ \Delta^{\varepsilon} f(x;u) .$$

Now, assume the contrary; there are open sets $\Omega_i \subset R^m$ such that $\Delta f(x;u) \cap \Omega_i \neq \emptyset$ $(i = 1, 2)$ and $\Delta f(x;u) \subset \Omega_1 \cup \Omega_2$, $\Omega_1 \cap \Omega_2 = \emptyset$.

Since $\Delta f(x;u)$ is compact there is some $\varepsilon > 0$ such that $\Delta^{\varepsilon} f(x;u) \subset \Omega_1 \cup \Omega_2$. Because of $\Delta^{\varepsilon} f(x;u) \cap \Omega_i \neq \emptyset$ the set $\Delta^{\varepsilon} f(x;u)$ is not connected, again.

To construct a contradiction, consider any two elements a' and a'' in $\Delta^{\varepsilon} f(x;u)$ generated by the pairs (λ', y') and (λ'', y''), respectively. Setting, for $0 \le t \le 1$,

$\lambda(t) = t\lambda' + (1-t)\lambda''$, $y(t) = ty' + (1-t)y''$,

$a(t) = [f(y(t) + \lambda(t)u) - f(y(t))] \cdot \lambda(t)^{-1}$

an arc connecting a' and a'' is formed. Since, obviously, $a(t) \in \Delta^{\varepsilon} f(x;u)$ $(\forall t)$, this set is connected. □

(P7) $v \in \Delta f(x;u)$ if there are sequences $x^k \to x$, $u^k \to u$ and $\lambda_k \downarrow 0$ such that $v = \lim (f(x^k + \lambda_k u^k) - f(x^k))/\lambda_k$.

(P8) Let $f(.) = g(h(.))$ where g and h map R^m into R^p and R^n into R^m, respectively. Then

(i) $\Delta f(x;u) \subset \Delta g(h(x) ; \Delta h(x;u))$

(ii) If g is a C^1-function, then (i) holds as equation:
$\Delta f(x;u) = Dg(h(x)) \Delta h(x;u)$.

To prove the first statement let $a = \lim a^k$ where

(2.1) $a^k = [g(h(x^k + \lambda_k u)) - g(h(x^k))] / \lambda_k$, $x^k \to x$, $\lambda_k \downarrow 0$.

Since the sequence

(2.2) $v^k := (h(x^k + \lambda_k u) - h(x^k)) / \lambda_k$

is bounded, we may assume that $v^k \to v \in \Delta h(x;u)$.

Substituting $h(x^k + \lambda_k u) = h(x^k) + \lambda_k v^k$ in (2.1) we observe

(2.3) $a^k = [g(h(x^k) + \lambda_k v^k) - g(h(x^k))] / \lambda_k$

and, in view of (P7), $a \in \Delta g(h(x) ; v)$.

Consider the second statement, and let $v \in \Delta h(x;u)$ be given. Now, there exist sequences $x^k \to x$, $\lambda_k \downarrow 0$ such that v^k (2.2) converge to v. We define a^k via (2.1) and write these points in the form (2.3). Since $g \in C^1$, we may then estimate

$$\| a^k - Dg(h(x))v \| < \varepsilon_k , \quad \text{where} \quad \varepsilon_k \to 0 .$$

Together with (2.1) this leads to $Dg(h(x))v \in \Delta f(x;u)$. □

The inclusion (i) of (P8) will not necessarily hold as equation if the inner function h belongs to C^1. In order to see this, we suggest to study the following example.

<u>Example 1.</u> $f(x) = g(h(x))$, $x \in R^1$, $g: R^2 \to R^1$ where $h(x) = (x , 0)$ and

$$g(y_1 , y_2) = \begin{cases} 0 & \text{if} \quad y_1 \le 0 \\ y_1 & \text{if} \quad 0 \le y_1 \le |y_2| \\ |y_2| & \text{otherwise} \end{cases}$$

Take $x = 0$ and $u = 1$.

2.3. The inverse function theorem

The first statement of Theorem 1 is, obviously, a direct consequence of Lemma 1, Lemma 2 and property (P1).

To verify the second one, let $v \in \Delta f(x;u)$, and let z^k, x^k, λ_k be sequences as in Def. 4 where $z^k \to v$. Then

$$f(x^k + \lambda_k u) = f(x^k) + \lambda_k z^k \quad \text{and}$$

$$x^k + \lambda_k u = f^{-1}(f(x^k) + \lambda_k z^k) ,$$

$$u = [f^{-1} (f(x^k) + \lambda_k z^k) - f^{-1}(f(x^k))] / \lambda_k .$$

In view of (P7) the latter means $u \in \Delta f^{-1}(f(x) ; v)$.

Conversely, let $u \in \Delta f^{-1}(f(x) ; v)$, and consider sequences $x^k \to f(x)$, $u^k \to u$ and $\lambda_k \downarrow 0$ such that

$$u^k = [f^{-1}(x^k + \lambda_k v) - f^{-1}(x^k)] / \lambda_k ,$$

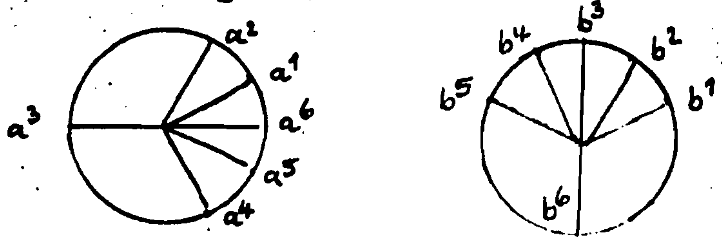
$$f^{-1}(x^k + \lambda_k v) = f^{-1}(x^k) + \lambda_k u^k.$$

By studying the f-image of both sides, the relation $v \in \Delta f(x;u)$ follows as above. □

Now, we present an example that will clarify the relation between F.H. Clarke's [2] inverse function theorem and Theorem 1. It shows, additionally, that the connected sets $\Delta f(x;u)$ may be non-convex and that the inclusions (1.1) can be proper ones.

<u>Example 2.</u> We define a piecewise linear homeomorphism f of R^2 into itself satisfying $0 \in \partial f(0)$.

Let a^i and b^i ($i = 1,\dots,6$) be vectors on the sphere which are arranged as follows.



Additionally, put $a^7 = a^1$; $b^7 = b^1$. The important properties of the figure are

(i) $a^1 = b^1$, $a^2 = b^2$; $a^4 = -b^4$, $a^5 = -b^5$

(ii) the common turning sense of the vectors a^i and b^i

(iii) $\sphericalangle(a^i , a^{i+1}) < \pi$, $\sphericalangle(b^i , b^{i+1}) < \pi$.

For $i = 1,\dots,6$, we put $A^i = (a^i, a^{i+1})$, $B^i = (b^i, b^{i+1})$ which are regular matrices as well as $F^i = B^i(A^i)^{-1}$.

Finally, we define the cones

$K^i =$ con (a^i, a^{i+1}) , $P^i =$ con (b^i, b^{i+1})

and the function f as

$f(x) = F^i x$ if $x \in K^i$.

Since F^i maps K^i onto P^i, we see without any difficulties that f establishes a homeomorphism of R^2 onto itself. Thus, f^{-1} exists and is Lipschitz.

Because of $F^1 = E \in \partial f(0)$, $F^4 = -E \in \partial f(0)$, we have $0 \in \partial f(0)$.

The non-convexity of $\Delta f(0;u)$ follows via

$u \in \Delta f(0;u)$, $-u \in \Delta f(0;u)$ and $0 \notin \Delta f(0;u)$.

3. Implicit functions and mean-values

3.1. Implicit functions

The aim of the next theorem is to show that, under weak assumptions concerning the set G of variations, the notions "Lipschitzian invertible", "weakly stable" and "strongly stable" are equivalent.

Theorem 2. Let $f: R^n \to R^n$ be continuous, and let G be some subset of $C^{0,1}(R^n,R^n)$. Then, at any fixed $x \in R^n$:

(i) f is strongly stable (with resp. to G) whenever it is Lipschitzian invertible;

(ii) f is Lipschitzian invertible whenever it is weakly stable (with resp. to G) and, additionally, the set G includes at least all affine functions g of the type $g(y) = a + Ay$ where $a \in R^n$ and rank $A \leq 1$.

We note that the first assertion is already shown for zeros of multifunctions in Banach-spaces, see[18] , Lemma 3.1. A classical proof can simply apply Banach's fixed point theorem to the mapping $F(y) = f^{-1}(f(x) - g(y))$.
We verify assertion (ii).

Since f is weakly stable and G includes all constant functions, the inverse f^{-1} exists near (x, f(x)). If it is not Lipschitz there, one finds sequences $x^k \to x$, $y^k \to x$, $x^k \neq y^k$ such that the points
$z^k = (f(y^k) - f(x^k))/\| y^k - x^k \|$
converge to zero. Now, define function g^k by setting
$g^k(y) = f(x) - f(x^k) - \langle y-x^k, y^k-x^k \rangle \cdot z^k \| y^k - x^k \|^{-1}$
Each g^k maps R^n onto a line, hence $g^k \in G$. The convergence $z^k \to 0$ implies
$|g^k|^{0,1}_{B(x,\varepsilon)} \to 0$ for each fixed $\varepsilon > 0$.
Moreover, the definition of g^k ensures that x^k and y^k are two different solutions of the equation (in y)
$f(y) + g^k(y) = f(x)$,
both converging to x. This contradiction proves the theorem. □

For the proof of the second part of the theorem, we have not used the Lemma 1. Therefore, the statement (ii) can be generalized by considering more general spaces.

3.2. Mean-value theorems

The next Theorem 3 (i) coincides with Theorem 2.3.7 in[3] in the case of finite dimension. However, we present the proof in order to illustrate the simple way how the directional derivatives can be used and to show that Rademacher's theorem is not needed in this context.

Theorem 3. Let $f \in C^{0,1}(R^n, R^m)$, and let $x, u \in R^n$.
(i) If m=1, then there is some $\theta \in (0 , 1)$ such that
$f(x+u) - f(x) \in \Delta f(x+\theta u ; u)$.
(ii) If $m \geq 1$ then $f(x+u) - f(x) \in \text{conv} (\bigcup_{0 \leq \theta \leq 1} \Delta f(x+\theta u ; u))$.

(iii) (Taylor expansion) If $m=n$ and $f = Dh$ where $h \in C^{1,1}(R^n, R^1)$, then there is some $\theta \in (0, 1)$ such that $h(x+u) - h(x) \in \langle f(x), u \rangle + \frac{1}{2} \langle u, \Delta f(x+\theta u ; u) \rangle$.

[7] uses generalized Hessians to make obvious a similar statement as (iii).

Proof: Because of (P8) we may restrict ourselves to the case $f(x+u) = f(x)$ for proving (i) and (ii).

(i) Consider the function

(3.1) $g(t) = f(x + tu), \quad 0 \leq t \leq 1$.

Since, obviously, $\Lambda g(\theta ; 1) \subset \Lambda f(x+\theta u ; u)$, it suffices to show that $0 \in \Lambda g(\theta ; 1)$ for some θ in $(0, 1)$. Omitting the trivial case g = constant, one may assume $\min_{[0,1]} g(.) < f(x)$ (otherwise consider the maximum).

Let $\theta \in \arg\min_{[0,1]} g(.)$. Then, $\theta \in (0, 1)$.

With $t < \theta$, $t \to \theta$ and $\lambda = \theta - t$ we obtain $(g(t+\lambda) - g(t))/\lambda \leq 0$ and $\Lambda g(\theta ; 1) \cap R_- \neq \emptyset$.

With $t > \theta$, $t \to \theta$ and $\lambda = t - \theta$ we observe $(g(\theta+\lambda) - g(\theta))/\lambda \geq 0$ and $\Lambda g(\theta ; 1) \cap R_+ \neq \emptyset$.

Therefore, the connected set $\Lambda g(\theta ; 1)$ (in R) contains the origin.

(ii) Let $C = \operatorname{conv} (\bigcup_{0 \leq \theta \leq 1} \Delta f(x+\theta u ; u))$.

Assume $0 \notin C$.

Since C is non-empty, convex and compact there is some $b \in R^m$ that separates C from the origin:

(3.2) $0 < \langle b, c \rangle \quad \forall c \in C$.

Let $G: R^n \to R$ be defined as $G(z) = \langle b, f(z) \rangle$, and put $F(\theta) = \langle b, \Delta f(x+\theta u ; u) \rangle$.

Using (P8) with $g = \langle b, . \rangle$, $h = f$, we obtain $F(\theta) = \Delta G(x+\theta u ; u)$.

Because of (i) there is some $\theta \in (0, 1)$ satisfying $\langle b, 0 \rangle = G(x+u) - G(x) \in F(\theta) = \langle b, \Delta f(x+\theta u;u) \rangle$.

This contradicts (3.2) and indicates $0 \in C$.

(iii) We define a function g by

(3.3) $g(t) = h(x+tu) - t (h(x+u) - h(x))$.

It satisfies $g(0) = g(1)$, $Dg(t) = \langle f(x+tu), u \rangle - h(x+u) + h(x)$

and, again by (P8)

$\Delta Dg(\tau ; 1) \subset \langle u , \Delta f(x+\tau u ; u)\rangle$.

Hence, it is enough to verify

(3.4) $\quad 0 \in Dg(0) + \frac{1}{2} \Delta Dg(\theta ; 1)$ for some $\theta \in (0 , 1)$.

To do so we introduce a second function r as

$r(t) = g(t) + Dg(0) (t - \frac{1}{2})^2$.

It fulfils the equations

$$r(0) = r(1), \quad Dr(0) = 0$$

(3.5) $\quad Dr(t) = Dg(t) + 2 Dg(0) (t - 1/2)$

$$\Delta Dr(t ; 1) = \Delta Dg(t ; 1) + 2 Dg(0) .$$

The application of statement (i) yields

$$\exists \tau \in (0, 1): \quad r(1) - r(0) = Dr(\tau)$$

$$\exists \theta \in (0, \tau): \quad Dr(\tau) - Dr(0) \in \Delta Dr(\theta ; \tau) .$$

Because of (3.5) we thus obtain

$$0 = r(1) - r(0) \in Dr(0) + \tau \Delta Dr(\theta ;1) = \tau \Delta Dr(\theta ; 1),$$

$$0 \in \tau (\Delta Dg(\theta ; 1) + 2 Dg(0)).$$

Since $\tau > 0$, the formula (3,4) is true. □

The third part of Theorem 3 reveals some information about a critical point x of a $C^{1,1}$- function h $(Dh(x) = 0)$: If $\langle u , \Delta Dh(x;u)\rangle > 0$ for all $u \neq 0$, then there is some $\varepsilon > 0$ such that

$$h(y) - h(x) \geq \varepsilon \| y - x \|^2 \quad \text{for all} \quad y \in B(x, \varepsilon).$$

This condition, however, is not a necessary one even if h is convex (why ?).

4. Chain rules and simple Lipschitz functions

The property (P8) may be seen as a first and useful chain rule for the directional derivatives being under consideration. It is, unfortunately, not sufficient for our aim of considering solutions of perturbed Karush-Kuhn-Tucker systems. We need some formula for the directional derivatives of a function $F(x,z) = f(x, g(z))$ under the following assumptions:

(4.1) $\quad f \in C^{0,1}(R^{n+m}, R^q)$, $g \in C^{0,1}(R^p, R^m)$, and the partial derivatives $D_y f(. , .)$ with respect to the second variable exist and are Lipschitzian.

The desired formula is

(4.2) $\Delta F((\bar{x},\bar{z});(u,w)) = \Delta_x f((\bar{x},g(\bar{z}));u) + D_y f(\bar{x},g(\bar{z}))\Delta g(\bar{z};w)$.

By Δ_x, we denote the partial directional derivative with respect to x. We will see that (4.2) is true whenever g is a <u>simple</u> function.

<u>Def. 5.</u> A function $g \in C^{0,1}(R^p,R^m)$ is said to be <u>simple</u> (at z) if for all sequences $\lambda_k \downarrow 0$ and all given pairs (v,w) satisfying $v \in \Delta g(z;w)$ there is some sequence $z^k \to z$ such that v becomes an accumulation point of

$$v^k = (g(z^k + \lambda_k w) - g(z^k)) / \lambda_k .$$

<u>Step 1</u> We investigate the formula

(4.3) $\Delta f((\bar{x},\bar{y});(u,v)) = \Delta_x f((\bar{x},\bar{y});u) + \Delta_y f((\bar{x},\bar{y});v)$,

which is not true, in general.

The left-hand side consists of all limits of the kind

(4.4) $a^k = [f(x^k+\lambda_k u, y^k+\lambda_k v) - f(x^k,y^k)] / \lambda_k$,
$(x^k,y^k) \to (\bar{x},\bar{y})$, $a^k \to a$, $\lambda_k \downarrow 0$.

The right-hand side is formed by sums $b^k = b_1^k + b_2^k$ with

(4.5) $b^k = (f(x^k+\alpha_k u,\bar{y})-f(x^k,\bar{y}))/\alpha_k +$
$+ (f(\bar{x}, y^k+\beta_k v)-f(\bar{x}, y^k))/\beta_k$,
$(x^k,y^k) \to (\bar{x},\bar{y})$, $b^k \to b$, $\alpha_k, \beta_k \downarrow 0$.

Suppose the sequences in (4.4) to be given. To obtain $a = \lim b^k$ we may put $\alpha_k = \beta_k = \lambda_k$ and estimate the difference

(4.6) $c^k = a^k - b^k$, $(\alpha_k = \beta_k = \lambda_k)$

If $c^k \to 0$, then (4.3) holds as inclusion "$\subset$".

Conversely, suppose the sequences in (4.5) to be given. Then, we can try to replace the second term b_2^k by

$p_2^k = (f(\bar{x}, \eta^k+\alpha_k v) - f(\bar{x}, \eta^k))/\alpha_k$ where $\eta^k \to \bar{y}$,

and

(4.7) $\| p_2^k - b_2^k \| \to 0$ (at least for some subsequence).

If this is possible, then we can put $\lambda_k = \alpha_k$, $y^k = \eta^k$ and, with the resulting a^k (4.4), again estimate the difference (4.6). If $c^k \to 0$, the inclusion "$\supset$" is now true in (4.3).

Summarizing, two questions remain crucial:

(i) When does c^k (4.6) tend to zero ?

(ii) When can b_2^k be replaced by p_2^k with given α_k such that (4.7) is true ?

Step 2. Recall the suppositions (4.1) and consider formula (4.3) in this particular case. Since $f(\bar{x}, .)$ is continuously differentiable (and simple, Def.5), question (ii) finds a positive answer.

To answer (i) we set

$x' = x^k + \lambda_k u, \quad y' = y^k + \lambda_k v, \quad x = x^k, \quad y = y^k, \quad \lambda = \lambda_k$.

The difference $c = c^k$ (4.6) then becomes

$$c = [(f(x', y') - f(x', \bar{y})) - (f(x, y') - f(x, \bar{y}))]/\lambda$$
$$+ [(f(x, y') - f(x, y)) - (f(\bar{x}, y') - f(\bar{x}, y))]/\lambda$$
$$= c_1/\lambda + c_2/\lambda .$$

Setting $\Theta = \bar{y} + t(y' - \bar{y})$ we may write

$$c_1 = \int_0^1 (D_y f(x', \Theta) - D_y f(x, \Theta)) (y' - \bar{y}) \, dt$$

and estimate

$$\| c_1 \| \leq L \| x' - x \| \| y' - \bar{y} \| = L \| \lambda_k u \| \| y' - \bar{y} \| .$$

A similar estimation for c_2 shows that $c = c^k$ tends to zero.

Thus, (4.3) holds for f (4.1) with

$$\Delta_y f((\bar{x}, \bar{y}); v) = D_y f(\bar{x}, \bar{y}) v .$$

The latter implies that (4.2) is true as inclusion "$\subset$".

Step 3 Finally, suppose g in (4.2) to be simple, and let $b = b_1 + D_y f(\bar{x}, g(\bar{z})) v$ ($v \in \Delta g(\bar{z} ; w)$) be some element of the right-hand side in (4.2).

We write b_1 as a limit of b_1^k where

$$b_1^k = [f(x^k + \lambda_k u, g(\bar{z})) - f(x^k, g(\bar{z}))]/\lambda_k ; \quad x^k \to x, \lambda_k \downarrow 0.$$

Since g is simple, there are sequences y^k and z^k such that $y^k = g(z^k)$, $g(z^k + \lambda_k w) = y^k + \lambda_k v^k$, where v^k is some sequence having the accumulation point v, and $z^k \longrightarrow \bar{z}$.

Further, put $b_2^k = (f(\bar{x}, y^k + \lambda_k v) - f(\bar{x}, y^k))/\lambda_k$ and

$$a^k = (f(x^k + \lambda_k u, y^k + \lambda_k v) - f(x^k, y^k))/\lambda_k .$$

Then, there is some accumulation point a of a^k in the set $\Delta F((\bar{x}, \bar{z}); (u, w))$, and the sequence $b^k = b_1^k + b_2^k$ converges to $b_1 + D_y f(\bar{x}, g(\bar{z})) v$. Recalling the integral estimation used, in step 2 for $c = c^k$, we obtain $b = a$, and (4.2) is verified. □

It remains to clarify which functions are simple.

The family of simple functions forms a proper subclass of $C^{0,1}(R^n,R^m)$ even if $n = 1$ and $m = 2$. The problem to construct some corresponding non-simple example is left to the reader. We will concentrate on the next two "positive" statements.

<u>Lemma 3.</u> Every functional $f \in C^{0,1}(R^n,R)$ is simple (at each point).

<u>Proof:</u> Let $v \in \Delta f(x;u)$ and $\lambda_k \downarrow 0$ be given. We write $v = \lim v^k$ where, with certain related sequences

(4.8) $v^k = (f(x^k+\alpha_k u) - f(x^k))/\alpha_k$, $x^k \to x$, $\alpha_k \downarrow 0$.

Fix any k and consider the line-segment $I_k = [x^k, x^k+\alpha_k u]$, where $\| u \| = 1$ may be assumed. The lemma is true if the following can be shown:

For sufficiently large $t > t(k)$, there exist $y^t \in I_k$ such that

$w^t := (f(y^t+ \lambda_t u) - f(y^t))/\lambda_t$

satisfy $| w^t - v^k| \leq 1/k$.

Let L be some Lipschitz module of f near x, and let t(k) be chosen such that

(4.9) $\lambda_t < \alpha_k (2kL + 1)^{-1}$ for all $t > t(k)$.

Assume that, for some $t > t(k)$, such point y^t does not exist. Considering w^t as a continuous function of y^t either $w^t > v^k + 1/k$ or $w^t < v^k - 1/k$ must be true for all points in I_k. We may suppose the first case. By using this inequality successively for

$y^t = x^k + s \lambda_t u$, $s= 0, 1, \ldots, [\alpha_k / \lambda_t] =: N$

and applying (4.8) we obtain ,after some elementary calculations, that $z := x^k + N \lambda_t u$ fulfils

$f(z) - f(x^k + \alpha_k u) > \lambda_t(N/k - L)$

where $|v^k| \leq L$ is assumed. Because of (4.9) we have

$N > (\alpha_k/\lambda_t) - 1 > 2kL$

which, together with $\|z - (x^k + \alpha_k u)\| < \lambda_t$, leads to a contradiction regarding the Lipschitz assumption. □

Note that the lemma, particularily, says that F.H. Clarke's directional derivative may be written as

$$f^o(x;u) = \limsup_{x^k \to x} k(f(x^k+ u/k) - f(x^k)) .$$

We finish this section with

Lemma 4. The function $g: R \to R^2$, defined by
$g(t) = (t^\ominus, t^\oplus)$, $t^\ominus = \min\{0,t\}$ $t^\oplus = \max\{0,t\}$,
is simple. The set $\Delta g(t;v)$ consists of all
$a = (a^-, a^+) \in R^2$ satisfying

(4.10) $a^- t^\oplus = a^+ t^\ominus = 0, \quad a^- + a^+ = v, \quad a^- \cdot a^+ \geq 0.$

Proof: The set A defined by (4.10) contains $\Delta g(t;v)$ because of (P4) and $A = \partial g(t)v$. We note that this argumentation is not consequently concerning the remark after the proof of (P4), but it gives more information than the (easily possible) elementary proof.
Now, let $a \in A$, and let $\lambda_k \downarrow 0$ be given. The construction of $x^k \to t$, such that

$$a^k := (g(x^k+\lambda_k v) - g(x^k)) / \lambda_k$$

converge to a, is then very easy.
If $t \neq 0$ then $x^k = t$. Let $t = 0$. Then, we put

$$x^k = \begin{cases} 0 & \text{if } v = 0 \\ -\lambda_k a^- & \text{if } v > 0 \\ -\lambda_k a^+ & \text{if } v < 0 \end{cases} . \quad \square$$

5. Strongly stable critical points in $C^{1,1}$- optimization

Consider an optimization problem

$$P(a,b,c): \quad \inf \; f(x) + \langle a, x \rangle \quad \text{s. t.}$$
$$g_i(x) \leq b_i \quad i = 1,\dots,m$$
$$h_j(x) = c_j \quad j = 1,\dots,p$$

where a, b, c are regarded as (small) parameters and f, g_i, h_j are supposed to belong to $C^{1,1}(R^n,R)$.
Following M. Kojima [11] we assign to P the function

$$(5.1) \quad F(x,y,z) = \begin{pmatrix} Df(x) + \sum y_i^\oplus Dg_i(x) + \sum z_j \, Dh_j(x) \\ y_i^\ominus - g_i(x) \\ - h_j(x) \end{pmatrix}$$

which lies in $C^{0,1}(R^{n+m+p}, R^{n+m+p})$.
The critical points of $P(a,b,c)$ are described by equation

(5.2) $F(x,y,z) = -(a,b,c)$.

Their relation to Karush-Kuhn-Tucker points (KKTP) is given by the (Lipschitzian) correspondence

(x,y,z) critical $\Longrightarrow$ $(x, y^{\oplus}, z)$ KKTP
(x,y,z) KKTP $\Longrightarrow$ $(x,y+g(x),z)$ critical.

Let $\bar{s} = (\bar{x},\bar{y},\bar{z})$ be some fixed critical point of $P(0,0,0)$. We ask after the existence of some neighbourhoods $N(\bar{s})$ and $U(0)$ such that $N(\bar{s})$ contains exactly one critical point s of $P(a,b,c)$ whenever $(a,b,c) \in U(0)$ where, additionally, the resulting function is Lipschitz. This property (often called "strongly stable") means obviously that F (5.1) is Lipschitzian invertible at $\bar{s}$. Consequences of this fact as well as sufficient conditions for C^2-problems are to be found in several papers; the ones closest to the present are (perhaps) [8], [11] and [17] whereas similar properties for local minimizers are the subject of [9] and [10].

Recalling Theorem 1 we have to clarify whether condition

(5.3) $\quad 0 \notin \Delta F(\bar{s}; r)$ for all $r = (u,v,w) \neq 0$

holds.

In order to use the chain rule (4.2) it is convenient to write F in the form

$$F(x,y,z) = \begin{pmatrix} Df & Dg_1 & \dots & Dg_m & 0 \dots 0 & Dh_1 & \dots & Dh_p \\ -g & 0 & \dots & 0 & E & 0 & \dots & 0 \\ -h & 0 & \dots & 0 & 0 \dots 0 & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} 1 \\ y^{\oplus} \\ y^{\ominus} \\ z \end{pmatrix}.$$

Here, all vectors included are considered as columns, and the matrix depends on x only. We denote it by $M(x)$, and write $b(y,z)$ for the vector on the right.

Applying Lemma 4 to the m functions $y_i \mapsto (y_i^{\oplus}, y_i^{\ominus})$ of independent variables we see that b is simple and that

$\Delta b((\bar{y},\bar{z});(v,w))$ consists of all vectors $(0,a^+,a^-,w)$ satisfying

(5.4) $\quad a^- + a^+ = v, \quad a_i^- = 0 \ (i \in I_+), \quad a_i^+ = 0 \ (i \in I_-)$ and

$a_i^- \cdot a_i^+ \geq 0$,

where

$I_- = \{i: \ \bar{y}_i < 0\}, \ I_+ = \{i: \ \bar{y}_i > 0\}, \ I_0 = \{i: \ \bar{y}_i = 0\}.$

By the help of formula (4.2) we obtain

$$\Delta F(\bar{s};r) = \Delta_x(Mb)(\bar{x}, b(\bar{y},\bar{z}); u) + D_b(Mb)(\bar{x},b(\bar{y},\bar{z}))\ \Delta b((\bar{y},\bar{z});(v,w)) .$$

The first term of the right-hand side consists of all vectors of the kind

$$\begin{pmatrix} \xi \\ -Dg(\bar{x})^T u \\ -Dh(\bar{x})^T u \end{pmatrix} \text{ where } \xi \in \Delta_x(Df + (\bar{y}^{\oplus})^T Dg + \bar{z}^T Dh)(\bar{x}; u) =: H(u) .$$

Note that, for C^2- functions, the set $H(u)$ is simply Hu if H denotes the Hessian of the Lagrangian with respect to x at $\bar{s}$.

The second term of the right-hand side has the form $M(\bar{x})\,\Delta b$ and consists of vectors of the kind

$$\begin{pmatrix} \sum_i a_i^+ Dg_i(\bar{x}) + \sum_j w_j Dh_j(\bar{x}) \\ a^- \\ 0 \end{pmatrix}$$

where a^-, a^+ are restricted to (5.4).

Summarizing we obtain: Condition (5.3) holds true iff the system

$$\begin{aligned} -(\textstyle\sum a_i^+ Dg_i(\bar{x}) + \sum w_j Dh_j(\bar{x})) &\in H(u) \\ Dg(\bar{x})^T u &= a^- \\ Dh(\bar{x})^T u &= 0, \end{aligned} \tag{5.5}$$

with a^+, a^- according to (5.4), has the trivial solution $(u,v,w) = 0$ only.

Let us introduce some second, more convenient system with variables u, α, β by the conditions

$$\begin{aligned} \sum_{i \in I_0 \cup I_+} \alpha_i\, Dg_i(\bar{x}) + \sum \beta_j\, Dh_j(\bar{x}) &\in H(u) \\ \alpha_i\, Dg_i(\bar{x})^T u &\leq 0 \qquad \forall\, i \in I_0 \\ Dg_i(\bar{x})^T u &= 0 \qquad \forall\, i \in I_+ \\ Dh(\bar{x})^T u &= 0 . \end{aligned} \tag{5.6}$$

<u>Theorem 4.</u> Condition (5.3) of "strong stability" holds true if and only if the system (5.6) has only the trivial solution $(u, \alpha, \beta) = 0$

<u>Proof:</u> Indeed, if (u,v,w) solves (5.5) with certain a^+, a^- from (5.4) then we find that (u, α, β) with $\alpha = -a^+$, $\beta = -w$ solves (5.6) (We remove the components α_i, $i \in I_-$). Conversely, if (u, α, β) solves (5.6) we construct a solution of (5.5) as follows:

$i \in I_-$: $v_i = Dg_i(\bar{x})^T u$, $a_i^+ = 0$, $a_i^- = Dg_i(\bar{x})^T u$

$i \in I_0$: $v_i = Dg_i(\bar{x})^T u - \alpha_i$, $a_i^+ = -\alpha_i$, $a_i^- = Dg_i(\bar{x})^T u$

$i \in I_+$: $v_i = -\alpha_i$, $a_i^+ = -\alpha_i$, $a_i^- = 0$.

Setting $w = -\beta$ the vector (u,v,w) solves (5.5). Since, finally, nontrivial solutions are preserved under the given transformations (which is not difficult to see), nothing remains to prove. □

In order to interprete this result we define the tangent space associated with $(\bar{x},\bar{y})$ as

$$T(\bar{x},\bar{y}) = \{u : Dh(\bar{x})^T u = 0, \; Dg_i(\bar{x})^T u = 0 \;\; \forall i \in I_+\}.$$

To each $u \in T(\bar{x},\bar{y})$ there corresponds some (normal-) cone

$$K(u) = \{z : z = \sum_{i \in I_0 \cup I_+} \alpha_i Dg_i(\bar{x}) + \sum_j \beta_j Dh_j(\bar{x}) \text{ where } \alpha_i Dg_i(x)^T u \leq 0 \;\; \forall i \in I_0\}.$$

The theorem then says that condition (5.3) is equivalent to the two requirements

(i) The gradients $Dg_i(\bar{x})$ $(i \in I_0 \cup I_+)$ and $Dh_j(\bar{x})$ are linearly independent (LICQ)

(ii) $K(u) \cap H(u) = \emptyset$ for each $u \in T(\bar{x},\bar{y})$, $u \neq 0$.

It seems natural to understand (ii) as a second-order condition.

If the involved functions are C^2 then system (5.6) describes a linear complementarity problem. The question whether such problems have non-trivial solutions was completely solved by S.M. Robinson [17].

Since $\langle z,u \rangle \leq 0$ for $z \in K(u)$, condition (ii) holds true whenever $u \in T(x,y) \setminus \{0\}$ and $v \in H(u)$ imply $\langle v,u \rangle > 0$.

Concluding remarks

In comparison with the C^1-case, we note two important unpleasant properties of $C^{0,1}$-functions:

1. In spite of (P2) the multifunction $y \mapsto \Delta_x f((x,y);u)$ may fail to be closed; see Example 1.
2. Suppose $f: R \to R$ to have a Lipschitzian inverse and try to determine its only zero by any iterative procedure which coincides with Newton's method at all points where f is differentiable. Then, the procedure may fail to converge (locally) as it generates an alternating sequence with almost all initial points; see [14], §2.3.

Acknowledgement: Many fruitful discussions with my colleagues D. Klatte and K. Tammer as well as with H.Th. Jongen and

F. Nožička have influenced the present investigations in a very constructive manner.

References

[1] Aubin, J.P.; Ekeland, I.: Applied Nonlinear Analysis. New York: Wiley, 1984

[2] Clarke, F.H.: On the inverse function theorem. Pacific J. of Mathematics 64 No.1 (1976), 97-102

[3] Clarke, F.H.: Optimization and Nonsmooth Analysis. New York: Wiley, 1983

[4] Demyanov, V.F.; Rubinov, A.M.: Quasidifferential Calculus, Optimization Software. New York, Berlin: Springer, 1986

[5] Hiriart-Urruty, J.B.: Tangent cones, generalized gradients and mathematical programming in Banach spaces. Math. Oper. Res. 4 (1979), 79-97

[6] Hiriart-Urruty, J.B.: Characterizations of the plenary hull of the generalized Jacobian matrix. Math. Programming Study 17 (1982), 1-12

[7] Hiriart-Urruty, J.B.; Strodiot, J.J.; Hien Nguyen, V.: Generalized Hessian matrix and second-order optimality conditions for problems with $C^{1,1}$-data. Appl. Math. Optimization 11 (1984), 43-56

[8] Jongen, H.Th.; Klatte, D.; Tammer, K.: Implicit functions and sensitivity of stationary points. Aachen, Techn.Hochschule, Preprint No.1 (1988)

[9] Klatte, D.: On strongly stable local minimizers in nonlinear programs. in: Advances in Mathematical Optimization/ ed. Guddat, J. et al.- Berlin: Akademie Verlag, 1988, 104-113

[10] Klatte, D.; Tammer, K.: On second-order sufficient optimality conditions for $C^{1,1}$-optimization problems. Optimization 19 (1988) 2 , 169-180

[11] Kojima, M.: Strongly stable stationary solutions in nonlinear programs. in: Analysis and Computation of Fixed Points/ ed. Robinson, S.M.- New York: Academic Press, 1980

[12] Kruger, A.Y.: Properties of generalized differentials. Sib. Math. J. 24 (1983), 822-832

[13] Kruger, A.Y.; Mordukhovich, B.S.: Extreme points and the Euler equations in nondifferentiable optimization problems. Dokl. Akad. Nauk BSSR 24 (1980), 684-687

[14] Kummer, B.: Newton's method for non-differentiable functions. in: Advances in Mathematical Optimization/ ed. Guddat, J. et al.- Berlin: Akademie Verlag, 1988, 114-125

[15] Mordukhovich, B.S.: Metody approksimacij v zadacach optimizacij i upravlenija. Moskva: Nauka , 1988

[16] Mordukhovich, B.S.: Metric approximations and necessary optimality conditions for general classes of nonsmooth extremal problems. Sov. Math. Dokl. 22 (1980), 526-530

[17] Robinson, S.M.: Strongly regular generalized equations. Math. Oper. Res. 5 (1980), 43-62

[18] Robinson, S.M.: An implicit-function theorem for B-differentiable functions. IIASA working paper WP- 88-67 (1988)

[19] Rockafellar, R.T.: Directionally Lipschitzian functions and subdifferential calculus. Proc. Lond. Math. Soc. 39 (1979), 331-355

[20] Rockafellar, R.T.: Extensions of subgradient calculus with applications to optimization. Nonlinear Anal. Theory Methods Appl. 9 (1985), 665-698

[21] Thibault, L.: On generalized differentials and subdifferentials of Lipschitz vector-valued functions. Nonlinear Anal. Theory Methods Appl. 6 (1982) 1037-1053

[22] Thibault, L.: Subdifferentials of compactly Lipschitzian vector-valued functions. Annali di Matematica pura ed applicata (IV) (1980), Vol. CXXV , 157-192

[23] Ward, D.E.; Borwein, J.M.: Nonsmooth calculus in finite dimension. SIAM J. Control and Optimization 25, No. 5 (1987), 1312-1340