An extension of the inverse function theorem

By V. Novo and L. Rodriguez Marin*

Recibido: 14 de Noviembre de 1990

Presentado por el Académico D. Pedro Jiménez Guerra.

Abstract

The purpose of this work is to give conditions under wich a G–differentiable function admits locally a G–differentiable inverse. The classical result that gives conditions under which a C^r function admits locally a C^r inverse is a special case of this theorem.

El propósito de este trabajo es dar condiciones suficientes para que una función G–diferenciable admita (localmente) una inversa G–diferenciable. El resultado clásico que da condiciones suficientes para que una función de clase C^r admita (localmente) inversa de clase C^r es un caso especial de este teorema.

Clasificación A.M.S. :.26A24

1. Introduction and notations

There have been a number of approaches recently toward developing a set–valued derivative of convex or Lipschitz functions which generalize the usual notion of derivative in such a way that the theorems of differential calculus also extend. The present interest in problems related to the optimization of non–smooth functions brought about the development of this news generalized differentiation theories.

With this aim T. Rockafellar [11] studied real convex functions on $\mathbb{R}^n$ introducing the subdifferential. In the same way F.H. Clarke [3] broaden the kind of functions considered by Rockafellar extending the theory to real locally lipschitz functions on $\mathbb{R}^n$ by defining the generalized gradient. Clarke latter included in his theory both functions from $\mathbb{R}^n$ into $\mathbb{R}^m$ [4] and real functionals on Banach spaces [5]. Futher development of this theories can be find in a number of references such as [1], [2], [5], [6] or [10].

In different previous works ([8], [9]) we extended the class of functions used by Clarke defining a new generalized derivative called G–derivative. This work aims to give conditions under which a G–differentiable function admits locally a G–differentiable inverse in such a way that the classical inverse function theorem is a special case.

Some previous results of [8] and [9] are given. f is a real–valued function on some interval $I \subset \mathbb{R}$; $a,\ x,\ x_n \in I$ and $(x_n) \longrightarrow a$. $F(a,x)$ and $l(f,a,x_n)$ will mean

$$F(a,x) = [f(x) - f(a)]/(x-a);\ l(f,a,x_n) = \lim_{n\to\infty} F(a,x_n).$$

*E.T.S. Ingenieros Industriales. U.N.E.D.

$(x_n) \longrightarrow a$ is said to be a G–derivability sequence of f at a if there exist $l(f,a,x_n)$. $S(f,a)$ will denote the G–derivability sequences set of f at a. The G–derivative of f at a is the set

$$\partial f(a) = \mathrm{co}\{l(f,a,x_n);\ (x_n) \in S(f,a)\}.$$

If $f \in C_S(I)$ (def. 2.1) $\partial f(a)$ is a non–empty convex compact set. Derivation and chain rules extend at this context, condition necessary of local extremum is now $0 \in \partial f(a)$ and we have next generalized mean value theorem. If $f \in C_S[a,b]$, then there exists $c \in (a,b)$ and $A \in \partial f(c)$ such that $f(b) - f(a) = A(b-a)$.

In the following U is an open set of $\mathbb{R}^n$, $\|x\|$ denotes the usual Euclidean norm and $\|\xi\|$ the supremum norm in $L(\mathbb{R}^n,\mathbb{R})$ wich is the usual topological dual of $\mathbb{R}^n$. We topologize the vector space $M_{m\times n}$ of $m \times n$ matrices with the norm $\|M\| = \max|m_{ij}|$ where $M = (m_{ij})$ and $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$. $<.,.>$ is the duality pairing between $L(\mathbb{R}^n,\mathbb{R})$ and $\mathbb{R}^n$, $\mathrm{co}(A)$ is the subset A convex hull and $F(a,v,t_n)$ and $l(f,a,v,t_n)$ will mean

$$F(a,v,t_n) = \frac{1}{t_n}\Big[f(a+t_n v) - f(a)\Big];\ l(f,a,v,t_n) = \lim_{n\to\infty} F(a,v,t_n)$$

where $f : U \longrightarrow \mathbb{R}$; $a \in U$, $v \in \mathbb{R}^n$, $v \neq 0$ and $\{t_n\} \longrightarrow 0$ when $n \longrightarrow \infty$ is a real number sequence.

2. G–differential. Basic properties

Let E and F be normed linear spaces, U be and open set in E and $f : U \longrightarrow F$ be a given mapping.

2.1 Definition

We will call f "strong–continuous" $(s-c)$ at $a \in U$ if there exists a neighborhood V of a and a constant $k > 0$ such that

$$\|f(x) - f(a)\| \leqslant k\|x-a\|, \quad \text{for every} \quad x \in V.$$

f is $s-c$ in U if it is $s-c$ at each point of U, we will denote $f \in C_S(U;F)$.

Next relation is inmediate: $LL(U,F) \subset C_S(U,F) \subset C(U,F)$, where LL denotes locally lipschitz functions and C denotes continuous functions. It is a easy consequence of 2.1 that $f+g$ and λf are $s-c$ functions at a if $f,g \in C_S(a)$ and $\lambda \in \mathbb{R}$. $C_S(U,F)$ is a real linear space and the composition of $s-c$ functions is a $s-c$ function.

Let $f : U \longrightarrow \mathbb{R}$ be a continuous function. The function $t \longrightarrow a+tv$ from $\mathbb{R}$ into $\mathbb{R}^n$ is continuous for each $v \in \mathbb{R}^n$ fixed and consequently $D = \{t \in \mathbb{R};\ a+tv \in U\}$ is an open set such that $0 \in D$. Next proposition is inmediate from 2.1.

2.2 Proposition

If f is a $s-c$ function at a, then for each $v \in \mathbb{R}^n$, $v \neq 0$, $g_v(t) = f(a+tv)$ is a $s-c$ function at $t = 0$.

2.3 Definition

The directional G–derivative of f at a with respect to a vector $v \in \mathbb{R}^n$, $v \neq 0$ or G_v–derivative of f at a denoted by $\partial_v f(a)$ is defined to be the set

$$\partial_v f(a) = \mathrm{co}\left\{\lim_{n\to\infty} F(a, v, t_n);\ \{t_n\} \in S(g_v, 0)\right\};\ \partial_0(a) = \{0\}$$

where $S(g_v, 0)$ denotes the G–derivability sequences set of g_v at 0. Note that $\partial_v f(a)$ is the G–derivative of g_v at $t = 0$.

2.4 Theorem

If f is a $s-c$ function at $a \in U$ then for each $v \in \mathbb{R}^n$

i) $\partial_v f(a)$ is a non–empty convex compact subset in $\mathbb{R}$.

ii) There exists $k > 0$ such that $\partial_v f(a) \subset [-k\|v\|, k\|v\|]$.

Proof. (i) From the fact that f is $s-c$ at a it follows that g_v is $s-c$ at $t = 0$, and consequently $\partial_v f(a) = \partial g_v(0)$ is a non–empty convex compact subset in $\mathbb{R}$.
(ii) Because f es a $s-c$ function at a, there exists $k > 0$ and V such that

$$\left|f(x) - f(a)\right| \leqslant k\|x-a\|; \quad \text{for every} \quad x \in V.$$

For $v \in \mathbb{R}^n$ and $\{t_n\} \in S(g_v, 0)$ fixed, there exists $n_0 \in \mathbb{N}$ such that $a + t_n v \in V$ for all $n > n_0$, then $\left|f(a+t_n v) - f(a)\right| \leqslant k|t_n|\|v\|$ and we have that $\left|F(a, v, t_n)\right| \leqslant k\|v\|$ and for each $\{t_n\} \in S(g_v, 0)$

$$-k\|v\| \leqslant \lim_{n\to\infty} F(a, v, t_n) \leqslant k\|v\|$$

Finally from the fact that $\partial_v f(a)$ is a convex set we deduce that

$$\partial_v f(a) \subset \Big[-k\|v\|, k\|v\|\Big].$$

2.5 Proposition

If f is $s-c$ at $a \in U$, then the set–valued mapping from $\mathbb{R}^n$ into $\mathbb{R}$ defined by $T(v) = \partial_v f(a)$ is a bounded odd prefan in the Ioffe's terminology [7].
Proof. We will show that T is a prefan. If $v = 0$, then $T(0) = \{0\}$. Let $l(f, a, \lambda v, t_n) \in T(\lambda v)$ where $\lambda \in \mathbb{R}^+$. Because

$$l(f, a, \lambda v, t_n) = \lim_{n\to\infty} F(a, \lambda v, t_n) = \lambda \lim_{n\to\infty} F(a, v, t_n)$$

we have that $T(\lambda v) \subset \lambda T(v)$. Moroever let $l(f, a, v, t_n) \in T(v)$ and $\{t'_n\} = \{t_n/\lambda\}$, we have

$$\lambda \lim_{n\to\infty} F(a, v, t_n) = \lambda \lim_{n\to\infty} F(a, v, \lambda t'_n) = \lambda \lim_{n\to\infty} F(a, v, \lambda t'_n) \in T(\lambda v)$$

hence $\lambda T(v) = T(\lambda v)$ for every $\lambda > 0$.

From 2.4.(i) we have that $T(v)$ is a convex compact subset for each $v \in \mathbb{R}^n$ and because $T(-v) = -T(v)$ we have that T is a odd prefan. Finally T is bounded because $T(v) \neq \emptyset$ for each $v \in \mathbb{R}^n$ and from 2.4.(ii) $T(v) \subset \left[-k\|v\|, k\|v\|\right]$.

T is not a fan. Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be defined by

$$f(x, y) = x|y| \Big/ \sqrt{x^2 + y^2} \quad \text{if} \quad (x, y) \neq (0, 0); f(0, 0) = 0.$$

It can be easily proved that for $a = (0, 0)$ T is not a fan because $T(u + v) \not\subset T(u) + T(v)$.

2.6 Definition

f is G–differentiable at $a \in U$ if for each $v \in \mathbb{R}^n$ and each $l \in \partial_v f(a)$ there exists a linear selection $\xi \in L(\mathbb{R}^n, \mathbb{R})$ of the prefan T such that $\xi(v) = l$. The set of this selections is called the G–differential of f at a and is denoted by $\partial f(a)$.

It follows inmediatly from this definition that $\partial f(a)(v) = \partial_v f(a)$ for each $v \in \mathbb{R}^n$.

2.7 Proposition

If f is G–differentiable at $a \in U$ then $\partial f(a)$ is a compact convex set in $L(\mathbb{R}^n, \mathbb{R})$.

Proof. Let $\xi, \eta \in \partial f(a)$ and $\alpha, \beta \geqslant 0$ such that $\alpha + \beta = 1$. For each $v \in \mathbb{R}^n$ $\xi(v) \in T(v)$ and $\eta(v) \in T(v)$ and because $T(v)$ is a convex set we have $\alpha\xi + \beta\eta \in \partial f(a)$ and $\partial f(a)$ is a convex set in $L(\mathbb{R}^n, \mathbb{R})$.

Let $\xi \in \partial f(a)$, from 2.4.(ii) we deduce that

$$\|\xi\| = \sup_{\|v\|=1} |\xi(v)| \leqslant \sup_{\|v\|=1} |T(v)| \leqslant k$$

and consequently $\partial f(a)$ is a bounded set. Moroever $\partial f(a)$ is closed because $T(v)$ is a closed set in $\mathbb{R}$ for each v and

$$\partial f(a) = \bigcap_{v\in\mathbb{R}^n} < ., v >^{-1} T(v).$$

For real functions on $\mathbb{R}$, strong–continuity is equivalent to G–derivability. In this case strong–continuity is not a sufficient condition to G–differentiability. Next theorem gives a necessary and sufficient condition to G–differentiability. Proof of Lemma 2.8 is a inmediate consequence from 2.5.

2.8 Lemma

If f is $s-c$ at a, the following are equivalent:

i) $<\xi,\ v>\in T(v)$ for every $v\in\mathbb{R}^n$.

ii) $<\xi,\ v>\leqslant \sup\ T(v)$ for every $v\in\mathbb{R}^n$.

2.9 Theorem

Let f be $s-c$ at a, then the following propositions are equivalent:

i) T is a set–valued fan.

ii) f is G–differentiable at a.

Proof. (i)⇒(ii) Let p from $\mathbb{R}^n$ into $\mathbb{R}$ defined by $p(v)=\sup\ T(v)$. We will prove that p is positively homogeneous and subadditive. If $\lambda>0$ then

$$p(\lambda v)=\sup\ T(\lambda v)=\sup\ \lambda T(v)=\lambda \sup\ T(v)=\lambda p(v).$$

From the fact that T is a fan we have $T(u+v)\subset T(u)+T(v)$ for all $u,v\in\mathbb{R}^n$ and

$$p(u+v)=\sup\ T(u+v)\leqslant \sup\ T(u)+\sup\ T(v)=p(u)+p(v).$$

Suppose now that $u\in\mathbb{R}^n$ and $l\in T(u)$, then there exists $\xi\in L(\mathbb{R}^n,\mathbb{R})$ such that $\xi(u)=l$. Because T is a odd fan it is homogeneous and consequently for each $\lambda\in\mathbb{R}$ and $v=\lambda u$ we have $\lambda<\xi,u>\in\lambda T(u)=T(\lambda u)$, $<\xi,v>\in T(v)$ and $<\xi,v>\leqslant p(v)$ for all $v\in S$, where S is the linear subspace $S=\{v\in\mathbb{R}^n;\ v=\lambda u,\ \lambda\in\mathbb{R}\}$. It follows from Hahn–Banach theorem that there exists at least a linear function $\eta\in L(\mathbb{R}^n,\mathbb{R})$ satisfying:

$$\eta(v)=\xi(v)\quad\text{for every}\quad v\in S\quad\text{and}\quad \eta(v)\leqslant p(v)\quad\text{for every}\quad v\in\mathbb{R}^n$$

from 2.8 we deduce that f is G–differentiable at a.
(ii)⇒(i) If f is G–differentiable at a, then for every $v\in E$ we have $\partial f(a)(v)=\partial_v f(a)$. Suppose that $l\in T(u+v)$, then there is $\xi\in\partial f(a)$ such that $<\xi,u+v>=l$, hence

$$l=<\xi,u+v>=<\xi,u>+<\xi,v>\in T(u)+T(v)$$

and T is a set–valued fan.

As a consequence we have that if $f\in LL(U,\mathbb{R})$ then f is G–differentiable at each point of U. Now we will prove validity of the generalized mean value theorem in this case.

2.10 Theorem

If f is G–differentiable at U and $[a, a+th] \subset U$ for all $t \in [0,1]$ and $h \in \mathbb{R}^n$, then there exists $\theta \in (0,1)$ and $\xi \in \partial f(a+\theta h)$ such that

$$f(a+h) - f(a) = \xi(h).$$

Proof. Suppose that we have $\varphi : [0,1] \longrightarrow U$ defined by $\varphi(t) = a + th$. φ is differentiable at $(0,1)$ and $\varphi'(t) = h$. Let $f : [0,1] \longrightarrow \mathbb{R}$ defined by $g(t) = (f \circ \varphi)(t)$. Strong–continuity of f and φ implies strong–continuity of g at [0,1]. By mean value theorem (section 1) there exists $\theta \in (0,1)$ and $c \in \partial g(\theta)$ such that $g(1) - g(0) = c$. We now show that

$$\partial g(\theta) \subset \partial f(a+\theta h) \circ \varphi'(\theta).$$

Let $l \in \partial g(\theta)$ and $\{t_n\} \in S(g,\theta)$ such that

$$l = \lim_{n\to\infty} \frac{1}{t_n}\Big[g(\theta + t_n) - g(\theta)\Big] = \lim_{n\to\infty} \frac{1}{t_n}\Big[f\,[\varphi(\theta + t_n)] - f[\varphi(\theta)]\Big].$$

From the fact that for each $n \in \mathbb{N}$, $\varphi(\theta + t_n) = \varphi(\theta) + t_n\varphi'(\theta)$, we have

$$l = \lim_{n\to\infty} \frac{1}{t_n}\Big[f\,[\varphi(\theta) + t_n\varphi'(\theta)] - f\,[\varphi(\theta)]\Big]$$

it follows that $l \in \partial_{\varphi'(\theta)} f\,[\varphi(\theta)] = \partial f(a+\theta h) \circ \varphi'(\theta)$ and consequently $\partial g(\theta) \subset \partial f(a+h) \circ \varphi'(\theta)$. Because $g(1) = f(a+h)$, $g(0) = f(a)$ and $c \in \partial g(\theta)$ there exists $\xi \in \partial f(a+\theta h)$ such that $\xi(h) = c$ and we deduce that

$$f(a+h) - f(a) = \xi(h).$$

Next theorem gives a necessary condition for local extremum of G–differentiable functions.

2.11 Theorem

If f is a G–differentiable function from U into $\mathbb{R}$ and f attains a local extremum at $a \in U$ then $0 \in \partial f(a)$.

Proof. Since f attains a local extremum at a, then for each $v \in \mathbb{R}^n$, g_v attains a local extremum at 0, hence $0 \in \partial g_v(0) = \partial_v f(a)$. It's clear that $< 0, v >= 0 \in \partial_v f(a)$ for each $v \in \mathbb{R}^n$ and consequently $0 \in \partial f(a)$.

Definition of G–differential extend easily to functions from $\mathbb{R}^n$ into $\mathbb{R}^m$. We will suppose that $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is defined by $f = (f_1, f_2, ..., f_m)$, where $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ for each $i = 1, 2, \ldots, m$. It is easily proved that strong–continuity of f at a is equivalent to strong–continuity of each f_i at $a, i = 1, 2, ..., m$.

2.12 Definition

The G–derivative of f at a with respect to $v \in \mathbb{R}^n$, denoted by $\partial_v f(a)$ is the set $\left(\partial_v f_1(a), \partial_v f_2(a), ..., \partial_v f_m(a)\right)$.

From this definition and theorem 2.4, we have inmediatly that if f is $s-c$ at a then for each $v \in \mathbb{R}^n$, $\partial_v f(a)$ is a non–empty convex compact set and

$$\partial_v f(a) \subset \prod_{i=1}^{m} [-k\|v\|, k\|v\|].$$

2.13 Definition

f is said to be a G–differentiable function at a if for each $v \in \mathbb{R}^n$ and each $l \in \partial_v f(x)$ there exists a linear selection $\xi \in L(\mathbb{R}^n, \mathbb{R}^m)$ of the set–valued function $T : v \longrightarrow \partial_v f(a)$ such that $\xi(v) = l$. The set of this selections is called the G–differential of f at a and is denoted by $\partial f(a)$.

It can be proved that if f is G–differentiable at a, then $\partial f(a)$ is a convex compact set in $L(\mathbb{R}^n, \mathbb{R}^m)$, also that G–differentiability of f at a is equivalent to G–differentiability of each f_i at a and

$$\partial f(a) = \prod_{i=1}^{m} \partial f_i(a).$$

From G–derivability properties it follows that $f+g$ and λf are G–differentiable functions if f and g are and $\lambda \in \mathbb{R}$. Moroever $\partial(\lambda f)(a) = \lambda \partial f(a)$ and $\partial(f+g)(a) \subset \partial f(a) + \partial g(a)$. However there exists G–differentiable functions f and g such that $g \circ f$ is not G–differentiable. Next theorem provide a chain rule in a special case.

2.14 Theorem

Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ be open sets, $f : U \longrightarrow V$ be a G–differentiable function on U, $g : V \longrightarrow \mathbb{R}$ be a C^1 function and $h = g \circ f$. For each $v \in \mathbb{R}^n$ and each $x \in U$ we have

i) There exists $\partial_v h(x) \subset < Dg[f(x)], \partial_v f(x) >$.

ii) h is G–differentiable at x and $\partial h(x) = Dg[f(x)] o \partial f(x)$.

Proof. (i) Let $\{t_n\} \in S(f, x, v)$ and suppose that $\left[f(x), f(x+t_n v)\right] \subset V$ for each $n \in \mathbb{N}$. Because $g \in C^1(V)$, from mean value theorem for each $n \in \mathbb{N}$ we have that there exists $c_n \in \left[f(x), f(x+t_n v)\right]$ such that

$$g\left[f(x+t_n v)\right] - g\left[f(x)\right] =< Dg(c_n), \left[f(x+t_n v) - f(x)\right] > .$$

If $\{t_n\} \longrightarrow 0$, then $\{c_n\} \longrightarrow f(x)$ and because $g \in C^1(V)$ we have $\lim_{n\to\infty} Dg(c_n) = Dg\left[f(x)\right]$ and

$$\lim_{n\to\infty} \frac{1}{t_n}\left[(g \circ f)(x+t_n v) - (g \circ f)(x)\right] =< Dg\left[f(x)\right], \lim_{n\to\infty} \frac{1}{t_n}\left[f(x+t_n v) - f(x)\right] >$$

and consequently $\partial_v h(x) \subset < Dg\left[f(x)\right], \partial_v f(x) >$.
(ii) Let $v \in \mathbb{R}^n$ and $l \in \partial_v h(x)$. From (i) we have that there exists $l' \in \partial_v f(x)$ such that $l =< Dg\left[f(x)\right], l' >$ and because f is G–differentiable at x, there is a linear selection ξ of the set–valued function T such that $\xi(v) = l'$. It follows that $l =< Dg\left[f(x)\right], \xi(v) >$ and $\eta = Dg\left[f(x)\right] \circ \xi \in L(\mathbb{R}^n, \mathbb{R})$ is a linear selection of $v \longrightarrow \partial_v h(x)$, h is a G–differentiable function at x and $\partial h(x) \subset Dg\left[f(x)\right] \circ \partial f(x)$. From (i) the other inclusion is inmediate.

It can be easily proved next extension of this theorem. If $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ are open sets, $f : U \longrightarrow V$ is G–differentiable on U, $g : v \longrightarrow \mathbb{R}^p$ is a C^1 function on V and $h = g \circ f$, then h is G–differentiable in U and $\partial h(x) = Dg\left[f(x)\right] \circ \partial f(x)$.

Definition of G–differential can be extended to a compact subset $K \subset U$. From this extension we have the next generalized mean value theorem.

2.15 Definition

f is said to be G–differentiable on K if it is G–differentiable at each point of K. The G–differential of f at K is the set

$$\partial f(K) = \overline{co} \bigcup_{x \in K} \partial f(x)$$

2.16 Proposition

If f is a continuous mapping from $[a, b]$ into $\mathbb{R}^k$ and G–differentiable on $(c, d) \subset [a, b]$, then

$$f(d) - f(c) \in (d - c)\,\overline{co} \bigcup_{t \in [0,1]} \partial f\left[c + t(d - c)\right].$$

Proof. Suppose that $F \in L(\mathbb{R}^k, \mathbb{R})$. Let $F \circ f$ from $[a, b]$ into $\mathbb{R}$. $F \circ f$ is continuous in $[a, b]$ and from 2.14 $F \circ f$ is G–derivable in (c, d). From generalized mean value theorem, we have that there exists θ_F, $0 < \theta_F < 1$ such that

$$(F \circ f)(d) - (F \circ f)(c) = (F \circ \xi)(d - c)$$

where $\xi \in \partial f(c + \theta_F(d - c))$. Then for all $F \in L(\mathbb{R}^k, \mathbb{R})$ it follows that

$$F\left[f(d) - f(c)\right] \in F\left[\bigcup_{t \in [0,1]} \partial f(c + t(d - c))(d - c)\right]$$

and from Hahn–Banach theorem we deduce that

$$f(d) - f(c) \in (d - c)\,\overline{co} \bigcup_{t \in [0,1]} \partial f(c + t(d - c)).$$

2.17 Proposition

Let U be an open subset in $\mathbb{R}^n$, $a \in U$, $h \in \mathbb{R}^n$ with $[a, a+h] \subset U$. If $f : U \longrightarrow \mathbb{R}^k$ is a G–differentiable function on $\{x \in V;\, x = a+th,\ t \in [0,1]\}$, then

$$f(a+h) - f(a) \in \overline{co} \bigcup_{t\in[0,1]} \partial f(a+th)(h)$$

Proof. It's a easy consequence of 2.17 using $g : [0,1] \longrightarrow \mathbb{R}^k$ defined by $g(t) = f(a+th)$.

3. Inverse function theorem

In this section we will assume that f is a G–differentiable function from the open set $U \subset \mathbb{R}^n$ into $\mathbb{R}^k$ $(n \geqslant k)$. Let us call a subset $A \subset L(\mathbb{R}^n, \mathbb{R}^k)$ surjective if each $\xi \in A$ is surjective. The set valued mapping M from U into $P_C\Big[L(\mathbb{R}^n, \mathbb{R}^k)\Big]$ (space of compacts with the Hausdorff's metric) is said to be semicontinuous at $a \in U$ if for every sequence $\{x_i\} \longrightarrow a$, $x_i \in U$, and all sequence $\{\xi_i\} \longrightarrow \xi$ with $\xi_i \in \partial f(x_i)$ for each i, we have $\xi \in \partial f(a)$. We will call ∂f is bounded if it transforms bounded sets of U into bounded sets of $L(\mathbb{R}^n, \mathbb{R}^k)$.

3.1 Lemma

If the set–valued mapping ∂f is bounded and semicontinuous at $a \in U$, then for each $\varepsilon > 0$, there is $\alpha > 0$ such that

$$\partial f(x) \subset B\Big(\partial f(a), \varepsilon\Big) \quad \text{for every} \quad x \in B(a, \alpha)$$

where $B\Big(\partial f(x), \varepsilon\Big) = \bigcup_{\xi\in\partial f(x)} B(\xi, \varepsilon)$.

Proof. Suppose that there exists $\varepsilon > 0$ such that for all $\delta > 0$ there is $x \in B(a, \delta)$ with $\partial f(x) \not\subset B\Big(\partial f(a), \varepsilon\Big)$. Let $\delta = 1/n$, $n \in \mathbb{N}$ for each n there is $x_n \in B(a, 1/n)$ such that $\partial f(x_n) \not\subset B\Big(\partial f(a), \varepsilon\Big)$. It's clear that $\{x_n\} \longrightarrow a$ when $n \longrightarrow \infty$. Let $\{\xi_n\}$ be a sequence with $\xi_n \in \partial f(x_n)$ and $\xi_n \notin B\Big(\partial f(a), \varepsilon\Big)$ for each $n \notin \mathbb{N}$. Because ∂f is bounded, we have that $\partial f(\{x_n\})$ is a bounded set in $L(\mathbb{R}^n, \mathbb{R}^k)$, then $\overline{\partial f(\{x_n\})}$ is a compact set and consequently there exists a subsequence $\{\xi_k\} \subset \{\xi_n\} \subset \overline{\partial f(\{x_n\})}$ such that $\{\xi_k\} \longrightarrow \xi_0$ when $n \longrightarrow \infty$.

We now show that $\xi_0 \in \partial f(a)$. Because $\xi_k \notin B(\partial f(a), \varepsilon)$ we have $\|\xi_k - \xi\| \geqslant \varepsilon$ for every $\xi \in \partial f(a)$ and because $\{\xi_k\} \longrightarrow \xi_0$ there is $n_0 \in \mathbb{N}$ such that $\|\xi_k - \xi_0\| < \varepsilon/2$ for all $k > n_0$, then

$$\|\xi_0 - \xi\| \geqslant \|\xi_k - \xi\| - \|\xi_k - \xi_0\| > \varepsilon/2,$$

for every $\xi \in \partial f(a)$ and $\xi_0 \notin B(\partial f(a), \varepsilon/2)$ contradicting the semicontinuity of ∂f at a.

3.2 Lemma

If $\partial f(a)$ is surjective, then

i) There is $\delta > 0$ such that $d[\partial f(a)(S), 0] \geqslant 2\delta$.

ii) There is $\varepsilon > 0$ such that $d\Big[B(\partial f(a), \varepsilon)(S), 0\Big] \geqslant \delta$

where $S = \{x \in \mathbb{R}^n;\ \|x\| = 1\}$.
Proof. (i) Because $\partial f(a)$ is surjective, $0 \notin \partial f(a)$ and from the fact that $\partial f(a)(S) \subset \mathbb{R}^k$ is a compact set there is $\delta > 0$ such that $d[\partial f(a)(S), 0] \geqslant 2\delta$.
(ii) Suppose that for each $n \in \mathbb{N}$ there exists $\xi_n \in B(\partial f(a), 1/n)$ such that $d[\xi_n(S), 0] < \delta$. Let $\eta_n \in \partial f(a)$ with $\xi_n \in B(\eta_n, 1/n)$ for each $n \in \mathbb{N}$. Since $\{\eta_n\} \subset \partial f(a)$ there exists a subsequence $\{\eta_k\} \longrightarrow \eta_0 \in \partial f(a)$, then the subsequence $\{\xi_k\} \subset \{\xi_n\}$ also is convergent to η_0. Let $\delta/2 > 0$, there is $k_0 \in \mathbb{N}$ with $d[\xi_0(S), \eta_0(S)\} < \delta/2$. From (i) it follows that $d[\eta_0(S), 0] \geqslant 2\delta$ and we have that

$$2\delta \leqslant d[\eta_0(S), 0] \leqslant d[\eta_0(S), \xi_{k_0}(S)] + d[\xi_{k_0}(S), 0] < 3\delta/2$$

wich is a contradiction.

3.3 Lemma

If $\partial f(a)$ is surjective, ∂f is a bounded set–valued mapping semicontinuous at a, then given any unit vector $v \in \mathbb{R}^n$ there are real numbers $\alpha > 0$ and $\delta > 0$ and a unit vector $u \in \mathbb{R}^k$ such that whenever $x \in B(a, \alpha)$ and $\xi \in \partial f(x)$, $< u, \xi(v) > \geqslant \delta$. In consequence $\partial f(x)$ is surjective for each $x \in B(a, \alpha)$.
Proof. From 3.2 there are $\delta > 0$ and $\varepsilon > 0$ such that $d\Big[B(\partial f(a), \varepsilon)(S), 0\Big] \geqslant \delta$ and from 3.1 there is $\alpha > 0$ such that $\partial f(x) \subset B(\partial f(a), \varepsilon)$ for every $x \in B(a, \alpha)$. Let $v \in S$, because $B(\partial f(a), \varepsilon)$ is a convex set we have that $B(\partial f(a), \varepsilon)(v)$ is a convex set and

$$d\Big[B(\partial f(a), \varepsilon)(v), 0\Big] \geqslant \delta$$

By the usual separation theorem for convex sets (see [11]), there is a unit vector $u \in \mathbb{R}^k$ such that $< u, \xi(v) > \geqslant \delta$ for every $\xi \in B(\partial f(a), \varepsilon)$. Results follows from the fact that $\partial f(x) \subset B(\partial f(a), \varepsilon)$ for all $x \in B(a, \alpha)$.

3.4 Lemma

If $\partial f(a)$ is surjective, ∂f is a bounded and semicontinuous set–valued mapping, then for every $x, y \in \bar{B}(a, \alpha)$ we have

$$\|f(x) - f(y)\| \geqslant \delta \|x - y\|.$$

Proof. If $x = y$ it's evident. Suppose that $x \neq y$. We will show it only for $x, y \in B(a, \alpha)$ (f and $\| \ \|$ are continuous). Let $v = (y-x)/\|y-x\| \in \mathbb{R}^n$ and $\beta = \|y-x\|$, then $y = x + \beta v$. Since $x, y \in B(a, \alpha)$, $[x, y] \subset B(a, \alpha)$. From generalized mean value theorem we have that there exists $\xi \in \partial f([x, y])$ such that $f(x + \beta v) - f(x) = \xi(\beta v)$.

From 3.3 it follows that there exists a unit vector $u \in \mathbb{R}^k$ and a real number $\delta > 0$ such that $< u, \xi(v) > \geqslant \delta$, then

$$< u, f(x + \beta v) - f(x) > = < u, \xi(\beta v) > = \beta < u, \xi(v) > \geqslant \beta\delta$$

we deduce that $\|f(x + \beta v) - f(x)\| \geqslant \beta\delta$ and consequently $\|f(y) - f(x)\| \geqslant \delta\|x - y\|$.

Next theorem is a extension of Banach's interior mapping theorem.

3.5 Theorem

If $\partial f(a)$ is surjective and ∂f is a bounded and semicontinuous set–valued mapping, then $f(a) \in \operatorname{Int} f(U)$.

Proof. We will show that $B(f(a), \alpha\delta/2) \subset f(B(a, \alpha))$, where α, δ are values of lemmas 3.1 and 3.2. Let for each $y \in B(f(a), \alpha\delta/2)$ fixed the function h from U into $\mathbb{R}$ defined by $h(z) = \|y - f(x)\|^2$. h attains it minimum in the compact set $\bar{B}(a, \alpha)$ at some point $x \in \bar{B}(a, \alpha)$. We claim x belongs to $B(a, \alpha)$ and $y = f(x)$. If $x \in Fr[\bar{B}(a, \alpha)]$ then $\|x - a\| = \alpha$ and from lemma 3.4 we have

$$\alpha\delta/2 > \|y - f(a)\| \geqslant \|f(x) - f(a)\| - \|y - f(a)\| \geqslant \delta\|x - a\| - \|y - f(x)\| \geqslant$$

$$\geqslant \delta\alpha - \|y - f(a)\| > \delta\alpha - \delta\alpha/2 = \delta\alpha/2$$

which is a contradiction, then $x \in B(a, \alpha)$.

Because x is a minimum of h we have $0 \in \partial h(x)$ (th. 2.11) and consequently $0 \in \partial(\|y - f(x)\|^2)$ and $0 \in \partial f(x) \left[2(y - f(x))\right]$. From 3.3 $\partial f(x)$ is surjective for each $x \in B(a, \alpha)$ and we deduce that $y = f(x)$, then $B(f(a), \alpha\delta/2) \subset f[B(a, \alpha)]$ and $f(a) \in \operatorname{Int} f(U)$.

3.6 Theorem

Let $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be G–differentiable on U. If ∂f is a bounded semicontinuous set–valued mapping and $\partial f(a)$ is surjective then there exists a neighborhood V, $a \in V \subset U$ and a function g defined from $W = f(V)$ into V such that

i) $g \circ \left(f|_V\right) = 1_V$ and $\left(f|_V\right) \circ g = 1_W$.

ii) g is G–differentiable at $y_0 = f(a)$.

Proof. From 3.2 we deduce that $\partial f(x)$ is surjective for every $x \in B(a,\alpha)$. Suppose that $V = B(a,\alpha) \subset U$; from 3.5 we have that $W = f(V)$ is an open set. Let $x_1, x_2 \in V$ there exists $\xi \in \partial f([x_1, x_2])$ (Prop. 2.17) such that $f(x_2) - f(x_1) = \xi(x_2 - x_1)$.

Since $\xi \in L(\mathbb{R}^n, \mathbb{R}^n)$ is surjective it's a biyection and consequently $\mathrm{Ker} f = \{0\}$ and f is injective. Let g be defined as follows. For each $y \in W$ $g(y)$ is the point $x \in V$ such that $f(x) = y$. It's clear that g verifies (i).

We will show that g is lipschitzian on $f(V)$ and consequently G–differentiable. From 3.4 we have

$$\|f(x_1) - f(x_2)\| \geqslant \delta \|x_1 - x_2\|, \quad \text{for every} \quad x_1, x_2 \in B(a,\alpha)$$

making $y_1 = f(x_1)$, $y_2 = f(x_2)$ and $x_1 = g(y_1)$, $x_2 = g(y_2)$ we deduce that

$$\|g(y_1) - g(y_2)\| \leqslant \frac{1}{\delta}\|y_1 - y_2\|, \quad \text{for every} \quad y_1, y_2 \in W.$$

This theorem reduces to the classical one if f is a C^1 function. A simple example to wich this theorem applies is the following. Classic result is not valid because f is not differentiable at (0,0).

3.7 Example

Let f be the function from $\mathbb{R}^2$ into $\mathbb{R}^2$ defined by $f(x,y) = (|x| + 2y, x + |y|)$.

$$\partial_v f_1(0,0) = \Big[-|v_1| + 2v_2, |v_1| + 2v_2 \Big]; \; \partial_v f_2(0,0) = \Big[v_1 - |v_2|, v_1 + |v_2| \Big]$$

where $v = (v_1, v_2) \in \mathbb{R}^2$.

$$\partial f(0,0) = \left\{ \begin{pmatrix} a & 2 \\ 1 & b \end{pmatrix}; \; -1 \leqslant a \leqslant 1, \; -1 \leqslant b \leqslant 1 \right\}$$

It can be easily proved that $\partial f(0,0)$ is surjective, ∂f is a bounded and semicontinuous set–valued mapping, then from 3.6 there exists local inverse function of f.

Next theorem gives a extension of the implicit function theorem. We will denote points of $\mathbb{R}^n \times \mathbb{R}^k$ by (x,y) where $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^k$. If U is an open set in $\mathbb{R}^n \times \mathbb{R}^k$ and $f : U \longrightarrow \mathbb{R}^k$, $\partial_2 f(x,y)$ will mean the G–differential of $f(x,.) : \mathbb{R}^k \longrightarrow \mathbb{R}^k$ for each $(x,y) \in U$.

3.8 Theorem

Let $f : U \subset \mathbb{R}^n \times \mathbb{R}^k \longrightarrow \mathbb{R}^k$, $(a,b) \in U$. If $f(a,b) = 0$, f is G–differentiable on U, ∂f is bounded and semicontinuous at each point of U and $\partial_2 f(a,b)$ is surjective, then there exists an open set $V \subset \mathbb{R}^n$ with $a \in V$ and a function $g : V \longrightarrow \mathbb{R}^k$ such that

i) g is G–differentiable in V.

ii) $g(a) = b$

iii) $(x, g(x)) \in U$ and $f(x, g(x)) = 0$ for every $x \in V$.

Proof. Let $h : U \longrightarrow \mathbb{R}^n \times \mathbb{R}^k$ be defined by $h(x,y) = (h_1(x,y), h_2(x,y))$, where $h_1(x,y) = x$ and $h_2(x,y) = f(x,y)$. It's clear that

$$\partial h(a,b) = \left(\frac{I \,|\, 0}{\partial f(a,b)} \right)$$

where I denotes the unit matrix of $M_{n\times n}$ and 0 denotes the zero matrix of $M_{n\times k}$. We will show that $\partial h(a,b)$ is surjective. Let $(p,q) \in \mathbb{R}^n \times \mathbb{R}^k$, $(p,q) = (0,0)$ iff $\partial h(a,b)(p,q) = 0$. If $(p,q) = (0,0)$ is inmediate. Suppose that $\partial h(a,b)(p,q) = 0$. From the linear system

$$\left(\frac{I \,|\, 0}{\partial f(a,b)} \right) \begin{pmatrix} p \\ q \end{pmatrix} = 0$$

it follows that $p = 0 \in \mathbb{R}^k$ and results a homogeneous system with k equations and k variables $\partial_2 f(a,b)\, q = 0$, but because $\partial_2 f(a,b)$ is surjective each element of $\partial_2 f(a,b)$ have rank k and consequently $q = 0$ and $\partial h(a,b)$ is surjective.

Now, because h_1 is a C^1 function on U and ∂h_2 is semicontinuous on U we have that ∂h is semicontinuous on U. From theorem 3.6 there exists an open set $U_1 \subset U$ with $(a,b) \in U_1$ and a inverse function

$$\varphi = (\varphi_1, \varphi_2) : h(U_1) \longrightarrow U_1.$$

Let $V \subset \mathbb{R}^n$ be the open set defined by $V = \{x \in \mathbb{R}^n; (x,0) \in h(U_1)\}$.

Since $f(a,b) = 0$ we have $h(a,b) = (a,0)$ and $a \in V$. Let

$$g : V \subset \mathbb{R}^n \longrightarrow \mathbb{R}^k$$

define by $g(x) = \varphi_2(x,0)$. It's clear that g is G–differentiable on V and

$$g(a) = \varphi_2(a,0) = \varphi_2[h(a,b)] = b$$

Moroever $(x, g(x)) = (x, \varphi_2(x,0)) \in U$ for every $x \in V$. Finally if $x \in V$ then

$$\varphi_1(x,0) = \varphi_1\,[h(x',y')] \quad \text{for some} \quad (x',y') \in U_1$$

but by definition of h, $x' = x$, then $\varphi_1(x,0) = \varphi_1\,[x, f(x,y')] = x$ because $\varphi\,[x, f(x,y')] = (x,y')$. For all $x \in V$ we have that

$$(x,0) = h|_{U_1} \circ \varphi(x,0) = h|_{U_1}\Big[\varphi_1(x,0), \varphi_2(x,0)\Big] =$$

$$= h|_{U_1}(x, g(x)) = \Big(x, f(x, g(x))\Big)$$

and we deduce that $f(x, g(x)) = 0$ for every $x \in V$.

REFERENCES

[1] BORWEIN, J. M. : The differentiability of real functions on normed linear spaces using generalized subgradients. *J. Math. Anal. Appl 128, num. 2(1987), 512–534.*

[2] BUTTLER, G. J., TIMOURIAN, J. G., VIGER, C. : The rank theorem for locally lipschitz continuous functions. *Canad. Math. Bull. 31 num. 2 (1988), 217–226.*

[3] CLARKE, F. H.: Generalized gradients and applications. *Trans. Amer. Math. Soc. 305 (1975), 246–262.*

[4] CLARKE, F. H.: On the inverse function theorem. *Pacif. J. Math. 64 num. 1 (1976), 97–102.*

[5] CLARKE, F. H.: Generalized gradients of Lipschitz functionals. *Adv. in Math. 40 (1981), 52–67.*

[6] HIRIART–URRUTY, J. B,:Miscellanies of non–smooth analysis and optimization. *Lect. Notes in Econ. and Math. Syst. 255 (1985), 8–24.*

[7] IOFFE, A.: Nonsmooth analysis: differential calculus of nondifferentiable mappings. *Trans. Amer. Math. Soc. 266 num 1 (1981), 1–56.*

[8] NOVO, V.: Diferenciación generalizada. Tésis doctoral. *Ed. Serv. Invest. UNED.*

[9] NOVO, V., RODRIGUEZ MARIN, L.: Derivada generalizada en $\mathbb{R}$. Aplicación al estudio de extremos. *Actas XV Jorn. Luso–Esp. de Mat. Univ. de Evora (1990).*

[10] POURCIAU, B. H.: Global invertibility of nonsmooth mapping. *J. Math. Anal. Appl. 131 (1988), 170–179.*

[11] ROCKAFELLAR, T.: Convex Analysis. *Princeton Univ. Press (1970).*