L. K. Raut

ISI LECTURE NOTES

CONVEX ANALYSIS

By

Lakshmi K. Raut

Indian Statistical Institute
7 SJS sansanwal Marg, New Delhi.

C O N T E N T S

NOTES AND REMARKS

Basic Concepts:

The theory of convex sets and convex functions originated in the early years of this century mainly by Minkwaski, For a finite dimensional extensive treatment of the subject, the reader should refer to Rockafellar (2). The materials on infinite dimensional spaces can be found in any functional analysis book. But the lecture notes of Moreau (9) is an excellent treatment of the subject in topological vector spaces of arbitrary dimension. In this connection, one can see the survey paper by Klee (11); see Kelley and Namioka for other generalizations.

There are many problems that arise in connection with the identification of the extreme points of a given convex set and there is an extensive literature in mathematics addressed to this problem for specific convex sets. See Kothe (10) pp. 333-337. Brønsted (12) tried to extend the Krein-Millman theorem and its converse to convex functions by using what is called affine minorants.

We did not have time and space to discuss about different types of convexity. A good account of it will be available in Ponstein (13).

For some more results on separation, one should refer to Klee (16(a)-(b)-(c)-(d)-(e)).

Differential and sub-differential Theory

On finite dimensional spaces, Rockafellar (2) deals very extensively with the subject. Surveys of infinite dimensional results are contained in Asplund (18), Ioffaand Tihomirov (19) Ekeland and Temam (20).

For the theory of conjugate functions, the reader is referred to Brønsted and Rockafellar (21), Brønsted (22) Fenchel (23). We mention here a work of Kutateladze and Rabinov (24) who pointed out that a natural language of various duality schemes is the language of K-spaces (Banach algebra). In (24), many analysis and interesting results are noted.

For decomposition theorem of sub-differentials on $\mathbb{R}^n$ which says that under certain general conditions on a convex function, every $x_0^* \in \partial f(x_0)$ can be decomposed into

$$x_0^* = \alpha_1 x_1^* + \alpha_2 x_2^* + \cdots + \alpha_r x_r^*$$

where $r \leq n+1$, $\sum_{i=1}^{r} \alpha_i = 1$, $\alpha_i > 0$, $x_i^* \in \partial f_{s_i}(x_0)$, $s_i \in S_0(x_0)$ for definition of the notations see Theorem 5.20, see (19(b)). Decomposition theorems are helpful to approximation theory. In this regard, a good reference is Levin (26).

For some characterisations of Banach spaces on which all continuous convex functions are Frechet differentiable (such spaces are called Asplund space) see Phelps (27) and for some characterisations of weak Asplund spaces, i.e. spaces on which all continuous convex functions are Gauteux differentiable see (28).

Acknowledgements

I like to thank specially to Mehar Lal and V.P.Sharma who managed to type this with great patience and care; I also like to thank T.Parthasarathy and Students of Indian Statistical Institute who encouraged me quite a lot by attending these lectures.

June, 1980

New Delhi

L.K.Raut

Lecture 1

Convex Sets, Affine Sets, Separation Theorems, Krein-Milman Theorem

Definition 1.1: A vector space over a field K (also called a linear space over a field K), is a set L having two operations - vector addition and scalar multiplication - satisfying the following: $\forall\ x,y,z \in L,\ \alpha\ \beta \in K$

(i) $x+y \in L$

(ii) $(x+y)+z = x+(y+z)$

(iii) there exists $0 \in L$ s.t. $x+0 = x \quad \forall\ x \in L$

(iv) $\forall\ x \in L,\ \exists (-x) \in L$ s.t. $x+(-x) = 0$

(v) $x+y = y+x$

(vi) $\alpha\ x \in L$

(vii) $\alpha\ (x+y) = \alpha x+ \alpha y$

(viii) $(\alpha+\beta)x = \alpha x + \beta x$

(ix) $(\alpha\ \beta)\ x = \alpha(\beta x)$

(x) $1\ x = x$.

Note that $(L, +)$ is an additive abelian group because of (i) through (v). In our lectures, we shall take $K = \mathbb{R}$.

Example 1.2: $\mathbb{R}^n$ over $\mathbb{R}$, $C_{[0,1]}$ over $\mathbb{R}$ etc.

Definition 1.3: A subspace of L is a subset M s.t. M itself is a linear space w.r.t. to the operations of L.

Definition 1.4: $A \subseteq L$ is said to be a convex set if whenever $x,y \in A$ and $0 \leq \alpha \leq 1$, then $\alpha x+(1-\alpha)y \in A$ and $A \subseteq L$ is said to be affine set if $x,y \in A$ and $\alpha \in \mathbb{R} \implies \alpha\ x+(1-\alpha)y \in A$.

Note that an affine set is also a convex set;but not theother way.

<u>Definition 1.5</u>: Let L and M be two linear spaces over $\mathbb{R}$. A transformation $T : L \to M$ is said to be <u>linear</u> if $T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$ $\forall\, x,y \in L$ and $\alpha, \beta \in \mathbb{R}$. If $M = \mathbb{R}$, we shall call this linear transformation as <u>linear functional</u>.

<u>Definition 6</u>. A transformation $T : L \to M$ (L,M are linear spaces over $\mathbb{R}$) is said to be <u>affine transformation</u> if

$$T(\alpha x + (1-\alpha)y) = \alpha T(x) + (1-\alpha)\, T(y) \quad \forall\, x,y \in L \text{ and } \alpha \in \mathbb{R}.$$

Note that a mapping which is linear is also affine but not the other way.

<u>Proposition 1.7</u>. Let L and M be linear spaces over $\mathbb{R}$ and $T : L \to M$ be a transformation.

T is affine <===> $T(x) = A(x)+b$ for some linear transformation $A : L \to M$ and some $b \in M$.

<u>Proof</u>: [===>] : Suppose T is affine. Set $b = T(0)$ and $A(x) = T(x)-b$. Note that

$$A(\alpha x) = T(\alpha x+(1-\alpha)\, 0) - b = \alpha T(x) + (1-\alpha - 1)\, T(0) = \alpha A(x)$$

$$A(x+y) = 2\, A\left(\frac{x+y}{2}\right) = 2\left(\tfrac{1}{2} T(x) + \tfrac{1}{2} T(y) - b\right)$$

$$= T(x) - b + T(y) - b$$

$$= A(x) + A(y)$$

Hence A is linear.

[<===]: Let $x,y \in L$, $\alpha \in \mathbb{R}$. Then

$$T(\alpha x+(1-\alpha)y) = \alpha A(x) + (1-\alpha)\, A(y) + (\alpha+(1-\alpha))\, b = \alpha T(x)+(1-\alpha)T(Y)$$

Q.E.D.

<u>Definition 1.8</u>: Let N be a proper subspace (affine subspace) of L. N is said to be <u>maximal proper subspace (maximal affine proper subspace)</u> if whenever K is a subspace (affine subspace) of L s.t. $N \subset K \subset L$, then either $K = L$ or $K = N$.

<u>Definition 1.9</u>: Let f be a linear functional on L. Then the "subspace" $N_f = \{x: f(x) = 0\}$ is called the <u>null space of f</u>.

<u>Proposition 1.10</u>: N is maximal proper subspace of L <==> N is the null space of a non-trivial linear functional.

<u>Proof</u>: [===>]: Suppose N is a maximal proper subspace. Then there exists $y \notin N$ in L. Set $K = \{\alpha y+x : x \in N, \alpha \in \mathbb{R}\}$. Note that K is a linear subspace with N as proper subspace. So, $K = L$. Define the functional

$$f(\alpha y+x) = \alpha \quad \forall x \in N \text{ and } \alpha \in \mathbb{R}$$

Note that f is a non-trivial linear functional with $N_f = N$.

[<====] : Suppose N is the null space of a non-trivial linear functional f and let $K \supsetneq N$ be a linear subspace. We shall show that $K = L$. $\exists\, y \in K$ s.t. $f(y) = \alpha \neq 0$ i.e. $f(\frac{y}{\alpha}) = 1$. Let $y/\alpha = y_o$. Observe that $\{\beta y_o+x : \beta \in \mathbb{R}, x \in N\} \subsetneq K \subseteq L$. We claim that $\{\beta y_o+x : \beta \in \mathbb{R}, x \in N\} \supset L$ for, let $z \in L$; define $x = z-f(z).y_o$. So

$$f(x) = f(z) - f(z)\, f(y_o) = 0 \implies x \in N \implies x + f(z).y_o \in \{\beta y_o+x : \beta \in \mathbb{R}, x \in N\}$$

Thhus $K = L$.

Q.E.D.

<u>Remark 1.11</u>: So there is a 1-1 correspondence between maximal proper subspaces and null spaces of non-trivial functionals on L. Something more is true; in fact, a maximal proper subspace is closed if and only if the corresponding f is continuous (Topology is norm-topology).

Warning: f is linear $\not\Rightarrow$ f is continuous unless L is finite dimensional.

The following is a characterisation of the maximal proper affine subspaces.

Proposition 1.12: H is maximal proper affine subspace of L <==> H = M+a, M is some maximal proper subspace of L and a is some vector in L.

Proof : Straightforward.

Definition 1.13: We define a hyperplane H as a maximal proper affine subset of L.

Question: Can we represent hyperplane H in terms of linear functionals as we do in $\mathbb{R}^2$, $\mathbb{R}^3$ by $a_1x_1 + a_2x_2 = b$; $a_1x_1 + a_2x_2 + a_3x_3 = b$ respectively? The following gives the positive answer.

Proposition 1.14: H is a hyperplane in L <===> H = $\{z \in L : f(z) = \alpha\}$ for some $\alpha \in \mathbb{R}$ and some non-trivial linear functional f on L.

Proof: [===>] Suppose H is a hyperplane in L. By definition, then H is a maximal proper affine subspace of L <==> H = M+a, M is some maximal proper subspace of L and a is some vector in L (by Proposition 1.12).

===> M = $\{x \in L : f(x) = 0\}$ for some non-trivial linear functional f.

===> H = $\{z = x+a \in L : f(z) = f(a) = \alpha\}$, where $\alpha = f(a)$.

[<===] :Select a $z_0 \in H$ and fix it.

Define N = $\{x \in L : x = z-z_0,\ z \in H\}$. Note that N = null space of f and hence a maximal proper subspace of L and H = N + z_0. Q.E.D.

Corollary 1.15: We know that there is a 1-1 correspondence between linear functionals f's on $\mathbb{R}^n$ and the vectors b $\in \mathbb{R}^n$. In $\mathbb{R}^n$ we can characterise the hyperplanes by H = $\{x \in \mathbb{R}^n \mid \langle x,b\rangle = \alpha\}$ for some non-zero $b \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ where $\langle x,b\rangle = \sum_{i=1}^{n} x_i b_i$ <==> H is a hyperplane. Moreover, every hyperplane in $\mathbb{R}^n$ can be represented this way with b and α unique upto a common non-zero multiple. The above vector b is called the normal to the hyperplane H. Note that every other normal to the hyperplane H is either a positive or negative scalar multiple of b.

Exercise 1.16: In $\mathbb{R}^n$, $M = \{x \in \mathbb{R}^n : Bx = b\}$ for some $b \in \mathbb{R}^m$ and B is some mxn real matrix <===> M is an affine set in $\mathbb{R}^n$. [Hint: M_1, M_2 affine sets ===> $M_1 \cap M_2$ is affine.]

Remark 1.17: (1.15) and (1.16) are true for any finite dimensional real vector space.

Definition 1.18: (1) By the term half spaces determined by the hyperplane $H(f,\alpha)$ we mean the sets $\{z \in L : f(z) \leq \alpha\}$ and $\{z \in L : f(z) \geq \alpha\}$.

(2) One says that $H(f,\alpha)$ separates U and V two subsets of L if U and V lie in opposite half spaces determined by $H(f,\alpha)$; and

(3) One says $H(f,\alpha)$ strongly separates U and V if there exist $\alpha_1, \alpha_2 \in \mathbb{R}$ s.t. $\alpha_1 < \alpha < \alpha_2$ and both $H(f, \alpha_1)$ and $H(f, \alpha_2)$ separate U and V.

Remark 1.19: Observe that strong separability requires U and V to be disjoint. The natural question arises: Is disjointness sufficient for it? The answer is, in general, no. Following is an example to justify it:

Example 1.20: Let $L = \mathbb{R}^2$

$$U = \{(r,s) \mid r > 0,\ s > 1/r\}$$

$$V = \{(r,s) \mid r > 0\ \ s \leq -1/r\}$$

Note that U and V are disjoint. But strong separation is not possible

Figure below gives the geometrical idea.

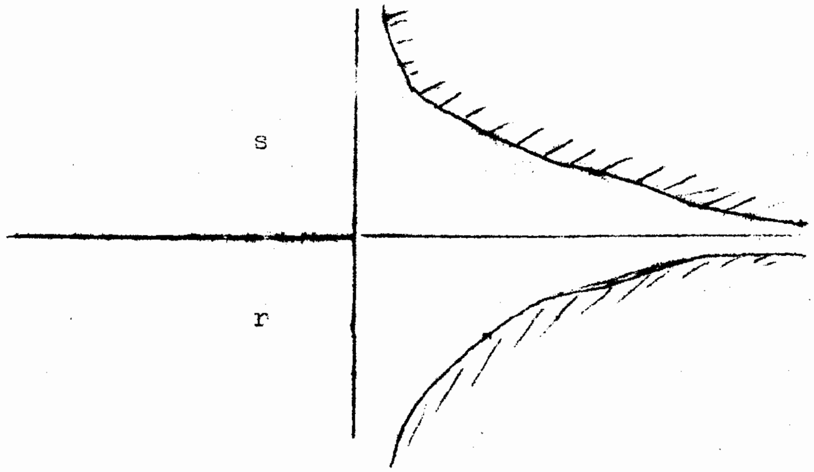


Figure 1.1

Any topological statements to be made here-after will be norm topology, unless otherwise mentioned.

The following theorem gives the sufficient condition for the existence of strongly separating hyperplanes for two sets U and V in L:

Theorem 1.21. (Strong Separation Theorem). Let U be a closed convex set disjoint with a compact convex set V, then there exist a hyperplane $H(f,\alpha)$ separating U and V strongly.

Proof: See [3].

Let U be a set in L. Then denote by

U^{o} = larget open set contained in U

= $\{x \varepsilon U : \exists$ an open set containing x and contained in U$\}$

and $\overline{U} = \bigcap C$, where intersection is taken over all closed sets C that contain U.

It's easy to varify that if V is convex then V^{o} and $\overline{V}$ are also convex. The following is known as separation theorem.

Theorem 1.22: (Separation Theorem); Let U and V be two convex subsets of L with $U^o \neq \phi$ and $U^o \cap V = \phi$. Then there is a (closed) hyper plane that separates U and V.

Proof: Follows from Hahn-Banach Theorem: See (3). The figure below gives the geometrical content of the theorem:

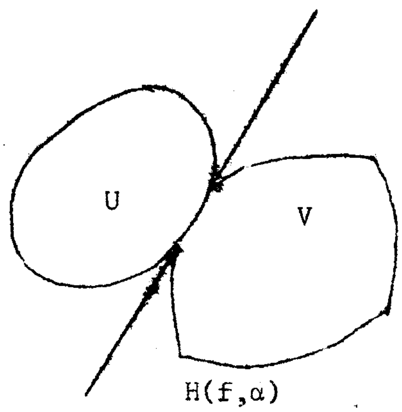


Figure 1.2.

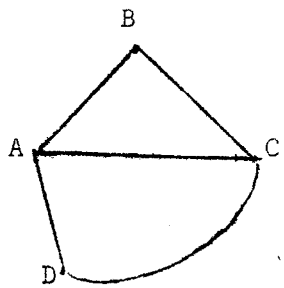
Definition: 1.23: Let U be a convex set in L. We say that a hyperplane H supports U at $x_o \in U$ if $x_o \in H$ and U is a subset of one of the half spaces determined by H

Corollary 1.24: (Support Theorem): Let U with $U^o \neq \phi$ be a convex set inL. Let $x_o \in U-U_o$. Then there is a (closed) supporting hyperplane H for U at x_o.

Proof: Take $V = \{x_o\}$ and U = U and appeal to theorem 1.22.

For x_1, $x_2 \in L$, let us denote $[x_1,x_2] = \{\alpha\, x_1+(1-\alpha)\, x_2 : 0 \leq \alpha \leq 1\}$

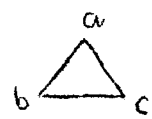
Definition 1.25: Let K be a non-void subset of L. A set $M \subset K$ is said to be extremal subset of K if an interior point of $[x_1,x_2]$, $x_1,x_2 \in K$, lies in K ===> $x_1,x_2, \in M$. An extremal subset of K consisting of single point is called an extreme point of K.



Point B, Δ ABC are extremal subsets the convex set K = ABCD. Furthermore the set consisting of point B is an extrme point of K.

Figure 1.3.

Example 1.26 (1) Let $K = [a,b]$, in $\mathbb{R}$. Extremal subsets are : K, {a} , {b}. Furthermore {a} , {b} are the two extreme points.

(2) $K = \Delta =$  in $\mathbb{R}^2$

Extremal subsets are : K,[a,b], [b,c], [a,c], {a} , {b} {c}

extreme points are : {a} , {b} {c}.

(3) $K = S^2$ = closed ball in $\mathbb{R}^3$.

Extremal subsets are : K, {x} , $x \in K - K^o$.

Extreme points are : {x} , $x \in K - K^o$.

Proposition 1.27: All [—] , Δ , S^2 and their finite dimensional analogues are compact and convex and they have extreme points. Any point in these sets can be written as the suitable convex combination of extreme points.

Proof: See Naimark (1).

Remark 1.28: In infinite dimensional linear space analogues of above sets need not be closed.

Define convex hull of E by smallest convex set containing E.

Theorem 1.29: (Krein-Milman): Every non-void compact convex subset K of L contains at least one extreme point.

Proof: See Naimark (1) for a proof.

Corollary 1.30: Every above type of K (i.e. compact convex) is the closed convex hull of its extreme points.

LECTURE 2

Convex Functions

In this lecture we do not need any topology on the linear space L.

Definition 2.1: Let U be a convex [convexity is needed for the definition] set in L. A functional $f : U \to \mathbb{R} \cup \{\infty\}$ is said to be convex if $\forall\, \alpha \in [0,1],\ x,y \in U,\ f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha)f(y)$.

Definition 2.2: (1) The effective domain (or simply domain) of f is the set $\text{dom } f = \{x \in U : f(x) < \infty\}$.

(2) The epigraph of f is the set

$E_f = \{(x,y) \in U \times \mathbb{R} : x \in U,\ y \in \mathbb{R} \text{ and } y \geq f(x)\}$.

The following is the characterisation of convex -functions in terms of convex sets.

Proposition 2.3: Let U be a convex set in L and f be a functional on U. Then f is convex $\Longleftrightarrow$ E_f is convex.

Proof: Straightforward.

2.1 SOME PROPERTIES OF CONVEX FUNCTIONS

Proposition 2.4: Let U be convex subset of L and f be a convex functional on U. Then dom f and $\{x \in U : f(x) \leq \alpha\}$ are convex, where $\alpha \in \mathbb{R}$.

Proof: Immediate.

Remark 2.5: For continuity properties and for some differentiability properties see the subsequent lectures.

Definition 2.6: One says that a convex function f is proper if $E_f \neq \phi$.

Exercise 2.7: Let $L = \mathbb{R}$ and $U = \mathbb{R}$. Show that the following functions are proper convex.

(i) $f(x) = e^{\alpha x} \quad -\infty < \alpha < \infty$

(ii) $f(x) = \begin{cases} x^p & \text{if } x \geq 0 \\ \infty & \text{if } x < 0 \end{cases} \quad 1 \leq p < \infty$

(iii) $f(x) = \begin{cases} \dfrac{1}{\alpha^2 - x^2} & \text{if } |x| < \alpha \\ \infty & \text{if } |x| \geq \alpha \end{cases} \quad \alpha > 0$

2.2 GENERATION OF CONVEX FUNCTIONS

Many operations on convex functions preserve convexity. In this section we investigate a few such.

Proposition 2.8: Let U be a convex subset of L, $f : U \to \mathbb{R} \cup \{\infty\}$ be convex, and $\phi : \mathbb{R} \cup\{\infty\} \to \mathbb{R} \cup \{\infty\}$ be convex non-decreasing with $\phi(\infty) = \infty$. Then $h = \phi o f$ is a convex functional on U.

Proof: Let $x,y \in U$ and $0 \leq \lambda \leq 1$. Since f is convex, we have $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$. Now apply ϕ to both sides of this inequality and get the result. Q.E.D.

Proposition 2.9: If f_1 and f_2 are proper convex functions on U then for $\lambda_1,\lambda_2 \geq 0$, $\lambda_1 f_1 + \lambda_2 f_2$ is convex.

Proof: Trivial.

Proposition 2.10: Let I be any index set and f_α for each $\alpha \in I$ is convex on U then $f(x) = \sup \{f_\alpha(x) : \alpha \in I\}$ is also convex.

Proof: Straightforward.

Proposition 2.11: Let $f_1, f_2, \ldots, f_m$ be proper convex functions on U a convex subset of L. Then following are also convex function on U:

(i) $f(x) = \inf \{\max \{f_1(x_1),\ldots,f_m(x_m)\} : x_1+\ldots+x_m = x\}$

(ii) $g(x) = \inf \{(f_1\lambda_1 + \ldots + f_m\lambda_m)(x) : \lambda_i \geq 0, \forall_i \text{ and } \lambda_1+\ldots+\lambda_m=1\}$

(iii) $h(x) = \inf \{\max \{(\lambda_1 f_1)(x),\ldots,(\lambda_m f_m)(x)\} : \lambda_i \geq 0, \forall_i \text{ and } \lambda_1+\ldots+\lambda_m=1\}$

(iv) $k(x) = \inf \{\max \{\lambda_1 f_1(x_1),\ldots, \lambda_m f_m x_m\}\}$,

where 'inf' is taken over all representations of x as a convex combination $x = \lambda_1 x_1 + \ldots + \lambda_m x_m$.

Proof: See R.Rockafellar (2).

LECTURE 3

Continuity of Convex Functions

3.1. Introduction

In this lecture we shall explore some ideas related to the continuity of convex functions on a NLS. Convex function of one real variable, is always continuous. The result is not true for functions on infinite dimensional vector spaces; however, any convex function on an open subset of $\mathbb{R}^n$ is always continuous. Then, naturally, one would like to ask : what is the sufficient condition for the continuity of a convex function on an open subset of a NLS? We shall investigate all these.

3.2 Continuity through restriction on f.

L will be understood as NLS unless otherwise stated.

Proposition 3.1. Let U be an open set in L and $f : U \to \mathbb{R}$ be convex. If f is bounded from above in a neighbourhood of one point $x_o \in U$, then it is locally bounded; i.e, each $x \in U$ has a neighbourhood on which f is bounded.

Proof: See A.W.Roberts and D.E.Varberg pp. 91-92.

Definition 3.2: A function f defined on an open subset U of L is said to be locally Lipschitz if for each $x \in U$ there exists a neighbourhood $N_\varepsilon(x)$ and a constant K(x) s.t. $y,z \in N_\varepsilon(x)$

$$\Longrightarrow \quad |f(y) - f(z)| \leq K(x)||y-z||$$

and said to be Lipschitz on $V \subset U$ if $\exists K$, independent of x s.t.

$$|f(y) - f(z)| \leq K||y-z|| \quad \forall y,z \in V.$$

Proposition 3.3: Let f be convex on an open set U in L. If f is bounded from above in a neighbourhood of one point of U, then

(i) f is locally Lipschitz in U,

(ii) f is continuous on U; and

(iii) f is Lipschitz on any compact subset of U.

Proof: Proposition 3.1 ==> f is locally bounded. So, given $x_o \in U$, we can find a neighbourhood $N_{2\varepsilon}(x_o)$ s.t. f(x) is bounded in this neighbourhood say by M. Now, note that f is Lipschitz on $N_\varepsilon(x_o)$.

(i) ==> (ii) trivially.

To prove (iii). Let $K \subset U$ be compact. (i) ==> $\forall x \in U$; $\exists$ open neighbourhood $N_{r_x}(x)$ and K(x) s.t. $|f(w)-f(y)| < K(x)||w-y|| \forall w,y \in N_{r_x}(x)$. Note that $\{N_{r_x}(x) \mid x \in U\}$ is an open covering for K and hence there is a finite sub cover, say, $\{N_{r_{x_1}}(x_n)\}$. Define

$$K = \max \{K(x_1), \ldots, K(x_n)\}$$

and note that $\forall x,y \in K$, one has

$$|f(x) - f(y)| \leq K||x-y||$$

Q.E.D.

Exercise 3.4: Let f be convex on the open set $U \subset \mathbb{R}^n$. Then f is Lipschitz on every compact subset of U and continuous on U.

Exercise 3.5: A convex function f: $(a,b) \to \mathbb{R}$ is absolutely continuous on any closed subinterval of (a,b). For several suggestions on the extension of the notion of absolute continuity to function of several variables. In the ~~connection~~ context of convex functions, see Friedman (1940).

3.3 Continuity through restrictions on the underlying space.

Definition 3.6: A function $f : U \to \mathbb{R}$ is said to be lower semicontinuous at $x_o \in U \subset L$ if, given $\varepsilon > 0$, $\exists$ $N(x_o)$ s.t. $\forall x \in N(x_o)$, one has $f(x) > f(x_o) - \varepsilon$.

Proposition 3.7: Let U be open subset of a Banach space L and $f : U \to \mathbb{R}$ be lower semi continuous. Then f is continuous on U.

Proof: Same type of argument will do as in Proposition 3.9.

Definition 3.8: $f : U \to \mathbb{R}$ is said to have support at $x_o \in U$ if there is an affine function $A : L \to \mathbb{R}$ s.t. $A(x_o) = f(x_o)$ and $A(x) \leq f(x)$ $\forall x \in U$.

Proposition 3.9. Suppose $f : U \to \mathbb{R}$ has continuous support at each point of a convex open set U in a Banach space L. Then f is a continuous convex function.

Proof: See A.W.Roberts and D.E. Varberg, pp. 109.

LECTURE 4

Differential Theory of Convex Functions

4.1 Introduction

In this lecture, we shall discuss what is differentiation of a mapping from one normed linear space to another normed linear space. We shall mention different types of differential concepts and their inter-relationships. These concepts and results will be useful in studying extremum problems of convex functions under differentiability assumption and -under without differentiability assumption. We shall deduce some interesting differential properties possessed by convex functions.

4.2 Different types of derivatives

Let $(L, ||\cdot||)$ and $(M, ||\cdot||)$ be the normed linear spaces.

Definition 4.1: Let $f : L \to \mathbb{R} \cup \{\infty\}$ be any function. Let $x_o \in$ r i dom f, $v \in L$ then the one sided directional derivative of f at x_o in the direction v is

$$f'_+(x_o,v) = \lim_{\lambda \downarrow 0} \frac{f(x_o + \lambda v) - f(x_o)}{\lambda}$$

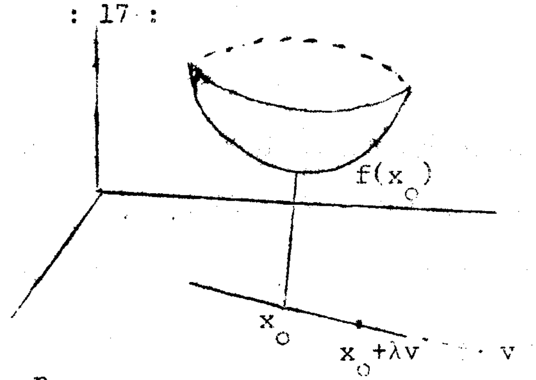
if it exists ($+\infty$ and $-\infty$ are allowed as limits); and the two sided directional derivative of f at x_o in the direction v is

$$f'(x_o,v) = \lim_{\lambda \to 0} \frac{f(x_o + \lambda v) - f(x_o)}{\lambda}$$

if it exists.

Note that when $f'(x_o,v)$ exists, one has $f'(x_o,v) = -f'(x_o-v)$

The following figure may help to visualise the above concepts.



<u>Special Case 4.2.</u> Let L = $\mathbb{R}^n$ and v be one from $\{e_1, e_2, \ldots, e_n\}$, the standard basis. Then the corresponding directional derivatives are denoted by $\frac{\partial f}{\partial x_i}(x_o)$ or by $f^i(x_o)$ when $v = e_i$ and this is called the <u>i-th partial derivative</u> of f at x_o.

<u>Remark 4.3:</u> From the study of functions of one real variable, we have learned that the existence of onesided derivatives are not enough to ensure the smoothness of the curve at a point. E.g. $f(x) = |x|, x \in \mathbb{R}$ and at the point x = 0. So, the existence of two sided derivative is necessary to ensure smoothness of the graph at a point in the sense of making it possible to "approximate" the function in a very small neighbourhood of the point x_o by a function of the form $A(x) = f(x_o) + m(x-x_o)$. The concept of above type of smoothness gives rise to a concept of differentiability called Frechet derivative defined below:

But for functions of several variables even the existence of the directional derivatives in all the directions at a point x_o, is not enough to ensure the smoothness of the function at that point in the sense stated above. Following exercise is to justify this.

<u>Exercise 4.4:</u> Let

$$f(x,y) = \frac{x^2 y}{x^2+y^2} \quad \text{if } (x,y) \neq (0,0)$$
$$= 0 \quad \text{if } (x,y) = (0,0)$$

Show that at (0,0) all the directional derivatives exist yet f'(0) [defined below] does not exist.

Question 4.5: What is the sufficient condition under which $f'(x)$ exists? For partial answer, see (*) of remark 4.14.

Definition 4.6. Let U be an open subset of L and $f : U \to M$. We say f is Frechet differentiable at $x_o \in U$ if there is a linear transformation $T : L \to M$ s.t. for all $h \in L$ of sufficiently small norm, one has

$$f(x_o+h) = f(x_o) + Th + ||h||\, e(x_o,h)$$

where $||e(x_o,h)|| \to 0$ as $||h|| \to 0$. This linear transformation, if it exists, is called Fréchet derivative of f at x_o and is denoted by $f'(x_o)$. Sometimes we denote it by $D\,f(x_o)$. Equivalently, if there exists a linear transformation $T : L \to M$ s.t.

$$\lim_{h \to 0} \frac{||f(x_o+h) - f(x_o) - TH||_M}{||h||_L} = 0$$

Then f is said to have Fréchet derivative at x_o. If f is differentiable at every $x \in U$. Then f is said to be differentiable in U. One can show easily (the following exercise) that when such a linear transformation exists, it is unique.

Exercise 4.7.: Show that T in the definition 4.6 is unique whenever it exists.

Remark 4.8: Let Lin (L,M) be the set of all linear transformations from the normed linear space L to the norm linear space M, and let us denote by B(L,M) the class of bounded (equivalently continuous) linear transformations from L to M. Then f' is nothing but a mapping

$$f' : U \to \text{Lin}\,(L,M).$$

Note that for $x_o \in U$, $f'(x_o)$ may not be continuous. Natural question arises: when is $f'(U) \subset B(L,M)$? The following theorems answers it.

Warning: We are not talking about the continuity of the differential operator f'.

Theorem 4.9: $f'(x_o) \in B(L,M) \Longleftrightarrow f$ is continuous at x_o.

Proof: [$\Longrightarrow$] suppose $f'(x_o) \in B(L,M)$. So there exist $\alpha \in \mathbb{R}$ s.t. $||f'(x_o)|| \leq \alpha$ Now, by definition of $f'(x_o)$ for h of small norm in L, we have

$$f(x_o+h) - f(x_0) = f'(x_o)h + ||h||_L \, e(x_o,h)$$

i.e.

$$||f(x_o+h) - f(x_o)||_M \leq ||f'(x_o)||_B \, ||h||_L + ||h||_L \, ||e(x_o,h)||_M$$

$$\leq (\alpha + ||e(x_o,h)||)||h||$$

So whenever $||h|| \to 0$ $||f(x_o+h) - f(x_o)|| \to 0$. Hence f is continuous at x_o.

[$\Longleftarrow$] Suppose f is continuous at x_o. Suppose $f'(x_o)$ exists. Let h_n be very near to 0.

$$f(x_o + h_n) - f(x_o) = f'(x_o)\, h_n + ||h_n||\, e(x_o,h_n)$$

i.e.

$$||f'(x_o)\, h_n|| \leq ||f(x_0+h_n) - f(x_0)|| + ||h_n||\, ||e(x_o,h_n)||$$

So, $f'(x_o)h_n \xrightarrow{||\,.\,||_M} 0$ as $h_n \xrightarrow{||\;||_L} 0$. Hence $f'(x_o)$ is continuous at 0 and thus continuous $\forall\, h \in L$.

Q.E.D.

Remark 4.10: Theorem 4.9 is relevant if one can produce an example of a function which is not continuous at some point but differentiable. One such function is as given in example 4.11. In this connection, we should note that since Lin $(L,M) = B(L,M)$ if L and M are finite dimensional, so the differentiability of f at a point implies continuity.

Example 4.11: Let $L = \ell_2 = \{\underset{\sim}{x} = \{x_i\} : \sum_{i=1}^{\infty} |x_i| < \infty\}$ with the "norm" $||\{x_i\}|| = \sup\{|x_i|, i = 1,2,\ldots\}$. Let $T : \ell_2 \longrightarrow \mathbb{R}$ be defined by $T(\{x_i\}) = \sum_1^{\infty} x_i$. It is easy to check that T is linear. T is not bounded, for consider the sequence $\{\underset{\sim}{x}^n\}$,

$$\underset{\sim}{x}^n = \{\underbrace{1,1,\dots 1}_{n},\ 0,0,\dots\}$$

So $||\underset{\sim}{x}|| = 1$ and $||T\,\underset{\sim}{x}^n|| = n$. So $||T|| = \sup \dfrac{||T\,\underset{\sim}{x}||}{||x||} = \infty$. So T is not continuous . But $T'(x_o) = T$.

So this is an example of a function not continuous at a point but differentiable.

The following proposition relates Fréchet derivative to directional derivative.

<u>Proposition 4.12</u>: Let U be open in L and $f : U \to \mathbb{R}$. If f has Fréchet derivative at $x_o \in U$ then f has all two-sided directional derivatives and

$$f'(x_o;v) = f'(x_o)(v)$$

<u>Proof</u>: By definition

$$f'(x_o;v) = \lim_{\lambda \to 0} \frac{f(x_o + \lambda v) - f(x_o)}{\lambda}$$

$$= \lim_{\lambda \to 0} \frac{f'(x_o)(\lambda v) + ||\lambda v||\ e(x_o, \lambda v)}{\lambda}$$

$$= f'(x_o)(v) + 0.$$

Q.E.D.

Henceforth, whenever we talk about derivative of f, we assume f is continuous unless otherwise stated. Note that $f'(x_o)$ as x_o varies over U can be thought of as a mapping $f' : U \longrightarrow B(L,M)$. Domain of f' is the set of those $x \in U$ for which $f'(x_o)$ exists and bounded. This mapping will, henceforth, be called "<u>differential operator</u>". The question that naturally comes to one's mind: Is the differential operator continuous w.r.t. the induced norm topology on B(L,M)? In this connection see Remark 4.14.

<u>Definition 4.13</u>: If for a $f: U \to M$, the differential operator f' is continuous we say that f is of class $C^1(U)$ or simply $f \in C^1$.

Remark 4.14. For $f : \mathbb{R}^n \to \mathbb{R}^m$, we can take any norm for $B(\mathbb{R}^n, \mathbb{R}^m)$ (since for finite dimensional vector spaces any two norms are "equivalent" - i.e. induce the same topology). A convenient way of giving a norm to $f'(x_o)$ is to consider the standard basis and define

$$||f'(x_o)|| = \max_{i,j} \left| \frac{\partial f_i}{\partial x_j}(x_o) \right|$$

$||.||$ is a norm in $B(\mathbb{R}^n, \mathbb{R}^m)$. Whenever all partial derivatives $\frac{\partial f_i}{\partial x_j}$ exist and are continuous on U, then f' exist on U and by above choice of the norm, $f \in C^1(U)$. (*) So, $f \in C^1(U)$ whenever all partial derivatives $\frac{\partial f_i}{\partial x_j}$ exist and are continuous on U.

4.3 Higher Order derivatives

Suppose $f \in C^1(U)$. Note that B(L,M) is a normed linear space, so we can talk about the differentiability of $f': U \longrightarrow B(L,M)$. If $f'' = (f')'$ exists at a point $x_o \in U$ then $f''(x_o)$ will be an element of B(L,B(L,M)) which is unique. Such an operator is called the 2nd order derivative of f at x_o. If, moreover, $f'': U \longrightarrow B(L,B(L,M))$ is continuous then one says that $f \in C^2(U)$ or f is of C^2-class. In the same way one can define the r-th order derivative of f (denote by $f^{(r)}$). But observe that $f^{(r)}(x_o)$ will be an element in a complicated space like.

$$B(L,B(L,\dots B(L,B(L,M))\dots)),$$

which is very cumbersome to deal with. But life is not so bad. We shall establish a linear isomorphism between the above space and a simpler space namely the space of r-multilinear mappings as defined below.

Definition 4.15. Let L, $L_1, L_2, \dots L_n$ are n+1 linear spaces. An map

$M : L_1 \times L_2 \times \dots \times L_n \longrightarrow L,$

which is linear in each factor separately (i.e. when all the components except one are held fixed, it is linear in the independent variable), is called an <u>n-multilinear mapping</u>.

Example 4.16: Let $L_1 = L_2 = \ldots = L_n = L = \mathbb{R}$. It's elementary to check that the mapping $(x_1, x_2, \ldots, x_n) \longmapsto x_1 \cdot x_2 \ldots x_n$ is a n-multilinear map.

<u>Example 4.17.</u> Consider nxn matrices as elements from $\mathbb{R}^n \times \mathbb{R}^n \times \ldots \times \mathbb{R}^n$. And consider the mapping

$$M : \mathbb{R}^n \times \mathbb{R}^n \times \ldots \times \mathbb{R}^n \longrightarrow \mathbb{R} \text{ defined by}$$

$$M(.) = \det(.)$$

Then M is an n-multilinear mapping.

[Hint: Use the fact: $\det(A) = a_{11} A_{11} + a_{21} A_{21} + \ldots + a_{n1} A_{n1} = a^{1'} A.1,$] where a^1 is the first column of A and $A'.1 = (A_{11}, A_{21}, \ldots, A_{n1})$ vector of co-factors $(a_{11}, \ldots, a_{n1})$. Note that M is not linear.]

Now we shall prove the linear isomorphism between the above said spaces for r = 2. For general case also the same proof goes through. Let $M_2 (L_1 \times L_2; L)$ be the "linear space" (with pointwise addition and scalar multiplication in the natural way) of continuous bi-linear maps from $L_1 \times L_2$ to L, where L_1, L_2, L are NLS.

<u>Theorem 4.18</u>: There exists a linear isomorphism between $M_2(L_1 \times L_2, L)$ and $B(L_1, B(L_2, L))$.

<u>Proof</u>: Let us construct a mapping $\phi : B(L_1, B(L_2, L)) \to M_2(L_1 \times L_2, L)$ by $A \to \phi(A) = \bar{A}$, where $\bar{A}$ is defined in the following way: for given $x_1 \in L_1$, A will associate an element say $A(x_1)$ in $B(L_2, L)$. Now, let $x_2 \in L_2$ be taken to $A(x_1)x_2$ in L define $\bar{A}(x_1, x_2) = A(x_1)x_2$ as obtained by the above way.

Claim 1: $\overline{A} \in M_2(L_1 \times L_2, L)$ which is trivial to varify.

Claim 2: ϕ is linear. For, let $A_1, A_2 \in B(L_1, B(L_2, L))$ and $\alpha \in \mathbb{R}$

$$\phi(A_1+A_2) = (A_1+A_2)(x_1)x_2 = (A_1(x_1) + A_2(x_1))\, x_2$$

$$= A_1(x_1)x_2 + A_2(x_1)x_2$$

$$= \overline{A}_1 + \overline{A}_2 = \phi(A_1) + \phi(A_2)$$

Similarly, $\phi(\alpha A) = \alpha\phi(A)$.

Claim 3: ϕ is 1-1 and onto. Varification is pathological.

Claim 4: $||A|| = ||\phi(A)||$ for,

$$||A|| = \sup_{\substack{x_1 \in L_1 \\ ||x_1|| \neq 0}} \frac{||Ax_1||}{||x_1||} = \sup_{\substack{x_1,\, x_2 \\ ||x_1|| \neq 0 \\ ||x_2|| \neq 0}} \frac{||A(x_1)x_2||}{||x_1||\,||x_2||} = ||\phi(A)||$$

Q.E.D

Remark 4.19: Because of above theorem, the 2nd order derivative of a map $f : U \longrightarrow M$ can be identified with a bilinear operator from $L \times L \longrightarrow M$. Similarly $f^{(r)}$ is an r-multilinear mapping $L \times L \times \ldots \times L \longrightarrow M$. The following theorem supplies even a simpler form for 2nd order derivative for a function $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ in terms of matrix. Let $\mathcal{M}_n$ = the space of all nxn real matrices.

Proposition 4.20: There is a "linear isomorphism" between $M_2(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R})$ and $\mathcal{M}_n$.

Proof: Construction : Let $\{e_1, \ldots, e_n\}$ be the standard basis for $\mathbb{R}^n$ and let $\underset{\sim}{h}, \underset{\sim}{k} \in \mathbb{R}^n$. Then there exist unique $h_i, k_i \in \mathbb{R}$, i=1,2,...,n s.t.

$$\underset{\sim}{h} = h_1 e_1 + \ldots + h_n e_n$$

$$\underset{\sim}{k} = k_1 e_1 + \ldots + k_n e_n$$

Let $M \in M_2(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R})$. Then

$$M(\underset{\sim}{h},\underset{\sim}{k}) = M(h_1e_1+\ldots+h_ne_n,k) = \sum h_i\ M(e_i,\ k_1e_1+\ldots+k_ne_n)$$
$$\sum_{i=1}^{n}\ \sum_{j=1}^{n} h_ik_j\ M(e_i,e_j)$$
$$\sum_{i=1}^{n}\ \sum_{j=1}^{n} h_i\ k_j\ a_{ij},\ \text{say},$$

where $a_{ij} = M(e_i,e_j)$, $j,i = 1,2,\ldots,n$ which are independent of $\underset{\sim}{k}$ and $\underset{\sim}{h}$.

Let $A = ((a_{ij}))_{mxn}$. Define the mapping

$$\phi: M_2(LxL,\ M) \to \mathcal{M}_n \text{ by}$$
$$M \to \phi(M) = A.$$

It's pathalogical to verify that ϕ is a linear bijection. ϕ is continuous owing to the fact that it is a linear map from one finite dimensional NLS to another finite dimensional NLS. Q.E.D

Exercise 4.21: Let $f : \mathbb{R}^n \to \mathbb{R}$. Suppose $f''(x)$ exists at x. Show that $f''(x)$ has the associated matrix.

$$H(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1\ \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_1\ \partial x_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial^2 f(x)}{\partial x_n\ \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n\ \partial x_n} \end{bmatrix}$$

$H(x)$ is called the Hessian matrix of f at x.

Definition 4.22: $A \in M_2(LxL,M)$ is said to be symmetric if $A(h,k) = A(k,h)$ $\forall$ k, h $\in$ L; and for $M = \mathbb{R}$ it is said to be positive definite (resp. non-negative definite) if $A(h,h) > 0$ (resp. ≥ 0) $\forall$ h $\in$ L.

The following Theorem is very useful.

Theorem 4.23: Let $f : U \to M$ be of C^1 class. Then $f''(x)$ is symmetric whenever it exists.

Proof: See (3)

<u>Theorem 4.24:</u> (Taylor's Theorem): Let U be a open convex subset of L, $f: U \to \mathbb{R}$ be $C^1(U)$ class and $f''(x)$ exists $\forall\, x \in U$. Then for any $x, x_o \in U$, there is an $s \in (0,1)$ s.t.

$$f(x) = f(x_o) + f'(x_o)h + \frac{1}{2} f''(x_o+sh)(h,h),$$

where $h = x-x_o$.

<u>Proof:</u> See K.R.Parthasarathy and Rajendra Bhatia, 152-153, where they give a proof for more generalized version of it.

4.4 <u>Convex Functions and Different Derivatives:</u>

In this section we shall present some theorems characterizing convex functions on $(L, ||.||)$ with the help of first and second order verivatives.

<u>Theorem 4.25</u>: Let U be an open convex subset of $(L, ||.||)$ and let $f : U \to \mathbb{R}$ be convex and differentiable at $x_o \in U$. Then $\forall\, x \in U$.

$$f(x) - f(x_o) \geq f'(x_o)(x-x_o)$$

<u>Moreover</u>, Suppose $f'(x)$ exists $\forall\, x \in U$ then $\forall\, x, x_o \in U$

$$f(x) - f(x_o) \geq f'(x_o)(x-x_o) \iff f \text{ is convex on } U.$$

(Strict inequality for strict convexity).

<u>Proof:</u> Suppose f is convex on U and differentiable at $x_o \in u$. Then for $\lambda \in (0,1)$ and $x \in U$,

$$f((1-\lambda)x_o + \lambda x) \leq (1-\lambda) f(x_o) + \lambda f(x)$$

i.e. $$f(x_o + \lambda h) - f(x_o) \leq \lambda(f(x_o+h) - f(x_0)), \text{ where } h = x-x_o.$$

i.e. $$\frac{f(x_o+ \lambda h) - f(x_0)}{\lambda} \leq f(x_o+h)-f(x_0)$$

i.e. $$\frac{f(x_o+ \lambda h) - f(x_0) - f'(x_o)(\lambda h)}{\lambda} \leq f(x_o+h)-f(x_o)-f'(x_o)\, h$$

Letting $\lambda \to 0$ establishes the 1st part of the theorem

2nd part: <=== is already proved. For ===> part, let x_1, x_2, ε U.

set $x_o = t\, x_1 + (1-t)x_2$. Want to show $f(x_o) \leq t\, f(x_1) + (1-t)f(x_2)$

Note that $f(x_o) = f(x_o) + f'(x_o)\,(t(x_1-x_o) + (1-t)(x_2-x_o))$

$$\Longrightarrow \quad f(x_o) = t(f(x_o)+f'(x_o)(x_1-x_o)) + (1-t)(f(x_o) + f'(x_o)(x_2-x_o))$$

$$\leq t\, f(x_1) + (1-t)\, f(x_2) \qquad \text{by hypothesis.}$$

Q.E.D.

Remark 4.26: The first part of the above theorem 4.25 can be made weaker in terms of directional derivatives (cf. theorem). We have, for convex functions of one real variable, a result that f"(x) is monotonically increasing function of x. Can't we have an analogue of this: yes, as follows:

Definition 4.27: We say that a map $F : L \to \text{Lin}(L, \mathbb{R})$ is monotonic increasing if $\forall\, x,y \in L$, we have

$$(F(x) - F(y))(x-y) \geq 0.$$

and strictly monotonic increasing if $\forall\, x,y \in L$, $x \neq y$

$$(F(x) - F(y))(x-y) > 0.$$

Note 4.28: This concept will be further generalized for 'relations' in the next lecture. See definition 5.7.

Theorem 4.29: Let U be open convex subset of L and let $f : U \to \mathbb{R}$ be continuous and differentiable on U. Then f is convex on U <=> $f'(\varepsilon B(L, \mathbb{R}))$ is monotonic increasing on U.

Proof [===>] part follows from theorem 4.25.

[<====]: Suppose f' is monotonic increasing. Fix $x,y \in U$ arbitrary. We want to show for $\alpha \in (0,1)$, $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)\, f(y)$.

Define $\phi: [0,1] \to \mathbb{R}$ by

$$\phi(\alpha) = f(\alpha x+(1-\alpha)\, y).$$

Claim 1: $\phi'(\alpha)$ is monotonic increasing, for, let $0 \leq \alpha_1 < \alpha_2 \leq 1$ and $u_1 = \alpha_1 x + (1-\alpha_1)y$ $u_2 = \alpha_2 x + (1-\alpha_2)y$. So, $u_2 - u_1 = (\alpha_2 - \alpha_1)(x-y)$ and

$0 \leq (f'(u_2) - f'(u_1)(u_2-u_1) = (\alpha_2 - \alpha_1)(f'(u_2) - f'(u_1))(x-y)$

$$\Longrightarrow \quad f'(u_1)(x-y) \leq f'(u_2)(x-y).$$

$$\Longrightarrow \quad \phi'(\alpha_1) = f'(u_1)(x-y) \leq f'(u_2)(x-y) = \phi'(\alpha_2) \text{ (by chain rule).}$$

Hence the claim 1.

Claim 1 $\Longrightarrow$ $\phi(\alpha)$ is convex. Thus

$$f(\alpha x + (1-\alpha)y) = \phi(\alpha) = \phi(\alpha.1 + (1-\alpha).0) \leq \alpha\phi(1) + (1-\alpha)\phi(0)$$
$$\leq \alpha f(x) + (1-\alpha) f(y)$$

Q.E.D

The following is a characterisation in terms of 2nd order derivative.

Theorem 4.30) Let U be an open convex set in L and let $f: U \to \mathbb{R}$ be s.t. $f \in C^1(U)$ and $f''(x)$ $(\in M_2(L \times L, \mathbb{R}))$ exists $\forall x \in U$. Then,

$$f \text{ is convex on } U \Longleftrightarrow f''(x) \text{ is non-negative definite } \forall x \in U$$

(strict convexity $\Longleftrightarrow$ positive definite).

Proof: [<===] Suppose $f''(x)$ is n.n.d. $\forall x \in U$. Let $x_o \in U$ be arbitrary. By Taylor's theorem (Theorem 4.24),

$$f(x_o+h) = f(x_o) + f'(x_o)h + \frac{1}{2} f''(x_o+sh)(h,h) \text{ for some } s \in (0,1).$$
$$\geq f(x_o) + f'(x_o)h \quad \text{(by hypothesis).}$$

Appeal to theorem 4.25 says that f is convex.

[===>] Suppose f is convex. Fix $x \in U$ and $h \in L$ arbitrarily and define a function $g : \mathbb{R} \to \mathbb{R}$ by

$$g(t) = f(x + th)$$

Observe that $g(t)$ is convex in some neighbourhood of 0. Now, by composition rule for differentiation

$$g'(t) = f'(x + th)h$$

$$g''(t) = f''(x+th)(h,h)$$

For convex functions $g(t)$ of one real variable we know that $g''(t) \geq 0$ in its domain. So, in particular,

$$g''(0) = f''(x)\,(h,h) \geq 0$$

Since x and h are arbitrary, the proof is complete. (For strict convexity all inequalities are to be replaced by strict inequality).

Q.E.D.

The following is the support characterisation of the convex functions.

Theorem 4.31. Let U be an open subset of L and $f : U \to \mathbb{R}$. Then, f is convex on U <===> f has support at each $x \in U$.

Proof: [<===]: Let $x, y \in U$ and for $\alpha \in (0,1)$, Let $x_o = \alpha x + (1-\alpha)\, y$. Let $A(x) = f(x_o) + T(x-x_o)$ be the affine function that support f at x_o, where T is a linear functional. Then

$$f(x_o) = f(\alpha x + (1-\alpha)y) = A(x_o) = \alpha A(x) + (1-\alpha)A(y)$$
$$\leq \alpha f(x) + (1-\alpha)\, f(y)$$

===> f is convex on U.

[<===]: Left as an exercise to the reader.

Exercise 4.32: Let f be convex on an open subset U of L and at $x_o \in U$, $f'(x_o)$ exists then f has the unique support at x_o given by $A(x) = f(x_o) + f'(x_o)(x-x_o)$.

Exercise 4.33: In the previous exercise if $L = \mathbb{R}^n$, then f has unique support at x_o <===> f is (Fréchet) differentiable at x_o.

So exercises 4.32 and 4.33 give a necessary and sufficient condition for Fréchet differentiability of a convex function on $\mathbb{R}^n$ and the following exercise ensures the Frechet differentiability a.e.

Exercise 4.34: Let $f : U \to \mathbb{R}$ be convex on an open set U in $\mathbb{R}^n$. Then f is (Fréchet) differentiable a.e x in U.

We conclude this Lecture by giving some more differential properties of convex functions in terms of directional derivatives. An important property of convex functions is the existence of directional derivatives at all points of their effective domains.

Theorem 4.35: Let f be convex on an open subset U of L. Then the following are true.

(i) $\forall\, x \in U,\ v \in L,\ f'_+(x;v)$ exists

(ii) $\forall\, x, x_o \in U,\ f(x)-f(x_o) \geq f'_+(x_o;\, x-x_o)$

(iii) $f'_-(x;v) \leq f'_+(x;v)$

(iv) $f'_+(x;v)$ is positively homogeneous sub-additive function in v, i.e.

$f'_+(x;\, \lambda v) = \lambda f'_+(x;v)\ \forall\ \lambda > 0,\ v \in L$ and

$f'_+(x;\, v_1+v_2) \leq f'_+(x;v_1) + f'_+(x;\, v_2)$.

Proof: (i) Let us prove first when $L = \mathbb{R}$. Let $t_1 < t_2 < t_3$. t_1 and t_2 are in dom f. Since f is convex,

$$f(t_2) \leq \frac{t_3-t_2}{t_3-t_1} f(t_1) + \frac{t_2-t_1}{t_3-t_1} f(t_3)$$

From which,

$$f(t_2)-f(t_1) \leq \frac{t_2-t_1}{t_3-t_1}(f(t_3) - f(t_1)) \text{ and}$$

$$f(t_3)-f(t_2) \geq \frac{t_3-t_2}{t_3-t_1} (f(t_3)-f(t_1))$$

Combining,

$$\frac{f(t_2)-f(t_1)}{t_2-t_1} \leq \frac{f(t_3) - f(t_1)}{t_3-t_1} \leq \frac{f(t_3)-f(t_2)}{t_3-t_2}$$

This implies, $\forall\ t \in$ dom f,

$$\frac{f(t+\lambda) - f(t)}{\lambda}$$

is decreasing as λ decreases (if t=right end point of dom f, then this quotent is identically equal to ∞ w.r.t.$\lambda > 0$). So $f_+'(t)=f_+'(t,1) = \lim_{\lambda\downarrow 0} \frac{f(t+\lambda)-f(t)}{\lambda}$ $\forall$ t ε dom f. For general space L, define for x ε dom f and y ε L,

$$\phi(t) = f(x+ty).$$

Note that $\phi_+'(0) = f_+'(x;y)$. Existence is provided by previous step. Hence (i) is proved.

(ii) By definition

$$f_+'(x_o, x-x_o) = \lim_{\lambda\downarrow 0} \frac{f(x_o + \lambda(x-x_o)) - f(x_o)}{\lambda}$$

$$= \lim_{\lambda\downarrow 0} \frac{f(\lambda x + (1-\lambda)x_o) - f(x_o)}{\lambda}$$

$$\leq \lim_{\lambda\downarrow 0} \frac{\lambda f(x) - \lambda f(x_o)}{\lambda} = f(x) - f(x_o).$$ Hence (ii).

(iii) is immediate.

(iv) positive homogeneity is trivial to varify. For sub-additivity;

$$f_+'(x,y+z) = \lim_{\lambda\downarrow 0} \frac{f(x+(\lambda/2)(y+z)) - f(x)}{\lambda/2}$$

$$= \lim_{\lambda\downarrow 0} \frac{f(\frac{(x+\lambda y)}{2} + \frac{x + \lambda z}{2}) - f(x)}{\lambda/2}$$

$$\leq \lim_{\lambda\downarrow 0} \frac{f(x + \lambda y) - f(x) + f(x + \lambda z) - f(x)}{\lambda}$$

$$= f_+'(x,y) + f_+'(x,z)$$ Hence (iv).

Q.E.D.

LECTURE 5

Sub-differential Theory of Convex Functions

5.1 Introduction:

In different applications we see that supporting hyperplanes to convex sets can be employed in situations where tangent hyperplanes, in the sense of the classical theory of smooth surfaces, do not exist. Similarly, sub-gradients of convex functions, which correspond to supporting hyperplanes to epigraphs rather than tangent hyperplanes to graph, are often useful where ordinary gradients do not exist.

The theory of sub-differentiation of convex functions is a fundamental tools in the analysis of extremum problems, For some elementary applications, see next lecture.

Notations

L : a topological vector space

L^*: dual space of L; i.e. the set of all continuous linear functionals.

$\langle x, x^* \rangle = x^*(x)$, $x^* \in L^*$, $x \in L$.

Definition 5.1: Let $f : L \to \mathbb{R} \cup \{\infty\}$ be a proper convex function. A subgradient of f at $x_o \in L$ is an $x^* \in L^*$ s.t.

$$f(y) \geq f(x_o) + \langle y - x_o, x_o^* \rangle \quad \forall\, y \in L.$$

This says that $f(x_o)$ is finite and that the graph of the affine function $A(y) = f(x_o) + \langle y-x_o, x_o^* \rangle$ is a 'non-vertical' supporting hyperplane at $(x_o, f(x_o))$. Or, in otherwords, $A(y)$ is a support of f at x_o. Figure below may help to visualise it:

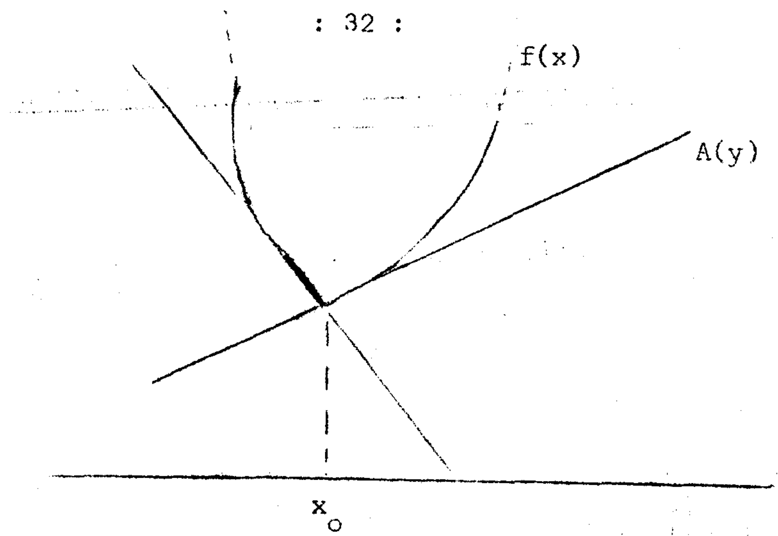


figure 5.1

Definition 5.2: Let $\partial f(x) = \{x^* \varepsilon L^* : x^*$ is a subgradient of f at $x \varepsilon L\}$. If $\partial f(x) \neq \phi$ then f is said to be sub-differentiable at x.

The sub-differential of f is the multivalued mapping (i.e., a relation) which assigns the set $\partial f(x)$ to each x. i.e. $\partial f: L \to L^*$ is a multiple valued mapping. We shall call this mapping as subdifferential (or subgradient) operator of f.

Definition 5.3: If the above subgradient operator is single valued then it will be called the gradient operator of f and $\partial f(x)$ will be called the gradient of f at x.

Remark 5.4: (1) If L is a Banach space, then the notion of gradient is equivalent to that of the Fréchet derivative.

(2) Note that $\partial f(x)$ is weak* - closed convex set in L^*.

Remark 5.5: The sub-differential at x is also sometimes called general differential at x. The advantage of the present treatment is that L need not be a Banach space, and f need not be Fréchet differential. E.g. $f(x) = ||x||$, x ε is convex but not Fréchet differentiable in many of the Banach spaces.

We shall investigate here the following problems:

(1) When $\partial f(x)$ exists, i.e. $\neq \phi$?

(2) What is the relation of $\partial f(x)$ with directional and with Fréchet derivatives in NLS.

(3) What are the properties of the sub-differential mapping?

Also, in this connection, we shall investigate the problems of characterisation of sub-differential mappings as some special types of "binary relations" in L x L*. Or, in otherwords, necessary and sufficient conditions will be sought in order that a multi-valued mapping will be the sub-differential mapping of a proper convex function.

5.2 Properties and a characterisation of the Sub-differential map.

Definition 5.6: A convex function f is said to be closed if for each $\alpha \in \mathbb{R}$, $\{x \in L : f(x) \leq \alpha\}$ is closed.

The following is a geometrical investigation for convex functions of one real variable. The following are sketches of three convex functions and their sub-dffferentials (identified as the slopes of the tangent lines)

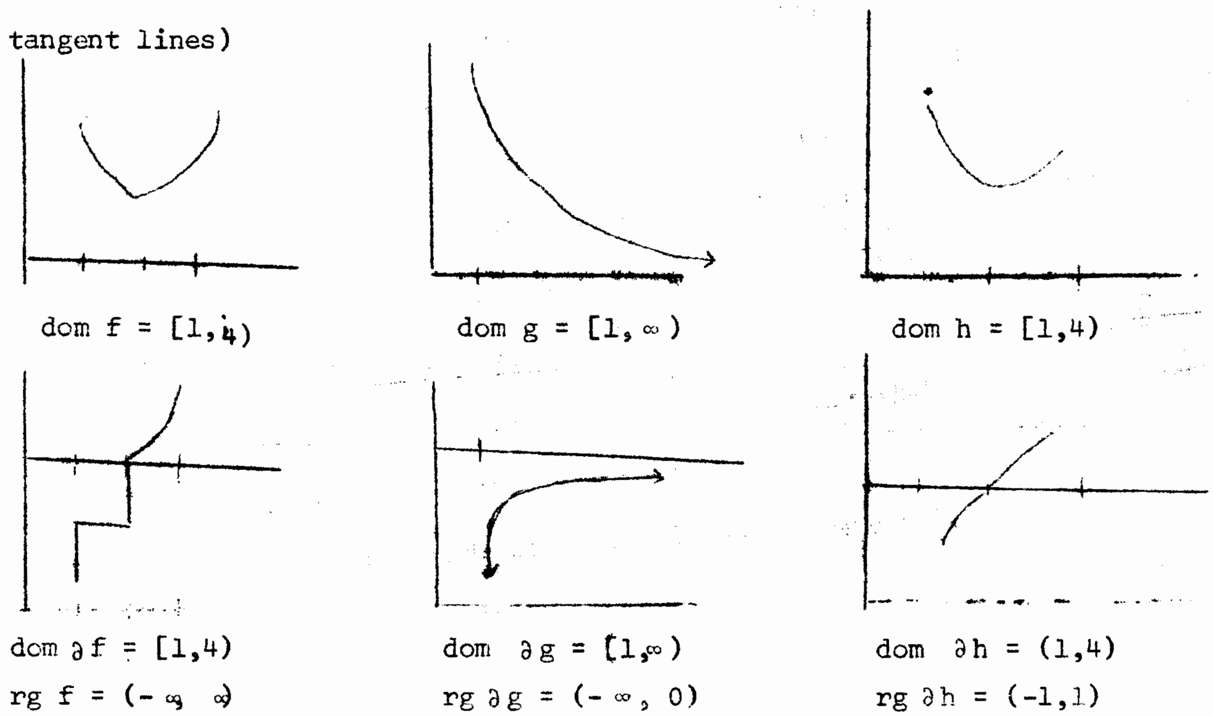


Figure 5.2

Note that f and g are closed whereas h is not. Intuitively one can see that the property of closedness of a convex function has something to do with its behaviour at the boundary of its domain. (In fact, a convex function f on I is closed <==> f is continuous at each end point, where I is an interval of the real line). About sub-differential mappings, note that in each case -the graph of the sub-differential mapping is a continuous increasing curve with verticle as well as horizontal segments. For ∂f and ∂g, both ends of the curve recede upto infinity whereas, for ∂h it is not. This phenomenon of convex functions is related to the closedness of the functions. To make it possible to investigate above types of phenomenon applicable to more general spaces. Let us proceed as follows:

Definition 5.6: A relation ρ on $L \times L^*$ is said to be monotone if

$$\langle y-x, y^* - x^* \rangle \geq 0$$

whenever $x^* \in \rho(x)$ and $y^* \in \rho(y)$; and it is said to be maximal monotone relation if its graph is not properly contained in the graph of any other monotone relation.

Remark 5.8: If L is a Banach space, then ∂f of each lower semi continuous proper convex function f on L is a monotone relation, even it is maximal. But not every monotone relation arises from a convex function. (In this connection, see theorem 4.29 lecture 4). For instance, every positive definite linear mapping ρ on a real Hilbert space is a (single valued) monotone relation. But, such a mapping is the sub-differential of a proper convex function if and only if it is also self adjoint. Then natural question arises: What are the properties that characterise the relations as the sub-differentials of some proper convex functions? Or, in otherwords, for which relations ρ, we have $\rho = \partial f$ for some proper convex function f. To that end, Definition 5.9: A relation ρ on $L \times L^*$ is said to be monotone of degree n if it satisfies

$$0 \geq \langle x_o - x_n, x_n^* \rangle + \ldots + \langle x_2 - x_1, x_1^* \rangle + \langle x_1 - x_o, x_o^* \rangle$$

for every set of n+1 pairs (x_i, x_i^*), $x_i \in L$, $x_i^* \in \rho(x_i)$, $\forall i = 0,1,\ldots, n$; ρ is said to be cyclically monotone if it is monotone of degree n, for all $n > 0$. Note that ρ is monotone if and only if ρ is monotone of degree 1 In fact, ρ is cyclically monotone implies ρ is monotone whereas ρ is monotone $=\not=> \rho$ is cyclically monotone unless L is an one-dimensional space. This leads one to the following open problem.

Conjecture 5.10: In the case of $L = \mathbb{R}^n$ each relation on $L \times L^*$ which is monotone of degree n is actually monotone of all degrees i.e, cyclically monotone.

The following theorem gives the necessary and sufficient condition for imbedding a relation into the sub-differential mapping of some proper convex functions.

Theorem 5.11: Let L be a topological vector space, and ρ be a relation on $L \times L^*$. Then $\rho \subset \partial f$ for some proper convex function f on L <===> ρ is cyclically monotone.

Proof: [===>] part follows from the definition

[<===] part : w.l.g. suppose $\rho \neq \phi$. Fix some $x_o \in L$ and $x_o^* \in L^*$ with $x_o^* \in \rho(x_o)$. Define, for each $x \in L$,

$$f(x) = \sup \{\langle x - x_n, x_n^* \rangle + \ldots + \langle x_1 - x_o, x_o^* \rangle\}$$

where $x_i^* \in \rho(x_i)$ for $i = 1,2,\ldots,n$, and the sup is taken over all possible finite sets of such pairs (x_i, x_i^*).

Claim 1: f is proper convex with $\partial f \supset \rho$. To prove this first note that $\langle x-x_n, x_n^*\rangle + \ldots + \langle x_1-x_o, x_o^*\rangle$ is an affine function in x for all such finite pairs, so $f(x) > -\infty \ \forall x \in L$. To show that f is not identically equal to ∞, note that ρ is cyclical $\Longrightarrow f(x_o) \leq 0$. In fact, $f(x_o) = 0$. Convexity of f follows from the sub-additivity property of 'sup' operation.

To show $\partial f \supset \rho$: Choose $\bar{x}$ and $\bar{x}^*$ with $\bar{x}^* \in \rho(\bar{x})$, arbitrarily and fix it. We want to show $\bar{x}^* \in \partial f(\bar{x})$. Given $\varepsilon > 0$, by definition of f it follows that there exist a finite number of pairs (x_i, x_i^*) with $x_i^* \in \rho(x_i)$, $\forall i = 0,1,2,\ldots, k$ s.t.

$$\langle \bar{x} - x_k, x_k^*\rangle + \ldots + \langle x_1-x_o, x_o^*\rangle \geq f(\bar{x}) - \varepsilon$$

Set $x_{k+1} = \bar{x}$ and $x_{k+1}^* = \bar{x}^*$. Then by definition of f, we have for each $x \in L$,

$$f(x) \geq \langle x-x_{k+1}, x_{k+1}^*\rangle + \langle x_{k+1}-x_k, x_k^*\rangle + \ldots + \langle x_1-x_o, x_o^*\rangle$$

$$\geq \langle x-x_{k+1}, x_{k+1}^*\rangle + f(x_{k+1}) - \varepsilon.$$

Now $\varepsilon > 0$ is arbitrary. Hence

$$f(x) \geq \langle x-\bar{x}, \bar{x}^*\rangle + f(\bar{x}). \text{ So } \bar{x}^* \in \partial f(\bar{x}).$$

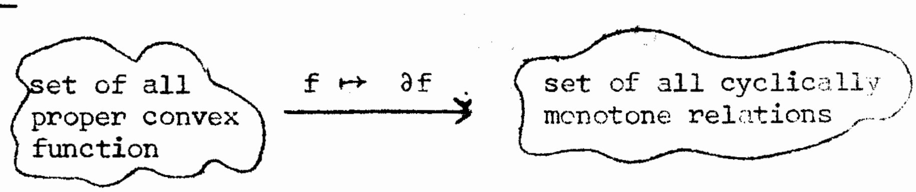
Q.E.D.

Corollary 5.12: If ρ is a maximal cyclically monotone relation on $L \times L^*$, then $\rho = \partial f$ for some proper convex function f on L.

Proof: Trivial.

Corollary 5.13: If f_1 is any proper convex function on L, then there exists a proper function f_2 on L s.t. $\partial f_1 \subset \partial f_2$ and ∂f_2 is a maximal cyclically monotone relation.

Remark 5.14: So what we have established is



that ∂f is a mapping from the set of all proper convex functions into the set of all cyclically monotone relations on L x L*. Now, the natural questions arise: When is such a mapping 1-1 and onto? We are not aware of any general result. However, if we put some restrictions on the domain of the sub-differential operator ∂ and also on the underlying space L and then we can have an 1-1 and onto ∂ map-summarised in the theorem below. Let us define a relation $\sim$ on the set of all l.s.c. proper convex function by $f(x) \sim g(x)$ if $f(x)-g(x) = c$, some constant. Note that $\sim$ is an equivalence relation. Denote by

$$\mathcal{F} = \{\text{all l.s.c. proper convex function on } L\} / \sim$$

and by

$$\mathcal{R} = \text{set of all cyclically maximal monotone relations on } L \times L^*.$$

Theorem 5.15: Let L be a Banach space. Then the sub-differential mapping $\partial : \mathcal{F} \to \mathcal{R}$ is an one-one and onto map.

Proof: See Rockafellar (1970) for the corrected version of his previous proof appeared in Pac. J. Math (1966).

Remark 5.16: What about algebraic and topological properties of the operator ∂ ? Some are discussed in the later sections. But, one should investigate for more possibilities.

5.3 Sub-gradients and Directional Derivatives:

Set of conjugate functions are some sort of dual space to the space of convex functions.

Definition 5.17: The conjugate to or Young - Fenchel transform of f on L is the function f* on L* defined by, for each $x^* \in L^*$

$$f^*(x^*) = \sup \{\langle x, x^* \rangle - f(x) : x \in L\}$$

<u>Example 5.18</u>: Let $A \subset L$ and $f(x) = \delta(x|A)$, the indicator function of A. i.e., $\delta(x|A) = 0$ if $x \in A$

$= \infty$ if $x \in A$

Then

$$f^*(x^*) = \sup \{\langle x,x^*\rangle : x \in A\} \overset{def}{=} s(x^*|A),$$ called the support function of the set A.

<u>Exercise 5.19:</u> Let A be a convex subset of L and let $\mu_A(x)$ be the Minkowski functional of the set A defined by

$$\mu_A(x) = 0 \quad \text{if } x = 0$$
$$= \inf \{\alpha > 0 | \alpha^{-1} x \in A\} \quad \text{if } x \neq 0.$$

Show that $f^*(x^*) = \delta(x^*|A^o)$, where A^o is the polar of the set A defined by

$$A^o = \{x^* \in L^* : s(x^*|A) \leq 1\} .$$

where $s(x^*|A)$ is the support function of the set A as defined in the previous example.

We do not have enough space here to present many facets of the conjugate function of a function. However, we shall state an important theorem about conjugate functions without proof. Interested readers should refer to A.D.Loffe and Tihomirov (1961), 171-180. Let $f^{**} = (f^*)^*$, called second conjugate of f.

<u>Theorem 5.20:</u> (Fenchel - Moreau Theorem): Let f be a function on L and $f(x) > -\infty \ \forall x \in L$. Then $f = f^{**} \Longleftrightarrow f$ is closed and convex.

<u>Proof:</u> See A.Loffe and V.Tihomirov (1961) pp. 175.

<u>Theorem 5.21</u>: Let $C^* \subset L^*$ be a weak* -closed convex set. Let, $\forall y \in L$, $\sigma(C^*,y) = \sup \{\langle y,x^*\rangle : x^* \in C^*\}$, the support function of C^* on L. Then $\sigma(C^*,y)$ is a positively homogeneous l.s.c. proper convex function on L (By positively homogeneous, we mean $\sigma(C^*,\lambda y) = \lambda\, \sigma(c^*,y) \ \forall \lambda > 0$ and $y \in L$.)

<u>Proof:</u> Left as an exercise to the reader.

The following theorem gives the different relationships among some of the concepts we have already introduced; in particular, relationship between directional derivatives and generalised derivatives.

Theorem 5.22: Let $x \in \text{dom } f$ and $x^* \in L^*$. Then

(i) $x^* \in \partial f(x) \iff f'_+(x;y) \geq \langle y, x^* \rangle \quad \forall y \in L,$

(ii) $f'_+(x;y) \geq \sigma(\partial f(x), y) \quad \forall y \in L$ and $x \in L$ s.t. $\partial f(x) \neq \phi.$

(iii) $\sigma(\partial f(x), y) = \liminf_{z \to y} f'_+(x;z) \; \forall x \in L$ when $\partial f(x) \neq \phi.$

Proof: (i) $x^* \in \partial f(x) \iff f(y) - f(x) \geq \langle y-x, x^* \rangle \quad \forall y \in L$

$$\iff \frac{f(x+\lambda z) - f(x)}{\lambda} \geq \langle z, x^* \rangle \; \forall z \in L \text{ and } \lambda > 0$$

obtained by setting $y = x + \lambda z$.

$$\iff \lim_{\lambda \downarrow 0} \frac{f(x+\lambda z) - f(x)}{\lambda} \geq \langle z, x^* \rangle \quad \forall z \in L$$

(ii) follows from (i)

(iii) using the theory of conjugate -functions, it can be proved. It is avoided here.

The following theorem 5.24 gives the limiting relationship between approximate subgradients and directional derivatives.

Definition 5.23: For given $\varepsilon > 0$, we define an approximate subgradient relation by

$$\partial_\varepsilon f(x) = \{x^* \in L^* : f(y) \geq f(x) - \varepsilon + \langle y-x, x^* \rangle \quad \forall y \in L\}.$$

Note that $\partial_\varepsilon f(x)$ is a weak*-closed convex subset of L^* for every $\varepsilon > 0$; and $\partial_\varepsilon f(x) \downarrow \partial f(x)$ as $\varepsilon \downarrow 0$.

Theorem 5.24: Let f be a l.s.c. proper convex function on L and let $x \in \text{dom } f$. Then

$$\sigma(\partial_\varepsilon f(x), y) \downarrow f'_+(x;y) \text{ as } \varepsilon \downarrow 0 \quad \forall y \in L.$$

Proof: Let $C^*_\beta = \{x^* \varepsilon L^* : f^*(x^*) - \langle x,x^*\rangle \leq \beta\}$ where f^* is the conjugate function of f and β be such that

$$\infty > \beta > \inf \{f^*(x^*) - \langle x,x^*\rangle \mid x^* \varepsilon L^*\} > -\infty ,$$

be a nonempty weak*-closed convex set in L*. Then the support function of C* is

$$\sigma(C^*,y) = \inf_{\lambda > 0} \frac{f(x + \lambda y) + \beta}{\lambda} \quad \text{(See Rockafellar (8)).}$$

Now, by the definition of $\partial_\varepsilon f$ and f^*, it follows that

$$x^* \varepsilon \partial_\varepsilon f(x) \iff f(x) - \varepsilon - \langle x,x^*\rangle \leq \inf \{f(y) - \langle y,x^*\rangle \mid y \varepsilon L\} = -f^*(x^*)$$

Set $\beta = \varepsilon - f(x)$ then note that $C^*_\beta = \partial_\varepsilon f(x)$, since

$$f(x) = \sup \{ \langle x,x^*\rangle - f^*(x^*) : x^* \varepsilon L^*\} .$$

Hence,

$$\sigma(\partial_\varepsilon f(x), y) = \inf_{\lambda>0} \frac{f(x+ \lambda y) - f(x) + \varepsilon}{\lambda} \qquad \forall \ \varepsilon > 0.$$

Now, the facts $\frac{f(x+\lambda y) - f(x)}{\lambda} \downarrow$ as $\lambda \downarrow 0$ and above is true for all $\varepsilon > 0$ imply that

$$\sigma(\partial_\varepsilon f(x),y) \downarrow \inf_{\lambda>0} \frac{f(x+\lambda y) - f(x)}{\lambda} = f'_+(x,y)$$

Q.E.D

The following theorem reveals some of the duality features of f and f* in connection with the generalised derivatives.

Theorem 5.25: Let f be a closed convex function on L. Then the following are equivalent:

(i) $x^* \varepsilon \partial f(x)$

(ii) $x \ \varepsilon \ \partial f^*(x^*)$

(iii) $f(x) + f^*(x^*) = \langle x,x^*\rangle$

and similarly, the following are also equivalent (i') $x^* \in \partial_\varepsilon f(x)$

(ii') $x \in \partial_\varepsilon f^*(x^*)$

(iii') $f(x) + f^*(x^*) \leq \langle x,x^*\rangle + \varepsilon$

Proof: (i) <===> (ii) follows from the fact that $f^{**} = f$ which, in turn, is followed by Fenchel-Moreau theorem (Theorem 5.28). Now we shall show that

[(i) ===> (iii)]: Let $x^* \in \partial f(x)$. Then by definition of sub-gradient, we have

$$f(y) \geq f(x) + \langle y-x, x^*\rangle \quad \forall y \in L$$

i.e. $$\langle y,x^*\rangle - f(y) \leq \langle x,x^*\rangle - f(x) \quad \forall y \in L.$$

So $f^*(x^*) = \sup \{ \langle y,x^*\rangle - f(y): y \in L\} \leq \langle x,x^*\rangle - f(x)$

i.e. $$f(x) + f^*(x^*) \leq \langle x,x^*\rangle .$$

On the otherhand, $f^*(x^*) = \sup \{\langle x,x^*\rangle - f(x) \mid x \in L\}$

$$\geq \langle x,x^*\rangle - f(x).$$

These two inequalities give (iii).

[(iii) ===> (ii)]: Suppose $f^*(x^*) + f(x) = \langle x,x^*\rangle$.

Now, since, $f(x+\varepsilon z) \geq \langle x+\varepsilon z,x^*\rangle - f^*(x^*)$, we have

$$\frac{f(x+\varepsilon z) - f(x)}{\varepsilon} \geq \frac{\langle x+\varepsilon z, x^*\rangle - f^*(x^*)-f(x)}{\varepsilon} = \frac{\langle \varepsilon z,x^*\rangle}{\varepsilon} = \langle z,x^*\rangle$$

$\forall \ \varepsilon > 0$ and $z \in L$.

Hence $f'_+(x;z) \geq \langle z,x^*\rangle$ ====> $x^* \in \partial f(x)$ by Theorem 5.22(i).

Q.E.D.

Now we shall state a few theorems without proof. The proofs can be found in (1982), pp. 199-204. These results are about the sub-differentials of functions – obtained after some operations on some initial convex functions – – interms of the sub-differentials of the original functions.

5.4 Existence of sub-differentials

The following theorems give the conditions on the functions so that their sub-differentials exist.

Theorem 5.29: A proper convex function f is sub-differentiable at a point $x \in \text{dom } f \iff f'_+(x,z)$ is l.s.c. at $z = 0$.

Proof: See (19b), pp. 198-199.

Theorem 5.30) If $f: L \to \mathbb{R} \cup \{\infty\}$ is a proper convex function and it is continuous at x_o then $\partial f(x_o) \neq \phi$.

Proof: It is a corollary to a general theorem due to Minty, See (14), pp.244.

References

1. Naimark (1972): "Normed Algebras" , Wolters Noordhoff Pub. Groningen, the Netherlands.

2. R.T.Rockafellar (1972): "Convex Analysis", Princeton University Press.

3. A.W.Roberts and D.E.Varberg (1973): "Convex functions", Aca. Press.

4. B.Friedman (1940): "A note on convex functions", Bull. Amer. Math. Soc. 46, 473-474.

5. K.R.Parthasarathy, and R.Bhatia (1978): "Lectures on Functional Analysis", Macmillan, New Delhi.

6. R.T.Rockafellar (1966): "Pac. J. Math.

7. ________________(1970): Pac. J. Math.

8. ________________(1973): "Level sets and Continuity of Conjugate convex functions" - Trans. Amer. Math. Soc.

9. Moreau (1966): "Fonctionelles Convexes" lecture notes, seminare Equations aux derivee partielle" College de France.

10. G.Kothe (1969): "Topological vector spaces" Springer - Verlag.

11. V.L.Klee (1969): Separation and support properties of convex sets - a survey. In "Control Theory and the Calculus of variations (A.V.Balakrishnan edited) 235-305, Acad. Press, ; New York.

12. A.Brønsted (1966): "Milman's Theorem for convex functions", Math. Scand. 19, 5-10.

13. Ponstein (1967): "Seven kinds of convexity" - SIAM REV. 9, 115-119.

14. Minty (1964): "On the monotonicity of Gradient of a convex function" Pac. J. Math. Vol. 14, No. 1, 243-247.

15. J.L.Kelley and I.Namioka (1963): "Linear topological spaces", Von-Nostrand.

16. (a) V.L.Klee (1951):"Convex sets in linear spaces", Duke Math. J. 18, 18, 443-446.

    (b) _______ '(1951): "Convex sets in Linear spaces II", Duke Math. J. 18, 875-883.

    (c) ________(1953): "Convex sets in linear spaces III", Duke Math. j. 20, 105-112.

    (d) ________(1956): "Strict separation of convex sets", Pro. Amer. Math. Soc. 7, 735-737.

    (e) ________(1968): "Maximal separation theorems for convex sets", Tran. Amer. Math. Soc. 134, 133-148.

17. W.Fenchel (1951): "Convex sets, cones and functions", memeographed lecture notes, Princeton University.

18. E.Asplund (1969): "Topics in the theory of convex functions" in theory and applications of monotone operators (Tipographia Ordrisi" Gubbio).

19. (a) A.D.Ioffe and V.M.Tihomirov (1968):"The duality of convex functions and extremal problems" Russ. Math. Survey, 23, 53-124.

__________(1979): "Theory of extrmal problems", North Holland Pub. CO.

20. I.Ekeland and R.Teman (1977):"Convex Analysis and variational problems" North Holland, Amsterdam.

21. A.Brønsted, and R.Rockafeller (1965):"On the sub-differentiability of convex functions" Bull. Amer. Math. Soc. 16, 605-611.

22. __________(1964): "Conjugate convex functions in topological vector spaces", Math. Fys. Medd. Dansk. Vid. Selsk. 34.

23. W.Fenchel (1949):"On conjugate convex functions", Cand. J. Math. 1, 73-77.

24. S.Kutateladze and A.Rubinov (1973):"Minkwaski duality and its applications" Russ. Math. Surveys, 27, 137-191.

25. E.G.Golshtein (1967): "Problems of best approximations by elements of a convex set; and some properties of support functionals", Soviet. Math. Dokl. 8, 504-507.

26. V.L.Levin (1969):"Application of E. Helleys theorem to convex programming, problems of best ;approximation and related questions" Math. USSR - Sb 8, 235-247.

27. R.R.Phelps (1978): "Differentiability of convex functions on Banach spaces" Lecture notes Univer. College, London.

28. Larman and Phelps (1979): "Gateaux differentiability of convex functions on Banach spaces" J. Lond. Math. Soc. 20, pt. 1, 115-127.