1

School of Graduate Studies
Addis Ababa University
Department of Mathematics

A Graduate Seminar Report ON Subdifferential of Finite Convex Function

By
Birilew Belayneh

Advisor
Prof. Dr.rer.nat.habil.R.Deumlich

Addis Ababa
June 2002

SEMINAR REPORT

ON

SUBDIFFERENTIAL OF FINITE CONVEX FUNCTIONS

Contents

Preface

This paper is a report of two seminars: namely Seminar I (Math 500) and Seminar II (Math 600), which are conducted in the first and second semesters respectively. It comprises five chapters that organized in order of dependence of one on the other. The chapters in this seminar report are divided into sections. Each section concludes with a set of definitions, propositions, lemmas, theorems and remarks that pertains to the topics of that section. After the first, the next two chapters are examined in the former while the last two chapters discussed in the latter one.

This seminar report is mainly focus on the generalization of the concept of derivative in general and subdifferentials in particular. Thus the subdifferentiation introduced to generalize the ordinary differentiation, one should therefore not be surprised to find counterparts of most of the results encountered in differential calculus such that first-order approximation, Mean-Value Theorem, Calculus rules, etc.

The first chapter provides basic definitions, theorems, and properties on convex sets and convex functions in $\mathbb{R}^n$ as well as some examples, which are essentially needed later for a better understanding of the following chapters.

In general speaking Chapter 2, Chapter 3 and Chapter 4 introduce the behavior of subdifferential of convex functions at a given fixed point, on the other hand the last chapter emphasis on its behavior with varying at a given point as well as the function itself. These behaviors\ properties of subdifferentials developed using the powerful apparatus of support functions, which is of the key ingredients to this paper.

Some of the important calculus rules are introduced in detail in the fourth chapter. Some operations on convex functions destroy differentiability and thereby find no place in differential calculus but preserves convexity. An important example is the maximum-operation.

First of all, I thank the almighty God, with the help of WHOM this seminar report took its present form.

I would like to express my heartfelt gratitude to my advisor and teacher, Prof. Dr. rer. nat. habil. R. Deumlich not only for his genuine advice, constructive comments, invaluable suggestions and provision of materials (reference books and diskettes) in preparing this seminar report but also in other social aspects showing me the bright future.

I also extend my deep appreciation to the Department of Mathematics for its material support and for giving me the chance to type the manuscript of this seminar myself.

Once again I will take the opportunity to address my thanks to all my previous teachers and friends who made fruitful my study and wish my progress. Among the many friends, Yibeltal Yitayew deserves worth mentioning. Of course a few words cannot adequately express the thanks I owe for their unreserved moral and material support.

Birilew Belayneh

1 Some Basic Definitions and Properties of Convex Functions

Here we state some results from the theory of convex functions and sets, which are important for the study of the next chapters. This chapter provides, on one hand, sufficient background for us to know basic properties of convex functions, in order to apply them to prove other propositions. On the other hand, it serves as an introduction to this seminar report.

1.1 Definitions and Some Results of Convex Functions

Definition 1.1.1: Let $S \subseteq \mathbb{R}^n$ be a nonempty set.

a) A set S is said to be *convex,* if the relation $x, y \in S, \lambda \geq 0$ implies $\lambda x + (1-\lambda)y \in S$. An element $y = \sum_{i=1}^{m} \lambda_i x_i$ where $\lambda_i \geq 0$ for all $i \in \{1,2,\cdots,m\}, \sum_{i=1}^{m} \lambda_i = 1$ is called a *convex combination* of elements $x_1, x_2, \cdots, x_m$. It is clear that a convex set contains any convex combination of its elements.

b) The intersection of all convex sets containing S is called the *convex hull* of S and is denoted by the symbol $convS$. In other words, $convS$ is the set of all convex combinations of elements of S. *i.e*

$$convS = \bigcap \{C : S \subseteq C \subseteq \mathbb{R}^n, \text{ C is convex}\}$$
$$= \{x \in \mathbb{R}^n : \exists m \in N, \exists \lambda_i \geq 0, \exists x_i \in S \text{ such that } x = \sum_{i=1}^{m} \lambda_i x_i, \sum_{i=1}^{m} \lambda_i = 1, \forall i \in \{1,2,\cdots m\}\}$$

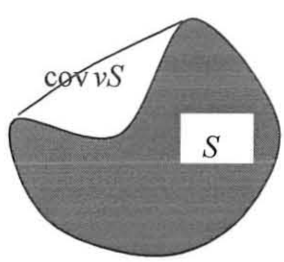

Figure 1.1.1 (convex hull of a set)

c) The *closed convex hull* of S is the intersection of all closed convex sets containing S, and is denoted by $\overline{convS}$.

Proposition 1.1.1: The closed convex hull of a nonempty set $S \subseteq \mathbb{R}^n$ is the closure of the convex hull of S. *i.e*

$$\overline{convS} = cl(convS).$$

Proof: Because $cl(convS)$ is a closed convex set containing S, it contains $\overline{convS}$ as well. On the other hand, take a closed convex set $C \subseteq \mathbb{R}^n$ containing S, being convex, C contains $convS$; being closed, it contains also the closure of $convS$.

Since C was arbitrary, we conclude that

$$\bigcap\{C : C \in \mathcal{C}\} \supseteq cl(convS)$$

where $\mathcal{C}$ is the collection of all closed convex sets containing S.
Therefore,

$$\overline{convS} = cl(convS). \ //$$

Remark 1.1.1: From Definition 1.1.1 c), the operation "*taking a hull*" is monotone. *i.e* if $S_1 \subseteq S_2$, then $convS_1 \subseteq convS_2$ and of course $\overline{concS_1} \subseteq \overline{convS_2}$. For a closed set S, $convS$ need not be closed. For instance, consider the set

$$S = \{(0,0)\} \bigcup \{(x,1) : x \geq 0, x \in \mathbb{R}\}$$

is closed but $convS$ is not closed because it misses the half line $(\mathbb{R}^+, 0) = \{(x,0) : x > 0\}$. Thus $cl(convS)$ is not necessarily equal to $conv(clS)$. This phenomenon can occur only when S is unbounded.

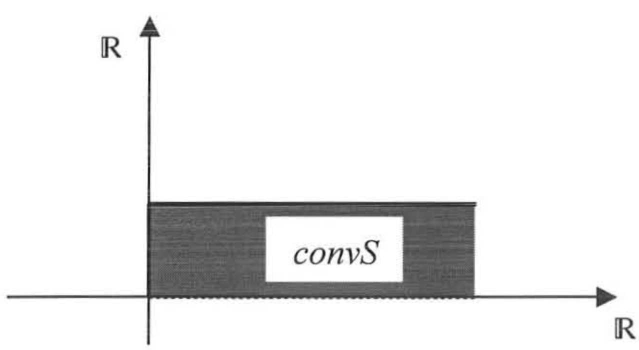

Figure 1.1.2 (closed and unbounded set)

In general, for nonempty set $S \subseteq \mathbb{R}^n$ the operations "*taking the closure*" and "*taking the convex hull*" do not commute, but these operations commute when S is bounded. *i.e* S is bounded in $\mathbb{R}^n$ implies

$$\overline{convS} = cl(convs) = conv(clS).$$

Theorem 1.1.1: Let I be an arbitrary index set and let $S, S_i \subseteq \mathbb{R}^n$ $i \in I$ be convex sets. Furthermore, let A: $\mathbb{R}^n \to \mathbb{R}$ be a linear transformation. Then

a) $\bigcap_{i \in I} S_i$ is a convex set;

b) $S_1 + S_2 = \{x \in \mathbb{R}^n : x = s_1 + s_2, s_1 \in S_1, s_2 \in S_2\}$ is a convex set;

c) clS and $\operatorname{int} S$ are convex sets;

d) $S_1 \times S_2$ is a convex set;

e) $A(S_i) := \{x \in \mathbb{R} : x = A(s_i), \text{for some } s_i \in S_i, i \in I\}$ is a convex set.

For the proof of these properties of convex sets apply Definition 1.1.1 a).

Definition 1.1.2: A function *f* defined on a nonempty convex set $C \subseteq \mathbb{R}^n$ is called *convex* on *C* when for all points $x, y \in C$ and all $\lambda \in [0,1]$ there hold:

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y). \tag{1.1.1}$$

We say that *f* is *strictly convex* on *C* when (1.1.1) holds as a strict inequality if $x \ne y$ and $\lambda \in (0.1)$. A function *g* is said to be *concave* if $-g$ is a *convex* function.

From this definition, the following inequality (called *Jensen's inequality*) holds:

$$f(\sum_{i=1}^{m} \lambda_i x_i) \le \sum_{i=1}^{m} \lambda_i f(x_i), \textit{ for } \text{all } x \in C, \textit{ for } \text{all } \lambda_i \ge 0, \text{and} \sum_{i=1}^{m} \lambda_i = 1.$$

To illustrate the geometric meaning of convexity consider the following in $\mathbb{R}^2$.

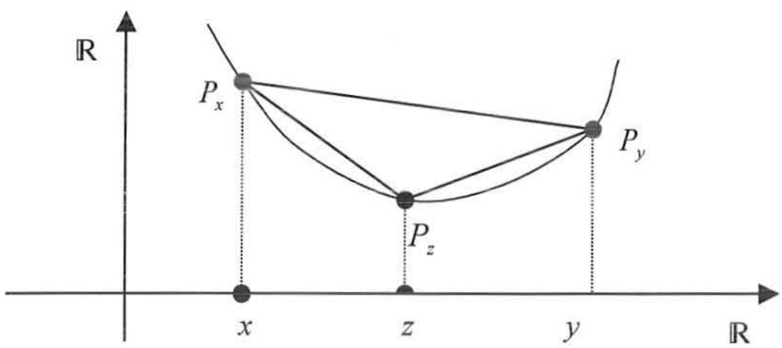

Figure 1.1.3 (convex function)

Let the line segment $P_x P_y$ joining the point $P_x = (x, f(x))$ to the point $P_y = (y, f(y))$. To say that *f* is convex is to say that, for all $x, y \in C$ and all z such that $z = \lambda x + (1-\lambda)y$ for $\lambda > 0$, the point $P_z = (z, f(z))$ lies below the segment $P_x P_y$.

Definition 1.1.3: Given a function *f* defined on $C \subseteq \mathbb{R}^n$ and taking its values in the set $\mathbb{R} \bigcup \{-\infty, +\infty\}$. Then the sets

$$domf := \{x \in C : f(x) < \infty\},$$
$$grf := \{(x,r) \in C \times \mathbb{R} : f(x) = r\},$$
$$epif := \{(x,r) \in C \times \mathbb{R} : f(x) \le r\},$$
$$hypof := \{(x,r) \in C \times \mathbb{R} : f(x) \ge r\},$$
$$S_r(f) := \{x \in C : f(x) \le r\}$$

are called the *domain, graph, epigraph, hypo graph* and the *sublevel-sets* of this function respectively.

Clearly from the above definition we have

$$(epif) \bigcap (hypof) = grf.$$

Theorem 1.1.2: A function *f* is convex if and only if its epigraph is a convex set.

Proof:($\Rightarrow$): Let (x,r), $(y,s) \in$ epif and $\lambda \in [0,1]$. Then by the convexity of f and by $f(x) \le r, f(y) \le s$ we have

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) \le \lambda r + (1-\lambda)s,$$

i.e

$$(\lambda x + (1-\lambda)y, \lambda r + (1-\lambda)s) = \lambda(x,r) + (1-\lambda)(y,s) \in epif .$$

Therefore *epif* is a convex set.

($\Leftarrow$): Suppose the set *epif* is convex. Then by Jensen's inequality (in particular for $m = 2$) we have $f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y)$ for each $x, y \in epif$ and $\lambda \ge 0$. Therefore *f* is convex. //

Theorem 1.1.3: Let $X \subseteq \mathbb{R}^n$ be a convex set and let $f: X \to \mathbb{R}$ be a function. Then

a) *f* is convex if and only if the function $\varphi_{xy} : [0.1] \to \mathbb{R}$ defined by

$$\varphi_{xy}(t) := f(tx + (1-t)y) \tag{1.1.2}$$

is convex for all $x, y \in X$

b) The function $\varphi : \mathbb{R}_+ \to \mathbb{R}$ defined by

$$\varphi(t) := \frac{f(x+td) - f(x)}{t} \quad \text{for } t \in I_d = \{t \in \mathbb{R}_+ : x + td \in X\},\ d \in \mathbb{R}^n$$

is a *monotone increasing* function whenever *f* is convex.

Proof: a) ($\Rightarrow$): Let *f* be convex. For $x, y \in X$, $t \in [0,1]$ and $\lambda_1, \lambda_2 \in [0,1]$ it follows

$$\begin{aligned}
\varphi_{xy}(t\lambda_1 + (1-t)\lambda_2) &= f([t\lambda_1 + (1-t)\lambda_2]x + [1 - (t\lambda_1 + (1-t)\lambda_2)]y) \\
&= f([t\lambda_1 + (1-t)\lambda_2]x + [(1-t) + t - t\lambda_1 - (1-t)\lambda_2)]y) \\
&= f(t\lambda_1 x + (1-t)\lambda_2 x + (1-t)y + ty - t\lambda_1 y - (1-t)\lambda_2 y) \\
&= f(t[\lambda_1 x + (1-\lambda_1)y] + (1-t)[\lambda_2 x + (1-\lambda_2)y]) \\
&\le tf(\lambda_1 x + (1-\lambda_1)y) + (1-t)f(\lambda_2 x + (1-\lambda_2)y) \\
&= t\varphi_{xy}(\lambda_1) + (1-t)\varphi_{xy}(\lambda_2)
\end{aligned}$$

So we have the convexity of φ_{xy}.

($\Leftarrow$): Let φ_{xy} be convex for all $x, y \in X$ and $t \in [0,1]$. Then we have

$$\begin{aligned}
f(tx + (1-t)y) = \varphi_{xy}(t) &= \varphi_{xy}(t.1 + (1-t).0) \le t\varphi_{xy}(t) + (1-t)\varphi_{xy}(0) \\
&= tf(x) + (1-t)f(y)
\end{aligned}$$

which establishes the convexity of *f*.

b) Consider the function $h(t) := f(x+td) - f(x)$. Then h is convex and $h(0) = 0$. Now let $0 < t_1 \le t_2, t_1, t_2 \in \mathbb{R}_+$. Then

$$h(t_1) = h(\frac{t_1}{t_2}t_2 + \frac{t_2 - t_1}{t_2}.0) \le \frac{t_1}{t_2}h(t_2) + \frac{t_2 - t_1}{t_2}h(0) = \frac{t_1}{t_2}h(t_2)$$

That means

$$\frac{h(t_1)}{t_1}=\frac{f(x+td_1)-f(x)}{t_1}\leq\frac{f(x+t_2d)-f(x)}{t_2}=\frac{h(t_2)}{t_2}$$

i.e $\varphi(t_1)\leq\varphi(t_2)$. //

Definition 1.1.4: Let $S\subseteq\mathbb{R}^n$ and $S\neq\emptyset$

(a) The set

$$H_{d,r}:=\{y\in\mathbb{R}^n:\langle d,y\rangle=r\},\ r\in\mathbb{R},\ d\in\mathbb{R}^n,\ d\neq 0,$$

is said to be a *hyperplane* in $\mathbb{R}^n$.

(b) The hyperplane $H_{d,r}$ is called a *supporting hyperplane* of *S* if and only if

(*i*) *S* is entirely contained in one of the two closed half-spaces:

$$H^+_{d,r}:=\{y\in\mathbb{R}^n:\langle d,y\rangle\geq r\}$$

and

$$H^-_{d,r}:=\{y\in\mathbb{R}^n:\langle d,y\rangle\leq r\}.$$

(*ii*) $S\cap H_{d,r}\neq\emptyset$.

$H_{d,r}$ is said to support *S* at $x\in S$ when $x\in H_{d,r}$, *i.e* $\langle d,x\rangle=r$.

c) The function $\sigma_S:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ defined by

$$\sigma_S(d):=\sup\{\langle s,d\rangle:s\in S\}.$$

is called the *support function* of *S*.

Given $S\subseteq\mathbb{R}^n$ and a support function σ_S of S. For $d\neq 0$ and $d\in dom\sigma_S$ we have

$$S\subseteq\{y\in\mathbb{R}^n:\langle y,d\rangle\leq r\}\ \text{for } r:=\sigma_S(d) \tag{1.1.3}$$

i.e *S* is contained in a closed half- space opposite to *d*.
Consequently when (1.1.3) holds, we can find *r* large enough so that

$$S\subseteq H^-_{d,r}:=\{y\in\mathbb{R}^n:\langle d,y\rangle\leq r\}.$$

Thus decreasing *r* as much as possible while keeping *S* in $H^-_{d,r}$ implies leaning the hyperplane $H_{d,r}$ onto *S*. So the smallest of those *r* is the value $\sigma_S(d)$.

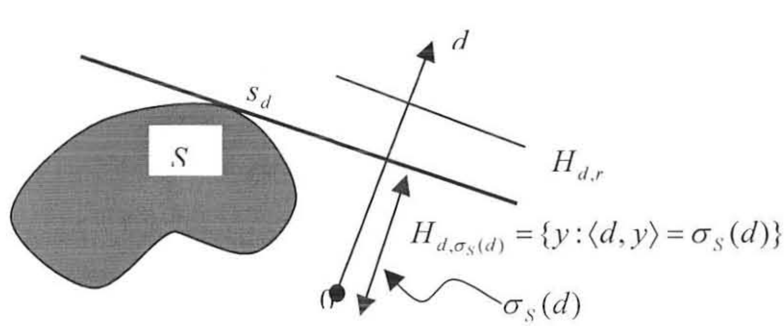

Figure 1.1.4 (supporting hyperplanes and support functions)

Remark 1.1.2: For a compact set S in $\mathbb{R}^n$, there is a closed half-space in which S is contained and the supremum of the linear form $\langle d,\cdot\rangle$ is attained on S, no matter how d is chosen. This means that, there is some (of course boundary point of S) such that s_d belongs to $H_{d,\sigma_S(d)}$. However, if S is not bounded in the oriented direction d, then we have $\sigma_S(d) = +\infty$ and therefore no closed half-space of this type. For instance, consider

$$S := \{(x,0) : x \geq 0\} \subseteq \mathbb{R}^2,$$

and let $d = (1,1)$, then no closed half-space of the form

$$H_{d,\sigma_S(d)} = \{y \in \mathbb{R}^2 : \langle d, y\rangle \leq \sigma_S(d)\}$$

can contain S, even if $\sigma_S(d)$ is increased to $+\infty$.

Definition 1.1.5: A function f: $\mathbb{R}^n \to \mathbb{R} \bigcup \{+\infty\}$ is said to be *sublinear* if it is subadditive:

$$f(x+y) \leq f(x) + f(y) \text{ for all } x, y \in \mathbb{R}^n$$

and positively homogenous *i.e*

$$f(\lambda x) = \lambda f(x) \text{ for all } \lambda \geq 0.$$

It is clear that a sublinear function is convex, and conversely a convex and positively homogenous function is sublinear. A function g is called *superlinear* if $f = -g$ is sublinear.

1.2 Continuity Property of Convex Functions

Definition 1.2.1: For a function f: $\mathbb{R}^n \to \mathbb{R}$ we say that f is *lower semi-continuous* if

$$\liminf_{y \to x} f(y) \geq f(x) \quad \text{for all } x \in \mathbb{R}^n.$$

The function f is said to be *upper semi-continuous* if and only if $-f$ is semi-continuous. Furthermore f is *continuous* if it is both lower and upper semi-continuous.

Proposition 1.2.1: For a function f: $\mathbb{R}^n \to \mathbb{R}$ the following three properties are equivalent:

a) f is lower semi-continuous on $\mathbb{R}^n$;

b) *epif* is a closed set in $\mathbb{R}^n \times \mathbb{R}$;

c) The sublevel-sets $S_r(f)$ for each $r \in \mathbb{R}$ are closed (possibly empty).

Proof:(a)$\Rightarrow$(b): Let $\{(y_k, r_k)\}$ be a sequence in *epif* converging to (x,r) for $k \to +\infty$. Since $f(y_k) \leq r_k$ for all k, by definition of epigraph

$$r = \lim_{k\to\infty} r_k \geq \liminf_{k\to\infty} f(y_k) \geq \liminf_{y \to x} f(y) \geq f(x)$$

This implies $(x,r) \in \text{epif}$. Thus *epif* is closed.

(b) $\Rightarrow$(c): Since for each $r \in \mathbb{R}$ *epif* $\cap \mathbb{R}^n \times \{r\} = \{x \in \mathbb{R}^n : f(x) \leq r, \text{ for } r \in \mathbb{R}\} = S_r(f)$ and the closed sets *epif* and $\mathbb{R}^n \times \{r\}$ have a closed intersection, (c) holds.

(c) $\Rightarrow$(a): Suppose not! *i.e* f is not lower semi-continuous at some x. Then there is a subsequence $\{y_k\}$ converges to x such that $f(y_k)$ converging to $s < f(x)$. Now choose $r \in (s, f(x))$. Then

$$f(y_k) \le r < f(x), \text{ for } k \text{ large enough}$$

This implies $S_r(f)$ contains $\{y_k\}$ but not its limit x. Consequently, this $S_r(f)$ is not closed. This is a contradiction. //

Definition 1.2.2: The function f: $\mathbb{R}^n \to \mathbb{R}$ is called *closed* if and only if one of the following conditions hold:

a) f is lower semi-continuous on $\mathbb{R}^n$;

b) *epif* is a closed set of $\mathbb{R}^n$;

c) The sublevel-sets $S_r(f)$ for each $r \in \mathbb{R}$ are closed.

Definition 1.2.3: Given $S \in \mathbb{R}^n$ and $S \neq \emptyset$, the function $I_S : \mathbb{R}^n \to \mathbb{R} \bigcup \{+\infty\}$ defined by

$$I_S(x) := \begin{cases} 0, & \text{if } x \in S \\ +\infty, & \text{else} \end{cases}$$

is called the *indicator function* of S. Clearly I_S is closed (and convex) if and only if S is closed (and convex). Indeed, $epiI_S = S \times \mathbb{R}_+$ by definition.

Example 1.2.1: Let f be a convex function and $domf = \mathbb{R}$ and let C be a nonempty closed interval. Then the convex restriction of f to C denoted by f_C and defined as

$$f_C(x) := \begin{cases} f(x), & \text{if } x \in C \\ +\infty, & \text{if } x \notin C \end{cases}$$

is closed and convex. Its epigraph is the intersection of *epif* with the vertical strip generated by C.

Example 1.2.2: Let C be a nonempty subset of $\mathbb{R}^n$. The indicator function of C is closed (and convex) if and only if C is closed (and convex). Its sublevel-sets are empty or C.

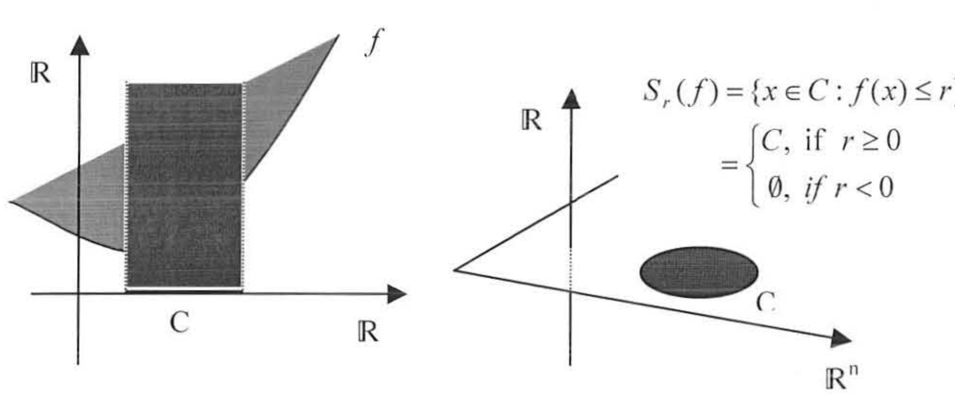

Figure 1.1.5 (closed functions)

Definition 1.2.4: A function f is said to be *Lipschitzian* on a nonempty set $S \subseteq \mathbb{R}^n$, if there is a constant (called Lipschitz constant) L such that

$$|f(x) - f(y)| \le L\|x - y\| \text{ for all } x, y \in S.$$

Lemma 1.2.1: Let f: $\mathbb{R}^n \to \mathbb{R} \bigcup \{+\infty\}$, not identically $+\infty$, be convex and suppose there are $x_0, \delta, m,$ and M such that

$$m \le f(x) \le M \text{ for all } x \in B(x_0, 2\delta)$$

Then f is Lipschitzian on $B(x_0, \delta)$, i.e for all $x, y \in B(x_0, \delta)$ it follows

$$|f(x) - f(y)| \le \frac{M - m}{\delta}\|x - y\| . \qquad (1.2.1)$$

Proof: Let $x, y \in B(x_0, \delta)$ such that $x \ne y$ and take z such that

$$z := y + r\frac{y - x}{\|y - x\|} \in B(x_0, 2\delta);$$

By construction, y lies on the line segment $[x, z]$, namely

$$y := \frac{\|y - x\|}{\delta + \|y - x\|}z + \frac{\delta}{\delta + \|y - x\|}x$$

By convexity of f and using bounded ness of f, we have

$$f(y) - f(x) \le \frac{\|y - x\|}{\delta + \|y - x\|}[f(z) - f((x)] \le \frac{1}{\delta}(M - m)\|y - x\|$$

Then it suffices to exchange the role of x and y to prove (1.2.1). //

Corollary 1.2.2: Let f: $\mathbb{R}^n \to \mathbb{R}$ be convex. Then f is continuous. Furthermore for each $x, y \in \text{int}\, dom f$, there is $L \ge 0$ such that

$$|f(x) - f(y)| \le L\|x - y\| . \qquad (1.2.2)$$

For the proof apply Lemma 1.2.1

The property (1.2.2) is called the Lipschitz continuity of f on $\mathbb{R}^n$. What Corollary 1.2.2 says is that f is locally Lipschitzian on the interior of its domain. Then it follows that the difference quotients $\frac{f(x) - f(y)}{x - y}$ for all $x, y \in \mathbb{R}^n$ are themselves locally bounded. *i.e* bounded on every bounded compact subset of int *domf*.

1.3 Differential Properties of Convex Functions

In this section we will consider the well-known notions of differentiability and we will investigate convex function with regard to these notions of differentiability. Monotonicity of the differences quotient of a convex function at a point x provides convex functions with rather astonishing properties of "*one-sided differentiability*", which allows the introduction of a substitute for the concept of derivative: the "*set of subderivatives*" of a convex function at a point of its domain.

Definition 1.3.1: Let f: $\mathbb{R} \to \mathbb{R} \bigcup \{+\infty\}$ be a convex function not identically $+\infty$. Then we say that f is differentiable at x_0 when the limit

$$Df(x_0) := \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \tag{1.3.1}$$

exists and is finite. $Df(x_0)$ is called the *derivative* of f at x_0.

Correspondingly, we define the *left derivative* and the *right derivative* of f at x_0 as

$$D_{-}f(x_0) := \lim_{x \uparrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = \sup_{x < x_0} \frac{f(x) - f(x_0)}{x - x_0} \tag{1.3.2}$$

and

$$D_{+}f(x_0) := \lim_{x \downarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = \inf_{x > x_0} \frac{f(x) - f(x_0)}{x - x_0} \tag{1.3.3}$$

respectively.
Furthermore it follows $D_{-}f(x_0) \le D_{+}f(x_0)$.

Theorem 1.3.1: Suppose f: $\mathbb{R} \to \mathbb{R} \bigcup \{+\infty\}$ be a function not identically $+\infty$ and be convex. Then at all $x_0 \in \operatorname{int} dom f$, the function f admits a finite left derivative and a finite right derivative.

Proof: By Theorem 1.1.3 b), the differences quotient (the slope function) involved in (1.3.2) and (1.3.3), call it φ is monotone increasing. For arbitrary two points say $x_1, x_2 \in \operatorname{int} dom f$ satisfying $x_1 < x_0 < x_2$, $\varphi(x_1)$ and $\varphi(x_2)$ are finite numbers satisfying $\varphi(x_1) \le \varphi(x_2)$. Furthermore, when $x_1 \uparrow x_0$ [resp. $x_2 \downarrow x_0$], $\varphi(x_1)$ increases [resp. $\varphi(x_2)$ decreases], hence they both converge. This completes the proof. //

Definition 1.3.2: Let $X \subseteq \mathbb{R}^n$ be nonempty, let $f : X \to \mathbb{R}$ and let $x_0 \in X$, $d \in \mathbb{R}^n$. Then

a) The function f is said to be *Gateaux- (or directional) differentiable* at x_0 in the direction of d if and only if there exists $\varepsilon > 0$ such that $[x_0 - \varepsilon d, x_0 + \varepsilon d] \subseteq X$ and the limit

$$f'(x_0, d) := \lim_{t \to 0} \frac{f(x_0 + td) - f(x_0)}{t}$$

exists and is finite for all $d \in \mathbb{R}^n$. In this case $f'(x_0, d)$ is known as *Gateaux-differential* of f at x_0 in the direction of *d*. The mapping $f'(x_0, \cdot) : \mathbb{R}^n \to \mathbb{R}$ is said to be *Gateaux-derivative* of f at x_0.

Correspondingly, we define

(i) the right-sided Gateaux- (directional) derivative of f by

$$f'_{+}(x_0, d) := \lim_{t \downarrow 0} \frac{f(x_0 + td) - f(x_0)}{t}, \quad [x_0, x_0 + \varepsilon d] \subseteq X \text{ for some } \varepsilon > 0$$

(ii) the left-sided Gateaux- (directional) derivative of f by

$$f'_-(x_0,d) := \lim_{t\uparrow 0}\frac{f(x_0+td)-f(x_0)}{t},\ [x_0-\varepsilon d, x_0]\subseteq \mathrm{X} \text{ for some } \varepsilon>0$$

b) The function f is said to be *Frechet-differentiable* at $x\in X$ ($X\subseteq \mathbb{R}^n$ be an open set) if and only if there exists a linear and continuous mapping $A:\mathbb{R}^n\to\mathbb{R}$ such that

$$\lim_{\|h\|\to 0,\ \|h\|\neq 0}\frac{|f(x+th)-f(x)-A(h)|}{\|h\|}=0.$$

$A(h)$ is called the *Frechet-derivative* of f at x. The mapping $f':U\to \vartheta(\mathbb{R}^n,\mathbb{R})$(the set of linear and continuous functional) is said to be the *Frechet-derivative* of f.

Remark 1.3.1: For a convex function $f:\mathbb{R}^n\to\mathbb{R}$ if we define the derivative $D:\mathbb{R}^n\to\mathbb{R}$ attached to f at x and consider

$$\frac{f(x+h)-f(x)-D(h)}{\|h\|}\xrightarrow{\|h\|\to 0}0 \tag{1.3.4}$$

Obviously (1.3.4) holds from the point of view of *Frechet*. Furthermore, from the point of view of *Gateaux-differentiability* it holds for $h:=td, d\in\mathbb{R}^n$ fixed, $t\to 0$ in $\mathbb{R}$.

Now some properties of the Gateaux-derivative of convex functions are given in the following.

Let us consider a convex function $f:X\to\mathbb{R}$, a fixed point $x\in X\subseteq\mathbb{R}^n$, a direction $d\in\mathbb{R}^n$ and $\varphi:\widetilde{X}\to\mathbb{R}$, where $\widetilde{X}:=\{t\in\mathbb{R}^+: x+td\in X\}$ defined by

$$\varphi(t):=f(x+td)\ .$$

Then it is clear that

$$f'_+(x,d)=\lim_{t\downarrow 0}\frac{1}{t}[f(x+td)-f(x)]=\lim_{t\downarrow 0}\frac{\varphi(t)-\varphi(0)}{t}=D_+\varphi(0)$$

and $f'_-(x,d)=D_-\varphi(0)$.

Furthermore $f'_-(x,-d)=-D_-\varphi(0)$.

Theorem 1.3.2: Let $X\subseteq\mathbb{R}^n$ be convex, let $f:X\to\mathbb{R}\bigcup\{+\infty\}$ be a convex function and let $x_0\in X$ be any algebraic interior point (*i.e* $\exists\varepsilon>0:[x_0-\varepsilon d, x_0+\varepsilon d]\subseteq X$ for all $d\in\mathbb{R}^n$) such that $f(x_0)\in\mathbb{R}$. Then

a) f is *left sided* and *right sided Gateaux-differentiable* at x_0 in the direction d for all $d\in\mathbb{R}^n$.

b) The inequality (known as *subgradient inequality*):

$$f'_+(x_0,x-x_0)\le f(x)-f(x_0),\ \text{for all } x\in X. \tag{1.3.5}$$

holds.

Proof: a) Let

$$\varphi(t) := \frac{f(x_0 + td) - f(x_0)}{t} \quad \text{for } t \in \mathbb{R}.$$

Since $f'_+(x_0, d) = \lim_{t \downarrow 0} \varphi(t)$ and φ is increasing (by Theorem 1.1.3 b)) the limit exists. Also $f'_-(x_0, d) = \lim_{t \uparrow 0} \varphi(t)$ exists. So we have a).

b) Now set $d := x - x_0$, then for $x \in X$ we have $1 \in I_{x-x_0} := \{t \in \mathbb{R}_+ : x_0 + t(x - x_0) \in X\}$. As φ is increasing by a) we get

$$f'_+(x_0, x - x_0) = \lim_{t \downarrow 0} \varphi(t) \leq \varphi(1) = f(x) - f(x_0). //$$

Remark 1.3.2: (*i*) A special case, for the Theorem 1.3.2 b) is that when *f* is differentiable at x_0 we have

$$f'_+(x_0, x - x_0) = \lim_{t \downarrow 0} \frac{f(x_0 + t(x - x_0)) - f(x_0)}{t} = \langle \nabla f(x_0), x - x_0 \rangle$$

and hence (1.3.5) becomes

$$\langle \nabla f(x_0), x - x_0 \rangle \leq f(x) - f(x_0) \quad \text{for all } x \in X$$

(*ii*) In case $f(X) \subseteq \mathbb{R}$, we have $f'_+(x_0, d)$ and $f'_-(x_0, d)$ for all $d \in \mathbb{R}^n$ are real values. Moreover,

$$f'_-(x_0, d) \leq f'_+(x_0, d) \text{ for all } d \in \mathbb{R}^n.$$

2. Definitions and Interpretations of the Subdifferentials

2.1 The Definitions of the Subdifferentials

Let $f:\mathbb{R}^n \to \mathbb{R}$ be a convex function and let $x, d \in \mathbb{R}^n$ be fixed. Then by Theorem 1.1.3, b) the differences quotient of f at x in the direction d is given by

$$s(t) := \frac{f(x+td)-f(x)}{t} \qquad t>0. \tag{2.1.1}$$

is increasing and bounded near zero. So by Theorem 1.3.2 the function f is *left sided* and *right sided* directional-differentiable at x in the direction of d for all $d \in \mathbb{R}^n$ *i.e* the limits

$$\lim_{t\downarrow 0} s(t) = \inf\{s(t) : t>0\}$$

and

$$\lim_{t\uparrow 0} s(t) = \sup\{s(t) : t<0\}$$

exist and are finite. In this case, we say that f is *right sided* and *left sided directionally differentiable* at x in the direction d.

Proposition 2.1.1: Let $f:\mathbb{R}^n \to \mathbb{R}$ be a convex function. For every fixed $x \in \mathbb{R}^n$

a) The mapping $f'_+(x,\cdot)$ is sublinear and the mapping $f'_-(x,\cdot)$ is superlinear.

b) If f is Lipschitz (with a constant L), then $f'_+(x,\cdot)$ is Lipschitz on $\mathbb{R}^n$ with the same constant L

Proof: a) .Let $d_1, d_2 \in \mathbb{R}^n$ and $\lambda_1, \lambda_2 > 0$ with $\lambda_1+\lambda_2=1$, then by definition

$$f'_+(x, \lambda_1 d_1 + \lambda_2 d_2) = \lim_{t\downarrow 0} \frac{1}{t}[f(x+t(\lambda_1 d_1 + \lambda_2 d_2)) - f(x)].$$

But due to the convexity of f one has

$$\begin{aligned} f(x+t(\lambda_1 d_1 + \lambda_2 d_2)) - f(x) &= f((\lambda_1+\lambda_2)x + t(\lambda_1 d_1 + \lambda_2 d_2)) - (\lambda_1+\lambda_2) f(x) \\ &= f(\lambda_1(x+td_1) + \lambda_2(x+td_2)) - \lambda_1 f(x) - \lambda_2 f(x) \\ &\le \lambda_1[f(x+td_1) - f(x)] + \lambda_2[f(x+td_2)-f(x)]. \end{aligned}$$

So we have

$$\begin{aligned} f'_+(x, \lambda_1 d_1 + \lambda_2 d_2) &= \lim_{t\downarrow 0} \frac{1}{t}[\lambda_1[f(x+td_1)-f(x)] + \lambda_2[f(x+td_2)-f(x)]] \\ &= \lambda_1 f'_+(x,d_1) + \lambda_2 f'_+(x,d_2) \end{aligned}$$

Thus $f'_+(x,\cdot)$ is convex.

To complete the proof let $\lambda>0$. Then the equality

$$f'_+(x,\lambda d) = \lim_{t\downarrow 0}\frac{1}{t}[f(x+t\lambda d)-f(x)] = \lambda \lim_{t\downarrow 0} \frac{1}{\lambda t}[f(x+t\lambda d) - f(x)] = \lambda f'_+(x,d)$$

implies $f'_+(x,\cdot)$ is positively homogenous. From this it follows, $f'_+(x,\cdot)$ is sublinear.

Analogously we can prove the second part of a).

b) For sufficiently small $t > 0$ we have

$$|f(x+td_1) - f(x+td_2)| \le tL\|d_1 - d_2\|$$

Therefore,

$$\left|\frac{f(x+td_1) - f(x)}{t} - \frac{f(x+td_2) - f(x)}{t}\right| \le L\|d_1 - d_2\|$$

Passing in the above expression to the limit as $t \downarrow 0$ we get

$$|f'_+(x,d_1) - f'_+(x,d_2)| \le L\|d_1 - d_2\|. \; //$$

Since the function $f'_+(x,\cdot)$ is sublinear and $f'_-(x,\cdot)$ is superlinear, we have the following relations

$$f'_+(x,d) = \max_{s\in U}\langle s,d\rangle, \text{ for all } d \in \mathbb{R}^n$$

$$f'_-(x,d) = \min_{\bar{s}\in V}\langle \bar{s},d\rangle, \text{ for all } d \in \mathbb{R}^n.$$

where U and V are convex compact set.

Thus, in view of the correspondence between finite sublinear function and compact convex sets, $f'_+(x,\cdot)$ can be expressed as

$$f'_+(x,d) := \max\{\langle s,d\rangle : s \in S\} \tag{2.1.2}$$

for all $d\in\mathbb{R}^n$ and for some nonempty compact convex set S. So it results directly from proposition 2.1.1 that $f'_+(x,\cdot)$ is a support function of some nonempty compact convex set.

So we have the following definition.

Definition 2.1.1: Let a function f be defined on $\mathbb{R}^n$ and be directionally differentiable at a point $x\in\mathbb{R}^n$. Then f is said to be

a) *subdifferentiable* at x if $f'_+(x,\cdot)$ is a sublinear function, *i.e* there exists a convex compact set U such that

$$f'_+(x,d) = \max_{s\in U}\langle s,d\rangle, \text{ for all } d \in \mathbb{R}^n$$

b) *superdifferentiable* at x if it is a function whose directional derivative $f'_+(x,\cdot)$ takes the form:

$$f'_+(x,d) = \min_{h\in V}\langle h,d\rangle, \text{ for all } d \in \mathbb{R}^n$$

where V is a convex compact set.

It is clear that a function f is superdifferentiable if and only if $-f$ is subdifferentiable. So all results concerning sublinear functions and their subdifferentials can also be reformulated (with appropriate alteration) in the case of superlinear functions and their superdifferentials.

These convex compact sets U and V are called the *subdifferential* (*superdifferential*) of a function f at the point x and are denoted by $\underline{\partial} f(x)$ (resp. $\overline{\partial} f(x)$). Now for the sake of simplicity we use the notation ∂f instead of $\underline{\partial} f$ throughout this seminar report.

Now we have the following definition for the subdifferential of a function at a fixed-point call it x.

Definition 2.1.2: Let f: $\mathbb{R}^n \to \mathbb{R}$ be a convex function and $x \in \mathbb{R}^n$ be fixed. Then the set

$$\partial f(x) := \{s \in \mathbb{R}^n : \langle s,d\rangle \le f'_+(x,d) \text{ for all } d \in \mathbb{R}^n\}, \tag{2.1.3}$$

is called the *subdifferential* of f at x. Obviously $\partial f(x)$ is nonempty and compact.

An element of the subdifferential $s \in \mathbb{R}^n$ is said to be a *subgradient* of the function f at the point x. Now from this definition it is easy to show that

$$f'_-(x,d) = -f'_-(x,-d) \le \langle s,d\rangle \le f'_+(x,d), \text{ for all } s \in \partial f(x), \text{ for all } d \in \mathbb{R}^n.$$

i.e $f'_+(x,\cdot)$ majorizes the function $f'(x,\cdot)$, and $f'_-(x,\cdot)$ minorizes the function $f'(x,\cdot)$.

Consider a function f, which is Gateaux-differentiable at a point x, then since $f'_+(x,\cdot)$ majorizes the function $f'(x,\cdot)$, we have

$$\langle \nabla f(x), d\rangle \le f'_+(x,d), \text{ for all } d \in \mathbb{R}^n$$

This means that the vector $\nabla f(x)$ is contained in the subdifferential of f at x. *i.e*

$$\nabla f(x) \in \partial f(x)$$

Figure 2.1.1 illustrates the notion of subdifferential of some convex functions at some fixed point.

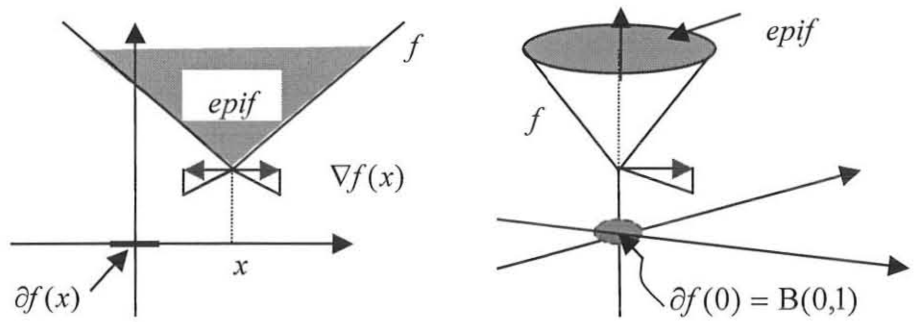

Figure 2.1.1 (Subdifferential of a convex function)

Example 2.1.1: Let f: $\mathbb{R} \to \mathbb{R}$ defined by $f(x) := |x-a|$ for $x \in \mathbb{R}$ and $a \in \mathbb{R}$.

Obviously f is convex. To compute $\partial f(x)$, we need to calculate the directional derivative $f'_+(x,\cdot)$ and to determine the set that it supports.

For any $d \in \mathbb{R}$, $d \neq 0$

$$f'_+(x,d) = \lim_{t \downarrow 0} \frac{1}{t}[|x+td-a| - |x-a|] = \begin{cases} d, if \;\; x > a \\ -d, if \;\; x < a \\ |d|, if \;\; x = a \end{cases} \qquad (2.1.4)$$

So we have

$$\partial f(x) := \{y \in \mathbb{R}: sd \leq f'_+(x,d) \text{ for all } d \in \mathbb{R}\}.$$

Then from $sd \leq f'_+(x,d)$ depending on the value of d we have

$$s \leq \frac{1}{d} f'_+(x,d) \text{ for } d>0 \text{ .and } s \geq \frac{1}{d} f'_+(x,d) \text{ for d} <0. \qquad (2.1.5)$$

Using (2.1.5) in connection with (2.1.4) we get

$$s \in \begin{cases} \{1\}, \; if \;\; x > a \\ \{-1\}, \; if \;\; x < a \\ [-1,1], \; if \; x = a \end{cases}$$

Therefore, the subdifferential of f at any $x \in \mathbb{R}$ is given by

$$\partial f(x) := \begin{cases} \{1\}, & if \ x > a \\ \{-1\}, & if \ x < a \\ [-1,1], & if \ x = a \end{cases}$$

and $\bar{\partial} f(x) = \{0\}$ for all $x \in \mathbb{R}^n$.

Remark 2.1.1: From Definition 2.1.2, the right-sided directional derivative $f'_+(x,\cdot)$ is characterized (whenever $\partial f(x) \neq \emptyset$) as

$$f'_+(x,d) = \sup\{\langle s,d\rangle : s \in \partial f(x)\}.$$

Then it follows that $f'_+(x,\cdot)$ is a support function of $\partial f(x)$.

Proposition 2.1.2: The sublinear function $\sigma(\cdot) := f'_+(x,\cdot)$ satisfies the following

(*i*) $\sigma'_+(0,z) = \sigma(z) = f'_+(x,z)$ for all $z \in \mathbb{R}^n$,

(ii) $\partial\sigma(0) = \partial f(x)$.

Proof: Because σ is positively homogenous and $\sigma(0) = f'_+(x,0) = 0$

$$\sigma'_+(0,z) = \lim_{t \downarrow 0} \frac{1}{t}[\ \sigma(0+tz) - \sigma(0)] = \sigma(z) = f'_+(x,z) \text{ for all } z \in \mathbb{R}^n \text{ and for } t > 0.$$

which establishes (i).

To prove the second part, apply the uniqueness of the supported set on the given functions:

$$\partial\sigma(0) = \{s \in \mathbb{R}^n : \langle s,z\rangle \le \sigma'_+(0,z) \text{ for all } z \in \mathbb{R}^n\}$$

$$= \{s \in \mathbb{R}^n : \langle s,z\rangle \le f'_+(x,z) \text{ for all } z \in \mathbb{R}^n\} = \partial f(x).\ //$$

The directional derivative $d \mapsto f'_+(x,d)$ (*i.e* $f'_+(x,\cdot)$) of a convex function *f* at a point *x* is sublinear (Proposition 2.1.1), its subdifferential $\partial[f'_+(x,\cdot)]$ at zero coincides with the subdifferential ∂f of *f* at the point *x* (Proposition 2.1.2).

Remark 2.1.2: A finite sublinear function σ has a subdifferential, just as any other convex function. The subdifferential of a sublinear function σ at zero is called the subdifferential of σ. Its subdifferential at 0 is defined by

$$\partial\sigma(0) := \{s \in \mathbb{R}^n : \langle s,d\rangle \le \sigma(d) \text{ for all } d \in \mathbb{R}^n\}.$$

Consequently, a finite sublinear function is the support of its subdifferential at zero.

Like wise for a superlinear function ψ, the set

$$\bar{\partial}\psi(0) := \{\bar{s} \in \mathbb{R}^n : \langle \bar{s}, d\rangle \geq \psi(d) \text{ for all } d \in \mathbb{R}^n\}$$

is called the *superdifferential* of ψ.

Example 2.1.2: Let S be a nonempty convex compact set, with support function σ_S. Then from the definition of subdifferentials and proposition 2.1.2 we have

$$\partial\sigma(0) = S \text{ and } (\sigma_S)'_+(0,.) = \sigma_S.$$

Thus, any compact convex set S can be considered as the subdifferential of some finite convex function f at some point. As an example take $f := \sigma_S$, $x = 0$.

Corollary 2.1.1: Let f: $\mathbb{R}^n \to \mathbb{R}$ be a convex function. For fixed $x \in \mathbb{R}^n$ let

$$C := \{s \in \mathbb{R}^n \;\; f(y) \geq f(x) + \langle s, y - x\rangle \text{ for all } y \in \mathbb{R}^n\} \tag{2.1.6}$$

Then $C = \partial f(x)$.

Proof: Let $s \in \partial f(x)$, *i.e*

$$\langle s, d\rangle \leq f'_+(x,d) \text{ for all } d \in \mathbb{R}^n.$$

Then by definition

$$f'_+(x,d) = \lim_{t \downarrow 0} \frac{1}{t}[f(x+td) - f(x)] = \inf\{\frac{1}{t}[f(x+td) - f(x)] : t > 0\}$$

$$\leq \frac{1}{t}[f(x+td) - f(x)] \text{ for all } d \in \mathbb{R}^n \text{ and } t > 0.$$

This implies

$$t\langle s, d\rangle \leq f(x+td) - f(x) \text{ for all } d \in \mathbb{R}^n \text{ and } t > 0.$$

So we get $f(x+td) \geq f(x) + \langle s, td\rangle$ for all $d \in \mathbb{R}^n$ and $t > 0$.

For any $d \in \mathbb{R}^n$ and any $t \in \mathbb{R}^+$, set $y := x + td$ be also arbitrary element in $\mathbb{R}^n$ and it follows that

$$f(y) \geq f(x) + \langle s, y - x\rangle, \text{ for all } d \in \mathbb{R}^n$$

This implies $s \in C$.

For the proof of the converse inclusion, put $td := y - x$ for any $d \in \mathbb{R}^n$ and $t > 0$ and then use definition of directional derivative $f'_+(x,\cdot)$ and Definition 1.1.4 c) to show $C \subseteq \partial f(x)$. //

Consequently, we have

$$\partial f(x) := \{s \in \mathbb{R}^n : f(y) - f(x) \geq \langle s, y - x \rangle \text{ for all } y \in \mathbb{R}^n). \qquad (2.1.7)$$

The linear form

$$\sigma_{s,x}(y) := f(x) + \langle s, y - x \rangle$$

is an affine function, which minimizes *f* and coincides with *f* for $y = x$. So (2.1.7) gives a necessary and sufficient condition for *x* to minimize *f*. That is *x* minimizes *f* if and only if $0 \in \partial f(x)$.

From (2.1.7), we mean that the elements of $\partial f(x)$ are the slopes of the hyperplanes supporting the epigraph of *f* at $(x, f(x)) \in \mathbb{R}^n \times \mathbb{R}$. This result illustrates as follows.

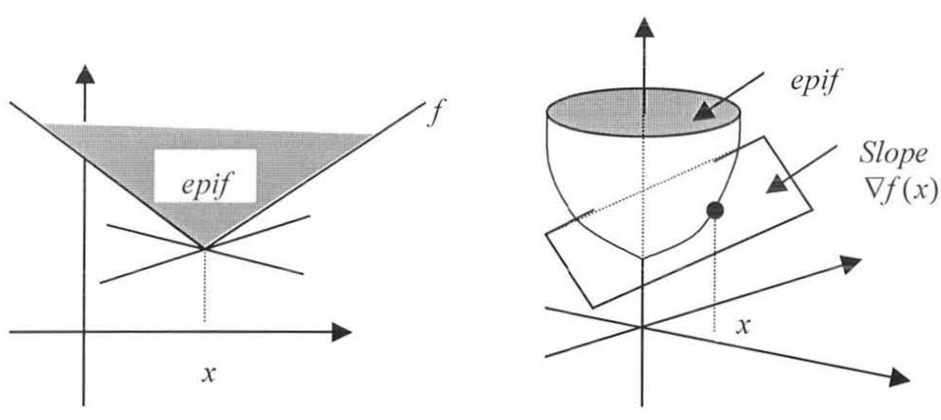

Figure 2.1.3 (Slopes of supporting hyperplanes to epigraph)

For the one dimensional convex function *f*, the subdifferential $\partial f(x)$ is the set of all $s \in \mathbb{R}$ such that $D_- f(x) \leq s \leq D_+ f(x)$ when $D_- f(x)$ and $D_+ f(x)$ are finite. *i.e*

$$\partial f(x) := [D_- f(x), D_+ f(x)]$$

If *f* is differentiable at *x*, then $D_- f(x) = D_+ f(x)$ and therefore $\partial f(x)$ is a singleton. *i.e* $\partial f(x) = \{Df(x)\}$.

Let f be a function defined on $\mathbb{R}^n$. Then

a) If f is convex and finite, then f is directionally differentiable at a point $x \in \mathbb{R}^n$ and

$$f'_+(x,d) = \max_{s \in \partial f(x)} \langle s, d \rangle \text{ for all } d \in \mathbb{R}^n$$

where $\partial f(x) := \{s \in \mathbb{R}^n : f(y) - f(x) \geq \langle s, y-x \rangle$ for all $y \in \mathbb{R}^n\}$ but $\bar{\partial} f(x) = \{0\}$ for all $x \in \mathbb{R}^n$.

This implies the subdifferentiable of a convex function.

b) Whereas, if f is concave, then $\bar{\partial} f(x) := \{\bar{s} \in \mathbb{R}^n : f(y) - f(x) \leq \langle \bar{s}, y-x \rangle$ for all $y \in \mathbb{R}^n$),

and $\partial f(x) = \{0\}$, for all $x \in \mathbb{R}^n$.

Thus, all properties of a convex function and its subdifferential hold (with appropriate alteration) in the case of a concave function and its superdifferential

Example 2.1.3: let f and g be functions defined on $\mathbb{R}$ such that $f(x) = |x|, x \in \mathbb{R}$ and $g(x) = -|x|, x \in \mathbb{R}$. Then their subdifferentials and superdifferentials are given by

$$\bar{\partial} g(x) = \partial f(x) = \begin{cases} \{1\}, & \text{if } x > 0 \\ \{-1\}, & \text{if } x < 0 \\ [-1,1], & \text{if } x = 0 \end{cases}$$

Whereas

$$\partial g(x) = \bar{\partial} f(x) = \{0\} \text{ , for all } x \in \mathbb{R}$$

Definition 2.1.3: A point x at which $\partial f(x)$ has more than one element is called a *kink* (or corner point) of f. In such case f is not differentiable at a point x. If x is not a kink, then clearly f is differentiable at x and its subdifferential is a singleton.

Example 2.1.4: Let a, $b \in \mathbb{R}$ such that $a < b$ and α, β be positive real numbers. If $f: \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) := \alpha|x-a| + \beta|x-b|,$$

then clearly f is convex as it is the sum of convex functions and differentiable at all x except at a and b. With out loss of generality assume that $\alpha \leq \beta$. Then we have the following graph for f

.

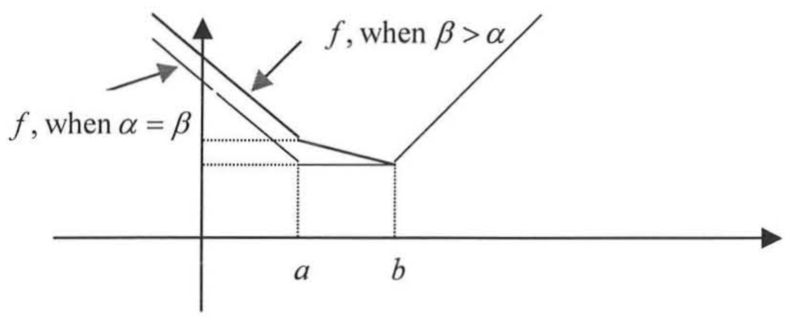

Figure 2.1.3 (Kink points of a function)

For $x \notin \{a,b\}$, f is differentiable at x and then

$$\partial f(x) = \begin{cases} \{-(\alpha+\beta)\}, if \ \ x < a \\ \{\alpha-\beta\}, if \ \ x \in (a,b) \\ \{\alpha+\beta\}, if \ \ x > b \end{cases}$$

But for $x \in \{a,b\}$ f is not differentiable and hence its kinks are at $x = a$ and $x = b$, with subdifferentials

$$\partial f(x) = \begin{cases} \{-\beta\} + [-\alpha,\alpha], if \ \ x = a \\ \{\alpha\} + [-\beta,\beta], if \ \ x = b \end{cases}$$

Now we can generalize this example as follows.

In general, suppose $b_j, j = 1,2,\cdots,m$ be real numbers and t_j, $j = 1,2,\cdots,m$ be positive real numbers. Then consider a function f defined by

$$f(x) := \sum_{j=1}^{m} t_j \left| x - b_j \right|.$$

Obviously, f is convex. This f is differentiable at all x except at $b_1, b_2, \cdots, b_m$. Then

$$\partial f(x) := \begin{cases} \{ \sum_{j:b_j<x} t_j - \sum_{j:b_j>x} t_j \}, if \ \ x \notin \{b_1, b_2, \cdots, b_m\} \\ \{ \sum_{j:b_j<x} t_j - \sum_{j:b_j>x} t_j \} + [-t_{j_0}, t_{j_0}], if \ \ x = b_{j_0} \end{cases}$$

Theorem 2.1.2: For all x the set $\partial f(x)$ described in (2.1.7) is nonempty, closed, convex and bounded.

Proof: The nonemptyness and convexity of the set $\partial f(x)$ follows immediately from the definition of subdifferential: the closed sublinear function $f'_+(x,\cdot)$ supports a nonempty set and by the fact that any convex function defined on $\mathbb{R}^n$ is minimized by some affine function. The closedness of the mapping $x \mapsto \partial f(x)$ implies the closedness of the set $\partial f(x)$. To prove the boundedness we use the local Lipschitz property of *f*. For this end let $s \in \partial f(x)$ such that $s \neq 0$ and take $y := x + \frac{\delta s}{\|s\|}$ for $\delta > 0$ be arbitrary. Since *f* is convex, by Theorem 1.2.2 there is a *Lipschitzian* constant L such that

$$|f(y) - f(x)| \leq L\|y - x\| = L\left\|\frac{\delta s}{\|s\|}\right\| = L\delta$$

i.e $f(x) - L\delta \leq f(y) \leq f(x) + L\delta$. (2.1.8)

On the other hand, substituting $y := x + \frac{\delta s}{\|s\|}$ for $\delta > 0$ in (2.1.7) gives

$$f(y) \geq f(x) + \left\langle s, \frac{\delta s}{\|s\|} \right\rangle = f(x) + \delta\|s\|. \qquad (2.1.9)$$

Thus (2.1.9) together with (2.1.8) yields

$$f(x) + \delta\|s\| \leq f(y) \leq f(x) + L\delta. //$$

As a result, for $x_0 \in \operatorname{int} dom f$, $\partial f(x_0)$ is a nonempty compact convex set. At the boundary point ∂f is certainly empty if *f* is not closed: while if *f* is closed, ∂f may be empty (case of vertical slope of the hyperplane supporting *epif*), otherwise it is nonempty. Moreover, for $x_0 \notin dom f$, $\partial f(x_0)$ is automatically empty.

2.2 Geometrical Interpretation of Subdifferentials

In order to investigate the geometric construction and interpretation of subdifferentials it is good to have the notion of a cone with some basic properties that we need very often in this section.

Definition 2.2.1: Let $K \subseteq \mathbb{R}^n$ be nonempty. Then

a) the set *K* is called a *cone* if and only if $\alpha K \subseteq K$ for all $\alpha \geq 0$.

b) Let *K* be a convex cone. The *polar cone* of *K* is denoted by K^0 and defined as

$$K^0 := \{s \in \mathbb{R}^n : \langle s, x \rangle \leq 0 \text{ for all } x \in K\}.$$

To illustrate the notion of a cone and a polar cone one consider the following two sets:

a). $K_1 := \{x : x = \sum_{j=1}^{3} \alpha_j x_j\ ,\ x_j \in \mathbb{R}^3,\ x_j \geq 0 \text{ for } j=1, 2,3\}$

b). $K_2 := \{(x, y) \in \mathbb{R}^2 : 3x - y \geq 0,\ x - 3y \leq 0,\ x, y \in \mathbb{R}, x \geq 0, y \geq 0\}$.

Then their polar cones are:

a) $K_1^0 = \{s \in \mathbb{R}^3 : \langle s, x_j \rangle \leq 0\ \ for\ j = 1,2,3\}$.

b) $K_2^0 = \{(x, y) \in \mathbb{R}^2 :\ 3x+y \leq 0, x+3y \leq 0, x, y \in \mathbb{R}, x \geq 0, y \geq 0\}$.

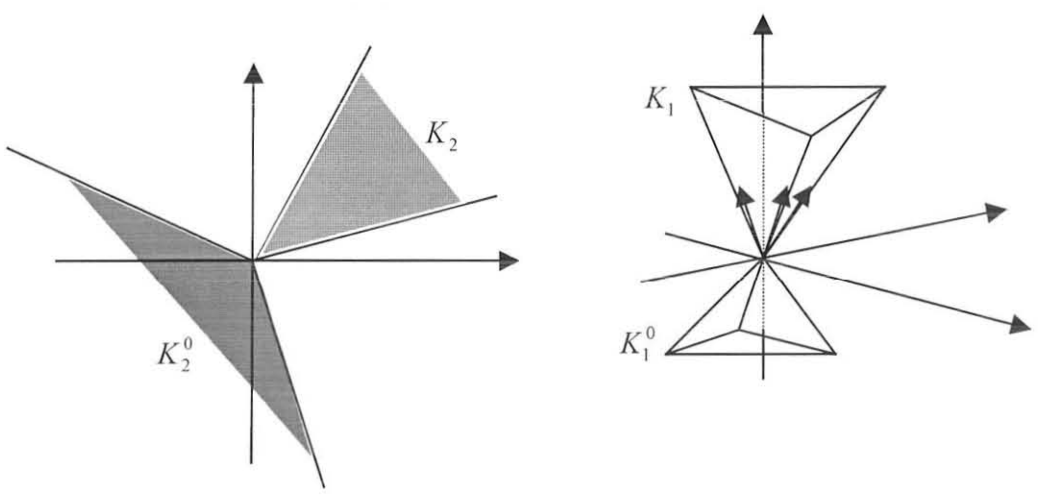

Figure 2.2.1 (Examples of polar cones)

Remark 2.2.1: In Definition 2.1.2 b) $(K^0)^0 = K$, whenever K is closed. If K is a subspace, then K^0 is its orthogonal. From this it follows that polarity generalizes orthogonality. As a particular case, the polar cone of the *nonnegative orthant*

$$\Omega_+ := \{x \in \mathbb{R}^n : x = (x_1, x_2, \cdots, x_n),\ x_j \geq 0,\ j = 1,2,\cdots, n\}.$$

is the *non-positive orthant*,

$$\Omega_- := (\Omega_+)^0 = \{s \in \mathbb{R}^n : s = (s_1, s_2, \cdots, s_n),\ s_j \leq 0,\ j = 1,2,\cdots, n\}.$$

Definition 2.2.2: Let S be nonempty closed convex set

a) We say that $d \in \mathbb{R}^n$ is a direction *tangent* to S at $x \in S$ when there exists a sequence $\{s_k\} \subset S$ and a sequence $\{t_k\}$ such that

$$x_k \xrightarrow{k\to\infty} x,\ t_k \downarrow 0,\ \frac{x_k - x}{t_k} \xrightarrow{k\to\infty} d.$$

The set of all such directions is called the *tangent cone* $T_S(x)$ to S at $x \in$ S.

b) The direction $s \in \mathbb{R}^n$ is said *normal* to S at $x \in S$ when

$$\langle s, y - x\rangle \le 0 \text{ for all } y \in \text{S}. \qquad 2.2.1)$$

The set of all such directions is called *normal cone* to S at x, denoted by $N_S(x)$. *i.e*

$$N_S(x) := \{s \in \mathbb{R}^n : \langle s, y - x\rangle \le 0 \text{ for all y} \in \text{S}\}.$$

This is illustrated as follows

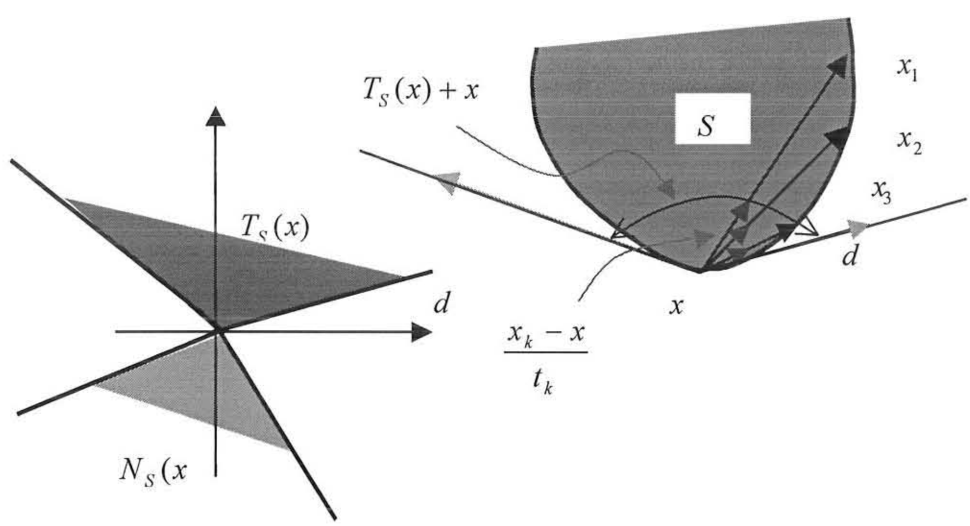

Figure 2.2.2 (Tangents and normals to S)

From (2.2.1), one can easily deduce $s \in N_S(x)$ is a vector in $\mathbb{R}^n$ such that the angle between s and $y - x$ is obtuse for each y in S.

Definition 2.2.3: A *conical combination* of elements $x_1, x_2, \cdots, x_n$ is an element of the form $\sum_{j=1}^{k} \lambda_j x_j$, where the coefficients λ_j are nonnegative. The set of all such conical combinations from a given nonempty set $S \subseteq \mathbb{R}^n$ is said to be the *conical hull* of S, denoted by *coneS*.

Now if we set $\overline{\lambda} := \sum_{j=1}^{k} \lambda_j$ is positive, then we can put $\beta_j := \frac{\lambda_j}{\overline{\lambda}}$ to realize that a conical combination of the type

$$\sum_{j=1}^{k} \lambda_j x_j = \overline{\lambda} \sum_{j=1}^{k} \beta_j x_j \text{ , with } \overline{\lambda}>0,\ \sum_{j=1}^{k} \beta_j = 1,\ \beta_j \geq 0 \text{ for } j = 1,2,\cdots.k\,.$$

is nothing but a convex combination, multiplied by an arbitrary positive coefficient.

Definition 2.2.4: The closed *conical hull* (or rather closed convex conical hull) of a nonempty set $S \subseteq \mathbb{R}^n$ is given by

$$\overline{coneS} := cl(coneS) = cl\{ \sum_{j=1}^{k} \lambda_j x_j : \lambda_j \geq 0, x_j \in \mathrm{S} \text{ for } j = 1,2,\cdots,k \}.$$

Proposition 2.2.1: Let $S \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Then

a) the tangent cone to S is closed.

b) the tangent cone to S at x is the closure of the cone generated by $S - \{x\}$:

$$\begin{aligned} T_S(x) &:= \overline{cone(S - x)} \\ &= cl\,\mathbb{R}^+(S - x) \\ &= cl\{d \in \mathbb{R}^n : d = \lambda(y - x), y \in \mathrm{S},\ \lambda \geq 0\}. \end{aligned}$$

Proof: a) Let a sequence $\{d_r\} \subset T_S(x)$ be converging to d. For each r consider sequences $\{x_{r,k}\}$ and $\{t_{r,k}\}$ associated with d_r in the sense of Definition 2.2.2 a). Fix $r > 0$, we can find k_r such that

$$\left\| \frac{x_{r,k_r} - x}{t_{r,k_r}} - d_r \right\| \leq \frac{1}{r}.$$

Letting $r \to \infty$, we then obtain the sequences $\{x_{r,k_r}\}$ and $\{t_{r,k_r}\}$ which defined as an element of $T_S(x)$. Therefore, $T_S(x)$ is closed.

b) Clearly $S - \{x\} \subseteq T_S(x)$. By a), $T_S(x)$ is a closed cone, then it immediately follows that $cl\,\mathbb{R}^+(S - x) \subseteq T_S(x)$

Conversely, let $d \in T_S(x)$ and take $\{x_k\}$ and $\{t_k\}$ as in the definition 2.2.2 (a). Then the point $\frac{x_k - x}{t_k}$ is in $\mathbf{R}^+(S - x)$, hence its limit d is in the closure of $\mathbf{R}^+(S - x)$. //

Proposition 2.2.2: The normal cone is the polar of the tangent cone and conversely.

More symbolically,

$$N_S(x) = [T_S(x)]^0 = \{d \in \mathbb{R}^n : \langle s,d \rangle \leq 0, \text{ for all } s \in T_S(x)\}, \tag{2.2.2}$$

$$T_S(x) = [N_S(x)]^0 = \{d \in \mathbb{R}^n : \langle s,d \rangle \leq 0, \text{ for all } s \in N_S(x)\}. \tag{2.2.3}$$

Proof: ($\Rightarrow$): Let $s\in N_S(x)$, then by definition, $\langle s, y-x\rangle \le 0$ for all $y\in S$. Set $d := y-x$ $d\in S-x$ be arbitrary and $\langle s,d\rangle \le 0$. But this is true for each $d\in \mathbf{R}^+(S-x)$, as well as for all $d\in cl\,\mathbf{R}^+(S-x)$. That means $s\in [T_S(x)]^0$.

On the other hand, let $s\in [T_S(x)]^0$ be arbitrary, then $\langle s,d\rangle \le 0$ for all $d\in T_S(x)$. In particular, this holds for all $d\in S-\{x\}\subseteq T_S(x)$. Thus for any $y\in S$, $\langle s,d\rangle=\langle s,y-x\rangle\le 0$ which is exactly 2.2.1. *i.e* $s\in N_S(x)$.

($\Leftarrow$): From proposition 2.2.1 a), we know that $T_S(x)$ is closed. Then by using (2.2.2) and Remark 2.2.1 it follows the result (2.2.3). //

In terms of tangent and normal cones, an element $s\in f(x)$ is expressed by the following proposition.

Proposition 2.2.3:

a) A vector $s\in\mathbb{R}^n$ is a subgradient of f at x if and only if $(s,-1)\in\mathbb{R}^n\times\mathbb{R}$ is normal to *epif* at $(x, f(x))$. *i.e*

$$N_{epif}(x,f(x)) := \{(\lambda s,-\lambda) : s\in\partial f(x),\ \ \lambda\ge 0\}.$$

b) The tangent cone to *epif* at $(x, f(x))$ is the epigraph of the right- sided directional derivative function $g(\cdot) := f'_+(x,\cdot)$:

$$T_{epif}(x,f(x)) := \{(d,r) : f'_+(x,d)\le r\}.$$

Proof: a) Suppose $(s,-1)\in\mathbb{R}^n\times\mathbb{R}$ is normal to *epif* at $(x, f(x))$. Then by Definition 2.2.2 b), s is normal to $\mathbb{R}^n$ at x and -1 is normal to $\mathbb{R}$ at $f(x)$.*i.e* $\langle s,y-x\rangle\le 0$ for all $y\in\mathbb{R}^n$ and $(-1)[r-f(x)]\le 0$, for all $r\in\mathbb{R}$.

Then we have

$$\langle s,y-x\rangle + (-1)[r-f(x)]\le 0,\ \text{for all } y\in\mathbb{R}^n \text{ and for all } r\in\mathbb{R}.$$

or

$$r\ge f(x)+\langle s,y-x\rangle,\ \text{for all } y\in\mathbb{R}^n \text{ and for all } r\in\mathbb{R}.$$

In particular, for $r=f(x)$, we have

$$f(y)\ge f(x)+\langle s,y-x\rangle \ \text{for all } y\in\mathbb{R}^n$$

Which is just (2.1.7) *i.e* $s\in\partial f(x)$.

Conversely, from (2.1.7) we get

$$\langle s,y-x\rangle + (-1)[f(y)-f(x)]\le 0 \ \text{for all } y\in\mathbb{R}^n$$

i.e

$$\langle s,y-x\rangle + (-1)[r-f(x)]\le 0 \ \text{for all } y\in\mathbb{R}^n \text{ for all } r\ge f(y) \tag{2.2.4}$$

For $x,y,s\in\mathbb{R}^n$ and $r\in\mathbb{R}$, $(s,-1)\in\mathbb{R}^n\times\mathbb{R}$ and $(y-x,r-f(x))\in\mathbb{R}^n\times\mathbb{R}$.

Therefore, we write (2.2.4) as

$$\langle (s,-1),(y-x,r-f(x))\rangle\le 0 \ \text{for all } y\in\mathbb{R}^n \text{ and for all } r\ge f(y)$$

This implies

$$\langle (s,-1),(y,r)-(r-f(x))\rangle \le 0 \text{ for all } y \in \mathbb{R}^n \text{ and for all } r \ge f(y)$$

Thus, $(s,-1) \in N_{epif}(x, f(x))$ and therefore, a) follows as the set of normals form a cone containing the origin.

b) By proposition 2.2.2, the tangent cone to *epif* is the polar of the normal cone to *epif*.
i.e $T_{epif}(x, f(x)) = \{(d,r) \in \mathbb{R}^n \times \mathbb{R} : \langle z,(d,r)\rangle \le 0 \text{ for all } z \in N_{epif}(x, f(x))\}$. Then we have $z = (\lambda s, -\lambda)$ for $s \in \partial f(x)$, $\lambda \ge 0$
And hence

$$\langle (\lambda s,-\lambda),(d,r)\rangle \le 0 \text{ for all } s \in \partial f(x),\ \lambda \ge 0.$$

or

$$\langle (\lambda s, d) + r(-\lambda) \le 0 \text{ for all } s \in \partial f(x), \lambda \ge 0. \qquad (2.2.5)$$

Clearly, (2.2.5) holds for $\lambda = 0$. If $\lambda > 0$, divide both sides by λ in (2.2.5) to get

$$\langle s,d\rangle - r \le 0 \text{ for all } s \in \partial f(x)$$

So we have

$$r \ge \langle s,d\rangle, \text{ for all } s \in \partial f(x).$$

This implies

$$r \ge \max\{\langle s,d\rangle : s \in \partial f(x)\} = f'_+(x,d).$$

i.e. $\mathrm{r} \ge f'_+(x,d)$
Therefore $T_{epif}(x, f(x)) := \{(d,r) : f'_+(x,d) \le r\}$. //

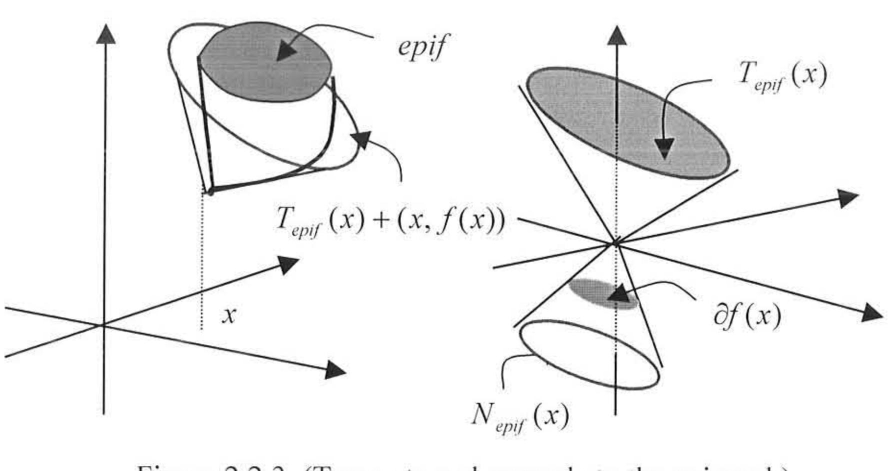

Figure 2.2.3 (Tangents and normals to the epigraph)

Definition 2.2.5: Let $x \in \mathbb{R}^n$ and f: $\mathbb{R}^n \to \mathbb{R}$ be convex. Then the set

$$Sf(x) := S_{f(x)}(f) = \{y \in \mathbb{R}^n : f(y) \le f(x)\}$$

is said to be the *level-set* passing through x.

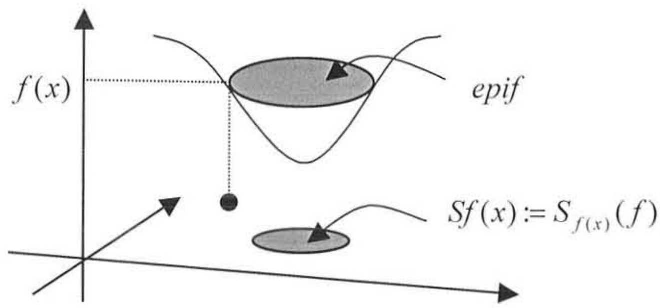

Figure 2.2.4 (Sublevel-set in $\mathbb{R}^3$)

Lemma 2.2.1: Let f: $\mathbb{R}^n \to \mathbb{R}$ be convex and $Sf(x)$ be the sublevel-set defined above. Then we have

$$T_{Sf(x)}(x) \subseteq \{d \in \mathbb{R}^n: f'_+(x,d) \leq 0\}. \tag{2.2.7}$$

Proof: Let $y \in Sf(x)$ be arbitrary, $t > 0$ and put $d := t(y-x)$. Then since $y \in Sf(x)$, $f(y) \leq f(x)$ for $y \in \mathbb{R}^n$ or $f(y) - f(x) \leq 0$. For $t > 0$, we have the following

$$0 \geq t[f(y) - f(x)] = t[f(x + \frac{d}{t}) - f(x)]$$

$$= \frac{f(x+\frac{d}{t}) - f(x)}{\frac{1}{t}} = \frac{f(x+\alpha d) - f(x)}{\alpha}, \quad \alpha = \frac{1}{t}$$

$$\geq \inf\{\frac{f(x+\alpha d) - f(x)}{\alpha} : \alpha > 0\} = f'_+(x,d).$$

Then we have

$$\mathbb{R}^+[Sf(x) - x] \subseteq \{d \in \mathbb{R}^n: f'_+(x,d) \leq 0\}.$$

As $f'_+(x,\cdot)$ is a closed function, $\{d \in \mathbb{R}^n: f'_+(x,d) \leq 0\}$ is closed and by Proposition 2.2.1, the tangent cone to $Sf(x)$ at x is the closure of the cone generated by $Sf(x) - \{x\}$:

$$T_{Sf(x)}(x) = \overline{cone}(Sf(x) - x) = cl\,\mathbb{R}^+[Sf(x) - x] \subseteq cl\{d \in \mathbb{R}^n: f'_+(x,d) \leq 0\}.$$
$$= \{d \in \mathbb{R}^n: f'_+(x,d) \leq 0\}.$$

Therefore,

$$T_{Sf(x)}(x) \subseteq \{d \in \mathbb{R}^n: f'_+(x,d) \leq 0\}. //$$

Remark 2.2.2: The converse inclusion in (2.2.7) need not be true. Of course, we need additional assumption to prove the converse inclusion. We will see this in the following theorem. As an example, consider $f(x)=\frac{1}{2}\|x\|^2$.

$$Sf(0)=\{y\in\mathbb{R}^n: f(y)\le f(0)\}=\{y\in\mathbb{R}^n: f(y)\le 0\}$$
$$=\{y\in\mathbb{R}^n:\frac{1}{2}\|y\|^2\le 0\}=\{0\}.$$

And then

$$T_{Sf(0)}(0)=cl\{d\in\mathbb{R}^n: d=t(y-0),\ y\in Sf(0),\ t>0\}$$
$$=cl\{d\in\mathbb{R}^n : d=ty,\ y=0,\ t>0\}$$
$$=cl\{d\in\mathbb{R}^n: d=0\}=\{0\}.$$

On the other hand, since $f_+'(0,d)=0$ for all $d\in\mathbb{R}^n$, we have

$$\{d\in\mathbb{R}^n: f_+'(0,d)\le 0\}=\mathbb{R}^n.$$

Therefore, in this example $\{0\}\subseteq\mathbb{R}^n$ but not the converse inclusion.

Proposition 2.2.4: Let g: $\mathbb{R}^n\to\mathbb{R}$ be convex and suppose that $g(x_0)<0$ for some $x_0\in\mathbb{R}^n$. Then

a) $cl\{z: g(z)<0\}=\{z: g(z)\le 0\}$,
b) $\{z: g(z)<0\}=\text{int}\{z: g(z)\le 0\}$,
c) $bd\{z: g(z)\le 0\}=\{z: g(z)=0\}$.

Proof: a) since each convex function on $\mathbb{R}^n$ is continuous the function g is continuous and hence it is both upper and lower semi- continuous.

Clearly $\{z: g(z)<0\}\subseteq\{z: g(z)\le 0\}$. From convexity and continuity of g, we have $\{z: g(z)\le 0\}$ is closed. Then taking the closure of both sides gives

$$cl\{z: g(z)<0\}\subseteq cl\{z: g(z)\le 0\}=\{z: g(z)\le 0\}.$$

Conversely, let $\bar{z}\in\{z: g(z)\le 0\}$ be arbitrary. Then $g(\bar{z})\le 0$. Now for $k>0$ and set $z_k:=\frac{1}{k}x_0+(1-\frac{1}{k})\bar{z}$

By convexity of g, $g(x_0)<0$ and $g(\bar{z})\le 0$ we have

$$g(z_k)=g(\frac{1}{k}x_0+(1-\frac{1}{k})\bar{z})\le\frac{1}{k}\,g(x_0)+(1-\frac{1}{k})\,g(\bar{z})<0$$

Then by continuity of g we get

$$\lim_{k\to\infty} g(z_k)=g(\lim_{k\to\infty} z_k)=g(\bar{z})<0.$$

Thus,

$$\bar{z}\in cl\{z: g(z)<0\}\text{ and hence }\{z: g(z)\le 0\}\subseteq cl\{z: g(z)<0\}.$$

b) Consider a) and then take its " int " in both sides *i.e*

$$\operatorname{int} cl\{z : g(z) < 0\} = \operatorname{int}\{z : g(z) \leq 0\}. \quad (2.2.8)$$

Now the left hand side in (2.2.8) is equal to $\operatorname{int}\{z : g(z) < 0\}$. Since g is upper semi-continuous, $\operatorname{int}\{z : g(z) < 0\} = \{z : g(z) < 0\}$.

So we have

$$\{z : g(z) < 0\} = \operatorname{int}\{z : g(z) \leq 0\}.$$

c) $bd\{z : g(z) \leq 0\} = cl\{z : g(z) \leq 0\} \setminus \operatorname{int}\{z : g(z) \leq 0\} = \{z : g(z) \leq 0\} \setminus \{z : g(z) < 0\}$
$= [\{z : g(z) < 0\} \bigcup \{z : g(z) = 0\}] \setminus \{z : g(z) < 0\} = \{z : g(z) = 0\}$. //

In the proposition 2.2.4, x_0 is called a *Slater* assumption. When this x_0 exists, taking *closures, interiors*, and *boundaries* of sublevel-sets accounts to imposing "$\leq$","$<$" and "=" in their respective definitions.

Theorem 2.2.2: Let f: $\mathbb{R}^n \to \mathbb{R}$ be convex and if $0 \notin \partial f(x)$. Then

a) For the sublevel set $Sf(x)$, we have

(i) $T_{Sf(x)}(x) = \{d \in \mathbb{R}^n : f'_+(x,d) \leq 0\}$,

(ii) $\operatorname{int}[T_{Sf(x)}(x)] = \{d \in \mathbb{R}^n : f'_+(x,d) < 0\} \neq \emptyset$.

c) A direction d in $\mathbb{R}^n$ is normal to $Sf(x)$ at x if and only if there is some $t \geq 0$ and come $s \in \partial f(x)$ such that $d = ts$:

$$N_{Sf(x)}(x) = \mathrm{R}^+ \partial f(x).$$

Proof: a) The forward inclusion of (i) is immediate from the Lemma 2.2.1. Now suppose $d \in \mathbb{R}^n$ satisfies the right hand side set of (i). Then by $0 \notin \partial f(x)$

$$f'_+(x,d) < 0 \text{ for some } d$$

This implies

$$f(x+td) < f(x) \text{ for } t > 0 \text{ small enough.}$$

But for $y := x + td \in Sf(x)$, our d is of the form $\dfrac{x+td-x}{t}$. Then we have

$$\{d \in \mathbb{R}^n : f'_+(x,d) < 0\} \subseteq \mathrm{R}^+[Sf(x) - x] \subseteq T_{Sf(x)}(x)) \quad (2.2.9)$$

Now by Proposition 2.2.4 a) with $g(\cdot) := f'_+(x,\cdot)$

$$cl\{d \in \mathbb{R}^n : f'_+(x,d) < 0\} = \{d \in \mathbb{R}^n : f'_+(x,d) \leq 0\}.$$

And then taking the closure of the sets in (2.2.9) gives

$$\{d \in \mathbb{R}^n : f'_+(x,d) \leq 0\} \subseteq T_{Sf(x)}(x).$$

Therefore,

$$T_{Sf(x)}(x) = \{d \in \mathbb{R}^n : f'_+(x,d) \leq 0\}. \quad (2.2.10)$$

To prove the second part of a), take the interior of both sides in (2.2.10) and use proposition 2.2.4 b) with $g(\cdot) := f'_+(x,\cdot)$.
b) Consider,

$$T_{Sf(x)}(x) = \{d \in \mathbb{R}^n : f'_+(x,d) \le 0\}.$$

For all $s \in \partial f(x)$ we have $\langle s,d \rangle \le f'_+(x,d) \le 0$ and hence

$$\begin{aligned} T_{Sf(x)}(x) &= \{d \in \mathbb{R}^n : \langle s,d \rangle \le 0 \text{ for all } s \in \partial f(x)\} \\ &= \{d \in \mathbb{R}^n : \langle \lambda s,d \rangle \le 0, \text{ for all } \lambda \ge 0 \text{ and for all } s \in \partial f(x)\} \\ &= [\,\mathrm{R}^+ \partial f(x)\,]^0. \end{aligned} \tag{2.2.11}$$

Since the normal cone is the polar of the tangent cone, taking the polar of both sides in (2.2.11) and using the fact that $\mathrm{R}^+ \partial f(x)$ is closed to obtain

$$N_{Sf(x)}(x) = cl[\,\mathrm{R}^+ \partial f(x)\,] = \mathrm{R}^+ \partial f(x).$$

This finishes the proof. //

Remark 2.2.3: The assumption $0 \notin \partial f(x)$ in Theorem 2.2.2 can be formulated in different ways. In view of Definition 2.1.2, it means $f'_+(x,d_0) < 0$ for some $d_0 \in \mathbb{R}^n$. On the other hand, in view of (2.1.7), there is some $x_0 \in \mathbb{R}^n$ such that $f(x_0) < f(x)$. Thus the existence of one point x with $0 \notin \partial f(x)$ allows the computation of the tangent and normal cone to the corresponding sublevel-set $Sf(x)$ at all its points.

3. Some Properties of Subdifferentials

3.1 First- Order Approximation of convex functions

Let f: $\mathbb{R}^n \to \mathbb{R}$ be convex and let $\varphi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$. Usually an approximation of a function f is performed with the aid of a function of two variables $\varphi(x,d)$ (here x being called a *point* while d is called a *direction* in $\mathbb{R}^n$). Speaking about approximation is usually has in mind that the difference of f at x: $f(x+td) - f(x) \approx t\varphi(x,d)$. The nature of this approximate equality and its accuracy depend on the type of approximation used and, of course on the complexity of the function under consideration.

One of the conditions frequently required from an approximation is the positive homogeneity with regard to the directions. In other words, an approximation $\varphi(x,d)$ must be such that

$$\varphi(x,td) = t\varphi(x,d) \quad \text{for all } t > 0$$

Example of such approximation is the directional derivative of f. By Remark 2.1.1, for a convex function f: $\mathbb{R}^n \to \mathbb{R}$, we have

$$f'_+(x,d) = \sup\{\langle s,d\rangle : s \in \partial f(x)\}.$$

From compactness of $\partial f(x)$, it follows that for any $d \in \mathbb{R}^n$ there is $s_d \in \partial f(x)$ such that

$$f(x+td) = f(x) + t\langle s_d, d\rangle + \varepsilon_d(t) \quad \text{for } t \geq 0. \tag{3.1.1}$$

Where $\varepsilon_d(\cdot) : X \to Y, \varepsilon_d \xrightarrow{t\to 0} 0$ (X, Y normed spaces) is called the generic function.

Now in (3.1.1) set $h := td$, then the difference $f(x+h) - f(x)$ can be approximated to first-order by a function $h\sigma(h) = \langle s_d, h\rangle$, which is sublinear. In general there exists a sublinear function σ_x defined on $\mathbb{R}^n$ such that

$$f(x+h) - f(x) = \sigma_x(h) + 0(\|h\|).$$

with $0(\|h\|)$ be define a function $h := \varphi(h)$ so that for each $\varepsilon > 0$ there is a $\delta > 0$ such that $\|h\| < \delta$ implies $\|\varphi(h)\| \leq \varepsilon\|h\|$.

Thus, a finite convex function enjoys a directional first-order approximation.

Lemma 3.1.1: Let f: $\mathbb{R}^n \to \mathbb{R}$ be convex and $x \in \mathbb{R}^n$. For each $\varepsilon > 0$, there exists $\delta > 0$ such that $\|h\| \leq \delta$ implies

$$|f(x+h) - f(x) - f'_+(x,h)| \leq \varepsilon\|h\|. \tag{3.1.2}$$

Proof: We now assume (indirect proof) that there exists $\varepsilon > 0$ and a sequence $\{h_k\}$ with $\|h_k\| =: t_k \leq \dfrac{1}{k}$ such that

$$|f(x+h_k) - f(x) - f'_+(x,h_k)| > \varepsilon t_k \quad \text{for } k \in \mathbb{N}$$

Now suppose that $\frac{h_k}{t_k} \xrightarrow{k\to\infty} d$ for some $d \in \mathbb{R}^n$ with $\|d\| = 1$. Then

$$\begin{aligned}\varepsilon t_k &< |f(x+h_k) - f(x) - f'_+(x,h_k)| \\ &= |f(x+h_k) - f(x+t_k d) + f(x+t_k d) - f(x) - f'_+(x+t_k d) + f'_+(x+t_k d) - f'_+(x,h_k)| \\ &\le |f(x+h_k) - f(x+t_k d)| + |f(x+t_k d) - f(x) - f'_+(x+t_k d)| + |f'_+(x+t_k d) - f'_+(x,h_k)|\end{aligned}$$

Because *f* is convex, *f* is locally Lipschitzian-continuous. So take a local Lipschitz constant L of *f* and then we have

$$\begin{aligned}\varepsilon t_k &< L\|h_k - t_k d\| + |f(x+t_k d) - f(x) - f'_+(x+t_k d)| + \|h_k - t_k d\| \\ &= 2L\|h_k - t_k d\| + |f(x+t_k d) - f(x) - f'_+(x+t_k d)|\end{aligned}$$

Dividing by $t_k > 0$ gives

$$\varepsilon < 2L\left\|\frac{h_k}{t_k} - d\right\| + \left|\frac{f(x+t_k) - f(x)}{t_k} - f'_+(x,d)\right|$$

Then taking the limit for $k \to \infty$ *i.e* $t_k \downarrow 0$, we get

$$\varepsilon < 2L\left\|\frac{h_k}{t_k} - d\right\| + \left|\frac{f(x+t_k) - f(x)}{t_k} - f'_+(x,d)\right| \xrightarrow{k\to\infty} 0$$

That means $\varepsilon \le 0$. But this is obviously a contradiction. //

Remark 3.1.1: We can also rewrite (3.1.2) as the first-order expansion. *i.e*

$$f(x+h) = f(x) + f'_+(x,h) + 0(\|h\|) \tag{3.1.3}$$

This shows that $f(x+h) - f(x)$ is approximated to first-order by a sublinear function $f'_+(x,\cdot)$.

Definition 3.1.1: Let C be a nonempty closed convex set. The set $F \subseteq C$ is called an exposed face of C if there is a supporting hyperplane of C such that $F = C \bigcap H_{d,r}$.

Let F be an exposed face of C and $H_{d,r}$ its associated supporting hyperplane. For any $s \in F$ by the definition we have

$\langle d, y\rangle \le \langle d, s\rangle$ for all $y \in C$

i.e $F := \{s \in C : \langle d, s\rangle = \sup\langle d, y\rangle$ for all $y \in C$.

Obviously $\sigma_C(d) := \sup\langle d, y\rangle$ for all $y \in C$ is the support function of C. Consequently for a nonempty closed convex set C, with support function σ_C and $x \neq 0$, the set

$$F_C(x) := \{ s \in C : \langle s, x\rangle = \sigma_C(x) \},$$

is called the *exposed face of C at x* or simply the *face exposed by x*.

Moreover, the face exposed by 0 is given by $F_C(0) = C$.

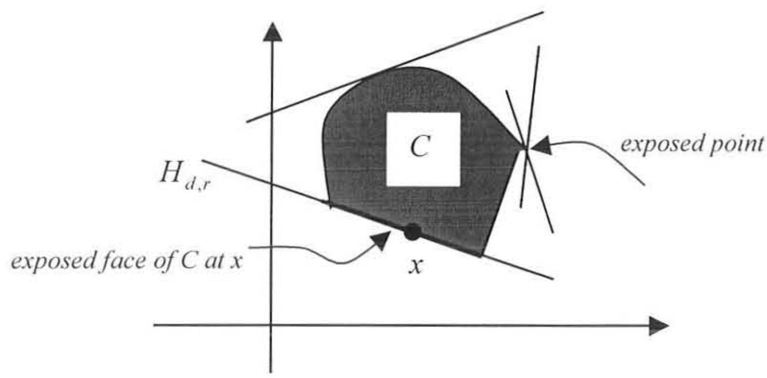

Figure 3.1.1 (supporting hyperplanes and exposed faces at various points)

Definition 3.1.2: Let C be a nonempty closed convex set in $\mathbb{R}^n$. The mapping P_C: $\mathbb{R}^n \to C$, which associates the unique solution $P_C(x)$, to each $x \in \mathbb{R}^n$ of the minimization problem:

$$\text{(P)} \qquad \inf\{\frac{1}{2}\|y-x\|^2 : y \in C\} \to \min. \tag{3.1.4}$$

is called a *projection operator*.

Theorem 3.1.2: A point $y_x \in C$ is the projection $P_C(x)$ if and only if

$$\langle x - y_x, y - y_x \rangle \leq 0 \text{ for all } y \in C \tag{3.1.5}$$

Proof: Suppose y_x is the solution of (3.1.4) and let $y \in C$ be arbitrary so that $y_x + \alpha(y - y_x) \in C$ for each $\alpha \in (0,1)$.

Then consider the function $f_x : \mathbb{R}^n \to \mathbb{R}$ defined by

$$f_x(y) := \frac{1}{2}\|y - x\|^2$$

which associates to $y \in \mathbb{R}^n$.

But from the supposition, since y_x is a solution of (3.1.4), we have

$$\begin{aligned} f_x(y_x) \leq f_x(y_x + \alpha(y - y_x)) &= \frac{1}{2}\|y_x - x + \alpha(y - y_x)\|^2 \\ &= \frac{1}{2}\langle y_x - x + \alpha(y - y_x), y_x - x + \alpha(y - y_x)\rangle \\ &= \frac{1}{2}[\|y_x - x\|^2 + 2\alpha\langle y_x - x, y - y_x\rangle + \alpha^2\|y - y_x\|^2] \\ &= \alpha\langle y_x - x, y - y_x\rangle + \frac{1}{2}\alpha^2\|y - y_x\|^2 \geq 0 \end{aligned}$$

Then dividing by $\alpha > 0$ and taking the limit for $\alpha \downarrow 0$ gives the result (3.1.5).

Conversely, assume that $y_x \in C$ satisfies (3.1.5). If $y_x = x$, then y_x is certainly solves (3.1.4). If not, for any $y \in C$ from *Cauchy-schwarz* inequality, we get

$$0 \geq \langle x - y_x, y - y_x \rangle = \langle x - y_x, y - x + x - y_x \rangle$$
$$= \|x - y_x\|^2 + \langle x - y_x, y - x \rangle \geq \|x - y_x\|^2 - \|x - y\|\|x - y_x\|.$$

Dividing by $\|x - y_x\| > 0$ gives

$$\|x - y_x\| \leq \|x - y\| \text{ for all } y \in C$$

which means y_x is a solution of (3.1.4) Therefore, the statement of the theorem holds. //

Geometrically, (3.1.5) expresses the fact that y_x is a vector in C such that the angle between $x - y_x$ and $y - y_x$ is obtuse. This comes from the definition of scalar product, for two nonzero vectors A and B:

$$\langle A, B \rangle = \|A\|\|B\| Cos\theta \leq 0$$

where θ is the angle between A and B.
Then $Cos\theta \in [-1,1]$ and hence θ is obtuse.

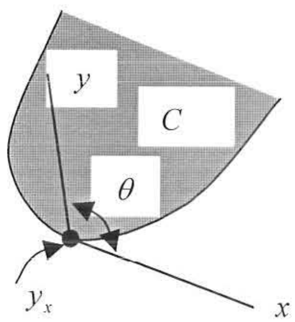

Figure 3.1.2. (The angle characterization of a *prpjection)*

Remark 3.1.2: Clearly for any $y \in C$, (3.1.5) implies

$$\langle x - P_C(x), y - P_C(x) \rangle \leq 0$$

i.e

$$\langle x - P_C(x), y \rangle \leq \langle x - P_C(x), P_C(x) \rangle$$

which means that $P_C(x) \in F_C(x - P_C(x))$. *i.e* $P_C(x)$ lies in the face of C exposed by $x - P_C(x)$.

Proposition 3.1.1: Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set. For $x \in C$ and $s \in \mathbb{R}^n$, the following properties are equivalent:

(a) $s \in N_C(x)$,

(b) $x \in F_C(s)$, *i.e* $\langle s, x \rangle = \max\{\langle s, y \rangle : y \in C\}$,

(c) $x = P_C(x + s)$.

Proof: To prove the proposition it suffices to show that (a)$\Rightarrow$(b)$\Rightarrow$(c)$\Rightarrow$(a). For this we often use the definition of normal cones, support hyperplanes, exposed faces, and projection.

(a)$\Rightarrow$(b): Let $s \in N_C(x)$, then $\langle s, y-x\rangle \leq 0$ for all y$\in C$. *i.e*

$$\langle s, y\rangle \leq \langle s, x\rangle \text{ for all } y \in C$$

That means

$$\langle s, x\rangle = \max\{\langle s, y\rangle : y \in C\}.$$

(b)$\Rightarrow$(c): Suppose $x \in F_C(s)$. *i.e* $\langle s, x\rangle = \max\{\langle s, y\rangle : y \in C\}$. Then we have

$$\langle s, x\rangle \geq \langle s, y\rangle \text{ for all } y \in C$$

But $\langle (x+s)-x, y-x\rangle = \langle s, y-x\rangle = \langle s, y\rangle - \langle s, x\rangle \leq 0$ for all $y \in C$

Therefore, $\langle (x+s)-x, y-x\rangle \leq 0$ for all $y \in C$ which implies that $x = P_C(x+s)$.

(c)$\Rightarrow$(a):Assume that (c) holds. Then $\langle (x+s)-x, y-x\rangle = \langle s, y-x\rangle \leq 0$ for all $y \in C$. But this implies that $s \in N_C(x)$. //

Now let us take $C := \partial f(x) \subseteq \mathbb{R}^n$ and compare (3.1.1) and (3.1.3) which can be rewritten with a subgradient s_d such that $\langle d, s_d\rangle \geq \langle d, s\rangle$ for all s$\in \partial f(x)$. Then by Proposition 3.1.1, s_d is an arbitrary element in $F_{\partial f(x)}(d)$. Equivalently $d \in N_{\partial f(x)}(s_d)$ and hence s_d is the projection of $s_d + d$ on to $\partial f(x)$. *i.e* $s_d = P_{\partial f(x)}(s_d + d)$. Thus the following corollary is just the same as Lemma 3.1.1.

Corollary 3.1.2: Let *f*: $\mathbb{R}^n \to \mathbb{R}$ be convex. If

$$s \in F_{\partial f(x)}(h) \Leftrightarrow h \in N_{\partial f(x)}(s) \Leftrightarrow s = P_{\partial f(x)}(s+h)$$

then at any *x*, we have

$$f(x+h) = f(x) + \langle s, h\rangle + 0(\|h\|).$$

As $\partial f(x)$ is compact, for any $h \in \mathbb{R}^n \setminus \{0\}$ the corresponding exposed face $F_{\partial f(x)}(h)$ is nonempty and covers the boundary of $\partial f(x)$. In particular for $\partial f(x)$ with only one exposed face, *i.e* only one element, there is some fixed s$\in \mathbb{R}^n$ such that

$$\lim_{t\downarrow 0} \frac{f(x+td)-f(x)}{t} = f'_+(x,d) = \langle s, d\rangle \text{ for all } d \in \mathbb{R}^n.$$

This implies

$$\left|\frac{f(x+td)-f(x)-\langle s, td\rangle}{t}\right| \xrightarrow{t\downarrow 0} 0. \tag{3.1.6}$$

which expresses precisely the *Gateaux*-differentiability of *f* at *x* for $h := td$, d fixed in $\mathbb{R}^n$, t$\downarrow$0. On the other hand, for $h \to 0$ (by Remark 1.3.1), (3.1.6) is equivalent to:

$$f(x+h) - f(x) = \langle s, h\rangle + 0(\|h\|).$$

i.e *f* is *Frechet*-differentiable at x.

Then we have the following corollary

Corollary 3.1.3: If the convex function f is *Gateaux*-differentiable at x, then $\partial f(x)$ is exactly a singleton ($\partial f(x)=\{\nabla f(x)\}$). Conversely, if $\partial f(x)$ contains only one element say s ($\partial f(x)=\{s\}$), then f is *Frechet*-differentiable at x, with $\nabla f(x)=s$.

In general, for all $x\in\mathbb{R}^n$ the equality $\partial f(x)=\{\nabla f(x)\}$ is not necessarily true. Otherwise f is *Gateaux*-differentiable at all x.

Since for a convex function f $f'_+(x,\cdot)$ is convex (Proposition 2.1.1), it has subdifferentials (Proposition 2.1.2). These subdifferentials are precisely the exposed faces of $\partial f(x)$. Therefore, for the general case where $\partial f(x)$ is not singleton, we have another way of defining faces. This is formulated in the following proposition.

Proposition 3.1.2: Let f: $\mathbb{R}^n\to\mathbb{R}$ be convex. For all $x,\ d\in\mathbb{R}^n$, we have

$$F_{\partial f(x)}(d)=\partial\,[f'_+(x,\cdot)](d).$$

Proof: Let $s\in\partial f(x)$. Then because $f'_+(x,\cdot)$ is the support function of $\partial f(x)$, we have

$$f'_+(x,z)\ge\langle s,z\rangle \text{ for all } z\in\mathbb{R}^n.$$

Moreover, if $s\in F_{\partial f(x)}(d)$ *i.e* $\langle s,z\rangle=f'_+(x,z)$, then we get

$$f'_+(x,z)\ge\langle s,z\rangle+f'_+(x,d)-\langle s,d\rangle=f'_+(x,d)+\langle s,z-d\rangle \text{ for all } z\in\mathbb{R}^n.$$

But this implies that $s\in\partial\,[f'_+(x,\cdot)\,](d)$.
Therefore,

$$F_{\partial f(x)}(d)\subseteq\partial\,[f'_+(x,\cdot)\,](d). \tag{3.1.7}$$

Conversely, let $r\in\partial\,[f'_+(x,\cdot)\,](d)$. Then by (2.1.7) we have

$$f'_+(x,z)\ge f'_+(x,d)+\langle r,z-d\rangle \text{ for all } z\in\mathbb{R}^n. \tag{3.1.8}$$

Now put $y:=z-d$ and by subadditivity of $f'_+(x,\cdot)$.

$$\begin{aligned}f'_+(x,d)+f'_+(x,y)\ge f'_+(x,z)&\ge f'_+(x,d)+\langle r,z-d\rangle\\ &=f'_+(x,d)+\langle r,y\rangle \text{ for all } y\in\mathbb{R}^n\end{aligned}$$

So we have

$$f'_+(x,y)\ge\langle r,y\rangle \text{ for all } y\in\mathbb{R}^n \tag{3.1.9}$$

Hence $r\in\partial f(x)$.
On the other hand, if we choose $z=0$ in (3.1.8), we get

$$f'_+(x,d)\le\langle r,d\rangle \text{ for all } d\in\mathbb{R}^n \tag{3.1.10}$$

Consequently, from (3.1.9) and (3.1.10) we obtain $f'_+(x,d)=\langle r,d\rangle$.
This implies, $r\in F_{\partial f(x)}(d)$. *i.e*

$$\partial\,[f'_+(x,\cdot)\,](d)\subseteq F_{\partial f(x)}(d) \tag{3.1.11}$$

By combining (3.1.7) and (3.1.11) we get

$$F_{\partial f(x)}(d)=\partial\,[f'_+(x,\cdot)](d)\,.\,//$$

Particularly, the subdifferential of $f'_+(x,\cdot)$ at the point td for $t > 0$ does not depend on t, but when $t \downarrow 0$, $td \to 0$ Proposition 2.1.2 confirms that

$$\partial\,[\,f'_+(x,\cdot)\,](0) = \partial f(x).$$

The result illustrates as follows.

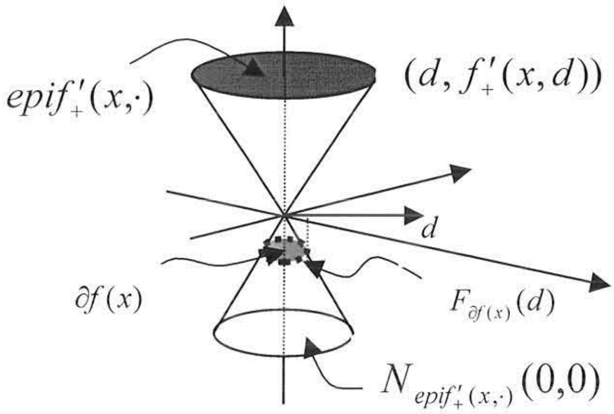

Figure 3.1.3 (Faces of subdifferentials)

Example 3.1.1: (cf. Example 2.1.2). Let C be a nonempty convex compact set, with support function σ_C. Then for $x \neq 0$ by Proposition 3.1.2 we have

$$\partial\sigma_C(x) = F_C(x) \text{ and } (\sigma_C)'_+(x,.) = \sigma_{F_C(x)}.$$

3.2 Conditions for a Minimum of Subdifferentiable Convex Functions

In (2.1.7), the linear form attached to a subdifferential defines an affine function that minimizes f. So (2.1.7) ensures x minimizes f if and only if the zero element belongs to the subdifferential of f at x.

Definition 3.2.1: A point $x_0 \in \mathbb{R}^n$ is called a *local minimum point* of the function f on the set $\mathbb{R}^n$, if there exists a $\delta > 0$ such that

$$f(x) \geq f(x_0) \quad \text{for all } x \in B(x_0, \delta) \cap \mathbb{R}^n \tag{3.2.1}$$

where $B(x_0,\delta) = \{x \in \mathbb{R}^n : \|x - x_0\| \leq \delta\}$. If, in addition, $\delta = \infty$, then the point x_0 is called a *global minimum point*. In this case (3.2.1) takes the form

$$f(x) \geq f(x_0) \quad \text{for all } x \in \mathbb{R}^n$$

Suppose a convex (hence locally Lipschitz) function f defined on $\mathbb{R}^n$ attains its local minimum at a point $x_0 \in \mathbb{R}^n$, then $f(x) \geq f(x_0)$ for all x sufficiently close to x_0. Now for any d and sufficiently small $t > 0$ the following inequality holds.

$$f(x_0 + td) - f(x_0) \geq 0$$

Hence

$$f'_+(x_0, d) = \lim_{t \downarrow 0} \frac{1}{t}[f(x_0 + td) - f(x_0)] \geq 0 \quad \text{for all } d \in \mathbb{R}^n.$$

This implies the sublinear function $d \mapsto f'_+(x_0, d)$ is nonnegative *i.e* $0 \leq f'_+(x_0, d)$ for all $d \in \mathbb{R}^n$. Therefore $0 \in \partial f(x_0)$. On the other hand, one can easily show that if $0 \in \partial f(x_0)$, then x_0 minimizes the function *f*. Hence to say *f* attains its smallest value on $\mathbb{R}^n$ at a point x_0 it is necessary and sufficient that

$$0 \in \partial f(x_0) \quad \text{or} \quad \max_{s \in \partial f(x_0)} \langle s, d \rangle \geq 0 \quad \text{for all } d \in \mathbb{R}^n. \tag{3.2.2}$$

As a result the problem of minimizing a convex function *f* at some point is equivalent to the problem of checking whether the subdifferential of this function at this point contains the origin or not.
Then we can summarize (3.2.2) in the following theorem.

Theorem 3.2.1: Let $f: \mathbb{R}^n \to \mathbb{R}$ be convex. Then the following properties are equivalent:

(a) f is minimized at $x \in \mathbb{R}^n$, *i.e* $f(y) \geq f(x)$ for all $y \in \mathbb{R}^n$,

(b) $0 \in f(x)$,

(c) $f'_+(x, d) \geq 0$ for all $d \in \mathbb{R}^n$.

Proof: Consider the case where f is minimized at x *i.e* $f(y) \geq f(x) + \langle 0, y - x \rangle$ for all $y \in \mathbb{R}^n$. Then by (2.1.7) we get $0 \in f(x)$.
To prove the second part of the theorem, let $0 \in f(x)$. Then by the property of scalar product and definition of subdifferential

$$0 = \langle 0, d \rangle \leq f'_+(x, d) \quad \text{for all } d \in \mathbb{R}^n$$

Therefore,

$$f'_+(x, d) \geq 0, \text{ for all } d \in \mathbb{R}^n$$

Finally assume $f'_+(x, d) \geq 0$ for all $d \in \mathbb{R}^n$. Then from the definition of directional derivative $f'_+(x, \cdot)$:

$$\lim_{t \downarrow 0} \frac{1}{t}[f(x + td) - f(x)] = \inf\{\frac{1}{t}[f(x + td) - f(x)] : t > 0\}.$$

But this means that

$$\frac{1}{t}[f(x + td) - f(x)] \geq 0 \text{ for } t > 0.$$

Now for any $d \in \mathbb{R}^n$ and $t \in \mathbb{R}^+$, set $y := x + td \in \mathbb{R}^n$ and then we have

$$f(y) - f(x) \geq 0 \text{ for all } y \in \mathbb{R}^n$$

i.e

$$f(y) \geq f(x) \text{ for all } y \in \mathbb{R}^n$$

Therefore

$$(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (a). \ //$$

When x is a local minimum of f, convexity implies that a local minimum is automatically global. A minimum point x is characterized by

$$0 \in \partial f(x) \text{ or by } f'_+(x, d) \geq 0 \text{ for all } d \in \mathbb{R}^n$$

While a non-minimal point x is characterized by the existence of an element $d \in \mathbb{R}^n$ with $f'_+(x, d) < 0$.

Proposition 3.2.1: A necessary and sufficient condition for the existence of $\varepsilon > 0$ such that

$$f(x+h) \geq f(x) + \varepsilon\|h\| \text{ for all } h \in \mathbb{R}^n$$

is $0 \in \operatorname{int} \partial f(x)$.

Proof: The condition $0 \in \operatorname{int} \partial f(x)$ implies that

$$B(0, \varepsilon) \subseteq \partial f(x) \text{ for some } \varepsilon > 0. \tag{3.2.3}$$

Then (3.2.3) can be expressed in terms of support function as

$$f'_+(x, \cdot) \geq \varepsilon\|\cdot\|. \tag{3.2.4}$$

From definition of directional derivative $f'_+(x, \cdot)$, for any $d \in \mathbb{R}^n$, (3.2.4) becomes

$$\lim_{t \downarrow 0} \frac{1}{t}[f(x+td) - f(x)] = \inf\{\frac{1}{t}[f(x+td) - f(x)] : t > 0\} \geq \varepsilon\|d\|.$$

But this implies,

$$\frac{1}{t}[f(x+td) - f(x)] \geq \varepsilon\|d\| \text{ for all } t > 0.$$

Now chose $h := td$, Then

$$f(x+td) - f(x) \geq t\varepsilon\|d\| \text{ for all } t > 0$$

or

$$f(x+h) - f(x) \geq \varepsilon\|h\|.$$

Therefore for any $d \in \mathbb{R}^n$, $h := td$ being arbitrary in $\mathbb{R}^n$ and the statement of the proposition holds. *i.e*

$$f(x+h) \geq f(x) + \varepsilon\|h\| \text{ for all } h \in \mathbb{R}^n. \text{ //}$$

3.3 Mean-Value Theorem

In this section, our aim is to answer some fundamental questions. For $x \neq y, x, y \in \mathbb{R}^n$ and for the subdifferential of a convex function $f: \mathbb{R}^n \to \mathbb{R}$ on the line segment (x, y) the questions are:

1. Can we evaluate $f(y) - f(x)$?
2. Is it possible to express f as the integral of its subdifferential?

Indeed, the problem reduces to that of one-dimensional convex function. To see this we consider the function

$$\varphi(t) := f(ty + (1-t)x) \text{ for all } t \in [0,1]. \tag{3.3.1}$$

Then $f(y)-f(x)=\varphi(1)-\varphi(0)$ is the trace of f on the line segment $[x,y]$. It is possible to express the subdifferential of φ at t in terms of the subdifferential of f at $ty+(1-t)x$ in the space $\mathbb{R}^n$.

Lemma 3.3.1: The subdifferential of a convex function defined in (3.3.1) is given by

$$\partial\varphi(t)=\{\langle s,y-x\rangle : s\in\partial f(x_t)\}.$$

or in a more symbolical form

$$\partial\varphi(t)=\langle\partial f(x_t),y-x\rangle, \tag{3.3.2}$$

where $x_t:=ty+(1-t)x$ for fixed $x,y\in\mathbb{R}^n$.

Proof: Since f is both convex and finite, it admits a finite left-sided derivative and a finite right-sided derivative (by Theorem 1.3.1) at any $x_0\in \operatorname{int} dom f$. Then

$$\begin{aligned}D_+\varphi(t)&=\lim_{t\downarrow 0}\frac{1}{\alpha}[f(x_t+\alpha(y-x))-f(x_t)]=f'_+(x_t,y-x)\\&=\max\{\langle s,y-x\rangle : s\in\partial f(x_t)\},\end{aligned}$$

and

$$\begin{aligned}D_-\varphi(t)&=\lim_{t\uparrow 0}\frac{1}{\alpha}[f(x_t+\alpha(y-x))-f(x_t)]=-f'_+(x_t,-(y-x))\\&=\min\{\langle s,y-x\rangle : s\in\partial f(x_t)\}\end{aligned}$$

So we have

$$\partial\varphi(t)=[D_-\varphi(t),D_+\varphi(t)]=\{\langle s,y-x\rangle : s\in\partial f(x_t)\}=\langle\partial f(x_t),y-x\rangle.\ //$$

Theorem 3.3.2: Let f: $\mathbb{R}^n\to\mathbb{R}$ be convex. For $x\neq y$, $x,y\in\mathbb{R}^n$, there exist $t\in(0,1)$ and $s\in\partial f(x_t)$ such that

$$f(y)-f(x)=\langle s,y-x\rangle$$

More generally,

$$f(y)-f(x)\in\bigcup_{t\in(0,1)}\{\langle\partial f(x_t),y-x\rangle\}..$$

where $x_t:=ty+(1-t)x$ for fixed $x,y\in\mathbb{R}^n$.

Proof: Let $\varphi(t):=f(ty+(1-t)x)$, $t\in[0,1]$ and consider the auxiliary function

$$g(t):=\varphi(t)-\varphi(0)-\frac{\varphi(1)-\varphi(0)}{1-0}(t-0)=\varphi(t)-\varphi(0)-(\varphi(1)-\varphi(0))t.$$

Then $g(0)=g(1)=0$ and g is continuous on $[0,1]$. So g is minimal at some $t_0\in(0,1)$ and we have

$$\begin{aligned}D_+g(t_0)&=\lim_{t\downarrow t_0}\frac{g(t)-g(t_0)}{t-t_0}=\lim_{t\downarrow t_0}[\frac{\varphi(t)-\varphi(t_0)}{t-t_0}-\frac{\varphi(1)-\varphi(0)}{1-0}]\\&=D_+\varphi(t_0)-(\varphi(1)-\varphi(0))\text{ for }t_0\in(0,1).\end{aligned}$$

Similarly

$$D_{-}g(t_0) = D_{-}\,\varphi(t_0) - (\varphi(1) - \varphi(0)) \text{ for } t_0 \in (0,1)$$

Then

$$\partial g(t_0) = [D_{-}g(t_0), D_{+}g(t_0)] = [D_{-}\,\varphi(t_0) - (\varphi(1) - \varphi(0)), D_{+}\,\varphi(t_0) - (\varphi(1) - \varphi(0))]$$
$$= \partial\varphi(t_0) - \{\varphi(1) - \varphi(0)\}$$

Since g is minimal at $t_0 \in (0,1)$ and $\varphi(1) - \varphi(0) = f(y) - f(x)$, by Theorem 3.2.1 we have $0 \in \partial g(t_0)$ and hence

$$f(y) - f(x) = \varphi(1) - \varphi(0) \in \partial\varphi(t_0) \text{ for } t_0 \in (0,1).$$

Then by Lemma 3.3.1 we have

$$f(y) - f(x) \in \{\langle s, y - x\rangle : s \in \partial f(x_{t_0})\}.$$

In other words,

$$f(y) - f(x) \in \bigcup_{t \in (0,1)} \{\langle s, y - x\rangle : s \in \partial f(x_t)\}. \ //$$

Now let $r(t) \in \partial\varphi(t)$ be an arbitrary selection. Then by Lemma 3.3.1 it takes of the form

$$r(t) := \langle s_t, y - x\rangle \text{ for } s_t \in \partial f(x_t).$$

This means, if $\{ s_t : t \in [0,1] \}$ is any selection of subgradients of f on the line segment $[x, y]$ *i.e* $s_t \in \partial f(x_t)$ for all $t \in [0,1]$, then $\langle s_t, y - x\rangle$ is also an arbitrary selection of subgradients of φ. Thus the integral $\int_0^1 \langle s_t, y - x\rangle dt$ is independent of the selection and its value is $\varphi(1) - \varphi(0)$, *i.e*

$$\varphi(1) - \varphi(0) = \int_0^1 \langle s_t, y - x\rangle dt = \int_0^1 \partial\varphi(t)dt = \int_0^1 \langle \partial f(x_t), y - x\rangle dt.$$

Therefore,

$$f(y) - f(x) = \int_0^1 \langle \partial f(x_t), y - x\rangle dt$$

Thus, the mean-Value Theorem can also be given in an integral form. *i.e* for a convex function $f: \mathbb{R}^n \to \mathbb{R}$ and for $x, y \in \mathbb{R}^n$, we have

$$f(y) - f(x) = \int_0^1 \langle \partial f(x_t), y - x\rangle dt$$

where $x_t := ty + (1-t)x$, $t \in (0,1)$.

4 Calculus Rules With Subdifferentials

When convex functions are combined so as to form a new convex function, their subdifferentials obey calculus rules resembling those of ordinary differential calculus. A difference, however, is that there are operations preserving convexity which do not preserve differentiability. If f is constructed from some other convex functions $f_j, j \in J$ the problem is to compute ∂f in terms of the ∂f_j.

4.1 Positive Combinations of Convex Functions

We start from some basic properties of support functions that are directly derived from their definition.

Proposition 4.1.1: Let $S \subseteq \mathbb{R}^n$ be nonempty. Then

$$\sigma_S = \sigma_{clS} = \sigma_{convS}$$

Furthermore,

$$\sigma_S = \sigma_{\overline{convS}} \tag{4.1.1}$$

Proof: By definition, for any $d \in \mathbb{R}^n$ we get

$$\sigma_S(d) = \sup\{\langle s,d\rangle : s \in S\} \le \sup\{\langle s,d\rangle : s \in cl\, S\}$$
$$\le \sup\{\langle s,d\rangle : s \in conv\, S\} \le \sup\{\langle s,d\rangle : s \in \overline{conv}\, S\} = \sigma_{\overline{convS}}(d)$$

This implies that

$$\sigma_S(d) \le \sigma_{\overline{convS}}(d) \text{ for all } d \in \mathbb{R}^n.$$

Conversely, the continuity (respectively linear, hence convexity) of the function, $\langle s,\cdot\rangle$, which is maximized over S, implies that $\sigma_S = \sigma_{clS}$ (respectively $\sigma_S = \sigma_{convS}$). Thus we have

$$\sigma_S = \sigma_{clS} = \sigma_{convS}.$$

But we know that $\overline{convS} = cl(convS)$ (Proposition 1.1.1). Then we get

$$\sigma_{\overline{convS}} = \sigma_{cl(convS)} = \sigma_{convS} = \sigma_S$$

Therefore, (4.1.1) holds. //

Theorem 4.1.1:

a) Let σ_{S_1} and σ_{S_2} be the support function of the nonempty closed convex sets S_1 and S_2. If $t_1, t_2 > 0$, then

$$t_1\sigma_{S_1} + t_2\sigma_{S_2} \text{ is the support function of } cl(t_1S_1 + t_2S_2).$$

b) Let $\{\sigma_{S_j}\}_{j\in J}$ be the support function of the family of nonempty closed convex sets $\{S_j\}_{j\in J}$, then

$$\sup_{j\in J}\sigma_{S_j} \text{ is the support function of } \overline{conv}\{\bigcup S_j : j \in J\}.$$

Proof: a) Let $S:=cl(t_1S_1+t_2S_2)$ the closed convex set. By definition, the support function of S is

$$\sigma_S(d)=\sup\{\langle t_1s_1+t_2s_2,d\rangle : s_1\in S_1, s_2\in S_2\}.$$

But s_1 and s_2 run independently in their index set S_1 and S_2 with t_1 and t_2 are positive. So we have

$$\sigma_S(d)=t_1\sup\{\langle s,d\rangle : s\in S_1\}+t_2\sup\{\langle s,d\rangle : s\in S_2\}=t_1\sigma_{S_1}+t_2\sigma_{S_2}.$$

Therefore, $\sigma_S=t_1\sigma_{S_1}+t_2\sigma_{S_2}$ is the support function of S=$cl(t_1S_1+t_2S_2)$.

b) Put $S:=\bigcup_{j\in J}S_j$, then the support function of S is

$$\sigma_S(d)=\sup_{s\in\bigcup S_j}\langle s,d\rangle=\sup_{j\in J}[\sup_{s_j\in S_j}\langle s_j,d\rangle]=\sup_{j\in J}\sigma_j(d).$$

But since $\sigma_S=\sigma_{\overline{convS}}$ (Proposition 4.1.1), we conclude that

$\sigma_S=\sup_{j\in J}\sigma_j$ is the support function of $\overline{conv}\{\bigcup S_j : j\in J\}$. //

Proposition 4.1.2: Suppose $f_k:\mathbb{R}^n\to\mathbb{R}$ for $k=1,2,\cdots,m$ be a family of convex functions and $t_1,t_2,\cdots,t_m$ >0, then the function

$$f(x)=\sum_{j=1}^{m}t_jf_j(x)$$

is convex.

Proof: Let $x, y\in\mathbb{R}^n$ and $\lambda\in[0,1]$. Then by the convexity of f_j for $j=1_j,2_j,\cdots,m$ we have

$$\begin{aligned}f(\lambda x+(1-\lambda)y)&=\sum_{j=1}^{m}t_jf_j(\lambda x+(1-\lambda)y)\le\sum_{j=1}^{m}t_j[\lambda f_j(x)+(1-\lambda)f_j(y)]\\&\le\lambda\sum_{j=1}^{m}t_jf_j(x)+(1-\lambda)\sum_{j=1}^{m}t_jf_j(y)\\&=\lambda f(x)+(1-\lambda)f(y)\end{aligned}$$

So we have the convexity of f. //

Theorem 4.1.2: Let f and g be two convex functions from $\mathbb{R}^n$ to $\mathbb{R}$ and t_1, t_2 >0. Then

$$\partial(t_1f+t_2g)(x)=t_1\partial f(x)+t_2\partial g(x)\ \text{ for all } x\in\mathbb{R}^n.$$

Proof: Let $S:=t_1\partial f(x)+t_2\partial g(x)$. Then S is a compact convex set. By Theorem 4.1.1 a) its support function is

$$\sigma_S(\cdot)=t_1f'_+(x,\cdot)+t_2g'_+(x,\cdot) \tag{4.1.2}$$

On the other hand, t_1f+t_2g is convex (Proposition 4.1.2) and by definition, the support function of $\partial(t_1f+t_2g)(x)$ is the directional derivative $(t_1f+t_2g)'_+(x,\cdot)$. Then for any $d\in\mathbb{R}^n$

$$(t_1 f + t_2 g)'_+(x,d) = \lim_{\alpha \downarrow 0} \frac{1}{\alpha}[\, (t_1 f + t_2 g)(x+\alpha d) - (t_1 f + t_2 g)(x)\,]$$

$$= t_1 \lim_{\alpha \downarrow 0} \frac{1}{\alpha}[\, f(x+\alpha d) - f(x)\,] + t_2 \lim_{\alpha \downarrow 0} \frac{1}{\alpha}[\, g(x+\alpha d) - g(x)]$$

$$= t_1 f'_+(x,d) + t_2 g'_+(x,d) \text{ for all } d \in \mathbb{R}^n$$

which coincides with (4.1.2).Therefore, the two compact convex sets $\partial\,(t_1 f + t_2 g)(x)$ and $t_1 \partial f(x) + t_2 \partial g(x)$ have the same support function. Hence they are equal. //

In general, if $f_k : \mathbb{R}^n \to \mathbb{R}$ for $k = 1,2,\cdots,m$ be convex functions and $t_1, t_2, \cdots, t_m > 0$, then using induction on m, for the function $f(x) := \sum_{j=1}^{m} t_j f_j(x)$,

we have

$$\partial f(x) = \sum_{j=1}^{m} t_j \partial f_j(x).$$

Example 4.1.1: In Example 2.1.4 take $\alpha = 1$, $\beta = 2$, $a = 1$, $b = 3$ and let $f(x) = |x - a|$, $g(x) = |x - b|$. Then for any $x \in \mathbb{R}$ one can show easily that

$$\partial f(x) + 2\partial g(x) = \partial(f + 2g))(x) = \begin{cases} \{-3\}, & \text{if } x < 1 \\ \{-2\} + [-1,1], & \text{if } x = 1 \\ \{-1\}, & \text{if } x \in (1,3) \\ \{1\} + [-2,2], & \text{if } x = 3 \\ \{3\}, & \text{if } x > 3 \end{cases} = \begin{cases} \{-3\}, & \text{if } x < 1 \\ [-3,-1], & \text{if } x = 1 \\ \{-1\}, & \text{if } x \in (1,3) \\ [-1,3], & \text{if } x = 3 \\ \{3\}, & \text{if } x > 3 \end{cases}$$

Consequently, it follows that the family of subdifferentiable functions is not a linear pace.

Remark 4.1.1: (*i*) In particular in Theorem 4.1.2 if we choose $t_1 = t_2 = 1$, then we have

- $\partial(f + g) = \partial f + \partial g$
- $\partial(\lambda f) = \lambda \partial f$ for all $\lambda > 0$.

(*ii*) The sign of t_1 and t_2 in this theorem is important to obtain a convex resulting function. For this consider the following example.

Example 4.1.2: Let f be convex and even function (*i.e* a function such that $f(x) = f(-x)$) defined on $\mathbb{R}^n$. Let us show the right-sided directional derivative of this function at zero is also an even function. Indeed,

$$f'_+(0,d) = \lim_{t \downarrow 0} \frac{1}{t}[f(0+td) - f(0)] = \lim_{t \downarrow 0} \frac{1}{t}[f(0-td) - f(0)] = f'_+(0,-d)$$

But on the other hand,

$$f'_+(0,-d) = \lim_{t \downarrow 0} \frac{1}{t}[f(0+t(-d)) - f(0)] = \lim_{t \downarrow 0} \frac{1}{t}[f(0+td) - f(0)]$$

$$= \lim_{t \downarrow 0} -(\frac{1}{t}[-f(0+td) + f(0)]) = -(-f)'_+(0,d)$$

So we have

$$f'_{+}(0,d) = -(-f)'_{+}(0,d)$$

This means that

$$\partial f(0) + (-1)\partial f(0) = \partial f(0) + \partial(-f)(0) = 2\partial f(0). \qquad (4.1.3)$$

At the same time we get $\partial(f + (-f))(0) = \{0\}$.

In particular, if we choose $t_1 = 1,\ t_2 = -1$ and consider the functions

a) $f, g : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = g(x) = |x|$, then on the right hand side of (4.1.3) we have the set $[-2,2]$, yet $\partial(f + (-g))(0) = \{0\}$.

b) $f, g : \mathbb{R}^n \to \mathbb{R}$ defined by $f(x) = g(x) = \|x\|$, then we get $f - g \equiv 0$ and hence $\partial(f + (-g))(0) = \{0\}$. However, on the right hand side of (4.1.3) we obtain the set $B(0,1)$ where $B(0,1)$ is the unit ball centered at the origin with radius 1.

Clearly in this case,

$$\partial f(0) + (-1)\partial g(0) = 2\,B(0,1) \neq \{0\} = \partial(f + (-g))(0).$$

4.2 Pre-Composition of Convex Functions With an Affine Mapping

Proposition 4.2.1: Let f: $\mathbb{R}^n \to \mathbb{R}$ be convex and let A: $\mathbb{R}^m \to \mathbb{R}^n$ be an affine mapping such that $imA \bigcap domf \neq \emptyset$. Then the function

$$(f \circ A)(x) = f(A(x)) \text{ for } x \in \mathbb{R}^m$$

is convex.

Proof: Clearly $(f \circ A)(x) > -\infty$ for all x and there exists by assumption $y := A(x) \in \mathbb{R}^n$ such that $f(y) < +\infty$. Because all affine functions are convex, for $\alpha \in [0,1]$ and $x, y \in \mathbb{R}^m$ we have

$$A(\alpha x + (1-\alpha)y) = \alpha A(x) + (1-\alpha)A(y).$$

Then from this it follows that

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

This shows that $f \circ A$ is convex. //

Proposition 4.2.2: Let A: $\mathbb{R}^m \to \mathbb{R}^n$ be a linear operator, $\mathbb{R}^m$ being equipped with a scalar product $\langle \cdot,\cdot \rangle$ for which A^* is the adjoint operator of A. If $\mathbb{R}^n \supseteq S \neq \emptyset$, then we have

$$\sigma_{clA(S)}(y) = \sigma_S(A^* y) \text{ for all } y \in \mathbb{R}^n.$$

Proof: By definitions,

$$\sigma_{A(S)}(y) = \sup_{s \in S} \langle As, y \rangle = \sup_{s \in S} \langle s, A^* y \rangle = \sigma_S(A^* y).$$

Then apply Proposition 4.1.1 to obtain

$$\sigma_{clA(S)}(y) = \sigma_S(A^* y) \text{ for all } y \in \mathbb{R}^n. //$$

Theorem 4.2.1: Let A: $\mathbb{R}^m \to \mathbb{R}^n$ be an affine mapping defined by

$$Ax := A_0 x + b,$$

where A_0 be linear and $b \in \mathbb{R}^n$ and let g: $\mathbb{R}^n \to \mathbb{R}$ be convex. Then

$$\partial(g \circ A)(x) = A_0^* \partial g(Ax) \text{ for all } x \in \mathbb{R}^m.$$

Proof: Let $d \in \mathbb{R}^m$ be arbitrary. Then by definition of right-sided directional derivative

$$(g \circ A)'_+(x,d) = \lim_{t \downarrow 0} \frac{1}{t}[(g \circ A)(x+td) - (g \circ A)(x)]$$

$$= \lim_{t \downarrow 0} \frac{1}{t}[g(A(x+td) - g(A(x))]$$

But since A_0 is linear,

$$A(x+td) = A_0(x+td) + b = A_0 x + tA_0 d + b = Ax + tA_0 d.$$

Then

$$(g \circ A)'_+(x,d) = \lim_{t \downarrow 0} \frac{1}{t}[g(Ax + tA_0 d) - g(Ax)] = g'_+(Ax, A_0 d) \text{ for all } d \in \mathbb{R}^m.$$

From Proposition 4.2.2, $g'_+(Ax, A_0 d)$ is the support function of the convex compact set $A_0^* \partial g(Ax)$. Therefore, $\partial(g \circ A)(x)$ and $A_0^* \partial g(Ax)$ have the same support function and hence they are equal. //

Example 4.2.3: (cf. Lemma 3.3.1). Let $\varphi(t) = f(ty + (1-t)x)$ for all $t \in [0,1]$ and consider the affine mapping A: $\mathbb{R} \to \mathbb{R}^n$ defined by

$$At := x + t(y-x) \text{ for fixed } x, y \in \mathbb{R}^n$$

Then $A_0 t = t(y-x)$ and its adjoint is defined by

$$A_0^* s = \langle s, y-x \rangle \text{ for all } s \in \mathbb{R}^n.$$

So we have

$$\varphi(t) = f(ty + (1-t)x) = f(x + t(y-x)) = f(At) = (f \circ A)(t).$$

Now apply Theorem 4.2.1 with m=1, $x = t$ and $g = f$, to obtain the subdifferential $\partial\varphi$ of Lemma 3.3.1.

$$\partial\varphi(t) = \partial(f \circ A)(t) = A_0^* \partial f(At) \text{ for } x_t = At \in \mathbb{R}^n$$

$$= \{ A_0^* s : s \in \partial f(x_t) \}$$

$$= \{ \langle s, y-x \rangle : s \in \partial f(x_t) \}.$$

4.3 Post-composition With an Increasing Convex Function of Several Variables

Definition 4.3.1: Let $x = (x^1, x^2, \cdots, x^n)$, $y = (y^1, y^2, \cdots, y^n) \in \mathbb{R}^n$. A function f: $\mathbb{R}^n \to \mathbb{R}$ is said to be *increasing component wise* if $f(y) \geq f(x)$ whenever $y^j \geq x^j$ for $j = 1, 2, \cdots, n$.

Theorem 4.3.1:Let f: $\mathbb{R}^n \to \mathbb{R}$ be convex and g: $\mathbb{R} \to \mathbb{R}$ be convex and increasing. Assume that there is $x_0 \in \mathbb{R}^n$ such that $f(x_0) \in dom g$. Then

$$(g \circ f)(x) = g(f(x)) \text{ for all } x \in \mathbb{R}^n$$

is convex.

Proof: Let x, $y \in \mathbb{R}^n$ and $\lambda \in [0,1]$. Then by the convexity of f and g, for which g is increasing, it holds

$$(g \circ f)(\lambda x + (1-\lambda)y) = g(f(\lambda x + (1-\lambda)y)) \le g(\lambda f(x) + (1-\lambda) f(y))$$
$$\le \lambda\ g(f(x)) + (1-\lambda) g(f(y)) = \lambda\ (g \circ f)(x) + (1-\lambda)(g \circ f)(y).$$

From this it follows that $g \circ f$ is convex. //

Thus, post-composition of convex functions with an increasing one-dimensional convex function preserves convexity. In this section our particular interest is the subdifferential of the result. So we generalize the problem, by considering a vector-valued version of this operation.

Theorem 4.3.2: Let $f_k : \mathbb{R}^n \to \mathbb{R}$ for $k = 1,2,\cdots,m$ be a collection of convex functions and let a mapping F: $\mathbb{R}^n \to \mathbb{R}^m$ given by

$$F(x) := (f_1(x), f_2(x), \cdots, f_m(x)) \text{ for } x \in \mathbb{R}^n$$

with the scalar product on $\mathbb{R}^m$ and let g: $\mathbb{R}^m \to \mathbb{R}$ be convex and increasing component wise. Then for all $x \in \mathbb{R}^n$

$$\partial(g \circ F)(x) = \{\sum_{i=1}^{m} p^i s_i : (p^1, p^2, \cdots, p^m) \in \partial g(F(x)), s_i \in \partial f_i(x) \text{ for } i = 1,2,\cdots,m.$$

Proof: From Theorem 4.3.1, we know that $(g \circ F)(x) := g(f_1(x), f_2(x), \cdots, f_m(x))$ for $x \in \mathbb{R}^n$ is convex and if $(p^1, p^2, \cdots, p^m) \in \partial g(y)$, then each p^i is nonnegative. For this, let $\{e_1, e_2, \cdots, e_m\}$ be the canonical basis of $\mathbb{R}^m$ and

$$g(y) \ge g(y - e_j) \ge g(y) + \sum p^i (-e_j)^i = g(y) - p^j$$

This implies

$$p^j \ge 0 \text{ for } j = 1,2,\cdots,m$$

Let $S := \{\sum_{i=1}^{m} p^i s_i : (p^1, p^2, \cdots, p^m) \in \partial g(F(x)),\ s_i \in \partial f_i(x) \text{ for } i = 1,2,\cdots,m.$

First, we need to show that the convexity and compactness of S. Compactness (*i.e.* boundedness and closedness) are coming from the fact that a subdifferential ∂g or ∂f_i is bounded and closed. Then it remains to prove convexity of S.

Let $x, y \in S$ such that

$$x := \sum_{i=1}^{m} p^i s_i \text{ and } y := \sum_{i=1}^{m} p'^i \dot{s}_i$$

Then their convex combination is say s given by

$$s=\alpha\sum_{i=1}^{m}p^{i}s_{i}+(1-\alpha)\sum_{i=1}^{m}p'^{i}s'_{i}=\sum_{i=1}^{m}[\alpha p^{i}s_{i}+(1-\alpha)\ p'^{i}s'_{i})] \tag{4.3.1}$$

Now we claim that s∈S. Since each p^{i} and p'' is nonnegative and the sum in (4.3.1) can be restricted to those terms such that $p''^{i}:=\alpha\ p^{i}+(1-\alpha)p'^{i}>0$, each such term expressed as

$$p''^{i}[\frac{\alpha p^{i}}{p''^{i}}s_{i}+\frac{(1-\alpha)p'^{i}}{p''^{i}}s'_{i}]$$

But $p''^{i}\in\partial g(F(x))$ and $\frac{\alpha p^{i}}{p''^{i}}s_{i}+\frac{(1-\alpha)p'^{i}}{p''^{i}}s'_{i}\in\partial f_{i}(x)$.

Thus s∈S. *i.e* S is convex.

Next to this let us compute the support function σ_{S} of *S*. *i.e* for $d\in\mathbb{R}^{n}$ denote

$$F'_{+}(x,d):=(f'_{1+}(x,d),f'_{2+}(x,d),\cdots,f'_{m+}(x,d))\in\mathbb{R}^{m}$$

and we need to show that

$$\sigma_{S}(d)=g'_{+}(F(x),F'_{+}(x,d)). \tag{4.3.2}$$

Let $s=\sum_{i=1}^{m}p^{i}s_{i}\in S$ be arbitrary, then by the property of scalar product on $\mathbb{R}^{m}$.

$$\langle s,d\rangle=\langle\sum_{i=1}^{m}p^{i}s_{i},d\rangle=\sum_{i=1}^{m}p^{i}\langle s_{i},d\rangle$$

Since $p^{i}\geq 0$ and by definition of $f'_{i+}(x,\cdot)=\sigma_{\partial f_{i}(x)}$ we have

$$\sum_{i=1}^{m}p^{i}\langle s_{i},d\rangle\leq\sum_{i=1}^{m}p^{i}f'_{i+}(x,d)$$

Again by the definition $g'_{+}(F(x),\cdot)=\sigma_{\partial g(F(x))}$, we get

$$\sum_{i=1}^{m}p^{i}f'_{i+}(x,d)\leq g'_{+}(F(x),F'_{+}(x,d)).$$

Therefore,

$$\langle s,d\rangle\leq g'_{+}(F(x),F'_{+}(x,d)). \tag{4.3.3}$$

On the other hand, the compactness of $\partial g(F(x))$ implies that there exist an element $(\overline{p}^{1},\overline{p}^{2},\cdots,\overline{p}^{m})\in\partial g(F(x))$ such that

$$g'_{+}(F(x),F'_{+}(x,d))=\sum_{i=1}^{m}\overline{p}^{i}f'_{i+}(x,d).$$

Like wise from compactness of each $\partial f_{i}(x)$ we have an $\overline{s}_{i}\in\partial f_{i}(x)$ such that

$$f'_{i+}(x,d)=\langle\overline{s}_{i},d\rangle \text{ for } i=1,2,\cdots,m.$$

Then if we denote $\overline{s}=\sum_{i=1}^{m}\overline{p}^{i}\overline{s}_{i}\in S$, we have

$$g'_{+}(F(x),F'_{+}(x,d))=\sum_{i=1}^{m}\overline{p}^{i}f'_{i+}(x,d)=\sum_{i=1}^{m}\overline{p}^{i}\langle\overline{s}_{i},d\rangle=\langle\sum_{i=1}^{m}\overline{p}^{i}\overline{s}_{i},d\rangle=\langle\overline{s},d\rangle$$

This implies equality in (4.3.3) holds and hence our clime (4.3.2) establishes. *i.e* $g'_+(F(x), F'_+(x,d))$ is the support function of S.

Finally, we have to show that this support function is really the directional derivative $(g \circ F)'_+(x,d)$. Now by definition, for $t > 0$

$$(g \circ F)'_+(x,d) = \lim_{t \downarrow 0} \frac{1}{t}[g(F(x+td)) - g(F(x))]$$

Convexity of g implies that g is locally Lipschitzian, and hence we have

$$F(x+td) = F(x) + tF'_+(x,d) + 0(t)$$

and then

$$g(F(x+td)) = g(F(x)) + tF'_+(x,d) + 0(t)) = g(F(x)) + tF'_+(x,d)) + 0(t)$$
$$= g(F(x)) + t\, g'_+(F(x), F'_+(x,d)) + 0(t)$$

So we have

$$(g \circ F)'_+(x,d) = \lim_{t \downarrow 0} \frac{1}{t}[tg'_+(F(x), F'_+(x,d)) + 0(t)] = g'_+(F(x), F'_+(x,d)).$$

Because $g'_+(F(x), F'_+(x,d))$ is the support function of S and it is also the directional derivative of $g \circ F$ at x, we get

$$\partial(g \circ F)(x) = S. \ //$$

Corollary 4.3.3: Let $f_k : \mathbb{R}^n \to \mathbb{R}$ for $k = 1,2,\cdots,m$ be convex functions and define the function

$$f(x) := \max\{f_k(x) : k \in \{1,2,\cdots,m\}\}$$

If $I(x) := \{i : f_i(x) = f(x)\}$ denotes the active index set, Then

$$\partial f(x) = conv\{\bigcup \partial f_i(x) : i \in I(x)\}$$

Proof: Now take $g(y) := \max\{g_1(y), g_2(y), \cdots, g_m(y)\}$ with $g_i(y) = y^i, i = 1,2,\cdots,m$. and let $\{e_i\}$ be the canonical basis of $\mathbb{R}^m$. Then $g_i(y) = y^i = \langle e_i, y\rangle$. To compute the subdifferential ∂g at a given y, for $x \in \mathbb{R}^m$ consider

$$g(x) = g(y) + max\{-s_i + \langle e_i, x-y\rangle : i \in \{1,2,\cdots,m\}\} \tag{4.3.4}$$

where $s_i = g(y) - y^i \geq 0, \ i = 1,2,\cdots,m$

Now put $x := y + td$ for any $d \in \mathbb{R}^m$. For $t > 0$ small enough, those i such that $s_i > 0$ do not count and then set

$$I(x) := \{i : s_i = 0\} = \{i : y^i = g(y)\}.$$

Then (4.3.4) becomes

$$g(y+td) = g(y) + t\, max\{\langle e_i, d\rangle : i \in I(x)\} \text{ for small } t > 0.$$

This implies

$$\lim_{t \downarrow 0} \frac{1}{t}[g(y+td) - g(y)] = g'_+(y,d) = \max_{i \in I(x)} \langle e_i, d\rangle = \max\{\langle e_i, d\rangle : y^i = g(y)\}.$$

Apply Theorem 4.1.1 b) and then use (2.1.7) to get $\partial g(y) = conv\{e_i : i \in I(x)\}$. Then by Theorem 4.3.2, $\partial g(F(x))$ can be written as:

$$\{(p^1, p^2, \cdots, p^m) : p^i = 0,\ if\ i \notin I(x),\ p^i \geq 0,\ if\ i \in I(x),\ \sum_{i=1}^{m} p^i = 1\}$$

And hence

$$\partial f(x) = \{\sum_{i \in I(x)} p^i \partial f_i(x) : p^i \geq 0, \text{for } i \in I(x),\ \sum_{i \in I(x)} p^i = 1\}. \tag{4.3.5}$$

By definition of a convex hull, (4.3.5) becomes

$$\partial f(x) = conv\{\bigcup \partial f_i(x) : i \in I(x)\}. //$$

4.4 Supremum of Convex Functions

In this section, we study the calculus rule, generalizing the result in Corollary 4.3.3.When the directional derivative $f_+'(x,d)$ is upper semi-continuous in x for every d, then f is a subdifferentiable. Therefore it is clear to show that the sum and the maximum (or point wise supremum) of finite number of subdifferentiable functions are subdifferentiable. At the same time the family of subdifferentiable functions is not a linear space.

Proposition 4.4.1: Let $\{f_i\}_{i \in I}, f_i : \mathbb{R}^n \to \mathbb{R},\ i \in I$ be an arbitrary family of convex functions for some index set I. Moreover, let $\sup_{i \in I} f_i(x) < +\infty$, for all $x \in \mathbb{R}^n$. Then

$$f(x) := \sup_{i \in I} f_i(x)$$

is convex.

Proof: Here it suffices to prove that $epif$ is convex. For this end, we have to show that

$$epif = \bigcap_{i \in I} epif_i. \tag{4.4.1}$$

1. $[\subseteq]$: By definition,

$$\begin{aligned} epif &= \{(x,r) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq r\} = \{(x,r) \in \mathbb{R}^n \times \mathbb{R} : \sup_{i \in I} f_i(x) \leq r\} \\ &\subseteq \{(x,r) \in \mathbb{R}^n \times \mathbb{R} : f_i(x) \leq r, \forall i \in I\} = \bigcap_{i \in I} \{(x,r) \in \mathbb{R}^n \times \mathbb{R} : f_i(x) \leq r\} \\ &= \bigcap_{i \in I} epif_i. \end{aligned} \tag{4.4.2}$$

2. $[\supseteq]$: Let $(y,s) \in \bigcap_{i \in I} epif_i$, then $(y,s) \in epif_i$ for all $i \in I$. i.e $f_i(y) \leq s$ for all $i \in I$. From this it follows that

$$\sup_{i \in I} f_i(y) \leq s \text{ or } f(y) \leq s$$

This implies that $(y,s) \in epif$. So we have

$$\bigcap_{i \in I} epif_i \subseteq epif. \tag{4.4.3}$$

From (4.4.2) and (4.4.3) we establish (4.4.1) and this completes the proof. //

Theorem 4.4.1: Let S_1 and S_2 be nonempty closed convex sets whose support functions are σ_{S_1} and σ_{S_2}, respectively. Then

$$S_1 \subseteq S_2 \Leftrightarrow \sigma_{S_1}(d) \le \sigma_{S_2}(d) \text{ for all } d \in \mathbb{R}^n.$$

Proof: Let S_1 and S_2 as stated in the theorem, then

$$\begin{aligned} S_1 \subseteq S_2 &\Leftrightarrow s \in S_2 \text{ for all } s \in S_1 \\ &\Leftrightarrow \sigma_{S_2}(d) \ge \langle s, d \rangle \text{ for all } s \in S_1 \text{ and for all } d \in \mathbb{R}^n \\ &\Leftrightarrow \sigma_{S_2}(d) \ge \sup_{s \in S_1} \langle s, d \rangle \text{ for all } d \in \mathbb{R}^n \\ &\Leftrightarrow \sigma_{S_2}(d) \ge \sigma_{S_1}(d) \text{ for all } d \in \mathbb{R}^n. \text{ //} \end{aligned}$$

Lemma 4.4.2: Let I be an arbitrary index set, $\{f_i\}_{i \in I}$, $f_i : \mathbb{R}^n \to \mathbb{R}$, $i \in I$ be a collection of convex functions and assume that

$$f(x) := \sup\{f_i(x) : i \in I\} < +\infty \text{ for all } x \in \mathbb{R}^n$$

Moreover, let $I(x) := \{i \in I : f_i(x) = f(x)\}$ at a given x is the active index set (possibly empty). Then

$$\partial f(x) \supseteq \overline{conv}\{\bigcup \partial f_i(x) : i \in I(x)\}. \tag{4.4.4}$$

Proof: Let $i \in I(x)$ and let $s \in \partial f_i(x)$. Then by (2.1.7) we get

$$f(y) \ge f_i(y) \ge f_i(x) + \langle s, y - x \rangle \text{ for all } y \in \mathbb{R}^n.$$

So we have

$$\partial f_i(x) \subseteq f(x) \text{ for all } i \in I(x)$$

Therefore

$$\bigcup_{i \in I(x)} \partial f_i(x) \subseteq \partial f(x).$$

From closedness and compactness of subdifferentials, it follows that $\partial f(x)$ also contains the closed convex hull of $\{\bigcup \partial f_i(x) : i \in I(x)\}$. *i.e* (4.4.4) holds. //

Remark 4.4.1: The converse inclusion of (4.4.4) need not be true. Otherwise it requires some additional assumptions. For this consider the following instances.

Example 4.4.1: Let $f_i : \mathbb{R} \to \mathbb{R}$ defined by

$$f_0(x) \equiv 0, \ f_i(x) = x - i \text{ for } i \in I = (0,1]$$

Then

$$f(x) = \sup\{f_i(x) : i \in [0,1]\} = x^+ = \max\{x, 0\} = \begin{cases} 0, & if \ x < 0 \\ x, & if \ x \ge 0. \end{cases}$$

So the active index set,

$$I(x) = \{i \in I : f_i(x) = f(x)\} = \begin{cases} \{0\}, & if \ x \le 0 \\ \emptyset, & if \ x > 0. \end{cases}$$

Here the function $i \mapsto f_i(x)$ is upper semi-continuous at $x = 0$ only. Thus $I(0) = \{0\}$ yields

$$\{\bigcup \partial f_i(0) : i \in I(0)\} = \partial f_0(0) = \{0\}. \text{ Yet } \partial f(0) = [0,1].$$

Therefore,

$\partial f_0(0) = \{0\} \subseteq [0,1] = \partial f(0)$, However, the converse inclusion is not true.

Example 4.4.2: $g_j : \mathbb{R} \to \mathbb{R}$ defined by

$$g_0(x) \equiv 0, \; g_j(x) = x - \frac{1}{j} \text{ for } j \in \mathbb{N}$$

Now J:=$\mathbb{N}$ which is closed and the function $j \mapsto g_j(x)$ is upper semi-continuous for all $x \in \mathbb{R}$ but not bounded. Then

$$g(x) = \sup\{g_j(x) : j \in \mathbb{N}\} = x^+ = \max\{x,0\} = \begin{cases} 0, & if \;\; x < 0 \\ x, & if \;\; x \geq 0. \end{cases}$$

And

$$J(x) = \{j \in J : g_j(x) = g(x)\} = \begin{cases} \{0\}, & if \;\; x \leq 0 \\ \emptyset, & if \;\; x > 0. \end{cases}$$

So $J(x)$ yields the same result as in Example 4.4.1 at $x = 0$.

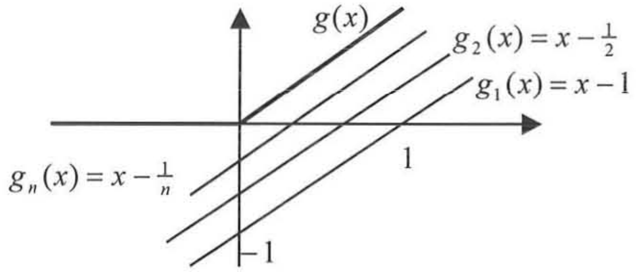

Figure 4.4.1 (supremum of functions)

From Example 4.4.1 and Example 4.4.2, we have seen that the three fundamental properties: the active index set J is closed, bounded and the functions $f_{(\cdot)}$ are upper semi-continuous are needed to prove the converse inclusion of (4.4.4). So together with these additional assumptions we have the following theorem.

Theorem 4.4.3: Let f_i, f, and I be given as in Lemma 4.4.2 and assume that I is a compact set (in some metric space), on which the functions $i \mapsto f_i(x)$ are upper semi-continuous for each $x \in \mathbb{R}^n$. Then

$$\partial f(x) = conv\{\bigcup \partial f_i(x) : i \in I(x)\}. \qquad (4.4.5)$$

Proof: Let $S := \{\bigcup \partial f_i(x) : i \in I(x)\}$. First, we need to establish the compactness of the set S. From our assumption, $I(x)$ is nonempty and compact. Because of (4.4.4), S is bounded. Then it remains to prove the closedness.

Let $\{s_k\}$ be a sequence in S with $s_k \xrightarrow{k\to\infty} s$ for each s_k and let $i_k \in I(x)$ such that $s_k \in \partial f_{ik}(x)$. *i.e*

$$f_{ik}(y) \geq f_{ik}(x) + \langle s_k, y - x\rangle \text{ for all } y\in\mathbb{R}^n.$$

By considering a subsequence so that $i_k \xrightarrow{k\to\infty} i \in I(x)$ we have $f_{ik}(x) \equiv f(x) = f_i(x)$ and from the upper semi-continuity of the function $f_{(\cdot)}(y)$, we get

$$f_i(y) \geq \limsup f_{ik}(y) \geq f_i(x) + \langle s, y - x\rangle \text{ for all } y\in\mathbb{R}^n.$$

This implies, $s \in \partial f_i(x) \subseteq S$ and therefore S is closed. Thus S is compact and hence its convex hull $convS$ is also compact.

So by Lemma 4.4.2 we have

$$conv\{\bigcup \partial f_i(x) : i \in I(x)\} \subseteq \partial f(x). \tag{4.4.6}$$

To prove the converse inclusion, it suffices to establish the corresponding inequality between support functions using calculus rules in Theorem 4.1.1 b). *i.e*

$$f'_+(x,d) \leq \sigma_S(d) = \sup\{f'_{i+}(x,d) : i \in I(x)\} \text{ for all } d\in\mathbb{R}^n \tag{4.4.7}$$

Let $\varepsilon > 0$ be arbitrary. From definition of $f'_+(x,d)$

$$\frac{1}{t}[f(x+td) - f(x)] > f'_+(x,d)\text{-}\varepsilon \;\; \forall\varepsilon > 0 \tag{4.4.8}$$

Then set

$$I_t := \{i \in I : \frac{1}{t}[f_i(x+td) - f(x)] \geq f'_+(x,d) - \varepsilon\}$$

From definition of $f(x+td)$ and (4.4.8), we see that $I_t \neq \emptyset$ and since I is compact and $f_{(\cdot)}(x+td)$ is upper semi-continuous, I_t is compact.

Now consider the function φ such that

$$\varphi(t) := \frac{1}{t}[f_i(x+td) - f_i(x)] + \frac{1}{t}[f_i(x) - f(x)] \text{ for all } t > 0.$$

Clearly $\frac{1}{t}[f_i(x+td) - f_i(x)]$ is the slope of a convex function and $\frac{1}{t}[f_i(x) - f(x)] \leq 0$, then it follows that φ is non-decreasing function whose supper level-set is I_t. Thus for $0 < t_1 \leq t_2$ we have $I_{t_1} \subseteq I_{t_2}$.

Because I_t is compact and $I_{t_1} \subseteq I_{t_2}$ with $0 < t_1 \leq t_2$ for each $\alpha \in (0,\alpha]$, there is some $i_\alpha \in I_\alpha$ such that the cluster point for $i_\alpha \in I_t$. *i.e* $i_\alpha \in I_\alpha \subseteq I_t$. So there exist some $i^* \in \bigcap_{t>0} I_t$ with

$$f_{i^*}(x+td) - f(x) \geq t[f'_+(x,d) - \varepsilon] \text{ for all } t > 0.$$

From continuity of the convex function f_{i^*} for $t \downarrow 0$ we have $i^* \in I(x)$. Then $f_{i^*}(x) = f(x)$ and then we get $\frac{1}{t}[f_{i^*}(x+td) - f_{i^*}(x)] \geq f'_+(x,d) - \varepsilon$ for all $t > 0$

From this it follows that

$$\lim_{t\downarrow 0}\frac{1}{t}[f_{i^*}(x+td)-f_{i^*}(x)]=f'_{i^*+}(x,d)\geq f'_+(x,d)-\varepsilon$$

But $\sigma_S(d)$ as defined in (4.4.7) is the support function of S. So we have

$$\sigma_S(d)\geq f'_{i^*+}(x,d)\geq f'_+(x,d)-\varepsilon\,. \tag{4.4.9}$$

Since $d\in\mathbb{R}^n$ and $\varepsilon>0$ were arbitrary, (4.4.9) becomes

$$\sigma_S(d)=\sup\{f'_{i+}(x,d):i\in I(x)\}\geq f'_+(x,d)$$

Then by Theorem 4.4.1, we get

$$\partial f(x)\subseteq S\subseteq convS\,. \tag{4.4.10}$$

Combining (4.4.6) and (4.4.10) gives (4.4.5) and hence the theorem is proved. //

When each f_i is differentiable, its only subgradient is its gradient ∇f_i. Thus the following corollary is a special case of Theorem 4.4.3.

Corollary 4.4.4: Assume that the notations and assumptions of Theorem 4.4.3 are satisfied and assume also that each f_i is differentiable, then

$$\partial f(x)=conv\{\nabla f_i(x):i\in I(x)\}\,.$$

For the proof apply first Corollary 3.1.2 and then Theorem 4.4.3.

Example 4.4.3: Let $f_k:\mathbb{R}^n\to\mathbb{R}$, for $k=1,2,\cdots,m$ be m differentiable convex functions defined by

$$f_i(x)=r_i+\langle s_i,x\rangle \text{ for } i=1,2,\cdots,m.$$

and let

$$f(x):=\max\{f_i(x):i=1,2,\cdots,m\}\,. \tag{4.4.11}$$

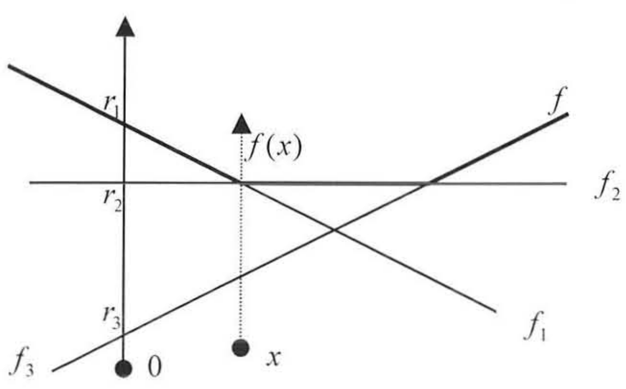

Figure 4.4.2 (maximum of differentiable functions)

Then from differentiability of each f_i for $i=1,2,\cdots,m$, we have

$$\partial f_i(x)=\{\nabla f_i(x)\}=\{s_i\},\ i=1,2,\cdots,m\,.$$

Since for $i \in \{1,2,\cdots,m\}$ those i satisfies $f(x) > f_i(x)$ do not count in the active index set at x, then the resulting active index set $I(x) = \{i : f_i(x) = f(x)\}$ is compact and the functions $f_{(\cdot)}(x)$ are obviously upper semi-continuous at $x \in \mathbb{R}^n$. Therefore, the supposition of Theorem 4.4.3 are satisfied and hence

$$\partial f(x) = conv\{\nabla f_i(x) : i \in I(x)\} = conv\{s_i : i \in I(x)\}. //$$

The above example illustrates the situation where we have finitely many differentiable convex functions; there are only finitely many active indices at x. In such a case $\partial f(x)$ is a compact convex polyhedron, generated by the active gradients at x.

4.5 Images of a Convex Function Under a Linear Mapping

Definition 4.5.1: Let A: $\mathbb{R}^m \rightarrow \mathbb{R}^n$ be linear and let g: $\mathbb{R}^m \rightarrow \mathbb{R} \cup \{\pm\infty\}$. The image of g under A is the function Ag: $\mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ defined by

$$(Ag)(x) := \inf\{g(y) : Ay = x\}. \tag{4.5.1}$$

(Note: convention, $\inf \emptyset = +\infty$)

Proposition 4.5.1: Suppose A: $\mathbb{R}^m \rightarrow \mathbb{R}^n$ be a linear mapping and C a convex set of $\mathbb{R}^m$. The image $A(C)$ of C under A is convex in $\mathbb{R}^n$.

Proof: Let $x, x' \in \mathbb{R}^m$. Then the image under A of the segment $[x, x']$ is clearly the segment $[A(x), A(x')] \subseteq \mathbb{R}^n$. So from convexity of C and linearity of A, the result follows. //

Now when $g = I_C$ (the indicator function on C), with C nonempty in $\mathbb{R}^m$, (4.5.1) rewritten as

$$(Ag)(x) := \begin{cases} 0, & if \;\; x = Ay, y \in C \\ +\infty, & othrerwise \end{cases}$$

In other words, $Ag := I_{A(C)}$ is the indicator function of the image of C under A, and by Proposition 4.5.1 this image is convex when C is convex.

Definition 4.5.2: Given a nonempty convex set $C \subseteq \mathbb{R}^n \times \mathbb{R}$. A function ℓ_C: $\mathbb{R}^n \rightarrow \mathbb{R}$ defined by

$$\ell_C(x) := \inf\{r \in \mathbb{R} : (x, r) \in C\}. \tag{4.5.2}$$

is called *lower-bound function* of C.

Theorem 4.5.1: Let C be a nonempty subset of $\mathbb{R}^n \times \mathbb{R}$ such that $\{r \in \mathbb{R} : (x, r) \in C\}$ is minorized for all $x \in \mathbb{R}^n$ and let its lower-bound function ℓ_C be defined by (4.5.2). Then ℓ_C is convex whenever C is convex.

Proof: Let $\varepsilon > 0$ be arbitrary, $\alpha \in (0,1)$ and $(x_1, r_1), (x_2, r_2) \in C$ such that

$$r_1 \le \ell_C (x_1) + \varepsilon \text{ and } r_2 \le \ell_C (x_2) + \varepsilon$$

From convexity of C, we have $(\alpha x_1 + (1-\alpha)x_2, \alpha r_1 + (1-\alpha)r_2) \in C$. Then it follows that

$$\ell_C (\alpha x_1 + (1-\alpha)x_2) \le \alpha r_1 + (1-\alpha)r_2 \le \alpha\, \ell_C (x_1) + (1\text{-}\alpha)\, \ell_C (x_2) + \varepsilon .$$

Since $\varepsilon > 0$ was arbitrary, the convexity of ℓ_C follows. //

Theorem 4.5.2: Let A: $\mathbb{R}^m \to \mathbb{R}^n$ be a surjective linear operator and g: $\mathbb{R}^n \to \mathbb{R}$ a convex function which is bounded from below on the inverse image

$$A^{-1}(x) := \{y \in \mathbb{R}^m: \ Ay = x\} \text{ for all } x \in \mathbb{R}^n.$$

Then Ag is convex on $\mathbb{R}^n$.

Proof: By supposition, Ag is nowhere $-\infty$; also $(Ag)(x) < +\infty$ whenever $x = Ay$, with $y \in dom g$. Now consider the extended operator $A' : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^m \times \mathbb{R}$, defined by

$$A'(y, r) := (Ay, r), \ y \in \mathbb{R}^m \text{ and } r \in \mathbb{R}.$$

Then the set $A'(epi g) =: C$ is convex in $\mathbb{R}^m \times \mathbb{R}$ and for given $x \in \mathbb{R}^n$, its lower-bound function is

$$\begin{aligned} \ell_C(x) &= \inf_r \{r \in \mathbb{R} : (x, r) \in C\} \\ &= \inf_{y,r} \{r \in \mathbb{R}: \ Ay = x, \ g(y) \le r\} \\ &= \inf_y \{ g(y) : \ Ay = x \}, \end{aligned}$$

and then Theorem 4.5.1 proves the convexity of $Ag := \ell_C$. //

Theorem 4.5.3: With all assumptions in Theorem 4.5.2. Assume also that x such that $Y(x)$ is nonempty where $Y(x)$ denotes the set of minimizers in (4.5.1) given by

$$Y(x) := \{y \in \mathbb{R}^m: \ Ay = x, \ g(y) = (Ag)(x)\}. \tag{4.5.3}$$

Then, for arbitrary $y \in Y(x)$ we have

$$\partial(Ag)(x) = \{s \in \mathbb{R}^n: \ A^* s \in \partial g(y)\} = (A^{*^{-1}})[\partial g(y)]. \tag{4.5.4}$$

Proof: Let $s \in \partial(Ag)(x)$. Then by (2.1.7 $s \in \partial(Ag)(x)$ if and only if

$$(Ag)(z) \ge (Ag)(x) + \langle s, z - x\rangle \ \text{ for all } z \in \mathbb{R}^n. \tag{4.5.5}$$

Since A is surjective and by the definition of Ag, for each $z \in \mathbb{R}^n$, $(Ag)(z) < +\infty$. Thus we can find $w \in \mathbb{R}^m$ such that

$$(Ag)(z) = g(w) \text{ and } Aw = z$$

Then (4.5.5) is equivalent to

$$\begin{aligned} g(w) &\ge g(y) + \langle s, Aw - Ay\rangle \ge g(y) + \langle s, A(w - y)\rangle \\ &\ge g(y) + \langle A^* s, w - y\rangle \ \text{ for all } \ w \in \mathbb{R}^m. \end{aligned}$$

So by (2.1.7) we get $A^* s \in \partial g(y)$. This completes the proof. //

Corollary 4.5.4: Define Ag and Y as in (4.5.1) and (4.5.3), Furthermore, if g is differentiable at some $y \in Y(x)$, then Ag is differentiable at x.

Proof: Because g is differentiable at some $y \in Y(x)$, by Corollary 3.1.2 we have $\partial g(y) = \{\nabla g(y)\}$. In (4.5.4), $A^* s \in \partial g(y)$ for some $s \in \partial(Ag)(x)$ implies that $A^* s = \nabla g(y)$. But from surjectivity of A, it follows that A^* is injective and conversely. Thus an equation $A^* s = \nabla g(y)$ in s has a unique solution, which means that $\partial(Ag)(x)=\{s\}$. Again by Corollary 3.1.2, Ag is differentiable at x with $\nabla(Ag)(x) = s$. //

A minimizer $y \in Y(x) \neq \emptyset$ is necessary to apply Theorem 4.5.3. This is illustrated in the following counter example.

Example 4.5.1: Let g: $\mathbb{R}^2 \to \mathbb{R}$ defined by $g(x,y) := \sqrt{x^2 + e^{2y}}$. Then g is convex. For this, we use the *Hesse-matrix*. Now for $(x,y) \in \mathbb{R}^2$ we get

$$\frac{\partial g(x,y)}{\partial x} = \frac{x}{\sqrt{x^2 + e^{2y}}} \qquad \frac{\partial g(x,y)}{\partial x^2} = \frac{e^{2y}}{\sqrt{(x^2 + e^{2y})^3}} \qquad \frac{\partial g(x,y)}{\partial y^2} = \frac{e^{2y}(2x^2+1)}{\sqrt{(x^2 + e^{2y})^3}}$$

$$\frac{\partial g(x,y)}{\partial y} = \frac{e^{2y}}{\sqrt{x^2 + e^{2y}}} \qquad \frac{\partial g(x,y)}{\partial x \partial y} = \frac{-xe^{2y}}{\sqrt{(x^2 + e^{2y})^3}} \qquad \frac{\partial g(x,y)}{\partial y \partial x} = \frac{-xe^{2y}}{\sqrt{(x^2 + e^{2y})^3}}$$

So our *Hesse-matrix* is

$$H(x,y) = \frac{e^{2y}}{\sqrt{(x^2 + e^{2y})^3}} \begin{pmatrix} 1 & -x \\ -x & 2x^2+1 \end{pmatrix}$$

with $a_{11} = \dfrac{e^{2y}}{\sqrt{(x^2 + e^{2y})^3}} > 0$ and $\det H(x,y) = \dfrac{(x^2+1)e^{2y}}{\sqrt{(x^2 + e^{2y})^3}} \geq 0$ for all $(x,y) \in \mathbb{R}^2$

Thus $H(x,y)$ is positive semi-definite and therefore g is convex.

Now consider the *marginal function f*, obtained by partial minimization of g:

$$f(x) := \inf\{g(y) : y \in \mathbb{R}\}.$$

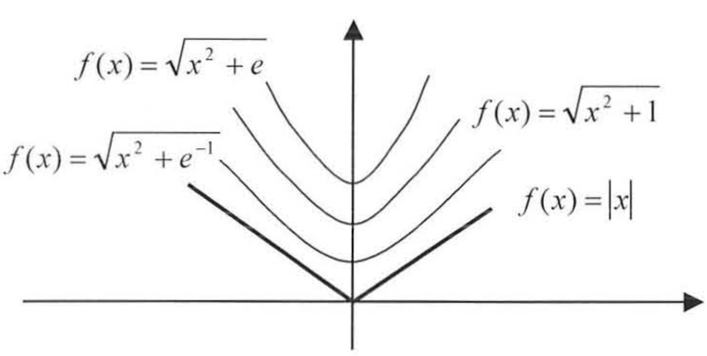

Figure 4.5.1 (Partial minimization of a function)

Then if we chose A: $\mathbb{R}^2 \to \mathbb{R}$ defined by $A(x, y) = x$, then f can be put under the form Ag. So the resulting marginal function is $f(x) = |x|$, which is not a smooth function, with subdifferential at $x = 0$:

$$\partial f(0) = \partial(Ag)(0) = [-1,1].$$

On the other hand, g is perfectly smooth and therefore $\partial g(x, y) = \nabla g(x, y)$ for all $(x, y) \in \mathbb{R}^2$, but minimal at *infinity* for all $x \in \mathbb{R}$. *i.e*

$$\begin{aligned} Y(x) &= \{(y_1, y_2) \in \mathbb{R}^2 \colon A(y_1, y_2) = x, g(y_1, y_2) = f(x)\} \\ &= \{(y_1, y_2) \in \mathbb{R}^2 \colon y_1 = x, \sqrt{y_1^2 + e^{2y_2}} = |x|\} \\ &= \{(y_1, y_2) \in \mathbb{R}^2 \colon \sqrt{x^2 + e^{2y_2}} = \sqrt{x^2}\} \\ &= \{(y_1, y_2) \in \mathbb{R}^2 \colon y_2 \to -\infty\} = \emptyset \text{ for all } x \in \mathbb{R}. \end{aligned}$$

In this case, the formula (4.5.4) is no of help.

5 The Subdifferential as a Multifunction

In the previous chapters we are mainly concerned with the behavior of $\partial f(x)$ for fixed x. Here, we study the behavior of this set varying with x and also with f.

Definition 5.0.1: Let $X\subseteq \mathbb{R}^n$ be nonempty. A mapping $F: X\rightarrow \mathbb{R}^n$ which associates $x\in X$ to a subset $F(x)$ of $\mathbb{R}^n$ is called be a *multi-valued*, or *set-valued* mapping, or more simply a *multifunction*.

The domain $domF$ of F is the set of $x\in X$ such that $F(x)\neq \emptyset$. Its image (or range) $F(X)$ and graph of F grF are the unions of the sets $F(x)\subseteq \mathbb{R}^n$ and $\{x\}\times F(x)\subseteq X\times \mathbb{R}^n$ respectively. Selection of F is a particular function f: $domF \rightarrow \mathbb{R}^n$ with $f(x)\in F(x)$ for all $x\in X$.

Example 5.0.1: The function F: $(0,\infty)\rightarrow \mathbb{R}$ given by $F(t) :=[0,\frac{1}{t})\subseteq \mathbb{R}$ for $t\in(0,\infty)$ is a multifunction.

5.1 Monotonicity Property of the Subdifferential

First we assume that f is differentiable on $C\subseteq \mathbb{R}^n$. Given $x_0\in C$, the sentence "*f is differentiable at x_0*" is meaningful only if f is at least defined in a neighborhood of x_0. Then it is clear to assume that C is contained in an open set Ω in which the function f is differentiable.

Theorem 5.1.1: Let f be a differentiable function on an open set $\Omega\subseteq \mathbb{R}^n$, and let C be a convex subset of Ω. Then

a) f is convex on C if and only if

$$f(x)\geq f(x_0)+\langle \nabla f(x_0), x-x_0\rangle \text{ for all } x, x_0\in C \tag{5.1.1}$$

b) f is strictly convex if and only if

$$f(x)> f(x_0)+\langle \nabla f(x_0), x-x_0\rangle \text{ for all } x\neq x_0 \text{ in } C \tag{5.1.2}$$

Proof: a) Let f be convex on C. For any $x, x_0\in C$ and $\alpha\in(0,1)$, we have

$$f(\alpha x+(1-\alpha)x_0)-f(x_0)\leq \alpha[f(x)-f(x_0)]$$

i.e

$$f(x_0+\alpha(x-x_0))-f(x_0)\leq \alpha[f(x)-f(x_0)] \tag{5.1.3}$$

Then divide (5.1.3) on both sides by α and taking the limit for $\alpha\downarrow 0$ to obtain

$$\langle \nabla f(x_0), x-x_0\rangle \leq f(x)-f(x_0)$$

and then (5.1.1) established.

Conversely, let $x_1, x_2\in C, \alpha\in(0,1)$ and define $x_0 :=\alpha x_1+(1-\alpha)x_2\in C$. Then by assumption,

$$f(x_i)\geq f(x_0)+\langle \nabla f(x_0), x_i-x_0\rangle \text{ for } i\in\{1,2\} \tag{5.1.4}$$

From convex combination, we get

$$\alpha f(x_1)+(1-\alpha)f(x_2)\geq f(x_0)+\langle\nabla f(x_0),\alpha x_1+(1-\alpha)x_2-x_0\rangle$$
$$=f(\alpha x_1+(1-\alpha)x_2),$$

which is just the definition of convexity.
b) Suppose f be strictly convex. Then for $x\neq x_0$ and $\alpha\in(0,1)$, we have

$$f(x_0+\alpha(x-x_0))-f(x_0)<\alpha[f(x)-f(x_0)] \tag{5.1.5}$$

But f is in particular convex. So by a) and (5.1.5), we get

$$\langle\nabla f(x_0),\alpha(x-x_0)\rangle\leq f(x_0+\alpha(x-x_0))-f(x_0)<\alpha[f(x)-f(x_0)].$$

Therefore, (5.1.2) follows after division by α.
For the converse, proceed as a) starting from strict inequality in (5.1.4). //

Thus, a differentiable function is convex when its graph lies above its tangent hyperplane. From the relation (5.1.1), for each x_0, f is minimized by its affine approximation

$$\sigma(x):=f(x_0)+\langle\nabla f(x_0),x-x_0\rangle,$$

which coincides with f at x_0. It is strict convex when the coincidence set is reduced to the singleton $(x_0,f(x_0))$.

Definition 5.1.2: Let $C\subseteq\mathbb{R}^n$ be convex. The mapping F: $C\to\mathbb{R}^n$ is said to be

a) *monotone* on C when for all $x, y\in C$
$$\langle F(x)-F(y),x-y\rangle\geq 0.$$
b) *strict monotone* on C when for all $x, y\in C$ and $x\neq y$
$$\langle F(x)-F(y),x-y\rangle>0$$

Theorem 5.1.2: Let f be a differentiable function on an open set $\Omega\subseteq\mathbb{R}^n$, and let C be a convex subset of Ω. Then

a) f is convex on C if and only if it's gradient mapping F: $=\nabla F$ is monotone on C.
b) f is strictly convex on C if and only if it's gradient mapping F: $=\nabla F$ is strictly monotone on C.

Proof: a) suppose f be convex on C, then by Theorem 5.1.1, for any $x, x_0\in C$

$$f(x)\geq f(x_0)+\langle\nabla f(x_0),x-x_0\rangle. \tag{5.1.6}$$
$$f(x_0)\geq f(x)+\langle\nabla f(x),x_0-x\rangle. \tag{5.1.7}$$

Adding (5.1.6) and (5.1.7) gives

$$\langle\nabla f(x_0)-\nabla f(x),x-x_0\rangle\leq 0$$

This implies

$$\langle\nabla f(x)-\nabla f(x_0),x-x_0\rangle\geq 0 \text{ for all } x,x_0\in C.$$

Conversely, let $x_0,x_1\in C$ and consider the function φ defined by $\varphi(t):=f(x_t)$ where $x_t:=x_0+t(x_1-x_0)$ for t in an open interval containing $[0,1], x_t\in\Omega$ and φ is well defined as well as differentiable; its derivative at t is $\varphi'(t)=\langle\nabla f(x_t),x_1-x_0\rangle$.
Thus for all $0\leq t'<t\leq 1$ we have

$$\varphi'(t)-\varphi'(t')=\langle \nabla f(x_t)-\nabla f(x_{t'}),x_1-x_0\rangle =\frac{1}{t-t'}\langle \nabla f(x_t)-\nabla f(x_{t'}),x_t-x_{t'}\rangle \qquad (5.1.8)$$

From monotonicity of ∇f

$$\varphi'(t)-\varphi'(t')=\langle \nabla f(x_t)-\nabla f(x_{t'}),x_t-x_{t'}\rangle \geq 0$$

which showed that φ' is increasing, and φ is therefore convex. So *f* is too.

b) ($\Rightarrow$): Proceed as a) using Theorem 5.1.1, starting from strict inequality in (5.1.6)& (5.1.7).
($\Leftarrow$): Set $t'=0$ in (5.1.8) and use the strict monotonicity relation to ∇f in order to found $\varphi'(t)-\varphi'(0)>0$.

Because the differentiable convex function φ is the integral of its derivative, we can write

$$\varphi'(t)-\varphi(0)-\varphi'(0)=\int_0^1[\varphi'(t)-\varphi'(0)]dt>0$$

This implies

$$\varphi(1)-\varphi(0)=f(x_1)-f(x_2)>\varphi'(0)=\langle \nabla f(x_0),x_1-x_0\rangle \text{ for all } x_1,x_0\in C.$$

Then by Theorem 5.1.1, *f* is strictly convex. //

Definition 5.1.3:Let *f* be a Lipschitz function on an open set $S\subseteq\mathbb{R}^n$ and $X\subseteq S$ be a compact set. The mapping $x\mapsto\partial f(x)$ is called the subdifferential mapping.

In Theorem 5.1.2, we have already seen that the gradient mapping of a differentiable convex function is monotone. In general, even in the absence of differentiability, the monotonicity property of the subdifferential mapping has its formulation.

Proposition 5.1.1: The subdifferential mapping $\partial f:\mathbb{R}^n\to\mathbb{R}^n$ is monotone when for all $x_1,x_2\in\mathbb{R}^n$,

$$\langle s_1-s_2,x_2-x_1\rangle\geq 0 \text{ for all } s_1\in\partial f(x_1) \text{ for all } s_2\in\partial f(x_2)$$

Proof: Let $x_1,x_2\in\mathbb{R}^n$ be arbitrary. Then by the subgradient inequality

$$f(x_1)\geq f(x_2)+\langle s_2,x_1-x_2\rangle \text{ for all } s_2\in\partial f(x_2). \qquad (5.1.9)$$

and

$$f(x_2)\geq f(x_1)+\langle s_1,x_2-x_1\rangle \text{ for all } s_1\in\partial f(x_1) \qquad (5.1.10)$$

Then adding (5.1.9) and (5.1.10) gives the result immediately. //

5.2 Continuity Property of the Subdifferential

In this section we study the continuity property of the set $\partial f(x)$. When *f* is differentiable convex function, its gradient ∇f is continuous as a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$. In the non-differentiable case, this gradient becomes a set ∂f. If *f* is nonsmooth (*i.e* for at least one *x* the set $\partial f(x)$ is not a singleton), then the mapping $\partial f(x)$ is not continuous. The concept of convergence in multifunction differs from that of single-valued case. The limit is going to be a set anyway.

Now consider the following two limit concepts.

Definition 5.2.1: Let $X\subseteq$ and consider a multifunction $F: X\to\mathbb{R}^n$

a) The limes exterior ($\lim ext.$) of $F(x)$ for $x\in X$ $x\to x_0$ is the set of all *cluster* or (accumulation) points of all selection (here $x_0\in cldomF$). *i.e* $y\in\lim\limits_{y\to x_0} ext\, F(x)$ means there exists a sequence $\{x_k, y_k\}$ such that $y_k\in F(x_k)$, $x_k \xrightarrow{k\to\infty} x_0$ and $y_k \xrightarrow{k\to\infty} y$

b) The limes interior ($\lim\text{int}.$) of $F(x)$ for $x\to x_0$ is the set of limits of all convergent selections. *i.e* $y\in\lim\limits_{y\to x_0}\text{int}\, F(x)$ means that $f(x)\in F(x)$ for all x and $f(x)\to$y when $x\to x_0$.

Example 5.2.1: (cf. Example 5.0.1). Then $\lim\limits_{t\downarrow 0} ext = \{0\} = \lim\limits_{t\downarrow 0}\text{int}\, F(t)$. Obviously, from Definition 5.2.1 we have $\lim\limits_{x\to x_0} ext\, F(x) \supseteq \lim\limits_{x\to x_0}\text{int}\, F(x)$. When these two sets are equal, the common set is the limit of $F(x)$ when $x\to x_0$.

Definition 5.2.2: The multifunction F is said to be:

a) *bounded-valued, closed-valued, convex-valued,* etc when the sets $F(x)$ are *bounded, closed, convex,* etc.

b) *locally bounded* near x_0 when

$$\begin{aligned}&\textit{for some neighborhood } N \textit{ of } x_0 \textit{ and}\\ &\textit{bounded set } B\subseteq\mathbb{R}^n, N\subseteq domF \textit{ and}\\ &F(N)\subseteq B.\end{aligned} \tag{5.2.1}$$

When F is locally bounded near every in a set S, we say that F is locally bounded on S. Then we have the following definition.

Definition 5.2.3: The multifunction F satisfying (5.2.1) is said to be:

a) *outer semi-continuous* at x_0 when

$$\lim_{x\to x_0} ext.F(x)\subseteq F(x_0).$$

b) *inner semi-continuous* at x_0 when

$$F(x_0)\subseteq \lim_{x\to x_0}\text{int}.F(x).$$

c) *continuous* when it is both outer and inner semi-continuous.

Proposition 5.2.1: A convex-compact-valued and locally bounded multifunction F from $\mathbb{R}^n$ to $\mathbb{R}^n$ is outer (respectively inner) semi-continuous at $x_0\in\text{int}\, domF$ if and only if its support function $x\to\sigma_{F(x)}(d)$ is upper (respectively lower) semi-continuous at x_0 for all $d\in\mathbb{R}^n$ of norm 1.

Proof: By calculus with support functions, our definition of outer semi-continuity is equivalent to

$$\forall \varepsilon > 0, \exists \delta > 0 : y \in B(x_0, \delta) \Rightarrow \sigma_{F(y)}(d) \le \sigma_{F(x_0)}(d) + \varepsilon\|d\| \text{ for all } d \in \mathbb{R}^n$$

and division by $\|d\|$ shows that the upper semi-continuity of the support function of F for $\|d\|=1$. Use the same proof for inner/lower semi-continuity. //

Thus a convex-compact-valued and locally bounded mapping F is both outer and inner semi-continuous at x_0 if and only if its support function $\sigma_{(\cdot)}(d)$ is continuous at x_0 for all d.

Proposition 5.2.2: Suppose f: $\mathbb{R}^n \to \mathbb{R}$ is convex. Then the graph of its subdifferential mapping is closed in $\mathbb{R}^n \times \mathbb{R}^n$. *i.e* the multi-valued mapping $x \mapsto \partial f(x)$ is closed.

Proof: Let $\{(x_k, s_k)\}$ be a sequence in $gr\partial f$ converging to $(x, s) \in gr\partial f$, then by definition, for all k we have

$$f(y) \ge f(x_k) + \langle s_k, y - x_k \rangle \text{ for all } y \in \mathbb{R}^n$$

Then from continuity of f and of the scalar product, taking the limit for $k \to \infty$ gives

$$f(y) \ge f(x) + \langle s, y - x \rangle \text{ for all } y \in \mathbb{R}^n.$$

Therefore $(x, s) \in gr\partial f$ and hence $gr\partial f$ is closed in $\mathbb{R}^n \times \mathbb{R}^n$. //

Proposition 5.2.3: The mapping ∂f is locally bounded. *i.e* the image $\partial f(B)$ of a bounded set $B \subseteq \mathbb{R}^n$ is a bounded set in $\mathbb{R}^n$.

Proof: Let $x \in B$ and $s \in \partial f(x) \setminus \{0\}$. Then in particular, for $y := x + \frac{s}{\|s\|}$ the subgradient inequality implies

$$f(x + \frac{s}{\|s\|}) \ge f(x) + \langle s, \frac{s}{\|s\|} \rangle = f(x) + \|s\|. \tag{5.2.2}$$

On the other hand, f is Lipschitz-continuous on the bounded set $B + B(0.1)$. Then

$$|f(y) - f(x)| \le L\|y - x\| = L\left\|\frac{s}{\|s\|}\right\| = L, \text{ for some } L>0. \tag{5.2.3}$$

From (5.2.2) and (5.2.3) it follows that

$$f(x) + \|s\| \le f(y) \le f(x) + L$$

i.e

$$\|s\| \le L \text{ for some } L.$$

Therefore, the mapping ∂f is locally bounded. //

Theorem 5.2.1: The subdifferential mapping $x \mapsto \partial f(x)$ of a convex function f: $\mathbb{R}^n \to \mathbb{R}$ is outer semi-continuous at any $x \in \mathbb{R}^n$. *i.e*

$$\forall \varepsilon > 0, \exists \delta > 0 : y \in B(x, \delta) \Rightarrow \partial f(y) \subseteq \partial f(x) + B(0, \varepsilon).$$

Proof: Suppose not! Then at some x there are $\varepsilon > 0$ and a sequence $\{(x_k, s_k)\}$ with $x_k \xrightarrow{k\to\infty} x$ and

$$s_k \in \partial f(x_k),\ s_k \notin \partial f(x) + B(0,\varepsilon),\ k \in \mathbb{N} \tag{5.2.4}$$

By Proposition 5.2.3, the sequence $\{s_k\}$ is bounded and by Proposition 5.2.2, it has a subsequence converging to s, which is contained in $\partial f(x)$. But this is a contradiction because (5.2.4) implies

$$s \notin \partial f(x) + B(0, \frac{1}{2}\varepsilon).\ //$$

Corollary 5.2.2 Let f: $\mathbb{R}^n \to \mathbb{R}$ is convex. Then for each fixed d the function

a) $f'_+(\cdot, d)$ (*i.e* $x \mapsto f'_+(x,d)$) is upper semi-continuous at all $x \in \mathbb{R}^n$. *i.e*

$$f'_+(x,d) = \limsup_{y\to x} f'_+(x,d) \text{ for all } d \in \mathbb{R}^n.$$

Furthermore, the mapping $x \mapsto \partial f(x)$ is upper semi-continuous.

b) the function $x \mapsto f'_-(x,d)$ is lower semi-continuous.

Proof: a) By Theorem 5 2.1, the subdifferential mapping of f is outer semi-continuous at all $x \in \mathbb{R}^n$. Then applying Proposition 5.2.1 to get its support function is upper semi-continuous at all $x \in \mathbb{R}^n$. But its support function is $f'_+(\cdot, d)$. We follow similar argument for the proof of b). //

In this chapter, we have seen that all the previous results concerned the behavior of $\partial f(x)$ at varying with x. The next two results concerned this behavior when f varies as well.

Theorem 5.2.3: Let $\{f_k\}$ be a sequence of (finite) convex functions converging point wise to f: $\mathbb{R}^n \to \mathbb{R}$ and let $\{x_k\}$ converges to $x \in \mathbb{R}^n$. Then for each $\varepsilon > 0$,

$$\partial f_k(x_k) \subseteq \partial f(x) + B(0,\varepsilon) \text{ for k large enough.} \tag{5.2.5}$$

Proof: Let $\varepsilon > 0$ be given. The point wise convergence of sequence of convex functions $\{f_k\}$ to f imply that f is convex and the convergence is uniform on every compact set of $\mathbb{R}^n$. First, we show that boundedness of $\partial f_k(x_k)$. For this, let $s_k \in \partial f_k(x_k) \setminus \{0\}$ be arbitrary. Then from subgradient inequality we have

$$f_k(x_k + \frac{s_k}{\|s_k\|}) \geq f_k(x_k) + \|s_k\|.$$

The uniform convergence of $\{f_k\}$ to f on $B(x,2)$ implies

$$\|s_k\| \leq f_k(x_k + \frac{s_k}{\|s_k\|}) - f_k(x_k) + \varepsilon. \tag{5.2.6}$$

From the Lipschitz property of f on $B(x,2)$ together with (5.2.6) we get $\|s_k\| < L$ for some $L > 0$ and for k large enough. Therefore $\{s_k\}$ is bounded.

Now suppose (5.2.5) is not true. *i.e* for some infinite subsequence, there is some $s_k \in \partial f_k(x_k)$ but not in $\partial f(x) + B(0,\varepsilon)$. Then for arbitrary $y \in \mathbb{R}^n$, we have

$$f_k(y) \geq f_k(x_k) + \langle s_k, y - x_k \rangle$$

and then taking the limit on a further subsequence such that $s_k \xrightarrow{k\to\infty} s$; point wise (respectively uniform) convergence of $\{f_k\}$ to f at y (respectively around x), and continuity of the scalar product gives

$$f(y) \geq f(x) + \langle s, y - x \rangle.$$

Since y was arbitrary, we get $s \in \partial f(x)$. But this is a contradiction. //

Corollary 5.2.4: Let $\{f_k\}$ be a sequence of (finite) differentiable convex functions converging to the differentiable function f: $\mathbb{R}^n \to \mathbb{R}$. Then ∇f_k converges to ∇f uniformly on every compact set of $\mathbb{R}^n$.

Proof: Assume $S \subseteq \mathbb{R}^n$ be compact and suppose for contradiction that there exist $\varepsilon > 0$, $\{x_k\} \subset S$ such that

$$\|\nabla f_k(x_k) - \nabla f(x_k)\| > \varepsilon \text{ for } k \in \mathbb{N}$$

For $x_k \xrightarrow{k\to\infty} x \in S$, Theorem 5.2.3 implies that both $\{\nabla f_k(x_k)\}$ and $\{\nabla f(x_k)\}$ converges to $\nabla f(x)$. This leads to a contradiction $0 \geq \varepsilon$. //

All results discussed so far are in the case of finite dimension. Most of these results can also be extended to the infinite dimensions. Here we need to introduce the notion of $K-space$. We say that an ordered vector space is a $K-space$ if any bounded from above set of this space has the least upper bound (supremum). If a $K-space$, in addition, a Banach space with a monotone norm (*i.e* the inequality $|x| \leq |y|$ implies $\|x\| \leq \|y\|$), then it is called a Banach $K-space$. Then we consider mappings defined in Banach spaces and whose values belong to Banach $K-space$.

Let X and Y be Banach spaces, Y is a Banach $K-space$. Then for a mapping $K: X \to Y$ then as usual the directional derivative and sublinearity of the mapping K defined the same as that of functions in a finite case.

Now we define some terms in the above stated space. That is let X and Y be Banach spaces, Y is a Banach $K-space$ and let $A: X \to Y$, $P: X \to Y$ be an operator. Then

- A linear operator A is called a support operator to a sublinear operator P if $A(x) \leq P(x)$ for all $x \in X$
- The set of all support operators to P is said to be the subdifferential of P and is denoted by ∂P. Clearly for any sublinear operator P the subdifferential ∂P is nonempty. Obviously from the definition we have

$$P(x) = \max\{Ax :\ A \in \partial P\}.$$

Furthermore the following relations hold

$$\partial(P_1 + P_2) = \partial P_1 + \partial P_2$$
$$\partial(\lambda P) = \lambda \partial P \text{ for all } \lambda \geq 0$$

For a Banach space X and a Banach $K-space$ Y, if we consider the following denotations

$\mathcal{L}(X,Y)$ be the family of linear continuous operators $A: X \to Y$

$P(X,Y)$ be the family of sublinear continuous operators $P: X \to Y$

Then

$$P \in P(X,Y) \Rightarrow \partial P \subseteq \mathcal{L}(X,Y).$$

Therefore in this case the support set can be defined as follows:

- A set $U \subseteq \mathcal{L}(X,Y)$ is called a support set if there exists an operator $P \in P(X,Y)$ such that $U = \partial P$. The family of this support subsets of $\mathcal{L}(X,Y)$ denoted by $\widetilde{M}(X,Y)$.

Clearly in the case $Y = \mathbb{R}$ the mapping $\varphi: P \to \partial P$ (called *Minkowski duality*) is a one to one correspondence between $P(X,Y)$ and $\widetilde{M}(X,y)$ like that of finite sublinear functions and compact convex sets.

References

[1] Hiriart-Urruty, J.B\ Lemarechal, C: Convex Analysis and Minimization Algorithms Volume I Springer Verlag Berlin, Heidelberg, New York, 1993

[2] Deumlich, R: Optimization and Theory of Approximation Textbook for the lecture, Addis Ababa, 1997

[3] Stoer, J, Witzgall, Ch: Convexity and Optimization in finite Dimensions Springer Verlag Berlin, Heidelberg, New York, 1997

[4] Vladimir F. , Demyanov\ Alexander M. , Rubinov: Constructive Nonsmooth Analysis Frunkfurt am Main, Berlin, Bern, New York, Paris, Wien, Lang, 1995