SOME REMARKS ON

CONJUGACY IN CONVEX ANALYSIS

A GRADUATE SEMINAR REPORT

By Tesfaye Tolu

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

School of Graduate Studies,
Addis Ababa University
Addis Ababa, 2003

PREFACE

Conjugacy is a notion defined on a certain class of functions. It was introduced in 1953 by W. Fenchel (Cf. [1]). Conjugating a function has a convexifying effect on the function. Furthermore, conjugacy has an important role in the duality theory of optimization. This is one motivation, among others, for the study of conjugacy.

In this seminar report, important properties of conjugacy will be investigated. Among other things, rules of conjugation and differentiability of the conjugate function are considered. Conjugates of some particular functions are also discussed.

I would like to express my heartfelt gratitude to my advisor Prof. Dr. rer. nat. habil. R. Deumlich not only for his willingness for consultation whenever I needed, but also for his all – round cooperation and enthusiasm to make his students perfect mathematicians. I also feel obliged to acknowledge his resourcefulness and excellent lectures, which inspired me so much.

Many friends and colleagues had contributed positively in different ways for the successful accomplishment of my study. Among them are Birhanu Tekle of Addis Ababa Uniniversity, Temesgen Feyissa of Nekemte Teachers' College, and all members in the Mathematics Section of Defence University College. I am very grateful to all of them. Finally, I would like to thank my friends Dawit Bulcha and Seifu Geleta, who were kind enough to type part of this seminar report on the computer.

Tesfaye Tolu.

CONTENTS Page

1. THE CONJUGATE OF A FUNCTION

1.1. INTRODUCTION

Let

$$f : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\} \text{ such that } f \not\equiv +\infty,$$
$$\text{and there is an affine function minorizing } f \text{ on } \mathbf{R}^n. \tag{1.1.1}$$

Moreover, let dom f: = $\{x \mid f(x) < +\infty\}$.

Definition 1.1.1: The *conjugate* of a function f satisfying (1.1.1) is the function

$f^* : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ defined by

$$f^*(s) = \sup_{x \in \text{dom } f} \{ \langle s, x \rangle - f(x) \}.$$

Remark 1.1:

1. If $x \notin$ dom f, then $f(x) = +\infty$. In such a case, $\langle s,x \rangle - f(x) = -\infty$. Since $f \not\equiv +\infty$,

   it follows that $\langle s,x \rangle - f(x) > -\infty$, for some $x \in \mathbf{R}^n$. Thus we may write

   $$f^*(s) = \sup_{x \in \mathbf{R}^n} \{ \langle s, x \rangle - f(x) \}. \tag{1.1.2}$$

2. The mapping $F : f \mapsto f^*$ shall be called the *conjugacy operation* or the *Legendre-Fenchel transform.*

From (1.1.2), we have

Fenchel's Inequality : Given $s \in \mathbf{R}^n$,

$$f^*(s) + f(x) \geq \langle s, x \rangle \text{ for all } x \in \mathbf{R}^n.$$

Next we show that f^* satisfies (1.1.1) .

Let the function ℓ given by $\ell(x) = \langle s_0, x \rangle + r_0$ be an affine function minorizing f, by (1.1.1). Then $f(x) \geq \langle s_{0,} x \rangle + r_0$.

This implies that

$$f(x) - \langle s_0, x \rangle \geq r_0 \quad \text{for all } x \in \mathbf{R}^n .$$

Now,

$$- f^*(s_0) = - \sup_{x \in \mathbf{R}^n} \{ \langle s_0, x \rangle - f(x) \}$$

$$= \inf_{x \in \mathbf{R}^n} \{ f(x) - \langle s_0, x \rangle \} \geq r_0 .$$

So $f^*(s_0) \leq - r_0 < +\infty$. This implies that $f^* \not\equiv +\infty$.

By Fenchel's inequality, we have

$$f^*(s) \geq \langle x, s \rangle - f(x) , \quad \text{for all} \quad x, s \in \mathbf{R}^n .$$

Fixing x, we notice that f* is minorized by the *affine function* g given by

$$g(x) := \langle s, x \rangle - f(x) .$$

Consequently, f* satisfies (1.1.1).

From the above discussion, we observe that dom f is the set of slopes of all affine functions minorizing f; dom f is the set of slopes of all affine functions minorizing f*.

Theorem 1.1.1: Let f satisfy (1.1.1). Then $f^* \in \overline{\mathrm{Conv}}\, \mathbf{R}^n$.

Proof: We know $\overline{\mathrm{Conv}}\, \mathbf{R}^n = \{ g : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\} | \text{ g is closed and convex } \}$.
Moreover, we know that f* satisfies (1.1.1). Now for each fixed x, the function ℓ where $\ell(s) = \langle x, s \rangle - f(x)$ is an affine function. So the conjugate function f* is a supremum of affine functions each of which is closed and convex.
But a supremum of convex and closed functions is closed and convex.
Thus $f^* \in \overline{\mathrm{Conv}}\, \mathbf{R}^n$. ///

Example 1.1.1: Let $f : \mathbf{R}^n \to \mathbf{R}$ be defined by $f(x) = \frac{1}{2}\|x\|^2$. Then

$$f^*(s) = \sup_{x \in \mathbf{R}^n} \left\{ \langle s, x \rangle - \tfrac{1}{2}\|x\|^2 \right\}$$

$$= \sup_{x \in \mathbf{R}^n} \left\{ \langle s, x \rangle - \frac{1}{2}\langle x, x \rangle \right\} = \sup_{x \in \mathbf{R}^n} \left\{ \left\langle s - \frac{1}{2}x, x \right\rangle \right\}$$

$$= -\inf_{x \in \mathbf{R}^n} \left\{ \left\langle \frac{1}{2}x - s, x \right\rangle \right\}$$

Putting $h(x) := \left\langle \frac{1}{2}x - s, x \right\rangle$, we get $\nabla h(x) = x - s$.

Thus, $\nabla h(x) = 0$ if and only if $x=s$. Since the second derivative of h is positive definite, h is convex and hence $x=s$ is its minimum point.

Consequently,

$$f^*(s) = \langle s, s \rangle - \frac{1}{2}\|s\|^2 = \frac{1}{2}s^2.$$

This means that f is its own conjugate.

Definition 1.1.2: Let $C \subseteq \mathbf{R}^n$, $C \neq \emptyset$.

(i) The indicator function of C is $I_C : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ defined by

$$I_C(x) := \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise} \end{cases}.$$

(ii) The support function of C is $\sigma_C : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ defined by

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

Example 1.1.2: Let $C \subseteq \mathbf{R}^n$, $C \neq \emptyset$. Then

$$(I_C)^*(s) = \sup_{x \in dom\ I_C} \{\langle s, x \rangle - I_c(x)\} = \sup_{x \in C} \{\langle s, x \rangle\} = \sigma_C(s)\cdot$$

In particular,

$$(I_{\mathbf{R}^n})^*(s) = \sup_{x \in \mathbf{R}^n} \{\langle s, x \rangle\} = \begin{cases} 0 & \text{if } s = 0 \\ +\infty & \text{otherwise} \end{cases} = I_{\{0\}}(s).$$

1.2. INTERPRETATIONS OF THE CONJUGATE FUNCTION

Let f satisfy (1.1.1) and let

$$F := \{a_{s,r} : \mathbf{R}^n \to \mathbf{R} \mid a_{s,r}(x) = \langle s, x\rangle - r, \quad s \in \mathbf{R}^n, r \in \mathbf{R}\},$$

$$M := \{a_{s,r} \in F \mid a_{s,r}(x) \le f(x)\ \forall x \in \mathbf{R}^n\},$$

$$\mathrm{grf} := \{(x, y) \in \mathbf{R}^n \times \mathbf{R} \mid y = f(x)\},$$

and

$$H_{s,r} := \{x \in \mathbf{R}^n \mid \langle s, x\rangle = r, a_{s,r} \in M\}.$$

Now for all $(x,r) \in H_{s,r}$, we have $\langle (s,-1),(x,r)\rangle = \langle s,x\rangle - r = 0$. Thus, $(s,-1) \perp H_{s,r}$, for all $s \in \mathbf{R}^n$, and for all $r \in \mathbf{R}$. By definition, we observe that $H_{s,r}$ lies below grf. To get $f^*(s)$, we lift $H_{s,r}$ as much as possible, but subject to supporting grf (See Fig. 1). In other words we translate $H_{s,r}$ by the vector $(0,z) \in \mathbf{R}^n \times \mathbf{R}$ until it touches grf. In this process we get $\langle s,x\rangle - r = z$ at each level $z \in \mathbf{R}$.

Let $(x_0, f(x_0))$ be the point of intersection of $H_{s,r}$ and grf.

Then $z = \langle s, x_0\rangle - r = f(x_0)$, so that $r = \langle s, x_0\rangle - f(x_0)$.

By Fenchel's inequality, we have

$$f(x) \ge \langle s,x\rangle - f^*(s) \quad \text{for all } s \in \mathbf{R}^n \text{ and for all } x \in \mathbf{R}^n.$$

So for $s \in \mathrm{dom}\ f^*$, we can put r=f*(s), since the affine hyperplane $H_{s, f^*(s)}$ can be lifted.

This means that the affine function $a_{s,r}$ given by $a_{s,r}(x) = \langle s, x\rangle - f^*(s)$ produces the "best hyperplane"

$$H = \{(x,z) \in \mathbf{R}^n \times \mathbf{R} \mid \langle s, x\rangle - f^*(s) = z\}$$

which intersects grf at $(x_0, f(x_0))$. This is because H is the hyperplane obtained by translating $H_{s, f^*(s)} = \{x \in \mathbf{R}^n \mid \langle s, x\rangle - f^*(s) = 0\}$ by the vector $(0,z) \in \mathbf{R}^n \times \mathbf{R}$.

Now putting x=0 in H, we observe H intersects the vertical axis in the graph space at $(0, -f^*(s))$.

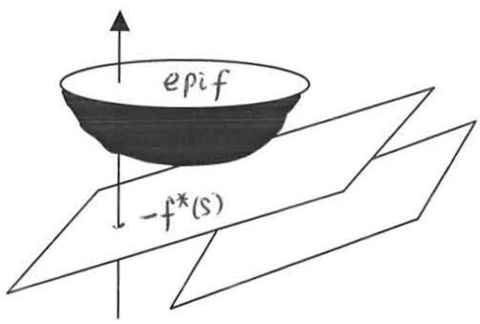

Fig.1.Geometrical Interpretations

Proposition 1.2.1: For all $s \in \mathbf{R}^n$ and for all u∈**R**, it holds that

(i) $f^*(s) = \sigma_{epi\ f}(s,-1);$

(ii) $$\sigma_{epi\ f}(s,-u) = \begin{cases} u\, f^*(\frac{1}{u}s), & if\ u > 0 \\ \sigma_{epi\ f}(s,0) = \sigma_{dom\ f}(s), & if\ u = 0 \\ +\infty, & if\ u < 0 \end{cases}.$$

Proof:

$$(i)\quad \sigma_{epi\ f}(s,-1) = \sup\{\langle (x,r),(s,-1)\rangle : (x,r) \in epi\ f\}$$

$$= \sup\{\langle s, x\rangle - r : (x,r) \in epi\ f\}$$

$$= \sup_{x \in \mathbf{R}^n} \sup_{r \geq f(x)} \{\langle s, x\rangle - r\} = \sup_{x \in \mathbf{R}^n} \{\langle s, x\rangle - f(x)\} = f^*(s)$$

(ii)Case1: u>0.

$$\sigma_{epi\ f}(s,-u) = \sigma_{epi\ f}(s,-1u) = \sigma_{epi\ f}(u\frac{s}{u},-1u)$$

$$= u\,\sigma_{epi\ f}(\frac{s}{u},-1)$$ since the support function is positive homogeneous.

$$= u f^* (\frac{s}{u}) \text{ by (i)} .$$

Case 2: u=0.

$$\sigma_{epif} (s,0) = \sup \{\langle (s,0), (x, r) \rangle \mid (x, r) \in epif \} = \sup \{\langle s, x\rangle - 0 \mid x \in \mathbf{R}^n \}$$
$$= \sup \{\langle s, x\rangle \mid x \in domf \} = \sigma_{domf} (s) .$$

Case 3: u<0.

$$\sigma_{epif} (s,-u) = \sup \{\langle (s,-u), (x,r)\rangle : (x,r) \in epif \} = \sup \{\langle s, x\rangle - ur : (x,r) \in epif \}$$
$$= \sup \{\langle s, x\rangle + |u| r : (x,r) \in epif \} = +\infty \quad ./\!/\!/$$

Definition 1.2.1: Let $f \in \overline{\text{Conv}}\,\mathbf{R}^n$. The *asymptotic function*, or recession function, or auto-deconvolution of *f* is the function $f'_\infty \in \overline{\text{Conv}}\,\mathbf{R}^n$ defined by

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

where $x_0 \in \text{dom } f$, arbitrary.

Proposition 1.2.2: For $f \in \overline{\text{Conv}}\,\mathbf{R}^n$, it holds that

$$\sigma_{domf} (s) = \sigma_{epif} (s,0) = (f^*)'_\infty (s) \text{ for all } s \in \mathbf{R}^n .$$

Proof: Easy consequence of the definition. ///

Now we consider the set $\text{epi } f \times \{-1\} \subseteq \mathbf{R}^n \times \mathbf{R} \times \mathbf{R}$, which is epi f translated by the vector (0, 0, -1) $\in \mathbf{R}^n \times \mathbf{R} \times \mathbf{R}$.

This set generates the cone

$$K_f := \{t(x, r, -1) \in \mathbf{R}^n \times \mathbf{R} \times \mathbf{R} \mid t > 0, (x, r) \in \text{epi } f \}.$$

The polar cone of K_f is

$$(K_f)^\circ := \{(s, \alpha, \beta) \in \mathbf{R}^n \times \mathbf{R} \times \mathbf{R} \mid \langle (s, \alpha, \beta), (tx, tr, -t)\rangle \le 0, \forall (x, r) \in \text{epi } f, t > 0 \}$$
$$= \{(s, \alpha, \beta) \in \mathbf{R}^n \times \mathbf{R} \times \mathbf{R} \mid \langle s, x\rangle + \alpha r \le \beta, \forall (x, r) \in \text{epi } f \}.$$

Putting $\alpha = 1$, we get

$$(\mathbf{R}^n \times \{-1\} \times \mathbf{R}) \cap (K_f)^\circ = \{(s,-1,\beta) \in \mathbf{R}^n \times \mathbf{R} \times \mathbf{R} \mid \langle s, x\rangle - r \le \beta, \forall (x,r) \in \text{epi} f\}$$

$$= \left\{(s,-1,\beta) \in \mathbf{R}^n \times \mathbf{R} \times \mathbf{R} \mid \sup_{(x,r)\in \text{epi } f} \{\langle s, x\rangle - r\} \le \beta\right\}$$

$$= \{(s,-1,\beta) \in \mathbf{R}^n \times \mathbf{R} \times \mathbf{R} \mid \sigma_{\text{epi } f}(s,-1) \le \beta\}$$

$$= \{(s,-1,\beta) \in \mathbf{R}^n \times \mathbf{R} \times \mathbf{R} \mid f^*(s) \le \beta\}$$ by Proposition 1.2.1(i).

Therefore,

$(\mathbf{R}^n \times \{-1\} \times \mathbf{R}) \cap (K_f)^\circ$ is epi f* translated by (0, -1,0).

Moreover,

$$K_f \cap (\mathbf{R}^n \times \mathbf{R} \times \{-1\}) = \{(x, r, -1) \in \mathbf{R}^n \times \mathbf{R} \times \{-1\} \mid (x, r) \in \text{epi} f\} \quad .$$

Hence we have

Remark 1.2: If $K_f := \{t(x,r,-1) \mid t > 0, (x,r) \in \text{epi} f\}$ and $(K_f)^\circ$ is the polar cone of K_f, then

(i) $K_f \cap (\mathbf{R}^n \times \mathbf{R} \times \{-1\}) = \text{epi } f$;

(ii) $(K_f)^\circ \cap (\mathbf{R}^n \times \{-1\} \times \mathbf{R}) = \text{epi} f^*$.

1.3. SIMPLE PROPERTIES OF CONJUGACY

A. Elementary Calculus Rules for Conjugacy

Direct application of the definition gives the properties in the following proposition:

Proposition 1.3.1: Let f, h_1, h_2 satisfy (1.1.1). Then

a) $g(x) := f(x) + \alpha, \alpha \in \mathbf{R}, \textit{implies} \quad g^*(s) = f^*(s) - \alpha;$

b) $g(x) := f(\alpha x),\ \alpha \in \mathbf{R},\ \alpha \neq 0,\ \textit{implies} \quad g^*(s) = f^*(\frac{s}{\alpha});$

c) $g(x) := \alpha f(x),\ \alpha > 0,\ \textit{implies} \quad g^*(s) = \alpha f^*(\frac{s}{\alpha});$

d) If $A : \mathbf{R}^n \to \mathbf{R}$ is an invertible linear operator and $(A^{-1})^*$ is the adjoint of the inverse of A, then

$$(f \circ A)^* = f^* \circ (A^{-1})^* ;$$

e) $g(x) := f(x) + \langle s_0, x \rangle$ *implies* $g^*(s) = f^*(s - s_0)$;

f) $g(x) := f(x - x_0)$ *implies* $g^*(s) = f^*(s) + \langle s, x_0 \rangle$;

g) $h_1 \le h_2$ *implies* $h_1^* \ge h_2^*$;

h) $(dom\ h_1) \cap (dom\ h_2) \ne \emptyset$ *and* $\alpha \in (0,1)$ *implies*

$$[\alpha h_1 + (1-\alpha) h_2]^* \le \alpha h_1^* + (1-\alpha) h_2^* ;$$

i) Let $f_j : \mathbf{R}^{n_j} \to \mathbf{R} \cup \{+\infty\}$, $j = 1,2,\ldots, m$, satisfy (1.1.1).

If $\mathbf{R}^n := \mathbf{R}^{n_1} \times \mathbf{R}^{n_2} \times \ldots \times \mathbf{R}^{n_m}$ and $h(x) := \sum_{j=1}^{m} f_j^*(s_j)$, then

$$h^*(s_1, s_2, \ldots, s_m) = \sum_{j=1}^{m} f_j^*(s_j)$$

for $\mathbf{R}^n$ equipped with scalar product of the product space.

Proposition 1.3.2: Let f satisfy (1.1.1), let H be a subspace of $\mathbf{R}^n$, and the operator P_H be the orthogonal projection onto H. Suppose there is $h \in H$ such that $f(h) < +\infty$.

Then (i) $f + I_H$ satisfies (1.1.1);

(ii) $(f + I_H)^* = (f \circ P_H)^* \circ P_H$.

Proof: (i) By hypothesis, there is $h \in H$ such that $f(h) < +\infty$.

Now, $(f+I_H)(h) = f(h) + I_H(H) = f(h) < +\infty$.So $f + I_H \not\equiv +\infty$. Moreover,

$f + I_H : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ is minorized by the affine functions which minorize f.

Consequently, $f + I_H$ satisfies (1.1.1).

(ii) $(f + I_H)^*(s) = \sup \{\langle s, x \rangle - (f + I_H)(x) \mid x \in H\}$

since $x \in H \Rightarrow \langle s, x \rangle - (f + I_H)(x) = -\infty$;

$= \sup \{\langle s, x\rangle - f(x) \mid x \in H\}$ since $I_H(x)=0$ for all $x \in H$;

$= \sup \{\langle s, P_H y\rangle - f(P_H y) \mid y \in \mathbf{R}^n\}$;

because for all $x \in H$, there is a $y \in \mathbf{R}^n$ such that $P_H y = x$;

$= \sup \{\langle P_H s, y\rangle - f(P_H y) \mid y \in \mathbf{R}^n\}$ since P_H is a symmetric operator;

$= (f \circ P_H)^* (P_H s)$ for all $s \in \mathbf{R}^n$.

Therefore, $(f + I_H)^* = (f \circ P_H)^* \circ P_H$. ///

Proposition 1.3.3: For f satisfying (1.1.1), let a subspace V contain the subspace parallel to *aff dom f* and set U: = $V^{\perp}$. For any $z \in \textit{aff dom } f$ and $s \in \mathbf{R}^n$ such that $s = s_u + s_v$, it holds

$$f^*(s) = \langle s_u, z\rangle + f^*(s_v).$$

Proof: Since V contains a subspace parallel to $\textit{aff domf}$, there is $z \in \mathbf{R}^n$ such that $z + V \supset \textit{aff domf}$. From this it follows that for all $x \in dom\, f$ there is $v \in V$ such that $x = z + v$. Now,

$$f^*(s) = \sup_{x \in domf} \{\langle s, x\rangle - f(x)\} = \sup_{v \in V} \{\langle s_u + s_v, z + v\rangle - f(z + v)\}$$

$$= \sup_{v \in V} \{\langle s_u, z\rangle + \langle s_u, v\rangle + \langle s_v, z + v\rangle - f(z + v)\}$$

$$= \langle s_u, z\rangle + \sup_{v \in V} \{\langle s_v, z + v\rangle - f(z + v)\}$$

$$= \langle s_u, z\rangle + f^*(s_v).$$ ///

B. The Biconjugate of a Function

Definition 1.3.1: The biconjugate of a function f on $\mathbf{R}^n$ satisfying (1.1.1) is the function f^{**} defined on $\mathbf{R}^n$ by

$$f^{**}(x) := (f^*)^*(x) = \sup_{s \in \mathbf{R}^n} \{\langle x, s\rangle - f^*(s)\}.$$

Theorem 1.3.1: Let f satisfy (1.1.1).

Then f** is the pointwise supremum of all affine functions on $\mathbf{R}^n$ majorized by f.

In other words,

$$\text{epi}(f^{**}) = \overline{\text{co}}\,(\text{epi } f)\,.$$

Proof: Let $E = \{(s,r) \in \mathbf{R}^n \times \mathbf{R} \mid a_{s,r}(x) := \langle x, s\rangle - r \le f(x)\}$.

Now, we have

$$(s,r) \in E \Leftrightarrow f(x) \ge \langle s, x\rangle - r \ \forall x \in \mathbf{R}^n;$$

$$\Leftrightarrow r \ge \langle s, x\rangle - f(x) \ \forall x \in \mathbf{R}^n;$$

$$\Leftrightarrow r \ge \sup\{\langle s, x\rangle - f(x) \mid x \in \mathbf{R}^n\} = f^*(s) \text{ and } s \in dom\ f^*.$$

Then for all $x \in \mathbf{R}^n$,we have

$$\overline{\text{co}}\, f(x) := \sup_{(s,r)\in E}\{\langle s, x\rangle - r\} = \sup\{\langle s, x\rangle - r \mid r \ge f^*(s)\}$$

$$= \sup\{\langle s, x\rangle - f^*(s) \mid s \in \text{dom} f^*\} = f^{**}(x)$$

Hence $\text{epi}(f^{**}) = \text{epi}(\overline{\text{co}} f) = \overline{\text{co}}\ \text{epi} f$. ///

Remark 1.3.1: From the above theorem, it follows that

$$f^{**} = \overline{\text{co}}\, f := \text{cl } f$$

Thus we may use the notation $\overline{\text{co}}\, f := f^{**}$ for f satisfying (1.1.1).

Corollary 1.3.1: If f and g are functions satisfying (1.1.1) such that

$\overline{\text{co}}\, f \le g \le f$, then $g^* = f^*$.

Moreover, $f = f^{**}$ iff $f \in \overline{\text{Conv}}\, \mathbf{R}^n$.

Proof:(i) Since $\overline{\text{co}} f \le g \le f$, it follows that $\overline{\text{co}}\, f \ge g^* \ge f^*$, by Proposition 1.3.1(g).

On the other hand,

$$g(x) \geq (\overline{\text{co}}\, f)(x) = f^{**}(x) = \sup_{s \in \mathbf{R}^n} \{\langle x, s\rangle - f^*(s)\} \text{ for all } x \in \mathbf{R}^n.$$

Then $\quad g(x) \geq \langle x, s\rangle - f^*(s)$ for all $s, x \in \mathbf{R}^n$.

So $\quad f^*(s) \geq \langle x, s\rangle - g(x)$ for all $s, x \in \mathbf{R}^n$.

This implies that $f^*(s) \geq \sup_{x \in \mathbf{R}^n} \{\langle s, x\rangle - g^*(x)\} = g^*(s)$.

So, $g^* \geq f^*$. Consequently, $f^* = g^*$.

(ii) $f = f^{**}$ iff $f = \overline{\text{co}}\, f \in \overline{\text{Conv}}\, \mathbf{R}^n$. ///

C. Conjugacy and Coercivity

Definition 1.3.2: A function f satisfying (1.1.1) is said to be

(a) $0-$*coercive* if and only if $\lim_{\|x\| \to +\infty} f(x) = +\infty$;

(b) $1-$*coercive* if and only if $\lim_{\|x\| \to +\infty} \dfrac{f(x)}{\|x\|} = +\infty$

Proposition 1.3.4: If f satisfying (1.1.1) is 1-coercive, then

$$f^*(s) < +\infty \quad \text{for all } s \in \mathbf{R}^n.$$

Proof: Suppose f is 1-coercive. Then

$$\lim_{\|x\| \to +\infty} \frac{f(x)}{\|x\|} = +\infty.$$

Thus, given $s \in \mathbf{R}^n$, there exists $R \in \mathbf{R}$ such that

$\|x\| \geq R$ implies $\dfrac{f(x)}{\|x\|} \geq \|s\| + 1$, which means

$$f(x) \geq \|x\|\|s\| + \|x\|, \quad \text{for all } x \in \mathbf{R}^n \text{ with } \|x\| \geq R;$$

$$\geq |\langle x, s\rangle| + \|x\| \quad \text{by Cauchy - Schwarz inequality.}$$

Then $\quad \langle s, x\rangle - f(x) \leq -\|x\| \quad$ for all $x \in \mathbf{R}^n$ with $\|x\| \geq R$.

Case 1: $\|x\| \geq R$. Then

$$\sup\{ \langle s, x\rangle - f(x) \mid \|x\| \geq R\} \leq \sup\{ -\|x\| : \|x\| \geq R\} \leq -R < +\infty .$$

Case 2: $\|x\| < R$. By (1.1.1), there exists $(s_0, r_0) \in \mathbf{R}^n \times \mathbf{R}$ such that

$$f(x) \geq \langle s_0, x\rangle - r_0 \quad \forall x \in \mathbf{R}^n .$$

This means that $-f(x) \leq -\langle s_0, x\rangle + r_0$ for all $x \in \mathbf{R}^n$.

So,

$$\begin{aligned}
\sup\{ \langle s, x\rangle - f(x) : \|x\| < R\} &\leq \sup\{ \langle s, x\rangle - \langle s, x_0\rangle + r_0 : \|x\| < R \} \\
&= \sup\{ \langle s - s_0, x\rangle + r_0 : \|x\| < R\} \\
&\leq \sup\{ \|s - s_0\| \|x\| + r_0 : \|x\| < R\} \\
&\qquad \text{by Cauchy- Schwarz inequality} \\
&< \|s - s_0\| R + r_0 < +\infty .
\end{aligned}$$

Consequently,

$$f^*(s) = \sup\{ \langle s, x\rangle - f(x) : x \in \mathbf{R}^n\} < +\infty , \text{ for all } s \in \mathbf{R}^n . ///$$

Proposition 1.3.5: Let f satisfy (1.1.1). Then

(i) $x_0 \in \text{int dom } f \Rightarrow f^* - \langle x_0, \cdot\rangle$ is $0-$coercive ;

(ii) $\text{dom } f = \mathbf{R}^n \Rightarrow f^*$ is $1-$coercive .

Proof: (i) By proposition 1.2.2, $\sigma_{\text{domf}} = (f^*)'_\infty$. Let now $x_0 \in \text{int domf}$.

(Clearly, $\text{int domf} \subset \text{int(co domf}$.))

Then $(f^*)'_\infty - \langle x_0, s\rangle = \sigma_{\text{domf}}(s) - \langle x_0, s\rangle > 0$ for all $s \neq 0$.

But $(f^* - \langle x_0, \cdot\rangle)'_\infty = (f^*)'_\infty - \langle x_0, \cdot\rangle$ by the definition asymptotic function.

So $(f^* - \langle x_0, \cdot\rangle)'_\infty$ is positive at each $s \neq 0$. But a function in $\overline{\text{Conv}}\, \mathbf{R}^n$ is 0-coersive if and only if its asymptotic function is positive at each nonzero point in $\mathbf{R}^n$. (Cf. [4].)

Since $f^* - \langle x_0, \cdot\rangle \in \overline{\text{Conv}}\, \mathbf{R}^n$, too, it follows that it is 0-coercive.

(ii) Since $\text{dom } f = \mathbf{R}^n$, every point $x \in \text{dom } f$ is an interior point.

By (i), $f^* - \langle x, \cdot \rangle$ is 0-coercive, which means

$$\lim_{\|s\| \to +\infty} [f^*(s) - \langle x, s \rangle] = +\infty > 0.$$

This implies that

$$\lim_{\|s\| \to +\infty} f^*(s) > \lim_{\|s\| \to +\infty} \langle x, s \rangle,$$

from which we have

$$\lim_{\|s\| \to +\infty} \frac{f^*(s)}{\|s\|} > \lim_{\|s\| \to +\infty} \left\langle x, \frac{s}{\|s\|} \right\rangle.$$

Since x is arbitrary, we have

$$\lim_{\|s\| \to +\infty} \frac{f^*(s)}{\|s\|} = +\infty .$$

Thus f* is 1- coercive. ///

Remark 1.3.2:

(a) From the second part of the above proof, it follows that

$$f^* - \langle x_0, . \rangle \text{ is } 0 - \text{coercive} \Rightarrow f^* \text{ is } 1 - \text{coercive} .$$

(b) If $f \in C\overline{\text{onv}}\ \mathbf{R}^n$, then we have

$$\begin{cases} (i)\ x_0 \in \text{int dom} f \Leftrightarrow f^* - \langle x_0, . \rangle \text{ is } 0 - \text{coercive} & ; \\ (ii)\ \text{dom} f = \mathbf{R}^n \Leftrightarrow f^* \text{ is } 1 - \text{coercive.} \end{cases}$$

1.4. SUBDIFFERENTIABILITY OF EXTENDED-VALUED FUNCTIONS

Definition 1.4.1: Let f be a function satisfying (1.1.1). The *subdifferential* of f at x is the set $\partial f(x) := \{ s \in \mathbf{R}^n \mid f(y) \geq f(x) + \langle s, y - x \rangle, \forall y \in \mathbf{R}^n \}$.

An element s of $\partial f(x)$ is called a *subgradient.*

Equivalently, the *subdifferential* can defined if *f is convex* as

$$\partial f(x) := \{ s \in \mathbf{R}^n \mid \langle s, d \rangle \leq f'(x, d) \quad \forall d \in \mathbf{R}^n \} .$$

Observe that $\partial f(x) = \varnothing$ if $x \notin \operatorname{dom} f$. To see this, take $y \in \operatorname{dom} f$.

Observe also that a subgradient $s \in \partial f(x)$ is the slope of an affine function minorizing f and coinciding with f at x.

Theorem 1.4.1: Let $s \in \mathbf{R}^n$ and let f satisfy (1.1.1). Then

$$s \in \partial f(x) \Leftrightarrow f^*(s) + f(x) - \langle s, x \rangle \leq 0 .$$

Proof: By definition,

$s \in \partial f(x)$ if and only if $\langle s, y \rangle - f(y) \leq -f(x) + \langle s, x \rangle$ for all $y \in \operatorname{dom} f$.
This is true if and only if

$$f^*(s) := \sup_{y \in \operatorname{dom} f} \{ \langle s, y \rangle - f(y) \} \leq -f(x) + \langle s, x \rangle. \quad ///$$

Remark 1.4.1: Combining Theorem 1.4.1 and Fenchel's inequality, we obtain the statement:

$$s \in \partial f(x) \Leftrightarrow f^*(s) + f(x) - \langle s, x \rangle = 0. \qquad (1.4.1)$$

From Theorem 1.4.1,it follows that

$$\partial f(x) = \{ s \in \mathbf{R}^n : f^*(s) - \langle s, x \rangle \leq -f(x) \} \qquad (1.4.2)$$

is the sublevel set of $f^* - \langle \cdot, x \rangle$ at level -f(x).

Since $f^* - \langle \cdot, x \rangle$ is convex and closed, $\partial f(x)$ is also convex and closed.

Theorem 1.4.2: For $f \in \overline{\text{Conv}}\ \mathbf{R}^n$, it holds that

$$x \in \text{ri dom } f \text{ implies that } \partial f(x) \neq \varnothing.$$

Proof: We know that any $f \in \overline{\text{Conv}}\ \mathbf{R}^n$ is minorized by some affine function ;

i.e. , for all $x \in ri\ dom\ f$, there is an $s \in aff\ (dom\ f) - x$ such that

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

[Notice that $aff\ (dom\ f) - x$ is the subspace parallel to $aff\ (dom\ f)$.]

This implies that $s \in \partial f(x)$. Consequently, $\partial f(x) \neq \varnothing$. ///

Proposition 1.4.1: Let f satisfy (1.1.1). Then

$$(i)\ \partial f(x) \neq \varnothing \Rightarrow \overline{co} f(x) = f(x);$$

$$(ii)\ [\overline{co} f \leq g \leq f \text{ and } g(x) = f(x)] \Rightarrow \partial f(x) = \partial g(x);$$

$$(iii)\ s \in \partial f(x) \Rightarrow x \in \partial f^*(s).$$

Proof: **(i)** Let $s \in \partial f(x)$. Then $f(y) \geq f(x) + \langle s, y - x \rangle$ *for all* $y \in \mathbf{R}^n$.

This implies that ℓ_s given by $\ell_s(y) := f(x) + \langle s, y - x \rangle$ is an affine function minorizing f. This in turn implies $\ell_s \leq \overline{\text{co}} f \leq f$ since

$$(\overline{\text{co}}\, f)(y) = \sup \left\{ \langle s, y \rangle - r \mid \langle s, z \rangle - r \leq f(z), \quad \forall z \in \mathbf{R}^n \right\}.$$

But $\ell_s(x) = f(x)$. So we have $\ell_s(x) = (\overline{co} f)(x) = f(x)$.

(ii) Let $(\overline{co} f) \leq g \leq f$ *and* $g(x) = f(x)$. Then f* = g* by Corollary 1.3.1.

Now by Remark1.4.1,

$$s \in \partial f(x) \Leftrightarrow f^*(s) + f(x) - \langle s, x \rangle = 0.$$

This is true if and only if $g^*(s) + g(x) - \langle s, x \rangle = 0$.

This holds if and only if $s \in \partial f(x)$. Thus, $\partial f(x) = \partial g(x)$.

(iii) Let $s \in \partial f(x)$. Then $(\overline{co} f)(x) = f(x)$ by (i).

But then,

$$f^{**}(x)=f(x) \text{ by Remark 1.3.1}$$

This implies that

$f^*(s)+(f^*)^*(x)-\langle s,x\rangle=0,$ since $f^*(s)+f(x)-\langle s,x\rangle=0$

by Remark 1.4.1.

Then $x \in \partial f^*(s)$ by Remark 1.4.1 again. ///

Remark 1.4.2: From Proposition 1.4.1(i) and (ii), we have:

(a) If $\partial f(x) \neq \varnothing$ for all $x \in \mathbf{R}^n$, then

$$f \in \overline{\text{Conv}}\,\mathbf{R}^n \text{ and f is finite - valued on } \mathbf{R}^n;$$

(b) If $\partial f(x) \neq \varnothing$, then $\partial f(x) = \partial(\overline{\text{co}}\, f)(x)$.

Corollary 1.4.1: Let $f \in \overline{\text{Conv}}\,\mathbf{R}^n$. Then

$$f(x) + f^*(s) - \langle s, x\rangle = 0 \Leftrightarrow s \in \partial f(x) \Leftrightarrow x \in \partial f^*(s).$$

Proof: Immediate from Remark 1.4.1, Proposition 1.4.1(iii) and the symmetric role of *f* and *f*** when $f \in \overline{Conv}\,\mathbf{R}^n$. ///

Let $s \in \partial f(x)$. Then $x \in \partial f^*(s)$, which implies that $s \in dom f^*$.

Thus

$$\partial f(x) \subset dom f^*.$$

Equality, however, need not hold. As a counter-example, take

$$f:\mathbf{R}\to\mathbf{R} \quad \textit{defined} \quad \textit{by} \quad f(x)=e^x.$$

Then $\partial f(x)=\{f'(x) \mid x\in\mathbf{R}\}=\{e^x \mid x\in\mathbf{R}\}$.

Now $f^*(0)=0$. So $0 \in \text{dom} f^*$, but $0 \notin \partial f = \{f'(x) \mid x \in \mathbf{R}\}$.

Remark 1.4.3: Let f satisfy (1.1.1). Then,

(i) x minimizes f $\Leftrightarrow 0 \in \partial f(x)$;

(ii) $x \in \partial f^*(0) \Leftrightarrow x \in \text{Argminf} := \left\{ x \in \mathbf{R}^n \mid f(x) = \inf_{y \in \mathbf{R}^n} f(y) \right\}$;

i.e., $\partial f^*(0) = \text{Argminf}$.

1.5. CONVEXIFICATION AND SUBDIFFERENTIABILITY

Let f satisfy (1.1.1). By Corollary 1.3.1, $f^* = (\overline{co}\, f)^*$.

Now,

$$\inf\{ f(x) \mid x \in \mathbf{R}^n \} = -f^*(0) = -(\overline{co}\, f)^*(0) = \inf\{(\overline{co}\, f)(x) \mid x \in \mathbf{R}^n\}.$$

Suppose x minimizes f. Then $0 \in \partial f(x)$ by Remark 1.4.3, which means $\partial f(x) \neq \varnothing$.

This implies by Remark 1.4.2(b) that $\partial f(x) = \partial(\overline{co}\, f)(x)$.

So $0 \in \partial(\overline{co}\, f)(x)$, which in turn implies that x minimizes $\overline{co}\, f$.

Consequently,

$$\text{Argmin} f \subset \text{Argmin}(\overline{co} f).$$

But Argmin $(\overline{co} f)$ is convex and closed as a sublevel set of the convex and closed function $\overline{co}\, f$ at the level $\inf \overline{co} f$.

Thus

$$\overline{co}\, \text{Argmin} f \subset \text{Argmin}(\overline{co} f).$$

This inclusion is strict as the above reasoning need not be reversible.

Proposition 1.5.1: Suppose f satisfying (1.1.1) is Gateaux-differentiable at x and $\partial f(x) \neq \varnothing$.

Then $\partial f(x) = \{\nabla f(x)\} = \partial(\overline{co} f)(x)$.

Proof: Since $\partial f(x) \neq \varnothing$, we have $\partial f(x) = \partial(\overline{co} f)(x)$ by Remark 1.4.2(b). Moreover, the convex function $\overline{co} f$, which is Gateaux-differentiable at x has only one subgradient, namely, $\nabla f(x)$.Consequently, $\partial f(x) = \{\nabla f(x)\} = \partial(\overline{co} f)(x)$. ///

Corollary 1.5.1: Let f satisfying (1.1.1) be Gateaux-differentiable on $\mathbf{R}^n$.

Then, x is a global minimum point of f on $\mathbf{R}^n$ if and only if

$$(i)\ \nabla f(x) = 0,$$

and $$(ii)\ (\overline{co} f)(x) = f(x)$$

In such a case, $\overline{co} f$ is differentiable at x, and $\nabla(\overline{co} f)(x) = 0$.

Proof: Suppose x is a global minimum point of f on $\mathbf{R}^n$. Then $0 \in \partial f(x)$ by Remark 1.4.3(i). So $\partial f(x) \neq \emptyset$. Now (ii) follows from Proposition 1.4.1(i). Since f is Gateaux-differentiable, it follows by Proposition 1.5.1 that $\nabla f(x) = 0$. So (i) holds. Conversely, suppose $x \in \mathbf{R}^n$ satisfies (i) and (ii). Then $0 = \nabla f(x) \in \partial f(x)$ by (i). Since f is Gateaux-differentiable, we have $\partial f(x) = \{\nabla f(x)\} = \partial(\overline{co} f)(x)$ by Proposition 1.5.1. So $0 \in \partial(\overline{co} f)(x)$. Then x minimizes $\overline{co} f$ by Remark 1.4.3(i). Hence x minimizes f on $\mathbf{R}^n$. This completes the proof. ///

Lemma 1.5.1: Let f satisfy (1.1.1), be lower semicontinuous and 1-coercive. Then *co (epif)* is a closed set.

For the proof Cf. [4].

Proposition 1.5.2: Let f satisfy (1.1.1), be lower semicontinuous and 1-coercive. Then

(i) $\mathrm{cof} = \overline{\mathrm{co}}\mathrm{f}$ (Hence $\mathrm{cof} \in \overline{\mathrm{Conv}}\,\mathbf{R}^n$.)

(ii) For all $x \in \mathrm{dom}\ f = \mathrm{co}\ \ \mathrm{dom}\ f$, there exist

$$x_j \in \mathrm{domf},\ \alpha_j \in \mathbf{R},\ \alpha_j \geq 0,\ j = 1,2,3,\cdots,n+1,\ \sum_{j=1}^{n+1} \alpha_j = 1$$

such that $$x = \sum_{j=1}^{n+1} \alpha_j x_j \quad \text{and} \quad (\overline{\mathrm{co}}\ f)(x) = \sum_{j=1}^{n+1} \alpha_j x_j\ .$$

Proof: **(i)** By Lemma 1.5.1, $co\,(epif) = \overline{co}\,(epif)$. By the definition of convex hull, we have that $epi(\overline{co}f) = \overline{co}(epif)$.

Now, $epi\,(cof\,) \subset epi\,(\overline{co}\,f) = \overline{co}\,(epif\,) = co\,(epif\,) \subset \overline{co}\,(epif\,)$.

Then it follows that $\text{epi}\ (\overline{\text{co}}\ \text{f}) = \text{epi}\ (\text{cof})$. Hence $\overline{\text{co}}\text{f} = \text{cof}$.

(ii) *co (epif)* is closed, by Lemma 1.5.1. Notice that the point (x, (cof)(x)) is a boundary point of *co (epi f)*. So (x, (cof)(x)) can be described as a convex combination of n+1 points of $epif = epi(\overline{co}f)$. That means there are

$$x_j \in \text{domf}\ , \alpha_j \in \mathbf{R}, \alpha_j \geq 0,\ j = 1,2,3,\cdots,n+1,\ \sum_{j=1}^{n+1} \alpha_j = 1$$

such that $(x,(cof)(x)) = \sum_{j=1}^{n+1} \alpha_j (x_j, r_j)$, where $f(x_j) \leq r_j$.

Now each $r_j = f(x_j)\ for\ \alpha_j > 0$, since $\sum_{j=1}^{n+1} \alpha_j (x_j, f(x_j)) \in co\ epif$.

Thus, $$(x,\ (cof)(x)) = \sum_{j=1}^{n+1} (\alpha_j x_j, \alpha_j f(x_j)) = (\sum_{j=1}^{n+1} \alpha_j x_j, \sum_{j=1}^{n+1} \alpha_j f(x_j)).///$$

Theorem 1.5.1: Let f satisfy (1.1.1). Suppose that for a given $x \in domf$, there is a family $\{(x_j, \alpha_j) \mid j = 1,2,3,\cdots,n+1\}$ satisfying Proposition 1.5.2(ii).

Let $J := \{j \mid \alpha_j > 0\}$.

Then $(i) f(x_j) = (cof)(x_j)\ for\ all\ j \in J$;

$(ii) cof$ is an affine function on the polyhedron $P := co\{x_j \mid j \in J\}$.

Proof: (i) Since cof is convex and minorizes f, we have that

$$(co\ f)(x) \leq \sum_{j=1}^{n+1} \alpha_j (co\ f)(x_j) \leq \sum_{j=1}^{n+1} \alpha_j f(x_j) = (co\ f)(x).$$

This implies that $\sum_{j=1}^{n+1} \alpha_j (\text{co f})(x_j) = \sum_{j=1}^{n+1} \alpha_j \text{f}(x_j)$.

Since $(\text{co f})(x_j) \leq \text{f}(x_j)$ for all x_j and $\alpha_j > 0$, for all $j \in J$, it follows that

$$\alpha_j (\mathrm{co}\ f)(x_j) = \alpha_j f(x_j) \text{ for all } j \in J.$$

Consequently, $f(x_j) = (\mathrm{co}\ f)(x_j)$ for all $j \in J$.

(ii) Let $x' \in P$ such that $x' = \sum_{j\in J} \beta_j x_j$.

All elements of P can be expressed similarly with same x_j; only the β_j's vary.

Let $\ell(x') = \sum_{j\in J} \beta_j f(x_j)$. Then ℓ is an affine function.

Now,

$$(\mathrm{co}\ f)(x') = (\mathrm{co}\ f)(\sum_{j\in J} \beta_j x_j) \le \sum_{j\in J} \beta_j (\mathrm{co}\ f)(x_j) \text{ by the convexity of co f;}$$

So, we have

$$(\mathrm{co}\ f)(x') \le \sum_{j\in J} \beta_j f(x_j) \text{ by (i);}$$
$$= \ell(x').$$

So, $\mathrm{co}\ f \le \ell$ on P.

Now let $x \in P$ and $x \neq x'$. Let the affine line passing through x and x' intersect $rbd\ P$ at y_1 and y_2 respectively. Then $y_1 \neq x$ and $y_2 \neq x$ since $x \in ri\ P$.

Then we can write

$$x = \lambda x' + (1-\lambda) y_1 \text{ with } \lambda \in (0,1).$$

Now,

$$(co\ f)(x) \le \lambda (co\ f)(x') + (1-\lambda)(co\ f)(y_1)$$
$$\le \lambda \ell(x') + (1-\lambda)\ell(y_1) = \ell(x).$$

But by the hypothesis, we have $(co\ f)(x) = \sum_{j\in J} \beta_j f(x_j) = \ell(x)$.

Thus we have

$$(co\ f)(x) = \lambda (co\ f)(x') + (1-\lambda)(co\ f)(y_1)$$
$$= \lambda \ell(x') + (1-\lambda)\ell(y_1) = \ell(x).$$

Then $\lambda[(co\ f)(x') - \ell(x')] + (1-\lambda)[co\ f(y_1) - \ell(y_1)] = 0$.

Since $co\, f \le \ell$ $on\, P$, we have

$$(co\ f)(x') = \ell(x') \text{ and } (co\ f)(y_1) = \ell(y_1).$$

Hence $co\, f$ is affine on P since ℓ is affine on P. ///

Theorem 1.5.2: Suppose the hypotheses of Theorem 1.5.1 hold. Then

$(i)\ \partial(co\ f)(x) = \bigcap_{j\in J} \partial f(x_j)$;

$(ii) For all\ s \in \partial(co f)(x),$ it holds that

$$\langle s,x\rangle - (co\ f)(x) = \langle s,x_j\rangle - f(x_j), for\ all\ j \in J.$$

Proof: (i) First we observe that

$$f^*=(co\ f)^*, \qquad (1.5.1)$$

by Corollary 1.3.1.

Let $s \in \partial(co\, f)(x)$. This is true if and only if

$$(co\, f)^*(s) + (co\, f)(x) - \langle s,x\rangle = 0, \qquad (1.5.2)$$

by Remark 1.4.1.

This is again true if and only if

$$f^*(s) + \sum_{j\in J} \alpha_j f(x_j) - \left\langle s, \sum_{j\in J} \alpha_j \right\rangle = 0 \quad ,$$

which is equivalent to

$$\sum_{j\in J} \alpha_j[f^*(s) + \alpha_j f(x_j) - \langle s, x_j\rangle] = 0 \quad \text{for all } j \in J,$$

since $\quad f^*(s) = \sum_{j\in J} \alpha_j f^*(s)$.

This, in turn, is equivalent to

$$f^*(s) + f(x_j) - \langle s, x_j\rangle = 0 \quad \forall j \in J, \qquad (1.5.3)$$

since by Fenchel's inequality, $\quad f^*(s) + f(x_j) - \langle s, x_j\rangle \geq 0, \forall j \in J.$

Now (1.5.3) is equivalent to $s \in \partial f(x_j)$ for all $j \in J$,

which is true if and only if $s \in \bigcap_{j\in J} \partial f(x_j)$.

(ii) From (1.5.1), (1.5.2) and (1.5.3), we have that

$$\langle s, x\rangle - (co\ f)(x) = -(co\ f)^*(s) = -f^*(s) = \langle s, x_j\rangle - f(x_j). \ ///$$

Corollary 1.5.2: Let $f : \mathbf{R}^n \to \mathbf{R}$ be lower semicontinuous, Gateaux-differentiable and 1-coercive. Then

(i) $\operatorname{co} f = \overline{\operatorname{co}} f$ is continuously differentiable on $\mathbf{R}^n$;

(ii) $\nabla(\operatorname{co} f)(x) = \nabla f(x_j)$ for all j, where x_j, x and J are as in Proposition 1.5.2 (ii).

Proof: **(i)** By Proposition 1.5.2(i), $\operatorname{co} f = \overline{\operatorname{co}} f$. But the convex function $\overline{\operatorname{co}} f$ (which is Gateaux-differentiable by the hypothesis) has a single subgradient

$$\nabla(\overline{\operatorname{co}} f)(x) = \nabla(\operatorname{co} f)(x),$$

which means $\overline{co} f$ is differentiable. Since x is arbitrary over all the open set $\mathbf{R}^n$, it follows that f is continuously differentiable on $\mathbf{R}^n$.

(iii) Follows from the differentiability and Theorem 1.5.2. ///

2.CALCULUS RULES ON THE CONJUGACY OPERATION

2.1.Image of a Function Under a Linear Mapping

Definition 2.1.1: Let $g : \mathbf{R}^m \to \mathbf{R} \cup \{+\infty\}$ satisfy (1.1.1), and let $A : \mathbf{R}^m \to \mathbf{R}^n$ be a linear mapping .The image of g under A, denoted by Ag, is $Ag : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ defined by

$$(Ag)(x) := \inf\{ g(y) : Ay = x\} .$$

Theorem 2.1.1:Let A^* be the adjoint of the operator $A : \mathbf{R}^m \to \mathbf{R}$, and let $g : \mathbf{R}^m \to \mathbf{R} \cup \{+\infty\}$ satisfy (1.1.1). Suppose $im\, A^* \cap dom\ g^* \neq \varnothing$. Then

(i) Ag satisfies (1.1.1);

(ii) $(Ag)^* = g^* \circ A^*$.

Proof: (i) Let $y \in domg$ such that x=Ay. Then $g(y) < +\infty$ and hence $Ag \not\equiv +\infty$.

Now let $s_0 \in imA^* \cap dom g$. Then $g^*(s_0) < +\infty$, and hence there is $t_0 \in \mathbf{R}^n$ such that $A^* t_0 = s_0$.

By Fenchel's inequality, we have for all $y \in \mathbf{R}^n$ that

$$g(y) \geq \langle s_0, y\rangle - g^*(s_0) = \langle A^* t_0, y\rangle - g^*(s_0) = \langle t_0, Ay\rangle - g^*(s_0).$$

Thus,

$$(Ag)(s) = \inf\{g(y) : Ay = x\} \geq \langle t_0, x\rangle - g^*(s_0) \textit{ for all } x \in \mathbf{R}^n.$$

This means that the affine function $\langle t_0, .\rangle - g^*(s_0)$ minorizes

Ag on $\mathbf{R}^n$. Consequently, Ag satisfies (1.1.1).

(ii) $(Ag)^*(s) = \sup_{x \in \mathbf{R}^n} \{\langle s, x\rangle - (Ag)(x)\} = \sup_{x \in \mathbf{R}^n} \{\langle s, x\rangle - \inf\{g(y) \mid Ay = x\}\}$

$$= \sup_{x \in \mathbf{R}^n} \{\langle s, x\rangle - g(y) \mid Ay = x\}$$

$$= \sup_{x \in \mathbf{R}^n} \{\langle s, Ay\rangle - g(y)\} = \sup_{x \in \mathbf{R}^n} \{\langle A^* s, y\rangle - g(y)\} = g^*(A^* s) = (g^* \circ A^*)(s)$$

Therefore,

$$(Ag)^* = g^* \circ A^*. \quad ///$$

Definition 2.1.2: Let $g : \mathbf{R}^n \times \mathbf{R}^p \to \mathbf{R} \cup \{+\infty\}$ satisfy (1.1.1).

The marginal function of g is $\gamma : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ defined by

$$\gamma(x) := \inf\{ g(x, y) \mid y \in \mathbf{R}^p\}.$$

Corollary 2.1.1: Let $\mathbf{R}^m := \mathbf{R}^n \times \mathbf{R}^p$; *let* $g : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$, and $g \not\equiv +\infty$.

Let g* be associated with a scalar product preserving the structure product space: $\langle .,.\rangle_m = \langle .,.\rangle_n + \langle .,.\rangle_p$. Suppose there is $s_0 \in \mathbf{R}^n$ such that

$(s_0, 0) \in dom\, g^*$.

Then the conjugate of the marginal function f of g is given by

$$f^*(s) = g^*(s, 0).$$

Proof: Consider the projection $A:\mathbf{R}^n \times \mathbf{R}^p \to \mathbf{R}^n$ *defined by* $A(x,z)=x$.

First we notice that for $(x, z) \in \mathbf{R}^n \times \mathbf{R}^p$, we have

$$(Ag)(x,z)=\inf\{g(y,w) \mid A(y,w)=(x,z)\}=\inf\{g(y,w) \mid (y,0)=(x,z)\}$$

$$=\inf\{g(x,w) \mid w \in \mathbf{R}^p\}=f(x) \qquad (2.1.1)$$

Next, $\langle (x,z), A^* x'\rangle = \langle A(x,z), x'\rangle_n = \langle x, x'\rangle_n = \langle x, x'\rangle_n + \langle z,0\rangle_p = \langle (x,z),(x',0)\rangle_m$

So, $\qquad A^* x' = (x',0). \qquad (2.1.2)$

Now, for all $s \in \mathbf{R}^n$,

$$f^*(s) = \sup_{(x,z)\in \mathbf{R}^n \times \mathbf{R}^n} \{\langle (s,0),(x,z)\rangle - \inf\{g(x,z) \mid z \in \mathbf{R}^p\}\}$$

$$= \sup_{(x,z)\in \mathbf{R}^n \times \mathbf{R}^n} \{\langle (s,0),(x,z)\rangle - (Ag)(x,z) |\}\}, \text{ from (2.1.1);}$$

$= (Ag)^*(s,0) = (g^* \circ A^*)(s,0)$, by Theorem 2.1.1,

$= g^*(A^*(s,0)) = g^*(s,0)$ by (2.1.2). ///

Definition 2.1.3: Let $f_1, f_2 : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$. The infimal convolution of f_1 *and* f_2 is $f_1 \underset{\vee}{+} f_2 : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ defined by

$$(f_1 \underset{\vee}{+} f_2)(x) = \inf\{f_1(x_1)+f_2(x_2) \mid x_1+x_2=x\}.$$

Corollary 2.1.2: Let $f_1, f_2 : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$, $f_1 \not\equiv +\infty$, $f_2 \not\equiv +\infty$ and $dom f_1 \cap dom f_2 \neq \emptyset$. Then

(*i*) $f_1 \underset{\vee}{+} f_2$ satisfies(1.1.1);

(*ii*) $(f_1 \underset{\vee}{+} f_2)^* = f_1^* + f_2^*$.

Proof: **(i)**Since $f_1 \not\equiv +\infty$, $f_2 \not\equiv +\infty$,it follows that $f_1 \underset{\vee}{+} f_2 \not\equiv +\infty$. Let $s \in dom f_1^* \cap dom f_2^*$.

Then by Fenchel's inequality, for all $x = x_1 + x_2 \in \mathbf{R}^n$, we have

$$f_1(x_1) \geq f_1^*(s) + \langle s, x_1 \rangle \text{ , and } f_2(x_2) \geq f_2^*(s) + \langle s, x_2 \rangle$$

so that

$$f_1(x_1) + f_2(x_2) \geq f_1^*(s) + f_2^*(s) + \langle s, x_2 + x_2 \rangle.$$

Then

$$(f_1 \underset{\vee}{+} f_2)(x) = \inf\{ f_1(x_1) + f_2(x_2) \mid x_1 + x_2 = x\}$$

$$\geq \langle s, x \rangle + f_1^*(s) + f_2^*(s) \text{ for all } x \in \mathbf{R}^n.$$

Therefore, $f_1 \underset{\vee}{+} f_2$ is minorized by an affine function .

Consequently,

$f_1 \underset{\vee}{+} f_2$ satisfies (1.1.1).

(ii) Let $s_1 \in \mathbf{R}^n$.Then

$$(f_1 \underset{\vee}{+} f_2)(s) = \sup_{x_1, x_2 \in \mathbf{R}^n} \{\langle s, x \rangle - \inf\{ f_1(x_1) + f_2(x_2) \mid x_1 + x_2 = x\}\}$$

$$= \sup_{x_1, x_2 \in \mathbf{R}^n} \{\langle s, x_1 \rangle + \langle s, x_2 \rangle - f_1(x_1) - f_2(x_2)\}.$$

(inf is omitted since introducing larger values wouldn't affect the supremum here.)

So,

$$(f_1 \underset{\vee}{+} f_2)(s) = \sup_{x_1 \in \mathbf{R}^n} \{\langle s, x_1 \rangle - f_1(x_1)\} + \sup_{x_2 \in \mathbf{R}^n} \{\langle s, x_2 \rangle - f_2(x_2)\}$$

$$= f_1^*(s) + f_2^*(s) \quad ./\!/\!/$$

2.2.Pre-Composition with an Affine Mapping

Theorem 2.2.1: For $g \in \overline{Conv}\, \mathbf{R}^n$, $A_0 : \mathbf{R}^n \to \mathbf{R}^m$ be linear and $A(x) := A_0 x + y_0$, suppose $A(\mathbf{R}^n) \cap dom g \neq ø$.Then

$(i)\ g \circ A \in \overline{Conv}\, \mathbf{R}^n$;

$(ii)\ (g \circ A)^* = cl\ k$, where k is the convex function $k : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ defined by $k(s) := \inf\{g^*(p) - \langle y_0, p \rangle_m \mid A_0^* p = s\}$.

Proof: **(i)** Immediate since $g \in \overline{Conv}\,\mathbf{R}^n$ and A is affine.

(ii) Case 1: $y_0 = 0$. Then A is linear. Now suppose that $h \in \overline{Conv}\,\mathbf{R}^n$ such that $im\, A_0 \cap dom\, h \neq \emptyset$. By Theorem 2.1.1, it follows that $(A_0 * h^*)^* = h \circ A_0$. Now $(h \circ A_0)^* = (A_0 * h^*)^{**} = cl(A_0 * h^*)$. (See Remark 1.3.1.)

Case 2: $y_0 \neq 0$.

Let $h := g(. + y_0) \in \overline{C\,onv}\,\mathbf{R}^m$.

By Proposition 1.3.1(b), we have that $h^* = g^* - \langle y_0, . \rangle_m$.

Moreover, $(g \circ A)(x) = g(A_0 x + y_0) = h(A_0 x) = (h \circ A_0)(x)$.

Then $(g \circ A)^*(s) = (h \circ A_0)^*(s) = [cl(A_0 * h^*)](s)$.

On the other hand,

$$(A_0 * h^*)(s) = \inf\{h^* y \mid A^* y = s\} = \inf\{g^*(y) - \langle y_0, y\rangle \mid A^* y = s\} = k(s).$$

Consequently, $(g \circ A)^* = cl\,k$. ///

Lemma 2.2.1: Let $g \in \overline{\mathrm{Conv}}\,\mathbf{R}^m$ such that $0 \in \mathrm{dom}\ g$, and let $A_0 : \mathbf{R}^n \to \mathbf{R}^m$ be linear. Suppose $\mathrm{im}\, A_0 \cap \mathrm{ri\,dom}\, g \neq \emptyset$. Let $A_0 * g^*$ be understood in the sense of Definition 2.1.1. Then

(i) $(g \circ A)^* = A_0 * g^*$

(ii) For all $s \in \mathrm{dom}\ (g \circ A_0)^*$, the problem

$$\inf\{g^*(p) \mid A_0 * p = s\}$$

has an optimal solution $\bar{p}$ so that $g^*(\bar{p}) = (g \circ A_0)^*(s) = (A_0 * g^*)(s)$.

For the proof Cf.[4].

Theorem 2.2.2: Let $g \in \overline{\text{Conv}}\, \mathbf{R}^n$, and $A_0 : \mathbf{R}^n \to \mathbf{R}^m$ be linear. Let $Ax = A_0 x + y_0$ such that $A(\mathbf{R}^m) \cap \text{ri dom } g \neq \emptyset$. Then for every $s \in \text{dom}(g \circ A_0)^*$,

the following optimization problem has a solution:

$$\min\{g^*(p) - \langle p, y_0 \rangle \mid A_0^* p = s\} = (g \circ A)^*(s) .$$

Proof: Since $A(\mathbf{R}^m) \cap \text{ri dom } g \neq \emptyset$, there is

$$\bar{x} \in \mathbf{R}^n \text{ such that } \bar{y} = A(\bar{x}) \in \text{ri dom } g .$$

Let $\bar{g} := g(\bar{y} + .)$. Then

$$(g \circ A)(x) = g(A(x)) = \bar{g}(A(x) - \bar{y}) = \bar{g}(A(x) - A(\bar{x})) = \bar{g}(A_0 x + y_0 - A\bar{x} - y_0)$$

$$= (\bar{g} \circ A_0)(x - \bar{x}).$$

Now by Proposition 1.3.1(f), we have

$$(g \circ A)^*(s) = (\bar{g} \circ A)^*(s) + \langle s, \bar{x} \rangle. \qquad (2.2.1)$$

Since $\bar{y} \in A(\mathbf{R}^m) \cap \text{ri dom } g$ and $\bar{g}(0) = g(\bar{y})$, it follows that $g(\bar{y}) < +\infty$.

Thus $0 \in \text{ri dom}\, \bar{g} \cap \text{im}\, A_0$.

Then by Lemma 2.2.1, for all $s \in \text{dom}(\bar{g} \circ A_0)^* = \text{dom}(g \circ A_0)^*$, we have that

$$(\bar{g} \circ A_0)^*(s) = \min_p\{\bar{g}^*(p) \mid A_0^* p = s\} .$$

Now, using (2.2.1), we have

$$(g \circ A)^*(s) = \min_p \{\bar{g}^*(p) \mid A_0^* p = s\} + \langle s, \bar{x} \rangle$$

$$= \min_p \{g^*(p) - \langle p, \bar{y} \rangle \mid A_0^* p = s\} + \langle s, \bar{x} \rangle$$

$$= \min_p \{g^*(p) - \langle p, A\bar{x} \rangle \mid A_0^* p = s\} + \langle s, \bar{x} \rangle$$

$$= \min_p \{g^*(p) - \langle p, A_0\bar{x} + y_0 \rangle \mid A_0^* p = s\} + \langle s, \bar{x} \rangle$$

$$= \min_p \{g^*(p) - \langle p, A_0\bar{x} \rangle - \langle p, y_0 \rangle \mid A_0^* p = s\} + \langle s, \bar{x} \rangle$$

$$= \min_p \{g^*(p) - \langle A_0^* p, \bar{x} \rangle - \langle p, y_0 \rangle \mid A_0^* p = s\} + \langle s, \bar{x} \rangle$$

$$= \min_p \{g^*(p) - \langle s, \bar{x} \rangle - \langle p, y_0 \rangle \mid A_0^* p = s\} + \langle s, \bar{x} \rangle$$

$$= \min_p \{g^*(p) - \langle p, y_0 \rangle \mid A_0^* p = s\} . \qquad ///$$

Corollary 2.2.1: Let $g \in \overline{Conv}\, \mathbf{R}^n$ and let $A : \mathbf{R}^n \to \mathbf{R}^m$ be linear such that $im\, A \cap dom\, g \neq \emptyset$ .Then

$(i)\ (g \circ A)^* = A^*(g + I_{imA})^*$;

(ii) *for every* $s \in dom\,(g \circ A)^*$, *the problem* $\inf_p \{(g + I_{imA})^*(p) \mid A_0^* p = s\}$ *has an optimal solution* $\bar{p}$ *satisfying* $(g + I_{imA})^*(\bar{p}) = A^*(g + I_{imA})^*(s) = (g \circ A)^*(s)$.

Proof: Since the convex closed functions $g \circ A$ *and* $(g + I_{imA}) \circ A$ are identical on $\mathbf{R}^n$, it follows that

$$(g \circ A)^* = [\ (g + I_{imA}) \circ A]^* .$$

Moreover, $g + I_{imA}$ is a closed convex function such that $dom(g + I_{imA}) \subseteq im\, A$. Thus,

$$[ri\,dom(g + I_{imA})] \cap im\, A = ri\,dom(g + I_{imA}) \neq \emptyset .$$

Then

$(g \circ A)^* = [(g + I_{imA}) \circ A]^* = A^*(g + I_{imA})^*$ by Lemma 2.2.1.

And (ii) follows from Theorem 2.2.2. ///

Example 2.2.1: Let $g \in \overline{Conv}\, \mathbf{R}^n$, $x_0 \in dom\ g$, $0 \neq d \in \mathbf{R}^n$.

Let $L(t) := g(x_0 + td)$ *for all* $t \in \mathbf{R}$.

To find L^*, we observe that $L = g \circ A$ where A is the affine mapping

$A : \mathbf{R} \to \mathbf{R}$ given by $At = x_0 + td$.

This means

$At = A_0 t + x_0$ *where* $A_0 : \mathbf{R} \to \mathbf{R}$ such that $A_0 t = td$.

So $A_0 * p = \langle d, p\rangle$, since

$$\langle A_0 t, p\rangle = \langle t, A_0 * p\rangle \text{ iff } t\langle d, p\rangle = tA_0 * p .$$

Now by Theorem 2.2.1,we have

$$L*(s) = \min_p \{g*(p) - \langle x_0, p\rangle \mid A_0 * p = s\}$$

$$= \min_p \{g*(p) - \langle x_0, p\rangle \mid \langle d, p\rangle = s\}$$

2.3. Conjugate of Sum of Two Functions

Theorem 2.3.1: Let $g_1, g_2 \in \overline{Conv}\,\mathbf{R}^n$ *such that* $dom\, g_1 \cap domg_2 \neq \varnothing$.

Then $(g_1 + g_2)^* = cl\,(g_1 * \underset{\vee}{+} g_2 *)$.

Proof: Put $f_1 * = g_1$, $f_2 * = g_2$. Then by Corollary 2.1.3,we have

$(g_1 * \underset{\vee}{+} g_2 *)^* = g_1 ** + g_2 ** = cl\,(g_1) + cl\,(g_2)$ by Remark 1.3.1,

$= g_1 + g_2$, because $g_1, g_2 \in \overline{Conv}\,\mathbf{R}^n$.

This imples that

$$cl(g_1 * \underset{\vee}{+} g_2 *) = (g_1 * \underset{\vee}{+} g_2 *)** = (g_1 + g_2)*. \quad ///$$

Theorem 2.3.2: Let $g_1, g_2 \in \overline{\mathrm{Conv}}\,\mathbf{R}^n$. Suppose $(\mathrm{ri\,dom\,g_1}) \cap (\mathrm{ri\,dom\,g_2}) \neq \emptyset$,

or equivalently, $0 \in \mathrm{ri}(\mathrm{dom\,g_1} - \mathrm{dom g_2})$.

Then

(i)$(\mathrm{g_1 + g_2})^* = \mathrm{g_2} * \underset{\vee}{+} \mathrm{g_2} *$;

(ii)for every $\mathrm{s} \in \mathrm{dom}(\mathrm{g_1 + g_2})^*$, the problem

$$\inf\{\mathrm{g_1} * (\mathrm{p}) + \mathrm{g_2} * (\mathrm{q}) \mid \mathrm{p + q = s}\}$$

has an optimal solution $(\bar{\mathrm{p}}, \bar{\mathrm{q}})$ such that

$$g_1 * (\bar{p}) + g_2 * (\bar{q}) = (g_1 * \underset{\vee}{+} g_2 *)(s) = (g_1 + g_2) * (s) .$$

Proof: Let $g_1, g_2 \in \overline{\text{Conv}}(\mathbf{R}^n \times \mathbf{R}^n)$ such that $g(x_1, x_2) := g_1(x_1) + g_2(x_2)$, and let $A : \mathbf{R}^n \to \mathbf{R}^n \times \mathbf{R}^n$ such that $Ax := (x, x)$. Then $g_1 + g_2 = g \circ A$.

By Proposition 1.3.1, we have $g^*(p,q) = g_1^*(p) + g_2^*(q)$.

Moreover, $A^*(p,q) = p + q$,

since $\langle x, A^*(p, q)\rangle = \langle Ax, (p, q)\rangle$ iff $\langle x, A^*(p, q)\rangle = \langle x, p + q\rangle$.

To apply Theorem 2.2.2, notice that $\text{dom} g = \text{dom} g_1 \times \text{dom} g_2$ and

$\text{im} A = \{(s, s) \mid s \in \mathbf{R}^n\}$.

Then $(x, x) \in \text{ri dom } g_1 \times \text{ri dom } g_2 = \text{ri}(\text{dom } g_1 \times \text{dom } g_2)$,

and hence $\text{im} A \cap \text{ri dom} g \neq \emptyset$.

Now,

$$\begin{aligned}(g_1 + g_2)^*(s) &= (g \circ A)^*(s) = (A^* g^*)(s) \\ &= \inf_{p, q} \{g_1^*(p) + g_2^*(q) \mid p + q = s\} \qquad (2.3.1) \\ &= (g_1^* \underset{\vee}{+} g_2^*)(s)\end{aligned}$$

This proves (i).

Since $A^*(p,q) = p + q$ and $g^*(p,q) = g_1^*(p) + g_2^*(q)$, the problem (2.3.1) has an optimal solution by Theorem 2.2.2. This proves (ii). ///

Now we have

Fenchel's Duality Theorem : Let $g_1, g_2 \in \overline{\text{Conv}}\, \mathbf{R}^n$ such that

$(\text{ri dom} g_1) \cap (\text{ri dom} g_2) \neq \emptyset$ and

$\mu := \inf\{g_1(x) + g_2(x) \mid x \in \mathbf{R}^n\} \in \mathbf{R}$. Then

$$-\mu = (g_1 + g_2)^*(0) = \min_{s \in \mathbf{R}^n} \{g_1^*(s) + g_2^*(-s)\}. \qquad (2.3.2)$$

Corollary 2.3.1: Let $f_1, f_2 \in \overline{\text{Conv}}\, \mathbf{R}^n$ such that f_1 is 0-coercive and f_2 is bounded from below. Then

(i) the problem $(f_1 \underset{\vee}{+} f_2)(x) := \inf\{f_1(x) + f_2(x_2) \mid x_1 + x_2 = x\}$

has a nonempty compact set of solutions;

(ii) $f_1 \underset{\vee}{+} f_2 \in \overline{\text{Conv}}\ \mathbf{R}^n$.

Proof: Let μ be a lower bound for f_2. Then

$f_1(x_1) + f_2(x - x_1) \geq f_1(x_1) + \mu$ for all $x_1 \in \mathbf{R}^n$. So there exist $\bar{x}_1, \bar{x}_2 \in \mathbf{R}^n$ such that $\bar{x} := \bar{x}_1 + \bar{x}_2$ with

$$(f_1 \underset{\vee}{+} f_2)(\bar{x}) = \inf\{f_1(x_1) + f_2(x_2) \mid x_1 + x_2 = x\}.$$

Hence (i) holds.

Let now $g_1 := f_1{}^*$ and $g_2 := f_2{}^*$. Since $g_1{}^* = f_1{}^{**} = \text{cl} f_1$ is 0-coercive, it follows by Remark 1.3.2(b) that

$$0 \in \text{int dom}\, g_1 \ .$$

Furthermore,

$$g_2(0) = f_2{}^*(0) = \sup_{x \in \mathbf{R}^n} \{ \langle 0, x \rangle - f(x) \} \leq -\mu$$

Then $(\text{intdom} g_1) \cap (\text{intdom} g_2) \neq \emptyset$.

Now by Theorem 2.3.2, we have

$$(f_1 \underset{\vee}{+} f_2)(x) := \inf\{f_1(x_1) + f_2(x - x_1) \mid x_1 \in \mathbf{R}^n\} = \inf\{f_1{}^{**}(x_1) + f_2{}^{**}(x - x_1) \mid x_1 \in \mathbf{R}^n\}$$

$$= f_1{}^{**}(\bar{x}) + f_2{}^{**}(x - \bar{x}) \quad \text{for some} \quad \bar{x} \in \mathbf{R}^n$$

$$= (f_1{}^* + f_2{}^*)^*(x) \ .$$

Hence $f_1 \underset{\vee}{+} f_2 = (f_1{}^* + f_2{}^*)^* \in \overline{\text{Conv}}\ \mathbf{R}^n$. So (ii) holds. ///

2.4. Conjugate of Suprema and Infima

We shall mean

$$(\sup_{j\in J} f_j)(x) := \sup_{j\in J} f_j(x)\ , \quad (\inf_{j\in J} f_j)(x) := \inf_{j\in J} f_j(x)\ ,$$

$$(\max_{j\in J} f_j)(x) := \max_{j\in J} f_j(x)\ , \text{ and } (\min_{j\in J} f_j)(x) := \min_{j\in J} f_j(x)$$

Theorem 2.4.1: Let $\{f_j\}_{j\in J}$ be a collection of functions satisfying (1.1.1) such that there is a common affine function minorizing f_j for all $j \in J$.

Then (a) $f := \inf_{j\in J} f_j$ satisfies (1.1.1);

(b) $(\inf_{j\in J} f_j)^* = \sup_{j\in J} f_j{}^*$.

Proof: (a) follows directly from the hypothesis.

(b)
$$(\inf_{j\in J} f_j)^*(s) = \sup_{x\in \mathbf{R}^n} \{ \langle s, x\rangle - \inf_{j\in J} f_j(x) \}$$
$$= \sup_{x\in \mathbf{R}^n} \sup_{j\in J} \{ \langle s, x\rangle - f_j(x) \}$$
$$= \sup_{j\in J} \sup_{x\in \mathbf{R}^n} \{ \langle s, x\rangle - f_j(x) \}$$
$$= \sup_{j\in J} f_j{}^*(s)$$ ///

Example 2.4.1 : Let $f \in \overline{\mathrm{Conv}}\ \mathbf{R}^n$ be Gateaux - differentiable and let $x_0 \in \mathbf{R}^n$ such that $\partial f(x_0) \neq \emptyset$.

Let $f_t : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ be defined by

$$f_t(d) := \frac{f(x_0 + td) - f(x_0)}{t} \quad \text{for } t > 0.$$

All the functions f_t are minorized by $\langle s_0, .\rangle$ for $s_0 \in \partial f(x)$, since

$$\langle s_0, .\rangle \leq f'(x_0, .) \leq f_t \quad \text{for all } t > 0 .$$

f_t is increasing for each t>0 since f is convex.

Thus,

$$\inf_{t>0} f_t(d) = \lim_{t\downarrow 0} f_t(d)$$

$$= \lim_{t\downarrow 0} \frac{f(x_0 + td) - f(x_0)}{t}$$

$$= f'(x_0, d) .$$

Now,

$$[f'(x_0, d)]^*(s) = \sup_{t>0} f_t^*(s) \text{ by Theorem 2.4.1.}$$

But

$$(f_t^*)(s) = \sup_{d\in \mathbf{R}^n} \{\langle s, d\rangle - f_t(d)\}$$

$$= \sup_{d\in \mathbf{R}^n} \left\{\langle s, d\rangle - \frac{f(x_0 + td) - f(x_0)}{t}\right\}$$

$$= \frac{1}{t} \sup_{d\in \mathbf{R}^n} \{\langle s, td\rangle - f(x_0 + td) + f(x_0)\}$$

$$= \frac{1}{t} h^*(s) ,$$

where $h(td) := f(x_0 + td) - f(x_0)$ so that

$$h^*(s) = f^*(s) - \langle s, x_0\rangle + f(x_0) \text{ by Proposition 1.3.1.}$$

Therefore,

$$[f'(x_0, d)]^*(s) = \sup_{t>0} \frac{f^*(s) - \langle s, x_0\rangle + f(x_0)}{t}.$$

By Fenchel's inequality,

$$\frac{f^*(s) - \langle s, x_0\rangle + f(x_0)}{t} \geq 0,$$

and by Remark 1.4.1,

$$\frac{f^*(s) - \langle s, x_0\rangle + f(x_0)}{t} = 0 \quad \text{if and only if} \quad s \in \partial f(x_0).$$

So ,

$$[f'(x_0,d)]^*(s) = \begin{cases} 0 & \text{if } s \in \partial f(x_0) \\ +\infty & \text{otherwise} \end{cases}$$

$$= I_{\partial f(x_0)}(s) \,.$$

Consequently, we have that

$$f'(x_0,.)^* = I_{\partial f(x_0)},$$

and hence,

$$\mathrm{cl}(f'(x_0,.)) = I_{\partial f(x_0)}^* \,.$$

$$= \sigma_{\partial f(x_0)}. \qquad \text{(See Example 1.1.2.)}$$

Theorem 2.4.2: Let $\{g_j\}_{j\in J}$ be a collection of functions in $\overline{\mathrm{Conv}}\,\mathbf{R}^n$.

If $g := \sup_{j\in J} g_j \not\equiv +\infty$, then

(a) $\sup_{j\in J} g_j \in \overline{\mathrm{Conv}}\,\mathbf{R}^n$;

(b) $(\sup_{j\in J} g_j)^* = \overline{\mathrm{co}}(\inf_{j\in J} g_j^*)$.

Proof: Let $f_j := g_j^*$ for all $j \in J$.

Then $f_j^* := g_j$ since $g_j \in \overline{\mathrm{Conv}}\,\mathbf{R}^n$ for all $j \in J$.

Now,

$$g = \sup_{j\in J} g_j = \sup_{j\in J} f_j^*$$

$$= (\inf_{j\in J} f_j)^* \qquad \text{by Theorem 2.4.1(b).}$$

Thus, it follows that $g \in \overline{\mathrm{Conv}}\,\mathbf{R}^n$.

Furthermore,

$$g^* = (\inf_{j\in J} f_j)^{**} = \overline{\mathrm{co}}(\inf_{j\in J} g_j^*) \quad . \quad ///$$

Example 2.4.2 (Asymptotic Function): Let $f \in \overline{\mathrm{Conv}}\, \mathbf{R}^n$, $x_0 \in \mathrm{dom} f$ and

$f_t : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ such that $f_t(d) := \dfrac{f(x_0 + td) - f(x_0)}{t}$ for $t > 0$.

Then $(f_\infty^{'})(d) = \sup\{ f_t \mid t > 0\}$. Clearly, $f_t \in \overline{\mathrm{Conv}}\, \mathbf{R}^n$.

Now by Theorem 2.4.2, we have that $f_\infty{}' \in \overline{\mathrm{Conv}}\, \mathbf{R}^n$,

and

$$(f_\infty^{'})^* = \overline{\mathrm{co}}\ (\inf_{j\in J} f_j^*) = \overline{\mathrm{co}}\ \left[\frac{f^* + f(x_0) - \langle ., x_0 \rangle}{t}\right] = \overline{\mathrm{co}}\ I_{\mathrm{dom}\, f},$$

since $\inf \dfrac{f^*(s) + f(x_0) - \langle s, x_0 \rangle}{t} = \begin{cases} 0 & \text{if } s \in \mathrm{dom}\ f^* \\ +\infty & \text{otherwise} \end{cases}$.

Therefore,

$$f_\infty^{'} = \overline{\mathrm{co}}\ f_\infty^{'} = (f_\infty^{'})^{**} = \overline{\mathrm{co}}\ I_{\mathrm{dom}\, f^*} = I_{\mathrm{dom}\, f^*}.$$

Theorem 2.4.3: $f_i : \mathbf{R}^n \to \mathbf{R}$, i=1,2, ... , p, be convex.

Let $f := \max_j f_j$ and $m := \min\{p, n+1\}$.

Then for every $s \in \mathrm{dom} f^* = \mathrm{co}\left(\bigcup_{j=1}^{p} \mathrm{dom}\ f_j^* \right)$, there exist m vectors

$s_j \in \mathrm{dom} f_j^*$ and convex multipliers α_j such that

$$s = \sum_j \alpha_j s_j \ \text{ and } \ f^*(s) = \sum_j \alpha_j f^*(s_j);$$

i.e., the minimization problem

$$\inf\left\{ \sum_{j=1}^{n+1} \alpha_j f_j^*(s_j) \mid \alpha_j \ge 0, j = 1,2, ..., n+1,\ s_j \in \mathrm{dom} f_j^*,\ \sum_{j=1}^{n+1} \alpha_j s_j = s \right\} \quad (2.2.1)$$

has an optimal solution, namely, f*(s).

Proof: Let $g := \min_j f_j{}^*$. Then $f^* = (\max_j f_j)^* = \overline{co}\,(\inf_j f_j{}^*) = \overline{co}\,g$,

and $\operatorname{dom} g = \operatorname{dom}\,(\min_j f_j{}^*) = \bigcup_j \operatorname{dom} f_j{}^*$.

Since $\operatorname{dom} f_g = \mathbf{R}^n$, it follows that $f_j{}^*$ is 1-coercive by Proposition 1.3.5. Then g is closed and 1-coercive. Now we apply Proposition 1.5.2 on g and conclude that $co\,g = \overline{co}\,g$. Moreover, for every $s \in \operatorname{dom} f^* = \operatorname{dom}(co\,g) = co\left(\bigcup_j \operatorname{dom} f_j{}^*\right)$,

there are $s_j \in \operatorname{dom} g = \operatorname{dom} f^*$ and convex multipliers α_j, $j = 1, 2, \cdots, n+1$, such that

$$s = \sum_{k=1}^{n+1} \alpha_k s_k \quad \text{and} \quad f^*(s) = (co\ g)(s) = \sum_{k=1}^{n+1} \alpha_k\, g(s_k) = \sum_{k=1}^{n+1} \alpha_k\, (co\ g)(s_k)$$

$$= \sum_{k=1}^{n+1} \alpha_k\, f^*(s_k) = \sum_i \sum_k \alpha_k \min_i f_i{}^*(s_k) \cdot$$

If we replace $\min_i f_i{}^*(s_k)$ by the corresponding values and then rearrange (several s_k 's may have the same f_i^*), we get

$$s = \sum_j \alpha_j s_j \quad \text{and} \quad f^*(s) = \sum_j f_j{}^*(s_j). \qquad ///$$

2.5. Post-Composition with an Increasing Function

Theorem 2.5.1: Let $f \in \overline{Conv}\,\mathbf{R}^n$ and $g \in \overline{Conv}\,\mathbf{R}$ such that g is increasing and $f(\mathbf{R}^n) \cap (\operatorname{int\,dom}\ g) \neq \emptyset$. Let $\psi_s : \operatorname{dom}\,(g \circ f)^* \to \mathbf{R} \cup \{+\infty\}$ defined by

$$\psi_s(\alpha) := \begin{cases} \alpha f^*(\frac{1}{\alpha}s) + g^*(\alpha), & \text{if } \alpha > 0 \\ \sigma_{\operatorname{dom} f}(s) + g^*(0), & \text{if } \alpha = 0 \\ +\infty, & \text{if } \alpha < 0 \end{cases}$$

Then $\qquad (g \circ f)^*(s) = \min_{\alpha \in R} \psi_s(\alpha)$.

Proof: $(g\circ f)^*(s) = -\inf_x \{ g(f(x)) - \langle s,x\rangle \} = -\inf_{x,r} \{ g(r)) - \langle s,x\rangle \mid f(x)\le r \}$

$$= -\inf_{x,r} \{ g(r) - \langle s,x\rangle + I_{\mathrm{epi}\, f}(x,r) \}.$$

Now let $f_1 : \mathbf{R}^n \times \mathbf{R} \to \mathbf{R} \cup \{+\infty\}$ such that $f_1(x,r) := g(r) - \langle s,x\rangle$ and $f_2 := I_{\mathrm{epi} f}$.

To apply Fenchel's Duality theorem, observe that

$$(\mathrm{int\ dom}\ f_1) \cap (\mathrm{dom}\ f_2) = (\mathbf{R}^n \times \mathrm{int\ dom}\ g) \cap \mathrm{epi}\ f \ne \emptyset.$$

Then,

$$(g\circ f)^*(s) = (f_1 + f_2)^*(0) = \min_{(p,\alpha)\in \mathbf{R}^n \times \mathbf{R}} \{ f_1^*(-p,\alpha) + f_2^*(p,-\alpha) \}.$$

On the other hand,

$$f_1^*(-p,\alpha) = \sup_{(x,r)\in \mathbf{R}^n \times \mathbf{R}} \{ \langle (-p,\alpha), (x,r) \rangle - f_1(x,r) \}.$$

So,

$$\begin{aligned} f_1(-p,\alpha) &= \sup_{(x,r)\in \mathbf{R}^n \times \mathbf{R}} \{ \langle -p, x \rangle + \alpha r - g(r) + \langle s, x\rangle \} \\ &= \sup_{r\in\mathbf{R}} \{ \alpha r - g(r) \} + \sup_{x\in\mathbf{R}^n} \{ \langle s,x\rangle - \langle p,x\rangle \} \\ &= g^*(\alpha) + I_{\{-s\}}(-p), \end{aligned}$$

and

$$f_2^*(p,\alpha) = \sigma_{\mathrm{epi} f}(p,-\alpha),$$

by Example 1.1.2.

Thus,

$$(g\circ f)^*(s) = \min_{(p,\alpha)\in \mathbf{R}^n \times \mathbf{R}} \{ g^*(\alpha) + I_{\{-s\}}(-p) + \sigma_{\mathrm{epi}\ f}(p,-\alpha) \}.$$

Now applying Proposition 1.2.1(ii), we get

$$(g\circ f)^*(s) = \min_{\alpha\in\mathbf{R}} \psi_s(\alpha).$$ ///

Remark 2.5.1: From Theorem 2.5.1,it follows that ψ_s always attains its minimum at some $\bar{\alpha} \geq 0$ under the given assumptions.

2.6. Biconjugate Calculus

Remark 2.6.1: Proposition 1.3.1 gives the following corresponding rules for the biconjugate; only part (h)has no corresponding rule:

Let h_1, h_2 and f satisfy (1.1.1). Then

(a) $g(x) := f(x) + \alpha$, $\alpha \in \mathbf{R}$, implies $(\overline{\mathrm{co}}\ g)(x) = (\overline{\mathrm{co}}\ f)(x) + \alpha$);

(b) $g(x) := f(\alpha x)$, $\alpha \neq 0$, implies $(\overline{\mathrm{co}}\ g)(x) = (\overline{\mathrm{co}}\ f)(\alpha x)$;

(c) $g(x) := \alpha f(x)$, $\alpha > 0$, implies $(\overline{\mathrm{co}}\, g)(x) = \alpha(\overline{\mathrm{co}}\, f)(x)$;

(d) $\overline{\mathrm{co}}\ (f \circ A) = (\overline{\mathrm{co}}\, f) \circ A$ for an invertible linear operator A;

(e) $g(x) := f(x) + \langle s_0, x\rangle$ implies $(\overline{\mathrm{co}}\, g)(x) = (\overline{\mathrm{co}}\, f)(x) + \langle s_0, x\rangle$;

(f) $g(x) := f(x - x_0)$ implies $(\overline{\mathrm{co}}\, g)(x) = (\overline{\mathrm{co}}\, f)(x - x_0)$;

(g) $h_1 \leq h_2$ implies $\overline{\mathrm{co}}\, h_1 \leq \overline{\mathrm{co}}\, h_2$;

(i) Let $f_j : \mathbf{R}^{n_j} \to \mathbf{R} \cup \{+\infty\}$, j=1,2, …, m, satisfy (1.1.1).

If $\mathbf{R}^n = \mathbf{R}^{n_1} \times \mathbf{R}^{n_2} \times \cdots \times \mathbf{R}^{n_m}$ and $h(x) := \sum_{j=1}^{m} f_j(x_j)$, then

$$(\overline{\mathrm{co}}\ f)(x) = \sum_{j=1}^{m} \overline{\mathrm{co}}\ f_j(x_j) \cdot$$

Let f_1 and f_2 satisfy (1.1.1) such that $\mathrm{dom}\ f_1 \cap \mathrm{dom}\ f_2 \neq \emptyset$. Then $\overline{\mathrm{co}}\ f_1 + \overline{\mathrm{co}}\ f_2 \in \overline{\mathrm{C\,onv}}\ \mathbf{R}^n$ such that $\overline{\mathrm{co}}\ f_1 + \overline{\mathrm{co}}\ f_2 \leq \overline{\mathrm{co}}\ (f_1 + f_2) \leq f_1 + f_2$. Let now $\{f_j \mid j \in J\}$ be a collection of functions satisfying (1.1.1). Then using Remark 2.6.1(g) and the property of supremum, we get

$$\sup_{j\in J} (\overline{co}\, f_j) \le \overline{co}\, (\sup_{j\in J} f_j) \cdot$$

Proposition 2.6.1: Let $\{f_j \mid j \in J\}$ be a collection of functions satisfying (1.1.1) such that there is a common affine function minorizing all of them.

Then

$$\overline{co}\,(\inf_{j\in J} f_j(x)) = \overline{co}\,(\inf_{j\in J} \overline{co}\, f_j) \le \inf_{j\in J} \overline{co}\, f_j \cdot$$

Proof: By Theorem 2.4.1, $(\inf_{j\in J} f_j)^* = \sup_{j\in J} f_j{}^* \quad (\in \overline{C\,onv}\, \mathbf{R}^n)$.

Then by Theorem 2.4.2,we have

$$\overline{co}\,(\inf_{j\in J} f_j) = (\inf_{j\in J} f_j)^{**} = \overline{co}\,(\inf_{j\in J} f_j) =: \overline{co}\,(\inf_{j\in J} \overline{co}\, f_j) \cdot$$

Thus $\overline{co}\,(\inf_{j\in J} f_j) = \overline{co}\,(\inf_{j\in J} \overline{co}\, f_j) \le \inf_{j\in J} \overline{co}\, f_j \cdot$ ///

Proposition 2.6.2: Let $g : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ satisfy (1.1.1), and let $A : \mathbf{R}^n \to \mathbf{R}^m$ be linear such that $(\text{im } A^*) \cap \text{ri}(\text{dom } g^*) \ne \emptyset$.

Then $\overline{co}\,(Ag) = A(\overline{co}\, g)$.

Proof: Since $(\text{im } A^*) \cap \text{dom } g^* \ne \emptyset$, we have by Theorem 2.1.1 that

$$(Ag)^* = g^* \circ A^*.$$

Then $(\overline{co}\, Ag)(x) := (Ag)^{**}(x) = (g^* \circ A^*)^*(x)$

$$= \min_y \{g^{**}(y) \mid A^{**}y = x\} = \min_y \{\overline{co}\, g(y) \mid Ay = x\}$$

$$= A(\overline{co}\, g)(x) .$$

Therefore, $\overline{co}\,(Ag) = A(\overline{co}\, g)$. ///

3. SOME APPLICATIONS

3.1. Some Results on the Euclidean Distance to a Closed Set

Definition 3.1.1: Let $S \subseteq \mathbf{R}^n$, $S \neq \emptyset$. The distance from $x \in \mathbf{R}^n$ to S is defined to be $d_S(x) := \inf_{y \in S} \|y - x\|$.

Let now $\varphi_s(x) := \frac{1}{2}[\|x\|^2 - d_S{}^2]$. Then

$$d_S{}^2(x) := \inf_{c \in S} \|c - x\|^2 = \inf_{c \in S} \{\langle c - x, c - x \rangle\}$$

$$= \inf_{c \in S} \{\|x\|^2 - 2\langle c, x\rangle + \|c\|^2\} = \|x\|^2 + \inf_{c \in S}\{-2\langle c, x\rangle + \|c\|^2\}$$

$$= \|x\|^2 - 2\sup_{c \in S}\left\{\langle c, x\rangle - \frac{1}{2}\|c\|^2\right\}.$$

So
$$\varphi_s(x) = \frac{1}{2}[\|x\|^2 - d_S{}^2] = \sup_{c \in S}\left\{\langle c, x\rangle - \frac{1}{2}\|c\|^2\right\} \qquad (3.1.1)$$

Let P_S be the projection operator onto S defined by

$$P_S(x) := \{ y \in S \mid d_S(x) = \|y - x\|\}$$

Proposition 3.1.1: Let $S \subseteq \mathbf{R}^n$ be closed and $x \in S$; let $f : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ such that $f := I_S + \frac{1}{2}\|\cdot\|^2$; i.e., $f(x) = \begin{cases} \frac{1}{2}\|x\|^2 & \text{if } x \in S \\ +\infty & \text{otherwise} \end{cases}$.

Then

(a) $t \in \partial f(x) \Leftrightarrow x \in P_S(t)$;

i.e. , $\partial f(x) = \{ t \in \mathbf{R}^n \mid x \in P_S(t) \} = \{ t \in \mathbf{R}^n \mid d_S(t) = \|t - x\|\}$

In particular, $x \in \partial f(x)$.

(b) $(\text{co } f)(x) = f(x) = \frac{1}{2}\|x\|^2$ and $\partial(\text{co } f)(x) = \partial f(x)$.

Proof: $f^*(t) = \sup\left\{\langle t, x\rangle - \frac{1}{2}\|x\|^2 \mid x \in S\right\} = \varphi_S(t)$ by (3.1.1).

$$= \frac{1}{2}[\,\|t\|^2 - d_S{}^2(t)\,]\cdot$$

Now, $t \in \partial f(x)$ if and only if $(x) + f^*(t) - \langle s, x\rangle = 0$,

by Remark 1.4.1, which is true if and only if

$$I_S(x) + \frac{1}{2}\|x\|^2 + \frac{1}{2}\|t\|^2 - \frac{1}{2}d_S{}^2 - \langle t, x\rangle = 0\cdot$$

This, in turn, is true if and only if

$\|x\|^2 - 2\langle t, x\rangle + \|t\|^2 = d_S{}^2$, since $I_S(x) = 0$.

But $\|x\|^2 - 2\langle t, x\rangle + \|t\|^2 = \|t - x\|^2$ so that we get

$d_S(t) = \|t - x\|$, which is true if and only if $x \in P_S(t)$.

In particular, putting t=x, we have that $x \in P_S(x) = \{x\}$. So we have (a).

(b) follows from Proposition 1.4.1. ///

Proposition 3.1.2: Let f be as in Proposition 3.1.1. Then

$\partial f^*(t) = (\text{co } P_S)(t)$ for all $t \in \mathbf{R}^n$.

Proof: f is lower semicontinuous and 1-coercive. (So Proposition 1.5.2 can be applied .) Now let $\text{co } f \in C\overline{\text{onv}}\,\mathbf{R}^n$ so that by Corollary 1.4.1,

$$x \in \partial f^*(t) \Leftrightarrow t \in \partial(\text{co } f)(x) \;\cdot$$

There exist $x_1, x_1, \ldots, x_{n+1} \in S$ and convex multipliers $\alpha_1, \alpha_1, \ldots, \alpha_{n+1}$ such that

$x = \sum_{j=1}^{n+1} \alpha_j x_j$ and $t \in \partial(\text{co } f)(x) = \bigcap\left\{\partial f(x_j) \mid \alpha_j > 0\right\}$ by Theorem 1.5.2(i).

Now by Proposition 3.2.1(a), we have that

$t \in \partial f(x_j) \Leftrightarrow x_j \in P_S(t)$ for all j=1,2, ... , n+1,

which means

$$t \in \bigcap_{j=1}^{n+1} \partial f(x_j) = \partial(\text{co } f)(x) \Leftrightarrow x = \sum_{j=1}^{n+1} \alpha_j x_j \in (\text{co } P_S)(t).$$

Thus $x \in \partial f^*(t) \Leftrightarrow x \in (\mathrm{co}\, P_S)(t)$, meaning that

$$\partial f^*(t) = (\mathrm{co}\, P_S)(t) \text{ for all } t \in \mathbf{R}^n. \quad ///$$

Theorem 3.1.1: Let $S \subseteq \mathbf{R}^n$ be closed and $S \neq \emptyset$. Then the following statements are equivalent:

(i) S is convex;

(ii) P_S is a function on $\mathbf{R}^n$;

(iii) $d_S{}^2$ is differentiable on $\mathbf{R}^n$;

(iv) $f := I_S + \frac{1}{2}\|.\|^2$ is convex.

Proof: Clearly we have $(i) \Rightarrow (ii)$. Now suppose (ii) holds.

Then $\partial f^*(t) = (\mathrm{co}\, P_S)(t)$ is a singleton for each t. So $f^* := \frac{1}{2}(\|.\|^2 - d_S{}^2)$ is differentiable so that $d_S{}^2$ is differentiable. So $(ii) \Rightarrow (iii)$.

Also clearly , we have $(iv) \Rightarrow (i)$.

To finish the proof, we show that $(iii) \Rightarrow (iv)$. Suppose (iii) holds.

Let $x \in \mathrm{ri}\,(\mathrm{dom\, co}\, f)$. Then there is $s \in \partial(\mathrm{co}\, f)(x)$ by Theorem 1.4.2.Now by Theorem 1.5.2, there exist $x_j \in S$ and positive convex multipliers α_j such that

$$x = \sum_j \alpha_j x_j, \quad (\mathrm{co}\, f)(x) = \sum_j \alpha_j f(x_j), \quad s \in \bigcap_j \partial f(x_j).$$

Then (co f)(x)=f (x) for all $x \in \mathrm{ri}(\mathrm{dom\, co}\, f)$ by Theorem 1.5.1. Now let $x \in (\mathrm{dom\, co}\, f) \setminus \mathrm{ri}(\mathrm{dom\, co}\, f)$. Then there is a sequence (x_k) in ri (dom f) such that $\lim_{k \to +\infty} x_k = x$.

We have

$$f(x) \geq (\mathrm{co}\ f)(x) = \lim_{k \to +\infty} (\mathrm{co}\ f)(x) = \lim_{k \to +\infty} f(x_k)$$

$$\geq f(x_k) \text{ since f is lower emicontinuous.}$$

Thus co f = f, which means that f is convex. Hence $(iii) \Rightarrow (iv)$. ///

3.2. Conjugate of a Partially Quadratic Function

Definition 3.2.1: Let H be a subspace of $\mathbf{R}^n$, and let B be a linear symmetric positive-semidefinite operator on $\mathbf{R}^n$. Then the function g defined by

$$g(x) := \begin{cases} \frac{1}{2}\langle Bx, x\rangle & \text{if } x \in H \\ +\infty & \text{otherwise} \end{cases} \qquad (3.2.1)$$

is called a *partially quadratic function.*

Let $P_H : \mathbf{R}^n \to \mathbf{R}^n$ be the orthogonal projection onto H defined by

$$P_H x := \begin{cases} 0 & \text{if } x \in H^{\perp} \\ x & \text{if } x \in H \end{cases}.$$

Definition 3.2.2: Let $A : \mathbf{R}^n \to \mathbf{R}^n$ be a symmetric linear operator. Then the linear mapping $A^- : \mathbf{R}^n \to \text{im } A$ given by

$$A^- y = x \quad \text{where} \quad Ax = P_{\text{im}A}\, y$$

is called the *pseudo-inverse*, or the *generalized inverse*, of A.

To help us evaluate the conjugate g* of the function g in (3.2.1), let's find f* where f is given by

$$f(x) := \frac{1}{2}\langle Bx, x\rangle \text{ for all } x \in \mathbf{R}^n.$$

$$f^*(s) = \sup\left\{ \langle s, x\rangle - \frac{1}{2}\langle Bx, x\rangle \mid x \in \mathbf{R}^n \right\} = \sup\left\{ \left\langle s - \frac{1}{2}Bx, x\right\rangle \mid x \in \mathbf{R}^n \right\}$$

$$= \frac{1}{2}\langle s, B^{-1}s\rangle.$$

Then $f^*(Bx) = \frac{1}{2}\langle Bx, x\rangle$ so that we have

$$f^*(s) := \begin{cases} \frac{1}{2}\langle s, B^{-1}s\rangle & \text{if } s \in \text{im}B \\ +\infty & \text{otherwise} \end{cases} = \begin{cases} \frac{1}{2}\langle s, B^{-}s\rangle & \text{if } s \in \text{im}B \\ +\infty & \text{otherwise} \end{cases} \qquad (3.2.2)$$

Proposition 3.2.1: The conjugate of g in (3.2.1) is given by

$$g^*(s) = \begin{cases} \frac{1}{2}\langle s, (P_H \circ B \circ P_H)^{-} s\rangle & \text{if } s \in \text{im } B + H^{\perp} \\ +\infty & \text{otherwise} \end{cases}$$

where P_H is the orthogonal projection onto H.

Proof: We have $g(x) = \frac{1}{2}\langle Bx, x\rangle + I_H(x)$ for each $x \in \mathbf{R}^n$.

Then by Proposition 1.3.2, we have $g^* = (f + I_H)^* = (f + I_H)^* \circ P_H$,

where $f(x) := \frac{1}{2}\langle Bx, x\rangle.$

But $(f \circ P_H)(x) = f(P_H x) = \frac{1}{2}\langle (B \circ P_H)x, P_H x\rangle = \frac{1}{2}\langle (P_H \circ B \circ P_H)x, x\rangle$

since P_H is symmetric.

Then by (3.2.2), we have

$$g^*(s) = (f \circ P_H)^*(P_H s) = \begin{cases} \frac{1}{2}\langle s, (P_H \circ B \circ P_H)^{-} s\rangle & \text{if im } (P_H \circ B \circ P_H) \\ +\infty & \text{otherwise} \end{cases}$$

$$= \begin{cases} \frac{1}{2}\langle s, (P_H \circ B \circ P_H)^{-} s\rangle & \text{if im}B + H^{\perp} \\ +\infty & \text{otherwise} \end{cases}. \quad ///$$

3.3. Polyhedral Functions

Proposition 3.3.1: For given $(s_i, b_i) \in \mathbf{R}^n \times \mathbf{R}$, i=1,2, ..., k, let

$f_i(x) := \langle s_i, x\rangle - b_i$ and let

$f(x) := \max\{ f_i(x) | \; i = 1,2, \ldots, k \}$. (3.3.1)

Then for each $s \in \operatorname{co}\{s_1, s_2, \dots, s_k\} = \operatorname{dom} f^*$,

$$f^*(s) = \min\left\{ \sum_{i=1}^{k} \alpha_i b_i \mid \alpha_i \in \Delta_k,\ \sum_{i=1}^{k} \alpha_i s_i = s \right\}, \tag{3.3.2}$$

where $\Delta_k = \{ \alpha_i \mid \alpha_i \geq 0,\ i = 1, 2, \dots, k,\ \text{and } \sum_{i=1}^{k} \alpha_i = 1 \}$ is the unit simplex.

Proof: First observe that $f_i^* = I_{\{s_i\}} + b_i$. Then apply Theorem 2.4.3 with the given s_i being as in (2.4.1) so that we have

$$f^*(s) = \min\left\{ \sum_{i=1}^{k} \alpha_i f_i^*(s_i) \mid s_i \in \operatorname{dom} f_i^*,\ \alpha_i \in \Delta_k,\ \sum_{i=1}^{k} \alpha_i s_i = s \right\}$$

$$= \min\left\{ \sum_{i=1}^{k} \alpha_i b_i \mid \alpha_i \in \Delta_k,\ \sum_{i=1}^{k} \alpha_i s_i = s \right\}. \qquad ///$$

Definition 3.3.1:

(a) The function in (3.1.1) is called a *piecewise affine function.*

(b) A *polyhedral function* is a function f whose epigraph is a closed convex polyhedron.

Remark 3.3.1: A polyhedral function has the general form $g = f + I_P$ where P is a closed convex polyhedron and f is defined by (3.3.1). Thus $g^* = f^* \underset{\vee}{+} \sigma_P$ by Theorem 2.3.2. By Proposition 3.3.1, we have

$$g^*(s) = \min_{\alpha \in \Delta_k} \left\{ \sum_{i=1}^{k} \alpha_i b_i + \sigma_P\left(s - \sum_{i=1}^{k} \alpha_i s_i\right) \right\}. \tag{3.3.3}$$

4.DIFFERENTIABILITY OF A CONJUGATE FUNCTION

4.1. First-Order Differentiability

Theorem 4.1.1: Let $f \in \overline{\text{Conv}}\,\mathbf{R}^n$ be strictly convex. Then

(a) int dom f* $\neq \emptyset$;

(b) f* is continuously differentiable (in the usual sense) on int dom f*.

Proof: Let $x \in \operatorname{dom} f$, arbitrary, and $d \in \mathbf{R}^n$, $d \neq 0$. By convexity of f, we have

$$0 < \frac{f(x_0 - td) - f(x_0)}{t} + \frac{f(x_0 + td) - f(x_0)}{t} \quad \text{for all } t>0, \qquad (4.1.1)$$

since $\dfrac{f(x_0) - f(x_0 - td)}{t} < \dfrac{f(x_0 + td) - f(x_0)}{t}$ for all t>0,

because of increasing slopes.

Taking suprema in (4.1.1), we get $0 < f'_\infty(-d) + f'_\infty(d)$.

But $f_\infty' = \sigma_{\operatorname{dom} f^*}$ (Proposition 1.1.2).

So we have

$$\sigma_{\operatorname{dom} f^*}(d) + \sigma_{\operatorname{dom} f^*}(-d) > 0 ,$$

which means that

$$\sigma_{\operatorname{dom} f^*}(d) - \inf_{s \in \operatorname{dom} f^*}\{\langle s, d\rangle\} > 0 .$$

Then there exist $s \in \operatorname{dom} f^*$ such that $\sigma_{\operatorname{dom} f^*}(s) > \langle s, d\rangle$ for each $d \neq 0$,

since dom f* is closed as $f^* \in \overline{\text{Conv}}\,\mathbf{R}^n$.

But, $\sigma_{\operatorname{dom} f^*}(s) > \langle s, d\rangle$ for all $d \neq 0$ iff $s \in \operatorname{int} \operatorname{dom} f^*$.

So we have that int dom f* $\neq \emptyset$.

We prove (b) by contradiction.

Suppose that $s \in \operatorname{int} \operatorname{dom} f^*$ such that there are $x_1, x_2 \in \partial f^*(s)$, $x_1 \neq x_2$.

Then $s \in \partial f(x_1) \cap \partial f(x_1)$, so that

$$f^*(s) + f(x_1) = \langle s, x_1 \rangle \text{ and } f^*(s) + f(x_2) = \langle s, x_2 \rangle.$$

Then $(\alpha_1 + \alpha_2) f^*(s) + \sum_{i=1}^{2} \alpha_i f(x_i) = \left\langle s, \sum_{i=1}^{2} \alpha_i x_i \right\rangle$,

for $\alpha_1, \alpha_2 \geq 0$ and $\alpha_1 + \alpha_2 = 1$.

This implies that

$$f^*(s) + \sum_{i=1}^{2} \alpha_i f(x_i) = \left\langle s, \sum_{i=1}^{2} \alpha_i x_i \right\rangle \leq f^*(s) + f[\sum_{i=1}^{2} \alpha_i s_i],$$

by Fenchel's inequality.

From this and the convexity of f, it follows that

$$\sum_{i=1}^{2} \alpha_i f(x_i) \leq f[\sum_{i=1}^{2} \alpha_i x_i] \leq \sum_{i=1}^{2} \alpha_i f(x_i) \text{ so that}$$

$f[\sum_{i=1}^{2} \alpha_i x_i] = \sum_{i=1}^{2} \alpha_i f(x_i)$, meaning that f is linear on $[x_1, x_2]$.

But then dom f* is a singleton (f* is the indicator function of a singleton set).

Then int dom f* = ∅, a contradiction to (a). Hence $x_1 = x_2$ so that ∂f^* is single-valued

on int dom f*.

Consequently, f* is continuously differentiable on int dom f*. ///

Theorem 4.1.2: Let $f \in \overline{\text{Conv}}\, \mathbf{R}^n$ be differentiable on the set T: =int dom f.

Then f* is strictly convex on each convex subset $C \subset \nabla f(T)$.

Proof: We prove by contradiction that there is no interval on which f* is affine.

Let $s_1, s_2 \in C$, $s_1 \neq s_2$ such that f* is affine on $[s_1, s_2]$. Let $s := \frac{1}{2}[s_1 + s_1]$.

Then $s \in C \subset \nabla f(T)$ since C is convex.

This implies that $x \in T$ such that $\nabla f(x) = s$; i.e., $x \in \partial f^*(s)$.

Then $0 = f(x) + f^*(s) - \langle s, x \rangle = \frac{1}{2} \sum_{i=1}^{2} [\ f(x) + f^*(s_i) - \langle s_i, x \rangle\ ]$.

But for each i, $f(x)+f^*(s_i)-\langle s_i,x\rangle \geq 0$ by Fenchel's inequality.

Thus $f(x)+f^*(s_i)-\langle s_i,x\rangle = 0,\ i=1,\ 2$.

This implies that $x \in \partial f^*(s_i)$ for each $i=1,2$.

So $s_1, s_2 \in \partial f(x)$, contradicting the hypothesis that f is differentiable. ///

Corollary 4.1.1: Let $f : \mathbf{R}^n \to \mathbf{R}$ be strictly convex, differentiable and 1-coercive.

Then (a) f* is finite-valued on $\mathbf{R}^n$, strictly convex and 1-coercive;

(b) $\nabla f : \mathbf{R}^n \to \mathbf{R}^n$ is bijective and continuous with a continuous inverse;

(c) $f^*(s) = \langle s, (\nabla f)^{-1}(s)\rangle - f((\nabla f)^{-1}(s))$ for all $s \in \mathbf{R}^n$.

Proof: **(a)** $f^*(s) < +\infty$ for all $s \in \mathbf{R}^n$ by Proposition 1.3.4. f* is strictly convex by Theorem 4.1.2 since $f \in \overline{\mathrm{Conv}}\, \mathbf{R}^n$ is differentiable. f* is1-cercive by Proposition 1.3.5 since $\mathrm{dom}\ f = \mathbf{R}^n$.

(b) Suppose there are $x_1, x_2 \in \mathbf{R}^n$ such that $\nabla f(x_1) = \nabla f(x_2)$.

Then $x_1, x_2 \in \partial f^*(s)$ for some $s \in \partial f(x_1) \cap \partial f(x_2)$.

But f* is continuously differentiable by Theorem 4.1.1. This implies that $\partial f^*(s)$ is a singleton, meaning $x_1 = x_2$.

Thus ∇f is one-to-one and hence invertible. Since f^* is strictly convex and f is closed, it follows that ∇f is continuous on $\mathbf{R}^n = \mathrm{int\ dom}\ f$.

By Corollary 1.4.1, $s \in \partial f(x) \Leftrightarrow x \in \partial f^*(s)$. Since both f and f* are differentiable on $\mathbf{R}^n$, this means

$$s = \nabla f(x) \Leftrightarrow x = \nabla f^*(s),\ \forall s \in \mathbf{R}^n.$$

This implies that $(\nabla f)^{-1}(s) = x = \nabla f^*(s)$. But ∇f^* is continuous since f* is continuously differentiable. Thus $(\nabla f)^{-1}$ is continuous.

(c) Immediate from the definition of conjugate. ///

4.2. Second – Order Differentiability

A. Lipschitz Continuity of Gradient Mapping

Definition 4.2.1: Let A be a nonempty convex set in $\mathbf{R}^n$. A function $f : A \to \mathbf{R}$ is said to be *strongly convex with modulus* C on A iff

$$f[\alpha x_1 + (1-\alpha)x_2] \leq \alpha f(x_1) + (1-\alpha)f(x_2) - \frac{1}{2}C\alpha(1-\alpha)\|x_1 - x_2\|^2 \qquad (4.2.1)$$

for all $x_1, x_2 \in A$ and $\alpha \in (0,1)$

Remark 4.2.1: Inequality (4.2.1) can be written as

$$f(x_2) \geq f(x_1) + \langle s, x_2 - x_1 \rangle + \frac{1}{2}C\,\|x_2 - x_1\|^2 \text{ for all } s \in \partial f(x_1) \qquad (4.2.2)$$

or equivalently,

$$\langle s_2 - s_1, x_2 - x_1 \rangle \geq C\,\|x_2 - x_1\|^2 \quad \text{for all } s_1 \in \partial f(x_1), s_2 \in \partial f(x_2) \qquad (4.2.3)$$

Definition 4.2.2: A function $g : \mathbf{R}^n \to \mathbf{R}$ is said to be *Lipschitz-continuous* (or a *Lipschitzian*) on $A \subseteq \mathbf{R}^n$ with constant $L \in \mathbf{R}$ iff

$$|g(x_1) - g(x_2)| \leq L\|x_1 - x_2\| \qquad \text{for all } x_1, x_2 \in A.$$

Theorem 4.2.1: Let $f : \mathbf{R}^n \to \mathbf{R}$ be strongly convex with modulus C >0 on $\mathbf{R}^n$.

Then (a) $\operatorname{dom} f^* = \mathbf{R}^n$;

(b) ∇f^* is Lipschitizian with constant $\frac{1}{C}$ on $\mathbf{R}^n$.

Proof: **(a)** Let $x_0 \in \mathbf{R}^n$ and $s_0 \in \partial f(x_0)$, both fixed.

Then for all $d \in \mathbf{R}^n$, $d \neq 0$, $t \geq 0$, it holds that

$$f(x_0 + td) \geq f(x_0) + \langle s_0, x_0 + td - x_0 \rangle + \frac{1}{2}C\|x_0 + td - x_0\|^2$$

$$= f(x_0) + t\langle s_0, d\rangle + \frac{1}{2}Ct^2\|d\|^2.$$

Since $s_0 \in \partial f(x_0)$, and x_0 is arbitrary in $\mathbf{R}^n$, it follows that $f \in \overline{\mathrm{Conv}}\,\mathbf{R}^n$.

So, $f = f^{**}$.

By proposition 1.2.2, $\sigma_{\mathrm{dom}\, f^*}(d) = (f^{**})'_\infty(d) = f'_\infty(d) = +\infty$.

Then $\mathrm{dom}\ f^* = \mathbf{R}^n$;

(b) f is strictly convex since it is strongly convex . Then by theorem 4.1.1, f* is continuously differentiable on $\mathrm{int\ dom}\ f^* = \mathbf{R}^n$. Using the description (4.2.3) of strong convexity, we have

$$\langle s_1 - s_2, x_1 - x_2\rangle \geq C\|x_1 - x_2\|^2 \text{ with } s_i \in \partial f(x_i),\ i = 1,2.$$

Then $x_1 \in \partial f^*(s_1)$ and $x_2 \in \partial f^*(s_2)$, meaning that

$$x_1 = \nabla f^*(s_1) \text{ and } x_2 = \nabla f^*(s_2).$$

By Cauchy-Schwarz inequality ,we have

$$\|s_1 - s_2\|\,\|x_1 - x_2\| \geq \langle s_1 - s_2, x_1 - x_2\rangle \geq C\|x_1 - x_2\|^2.$$

Thus $\|\nabla f^*(s_1) - \nabla f^*(s_2)\| = \|x_1 - x_2\| \leq \frac{1}{C}\|s_1 - s_2\|$.

So ∇f^* is Lipschitzian with constant $\frac{1}{C}$. ///

Theorem 4.2.2: Let $f : \mathbf{R}^n \to \mathbf{R}$ be convex such that ∇f is Lipschitzian with constant L>0 on $\mathbf{R}^n$.Then f* is strongly convex with modulus 1/L on each convex subset $C \subset \mathrm{dom}\,\partial f^*$. In particular,

$$\langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2\rangle \geq \frac{1}{L}\|\nabla f(x_1) - \nabla f(x_2)\|^2, \tag{4.2.4}$$

for all $x_1, x_2 \in \mathbf{R}^n$.

Proof:

Let $s_1, s_2 \in \operatorname{dom} \partial f^* \subset \operatorname{dom} f^*$; let $s, s' \in [s_1, s_2]$.

Now $f(y) = f(x) + \int_0^1 \langle \nabla f[x + t(y-x)], y - x \rangle dt$

$$= f(x) + \langle \nabla f(x), y - x \rangle + \int_0^1 \langle \nabla f[x + t(y-x)] - \nabla f(x), y - x \rangle dt$$

But $\int_0^1 \langle \nabla f[x + t(y-x)] - \nabla f(x),\ y - x \rangle dt \le \int_0^1 \| \nabla f[x + t(y-x)] - \nabla f(x) \| \, \| y - x \| dt$,

by Cauchy-Schwarz inequality,

$$\le \int_0^1 Lt \| y - x \| \| y - x \| dt \ = \frac{1}{2} L \| y - x \|^2 .$$

So $f(y) \le f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} L \| y - x \|^2 = \underbrace{f(x) - \langle s, x \rangle}_{=-f^*(s)} + \langle s, y \rangle + \frac{1}{2} L \| y - x \|^2$,

taking $x \in \partial f^*(s)$,

$$= - f^*(s) + \langle s, y \rangle + \frac{1}{2} L \| y - x \|^2 .$$

Now adding $\langle s', y \rangle$ to both side and taking suprema, we get

$$f^*(s) + \sup_y \left\{ \langle s' - s, y \rangle - \frac{1}{2} L \| y - x \|^2 \right\} \le f^*(s') . \qquad (4.2.5)$$

But $\sup_y \left\{ \langle s' - s, y \rangle - \frac{1}{2} L \| y - x \|^2 \right\} = h^*(s' - s)$, where $h(y) := \frac{1}{2} L \| y - x \|^2$.

By Proposition 1.3.1(b) and (f), we have

$$h^*(s' - s) = \langle s' - s, x \rangle + \frac{1}{2L} \| s - s \|^2 .$$

So (4.2.5) becomes

$$f^*(s') \ge f^*(s) + \langle s' - s, x \rangle + \frac{1}{2L} \| s' - s \|^2 \qquad (4.2.6)$$

for all $s, s' \in [s_1, s_2]$ and for all $x \in \partial f^*(s)$.

Now replacing s' in (4.2.6) by s_1 and s_2 we get, respectively,

$$f^*(s_1) \geq f^*(s) + \langle s_1 - s, x \rangle + \frac{1}{2L} \|s_1 - s\|^2 \qquad (4.2.7)$$

and

$$f^*(s_2) \geq f^*(s) + \langle s_2 - s, x \rangle + \frac{1}{2L} \|s_2 - s\|^2 \qquad (4.2.8)$$

Convex combination of (4.2.7) and (4.2.8) gives

$$\alpha f^*(s_1) + (1-\alpha) f^*(s_2) \geq f^*(s + \frac{1}{2L}[\alpha\|s_1 - s\|^2 + (1-\alpha)\|s_2 - s\|^2$$

$$= f^*[\alpha s_1 + (1-\alpha)s_2] + \frac{1}{2L}\left[\alpha(1-\alpha)^2 \|s_1 - s_2\|^2 + (1-\alpha)\alpha^2 \|s_1 - s_2\|^2\right]$$

$$= f^*[\alpha s_1 + (1-\alpha)s_2] + \frac{1}{2L}\left[\alpha(1-\alpha)^2 + (1-\alpha)\alpha^2\right]\|s_1 - s_2\|^2$$

$$= f^*[\alpha s_1 + (1-\alpha)s_2] + \frac{1}{2L}\alpha(1-\alpha)\|s_1 - s_2\|^2.$$

So f* is strongly convex.

To prove the particular case (4.2.4), we replace (s, s') by (s_1, s_2) and (s_2, s_1) in (4.2.6) and get respectively

$$f^*(s_2) \geq f^*(s_1) + \langle s_2 - s_1, x_1 \rangle + \frac{1}{2L}\|s_2 - s_1\|^2 \qquad (4.2.9)$$

$$f^*(s_1) \geq f^*(s_2) + \langle s_1 - s_2, x_2 \rangle + \frac{1}{2L}\|s_1 - s_2\|^2 \qquad (4.2.10)$$

for $x_i \in \partial f^*(s_i)$, $i = 1, 2$.

Adding (4.2.9) and (4.2.10), we get

$$\langle x_1 - x_2, s_1 - s_2 \rangle \geq \frac{1}{L} \|s_1 - s_2\|^2.$$

This is just (4.2.4) since f is differentiable. It holds for all $x_1, x_2 \in \mathbf{R}^n = \operatorname{dom} \nabla f$. ///

B. Second- Order Approximations

Definition 4.2.3: A function $\varphi \in \overline{\mathrm{Conv}}\, \mathbf{R}^n$ is said to be *directionally quadratic* if and only if it is positively homogeneous of degree 2, i.e., iff $\varphi(tx) = t^2 \varphi(x) \quad \text{for all} \quad x \in \mathbf{R}^n, \quad t > 0$.

Remark 4.2.2: If φ is directionally quadratic, then so φ^*. Moreover, φ (0)=0=φ^*(0).

Lemma 4.2.1: For a directionally quadratic function φ, the following properties are equivalent:

(i) φ is finite everywhere;

(ii) $\nabla\varphi$ (0) exists (and is equal to 0);

(iii) There is $C \geq 0$ such that $\varphi(x) \leq \frac{1}{2}C\|x\|^2$ for all $x \in \mathbf{R}^n$;

(iv)There is $C \geq 0$ such that $\varphi^*(s) \geq \frac{1}{2}C\|s\|^2$ for all $s \in \mathbf{R}^n$;

(v) $\varphi^*(s) > 0$ for all s $\neq 0$.

Proof: First we prove $(i) \Leftrightarrow (ii) \Rightarrow (iii)$.

Suppose (i) holds. Since φ is convex and $\mathrm{dom}\ \varphi = \mathbf{R}^n$, it follows that φ is continuous on $\mathbf{R}^n$. Then φ has a maximum value, say $\frac{1}{2}C \geq 0$, on the unit sphere so that

$$\varphi(x) = \varphi(\|x\|\frac{x}{\|x\|}) = \|x\|^2\varphi(\frac{x}{\|x\|}) \leq \|x\|^2\frac{1}{2}C,$$

since f is directionally quadratic.

So we have (iii).

Now, $\varphi'(0, .) \equiv 0$. But $\varphi'(0, .) = \langle\nabla\varphi(0), .\rangle$ since f is differentiable.

Then $\nabla\varphi(0) = 0$, and hence (ii).

Conversely, if (ii) holds, then the existence of $\nabla\varphi(0)$ implies finiteness of φ in the neighborhood of 0, and by homogeneity, we have finiteness on the whole space.

Now we prove $(iii) \Rightarrow (iv) \Rightarrow (v) \Rightarrow (i)$.

(iii)$\Rightarrow$(iv) by Proposition 1.3.1, and (iv)$\Rightarrow$(v) trivially.

Being directionally quadratic, φ^* is 1-coercive. Then $\varphi = (\varphi^*)^*$ is finite everywhere by Proposition 1.3.4. So (v) $\Rightarrow$(i). ///

Definition 4.2.4: A directionally quadratic function φ_s is said to define a minorization *to second order* of $f \in \overline{\text{Conv}}\,\mathbf{R}^n$ at $x_0 \in \operatorname{dom} f$, and *associated with* $s \in \partial f(x_o)$

if and only if $f(x_0 + h) \geq f(x_0) + \langle s, h \rangle + \varphi_s + o(\|h\|^2)$.

We say, likewise, that φ_s defines a *majorization to second order* iff

$$f(x_0 + h) \leq f(x_0) + \langle s, h \rangle + \varphi_s + o(\|h\|^2).$$

Proposition 4.2.1: Suppose there is a directionally quadratic function φ_s such that there is C>0 with

$$\varphi_s(h) \geq \frac{1}{2} C \|h\|^2 \quad \text{for all } h \in \mathbf{R}^n, \qquad (4.2.11)$$

and defining a minorization to second order of $f \in \overline{\text{Conv}}\,\mathbf{R}^n$ associated with $s \in \partial f(x_0)$.

Then φ_s * defines a majorization of f* at s, associated with x_0;

i.,e., $f^*(s+p) \leq f^*(s) + \langle x_0, p \rangle + \varphi_s^*(p) + o(\|p\|)^2$

In particular, $\nabla f^*(s) = x_0$.

Proof: Since $\varphi_s(h) \geq \frac{1}{2} C \|h\|^2$ for all $h \in \mathbf{R}^n$, we have by Proposition 1.3.1(g),

$$\varphi_s^*(p) \leq \frac{1}{2}\frac{1}{C} \|p\|^2 \text{ for all } p \in \mathbf{R}^n.$$

Then φ_s * is finite everywhere by Lemma 4.2.1. If we show the majorization of f* at s, then it will follow that f* is differentiable at s, and that $\nabla f^*(s) = x_0$.

Define g on $\mathbf{R}^n$ by $g(h) := f(x_0 + h) - f(x_0) - \langle s, h \rangle$. Then

$$g(h) \geq \varphi_s(h) + o(\|h\|^2), \qquad (4.2.12)$$

since φ_s defines a minoriazition of f to second order.

Now,

$$g^*(p) = f^*(s+p) - \langle s+p, x_0 \rangle + f(x_0) - \langle p, x_0 \rangle$$

$$= f^*(s+p) - f^*(s) - \langle p, x_0 \rangle .$$

So we need only show that $g^*(p) \le \varphi_s{}^*(p) + 0(\|h\|^2)$ for all $p \in \mathbf{R}^n$.

Step1: By (4.2.12), for any $\varepsilon > 0$, there is a $\delta > 0$ such that

$$\|h\| < \delta \Rightarrow g(h) \ge g(h) \ge \varphi_s(h) - \varepsilon\|h\|^2.$$

Then by (4.2.11), we have that

$$\|h\| < \delta \Rightarrow g(h) \ge (\frac{1}{2}C - \varepsilon)\|h\|^2 \qquad (4.2.13)$$

or

$$\|h\| < \delta \Rightarrow g(h) \ge \varphi_s(h) - \frac{2\varepsilon}{C}\varphi_s(h) = (1 - \frac{2\varepsilon}{C})\varphi_s(h) \qquad (4.2.14)$$

Step2: Let $\|h\| > \delta$. Putting $\widetilde{h} := \frac{\delta h}{\|h\|}$, we get

$$\|\widetilde{h}\| = \delta \le \delta,$$

so that by (4.2.13), we have

$$g\left(\frac{\delta h}{\|h\|}\right) \ge (\frac{1}{2}C - \varepsilon)\delta^2 .$$

Since *g* is convex on [0,h] and g (0)=0, we have

$$g(\frac{\delta}{\|h\|}h) \le \frac{\delta}{\|h\|}g(h) .$$

This means that $\frac{\delta}{\|h\|}g(h) \ge (\frac{1}{2}C - \varepsilon)\delta$ so that

$$g(h) \ge (\frac{1}{2}C - \varepsilon)\|h\| \qquad (4.2.15)$$

Now if $\|p\| \le (\frac{1}{2}C - \varepsilon)\delta =: \delta'$, then we have

$$\langle p,h\rangle - g(h) \le \|p\|\|h\| - g(h) \le (\frac{1}{2}c - \varepsilon)\delta\|h\| - g(h),$$

by Cauchy-Schwarz inequality and (4.2.15)

$$\le (\frac{1}{2}c - \varepsilon)\|h\|^2 - g(h) \le 0 \quad \text{whenever } \|h\| > \delta.$$

Since $g^* \ge 0$, we have

$$g^*(p) = \sup_{\|h\|\le\delta} \{\langle p, h\rangle - g(h)\} \le \sup_{\|h\|\le\delta} \left\{\langle p, h\rangle - (1 - \frac{2\varepsilon}{C})\varphi_s(h)\right\}, \quad \text{from (4.2.14)},$$

$$\le \left(\left(1 - \frac{2\varepsilon}{C}\right)\varphi_s\right)^*(p) \qquad \text{for } \|p\| \le \delta'.$$

Step3: We have for $p \in B(0,\delta')$ that

$$g^*(p) = (1 - \frac{2\varepsilon}{C})\varphi_s^*(p) = (1 - \frac{2\varepsilon}{C})\varphi_s^*\left(\frac{p}{1 - \frac{2\varepsilon}{C}}\right)$$

$$= [\frac{1}{1 - 2\varepsilon/C}]\ \varphi_s^*\left(\frac{p}{1 - \frac{2\varepsilon}{C}}\right)$$

$$= (\frac{C}{C - 2\varepsilon})\varphi_s^*(p)$$

$$\le \varphi_s^*(p) + (\frac{2\varepsilon}{C - 2\varepsilon})\frac{1}{2C}\|p\|^2,$$

by conjugating both side of (4.2.11) and using Proposition 1.3.1.

Given $\varepsilon' > 0$, choose ε in step 1 such that

$$\frac{\varepsilon}{C(C - 2\varepsilon})\le \varepsilon'.$$

This gives $\delta > 0$ and $\delta' > 0$ such that

$$g^*(p) \le \varphi_s^*(p) + \varepsilon'\|p\|^2 \quad \text{for all } p \in B(0,\delta'). \quad ///$$

Proposition 4.2.2: Suppose there is a directionally quadratic function φ_s such that there is a c>0 with

$$\varphi_s(h) \geq \frac{1}{2}C\|h\|^2 \text{ for all } h \in \mathbf{R}^n, \qquad (4.2.15)$$

and defining a majorization to a second order of $f \in \overline{\mathrm{Conv}}\,\mathbf{R}^n$, associated with $s \in \partial f(x_0)$.

Then φ_s * defines a minorization to a second order of f* at s, associated with x_0.

Proof: Similar to the proof of Proposition 4.2.1. (Cf. [4]).

Definition 4.2.5: If a directionally quadratic function φ_s defines both a majorization and a minorization to a second order of $f \in \overline{\mathrm{Conv}}\,\mathbf{R}^n$ at x_0, associated with s, then φ_s is said to define an *approximation to second order* of f.

Propositions 4.2.1 and 4.2.2 result in:

Corollary 4.2.1: Let $f \in \overline{\mathrm{Conv}}\,\mathbf{R}^n$ have a quadratic (second order) approximation at x_0; i.e., suppose for $s := \nabla f(x_0)$ and for some symmetric positive semidefinite linear operator A, it holds that

$$f(x_0+h) = f(x_0) + f(x_0 + h) = f(x_0) + \langle s, h \rangle + \frac{1}{2}\langle Ah, h \rangle + o(\|h\|^2).$$

If A is positive definite, then f* has also a quadratic approximation at s, namely,

$$f^*(s+p) = f^*(s) + \langle p, x_0 \rangle + \frac{1}{2}\langle A^{-1}p, p \rangle + o(\|p\|^2).$$

Corollary 4.2.1 can be generalized as follows:

Corollary 4.2.2: Let $f : \mathbf{R}^n \to \mathbf{R}$ be convex, twice differentiable and 1-coercive such that $\nabla^2 f(x)$ is positive definite for all $x \in \mathbf{R}^n$. Then f* is also twice differentiable, 1-coercive and $\nabla^2 f^*(s) = \left[\nabla^2 f(\nabla f^{-1}(s))\right]^{-1}$ for all $s \in \mathbf{R}^n$.

Proof: Since f is 1-coerciive, it follows by Proposition 1.3.4 that $\operatorname{dom} f^* = \mathbf{R}^n$.

So $\operatorname{im} \nabla f = \operatorname{dom} f^* = \mathbf{R}^n$. Since $\operatorname{dom} f^* = \mathbf{R}^n$, we have by Proposition 1.3.5(ii) that f* is 1-coercive.

Since f is twice differentiable, we have

$$f(x+h) = f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2}\langle \nabla^2 f(x)h, h \rangle + o\left(\|h\|^2\right) \qquad \text{for all } s \in \mathbf{R}^n.$$

Then by Corollary 4.2.1, with $A = \nabla^2 f(x)$, we have

$$f^*(s+p) = f^*(s) + \langle p, x \rangle + \frac{1}{2}\left\langle \left(\nabla^2 f(x)\right)^{-1} p,\ p \right\rangle + o\left(\|p\|^2\right).$$

This means that f* is twice differentiable.

Moreover, $\nabla^2 f^*(s) = \left[\nabla^2 f\left(\nabla f^{-1}(s)\right)\right]^{-1}$ for all $s \in \mathbf{R}^n$; x is replaced by $\nabla f^{-1}(s)$ since $s = \nabla f(x)$. ///

REFERENCES

[1] Deumlich , R.: Optimization and Theory of Approximation, Textbook for Lecture, Addis Ababa University, Addis Ababa, 1997.

[2] Deumlich, R.: Functional Analysis I, Textbook for Lecture, Addis Ababa University, Addis Ababa, 1998.

[3] Deumlich, R.: Functional Analysis II, Textbook for Lecture, Addis Ababa University, Addis Ababa, 1998.

[4] Hiriart-Urrty,J.-B. and Lemarechal, C.: Convex Analysis and Minimization Algorithms, Springer-Verlag, Volumes I and II, Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, Barcelona, Budapest, 1993.

[5] Mital, K.V.: Optimization Methods in Operations Research and Systems Analysis, 2^{nd} ed. , Wiley Eastern Ltd, 1983.