2.3 Conjugate Functions

In the next chapter, we will consider the concept of duality. One notion that is crucial in the theory of duality is the conjugate function.

Definition:(Conjugate function) Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function. The *conjugate function* of f is the function $f^* : \mathbb{R}^n \to [-\infty, \infty]$ defined by

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

Note that f^* is convex even if f is not convex.

Examples of conjugate functions

1. $f(x) = ||x||_1$

$$\begin{aligned} f^*(a) &= \sup_{x \in \mathbb{R}^n} \langle x, a\rangle - ||x||_1 \\ &= \sup \sum (a_n x_n - |x_n|) \\ &= \begin{cases} 0 & ||a||_\infty \leq 1 \\ \infty & \text{otherwise} \end{cases} \end{aligned}$$

2. $f(x) = ||x||_\infty$

$$\begin{aligned} f^*(a) &= \sup_{x \in \mathbb{R}^n} \sum a_n x_n - \max_n |x_n| \\ &\leq \sup \sum |a_n||x_n| - \max_n |x_n| \\ &\leq \max_n |x_n| ||a||_1 - \max_n |x_n| \\ &\leq \sup ||x||_\infty (||a||_1 - 1) \\ &= \begin{cases} 0 & ||a||_1 \leq 1 \\ \infty & \text{otherwise} \end{cases} \end{aligned}$$

If $||a||_1 \leq 1$, $\langle 0, a\rangle - ||0||_\infty = 0$, $f^*(a) \geq 0$ in this case.
If $||a||_1 > 1$, then $\langle x, a\rangle - ||x||_\infty$ is unbounded. Hence

$$f^*(a) = \begin{cases} 0 & ||a||_1 < 1 \\ \infty & \text{otherwise} \end{cases}$$

We can also consider the conjugate of f^* (double conjugate of f). It is given by

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

It is natural to ask whether $f = f^{**}$. Indeed, this is true under some conditions.

Theorem: Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function. Then:

1. $f(x) \geq f^{**}(x)$ for all $x \in \mathbb{R}^n$.
2. If f is closed, proper and convex, then $f(x) = f^{**}(x)$.

Proof. 1 For all x and y, we have

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

So $f(x) \geq \langle x, y \rangle - f^*(y)$ for all x, y. (*)
Therefore, $f(x) \geq \sup\{\langle x, y \rangle - f^*(y)\} = f^{**}(x)$.
2 By (1), we have epi$f \subseteq$ epif^{**}. We need to show epi$f^{**} \subseteq$ epi$f$.
It suffices to show that $(x, f^{**}(x)) \in$epi$f$. So suppose not.
Since epi$f$ is a closed convex set, $(x, f^{**}(x))$ can be strictly separated from epi$f$.
Hence

$$\langle y, z \rangle + bs < c < \langle y, x \rangle + bf^{**}(x)$$

for some y, b, c, and for all $(z, s) \in$ epi$f$.
We may assume $b \neq 0$ (If not, add $\epsilon(\overline{y}, -1)$ to (y, b) for some $\overline{y} \in$domf^*).
We must have $b < 0$. Since if $b > 0$, we have a contradiction by choosing s large.
Therefore, we further assume $b = -1$. Hence, in particular, we have

$$\langle y, z \rangle - f(z) < c < \langle y, x \rangle - f^{**}(x)$$

Then taking supremum over z, we have

$$f^*(y) + f^{**}(x) < \langle x, y \rangle$$

This is a contradiction to (*). Hence epif^{**} = epi$f$.
Therefore, $f = f^{**}$. □

2.4 Subgradient of Convex Function

In this section, we introduce the crucial concept of subgradient for convex functions. It acts as generalized derivative for nonsmooth functions and has many applications in optimization theory.

Definition:(Subgradient) Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function and let $\overline{x} \in$ dom$f$. An element $g \in \mathbb{R}^n$ is called a *subgradient* of f at $\overline{x}$ if

$$f(x) - f(\overline{x}) \geq \langle g, x - \overline{x} \rangle \text{ for all } x \in \mathbb{R}^n$$

The collection of all subgradients of f is denoted by $\partial f(\overline{x})$.

Proposition: Let f be a convex function and let $\overline{x} \in$ int(dom$f$), then $\partial f(\overline{x})$ is nonempty and compact.

Proof. Since f is convex, epi$f$ is a convex set.
By the supporting hyperplane theorem to epi$f$ and the point $(\overline{x}, f(\overline{x}))$, there exists $(a, b) \neq 0$ such that

$$\left\langle \begin{bmatrix} a \\ b \end{bmatrix}, \left(\begin{bmatrix} x \\ t \end{bmatrix} - \begin{bmatrix} \overline{x} \\ f(\overline{x}) \end{bmatrix} \right) \right\rangle \leq 0, \text{ for all } (x, t) \in \text{epi} f$$

By considering $(\overline{x}, t) \in \text{epi}f$, we must have $b \leq 0$. Also

$$\langle a, x - \overline{x}\rangle + b(f(x) - f(\overline{x})) \leq 0 \text{ for all } x$$

Suppose $b = 0$, this implies $\langle a, x - \overline{x}\rangle \leq 0$.
This is impossible since $\overline{x} \in \text{int}(\text{dom}f)$. Hence, $b < 0$. Then

$$\Big\langle -\frac{a}{b}, x - \overline{x}\Big\rangle \leq f(x) - f(\overline{x})$$

Therefore, $-\frac{a}{b} \in \partial f(\overline{x}) \neq \emptyset$.
Recall that a function is locally Lipschitz continuous on the $\text{int}(\text{dom}f)$.
So there exists $\epsilon > 0$ such that

$$f(x) - f(y) \leq L||x - y||, \text{ for all } x, y \in B(\overline{x}; \epsilon)$$

Let $g \in \partial f(\overline{x})$. Consider $x = \overline{x} + \frac{\epsilon g}{||g||}$, then

$$\epsilon||g|| = \langle g, x - \overline{x}\rangle \leq f(x) - f(\overline{x}) \leq L||x - \overline{x}|| = L\epsilon$$

Then we have $||g|| \leq L$. Therefore $\partial f(\overline{x})$ is bounded.
It follows from the definition that $\partial f(\overline{x})$ is closed and hence compact. □

For a differentiable convex function, the subdifferential is just the usual gradient.
Proposition: Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex and differentiable at $\overline{x} \in \text{int}(\text{dom}f)$. Then $\partial f(\overline{x}) = \{\nabla f(\overline{x})\}$.

Proof. Since f is convex, we have

$$\langle \nabla f(\overline{x}), x - \overline{x}\rangle \leq f(x) - f(\overline{x}) \text{ for all } x \in \mathbb{R}^n$$

So $\nabla f(\overline{x}) \in \partial f(\overline{x})$.
Conversely, suppose $g \in \partial f(\overline{x})$. Since f is differentiable at $\overline{x}$, then for all $\epsilon > 0$, there exists $\delta > 0$ such that

$$f(x) - f(\overline{x}) - \langle \nabla f(\overline{x}), x - \overline{x}\rangle \leq \epsilon||x - \overline{x}|| \text{ for all } x \text{ with } ||x - \overline{x}|| < \delta$$

Then

$$\langle g - \nabla f(\overline{x}), x - \overline{x}\rangle \leq \epsilon||x - \overline{x}|| \text{ for all } x \text{ with } ||x - \overline{x}|| < \delta$$

Hence $||g - \nabla f(\overline{x})|| \leq \epsilon$. Since ϵ is arbitrary, this shows that $g = \nabla f(\overline{x})$.
Therefore, $\partial f(\overline{x}) = \{\nabla f(\overline{x})\}$. □

Example: Let $f : \mathbb{R} \to \overline{\mathbb{R}}$ be defined by

$$f(x) := \begin{cases} 0 & x \in [-1, 1] \\ |x| - 1 & x \in [-2, 1) \cup (1, 2] \\ \infty & x \in (-\infty, -2) \cup (2, \infty) \end{cases}$$

For $x \in (-2,1)$, $(-1,1)$ and $(1,2)$, f is differentiable, hence $\partial f(x) = \{\nabla f(x)\}$.
For $x \in (-\infty,-2) \cup (2,\infty)$, $f(x) = \infty$, hence $\partial f(x) = \emptyset$.
For $x = 1$, we show that $\partial f(x) = [0,1]$. Let $g \in \partial f(1)$. Then

$$f(y) \geq g(x-1) \text{ for all } y$$

If $y \in [1,2]$, then $x - 1 \geq g(x-1)$, that is $1 \geq g$.
If $y \in [-1,1]$, then $0 \geq g(x-1)$, so $g(1-x) \geq 0$ and $g \geq 0$.
It is easy to check that for $g \in [0,1]$, g satisfies

$$f(y) \geq f(1) + g(x-1) \text{ for all } y$$

Hence, $\partial f(1) = [0,1]$.
The subdifferential of other points can be found similarly.
We have

$$\partial f(x) = \begin{cases} \emptyset & x \in (-\infty,-2) \cup (2,\infty) \\ (-\infty,-1] & x = -2 \\ \{-1\} & x \in (-2,-1) \\ [-1,0] & x = -1 \\ \{0\} & x \in (-1,1) \\ [0,1] & x = 1 \\ \{1\} & x \in (1,2) \\ [1,\infty) & x = 2 \end{cases}$$

The following results show the relationship between subgradients and conjugate of convex functions.

Proposition: Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function with $\text{dom} f \neq \emptyset$. Then

$$\langle x, y \rangle \leq f(x) + f^*(y) \text{ for all } x, y$$

Proof. By the definition of conjugate function, $f^*(y) \geq \langle x, y \rangle - f(x)$. □

Theorem: Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex with $x \in \text{dom} f$. Then $g \in \partial f(x)$ if and only if

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

Proof. Suppose $g \in \partial f(x)$, then

$$f(x) + \langle g, y \rangle - f(y) \leq \langle g, x \rangle, \text{ for all } y$$

Then $f(x) + f^*(g) \leq \langle g, x \rangle$. Hence by the above proposition, we have

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

Suppose $f(x) + f^*(g) = \langle g, x \rangle$, then by the definition of conjugate function,

$$f^*(g) \geq \langle g, y \rangle - f(y) \text{ for all } y$$

Since $f^*(g) = \langle g, x \rangle - f(x)$, we have

$$\langle g, x \rangle - f(x) \geq \langle g, y \rangle - f(y) \text{ for all } y$$

Therefore, $g \in \partial f(x)$. $\square$