Notes 04: Conjugate function

1. Convex functions

**Definition 1.1.** A real-valued function \( f : \mathbb{R}^n \to \mathbb{R} \) on \( n \)-dimensional Euclidean space is convex if for every \( x, y \in \mathbb{R}^n \) and every \( 0 \leq \lambda \leq 1 \), we have
\[
f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).
\]

In other words, if we consider the graph of a function, defined as \( \{(x, f(x)) \mid x \in \mathbb{R}^n\} \subseteq \mathbb{R}^n \times \mathbb{R} \), then \( f \) is convex if the line segment between any two points of the graph lies above or on the graph.

2. Conjugate function

We now define a dual object for every function \( f : \mathbb{R}^n \to \mathbb{R} \), called its conjugate.

We have defined dual objects for sets, using support functions. To define a dual object for a function, we want to first turn \( f : \mathbb{R}^n \to \mathbb{R} \) into a set.

Given a function \( f : \mathbb{R}^n \to \mathbb{R} \) (not necessarily convex), its epigraph is \( \text{epi} \ f = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq t\} \).

Note that a function is convex if and only if its epigraph is a convex set, as can be easily checked.

The conjugate of a function \( f \) is essentially the support function of \( \text{epi} \ f \), “simplified”.

The support function of \( \text{epi} \ f \) is \( S_{\text{epi} f}(y, s) = \sup \{\langle y, x \rangle + st \mid x \in \mathbb{R}^n, f(x) \leq t \} \).

But if \( s > 0 \), \( S_{\text{epi} f} \) says nothing about \( f \), because the supremum is \( + \infty \) by taking arbitrarily large \( t \). If \( s = 0 \), \( S_{\text{epi} f} \) also says nothing about \( f \). Only when \( s < 0 \) does \( S_{\text{epi} f} \) capture information about \( f \). In this case we always choose \( t = f(x) \) in the supremum without changing the outcome. Given any \( (y, s) \) with \( s < 0 \), we can renormalize \( (y, s) \) so that \( s = -1 \). This motivates the following definition.

**Definition 2.1.** Given a function \( f : \mathbb{R}^n \to \mathbb{R} \), its conjugate \( f^* : \mathbb{R}^n \to \mathbb{R} \) is defined as
\[
f^*(y) = \sup \{\langle y, x \rangle - f(x) \mid x \in \mathbb{R}^n\}.
\]

Turns out \( f^* \) is always convex even when \( f \) is not, since it is the pointwise supremum of convex (in this case, affine) functions of \( y \).

Under an additional technical assumption, we can indeed recover \( f \) as the conjugate of \( f^* \).

**Theorem 2.2.** If \( F \) is convex and its epigraph is a closed set, then \( f^{**} = f \).

We will not prove this theorem; see [BV, Exercise 3.39].

In fact \( f^{**} \) is the lower semi-continuous envelop of \( f \), that is, the largest lower semi-continuous function upper-bounded by \( f \). (We will not define semi-continuous here; just think of it as a weaker notion than continuity.)

**Proposition 2.3** (Fenchel inequality). For any \( x, y \in \mathbb{R}^n \), \( \langle y, x \rangle \leq f^*(y) + f(x) \).

The proof follows from the definition of conjugate.

Examples of functions and their conjugates:

- **Negative entropy.** \( f(x) = x \log x \), defined for \( x \geq 0 \). Then \( f^*(y) = \sup_{x \geq 0} yx - x \log x \)
  
The supremum is achieved when \( 0 = \frac{d}{dx}(yx - x \log x) = y - x(\frac{1}{2}) - \log x \iff x = e^{y-1} \)
  
  Hence \( f^*(y) = ye^{y-1} - e^{y-1}(y - 1) = e^{y-1} \)

- **Strictly convex quadratic form.** \( f(x) = \frac{1}{2}x^\top Qx \), where \( Q \) is a symmetric positive definite matrix. Then \( f^*(y) = \sup_x y \top x - \frac{1}{2}x^\top Qx \).
  
The supremum is achieved when \( 0 = \nabla(y \top x - \frac{1}{2}x^\top Qx) = y - Qx \iff x = Q^{-1}y \)
  
  Hence \( f^*(y) = y \top Q^{-1}y - \frac{1}{2}(y \top Q^{-1}y)Q(Q^{-1}y) = \frac{1}{2}y \top Q^{-1}y \)

- **Log-sum-exp.** \( f(x) = \log(\sum_{1 \leq i \leq n} e^{x_i}) \). [BV, Example 3.25] shows that \( f^*(y) = \sum_i y_i \log y_i \), the negative entropy function, restricted to the probability simplex \( (y \geq 0, \sum_{1 \leq i \leq n} y_i = 1) \).