Optimization Theory and Algorithm **Lecture 3 - 09/22/2021**

Lecture 3

Lecturer:Xiangyu Chang *Scribe: Xiangyu Chang*

Edited by: Junbo Hao

1 Conjugate Function

Definition 1.1. Let $f : \mathbb{R}^n \to \mathbb{R}$, the function $f^* : \mathbb{R}^n \to \mathbb{R}$, defined as

$$f^*(\mathbf{y}) = \sup_{\mathbf{x} \in \mathrm{dom}(f)} \{\mathbf{y}^\top \mathbf{x} - f(\mathbf{x})\}, \tag{1}$$

is called the *conjugate* of the function f.

Remark 1.2.

- *f^* is a convex function. This is true whether or not f is convex.*
- *The domain of conjugate function consists of $\mathbf{y} \in \mathbb{R}^n$ for which the supermom is finite.*
- *Geometric Interpretation of conjugate function:*

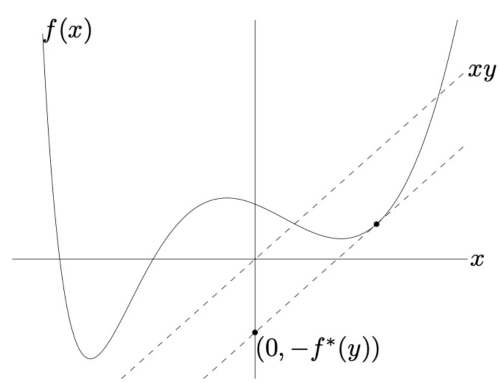

Figure 1: Geometric Interpretation of conjugate function

1.1 Examples of Conjugate Function

Example 1.3.

- $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} + b$. Then

$$
\begin{aligned}
f^*(\mathbf{y}) &= \sup_{\mathbf{x}\in \mathrm{dom}(f)} \{\mathbf{y}^\top \mathbf{x} - \mathbf{a}^\top \mathbf{x} - b\} = \sup_{\mathbf{x}\in \mathrm{dom}(f)} \{(\mathbf{y}-\mathbf{a})^\top \mathbf{x} - b\} \\
&= \begin{cases} -b, \ \mathbf{y} = \mathbf{a}, \\ \infty, \ \textit{otherwise}. \end{cases}
\end{aligned}
$$

- Exponential Function: $f(x) = \exp(x)$, then

$$
f^*(y) = \begin{cases} y\log(y) - y, \ y > 0, \\ 0, \ y = 0. \end{cases}
$$

- Negative Logarithm: $f(x) = -\log(x)$. Then

$$
f^*(y) = \begin{cases} \log(-1/y) - 1, \ y < 0, \\ \infty, \ y \geqslant 0. \end{cases}
$$

- Negative Entropy: $f(x) = x\log(x), (x \geqslant 0, 0\log(0) = 0)$, then

$$
f^*(y) = \exp(y-1).
$$

Q: Why we need the conjugate function? The following three examples may indicate reasons.

Example 1.4. Consider a naive problem:

$$
\begin{aligned}
&\min_{\mathbf{x}} f(\mathbf{x}), \\
&s.t. \ \mathbf{x} = 0.
\end{aligned}
$$

Then, its Lagrange dual function

$$
\begin{aligned}
g(\boldsymbol{\nu}) &= \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\nu}) = \inf\{f(\mathbf{x}) + \boldsymbol{\nu}^\top \mathbf{x}\} \\
&= -\sup_{\mathbf{x}}\{(-\boldsymbol{\nu})^\top \mathbf{x} - f(\mathbf{x})\} \\
&= -f^*(-\boldsymbol{\nu}).
\end{aligned}
$$

Example 1.5. More general cases:

$$
\begin{aligned}
&\min_{\mathbf{x}} f_0(\mathbf{x}), \\
&s.t. \ A\mathbf{x} \succeq \mathbf{b}, \\
&\quad\ C\mathbf{x} = \mathbf{d}.
\end{aligned}
$$

The Lagrange dual function is

$$
\begin{aligned}
g(\boldsymbol{\lambda},\boldsymbol{\nu}) &= \inf_{\mathbf{x}} L(\mathbf{x},,\boldsymbol{\lambda},\boldsymbol{\nu}) \\
&= \inf_{\mathbf{x}}\{f_0(\mathbf{x}) + \boldsymbol{\lambda}^\top(A\mathbf{x}-\mathbf{b}) + \boldsymbol{\nu}^\top(C\mathbf{x}-\mathbf{d})\} \\
&= -\boldsymbol{\lambda}^\top\mathbf{b} - \boldsymbol{\nu}^\top\mathbf{d} + \inf_{\mathbf{x}}\{(\boldsymbol{\lambda}^\top A + \boldsymbol{\nu}^\top C)\mathbf{x} + f_0(\mathbf{x})\} \\
&= -\boldsymbol{\lambda}^\top\mathbf{b} - \boldsymbol{\nu}^\top\mathbf{d} - \sup_{\mathbf{x}}\{(-\boldsymbol{\lambda}^\top A - \boldsymbol{\nu}^\top C)\mathbf{x} - f_0(\mathbf{x})\} \\
&= -\boldsymbol{\lambda}^\top\mathbf{b} - \boldsymbol{\nu}^\top\mathbf{d} - f_0^*(-\boldsymbol{\lambda}^\top A - \boldsymbol{\nu}^\top C).
\end{aligned}
$$

So, it has the Lagrange dual problem as:

$$
\begin{aligned}
&\max_{\boldsymbol{\lambda},\boldsymbol{\nu}} \ -\boldsymbol{\lambda}^\top\mathbf{b} - \boldsymbol{\nu}^\top\mathbf{d} - f_0^*(-\boldsymbol{\lambda}^\top A - \boldsymbol{\nu}^\top C), \\
&\quad s.t.\ \boldsymbol{\lambda} \succeq 0.
\end{aligned}
$$

Example 1.6. Generalized Linear Model:

- LS:
$$\min_{\mathbf{x}} \left\{ \frac{1}{2}\|A\mathbf{x}\|^2 - (\mathbf{b}^\top A)\mathbf{x} \right\}.$$
- LR:
$$\min_{\mathbf{x}} \left\{ \sum_{i=1}^{n} \log(1+\exp(\mathbf{a}_i^\top \mathbf{x})) - (\mathbf{b}^\top A)\mathbf{x} \right\}.$$
- PR:
$$\min_{\mathbf{x}} \left\{ \sum_{i=1}^{n} \exp(\mathbf{a}_i^\top \mathbf{x}) - (\mathbf{b}^\top A)\mathbf{x} \right\}.$$
- General form:

$$\min_{\mathbf{x}}\{f_0(\mathbf{x}) - (\mathbf{b}^\top A)\mathbf{x}\} = -\sup\{(A^\top\mathbf{b})^\top\mathbf{x} - f_0(\mathbf{x})\} = -f_0^*(A^\top\mathbf{b}).$$

Next, we will show some useful and advanced examples:

Example 1.7. Quadratic Function:
$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top Q\mathbf{x},\ Q \succ 0.$$
Then
$$f^*(\mathbf{y}) = \frac{1}{2}\mathbf{y}^\top Q^{-1}\mathbf{y}.$$
Consider the special case $Q = I$, then $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2$ and $f^*(\mathbf{y}) = \frac{1}{2}\|\mathbf{y}\|^2$.

Example 1.8. Indicator Function:
$$\delta_C(\mathbf{x}) = \begin{cases} 0,\ \mathbf{x} \in C, \\ \infty,\ \textit{otherwise}. \end{cases}$$

Thus,

$$\delta_C^*(\mathbf{y}) = \sup_{\mathbf{x}} \{\mathbf{y}^\top \mathbf{x} - \delta_C(\mathbf{x})\} = \sup_{\mathbf{x} \in C} \{\mathbf{y}^\top \mathbf{x}\} = \sigma_C(\mathbf{y})$$

where $\sigma_C(\mathbf{y})$ is the *support function* of C.

We consider a special case and take $C = \mathbb{R}_+^n = \{\mathbf{x} | \mathbf{x} \succeq 0\}$. Then

$$\delta_{\mathbb{R}_+^n}^*(\mathbf{y}) = \sigma_{\mathbb{R}_+^n}(\mathbf{y}) = \sup_{\mathbf{x} \in \mathbb{R}_+^n} \{\mathbf{y}^\top \mathbf{x}\} = \begin{cases} 0, \ \mathbf{y} \in \mathbb{R}_-^n, \\ \infty, \ \textit{otherwise}, \end{cases} = \delta_{\mathbb{R}_-^n}(\mathbf{y}).$$

Fact: The conjugate function of $\delta_{\mathbb{R}_+^n}$ is $\delta_{\mathbb{R}_-^n}$.