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:*

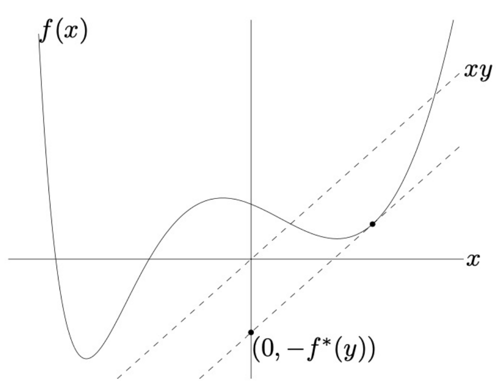

Figure 1: Geometric Interpretation of conjugate function

1.1 Examples of Conjugate Function

Example 1.3.

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

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

- 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

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

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

Example 1.4. Consider a naive problem:

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

Then, its Lagrange dual function

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

Next, we will show some useful and advanced examples:

Example 1.5. 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 1.6. 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 $\delta_{\mathbb{R}^n_-}$.

Example 1.7. Explanation of Lagrange: The general optimization formulation (**??**) has the equivalent form as

$$\min_{\mathbf{x}} \left\{ f_0(\mathbf{x}) + \sum_i \delta_{\mathbb{R}_-}(f_i(\mathbf{x})) + \sum_j \delta_0(h_j(\mathbf{x})) \right\}$$

and

$$\delta_{\mathbb{R}_-}(f_i(\mathbf{x})) = \sup_{\lambda_i \geqslant 0} \lambda_i f_i(\mathbf{x}),$$
$$\delta_0(h_j(\mathbf{x})) = \sup_{\nu_j} \nu_j h_j(\mathbf{x}).$$

Thus,

$$\min_{\mathbf{x}} \left\{ f_0(\mathbf{x}) + \sum_i \delta_{\mathbb{R}_-}(f_i(\mathbf{x})) + \sum_j \delta_0(h_j(\mathbf{x})) \right\} \Longleftrightarrow \min_{\mathbf{x}} \left\{ f_0(\mathbf{x}) + \sum_i \sup_{\lambda_i \geqslant 0} \lambda_i f_i(\mathbf{x}) + \sum_j \sup_{\nu_j} \nu_j h_j(\mathbf{x}) \right\}$$

$$\Longleftrightarrow \min_{\mathbf{x}} \sup_{\boldsymbol{\lambda}\succeq 0, \boldsymbol{\nu}} \left\{ f_0(\mathbf{x}) + \sum_i \lambda_i f_i(\mathbf{x}) + \sum_j \nu_j h_j(\mathbf{x}) \right\}.$$

Before to state the next example, we would like to introduce a widely used concept "dual norm" in advance.

Definition 1.8. In $\mathbb{R}^n$, $\|\mathbf{z}\|_* = \sup\{|\langle \mathbf{z}, \mathbf{x} \rangle| \,|\, \|\mathbf{x}\| \leqslant 1\}$ is the *dual norm* of $\|\cdot\|$.

Theorem 1.9. *According to the definition of dual norm, the following facts hold:*

(i) $|\langle \mathbf{z}, \mathbf{x} \rangle\,| \leqslant \|\mathbf{x}\| \|\mathbf{z}\|_*$.

(ii) The dual norm is the operator norm of $\mathbf{z}^\top$.

(iii) The dual norm of $\|\cdot\|_p$ *is* $\|\cdot\|_q$*, where* $\frac{1}{p}+\frac{1}{q}=1, p,q>0$*.*

Proof. For (i),

$$|\langle \mathbf{z},\mathbf{x}\rangle| = \|\mathbf{x}\| \left| \left\langle \mathbf{z}, \frac{\mathbf{x}}{\|\mathbf{x}\|} \right\rangle \right| \leqslant \|\mathbf{x}\|\|\mathbf{z}\|_*,$$

where the last inequality due to the definition of dual norm.

For (ii), we know that

$$\|\mathbf{z}\|_* = \sup\{(\mathbf{z}^\top)\mathbf{x} | \|\mathbf{x}\| \leqslant 1\}$$

is the operator norm of matrix $\mathbf{z}^\top$.

For (iii), we first recall the Holder inequality as

$$\langle \mathbf{z},\mathbf{x}\rangle \leqslant \|\mathbf{x}\|_p\|\mathbf{z}\|_q.$$

Thus,

$$\|\mathbf{z}\|_* = \sup_{\|\mathbf{x}\|_p\leqslant 1} \mathbf{z}^\top\mathbf{x} \leqslant \|\mathbf{x}\|_p\|\mathbf{z}\|_q \leqslant \|\mathbf{z}\|_q.$$

Let $\tilde{x}_i = \frac{|z_i|^{q-2}z_i}{\|\mathbf{z}\|_q^{q-1}}$, then

$$\|\tilde{\mathbf{x}}\|_p^p = \sum_i |\tilde{x}_i|^p = \frac{1}{\|\mathbf{z}\|_q^{(q-1)p}} \sum_i |\tilde{z}_i|^{p(q-1)} = \frac{1}{\|\mathbf{z}\|_q^q} \sum_i |\tilde{z}_i|^q = 1,$$

due to $p(q-1)=q$. So,

$$\|\mathbf{z}\|_* \geqslant \sum_i \tilde{x}_i z_i = \frac{1}{\|\mathbf{z}\|_q^{q-1}} \sum_i |z_i|^q = \frac{\|\mathbf{z}\|_q^q}{\|\mathbf{z}\|_q^{q-1}} = \|\mathbf{z}\|_q.$$

■

Example 1.10. Norm function: $f(\mathbf{x}) = \|\mathbf{x}\|$.

We denote $\ell(\mathbf{x},\mathbf{y}) = \langle \mathbf{x},\mathbf{y}\rangle - \|\mathbf{x}\|$ and $f^*(y) = \sup_{\mathbf{x}} \ell(\mathbf{x},\mathbf{y})$. According to the property of dual norm, it has

$$\ell(\mathbf{x}) \leqslant \|\mathbf{x}\|\|\mathbf{y}\|_* - \|\mathbf{x}\| = \|\mathbf{x}\|(\|\mathbf{y}\|_* - 1).$$

Thus, if $\|\mathbf{y}\|_* \leqslant 1$, then $f^*(\mathbf{y}) = 0$. If $\|\mathbf{y}\|_* > 1$, then by the definition of of dual norm, there exists $\mathbf{z}$ such that $\mathbf{z}^\top\mathbf{y} > 1, \|\mathbf{z}\| \leqslant 1$. Let $\mathbf{x} = t\mathbf{z}$, then

$$\ell(\mathbf{x}) = \langle \mathbf{y},\mathbf{x}\rangle - \|\mathbf{x}\| = t\langle \mathbf{y},\mathbf{z}\rangle - t\|\mathbf{z}\| = t(\langle \mathbf{y},\mathbf{z}\rangle - \|\mathbf{z}\|) \to \infty,$$

as $t\to\infty$. Finally, we have

$$f^*(\mathbf{y}) = \begin{cases} 0, \ \|\mathbf{y}\|_* \leqslant 1, \\ \infty, \ \textit{otherwise}, \end{cases} = \delta_{B_{\|\cdot\|_*}}(\mathbf{y}).$$

The conjugate of norm function is the indicator function of the dual norm unit ball.

Example 1.11. Square norm function: $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2$. We denote $\ell(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle - \frac{1}{2}\|\mathbf{x}\|^2$ and $f^*(y) = \sup_{\mathbf{x}} \ell(\mathbf{x}, \mathbf{y})$. Then,

$$\ell(\mathbf{x}, \mathbf{y}) \leqslant \|\mathbf{y}\|_* \|\mathbf{x}\| - \frac{1}{2}\|\mathbf{x}\|^2.$$

Minimize the right hand size with respect to $\|\mathbf{x}\|$, we can obtain that

$$\ell(\mathbf{x}, \mathbf{y}) \leqslant \frac{1}{2}\|\mathbf{y}\|_*,$$

the equality holds when $\|\mathbf{x}\| = \|\mathbf{y}\|_*$, so $f^*(\mathbf{y}) = \frac{1}{2}\|\mathbf{y}\|_*$.