Optimization Theory and Algorithm	Lecture 4 - 09/24/2021
Lecture 4	
Lecturer:Xiangyu Chang	*Scribe: Xiangyu Chang*

Edited by: Junbo Hao

1 Conjugate Function II

1.1 Property of Conjugate Function

- $g(\mathbf{x}) = af(\mathbf{x}) + b$, then $g^*(\mathbf{y}) = af^*(\mathbf{y}/a) - b$.

- $g(\mathbf{x}) = f(A\mathbf{x} + \mathbf{b})$, then $g^*(\mathbf{y}) = f^*(A^{-\top}\mathbf{y}) - \mathbf{b}^\top A^{-\top}\mathbf{y}$.

- $f(u, v) = f_1(u) + f_2(v)$. Then, $f^*(w, x) = f_1^*(w) + f_2^*(x)$.

- Fenchel's Inequality:

$$f(\mathbf{x}) + f^*(\mathbf{y}) \geqslant \langle \mathbf{x}, \mathbf{y} \rangle . \tag{1}$$

- Define that $f^{**}(\mathbf{x}) = \sup_{\mathbf{y}}\{\mathbf{x}^\top \mathbf{y} - f^*(\mathbf{y})\}$. Obviously, we can justify that $f^{**}(\mathbf{x}) \leqslant f(\mathbf{x})$ due to the Fenchel's inequality (1). In addition, if f is convex and closed, then $f^{**} = f$. The poof can be found at Page 61. A closed function means $\text{epi}(f)$ is a closed set.

1.2 Using conjugate to derive Lagrange dual function

Recall:

Example 1.1. 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

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

So, it has the Lagrange dual problem as:

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

Two special cases:

$$\begin{aligned} &\min_{\mathbf{x}} \|\mathbf{x}\|, \\ &s.t.\ A\mathbf{x} = \mathbf{b}. \end{aligned}$$

We know that

$$g(\boldsymbol{\nu}) = -\mathbf{b}^\top \boldsymbol{\nu} - f_0^*(-A^\top \boldsymbol{\nu}) = \begin{cases} -\mathbf{b}^\top \boldsymbol{\nu},\ \|A^\top \boldsymbol{\nu}\|_* \leqslant 1, \\ \infty,\ otherwise. \end{cases}$$

Thus, the Lagrange dual problem is

$$\begin{aligned} &\max_{\boldsymbol{\nu}}\ -\mathbf{b}^\top \boldsymbol{\nu}, \\ &s.t.\ \|A^\top \boldsymbol{\nu}\|_* \leqslant 1. \end{aligned}$$

Example 1.2.

$$\min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{x}).$$

This problem has the equivalent formulation as:

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

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

Example 1.3. We consider the LASSO problem:

$$\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x}-\mathbf{b}\|^2 + \lambda\|\mathbf{x}\|_1.$$

Let $= A\mathbf{x}-\mathbf{b}$, then

$$\min_{\mathbf{x},} \frac{1}{2}\|\|^2 + \lambda\|\mathbf{x}\|_1,$$
$$s.t.\ A\mathbf{x}-\mathbf{b} = .$$

The Lagrangian is

$$L(\mathbf{x},,\boldsymbol{\nu}) = \frac{1}{2}\|\|^2 + \lambda\|\mathbf{x}\|_1 + \boldsymbol{\nu}^\top(A\mathbf{x}-\mathbf{b}-) = (\frac{1}{2}\|\|^2 - \boldsymbol{\nu}^\top) + (\lambda\|\mathbf{x}\|_1 + \boldsymbol{\nu}^\top A\mathbf{x}) - \boldsymbol{\nu}^\top\mathbf{b}.$$

Because that $\min_{(}\frac{1}{2}\|\|^2 - \boldsymbol{\nu}^\top) = -\frac{1}{2}\|\boldsymbol{\nu}\|^2$, and

$$\|\mathbf{x}\|_1 + \boldsymbol{\nu}^\top A\mathbf{x}/\lambda \geqslant (1 - \frac{\|A^\top\boldsymbol{\nu}\|_\infty}{\lambda})\|\mathbf{x}\|_1.$$

Thus, $g(\boldsymbol{\nu}) = \inf_{\mathbf{x},} L(\mathbf{x},,\boldsymbol{\nu}) = -\frac{1}{2}\|\boldsymbol{\nu}\|^2 - \boldsymbol{\nu}^\top\mathbf{b}$, when $\|A^\top\boldsymbol{\nu}\|_\infty \leqslant \lambda$. Finally, the Lagrange dual problem is

$$\max_{\boldsymbol{\nu}}\ -\frac{1}{2}\|\boldsymbol{\nu}\|^2 - \boldsymbol{\nu}^\top\mathbf{b},$$
$$s.t.\ \|A^\top\boldsymbol{\nu}\|_\infty \leqslant \lambda.$$

Example 1.4. Consider the Lagrange dual problem of

$$\begin{aligned}
\min_{\mathbf{x}}\{\langle\mathbf{c},\mathbf{x}\rangle + h(\mathbf{b}-A\mathbf{x}) + k(\mathbf{x})\} &= \min_{\mathbf{x}}\left\{\langle\mathbf{c},\mathbf{x}\rangle + \sup_{\mathbf{y}}\{\langle\mathbf{b}-A\mathbf{x},\mathbf{y}\rangle - h^*(\mathbf{y})\} + k(\mathbf{x})\right\} \\
&= \min_{\mathbf{x}}\sup_{\mathbf{y}}\left\{\left\langle\mathbf{c}-A^\top\mathbf{y},\mathbf{x}\right\rangle + k(\mathbf{x}) + \langle\mathbf{b},\mathbf{y}\rangle - h^*(\mathbf{y})\right\} \\
&\geqslant \sup_{\mathbf{y}}\left\{\min_{\mathbf{x}}(\left\langle\mathbf{c}-A^\top\mathbf{y},\mathbf{x}\right\rangle + k(\mathbf{x})) + \langle\mathbf{b},\mathbf{y}\rangle - h^*(\mathbf{y})\right\} \\
&= \sup_{\mathbf{y}}\left\{-\sup_{\mathbf{x}}(\left\langle A^\top\mathbf{y}-\mathbf{c},\mathbf{x}\right\rangle - k(\mathbf{x})) + \langle\mathbf{b},\mathbf{y}\rangle - h^*(\mathbf{y})\right\} \\
&= \sup_{\mathbf{y}}\{-k^*(A^\top\mathbf{y}-\mathbf{c}) - h^*(\mathbf{y}) + \langle\mathbf{b},\mathbf{y}\rangle\}\ \text{(Lagrange dual problem)}.
\end{aligned}$$