SYS 6003: Optimization **Fall 2016**

Lecture 24

Instructor: Quanquan Gu *Date: Nov 21^{st}*

Now we introduce the calculation rules for conjugate functions.

1. **Separable sum**

$$f\left(\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix}\right) = g(\mathbf{x}_1) + h(\mathbf{x}_2),$$

we have

$$f^*\left(\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix}\right) = g^*(\mathbf{y}_1) + g^*(\mathbf{y}_2).$$

Remark 1 $f(\mathbf{x}) = g(\mathbf{x}) + h(\mathbf{x})$ *does not imply* $f^*(\mathbf{y}) = g^*(\mathbf{y}) + h^*(\mathbf{y})$.

2. **Scaling Multiplication**
For any $\alpha > 0, f(\mathbf{x}) = \alpha g(\mathbf{x})$, we have

$$f^*(\mathbf{y}) = \alpha g^*\left(\frac{\mathbf{y}}{\alpha}\right).$$

"Right" Scalar Multiplication
For any $\alpha > 0$,

$$f(x) = \alpha g\left(\frac{\mathbf{x}}{\alpha}\right),$$

we have

$$f^*(\mathbf{y}) = \alpha g^*(\mathbf{y}).$$

3. **Addition to Affine Function**

$$f(\mathbf{x}) = g(\mathbf{x}) + \mathbf{a}^\top \mathbf{x} + b,$$

we have

$$f^*(\mathbf{y}) = g^*(\mathbf{y} - \mathbf{a}) - b.$$

4. **Translation of Argument**

$$f(\mathbf{x}) = g(\mathbf{x} - \mathbf{b}),$$

we have

$$f^*(\mathbf{y}) = g^*(\mathbf{y}) - \mathbf{b}^\top \mathbf{y}.$$

5. **Composition with Invertible Linear Mapping**
   $\mathbf{A}$ is an invertible matrix,
   $$f(\mathbf{x}) = g(\mathbf{A}\mathbf{x}),$$
   we have
   $$f^*(\mathbf{y}) = g^*\Big([\mathbf{A}^{-1}]^\top \mathbf{y}\Big).$$

6. **Infimal Convolution**
   $$f(\mathbf{x}) = \inf_{\mathbf{x}=\mathbf{u}+\mathbf{v}} \big(g(\mathbf{u}) + h(\mathbf{v})\big),$$
   we have
   $$f^*(\mathbf{y}) = g^*(\mathbf{y}) + h^*(\mathbf{y}).$$

Next we introduce the second conjugate functions (i.e., the conjugate of conjugate function).

Definition 1 (The Second Conjugate)
$$f^{**}(\mathbf{x}) = \sup_{\mathbf{y}\in\mathbf{dom} f^*} \mathbf{x}^\top \mathbf{y} - f^*(\mathbf{y})$$
is called the second conjugate function of f.

Example 1
$$f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2, \quad f^*(\mathbf{y}) = \frac{1}{2}\|\mathbf{y}\|_2^2, \quad f^{**}(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2.$$

Theorem 1 *We have*

1. *f^{**} is closed and convex,*
2. *$f^{**}(\mathbf{x}) \le f(\mathbf{x})$ for all $\mathbf{x}$,*
3. *If f is closed and convex, then*
   $$f^{**}(\mathbf{x}) = f(\mathbf{x}) \text{ for all } \mathbf{x}.$$

Proof:

1. Due to the property of conjugate function when considering f^*.

2. By Fenchel's inequality, for all $\mathbf{x}, \mathbf{y}$ we have

$$f(\mathbf{x}) + f^*(\mathbf{y}) \geq \mathbf{x}^\top \mathbf{y}.$$

$$\Rightarrow \mathbf{x}^\top \mathbf{y} - f^*(\mathbf{y}) \leq f(\mathbf{x}). \tag{1}$$

Be definition,

$$\begin{aligned} f^{**}(\mathbf{x}) &= \sup_{\mathbf{y} \in \mathbf{dom} f^*} \mathbf{x}^\top \mathbf{y} - f^*(\mathbf{y}) \\ &= \max_{\mathbf{y} \in \mathbf{dom} f^*} \mathbf{x}^\top \mathbf{y} - f^*(\mathbf{y}) \\ &= \mathbf{x}^\top \mathbf{y}^* - f^*(\mathbf{y}^*) \end{aligned}$$

By (1) we have

$$\mathbf{x}^\top \mathbf{y}^* - f^*(\mathbf{y}^*) \leq f(\mathbf{x}). \tag{2}$$

Thus we have

$$f^{**}(\mathbf{x}) \leq f^*(\mathbf{x}), \text{ for all } \mathbf{x}.$$

3. By Separating Hyperplane Theorem. We omit its proof here.

■

Next we present a theorem regarding the connection between conjugate function and subdifferential.

Theorem 2 *If f is closed and convex, then*

$$\mathbf{y} \in \partial f(\mathbf{x}) \iff \mathbf{x} \in \partial f^*(\mathbf{y}) \iff \mathbf{x}^\top \mathbf{y} = f(\mathbf{x}) + f^*(\mathbf{y}).$$

Proof: If $\mathbf{y} \in \partial f(\mathbf{x})$, we have

$$f^*(\mathbf{y}) = \sup_{\mathbf{u} \in \mathbf{dom} f} \mathbf{y}^\top \mathbf{u} - f(\mathbf{u}) = \max_{\mathbf{u} \in \mathbf{dom} f} \mathbf{y}^\top \mathbf{u} - f(\mathbf{u}),$$

where the second equality is due to the closeness and convexity of f. By the optimality condition of the above optimization problem, $\mathbf{u}^* = \mathbf{x}$ is the global minimizer if and only if

$$\mathbf{0} \in \partial[\mathbf{y}^\top \mathbf{x} - f(\mathbf{x})] = \mathbf{y} - \partial f(\mathbf{x}) \iff \mathbf{y} \in \partial f(\mathbf{x}).$$

Therefore, $f^*(\mathbf{y}) = \mathbf{y}^\top \mathbf{x} - f(\mathbf{x})$.

For any $\mathbf{v}$, we have

$$\begin{aligned} f^*(\mathbf{v}) = \sup_{\mathbf{u}} \mathbf{v}^\top \mathbf{u} - f(\mathbf{u}) &\geq \mathbf{v}^\top \mathbf{x} - f(\mathbf{x}) = \mathbf{v}^\top \mathbf{x} - \mathbf{y}^\top \mathbf{x} + \mathbf{y}^\top \mathbf{x} - f(\mathbf{x}) \\ &= f^*(\mathbf{y}) + (\mathbf{v} - \mathbf{y})^\top \mathbf{x}. \end{aligned}$$

In other words, for any $\mathbf{v}$, we have

$$f^*(\mathbf{v}) \geq f^*(\mathbf{y}) + (\mathbf{v} - \mathbf{y})^\top \mathbf{x}. \tag{3}$$

By the definition of subgradient, we have $\mathbf{x} \in \partial f^*(\mathbf{y})$.

Now we have shown that $\mathbf{y} \in \partial f(\mathbf{x}) \Longrightarrow f(\mathbf{x}) + f^*(\mathbf{y}) = \mathbf{x}^\top \mathbf{y} \Longrightarrow \mathbf{x} \in \partial f^*(\mathbf{y})$. It remains to show $\mathbf{x} \in \partial f^*(\mathbf{y}) \Longrightarrow \mathbf{y} \in \partial f(\mathbf{x})$. To show this, by second conjugate function and its property, $f^{**}(\mathbf{x}) = f(\mathbf{x})$. By repeating the above argument, we have

$$\mathbf{x} \in \partial f^*(\mathbf{y}) \Longrightarrow \mathbf{y} \in \partial f^{**}(\mathbf{x}) = \partial f(\mathbf{x}).$$

This completes the proof. ■

The conjugate function of a strongly convex function is a smooth function (i.e., has Lipschitz continuous gradient). The following theorem shows this point.

Theorem 3 *If f is closed and μ-strongly convex, then*

(1) f^* is defined for all $\mathbf{y}$, i.e., $\mathbf{dom} f^* = \mathbb{R}^d$.

(2) f^* is differentiable everywhere, and

$$\nabla f^*(\mathbf{y}) = \arg\max_{\mathbf{x}} (\mathbf{y}^\top \mathbf{x} - f(\mathbf{x})).$$

(3) ∇f^* is Lipschitz continuous with parameter $1/\mu$, i.e.,

$$\|\nabla f^*(\mathbf{y}) - \nabla f^*(\mathbf{y}')\|_2 \leq \frac{1}{\mu}\|\mathbf{y} - \mathbf{y}'\|_2, \quad \forall\ \mathbf{y}, \mathbf{y}'.$$

Proof: (1) By the definition of conjugate function,

$$f^*(\mathbf{y}) = \sup_{\mathbf{x} \in \mathbf{dom} f} \mathbf{y}^\top \mathbf{x} - f(\mathbf{x}) = \max_{\mathbf{x} \in \mathbf{dom} f} \mathbf{y}^\top \mathbf{x} - f(\mathbf{x}).$$

For any $\mathbf{y}$, the above maximization has a unique maximizer, since f is strongly convex. Therefore, f^* exists for every $\mathbf{y}$.

(2) By Theorem 2, $\mathbf{x}$ maximizes $\mathbf{y}^\top \mathbf{x} - f(\mathbf{x})$ if and only if $\mathbf{y} \in \partial f(\mathbf{x})$, which is equivalent to $\mathbf{x} \in \partial f^*(\mathbf{y})$. Since $\mathbf{x}$ is a unique maximizer, $\partial f^*(\mathbf{y})$ contains only one element, which

$$\partial f^*(\mathbf{y}) = \{\nabla f^*(\mathbf{y})\} = \{\mathbf{x}\}.$$

This implies that f^* is a differentiable function, and

$$\nabla f^*(\mathbf{y}) = \mathbf{x} = \arg \max_{\mathbf{u} \in \mathbf{dom} f} \mathbf{y}^\top \mathbf{u} - f(\mathbf{u}).$$

(3) Since $f(\mathbf{x})$ is μ-strongly convex, we have

$$[\mathbf{y} - \mathbf{y}']^\top (\mathbf{x} - \mathbf{x}') \geq \mu \|\mathbf{x} - \mathbf{x}'\|_2^2, \qquad \forall\ \mathbf{x}, \mathbf{x}', \mathbf{y} \in \partial f(\mathbf{x}), \mathbf{y}' \in \partial f(\mathbf{x}'). \tag{4}$$

By Theorem 2 we have $\mathbf{x} \in \partial f^*(\mathbf{y}), \mathbf{x}' \in \partial f^*(\mathbf{y}')$. In addition, by (2), we must have $\mathbf{x} = \nabla f^*(\mathbf{y})$ and $\mathbf{x}' = \nabla f^*(\mathbf{y}')$. Submitting this into (4) gives

$$(\mathbf{y} - \mathbf{y}')^\top (\nabla f^*(\mathbf{y}) - \nabla f^*(\mathbf{y}')) \geq \mu \|\nabla f^*(\mathbf{y}) - \nabla f^*(\mathbf{y}')\|_2^2.$$

By Cauchy-Schwartz inequality, we have

$$\|\nabla f^*(\mathbf{y}) - \nabla f^*(\mathbf{y}')\|_2 \cdot \|\mathbf{y} - \mathbf{y}'\|_2 \geq (\mathbf{y} - \mathbf{y}')^\top (\nabla f^*(\mathbf{y}) - \nabla f^*(\mathbf{y}')) \geq \mu \|\nabla f^*(\mathbf{y}) - \nabla f^*(\mathbf{y}')\|_2^2.$$

which immediately yields

$$\|\nabla f^*(\mathbf{y}) - \nabla f^*(\mathbf{y}')\|_2 \leq \frac{1}{\mu} \|\mathbf{y} - \mathbf{y}'\|_2.$$

This completes the proof. ■