SYS 6003: Optimization	**Fall 2016**

Lecture 23

Instructor: Quanquan Gu — *Date: Nov 16^{th}*

First we review the convergence rate of proximal gradient descent algorithm for solving: $\min_{\mathbf{x}} f(\mathbf{x}) + h(\mathbf{x})$.

Table 1: Convergence rate of proximal gradient descent algorithms for different convex functions.

$f(\mathbf{x})$	$h(\mathbf{x})$	Convergence rate
convex and smooth	simple convex	$O(1/T)$
strongly convex and smooth	simple convex	$O\Big(\big(1-\frac{\mu}{L}\big)^T\Big)$

Recall that we have introduced the definition of conjugate function:

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

Now we introduce an important inequality describing the properties of conjugate function.

Lemma 1 (Fenchel's Inequality) *For any $\mathbf{x} \in$ **dom** f, $\mathbf{y} \in$ **dom** f^* we have*

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

Proof: By definition,

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

Thus

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

■

Following is an example of Lemma 1.

Example 1

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

By Fenchel's inequality we have

$$\frac{1}{2}\|\mathbf{x}\|_2^2 + \frac{1}{2}\|\mathbf{y}\|_2^2 \geq \mathbf{x}^\top \mathbf{y}$$

$$\iff \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 \geq 0.$$

In the remaining of this lecture, we continue to show the Fenchel conjugate of some frequently encountered functions.

Example 2 (Indicator Function of Convex Set C)

$$\delta_C(\mathbf{x}) = \begin{cases} 0 & , \textit{ if } \mathbf{x} \in C \\ +\infty & , \textit{ if } \mathbf{x} \notin C \end{cases}, \qquad \delta_C^*(\mathbf{y}) = \sup_{\mathbf{x} \in C} \; \mathbf{x}^\top \mathbf{y},$$

where C is a convex set.

Proof: By definition,

$$\begin{aligned} \delta_C^*(\mathbf{y}) &= \sup_{\mathbf{x} \in \mathbf{dom}\ \delta_C} \; \mathbf{x}^\top \mathbf{y} - \delta_C(\mathbf{x}) \\ &= \max\left\{ \sup_{\mathbf{x} \in C} \mathbf{x}^\top \mathbf{y}, \; \sup_{\mathbf{x} \notin C} \mathbf{x}^\top \mathbf{y} - \infty \right\} \\ &= \sup_{\mathbf{x} \in C} \; \mathbf{x}^\top \mathbf{y}. \end{aligned}$$

■

Before we show another example of calculating the conjugate function, we first introduce the dual norm.

Definition 1 (Dual norm) *The dual norm of norm $\|\|$, denoted by $\|\|_*$, is defined as*

$$\|\mathbf{y}\|_* = \sup_{\|\mathbf{x}\| \le 1} \mathbf{x}^\top \mathbf{y}.$$

By the definition of dual norm, we can prove that $\|\cdot\|_2$ is the dual norm of itself.
Proof: Let $\|\cdot\| = \|\cdot\|_2$. Then for any $\mathbf{y} \in \mathbb{R}^d$,

$$\|\mathbf{y}\|_* = \sup_{\|\mathbf{x}\|_2 \le 1} \mathbf{x}^\top \mathbf{y} \le \sup_{\|\mathbf{x}\|_2 \le 1} \|\mathbf{x}\|_2 \cdot \|\mathbf{y}\|_2 \le \sup_{\|\mathbf{x}\|_2 \le 1} \|\mathbf{y}\|_2 = \|\mathbf{y}\|_2,$$

where the first inequality follows from Cauchy-Schwartz inequality and the second inequality follows from the constraint that $\|\mathbf{x}\|_2 \le 1$. It remains to show $\|\mathbf{y}\|_* \ge \|\mathbf{y}\|_2$. To show this, we choose $\mathbf{x} = \mathbf{y}/\|\mathbf{y}\|_2$. It is easy to show that $\|\mathbf{x}\|_2 = 1$. Then

$$\|\mathbf{y}\|_* = \sup_{\|\mathbf{x}\|_2 \le 1} \mathbf{x}^\top \mathbf{y} \ge \mathbf{x}^\top \mathbf{y} = \frac{\mathbf{y}^\top}{\|\mathbf{y}\|_2} \mathbf{y} = \frac{\|\mathbf{y}\|_2^2}{\|\mathbf{y}\|_2} = \|\mathbf{y}\|_2.$$

Since we have both $\|\mathbf{y}\|_* \le \|\mathbf{y}\|_2$ and $\|\mathbf{y}\|_* \ge \|\mathbf{y}\|_2$, we must have

$$\|\mathbf{y}\|_* = \|\mathbf{y}\|_2.$$

This completes the proof. ■

Example 3 (Norm) *Let $f(\mathbf{x}) = \|\mathbf{x}\|$. Then its conjugate function is*

$$f^*(\mathbf{y}) = \begin{cases} 0, & \text{if} \quad \|\mathbf{y}\|_* \leq 1 \\ +\infty, & \text{if} \quad \|\mathbf{y}\|_* > 1 \end{cases},$$

where $\|\cdot\|_*$ is the dual norm of $\|\|$,

Remark 1 *It is worth noting that in this example, $f^*(\mathbf{y})$ is the indicator function of the unit norm ball of $\|\cdot\|_*$, which is defined as $\{\mathbf{y} : \|\mathbf{y}\|_* \leq 1\}$.*

Remark 2 *If $\|\cdot\| = \|\cdot\|_2$, then $\|\cdot\|_* = \|\cdot\|_2$. Thus*

$$f^*(\mathbf{y}) = \begin{cases} 0, & \text{if} \quad \|\mathbf{y}\|_2 \leq 1 \\ +\infty, & \text{if} \quad \|\mathbf{y}\|_2 > 1 \end{cases}.$$

If $\|\cdot\| = \|\cdot\|_1$, then $\|\cdot\|_* = \|\cdot\|_\infty$. Thus

$$f^*(\mathbf{y}) = \begin{cases} 0, & \text{if} \quad \|\mathbf{y}\|_\infty \leq 1 \\ +\infty, & \text{if} \quad \|\mathbf{y}\|_\infty > 1 \end{cases}.$$

In general, if $\|\cdot\| = \|\cdot\|_p$, then $\|\cdot\|_* = \|\cdot\|_q$, where $1/p + 1/q = 1$. Therefore

$$f^*(\mathbf{y}) = \begin{cases} 0, & \text{if} \quad \|\mathbf{y}\|_q \leq 1 \\ +\infty, & \text{if} \quad \|\mathbf{y}\|_q > 1 \end{cases}.$$

Proof:[Proof of Example 3] By the definition of Fenchel conjugate,

$$f^*(\mathbf{y}) = \sup_{\mathbf{x} \in \mathbb{R}^d} \mathbf{x}^\top \mathbf{y} - \|\mathbf{x}\|. \tag{1}$$

Note that the domain of a norm is the whole space. By Hölder's inequality, $\mathbf{x}^\top \mathbf{y} \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|_*$. Substituting this result to (1) gives rise to

$$f^*(\mathbf{y}) \leq \sup_{\mathbf{x} \in \mathbb{R}^d} \|\mathbf{x}\| \cdot \|\mathbf{y}\|_* - \|\mathbf{x}\| = \|\mathbf{x}\|(\|\mathbf{y}\|_* - 1).$$

Case 1: $\|\mathbf{y}\|_* \leq 1$. Immediately we have $f^*(\mathbf{y}) \leq \|\mathbf{x}\|(\|\mathbf{y}\|_* - 1) \leq 0$ since $\|\mathbf{x}\| \geq 0$. Next we show that this bound is achievable: choosing $\mathbf{x} = \mathbf{0}$, by (1) we have

$$f^*(\mathbf{y}) = \sup_{\mathbf{x} \in \mathbb{R}^d} \mathbf{x}^\top \mathbf{y} - \|\mathbf{x}\| \geq \mathbf{x}^\top \mathbf{y} - \|\mathbf{x}\| = \mathbf{0}^\top \mathbf{y} - 0 = 0.$$

Hence we have $f^*(\mathbf{y}) = 0$ for any $\|\mathbf{y}\|_* \leq 1$.

Case 2: $\|\mathbf{y}\|_* > 1$. By the definition of dual norm we have $\|\mathbf{y}\|_* = \sup_{\|\mathbf{x}\| \leq 1} \mathbf{x}^\top \mathbf{y} > 1$. Therefore, there exists an $\mathbf{x}_0$ such that $\|\mathbf{x}_0\| \leq 1$ and $\mathbf{x}_0^\top \mathbf{y} > 1$. Hence $\mathbf{x}_0^\top \mathbf{y} - \|\mathbf{x}_0\|$ is a number strictly larger than 0.

Let $\mathbf{x}' = t\mathbf{x}_0$, where t is a positive number. By the definition of $f^*(\cdot)$, for any $t > 0$,

$$\begin{aligned} f^*(\mathbf{y}) &= \sup_{\mathbf{x} \in \mathbb{R}^d} \mathbf{x}^\top \mathbf{y} - \|\mathbf{x}\| \geq \mathbf{x}'^\top \mathbf{y} - \|\mathbf{x}'\| = t\mathbf{x}_0^\top \mathbf{y} - \|t\mathbf{x}_0\| \\ &= t(\mathbf{x}_0^\top \mathbf{y} - \|\mathbf{x}_0\|). \end{aligned} \tag{2}$$

Because $\mathbf{x}_0^\top \mathbf{y} - \|\mathbf{x}_0\|$ is strictly positive, when $t \to \infty$, the right hand side of (2) tends to be positive infinity. Therefore $f^*(\mathbf{y}) = +\infty$.

Combining Case 1 and Case 2, we have

$$f^*(\mathbf{y}) = \begin{cases} 0, & \text{if} \quad \|\mathbf{y}\|_* \leq 1 \\ +\infty, & \text{if} \quad \|\mathbf{y}\|_* > 1 \end{cases} .$$

■