SYS 6003: Optimization	**Fall 2016**

Lecture 21

Instructor: Quanquan Gu *Date: Nov 9^{th}*

One important property of proximal operator is that: the minimizer of a function is the fixed point of the proximal operator of this function, and vice versa.

Lemma 1 (Fixed Point) *The point $\mathbf{x}^*$ minimizes a convex function $f(\cdot)$ if and only if $\mathbf{x}^* = \mathrm{Prox}_f(\mathbf{x}^*)$*

Proof: "$\Rightarrow$" direction: If $\mathbf{x}^*$ minimizes f, for any $\mathbf{x} \in \mathrm{dom} f$, we have $f(\mathbf{x}^*) \leq f(\mathbf{x})$. Furthermore we have

$$f(\mathbf{x}^*) + \frac{1}{2}\|\mathbf{x}^* - \mathbf{x}^*\|_2^2 \leq f(\mathbf{x}) + \frac{1}{2}\|\mathbf{x} - \mathbf{x}^*\|_2^2. \tag{1}$$

Recall that

$$\mathrm{Prox}_f(\mathbf{x}^*) := \mathrm{argmin}_{\mathbf{u}} \frac{1}{2}\|\mathbf{u} - \mathbf{x}^*\|_2^2 + f(\mathbf{u}). \tag{2}$$

By (1) it can be seen that $\mathbf{x}^*$ is the minimizer of the optimization problem (2). Hence we have $\mathbf{x}^* = \mathrm{Prox}_f(\mathbf{x}^*)$.

"$\Leftarrow$" direction: Recall that $\mathrm{Prox}_f(\mathbf{x}) := \mathrm{argmin}_{\mathbf{u}} \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|_2^2 + f(\mathbf{u})$. Hence we have for any $\mathbf{x} \in \mathrm{dom} f$, $\tilde{\mathbf{x}} = \mathrm{Prox}_f(\mathbf{x})$ if and only if $\mathbf{x} - \tilde{\mathbf{x}} \in \partial f(\tilde{\mathbf{x}})$. Let us choose $\mathbf{x} = \tilde{\mathbf{x}} = \mathbf{x}^*$. Then we have $\mathbf{0} \in \partial f(\mathbf{x}^*)$, which means that $\mathbf{x}^*$ is a minimizer of $f(\cdot)$. ■

We are going to introduce Monreau Decomposition. First we lay out the definition of closed functions.

Definition 1 (Closed Function) *A function $f(\cdot)$ is closed if its epigraph $\{(\mathbf{x}, t) | f(\mathbf{x}) \leq t\}$ is a closed set.*

In order to show that a function is closed, we can check whether its epgraph is closed according to the definition. We can also check whether its all sublevel sets are closed. This is suggested by the following lemma.

Lemma 2 *A function $f(\cdot)$ is closed if and only if all its sublevel sets $S_t(f) = \{\mathbf{x} \in \mathrm{dom} f | f(\mathbf{x}) \leq t\}$ are closed sets.*

Now we introduce the definition of Fenchel conjugate functions. In the rest of this class, we simply refer to Fenchel conjugate functions as conjugate functions.

Definition 2 (Conjugate Function) *The conjugate of a function $f : \mathbb{R}^d \to \mathbb{R}$ is*

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

In Definition 2, there is no requirement that f should be convex. In fact, we will show that the conjugate f^* is a convex function no matter whether f is convex.

Lemma 3 *For any function $f : \mathbb{R}^d \to \mathbb{R}$, $f^*(\cdot)$ is closed and convex.*

Proof:[Proof of the convexity of f^*] Define $h_{\mathbf{x}}(\mathbf{y}) = \mathbf{y}^\top \mathbf{x} - f(\mathbf{x})$. It can be seen that $h_{\mathbf{x}}(\mathbf{y})$ is convex for any $\mathbf{x}$ since it is a linear function of $\mathbf{y}$. Recall that, pointwise supremum is a convexity-preserving operator. Thus, we have $\sup_{\mathbf{x}\in\mathrm{dom} f} h_{\mathbf{x}}(\mathbf{y})$ is convex in $\mathbf{y}$. This immediately implies that

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

is convex in $\mathbf{y}$. ■

Following are some examples of conjugate functions.

Example 1 *$f(\mathbf{x}) = \|\mathbf{x}\|_2^2/2$. Then*

$$f^*(\mathbf{y}) = \sup_{\mathbf{x}} \mathbf{y}^\top \mathbf{x} - \frac{1}{2}\|\mathbf{x}\|_2^2. \tag{3}$$

Denote $g(\mathbf{x}) = \mathbf{y}^\top \mathbf{x} - \frac{1}{2}\|\mathbf{x}\|_2^2$ and $\mathbf{x}^* = \mathrm{argmax}_{\mathbf{x}} g(\mathbf{x})$. We have $\nabla g(\mathbf{x}^*) = \mathbf{0}$. By calculation we know $\nabla g(\mathbf{x}) = \mathbf{y} - \mathbf{x}$. Hence we have $\mathbf{y} - \mathbf{x}^* = \mathbf{0}$, i.e., $\mathbf{x}^* = \mathbf{y}$. Substituting this into (3) yields

$$f^*(\mathbf{y}) = \mathbf{y}^\top \mathbf{y} - \frac{1}{2}\|\mathbf{y}\|_2^2 = \frac{1}{2}\|\mathbf{y}\|_2^2.$$

It is worth noting that in this example, $f = f^*$.

Example 2 (Quadratic Function) *If $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{x} + c$, where A is positive definite, then*

$$f^*(\mathbf{y}) = \frac{1}{2}(\mathbf{y} - \mathbf{b})^\top \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}) - c.$$

Remark 1 *When $\mathbf{b} = \mathbf{0}, \mathbf{c} = \mathbf{0}$, $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x}/2 = \|\mathbf{x}\|_{\mathbf{A}}^2/2$, where $\|\mathbf{x}\|_{\mathbf{A}}$ is the Mahalanobis distance from $\mathbf{x}$ to $\mathbf{0}$ with matrix $\mathbf{A}$. Then $f^*(\mathbf{y}) = \mathbf{y}^\top \mathbf{A}^{-1}\mathbf{y}/2 = \|\mathbf{y}\|_{\mathbf{A}^{-1}}^2/2$.*

Proof: By the definition we have

$$f^*(\mathbf{y}) = \sup_{\mathbf{x}} \left[\mathbf{x}^\top \mathbf{y} - \left(\frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{x} + c\right)\right]. \tag{4}$$

Define $g(\mathbf{x}) := \mathbf{x}^\top \mathbf{y} - (\frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{x} + c)$ and $\mathbf{x}^* := \mathrm{argmax}_{\mathbf{x}} g(\mathbf{x})$. By the first order optimality condition we have

$$\mathbf{0} = \nabla \mathbf{g}(\mathbf{x}^*) = \mathbf{y} - \mathbf{A}\mathbf{x}^* - \mathbf{b},$$

which yields that $\mathbf{x}^* = \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})$. Substituting this into (4) gives

$$\begin{aligned} f^*(\mathbf{y}) &= \mathbf{x}^{*\top}\mathbf{y} - \left(\frac{1}{2}\mathbf{x}^{*\top}\mathbf{A}\mathbf{x}^* + \mathbf{b}^\top\mathbf{x}^* + c\right) \\ &= (\mathbf{y} - \mathbf{b})^\top\mathbf{A}^{-1}\mathbf{y} - \left(\frac{1}{2}(\mathbf{y} - \mathbf{b})^\top\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}) + \mathbf{b}^\top\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}) + c\right) \\ &= \frac{1}{2}(\mathbf{y} - \mathbf{b})^\top\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}) - c. \end{aligned}$$

This completes the proof. ■

Example 3 (Negative Entropy) *If $f : \mathbb{R}^d \to \mathbb{R}$ is defined as*

$$f(\mathbf{x}) = \sum_{i=1}^{d} x_i \log x_i, \qquad \mathrm{dom} f = \{\mathbf{x} | x_i > 0, i = 1, 2, \ldots, d\},$$

then $f^*(\mathbf{y}) = \sum_{i=1}^{d} e^{y_i - 1}$.

Proof: By definition we have

$$\begin{aligned} f^*(\mathbf{y}) &= \sup_{\mathbf{x}} \left(\mathbf{x}^\top\mathbf{y} - \sum_{i=1}^{d} x_i \log x_i\right) \\ &= \sup_{\mathbf{x}} \left(\sum_{i=1}^{d} x_i y_i - \sum_{i=1}^{d} x_i \log x_i\right) \\ &= \sum_{i=1}^{d} \sup_{x_i} [x_i y_i - x_i \log x_i]. \end{aligned} \tag{5}$$

Define $g(x_i) = x_i y_i - x_i \log x_i$ and $x_i^* = \mathrm{argmax}_{x_i} g(x_i)$. By the first order optimality condition we have

$$0 = g'(x_i^*) = y_i - \log x_i^* - x_i^* \frac{1}{x_i^*} = y_i - \log x_i^* - 1.$$

Hence we have $x_i^* = e^{y_i - 1}$. Substituting this into (5) gives rise to

$$\begin{aligned} f^*(\mathbf{y}) &= \sum_{i=1}^{d} y_i e^{y_i - 1} - e^{y_i - 1} \log e^{y_i - 1} \\ &= \sum_{i=1}^{d} e^{y_i - 1}(y_i - y_i + 1) = \sum_{i=1}^{d} e^{y_i - 1}. \end{aligned}$$

■

Example 4 (Negative Logarithm) *Let* $f(\mathbf{x}) = -\sum_{i=1}^{d} \log x_i$. *Then*

$$f^*(\mathbf{y}) = -\sum_{i=1}^{d} \log(-y_i) - d.$$

Proof: By definition of $f^*(\mathbf{y})$ we have

$$\begin{aligned} f^*(\mathbf{y}) &= \sup_{\mathbf{x}} \mathbf{x}^\top \mathbf{y} + \sum_{i=1}^{d} \log x_i \\ &= \sum_{i=1}^{d} \sup_{x_i} x_i y_i + \log x_i. \end{aligned} \tag{6}$$

Define $g(x_i) := x_i y_i + \log x_i$ and $x_i^* := \operatorname{argmax}_{x_i} g(x_i)$. By the first order optimality condition we have

$$0 = \nabla g(x_i^*) = y_i + \frac{1}{x_i^*},$$

which yields that $x_i^* = -1/y_i$. Substituting this into (6) gives

$$\begin{aligned} f^*(\mathbf{y}) &= \sum_{i=1}^{d} \sup_{x_i} x_i y_i + \log x_i \\ &= \sum_{i=1}^{d} \Big(-1 + \log\Big(-\frac{1}{y_i}\Big)\Big) \\ &= \sum_{i=1}^{d} \big(-1 - \log(-y_i)\big) \\ &= -\sum_{i=1}^{d} \log(-y_i) - d. \end{aligned}$$

This completes the proof. ■

Example 5 (Matrix Logarithm) *Let* $f(\mathbf{X}) = -\log\det(\mathbf{X})$. *Then*

$$f^*(\mathbf{Y}) = -\log\det(-\mathbf{Y}) - d.$$

Proof: By definition of $f^*(\mathbf{Y})$ we have

$$f^*(\mathbf{Y}) = \sup_{\mathbf{X}} \ \big[\langle \mathbf{X}, \mathbf{Y}\rangle + \log\det(\mathbf{X})\big] \tag{7}$$

Define $g(\mathbf{X}) := \langle \mathbf{X}, \mathbf{Y}\rangle + \log\det(\mathbf{X})$ and $\mathbf{X}^* := \operatorname{argmax}_{\mathbf{X}} g(\mathbf{X})$. By the first order optimality condition we have

$$\mathbf{0} = \nabla \mathbf{g}(\mathbf{X}^*) = \mathbf{Y} + \mathbf{X}^{*-\mathbf{1}},$$

which yields that $\mathbf{X}^* = -\mathbf{Y}^{-1}$. Substituting this into (7) gives

$$\begin{aligned} f^*(\mathbf{Y}) &= \langle \mathbf{X}^*, \mathbf{Y} \rangle + \log\det(\mathbf{X}^*) \\ &= \langle -\mathbf{Y}^{-1}, \mathbf{Y} \rangle + \log\det(-\mathbf{Y}^{-1}) \\ &= -\mathrm{tr}(\mathbf{I}_{d\times d}) + \log[\det(-\mathbf{Y})]^{-1} \\ &= -\log\det(-\mathbf{Y}) - d. \end{aligned}$$

This completes the proof. ■

Based on the definitions and lemmas mentioned above, we are now ready to introduce the definition of Moreau Decomposition.

Theorem 1 (Moreau Decomposition) *If f is a convex function, then for any $\mathbf{x} \in \mathrm{dom} f$*

$$\mathrm{Prox}_f(\mathbf{x}) + \mathrm{Prox}_{f^*}(\mathbf{x}) = \mathbf{x},$$

where $\mathrm{Prox}_{f^*}(\mathbf{x})$ is the proximal operator of the conjugate function of f, i.e., f^*.

Theorem 1 reveals that Prox_f and Prox_{f^*} are analogous to the orthogonal projection of a point $\mathbf{x}$ on to a linear subspace S and its orthogonal complement $S^\perp$.