SYS 6003: Optimization	**Fall 2016**

Lecture 25

Instructor: Quanquan Gu | *Date: Nov 28^{th}*

Now we review the Moreau Decomposition and prove it.

Theorem 1 (Moreau Decomposition)

$$\mathbf{x} = \mathrm{Prox}_f(\mathbf{x}) + \mathrm{Prox}_{f^*}(\mathbf{x}) \textit{ for all } \mathbf{x}.$$

Proof: Let $\mathbf{u} = \mathrm{Prox}_f(\mathbf{x})$

$$\begin{aligned}
&\iff \mathbf{x} - \mathbf{u} \in \partial f(\mathbf{u}) \\
&\iff \mathbf{u} \in \partial f^*(\mathbf{x} - \mathbf{u}) \\
&\iff \mathbf{x} - (\mathbf{x} - \mathbf{u}) \in \partial f^*(\mathbf{x} - \mathbf{u}) \\
&\iff \mathbf{x} - \mathbf{u} = \mathrm{Prox}_{f^*}(\mathbf{x}) \\
&\iff \mathbf{x} = \mathbf{u} + \mathrm{Prox}_{f^*}(\mathbf{x}) = \mathrm{Prox}_f(\mathbf{x}) + \mathrm{Prox}_{f^*}(\mathbf{x}).
\end{aligned}$$

■

Theorem 2 (Extended Moreau Decomposition) *For any $\lambda > 0$,*

$$\mathbf{x} = \mathrm{Prox}_{\lambda f}(\mathbf{x}) + \lambda \mathrm{Prox}_{\lambda^{-1} f^*}(\mathbf{x}/\lambda) \textit{ for all } \mathbf{x}.$$

Proof:

$$\begin{aligned}
\mathbf{x} &= \mathrm{Prox}_{\lambda f}(\mathbf{x}) + \mathrm{Prox}_{(\lambda f)^*}(\mathbf{x}) \\
&= \mathrm{Prox}_{\lambda f}(\mathbf{x}) + \lambda \mathrm{Prox}_{\lambda^{-1} f^*}(\mathbf{x}/\lambda),
\end{aligned}$$

where the last equality follows from "Right" Scalar Multiplication rule. ■

Example 1

$$f(\mathbf{x}) = \|\mathbf{x}\|_2$$

$$\mathrm{Prox}_{tf}(\mathbf{x}) = \begin{cases} (1 - t/\|\mathbf{x}\|_2)\mathbf{x}, & \textit{if } \|\mathbf{x}\|_2 > t \\ 0, & \textit{otherwise} \end{cases}$$

Proof:

$$\begin{aligned}
\mathbf{x} &= \mathrm{Prox}_{tf}(\mathbf{x}) + \mathrm{Prox}_{(tf)^*}(\mathbf{x}) \\
&= \mathrm{Prox}_{tf}(\mathbf{x}) + t \cdot \mathrm{Prox}_{(1/t)f^*}(\mathbf{x}/t).
\end{aligned}$$

$$\Rightarrow \mathrm{Prox}_{tf}(\mathbf{x}) = \mathbf{x} - t \cdot \mathrm{Prox}_{(1/t)f^*}(\mathbf{x}/t).$$

Since we know

$$f^*(\mathbf{y}) = \delta_C(\mathbf{y}) = \begin{cases} 0, & \textit{if } \|\mathbf{y}\|_2 \le 1 \\ +\infty, & \textit{otherwise} \end{cases},$$

where $C = \{\mathbf{y} : \|\mathbf{y}\|_2 \le 1\}$. *Note that for* $t > 0$, $(1/t)f^*(\mathbf{y}) = (1/t)\delta_C(\mathbf{y}) = \delta_C(\mathbf{y})$, *we have*

$$\mathrm{Prox}_{(1/t)f^*}(\mathbf{x}/t) = \mathrm{Prox}_{\delta_C}(\mathbf{x}/t) = \Pi_C(\mathbf{x}/t) = \begin{cases} \mathbf{x}/\|\mathbf{x}\|_2 & , \textit{if } \|\mathbf{x}\|_2 \ge t \\ \mathbf{x}/t & , \textit{otherwise} \end{cases}.$$

Thus we have

$$\begin{aligned} \mathrm{Prox}_{tf}(\mathbf{x}) &= \mathbf{x} - t \cdot \mathrm{Prox}_{(1/t)f^*}(\mathbf{x}/t) \\ &= \begin{cases} \mathbf{x} - t \cdot \mathbf{x}/\|\mathbf{x}\|_2, & \textit{if } \|\mathbf{x}\|_2 \ge t \\ \mathbf{x} - t \cdot \mathbf{x}/t, & \textit{otherwise} \end{cases} \\ &= \begin{cases} \mathbf{x}(1 - t/\|\mathbf{x}\|_2), & \textit{if } \|\mathbf{x}\|_2 \ge t \\ 0, & \textit{otherwise} \end{cases}. \end{aligned}$$

■

Next we introduce the duality theory.

The standard form of constraint optimization problem:

$$\begin{aligned} p^* = &\min_{\mathbf{x}} f(\mathbf{x}) \\ \text{subject to } & g_i(\mathbf{x}) \le 0, i = 1, 2, \ldots, m, \\ & h_i(\mathbf{x}) = 0, i = 1, 2, \ldots, n. \end{aligned} \tag{1}$$

where

$$\begin{aligned} g_i(\mathbf{x}) &: \text{ inequality constraint} \\ h_i(\mathbf{x}) &: \text{ equality constraint} \end{aligned}$$

Remark 1 f, g_i, h_i *do not need to be convex.*

Definition 1 (Lagrangian Function) *The Lagrangian function is defined as*

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{n} \nu_i h_i(\mathbf{x})$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n)^\top$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^\top$ *with*

$$\begin{aligned} \lambda_i &: \textit{ Lagrangian multiplier associated with } g_i(\mathbf{x}) \le 0 \\ \nu_i &: \textit{ Lagrangian multiplier associated with } h_i(\mathbf{x}) = 0 \end{aligned}$$

Definition 2 (Lagrangian Dual Function) *The Lagrangian dual function is defined as*

$$d(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}), \tag{2}$$

where $\mathcal{D} = \mathbf{dom} f \cap_{i=1}^{m} \mathbf{dom} g_i \cap_{j=1}^{n} \mathbf{h}_j$ *is the intersection of the domains of* f*,* g_i*'s and* h_j*'s.*

Remark 2 *The Lagrangian dual function* $d(\boldsymbol{\lambda}, \boldsymbol{\nu})$ *is concave in* $(\boldsymbol{\lambda}, \boldsymbol{\nu})$*. It can be proved by using the fact that pointwise infimum preserves concavity.*

Theorem 3 (Lower Bound Property) *If* $\boldsymbol{\lambda} \geq 0$*, then* $d(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*$*. Here* $\boldsymbol{\lambda} \geq 0$ *means* $\lambda_i \geq 0$ *for any* $1 \leq i \leq m$*.* p^* *is the minimum value of the constrained minimization problem.*

Proof: For all feasible solution $\mathbf{x}$, for any $\boldsymbol{\lambda} \geq 0$, we have

$$f(\mathbf{x}) \geq L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \geq \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = d(\lambda, \nu),$$

where the first inequality follows from the fact that $g_i(\mathbf{x}) \leq 0$, $\lambda_i \geq 0$ and $h_i(\mathbf{x}) = 0$. Taking minimization over all feasible $\mathbf{x}$ of $f(\mathbf{x})$, we get

$$p^* \geq d(\boldsymbol{\lambda}, \boldsymbol{\nu}).$$

■

Next we introduce the Lagrangian dual problem (or dual problem for short), which is based on Lagrangian dual function.

Definition 3 (Lagrangian dual problem) *For a constrained minimization problem as in* (1) *and the corresponding dual function as in* (2)*, the dual problem is*

$$d^* = \max_{\boldsymbol{\lambda}, \boldsymbol{\nu}} \quad d(\boldsymbol{\lambda}, \boldsymbol{\nu}), \qquad \textit{subject to} \qquad \boldsymbol{\lambda} \geq 0. \tag{3}$$

Remark 3

1. *By lower bound property of* $d(\boldsymbol{\lambda}, \boldsymbol{\nu})$*, the dual problem finds the best/tightest lower bound for* p^**.*
2. *The dual optimal value* d^* *exists.*
3. *We say* $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ *are feasible dual variables if* $\boldsymbol{\lambda} \geq 0$ *and* $(\boldsymbol{\lambda}, \boldsymbol{\nu}) \in \mathbf{dom}\ d$*.*

Theorem 4 (Weak Duality) *The optimal value of the dual problem in* (3) *is always less than or equal to the optimal value of the primal problem* (1)*. In other wrods,* $d^* \leq p^*$*.*

Proof: By Theorem 3, we have $d(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*$. Taking maximization over all feasible $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ on the left hand side of the inequality yields that $d^* \leq p^*$. ■

Note that weak duality holds in general.

Definition 4 (Strong Duality) *We say the strong duality holds if* $d^* = p^*$ *holds, w*

A natural question is when the strong duality holds. We will come back to this topic later.

Example 2 (Least Norm Solution) *Consider the constrained minimization problem*

$$\min_{\mathbf{x}} \quad \|\mathbf{x}\|_2^2 \quad \text{subjectto} \quad \mathbf{A}\mathbf{x} = \mathbf{b}.$$

The Lagrangian function is

$$L(\mathbf{x}, \boldsymbol{\nu}) = \|\mathbf{x}\|_2^2 + \boldsymbol{\nu}^\top (\mathbf{A}\mathbf{x} - \mathbf{b}).$$

Taking minimization over $\mathbf{x}$*, we get the Lagrangian dual function*

$$d(\boldsymbol{\nu}) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\nu}) = \min_{\mathbf{x}} \|\mathbf{x}\|_2^2 + \boldsymbol{\nu}^\top (\mathbf{A}\mathbf{x} - \mathbf{b}). \tag{4}$$

By the first order condition, the minimizer $\mathbf{x}^*$ *satisfies* $2\mathbf{x}^* + \mathbf{A}^\top \boldsymbol{\nu} = 0$*, which yields that* $\mathbf{x}^* = -\mathbf{A}^\top \boldsymbol{\nu}/2$*. Substituting this solution to the optimization problem in* (4)*, we get*

$$d(\nu) = \frac{1}{4} \boldsymbol{\nu}^\top \mathbf{A}\mathbf{A}^\top \boldsymbol{\nu} + \boldsymbol{\nu}^\top (-\mathbf{A}\mathbf{A}^\top \boldsymbol{\nu}/2 - \mathbf{b}) = -\frac{1}{4} \boldsymbol{\nu}^\top \mathbf{A}\mathbf{A}^\top \boldsymbol{\nu} - \mathbf{b}^\top \boldsymbol{\nu}.$$