Lecture #13 — April 20, 2022

Lecturer: Yura Malitsky *Scribe: Guang Wei*

1 Moreau's identity

Theorem 1. *For a convex and closed function $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, the following expression holds:*

$$\mathrm{prox}_f \mathbf{x} + \mathrm{prox}_{f^*} \mathbf{x} = \mathbf{x}, \quad \forall\, \mathbf{x} \in \mathbb{R}^n \tag{1}$$

Proof. Let $\mathbf{u} = \mathrm{prox}_f \mathbf{x}$, then one needs to show:

$$\mathrm{prox}_{f^*} \mathbf{x} = \mathbf{x} - \mathbf{u} \tag{2}$$

From the expression of $\mathbf{u}$

$$\mathbf{u} = \mathrm{prox}_f \mathbf{x} \tag{3}$$

and recall the definition of proximal operator

$$\mathrm{prox}_f = \arg\min_{\mathbf{u}} \{ f(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \} \tag{4}$$

By directly calculating the subgradient, one can obtain:

$$\begin{aligned}
\mathbf{0} &\in \partial f(\mathbf{u}) + \mathbf{u} - \mathbf{x} && (5)\\
\iff \mathbf{u} &= (\mathbf{I} + \partial f)^{-1} \mathbf{x} && (6)\\
\iff \mathbf{x} &\in \mathbf{u} + \partial f(\mathbf{u}) && (7)\\
\iff \mathbf{x} - \mathbf{u} &\in \partial f(\mathbf{u}) && (8)
\end{aligned}$$

Recall the theorem that for a convex and closed function $f\colon \mathbf{X} \to \mathbb{R} \cup \{\infty\}$, $\mathbf{p} \in \partial f(\mathbf{x}) \Leftrightarrow \mathbf{x} \in \partial f^*(\mathbf{p})$. Equation 8 can further indicates

$$\begin{aligned}
\mathbf{u} &\in \partial f^*(\mathbf{x} - \mathbf{u}) && (9)\\
\iff \mathbf{x} &\in \mathbf{x} - \mathbf{u} + \partial f^*(\mathbf{x} - \mathbf{u}) && (10)\\
\iff \mathbf{x} &= (\mathbf{I} + \partial f^*)(\mathbf{x} - \mathbf{u}) && (11)\\
\iff \mathbf{x} - \mathbf{u} &= (\mathbf{I} + \partial f^*)^{-1}(\mathbf{x}) && (12)\\
\iff \mathbf{x} - \mathbf{u} &= \mathrm{prox}_{f^*} \mathbf{x} && (13)
\end{aligned}$$

Equation 13 is identical to Equation 2, which ends the proof. □

This theorem has a more generalized form:

Theorem 2. *For a convex function $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, the following expression holds:*

$$\mathrm{prox}_{\alpha f} \mathbf{x} + \alpha \mathrm{prox}_{\frac{1}{\alpha} f^*} \frac{\mathbf{x}}{\alpha} = \mathbf{x}, \quad \forall\, \mathbf{x} \in \mathbb{R}^n, \alpha > 0 \tag{14}$$

The proof is similar as that from **Theorem 1** and α represents the step size in relevant algorithms.

2 Weak and strong duality

For convex functions f and g: $\mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, consider

$$p_* = \inf_{\mathbf{x}} \{f(\mathbf{x}) + g(\mathbf{A}\mathbf{x})\} \tag{15}$$

According to Legendre-Fenchel Transform,

$$\begin{aligned} p_* &= \inf_{\mathbf{x}} \{f(\mathbf{x}) + \sup_{\mathbf{y}} \{\langle \mathbf{A}\mathbf{x}, \mathbf{y}\rangle - g^*(\mathbf{y})\}\} && (16) \\ &= \inf_{\mathbf{x}} \sup_{\mathbf{y}} \{f(\mathbf{x}) + \langle \mathbf{A}\mathbf{x}, \mathbf{y}\rangle - g^*(\mathbf{y})\} && (17) \\ &\geq \sup_{\mathbf{y}} \inf_{\mathbf{x}} \{f(\mathbf{x}) + \langle \mathbf{A}\mathbf{x}, \mathbf{y}\rangle - g^*(\mathbf{y})\} && (18) \\ &= \sup_{\mathbf{y}} \{-\sup_{\mathbf{x}} \{-f(\mathbf{x}) - \langle \mathbf{A}\mathbf{x}, \mathbf{y}\rangle\} - g^*(\mathbf{y})\} && (19) \\ &= \sup_{\mathbf{y}} \{-\sup_{\mathbf{x}} \{\langle -\mathbf{A}^{\mathbf{T}}\mathbf{y}, \mathbf{x}\rangle - f(\mathbf{x})\} - g^*(\mathbf{y})\} && (20) \end{aligned}$$

Recall that

$$\phi^*(\mathbf{y}) = \sup_{\mathbf{x}} \{\langle \mathbf{y}, \mathbf{x}\rangle - \phi(\mathbf{x})\} \tag{21}$$

Equation 20 implies

$$p_* \geq \sup_{\mathbf{y}} \{-f^*(-\mathbf{A}^{\mathbf{T}}\mathbf{y}) - g^*(\mathbf{y})\} := d_* \tag{22}$$

NOTE: $p_* \geq d^*$ represents weak duality. On the other hand, $p_* = d_*$ indicates strong duality which has more sufficient condition (such as Slater's condition) which will not be further involved here. In the following parts in this lecture, we only use the conclusion of strong duality: $p_* = d_*$.

3 Examples

3.1 Example 1

Primal:

$$\begin{aligned} &\min_{\mathbf{x}} \langle \mathbf{c}, \mathbf{x}\rangle && (23) \\ &s.t.\ \mathbf{A}\mathbf{x} = \mathbf{b} && (24) \\ &\mathbf{x} \geq \mathbf{0} && (25) \end{aligned}$$

This primal problem can be rewritten as

$$\min_{\mathbf{x}} \{f(\mathbf{x}) + g(\mathbf{A}\mathbf{x})\} \tag{26}$$

where

$$\begin{aligned} f(\mathbf{x}) &= \langle \mathbf{c}, \mathbf{x}\rangle + \delta_{\geq \mathbf{0}}(\mathbf{x}) && (27) \\ g(\mathbf{z}) &= \delta_{\{\mathbf{b}\}}(\mathbf{z}) && (28) \end{aligned}$$

Equation 27 indicates that if $\mathbf{x} \geq \mathbf{0}$ is not satisfied, the value of f goes to $+\infty$, which makes Expression 26 goes to $+\infty$. Equation 27 demonstrates that Condition 25 must be satisfied. A similar interpretation can be made for Equation 28. By strong duality theorem,

$$\begin{aligned} f^*(\mathbf{y}) &= \sup_{\mathbf{x}} \{\langle \mathbf{x}, \mathbf{y}\rangle - f(\mathbf{x})\} && (29) \\ &= \sup_{\mathbf{x}} \{\langle \mathbf{x}, \mathbf{y}\rangle - \langle \mathbf{c}, \mathbf{x}\rangle - \delta_{\geq \mathbf{0}}(\mathbf{x})\} && (30) \\ &= \sup_{\mathbf{x} \geq \mathbf{0}} \langle \mathbf{x}, \mathbf{y} - \mathbf{c}\rangle && (31) \end{aligned}$$

Note that for any index i such that $y_i > c_i$, a feasible x will make $f^*(\mathbf{y})$ goes to $+\infty$. Equation 31 indicates

$$f^*(\mathbf{y}) = \begin{cases} +\infty & y_i > c_i,\ for\ some\ i \\ 0, & otherwise \end{cases} \tag{32}$$

$$= \delta_{\leq \mathbf{c}}(\mathbf{y}) \tag{33}$$

For the function g,

$$g^*(\mathbf{y}) = \sup_{\mathbf{z}}\{\langle \mathbf{y}, \mathbf{z}\rangle - \delta_{\{\mathbf{b}\}}(\mathbf{z})\} \tag{34}$$

$$= \sup_{\mathbf{z}=\mathbf{b}} \langle \mathbf{y}, \mathbf{z}\rangle \tag{35}$$

$$= \langle \mathbf{y}, \mathbf{b}\rangle \tag{36}$$

The dual problem can be formulated by

$$\max_{\mathbf{y}}\{-f^*(-\mathbf{A}^{\mathbf{T}}y) - g^*(\mathbf{y})\} \tag{37}$$

$$= \max_{\mathbf{y}}\{-\delta_{\leq \mathbf{c}}(\mathbf{y}) - \langle \mathbf{y}, \mathbf{b}\rangle\} \tag{38}$$

Dual:

$$\max_{\mathbf{y}} \quad -\langle \mathbf{y}, \mathbf{b}\rangle \tag{39}$$

$$s.t. \quad -\mathbf{A}^{\mathbf{T}}\mathbf{y} \leq \mathbf{c} \tag{40}$$

By substitution of variable,
Dual:

$$\max_{\mathbf{y}} \quad \langle \mathbf{y}, \mathbf{b}\rangle \tag{41}$$

$$s.t. \quad \mathbf{A}^{\mathbf{T}}\mathbf{y} \leq \mathbf{c} \tag{42}$$

3.2 Example 2

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x} - \mathbf{u}\|^2 + \lambda\|\mathbf{A}\mathbf{x}\|_1 \tag{43}$$

Prerequisite knowledge: For the function $\phi(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$,

$$\mathrm{prox}_{\phi}\mathbf{x} = \begin{cases} x_i - \lambda & x_i > \lambda,\ for\ some\ i \\ x_i + \lambda & x_i < -\lambda,\ for\ some\ i \\ 0, & otherwise \end{cases} \tag{44}$$

From the optimization problem 43, functions f and g are set to be

$$f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - \mathbf{u}\|^2 \tag{45}$$

$$g(\mathbf{z}) = \lambda\|\mathbf{z}\|_1 \tag{46}$$

By strong duality theorem,

$$f^*(\mathbf{y}) = \sup_{\mathbf{x}}\{\langle \mathbf{x}, \mathbf{y}\rangle - \frac{1}{2}\|\mathbf{x} - \mathbf{u}\|^2\} \tag{47}$$

By calculating the subgradient directly, one can obtain $\mathbf{x} = \mathbf{y} + \mathbf{u}$.

$$
\begin{aligned}
f^*(\mathbf{y}) &= \langle \mathbf{y}, \mathbf{y} + \mathbf{u} \rangle - \frac{1}{2}\|\mathbf{y}\|^2 && (48)\\
&= \frac{1}{2}\|\mathbf{y}\|^2 + \langle \mathbf{y}, \mathbf{u} \rangle && (49)
\end{aligned}
$$

For function g,

$$
\begin{aligned}
g^*(\mathbf{y}) &= \sup_{\mathbf{z}} \{\langle \mathbf{y}, \mathbf{z} \rangle - \lambda \|\mathbf{z}\|_1\} && (50)\\
&= \begin{cases} +\infty & |y_i| > \lambda, \ for\ some\ i \\ 0, & otherwise \end{cases} && (51)\\
&= \delta_{B_\infty(0,\lambda)} && (52)
\end{aligned}
$$

where $\delta_{B_\infty(0,\lambda)}$ represents a ball with infinity-norm with center at $\mathbf{0}$ and radius of λ.
Dual:

$$
\begin{aligned}
\max_{\mathbf{y}} \quad & -\frac{1}{2}\|\mathbf{A}^{\mathbf{T}}\mathbf{y}\|^2 + \langle \mathbf{A}^{\mathbf{T}}\mathbf{y}, \mathbf{u} \rangle && (53)\\
s.t. \quad & \|y_\infty\| < \lambda && (54)
\end{aligned}
$$

By substitution of variable,
Dual:

$$
\begin{aligned}
\max_{\mathbf{y}} \quad & -\frac{1}{2}\|\mathbf{A}^{\mathbf{T}}\mathbf{y}\|^2 - \langle \mathbf{A}^{\mathbf{T}}\mathbf{y}, \mathbf{u} \rangle && (55)\\
s.t. \quad & \|y_\infty\| < \lambda && (56)
\end{aligned}
$$

3.3 Example 3 (exercise)

Primal:

$$
\begin{aligned}
&\min_{\mathbf{x}} |\mathbf{x}\|_1 && (57)\\
&s.t.\ \mathbf{A}\mathbf{x} = \mathbf{b} && (58)
\end{aligned}
$$

Functions f and g are set to be

$$
\begin{aligned}
f(\mathbf{x}) &= \|\mathbf{x}\|_1 && (59)\\
g(\mathbf{z}) &= \delta_{\{\mathbf{b}\}}(\mathbf{z}) && (60)
\end{aligned}
$$

From the first two examples, we know that

$$
\begin{aligned}
f^*(\mathbf{x}) &= \delta_{B_\infty(0,1)} && (61)\\
g^*(\mathbf{y}) &= \langle \mathbf{y}, \mathbf{b} \rangle && (62)
\end{aligned}
$$

Dual:

$$
\begin{aligned}
\max_{\mathbf{y}} \quad & -\langle \mathbf{y}, \mathbf{b} \rangle && (63)\\
s.t. \quad & |-\mathbf{A}^{\mathbf{T}}\mathbf{y}\|_\infty \leq 1 && (64)
\end{aligned}
$$