Duality Recovery

1. Norm $f(x) = \|x\|$

$\downarrow$

2. $\boxed{\text{Dual Norm}}$ → measure the length of induced norm (operator norm)

$\forall x \in \mathbb{R}^n \longrightarrow$ induce a Linear Functional

$x \longleftrightarrow \boxed{f_x(z) := x^T z}$

consider Operator Norm of f_x

$$\|f_x\| := \max_{y \neq 0} \frac{\|f_x(y)\|}{\|y\|}$$

$$= \max_{\|y\| \leq 1} \langle x, y \rangle$$

$$:= \|x\|_*$$

$$\|x\|_* = \max \{ \langle x, y \rangle : \|y\| \leq 1 \}$$

Property:

① $\forall x, z \in \mathbb{R}^n$, we have $\|x\| \, \|z\|_* \geq \langle x, z \rangle$

Pf: $\|z\|_* := \sup \{ \langle x, z \rangle : \|x\| \leq 1 \}$

$$= \sup \left\{ \frac{\langle x, z \rangle}{\|x\|} : \forall x \neq 0 \in \mathbb{R}^n \right\}$$

$\Rightarrow \forall x \neq 0 \in \mathbb{R}^n$, we have $\|z\|_* \geq \dfrac{\langle x, z \rangle}{\|x\|}$

$\Leftrightarrow \|x\| \, \|z\|_* \geq \langle x, z \rangle$

② $f(x) = \|x\|_p \Rightarrow \|x\|_* = \|x\|_q$ s.t $\frac{1}{p} + \frac{1}{q} = 1$

☆ Pf: Holder Ineq $\Rightarrow x^T z \leq \|x\|_q \|z\|_p$

$\boxed{\text{Dual Norm of p-norm}}$ is $\boxed{\text{q - norm}}$

$$\Rightarrow \|x\|_* = \sup_z \frac{z^T x}{\|z\|_p} = \|x\|_q$$

$\boxed{\text{dual norm of p-norm}} \longleftrightarrow \boxed{\text{q-norm}}$

③ $\boxed{\|x\|_{**} = \|x\|}$

Pf: $\|x\| = \begin{cases} \min_{y} & \|y\| \\ \text{s.t} & y = x \end{cases}$ (1)

$\boxed{\text{triangle inequality of norm}}$

Dual Problem of $\boxed{(1)}$ convex since $\|\lambda x + (1-\lambda) y\| \leq \lambda \|x\| + (1-\lambda)\|y\|$

$\Downarrow$

$L(y; u) = \|y\| + u^T(x-y)$

consider $\min_{y} L(y,u) = \min_{y} \underline{\|y\| - u^T y} + u^T x$

Here, 2 ways of analysis:

1. Observation: $\min_{y} \|y\| - u^T y = -f^*(u)$ $\boxed{f(\cdot) = \|\cdot\|}$

$= -\delta_{B_1^*}(u)$

$= \begin{cases} -\infty, & \text{o/w} \\ 0, & \|u\|_* \leq 1 \end{cases}$

2. Direct Calculation: Consider $\|u\|_* = \sup\{\langle u, y\rangle : \|y\| \leq 1\}$

$\rightarrow$ if $\|u\|_* > 1$, then $\exists \tilde{y}$ s.t

$\|\tilde{y}\| - u^T\tilde{y} < 0 \rightarrow -\infty$

$\rightarrow$ if $\|u\|_* \leq 1 \Rightarrow \forall y, \dfrac{\langle y, u\rangle}{\|y\|} \leq 1$

$\Rightarrow \langle y, u\rangle \leq \|y\| \quad \forall y \in \mathbb{R}^n$

$\Rightarrow$ $\underline{\min = 0}$

Thus, Dual Problem $\longrightarrow \max_{u} \ \theta(u)$

$\Leftrightarrow \max_{u} \ u^T x$

s.t $\|u\|_* \leq 1$

$\Leftrightarrow (\|x\|_*)_*$

$\boxed{\text{Weak Slater's condition}}$

$\downarrow$

Strong Duality $\Rightarrow \underline{(\|x\|_*)_* = \|x\|}$

3. Conjugate Function

$$\boxed{f^*(u)} := \sup_x \{ f_u(x) - f(x) \}$$

$u \longrightarrow$ induce $f_u(x) = u^T x$

$f^*(u)$ measures the distinguish between $f_u(x)$ & $f(x)$

$$= \sup_x \{ u^T x - f(x) \}$$

Application:

① $f(x) = \|x\| \longrightarrow$ Positive Homogeneous

a) $(f^*)^* = f$ if f is closed & proper & convex

b) $\partial f(x) = \{ y : \|y\|_* \leq 1 ,\ \langle x, y \rangle = \|x\| \}$

Pf: $y \in \partial f(x)$

$\Leftrightarrow \forall u \in \mathbb{R}^n,\ \|u\| \geq \|x\| + y^T(u-x)$

$\Rightarrow \begin{cases} u = 2x \Rightarrow \|x\| \geq y^T x \\ u = 0 \Rightarrow \|x\| \leq y^T x \end{cases} \Rightarrow \langle x, y \rangle = \|x\|$ (necessary condition)

$\Leftrightarrow \forall u \in \mathbb{R}^n,\ \|u\| \geq y^T u$ & $\|x\| = \langle x, y \rangle$

$\Leftrightarrow \|y\|_* = \sup_u \frac{\langle y, u \rangle}{\|u\|} \leq 1$ & $\|x\| = \langle x, y \rangle$

$\Rightarrow \partial f(x) = \{ y : \|y\|_* \leq 1 ,\ \|x\| = \langle x, y \rangle \}$

$\Rightarrow \partial f(0) = B_*^1 := \{ y : \|y\|_* \leq 1 \}$

c) $f^*(y) = \delta_{B_*^1}(y)$

Generally speaking, $f^*(y) = \delta_{\partial f(0)}(y)$

Pf Sketch: 1. f^* can only be 0 & $+\infty$

2. consider $C = \{ y : f^*(y) = 0 \}$

$y \in C \Leftrightarrow f^*(y) = 0$

$\Leftrightarrow \sup_x \{ \langle y, x \rangle - f(x) \} = 0$

$\Leftrightarrow \forall x,\ y^T x \leq f(x)$

$\Leftrightarrow \forall x,\ y^T(x - 0) + f(0) \leq f(x)$

$$\Leftrightarrow \quad y \in \partial f(0)$$

$$\Rightarrow C = \partial f(0)$$

3. from 1 & 2. Firstly, f^* can be written as

$$f^*(y) = \delta_C(y)$$

Then, since $C = \partial f(0)$,

$$\underline{f^*(y) = \delta_{\partial f(0)}(y)}$$

② $f(x) = \delta_C(x)$ $\quad$ C is <u>convex & closed</u>

a) $\delta_C^*(y) = \sup\limits_{x} \{ \langle y, x \rangle - \delta_C(x) \}$

$$= \sup_{x \in C} \{ \langle y, x \rangle \}$$

b) $(\delta_C^*)^* = \delta_C$

c) $\partial \delta_C(x) = N_C(x) := \{ y : y^T(u-x) \le 0 \ \ \forall u \in C \}$

<u>Pf</u>: $y \in \partial \delta_C(x) \Leftrightarrow \forall u, \ \ \delta_C(u) \ge \delta_C(x) + y^T(u-x)$

$$\Leftrightarrow \forall u \in C, \quad 0 \ge y^T(u-x)$$

$$\Leftrightarrow y \in N_C(x)$$

4. <u>Some Application</u>:

a) <u>Toy - Example</u> <u>(use conjugate function to do simplification)</u>

$$\begin{cases} \min & \sum_{i=1}^n f_i(x_i) \\ \text{s.t} & a^T x = b \end{cases} \xrightarrow{\text{dual}} L(x; u) = \sum_{i=1}^n f_i(x_i) + \nu(b - a^T x)$$

$$\theta(u) := \min_x \sum_{i=1}^n f_i(x_i) + \nu b - \sum_{i=1}^n \nu a_i x_i$$

$$= \nu b + \sum_{i=1}^{n} \min_{x_i} \{ f_i(x_i) - \nu a_i x_i \}$$

$$= \nu b - \sum_{i=1}^{n} f_i^*(\nu a_i)$$

b) LASSO

$$\min_{\beta} \frac{1}{2} \| X\beta - y \|_2^2 + \lambda \|\beta\|_1$$

(Proximal Gradiant)

soft-threshold

(Actually we can derive iterative algorithm to solve for $\hat{\beta}$, but here we focus on, understanding the property of minimizer $\hat{\beta}$)

$$\begin{cases} \min_{z,\beta} \frac{1}{2} \| z - y \|_2^2 + \lambda \|\beta\|_1 \\ \text{s.t} \quad z = X\beta \end{cases}$$

$\rightarrow$ $\boxed{\text{weak Slater's condition}}$

$\rightarrow L(z, \beta ; u) = \frac{1}{2} \| z - y \|_2^2 + \lambda \|\beta\|_1 + u^T (z - X\beta)$

$\rightarrow \theta(u) = \min_{z,\beta} L(z, \beta ; u) = \min_{z,\beta} \frac{1}{2} \| z - y \|_2^2 + \lambda \|\beta\|_1 + u^T (z - X\beta)$

$$= \min_{z} \frac{1}{2} \| z - y \|_2^2 + u^T z \qquad \boxed{\hat{z} = y - u}$$

$$+ \min_{\beta} \lambda \|\beta\|_1 - (X^T u)^T \beta$$

Here, $g(\cdot) = \|\cdot\|_1$

$\Rightarrow g^*(\cdot) = \delta_{B_1^\infty}(\cdot)$

$$= \frac{1}{2} u^T u + u^T (y - u)$$

$$- \lambda g^* \left(\frac{X^T u}{\lambda} \right)$$

$\rightarrow$ Dual Problem : $\max_u \theta(u)$

$$\Leftrightarrow \begin{cases} \max_u \frac{1}{2} (\|y\|_2^2 - \|y - u\|_2^2) \\ \text{s.t} \quad \left\| \frac{X^T u}{\lambda} \right\|_\infty \le 1 \end{cases} \Leftrightarrow \boxed{\|X^T u\|_\infty \le \lambda}$$

$$\xleftrightarrow[\text{w.r.t}]{\text{argmax}} \quad \begin{cases} \max\limits_{u} & -\frac{1}{2}\|y-u\|_2^2 \\ \text{s.t} & \|X^T u\|_\infty \le \lambda \end{cases} \qquad \boxed{P_{\delta_C}(y)}$$

$$\Rightarrow \hat{u} = \underset{X^T u \in B_\lambda^\infty}{\arg\min} \|y-u\|_2^2 = \underset{u}{\arg\min} \frac{1}{2}\|u-y\|_2^2 + \delta_C(u)$$

$$= \Pi_C(y) \qquad (C := \{u : \|X^T u\|_\infty \le \lambda\})$$

Since $\underline{\hat{z} = y - u = X\hat{\beta}}$ from stationary condition,

then $\hat{u} = y - X\hat{\beta}$ represents the residual !

This can explain the Robustness of LASSO :

Projection to B_λ^∞-ball is a robust operation !