Convex Optimization.

1. Lagrangian. (Conjugate function)

2. Dual problem

3. KKT

4. What is Lasso.

Lagrangian.

standard problem is

$$\min f_0(x) \qquad (*)$$
$$\text{s.t. } f_i(x) \le 0 \quad i = 1 \cdots m$$
$$h_i(x) = 0 \quad i = 1 \cdots p$$

Lagrangian of $(*)$ is

$$f_0(x) + \underline{\sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x)} := L$$

where L is function of $L(x, \lambda, \mu)$.

Lagrangian dual function. $g: \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

$$g(\lambda, \mu) = \inf_{x \in D} L(x, \lambda, \mu)$$

Rmk: g is always concave.
and g could be $-\infty$

Rmk: $\lambda \succeq 0$ then $g(\lambda, \mu) \le p^*$.
where p^* is the optimal value of original optimization problem.

$x \succeq_K y \Rightarrow x - y \in K$.

Thm: If $\lambda \succeq 0$. then $g(\lambda, \mu) \le p^*$. (lower bound of optimization)

Pf: If $\tilde{x}$ is a feasible point,
i.e. $f_i(\tilde{x}) \le 0$, and $h_i(\tilde{x}) = 0$, then.

$$f_0(\tilde{x}) \ge L(\tilde{x}, \lambda, \mu) \quad \text{for any } \lambda, \mu \text{ if } \lambda \succeq 0$$
$$\ge \inf_x L(\tilde{x}, \lambda, \mu)$$
$$\stackrel{\triangle}{=} g(\lambda, \mu).$$

This provide us a nontrivial lower bound. If $\lambda \succeq 0$, and $g(\lambda, \mu) > -\infty$, (λ, μ) is called dual feasible.

Ex) least norm problem.

$$\begin{cases} \min & x^T x \\ \text{s.t} & Ax = b \end{cases} \quad (*)$$

dual function of $(*)$ is:

$L(x, \mu) = \underline{x^T x + \mu^T (Ax - b)} \rightarrow$ quadratic so it is minimized at $x = -\frac{1}{2} A^T \mu$

$g(\mu) = \inf_x L(x, \mu)$

$= L(-\frac{1}{2} A^T \mu, \mu)$

$= \frac{1}{4} \mu^T A A^T \mu + \mu^T (-\frac{1}{2} A A^T \mu - b)$

$= -\frac{1}{4} \mu^T A A^T \mu + (-b^T \mu)$

$\rightarrow$ so $p^* \geq -\frac{1}{4} \mu^T A A^T \mu - b^T \mu$. for all μ.

Ex) LP

$\min c^T x$

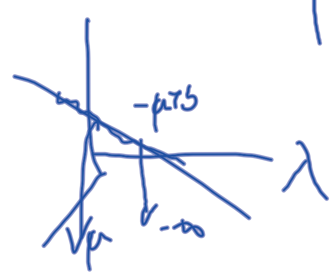
s.t $Ax = b$, $x \geq 0$

Lagrangian is $L(x, \lambda, \mu) = c^T x + \mu^T (Ax - b) - \lambda^T x$
$= -\mu^T b + (c + A^T \mu - \lambda)^T x$

$\Rightarrow g(\lambda, \mu) = \inf_x L(x, \lambda, \mu)$

$$= \begin{cases} -\mu^T b & c + A^T \mu - \lambda = 0 \\ -\infty & \text{o.w.} \end{cases}$$

$\rightarrow$ affine space



Ex1 (Partition)

$$\begin{cases} \min & x^T W x = \sum w_{ij} x_i x_j \\ \text{s.t.} & x_i^2 = 1 \quad i = 1 \dots n. \end{cases}$$

Lagrangian:

$$L(x, \mu) = x^T W x + \sum \mu_i (x_i^2 - 1)$$

so
$$g(\mu) = \inf_x \{ x^T W x + \sum \mu_i (x_i^2 - 1) \}$$
$$= \inf_x \{ x^T W x + \underline{\sum \mu_i x_i^2} - \sum \mu_i \}$$
$$= \inf \{ x^T W x + x^T \operatorname{diag}(\mu) x - \mu^T \mathbb{1} \}$$
$$= \inf_x \{ x^T (W + \operatorname{diag}(\mu)) x \} - \mu^T \mathbb{1}$$

$$= \begin{cases} -\mu^T \mathbb{1} & \text{if } W + \operatorname{diag}(\mu) \succeq 0 \\ -\infty & \text{o.w} \end{cases}$$

If μ is $-\min_i \lambda_i(W) \cdot \mathbb{1}$, then $W + \operatorname{diag}(\mu) \succeq 0$

$\Rightarrow p^* \geq n \cdot \lambda_0(W)$

Dual & Conjugate function

Def: For any $f: \mathbb{R}^n \to \mathbb{R}$, the conjugate is defined by $f^*(y) = \max_{x \in \operatorname{dom}(f)} \{ y^T x - f(x) \}$ (sup)

If f is continuous, f^* is legendre transform

Ex) $f(x) = \frac{1}{2}x^TQx$ for $Q \succ 0$

$y^Tx - \frac{1}{2}x^TQx$ is concave,

and $f^*(y) = \max_x (y^Tx - \frac{1}{2}x^TQx)$
$= \frac{1}{2}y^TQ^{-1}y$

is the conjugate to $f(x)$.

Ex) $f(x) = I_S(x)$, S is some set on $\mathbb{R}^n$

$f^*(y) = I_S^*(y) = \max_{x \in S}\{y^Tx\}$

is called support function of S

Ex) $f(x) = \|x\|$. $\|\cdot\|$ is any norm

$$f^*(y) = \begin{cases} 0 & \text{if } \|y\|_* \le 1 \\ \infty & \text{o.w.} \end{cases}$$

you can show this

is the indicator function of unit ball of dual norm

$(\|y\|_* = \sup\{z^Ty \mid \|z\| \le 1\})$

Actually. If $\|\cdot\|$ is ℓ^2, $\|\cdot\|_*$ is ℓ^2

$\|\cdot\|$ is ℓ^p, $\|\cdot\|_*$ is ℓ^q. $(\frac{1}{p} + \frac{1}{q} = 1)$

Lasso $\Leftarrow$ $\|\cdot\|$ is ℓ^1, $\|\cdot\|_*$ is $\|\cdot\|_\infty$

$\|\cdot\|$ is nuclear, $\|\cdot\|_* = \sigma_{\max}(\cdot)$

Conjugate & dual problems

$$\min_x f(x) + g(x)$$

Trick: $\min_{x,z} f(x) + g(z)$

s.t. $z = x$

Lagrange dual is

$$\inf_{x,z} f(x) + g(z) + \mu^T(z - x)$$

$$= \inf_x \{f(x) - \mu^T x\} + \inf_z \{g(z) + \mu^T z\}$$

$$= -f^*(\mu) - g^*(-\mu)$$

f^*, g^* all conjugate.

Exl. $\min \|x\|$

s.t. $Ax = b$

dual function is

$$g(\mu) = \inf_x \{\|x\| + \lambda(Ax - b)\}$$

$$= \begin{cases} b^T v & \|A^T v\|_* \leq 1 \\ -\infty & \text{o.w.} \end{cases}$$

(you can try this your self)

Dual problem

$\max_{\mu, \lambda} g(\mu, \lambda)$ is called the dual problem.

$= d^*$

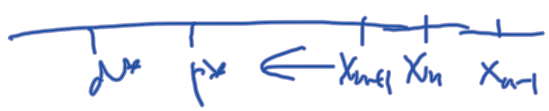
Weak duality: $d^* \le p^*$

d^*: dual problem.

p^*: solution of original problem

$(g(\mu,\lambda) \le p^*)$

[It is used to determine stopping criterion! If $|f(x_{n+1}) - d^*| < \varepsilon$, then: $|f(x_{n+1}) - p^*|$ is also small].



Strong duality: $d^* = p^*$.

Rmk: Strong duality hold for convex optimization. (usually)

Slater's condition (theorem)

Consider
$$\begin{aligned} &\min f_0(x) \\ &\text{s.t. } Ax = b \\ &\quad f_i(x) \le 0 \quad i = 1 \cdots m \end{aligned}$$

If the problem is strictly feasible then, the strong duality is always true. i.e. if $\exists x \in \mathrm{dom}(f_i)$. s.t. $f_i(x) < 0$ $\forall i$; $d^* = p^*$.

Ex) LP. primal

$$\left.\begin{array}{l} \min\ c^T x \\ \text{s.t. } Ax \le b \end{array}\right\} (*)$$

dual function.

$$g(\lambda) = \inf_x \left(c^T x + \lambda (Ax-b)\right)$$

$$= \begin{cases} -b^T\lambda & \text{if } A^T\lambda + c = 0 \\ -\infty & \text{o.w.} \end{cases}$$

dual problem :

$$\left.\begin{array}{l} \min\ b^T\lambda \\ \text{s.t. } A^T\lambda + c = 0,\ \lambda \ge 0. \end{array}\right\} (**)$$

if $\exists \tilde{x}$. $A\tilde{x} < b$, p^* of $(*)$

$= d^*$ of $(**)$

Ex) Quadratic problem.

$$\left\{\begin{array}{l} \min\ x^T P x \\ \text{s.t. } Ax \le b \end{array}\right. \qquad P > 0.$$

dual function.

$$g(\lambda) = \inf_x \{x^T P x + \lambda (Ax-b)\}$$

$$= -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T\lambda.$$

dual problem. is

$$\max_{\lambda \ge 0}\ -\frac{1}{4}\left(\lambda^T A P^{-1} A^T \lambda\right) - b^T\lambda.$$

KKT (Karush-Kuhn-Tucker)

Complementary-Slackness.

$$f_0(x^*) = g(\mu^*, \lambda^*)$$
$$= \inf_x \{ f_0(x) + \sum \lambda_i^* f_i(x) + \sum \mu_i^* h_i(x) \}$$
$$\le f_0(x^*) + \underbrace{\sum \lambda_i^* f_i(x^*)}_{\le 0} + \underbrace{\sum \mu_i^* h_i(x^*)}_{=0}$$
$$\le f_0(x^*)$$

$\Rightarrow \sum \lambda_i^* f_i(x^*) = 0 \qquad \lambda_i \ge 0$

$\Rightarrow \begin{cases} \text{if } \lambda_i > 0 & f_i(x^*) = 0 \\ \text{if } f_i(x^*) < 0 & \lambda_i = 0. \end{cases}$

KKT condition

i) $0 \in \partial f(x) + \sum_{i=1}^{m} \lambda_i \partial f_i(x) + \sum_{i=1}^{p} \mu_i \partial h_i(x)$. (Stationary)

ii) $\lambda_i \ge 0$ for all i (dual feasibility)

iii) $f_i(x) \le 0$, $h_i(x) = 0 \quad \forall i$. (primal feasibility)

iv) $\lambda_i f_i(x) = 0$ for all i (C-S)

Pf: (Boyd, C.O.)

Rmk: If $\bar{x}, \bar{\lambda}, \bar{\mu}$ are KKT point, then, for a convex problem, they are optimal.

Rmk: if Slater's condition hold.
$\bar{x}$ is optimal iff $\exists$ KKT point λ, μ.