(HW2)

(Ex1) (LP Duality)

(1.) (P) $\begin{cases} \min_x & c^T x \\ \text{s.t} & Ax = b \\ & x \geq 0 \Leftrightarrow -x \leq 0 \end{cases}$

$$L(\underset{\in \mathbb{R}^d}{x}, \underset{\in \mathbb{R}^d}{\lambda}, \underset{\in \mathbb{R}^m}{\nu}) = c^T x - \lambda^T x + \nu^T (Ax - b)$$

$$= (c^T - \lambda^T + \nu^T A) x - \nu^T b$$

$$= (c + A^T \nu - \lambda)^T x - \nu^T b$$

L is linear in x, thus:

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} -\nu^T b & \text{if } A^T \nu - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

b/c we can take $x_i^{(m)} = -\delta_{ij} m$ where $(A^T \nu - \lambda + c)_j > 0 \Rightarrow L(x^{(m)}, \lambda, \nu) \xrightarrow[m \to +\infty]{} -\infty$

$\Rightarrow (P_{Dual}) \begin{cases} \max & g(\lambda, \nu) \\ \text{s.t} & \lambda \geq 0 \\ & \nu \in \mathbb{R}^m \end{cases}$

However, $\max_{\substack{\lambda \geq 0 \\ \nu \in \mathbb{R}^m}} g(\lambda, \nu) = \max \Big(\underbrace{\max_{\substack{\nu \in \mathbb{R}^m \\ \lambda \geq 0 \\ A^T \nu + c = \lambda}} g(\lambda, \nu)}_{-\nu^T b}, \underbrace{\max_{\substack{\nu \in \mathbb{R}^m \\ \lambda \geq 0 \\ A^T \nu + c \neq 0}} g(\lambda, \nu)}_{= -\infty} \Big)$

$$= \max_{\substack{\nu \in \mathbb{R}^m \\ A^T \nu + c \geq 0 \\ A^T \nu + c = \lambda}} g(\lambda, \nu)$$

$g(\lambda, \nu)$ doesn't depend on λ

$$= \max_{\substack{\nu \in \mathbb{R}^m \\ A^T \nu + c \geq 0}} -\nu^T b$$

$\nu^T b \in \mathbb{R}$

$\Rightarrow (\nu^T b)^T = \nu^T b$

$$= \max_{\substack{\nu \in \mathbb{R}^m \\ A^T \nu + c \geq 0}} -b^T \nu$$

Thus: $\boxed{(P_{Dual}) \begin{cases} \max & -b^T \nu \\ \text{s.t} & A^T \nu + c \geq 0 \\ & \nu \in \mathbb{R}^m \end{cases}}$

(2.)

(D) $\begin{cases} \max_y & b^T y \\ \text{s.t} & A^T y \leq c \end{cases}$

$$L(\underset{\in \mathbb{R}^m}{y}, \underset{\in \mathbb{R}^d}{\lambda}) = b^T y + \lambda^T (-A^T y + c)$$

$$= (b^T - \lambda^T A^T) y + \lambda^T c$$

$\lambda^T c \in \mathbb{R}$

$$= (-A\lambda + b)^T y + c^T \lambda$$

$\Rightarrow$ L is linear in y, so:

$$g(\lambda) = \sup_y L(y, \lambda) = \begin{cases} +c^T \lambda & \text{if } -A\lambda + b = 0 \\ +\infty & \text{otherwise} \end{cases}$$

Thus similarly to (1.), we prove:

$$\min_{\lambda \geq 0} g(\lambda) = \min \Big(\underbrace{\min_{\substack{\lambda \geq 0 \\ A\lambda + b = 0}} g(\lambda)}_{= -c^T \lambda}, \underbrace{\min_{\substack{\lambda \geq 0 \\ A\lambda + b \neq 0}} g(\lambda)}_{= +\infty} \Big)$$

$$= \min_{\substack{\lambda \geq 0 \\ -A\lambda + b = 0}} +c^T \lambda$$

Hence: $\boxed{(D_{dual}) \begin{cases} \min & -c^T \lambda \\ \text{s.t} & A\lambda = +b \\ & \lambda \geq 0 \end{cases}}$

(3.) We denote by:

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^{d+m}, \quad a = \begin{pmatrix} c \\ -b \end{pmatrix} \in \mathbb{R}^{d+m}$$

$$B_1 = (A \quad 0) \in \mathbb{R}^{m \times (m+d)}, \quad B_2 = \begin{pmatrix} -I_d & 0 \\ 0 & A^T \end{pmatrix} \in \mathbb{R}^{2d \times (m+d)}$$

$$\gamma = \begin{pmatrix} 0 \\ c \end{pmatrix} \in \mathbb{R}^{2d}$$

Thus, one can write:

$$(\text{Self-Dual})\begin{cases}\min\limits_{z} & a^T z \\ \text{s.t} & B_1 z = b \\ & B_2 z \leq \gamma\end{cases}$$

$$L(\underset{\in\mathbb{R}^{d+m}}{z}, \underset{\in\mathbb{R}^{2d}}{\lambda}, \underset{\in\mathbb{R}^{m}}{\nu}) = a^T z + \lambda^T(B_2 z - \gamma) + \nu^T(B_1 z - b)$$

$$= (a^T + \lambda^T B_2 + \nu^T B_1) z - \lambda^T\gamma - \nu^T b$$

$$= (a + B_2^T\lambda + B_1^T\nu) z - \lambda^T\gamma - \nu^T b$$

We show like in (1.) and (2.) that for $g(\lambda, \nu) = \inf\limits_{z} L(z, \lambda, \nu)$ since L is linear in z that:

$$(\text{Self-Dual}_{\text{Dual}}): \begin{cases}\max & -\gamma^T\lambda - b^T\nu \\ \text{s.t} & -A^T\nu + \lambda_1 = c \\ & A\lambda_2 = b \\ & \lambda \geq 0 \\ & \nu \in \mathbb{R}^m\end{cases}$$

So we can substitute ν by $-\nu$ ($\nu \in \mathbb{R}^m$)

$$\equiv \begin{cases}\max & -\gamma^T\lambda + b^T\nu \\ \text{s.t} & A^T\nu + \lambda_1 = c \\ & A\lambda_2 = b \\ & \lambda \geq 0 \\ & \nu \in \mathbb{R}^m\end{cases}$$

$$\equiv \begin{cases}-\min & -(-\gamma^T\lambda + b^T\nu) = \gamma^T\lambda - b^T\nu \\ \text{s.t} & A^T\nu + \lambda_1 = c \\ & A\lambda_2 = b \\ & \lambda \geq 0 \\ & \nu \in \mathbb{R}^m\end{cases}$$

$\gamma^T\lambda = (0\ c^T)\lambda = c^T\lambda$ ~~(we write~~ $\lambda = (\underset{\in\mathbb{R}^d}{\ast}\ \underset{\in\mathbb{R}^d}{\lambda})$)

~~Let~~ $S_1 = \{(\lambda, \nu) \mid A^T\nu + \lambda = c,\ A\lambda = b,\ \lambda \geq 0\}$

$S_2 = \{(\lambda, \nu) \mid A^T\nu \leq c,\ A\lambda = b,\ \lambda \geq 0\}$

~~Obviously, one can write~~ $S_1 \subset S_2$

~~So~~ $\underbrace{\min\limits_{S_2} c^T\lambda - b^T\nu}_{= p^* \text{ optimal value of the (Self-Dual)}} \leq \underbrace{\min\limits_{S_1} c^T\lambda - b^T\nu}_{= d^* \text{ optimal value of the (Self-Dual}_{\text{Dual}})}$

~~However, by the lower bound property we know:~~ $p^* \geq d^*$ ~~and thus~~

$p^* = d^*$

~~Meaning that one can write:~~

$$(\text{Self-Dual}_{\text{Dual}})\begin{cases}\min & c^T\lambda - b^T\nu \\ \text{s.t} & A^T\nu \leq c \\ & A\lambda = b \\ & \lambda \geq 0\end{cases} \equiv (\text{Self-Dual})$$

~~Thus, (Self-Dual) is a self-Dual pb~~

Note that we write $\lambda = (\underset{\in\mathbb{R}^d}{\lambda_1^T}\ \underset{\in\mathbb{R}^d}{\lambda_2^T})^T$

That's why we found the constraints from:

$$a + B_2^T\lambda + B_1^T\nu = 0_{d+m}$$

$$\begin{pmatrix} c \\ -b\end{pmatrix} + \begin{pmatrix} -I_d & 0 \\ 0 & A\end{pmatrix}\begin{pmatrix}\lambda_1 \\ \lambda_2\end{pmatrix} + \begin{pmatrix} A^T \\ 0\end{pmatrix}\nu = \begin{pmatrix} 0_d \\ 0_m\end{pmatrix}$$

$$\Longrightarrow \begin{cases} c - \lambda_1 + A^T\nu = 0_d \\ -b + A\lambda_2 + 0_m = 0_m\end{cases}$$

We notice that in the constraints that $\lambda_1 \geq 0$ only plays the role of a deviation variable thus we can replace $\begin{cases}A^T\nu + \lambda_1 = c \\ \begin{pmatrix}\lambda_1 \\ \lambda_2\end{pmatrix} \geq 0\end{cases}$ by $\begin{cases}A^T\nu \leq c \\ \lambda_2 \geq 0\end{cases}$

Finally, we obtain:

$$\boxed{(\text{Self-Dual}_{\text{Dual}})\begin{cases}-\min & c^T\lambda_2 - b^T\nu \\ \text{s.t} & A^T\nu \leq c \\ & A\lambda_2 = b \\ & \lambda_2 \geq 0 \\ & \nu \in \mathbb{R}^m\end{cases}}$$

Ex 1)

4.) (Self-Dual) has an optimal sol° so his dual has as well (it's the same pb) and they have the same optimal value i.e. $p^* = d^* \Rightarrow$ we have strong duality

In particular, we have:

Primal feasibility: $Ax^* = b$, $x^* \geq 0$
$A^T y^* \leq c$

Complementarity slackness: $\lambda_i^* (B_2 z^* - \gamma)_i = 0$
$\forall \ 1 \leq i \leq 2d$

$$\Longrightarrow \forall d+1 \leq i \leq 2d \quad \lambda_i^* (B_2 z^* - \gamma)_i = 0$$

$$\Longrightarrow \forall \ 1 \leq i \leq d \quad x_i^* (A^T y^* - c)_i = 0$$

Primal feasibility implies that x^* and y^* are feasible for respectively (P) and (D). However,

$$\min_{x,y} \underbrace{c^T x}_{\text{doesn't depend on } y} - \underbrace{b^T y}_{\text{doesn't depend on } x} = \min_x c^T x - \max_y b^T y$$

Thus, $\boxed{\begin{cases} x^* \text{ is an optimal sol° of (P)} \\ y^* \text{ is an optimal sol° of (D)} \end{cases}}$

Moreover, using always primal feasibility, we can write

$$p^* = c^T x^* - b^T y^*$$
$$= c^T x^* - (Ax^*)^T y^*$$
$$= c^T x^* - \underbrace{x^{*T} A^T y^*}_{\in \mathbb{R} \text{ as we can transpose it}}$$
$$= c^T x^* - y^{*T} A^T x^*$$
$$= (c^T - y^{*T} A^T) x^* \in \mathbb{R}$$
$$= x^{*T} (c - A^T y^*) \underset{\uparrow}{=} 0$$

Complementarity slackness

Thus, $\boxed{p^* = 0}$

Ex 2)

1.) Let $u \in \mathbb{R}^d$

$$\|\cdot\|_1^*(u) = \sup_{x \in \mathbb{R}^d} \{ u^T x - \|x\|_1 \}$$

$\underline{\|u\|_\infty > 1}$: $\exists \, i_0 ; \ |u_{i_0}| = \|u\|_\infty > 1$

WLOG, assume $u_{i_0} > 0 \Rightarrow u_{i_0} > 1$

We define $(x^{(n)})$ a seq in $\mathbb{R}^d$ s.t.

$\forall n \geq 1, \ x^{(n)} = (0 \dots 0 \ n \ 0 \dots 0)$ (n is the i_0th coefficient)

$$\forall n \geq 1, \ \|\cdot\|_1^*(u) \geq u^T x^{(n)} - \|x^{(n)}\|_1$$
$$= n \underbrace{(u_{i_0} - 1)}_{>0} \xrightarrow[n \to +\infty]{} +\infty$$

$$\Longrightarrow \|\cdot\|_1^*(u) = +\infty$$

$\underline{\|u\|_\infty \leq 1}$

Let $x \in \mathbb{R}^d$

$$u^T x - \|x\|_1 = \sum_{i=1}^d [u_i x_i - |x_i|]$$
$$\leq \sum_{i=1}^d [|u_i x_i| - |x_i|]$$
$$\leq \underbrace{(\|u\|_\infty - 1)}_{\leq 0} \underbrace{\sum_{i=1}^d |x_i|}_{\geq 0} \leq 0$$

$$\xrightarrow{\sup_{x \in \mathbb{R}^d}} \|\cdot\|_1^*(u) \leq 0$$

For $x = 0_d$, $u^T x - \|x\|_1 = 0$

$$\Longrightarrow \|\cdot\|_1^*(u) \geq 0$$

Thus, $\|\cdot\|_1^*(u) = 0$

Conclusion:

$$\boxed{\|\cdot\|_1^*(u) = \begin{cases} 0 & \text{if } \|u\|_\infty \leq 1 \\ +\infty & \text{if } \|u\|_\infty > 1 \end{cases}}$$

Scanné avec CamScanner

(Q.1) $\min_x \|Ax-b\|_2^2 + \|x\|_1$

$$\equiv \min_{\substack{x \\ Ax-b=y}} \|y\|_2^2 + \|x\|_1$$

Let us compute the dual of that latter form.

$$L(x,y,\lambda^{\in\mathbb{R}^m}) = \|y\|_2^2 + \|x\|_1 + \lambda^T(y+b-Ax)$$

$$\inf_{x,y} L(x,y,\lambda) = \inf_{x,y} \|y\|_2^2 + \lambda^T y - [\lambda^T A x - \|x\|_1] + \lambda^T b$$

$$\Rightarrow g(\lambda) = \inf_y \underbrace{[\|y\|_2^2 + \lambda^T y]} - \sup_x [\lambda^T A x - \|x\|_1] + \lambda^T b$$

$\nabla(\ldots) = 0 \Rightarrow 2y + \lambda = 0$

$\Rightarrow y = -\frac{\lambda}{2}$

Since $y : y \mapsto \|y\|_2^2 + \lambda^T y$ is strictly convex $\Rightarrow$ it has a ! minimum over $\mathbb{R}^m$

$$= \frac{\|\lambda\|_2^2}{4} + \frac{\lambda^T(-\lambda)}{2} - \|.\|_1^*(A^T\lambda) + \lambda^T b$$

$$= \begin{cases} -\frac{\|\lambda\|_2^2}{4} + \lambda^T b & \text{if } \|A^T\lambda\|_\infty \leq 1 \\ -\infty & \text{if } \|A^T\lambda\|_\infty > 1 \end{cases}$$

Thus, the dual of RLS is:

$$\begin{cases} \max & -\frac{\|\lambda\|_2^2}{4} + \lambda^T b \\ \text{s.t.} & \|A^T\lambda\|_\infty \leq 1 \\ & \lambda \in \mathbb{R}^m \end{cases}$$

(Ex 3)

(1.) (Sep1): $\min_w \frac{1}{m}\sum_{i=1}^m \max(0, 1-y_i(w^T x_i)) + \frac{\tau}{2}\|w\|_2^2$

$$\equiv \tau \min_w \frac{1}{m\tau}\sum_{i=1}^m \max(0, 1-y_i(w^T x_i)) + \frac{1}{2}\|w\|_2^2$$

The idea is to characterize $\max(a,b)$ as follows:

$$\max(a,b) = \min_{\substack{\alpha \geq a \\ \alpha \geq b}} \{\alpha\}$$

$$\Rightarrow (\text{Sep1}) \equiv \tau \min_w \frac{1}{m\tau}\sum_{i=1}^m \min_{\substack{z_i \geq 1-y_i(w^T x_i) \\ z_i \geq 0}} z_i + \frac{1}{2}\|w\|_2^2$$

indep of z_i — indep of z_i

$$\equiv \tau \min_{\substack{w \\ z_i \geq 1-y_i(w^T x_i) \\ z_i \geq 0}} \frac{1}{m\tau}\sum_{i=1}^m z_i + \frac{1}{2}\|w\|_2^2$$

$$\equiv \tau \underbrace{\min_{\substack{w \\ z_i \geq 1-y_i(w^T x_i) \\ z_i \geq 0}} \frac{1}{m\tau} 1^T z + \frac{1}{2}\|w\|_2^2}_{(\text{Sep2})}$$

Thus, $\boxed{(\text{Sep2}) \text{ solves } (\text{Sep1})}$

(2.) $L(w, z, \lambda^{\in\mathbb{R}^m}, \pi^{\in\mathbb{R}^m})$

$$= \frac{1}{m\tau} 1^T z + \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^m \lambda_i(1 - y_i w^T x_i - z_i) - \pi^T z$$

$$= \sum_{i=1}^m \left(\frac{1}{m\tau} - \lambda_i - \pi_i\right) z_i + \underbrace{\frac{1}{2}\|w\|_2^2 - w^T \sum_{i=1}^m \lambda_i y_i x_i} + \sum_{i=1}^m \lambda_i$$

$$g(\lambda, \pi) = \inf_{w,z} L$$

$\nabla(\ldots) = 0 \Rightarrow w - \sum_{i=1}^m \lambda_i y_i x_i = 0$

Since it's a quadratic fct it's minimum is global and unique

(Ex 3) (Follow-up)

2.) So,

$$g(\lambda, \pi) = \inf_{\omega, 3} L(\omega, 3, \lambda, \pi)$$

$$= \begin{cases} \frac{1}{2} \|\sum_{i=1}^{m} \lambda_i y_i x_i\|_2^2 - \|\sum_{i=1}^{m} \lambda_i y_i x_i\|_2^2 + \sum_{i=1}^{m} \lambda_i & \text{if } \lambda_i + \pi_i = \frac{1}{m2} \\ -\infty & \text{otherwise} \end{cases}$$

Then, the dual prb is :

$$\max_{\substack{\lambda \geq 0 \\ \pi \geq 0 \\ \lambda + \underbrace{\pi}_{} = \frac{1}{m2}\mathbb{1}}} -\frac{1}{2} \|\sum_{i=1}^{m} \lambda_i y_i x_i\|_2^2 + \sum_{i=1}^{m} \lambda_i$$

$\hookrightarrow$ not in the objective fct and $\pi \geq 0$

$\Rightarrow$ can be seen as a deviation variable

$$\equiv \boxed{\max_{\substack{\lambda \geq 0 \\ \lambda \leq \frac{1}{m2}\mathbb{1}}} -\frac{1}{2} \|\sum_{i=1}^{m} \lambda_i y_i x_i\|_2^2 + \sum_{i=1}^{m} \lambda_i}$$

(Ex 5)

(1.) $$\min_{\substack{Ax \leq b \\ x_i(1-x_i) = 0 \quad \forall i = 1, \dots, n}} c^T x$$

$$L(x, \overset{\in \mathbb{R}^m}{\lambda}, \overset{\in \mathbb{R}^n}{\partial}) = c^T x + \lambda^T (Ax - b) + \sum_{i=1}^{n} \partial_i x_i (1 - x_i)$$

$$= (c^T + \lambda^T A - \partial^T) x - \sum_{i=1}^{n} \partial_i x_i^2 - \lambda^T b$$

$D = \text{diag}(\partial) = \begin{pmatrix} \partial_1 & (0) \\ (0) & \partial_n \end{pmatrix}$

$$\overset{J}{=} (c + A^T \lambda - \partial)^T x - \langle Dx, x \rangle - b^T \lambda$$

$$\nabla_x L = c + A^T \lambda - \partial - 2Dx = 0$$

$$\Longrightarrow x = \frac{1}{2} D^{-1} (c + A^T \lambda - \partial)$$

where $D^{-1} = \begin{pmatrix} 1/\partial_1 & (0) \\ (0) & \frac{1}{\partial_n} \end{pmatrix}$

$$\Rightarrow g(\lambda, \partial) = \min_x L(x, \lambda, \partial)$$

$$= \frac{1}{2} \langle D^{-1}(c + A^T\lambda - \partial), c + A^T\lambda - \partial \rangle - \langle c + A^T\lambda - \partial, D^{-1}(c + A^T\lambda - \partial) \rangle - b^T \lambda$$

$$= -\frac{1}{2} \langle D^{-1}(c + A^T\lambda - \partial), c + A^T\lambda - \partial \rangle - b^T \lambda$$

The dual pb is:

$$\max_{\substack{\lambda \geq 0 \\ \partial \in \mathbb{R}^n}} -\frac{1}{2} \langle D^{-1}(c + A^T\lambda - \partial), c + A^T\lambda - \partial \rangle - b^T \lambda$$

$$\langle D^{-1}(c + A^T\lambda - \partial), c + A^T\lambda - \partial \rangle = \frac{(c_i + A_i^T \lambda - \partial_i)^2}{\partial_i}$$