Homework 2 I ① The dual of (P) is given by:

$$\begin{cases} \max_{\lambda,\nu} \inf_x c^Tx - \lambda^T x + \nu^T(Ax-b) \\ \text{s.t. } \lambda \geq 0 \end{cases} = \begin{cases} \max_{\lambda,\nu} -\nu^T b \\ \text{subject to} \begin{cases} \lambda = c + A^T\nu \\ \lambda \geq 0 \end{cases} \end{cases} = \begin{cases} \max -\nu^T b \\ \text{s.t. } A^T\nu + c \geq 0 \end{cases}$$

② (D) is equivalent to $\begin{cases} \min_y -b^Ty \\ \text{s.t. } A^Ty - c \leq 0 \end{cases}$. Its dual is:

$$\begin{cases} \max_\lambda g(\lambda) \\ \text{s.t } \lambda \geq 0 \end{cases} \text{ with } g(\lambda) = \inf_y -b^Ty + \lambda^T(A^Ty - c) = \inf_y (A\lambda - b)^T y - \lambda^T c = \begin{cases} -\lambda^T c \text{ if } A\lambda = b \\ -\infty \text{ otherwise} \end{cases}$$

$$\Updownarrow$$

$$\begin{cases} \max -\lambda^T c \\ \text{s.t.} \begin{cases} \lambda \geq 0 \\ A\lambda = b \end{cases} \end{cases} \Leftrightarrow \begin{cases} \min c^T\lambda \\ \text{s.t.} \begin{cases} \lambda \geq 0 \\ A\lambda = b \end{cases} \end{cases} \Leftrightarrow (P).$$

③ We first compute the dual of (Self-Dual), which is given by:

$$\begin{cases} \max_{\lambda,\mu,\nu} g(\lambda,\nu,\mu) \\ \text{s.t. } \mu \geq 0, \nu \geq 0 \end{cases} \text{ with } g(\lambda,\mu,\nu) = \inf_{x,y} (c + A^T\lambda - \nu)^T x + (A\mu - b)^T y - \lambda^T b - \mu^T c$$

$$= \begin{cases} -\lambda^T b - \mu^T c \text{ if } A^T\lambda = \nu - c \text{ and } A\mu = b \\ -\infty \text{ otherwise} \end{cases}$$

$$\Updownarrow$$

$$\begin{cases} \max_{\lambda,\mu,\nu} -\lambda^T b - \mu^T c \\ \text{s.t. } A^T\lambda = \nu - c, A\mu = b, \mu \geq 0, \nu \geq 0 \end{cases} \Leftrightarrow \begin{cases} \min_{\lambda,\mu} \lambda^T b + \mu^T c \\ \text{s.t. } A\mu = b, \mu \geq 0, A^T\lambda + c \geq 0 \end{cases} =: (SD)'$$

Then we compute the dual of (SD)':

$$\begin{cases} \max_{z_1,z_2,z_3} g(z_1,z_2,z_3) \\ \text{s.t. } z_2 \geq 0, z_3 \geq 0 \end{cases} \text{ with } g(z_1,z_2,z_3) = \inf_{\lambda,\mu} \underbrace{\lambda^T b + \mu^T c + z_1^T(A\mu - b) - z_2^T\mu - z_3^T(A^T\lambda + c)}$$

$$= (b - Az_3)^T\lambda + (c + A^Tz_1 - z_2)^T\mu - z_1^T b - z_3^T c$$

$$= \begin{cases} -z_1^T b - z_3^T c \text{ if } Az_3 = b \text{ and } z_2 = A^Tz_1 + c \\ -\infty \text{ otherwise} \end{cases}$$

$$\Updownarrow$$

$$\begin{cases} \max_{z_1,z_2,z_3} -z_1^T b - z_3^T c \\ \text{s.t.} \begin{cases} Az_3 = b, z_2 = A^Tz_1 + c, \\ z_2 \geq 0, z_3 \geq 0 \end{cases} \end{cases} \Leftrightarrow \begin{cases} \max_{z_1,z_3} -z_1^T b - z_3^T c \\ \text{s.t. } Az_3 = b, z_3 \geq 0, A^Tz_1 + c \geq 0 \end{cases}$$

which is exactly (Self-Dual) (change of variable $z_1 \to -z_1$).
So (Self-Dual) is indeed self-dual.

④ These assumptions imply that (P) and (D) are bounded and feasible themselves.
Then:

$$\begin{cases} \min_{x,y} c^Tx - b^Ty \\ \text{s.t. } Ax = b, x \geq 0, A^Ty \leq c \end{cases} = \begin{cases} \min c^Tx \\ \text{s.t } Ax = b, x \geq 0 \end{cases} - \begin{cases} \max b^Ty \\ \text{s.t. } A^Ty \leq c \end{cases} = p^* - d^*.$$

The strong duality imply $p^* = d^*$ so the optimal value of (Self-Dual) is 0.

(II) ① We define $g(y) := \sup_{\|x\|_1 = 1} y^T x$

Let $y \in \mathbb{R}^d$. We have $y^T x = \|x\|_1 \, y^T\left(\frac{1}{\|x\|_1} x\right) \le \|x\|_1 \, g(y) \le \|x\|_1$ if $g(y) \le 1$

. So if $g(y) \le 1$ we have $y^T x - \|x\|_1 \le 0$ with equality if $x = 0$.

. If $g(y) > 1$, then there exists $x \in \mathbb{R}^d$ s.t. $\begin{cases} y^T x > 1 \\ \|x\|_1 = 1 \end{cases}$

so $y^T x > \|x\|_1$

so $y^T x - \|x\|_1 > 0$ and by taking tx with $t > 0$ we have $y^T(tx) - \|tx\|_1$
$= t(y^T x - \|x\|_1)$
$\xrightarrow[t \to +\infty]{} +\infty$.

. Finally, we have $g(y) = \|y\|_\infty$. Indeed; for $\|x\|_1 = 1$ we have:
$y^T x = \sum_{i=1}^{d} y_i x_i \le \sum_{i=1}^{d} |y_i| . |x_i| \le \underbrace{\sum_{i=1}^{d} |x_i|}_{=1} \|y\|_\infty$. And the supremum is attained for the point $(0, .., 0, \text{sign}(y_i), 0, .., 0)$ where $i := \underset{1 \le i \le d}{\text{argmax}} |y_i|$.

. So: $\|y\|_1^* = \begin{cases} 0 & \text{if } \|y\|_\infty \le 1 \\ +\infty & \text{otherwise} \end{cases}$

② (RLS) is equivalent to (RLS)': $\begin{cases} \min_{x,y} \|y\|_2^2 + \|x\|_1 \\ \text{s.t. } y = Ax - b \end{cases}$. We will compute the dual of (RLS)':

$$\max_\lambda \underbrace{\inf_{x,y} \left(\|y\|_2^2 + \|x\|_1 + \lambda^T(y - Ax + b) \right)}$$

$$= \underbrace{\inf_x \left(\|x\|_1 - \lambda^T A x \right)} + \underbrace{\inf_y \left(\|y\|_2^2 + \lambda^T y \right)} + \lambda^T b$$

$= -\sup_x \left((A^T\lambda)^T x - \|x\|_1 \right)$

$\overset{①}{=} -\begin{cases} 0 & \text{if } \|A^T\lambda\|_\infty \le 1 \\ -\infty & \text{otherwise} \end{cases}$

$= -\sup_y \left((-\lambda)^T y - \|y\|_2^2 \right)$

$= -2 \sup_y \left(\left(-\frac{\lambda}{2}\right)^T y - \frac{1}{2}\|y\|_2^2 \right)$

$\Rightarrow$ let's compute the conjugate of $x \overset{f}{\mapsto} \frac{1}{2}\|x\|_2^2$

Let $h(y) := \sup_{\|x\|_2 = 1} y^T x$. We have: $y^T x \le \|x\|_2 \, h(y)$

so: $y^T x - \frac{1}{2}\|x\|_2^2 \le \|x\|_2 \, h(y) - \frac{1}{2}\|x\|_2^2 \le \frac{h(y)^2}{2}$ attained for $\|x\|_2 = h(y)$

So: $f^*(y) \le \frac{h(y)^2}{2}$

Now let $(x_n) \in (\mathbb{R}^d)^{\mathbb{N}}$ s.t. $\forall n \in \mathbb{N}$, $\|x_n\|_2 = 1$ and $y^T x_n \xrightarrow[n \to +\infty]{} h(y)$. We have:
$y^T x_n - h(y) \|x_n\|_2 \xrightarrow[n \to +\infty]{} 0$ ($\|x_n\|_2 = 1$)

so $y^T(h(y) x_n) - \underbrace{h(y)^2 \|x_n\|_2}_{= \underbrace{\frac{1}{2} h(y)^2 \underbrace{\|x_n\|_2}_{=\|x_n\|_2^2}}_{= \|h(y) x_n\|_2^2} + \frac{1}{2} h(y)^2 \underbrace{\|x_n\|_2}_{=1}} \xrightarrow[n \to +\infty]{} 0$

so $y^T(h(y) x_n) - \frac{1}{2}\|h(y) x_n\|_2^2 \xrightarrow[n \to +\infty]{} \frac{1}{2} h(y)^2$

so $f^*(y) \ge \frac{1}{2} h(y)^2$

Finally: $f^*(y) = \frac{1}{2} h(y)^2$.

Moreover, it is easy to see that $h(y) = \|y\|_2$ (Cauchy-Schwarz).
So the dual of $(RLS)'$ is:

$$\begin{cases} \max_{\lambda} -2\times\frac{1}{2}\left\|\frac{\lambda}{2}\right\|_2^2 + \lambda^T b \\ \text{s.t. } \|A^T\lambda\|_\infty \leq 1 \end{cases} \iff \begin{cases} \max \lambda^T b - \frac{1}{4}\|\lambda\|_2^2 \\ \text{s.t. } \|A^T\lambda\|_\infty \leq 1 \end{cases}$$

III ① Instead of solving (Sep. 1), we can solve τ (Sep. 1), it will not change the optimal (w^*, z^*) (if it exists). Then τ (Sep. 1) is equivalent to:

$$\begin{cases} \min_{w,z} \frac{1}{n}\sum_{i=1}^{n} z_i + \frac{\tau}{2}\|w\|_2^2 \\ \text{s.t. } z_i = L(w, x_i, y_i) \quad (= \max(0, 1 - y_i(w^T x_i))) \end{cases} \quad (*).$$

Then problem $(*)$ solves problem (Sep. 2). Indeed, let (w,z) be an optimal solution of (Sep. 2). Since $(w,z) \overset{f_0}{\mapsto} \frac{1}{n\tau}\mathbb{1}^T z + \frac{1}{2}\|w\|_2^2$ is strictly increasing in the second variable (in the following sense: $\forall w,\ z_1 < z_2 \Rightarrow f_0(w,z_1) < f_0(w,z_2)$), we know that the optimal z satisfies: $\forall i,\ z_i = \max(0, 1 - y_i(w^T x_i))$.
(otherwise we could find i and $\varepsilon > 0$ such that $f_0\left(w, \begin{pmatrix} z_1 \\ \vdots \\ z_i - \varepsilon \\ \vdots \\ z_n \end{pmatrix}\right) < f_0(w,z)$ and $\left(w, \begin{pmatrix} z_1 \\ \vdots \\ z_i - \varepsilon \\ \vdots \\ z_n \end{pmatrix}\right)$ feasible for (Sep. 2): contradiction).

② The dual of (Sep. 2) is given by:

$$\begin{cases} \max_{\lambda,\pi} \inf_{w,z} \left(\frac{1}{n\tau}\mathbb{1}^T z + \frac{1}{2}\|w\|_2^2 + \underbrace{\sum_{i=1}^{n}\lambda_i(1 - y_i(w^T x_i) - z_i)}_{= \mathbb{1}^T\lambda - \lambda^T z - w^T\tilde{x} \text{ with } \tilde{x} = \sum_{i=1}^{n}\lambda_i y_i x_i} - \pi^T z \right) \\ \quad = \inf_z \left(\frac{1}{n\tau}\mathbb{1} - \lambda - \pi\right)^T z + \underbrace{\inf_w \left(\frac{1}{2}\|w\|_2^2 - w^T\tilde{x}\right)}_{= -\underbrace{\sup_w \left(w^T\tilde{x} - \frac{1}{2}\|w\|_2^2\right)}_{= \frac{1}{2}\|\tilde{x}\|_2^2 \text{ (see II②)}}} + \mathbb{1}^T\lambda \\ \quad = \begin{cases} \mathbb{1}^T\lambda - \frac{1}{2}\|\tilde{x}\|_2^2 & \text{if } \frac{1}{n\tau}\mathbb{1} = \lambda + \pi \\ -\infty & \text{otherwise} \end{cases} \\ \text{s.t. } \lambda \geq 0,\ \pi \geq 0 \end{cases}$$

$\Updownarrow$

$$\begin{cases} \max_{\lambda,\pi} \mathbb{1}^T\lambda - \frac{1}{2}\|\tilde{x}\|_2^2 \\ \text{s.t. } \lambda \geq 0,\ \pi \geq 0,\ \frac{1}{n\tau}\mathbb{1} - \lambda = \pi \end{cases} \iff \begin{cases} \max_{\lambda} \mathbb{1}^T\lambda - \frac{1}{2}\|\tilde{x}\|_2^2 \quad \text{with } \tilde{x} = \sum_{i=1}^{n}\lambda_i y_i x_i \\ \text{st } 0 \leq \lambda \leq \frac{1}{n\tau}\mathbb{1} \end{cases}$$