Exercise 1 (LP Duality) For a given $c \in \mathbb{R}^d$, $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times d}$ consider the two following linear optimization problems,

$$\min_x c^T x \quad \text{s.t. } Ax = b, \; x \geq 0 \qquad (P)$$

and

$$\max_y b^T y \quad \text{s.t. } A^T y \leq c \qquad (D)$$

1. Compute the dual of problem (P) and simplify it if possible.

At first, let's write standard form: $\min_x c^T x$, s.t. $Ax - b = 0$, $-x \leq 0$. The lagrangian function $\mathcal{L}(x, \lambda, \nu) = c^T x + \nu^T(Ax-b) - \lambda^T x = (\nu^T A + c^T - \lambda^T)x - \nu^T b = (A^T\nu + c - \lambda)^T x - b^T \nu$. The dual function: $g(\lambda, \nu) = \inf_{x \in D} \mathcal{L}(x, \lambda, \nu)$, where $D = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$. The lagrangian is linear by x, so

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

g is linear in (λ, ν) on the set $\{(\lambda, \nu) \mid A^T\nu - \lambda + c = 0\}$ - affine domain. Therefore, g is concave. We have a dual problem

$\max -b^T\nu$ s.t. $A^T\nu + c - \lambda = 0$, $\lambda \geq 0 \Rightarrow -b^T\nu$ s.t. $A^T\nu + c \geq 0 \Rightarrow$

$\Rightarrow \max -b^T \nu$ s.t. $-A^T\nu - c \leq 0$ — standard form $\Rightarrow \nu = -\nu$

$\Rightarrow \max b^T \nu$ s.t. $A^T\nu - c \leq 0$, if $\nu = y$, then we get (D).

2. Compute the dual of problem (D).

At first, let's write standard form: $\min_y -b^T y$, s.t. $A^T y - c$

The lagrangian: $\mathcal{L}(y,\lambda) = -b^Ty + \lambda^T(A^Ty - c) = -b^Ty + \lambda^TA^Ty -$
$-\lambda^Tc = (A\lambda - b)^Ty - c^T\lambda$. The dual function: $g(\lambda) = \inf_{y\in D} \mathcal{L}(y,\lambda)$,
where $D = \{y \in \mathbb{R}^n \mid A^Ty - c \le 0\}$. $\mathcal{L}$ is linear by y $\Rightarrow$
$\Rightarrow g(\lambda) = \begin{cases} -c^T\lambda & \text{if } A\lambda - b = 0 \\ -\infty & \text{otherwise} \end{cases}$ — linear on the $\{\lambda \mid A\lambda - b = 0\}$ —
affine $\Rightarrow$ we can write next transformation:
$\max_\lambda -c^T\lambda$ s.t. $A\lambda - b = 0, \lambda \ge 0 \iff \max_\lambda -c^T\lambda, A\lambda = b, \lambda \ge 0 \iff$
$\iff \min_\lambda c^T\lambda$ s.t. $A\lambda = b, \lambda \ge 0$ — (P).

3. A problem is called self-dual if its dual is the problem itself. Prove that the following problem is self-dual.

$$\begin{aligned} \min_{x,y} \ & c^Tx - b^Ty \\ \text{s.t.} \ & Ax = b \\ & x \ge 0 \\ & A^Ty \le c \end{aligned} \qquad \text{(Self-Dual)}$$

The lagrangian: $\mathcal{L}(x,\lambda,\vartheta) = c^Tx - b^Ty - \lambda_1^Tx + \lambda_2^T(A^Ty - c) +$
$+\vartheta^T(Ax - b) = (A^T\vartheta + c - \lambda_1)^Tx - \lambda_2^Tc + (A\lambda_2 - b)^Ty - b^T\vartheta$ — linear in (x,y). The dual function $g(\lambda,\vartheta) = \inf_D \mathcal{L}(x,\lambda,\vartheta) =$
$= \begin{cases} -\vartheta^Tb - \lambda_2^Tc & \text{if } A^T\vartheta + c - \lambda_1 = 0 \text{ and } A\lambda_2 - b = 0 \\ -\infty & \text{, otherwise} \end{cases}$. Where
$D = \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^n \mid Ax = b, x \ge 0, A^Ty \le c\}$. g is linear $\Rightarrow$
$\Rightarrow g$ is concave. Hence, next transformation can be performed:
$\max_{\vartheta,\lambda_2} -b^T\vartheta - c^T\lambda_2$, $A^T\vartheta + c - \lambda_1 = 0$, $A\lambda_2 - b = 0$, $\lambda_1 \ge 0$, $\lambda_2 \ge 0 \iff$
$\iff \max_{\vartheta,\lambda_2} -b^T\vartheta - c^T\lambda_2$, $A^T\vartheta + c \ge 0$, $A\lambda_2 = b, \ge 0 \iff \max_{\vartheta,\lambda_2} -b^T\vartheta - c^T\lambda_2$,

$-A^T\nu \le c,\ A\lambda_2 = b,\ \lambda_2 \ge 0 \Leftrightarrow \nu = -\nu \Leftrightarrow \max\limits_{\nu,\lambda_2} b^T\nu - c^T\lambda_2,\ A^T\nu \le c,$
$A\lambda_2 = b,\ \lambda_2 \ge 0 \Rightarrow \min\limits_{\lambda_2 \ge 0} c^T\lambda_2 - b^T\nu,\ A\lambda_2 = b,\ A^T\nu \le c,\ \lambda_2 \ge 0 \Rightarrow$
$\Leftrightarrow \min\limits_{x,y} c^Tx - b^Ty,\ Ax = b,\ x \ge 0,\ A^Ty \le c$ - (Self-Dual) $\Rightarrow$
$\Rightarrow$ The problem is self-dual.

4. Assume the above problem feasible and bounded, and let $[x^*; y^*]$ be its optimal solution. Using the strong duality property of linear programs, show that
- the vector $[x^*; y^*]$ can also be obtained by solving (P) and (D)
- the optimal value of (Self-Dual) is exactly 0.

The problem is easily separably by two variables: x and y, because constraints are disjoint, so we can divide problems into two. $\Rightarrow \min\limits_{x} c^Tx$, s.t. $Ax = b, x \ge 0$ - which is (P) and $\min\limits_{y} -b^Ty$ s.t. $A^Ty \le c \Leftrightarrow \max\limits_{y} b^Ty$ s.t. $A^Ty \le c$ - which is (D), so is we solve (P) and (D) we will get x^*, y^* which are solution $[x^*, y^*]$ for (Self-Dual). ~~Beca~~ We can use strong duality property, because (P) and (D) are dual, linear and convex, so their optimal values of the their objective functions are equal, so we have $P^* = c^Tx^* = D^* = b^Ty^* \Rightarrow c^Tx^* - b^Ty^* = SF^* = 0$. So, the optimal value of (Self-Dual) is exacty 0.

Exercise 2 (Regularized Least-Square)

For given $A\in\mathbb{R}^{n\times d}$ and $b\in\mathbb{R}^n$, consider the following optimization problem,

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

1. Compute the conjugate of $\|x\|_1$.

$f(x)=\|x\|_1$, $f^*(y)=\sup_x(y^Tx-\|x\|_1)=\sup_x\left(\sum_{i=1}^n y_ix_i-\sum_{i=1}^n|x_i|\right)\le$

$\le \sup_x\left(\left(\sum_{i=1}^n x_i\right)\max_i|y_i|-\sum_{i=1}^n|x_i|\right)$. Let's $\max_i|y_i|=M$.

if $M\le1$, then $f^*(y)\le\sup_x\left(\sum_{i=1}^n x_i-\sum_{i=1}^n|x_i|\right)$. $\sum_{i=1}^n x_i\le\sum_{i=1}^n|x_i|\Rightarrow$

$\Rightarrow \sum_{i=1}^n x_i-\sum_{i=1}^n|x_i|\le 0 \Rightarrow \sup_x\left(\sum_{i=1}^n x_i-\sum_{i=1}^n|x_i|\right)=0$ when $\forall i=\overline{1,n}$:

$|x_i|=x_i$, $x\ge0$, so $f^*(y)=0$.

If $M>1\Rightarrow \exists i: |y_i|>1$. If $y_i>1$. We can choose x like $x_i=a>0$, $x_j=0$, when $j\ne i$. $\Rightarrow$

$\Rightarrow f^*(y)=\sup_x\left(\sum_{i=1}^n y_ix_i - a\right)=\sup_a(y_ia-a)=\sup_a a(y_i-1)$, where $a>0$ and $y_i>0 \Rightarrow$ if $a\to+\infty$, then $f^*(y)=+\infty$.

if $y_i<-1$, then

$x_i=a<0$, $x_j=0\ \forall j\ne i$. $f^*(y)=\sup_x\left(\sum_{i=1}^n x_iy_i - a\right)=$

$=\sup_a(ay_i-|a|)=\sup_a a(y_i+1)$, $a<0$ $y_i+1<0\Rightarrow$

$\Rightarrow$ if $t\to-\infty\Rightarrow f^*(y)=\infty$. We can ensemble the conjugate function $f^*(y)=\begin{cases}0, & \text{if } \|y\|_\infty\le1\\ +\infty, & \text{otherwise}\end{cases}$

2. Compute the dual of (RLS).

Let's transform (RLS) to standard form. $\min_x\|Ax-b\|_2^2+\|x\|_1 \Leftrightarrow \min_{x,y}\|y\|_2^2+\|x\|_1$ s.t. $Ax-b-y=0 \Leftrightarrow$

$\Leftrightarrow \min_{x,y} y^T y + \|x\|_1$. The lagrangian $\mathcal{L}(x,y,\vartheta) = y^T y + \|x\|_1 + \vartheta^T(Ax - b - y)$. The dual function $g(\vartheta) = \inf_{x,y \in D} \mathcal{L}(x,y,\vartheta)$, where $D = \{(x,y) \mid Ax - b - y = 0\}$. $g(\vartheta) = \inf_{x,y}(y^T y + \|x\|_1 + \vartheta^T Ax - \vartheta^T b - \vartheta^T y)$.

This expression can be divided into parts of x and y.

$g(\vartheta) = \inf_x(\|x\|_1 + \vartheta^T Ax) + \inf_y(y^T y - \vartheta^T y) - \vartheta^T b$

Let's find each inf separately. $\inf_x(\|x\|_1 + \vartheta^T Ax) = -\sup_x(-\|x\|_1 - \vartheta^T Ax)$. The inner part looks like conjugate of $\|x\|_1$, if $a = -A^T\vartheta$, then $a^T = -\vartheta^T A$, so $-\sup_x(a^T x - \|x\|_1) = \begin{cases} 0, & \text{if } \|A^T\vartheta\|_\infty \le 1 \\ -\infty, & \text{otherwise} \end{cases}$

$\inf_y(y^T y - \vartheta^T y)$. We have $y^T y - \vartheta^T y \Rightarrow$ we can get $\frac{\partial}{\partial y}(y^T y - \vartheta^T y) = 2y - \vartheta = 0 \Rightarrow y = \frac{1}{2}\vartheta \Rightarrow \inf_y(y^T y - \vartheta^T y) = \frac{1}{4}\vartheta^T\vartheta - \frac{1}{2}\vartheta^T\vartheta = -\frac{1}{4}\vartheta^T\vartheta$. If we join the expressions, we will get dual function $g(\vartheta) = \begin{cases} -\frac{1}{4}\vartheta^T\vartheta - \vartheta^T b, & \text{if } \|A^T\vartheta\| \le 1 \\ -\infty & \text{otherwise} \end{cases}$

g is concave on the set $\{\vartheta \mid \|A^T\vartheta\| \le 1\}$, because g is quadratic with $-$ before $\vartheta^T\vartheta$, and the set is convex. So, the dual of (RLS) $\max_{\vartheta} -\frac{1}{4}\vartheta^T\vartheta - \vartheta^T b$, $\|A^T\vartheta\|_\infty \le 1$.

Exercise 3. (Data Separation)
Assume we have n data points $x_i \in \mathbb{R}^d$, with label $y_i \in \{-1, 1\}$.
We are searching for an hyper-plane defined by its normal w, which separates the points according to their label. Ideally, we would like to have

$$w^T x_i \leq -1 \Rightarrow y_i = -1 \text{ and } w^T x_i \geq 1 \Rightarrow y_i = 1$$

Unfortunately, this condition is rarely met with real-life problems. Instead, we solve an optimization problem which minimizes the gap between the hyperplane and the miss-classified points. To do so, we will use a specific loss function

$$\mathcal{L}(w, x_i; y_i) = \max\{0, 1 - y_i(w^T x_i)$$

which is equal to 0 when the point x_i is well-classified (the sign of $w^T x_i$ and y_i is the same), but is strictly positive when the sign of $w^T x_i$ and y_i is different.
To improve the permormances, instead of minimizing the loss function alone, we also use the quadratic regulizer as follow

$$\min_w \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(w, x_i; y_i) + \frac{\tau}{2} \|w\|_2^2 \qquad \text{(Sep. 1)}$$

where τ is the regularization parameter.
1. Consider the following quadratic optimization problem ($\mathbf{1}$ is a vector full of ones),

$$\min_{w,z} \frac{1}{n\tau} \mathbf{1}^T z + \frac{1}{2}\|w\|_2^2$$

s.t. $z_i \geq 1 - y_i(w^T x_i)\ \forall i = 1, \dots, n$ $\quad(\lambda_i)$ (Sep. 2)

$z \geq 0$ $\quad(\pi)$

Explain why problem (Sep.2) solves problem (Sep. 1).

At first, in (Sep.1) we divide the whole expression by $\tau \neq 0$. We get $\min_{w} \left(\frac{1}{n\tau} \sum_{i=1}^{n} \mathcal{L}(w, x_i; y_i) + \frac{1}{2}\|w\|_2^2\right) = \min_{w} \frac{1}{n\tau} \sum_{i=1}^{n} \max\{0, 1 - y_i(w^T x_i)\} + \frac{1}{2}\|w\|_2^2$.

Let's work with first part of expression, it can be rewritten as $\min_{z_i} z_i$, s.t. $0 \leq z_i$, $1 - y_i(w^T x_i) \leq z_i$, where $i = 1, \dots, n$. Therefore, we have transformed optimization problem: $\min_{w,z} \frac{1}{n\tau} \sum_{i=1}^{n} z_i + \frac{1}{2}\|w\|_2^2$, $z_i \geq 0$, $z_i \geq 1 - y_i(w^T x_i)$, $i = 1, \dots, n$. The $\sum_{i=1}^{n} z_i = \mathbf{1}^T z$, so we have $\min_{w,z} \frac{1}{n\tau} \mathbf{1}^T z + \frac{1}{2}\|w\|_2^2$ s.t. $z_i \geq 1 - y_i(w^T x_i)$, $z \geq 0$ $\forall i = 1, \dots, n$. So, problems are equivalent and (Sept. 2) solves (Sept. 1).

2. Compute the dual of (Sep. 2), and try to reduce the number of variables. Use the notations λ_i and π for the dual variables.

The lagrangian $\mathcal{L}(w, z, \lambda, \pi) = \frac{1}{n\tau} \mathbf{1}^T z + \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^{n} \lambda_i (1 - y_i(w^T x_i) - z_i) - \pi^T z = \frac{1}{n\tau} \mathbf{1}^T z + \frac{1}{2} w^T w + \mathbf{1}^T \lambda - \left(\sum_{i=1}^{n} \lambda_i y_i x_i^T\right) w - \lambda^T z - \pi^T z = \left(\frac{1}{n\tau} \mathbf{1} - \lambda - \pi\right)^T z + \frac{1}{2} w^T w - \left(\sum_{i=1}^{n} \lambda_i y_i x_i^T\right) w +$

$+ 1_n^T\lambda$. The dual function: $g(\lambda, \pi) = \inf_{(y,z)\in D} (\frac{1}{n\tau}1 - \lambda - \pi)^T z + \frac{1}{2}w^Tw -$
$- (\sum_{i=1}^n \lambda_i y_i x_i^T) w + 1^T\lambda$. The dual function is clearly separated by two ~~variables~~ variables w and z: $g(\lambda,\pi) = \inf_z \underbrace{(\frac{1}{n\tau}1 - \lambda - \pi)}_{(1)}^T z +$
$+ \inf_w (\frac{1}{2}w^Tw - (\sum_{i=1}^n \lambda_i y_i x_i^T) w) + 1^T\lambda$. if $(1) \neq 0 \Rightarrow$
$\Rightarrow g$ is linear in z, so we can make this inf equal to $-\infty$. if $(1) = 0$, then we work with second part of the expression. $\inf_w (\frac{1}{2}w^Tw - (\sum_{i=1}^n \lambda_i y_i x_i^T) w)$ ~~...~~, in order to find the inf we find where $\frac{\partial f(w)}{\partial w} = 0$. $\frac{\partial f(w)}{\partial w} =$
$= w - \sum_{i=1}^n \lambda_i y_i x_i^T = 0 \Rightarrow w = \sum_{i=1}^n \lambda_i y_i x_i^T \Rightarrow \inf_w (\frac{1}{2}w^Tw -$
$- (\sum_{i=1}^n \lambda_i y_i x_i) w) = -\frac{1}{2} \|\sum_{i=1}^n \lambda_i y_i x_i^T\|_2^2$. Thus, taking together all parts, we get

$$g(\lambda, \pi) = \begin{cases} -\frac{1}{2} \|\sum_{i=1}^n \lambda_i y_i x_i^T\|_2^2 + 1^T\lambda & \text{if } \frac{1}{n\tau}1 - \lambda - \pi = 0 \\ -\infty & \text{otherwise} \end{cases}$$

g is concave, because it is quadratic in λ with neg. coef., $\{(\pi,\lambda) \mid \frac{1}{n\tau}1 - \pi - \lambda = 0\}$ - affine domain in (π, λ). We get the dual problem

$$\max_\lambda -\frac{1}{2}\|\sum_{i=1}^n \lambda_i y_i x_i\|_2^2 + 1^T\lambda, \text{ s.t } \frac{1}{n\tau} - \lambda - \pi = 0, \lambda \geq 0, \pi \geq 0 \Leftrightarrow$$

$$\max_\lambda -\frac{1}{2}\|\sum_{i=1}^n \lambda_i y_i x_i\|_2^2 + 1^T\lambda, \text{ s.t } \frac{1}{n\tau} - \lambda \geq 0, \lambda \geq 0.$$