Convex Optimization
Homework 2

Realized by: Eya Ghamgui

Exercice n 1: LP Duality:

for $c \in \mathbb{R}^d$, $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times d}$, we have:

$$(P) \begin{cases} \text{Min } c^T x \\ \text{st} \\ Ax = b,\ x \geq 0 \end{cases} \qquad (D) \begin{cases} \text{Max } b^T y \\ \text{st} \\ A^T y \preccurlyeq c \end{cases}$$

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

* standard form of (P): $\begin{cases} \text{Min } c^T x \\ \text{st} \\ Ax - b = 0 \ ;\ -x \preccurlyeq 0 \end{cases}$

* The lagrangian function:

$$\begin{aligned} L(x, \lambda, \upsilon) &= c^T x + \lambda^T(-x) + \upsilon^T(Ax - b) \\ &= c^T x - \lambda^T x + \upsilon^T(Ax - b) \\ &= (c + A^T\upsilon - \lambda)^T x - \upsilon^T b \\ &= (c + A^T\upsilon - \lambda)^T x - b^T \upsilon \end{aligned}$$

* Let suppose $D = \{ x \in \mathbb{R}^n \mid Ax = b \ ;\ x \geq 0 \}$

The dual lagrangian function:

$$\begin{aligned} g(\lambda, \upsilon) &= \inf_{x \in D} L(x, \lambda, \upsilon) \\ &= \inf_{x \in D} \{ (c + A^T\upsilon - \lambda)^T x - b^T \upsilon \} \end{aligned}$$

We have $L(x, \lambda, \upsilon)$ is linear in x, Thus,

$$\inf_{x \in D} L(x, \lambda, \upsilon) = \begin{cases} -b^T\upsilon & \text{if } A^T\upsilon + c - \lambda = 0 \\ -\infty & \text{otherwise} \end{cases}$$

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

* We have:

- $g(\lambda, \upsilon)$ is linear in (λ, υ)
- $\{ (\lambda, \upsilon) \mid A^T\upsilon - \lambda + c = 0 \}$ is an affine domain

$\Rightarrow$ g is a concave function.

$\Rightarrow$ The dual problem of (P):

(change υ with $-\upsilon$)

$$\begin{cases} \text{Max } -b^T\upsilon \\ \text{st} \\ \begin{cases} A^T\upsilon - \lambda + c = 0 \\ \lambda \geq 0 \end{cases} \end{cases} \Leftrightarrow \begin{cases} \text{Max } -b^T\upsilon \\ \text{st} \\ A^T\upsilon + c \geq 0 \end{cases} \Leftrightarrow \begin{cases} \text{Max } b^T\upsilon \ : (D) \\ \text{st} \\ A^T\upsilon \preccurlyeq c \end{cases}$$

2- Compute the dual of problem (D):

* The standard form of (D): $\begin{cases} \text{Max } b^T y \\ \text{st} \\ A^T y \leq c \end{cases} \Leftrightarrow \begin{cases} \underset{y}{\text{Min}} \ -b^T y \\ \text{st} \\ A^T y - c \leq 0 \end{cases}$

* The lagrangian function:

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

* Let suppose $D = \{ y \in \mathbb{R}^n \mid A^T y - c \leq 0 \}$

The dual lagrangian function:

$$g(\lambda) = \inf_{y \in D} \{ L(y,\lambda) \} = \inf_{y \in D} \{ (A\lambda - b)^T y - c^T \lambda \}$$

We have $L(y,\lambda)$ is linear in $y \Rightarrow$

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

* We have: - g is a linear function in function of λ.
  - $\{ \lambda \mid A\lambda - b = 0 \}$ is an affine domain

  $\Rightarrow g$ is a concave function.

* The dual problem $\Leftrightarrow \begin{cases} \text{Max } -c^T \lambda \\ \text{st} \\ \begin{cases} A\lambda - b = 0 \\ \lambda \geq 0 \end{cases} \end{cases} \Leftrightarrow \begin{cases} \text{Max } -c^T \lambda \\ \text{st} \\ \begin{cases} A\lambda = b \\ \lambda \geq 0 \end{cases} \end{cases}$

$\Leftrightarrow \begin{cases} \text{Min } c^T \lambda \\ \text{st} \\ \begin{cases} A\lambda = b \\ \lambda \geq 0 \end{cases} \end{cases} = (P)$

3- Prove that the following problem is self-dual:

(self dual) $\begin{cases} \underset{x,y}{\text{Min}} \ c^T x - b^T y \\ \text{st} \begin{cases} Ax = b \\ x \geq 0 \\ A^T y \leq c \end{cases} \end{cases}$

* The standard form of the problem:

$\begin{cases} \underset{x,y}{\text{Min}} \ c^T x - b^T y \\ \text{st} \\ \begin{cases} Ax - b = 0 \\ -x \leq 0 \\ A^T y - c \leq 0 \end{cases} \end{cases}$

* The lagrangian function:

$$L(x, y, \lambda_1, \lambda_2, \nu) = c^T x - b^T y - \lambda_1^T x + \lambda_2^T (A^T y - c) + \nu^T (Ax - b)$$
$$= (c - \lambda_1 + A^T \nu)^T x + (A\lambda_2 - b)^T y - \lambda_2^T c - \nu^T b$$
$$= (c - \lambda_1 + A^T \nu)^T x + (A\lambda_2 - b)^T y - (\lambda_2^T c + \nu^T b)$$

* $D = \{(x, y) \mid Ax = b \,;\, x \geq 0 \,;\, A^T y \leq c\}$
* The dual function

$$g(\lambda_1, \lambda_2, \nu) = \inf_{(x,y) \in D} L(x, y, \lambda_1, \lambda_2, \nu)$$
$$= \inf_{x,y \in D} \left\{ (c - \lambda_1 + A^T \nu)^T x + (A\lambda_2 - b)^T y - (\lambda_2^T c + \nu^T b) \right\}$$
$$= \begin{cases} \inf_y (A\lambda_2 - b)^T y - (\lambda_2^T c + \nu^T b) & \text{if } c - \lambda_1 + A^T \nu = 0 \\ -\infty & \text{otherwise} \end{cases}$$
$$= \begin{cases} -(c^T \lambda_2 + b^T \nu) & \text{if } c - \lambda_1 + A^T \nu = 0 \text{ and } A\lambda_2 - b = 0 \\ -\infty & \text{otherwise} \end{cases}$$

* - g is a linear function in $(\lambda_1, \lambda_2, \nu)$ $\{-c^T\lambda_2 - b^T \nu$ is linear in λ_2 and $\nu\}$
  - $\{(\lambda_1, \lambda_2, \nu) \mid A\lambda_2 - b = 0 \,;\, c - \lambda_1 + A^T \nu = 0\}$ is a linear domain

  $\Rightarrow$ g is a concave function.

* The dual of the problem:

$$\begin{cases} \max_{\lambda_1, \lambda_2, \nu} \; -c^T \lambda_2 - b^T \nu \\ \text{st} \\ \begin{cases} c - \lambda_1 + A^T \nu = 0 \\ A\lambda_2 - b = 0 \\ \lambda_1 \geq 0 \;;\; \lambda_2 \geq 0 \end{cases} \end{cases} \Leftrightarrow \begin{cases} \max_{\lambda_1, \lambda_2, \nu} \; -c^T \lambda_2 - b^T \nu \\ \text{st} \\ \begin{cases} A\lambda_2 - b = 0 \\ A^T \nu + c = \lambda_1 \geq 0 \\ \lambda_2 \geq 0 \end{cases} \end{cases}$$

(objective function independent of λ_1)

$$\Leftrightarrow \begin{cases} \max_{\lambda_2, \nu} \; -c^T \lambda_2 - b^T \nu \\ \text{st} \\ \begin{cases} A\lambda_2 - b = 0 \\ A^T \nu + c \geq 0 \\ \lambda_2 \geq 0 \end{cases} \end{cases} \overset{(\nu = -V)}{\Leftrightarrow} \begin{cases} \max_{\lambda_2, V} \; -c^T \lambda_2 + b^T V \\ \text{st} \\ \begin{cases} A\lambda_2 - b = 0 \\ -A^T V + c \geq 0 \\ \lambda_2 \geq 0 \end{cases} \end{cases}$$

$$\Leftrightarrow \left\{ \begin{array}{l} \text{Max} \ -C^T d_2 + b^T \upsilon \\ \text{st} \\ \left\{ \begin{array}{l} A d_2 - b = 0 \\ A^T \upsilon \leq C \\ d_2 \geq 0 \end{array} \right. \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} \underset{d_2, \upsilon}{\text{Min}} \ C^T d_2 - b^T \upsilon \\ \text{st} \\ \left\{ \begin{array}{l} A d_2 - b = 0 \\ A^T \upsilon \leq C \\ d_2 \geq 0 \end{array} \right. \end{array} \right. = (\text{self-dual})$$

$\Rightarrow$ we can say that this problem is self dual because we found that its dual is the problem itself.

4 - Assume the above problem feasible and bounded, (x^*, y^*) the 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.

* we have that the (self-dual) problem is feasible and bounded and (x^*, y^*) its optimal solution.

We know that the constraint of (self-dual) problem are disjoint and it can be written as:

$$\{(x, y) \mid An = b; \ n \geq 0; A^T y \leq c\} = \{x \mid An = b; \ x \geq 0\} \cup \{y \mid A^T y \leq c\}$$

Thus, the (self-dual) problem can be decomposed into two problems:

$$\left\{ \begin{array}{l} \text{Min} \ C^T n - b^T y \\ \text{st} \\ \left\{ \begin{array}{l} An = b \\ n \geq 0 \\ A^T y \leq C \end{array} \right. \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} \underset{x}{\text{Min}} \ C^T x \\ \text{st} \\ \left\{ \begin{array}{l} An = b \\ n \geq 0 \end{array} \right. \end{array} \right. + \left\{ \begin{array}{l} \underset{y}{\text{Min}} \ -b^T y \\ \text{st} \\ A^T y \leq C \end{array} \right.$$

$$\Leftrightarrow \left\{ \begin{array}{l} \underset{x}{\text{Min}} \ C^T n \\ \text{st} \\ \left\{ \begin{array}{l} An = b \\ x \geq 0 \end{array} \right. \end{array} \right. + \left\{ \begin{array}{l} \underset{y}{\text{Max}} \ b^T y \\ \text{st} \\ A^T y \leq C \end{array} \right.$$

$$\Leftrightarrow \quad (P) \quad + \quad (D)$$

We have that (x^*, y^*) is a solution of the (self-dual) problem. Thus, we can say that x^* is an optimal solution for (P) and y^* is an optimal solution for (D).

* We have (P) and (D) are linear problems.

The objective function is a linear function.
The constraints are linear.

Therefore, (P) and (D) are convex problems.

We also know that the (self-dual) is a feasible and bounded problem. We can say that the problem (P) is then feasible and bounded. Then, it is a strictly feasible problem. We can say that the qualification constraints are verified for problem (P) which is a convex problem. So, the strong duality is verified for problem (P).
We have also the dual problem of (P) is (D) from the first question. Then, $p^* = d^* \iff C^T x^* = b^T y^*$

$$\iff C^T x^* - b^T y^* = 0$$

$\Rightarrow$ we found that $c^T x^* - b^T y^* = 0$ which is the optimal value for the (self-dual) problem.

$\Rightarrow$ In this question, we verified that the vector (x^*, y^*) can be also obtained by solving (P) and (D) and the optimal value of (self-dual) is exactly 0.

Exercice n 2: Regularized least-square:

We have : $\begin{cases} A \in \mathbb{R}^{n \times d} \\ b \in \mathbb{R}^{n} \end{cases}$

(RLS) : $\min_{x} \| Ax - b \|_2^2 + \| x \|_1$

1 - Compute the conjugate of $\| x \|_1$:

$\| x \|_1 = \sum_{i=1}^{d} |x_i|$

The conjugate function of $\| x \|_1$:

$$f^*(y) = \sup_{x} \{ y^T x - \| x \|_1 \}$$
$$= \sup_{x} \left\{ y^T x - \sum_{i=1}^{d} |x_i| \right\}$$
$$= \sup_{x} \left\{ \sum_{i=1}^{d} x_i y_i - \sum_{i=1}^{d} |x_i| \right\}$$

* if $|y_i| \leqslant 1$:

$$x_i y_i \leqslant |x_i y_i| \leqslant |x_i|$$
$$\Rightarrow \sum_{i=1}^{d} x_i y_i \leqslant \sum_{i=1}^{d} |x_i|$$
$$\Rightarrow y^T x - \| x \|_1 \leqslant 0$$
$$\Rightarrow \sup_{x} \{ y^T x - \| x \|_1 \} = 0$$

$\Rightarrow \forall i \quad |y_i| \leqslant 1 \quad \Rightarrow \max |y_i| \leqslant 1$

$\Rightarrow \| y \|_\infty \leqslant 1 \quad \Rightarrow f^*(y) = 0$

* if $y_i > 1$:

Let's take as an example:

$\begin{cases} x_i = t > 0 \\ x_k = 0 \quad \forall k \neq i \end{cases}$

$$\Rightarrow y^T x - \| x \|_1 = y_i t - t = t(y_i - 1) > 0 \xrightarrow[t \to +\infty]{} +\infty$$

Conclusion, if $\exists i$ such that $y_i > 1$, then,

$$\sup_{x} \{ y^T x - \| x \|_1 \} = +\infty$$

* if $\exists i$ such that $y_i < -1$.

We take as an example : $\begin{cases} t < 0 \\ x_i = t \\ x_k = 0 \quad \forall k \neq i \end{cases}$

Then, $y^T x - \|x\|_1 = y_i t - (-t) = y_i t + t$
$= t(y_i + 1) < 0$

$\xrightarrow[t \to -\infty]{} +\infty$

Therefore, $\sup_x \{ y^T x - \|x\|_1 \} = +\infty$

Which makes, if $|y_i| > 1 \Rightarrow \sup_x \{ y^T x - \|x\|_1 \} = +\infty$

$\Rightarrow$ if $\|y\|_\infty > 1$ then, $f^*(y) = +\infty$

Conclusion : $f^*(y) = \begin{cases} 0 & \text{if } \|y\|_\infty \leq 1 \\ +\infty & \text{otherwise} \end{cases}$

2- Compute the dual of (RLS):

$$(RLS): \left\{ \min_x \; \|Ax - b\|_2^2 + \|x\|_1 \right.$$

Let's start by adding and introducing a new variable and new constraint.

$$(RLS) \Leftrightarrow \begin{cases} \min_{x,y} \; \|y\|_2^2 + \|x\|_1 \\ \text{st} \\ Ax - b = y \end{cases}$$

$$\Leftrightarrow \begin{cases} \min_{x,y} \; \|y\|_2^2 + \|x\|_1 \\ \text{st} \\ Ax - y - b = 0 \end{cases}$$

* The lagrangian function: $L(x,y,\upsilon) = \|y\|_2^2 + \|x\|_1 + \upsilon^T(Ax - y - b)$

* $D = \{ (x,y) \mid Ax - y - b = 0 \}$

The dual lagrangian function: $g(\upsilon) = \inf_{x,y \in D} \{ \|y\|_2^2 + \|x\|_1 + \upsilon^T(Ax - y - b) \}$

$$g(\upsilon) = \inf_{x,y \in D} \{ \|x\|_1 + \upsilon^T A x + \|y\|_2^2 - \upsilon^T y - \upsilon^T b \}$$
$$= \inf_{x,y \in D} \{ \|x\|_1 + \upsilon^T A x + y^T y - \upsilon^T y - \upsilon^T b \}$$
$$= \inf_x \{ \inf_y \{ \|x\|_1 + \upsilon^T A x + y^T y - \upsilon^T y - \upsilon^T b \} \}$$

$$= \inf_x \left\{ \|x\|_1 + \upsilon^T A x - b^T \upsilon + \inf_y \{ y^T y - \upsilon^T y \} \right\}$$

$$= \inf_x \left\{ \|x\|_1 + \upsilon^T A x \right\} + \inf_y \{ y^T y - \upsilon^T y \} - b^T \upsilon$$

We suppose $g_1 : y \longmapsto y^T y - \upsilon^T y$ is a quadratic function, differentiable

$\Rightarrow \nabla g_1 = 2 y^T - \upsilon^T = 0$

$\Rightarrow \quad y = \upsilon/2$

$\Rightarrow \inf_y \{ y^T y - \upsilon^T y \} = \dfrac{\upsilon^T \upsilon}{4} - \dfrac{\upsilon^T \upsilon}{2} = -\upsilon^T \upsilon / 4$

In the other hand, $\inf_x \{ \|x\|_1 + \upsilon^T A x \}$

$= - \sup_x \{ -\|x\|_1 - \upsilon^T A x \}$

$= - \sup_x \{ (-A^T \upsilon)^T x - \|x\|_1 \}$

Using the previous question:

$$\sup_x \{ (-A^T \upsilon)^T x - \|x\|_1 \} = \begin{cases} 0 & \text{if } \|A^T \upsilon\|_\infty \leq 1 \\ +\infty & \text{otherwise} \end{cases}$$

$$\Rightarrow \inf_x \{ \|x\|_1 + \upsilon^T A x \} = \begin{cases} 0 & \text{if } \|A^T \upsilon\|_\infty \leq 1 \\ -\infty & \text{otherwise} \end{cases}$$

$$\Rightarrow g(\upsilon) = \begin{cases} -\upsilon^T \upsilon / 4 - b^T \upsilon & \text{if } \|A^T \upsilon\|_\infty \leq 1 \\ -\infty & \text{otherwise} \end{cases}$$

* we have g is a quadratic function with negative coefficients. Thus, g is a concave function.
* $\begin{cases} A^T \upsilon \text{ is a linear function} \Rightarrow \text{convex function} \\ \|\cdot\|_\infty \text{ is a norm which is a convex function} \end{cases}$

$\Rightarrow \{ \upsilon \mid \|A^T \upsilon\|_\infty \leq 1 \}$ is a convex set.

$\Rightarrow$ g is a concave function on the set $\{ \upsilon \mid \|A^T \upsilon\|_\infty \leq 1 \}$.
Thus, we can write the dual function as:

$$(\text{dual}) \Leftrightarrow \begin{cases} \max\limits_{\upsilon} \; -\dfrac{\upsilon^T \upsilon}{4} - \upsilon^T b \\ \text{st} \\ \|A^T \upsilon\|_\infty \leq 1 \end{cases} \Leftrightarrow \begin{cases} \min\limits_{\upsilon} \; \dfrac{\upsilon^T \upsilon}{4} + b^T \upsilon \\ \text{st} \\ \|A^T \upsilon\|_\infty \leq 1 \end{cases}$$

Exercice n 3 : Data Seperation :

* n data points $x_i \in \mathbb{R}^d$ with label $y_i \in \{-1, 1\}$.

we would like to have : $\begin{cases} w^T x_i \leq -1 \Rightarrow y_i = -1 \\ w^T x_i \geq 1 \Rightarrow y_i = 1 \end{cases}$

We solve an optimization problem which minimizes the gap between the hyper-plane and miss-classified points.

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

We also use a quadratic regularization as follow:

$$\min_w \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(w, x_i, y_i) + \frac{\tau}{2} \|w\|_2^2$$, τ regularization parameter

1_ Consider the following quadratic optimization problem

$$\begin{cases} \min_{w, z} \frac{1}{n\tau} 1^T z + \frac{1}{2}\|w\|_2^2 \\ \text{st} \\ z_i \geq 1 - y_i(w^T x_i) \quad \forall i = 1, \dots, n \\ z \geq 0 \end{cases}$$

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

$$\min_w \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(w, x_i, y_i) + \tau/2 \|w\|_2^2 \quad : \quad (\text{sep } 1)$$

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

Now, we will write the problem as its epigraph form:

$$\max\{0, 1 - y_i(w^T x_i)\} = z_i \ \forall i$$

* if $z_i = 0 \Rightarrow 1 - y_i(w^T x_i) \leq 0 = z_i$
* if $z_i = 1 \Rightarrow 1 - y_i(w^T x_i) = 1 = z_i$

$$\Rightarrow \max\{0, 1 - y_i(w^T x_i)\} = z_i \quad \forall i$$

$$\Leftrightarrow \begin{cases} 1 - y_i(w^T x_i) \leq z_i \\ z_i \geq 0 \end{cases}$$

$$\Rightarrow \min_w \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(w, x_i, y_i) + \tau/2 \|w\|_2^2$$

$$\Leftrightarrow \begin{cases} \min_{w, z_i} \frac{1}{n} \sum_{i=1}^{n} z_i + \tau/2 \|w\|_2^2 \\ \text{st} \\ \begin{cases} 1 - y_i(w^T x_i) \leq z_i \\ z_i \geq 0 \end{cases} \end{cases}$$

$$\Leftrightarrow \begin{cases} \min_{w, z} \frac{1}{n\tau} 1^T z + \frac{1}{2}\|w\|_2^2 \quad : (\text{Sep } 2) \\ \begin{cases} 1 - y_i(w^T x_i) \leq z_i \quad \forall i = 1, \dots, n \\ z \geq 0 \end{cases} \end{cases}$$

$\Rightarrow$ (Sep 2) solves problem (Sep 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.

We have (Sep 2):
$$\begin{cases} \underset{w,z}{\text{Min}} & \frac{1}{n\tau} \mathbb{1}^T z + \frac{1}{2}\|w\|_2^2 \\ \text{st} & \begin{cases} -z \leq 0 \\ 1 - y_i(w^T x_i) - z_i \leq 0 \quad \forall i = 1, \dots n \end{cases} \end{cases}$$

The lagrangian function:
$$\begin{aligned} \mathcal{L}(w, z, \lambda_i, \pi) &= \frac{1}{n\tau} \mathbb{1}^T z + \frac{1}{2}\|w\|_2^2 - \pi^T z + \sum_{i=1}^n \lambda_i(1 - z_i - y_i(w^T x_i)) \\ &= \frac{1}{n\tau} \mathbb{1}^T z - \pi^T z - \sum_{i=1}^n \lambda_i z_i + \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^n \lambda_i(1 - y_i(w^T x_i)) \\ &= \left[\frac{1}{n\tau} - \pi - \lambda\right]^T z + \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^n \lambda_i(1 - y_i(w^T x_i)) \\ &= \left[\frac{1}{n\tau} - \pi - \lambda\right]^T z + \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^n \lambda_i(1 - y_i(x_i^T w)) \end{aligned}$$

The dual lagrangian function: $g(\lambda, \pi) = \underset{z, w \in D}{\inf} \{\mathcal{L}(w, z, \lambda_i, \pi)\}$

with $D = \{(z, w) \mid -z \leq 0 \ ; \ 1 - y_i(w^T x_i) - z_i \leq 0 \quad \forall i = 1, \dots n\}$

$$\begin{aligned} \Rightarrow g(\lambda, \pi) &= \inf_w \left\{ \inf_z \left\{ \left(\frac{1}{n\tau} - \pi - \lambda\right)^T z + \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^n \lambda_i(1 - y_i(w^T x_i)) \right\}\right\} \\ &= \inf_w \left\{ \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^n \lambda_i(1 - y_i(w^T x_i)) \right\} + \inf_z \left\{ \left(\frac{1}{n\tau} - \pi - \lambda\right) z \right\} \\ &= \inf_w \left\{ \frac{1}{2} w^T w - \sum_{i=1}^n \lambda_i y_i (x_i^T w) \right\} + \inf_z \left\{ \left(\frac{1}{n\tau} - \pi - \lambda\right)^T z \right\} + \mathbb{1}^T \lambda \end{aligned}$$

* $g_1 : w \longmapsto \frac{1}{2} w^T w - \sum_{i=1}^n \lambda_i y_i (x_i^T w)$ is a quadratic and differentiable function.

$\nabla g_1 = w - \sum_{i=1}^n \lambda_i y_i x_i = 0 \Rightarrow w = \sum_{i=1}^n \lambda_i y_i x_i$

$$\begin{aligned} \Rightarrow \inf_w \left\{ \frac{1}{2} w^T w - \sum_{i=1}^n \lambda_i y_i x_i^T w \right\} &= \frac{1}{2} \left\| \sum_{i=1}^n \lambda_i y_i x_i \right\|_2^2 - \left\| \sum_{i=1}^n \lambda_i y_i x_i \right\|_2^2 \\ &= -\frac{1}{2} \left\| \sum_{i=1}^n \lambda_i y_i x_i \right\|_2^2 \end{aligned}$$

* $g_2 : z \longmapsto \left(\frac{1}{n\tau} - \pi - \lambda\right)^T z$ is a linear function.

$$\Rightarrow \inf_z \left\{ \left(\frac{1}{n\tau} - \pi - \lambda\right)^T z \right\} = \begin{cases} 0 & \text{if } \frac{1}{n\tau} - \pi - \lambda = 0 \\ -\infty & \text{otherwise} \end{cases}$$

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

We have $-\frac{1}{2} \| \sum_{i=1}^{n} \lambda_i y_i x_i \|_2^2 + 1^T \lambda$ is a quadratic function with negative coefficients on $\{ (\pi, \lambda) \mid \frac{1}{n\tau} - \pi - \lambda \}$ which is an affine domain in (λ, π). Thus, $g(\lambda \pi)$ is a concave function.

Dual problem $\Leftrightarrow$
$$\begin{cases} \max_{\lambda, \pi} & -\frac{1}{2} \| \sum_{i=1}^{n} \lambda_i y_i x_i \|_2^2 + 1^T \lambda \\ \text{st} & \begin{cases} \frac{\mathbb{1}}{n\tau} - \pi - \lambda = 0 \\ \lambda \geq 0 \\ \pi \geq 0 \end{cases} \end{cases}$$

$$\Leftrightarrow \begin{cases} \max_{\lambda, \pi} & -\frac{1}{2} \| \sum_{i=1}^{n} \lambda_i y_i x_i \|_2^2 + 1^T \lambda \\ \text{st} & \begin{cases} \frac{\mathbb{1}}{n\tau} - \lambda = \pi \geq 0 \\ \lambda \geq 0 \end{cases} \end{cases}$$

(the objective function is independent of π)

$$\Leftrightarrow \begin{cases} \max_{\lambda} & -\frac{1}{2} \| \sum_{i=1}^{n} \lambda_i y_i x_i \|_2^2 + 1^T \lambda \\ \text{st} & \begin{cases} \frac{1}{n\tau} - \lambda \geq 0 \\ \lambda \geq 0 \end{cases} \end{cases}$$

$$\Leftrightarrow \begin{cases} \max_{\lambda} & -\frac{1}{2} \| \sum_{i=1}^{n} \lambda_i y_i x_i \|_2^2 + 1^T \lambda \\ \text{st} & \begin{cases} \frac{1}{n\tau} \geq \lambda \\ \lambda \geq 0 \end{cases} \end{cases}$$