Assignment 4

Songtuan Lin u6162630

May 3, 2019

Question 1

(a)

According to the definition of conjugate function: $f^*(\mathbf{y}) = \sup_{\mathbf{x}}\{\mathbf{x}^T\mathbf{y} - f(\mathbf{x})\}$, for any $\mathbf{y}$ and $\mathbf{x} \in \mathcal{D}$, we have:

$$\begin{aligned} f^*(\mathbf{y}) &= \sup_{\mathbf{x}}\{\mathbf{x}^T\mathbf{y} - f(\mathbf{x})\} \\ &\geq \mathbf{x}^T\mathbf{y} - f(\mathbf{x}) \end{aligned}$$

By moving the term $f(\mathbf{x})$ within above equation from right-hand side to left-hand side, we can verify that:

$$f^*(\mathbf{y}) + f(\mathbf{x}) \geq \mathbf{x}^T\mathbf{y}$$

(b)

By substituting $\mathbf{y} = \mathbf{0}$ to $f^*(\mathbf{y})$, we have:

$$f^*(0) = \sup_{\mathbf{x}}\{-f(\mathbf{x})\}$$

We can multiple both side of above equation with -1, then, we get:

$$\begin{aligned} -f^*(0) &= -\sup_{\mathbf{x}}\{-f(\mathbf{x})\} \\ &= \inf_{\mathbf{x}}\{f(\mathbf{x})\} \end{aligned}$$

(c)

We can write the conjugate function of $f(\mathbf{x})$ as:

$$f^*(\mathbf{y}) = \sup_{\mathbf{x}}\{\sum_{i=1}^{n}(x_i y_i - \alpha_i \log x_i)\}$$

Suppose there exist a $\alpha_i \geq 0$, then, for any y_i, the term $x_i y_i - \alpha_1 \log x_i$ can always reach infinity, *i.e.* for any y_i, $(x_i y_i - \alpha_1 \log x_i) \to \infty$ when $x_i \to 0$. Hence, we can always choose a $\mathbf{x}$ with $x_i \to 0$ and make $(x_i y_i - \alpha_1 \log x_i) \to \infty$, which means, $f^*(\mathbf{y}) < \infty$ if $\alpha_i < 0$ for each i.

Now we can start to discuss the domain of f^* under the situation $\alpha \prec 0$: Indeed, if there exist a $y_i > 0$, then, the term $x_i y_i - \alpha_1 \log x_i$ will goes to infinity when $x_i \to \infty$. This is because $x_i y_i \to \infty$ along with $x_i \to \infty$ and $-\alpha_i \log x_i$ always greater than zero. Based on this observation, we can conclude the domain of f^* is $\mathbf{y} \prec 0$.

Finally, we can start to find the conjugate function. We first notice that when $y_i < 0$ and $\alpha_i > 0$, the term $x_i y_i - \alpha_1 \log x_i$ is a concave function over x_i. As a result, the function $g(\mathbf{x}) = \sum_{i=1}^{n}(x_i y_i - \alpha_i \log x_i)$ is also concave as it is a non-negative sum of concave functions, which means, $g(\mathbf{x})$ will reach the maximum at the point $\mathbf{x}^*$ that satisfied:

$$\nabla_{\mathbf{x}} g(\mathbf{x}^*) = 0$$

The above equation is equivalent to:

$$y_i - \alpha_i \frac{1}{x_i} = 0 \quad i = 1, 2, \cdots, n \tag{1}$$

Solving Equation 1, we get:

$$x_i = \frac{\alpha_i}{y_i}$$

Therefore, the conjugate function is:

$$\begin{aligned} f^*(\mathbf{y}) &= \sup_{\mathbf{x}}(g(\mathbf{x})) \\ &= \sum_{i=1}^{n}(\alpha_i - \alpha_i \log \frac{\alpha_i}{y_i}) \end{aligned}$$

Question 2

(a)

The original problem can be expressed as:

$$\min_x \quad \sum_{i=1}^{n} |x - b_i|$$

with the objective function $f(x) = \sum_{i=1}^{n} |x - b_i|$ and domain $x \in \mathcal{R}$. Indeed, we can reorder the terms within $f(x)$ as:

$$f(x) = |x - b_1^*| + |x - b_2^*| + \cdots + |x - b_n^*|$$

Where $b_1^*, b_2^* \cdots b_n^*$ is a reordered sequence generated from $b_1, b_2 \cdots b_n$ and satisfied $b_i^* \leq b_j^*$ if $i < j$. By observing $f(x)$, we noticed that when x satisfied $b_i^* \leq x < b_{i+1}^*$, there is:

$$\begin{aligned} f(x) &= (x - b_1^*) + (x - b_2^*) + \cdots + (x - b_i^*) + (b_{i+1}^* - x) + (b_{i+2}^* - x) + \cdots + (b_n^* - x) \\ &= (2i - n)x + (b_{i+1}^* + b_{i+2}^* + \cdots + b_n^*) - (b_1^* + b_2^* + \cdots + b_i^*) \\ &= (2i - n)x + \sigma \end{aligned}$$

Where we have defined $\sigma = (b_{i+1}^* + b_{i+2}^* + \cdots + b_n^*) - (b_1^* + b_2^* + \cdots + b_i^*)$. By observing $f(x)$, we first notice that when $i \leq \frac{n}{2}$, $f(x)$ is always a non-increasing function and when $i > \frac{n}{2}$, $f(x)$ is always non-decreasing. Additionally, i is the index and hence, it must be an integer. Therefore, we can conclude that:

1. $f(x)$ is non-increasing when $i \leq \lfloor \frac{n}{2} \rfloor$.
2. $f(x)$ is non-decreasing when $i \geq \lceil \frac{n}{2} \rceil$.

Furthermore, since $i \leq \lfloor \frac{n}{2} \rfloor$ corresponding to the interval $x < b^*_{\lfloor \frac{n}{2} \rfloor + 1} = b^*_{\lceil \frac{n}{2} \rceil}$ and $i \geq \lceil \frac{n}{2} \rceil$ corresponding to the interval $x \geq b^*_{\lceil \frac{n}{2} \rceil}$, we can conclude that $f(x)$ is non-increasing when $x < b^*_{\lceil \frac{n}{2} \rceil}$ and is non-decreasing when $x \geq b^*_{\lceil \frac{n}{2} \rceil}$. As a result, $f(x)$ reach the minimum at the optimal point $x = b^*_{\lceil \frac{n}{2} \rceil}$.

(b)

We first notice that the function $\|x\mathbf{1} - \mathbf{b}\|_2$ have the same optimal point as $\|x\mathbf{1} - \mathbf{b}\|_2^2$, which means, they both reach the minimum at the same point x. Therefore, we can find this optimal point by solving the optimal problem:

$$\min_x \quad \|x\mathbf{1} - \mathbf{b}\|_2^2$$

Which is the same as:

$$\min_x \quad \sum_{i=1}^{n}(x - b_i)^2$$

It is easy to verify that the function $f(x) = \sum_{i=1}^{n}(x - b_i)^2$ is convex by checking $\nabla_x^2 f(x) = 2n > 0$. Therefore, according to the first-order condition, $f(x)$ will reach the minimum at the optimal x^* which satisfied:

$$\begin{aligned} \nabla_x f(x^*) &= \sum_{i=1}^{n} 2(x^* - b_i) \\ &= 0 \end{aligned}$$

The solution of this equation is:

$$x^* = \frac{1}{n}\sum_{i=1}^{n} b_i$$

Hence, the optimal point for the original problem is also $x^* = \frac{1}{n}\sum_{i=1}^{n} b_i$.

(c)

By using the similar method as in (a), we can reorder the elements within $\mathbf{b}$ and reform the problem as:

$$\min_x \quad \max_i\{|x - b_1^*|, |x - b_2^*|, \cdots, |x - b_n^*|\}$$

We first check the situation that $x \leq b_1^*$, we have:

$$\begin{aligned}
\|x\mathbf{1} - \mathbf{b}\|_\infty &= \max\{b_1^* - x, b_2^* - x, \cdots, b_n^* - x\} \\
&= b_n^* - x \\
&\geq b_n^* - b_1^*
\end{aligned}$$

It can be seen that under this situation, $\|x\mathbf{1} - \mathbf{b}\|_\infty$ reach the minimum $b_n^* - b_1^*$ when $x = b_1^*$. After that, we can check the situation that $b_1^* \leq x \leq b_i^*$, we have:

$$\begin{aligned}
\|x\mathbf{1} - \mathbf{b}\|_\infty &= \max\{x - b_1^*, x - b_2^*, \cdots, x - b_{i-1}^*, b_i^* - x, b_{i+1}^* - x, \cdots, b_n^* - x\} \\
&= \max\{x - b_1^*, b_n^* - x\}
\end{aligned}$$

We first notice that the above equation hold for any b_i^*, hence, we can expand this equation to the condition that $b_1^* \leq x \leq b_n^*$:

$$\|x\mathbf{1} - \mathbf{b}\|_\infty = \max\{x - b_1^*, b_n^* - x\}$$

Furthermore, we realize that when $x > \frac{b_1^*+b_n^*}{2}$, $x - b_1^* > b_n^* - x$ and when $x \leq \frac{b_1^*+b_n^*}{2}$, $x - b_1^* \leq b_n^* - x$. Therefore, we have:

$$\|x\mathbf{1} - \mathbf{b}\|_\infty = \begin{cases} b_n^* - x & x \leq \frac{b_1^*+b_n^*}{2} \\ x - b_1^* & x > \frac{b_1^*+b_n^*}{2} \end{cases}$$

Hence, $\|x\mathbf{1} - \mathbf{b}\|_\infty$ reach the minimum value $\frac{b_n^*-b_1^*}{2}$ at $x = \frac{b_n^*+b_1^*}{2}$. Finally, we can check the situation that $x \geq b_n^*$. Under this situation, $\|x\mathbf{1} - \mathbf{b}\|_\infty$ has the minimum $b_n^* - b_1^*$. Since $\frac{b_n^*-b_1^*}{2} < b_n^* - b_1^*$, we can then conclude that for any $x \in \mathcal{R}$, the optimal point for function $\|x\mathbf{1} - \mathbf{b}\|_\infty$ is:

$$x = \frac{b_n^* + b_1^*}{2}$$

with the optimal value $\frac{b_n^*-b_1^*}{2}$.

(d)

As the similar method been used in (b), we can first find the optimal point for the optimal problem:

$$\min_x \quad \|x\mathbf{a} - \mathbf{b}\|_2^2$$

Which is equivalent to the problem:

$$\min_x \quad \sum_i (a_i x - b_i)^2$$

It is easy to verify that the function $f(x) = \sum_i (a_i x - b_i)^2$ is convex as $\nabla_x^2 f(x) = \sum_i 2a_i^2 > 0$. Therefore, the optimal point x^* must satisfied the first-order condition:

$$\begin{aligned}
\nabla_x f(x^*) &= \sum_i 2a_i(a_i x - b_i) \\
&= 0
\end{aligned}$$

The solution of this equation is:

$$x^* = \frac{\sum_i a_i b_i}{\sum_i a_i^2}$$

Which means, the optimal point of the original problem is $x^* = \frac{\sum_i a_i b_i}{\sum_i a_i^2}$.

Q3

(a)

We can first the Lagrangian as:

$$\begin{aligned}\mathcal{L}(\mathbf{x}, \mathbf{r}) &= \sum_i \phi(r_i) - \nu^T(\mathbf{r} - \mathcal{A}\mathbf{x} + \mathbf{b}) \\ &= \sum_i \phi(r_i) - \nu^T\mathbf{r} + \nu^T\mathcal{A}\mathbf{x} - \nu^T\mathbf{b}\end{aligned}$$

According to the Lagrangian, we can further write the Lagrange Dual Function as:

$$g(\nu) = \inf_{\mathbf{r},\mathbf{x}}\{\sum_i \phi(r_i) - \nu^T\mathbf{r} + \nu^T\mathcal{A}\mathbf{x} - \nu^T\mathbf{b}\}$$

By inspecting the Lagrange Dual Function, we can find that if $\nu^T\mathcal{A} \neq 0$, $g(\nu)$ can easily approach to $-\infty$ as $\mathbf{x} \to \infty$ or $-\infty$. Therefore, the first condition that ν need to satisfied is:

$$\nu^t\mathcal{A} = 0 \tag{2}$$

Therefore, the Lagrange Dual Function can be simplified as:

$$\begin{aligned}g(\nu) &= \inf_{\mathbf{r}}\{\sum_i \phi(r_i) - \nu^T\mathbf{r} - \nu^T\mathbf{b}\} \\ &= \inf_{\mathbf{r}}\{\sum_i \phi(r_i) - \nu^T\mathbf{r}\} - \nu^T\mathbf{b} \\ &= \inf_{\mathbf{r}}\{\sum_i (\phi(r_i) - v_i r_i)\} - \nu^T\mathbf{b}\end{aligned}$$

As a result, the first thing we need to do to construct the Dual Problem is to solve $\inf_{\mathbf{r}}\{\sum_i (\phi(r_i) - v_i r_i)\}$. In order to do so, we first notice that the element r_i within $\mathbf{r}$ is independent with each other, therefore, we can first find the minimum of each single term: $\phi(r_i) - v_i r_i$ and then sum them together. Indeed, we can discuss the minimum of $\phi(r_i) - v_i r_i$ based on the following three cases:

1. $r_i \leq -1$: Under this condition, the term $\phi(r_i) - v_i r_i$ become:

$$\begin{aligned}\phi(r_i) - v_i r_i &= -r_1 - 1 - \nu_i r_i \\ &= (-\nu_i - 1)r_i - 1\end{aligned}$$

   Based on this equation, it is easy to find out that of $\nu_i \geq -1$, $\phi(r_i) - v_i r_i$ reach the minimum ν_i at $r_i = -1$. By contrast, if $\nu_i < -1$, $\phi(r_i) - v_i r_i \to -\infty$ as $r_i \to -\infty$. Therefore, the second condition that ν must satisfied is: For each ν_i:

$$\nu_i \geq -1 \tag{3}$$

2. $r_i \geq 1$: Under this condition, the term $\phi(r_i) - v_i r_i$ become:

$$\phi(r_i) - v_i r_i = (1 - \nu_i)r_i - 1$$

   By inspecting this equation, if $\nu_i \leq 1$, $\phi(r_i) - v_i r_i$ has the minimum $-\nu_i$ at the point $r_i = 1$. By contrast, if $\nu_i > 1$, $\phi(r_i) - v_i r_i$ goes to $-\infty$ along with $r_i - \to \infty$. As a result, the third condition that ν must satisfied is: For each ν_i

$$\nu_i \leq 1 \tag{4}$$

3. $-1 \le r_i \le 1$: Under this condition, the term $\phi(r_i) - v_i r_i$ become:

$$\phi(r_i) - v_i r_i = -\nu_i r_i$$

Indeed, under this condition, it is straightforward to figure out the minimum of $\phi(r_i) - v_i r_i$ is either ν_i or $-\nu_i$ according to the sign of ν_i.

As a result, we can find the minimum of $\phi(r_i) - v_i r_i$ over the entire range of r_i by combining the above three conditions:

$$\begin{aligned} \inf_{r_i}\{\phi(r_i) - v_i r_i\} &= \min\{-\nu_i, \nu_i\} \\ &= -|\nu_i| \end{aligned} \tag{5}$$

With the constrain:

$$-1 \le \nu_i \le 1$$

Which is equivalent to:

$$|\nu_i| \le 1 \tag{6}$$

and

$$\nu^T \mathcal{A} = 0 \tag{7}$$

As a result, the Dual Function, accoeding to Equation 5, 6 and 7, can be written as:

$$\begin{aligned} g(\nu) &= \sum_i -|\nu_i| - \nu^T \mathbf{b} \\ &= -\|\nu\|_1 - \nu^T \mathbf{b} \end{aligned} \tag{8}$$

With the constrains: $|\nu_i| \le 1$ for each ν_i and $\nu^T \mathcal{A} = 0$. Furthermore, since for each ν_i in ν, there is $|\nu_1| \le 1$, we can rewrite this constrain as $\|\nu\|_\infty \le 1$. As a result, the dual problem is:

$$\begin{aligned} \max_{\nu} \quad & -\|\nu\|_1 - \nu^T \mathbf{b} \\ s.t. \quad & \|\nu\|_\infty \le 1 \\ & \nu^T \mathcal{A} = 0 \end{aligned}$$

(b)

As what we did in (a), we can first construct the Lagrangian:

$$\sum_i \phi(r_i) - \nu^T \mathbf{r} + \nu^T \mathcal{A}\mathbf{x} - \nu^T \mathbf{b}$$

and the Lagrange Dual Function:

$$g(\nu) = \inf_{\mathbf{r}}\{\sum_i (\phi(r_i) - v_i r_i)\} - \nu^T \mathbf{b}$$

with constrain:

$$\nu^T \mathcal{A} = 0$$

Indeed, in order to find the minimum of $\sum_i (\phi(r_i) - v_i r_i)$, we can also start with finding the minimum of each $\phi(r_i) - \nu_i r_i$ and then summing them together. As in (a), we also separate the discussion into following cases:

1. $r_i \le -1$: Under this condition, the term $\phi(r_i) - \nu_i r_i$ becomes:

$$\phi(r_i) - \nu_i r_i = (-2 - \nu_i) r_i - 1$$

By inspecting this equation, we can find out that if $\nu_i \ge -2$, $\phi(r_i) - \nu_i r_i$ reach the minimum $\nu_i + 1$ at point $r_i = -1$. By contrast, if $\nu_i < -2$, $\phi(r_i) - \nu_i r_i$ will goes to $-\infty$ along with $r_i \to -\infty$. Therefore, for each ν_i, it must satisfied:

$$\nu_i \ge -2 \tag{9}$$

2. $r_i \ge 1$: Under this condition, the term $\phi(r_i) - \nu_i r_i$ becomes:

$$\phi(r_i) - \nu_i r_i = (2 - \nu_i) r_i - 1$$

By inspecting this equation, we can find out that if $\nu_i \le 2$, $\phi(r_i) - \nu_i r_i$ reach the minimum at point $r_i = 1$. By contrast, if $\nu_i > 2$, $\phi(r_i) - \nu_i r_i$ goes to $-\infty$ along with $r_i \to \infty$. Therefore, for each ν_i, it also need to satisfied:

$$\nu_i \le 2 \tag{10}$$

3. $-1 \le r_i \le 1$: Under this condition, the term $\phi(r_i) - \nu_i r_i$ becomes:

$$\phi(r_i) - \nu_i r_i = r_i^2 - \nu_i r_i \tag{11}$$

We first notice that if we do not consider the range of r_i, $r_i^2 - \nu_i r_i$ will reach the minimum $-\frac{\nu_i^2}{4}$ at the point $r_i = \frac{\nu_i}{2}$. Indeed, from constrain 9 and 10, we have $-2 \le \nu_i \le 2$. Therefore,

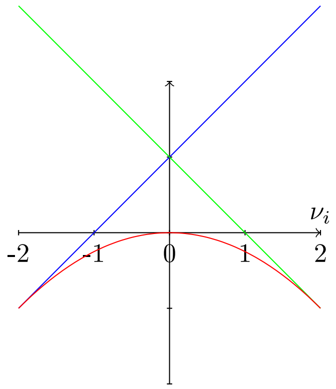
$$-1 \le \frac{\nu_i}{2} \le 1$$

which located inside out assumption range of r_i. Hence, $\phi(r_i) - \nu_i r_i$ reach the minimum $-\frac{\nu_i^2}{4}$ at the point $r_i = \frac{\nu_i}{2}$.

We can construct the minimum of $\phi(r_i) - \nu_i r_i$ over the entire range of r_i by combining the above three cases:

$$\inf_{r_i}\{\phi(r_i) - \nu_i r_i\} = \min\{\nu_i + 1, 1 - \nu_i, -\frac{\nu_i^2}{4}\}$$

with the constrain $-2 \le \nu_i \le 2$ and $\nu^T \mathcal{A} = 0$. Indeed, we can plot the graph for these three terms as:



It can be seen that within the interval $-2 \le \nu_i \le 2$, we have $\min\{\nu_i + 1, 1 - \nu_i, -\frac{\nu_i^2}{4}\} = -\frac{\nu_i^2}{4}$. As a result, we have:

$$\inf_{r_i}\{\phi(r_i) - \nu_i r_i\} = -\frac{\nu_i^2}{4}$$

Therefore, the Dual Function is:

$$\begin{aligned} g(\nu) &= \sum_i -\frac{\nu_i^2}{4} - \nu^T \mathbf{b} \\ &= -\frac{1}{4}\|\nu\|_2^2 - \nu^T \mathbf{b} \end{aligned}$$

Hence, the Dual Problem is:

$$\begin{aligned} \max_{\nu} \quad & -\frac{1}{4}\|\nu\|_2^2 - \nu^T \mathbf{b} \\ s.t. \quad & \|\nu\|_\infty \le 2 \\ & \nu^T \mathcal{A} = 0 \end{aligned}$$