Convex Optimizations – Homework 2

Q1.

1. Given:

The objective function is

$$\min_{w,x} \sum_{i=1}^{m} g(z_i)$$

$$\text{s.t } z_i = y_i - w^t x_i \qquad\qquad i = 1, \dots, m$$

Writing the above in Lagragian form, we get

$$L(w, z, v) = \sum_{i=1}^{m} g(z_i) + \sum_{i=1}^{m} v_i(z_i - y_i + w^t x_i)$$

Writing the second term in vector form, we get,

$$L(w, z, v) = \sum_{i=1}^{m} g(z_i) + v^T Z - \ v^T Y + \ v^T X^T W$$

To find the dual function we need to find $G(v) = \ \inf\limits_{w,x} L(w, z, v)$

$$G(v) = \ \inf_{w,x} \sum_{i=1}^{m} (g(z_i) + v^T Z - \ v^T Y + \ v^T X^T W$$

$$G(v) = -v^T Y + \ \inf_{z} \textstyle\sum_{i=1}^{m} g(z_i) + \ v^T Z + \ \inf\limits_{w} v^T X^T W$$

$$G(v) = -v^T Y - \left(\sup_{z} - \sum_{i=1}^{m} g(z_i) - \ v^T Z \right) + \ \inf_{w} v^T X^T W \tag{1}$$

Consider, the part, $\sup\limits_{z} - \sum_{i=1}^{m} g(z_i) - \ v^T Z$, this is the fenchel conjugate of Z

$$\therefore \ \sup_{z} - \sum_{i=1}^{m} g(z_i) - \ v_i^T z_i = g^*(-v) \tag{2}$$

To find infimum w.r.t w we can take derivative of $L(w, z, v)$ and equate it to 0

$$\therefore we\ get, v^T X^T = 0$$

$$\therefore \inf_{w} v^T X^T W = \begin{cases} 0 \quad , & v^T X^T = 0 \\ -\infty & , otherwise \end{cases} \tag{3}$$

From $(1), (2)\ and\ (3)\ we\ get$

$$G(v) = \begin{cases} -{v^*}^T Y - \sum_{i=1}^{m} g^*(-v_i^*) \quad , & v^T X^T = 0 \\ -\infty & , otherwise \end{cases}$$

1.2 In the current problem we don't have any inequality constraints, therefore, we will have to satisfy less number of KKT conditions. The KKT conditions for the current problem are

$$z_i^* - y_i + {w^*}^T x_i = 0 \qquad i = 1, \dots, m \tag{4}$$

$$\sum_{i=1}^{m} \nabla_z g({z^*}_i) + \ v_i^* = 0 \tag{5}$$

$${v^*}^T X^T = 0 \tag{6}$$

$\therefore (4), (5)\ and\ (6)$ are the KKT conditions for the pair of primal and dual optimal solutions for the current problem.

When the primal problem is convex, the KKT conditions are also sufficient for the points to be primal and dual optimal. To see this, note that the condition (4) notes that $Z^*\ and\ W^*$ are primal feasible. In the case where we also have p inequality constraints in the problem, we need to satisfy the following extra KKT conditions.

$$f_i(W^*, Z^*) \leq 0, \qquad i = 1, \dots, p$$

$$\lambda^* \geq 0, \qquad i = 1, \dots, p$$

$$\lambda_i^* f_i(W^*, Z^*) = \ 0, \qquad i = 1, \dots, p$$

Since, $\lambda^* \geq 0$, $L(W, Z, \lambda^*, \nu^*)$ is convex in (W, Z), the conditions $(5)\ and\ (6)$ state that its gradient with respect to W and Z vanish when $W = W^*\ and\ Z = Z^*$, so it follows that $W^* and\ Z^*$ minimize $L(W, Z, \lambda^*, \nu^*)$

From this we can conclude that, where $f_0(W, Z)$ is the objective function.

$$\begin{aligned} G(\lambda^*, \nu^*) &= L(W^*, Z^*, \lambda^*, \nu^*) \\ &= f_0(W^*, Z^*) + \sum_{i=1}^{p} \lambda^* f_i(W^*, Z^*) + \sum_{i=1}^{m} \nu_I^* h_i(W^*, Z^*) \\ &= f_0(W^*, Z^*) \end{aligned}$$

Where in the last line we use $h_i(W^*, Z^*) = 0$ and $\lambda_i^* f_i(W^*, Z^*) = \ 0$. This shows that W^*, Z^* and (λ^*, ν^*) have zero duality gap, and therefore the primal and dual optimal. In summary, for any convex optimization problem with differentiable objective and constraint functions, any points that satisfy the KKT conditions are primal and dual optimal, and have zero duality gap.

1.3

a) We need to derive the Fenchel conjugate, first we consider $g(Z) = \frac{1}{2}Z^2$

Using the formula for Fenchel conjugate, we get

$$g^*(Y) = \sup_Z Z^T Y - \frac{1}{2}Z^2 \tag{7}$$

To find the sup, we take the derivate of (7) w.r.t Z and equate it to 0.

$$\therefore Y - Z = 0$$
$$\therefore Z^* = Y$$

Substitute this in the equation (7),

$$\therefore g^*(Y) = Y^T Y - \frac{1}{2}Y^2$$
$$\therefore g^*(Y) = \left||Y|\right|^2 - \frac{1}{2}Y^2$$
$$\therefore g^*(Y) = \begin{cases} \left||Y|\right|^2 - \frac{1}{2}Y^2 , Y = Z^* \\ \infty , \quad otherwise \end{cases}$$

Hence, we found the fenchel conjugate for the function.

b) Consider the function $g(z) = \max(|z| - 1{,}0)$

Using the formula for Fenchel conjugate, we get

$$g^*(y_i) = \sup_Z z_i y_i - \max(|z_i| - 1{,}0) \tag{8}$$

To find the sup, we take the derivate of (8) w.r.t Z and equate it to 0.

Consider the case, $\mathrm{z_i} > 1$

$$\therefore g^*(y_i) = \sup_Z z_i y_i - \mathrm{z_i} + 1$$

Taking derivative and equate to 0 we get,

$$y_i^* = 1$$

$$g^*(y_i) = \begin{cases} 1, & y_i^* = 1 \\ \infty, & otherwise \end{cases} \tag{9}$$

Consider the case, $z_i < 1$

$$\therefore g^*(y_i) = \sup_{Z} z_i y_i - (-\mathrm{z_i} - 1)$$

$$= \sup_{Z} z_i y_i + z_i + 1$$

Taking derivative and equate to 0 we get,

$$y_i^* = -1$$

$$g^*(y_i) = \begin{cases} 1, & y_i^* = -1 \\ \infty, & otherwise \end{cases} \tag{10}$$

Case: $-1 \le z_i \le 1$

$$\therefore g^*(y_i) = \sup_{Z} z_i y_i$$

Taking derivative and equate to 0 we get,

$$y_i^* = 0$$

$$g^*(y_i) = \begin{cases} 0, & y_i = 0 \\ \max(y_i, -y_i), & otherwise \end{cases}$$

Therefore the Fenchel conjugate of the original function is,

$$g^*(Y) = \sum_{i=1}^{m} g^*(y_i)$$

Where, $$g^*(y_i) = \begin{cases} 1,\ (y_i = 1 \ or \ y_i = -1) \ and \ |z_i| > 1 \\ 0, -1 \le z_i \le 1 \ and \ y_i = 0 \\ \max(y_i, -y_i), -1 \le z_i \le 1 \ and \ y_i \ne 0 \\ \infty, otherwise \end{cases}$$

1.4

$$g(w,z) = \sum_{i=1}^{m}(z_i - y_i + w^T x_i) + < 0, w >$$

$$g(\,\xi,\nu) = \sum_{i=1}^{m} g^*(\nu_i) + \begin{cases} 0, \xi = 0 \\ \infty, \xi \neq 0 \end{cases}$$

1.4.1 Consider $\mathrm{g(z)} = \frac{1}{2}z^2$

$$g(\xi,\nu_i) = \sum_{i=1}^{m} g^*(\nu_i) + \begin{cases} 0, \xi = 0 \\ \infty, \xi \neq 0 \end{cases}$$

$$\therefore \sum_{i=1}^{m} g^*(\nu_i) = \begin{cases} \left||\nu|\right|^2 - \frac{1}{2}\nu^2 \,, \nu = Z^* \\ \infty \,, \qquad otherwise \end{cases}$$

1.4.2 Consider $g(z) = \max(|z| - 1{,}0)$

$$g(\xi,\nu_i) = \ g^*(\nu_i) \begin{cases} 0, \xi = 0 \\ \infty, \xi \neq 0 \end{cases}$$

$$g^*(\xi,\nu_i) = \begin{cases} 1, \qquad (\nu_i = 1 \; or \; \nu_i = \; -1) \; and \; |z_i| > 1 \; and \; \xi = 0 \\ 0, \qquad -1 \leq z_i \leq 1 \; and \;\; \nu_i = 0and \; \xi = 0 \\ \max(\nu_i, -\nu_i)\,, -1 \leq z_i \leq 1 \; and \; \nu_i \neq 0and \; \xi = 0 \\ \infty, otherwise \end{cases}$$

$$g(\xi,\nu) = \sum_{i=1}^{m} g^*(\xi,\nu_i)$$