Convex Optimization. Duality.

1.

$$\min_x \; c^T x + \frac{1}{\mu}\sum_{i=1}^m \log(1+e^{\mu(a_i^T x - b_i)}). \quad x \in \mathbb{R}^n. \; \mu>0 \qquad (1)$$

(a) Show that (1) is equivalent to.

$$\min_x \; c^T x + \frac{1}{\mu}\sum_{i=1}^m \log(1+e^{\mu y_i}) \qquad (2)$$

$$\text{s.t.} \quad Ax - b \le y$$

Notice that. $\frac{1}{\mu}\log(1+e^{\mu y_i})$ is monotonously increasing., it can only be minimized when y^* is tight on its constraints.

Otherwice, suppose $\exists j$. $a_j^T x^* - b_j < y_j^*$, and (x^*, y^*) is minimizer

take $\tilde{y} = (y_1^* \ldots y_{j-1}^*, a_j^T x^* - b_j, y_{j+1}^*, \ldots)$.

~~We~~ Then $c^T x^* + \frac{1}{\mu}\sum_{i=1}^m \log(1+e^{\mu \tilde{y}_i}) < c^T x^* + \frac{1}{\mu}\sum_{i=1}^m \log(1+e^{\mu y^*})$

Contradicts to (x^*, y^*) is minimizer.

Hence. $y_j^* = a_j^T x^* - b_j$. $\forall j \in [m]$.

$\Rightarrow$ ~~minimizer (2) =~~

(2)'s minimizer $= (x^{*T}, (Ax^* - b)^T)^T$

Just consider the x. part.

x^* is then also the ~~of~~ minimizer of (1)

$\Rightarrow$ (1) and (2) are equivallent.

(b). $L(x,y,\lambda) = c^T x + \frac{1}{\mu}\sum_{i=1}^m \log(1+e^{\mu y_i}) + \lambda^T(Ax-b-y)$

$$d(\lambda) = \inf_{x,y} c^T x + \frac{1}{\mu}\sum_{i=1}^m \log(1+e^{\mu y_i}) + \lambda^T(Ax-b-y)$$

$$= \inf_x (c+A^T\lambda)^T x + \inf_y \left[\frac{1}{\mu}\sum_{i=1}^m \log(1+e^{\mu y_i}) - \lambda^T y\right] - \lambda^T b$$

Let $g(y) = \left[\frac{1}{\mu}\sum_{i=1}^m \log(1+e^{\mu y_i}) - \lambda^T y\right]$

i) $\lambda_i \ne 0$ $\frac{\partial g(y)}{\partial y_i} = 0 \Rightarrow y_i = \frac{1}{\mu}\ln\frac{\lambda_i}{1-\lambda_i}$. ~~($\lambda_i \ne 1$)~~ $\lambda_i \in (0,1)$

ii) $\lambda_i = 0$. $y = -\infty$. to reach infimum.

iii) $\lambda_i > 1$. $\inf_y g(y) = -\infty$

$0 \le \lambda < 1$

Hence. $d(\lambda) = \begin{cases} -\lambda^T b & c + A^T\lambda = 0 \wedge \cancel{y_i = \frac{1}{\mu}\ln\frac{\lambda_i}{1-\lambda_i}, i \in [m]} \wedge \cancel{\lambda_i \ne 1, i \in [m]} \; \cancel{\lambda \in (0,1)} \\ -\infty. & \text{Otherwise} \end{cases}$

the Lagrage dual problem is $\max \; -\lambda^T b$

s.t. $A^T\lambda + C = 0$

$1 > \lambda \ge 0$, ~~$\lambda_i \ne 1, i \in [m]$~~

(c) Denote the optimal value of dual Linear Programming as d^*.

Denote the optimal value of (2) as $\tilde{d}^*$.

Because dual LP has all entries smaller than 1.

$\Rightarrow \quad d^* = \tilde{d}^*$.

As strong duality holds for Linear programming

$$p^* = d^*.$$

For weak duality of convex optimization

$\Rightarrow \quad q^* \geq \tilde{d}^*$.

Hence $\quad q^* \geq \tilde{d}^* = d^* = p^*$.

$\Rightarrow \quad p^* \leq q^*$.

On the other hand

$$q^* = \min_x \; c^T x + \frac{1}{\mu} \sum_{i=1}^m \log(1 + e^{\mu(a_i^T x - b_i)})$$

$$\leq \min_x \; c^T x + \frac{1}{\mu} \sum_{i=1}^m \log(1 + e^{\mu(a_i^T x - b_i)}), \quad \text{s.t. } Ax \leq b. \quad \text{(still feasible)}$$

because of primal LP.

$$\leq \min_x \; c^T x + \min_x \frac{1}{\mu} \sum_{i=1}^m \log(1 + e^{\mu(a_i^T x - b_i)}), \quad \text{s.t } Ax \leq b$$

has finite optimum

$$\leq p^* + \frac{m}{\mu} \log 2$$

$$\Rightarrow \quad p^* \leq q^* \leq p^* + \frac{m}{\mu} \log 2 \qquad \square$$

2. Problem 2.

$$\min_{x_1, x_2} \quad e^{-x_2}$$

$$\text{s.t.} \quad \sqrt{x_1^2+x_2^2} - x_1 \le 0.$$

(a). The objective function is convex because.

i) e^x is convex

ii) $-x_2 = (0,-1)\cdot\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ is an affine function

The Constraint function is convex because

i) $\sqrt{x_1^2+x_2^2} = \|x\|_2$ is a norm, and is convex

ii) ~~$-x_1$ is ... $(0,-$~~

$-x_1 = (-1, 0)\cdot\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ is affine function ; hence convex

iii) Sum of Convex functions ~~are~~ is Convex

~~(b)~~ Feasible set:

$$\sqrt{x_1^2+x_2^2} \le x_1$$

$$\Rightarrow \quad x_1^2 + x_2^2 \le x_1^2$$

$$\Rightarrow \quad x_2 = 0.$$

The feasible set is $\{(x_1, x_2) \mid x_1 \ge 0, x_2 = 0\} = \{(R_+, 0)\}$,

Hence. the optimal value is 1.

(b)

$$L(x_1, x_2, \lambda) = e^{-x_2} + \lambda\left(\sqrt{x_1^2+x_2^2} - x_1\right)$$

$$= e^{-x_2} + \lambda \cdot \frac{x_2^2}{\sqrt{x_1^2+x_2^2} + x_1} \quad (x_1, x_2 \neq 0).$$

$$L(0,0,\lambda) = 1.$$

~~Also~~

Obviously, we have $L(x_1, x_2, \lambda) > 0$. (1)

Along the trajectory $x_1 = x_2^3$.

~~dx~~ $$L(x_1, x_2, \lambda) = e^{-x_2} + \lambda \frac{x_2^2}{\sqrt{x_2^6+x_2^2} + x_2^3}$$

$$= e^{-x_2} + \lambda \cdot \frac{1}{x_2 + \sqrt{x_2^2 + \frac{1}{x_2^2}}}$$

$$d(\lambda) = \inf_{x_1, x_2} L(x_1, x_2, \lambda)$$

$$\leq \inf_{x_1 = x_2^3, x_2} L(x_1, x_2, \lambda)$$

$$= \inf_{x_2} e^{-x_2} + \frac{\lambda}{x_2 + \sqrt{x_2^2 + \frac{1}{x_2^2}}}$$

$$= 0. \quad (2)$$

Hence. $d(\lambda) = 0$. based on (1) and (2)

dual problem

$$\max \; d(\lambda) = 0$$
$$\text{s.t. } \lambda > 0$$

gap is 1.

1 (c) No

The weak duality implies the duality gap $g \succeq 0$.

If the optimization problem is convex problem, and slater's condition is satisfied, then the strong duality holds. But slater's condition is just a sufficient condition for strong duality.

A necessary condition of strong duality is that we compute the ~~maximum~~ optimal value of dual problem d^*, and the optimal value of primal problem p^*, and it holds that $p^* = d^*$. we cannot prove the duality gap is nonzero without solving the dual problem.

3. $$\min_{x,y} \; \|y\|_2 + \gamma \|x\|_1$$
$$\text{s.t.} \quad Ax - b = y.$$

(a) $$L(x,y,z) = \|y\|_2 + \gamma\|x\|_1 + z^T(Ax-b-y)$$
$$= \left(\gamma\|x\|_1 + (A^Tz)^T x\right) + \left(\|y\|_2 - z^T y\right) - z^T b.$$

$$d(z) = \inf_{x,y} \left(\gamma\|x\|_1 + (A^Tz)^T x\right) + \left(\|y\|_2 - z^T y\right) - z^T b$$
$$= \inf_x \left(\gamma\|x\|_1 + (A^Tz)^T x\right) + \inf_y \left(\|y\|_2 - z^T y\right) - z^T b,$$
$$= -\gamma \left[\sup_x \left(\left(\frac{-A^Tz}{\gamma}\right)^T x - \|x\|_1\right)\right] - \sup_y \left(z^T y - \|y\|_2\right) - z^T b.$$
$$= -\gamma \cdot f^*\left(\frac{-A^Tz}{\gamma}\right) - g^*(z) - z^T b, \quad \text{where } f(x) = \|x\|_1,\; g(x) = \|y\|_2.$$
$$= -\gamma\, I\left\{\left\|\frac{-A^Tz}{\gamma}\right\|_\infty \le 1\right\} - I\{\|z\|_2 \le 1\} - z^T b.$$

Hence the dual problem is.

$$\max \quad -b^T z$$
$$\text{s.t.} \quad \|A^T z\|_\infty \le \gamma$$
$$\|z\|_2 \le 1.$$

Since x^* is an optimal solution, and (5) is a convex problem
x^* must fulfill the KKT conditions
$\Rightarrow$ $\nabla_y L(x,y,z) = 0$
$\Rightarrow$ $\frac{y^*}{\|y^*\|_2} - z^* = 0$,

Dual feasibility $\Rightarrow$ $\|A^T z^*\|_\infty \le \gamma$.
$\Rightarrow$ $\|A^T \frac{y^*}{\|y^*\|_2}\|_\infty \le \gamma$.
$\Rightarrow$ $\|A^T \frac{Ax^*-b}{\|Ax^*-b\|_2}\|_\infty \le \gamma$. (1)

$\nabla_x L(x,y,\lambda) = 0$
$\Rightarrow$ $z^{*T}A + \gamma\, \mathrm{sgn}(x) = 0$ Notice $z^* = \frac{y^*}{\|y^*\|_2} = \frac{Ax^*-b}{\|Ax^*-b\|_2} = r$
$\Rightarrow$ $r^T A x^* + \gamma \cdot \mathrm{sgn}(x) \cdot x^* = 0$
$\Rightarrow$ $r^T A x^* + \gamma \|x^*\|_1 = 0$ (2)

(1)+(2) $\Rightarrow$ proof done. $\square$

4.

$$g^*(s) = \sup_x \; s^T x - g(x)$$
$$= \sup_x \; s^T x - f(Lx+c)$$
$$= -\inf_x \left(f(Lx+c) - s^T x \right)$$
$$= -\inf_{x,y} \left[f(y) - s^T x + I_{\{0\}}(Lx+c-y) \right]$$
$$= -\inf_{x,y} \sup_\mu \left[f(y) - s^T x + \mu^T (Lx+c-y) \right]$$

$\mathrm{dom}(g) \neq \phi$, hence strong duality holds

$$= -\sup_\mu \left[\left(\inf_x \, (L^T\mu - s)^T x \right) + \left(\inf_y \, f(y) - \mu^T y \right) + \mu^T c \right]$$
$$= -\sup_\mu \left(-f^*(\mu) + \mu^T c \;:\; s = L^T \mu . \right)$$
$$= \inf_\mu \left(f^*(\mu) - \mu^T c \;:\; s = L^T \mu . \right) \quad \square$$