EE381K : Convex Optimization Fall 2016 Po-Cheng Pan

UT EID: pp22828

1. Condition number

Convergence condition: Bounded Hessian $\Rightarrow \nabla^2 f \preceq MI$ (From (9.12) on the textbook)

Strong convexity : $\nabla^2 f(x) \succeq mI \Rightarrow f(y) \geq f(x) + \nabla f(x)^T(y-x) + \frac{m}{2}\|y-x\|_2^2$

$\Rightarrow p^* \geq f(x) - \frac{1}{2m}\|\nabla f(x)\|_2^2$ (From (9.9) on the textbook)

For $x, y \in S$,

we have $f(y) = f(x) + \nabla f(x)^T(y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(z)(y-x)$ for some z on the line segment $[x, y]$

$\because \nabla^2 f \preceq MI$

$\Rightarrow f(y) \leq f(x) + \nabla f(x)^T(y-x) + \frac{M}{2}\|y-x\|_2^2$

Consider gradient algorithm,

let $y = x - t\nabla f(x)$

$\Rightarrow f(x - t\nabla f(x)) \leq f(x) - t\|\nabla f(x)\|_2^2 + \frac{Mt^2}{2}\|\nabla f(x)\|_2^2$

To satisfy $f(x - t\nabla f(x)) \leq f(x)$

$-t\|\nabla f(x)\|_2^2 + \frac{Mt^2}{2}\|\nabla f(x)\|_2^2 < 0$

$\Rightarrow t < \frac{2}{M} \Rightarrow M < \frac{2}{t}$

$\because \nabla^2 f \leq MI$

$\Rightarrow \nabla^2 f \leq \frac{2}{t} I$ is its converge condition #

For example, consider $f = \frac{4}{t}x^t x$, $\nabla^2 f = \frac{8}{t}I \npreceq \frac{2}{t}I$

$x_{k+1} = x_k - t\nabla f(x_k)$ ($\because$ fixed step size t)

$= x_k - t \cdot \frac{8}{t} \cdot x_k$

$= -7x_k$

$\Rightarrow$ It's easy to know from $x_{k+1} = -7x_k$ that implies this algorithm doesn't converge.

2. Decreasing Stepsize

step size satisfies $\lim_k t_k = 0$, $\sum_{k=0}^{\infty} t_k = \infty$

① From the upper bound on the Hessian,

We know $f(x_{k+1}) \le f(x_k) + \nabla f(x_k)^T (x_{k+1} - x_k) + \frac{M}{2}\|x_{k+1} - x_k\|_2^2$ (From (9.13) on the textbook)

$(\because x_{k+1} = x_k - t_k \nabla f(x_k))$

$$\Rightarrow f(x_{k+1}) \le f(x_k) - t_k \|\nabla f(x_k)\|_2^2 + \frac{M}{2} t_k^2 \|\nabla f(x_k)\|_2^2$$
$$= f(x_k) + \left(\frac{M}{2}t_k - 1\right) t_k \|\nabla f(x_k)\|_2^2$$
$$= f(x_k) - \left(1 - \frac{M}{2}t_k\right) t_k \|\nabla f(x_k)\|_2^2$$

$\because \lim_k t_k = 0$ $\therefore$ Consider k is large enough s.t. $\left(1 - \frac{M}{2}t_k\right) \le \frac{1}{2}$

$$\Rightarrow f(x_{k+1}) \le f(x_k) - \frac{1}{2} t_k \|\nabla f(x_k)\|_2^2 \quad — ⓐ$$ (let K be the index s.t. k is large enough to satisfy the condition)

② f is strongly convex $\Rightarrow f(y) \ge f(x) + \nabla f(x)^T(y - x) + \frac{m}{2}\|y - x\|^2$ (From (9.8) on the textbook)

Consider $y = x^*$

$$f(x^*) \ge f(x) + \nabla f(x)^T(x^* - x) + \frac{m}{2}\|x^* - x\|^2$$
$$\ge f(x) - \|\nabla f(x)\|_2 \|x^* - x\|_2 + \frac{m}{2}\|x^* - x\|_2^2$$

$\because f(x^*) \le f(x) \Rightarrow -\|\nabla f(x)\|_2 \|x^* - x\|_2 + \frac{m}{2}\|x^* - x\|_2^2 \le 0$

$$\Rightarrow \|x^* - x\|_2 \le \frac{2}{m}\|\nabla f(x)\|_2 \quad — ⓑ$$

Put ⓐ & ⓑ together,

$$f(x_{k+1}) \le f(x_k) - \frac{m t_k}{4}\|x^* - x_k\|_2 \quad \forall k \ge K \quad — ⓒ$$

To prove the gradient descent converges to the global optimal solution.

We want to show that $\lim_{k\to\infty} f(x_k) - f(x^*) = 0$

From ⓒ, $f(x_k) - f(x_{k+1}) \ge \frac{m t_k}{4}\|x^* - x_k\|_2$

For $k \ge K$

$$f(x_k) - f(x^*) \ge f(x_k) - f(x_{k+1}) \quad (\because f(x_{k+1}) \ge f(x^*))$$
$$\ge \frac{m t_k}{4}\|x^* - x_k\|_2$$

Then we want to show as $k \to \infty$, $\|x^* - x_k\|_2 \to 0$

$$\|x^* - x_k\|_2^2 - \|x^* - x_{k+1}\|_2^2 = \|x_k - x^*\|_2^2 - \|x_k - x^* + t_k \nabla f(x_k)\|_2^2$$
$$= 2t_k \nabla f(x_k)^T(x_k - x^*) + t_k^2\|\nabla f(x_k)\|_2^2 \quad (\because f(x^*) \le f(x_k) + \nabla f(x_k)(x^* - x_k))$$
$$\ge 2t_k(f(x_k) - f(x^*)) + t_k^2\|\nabla f(x_k)\|_2^2 \quad (\because f(x_{k+1}) \le f(x_k) - \frac{1}{2}t_k\|\nabla f(x_k)\|_2^2)$$
$$\ge 2t_k(f(x_k) - f(x^*)) + t_k^2 \cdot \frac{2}{t_k}(f(x_{k+1}) - f(x_k))$$
$$= 2t_k(f(x_{k+1}) - f(x^*)) > 0 \quad (\because f(x^*) < f(x_{k+1}))$$

$\because \|x^* - x_k\|_2^2 - \|x^* - x_{k+1}\|_2^2 > 0$

$\therefore$ As k grows, $\|x^* - x_k\|$ will decreases as well.

So, $\lim_{k\to\infty} \|x^* - x_k\| = 0$

We proves that we can find a global optimal solution

Q.E.D.

3. Convex functions

(a) From (1.3) on the textbook, if a function is convex

$$f(\alpha x_1+(1-\alpha)x_2) \le \alpha f(x_1)+(1-\alpha)f(x_2) \text{ for } \alpha \in [0,1]$$

In this question,

$$f(\alpha x_1+(1-\alpha)x_2) = \sup_i f_i(\alpha x_1+(1-\alpha)x_2)$$
$$\le \sup_i (\alpha f_i(x_1)+(1-\alpha)f_i(x_2)) \quad (\because f_i \text{ is convex})$$
$$\le \alpha \sup_i f_i(x_1) + (1-\alpha)\sup_i f_i(x_2) \quad (\because \sup_i \text{ is convex})$$
$$= \alpha f(x_1)+(1-\alpha)f(x_2)$$

$\Rightarrow f(x) := \sup_i f_i(x)$ is convex # Q.E.D.

(b) ① If λ is a eigenvalue of M, it's easy to know that $Mx = \lambda x$ for some x

Consider $M \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^{n\times 1}$, $\lambda \in \mathbb{R}$

$x^T M x = x^T \lambda x = \lambda \cdot \|x\|_2^2$ if we let $\|x\|_2^2 = 1 \Rightarrow x^T M x = \lambda$ for $\|x\|_2^2 = 1$

Let λ_{max} be the function to get the largest eigenvalue of a matrix

$$\lambda_{max}(M) = \max_{\|x\|_2^2=1} x^T M x$$

Then consider

$$\lambda_{max}(\alpha M_1+(1-\alpha)M_2) = \max_{\|x\|_2^2=1} x^T(\alpha M_1+(1-\alpha)M_2)x$$
$$= \max_{\|x\|_2^2=1} (\alpha x^T M_1 x + (1-\alpha)x^T M_2 x)$$
$$\le \alpha \max_{\|x\|_2^2=1} x^T M_1 x + (1-\alpha)\max_{\|x\|_2^2=1} x^T M_2 x \quad (\because \max_{\|x\|_2^2=1} \text{ is convex})$$
$$= \alpha \lambda_{max}(M_1) + (1-\alpha)\lambda_{max}(M_2)$$

$\Rightarrow \lambda_{max}$ is a convex function of the matrix. Q.E.D. #

② However, if we want to find the eigenvalue for largest magnitude

Consider M is negative-definite, that is, $x^T M x < 0$ for all x

Now, To get the eigenvalue for largest "magnitude" will become $\lambda_{min}(M) = \min_{\|x\|_2^2=1} x^T M x$

(Because the minimum eigenvalue has the largest magnitude)

Then, $\lambda_{min}(\alpha M_1+(1-\alpha)M_2)$

$$= \min_{\|x\|_2^2=1} x^T(\alpha M_1+(1-\alpha)M_2)x$$
$$= \min_{\|x\|_2^2=1} \alpha x^T M_1 x + (1-\alpha)x^T M_2 x$$
$$\ge \alpha \min_{\|x\|_2^2=1} x^T M_1 x + (1-\alpha)\min_{\|x\|_2^2=1} x^T M_2 x = \alpha\lambda_{min}(M_1) + (1-\alpha)\lambda_{min}(M_2)$$

$\Rightarrow \lambda_{min}$ isn't convex

Q.E.D. #

3. Convex function

(c) Suppose from a to b we can find a path p

$p = (p_1, p_2, \cdots, p_n)$ if we have n edges from a to b

$$f(\alpha w_1 + (1-\alpha) w_2) = \min_p \sum_{i=1}^{n} (\alpha w_1(p_i) + (1-\alpha) w_2(p_i)) \quad (\because f(w) = \min \sum_{p_i=1}^{n} w(p))$$

$$\geq \min_p \sum_{i=1}^{n} \alpha w_1(p_i) + \min_p \sum_{i=1}^{n} (1-\alpha) w_2(p_i) \quad (\because \min \text{ is concave})$$

$$= \alpha f(w_1) + (1-\alpha) f(w_2)$$

$\Rightarrow f$ is a concave function of w Q.E.D.

#

4. Convex functions : Jensen's Inequality

$f: \mathbb{R}^2 \to \mathbb{R}$, $\text{epi}(f) = \{(x, y) \in \mathbb{R}^{n+1} : y \geq f(x)\}$

(a) Consider $(x_1, y_1), (x_2, y_2) \in \text{epi}(f)$ — ①

$$\alpha y_1 + (1-\alpha) y_2 \geq \alpha f(x_1) + (1-\alpha) f(x_2) \quad (\because ①)$$

$$\geq f(\alpha x_1 + (1-\alpha) x_2) \quad (\because f \text{ is convex})$$

$\Rightarrow$ From this inequality, we know $(\alpha x_1 + (1-\alpha) x_2, \alpha y_1 + (1-\alpha) y_2) \in \text{epi}(f)$

So the line segment between (x_1, y_1) & (x_2, y_2) $\in \text{epi}(f)$

lies in $\text{epi}(f)$ $\Rightarrow \text{epi}(f)$ is convex Q.E.D.

#

(b) If p is a distribution on $\{x_1, \dots, x_m\}$ with weights $p_1, p_2, \cdots, p_m$

Consider $m=2$ first $(\because p_1 + p_2 = 1)$

$$E[f(x)] = p_1 f(x_1) + p_2 f(x_2) = p_1 f(x_1) + (1-p_1) f(x_2) \leq f(p_1 x_1 + (1-p_1) x_2) \quad (\because f \text{ is concave})$$

$$= f(E[x])$$

Suppose $m=k$, $E[f(x)] \leq f(E[x])$

Then consider $m=k+1$, $E[f(x)] = p_1 f(x_1) + p_2 f(x_2) + \cdots + p_k f(x_k) + p_{k+1} f(x_{k+1})$

$$= \sum_{i=1}^{k} p_i f(x_i) + p_{k+1} f(x_{k+1})$$

$$= \sum_{i=1}^{k} p_i \sum_{i=1}^{k} \frac{p_i}{\sum p_i} f(x_i) + p_{k+1} f(x_{k+1})$$

$$\leq \sum_{i=1}^{k} p_i f\left(\sum_{i=1}^{k} \frac{p_i}{\sum p_i} x_i\right) + p_{k+1} f(x_{k+1}) \quad (\because \text{For } m=k, E[f(x)] \leq f(E[x]))$$

(conti.)

4 Jensen's Inequality

(b) conti.

$$E[f(X)] \le \sum_{i=1}^{k} p_i\, f\left(\sum_{i=1}^{k} \frac{p_i}{\sum_{i=1}^{k} p_i} x_i\right) + \left(1-\sum_{i=1}^{k} p_i\right) f(x_{k+1}) \quad (\because p_1+p_2+\cdots+p_k=1)$$

$$\le f\left(\sum_{i=1}^{k} p_i \sum_{i=1}^{k} \frac{p_i}{\sum_{i=1}^{k} p_i} x_i + \left(1-\sum_{i=1}^{k} p_i\right) f(x_{k+1})\right) \quad (\because f \text{ is concave})$$

$$= f\left(\sum_{i=1}^{k+1} p_i x_i\right) = f(E[X])$$

By induction, we show that if f is any concave function,

then $E[f(X)] \le f(E(X))$ #

5. Projection

(a) $\mathcal{X} \in \mathbb{R}^d$ is a closed and bounded convex set. Let $y \in \mathcal{X}$ be any point in $\mathcal{X}$

The projection of y on $\mathcal{X}$ is defined by $\Pi_{\mathcal{X}}(y) = \arg\min_{x \in \mathcal{X}} \|y-x\|_2^2$

Because $y \in \mathcal{X}$, so we can choose $x = y \in \mathcal{X}$ to minimize $\|y-x\|_2^2$

then $\|y-x\|_2^2 = \|y-y\|_2^2 = 0$

However, if we choose $x \ne y$ but $x \in \mathcal{X}$

$\because |y-x| > 0 \Rightarrow \|y-x\|_2^2 > 0$ It means it wouldn't be the solution to $\Pi_{\mathcal{X}}(y)$

So the solution to $\Pi_{\mathcal{X}}(y)$ is $x=y$ & is unique #

(b) If $y \notin \mathcal{X}$ $\Pi_{\mathcal{X}}(y) = \arg\min_{x \in \mathcal{X}} \|y-x\|_2^2$

Then consider $s = y - x$ $b = \frac{\|y\|_2^2 - \|x\|_2^2}{2}$

& A hyperplane $s^T z - b = (y-x)^T(z - \frac{1}{2}(y-x))$

Then consider $z \notin \mathcal{X}$.

$$\Rightarrow \quad s^T z - b = (y-x)^T(z-y) + \frac{1}{2}\|x-y\|_2^2 \quad -①$$

If we suppose $s^T z - b \le 0$ From ① & $\frac{1}{2}\|x-y\|_2^2 \ge 0$ we know $(y-x)^T(z-y) < 0$

Then observe that

$$\frac{d}{dt}\|y+t(z-y)-x\|_2^2\Big|_{t=0} = 2(y-x)^T(z-y) \le 0$$

so for small $t > 0$

$$\|y+t(z-y)-x\|_2^2 \le \|y-x\|_2^2$$

(Continue)

5.(b) conti

That means we can find $y+t(z-y)$ which is closer to $\mathcal{X}$ than y is. — $\otimes$

$\because y \notin \mathcal{X}$ is a closed and bounded convex set $y+t(z-y)$ for $z \notin \mathcal{X}$ $\in (y \notin \mathcal{X})$

However, $\Pi_{\mathcal{X}}(y) = \underset{x \in \mathcal{X}}{\arg\min} \|y-x\|_2^2$ so $\otimes$ is impossible

By contradiction, we know the supposition $s^T z - b \le 0$ for $z \notin \mathcal{X}$ is wrong

So $s^T z - b > 0$ for $z \notin \mathcal{X}$

That means $s^T y - b > 0$ $(\because y \notin \mathcal{X})$

$\Rightarrow \langle s, y \rangle > b$ #

Then if we consider the same s & b

but suppose $-(s^T z - b) < 0$ where $z \in \mathcal{X}$

By the similar contradiction proof, we can show that $s^T z - b \le 0 \Rightarrow \langle s, z \rangle \le b$ #

Q.E.D.

(c) $x^{(k+1)} = \mathrm{Proj}_{\mathcal{X}}(x^{(k)} - t_k \nabla f(x^{(k)}))$

$= \underset{x \in \mathcal{X}}{\arg\min} \|x - (x^{(k)} - t_k \nabla f(x^{(k)}))\|_2^2$

$= \underset{x \in \mathcal{X}}{\arg\min} \|(x - x^{(k)}) + t_k \nabla f(x^{(k)})\|_2^2$

$= \underset{x \in \mathcal{X}}{\arg\min} \left(\|x - x^{(k)}\|_2^2 + 2\langle x - x^{(k)}, t_k \nabla f(x^{(k)}) \rangle + \|t_k \nabla f(x^{(k)})\|_2^2 \right)$

Because x is the only variant here, we can remove those terms which don't contain x

$\Rightarrow \underset{x \in \mathcal{X}}{\arg\min} \left(\|x - x^{(k)}\|_2^2 + 2\langle x, t_k \nabla f(x^{(k)}) \rangle \right)$

we can divide it by $2t_k$ due to t_k isn't a variant for this equation

$\Rightarrow \underset{x \in \mathcal{X}}{\arg\min} \left\{ \langle x, \nabla f(x^{(k)}) \rangle + \frac{1}{2t_k} \|x - x^{(k)}\|_2^2 \right\}$

#

6. Computing Projections

(a) $\mathcal{X} = \mathbb{R}^n_+$ $\qquad z \in \mathbb{R}^n$

$$\text{Proj}_{\mathcal{X}} z = \{\max(z_1, 0), \max(z_2, 0), \ldots, \max(z_n, 0)\} \quad \#$$

(b) Euclidean ball : $\{x : \|x\|_2 \le 1\}$

$z \in \mathbb{R}^n$

If $\|z\|_2 \le 1 \Rightarrow \text{Proj}_{\mathcal{X}} z = z$

Else, if $\|z\|_2 > 1 \Rightarrow \text{Proj}_{\mathcal{X}} z = \frac{z}{\|z\|_2}$ $\qquad \Rightarrow \text{Proj}_{\mathcal{X}} z = \frac{z}{\max\{1, \|z\|_2\}}$ #

(c) Positive semidefinite cone : $S_+^n = \{M \in S^n : x^T M x \ge 0, \forall x \in \mathbb{R}^n\}$.

To project a matrix Z onto S_+^n, we must assume Z is symmetric

By SVD, $Z = U \Sigma V^T$ $\qquad \because Z^T = Z$

$\therefore Z^T = V \Sigma^T U^T = V \Sigma U^T$

$\Rightarrow U^T = V^T \Rightarrow U = V \Rightarrow Z = U \Sigma U^T$

Without loss of generality, we can use Frobenius norm to be the distance between two matrices.

$$\text{Proj}_{\mathcal{X}} Z = \arg\min_{X \in S_+^n} \|X - Z\|_F$$

$$= \arg\min_{X \in S_+^n} \|U^T\|_F \|X - Z\|_F \|U\| \quad (\because U \text{ is unitary } \therefore \|U^T\| = \|U\| = 1)$$

$$= \arg\min_{X \in S_+^n} \|U^T (X - Z) U\|_F$$

$$= \arg\min_{X \in S_+^n} \|U^T X U - U^T Z U\|_F$$

$$= \arg\min_{X \in S_+^n} \|U^T X U - \Sigma\|_F \quad (\because Z = U \Sigma U^T \text{ \& } U \text{ is unitary})$$

To minimize it, we can find $X \in S_+^n$ s.t $X = U \Sigma_+ U^T$, Σ_+ is a diagonal matrix with nonnegative eigenvalues ($\because X \in S_+^n$)

$$\Rightarrow \text{Proj}_{\mathcal{X}} Z = \arg\min_{X \in S_+^n} \|\Sigma_+ - \Sigma\|_F$$

$$= X$$

$$= U \Sigma_+ U^T \quad \#$$

(U & U^T are same as the SVD of Z, Σ_+ is a diagonal matrix with nonnegative eigenvalues)

6. Computing Projections

(d) $\mathcal{X} = \{x : L_i \le x_i \le U_i, i=1,2,\cdots,n\}$

Consider $z \in \mathbb{R}^n$, after projection, each z_i should satisfy $L_i \le z_i \le U_i$ for $i=1,2,\cdots,n$

$\Rightarrow$ The projection on each z_i can be expressed as $\max(\min(z_i, U_i), L_i)$

$\Rightarrow \mathrm{Proj}_{\mathcal{X}} z = (\max(\min(z_1,U_1),L_1), \max(\min(z_2,U_2),L_2), \cdots, \max(\min(z_n,U_n)L_n))$ #

(e) 1-norm ball: $\{x : \sum_i |x_i| \le 1\}$

If $z \in \mathcal{X}$, that is, $\sum_i |z_i| \le 1 \Rightarrow \mathrm{Proj}_{\mathcal{X}} z = z$

But for $z \notin \mathcal{X}$, that is, $\|z\|_1 > 1$, the projection $x = \mathrm{Proj}_{\mathcal{X}} z$ should lie on the boundary $\{x : \sum_i |x_i| = 1\}$

So we want to $x = \mathrm{Proj}_{\mathcal{X}} z$ s.t. $\min_x \|x - z\|_2^2$ where $\sum_i |x_i| = 1$

Then we claim that $x_i z_i \ge 0$ when x is a optimal solution

We can suppose there exists i for which $x_i z_i < 0$

Let $\hat{x}$ be a vector s.t. $\hat{x}_i = 0$ & for all $i \ne j$ we have $\hat{x}_j = x_j$

$\|\hat{x}\|_1 = \|x\|_1 - |x_i| \le 1$ s.t. $\hat{x}$ is a feasible solution

However,

$\|x - z\|_2^2 - \|\hat{x} - z\|_2^2 = (x_i - z_i)^2 - (0 - z_i)^2$

$= x_i^2 - 2x_i z_i \ge x_i^2 \ge 0$ ($\because$ We suppose $x_i z_i < 0$)

This means that $\hat{x}$ is a solution more optimal than $x \Rightarrow$ Contradict with our claim

Hence, we make sure that $x_i z_i \ge 0$ when x is a optimal solution — ①

Then we consider the result from 6 (f) Probability Simplex

we know that when $\mathcal{X} = \{x : \sum_i x_i = 1, x_i \ge 0, i=1,\cdots,n\}$

$\mathrm{Proj}_{\mathcal{X}} z = \{z_1 - u, z_2 - u, \cdots, z_n - u\}$ where u is chosen for every $z_i - u \ge 0$

& $\|\mathrm{Proj}_{\mathcal{X}} z\|_1 = 1$ — ②

By combining ① & ②, we can get

$\mathrm{Proj}_{\mathcal{X}} z = \{\mathrm{sign}(z_1)(|z_1| - u), \mathrm{sign}(z_2)(|z_2| - u), \cdots, \mathrm{sign}(z_n)(|z_n| - u)\}$

where λ is chosen s.t. $\|\mathrm{Proj}_{\mathcal{X}} z\|_1 = 1$ for $z \notin \mathcal{X}$

$\mathrm{Proj}_{\mathcal{X}} z = z$ for $z \in \mathcal{X}$

#

6 Computing Projections

(f) Probability simplex : $\mathcal{X} = \{x : \sum_i x_i = 1, x_i \geq 0, i = 1, \cdots, n\}$

To compute the projection, we can consider it as an optimization problem

$\min \frac{1}{2}\|z - x\|_2^2$ s.t. $\sum_{i=1}^{n} x_i = 1$, $x_i \geq 0$, $i = 1, \cdots, n$

Then the partial Lagrangian formed by dualizing the constraint $\sum_{i=1}^{n} x_i = 1$ with the implicit constraint $x_i \geq 0$ for $i = 1, \cdots, n$

We can write it as

$$L(x, u) = \frac{1}{2}\|z - x\|_2^2 + u\left(\sum_{i=1}^{n} x_i - 1\right)$$ with the implicit constraint $x_i \geq 0$ for $i = 1, \cdots, n$

So $L(x, u)$ can rewritten as

$$L(x, u, a) = \frac{1}{2}\|z - x\|_2^2 + u\left(\sum_{i=1}^{n} x_i - 1\right) - \sum_{i=1}^{n} a_i x_i$$ where $a_i \geq 0$ for $i = 1, 2, \cdots, n$

At the optimal solution, the following KKT conditions hold:

$$\begin{cases} x_i - z_i + u - a_i = 0, & i = 1, \cdots, n \\ x_i \geq 0, & i = 1, \cdots, n \\ a_i \geq 0, & i = 1, \cdots, n \\ a_i x_i = 0, & i = 1, \cdots, n \\ \sum_{i=1}^{n} x_i = 1 & \end{cases}$$

Hence, if $x_i > 0$ $\because a_i x_i = 0$ $\therefore a_i = 0$ So $x_i - z_i + u = 0$

$x_i = z_i - u > 0$

if $x_i = 0$ $\Rightarrow a_i \geq 0$ So $x_i - z_i - a_i + u = 0 \Rightarrow x_i = z_i + a_i - u = 0$

$\Rightarrow z_i - u = -a_i \leq 0$

So $x_i \in \{z_i - u, 0\}$

By this, we can write

$$\mathrm{Proj}_{\mathcal{X}} z = \{z_1 - u, z_2 - u, \cdots, z_n - u\}$$

where u is chosen for every $z_i - u \geq 0$ for $i = 1, 2, \cdots, n$

$\& \|\mathrm{Proj}_{\mathcal{X}} z\|_1 = 1$

#