Leung Ko Tsun 20516087

1.01.a norm should fulfill 3 conditions,

① $\|x\|_M = \max\{\|x\|_A, \|x\|_B\} \geq \|x\|_A \geq 0$, $\forall x \in V$

$\geq \|x\|_B \geq 0$ $\forall x \in V$.

and $\|x\|_A = 0 \Leftrightarrow x=0$

$\|x\|_B = 0 \Leftrightarrow x=0$

$\therefore \max\{\|x\|_A, \|x\|_B\} = 0$ if $x=0$.

$\therefore \|x\|_M = 0 \Leftrightarrow x=0$.

② $\cancel{\|\alpha x\| = |\alpha|}$

$\|\alpha x\|_M = \max\{\|\alpha x\|_A, \|\alpha x\|_B\}$

$= \max\{|\alpha|\,\|x\|_A, |\alpha|\,\|x\|_B\}$

$= |\alpha| \max\{\|x\|_A, \|x\|_B\}$

$= |\alpha|\,\|x\|_M$

③. $\|x+y\|_M = \max\{\|x+y\|_A, \|x+y\|_B\}$

$\cancel{\leq \max\{\|x\|_A+\|y\|_A, \|x\|_B+\|y\|_B\}}$

$\cancel{\leq \max\{\|x\|_M, \|y\|_M}$

$\leq \|x\|_M + \|y\|_M$

(b) when ~~$\|x\|_A = \left(\int f(t)g(t)\right)^{\frac{1}{2}}$~~

~~$\|x\|_B = \|x\|$~~

$\min\{\|x\|_A, \|x\|_B\}$ does not ~~define~~ a norm on V.

when $\|x\|_C = \dfrac{2\|x\|_\infty + \|x\|_1}{3}$

and $\|x\|_2$,

after calculation we find it does not define a norm.

$\|x+\alpha y\|^2 = \langle x+\alpha y, x+\alpha y\rangle$

$= \|x\|^2 + \alpha\langle y,x\rangle + \alpha\langle x,y\rangle + |\alpha|^2\|y\|^2$

2a). Claim(*):
x, y are orthogonal if and only if $\|x+\alpha y\| = \|x-\alpha y\|$ $\forall \alpha \in \mathbb{R}$.

Proving "if", if $\|x+\alpha y\| = \|x-\alpha y\|$, x and y are orthogonal.

$$\|x+\alpha y\|^2 - \|x-\alpha y\|^2 = 0$$

$$\|x\|^2 + 2\langle x,\alpha y\rangle + \|\alpha y\|^2 - (\|x\|^2 - 2\langle x,\alpha y\rangle + \|\alpha y\|^2) = 0$$

$$4\alpha\langle x,y\rangle = 0$$

$$\langle x,y\rangle = 0.$$

$\Rightarrow$ x and y are orthogonal.

Proving "only if", if $\langle x,y\rangle = 0$, $\|x+\alpha y\| = \|x-\alpha y\|$.

Similarly, producing $\|x\|^2 - \|x\|^2 + \|\alpha y\|^2 - \|\alpha y\|^2 + 4\langle x,\alpha y\rangle = 0$,

we have

$$\|x+\alpha y\|^2 = \|x-\alpha y\|^2.$$

$$\|x+\alpha y\| = \|x-\alpha y\|.$$

So, putting $\alpha = 1$ in (*), we get $\|x+y\| = \|x-y\|$

Thus the statement is proved.

2b). Claim (**): x and y are orthogonal if and only if $\|x+\alpha y\| \geq \|x\| \quad \forall \alpha \in \mathbb{R}$.
First suppose $x \perp y$. Then $\langle x,y \rangle = 0$, $\forall \alpha \in \mathbb{R}$,

$$\begin{aligned}
\|x+\alpha y\|^2 &= \langle x+\alpha y, x+\alpha y \rangle \\
&= \|x\|^2 + \alpha\langle y,x \rangle + \alpha\langle x,y \rangle + |\alpha|^2\|y\|^2 \\
&\geq \|x\|^2
\end{aligned}$$

$\therefore \|x\| \leq \|x+\alpha y\|$.

Now suppose $\|x\| \leq \|x+\alpha y\| \ \forall \alpha \in \mathbb{R}$, Assume $\langle y,x \rangle \neq 0$ and let $\alpha = -\frac{\langle x,y \rangle}{\|y\|^2}$

$$\begin{aligned}
\|x+\alpha y\| &= \langle x+\alpha y, x+\alpha y \rangle \\
&= \|x\|^2 + \alpha\langle x,y \rangle + \alpha(\langle y,x \rangle + \alpha\|y\|^2) \\
&= \|x\|^2 - \frac{|\langle x,y \rangle|^2}{\|y\|^2} \\
&< \|x\|^2
\end{aligned}$$

Since $\frac{|\langle x,y \rangle|^2}{\|y\|^2} > 0$, contradiction occurs, proving the fact that $\|x\| \leq \|x+\alpha y\| \ \forall \alpha \in \mathbb{R}$.

~~Part~~ Since α is arbitrary, we put $-\alpha$ into (**) and get:

$\|x-\alpha y\| \geq \|x\| \quad \forall \alpha \in \mathbb{R}$ iff ~~x, y are~~ $x \perp y$.

~~3. We first prove that $\langle y-z, x\rangle = 0 \Rightarrow P_S(y) = z$. For any $x \in S$, we have~~

$$\cancel{\|x-y\| = \|x-z+z-y\|}$$

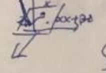
$$=$$

3. We first prove that if $z \in S$ is a solution of $\min_{x\in S} \|x-y\|$, then

$$\langle y-z, x\rangle = 0 \quad \forall x \in S.$$

Since z is a solution, $z \in S$, i.e. $x, z \in S$ implies $\alpha x + \beta z \in S$. $\forall \alpha, \beta \in \mathbb{R}$.

We know that $x, z \in S$, and the span of these two vectors also lie on same plane. $tx + (1-t)z \in S$. $\forall t \in \mathbb{R}$



Since V is a Hilbert space, from Riesz representation theorem, we know that any linear function ~~is~~ can be written as $\langle a, x\rangle$ for $a \in V$.

and $u, v \in S$ implies $\alpha u + \beta v \in S$. S is a ~~plane~~ hyperplane

Now consider the set $\{x \in S \mid \langle a, x\rangle = b\} \equiv S_{a,b}$,

then, $\forall x \in S$ and $t \in \mathbb{R}$,

$$\langle a, (1+t)z - tx\rangle = \cancel{(1+t)}\ (1+t)\langle a, z\rangle - t\langle a, x\rangle = b$$

$$\Rightarrow (1+t)z - tx \in S.$$

Since z is the closest to y on S,

$$\|z-y\|^2 \le \|(1+t)z - tx - y\|^2 = \|z-y+t(z-x)\|^2$$

$$= \|y-z\|^2 + t^2\|z-x\|^2 + 2t\langle z-y, z-x\rangle$$

$$= \|y-z\|^2 + t^2\|z-x\|^2 + 2t\langle y-z, x\rangle$$

$$\Rightarrow t\langle y-z, x\rangle \ge -\frac{t^2}{2}\|z-x\|$$

if we choose $t > 0$,

$$\langle y-z, x\rangle \ge -\frac{t}{2}\|z-x\|^2$$

let $t \to 0^+ \Rightarrow \langle y-z, x\rangle \ge 0 \quad \forall x \in S.$

if we choose $t < 0$,

$$\langle y-z, x\rangle \le -\frac{t}{2}\|z-x\|^2$$

let $t \to 0^- \Rightarrow \langle y-z, x\rangle \le 0$

3. cont'd). So, $\langle y-z, x\rangle = 0 \quad \forall x \in S$.

We then show: If $z \in S$, and $\langle y-z, x\rangle = 0$, $\forall x \in S$,

then $z = \underset{x \in S}{\operatorname{argmin}} \|x-y\|$

By direct calculation:

$\forall x \in S$,

$$\begin{aligned}
\|x-y\|^2 &= \| \cancel{z-x-(z-x)-(z-y)} \|^2 \\
&= \| x+z-z-y\|^2 \\
&= \| -(y-z)-(z-x)\|^2 \\
&= \| z-y\|^2 + \|z-x\|^2 - 2\langle z-x, z-y\rangle \\
&= \|z-y\|^2 + \|z-x\|^2 - 2\langle y-z, x\rangle \\
&\geq \|z-y\|^2
\end{aligned}$$

This together with $z \in S$ implies $z = \underset{x \in S}{\operatorname{argmin}} \|x-y\|$ //

Fix $x \in \mathbb{R}^n$

4a). $\nabla f(x) = \frac{\partial \sigma(a^Tx+b)}{\partial a^Tx+b} \frac{\partial a^Tx+b}{\partial x}$ (chain rule).

$= \sigma'(a^Tx+b)(a)$

$= a\sigma'(a^Tx+b)$.

b). $\frac{\partial}{\partial x} \sum_{i=1}^{m} (a_i^Tx)^4$

$= \sum_{i=1}^{m} \frac{\partial}{\partial x}(a_i^Tx)^4$

$= \sum_{i=1}^{m} \frac{\partial (a_i^Tx)^4}{\partial a_i^Tx} \frac{\partial a_i^Tx}{\partial x}$

$= 4 \sum_{i=1}^{m} (a_i^Tx)^3 (a_i)$

First, we prove $|x|$ is convex. It is ...

5. Given $S := \{ z \mid f(z) = \min_{x \in \mathbb{R}^n} f(x) \}$.

$$f(z) \le f(x) \quad \forall x \in \mathbb{R}^n$$

for any $u, v \in S$, $f(u) \le f(x)\ \forall x \in \mathbb{R}^n$, $f(v) \le f(x)\ \forall x \in \mathbb{R}^n$.

$$\begin{aligned} f(tu + (1-t)v) &\le t f(u) + (1-t) f(v) \quad \text{(Since } f \text{ is convex function)} \\ &\le t f(x) + (1-t) f(x) \\ &= f(x). \end{aligned}$$

So, we proved that $f(tu + (1-t)v) \le f(x) \quad \forall x \in \mathbb{R}^n$,

$\therefore$ $tu + (1-t)v$ is also in S.

$\therefore$ S is a convex set.

First, we prove $|x| =$

6. $K(x,y) = (x^Ty+1)^2$

$= (x^Ty)^2 + 2(x^Ty) + 1$ $\quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$

$= (x_1y_1 + x_2y_2)^2 + 2(x_1y_1 + x_2y_2) + 1$ $\quad x^T = [x_1\ x_2]$

$= (x_1y_1)^2 + 2x_1y_1x_2y_2 + (x_2y_2)^2 + 2x_1y_1 + 2x_2y_2 + 1$

So, ~~we could~~ $\langle \phi(x), \phi(y) \rangle$ should be $K(x,y)$, and from term by term comparison, we have $\phi(x) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \\ \sqrt{2}x_1x_2 \\ 1 \end{bmatrix}$

and so H is $\mathbb{R}^6$,

$\langle x,y \rangle_H = x^Ty$

$\phi(x) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \\ \sqrt{2}x_1x_2 \\ 1 \end{bmatrix}$

~~7.1~~

7. We use forward backward splitting method:

Let $g(x) = \frac{1}{2}\|y-x\|_2^2$, given $y \in \mathbb{R}^n$.

$f(x) = \lambda\left(\frac{1}{2}\|x\|_2^2 + \mu\|x\|_1\right)$

Since $f(x)$ is convex but not smooth, we use backward sub-gradient.

$$x^{(k+1)} = x^{(k)} - \alpha_k\left(\nabla g(x^{(k)}) + \lambda u^{(k+1)}\right), \quad u^{(k+1)} \in \partial f(x^{(k+1)})$$

$$\Updownarrow$$

$$x^{(k+1)} \in x^{(k)} - \alpha_k\left(\nabla g(x^{(k)}) + \lambda \partial f(x^{(k+1)})\right)$$

$$\Updownarrow$$

$$x^{(k)} - \alpha_k \nabla g(x^{(k)}) \in x^{(k+1)} + \alpha_k \lambda \partial f(x^{(k+1)})$$

$$\Updownarrow$$

$$0 \in x^{(k+1)} - [x^{(k)} - \alpha_k \nabla g(x^{(k)})] + \alpha_k \lambda \partial f(x^{(k+1)})$$

$$\Updownarrow$$

$$0 \in \partial\left(\frac{1}{2}\| x - [x^{(k)} - \alpha_k \nabla g(x^{(k)})]\|_2^2 + \alpha_k \lambda f(x)\right)\Big|_{x = x^{(k+1)}}$$

$$x^{(k+1)} = \underset{x \in \mathbb{R}^n}{\arg\min}\ \frac{1}{2}\| x - [x^{(k)} - \alpha_k \nabla g(x^{(k)})]\|_2^2 + \alpha_k \lambda f(x))$$

So, FBS is written as

$$\begin{cases} z^{(k+1)} = x^{(k)} - \alpha_k \nabla g(x^{(k)}) \\ x^{(k+1)} = \underset{x \in \mathbb{R}^n}{\arg\min}\ \frac{1}{2}\| x - z^{(k+1)}\|_2^2 + \alpha_k \lambda f(x). \end{cases}$$

where $\nabla g(x^{(k)}) =$

$$\text{FBS} = \begin{cases} z^{(k+1)} = x^{(k)} - \alpha_k (x^{(k+1)} - y) \\ x^{(k+1)} = \underset{x \in \mathbb{R}^n}{\arg\min}\ \frac{1}{2}\| x - z^{(k+1)}\|_2^2 + \alpha_k \lambda f(x). \end{cases}$$

Q7 (cont'd),

$$x^{(k+1)} = \arg\min_{x\in\mathbb{R}^n} \frac{1}{2}\|x - z^{(k+1)}\|_2^2 + \alpha_k\lambda\left(\frac{1}{2}\|x\|_2^2 + \mu\|x\|_1\right)$$

$\Updownarrow$ Fermat' lemma

$$0 \in x^{(k+1)} - z^{(k+1)} + \alpha_k\lambda\left(x^{(k+1)} + \mu\,\partial\|x^{(k+1)}\|_1\right)$$

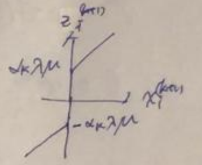
$\Updownarrow$

$$0 \in x_i^{(k+1)} - z_i^{(k+1)} + \alpha_k\lambda x_i^{(k+1)} + \alpha_k\lambda\mu\,\partial|x_i^{(k+1)}|, \quad i=1,2,\dots,n$$

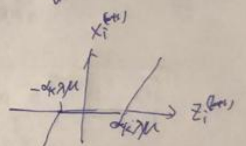
$\Updownarrow$

$$z_i^{(k+1)} \in (1+\alpha_k\lambda)\,x_i^{(k+1)} + \alpha_k\lambda\mu\,\partial|x_i^{(k+1)}|$$

where $\partial|x_i^{(k+1)}| = \begin{cases}\{-1\} & <0\\ [-1,1] & =0\\ \{1\} & >0\end{cases}$



$\Updownarrow$ switch $x_i \longleftrightarrow z_i$



$$x_i^{(k+1)} = \begin{cases} z_i^{(k+1)} + \alpha_k\lambda\mu & \\ 0 & \text{if} \\ z_i^{(k+1)} - \alpha_k\lambda\mu & \end{cases} \begin{matrix} z_i^{(k+1)} \le -\alpha_k\lambda\mu \\ |z_i^{(k+1)}| \le \alpha_k\lambda\mu \\ z_i^{(k+1)} \geq \alpha_k\lambda\mu \end{matrix}$$

$$|||$$

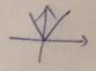
$$T_{\alpha_k\lambda\mu}(z_i^{(k+1)}) \qquad i=1,\dots,n.$$

$$\therefore \text{FBS}: \begin{cases} z^{(k+1)} = x^{(k)} - \alpha_k(x^{(k+1)} - y) \\ x^{(k+1)} = T_{\alpha_k\lambda\mu}(z^{(k+1)}) \end{cases}$$

8.1.

~~2 at.~~

8.(a). First, we prove $|x|$ is convex. It is obvious since the graph of $|x|$ is convex.



Second, from the lecture notes, we know that an affine function is convex. ~~shown here~~:

Thirdly, from the lecture notes, we know that $\|x\|_2^2$ is convex.

$$\|\alpha x+(1-\alpha)y\|_2^2 \le \alpha\|x\|_2^2+(1-\alpha)\|y\|_2^2 = \alpha\|x\|_2^2+(1-\alpha)\|y\|_2^2$$

$\therefore$ we know that $f\circ g(x)$ is convex if $f(x)$ and $g(x)$ are convex function.

Moreover, $f(x)+g(x)$ is also convex if $f(x)$ & $g(x)$ are convex,

So, F is convex from $\mathbb{R}^n \to \mathbb{R}$.

8 b) $\partial F(x) = \sum_{i=1}^{N} \partial |a_i^T x - b| + 2\lambda x$

$= \sum_{i=1}^{N} \partial |\langle a_i, x\rangle - b| + 2\lambda x$

where $\partial |\langle a_i, x\rangle - b|$ is: $\begin{cases} \{-a_i\} & \text{if } \langle a_i, x\rangle < b \\ [-a_i, a_i] & \text{if } \langle a_i, x\rangle = b \\ \{a_i\} & \text{if } \langle a_i, x\rangle > b \end{cases}$

And according to Fermat's lemma:

$x^{(*)} = \underset{x \in \mathbb{R}^n}{\arg\min}\ F(x) \iff 0 \in \partial F(x^{(*)})$

$\therefore$ optimality condition should be:

$0 \in \sum_{i=1}^{N} \partial |\langle a_i, x\rangle - b| + 2\lambda x,$

~~and we can~~

$-2\lambda x \in \sum_{i=1}^{N} \partial |\langle a_i, x\rangle - b|$

$\therefore [2\lambda x]_k \in [-a_i, a_i]_k$

if $|[2\lambda x]_i| < a_i$, then $[x^{(*)}]_i = 0$.

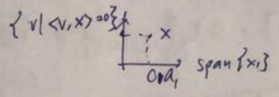
8 c). Similar to the proof of representer theorem

claim: For any $x \in \mathbb{R}^n$, x can be decomposed as $x = a_s + \sum_{i=1}^{N} c_i a_i$,

with $c = [c_1, c_2, \dots, c_N]^T \in \mathbb{R}^N$

~~then~~,

proof: consider $n=1$ for simplicity:



~~$S = \{v | \langle v, x\rangle = 0\}$ is a hyperplane~~. $S = \{v | \langle v, x \rangle = 0\}$

For any $x \in \mathbb{R}^N$,

$x = P_S a + (x - P_S a)$

where $P_S a$ is the projection of a onto S

by the property of projection,

$\langle P_S a - x, P_S a - 0 \rangle = 0$

$\langle P_S a - x, P_S a \rangle = 0$

$x - P_S a = c_1 a_1$

$x = P_S a + c_1 a_1$

$x = a_s + c_1 a_1$

$\Rightarrow$ we can generalize it to N, and get the result that

x can be decomposed into $x = a_s + \sum_{i=1}^{N} c_i a_i$.

Then, the objective function becomes

$$F(x) = \sum_{i=1}^{N} |\langle a_i, x\rangle - b| + \lambda\|x\|_2^2$$

$$\cancel{= \sum_{i=1}^{N} |\langle a_i, \sum_{j=1}^{N} a_j c_j \rangle}$$

$$= \sum_{i=1}^{N} |\langle a_s + \sum_{j=1}^{N} c_j a_j, a_i\rangle - b| + \lambda\|a_s\|_2^2 + \lambda\|\sum_{i=1}^{N} c_i a_i\|_2^2$$

$$= \underbrace{\sum_{i=1}^{N}\sum_{j=1}^{N} |c_j\langle a_j, a_i\rangle - b| + \lambda\|\sum_{i=1}^{N} c_i a_i\|_2^2}_{\text{depends on } c = \begin{bmatrix} c_1 \\ \vdots \\ c_N \end{bmatrix},\ F_1(c)} + \underbrace{\lambda\|a_s\|_H^2}_{\text{depends on } a_s \text{ only},\ F_2(a_s)}$$

So, $F(x) = F_1(c) + F_2(a_s)$

$$\min_{x\in\mathbb{R}^N} F(x) \iff \min_{\substack{c\in\mathbb{R}^N \\ a_s\in\mathbb{R}^N \\ \langle a_s, x_i\rangle = 0 \\ i=1\ldots N}} F_1(c) + F_2(a_s)$$

$$\iff \min_{c\in\mathbb{R}^N} F_1(c),\ \min_{\substack{a_s\in\mathbb{R}^N \\ \langle a_s, x_i\rangle = 0 \\ i=1,\ldots N}} F_2(a_s)$$

→ the minimizer is obviously $a_s = 0$.

So, $\min_{x\in\mathbb{R}^N} F(x) \iff \min_{c\in\mathbb{R}^N} F_1(c)$

and the solution of $\min_{x\in\mathbb{R}^N} F(x)$ is

$$x^{(*)} = 0 + \sum_{i=1}^{N} c_i a_i \quad \text{where}$$

c is the solution of $\min_{c\in\mathbb{R}^N} F_1(c)$.

Then, the objective function becomes

$F(x) = \sum^{N}$ ...

d). As finished in part (c), ~~F(x)~~.

$$\min_{x \in \mathbb{R}^n} F(x) \Longleftrightarrow \min_{c \in \mathbb{R}^N} F_1(c).$$

where $F_1(c)$ is

$$= \sum_{i=1}^{N} \left\{ \left(\sum_{j=1}^{N} |c_j \langle a_j, a_i \rangle - b| \right) + \lambda \| \sum_{i=1}^{N} c_i a_i \|_2^2 \right\}$$

e). $c = \arg\min_{c \in \mathbb{R}^N} F_1(c)$.

$\Updownarrow$ Fermat' Lemma

$$0 \in \sum_{i=1}^{N} \left\{ \left(\sum_{j=1}^{N} \partial |c_j \langle a_j, a_i \rangle - b| \right) - 2\lambda \sum_{i=1}^{N} c_i a_i \right\}$$

$$2\lambda \sum_{i=1}^{N} c_i a_i \in \sum_{i=1}^{N} \left(\sum_{j=1}^{N} \partial | c_j \langle a_j, a_i \rangle - b| \right) \quad -(*).$$

~~where~~ $\partial g(x)$ where $g(x) = |x|$

$$\partial g(x) : \begin{cases} \{-1\} & \text{if } x < 0 \\ [-1, 1] & \text{if } x = 0 \\ \{1\} & \text{if } x > 0. \end{cases}$$

$$\partial |c_j \langle a_j, a_i \rangle - b|$$

$$= \begin{cases} \{-\langle a_j, a_i \rangle\} & \text{if } c_j \langle a_j, a_i \rangle < b \\ [-\langle a_j, a_i \rangle, \langle a_j, a_i \rangle] & \text{if } c_j \langle a_j, a_i \rangle = b \\ \{\langle a_j, a_i \rangle\} & \text{if } c_j \langle a_j, a_i \rangle > b \end{cases}$$

So. $(*)$ gives the optimality condition that $c^{(*)}$ should satisfy.