Convex optimization
HW-02
Rifat Bin Rashid
ID: 237000174

Q1

$\boxed{\text{problem 12.1}}$

let, $g(t) = \|a + tb\|_2^2$

$= (a+tb)^T \cdot (a + tb)$

$= (a^T + b^T t)(a + tb)$

$= a^T a + a^T tb + tb^T a + t^2 b^T b$

as,
$x^T x = \|x\|_2^2$
$(M+N)^T = M^T + N^T$
$(MN)^T = N^T M^T$

(t scaler)

$= \|a\|_2^2 + 2t(a^T b) + t^2 \|b\|_2^2$

$[a^T b = b^T a]$

at $\inf_t g(t)$, $\frac{d}{dt} g(t) = 0$

$\Rightarrow 2a^T b + 2t \|b\|_2^2 = 0$

$t = -\frac{a^T b}{\|b\|^2}$

so, $\inf_t g(t) = \|a\|_2^2 - \frac{a^T b}{\|b\|_2^2}$

Now, $\inf_t g(t) \geq 0$

So, $\|a\|_2^2 - \dfrac{(a^Tb)^2}{\|b\|_2^2} \geq 0$

$\Rightarrow \|a\|_2^2 \geq \dfrac{(a^Tb)^2}{\|b\|_2^2}$

$$\boxed{\therefore |a^Tb| \leq \|a\|_2 \|b\|_2}$$

(b) let, $a = [1 \;\; 1 \;\; 1 \;\cdots\; 1]^T \in \mathbb{R}^n$

So, $\|x\|_1 = \sum_{k=1}^{n} |x_k| = |a^Tx|$

$$= \left| [1 \;\; 1 \;\; 1 \cdots 1] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right|$$

Now, $|a^Tx| \leq \|a\|_2 \|x\|_2$

So, $\|x\|_1 \leq \|a\|_2 \|x\|_2$

Now $\|a\|_2 = \sqrt{1+1+\cdots+1} = \sqrt{n}$

So, $\boxed{\|x\|_1 \leq \sqrt{n} \|x\|_2}$

(c)

let,
$$\underline{y} = (\sqrt{x_1} \quad \sqrt{x_2} \quad \cdots \quad \sqrt{x_n})^T$$
$$\underline{y}_h = \left(\frac{1}{\sqrt{x_1}}, \frac{1}{\sqrt{x_2}} \quad \cdots \quad \frac{1}{\sqrt{x_n}} \right)^T$$

Now, $\left| \underline{y}^T \underline{y}_h \right| \leq \|\underline{y}\|_2^2 \; \|\underline{y}_h\|_2^2$

$$\Rightarrow (1 + 1 + \cdots + 1) \leq \left((\sqrt{x_1})^2 + (\sqrt{x_2})^2 + \cdot (\sqrt{x_n})^2 \right) \cdot \left(\frac{1}{x_1} + \cdots + \frac{1}{x_n} \right)$$

$$\Rightarrow n \leq (x_1 + x_2 + \cdots + x_n) \left(\frac{1}{x_1} + \frac{1}{x_2} + \cdots \frac{1}{x_n} \right)$$

$$\Rightarrow n \leq \sum_{k=1}^{n} x_k \cdot \sum_{k=1}^{n} \frac{1}{x_k}$$

$$\Rightarrow 1 \leq \left(\frac{1}{n} \sum_{k=1}^{n} x_k \right) \cdot \left(\frac{1}{n} \sum_{k=1}^{n} \frac{1}{x_k} \right)$$

$$\Rightarrow 1 \leq (AM) \cdot (HM)^{-1}$$

$$\boxed{\therefore AM \geq HM}$$

AM = Arithmetic mean
HM = Harmonic mean

Q2 $\boxed{\text{Problem 13.2 (b)}}$

If $Z = X \cdot Y$ where $X \in \mathbb{R}^{m \times n}$, $Y \in \mathbb{R}^{n \times p}$

each entry of Z is calculated as,

$$Z_{ij} = \sum_{k=1}^{n} X_{ik} Y_{kj}$$

This needs n multiplication & $(n-1)$ addition

So, for each entry in Z, # flop count $= n + n - 1 = 2n - 1$

for $m \times p = mp$ entry in Z # total flop count $= mp(2n-1)$

Now, Matrix multiplications are associative,

meaning $(AB)C = A(BC)$

if, $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, $C \in \mathbb{R}^{p \times q}$

flop count

for $(AB)C$	for $A(BC)$
$\Rightarrow mp(2n-1) + mq(2p-1)$	$= nq(2p-1) + mq(2n-1)$

which are not same!

So, the statement is $\boxed{\text{False}}$

Scanned with CamScanner

Q3

Given, $f(x) = \|Ax - b\|_2^2$

where $A \in \mathbb{R}^{k \times n}$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^k$

So, $f(x) = (Ax-b)^T (Ax-b)$

$= [(Ax)^T - b^T] \cdot (Ax-b)$

$= [x^T A^T - b^T] (Ax-b)$

$(M+N)^T = M^T + N^T$

$(MN)^T = N^T M^T$

$f(x) = x^T A^T Ax - x^T A^T b - b^T Ax + b^T b$

Here, $x^T A^T b = b^T Ax$ (both scaler)

So,

$\cancel{Df(x) = 2A^T A}$

$f(x) = x^T A^T Ax - 2b^T Ax + b^T b$

$\therefore Df(x) = 2x^T A^T A - 2b^T A$

$\therefore \nabla f(x) = [Df(x)]^T$

$= [2x^T A^T A - 2b^T A]^T$

$= 2A^T Ax - 2A^T b = \boxed{2A^T(Ax-b)}$

Q4

From chain rule

if $h(x) = g(f(x))$; $x \in \mathbb{R}^n$

$$Dh(x) = D_{f(x)}[g(f(x))] \cdot D_x[f(x)]$$

Given,

$$f(x) = \log\left(\sum_{i=1}^{m} e^{a_i^T x}\right)$$

Let,
$$s = \sum_{i=1}^{m} e^{a_i^T x}$$

So, $f(x) = \log s$

So,
$$Df(x) = \frac{1}{s} D_x[s]$$

Now
$$D_x[s] = D_x\left(\sum_{i=1}^{m} e^{a_i^T x}\right)$$

$$= \sum_{i=1}^{m} e^{a_i^T x} \cdot a_i^T$$

So, $$Df(x) = \frac{1}{\sum_{i=1}^{m} e^{a_i^T x}} \sum_{i=1}^{m} e^{a_i^T x} . a_i^T$$

And, $$\nabla f(x) = [D f(x)]^T$$

$$= \frac{1}{\sum_{i=1}^{m} e^{a_i^T x}} \sum_{i=1}^{m} e^{a_i^T x} . a_i$$

$\boxed{Q5}$

Given, $f(x) = -\sum_{i=1}^{n} x_i \log x_i$

$= -x_1 \log x_1 - x_2 \log x_2 \cdots - x_n \log x_n$

if we define

$$[Df(x)]_{ij} = \frac{\partial f_i(x)}{\partial x_j} \quad ; \quad \text{where, } i=1,\ j=1,2,\dots n$$

Here,

$$[Df(x)]_{11} = \frac{\partial f(x)}{\partial x_1} = -\log x_1 - 1$$

$$[Df(x)]_{12} = \frac{\partial f(x)}{\partial x_2} = -\log x_2 - 1$$

\- - - - - - - -

$$[Df(x)]_{1n} = \frac{\partial f(x)}{\partial x_n} = -\log x_n - 1$$

So, $Df(x) = [(-\log x_1 - 1) \ (-\log x_2 - 1) \ \cdots \ (-\log x_n - 1)]$

So $\nabla f(x) = [Df(x)]^T = \begin{bmatrix} (-\log x_1 - 1) \\ (-\log x_2 - 1) \\ \vdots \\ (-\log x_n - 1) \end{bmatrix}$ (Ans)