AI 202311398 윤석민

2 Linear algebra

1.1 1. Prove each of below statements

- For matrix $A, B \in \mathbb{R}^{n\times n}$, $||AB|| \leq ||A||||B||$ ← spectral norm
- Function $||x||_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$ (known as p-norm) is norm (hint : hölder inequality)

Q1. $\|AB\| = \max_{\|x\|_2=1} \|ABx\|_2$ (spectral norm) we can use max, because it can have maximum value

- $\|Cx\|_2 \leq \|C\| \|x\|_2$ holds,

because $\|C\|$ means the greatest rate of that How significantly it changes the Euclidean norm of the vector x. (equal holds when C changes norm the bigest)

$\Rightarrow$ By that, $\|ABx\|_2 = \|A(Bx)\|_2 \leq \|A\| \|Bx\|_2 \leq \|A\|\|B\|\|x\|_2$

So, $\|AB\| = \max_{\|x\|_2=1} \|ABx\|_2 \leq \max_{\|x\|_2=1} \|A\| \|Bx\|_2 \leq \max_{\|x\|_2=1} \|A\|\|B\|\|x\|_2$

$= \|A\|\|B\|$

$\therefore \|AB\| \leq \|A\|\|B\|$ holds

Q2.

- $f(\boldsymbol{x}) = 0 \Rightarrow \boldsymbol{x} = \boldsymbol{0}$
- $f(\boldsymbol{x}+\boldsymbol{y}) \le f(\boldsymbol{x}) + f(\boldsymbol{y})$ (the *triangle inequality*)
- $\forall \alpha \in \mathbb{R}, f(\alpha \boldsymbol{x}) = |\alpha| f(\boldsymbol{x})$

⇐ condition of norm.

Target Func : $\|x\|_p = (\Sigma_{i=1}^n |x_i|^p)^{1/p}$

Assume $\vec{x} = [x_1, \dots, x_n]$ that

1. $\|x\|_p = (\Sigma_{i=1}^n |x_i|^p)^{\frac{1}{p}} = (|x_1|^p + \dots + |x_n|^p)^{\frac{1}{p}} = 0$

By the defination, they takes the absolute value.

To be Zero, $x_1 = \dots = x_n = 0$, so $\vec{x} = \vec{0}$.

2. $f(x+y) = (\Sigma_{i=1}^n |x_i+y_i|^p)^{\frac{1}{p}} = (|x_1+y_1|^p + \dots + |x_n+y_n|^p)^{\frac{1}{p}}$

$f(x) = (|x_1|^p + \dots + |x_n|^p)^{\frac{1}{p}}$ $\quad f(y) = (|y_1|^p + \dots + |y_n|^p)^{\frac{1}{p}}$

By hölder inequality,

$$\sum_{i=1}^n x_i (x_i+y_i)^{p-1} \le \left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^n (x_i+y_i)^p\right)^{\frac{1}{q}}$$

(where $\frac{1}{p}+\frac{1}{q}=1$)

$p = (p-1)q$

$$\sum_{i=1}^n y_i (x_i+y_i)^{p-1} \le \left(\sum_{i=1}^n y_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^n (x_i+y_i)^p\right)^{\frac{1}{q}}$$

sum two inequalities

$$\sum_{i=1}^n (x_i+y_i)^p \le \left\{\left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^n y_i^p\right)^{\frac{1}{p}}\right\} \left(\sum_{i=1}^n (x_i+y_i)^p\right)^{\frac{1}{q}}$$

$$\left(\sum_{i=1}^n (x_i+y_i)^p\right)^{1-\frac{1}{q}} \le \left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^n y_i^p\right)^{\frac{1}{p}}$$

$\frac{1}{p}+\frac{1}{q}=1 \Rightarrow \frac{1}{p}=1-\frac{1}{q}$

So. $\left(\sum_{i=1}^{n}|x_i+y_i|^p\right)^{\frac{1}{p}} \preceq \left(\sum_{i=1}^{n}x_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n}y_i^p\right)^{\frac{1}{p}}$

this holds when $x_i>0$, $y_i>0$, $x_i+y_i>0$ by the inequality's condition.

So $\left(\sum_{i=1}^{n}|x_i+y_i|^p\right)^{\frac{1}{p}} \preceq \left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n}|y_i|^p\right)^{\frac{1}{p}}$

$\Rightarrow f(x+y) \preceq f(x)+f(y)$ it holds.

3. $f(\alpha x) = \left(|\alpha x_1|^p+\dots+|\alpha x_n|^p\right)^{\frac{1}{p}}$

$= \left(|\alpha|^p\left(|x_1|^p+\dots+|x_n|^p\right)\right)^{\frac{1}{p}} = |\alpha|\left(|x_1|^p+\dots+|x_n|^p\right)^{\frac{1}{p}}$

$= |\alpha|\ f(x)$ (it holds for $\forall\alpha\in\mathbb{R}$)

3 conditions hold. so $\|x\|_p$ is norm.

2. (Linear independent) Show that the number of elements of a linearly independent set of n-vectors is at most n

assume $k > n$ a set $\{v_1, \dots, v_k\}$ in $\mathbb{R}^n$

Assume $\{v_1, \dots, v_k\}$ is linearly independent.

v_i is n vector, so it can be written like

$$v_i = a_{1i} e_1 + \dots + a_{ni} e_n \quad (0 \le i \le k).$$

$\{v_1, \dots, v_k\}$ set is linearly independent.

so the equation holds

$$: c_1 v_1 + \dots c_k v_k = 0 \quad (\text{only } c_1 = \dots = c_k = 0)$$

it can be written like

$$c_1 (a_{11} e_1 + \dots + a_{n1} e_n) + \dots + c_k (a_{1k} e_1 + \dots + a_{nk} e_n)$$

$$= (c_1 a_{11} + \dots + c_k a_{1k}) e_1 + \dots + (c_1 a_{n1} + \dots + c_k a_{nk}) e_n = 0$$

For the equation to be zero vector, e_i's coefficients must be zero

$$c_1 a_{11} + \dots + c_k a_{1k} = 0$$
$$\vdots$$
$$c_1 a_{n1} + \dots + c_k a_{nk} = 0$$

we have n equations

but $c_1, \dots, c_k$ are k. $(k > n)$.

So it has infinitely many solution,

which is a contradiction.

$\therefore$ So the statement holds.

3. (Matrix norm) Show that for matrix $A \in \mathbb{R}^{n\times n}$ and its maximum singular value σ_{max}, below holds

$$||A||^2 = \sigma_{max}^2$$

$$\|A\|^2 = \left(\sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}\right)^2 = \left(\max_{\|x\|_2 = 1} \|Ax\|_2\right)^2$$ it has maximum so we can use it.

By SVD, $A = U\Sigma V^T$ $\quad U, V \in \mathbb{R}^{n\times n}$ orthogonal matrix

$\Sigma \in \mathbb{R}^{n\times n}$ is diagonal matrix

We need to find $\|Ax\|_2$.

So. $Ax = U\Sigma V^T x$

$= U\Sigma y$ (where $y = V^T x$

$\|y\|_2 = \|x\|_2$ because V is orthogonal matrix)

and U is orthogonal matrix. So, $\|U\Sigma y\|_2 = \|\Sigma y\|_2$

then. $\|Ax\|_2 = \|\Sigma y\|_2$

$$\|\Sigma y\|_2 = \sqrt{\sum_{i=1}^{n} (\sigma_i y_i)^2} = \sqrt{(\sigma_1 y_1)^2 + \dots + (\sigma_n y_n)^2}$$

$$\leq \sqrt{\sigma_{max}^2 (y_1^2 + \dots + y_n^2)} = |\sigma_{max}| \sqrt{y_1^2 + \dots + y_n^2} = |\sigma_{max}| \, \|y\|_2$$

So, $\left(\max_{\|x\|_2 = 1} \|Ax\|_2\right)^2 = \sigma_{max}^2 \cdot (\|y\|_2)^2$ and $\|x\|_2 = \|y\|_2 = 1$.

So. $\|A\|^2 = \sigma_{max}^2$ holds.

4. (Ax = 0) Solve below problem with SVD

$$\underset{x}{\text{argmin}} \ \|Ax\|_2^2 \quad s.t. \quad \|x\| = 1$$

Assume $A \in \mathbb{R}^{m \times n}$

By SVD, $A = U\Sigma V^T$ $\quad U \in \mathbb{R}^{m \times m}, V \in \mathbb{R}^{n \times n}$ are orthogonal matrix

$\Sigma \in \mathbb{R}^{m \times n}$ is diagonal matrix.

like question 3.

$$Ax = U\Sigma V^T x$$

$$= U\Sigma y \quad (\text{where } y = V^T x, \ \|y\|_2 = \|x\|_2$$

$\because$ V is orthogonal matrix)

$\|Ax\|_2 = \|U\Sigma y\| = \|\Sigma y\|$ ($\because$ U is orthogonal matrix)

So, $\|Ax\|_2^2 = \|\Sigma y\|_2^2 = \sum_{i=1}^{r} (\sigma_i y_i)^2$ where $r = \min(m, n)$

$$\geq (\sigma_{min})^2 \cdot \sum_{i=1}^{r} y_i^2 = \sigma_{min}^2 \|y\|_2^2 = \sigma_{min}^2$$

($\because \|y\|_2 = \|x\|_2 = 1$)

To hold $\|Ax\|_2^2 = \|\Sigma y\|_2^2 = \sigma_{min}^2 = \sigma_r^2$, $y = e_r$ (where $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$)

$y = V^T x \rightarrow Vy = V \cdot V^T x = x.$ ($VV^T = I$

$\because$ V is orthogonal matrix)

$\therefore x = V \cdot e^r$

5. (Ax=b) Derive pseudo-inverse of A with A,b, and SVD components

assume $A \in \mathbb{R}^{m\times n}$

By SVD, $A = U\Sigma V^T$ $U \in \mathbb{R}^{m\times m}, V \in \mathbb{R}^{n\times n}$ are orthogonal matrix

$\Sigma \in \mathbb{R}^{m\times n}$ is diagonal matrix.

we need to find A^+ that $A^+A = I_n$, $AA^+ = I_m$ holds

pseudo inverse A^+ is defined as $A^+ = V\Sigma^+U^T$

where Σ^+ is pseudo inverse of Σ

If $m>n$

$$\Sigma = \underbrace{\overbrace{\begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix}}^{n}}_{} \Bigg\} m \qquad \text{then } \Sigma^+ = \overbrace{\begin{bmatrix} \frac{1}{\sigma_1} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & \frac{1}{\sigma_n} & 0 & \cdots & 0 \end{bmatrix}}^{m} \Big\} n$$

$\Sigma\Sigma^+ = I_m$ $\Sigma^+\Sigma = I_n$

If $m<n$

$$\Sigma = \overbrace{\begin{bmatrix} \sigma_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & \sigma_n & 0 & \cdots & 0 \end{bmatrix}}^{n} \Big\} m \qquad \text{then } \Sigma^+ = \overbrace{\begin{bmatrix} \frac{1}{\sigma_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma_n} \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix}}^{m} \Bigg\} n$$

$\Sigma\Sigma^+ = I_m$ $\Sigma^+\Sigma = I_n$

So $AA^+ = (U\Sigma V^T)(V\Sigma^+U^T) = U\Sigma\Sigma^+U^T = UU^T = I_m$

$A^+A = (V\Sigma^+U^T)(U\Sigma V^T) = V\Sigma^+\Sigma V^T = VV^T = I_n$

$\therefore A^+ = V\Sigma^+U^T$

3 Probability

1. Prove below statements for ~~random variable~~ events A,B.

   - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (Use only Probability axiom!)
   - If A,B are independent $P(A) = P(A|B^c)$

Q1. event $A \cup B$ includes all outcomes that are in A or B (or both).

event $A \cap B$ includes all outcomes that are in both A and B.

According to Axiom, if A and B are disjoint, then :

$$P(A \cup B) = P(A) + P(B)$$

But, A and B may not be mutually exclusive, so we need to account for the overlap.

Simply adding P(A) and P(B), we would double-count the outcomes that are in both A and B.

So, we need to subtract the probability of intersection.

$$\Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

By axiom, $P(A \cap B) \geq 0$. thus, subtracting $P(A \cap B)$ from $P(A) + P(B)$ ensures valid probability. if the values are appropriately defined.

$\Rightarrow$ So statement holds.

Q2. If A, B are independent, $P(A) = P(A|B^c)$

'independent' means $P(A,B) = P(A)P(B)$,

so, $P(A|B) = \dfrac{P(A,B)}{P(B)} = P(A)$ holds

By axiom $P(B^c) = 1 - P(B)$, and $P(A|B^c) = \dfrac{P(A,B^c)}{P(B^c)}$

$P(A)$ can be expressed : $P(A) = P(A,B) + P(A,B^c)$

$P(A,B) = P(A) \cdot P(B)$ holds, so $P(A) = P(A) \cdot P(B) + P(A,B^c)$

then $P(A,B^c) = P(A)(1 - P(B)) = P(A) \cdot P(B^c)$

thus, $P(A|B^c) = \dfrac{P(A) \cdot P(B^c)}{P(B^c)} = P(A)$

$\therefore$ statement holds.

2. (Bayes' theorem) There is a factory with machines M_1, M_2, M_3 accounting for ratio of 0.2, 0.3, 0.5 of entire production each. The failure rate of each machine is known to be 0.03, 0.02, 0.01 each. If one randomly chosen product is found to be a failure, find the probability for each machine for manufacturing the product.

$P(M_1) = 0.2 \quad P(M_2) = 0.3 \quad P(M_3) = 0.5$ ← choose machine.

$P(F|M_1) = 0.03 \quad P(F|M_2) = 0.02 \quad P(F|M_3) = 0.01$

↑ if chose M_i product, the probability of failure product.

⇒ we need to find $P(M_1|F)$, $P(M_2|F)$, $P(M_3|F)$

By Bayes' Theorem,

$$P(F) = P(F|M_1)\cdot P(M_1) + P(F|M_2)\cdot P(M_2) + P(F|M_3)\cdot P(M_3)$$

$$= 0.03\cdot 0.2 + 0.02\cdot 0.3 + 0.01\cdot 0.5$$

$$= 0.017$$

$$P(M_1|F) = \frac{P(F|M_1)\cdot P(M_1)}{P(F)} = \frac{0.03\cdot 0.2}{0.017} = \frac{6}{17}$$

$$P(M_2|F) = \frac{P(F|M_2)\cdot P(M_2)}{P(F)} = \frac{0.02\cdot 0.3}{0.017} = \frac{6}{17}$$

$$P(M_3|F) = \frac{P(F|M_3)\cdot P(M_3)}{P(F)} = \frac{0.01\cdot 0.5}{0.017} = \frac{5}{17}$$

So, the probability for each machine for manufacturing the chosen failure product are $6/17$, $6/17$, $5/17$

3. (Gaussian distributions) For random variable $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,1)$ show that $Z = X + Y$ is gaussian and state its variance.

$X \sim N(0,1)$ means $\mu_x = 0$, $\sigma_x^2 = 1$

$Y \sim N(0,1)$ also means $\mu_y = 0$, $\sigma_y^2 = 1$

in $Z = X + Y$

μ_z means the x value of peak of graph.

X and Y have same μ. so $\mu_z = \mu_x + \mu_y = 0$

σ means variance that $var[x] = E[x^2] - E[x]^2 = \sigma^2$

then, $\sigma_x^2 = E[x^2] - E[x]^2$, $\sigma_y = E[y^2] - E[y]^2$

we need to get σ_z^2.

$$E[(x+y)^2] - E[x+y]^2 = E[x^2 + y^2 + 2xy] - E[x+y]^2$$

Using Expectation func is linear function for continous function.

$\Rightarrow E[aX + bY] = aE[X] + bE[Y]$ holds.

And X and Y are independent $\Rightarrow$ So, $E[XY] = E[X][Y]$.

$$= E[x^2] + E[y^2] + 2E[x]E[y] - (E[x] + E[y])^2$$

$$= (E[x^2] - E[x]^2) + (E[y^2] - E[y]^2) + (2E[x]E[y] - 2E[x]E[y])$$

$$= \sigma_x^2 + \sigma_y^2 = \sigma_z^2 = 2$$

$\therefore$ $Z = X + Y$ is Gaussian with $Z \sim N(0,2)$

4. (Gaussian random vector) Show that for gaussian random vector X with covariance C, there always exist a matrix A that satisfies $C = AA^T$

$C \in \mathbb{R}^{D \times D}$ $X \sim N(\mu, C)$

C is defined as $C = E[(X - E[X]) \cdot (X - E[X])^T]$

$C_{ij} = E[(X_i - E[X_i]) \cdot (X_j - E[X_j])]$

$C_{ji} = E[(X_j - E[X_j])(X_i - E[X_i])]$

by properties of covariance

$C_{ij} = C_{ji}$, so it is symmetric.

and also positive semi-definite.

So. By SVD. C can be expressed as $C = U\Sigma U^T$ (symmetric)

where U is orthogonal matrix
Σ is diagonal matrix

We need to show $C = AA^T$

$\Rightarrow AA^T = U\Sigma U^T = U\sqrt{\Sigma}(\sqrt{\Sigma}U^T)$

we set $A = U\sqrt{\Sigma}$, where $\sqrt{\Sigma}$ is the diagonal matrix whose entries are square roots of Σ's entries.

C is positive semi-definite. So $A = U\sqrt{\Sigma}$ exists.

$\therefore$ statement holds.

5. (Conditional Expectation) There is a piece of wooden stick with a length of L. If we break the stick at random point 2 times, what is the average length of the remaining stick?

·) First Break. at random point $X \sim U[0, L]$

$\Rightarrow$ Result Piece 1 : Length $X \Rightarrow E[X] = \frac{L}{2}$

Piece 2 : Length $L - X \Rightarrow L - \frac{L}{2} = \frac{L}{2}$

and choose one piece randomly.

each piece has ½ probability of being chosen.

·) Breaking chosen piece

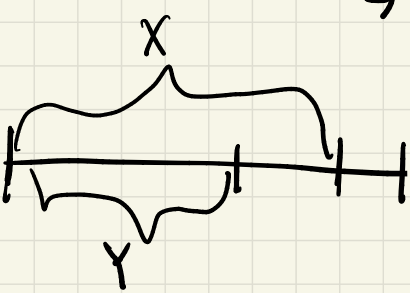
Case 1. Piece 1. (length X_1)

Break at random point $Y \sim [0, X]$

$\Rightarrow$ Result : Remaining piece : $Y \Rightarrow E[Y|X] = \frac{X}{2}$

break : $X - Y \Rightarrow X - \frac{X}{2} = \frac{X}{2}$

$\Rightarrow E[Y] = E[E[Y|X]] = E[\frac{X}{2}] = \frac{L}{4}$



$E[\text{total case } 1] = E[L-X] + E[Y] + E[X-Y]$

$= \frac{L}{2} + \frac{L}{4} + \frac{L}{4} = L$.

Case 2. Piece 2. (Length $L-X_1$)

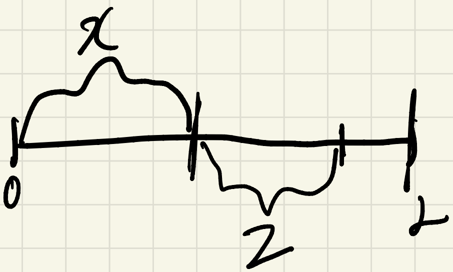
Break at random point Z in $[0, L-X]$

$\Rightarrow$ Result : Remaining piece : Z $\Rightarrow E[Z \mid L-X] = \frac{L-X}{2}$

break : $(L-X)-Z$ $\Rightarrow (L-X)-\left(\frac{L-X}{2}\right) = \frac{L-X}{2}$

$$\Rightarrow E[Z] = E\big[E[Z \mid L-X]\big] = E\left[\frac{L-X}{2}\right] = \frac{1}{2}E[L-X]$$

$$= \frac{L}{4}$$



So, $E[\text{total case2}]$

$= E[X] + E[Z] + E[L-X-Z]$

$= \frac{L}{2} + \frac{L}{4} + \frac{L}{4} = L$.

So. $E[\text{total}] = \frac{1}{2}E[\text{total case1}] + \frac{1}{2}E[\text{total case 2}]$

$= \frac{1}{2}L + \frac{1}{2}L = L$.

So total remaining stick's length is L.

So average is $L/3$.

4 Optimization

1. Show that if function f is differentiable, $D_v f(x) = \nabla_x f(x)^T v$ holds

Assume $x, v \in \mathbb{R}^n$ $f: \mathbb{R}^n \to \mathbb{R}$

$$D_v f(x) = \lim_{t \to 0} \frac{f(x+tv) - f(x)}{t}, \quad \nabla_x f(x)^T = \left[\frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n} \right]$$

We want to know how many f(x)'s value changes whan moving to the direction of v

thus,

By applying chain rule, we can get change of f(x) when x changes in the direction of v

$$\therefore \frac{d}{dt} f(x+tv) = \frac{d}{dx} f(x+tv) \cdot \frac{d(x(t))}{dt} = \nabla f(x+tv)^T \cdot v$$

At $t=0$, it becomes: $\left.\frac{d}{dt} f(x+tv)\right|_{t=0} = \nabla f(x)^T v$

$\lim_{t \to 0} \frac{f(x+tv) - f(x)}{t}$ is the same as evaluating the derivative of $f(x+tv)$ at $t=0$ $\Rightarrow D_v f(x) = \nabla f(x)^T v$

$\therefore$ So statement holds

2. (Constrained least squares) Show that for n-vector x, a, p-vector d and matrix $C \in \mathbb{R}^{p\times n}$ with linearly independent row vectors, the solution of the problem

$$\begin{array}{ll} \text{minimize} & \|x-a\|^2 \\ \text{subject to} & Cx = d \end{array} \tag{1}$$

is $\hat{x} = a - C^\dagger(Ca-d)$. *Hint.* Use directly below optimality condition(known as KKT condition)

$$\begin{bmatrix} 2I & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} 2a \\ d \end{bmatrix} \tag{2}$$

where p-vector z is lagrangian multiplier of formulation

$$L(\mathbf{x}, \mathbf{z}) = \|\mathbf{x} - a\|^2 + z_1(\mathbf{c}_1^T\mathbf{x} - d_1) + \cdots + z_p(\mathbf{c}_p^T\mathbf{x} - d_p)$$

$$C = \begin{bmatrix} -c_1- \\ \vdots \\ -c_p- \end{bmatrix} \qquad \mathcal{L}(x,z) = \|x-a\|^2 + z^T(Cx-d)$$

$$= (x-a)^T(x-a) + z^T(Cx-d)$$

• Stationarity : to minimize the Lagrangian.

① $$\frac{\partial \mathcal{L}(x,z)}{\partial x} = 2(x-a) + C^T z = 0$$

$$\therefore x = a - \frac{1}{2}C^T z$$

result of computing the matrix.

② $$\frac{\partial \mathcal{L}(x,z)}{\partial z} = Cx - d = 0$$

$\therefore Cx = d$ that is constraint

And, substitute $x = a - \frac{1}{2}C^T z$ into $Cx = d$, solve z.

$$C\left(a - \frac{1}{2}C^T z\right) = d \Rightarrow Ca - \frac{1}{2}CC^T z = d$$

$$\Rightarrow \frac{1}{2}CC^T z = Ca - d \Rightarrow \therefore z = 2(CC^T)(Ca-d)$$

So, $x = a - \frac{1}{2} C^T \cdot 2 (CC^T)^{-1} (Ca - d)$

$= a - C^T (CC^T)^{-1} (Ca - d)$

then, we need to show $C^+ = C^T (CC^T)^{-1}$

$CC^+ \overset{?}{\Rightarrow} CC^T (CC^T)^{-1} = (CC^T)(CC^T)^{-1} = I$

$C^+ C \overset{?}{\Rightarrow} C^T (CC^T)^{-1} C = C^T ((C^T)^{-1} C^{-1}) C$

$= C^T \cdot (C^T)^{-1} \cdot (C^{-1} \cdot C) = I$

thus, $C^+ = C^T (CC^T)^{-1}$

$\therefore$ the solution is $\hat{x} = a - C^+ (Ca - d)$

3. (Minimizing norms) Formulate the following problem as LP. Explain in detail the relation between the optimal solution of each problem and the solution of its equivalent LP. (LP form is something like below)

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & a_i^T x \leq b_i \; i = 1, \ldots, m \end{aligned} \tag{3}$$

- Minimize $||Ax - b||_\infty$
- Minimize $||Ax - b||_1$

Assume $c, x \in \mathbb{R}^n$ $A \in \mathbb{R}^{m \times n}$ $b \in \mathbb{R}^m$

Q1. $||Ax - b||_\infty = \max_i |a_i^T x - b_i|$ $1 \leq i \leq m$.

To minimize this, we use an auxiliary variable t, it represents the maximum boundary.

Rewriting the problem as LP.

minimize t subject to $a_i^T x - b_i \leq t$ $i = 1, \ldots, m$

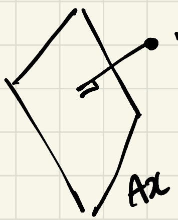
$-a_i^T x + b_i \leq t$ $i = 1, \ldots, m$

$t \geq 0$

Q2. $\|Ax-b\|_1 = \sum_{i=1}^{m} |a_i^T x - b_i|$

To minimizing that, we minimize sum of each absolute value.

geometrically, think b as point and Ax as hyperplane that x makes.



we find x that minimize the distance between b and Ax's hyper plane.

minimize $c^T z$

subject to $Mz \le d$, $z \ge 0$

where $z = \begin{bmatrix} x \\ u \\ v \end{bmatrix}$ $\quad x \in \mathbb{R}^n$

$u \in \mathbb{R}^m$ positive deviation

$v \in \mathbb{R}^m$ negative deviation

$c = \begin{bmatrix} 0_n \\ 1_m \\ 1_m \end{bmatrix}$ where 0_n: length n zero vector

1_m: length m. vector of ones

$M = \begin{bmatrix} A & -I & I \\ -A & -I & I \\ 0 & -I & 0 \\ 0 & 0 & -I \end{bmatrix}$ $\quad d = \begin{bmatrix} b \\ -b \\ 0 \\ 0 \end{bmatrix}$ where I is I_m.

LP rewrited as

$$\text{minimize} \quad \sum_{i=1}^{m} u_i + v_i$$

$$\text{s.t.} \quad \begin{bmatrix} A & -I & I \\ -A & -I & I \\ 0 & -I & 0 \\ 0 & 0 & -I \end{bmatrix} \begin{bmatrix} x \\ u \\ v \end{bmatrix} \leq \begin{bmatrix} b \\ -b \\ 0 \\ 0 \end{bmatrix} \qquad \begin{matrix} x \geq 0 \\ u \geq 0 \\ v \geq 0 \end{matrix}$$

By solving this LP, we obtain the optimal solution

$z^* = \begin{bmatrix} x^* \\ u^* \\ v^* \end{bmatrix}$, where x^* is the solution of problem