1) Normal equation

Theorem 1: Let $A \in \mathbb{C}^{m\times n}$, $m \geq n$ and $b \in \mathbb{C}^m$

define $r(x) := Ax - b$ = residual

then x solves LS:

$$x = \underset{\xi \in \mathbb{C}^n}{\arg\min} \ \|\underbrace{A\xi - b}_{r(\xi)}\|_2$$

If $r(x) \perp \text{range}(A)$

Note: $\mathbb{C}^m = \text{range}(A) \oplus \underbrace{\ker(A^*)}_{[\text{range}(A)]^\perp}$

Thus if $r(x) \perp \text{range}(A) \Rightarrow r(x) \in \ker(A^*)$

$A^* r(x) = A^*(Ax - b) = 0$

The LS solver must satisfy: $\boxed{A^*Ax = A^*b}$ normal equation.

P = orthogonal projector on range (A)

$$b = \underbrace{Pb}_{\in \text{range } A} + \underbrace{(I-P)b}_{\perp \text{ range } A} = b^{\parallel} + b^{\perp}$$

$\forall \xi \in \mathbb{C}^n$

$$r(\xi) = A\xi - b = \underbrace{A\xi - b^{\parallel}}_{r^{\parallel}(\xi)} - \underbrace{b^{\perp}}_{r^{\perp}(\xi)}$$

$$\|r(\xi)\|_2^2 = \|r^{\parallel}(\xi) + r^{\perp}(\xi)\|_2^2 = \|r^{\parallel}(\xi)\|_2^2 + \|r^{\perp}(\xi)\|_2^2$$

$$= \|A\xi - b^{\parallel}\|_2^2 + \|b^{\perp}\|_2^2$$

At minimizes: $r(x) = Ax - b^{\parallel} = 0$

$r^{\perp}(x) = -b^{\perp} \longrightarrow r(x) = -b^{\perp} \perp \text{range}(A)$

If A is full rank: $\text{rank}(A) = n$

A^*A is invertible $\qquad x = \underbrace{(A^*A)^{-1}A^*}_{A^+} b = A^+ b \qquad A^+ \in \mathbb{C}^{n\times m}$

A^+ pseudo inverse $\qquad A^+A = I_n$

If A not full rank: infinitely many minimizers:

$b^\perp + b'' \qquad Ax = b''$

$$\mathcal{S} := \{x \in \mathbb{C}^n : \|Ax - b\|_2 = \|b^\perp\|_2\}$$

$\boxed{\text{convex set}}$. i.e. $\forall x, x' \in \mathcal{S}$  line segment within set

$\exists$ unique $x \in \mathcal{S}$ of min. norm $\|x\|_2$ $\qquad x = \underbrace{A^+}_{\text{pseudo inverse}} b$

$\forall t \in [0, 1]$

$$tx + (1-t)x' \in \mathcal{S}$$

since $A[tx + (1-t)x'] = \underbrace{tAx}_{b''} + \underbrace{(1-t)Ax'}_{b''} = b''$

2) LS with SVD

Reduced SVD: $A = \hat{U}\hat{\Sigma}V^*$

$\hat{U} \in \mathbb{C}^{m\times n}$ orthonormal columns

$\hat{\Sigma} = \text{diag}(\sigma_1, \cdots \sigma_n)$

$V \in \mathbb{C}^{n\times n}$ unitary

Suppose $\text{rank}(A) = r < m \qquad \hat{\Sigma} = \text{diag}(\sigma_1, \cdots \sigma_r, 0, \cdots 0)$

$\|Ax - b\|_2^2 = \|\hat{U}\hat{\Sigma}V^*x - b\|_2^2 \qquad \hat{U} \in \mathbb{C}^{m\times n} \quad \hat{U}^* \in \mathbb{C}^{n\times m}$

$\hat{U}^*x \in \mathbb{C}^n$

$\hat{U}\hat{U}^*x = \hat{U}(\hat{U}^*x) \Rightarrow$ must be in range $\hat{U}$

$$\begin{cases} y := V^*x & (1) \\ \Pi := \hat{U}\hat{U}^* = \text{orthog projection on span}\{u_1, \cdots u_n\} \end{cases}$$

$$\|Ax-b\|_2^2 = \|\hat{U}\hat{\Sigma}y - \pi b - (I-\pi)b\|_2^2 \qquad b = (I-\pi)b + \pi b \qquad (2)$$

$$= \|\hat{U}\hat{\Sigma}y - \underset{\hat{U}\hat{U}^*}{\underline{\underline{\pi}}} b\|_2^2 + \|(I-\pi)b\|_2^2$$

let $\vec{v} = \hat{U}(\hat{\Sigma}y - \hat{U}^* b)$

$\|\vec{v}\|_2 = \vec{v}^*\vec{v}$

$= (\hat{\Sigma}y - \hat{U}^*b)^*\hat{U}^*\hat{U}(\hat{\Sigma}y - \hat{U}^*b)$

$= \|\hat{\Sigma}y - \hat{U}^*b\|_2$

$$= \|\hat{U}(\hat{\Sigma}y - \hat{U}^*b)\|_2^2 + \|(I-\pi)b\|_2^2$$

orthonormal columns

$$= \|\hat{\Sigma}y - U^*b\|_2^2 + \|(I-\pi)b\|_2^2$$

$$\hat{\Sigma}y - U^*b = \begin{pmatrix} \sigma_1 y_1 \\ \vdots \\ \sigma_r y_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} - \begin{pmatrix} u_1^* b \\ \vdots \\ u_r^* b \\ \vdots \\ u_n^* b \end{pmatrix}$$

(2) becomes $\|Ax-b\|_2^2 = \sum_{j=1}^{r} |\sigma_j y_j - u_j^* b|^2 + \sum_{j=r+1}^{n} |u_j^* b|^2 + \|(I-\pi)b\|_2^2$ (3)

At min: $y_j = \dfrac{u_j^* b}{\sigma_j}$ $\quad j = 1, \cdots r \qquad y_{r+1}, \cdots y_n$ unspecified.

$y = V^* x$

only x (y) can be minimized.

minimizes from (1): $x = Vy = \sum_{j=1}^{n} y_j v_j$

$$= \boxed{\sum_{j=1}^{r} v_j \frac{(u_j^* b)}{\sigma_j} + \sum_{j=r+1}^{n} v_j y_j}$$

$$Ax = \hat{U}\hat{\Sigma}y = \hat{U}\begin{pmatrix} \sigma_1 y_1 \\ \vdots \\ \sigma_r y_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \boxed{\sum_{j=1}^{r} \sigma_j y_j u_j}$$

$b = b^{\parallel} + b^{\perp}$

$\Downarrow$

$\hat{U}(:, 1:r)\hat{U}^*(:, 1:r)$

$$\forall x \in \mathcal{Y}: \quad Ax - b = -b^{\perp} = \boxed{\sum_{j=1}^{r} \sigma_j y_j u_j} - \hat{U}(:, 1:r)\hat{U}^*(:, 1:r)\, b - b^{\perp}$$

$$= \sum_{j=1}^{r} \sigma_j y_j u_j - \sum_{j=1}^{r} (u_j^* b) u_j - b^{\perp}$$

$$= -b^{\perp}$$

x is a span of columns of V.

$\|x\|_2^2 = \sum_{j=1}^{r} \left|\dfrac{u_j^* b}{\sigma_j}\right|^2 + \sum_{j=r+1}^{n} |y_j|^2$ $\qquad$ Min. norm of x: $\; x = \sum_{j=1}^{r} v_j \dfrac{u_j^* b}{\sigma_j}$

$x = A^+ b = V\Sigma^+ \hat{U}^* b$ where Σ^+ is $\mathrm{diag}\left(\frac{1}{\sigma_1}, \cdots \frac{1}{\sigma_r}, 0 \cdots 0\right)$

$$\boxed{A^+ = V\Sigma^+ \hat{U}^*}$$ definition of A^+ in SVD.

In particular case $A =$ full rank:

$$A^+ = V\Sigma^{-1}\hat{U}^* \qquad \Sigma^+ = \mathrm{diag}\left(\frac{1}{\sigma_1}, \cdots \frac{1}{\sigma_n}\right) = \hat{\Sigma}^{-1}$$

$$(A^*A)^{-1}A^* = (\underbrace{V\hat{\Sigma}\hat{U}^*}_{A^*}\underbrace{\hat{U}\hat{\Sigma}V^*}_{A})^{-1} V\hat{\Sigma}\hat{U}^*$$

$\hat{U}^*\hat{U} \downarrow I_n$

$$= \underbrace{(V\hat{\Sigma}^2 V^*)^{-1}}_{\text{take inverse}} V\hat{\Sigma}\hat{U}^*$$

$$= V\hat{\Sigma}^{-2}V^* V\hat{\Sigma}\hat{U}^* = V\hat{\Sigma}^{-2}\hat{\Sigma}\hat{U}^* = V\hat{\Sigma}^{-1}\hat{U}^*$$

$$AA^+ = \hat{U}\hat{\Sigma}V^*V\Sigma^+\hat{U}^* = \hat{U}\hat{\Sigma}\Sigma^+\hat{U}^* = \hat{U}\,\mathrm{diag}(1, \cdots 1, 0, \cdots 0)\,\hat{U}^*$$

$= \hat{U}(:, 1:r)\,\hat{U}^*(:, 1:r) =$ orthogonal projector on the range of A.

$$A^+A = V\hat{\Sigma}^+\hat{U}^*\hat{U}\hat{\Sigma}V^* = V\Sigma^+\hat{\Sigma}V^* = V\,\mathrm{diag}(1, \cdots 1, 0 \cdots 0)\,V^*$$

$= \hat{V}(:, 1:r)\,\hat{V}^*(:, 1:r) =$ orthog proj on range of A^*

If A is full rank: range $(A^*) = \mathbb{C}^n$, $A^+A = I_n$

<u>Remark</u>: Pseudo inverse A^+ is the unique solution of :

$$\min \| AX - I_m \|_F^2 \quad \text{s.t } \|X\|_F \text{ is minimum.}$$

$$A^+ = \mathrm{argmin}$$

$$\|AX - I_m\|_F^2 = \|A(x_1, \cdots x_m) - (e_1, \cdots e_m)\|_F^2$$
$$= \sum_{j=1}^m \|Ax_j - e_j\|_2^2$$

$$\|X\|_F^2 = \sum_{j=1}^m \|x_j\|_2^2$$

j by j: $\min \|Ax_j - e_j\|_2$

s.t. $\|x_j\|_2$ is min. $\quad x_j = A^+e_j$

$\Rightarrow X = A^+$

3) LS via QR factorization

$A \in \mathbb{C}^{m\times n}$, $A = \hat{Q}\hat{R}$: $\hat{Q} = (q_1, \cdots q_n)$ matrix with orthonormal columns $\in \mathbb{C}^{m\times n}$

$$\hat{R} = \begin{pmatrix} r_{11} & \cdots & r_{1n} \\ & \ddots & \\ 0 & & r_{nn} \end{pmatrix}$$

$$\|Ax - b\|_2^2 = \|\hat{Q}\hat{R}x - b\|_2^2$$

$$= \|\hat{Q}\hat{R}x - \hat{Q}\hat{Q}^* b - (I - QQ^*)b\|_2^2$$

$$= \|\hat{Q}(\hat{R}x - \hat{Q}^* b)\|_2^2 + \|(I - \hat{Q}\hat{Q}^*)b\|_2^2$$

$$= \|\hat{R}x - \hat{Q}^* b\|_2^2 + \|(I - \hat{Q}\hat{Q}^*)b\|_2^2$$

If A is full rank $r_{jj} \neq 0 \quad \forall j = 1, \cdots n$

$\Rightarrow$ LS solution is $x = \hat{R}^{-1}\hat{Q}^* b = A^+ b$

$$A^+ = (A^*A)^{-1}A^* = (\hat{R}^* \underbrace{\hat{Q}^*\hat{Q}}_{I_n} \hat{R})^{-1} \hat{R}^* \hat{Q}^*$$

$$= (\hat{R}^*\hat{R})^{-1} \hat{R}^* \hat{Q}^*$$

$$= \hat{R}^{-1}(\hat{R}^*)^{-1}\hat{R}^*\hat{Q}^* = \hat{R}^{-1}\hat{Q}^*$$

If A is not full rank, can't solve $\hat{R}x - \hat{Q}^* b$ uniquely

some r_j is 0.

eg. $\hat{R} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 0 \end{pmatrix}$ $\quad \hat{R}x = \begin{pmatrix} x_1 + 2x_2 + 3x_3 \\ 4x_4 + 5x_5 \\ 0 \end{pmatrix}$ minimize it

as close as you can to $\hat{Q}^* b$

$Ax = b \qquad A\backslash b \Rightarrow$ solves LS