2.2. Math Preparation.

Consider the linear system

$$Ax = b \qquad A \in \mathbb{R}^{m\times n}$$

with $A = U^T \Sigma V$ (SVD).

(Recall $\Sigma = \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_{r_0} & & 0 \\ & & & 0 & \\ & & & & \ddots \; 0 \end{pmatrix}$ if $n \geq m$.)

Without loss of generality, we assume $n \geq m$.

Let us look at

$$Ax = b$$

$$\Leftrightarrow \quad U^T \Sigma Vx = b \qquad \begin{array}{l} U \in \mathbb{R}^{m\times m} \\ V \in \mathbb{R}^{n\times n} \end{array} \text{ orthogonal}$$

$$U^T U = UU^T = id$$

$\times U$

$$\Leftrightarrow \quad \Sigma \underbrace{Vx}_{\overset{def}{=} \tilde{x}} = \underbrace{Ub}_{\overset{def}{=} \tilde{b}}$$

$$\Leftrightarrow \quad \Sigma \tilde{x} = \tilde{b}$$

$$\Leftrightarrow \begin{cases} \sigma_1 \tilde{x}_1 = \tilde{b}_1 \\ \sigma_2 \tilde{x}_2 = \tilde{b}_2 \\ \vdots \\ \sigma_r \tilde{x}_r = \tilde{b}_r \\ 0 = \tilde{b}_{r+1} \\ \vdots \\ 0 = \tilde{b}_m \end{cases}$$

r: rank of A

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$

Case 1: If $\tilde{b}_{r+1} = \cdots = \tilde{b}_m = 0$, there is at least one sol

$$\tilde{x}_1 = \frac{\tilde{b}_1}{\sigma_1}, \ldots, \tilde{x}_r = \frac{\tilde{b}_r}{\sigma_r}.$$

Case 2: If not all of $\tilde{b}_{r+1}, \ldots, \tilde{b}_n$ are zeros, there is no sol.

Another Point of View: If there is no solution, write $A = \begin{pmatrix} | & | & & | \\ a^{(1)} & a^{(2)} & \cdots & a^{(n)} \\ | & | & & | \end{pmatrix}$.

then $\quad Ax \neq b \quad \forall x,$

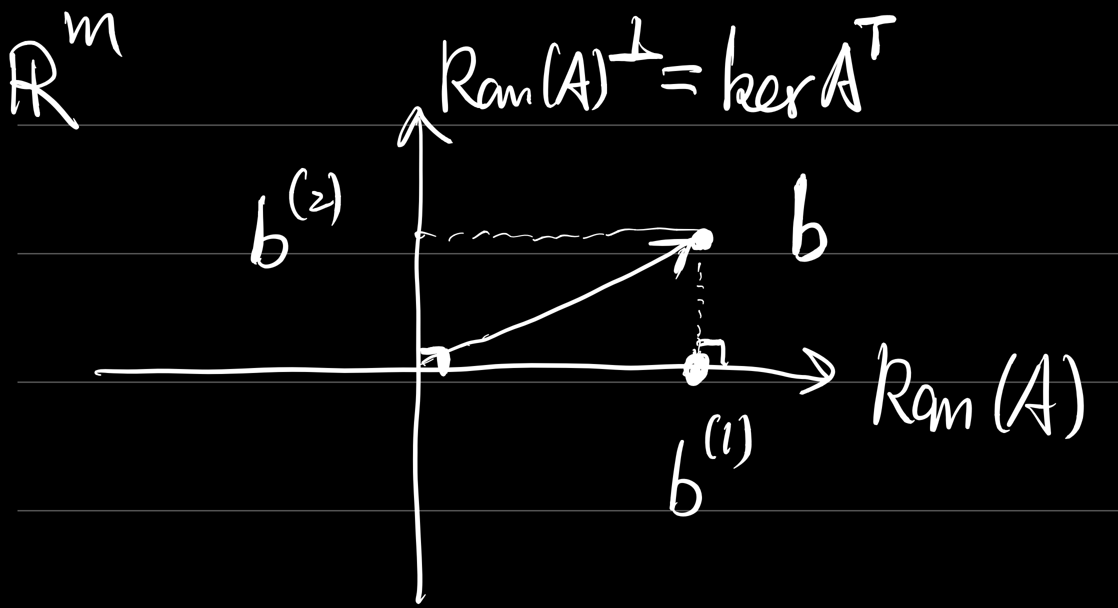
$$\Leftrightarrow \begin{pmatrix} | & & | \\ a^{(1)} & \cdots & a^{(n)} \\ | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \neq b$$

matrix multi.

$$\Leftrightarrow \quad x_1 \begin{matrix} | \\ a^{(1)} \\ | \end{matrix} + x_2 \begin{matrix} | \\ a^{(2)} \\ | \end{matrix} + \cdots + x_n \begin{matrix} | \\ a^{(n)} \\ | \end{matrix} \neq b$$

def of span

$$\Leftrightarrow \quad \mathrm{Ran}(A) = \mathrm{span}\{ a^{(1)}, \ldots, a^{(n)} \} \not\ni b$$



Recall:
$\mathrm{Ran}(A)^{\perp}$
$= \ker(A^T)$

Let $\quad b^{(1)} \overset{\text{def}}{=\!=} \mathrm{Proj}_{\mathrm{Ran}(A)}\, b \quad$ (orthogonal projection)

then $\quad b^{(1)} \in \mathrm{span}\{ a^{(1)}, \ldots, a^{(n)} \} = \mathrm{Ran}(A)$

$$\Leftrightarrow \quad \hat{x}_1 \begin{matrix} | \\ a^{(1)} \\ | \end{matrix} + \dots + \hat{x}_n \begin{matrix} | \\ a^{(n)} \\ | \end{matrix} = b^{(1)}$$

for some $\hat{x}$.

$$\Leftrightarrow \quad A\hat{x} = b^{(1)} \text{ for some } \hat{x}.$$

To find $\hat{x}$, notice

$$Ax = b = \underbrace{b^{(1)}}_{\in \mathrm{Ran}(A)} + \underbrace{b^{(2)}}_{\in \mathrm{Ran}(A)^{\perp} = \ker A^T}$$

apply A^T: $\quad A^T A \hat{x} = A^T b = A^T b^{(1)}$

This is the eqn for $\hat{x}$.

Def: The sol of

$$A^T A \hat{x} = A^T b \quad \text{(normal eqn)}$$

is called the least square sol.

Prop: $\hat{x} = \arg\min_{x} \|Ax - b\|^2$

Proof: For any x,

$$\|Ax-b\|^2 = \|\underbrace{Ax - b^{(1)}}_{\in \operatorname{Ran}(A)} + \underbrace{(-b^{(2)})}_{\in \operatorname{Ran}(A)^\perp}\|^2$$

$$\stackrel{\text{Pythagorean}}{=} \|Ax - b^{(1)}\|^2 + \|b^{(2)}\|^2$$

$$\geq \|b^{(2)}\|^2$$

and "=" occurs if and only if $Ax = b^{(1)}$, this is the eqn for $\hat{x}$

∎

2.3. Linear Regression.

Recall our linear system:

$$X^T \beta = y \qquad X \in \mathbb{R}^{p \times n}$$

We solve for the least square sol β^{ls}:

~~from~~ $$(XX^T)\beta = Xy$$

In particular, if X has full row rank, then XX^T is invertible, so

$$\beta^{ls} = (XX^T)^{-1} X y$$