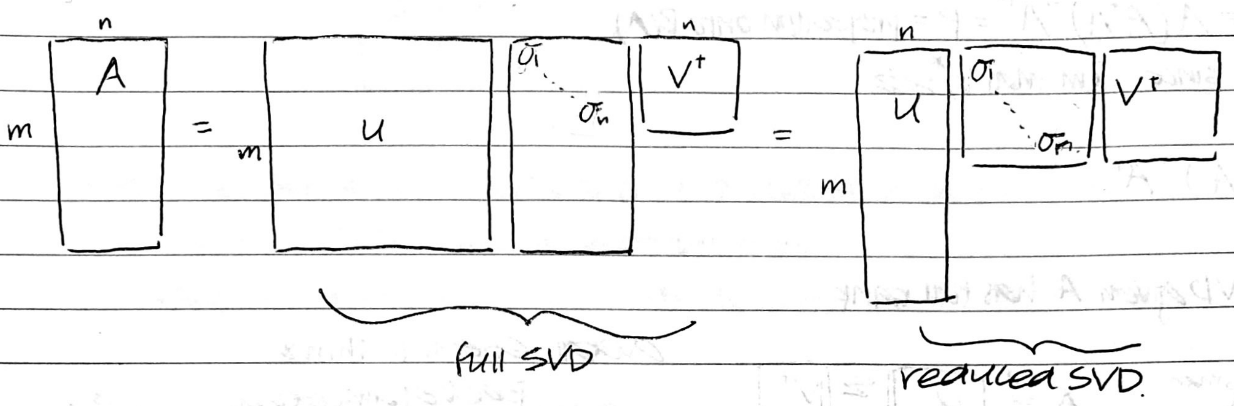
SVD recap w/ accuracy of Least Squares

Wed, Nov. 27th/19

Recall: SVD for full-rank, tall matrix $\underset{m}{\overset{n}{\boxed{A}}} \in \mathbb{R}^{m \times n}.$

↓

implication: n # of singular values (non-zero)



$$= \underbrace{\sum_{i=1}^{n} \sigma_i \boxed{u_i} \boxed{V^T}}_{\text{sum of rank 1}}$$

Terminology: u_i, v_i = <u>singular</u> <u>vectors</u> (~eigenvectors)

Q: In terms of SVD, what is P, the matrix which projects onto <u>R(A)</u>. What sing. vals of P & $A^{\dagger}$?

$$A^{\dagger} = \boxed{V} \boxed{\Sigma^{\dagger}} \boxed{U^{\dagger}}$$

$$\Sigma^{\dagger} = \begin{bmatrix} 1/\sigma_1 & & & \\ & \ddots & & 0 \\ & & 1/\sigma_n & \end{bmatrix}$$

Recall: $R(A) = \text{span}\underbrace{(u_1, \dots u_n)}_{\text{orthobasis}}$ (since rank(A) = n)

$$P = \sum_{i=1}^{n} u_i u_i^T = \boxed{U} \boxed{U^T} = \sum_{i=1}^{n} 1 \cdot u_i u_i^T$$

P has sing vals $\sigma_1 = \sigma_2 = \cdots = \sigma_n = 1$.

$A^+ = V\Sigma^+ U^T = \sum_{i=1}^{n} \frac{1}{\sigma_n} v_i u_i^T$

sing values of A^+ : $\frac{1}{\sigma_n}, \frac{1}{\sigma_{n-1}}, \dots \frac{1}{\sigma_1}$, where $\frac{1}{\sigma_n} > \frac{1}{\sigma_{n-1}}$

Accuracy of least squares assuming noisy linear model.

↳ not on final other than understanding of SVD.

$y = Ax + z$

↳ white noise (Gaussian/standard normal).

known

↳ unknown $x \in \mathbb{R}^n$.

↳ fourier

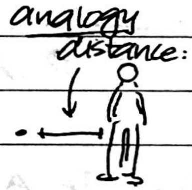
Def: Given a random scalar q, $\mathbb{E}q$ is the expected value

↳ expected value

$\mathbb{E}z_i = 0$ → expected value of z^{th} entries of noise

$\mathbb{E}z_i^2$ = noise level = 1. ∴ note: $(\mathbb{E}z_i)^2 \neq \mathbb{E}z_i^2$ →

↳ S.D. (standard deviation).



Lemma: Let $B \in \mathbb{R}^{k \times m}$.

$$\Rightarrow \mathbb{E}\,\|\underset{\substack{\uparrow \\ \text{white noise}}}{Bz}\|^2 = \sum_i \underbrace{\sigma_i^2(B)}_{\text{function of B}}$$

Pf ( - wavy)

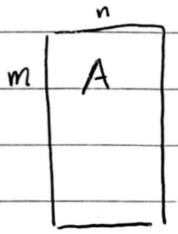
$$\mathbb{E}\,\|Bz\|^2 = \mathbb{E}\,\|\overbrace{\cancel{U}\Sigma \underbrace{V^T z}_{\sim z}}^{\text{B's SVD}}\| = \mathbb{E}\|\underbrace{\Sigma z}_{\text{diagonal case (reduces to this)}}\|^2$$

∴ noise z does not change behaviour after rotation ($V^T z$).

∴ orthogonal matrix U ~~or~~ preserves norm.

$= \mathbb{E}\sum_i (\sigma_i z_i)^2$

$= \sum_{i=1} \sigma_i^2 \cdot \cancel{\mathbb{E}z_i^2}^{\,1}$

Now, assume $m \geq n$, 

, rank(A) = n.,

Let $\hat{x} = \arg\min \| Ax' - y \|$ be least squares estimate. , $x' \in \mathbb{R}^n$

Goal: Determine $\underline{\mathbb{E}\|\hat{x} - x\|^2}$, also $\mathbb{E}\|A\hat{x} - Ax\|^2$

~ how precise is my MRI image?

↳ on average, how large is the difference between my actual data and perceived image.

Recall: $\hat{x} = (A^TA)^{-1}A^Ty = A^+y$

(from $A^TA\hat{x} = A^Ty$)

$= A^+(Ax + z)$

$= x + A^+z$

∴ projection: $A\hat{x}$

∴ when rank(A) ≠ full, multiple least squares slns exist → not good estimate

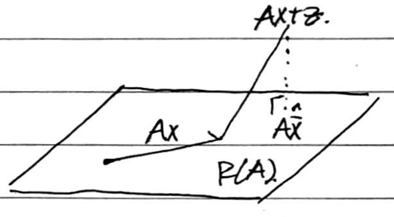
so, $\mathbb{E}\|\hat{x} - x\|^2 = \mathbb{E}\|x + A^+z - x\|^2$

$= \mathbb{E}\|A^+z\| = \sum_{i=1} (\sigma_i(A^+))^2$ by lemma,

$= \sum_{i=1}^{n} \frac{1}{\sigma_i^2(A)}$

Rmk: If A has small s.v.'s, it blows up the noise ~ badly conditioned A

Next, $\mathbb{E}\|A\hat{x} - Ax\|^2$



∴ worst case: noise lying in R(A)

∴ best case: noise orthogonal to R(A)

∴ $A\hat{x} \sim \text{proj}_{R(A)}(Ax + z)$

$= \mathbb{E}\|Ax + \underset{\substack{\downarrow \\ P}}{\underline{AA^+}}z - Ax\|^2.$

∴ if noise is highly dimensional, it is likely orthogonal to lower dimensional subspace

$= \mathbb{E}\|Pz\|^2.$

$= \sum_{i=1} \sigma_i(P)^2$

$= \sum_{i=1}^{n} 1^2 = n = \dim(R(A))$

Rmk: $\mathbb{E}\|z\|^2 = m$, $\mathbb{E}\|A\hat{x} - Ax\|^2 = m \cdot \frac{n}{m}$.

Elaboration on related research:

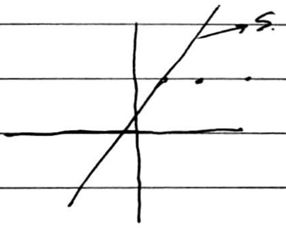
$y = Ax + z$, $x \in K$

Can we define $\sigma_i(A|_K)$ → A non-restricted to K?

→ projecting onto smaller-dimensional subspace, R(A). → kills noise

Intuition, K=S=subspace.

Is a subspace continuous? Yes, as it is defined by a set having close enough points (non-discrete)



Next class: RVW class w/ techniques on R(A), N(A) "intuitions."

Final - all-encompassing semi-proof test w/ more time