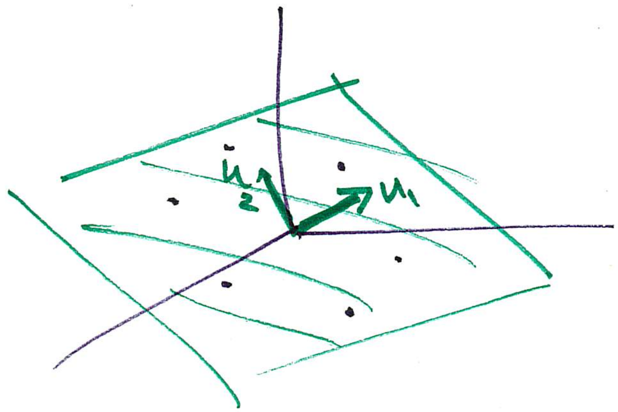
Lecture 13 - More SVD

if $\underline{x}_i \in \mathbb{R}^p$ lie in a subspace of r-dims.



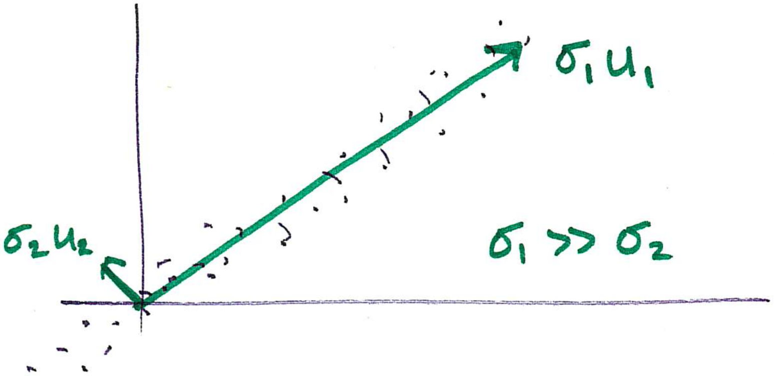
$$X = [x_1 \quad x_2 \quad \cdots \quad x_n] \in \mathbb{R}^{\underline{p \times n}}$$

$$= \underbrace{U}_{p\times r}\underbrace{\Sigma}_{r\times r}\underbrace{V^T}_{r\times n}$$

- U describes the subspace — orthog. basis
- ΣV^T describes where each x_i is within subspace.

$\sigma_1 \geq \sigma_2 \geq \cdots$

If x_i's are near subspace



$X = [x_1 \cdots x_n] \in \mathbb{R}^{p \times n}$ $(n > p)$

$\text{rank}(X) = p$

- $\underset{p \times p}{\underbrace{U}}$ gives an orthobasis for all of $\mathbb{R}^p$

for any $1 \le r \le p$

1^{st} r cols of U tells us a basis for the best r-dim subspace fit to data

- σ_i (diag elts of Σ) indicate how important each subspace dim is to representing / approximating data
- each u_i is direction of most variation among pts x_i's not already represented by $u_1, \ldots, u_{i-1}$
- 1^{st} r cols of V give locations of each x_i within subspace $[u_1, \ldots, u_r]$

Theorem (Subspace Approximation)

if $X \in \mathbb{R}^{p \times n}$ has rank $r > k$, then

$$\min_{Z\,:\,\mathrm{rank}(Z) = k} \|X - Z\|_F^2$$

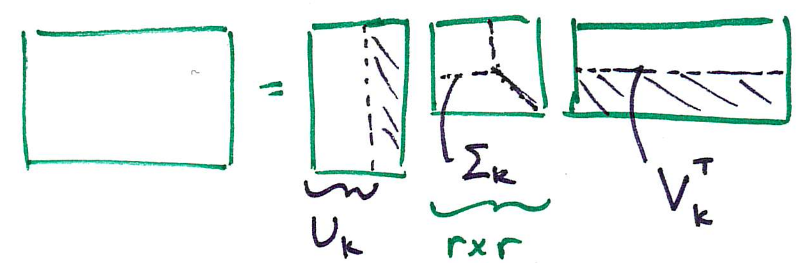
$$\|A\|_F = \Big(\sum_{i,j} A_{ij}^2\Big)^{1/2}$$

Frobenius norm

is given by $Z = X_k = U_k \Sigma_k V_k^T$

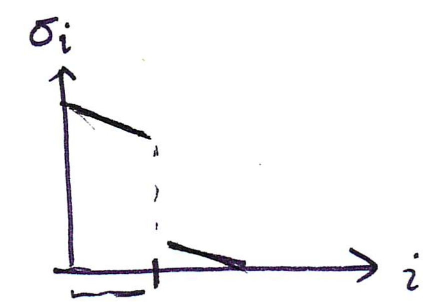
and $\|X - X_k\|_F^2 = \sum_{i=k+1}^{r} \sigma_i^2$

$$X = U \Sigma V^T$$



$$\Sigma = \begin{bmatrix} 100 & & & & & \\ & 90 & & & & \\ & & 80 & & & \\ & & & 3 & & \\ & & & & 2 & \\ & & & & & 1 \end{bmatrix}$$

"spectrum" of a matrix — ordered singular values



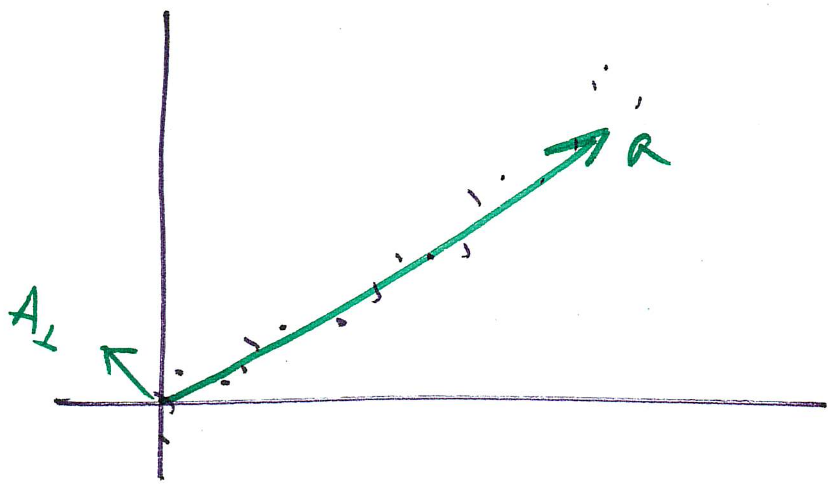
$$\frac{\sum_{i=k+1}^{r} \sigma_i^2}{\sum_{i=1}^{r} \sigma_i^2}$$

Last time:

have points $x_1, \dots, x_n \in \mathbb{R}^p$

find 1-d subspace (line) closest to points

line / subspace by vec $\underset{\sim}{a}$



$$d_i^2 = \| x_i - P_a x_i \|_2^2$$

$$= x_i^T \underbrace{\left(I - \frac{aa^T}{a^Ta} \right)^2}_{\equiv P_{A_\perp}} x_i$$

• why is $\left(I - \frac{aa^T}{a^Ta}\right)$ a projection matrix?

a matrix P is a proj matrix iff $P = P^2$

$$\left(I - \frac{aa^T}{a^Ta}\right)^2 = I - \frac{2aa^T}{a^Ta} + \frac{a\cancel{a^Ta}a^T}{a^Ta\,\cancel{a^Ta}} = I - \frac{aa^T}{a^Ta}$$

- how do we know that $I - \frac{aa^T}{a^Ta}$ projects onto $A_\perp$?

let S = subspace (r-dim) $= \{ u \in \text{span}(\text{cols}(A)) \}$ — A: pxr, LI cols; basis for subspace

$= \{ u = Aw,\ w \in \mathbb{R}^r \} = \{ u : A_\perp^T u = 0 \}$

$S_\perp$ = orthogonal complement to S

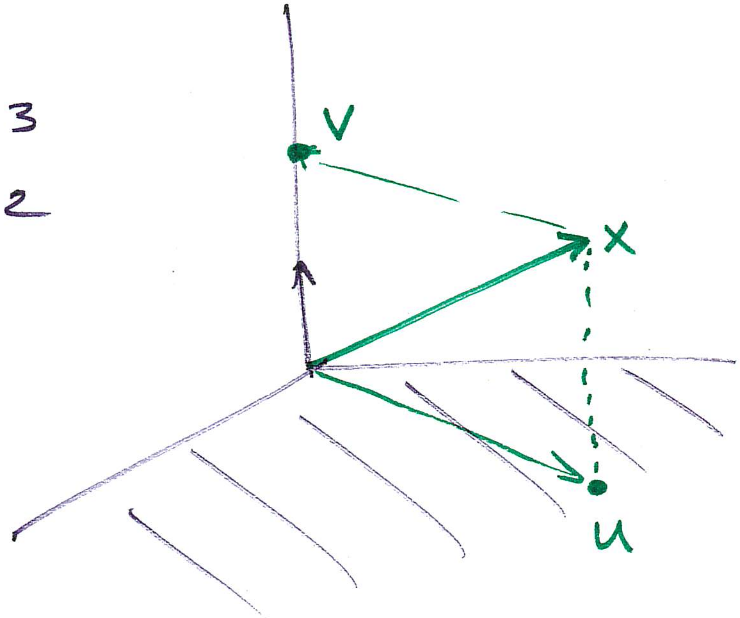
$= \{ v : u^T v = 0 \ \forall\ u \in S \}$

$= \{ v : \underline{A^T v = 0} \} = \{ v : A_\perp w,\ w \in \mathbb{R}^{p-r} \}$

let $\underline{A_\perp}$ be a basis for $S_\perp \implies A^T A_\perp = 0$

$p = 3$

$r = 2$



$$\operatorname{span}(\operatorname{col}([A \quad A_\perp])) = \mathbb{R}^p$$

for any $x \in \mathbb{R}^p$

$$\underset{}{x} = \underset{\in S}{u} + \underset{\in S_\perp}{v}$$

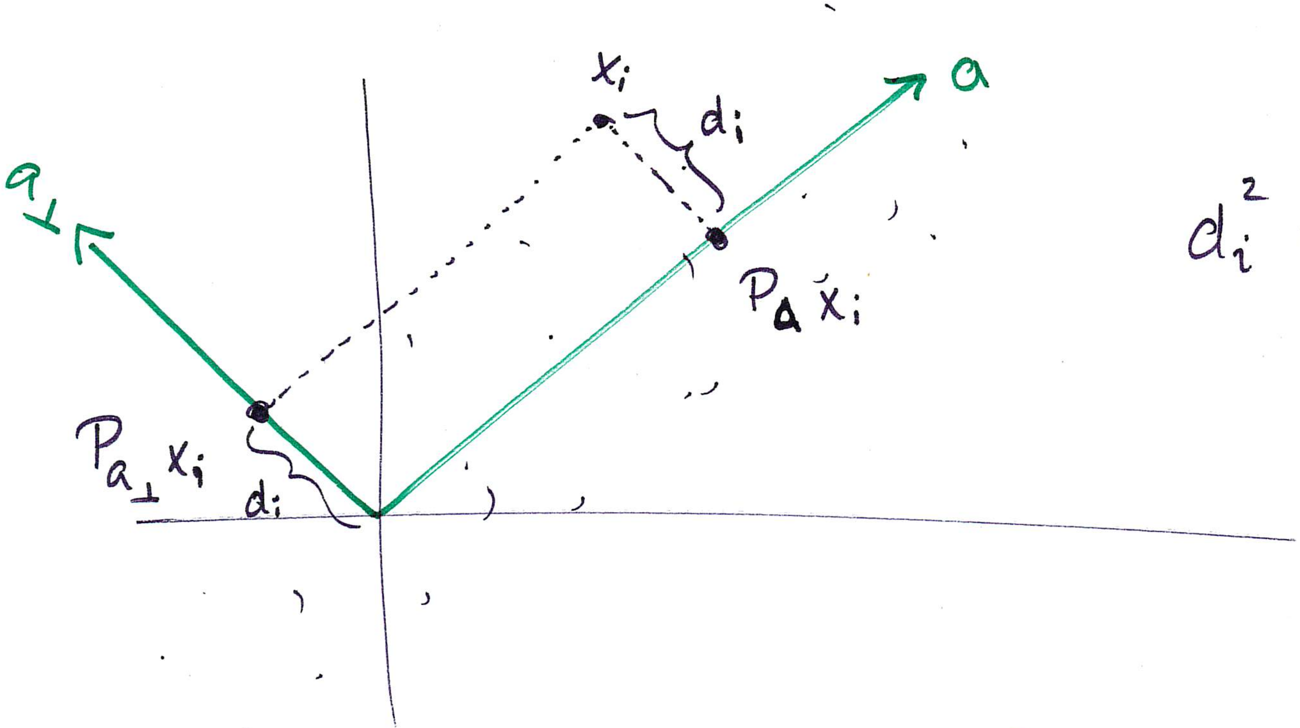
$$P_A x = A(A^TA)^{-1}A^T x$$

$$= A(A^TA)^{-1}A^T \underbrace{u}_{Aw} + A(A^TA)^{-1}\underbrace{A^T v}_{0}$$

$$= A\cancel{(A^TA)^{-1}A^TA}\, w = Aw = u$$

$$P_{A_\perp} x = (I - P_A)x = x - u = v$$

back to line fit



$$d_i^2 = x_i^T P_{a_\perp} x_i$$

let $X = [x_1 \ x_2 \ \cdots \ x_n] \in \mathbb{R}^{p \times n}$

$$\min_a \sum_i d_i^2 = \max_{a \neq 0} \frac{\|X^T a\|_2^2}{\|a\|_2^2} = \|X^T\|_2^2 = \|X\|_2^2 = \sigma_1^2$$

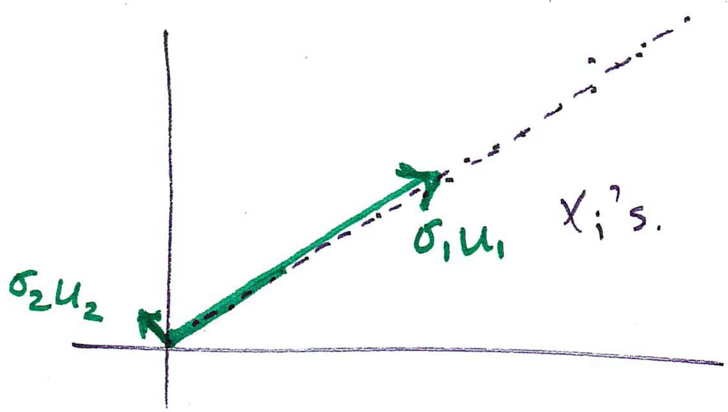
line a that achieves this is 1^{st} left singular vec of X

($u_1 = 1^{st}$ col of U is $X = U\Sigma V^T$)

example:

$p = 2$

$$X_i = \alpha_i \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \beta_i \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\text{basis for } \mathbb{R}^2} \underbrace{\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix}}_{\text{basis coeffs.}}$$



$\alpha_i \approx \beta_i \quad \forall \, i$

$$X = [x_1, x_2, \dots x_n] = U \Sigma V^T$$

$$= \underbrace{\frac{\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}}{\sqrt{2}}}_{U} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1/100 \end{bmatrix}}_{\Sigma} \underbrace{\frac{\begin{bmatrix} 1 & 3 & -1 & -2 \\ 1 & -1 & 3 & 2 \end{bmatrix}}{\sqrt{15}}}_{V^T}$$

u_1 = 1st left sing vec

u_2

$$x_i = \frac{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}{\sqrt{2}} 1 \cdot v_{i1} + \frac{\begin{bmatrix} -1 \\ 1 \end{bmatrix}}{\sqrt{2}} \frac{1}{100} v_{i2}$$

$$\approx \frac{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}{\sqrt{2}} v_{i1}$$

LS Regression / Classifacation + Tikhinov Reg. + SVD

imagine $y_i = \langle x_i, w \rangle + \varepsilon_i$

$$\Downarrow$$

$$y = X^T w + \varepsilon = A w + \varepsilon$$

$$X = [x_1 \cdots x_n] \in \mathbb{R}^{p \times n} = U \Sigma V^T$$

Let $A = X^T = V \Sigma U^T$ ($p \leq n$, cols A L.I.)

LS: $$\hat{w} = (A^T A)^{-1} A^T y$$

$$= (U \Sigma \underbrace{\cancel{V^T V}}_{I} \Sigma U^T)^{-1} U \Sigma V^T y$$

$$= (U \Sigma^2 U^T)^{-1} U \Sigma V^T y = U \Sigma^{-1} V^T y$$

$$= \cancel{U \Sigma^{-1} V^T V \Sigma U^T} w + U \Sigma^{-1} V^T \varepsilon$$

$$= w + U \underline{\Sigma^{-1}} \, \underline{V^T \varepsilon}$$

$$\Sigma^{-1} \to \begin{bmatrix} 1 & 0 \\ 0 & 100 \end{bmatrix} \leftarrow \text{noise amplification.}$$

Tikhinov Reg.

$$\hat{w} = (A^TA + \lambda I)^{-1} A^T y$$

$$= (U\Sigma V^T V \Sigma U^T + \lambda I)^{-1} U \Sigma V^T y$$

$$= U \underbrace{\Sigma (\Sigma^2 + \lambda I)^{-1}}_{\begin{bmatrix} \frac{\sigma_1}{\sigma_1^2+\lambda} & & & \\ & \frac{\sigma_2}{\sigma_2^2+\lambda} & & \\ & & \ddots & \\ & & & \frac{\sigma_p}{\sigma_p^2+\lambda} \end{bmatrix}} V^T y$$

$$= \underbrace{U \frac{\Sigma^2}{\Sigma^2 + \lambda I} U^T w}_{\text{want } \approx w} + \underbrace{U \frac{\Sigma}{\Sigma^2 + \lambda I}}_{\text{mitigate noise amplification}} V^T \varepsilon$$

~~mitiggat~~

e.g. $\Sigma = \begin{bmatrix} 100 & & & & & \\ & 90 & & & & \\ & & 80 & & & \\ & & & 10^{-1} & & \\ & & & & 10^{-2} & \\ & & & & & 10^{-10} \end{bmatrix} \Rightarrow \Sigma^{-1} = \begin{bmatrix} 1/100 & & & & & \\ & 1/90 & & & & \\ & & 1/80 & & & \\ & & & 10 & & \\ & & & & 100 & \\ & & & & & 10^{10} \end{bmatrix}, \frac{\Sigma}{\Sigma^2 + \lambda I} \approx \begin{bmatrix} 1/100 & & & & & \\ & 1/90 & & & & \\ & & 1/80 & & & \\ & & & 10 & & \\ & & & & 50 & \\ & & & & & 100 \end{bmatrix}$