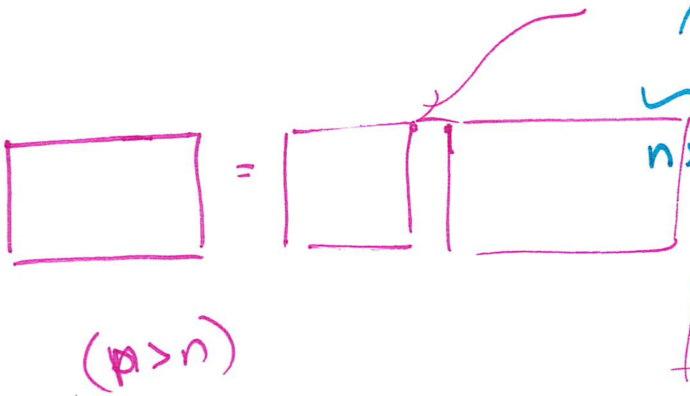
Lecture 12 – SVD and Principal Components Analysis (PCA)

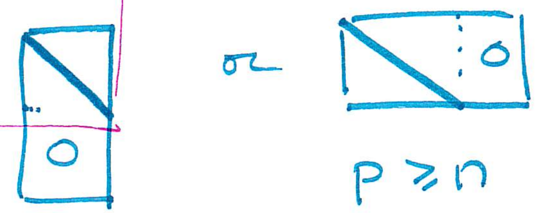
Recall the SVD: for any $X \in \mathbb{R}^{n \times p}$, there exists factorization

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

- $U^T U = U U^T = I$
- $V^T V = V V^T = I$
- Σ is diagonal



$(p > n)$

$\Sigma =$  or

$n \geq p$ $\qquad$ $p \geq n$

diagonal elements: $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq \sigma_{\min(n,p)}$

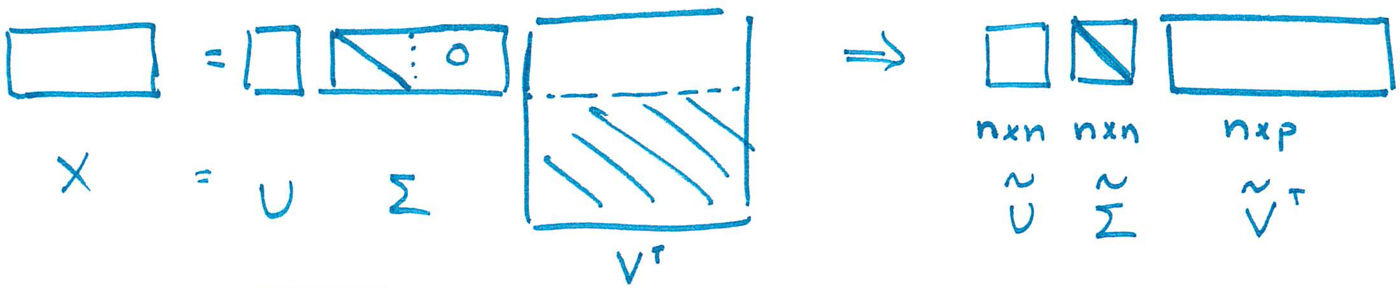
"singular values of X"

cols of U = "left singular vectors of X"

cols of V = "right singular vectors of X"

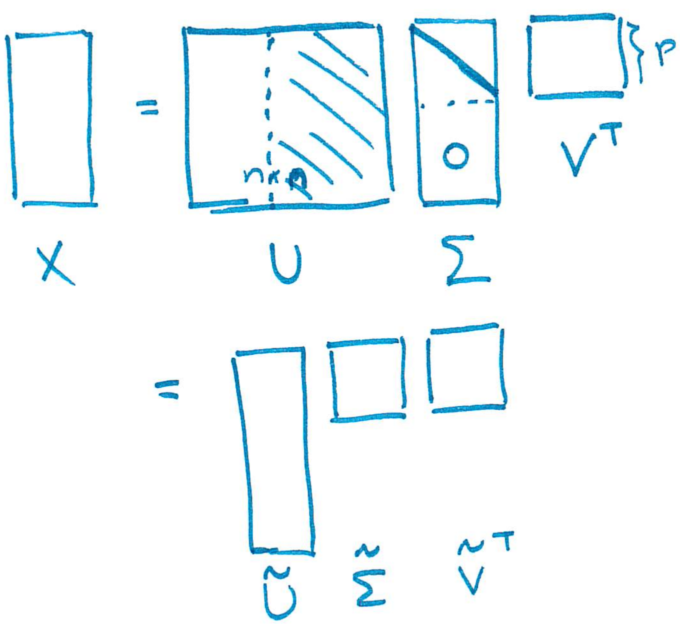
Alternative:

if $p > n$



- $\tilde{U}^T \tilde{U} = \tilde{U} \tilde{U}^T = I$ $(U = \tilde{U})$
- $\tilde{\Sigma} = \tilde{\Sigma}^T$
- $\tilde{V}^T \tilde{V} = I \neq \tilde{V} \tilde{V}^T$

if $n > p$

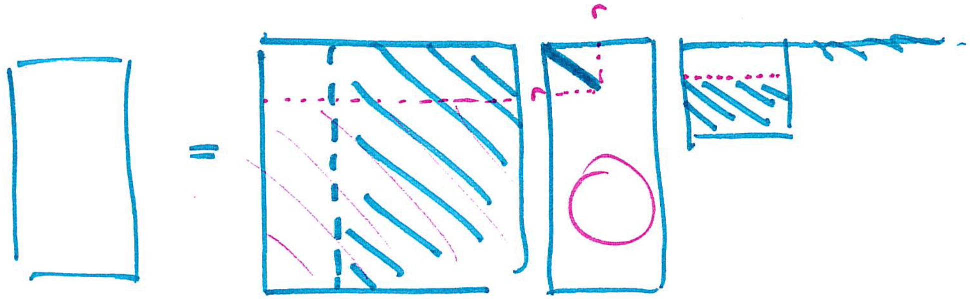


- $\tilde{V} = V$, $\tilde{V}^T \tilde{V} = \tilde{V} \tilde{V}^T = I$
- $\tilde{\Sigma} = \tilde{\Sigma}^T$
- $\tilde{U}^T \tilde{U} = I \neq \tilde{U} \tilde{U}^T$

what if $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_{\min(p,n)} = 0$

(ie. $\sigma_1, \dots, \sigma_r > 0$) for some $r < \min(p,n)$

imagine $n \geq p > r$



$$\underset{}{X} = \underset{n\times n}{U}\ \underset{n\times p}{\Sigma}\ \underset{p\times p}{V^T}$$

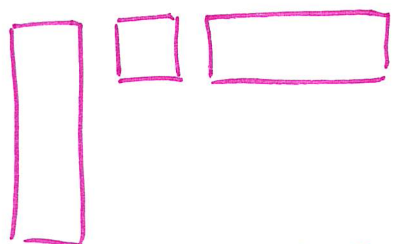
here:

$\tilde{U}^T \tilde{U} = I \neq \tilde{U}\tilde{U}^T$

$\tilde{V}^T \tilde{V} = I \neq \tilde{V}\tilde{V}^T$

$\tilde{\Sigma} = \tilde{\Sigma}^T$, diag elts > 0, $\tilde{\Sigma}$ invertible.

$\tilde{U}$ = basis for subspace spanned by cols of X



$$= \underset{n\times r}{\tilde{U}}\ \underset{r\times r}{\tilde{\Sigma}}\ \underset{r\times p}{\tilde{V}^T}$$

$\Longrightarrow \text{rank}(X) = r$

from Last time:

$$\hat{w} = \underbrace{(X^T X)^{-1} X^T}\, y$$

assume $X \in \mathbb{R}^{n \times p}$, $n \geq p$, cols X lin. indep
rank$(X) = p$.

if $X = U\Sigma V^T$ (Σ is square, $\Sigma = \Sigma^T$)

$U^T U = I$
$V^T V = V V^T = I$
$\Rightarrow V^T = V^{-1}$

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

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

$(ABC)^{-1} = C^{-1} B^{-1} A^{-1}$

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

$$= V \Sigma^{-2} \cancel{\underbrace{V^T V}_{I}} \Sigma U^T$$

$$= V \Sigma^{-1} U^T$$

if $\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_p \end{bmatrix}$

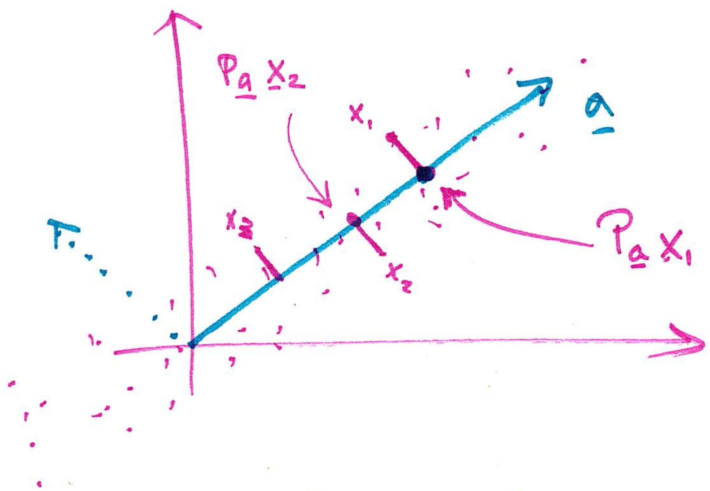
$\Sigma^{-2} = \begin{bmatrix} 1/\sigma_1^2 & & & \\ & 1/\sigma_2^2 & & \\ & & \ddots & \\ & & & 1/\sigma_p^2 \end{bmatrix}$

$$\hat{y} = X \hat{w} = \underbrace{X (X^T X)^{-1} X^T}_{\substack{\text{Projection matrix} \\ P_X}} y$$

$$\underline{P_X} = U \Sigma \cancel{\underbrace{V^T V}_{I}} \Sigma^{-1} U^T = U \cancel{\underbrace{\Sigma \Sigma^{-1}}_{I}} U^T = \underbrace{\underline{UU^T}}_{\text{we've seen this before}} = P_U$$

Geometric Interpretation

example: find line closest to set of points $\underline{x}_1, \ldots, \underline{x}_n \in \mathbb{R}^p$



find $\underline{a}$ to minimize sum of squared distances:

($\underline{a}$ defines a subspace)

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

$$P_{\underline{a}} = \underline{a}\,\underbrace{(\underline{a}^T\underline{a})^{-1}}_{\text{scalar}}\,\underline{a}^T$$

$$P_A^2 = A(A^TA)^{-1}\cancel{A^TA(A^TA)^{-1}}A^T$$

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

$$\boxed{P_A^2 = P_A}$$

dist from pt $\underline{x}_i$ to line $\underline{a}$:

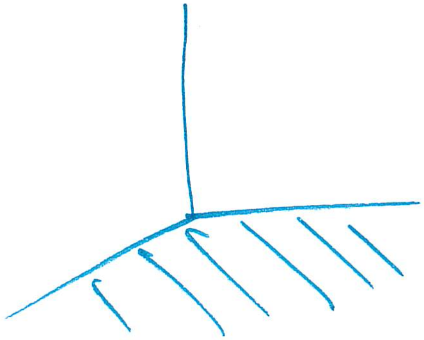
$$d_i^2 = \| \underline{x}_i - P_{\underline{a}} \underline{x}_i \|_2^2$$

$$= \| x_i - \underline{a}(a^Ta)^{-1}a^T x_i \|_2^2$$

$$= \| \underline{x}_i - \underline{a} \frac{a^T \underline{x}_i}{a^Ta} \|_2^2$$

$$= \| \left(I - \frac{aa^T}{a^Ta} \right) x_i \|_2^2$$

$$\left\|\left(I - \frac{aa^T}{a^Ta}\right)x_i\right\|_2^2 = x_i^T \underbrace{\left(I - \frac{aa^T}{a^Ta}\right)^T}_{\substack{\text{also a} \\ \text{Projection matrix} \\ \text{(Proj onto orthogonal} \\ \text{complement of } a\text{)}}} \left(I - \frac{aa^T}{a^Ta}\right) x_i$$



$$= x_i^T \left(I - \frac{aa^T}{a^Ta}\right) x_i$$

$$= x_i^T x_i - \frac{\cancel{x_i}^T a a^T \cancel{x_i}}{a^Ta}$$

$$= \underline{a^T x_i}\, \underline{x_i^T a}$$

want to minimize

$$\sum_{i=1}^{n} d_i^2 = \sum_i \left(\underbrace{x_i^T x_i}_{\text{constant with respect to } \underset{\sim}{a}} - \cancel{x_i^T}\, \frac{a^T x_i x_i^T a}{a^Ta} \right)$$

$$\underset{a}{\text{minimizing}} \sum_i d_i^2 \equiv \underset{a}{\text{maximizing}} \sum_i \frac{a^T x_i x_i^T a}{a^T a}$$

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

$$\sum_i \frac{a^T x_i x_i^T a}{a^T a} = \frac{a^T X X^T a}{a^T a}$$

$$\max_{a \neq 0} \frac{a^T X X^T a}{a^T a} = \max_{a \neq 0} \frac{\|X^T a\|_2^2}{\|a\|_2^2} = \|X^T\|_2^2 \equiv \|X^T\|_{op}^2$$

$$= \|X\|_2^2 = \sigma_1^2$$

value of a that maximizes $\dfrac{\|X^T a\|_2^2}{\|a\|_2^2}$ = 1st right singular vector of X^T

= 1st left singular vector of X

$X \in \mathbb{R}^{p \times n}$ (n points in p-dim space, each col of X = coords of one point)

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

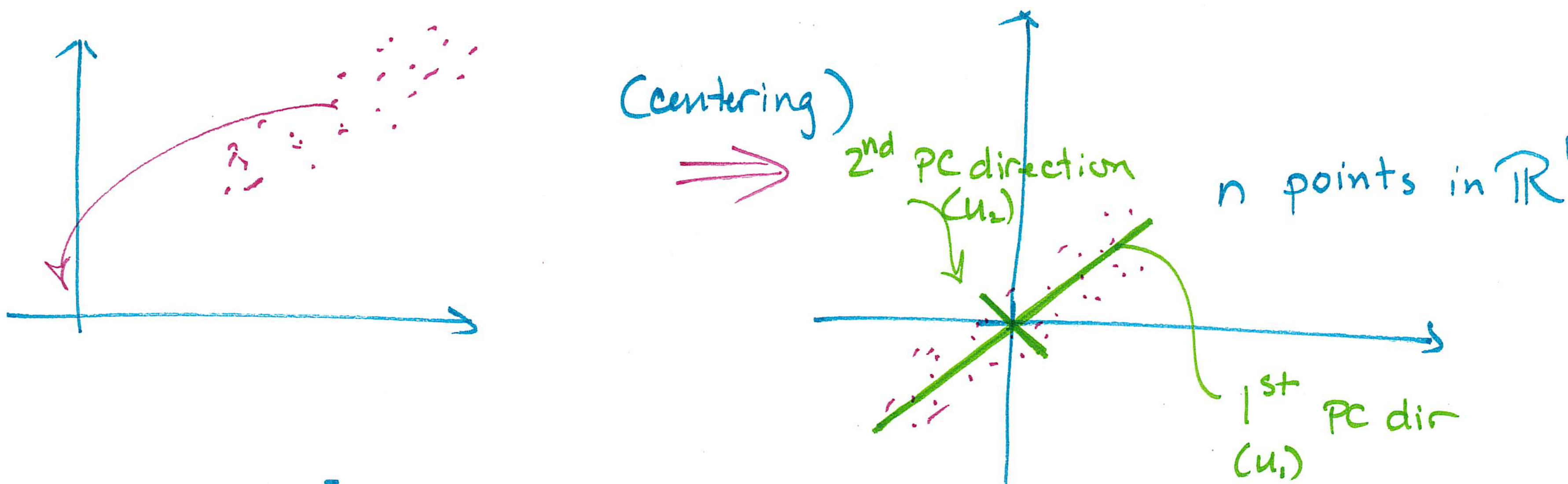
U gives a basis of $\mathbb{R}^p$ for our points

but U is a special basis:

1st basis vector is ~~thest~~ the best 1-d subspace fit to data

Principal components analysis (PCA)

Let $X \in \mathbb{R}^{p \times n}$ be data matrix, w/ rows "centered" to have average value of 0



Let $X = U \Sigma V^T$. Left singular vectors of X are called to "principal component directions" of X

u_1 = best 1d subspace approx/fit to all data

$\tilde{x}_i = x_i - P_{u_1} x_i$ = i^{th} residual

u_2 = best 1d subspace approx/fit to all $\tilde{x}_i$ (residuals)

$\tilde{\tilde{x}}_i = \tilde{x}_i - P_{u_2} \tilde{x}_i$ = i^{th} 2^{nd} resid

u_3 = best 1d subspace approx/fit to all $\tilde{\tilde{x}}_i$

Gram-Schmidt

$u_1 = x_1 / \|x_1\|_2$

$\tilde{x}_2 = x_2 - P_{u_1} x_1$

$u_2 = \tilde{x}_2 / \|\tilde{x}_2\|_2$

$\vdots$

ordering of pts matters

PCA

u_1 = best fit to all data

$\tilde{X}$ = residual matrix

u_2 = best fit to all $\tilde{X}$

order does not matter