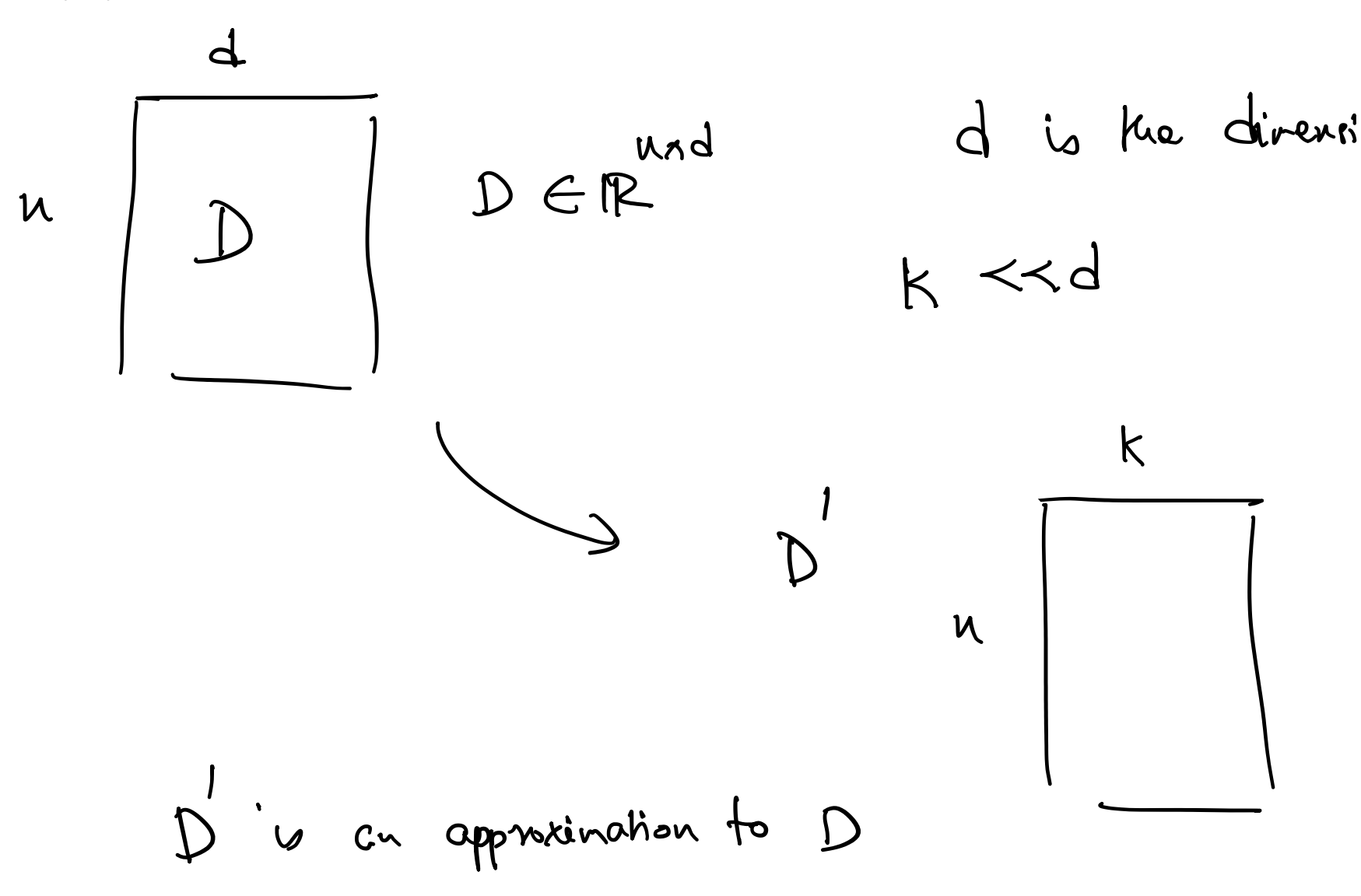
Lecture 5

Monday, September 16, 2024 10:02 AM



$D \in \mathbb{R}^{n \times d}$

d is the dimensionality

$k << d$

D' is an approximation to D

dimensionality reduction

---

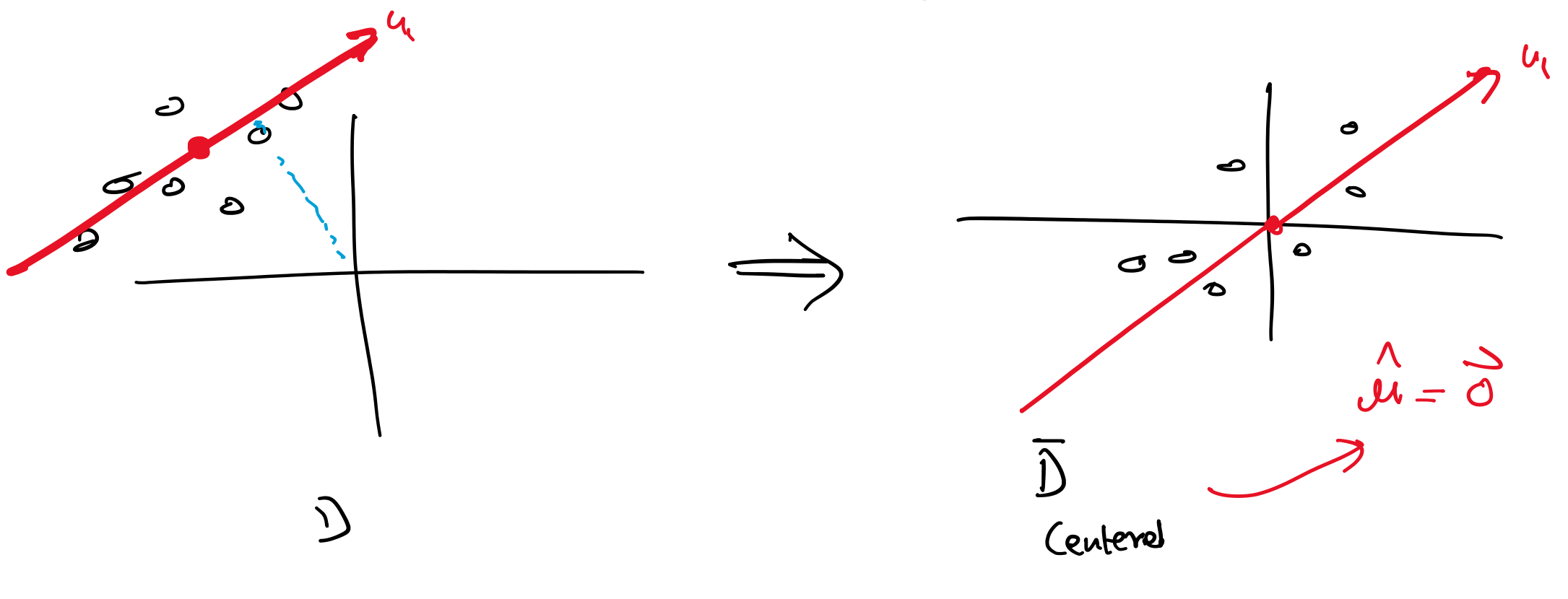
k=1, Q: what is the best "column" (MSE objective)

U_1 = linear combination of $A_1, A_2, \ldots, A_d$

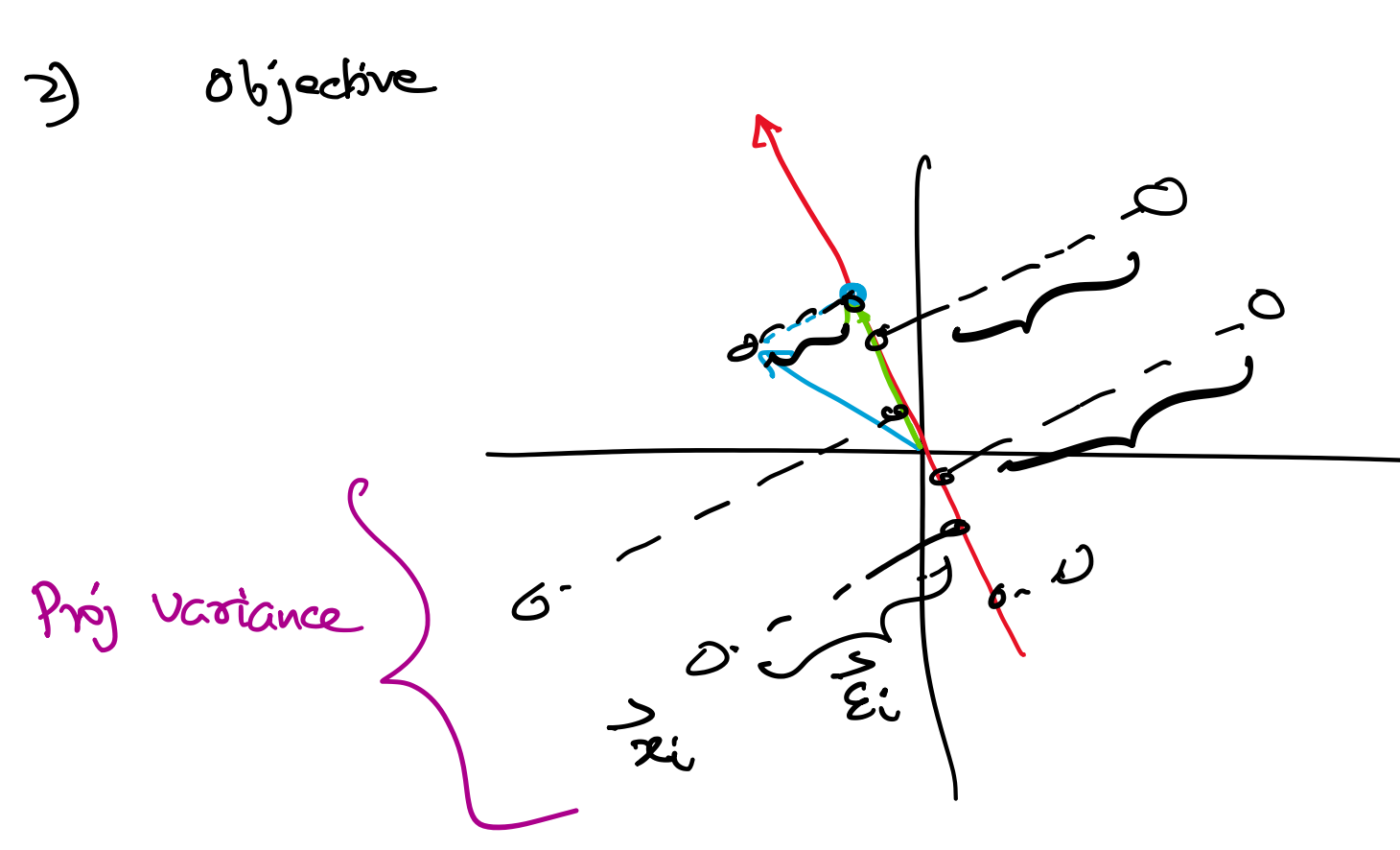
$= c_1 A_1 + c_2 A_2 + \ldots + c_d A_d$

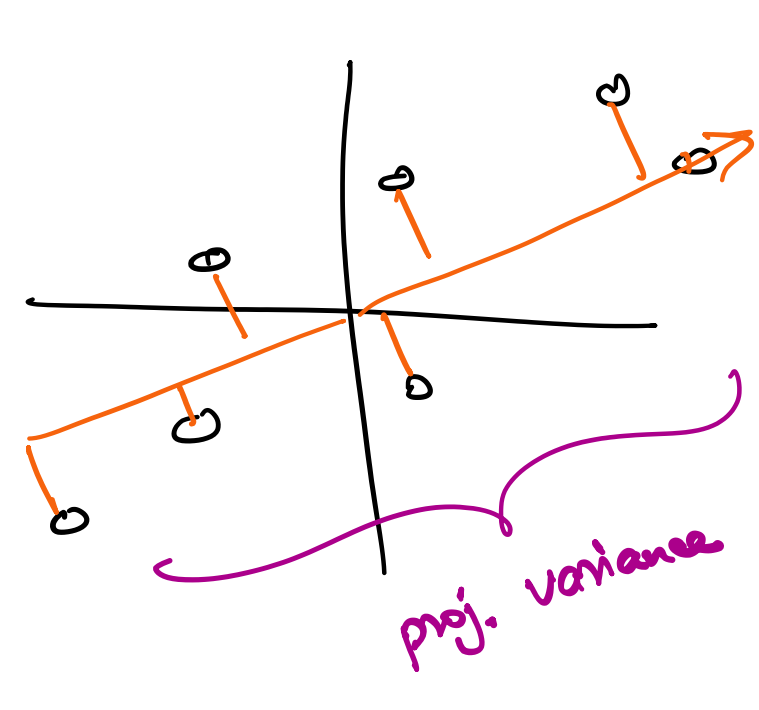
$\vec{u}_1 = (c_1 c_2 \ldots c_d)^T$ $\qquad \|u_1\| = 1$

eigenvector corresponding to largest eigenvalue or dominant eigenvector of Σ



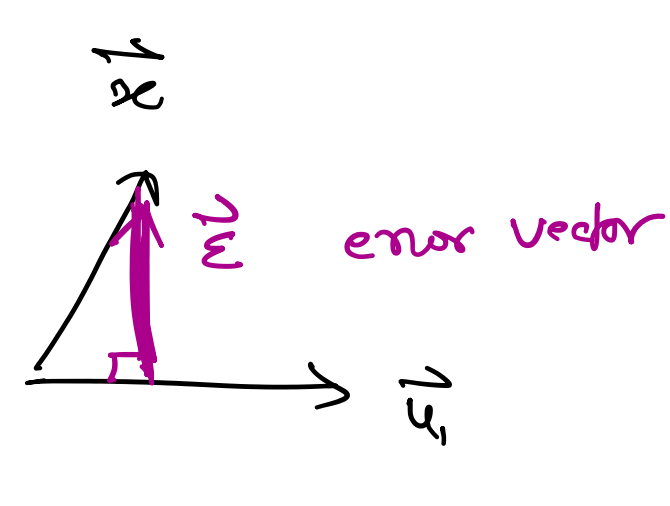
1) center the data matrix
2) objective





$$\frac{1}{n}\sum_{i=1}^{n}\|\vec{\varepsilon}_i\|^2 \equiv \text{minimum mean squared error}$$

$\equiv$ {maximum projected variance



$$MSE = \frac{1}{n}\sum_{i=1}^{n}\|\vec{\varepsilon}_i\|^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\|\vec{x}_i - \text{proj}_{\vec{u}}(\vec{x}_i)\,\vec{u}\|$$

$\vdots$

$\vec{x}_i$ are centered

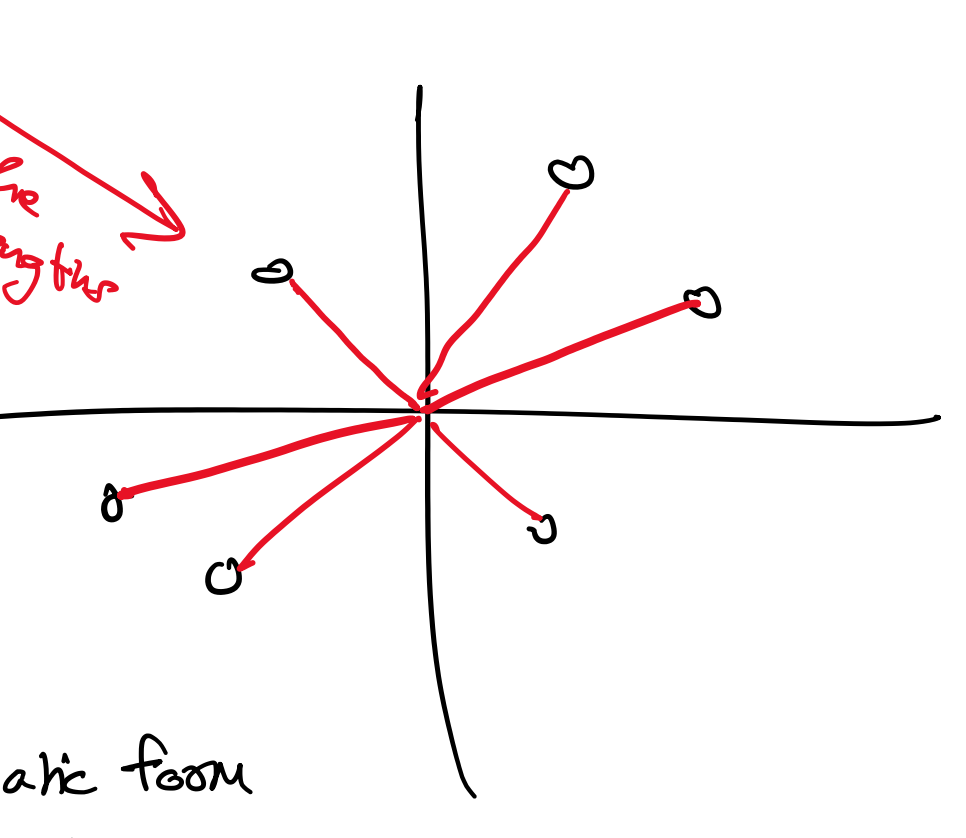
$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n}\|\vec{x}_i\|^2}_{\frac{1}{n}\sum\|\vec{x}_i - \hat{\mu}\|^2} - \vec{u}^T\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\vec{x}_i\vec{x}_i^T\right)}_{\Sigma \text{ Cov matrix}}\vec{u}$$

$\vec{x}_i = \vec{x}_i - \hat{\mu}$ ← centered

$var(D) \equiv$ total variance

fixed given D

Sum of the squared length



$$\min_{\vec{u}} MSE = \boxed{Var(D)} - \underbrace{\vec{u}^T\Sigma\vec{u}}$$

quadratic form (scalar)

$$\equiv \max_{\vec{u}} \vec{u}^T\Sigma\vec{u}$$

Constraint $\vec{u}^T\vec{u} = 1$ $\quad (\|\vec{u}\|^2 = 1)$

I'm only interested in the direction

final objective $\max_{\vec{u}} \underbrace{\vec{u}^T\Sigma\vec{u}}_{\text{primary objective}}$, such that $\underbrace{\vec{u}^T\vec{u} = 1}_{\text{Constraint}}$

Constrained optimization

$\Downarrow$

unconstrained objective

$\alpha\left(\underbrace{\vec{u}^T\vec{u} - 1}\right) = 0$

scalar (lagrange multiplier) $\alpha > 0$

$$\max_{\vec{u}} J = \vec{u}^T\Sigma\vec{u} - \alpha(\vec{u}^T\vec{u} - 1)$$

$$\boxed{\frac{\partial J}{\partial \alpha} = \vec{u}^T\vec{u} - 1 = 0 \Rightarrow \vec{u}^T\vec{u} = 1}$$

$$\frac{\partial J}{\partial \vec{u}} = 2\Sigma\vec{u} - 2\alpha\vec{u} = 0$$

$\frac{\partial}{\partial \vec{u}} \vec{u}^T\Sigma\vec{u} = 2\Sigma\vec{u}$

$\frac{\partial}{\partial \vec{u}} \vec{u}^T\vec{u} = 2\vec{u}$

$\boxed{\Sigma\vec{u} = \alpha\vec{u}}$ eigenvector equation

direction $\vec{u}$ must be an eigenvector of Σ

Σ has d of them

$$\Sigma \begin{cases} \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d \geq 0 \\ u_1 \quad u_2 \quad \ldots \quad u_d \end{cases}$$

max $\qquad$ min

$$\min_{\vec{u}} MSE = var(D) - \vec{u}^T\Sigma\vec{u}$$

$\searrow \max_{\vec{u}} \vec{u}^T\Sigma\vec{u}$

$\Sigma\vec{u} = \alpha\vec{u}$

case I: $\vec{u} = \vec{u}_1$

$\alpha = \lambda_1$

$\Sigma\vec{u}_1 = \lambda_1\vec{u}_1$

$\vec{u}_1^T\Sigma\vec{u}_1$

$= \vec{u}_1^T(\Sigma\vec{u}_1)$

$= \vec{u}_1^T(\lambda_1\vec{u}_1)$

$= \lambda_1 u_1^T u_1$

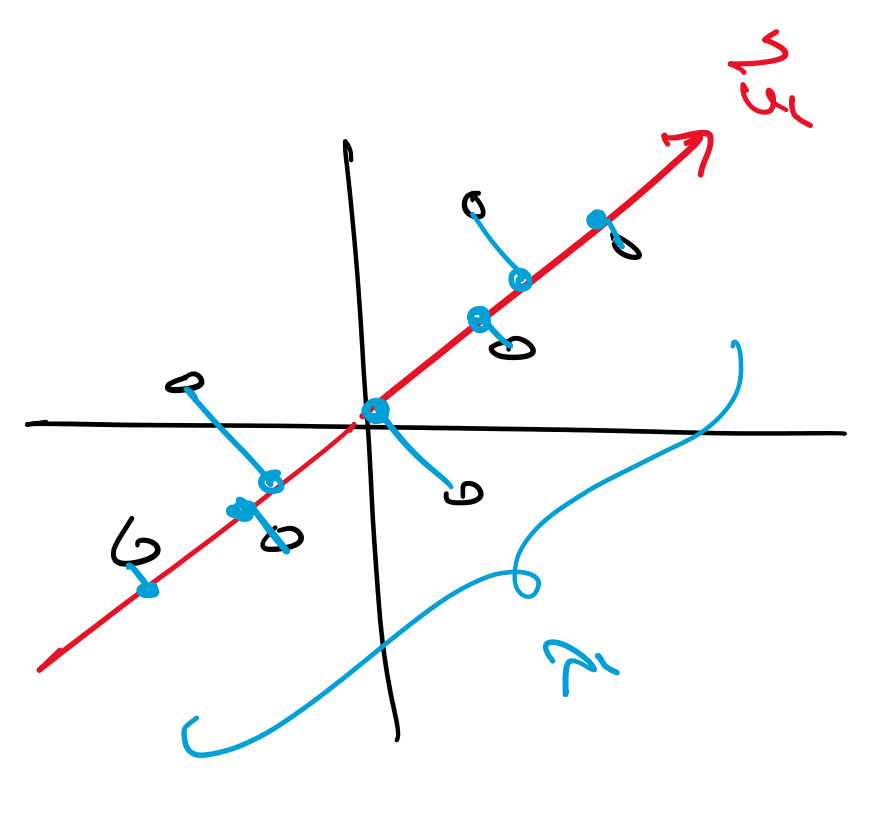
$= \lambda_1$

$MSE = Var(D) - \vec{u}_1^T\Sigma\vec{u}_1$

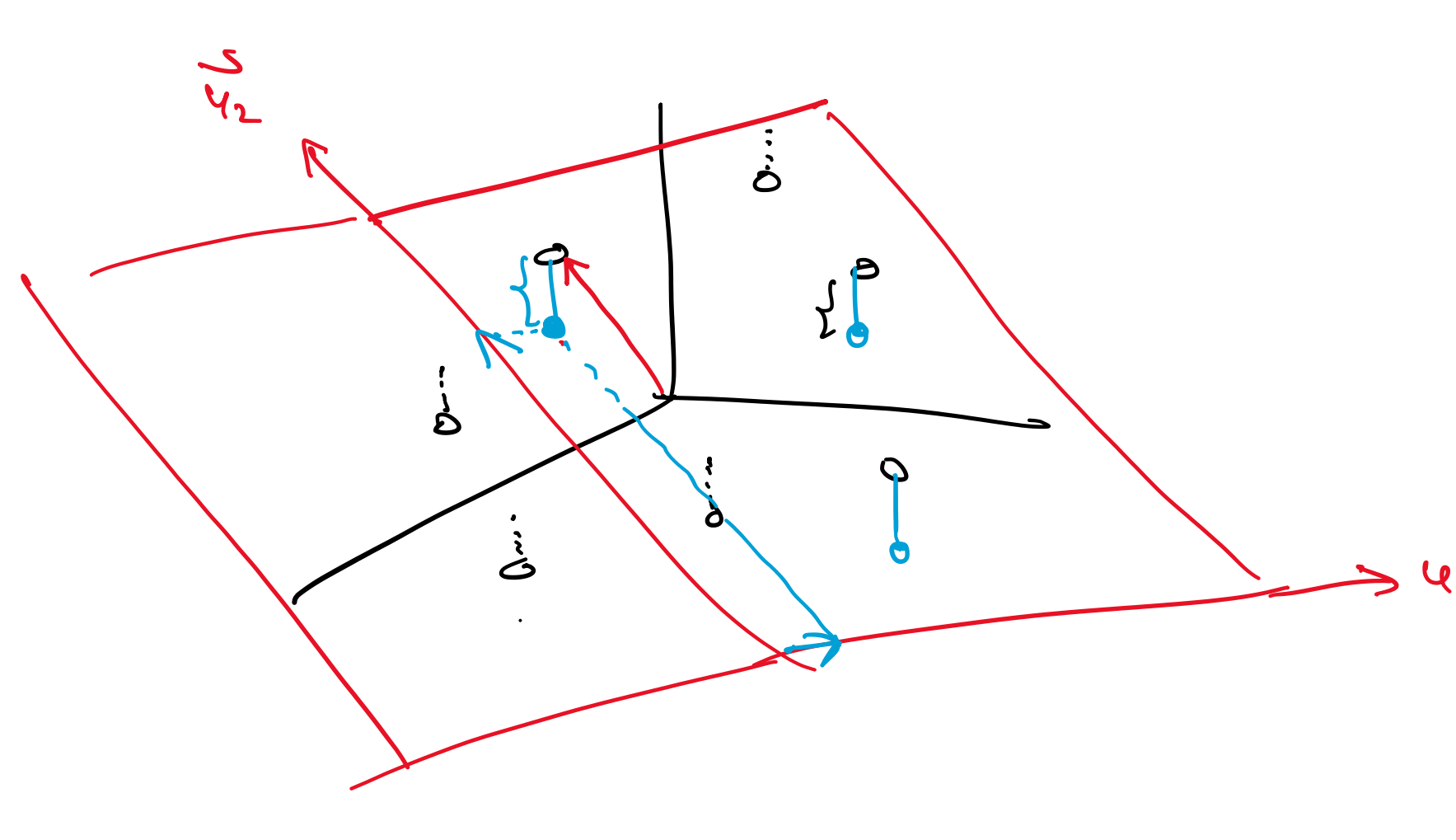
$$\boxed{MSE = \underbrace{Var(D)}_{\text{total Variance}} - \underbrace{\lambda_1}_{\text{projected variance}}}$$

PCA: Choose k directions that minimize MSE

e.g. k=2



projected Variance along u_1
a) scalar projections
b) compute the variance $\equiv \lambda_1$



k=2 ⇒ find $\vec{u}_1$ and $\vec{u}_2$

$\lambda_1 \quad \lambda_2$

$$\boxed{MSE = Var(D) - \lambda_1 - \lambda_2}$$

PCA: principal Component Analysis

1) $D \rightarrow \bar{D}$ center D
2) solve $\Sigma\vec{u} = \lambda\vec{u}$ (eg use eigh)

$$\begin{array}{ccccc} \lambda_1 \geq & \lambda_2 \geq & \lambda_3 \geq & \ldots \geq & \lambda_d \geq 0 \\ | & | & | & & | \\ \vec{u}_1 & \vec{u}_2 & \vec{u}_3 & & \vec{u}_d \end{array}$$

3) what k should be used? how many "dimensions"?

$\theta = 0.9$ (user defined)

smallest k that capture θ fraction of the total variance

$\bar{D} \rightarrow Var(\bar{D}) \equiv$ total variance $\quad \frac{1}{n}\sum\|\vec{x}_i - \hat{\mu}\|^2$

k	1	2	3	4				- -	-	d
f_k	0.81	0.85	0.91							

$$f_k = \frac{\sum_{i=1}^{k}\lambda_i}{Var(D)} = \frac{Var(k\ dim)}{total\ Var} \geq \theta$$

find the smallest k so that $f_k \geq \theta$

$$\boxed{Var(D) = \lambda_1 + \lambda_2 + \ldots + \lambda_d = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_d^2}$$

Sum of original variances along $A_1\ A_2 \ldots A_d$ is equal to sum of projected variances along $\vec{u}_1\ \vec{u}_2 \ldots \vec{u}_d$

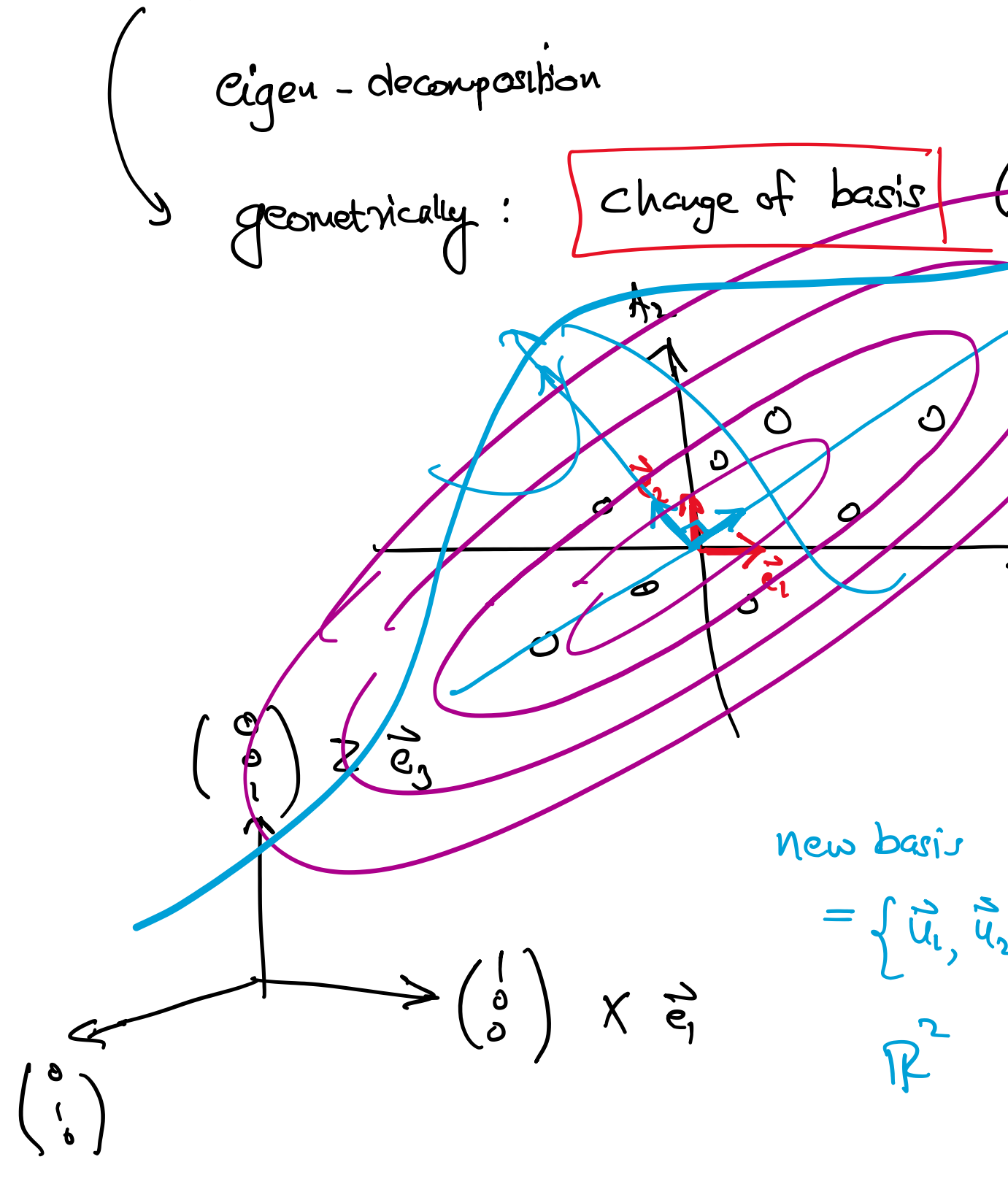
equal to ~~total~~ total variance (D)

PCA ≡ normal distribution

eigen-decomposition

geometrically: $\boxed{\text{change of basis}}$ (rotation of the basis)

↳ after centering


A_2
A_1
$\vec{e}_1$
$\vec{e}_2$
$\begin{pmatrix}0\\1\\0\end{pmatrix}$
$\begin{pmatrix}1\\0\\0\end{pmatrix}$
$\begin{pmatrix}0\\0\\1\end{pmatrix}$
$\vec{e}_3$
Y
$\vec{e}_2$

original axes = $\{A_1, A_2\}$

basis for $\mathbb{R}^2$ $= \left\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right\}$

$\vec{e}_1 \quad \vec{e}_2$

new basis $= \{\vec{u}_1, \vec{u}_2\}$ $\mathbb{R}^2$

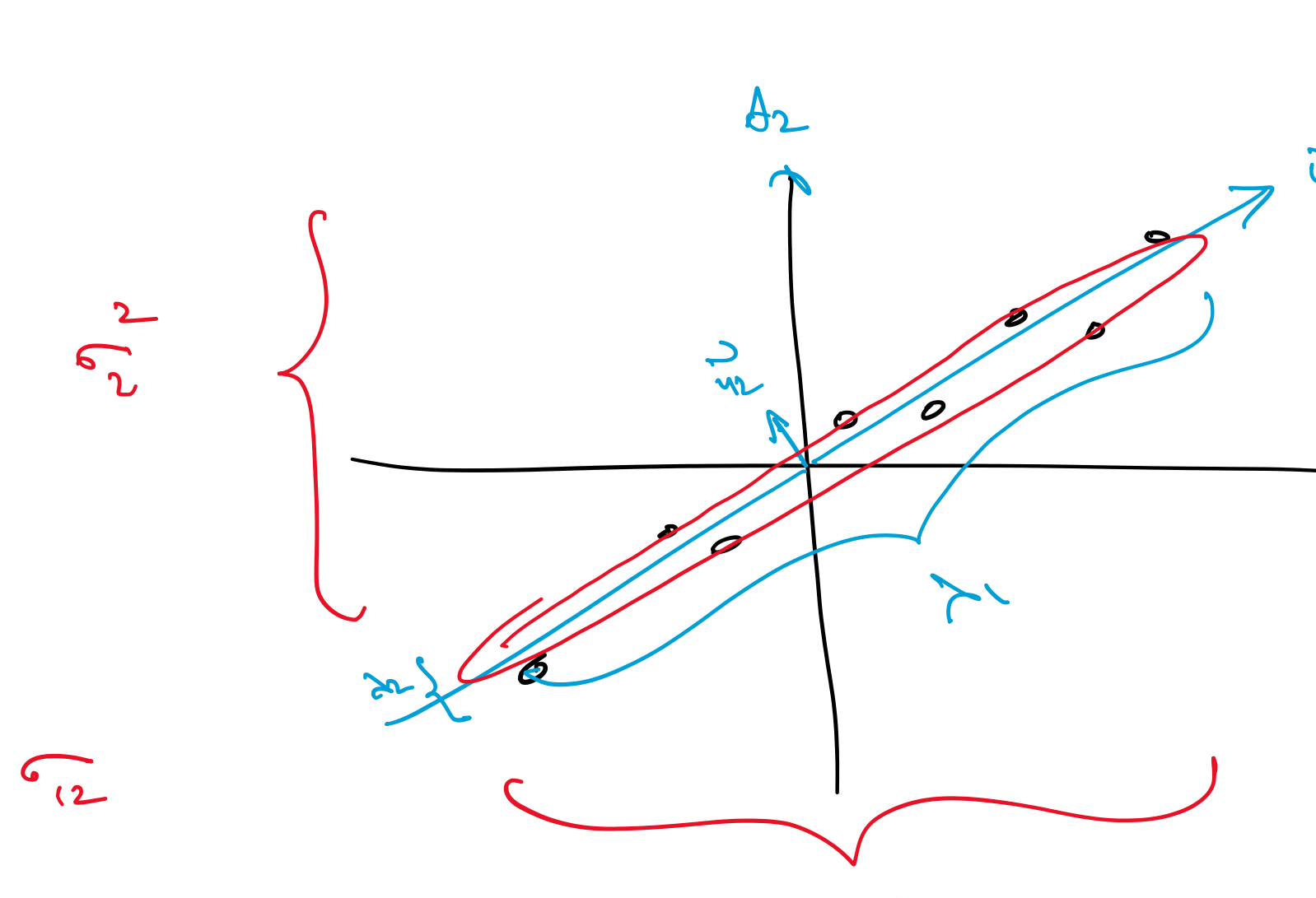
$\vec{x}_1 = (5, 3)^T$

$= 5\begin{pmatrix}1\\0\end{pmatrix} + 3\begin{pmatrix}0\\1\end{pmatrix}$

$O = \{\vec{e}_1, \vec{e}_2\}$ or $\{\vec{u}_1, \vec{u}_2\} = V$

any point in $\mathbb{R}^2$ is linear combination of O or U

eigenvector basis



$\boxed{\Sigma\vec{u} = \lambda\vec{u}}$ d different equations

$$\begin{cases} \Sigma\vec{u}_1 = \lambda_1\vec{u}_1 \\ \Sigma\vec{u}_2 = \lambda_2\vec{u}_2 \\ \vdots \\ \Sigma\vec{u}_d = \lambda_d\vec{u}_d \end{cases}$$

$\Downarrow$

matrix representation of these d equations

$$\underset{d \times d}{\Sigma}\ \underbrace{\overset{U}{\begin{bmatrix} | & | & & | \\ u_1 & u_2 & \ldots & u_d \\ | & | & & | \end{bmatrix}}}_{d \times d} = \underbrace{\begin{bmatrix} u_1 & u_2 & \ldots & u_d \end{bmatrix}}_{d \times d} \underbrace{\overset{\Lambda}{\begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_d \end{bmatrix}}}_{d \times d}$$

$$\Sigma U = U\Lambda$$

U = [all the eigenvectors as columns]

orthogonal matrix (all columns are unit vectors & are mutually orthogonal)

$\begin{cases} \|u_i\|^2 = 1 \\ u_i^T u_j = 0 \end{cases}$

$U^{-1} = U^T$

$\Sigma U = U\Lambda$

$U^T\Sigma U = U^T U\Lambda$

$\Rightarrow U^T\Sigma U = \Lambda$

Cov $\qquad$ new Cov matrix (diagonal)

$\Sigma \underbrace{U U^T}_{I} = U\Lambda U^T$

$$\boxed{\Sigma = U\Lambda U^T}$$

eigen decomposition

$$\boxed{\Sigma = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \ldots + \lambda_d u_d u_d^T}$$

$$\boxed{\Sigma} = \lambda_1 \boxed{u_1 u_1^T} + \lambda_2 \boxed{u_2 u_2^T} + \ldots + \lambda_d \boxed{u_d u_d^T}$$

outerproduct of $\vec{u}_1$

$\underbrace{(d \times 1)(1 \times d)}_{d \times d}$

k

smallest k

$$\text{approx error} = \left\|\Sigma - \underbrace{\sum_{i=1}^{k}\lambda_i u_i u_i^T}\right\|^2 \leq \theta$$

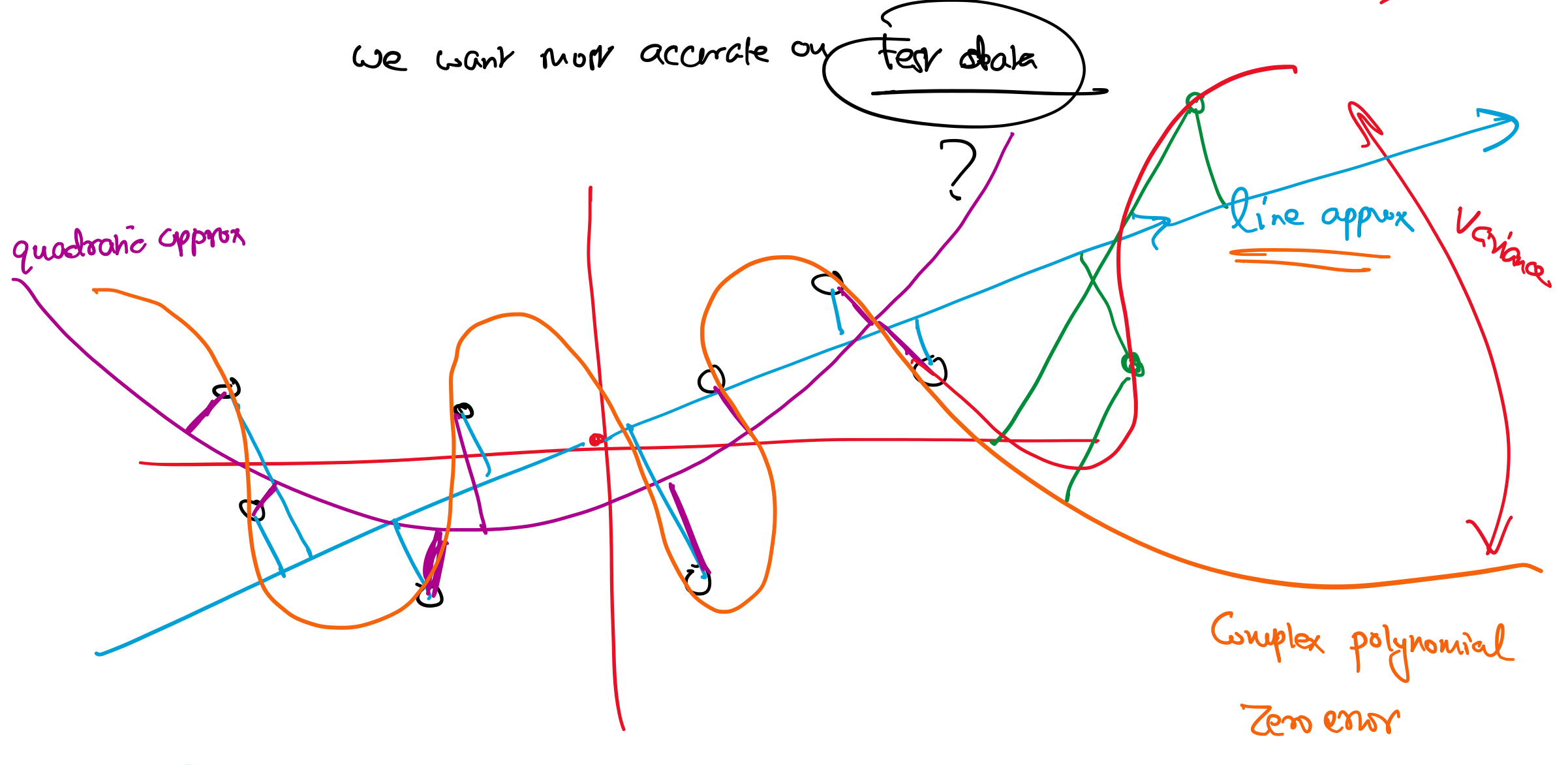
---

learning / mining:

$\boxed{\text{we don't want the most accurate model}}$ (on the training data)

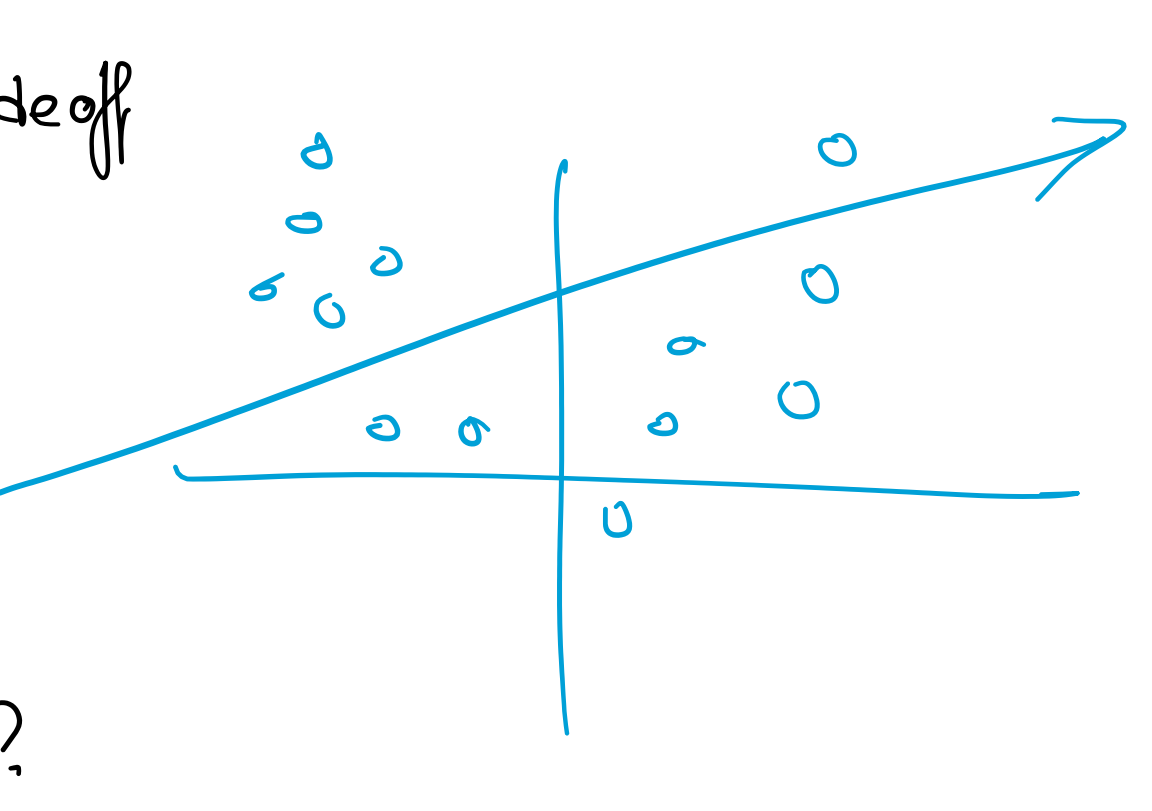
we want more accurate on test data

?



bias vs. variance tradeoff

over fitting



Complex models:

low bias
high variance + ?

Simple model

high bias
low variance + ?