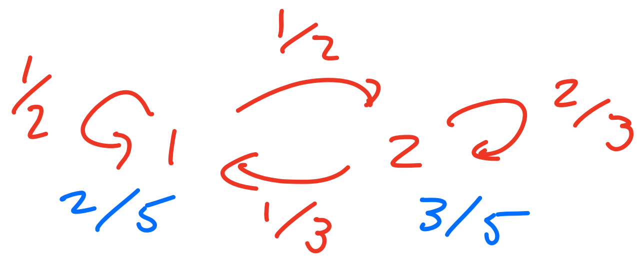
HW 6 due Tues Dec 13.
Choose 5 Problems.

Applications.

Last Time: Markov chains.

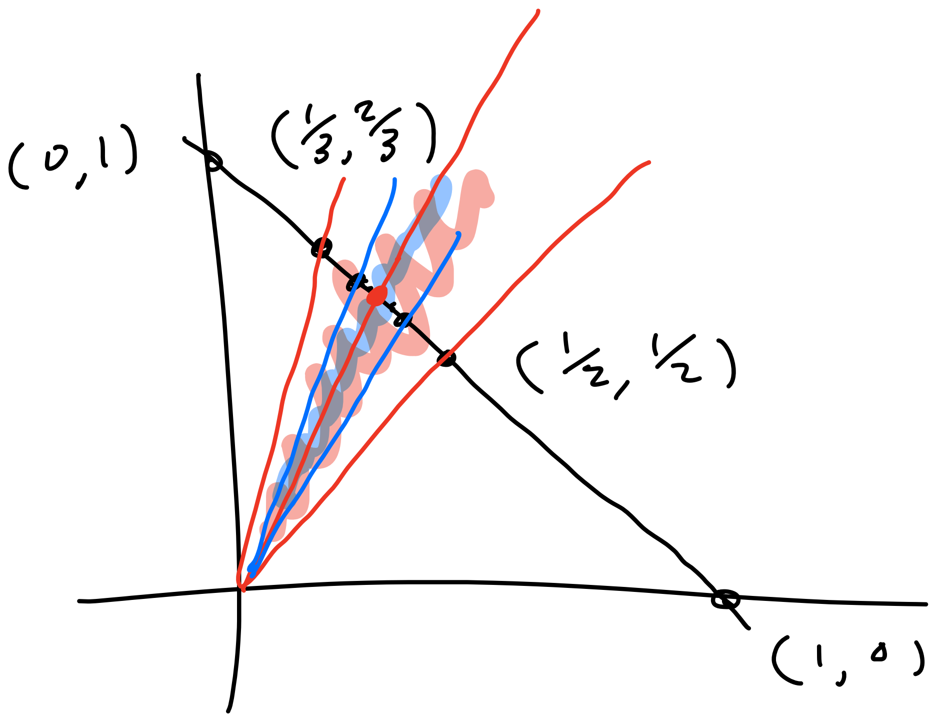
$$M = \begin{pmatrix} 1/2 & 1/3 \\ 1/2 & 2/3 \end{pmatrix}.$$



Saw $\exists$ unique equilibrium distribution:

$$M \begin{pmatrix} 2/5 \\ 3/5 \end{pmatrix} = \begin{pmatrix} 2/5 \\ 3/5 \end{pmatrix}.$$

Geometry: Look at how M^n acts on the standard basis of $\mathbb{R}^2$.



$$M = \begin{pmatrix} 1/2 & 1/3 \\ 1/2 & 2/3 \end{pmatrix}$$

$$M^2 = \begin{pmatrix} 5/12 & 7/18 \\ 7/12 & 11/18 \end{pmatrix}.$$

M acting on 1st quadrant is a "contraction mapping".

Contraction mappings in general converge to a unique limit.

Perron-Frobenius more general than Markov chains.

e.g. Google Page-Rank.

Tournament arrow "$A \to B$"
"A defeated B".

$$\begin{tikzcd}
A \arrow[r] \arrow[d] & B \arrow[d] \\
C \arrow[r] \arrow[ur] & D \arrow[ul]
\end{tikzcd}$$

Who is the winner? A ?

Rank the players.

Issue: D looks bad,
but D defeated A ...

Recursive: Assign random
initial scores $\vec{x}_0 = (1,1,1,1)$.

update the scores

your $(n+1)$st score $= \sum$ nth scores of people you defeated.

$$\vec{X}_{n+1} = \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{pmatrix} & 1 & 1 & \\ & & 1 & 1 \\ & & & 1 \\ 1 & & & \end{pmatrix} \end{array} X_n$$

$M.$

$$\vec{X}_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \vec{X}_1 = \begin{pmatrix} 2 \\ 2 \\ 1 \\ 1 \end{pmatrix}, \quad \vec{X}_2 = \begin{pmatrix} 3 \\ 2 \\ 1 \\ 2 \end{pmatrix}$$

$$\ldots, \quad \vec{X}_5 = \begin{pmatrix} 8 \\ 6 \\ 3 \\ 5 \end{pmatrix}$$

Rank:

$A > B > D > C.$

$$\ldots \quad \vec{X}_{100} = \begin{pmatrix} 1037 \\ 933 \\ 547 \\ 731 \end{pmatrix}$$

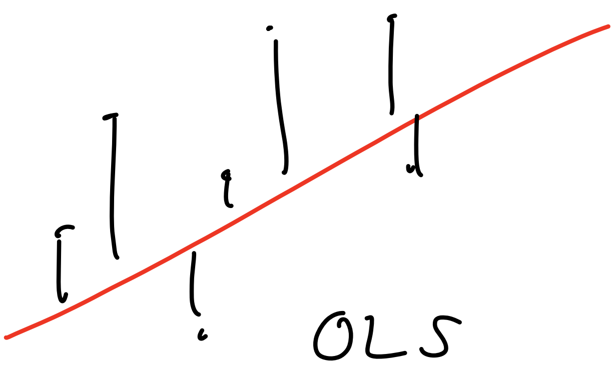
When can we stop?

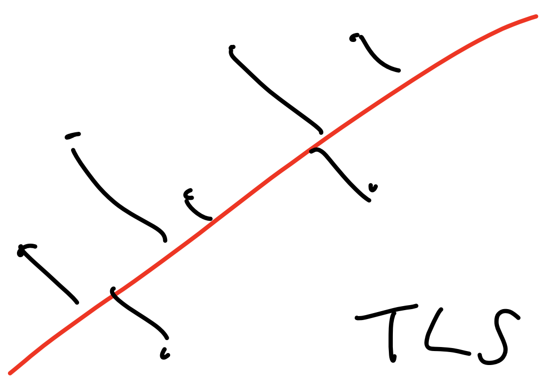
Perron-Frobenius: If you normalize at each step (make the scores sum to 1) then there is a unique equilibrium:

$$\vec{x}_n \longrightarrow \begin{pmatrix} 0.321\ldots \\ 0.283\ldots \\ 0.165\ldots \\ 0.230\ldots \end{pmatrix}$$

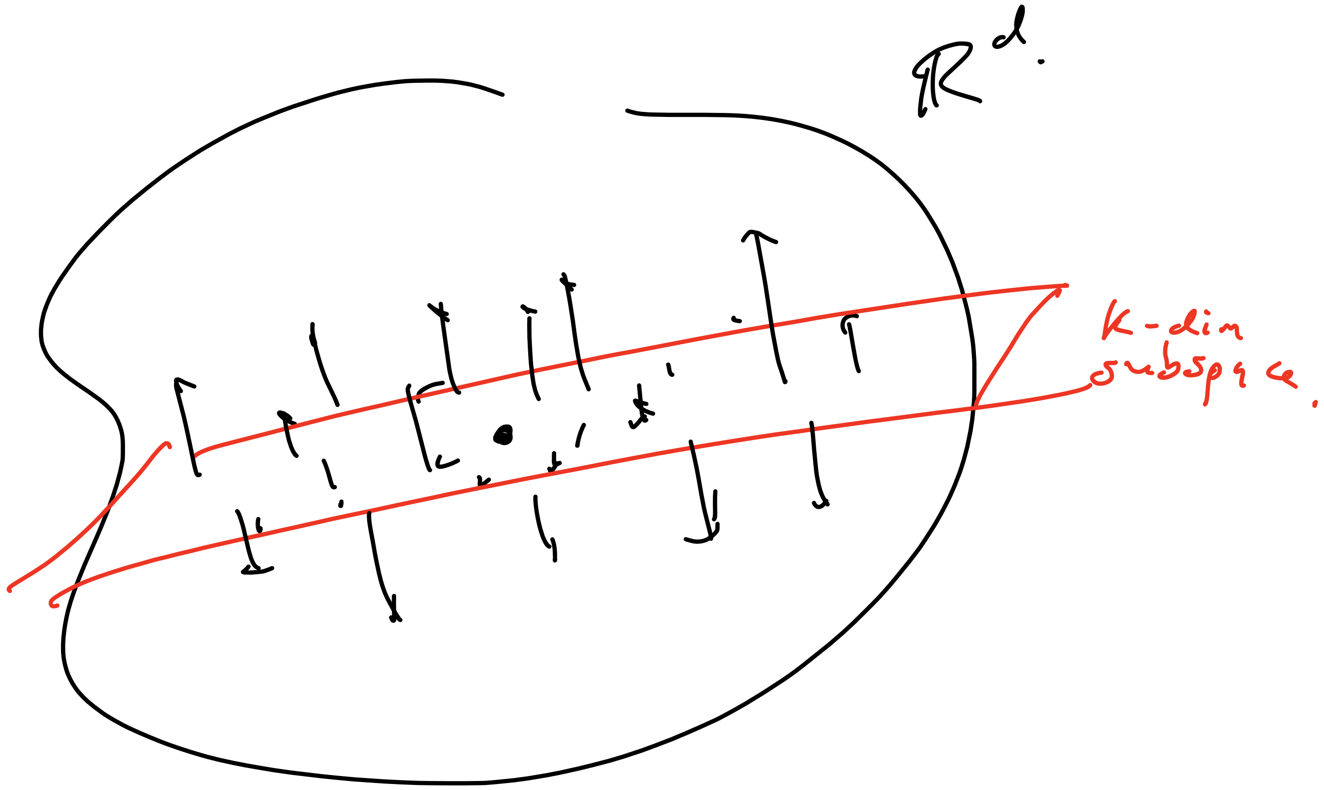
normalized.

Recall: Ordinary vs. Total Least Squares.





Problem: Given n data points in d-dim space $\mathbb{R}^d$.



Data: $\vec{x}_1, \dots, \vec{x}_n \in \mathbb{R}^d$.

Assume data is centered:

$$\vec{x}_1 + \vec{x}_2 + \dots + \vec{x}_n = \vec{0}.$$

Problem: Find the k-dim subspace of $\mathbb{R}^n$ that minimizes sum of squares of orthogonal distances to data points.

Form the data matrix:

$$X = \underbrace{\left(\vec{x}_1 \,\middle|\, \vec{x}_2 \,\middle|\, \cdots \,\middle|\, \vec{x}_n \right)}_{n} \Big\} d.$$

Assume $n >> d$,

so $\text{rank}(X) = d$.

Consider the invertible $d \times d$ matrix XX^T.

The eigenvalues are <u>real</u> & <u>positive</u>

symmetric pos. def

Typical to write e.values as

$$\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_d^2 > 0.$$

probably strict because data is noisy.

Jargon: $\sigma_1, \sigma_2, \dots, \sigma_d$ are the "singular values" of the data matrix X.

Since XX^T is symmetric, Spectral Theorem

$\Longrightarrow$ $\exists$ orthonormal basis

of eigenvectors:

$$\vec{u}_1, \vec{u}_2, \dots, \vec{u}_d \in \mathbb{R}^d$$

$$XX^T \vec{u}_k = \sigma_k^2 \vec{u}_k .$$

Theorem (Total Least Squares)

Best fit k-dim subspace ($k \le d$) is

$$\text{span}\{\vec{u}_1, \vec{u}_2, \dots, \vec{u}_k\}.$$

Special Case: Best fit <u>line</u>

For the data points is $t\vec{u}_1$ where $\vec{u}_1$ is any eigenvector for XX^T corresponding to the <u>largest</u> eigenvalue.

Very common method in data science called "principal component analysis".

Background Theory:

The SVD (Singular Value Decomposition).

Given $m \times n$ matrix A, rank r.

Recall

$$\text{rank}(A^TA) = \text{rank}(AA^T) = \text{rank}(A) = \text{rank}(A^T).$$

Eigenvalues of $A^T A$ are

$$\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2, \underbrace{0, 0, \ldots, 0}_{n-r \text{ zeros}}.$$

Spectral Theorem

$\Longrightarrow \exists$ o.n. basis of e.vectors

$$\vec{v}_1, \ldots, \vec{v}_r, \vec{v}_{r+1}, \ldots, \vec{v}_n \in \mathbb{R}^n$$

$$A^T A \vec{v}_k = \begin{cases} \sigma_k^2 \vec{v}_k & k \leq r, \\ 0 & \text{else}. \end{cases}$$

Claim that AA^T has eigenvalues

$$\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2, \underbrace{0, \ldots, 0}_{m-r \text{ zeros}}.$$

i.e. $\overset{n\times n}{A^T A}$ & $\overset{m \times m}{AA^T}$ have the same nonzero eigenvalues.

Furthermore, we can find an o.n. basis of e.vectors for AA^T.

$$\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m \in \mathbb{R}^m$$

such that

$$\boxed{\begin{aligned} A^TA\,\vec{v}_k &= \sigma_k^2\,\vec{v}_k \\ AA^T\,\vec{u}_k &= \sigma_k^2\,\vec{u}_k \\ A\,\vec{v}_k &= \sigma_k\,\vec{u}_k \end{aligned}}$$

SVD

for all $k \leq r$.

"simultaneously diagonalized A^TA & AA^T"

In matrix terms:

$$\underset{m\times n}{A} = \underset{m\times r}{\begin{pmatrix} \vec{u}_1 & | \cdots | & \vec{u}_r \end{pmatrix}} \underset{r\times r}{\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}} \underset{r\times n}{\begin{pmatrix} \vec{v}_1^{\,T} \\ \hline \vdots \\ \hline \vec{v}_r^{\,T} \end{pmatrix}}$$

We can also use the full bases for $\mathbb{R}^m$ & $\mathbb{R}^n$.

$$\underset{m\times n}{A} = \underset{m\times m}{U} \underset{m\times n}{\left(\begin{array}{c|c} \begin{matrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{matrix} & 0 \\ \hline 0 & 0 \end{array}\right)} \underset{n\times n}{V^T}$$

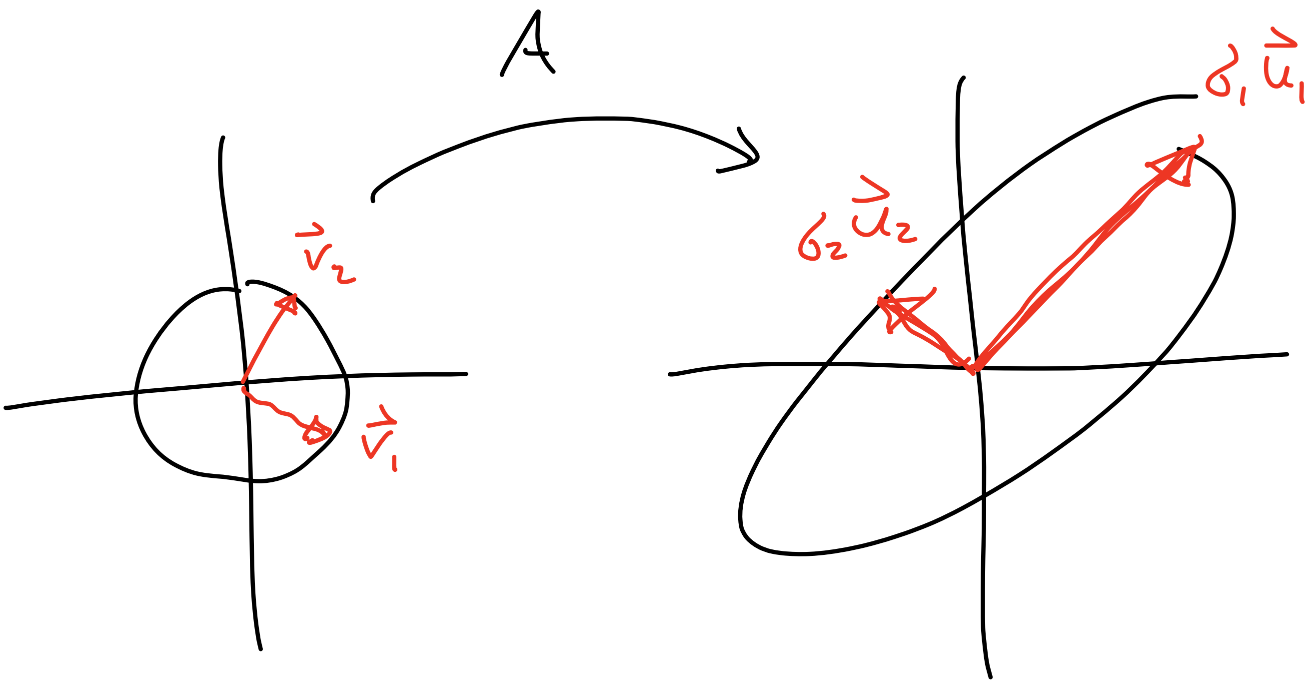
where U, V are square orthogonal.

Proof: For $k \leq r$ ($\sigma_k \neq 0$), define

$$\vec{u}_k = \frac{1}{\sigma_k} A \vec{v}_k$$

Check that it works. ✓

Picture: A is 2×2



Hard Theorem (Eckart-Young, 1936)

A $m \times n$, rank r.

$$\underset{m \times n}{A} = \underset{m \times r}{U} \ \underset{r \times r}{\Sigma} \ \underset{r \times n}{V^T}$$

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

SORT

$$= \left(\vec{u}_1 \middle| \cdots \middle| \vec{u}_r \right) \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} \begin{pmatrix} \vec{v}_1^{\,T} \\ \vdots \\ \vec{v}_r^{\,T} \end{pmatrix}.$$

$$= \underbrace{\sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^{\,T}}_{\text{sum of rank one matrices.}}$$

For $k \le r$ define

$$A_k = \sum_{i=1}^{k} \sigma_i \vec{u}_i \vec{v}_i^{\,T}$$

cut the smallest singular values.

Then $\operatorname{rank}(A_k) = k$.

For <u>any</u> X $m \times n$, rank k,

$$\| A - A_k \|_F \le \| A - X \|_F$$

Best rank k approximation to A.