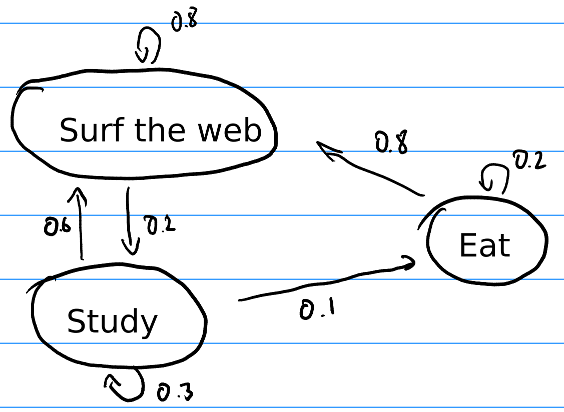
In what way can eigenvalue computation be applied to Markov chains?



Markov property:

The probability distribution of the next state only depends on the *current* state (and not anything longer ago).

Rewrite the state graph as a matrix:

$$A = \begin{array}{c} \text{current} \\ \begin{array}{ccc} \text{surf} & \text{study} & \text{eat} \end{array} \\ \begin{pmatrix} .8 & .6 & .8 \\ .2 & .3 & 0 \\ 0 & .1 & .2 \end{pmatrix} \end{array} \begin{array}{l} \\ \\ \text{surf} \\ \text{stu} \\ \text{eat} \end{array} \text{ next state}$$

The columns (= the probabilities in the current state) add up to 1.

$$A \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \leftarrow \text{surf} = \begin{pmatrix} .8 \\ .2 \\ 0 \end{pmatrix} \qquad A^2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \ldots$$

$$A \begin{pmatrix} 0.5 \\ 0.5 \\ 0 \end{pmatrix} = \begin{pmatrix} .4 + .3 \\ .1 + .15 \\ .05 \end{pmatrix} = \begin{pmatrix} .7 \\ .25 \\ .05 \end{pmatrix}$$

Really: Looking for the *equilibrium* distribution p.

$$Ap = p \qquad Ap = 1p$$

Hit this with power iteration.

8 **Singular Value Decomposition**

What is the Singular Value Decomposition?

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

orthgonal

diagonal, entries positive

How do I compute it? (for any matrix A)

symmetric positive definite (SPD)

(1) Compute eigenvalues of $A^T A$

$$A^T A v_1 = \lambda_1 v_1 \quad \cdots \quad A^T A v_n = \lambda_n v_n$$

(2) Make a matrix of the eigenvectors $V = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}$ ← orth.

eigenvalues are nonnegative, eigenvectors v_i orthogonal

(3) Make a diagonal matrix Σ of the square roots of the eigenvalues

$$\Sigma = \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix}$$

(4) $A = U \Sigma V^T \iff U\Sigma = AV$

$\iff U = A V \Sigma^{-1}$

columns: left singular vectors

singular values

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

columns of V: right singular vectors

What's another way of writing the SVD?

Starting from $A = U\Sigma V^T = \begin{array}{ccc} U & \Sigma & \\ \begin{pmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{pmatrix} & \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ & & \sigma_n \\ 0 & & \end{pmatrix} & \begin{pmatrix} — v_1 — \\ \vdots \\ — v_n — \end{pmatrix} \end{array}$

$$A = \begin{pmatrix} | & | & & | \\ \sigma_1 u_1 & \sigma_2 u_2 & \cdots & \sigma_n u_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} — v_1 — \\ \vdots \\ — v_n — \end{pmatrix} = \underbrace{\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T}_{\text{closest rank-2 matrix to } A} + \cdots + \sigma_n u_n v_n^T$$

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \sigma_n \geq 0$$

$$x y^T = \begin{pmatrix} x \end{pmatrix} \begin{pmatrix} — y — \end{pmatrix} = \begin{pmatrix} x_1 \cdot y_1 & & x_1 \cdot y_n \\ x_2 \cdot y_1 & & \vdots \\ x_n \cdot y_1 & & x_n \cdot y_n \end{pmatrix}$$

outer product

rank(outer product) = 1

Rank-k best-approximation $(k \leq n)$

$A_k = \sigma_1 u_1 v_1^T + \cdots + \sigma_k u_k v_k^T$ minimizes $\|A - B\|_2$ among all rank-2 matrices

What do the singular values mean? (in particular the first/largest one)

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

$$\|A\|_2 = \max_{\|x\|_2 = 1} \|Ax\|_2 = \max_{\|x\|_2=1} \|\overset{\text{orth}}{U}\Sigma V^T x\|_2 = \max_{\|x\|_2=1} \|\Sigma V^T x\|_2$$

$$\underset{\|V^T x\|_2 = \|x\|_2}{=} \max_{\|V^T x\|_2 = 1} \|\Sigma V^T x\|_2 = \max_{\|y\|_2 = 1} \|\Sigma y\|_2 = \max_i |\sigma_i|$$

$y = V^T x$

How expensive is it to compute the SVD?

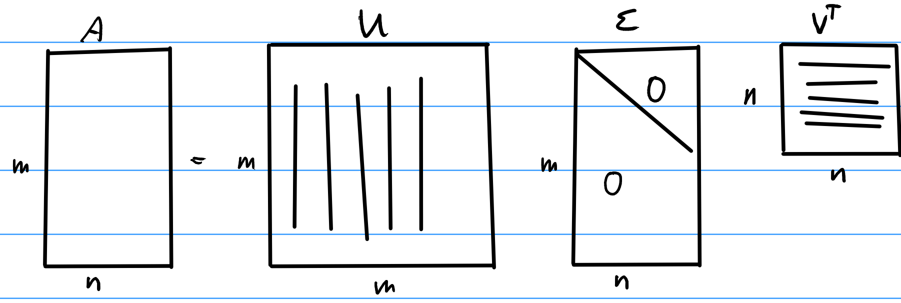
So why bother with the SVD if it is so expensive? I.e. what makes the SVD special?

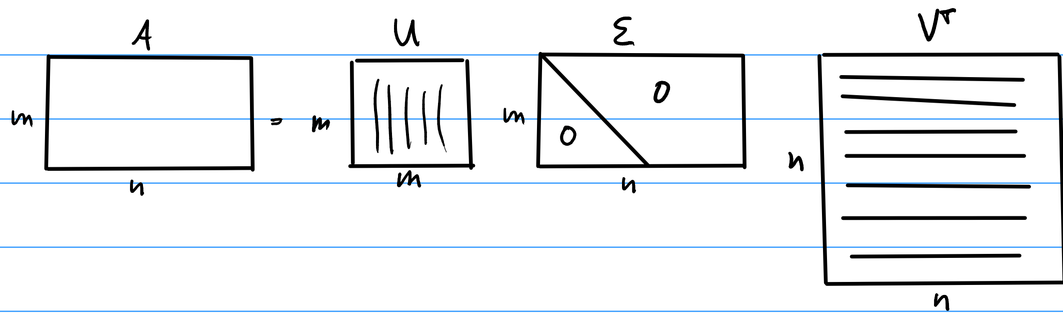
$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_n u_n v_n^T$$

What is the Frobenius norm of a matrix?

How about rank-k best-approximation in the Frobenius norm?

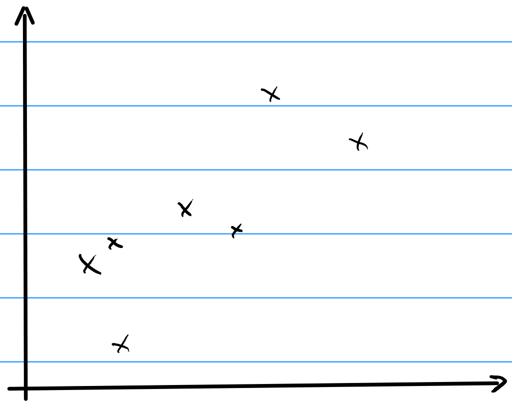
Is there a "reduced" transform for non-square matrices?





9 **Applications of the SVD**

(1) Rank-k approximation



So now how about rank-k approximation?

Give an example of where rank-k approximation does something useful.

(2) Computing the 2-norm

(3) Computing the 2-norm condition number