NORMALIZED CONDUCTANCE

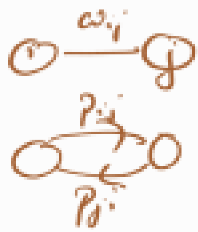
LECTURE 18
21/4/2021

Def. A Markov chain with TPM P and a stationary distribution π is called *time reversible* if

$$\forall i,j \quad \pi_i P_{ij} = \pi_j P_{ji}.$$

$\Rightarrow$ Underlying digraph is symmetric

$$(i) \to (j) \iff (i) \rightleftarrows (j)$$



$\Rightarrow$ G can be obtained from an undirected graph H by doubling each edge. (H may have self loops).

$\Rightarrow$ P can be encoded as weights on the edges of H.

$$w_{ij} = \pi_i P_{ij} \quad (= \pi_j P_{ji})$$

Claim: If I know all w_{ij} values then I can compute all P_{ij} values.

Qn. Can we find P from w_{ij}'s?

$$\pi_i = \sum_j w_{ij} \qquad (w_{ij} = \pi_i P_{ij} = \pi_j P_{ji})$$

$$\sum_j w_{ij} = \sum_j \pi_i P_{ij} = \pi_i \underbrace{\sum_j P_{ij}}_{1} = \pi_i$$

$$P_{ij} = w_{ij}/\pi_i$$

Some time w_{ij}'s may be scaled by an unknown constant ($w_{ij} = c\,\pi_i P_{ij}$).

In that case,

$$\sum_i \sum_j w_{ij} = \sum_i \sum_j c\,\pi_i P_{ij} = c \sum_i \pi_i \underbrace{\sum_j P_{ij}}_{1} = c \underbrace{\sum_i \pi_i}_{1} = c$$

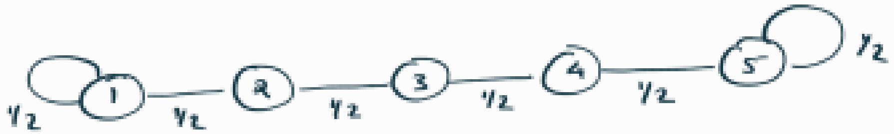
$$c = \sum_i \sum_j w_{ij}$$

$$w_i = \sum_j w_{ij} \qquad (\pi_i = w_i/c)$$

$$P_{ij} = w_{ij}/w_i$$

Hence the weighted undirected graph (with possibly self loops) defines the random walk.

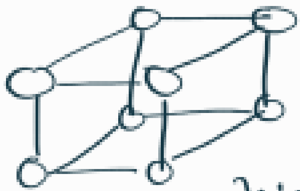
Eg. 1.



2.

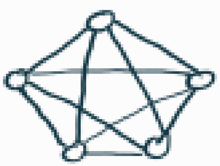
(No $w_{ij} \Rightarrow w_{ij} = 1$).

3.



"RANDOM WALKS ON WEIGHTED UNDIRECTED GRAPHS"

4.



(P need not be symmetric).

Defn. The normalised conductance $\phi(G)$ of a weighted undirected graph G is defined as

$$\phi(G) = \min_{\substack{S \subseteq V(G) \\ 0 < w(S) \le w(\bar{S})}} \left\{ \frac{\sum_{x \in S, y \in \bar{S}} w_{xy}}{w(S)} \right\}$$

where $w(S) = \sum_{x \in S} w_x$

$$= \min_{S \subseteq V(G)} \left\{ \frac{\sum_{x \in S, y \in \bar{S}} w_{xy}}{\min\{w(S), w(\bar{S})\}} \right\} \qquad w_x = \sum_y w_{xy}$$

Intution.

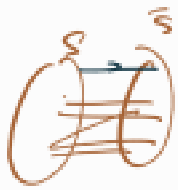
(Assume $\sum_{x,y} w_{xy} = 1$, so that $(c = 1)$

$\sum_y w_{xy} = \pi_x$ and $w_{xy} = \pi_x P_{xy}$)

$$P_{xy} = \Pr\left[\text{next state} = y \,/\, \text{Curr. st} = x\right]$$

$$= \Pr\left[\begin{matrix}\text{next move is} \\ \text{along the arc } xy\end{matrix} \,/\, \text{Curr. st} = x\right]$$

$$w_{xy} = \pi_x P_{xy} = \Pr\left[\begin{matrix}\text{next move is along} \\ \text{the arc } xy\end{matrix}\right] \text{ in steady state}$$



$$\sum_{x \in S, y \in \bar{S}} w_{xy} = \Pr\left[\begin{matrix}\text{next move is an} \\ \text{escape from } S\end{matrix}\right] \text{ in SS}$$

$$\frac{1}{w(S)} \sum_{x \in S, y \in \bar{S}} w_{xy} = \Pr\left[\begin{matrix}\text{Next state} \\ \text{is outside } S\end{matrix} \,/\, \begin{matrix}\text{Cur. St is} \\ \text{in } S\end{matrix}\right] \text{ in SS.}$$

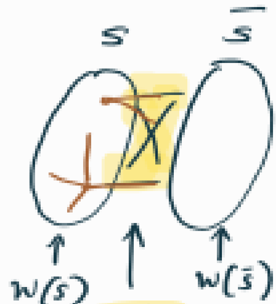
= "Escape probability from S."

ϕ = minimum escape prob among all sets $S \subseteq V(G)$ with $0 < w(S) \leq 1/2$

Defn. (Norm. Cond of a cut).

for $S \subseteq V(G)$ with $w(S) > 0$,

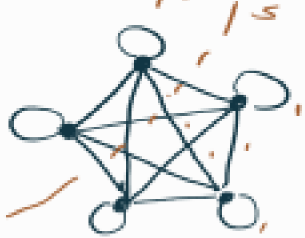
$$\underset{\phi(S)}{\phi(S, \bar{S})} := \frac{\sum_{x \in S, y \in \bar{S}} w_{xy}}{\min\{w(S), w(\bar{S})\}} \leq 1$$



$\phi(G) = \min \phi(S, \bar{S})$, where the minimum is over all non-trivial cuts of G.

EXAMPLES

1. K_n with self loops
   $w_{ij} = 1 \;\forall$ edges
   $w_i = n \;\forall$ vertices
   $w(S) = n|S|$



$| \; P_{ij} = \frac{w_{ij}}{w_i} = \frac{1}{n}$

Let $(S, \bar{S})$ be a cut with $k = |S| \leq |\bar{S}|$

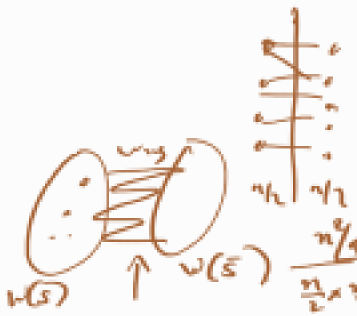
$$\phi(S, \bar{S}) = \frac{k \times (n-k)}{\min\{nk, n(n-k)\}}$$

k vertices

$$= \frac{k \times (n-k)}{nk} = \frac{n-k}{n}$$

$$\phi(G) = \min_{k \leq n/2} \left(\frac{n-k}{n}\right) = \underline{\underline{1/2}}$$

$$\sum_{x \in S, y \in \bar{S}} w_{xy} = |E(S, \bar{S})|$$

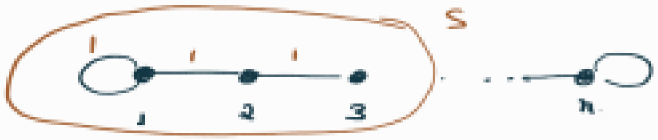


$\phi \leq 1$

$\frac{n^2/4}{\frac{n}{2} \times \frac{n}{2}}$

$\frac{n^2/4}{n^2/4} = 1$

2. P_n with end loops



$w_{ij} = 1 \quad \forall$ edges

Let $S \subseteq [n]$

$w(S) = 2|S|$

Hence $|S| < |\bar{S}| \Rightarrow w(S) \le w(\bar{S})$
$(|S| \le n/2)$

$$\phi(S,\bar{S}) = \sum_{x \in S, y \in \bar{S}} w_{xy} \Big/ w(S)$$

$$= \frac{|E(S,\bar{S})|}{2|S|}$$

Minimise

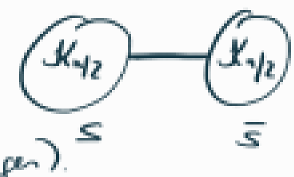
$|S| = k \le n/2$

$$w(G) = \min_{|S| \le n/2} \phi(S,\bar{S}) \approx \frac{1}{2 \cdot n/2} = \frac{1}{n}.$$

3. G disconnected $\Rightarrow \phi(G) = 0$.

Pick S to be $V(C_i)$ with $w(V(C_i))$ smallest.

4. Dumbbell



$(w_{ij} = 1 \quad \forall$ edges$)$

$w(S) = \frac{n}{2} \cdot \frac{n}{2} + 1 \approx n^2/4$

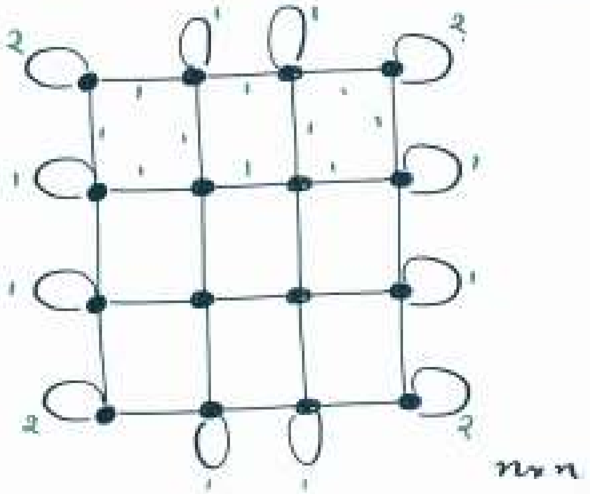
$E(S,\bar{S}) = 1.$

$\phi(S,\bar{S}) = 4/n^2$

$\phi(G) \le 4/n^2$ "Worse than path"

No monotonicity

5. 2D-lattice with end loops



$n \times n$

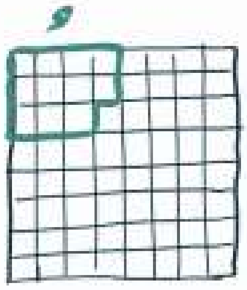
$|S| = k = m^2$

$w(S) = 4|S|$ hence

$|S| < |\bar{S}| \Rightarrow w(S) \le w(\bar{S})$.

Obs 1.

If $|S| \le \frac{1}{4}n^2$ then a corner square has least no. of escape edges relative to $|S|$



$$E(S,\bar{S}) \approx 2\sqrt{|S|}$$

Hence
$$\phi(S,\bar{S}) = E(S,\bar{S})/w(S) \approx \frac{2\sqrt{|S|}}{4|S|} = \frac{1}{2\sqrt{|S|}} \le \frac{1}{2\sqrt{n^2/4}} = \frac{1}{n}$$

Obs 2.

If $\frac{n^2}{4} < |S| \le \frac{n^2}{2}$, then a margin strip has least no. of escape edges.

$$E(S,\bar{S}) \approx n$$

Hence
$$\phi(S,\bar{S}) \approx n/4|S| \le \frac{n}{4 \cdot n^2/2} = \frac{1}{2n}$$

$\phi(G) = 1/2n$ $\quad P_n - 1/n$

$= 1/2\sqrt{|V(G)|}$

6. d-dimensional lattice with end loops

$|V(G)| = n^d$

$\phi(G) = 1/dn$ (Exercise)

$= 1/d \, |V(G)|^{1/d}$

Why is $\phi(G)$ important?

$$\Omega\left(\frac{1}{\phi(G)}\right) \leq \text{Mixing time} \leq O\left(\frac{1}{\phi^2(G)}\right)$$

MIXING TIME

"No. of steps by which any starting distribution gets trapped in an ε-neighbourhood of the stationary distribution".

Formal defn depends on

1. Distance measure used
2. We use $p(t)$ or $a(t)$.

Defn (BHK Defn 4.1)

Fix any $\varepsilon > 0$. The ε-mixing time of a random walk is the smallest integer t such that for any starting distribution p, $\|a(t) - \pi\|_1 < \varepsilon$, where π is the stationary distribution of the random walk.

LECTURE 19.
25/Apr/2021

MIXING TIME & NORMALIZED CONDUCTANCE

- $\phi(S) = \phi(S, \bar{S}) = \dfrac{\sum_{x \in S, y \in \bar{S}} w_{xy}}{\min\{w(S), w(\bar{S})\}}$

- $\phi = \min_{\emptyset \neq S \subseteq V(G)} \phi(S)$.

- $\tau_\varepsilon = \max_{\sigma} \min_{t} \|a_\sigma(t) - \pi\|_1 < \varepsilon$

$\pi_{min} = \min_i \pi_i$

Goal: TST

$$\Omega\left(\frac{1}{\phi}\right) \le \tau_\varepsilon \le O\left(\frac{\log(1/\pi_{min})}{\varepsilon^3 \phi^2}\right)$$

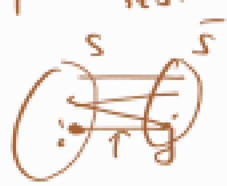
Tool: Probability flow across a cut $(S, \bar{S})$.

Let $q = pP$ (current distrbn p $\downarrow$ Next distrbn q).

for a set $S \subseteq V(G)$,

net prob. out-flow $= \Delta(S) = p(S) - q(S)$.

$p(S) = \sum_{i \in S} p_i$



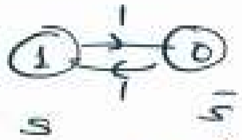
(Verify)
$$= \sum_{i \in S, j \in \bar{S}} (p_i P_{ij} - p_j P_{ji})$$

$$= \sum_{i \in S, j \in \bar{S}} \left(\frac{p_i}{\pi_i} w_{ij} - \frac{p_j}{\pi_j} w_{ij}\right)$$

$$= \sum_{i \in S, j \in \bar{S}} \left(\frac{p_i}{\pi_i} - \frac{p_j}{\pi_j}\right) w_{ij}$$

$$= \sum_{i \in S, j \in \bar{S}} (v_i - v_j) w_{ij} \qquad \left(v_i = p_i/\pi_i\right)$$

Obs: 1. $\Delta_p(S)$ can be as large as 1.
and as small as -1.



$p = (1,0)$
$q = (0,1)$

2. $\Delta_{\pi}(S) = 0 \quad \forall\, S \subseteq V(G)$

3. $\Delta_{a(t)}(S) \le \dfrac{2}{t} \quad \forall\, S \subseteq V(G)$
(∀ starting distribution σ)

Recall $\quad a(t) = \dfrac{1}{t}\left(p(0) + \dots + p(t-1)\right)$

$$\left(p(0) = \sigma,\quad p(i+1) = p(i)P\right)$$

$$a(t)P - a(t) = \frac{1}{t}\left(p(1) + \dots + p(t+1)\right) - \frac{1}{t}\left(p(0) + \dots\ p(t)\right)$$

$$= \frac{1}{t}\left(p(t+1) - p(0)\right)$$

$$\|a(t)P - a(t)\|_1 = \frac{1}{t}\,\|p(t+1) - p(0)\|_1 \le 2/t.$$

$\|p-q\|_1$
$p = (0,1,0)$
$q = (1,0,0)$

$$\boxed{\Delta_{a(t)}(S) \le 2/t}$$

$\underbrace{p_1\ p_2\ \dots}_{S}\ \ p_n$
$\underbrace{q_1\ q_2\ \dots}_{S}\ \ q_n$

$$\Delta(S) = \sum_{i\in S} p_i - \sum_{i \in S} q_i = \sum_{i\in S}(p_i - q_i) \le \sum_{i\in S}|p_i - q_i| \le \sum_i |p_i - q_i| = \|p-q\|_1.$$

Useful observation

$$\|p-q\|_1 = \sum_i |p_i - q_i| = \sum_{p_i > q_i}(p_i - q_i) + \sum_{p_i < q_i}(q_i - p_i)$$

Exercise

$$= 2\sum_{p_i > q_i}(p_i - q_i)$$

Direction 1. ($1/\phi \leq \tau_{1/2}$)

Let S: $\pi(S) \leq \pi(\bar{S})$

Consider the starting distribution to be $\sigma = \pi|_S$ "π restricted to S"

$$\sigma_i = \begin{cases} \pi_i/\pi(S) & , i \in S \\ 0 & , i \in \bar{S} \end{cases}$$

$(v_i = 1/\pi(S) = \sigma_i/\pi_i)$

$(v_i = 0)$

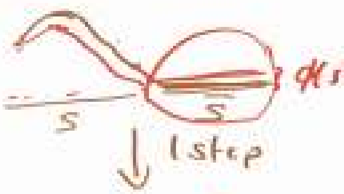
$$\begin{aligned}
\Delta_\sigma(S) &= \sum_{\substack{i \in S \\ j \in \bar{S}}} (v_i - v_j)\, w_{ij} \\
&= \sum_{i \in S, j \in \bar{S}} \left(1/\pi(S) - 0\right) w_{ij} \\
&= \frac{1}{\pi(S)} \sum_{i \in S, j \in \bar{S}} w_{ij} \\
&= \phi(S, \bar{S}) \qquad (\phi(S))
\end{aligned}$$

For the distribution $\sigma = \pi|_S$,

$\Delta_\sigma(S) = \phi(S, \bar{S})$

$\Delta_{\pi|_S}(S) = \phi(S, \bar{S})$

$\sigma(\bar{S}) = 0$



ϵ-close to π

$\|\sigma_t - \pi\|_1 < \epsilon$

$\Rightarrow \|\sigma_t(\bar{S}) - \pi(\bar{S})\| < \epsilon$

In the subsequent moves the flow is only smaller (why?)

$$\begin{aligned}
\therefore\ \tau_\epsilon &\geq \frac{\pi(\bar{S}) - \epsilon/2}{\phi(S)} \\
&\geq \frac{1/2 - \epsilon/2}{\phi(S)} \\
&\approx \frac{1}{2\phi(S)}
\end{aligned}$$

$$\tau_{1/2} \geq \frac{1}{2\phi(S)} \geq \frac{1}{2\phi}$$

Other Direction $\left(T_c \le O\left(\frac{\log(1/\sqrt{\pi_0})}{\varepsilon^2 \phi^2} \right) \right)$

Let $a = a(t)$.

Vertices in S are lucky.
" $\bar{S}$ " unlucky

$$S = \{ i : a_i \ge \sqrt{\pi}_i \}$$
$$= \{ i : v_i \ge 1 \}$$

$(v_i = a_i / \sqrt{\pi_i})$

$$\Delta_a(S) = \sum_{i \in S, j \in \bar{S}} (v_i - v_j)\, w_{ij}$$

$$> \sum_{i \in S, j \in \bar{S}} (v_i - 1)\, w_{ij} \qquad (v_j < 1 \ \forall j \in \bar{S})$$

$$= \sum_{i \in S} (v_i - 1) \sum_{j \in \bar{S}} w_{ij} \qquad -(1)$$

$$= \sum_{i \in S} (v_i - 1)\, w_i|_{\bar{S}} \qquad \left(w_i|_{\bar{S}} = \sum_{j \in \bar{S}} w_{ij} \right)$$

Over simplified idea

if $\forall i \in S \quad (v_i - 1) \approx c > 0$

Then $(1) \Rightarrow \Delta_a(S) \ge c \sum_{i \in S, j \in \bar{S}} w_{ij}$

$$\ge c\, \phi(S) \min\{\sqrt{\pi}(S), \sqrt{\pi}(\bar{S})\}$$

$$\ge c\, \phi(S) \min\{\sqrt{\pi}(S), \varepsilon/2\}$$

$\left(\sqrt{\pi}(\bar{S}) < \varepsilon/2 \Rightarrow \sum_{\sqrt{\pi}_i > a_i} (\sqrt{\pi}_i - a_i) < \varepsilon/2 \right.$
$\left. \Rightarrow \|\sqrt{\pi} - a\|_1 < \varepsilon \right)$

$$\ge \underline{c\, \phi(S)\, \sqrt{\pi}(S)\, \varepsilon/2}$$

But $\Delta_a(S) \le 2/t$

Hence $c \le \dfrac{4}{t\, \phi(S) \sqrt{\pi}(S)\, \varepsilon}$

$c\,\phi(S)\,\pi(S)\,\frac{\varepsilon}{2} \le \frac{2}{t}$

$c \le \dfrac{4}{t\, \phi(S) \sqrt{\pi}(S)\, \varepsilon}$

Now,

$$\|a-\pi\|_1 = \sum_i |a_i - \pi_i|$$

$$= 2\sum_{i\in S}(a_i-\pi_i) \qquad \left(\begin{matrix} a_i \geq \pi_i \\ \forall i \in S\end{matrix}\right)$$

$$= 2\sum_{i\in S}(v_i-1)\,\pi_i$$

(Over simplified)

$$\|a-\pi\|_1 \approx 2c\,\pi(S)$$

$$\leq \frac{8}{t\,\phi(S)\,\varepsilon}$$

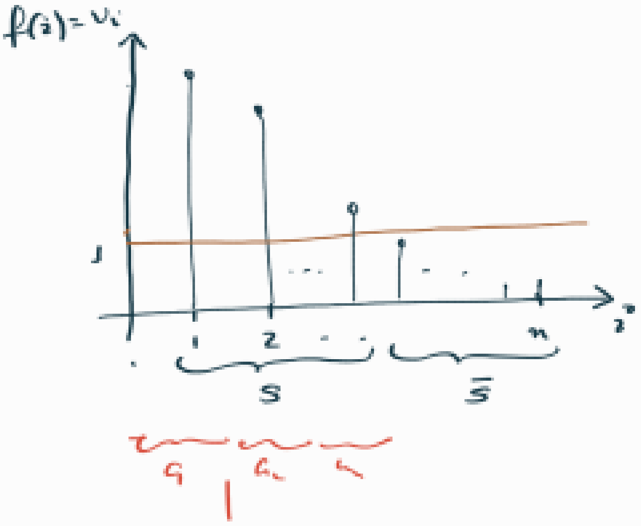
$$\leq \frac{8}{t\,\phi\,\varepsilon} \qquad (\phi = \phi(S))$$

If $t > \frac{8}{\phi\varepsilon^2}$, then $\frac{8}{t\phi\varepsilon} < \varepsilon$. $\quad \Big| \quad \frac{8}{t\phi\varepsilon} < \varepsilon$

The above $\overline{\text{over simplification}}$ is wrong. v_i's is never like a single step.

Relabel the vertices so that

$$v_1 \geq v_2 \geq \dots \geq v_n.$$



"A more careful analysis is required. And you end up with

$$t_\varepsilon \leq O\left(\frac{\log(1/\pi_{min})}{\varepsilon^3\phi^2}\right)$$

$$= O\left(\frac{\log n}{\varepsilon^3\phi^3}\right) \text{ for uniform } \pi$$

HITTING & COVERING TIMES

LECTURE 20
25/Apr/2021

Defn. For a random walk G on $V(G) = [n]$,

1. Hitting time from i to j. $h(i,j)$
   = expected time for the random walk to hit j for the first time if we start at i

2. Hitting time of j

$$h(j) = \max_{i \in [n]} h(i,j).$$

(worst case)

3. Hitting time of G

$$h(G) = \max_{i,j \in [n]} h(i,j).$$

4. Covering time of i $c(i)$
   expected time for the random walk starting at i to reach every node at least once.

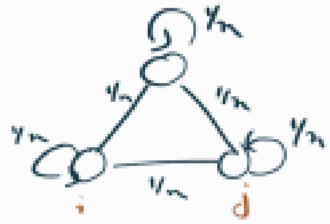
5. Covering time of G.

$$c(G) = \max_i c(i).$$

Mixing Time vs. Hitting Time vs. Covering Time.

Examples

1. K_n with self loops



(a) Mixing time

$$P = \begin{bmatrix} 1/n & 1/n & \cdots & 1/n \\ \vdots & & & \\ 1/n & & \cdots & 1/n \end{bmatrix}$$

For any starting distribution $p(0)$

$p(1) = (1/n, \cdots, 1/n) = \pi$ ← Verify.

$a(t) = \frac{1}{t}\left(p(0) + (t-1)\pi\right)$

$= \frac{1}{t}p(0) + \left(1 - \frac{1}{t}\right)\pi$

$\|a(t) - \pi\|_1 = \left\|\frac{1}{t}(p(0) - \pi)\right\|_1$

$< 2/t$

$\Rightarrow \boxed{\tau_\varepsilon < 2/\varepsilon}$ "constant"

$p = 1/n$ – 1 step
$(1-p)p$ – 2 step
P(you will hit j in k steps)
$(1-p)^{k-1}p$
↓
Geometric RV
Expectation $= 1/p$

$\frac{2}{t} \leq \varepsilon$
$\Uparrow$
$t \geq 2/\varepsilon$

(If we define mixing in terms of $p(t)$ rather than $a(t)$ then $\tau_\varepsilon = 1$)

(b) Hitting time.

Let $i \neq j$

$h(i,j) = E[\underbrace{\#\text{ steps to reach from } i \text{ to } j}_{X}]$

But X is geometric with $p = 1/n$.

$\Pr[X = k] = (1-p)^{k-1}p$

Hence $h(i,j) = \mathbb{E}[X] = 1/p$ (Refer & Verify)

$= n.$ $(\gg \tau_2)$

K_n $\boxed{h(K_n) = n}$

(c) Covering time:

Equivalent to coupon collector problem

Hence $C(K_n) \approx n \ln n \gg$ hitting time $\gg$ mixing time.

2. Path with end loops

(a) Hitting time

$$h(1,2) = \mathbb{E}\{\text{Geom. r.v with mean } 1/2\}$$

$$= 2.$$

$$h(2,3) = \frac{1}{2} + \frac{1}{2}(1 + h(1,3))$$

$h(2,3) = \frac{1}{2} + \frac{1}{2}(1 + h(1,2) + h(2,3))$

$\frac{1}{2}h(2,3) = \frac{1}{2} + \frac{1}{2}(1 + h(1,2))$

$h(2,3) = 2 + h(1,2)$

$= 4.$

$$h(i,i+1) = \frac{1}{2} + \frac{1}{2}[h(i-1,i+1) + 1]$$

$$= \frac{1}{2} + \frac{1}{2}[h(i-1,i) + h(i,i+1) + 1]$$

$$h(i,i+1) = 2 + h(i-1,i) \qquad (h(2,3) = 2+2 = 4)$$

$$= 2i$$

$$h(i,j) = h(i,i+1) + h(i+1,i+2) + \dots + h(j-1,j)$$

$(i<j)$

$$= 2i + 2(i+1) + \dots + 2(j-1)$$

$$= (i+j-1)(j-i)$$

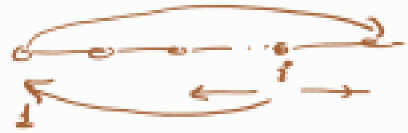
$$\max_{i,j} h(i,j) = h(1,n)$$
$$= 2(1+2+\cdots\ n-1)$$
$$= (n-1)n$$
$$\leq n^2$$

P_n: $\boxed{h(G) = n^2}$

(b) Covering time:

$$c(1) \leq h(1,n) \leq n^2$$

What about $c(i)$ vs $h(i,1)$ & $h(i,n)$?



$$c(i) \leq h(i,1) + h(1,n)$$
$$\leq 2h(1,n)$$
$$\leq 2n^2$$

Hence $c(G) \leq 2n^2$

(c) Mixing time:

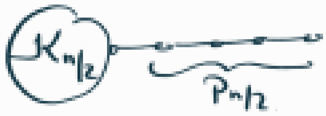
π = uniform (why?)

$= (1/n, \ldots, 1/n)$

$$\pi_{min} = 1/n$$
$$\phi = 1/n$$
$$\tau_\varepsilon \leq O\left(\frac{\log n}{\varepsilon^{3}(1/n)^2}\right)$$
$$\tau_\varepsilon = O\left(\frac{n^2 \log n}{\varepsilon^{3}}\right)$$

(Can be improved)

3. Lollipop graph



Hitting time $h(y,x) \approx n^2/4 = O(n^2)$.

But $h(x,y) = \Theta(n^3)$.

$$h(x,x+1) = \frac{1}{n/2}\,1 + \left(1-\frac{1}{n/2}\right)\times\left(\frac{n}{2} + h(x,x+1)\right)$$

$$= \frac{2}{n} + \frac{n}{2} - 1 + \left(1-\frac{2}{n}\right) h(x,x+1)$$

$$h(x,x+1)\cdot\frac{2}{n} = \frac{2}{n} + \frac{n}{2} - 1$$

$$h(x,x+1) = 1 + \frac{n^2}{4} - \frac{n}{2}$$

$$\approx n^2/4$$

$$h(x+i,x+i+1) = \frac{1}{2}\cdot 1 + \frac{1}{2}\left(h(x+i-1,x+i+1)+1\right)$$

$$= 1 + \frac{1}{2}\left(h(x+i-1,x+i) + h(x+i,x+i+1)\right)$$

$$= 2 + h(x+i-1,x+i)$$

$$\approx 2i + n^2/4$$

$$h(x,y) = \sum_{i=0}^{n/2-1}\left(2i + n^2/4\right)$$

$$= \frac{n^2}{4}\cdot\frac{n}{2} + O(n^2)$$

$$= \Theta(n^3)$$

$h(G) = \Theta(n^3)$.

Beyond Examples.

Thm 1. (Mean first reccurence theorem)

(weighted undirected graphs).

$$h(i,i) = 1/\pi_i \quad \forall i \in [n].$$

Proof: π_i = steady state prob for a random walk to be in node i

= proportion of time the walk is in node i

"ergodic property"

$N\pi_i$

$\Rightarrow$ Expected time b/n visit to i = $1/\pi_i$

Thm 2. If

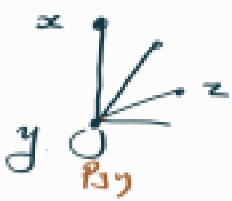
?? 1. All edge weights of G are 1, and

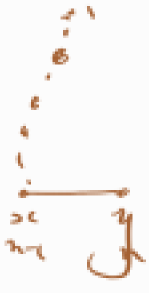
2. $x \sim y$ in G. (x in adj to y in G)

Then

$h(x,y) \le 2m$,

where $m = |E(G)|$





$2(n-1)$
$=2m$

Proof: $h(y,y) = P_{yy}\, 1 + \sum_{z \in N(y) \setminus \{y\}} P_{yz}(1 + h(z,y))$

Hence $P_{yx}\, h(x,y) \le h(y,y) = 1/\pi_y$

$$h(x,y) \le \frac{1}{\pi_y P_{yx}} = \frac{1}{\frac{d_y}{2m} \cdot \frac{1}{d_y}} = 2m$$

($\pi_y = d_y/2m$ Quiz 3)

Corollary: $h(G) \leq n^3$ for every graph on n vertices.

$$h(x,y) \leq 2m \cdot dist_G(x,y) \quad \forall x,y \in V(G)$$

$$h(G) \leq 2m \times diam(G)$$
$$\leq 2mn$$
$$\leq n^3.$$ //

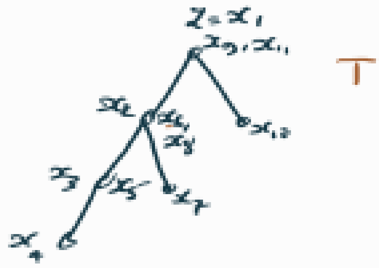
(Remember Lollipop).

Thm 3

$$c(G) \leq 4mn.$$

Proof. Let $z \in V(G)$ be arb.

Let T be a spanning tree of G rooted at z and let



$z = x_1, x_2, \ldots, x_{2n-1}$ be a DFS traversal of T.

(Each edge of T travelled twice).

$c(z) \leq$ Exp time to traverse T as above

$$\leq \sum_{i=1}^{2n-2} H(x_i, x_{i+1}) \quad \text{(Linearity of Expectation)}$$

$$\leq 2n \cdot 2m$$

$$= 4mn \ //$$

Since z was arbitrary,

$$c(G) \leq 4mn.$$

This does not mean $c(G) \leq 2h(G)$.

(Eg: K_n).

Thm 4.

$$c(G) \le h(G) \times 2\log_2 n.$$

Lemma. Let S be any subset of $V(G)$.

$x \in V(G)$,

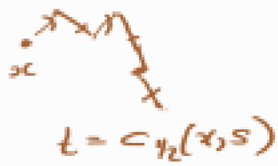
$c_{1/2}(x,S)$ = expected time to hit half the vertices of S starting from x

Then $c_{1/2}(x,S) \le 2 \overbrace{\max_{y\in S} h(x,y)}^{h_x} = 2h_x$

[Hence $c_{1/2}(x,S) \le 2h(G) \quad \forall x \in V(G)$.]

Proof:

(1) $\sum_{y\in S} h(x,y) \le |S|\, h_x$



(2) $\sum_{y\in S} h(x,y) \ge \frac{|S|}{2} c_{1/2}(x,S)$

(1)+(2) $\Rightarrow$ $c_{1/2}(x,S) \le 2h_x$. ∎

Recursively applying the lemma $\log_2 n$ times to the set of un-hit vertices gives Thm 4.

$$c(G) \le h(G)\left(1+\tfrac{1}{2}+\tfrac{1}{3}+\dots \tfrac{1}{n}\right). \quad \square$$

Lecture 21
3/5/2021

Singular Value Decomposition
SVD

$A_{m\times n}$

Consider n points $x_1, \dots, x_n$ in $\mathbb{R}^d$ $(d \gg 1)$.

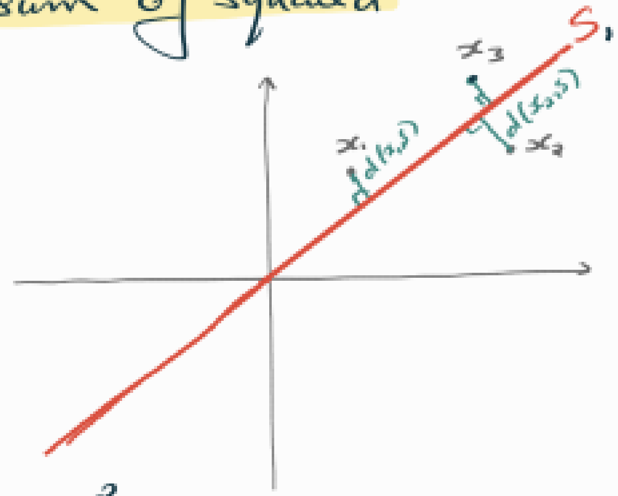
Can we find a lower (k) dimensional subspace S_k of $\mathbb{R}^d$ s.t. $x_1, \dots, x_n$ can be "approximated well" by points $y_1, \dots, y_n$ in S_k ?

$\rightsquigarrow$ "DIMENSIONALITY REDUCTION"

"Approximated well" can have various interpretations.

SVD - interpretation.

Choose the k-dimensional subspace which minimises the sum of squared Euclidean distances



That is,

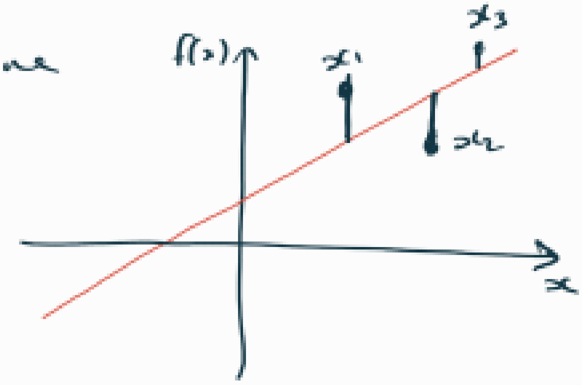
$$S_k = \underset{S \in \mathcal{S}}{\arg\min} \sum_{i=1}^{n} \operatorname{dist}^2(x_i, S),$$

where the minimization is over all k-dimensional subspaces of $\mathbb{R}^d$

Notes

1. It is different from the least squares regression line

   - vertical distance
   - not necessarily through origin.



2. Once you find S_k, $\forall i \in [n]$ let

$$y_i = \mathrm{Proj}_{S_k}(x_i) \text{ be the orthogonal projection of } x_i \text{ on } S_k.$$

Then the matrix $B_k = \begin{bmatrix} - y_1 - \\ \vdots \\ - y_n - \end{bmatrix}_{n \times d}$ ← rank$(B_k) \leq k$

is the best k-rank approximation for the data matrix $A = \begin{bmatrix} - x_1 - \\ \vdots \\ - x_n - \end{bmatrix}_{n \times d}$

That is

$$B_k = \underset{B \in \mathcal{B}}{\arg\min} \; \|A - B\|_F ,$$

where the minimization is over all $n \times d$ real matrices with rank $\leq k$.

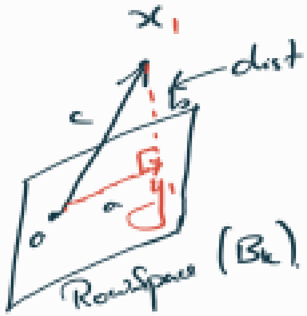
$$\left\lceil \|A\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2 = \sum_{i=1}^{m} \|\mathrm{row}_i(A)\|^2 = \sum_{j=1}^{n} \|\mathrm{col}_j(A)\|^2 \right\rfloor$$

Proof Idea

$$\|A-B_k\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{d} (x_{ij}-y_{ij})^2$$

$$= \sum_{i=1}^{n} \|x_i - y_i\|^2$$

$$= \sum_{i=1}^{n} \mathrm{dist}^2(x_i, S_k).$$



$a^2+b^2=c^2$

But how do we find S_k?

- Equivalently, we find an orthonormal basis $\{v_1, \dots, v_k\}$ for S_k

- In fact we do more. We find an orthonormal basis $\{v_1, \dots, v_d\}$ of $\mathbb{R}^d$ such that for any $k \in [d]$,

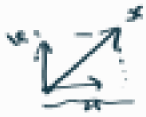
  $\{v_1, \dots, v_k\}$ gives the best-fit k-dim subspace for the data.

- $\underset{\text{Minimise}}{\mathrm{dist}^2(x_i, S_k)} + \underset{\text{Maximise}}{\|y_i\|^2} = \underset{\text{Independent of } S_k}{\|x_i\|^2}$ (Pythagoras)

$$\boxed{\underbrace{\sum_{i=1}^{n} dist^2(x_i, S)}_{\text{Minimise}} + \underbrace{\sum_{i=1}^{n} \|y_i\|^2}_{\text{Maximise}} = \underbrace{\sum_{i=1}^{n} \|x_i\|^2}_{\text{Ind of } S_k}}$$

$y_i = proj_{S_k}(x_i)$

- $$S_k = \arg\max_{S} \sum_{i=1}^{n} \|proj_S(x_i)\|^2$$



- $$\|proj_S(x_i)\|^2 = \|proj_{v_1}(x_i)\|^2 + \dots + \|proj_{v_k}(x_i)\|^2$$
$$= \langle x_i, v_1\rangle^2 + \dots + \langle x_i, v_k\rangle^2$$

- $$\sum_{i=1}^{n} \langle x_i, v_1\rangle^2 = \left\| \begin{bmatrix} \langle x_1, v_1\rangle \\ \vdots \\ \langle x_n, v_1\rangle \end{bmatrix} \right\|^2$$
$$= \left\| \begin{bmatrix} - x_1 - \\ \vdots \\ - x_n - \end{bmatrix} \begin{bmatrix} | \\ v_1 \\ | \end{bmatrix} \right\|^2$$
$$= \|A v_1\|^2$$

$a_1^2 + \dots a_n^2$

$\left\| \begin{bmatrix} a_1 \\ a_n \end{bmatrix} \right\|^2$

- $$\sum_{i=1}^{n} \|proj_S(x_i)\|^2 = \underbrace{\|Av_1\|^2 + \dots + \|Av_k\|^2}_{\text{Maximise}} \qquad -(1)$$

- Suppose $k = 1$.
- That is find the best fit line (through origin) for the data.

Then $$v_1 = \arg\max_{\substack{v \in \mathbb{R}^d \\ \|v\| = 1}} \|Av\|.$$

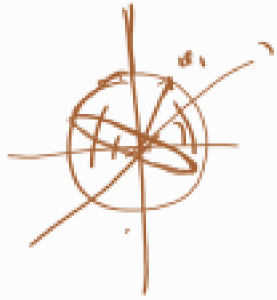
(ties can be broken arbitrarily).

and $S_1 = Span(\{v_1\})$.

Further $\max_{\|v\|=1} \|A(v)\| = \|Av_1\|$ is called the First Singular Value of A, denoted by $\sigma_1(A)$

Finding v_1 and σ_1 are therefore classic numerical analysis problems.

- Suppose $k = 2$.

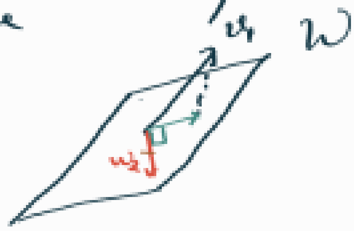


Let
$$v_1 = \underset{\|v\|=1}{\arg\max} \|Av\|$$

$$v_2 = \underset{\substack{\|v\|=1, \\ v \perp v_1}}{\arg\max} \|Av\|$$

$$V_2 = \text{span}(\{v_1, v_2\})$$

Claim: $S_2 = V_2$. (V_2 is the best-fit 2D subspace)



Proof: W be any 2D subspace.

- Pick $w_2 \in W$ s.t. $w_2 \perp v_1$ and $\|w_2\| = 1$.

- Pick $w_1 \in W$ s.t. $w_1 \perp w_2$ and $\|w_1\| = 1$.

So $\{w_1, w_2\}$ is an ONB for W

$$\|Aw_2\| \leq \max_{\substack{\|v\|=1 \\ v \perp v_1}} \|Av\| = \|Av_2\|.$$

$$\|Aw_1\| \leq \max_{\|v\|=1} \|Av\| = \|Av_1\|.$$

So $\|Aw_1\|^2 + \|Aw_2\|^2 \leq \|Av_1\|^2 + \|Av_2\|^2$

Exercise $\{v_1, \dots v_s\}$

SVD Algorithm

Input : $A = \begin{bmatrix} - x_1 - \\ - x_2 - \\ \vdots \\ - x_n - \end{bmatrix}_{n\times d} \in \mathbb{R}^{n\times d}$

Output : An ONB $\{v_1, \ldots, v_d\}$ of $\mathbb{R}^d$ s.t.
$\forall k$. $\sum_{i=1}^{k} \|Av_i\|^2 = \max \sum_{i=1}^{k} \|Aw_i\|^2$,
where the maximisation is over all sets of k orthonormal vectors $\{w_1, \ldots, w_k\}$.

$$v_1 = \arg\max_{\|v\|=1} \|Av\| \quad , \quad \sigma_1 = \|Av_1\|$$

$$v_2 = \arg\max_{\substack{\|v\|=1 \\ v \perp v_1}} \|Av\| \quad , \quad \sigma_2 = \|Av_2\|$$

$\sigma_1 \geq \sigma_2$
$\geq \sigma_3$
$\vdots$

$$v_3 = \arg\max_{\substack{\|v\|=1 \\ v \perp v_1, v_2}} \|Av\| \quad , \quad \sigma_3 = \|Av_3\|$$

$\vdots$

$$v_d = \arg\max_{\substack{\|v\|=1 \\ v \perp v_1, \ldots, v_{d-1}}} \|Av\| \quad , \quad \sigma_d = \|Av_d\|$$

Return $(\{\sigma_1, \ldots, \sigma_d\}, \{v_1, \ldots, v_d\})$.

Singular values — Singular vectors.

Proof of correctness :

Exercise. Hint : Induction on k.

Easy Observation:

$$\sigma_1^2 + \cdots + \sigma_d^2 = \|A\|_F^2$$

Proof: If $k = d$, "best-fit" = "perfect-fit"

$$\boxed{\underbrace{\sum_{i=1}^{n} \mathrm{dist}^2(x_i, S_k)}_{0} + \underbrace{\sum_{i=1}^{n} \|y_i\|^2}_{\sum_{i=1}^{d} \|Av_i\|^2 = \sum_{i=1}^{d} \sigma_i^2} = \underbrace{\sum_{i=1}^{n} \|x_i\|^2}_{\|A\|_F^2}}$$

Minimize — Maximise. — Ind of S

S_d

(by 1)

SVD (contd...)

LECTURE 22
4/5/2021

MATRIX MULTIPLICATION – Different pictures

I $\quad A_{m\times n}\, x_{n\times 1} = b_{m\times 1}$ $\qquad c_{ij} = \sum_k a_{ik} b_{jk}$

View 1

$$\begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}_{m\times n} \begin{bmatrix} x_1 \\ \\ \vdots \\ \\ x_n \end{bmatrix}_{n\times 1}$$

$$\downarrow$$

$$\begin{bmatrix} x_1 & x_2 & & x_n \end{bmatrix}$$

$$\begin{bmatrix} \times & & \times & & & & \times \\ | & & | & & & & | \\ a_1 & + & a_2 & + & \cdots & + & a_n \\ | & & | & & & & | \end{bmatrix} = \underbrace{\sum_{i=1}^{n} \underset{\text{coeff}}{x_i}\, \underset{\text{vector}}{\mathrm{col}_i(A)}}_{\text{lin. comb}}$$

View 2

$$\begin{bmatrix} — & a_1 & — \\ & \vdots & \\ — & a_m & — \end{bmatrix}_{m\times n} x_{n\times 1} = \begin{bmatrix} \langle a_1, x\rangle \\ \langle a_2, x\rangle \\ \vdots \\ \langle a_m, x\rangle \end{bmatrix}$$

II $\quad x_{1\times m}\, A_{m\times n} = b_{1\times n}$ $\qquad [b_1 \; - - \; b_n]$

$x\,[\quad\;]$

$$\begin{matrix} x_1 & — & a_1 & — \\ \vdots & & \vdots & \\ x_m & — & a_m & — \end{matrix}$$

Two views : 1. Linear combination of rows of A

2. A row vector of inner product of x with columns of A.

III $\quad A_{m\times n} B_{n\times p} = C_{m\times p}$

View 1.

$$A \begin{bmatrix} | & & | \\ b_1 & \cdots & b_p \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ c_1 & \cdots & c_p \\ | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ Ab_1 & Ab_2 & \cdots & Ab_p \\ | & | & & | \end{bmatrix}$$

where i-th col of C = $A \times$ i-th col of B

(2 views).

View 2.

i-th row of C = (i-th row of A) $\times B$

$$\begin{bmatrix} - & a_1 & - \\ & \vdots & \\ - & a_m & - \end{bmatrix} B = \begin{bmatrix} - & c_1 & - \\ & \vdots & \\ - & c_m & - \end{bmatrix} = \begin{bmatrix} - & a_1 B & - \\ & \vdots & \\ - & a_m B & - \end{bmatrix}$$

View 3.

$$\begin{bmatrix} - & a_1 & - \\ & \vdots & \\ - & a_m & - \end{bmatrix} \begin{bmatrix} | & & | \\ b_1 & \cdots & b_p \\ | & & | \end{bmatrix} = \overbrace{\begin{bmatrix} \langle , \rangle & \cdots & \langle , \rangle \\ & & \\ \langle , \rangle & \cdots & \langle , \rangle \end{bmatrix}}^{C}$$

$$c_{ij} = \langle a_i, b_j \rangle$$
$$= \langle \text{i-th row of } A, \text{ j-th col of } B \rangle$$

$c_{ij} = \sum a_{ik} b_{kj}$

$$\begin{bmatrix} | & & | \\ Ab_1 & \cdots & Ab_p \\ | & & | \end{bmatrix}$$

$$\sum_{i=1}^{n} \begin{bmatrix} | & & | \\ a_i & & a_i \\ | & & | \end{bmatrix}$$

View 4.

$$\begin{bmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{bmatrix} \begin{bmatrix} - & b_1 & - \\ & \vdots & \\ - & b_n & - \end{bmatrix} = \sum_{i=1}^{n} \underbrace{\underbrace{\begin{bmatrix} | \\ a_i \\ | \end{bmatrix}}_{m\times 1} \begin{bmatrix} - & b_i & - \end{bmatrix}_{1\times p}}_{m\times p}$$

PROJECTION MATRIX

Let • W be any k-dim subspace of $\mathbb{R}^d$ $(k \le d)$.

• $\{w_1, \dots, w_k\}$ be any <u>ONB</u> for W.

• $W = \begin{bmatrix} | & | & & | \\ w_1 & w_2 & \cdots & w_k \\ | & | & & | \end{bmatrix}_{d \times k}$

$x_1, \dots x_n \in \mathbb{R}^d$

$A = \begin{bmatrix} - & x_1 & - \\ & \vdots & \\ - & x_n & - \end{bmatrix}$

S_k

$y = P_W x$

$P_W = WW^T$

<u>Claim:</u>

$\forall x \in \mathbb{R}^d$,

$$\mathrm{proj}_W(x) = \underbrace{WW^T}_{P_W,\ \text{the projection matrix for } W.} x$$

<u>Proof</u>: (rearrangement)

$$\mathrm{proj}_W(x) = \langle x, w_1 \rangle w_1 + \cdots + \langle x, w_k \rangle w_k.$$

$$= \begin{bmatrix} | & & | \\ w_1 & \cdots & w_k \\ | & & | \end{bmatrix} \begin{bmatrix} \langle x, w_1 \rangle \\ \vdots \\ \langle x, w_k \rangle \end{bmatrix}$$

$$= \begin{bmatrix} | & & | \\ w_1 & \cdots & w_k \\ | & & | \end{bmatrix} \begin{bmatrix} - & w_1 & - \\ & \vdots & \\ - & w_k & - \end{bmatrix} \begin{bmatrix} | \\ x \\ | \end{bmatrix}$$

Check $W^TW = I$

$$= WW^Tx.$$

Corollary: If $\{w_1, \dots, w_d\}$ is an ONB for $\mathbb{R}^d$

then $WW^T = I$.

Claim 2.

$$WW^T = \sum_{i=1}^{k} w_i w_i^T$$

Proof:

Standard property of matrix multiplication.

$$C_{p \times q} = A_{p \times k} B_{k \times q}$$

$$c_{ij} = \sum_{l=1}^{k} a_{il} b_{lj}$$

$a_{il} b_{lj}$ = (i,j)-th entry of $col_l(A)\, row_l(B)$.

Combining

Let $\sigma_1, \dots, \sigma_d$ and $v_1, \dots, v_d$ be singular values & sing. vectors of a matrix $A_{n \times d}$.

Then,

- $VV^T = I$
- $VV^T = \sum_{i=1}^{d} v_i v_i^T$
- $$\begin{aligned} A &= AVV^T \\ &= A \sum_{i=1}^{d} v_i v_i^T \\ &= \sum_{i=1}^{d} (Av_i)\, v_i^T \\ &= \sum_{i=1}^{d} \sigma_i u_i v_i^T, \end{aligned}$$

(notation $Av_i = \underset{\sigma_i}{c}\, \underset{\|u_i\|=1}{u_i}$

where $u_i = \dfrac{Av_i}{\|Av_i\|}$ are called the left singular vectors.

Also

$$A = AVV^T$$

$$= \underbrace{A \begin{bmatrix} | & & | \\ v_1 & \cdots & v_d \\ | & & | \end{bmatrix}}_{\begin{bmatrix} | & & | \\ Av_1 & \cdots & Av_d \\ | & & | \end{bmatrix}} V^T$$

$$= \begin{bmatrix} | & & | \\ \sigma_1 u_1 & \cdots & \sigma_d u_d \\ | & & | \end{bmatrix} V^T$$

$$A = \begin{bmatrix} | & & | \\ a_1 & \cdots & a_d \\ | & & | \end{bmatrix}$$

$$C = \begin{bmatrix} \alpha_1 a_1 & \cdots & \alpha_n a_n \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} | & & | \\ u_1 & \cdots & u_d \\ | & & | \end{bmatrix}}_{\mathcal{U} \quad m \times d} \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_d \end{bmatrix}_{d \times d}}_{\Sigma} V^T$$

$$= \mathcal{U} \Sigma V^T$$

SVD(A)

Diagonal matrix

V: cols form an o.n.b

U:

$$A = \mathcal{U} \Sigma V^T$$

$$= \sum_{i=1}^{d} \sigma_i u_i v_i^T$$

singular vector

left sing. vector

sing. values

$$\mathcal{U} = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_d \\ | & & | \end{bmatrix}_{n \times d} \quad \Sigma = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ 0 & \cdots & \sigma_d \end{bmatrix}_{d \times d} \quad V = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_d \\ | & & | \end{bmatrix}_{n \times d}$$

Connection to Eigen Values?

$$AA^T = \mathcal{U}\Sigma V^T (\mathcal{U}\Sigma V^T)^T$$

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

$$= \mathcal{U}\Sigma^2 \mathcal{U}^T$$

$\|Av\|$

"Eigen decomposition" of AA^T ← real, sym., square.

$\sigma_1^2, \dots, \sigma_d^2$ are eigen values of AA^T.