# Random Walks and Mixing

15-859NN
2/8/21

## Random Walk as a matrix-vector

$G = (V, E, w)$; $A_{ij} = w_{ij}$; $D_{ii} = \sum_j w_{ij}$

Let $p_i^{(t)} \equiv$ prob at vertex $V_i$ at time $t$.

$p^{(0)} \equiv$ starting configuration.

---

<u>Claim</u> $A^T D^{-1} p^{(t)} = p^{(t+1)}$

<u>note</u> $\text{Prob}(V_i \text{ to } V_j) = \dfrac{w_{ij}}{d_i} = (A^T D^{-1})_{ji}$

$$\underbrace{\begin{bmatrix} | & & | & & | \\ w_1^T & \cdots & w_i^T & \cdots & w_n^T \\ | & & | & & | \end{bmatrix}}_{A^T} \cdot \underbrace{\begin{bmatrix} d_1^{-1} & & 0 \\ & \ddots & \\ 0 & & d_n^{-1} \end{bmatrix}}_{D^{-1}} = \begin{pmatrix} w_{i1}/d_i \\ \vdots \\ w_{in}/d_i \end{pmatrix}$$

<u>Note</u> All col sums in $A^T D^{-1} \equiv 1$

<u>Def</u> $A^T D^{-1} \equiv$ transition matrix

In our case

$$A = A^T \;\&\; \sum_i P_i^{(0)} = 1 \;\&\; P_i^{(0)} \geq 0$$

---

<u>Two Natural Questions</u>

1) $\exists$ dist $\bar{P}$ s.t. $AD^{-1}\bar{P} = \bar{P}$ ?

(Stationary dist)

2) $\forall P_0$ $\lim_{n \to \infty} (AD^{-1})^n P^{(0)} = \bar{P}$ ?

# Answers

1) Yes & no

Let $d = \sum d_i$ $\quad \pi = \begin{pmatrix} d_1/d \\ \vdots \\ d_n/d \end{pmatrix} = 1/d \, D \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

pf

$$A^T D^{-1} \pi = A^T D^{-1} (1/d) D \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} =$$

$$1/d \, A^T \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 1/d \, A \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 1/d \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix} = \pi$$

thus $\tilde{d} = \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix}$ is an eigen vector with value 1.

In general not unique

$G \cong$  $\quad A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ note $\lambda(A) = \{\pm 1\}$

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} = - \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

2) In general no.

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$ $\left( \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ will not converge} \right)$

<u>Thm</u> If $G$ is <u>not bipartite</u> and connected then

$\forall p^{(0)}$ $|p^{(0)}|_1 = 1$ & $p^{(0)} \geq 0$ then

$$\lim_{k \to \infty} (AD^{-1})^k p^{(0)} = \pi$$

2) We will prove convergence by answering a more general question:

<u>Question</u> How fast does a walk "mix" on $G$.

<u>Goal</u>: Use Spectral Thm.

<u>Prob</u>: $A$ symmetric but $AD^{-1}$ not!

We do a change of variables

$$AD^{-1} \longrightarrow \tilde{A} = D^{-1/2} A D^{-1/2} \quad (\tilde{A} \text{ sym})$$

$$p^{(k)} \longrightarrow \tilde{p}^{(k)} = D^{-1/2} p^{(k)}$$

$$\pi \longrightarrow \tilde{\pi} = D^{-1/2} \pi$$

---

<u>Claim</u> $AD^{-1}x = \lambda x$ iff $\tilde{A}y = \lambda y$

where $y = D^{-1/2} x$.

<u>Pf</u> $(\Rightarrow)$

$$\tilde{A}y = (D^{-1/2} A D^{-1/2})(D^{-1/2} x) = D^{-1/2} A D^{-1} x$$

$$= \lambda (D^{-1/2} x) = \lambda y$$

$(\Leftarrow)$ Same.

# Mixing as fcn of Error



Thus $\hat{A}\tilde{\pi} = \tilde{\pi}$

Def (New def of error!)

Error: $\tilde{\varepsilon}^{(k)} = \tilde{\pi} - \tilde{A}^{k}\tilde{p}^{(0)} = \tilde{\pi} - \tilde{p}^{(k)}$

$\left(\tilde{\varepsilon}^{(k)} = D^{-1/2}\varepsilon^{(k)}\right)$

Question?

How fast does $\tilde{\varepsilon}^{k}$ go to 0 with k?

(If at all!)

Note $\hat{A}\tilde{\varepsilon}^{(k)} = \hat{A}\tilde{\pi} - \hat{A}^{(k+1)}\tilde{p}^{(0)}$

$$= \tilde{\pi} - \tilde{A}^{(k+1)}\hat{p}^{(0)} = \tilde{\varepsilon}^{(k+1)}$$

Thus $\tilde{\varepsilon}^{(k)} = \tilde{A}^{(k)}\varepsilon^{(0)}$

A much simpler recurrence!

<u>Claim</u> $\tilde{\varepsilon}^{(0)} \perp \tilde{\pi}$ i.e. $\varepsilon^{(0)T}\tilde{\pi} = 0$

i.e. $(\tilde{\pi} - \tilde{p}^{(0)})^T \tilde{\pi} = 0$

to show: $\tilde{\pi}^T \tilde{\pi} = p^{(0)T} \tilde{\pi}$

$$\tilde{\pi} = D^{-1/2}\pi = \begin{pmatrix} \frac{1}{\sqrt{d_1}} & & \\ & \ddots & \\ & & 1/\sqrt{d_n} \end{pmatrix} \begin{pmatrix} d_1/d \\ \vdots \\ d_n/d \end{pmatrix} = \begin{pmatrix} \sqrt{d_1}/d \\ \vdots \\ \sqrt{d_n}/d \end{pmatrix}$$

$$\tilde{\pi}^T \tilde{\pi} = \sum \left( \sqrt{d_i}/d \right)^2 = \sum \frac{d_i}{d^2} = 1/d$$

---

$$\tilde{\pi}^T \tilde{p}^{(0)} = \left( \sqrt{d_1}/d, \cdots \sqrt{d_n}/d \right) \begin{pmatrix} p_1/\sqrt{d_1} \\ \vdots \\ p_n/\sqrt{d_n} \end{pmatrix}$$

$$= \sum p_i/d = \left(\sum p_i\right)/d = 1/d$$

$\left(\text{since } \sum p_i = 1\right)$ $\square$

# Spectral Thm

If $A$ is real sym matrix then

1) Eigenvalues of $A$ are real.

   i.e. $Ax = \lambda x \Rightarrow \lambda$ is real

2) If $Ax = \lambda x$ & $Ay = \mu y$ & $\lambda \neq \mu$

   then $x^T y = 0$ ie $x \perp y$.

3) $\exists$ orthonormal bases $y_1, \dots, y_n$ (eigenvectors)

   st
$$A = \begin{pmatrix} | & & | \\ y_1 & \cdots & y_n \\ | & & | \end{pmatrix} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \begin{pmatrix} - & y_1^T & - \\ & \vdots & \\ - & y_n^T & - \end{pmatrix}$$

4) $A = \sum_{i=1}^{n} \lambda_i y_i y_i^T$

# Perron-Frobenius Thm

Suppose $A^{n\times n} \geq 0$

Graph$(G)$ is strongly connected.

<u>Def</u> $z = a+ib \in \mathbb{C}$

$$|z| = \sqrt{z^* z} = \sqrt{a^2+b^2}$$

<u>Spectral Radius</u> $\rho(A) = \max_{\lambda \in \lambda(A)} |\lambda|$

e.g.

$$\begin{tikzcd} V_1 \arrow[rr] & & V_2 \arrow[dl] \\ & V_3 \arrow[ul] & \end{tikzcd} \qquad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

$w \equiv$ 3rd root of unity

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ w \\ w^2 \end{pmatrix} = \begin{pmatrix} w \\ w^3 \\ 1 \end{pmatrix} = w\begin{pmatrix} 1 \\ w \\ w^2 \end{pmatrix}$$

$\Rightarrow$ $w$ is en eigenvalue

$\lambda(A) = 1, w, w^2$

<u>Thm</u> (PF)

1) $\rho(A)$ is a simple eigenvalue of $A$.

If $x$ is its eigenvector then

$\text{sign}(x_i) = \text{sign}(x_j) \neq 0 \quad \forall i,j$

2) $\theta \in \lambda(A)$ & $|\theta| = \rho(A)$ then

$\theta / \rho(A)$ is an $m$th root of unity

and all cycles in $G(A)$ have

length a multiple of $m$.

3) Only non-neg eigenvector

is $x$.

pf Maybe?

Back to the symmetric case.

Suppose eigenvalues are

$$-1 < \lambda_1 \leq \cdots \leq \lambda_n = 1$$

( $-1 < \lambda_1$ true by Sym Perron-F, next lecture )

Orthonormal vectors are

$$Y_1, \ldots, Y_n = \tilde{\pi} \cdot d$$

We know that

$$\tilde{\varepsilon}^{(0)} = \alpha_1 Y_1 + \cdots + \alpha_{n-1} Y_{n-1}$$

$$\tilde{\varepsilon}^{(1)} = \tilde{A} \tilde{\varepsilon}^{(0)} = \lambda_1 \alpha_1 Y_1 + \cdots + \lambda_{n-1} \alpha_{n-1} Y_{n-1}$$

$$\tilde{\varepsilon}^{(k)} = \lambda_1^k \alpha_1 Y_1 + \cdots + \lambda_{n-1}^k \alpha_{n-1} Y_{n-1}$$

We will show that:
The mixing rate determined by

$$\lambda = \max\{|\lambda_1|, |\lambda_{n-1}|\}$$

What norm do we want our error?

$|\varepsilon|_1, |\varepsilon|_2, |\varepsilon|_\infty$ ?

Consider $|\tilde{\varepsilon}^{(k)}|_2$ (How does it compare to $|\varepsilon^{(k)}|_2$?)

Since $y_i$'s are orthonormal

$$|\varepsilon^{(0)}|_2 = \sqrt{\sum \alpha_i^2 |y_i|_2} = \sqrt{\sum \alpha_i^2}$$

pick $k$ s.t $\lambda^k \le 1/2$ (G not bipartite.)

$$|\tilde{\varepsilon}^{(k)}|_2 = \sqrt{\sum \lambda_i^{2k} \alpha_i^2} \le \sqrt{\sum \lambda^{2k} \alpha_i^2}$$

$$= \lambda^k \sqrt{\sum \alpha_i^2} \le \frac{1}{2}\sqrt{\sum \alpha_i^2} = \frac{1}{2}|\tilde{\varepsilon}^{(0)}|_2$$

---

Thus every $k$ steps error goes down by $1/2$.

**Def** Mixing rate

$$\min_k \; 2\|\varepsilon^{(k)}\| \le \|\varepsilon^{(0)}\|$$

# $L_1$ norm & Cauchy-Schwartz

(CS) $a, b \in \mathbb{R}^n$

$$(a^T b)^2 \le (a^T a)(b^T b)$$

---

<u>Def:</u> If $a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ then $|a| = \begin{pmatrix} |a_1| \\ \vdots \\ |a_n| \end{pmatrix}$

---

e.g. $b = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ $a$ arbitrary

$$|a|_1^2 = |a|^T \cdot b \le |a|^T |a| \cdot n$$

$$= a^T a \cdot n$$

Thus $|a|_1 \le \sqrt{n}\, |a|_2$

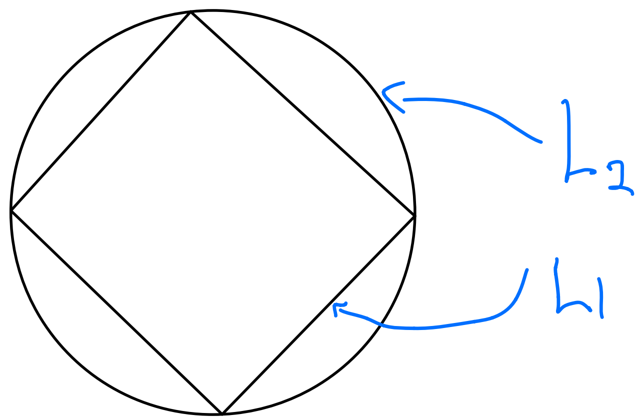
Claim: Mixing rate in $L_1$ is

$O(\log n \, (\text{Mixing in } L_2))$

(Better estimates for $L_1$ error are harder!)

Pf note

$|a|_2 \le |a|_1$




$$|\mathcal{E}_{\log n \cdot k}|_1 \le \sqrt{n}\, |\mathcal{E}_{\log n \cdot k}|_2$$

$$\le \left(1/2^{\log n}\right) \sqrt{n}\, |\mathcal{E}_0|_2$$

$$\le 1/2\, |\mathcal{E}_0|_2 \le 1/2\, |\mathcal{E}_0|_1$$

# PROOF OF SPECTRAL THEOREM

**Theorem 1** (Spectral Theorem). *Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix. Then*

(1) *All eigenvalues of $A$ are real.*
(2) *There exists an orthogonal matrix $Q$ and a diagonal matrix $\Lambda$ such that $A = Q\Lambda Q^T$.*

*Proof.* We have proved (1) in the class. Only need to prove (2). We make an induction on $n$.

When $n = 1$, the claim is obvious. Now assume that the claim is valid for $n = m$, that is, for any $m\times m$-real symmetric matrix $A$, there exists an orthogonal matrix $Q$ and diagonal matrix $\Lambda$ such that $A = Q\Lambda Q^T$. Let us consider $(m+1)\times(m+1)$-real symmetric matrix $A$. By (1), $A$ has a real eigenvalue $\lambda$ with eigenvector $\alpha$. We see that all entries of $\alpha$ must be real numbers. By Gram-Schmidt process, we may assume that there exists an *orthonormal* basis $q_1, \dots, q_n$ with $q_1 = \alpha$. Let $P := (q_1 q_2 \cdots q_n)$ and $C := P^TAP = (c_{ij})_{(m+1)\times(m+1)}$. We claim that $c_{11} = \lambda$ and $c_{i1} = 0$ for $i \neq 1$. In fact, note that $P$ is an orthogonal matrix, we have $AP = PC$, that is, $A(q_1q_2\cdots q_n) = (q_1q_2\cdots q_n)C$. Therefore, we have $Aq_1 = \sum_{i=1}^{m+1} c_{i1}q_i$. But $q_1 = \alpha$ is an eigenvector,so $\lambda q_1 = \sum_{i=1}^{m+1} c_{i1}q_i$. Since $q_1, \cdots, q_n$ are linearly independent. So $c_{11} = \lambda$ and $c_{i1} = 0$ for $i \neq 1$. So $C$ has four blocks like $\begin{pmatrix}\lambda & \star\\ 0 & \tilde{A}\end{pmatrix}$. Note that $C = P^TAP$ is symmetric(why ?), thus $\star = 0$. So $C = \begin{pmatrix}\lambda & 0\\ 0 & \tilde{A}\end{pmatrix}$ and $\tilde{A}$ has to be symmetric matrix with the size $m \times m$. By induction, there exists an orthogonal matrix $Q$ and diagonal matrix $\Lambda$ such that $\tilde{A} = Q\Lambda Q^T$. Therefore

$$C = \begin{pmatrix}\lambda & 0\\ 0 & \tilde{A}\end{pmatrix} = \begin{pmatrix}\lambda & 0\\ 0 & Q\Lambda Q^T\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & Q\end{pmatrix}\begin{pmatrix}\lambda & 0\\ 0 & \Lambda\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & Q\end{pmatrix}^T$$

Therefore

$$A = PCP^T = P\begin{pmatrix}1 & 0\\ 0 & Q\end{pmatrix}\begin{pmatrix}\lambda & 0\\ 0 & \Lambda\end{pmatrix}(P\begin{pmatrix}1 & 0\\ 0 & Q\end{pmatrix})^T$$

and we easily check $P\begin{pmatrix}1 & 0\\ 0 & Q\end{pmatrix}$ is an orthogonal matrix and we are done. $\square$

Consider Random walks on path graph $\equiv P_n$

We will show that $\lambda = \lambda(P_n) \approx (1 - 1/n^2)$

Note $(1 - 1/n^2)^{n^2} \approx 1/e$

Thus mixing rate for $P_n$ is $\approx n^2$!