Lecture 12

- linear algebra & mixing times
- Saving random bits via random walks

From last time:

Linear Algebra Review

<u>def</u> v is an eigenvector of A with corresponding eigenvalue λ iff

$$vA = \lambda v$$

<u>def</u> ℓ_2-norm of $v = (v_1 \cdots v_n) = \sqrt{\sum_{i=1}^{n} v_i^2}$

<u>def</u> $v^{(1)} \cdots v^{(m)}$ orthonormal if

$$\underbrace{v^{(i)} \cdot v^{(j)}}_{\text{inner product}} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

$$= \sum_{\ell} v_\ell^{(i)} \cdot v_\ell^{(j)}$$

<u>Thm</u> Transition matrix P real + symmetric

$\Rightarrow$ $\exists$ e-vecs $v^{(1)} \cdots v^{(n)}$

forming orthonormal basis with corresponding e-values $1 = \lambda_1 \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$

+ $v^{(1)} = \frac{1}{\sqrt{n}} (1 \ldots 1)$

chosen so that $\|v^{(1)}\|_2 = 1$

$\Rightarrow$ <u>any</u> vector w is expressible as linear combination of $v^{(i)}$'s

$$w = \sum \alpha_i v^{(i)}$$

+ L_2-norm of w is $\sqrt{\sum \alpha_i^2}$ (*)

From last time:

Useful Facts:

Assume P has all positive entries & e-vecs $v^{(1)} \dots v^{(n)}$ with corresponding e-vals $\lambda_1 \cdots \lambda_n$

Facts

(1) αP has e-vecs $v^{(1)} \dots v^{(n)}$ with corresponding evals $\alpha\lambda_1, \dots, \alpha\lambda_n$

(2) $P+I$ " " " " " " $\lambda_1+1, \dots, \lambda_n+1$

(3) P^k " " " " " " $\lambda_1^k, \dots, \lambda_n^k$

(4) P stochastic $\Rightarrow$ $|\lambda_i| \leq 1$ $\forall i$

Note: add self-loops; $\frac{P+I}{2}$ = "stay put with prob ½ & walk with prob ½"

$\Rightarrow$ new eigenvalues $\frac{\lambda_1+1}{2}, \dots, \frac{\lambda_n+1}{2}$

From last time:

Mixing Times

How long does it take to reach stationary distribution?

def. $\varepsilon > 0$

Mixing time, $T(\varepsilon)$, of M.C. A with stationary dist π is min t s.t.

$\forall\, \pi^{(0)}, \quad \|\pi - \pi^{(0)} A^t\|_1 < \varepsilon$

def. M.C. A is rapidly mixing if

$$T(\varepsilon) = \text{poly}(\log \underset{\uparrow}{n}, \log 1/\varepsilon)$$

\# states

<u>Thm</u> P is transition matrix of undirected,

can put self loop on each node $\rightarrow$ non k-partite, d-reg connected graph

π_0 is start dist.

π is stationary dist $= (\frac{1}{n}, \dots, \frac{1}{n})$

(so $\pi P = \pi$)

Then $\|\pi_0 P^t - \pi\|_2 \leq \underbrace{|\lambda_2|^t}$

exponentially decreasing dist if $1-\lambda_2$ is const!!

$\Rightarrow$ rapid mixing

<u>Proof</u>

P real, symmetric $\Rightarrow$

evecs $v^{(1)} \dots v^{(n)}$ are orthonormal basis

with e-vals $1 = \lambda_1 \geq |\lambda_2| \geq \dots \geq |\lambda_n|$

so <u>any</u> vector, in particular π_0,
can be expressed as lin comb
of $v^{(i)}$'s:

$$\pi_0 = \sum_{i=1}^{n} \alpha_i v^{(i)}$$

so $$\pi_0 P^t = \sum_{i=1}^{n} \alpha_i \underbrace{v^{(i)} \cdot P^t}_{= \lambda_i^t v^{(i)}}$$

$$= \underbrace{\alpha_1}_{=\frac{1}{\sqrt{n}}} \underbrace{\lambda_1^t}_{=1} v^{(1)} + \alpha_2 \lambda_2^t v^{(2)} + \ldots$$

What is α_1?

$v^{(1)} = \frac{1}{\sqrt{n}} (11 \ldots 1)$

$$\pi_0 \cdot v^{(1)} = \alpha \underbrace{v^{(1)} \cdot v^{(1)}}_{=1} + \sum_{i=2}^{n} \alpha_i \lambda_i^t \underbrace{v^{(i)} \cdot v^{(1)}}_{=0} = \alpha_1$$

also, $$\pi_0 \cdot v^{(1)} = \pi_0 \cdot \frac{1}{\sqrt{n}} (11 \ldots 1) = \frac{1}{\sqrt{n}} \underbrace{\pi_0 \cdot (11 \cdots 1)}_{=1}$$

so $$\alpha_1 = \frac{1}{\sqrt{n}}$$

note that this argument does not use any knowledge of π_0, other than it is a distribution.

Continuing...

$$\|\pi_0 \cdot P^t - \alpha_1 \cdot v^{(1)}\|_2 = \left\| \sum_{i=2}^{n} \alpha_i \lambda_i^t v^{(i)} \right\|_2$$

$$= \sqrt{\sum_{i=2}^{n} \alpha_i^2 \lambda_i^{2t}} \qquad \text{by } (*)$$

$$= |\lambda_2|^t \sqrt{\sum_{i=2}^{n} \alpha_i^2} \qquad \text{since } |\lambda_2| \geq |\lambda_3| \geq \dots$$

$$\leq |\lambda_2|^t \, \|\pi_0\|_2 \qquad \text{by } (*) \text{ + since } \sum_{i=1}^{n} \alpha_i^2 > \sum_{i=2}^{n} \alpha_i^2$$

$$\leq |\lambda_2|^t \qquad \text{since } \|w\|_2 \leq \|w\|_1 = 1 \text{ when entries} \leq 1$$

$\blacksquare$

since $|\lambda_2|^t \to 0$

$\alpha_1 \cdot v^{(1)} = \frac{1}{n} \cdot (1 \dots 1)$ has to be the stationary distribution

Reducing Randomness via Random Walks:

For language L,

let $\mathcal{A}$ be algorithm s.t.

(1) $\forall x \in L \quad \Pr_{\mathcal{A}\text{'s coins}}[\mathcal{A}(x)=1] \geq 99/100$ usually correct

(2) $\forall x \notin L \quad \Pr_{\mathcal{A}\text{'s coins}}[\mathcal{A}(x)=0] = 1$ alway correct

To get error $< 2^{-k}$

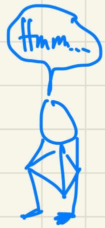
Method	# random bits used
1) run k times & output "$x \notin L$" if see 0 else output "$x \in L$"	$k \cdot r$
2) use pairwise ind random bits	$O(k+r)$
3) today! use random walks to choose bits	$r + O(k)$

Plan

- $\forall$ (random) string in $\{0,1\}^n$, assign it to node in graph G

- picking random n-bit string
  $\implies$ picking random node in G

  easier?

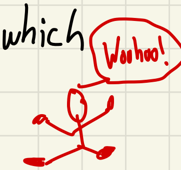


picking several random n-bit strings
$\implies$ picking several random nodes in G

easier?

picking several strings, one of which is "good"
$\implies$ picking several nodes, one of which is "good"

Easier!!



The graph G: ← we get to pick G!!!

- constant degree d-regular, connected, nonbipartite
- transition matrix P for r.w. on G has $|\lambda_2| \le \frac{1}{10}$

  d-reg $\Rightarrow$ stat dist π is uniform
- # nodes $= 2^r$ corresponds to all possible choices of r random bits

The Algorithm

<u>\# random bits</u>

- Pick random start node $w \in \{0,1\}^r$ $\quad$ r

- Repeat K times:

  $w \leftarrow$ random nbr of w $\quad$ $\underset{\substack{d \\ \text{is const}}}{O(1)} \times \underset{\#\text{loops}}{k}$

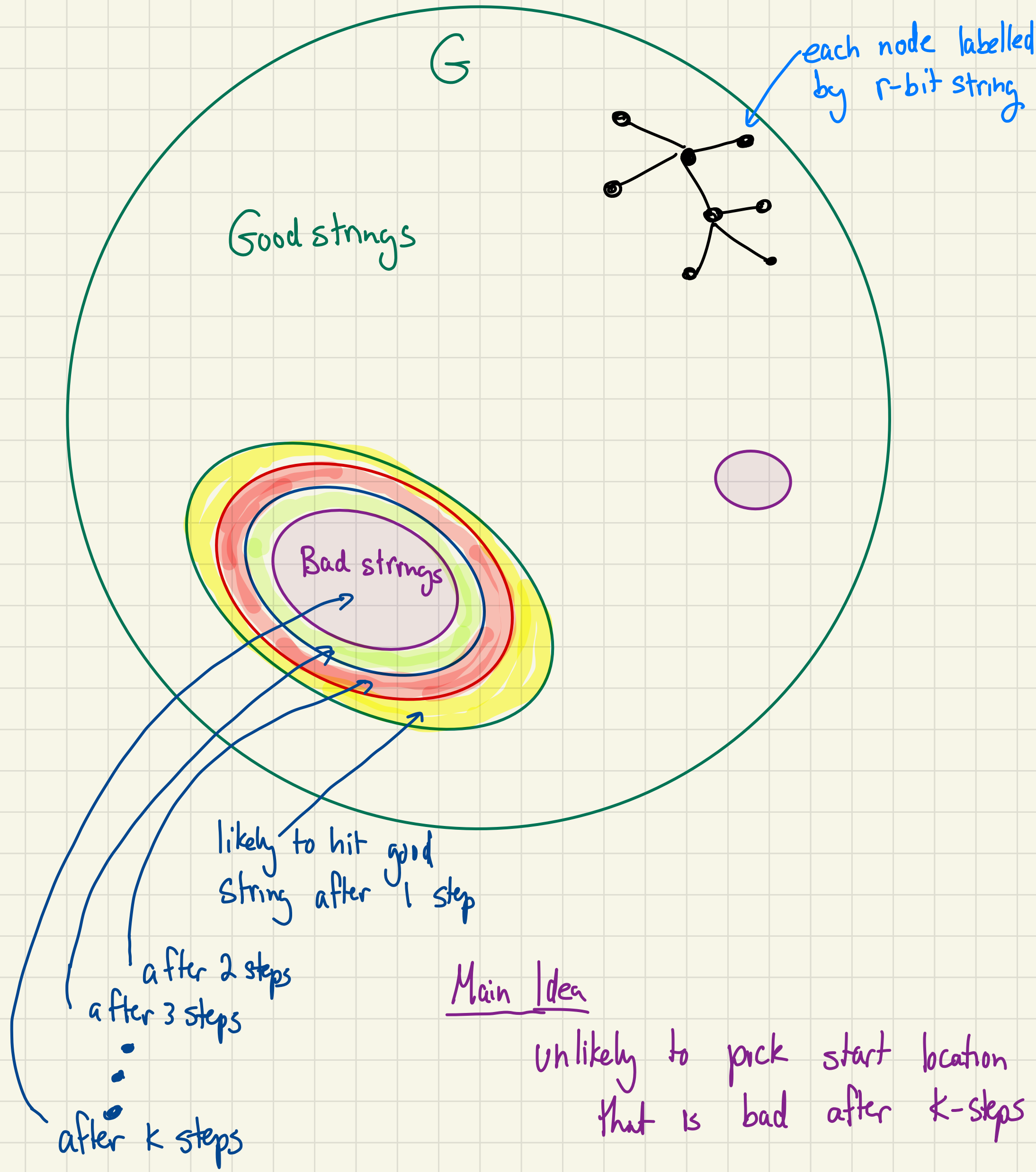
  run $\mathcal{A}(x)$ with w as random bits.

  If $\mathcal{A}(x)$ outputs "$x \in L$", output "$x \in L$" & halt
  else continue

- Output "$x \notin L$"

total: $r + O(k)$

Behavior: Claim: error of new algorithm is $\leq \left(\frac{1}{5}\right)^k$ for $x \in L$
(still 0 error for $x \notin L$)



Main Idea

unlikely to pick start location that is bad after k-steps

bad case: walk only on "bad strings" & never reach good strings

why is this possible if G arbitrary? e.g. line

Proof of Claim

$x \notin L$: algorithm <u>never</u> errs (no bad strings)

$x \in L$:

most random bits say $x \in L$: $\geq \frac{99}{100} \cdot 2^r$

define $B = \{ w \mid A(x)$ with random bits w is incorrect. i.e. says $x \notin L \}$

"bad w's"

$$|B| \leq \frac{2^r}{100}$$

need lin. alg. way of describing walks that stay in bad set:

define N diagonal matrix

$$N_w = \begin{cases} 1 & \text{if } w \in B \\ 0 & \text{o.w.} \end{cases}$$

$\leftarrow$ incorrect

$\leftarrow$ correct

Bad w's

$$N = \begin{pmatrix} 1\,1 & & & & & O \\ & 0\,0\,0\,1 & & & & \\ & & 0\,0\,0\,1\,1 & & & \\ & & & 0\,0\,0\,0 & & \\ & & & & 0\,0\,0\,1 & \\ O & & & & & 0\,0\,1 \end{pmatrix}$$

For q <u>any</u> probability dist:

$q \cdot N$ is??

example:

$q = \left(\frac{1}{4} \;\; \frac{3}{4}\right) \qquad N = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$

$q \cdot N = \left(\frac{1}{4} \;\; 0\right)$

this 0-entry zeroed out the 3/4

$q \cdot N$ deletes weight that finds a witness to $x \in L$

$$\|q \cdot N\|_1 = \Pr_{w \in q}[\, w \text{ is bad}\,]$$

Can compose:

$$\| q \cdot PN \|_1 = \Pr_{w \in q}[\text{start at } q, \text{ take a step \& land on "bad"}]$$

$\vdots$

$$\| q (PN)^i \|_1 = \Pr_{w \in q}[\text{start at } q, \text{ take } i \text{ steps \& each is "bad"}]$$

ignores whether start node bad. this just hurts us, so ok to ignore.

<u>Lemma</u> $\forall \pi \quad \| \pi PN \|_2 \leq \frac{1}{5} \| \pi \|_2$

First: how do we use lemma?

answer incorrect only if always see bad w's

$$\Rightarrow \Pr[\text{incorrect}] \leq \| p_0 (PN)^k \|_1$$

$$\leq \sqrt{2^r} \; \| p_0 (PN)^k \|_2$$

since $\|p\|_1 \leq \sqrt{\text{domain size}} \cdot \|p\|_2$

$$\le \cancel{\sqrt{2^r}}\ \underbrace{\|p_0\|_2}_{=\frac{1}{\cancel{\sqrt{2^r}}}} \left(\frac{1}{5}\right)^k$$

apply lemma k times

since start at uniform + L_2 norm of uniform $= \sqrt{\sum \left(\frac{1}{2^r}\right)^2} = \sqrt{\frac{1}{2^r}}$

$$= \left(\frac{1}{5}\right)^k$$

Proof of lemma:

let $v_1 \cdots v_{2^r}$ be e-vecs of P

+ v_1 is s.t. $\|v_1\|_2 = 1$ (so $v_1 = \left(\frac{1}{\sqrt{2^r}}, \dots, \frac{1}{\sqrt{2^r}}\right)$)

then $\pi = \sum_{i=1}^{2^r} \alpha_i v_i$

note: 1) $\|\pi\|_2 = \sqrt{\alpha_i^2}$ by (*) proved previously

2) $\forall\ w \quad \|wN\|_2 = \sqrt{\sum_{i \in B} w_i^2} \le \sqrt{\sum_i w_i^2} = \|w\|_2$

So:

$$\|\pi P N\|_2 = \left\| \sum_{i=1}^{2^n} \alpha_i v_i P N \right\|_2$$

since any π is lin comb of basis vectors

$$= \left\| \sum_{i=1}^{2^n} \alpha_i \lambda_i v_i N \right\|_2$$

$$\leq \underbrace{\| \alpha_1 \lambda_1 v_1 N \|_2}_{A} + \underbrace{\left\| \sum_{i=2}^{2^n} \alpha_i \lambda_i v_i N \right\|_2}_{B}$$

Cauchy-Schwarz

bound A:

$$\| \alpha_1 \lambda_1 v_1 N \|_2 = \| \alpha_1 v_1 N \|_2$$

since $\lambda_1 = 1$

$$= |\alpha_1| \cdot \sqrt{\sum_{i \in B} \left(\frac{1}{\sqrt{2^r}}\right)^2}$$

since $v_1 = \left(\frac{1}{\sqrt{2^r}}, \dots, \frac{1}{\sqrt{2^r}}\right)$ & $N = \begin{pmatrix} 0 \\ & \ddots & \\ & & 0 \end{pmatrix}$

uses that uniform dist is unlikely to be on a bad string

$$= |\alpha_1| \cdot \sqrt{\frac{|B|}{2^r}}$$

$$\leq \frac{|\alpha_1|}{10}$$

since $\frac{|B|}{2^r} \leq \frac{1}{100}$

$$\leq \frac{\|\pi\|_2}{10}$$

since $\|\pi\|_2 = \sqrt{\sum_{i=1} \alpha_i^2}$

bound B:

$$\left\| \sum_{i=2}^{2^n} \alpha_i \lambda_i v_i N \right\|_2 \leq \left\| \sum_{i=2}^{2^n} \alpha_i \lambda_i v_i \right\|_2$$

from note

Uses "mixing" of v_i's for $i > 2$. These could be "heavy" in bad areas, but won't stay for long !!

$$= \sqrt{\sum (\alpha_i \lambda_i)^2}$$

$$\leq \sqrt{\sum \alpha_i^2 \cdot \left(\frac{1}{10}\right)^2}$$

$\lambda_i \leq 1/10$

$$\leq \frac{1}{10} \cdot \|\pi\|_2$$

So: $\|\pi P N\|_2 \leq \dfrac{\|\pi\|_2}{5}$ ∎