3.5 Markov chain convergence & pagerank

Wednesday, January 22, 2020 7:56 PM

Thm. The ε-mixing time of a random walk on an undirected graph is

$$O\left(\frac{\ln(1/\pi_{min})}{\Phi^2 \varepsilon^3}\right),$$

where $\pi_{min} = \min_{x \in V} \pi_x$.

proof. Let $t = \underbrace{\frac{c \ln(1/\pi_{min})}{\Phi^2 \varepsilon^2}}$, for some suitable constant c.

Let $\vec{a} = \vec{a}(t) = \frac{1}{t}\left(\vec{p}(0) + \cdots + \vec{p}(t)\right)$.

Need to show $|\vec{a} - \vec{\pi}|_1 < \varepsilon$.

Let $v_i = \frac{a_i}{\pi_i}$. If $v_i > 1$, we call the state "heavy", because there is more mass in a_i than in the stationary π_i.

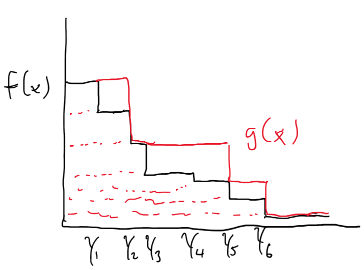
WLOG, reindex so $v_1 \ge v_2 \ge v_3 \ge \cdots$.

Let i_0 be the maximum i s.t. $v_i > 1$ (the last heavy state).

Then $|\vec{a} - \vec{\pi}|_1 = 2 \underbrace{\sum_{i=1}^{i_0} (v_i - 1)\pi_i}_{(a_i - \pi_i)^+} = 2 \underbrace{\sum_{i \ge i_0+1} (1 - v_i)\pi_i}_{(\pi_i - a_i)^+}$ (Prop 4.4)

Let $\gamma_i = \pi_1 + \cdots + \pi_i$.

Define $f : [0, \gamma_{i_0}] \to \mathbb{R}$ by $f(x) = v_i - 1$ for $x \in [\gamma_{i-1}, \gamma_i]$.



Then $\sum_{i=1}^{i_0} (v_i - 1)\pi_i = \int_0^{\gamma_{i_0}} f(x)\,dx$

Also, divide $\{1, \ldots, i_0\}$ into contiguous subsets $G_1, \ldots, G_r$, to be specified later.

Let $u_s = \max_{i \in G_s} v_i$.

Define $g : [0, \gamma_{i_0}] \to \mathbb{R}$ by $g(x) = u_s - 1$ for $x \in \bigcup_{i \in G_s} [\gamma_{i-1}, \gamma_i]$.

Since $g(x) \ge f(x)$, $\int_0^{\gamma_{i_0}} f(x)\,dx \le \int_0^{\gamma_{i_0}} g(x)\,dx$.

$$\int_0^{\gamma_{i_0}} g(x)\,dx = \sum_{s=1}^{r} \underbrace{\pi(U_1 \cup \cdots \cup U_r)}_{\text{width}} \underbrace{(u_s - u_{s+1})}_{\text{height}}$$ (WTS $\le \frac{\varepsilon}{2}$)

Note: If $2\sum_{i \geq i_0+1} (1-v_i)\pi_i \leq \varepsilon$, then we're already done

So consider case $\sum_{i \geq i_0+1} (1-v_i)\pi_i > \frac{\varepsilon}{2} \Rightarrow \sum_{i \geq i_0+1} \pi_i > \frac{\varepsilon}{2}$.

Then, for any subset A of heavy nodes,

$$\min\left(\pi(A), \pi(\bar{A})\right) \geq \frac{\varepsilon}{2} \cdot \pi(A)$$

incl all light nodes

$\left(\pi(\bar{A}) \geq \frac{\varepsilon}{2} \geq \frac{\varepsilon}{2} \cdot \underset{\leq 1}{\pi(A)}\right)$

$\left(\frac{\varepsilon}{2} < 1, \text{ so } \pi(A) \geq \frac{\varepsilon}{2} \cdot \pi(A)\right)$

Let's now define $G_1, \ldots, G_r$.

Let $G_1 = \{1\}$. If $G_1, \ldots, G_{s-1}$ have already been defined, let G_s start from $k = (\text{end of } G_{s-1}) + 1$. Let G_s end on an element ℓ, defined the largest integer $i \geq k$ and $i \leq i_0$ s.t.

$$\sum_{j=k+1}^{\ell} \pi_j \leq \frac{\varepsilon \Phi \gamma_k}{4} \qquad (\text{i.e. } \pi(\{k+1, \ldots, \ell\}) \leq$$

If $G_s = \{k, k+1, \ldots, \ell\}$,

Then by def.

$$\gamma_{\ell+1} = \gamma_k + \underbrace{\sum_{j=k+1}^{\ell+1} \pi_j}_{> \frac{\varepsilon \Phi \gamma_k}{4}} > \gamma_k + \frac{\varepsilon \Phi}{4}\gamma_k = \left(1 + \frac{\varepsilon \Phi}{4}\right)\gamma_k$$

because ℓ largest that's smaller

So $r \leq \ln_{\left(1+\frac{\varepsilon\Phi}{4}\right)} \frac{1}{\pi_1} + 2$ (because $\gamma_1 = \pi_1$, and each γ_i scales that up by $\left(1 + \frac{\varepsilon\Phi}{4}\right)$)

$$\gamma_{i_0} = \left(1 + \frac{\varepsilon\Phi}{4}\right)^{r-2} \pi_1 \quad , \quad \gamma_{i_0} \leq 1$$

$$\left(1 + \frac{\varepsilon\Phi}{4}\right)^{r-2} \pi_1 \leq 1$$

$$\left(1 + \frac{\varepsilon\Phi}{4}\right)^{r-2} \leq \frac{1}{\pi_1}$$

$$r \leq \ln_{\left(1+\frac{\varepsilon\Phi}{4}\right)} \frac{1}{\pi_1} + 2 = \frac{\ln \frac{1}{\pi_1}}{\ln\left(1+\frac{\varepsilon\Phi}{4}\right)} + 2 \leq \frac{\ln \frac{1}{\pi_1}}{\frac{\varepsilon\Phi}{4}} + 2 = \left(\ln \frac{1}{\pi_1}\right)\left(\frac{4}{\varepsilon\Phi}\right) + 2$$

Lemma: $\pi(G_1 \cup \cdots \cup G_r)(u_s - u_{s+1}) \leq \frac{8}{t\Phi\varepsilon}$

proof later

Then $\sum_{s=1}^{r} \pi(G_1 \cup \cdots \cup G_r)(u_s - u_{s+1}) \leq r \cdot \left[\frac{8}{t\Phi\varepsilon}\right]$

$$= \left[\left(\ln \frac{1}{\pi_1}\right) \cdot \left(\frac{4}{\varepsilon\Phi}\right) + 2\right] \left[\frac{8}{\Phi\varepsilon}\right] \left[\frac{c \ln\left(\frac{1}{\pi_{min}}\right)}{\Phi^2 \varepsilon^3}\right]^{-1}$$

$$= \left[\left(\ln \frac{1}{\pi_1}\right)\cdot\left(\frac{4}{\varepsilon\Phi}\right) + 2\right]\left[\frac{8}{\Phi\varepsilon}\right]\left[\frac{c\ln\left(\frac{1}{\pi_{min}}\right)}{\Phi^2\varepsilon^3}\right]^{-1}$$

$$= \left(\ln\frac{1}{\pi_1}\right)\cdot\left(\frac{4}{\varepsilon\Phi}\right)\cdot\left(\frac{8}{\Phi\varepsilon}\right)\cdot\frac{\Phi^2\varepsilon^3}{c\ln\left(\frac{1}{\pi_{min}}\right)} + \frac{16}{\Phi\varepsilon}\cdot\frac{\Phi^2\varepsilon^3}{c\ln\left(\frac{1}{\pi_{min}}\right)}$$

$$\le \frac{32}{c}\cdot\varepsilon + \frac{16\,\Phi\,\varepsilon^2}{c\ln\left(\frac{1}{\pi_{min}}\right)} < \varepsilon \quad \text{for some } c \qquad \left(\begin{array}{l}\Phi < 1\\ \ln\frac{1}{\pi_{min}} > 1\end{array}\right)$$

$\boxtimes$

proof of lemma: Uses probability flows, computed 2 ways, from heavy states to light states.

Consider a starting prob. $\vec{a}$. $\vec{a} - \vec{a}P$ is the net loss in prob. for each state after 1 step.

Consider a group $G_s = \{k, k+1, \dots, \ell\}$, $k < i_0$.

let $A = \{1, 2, \dots, k\}$.

The net loss for A in one step at time t is

$$\sum_{i=1}^{k}\left(a_i - (\vec{a}P)_i\right) \le \frac{2}{t} \qquad \left(\text{by convergence proof, since } \vec{a}(t)P - \vec{a}(t) = \frac{1}{t}(\vec{p}(t) - p(0))\right)$$

Consider another way to measure prob. loss. Take the difference of flow from A to $\bar{A}$ and the flow from $\bar{A}$ to A.

For any $i < j$

$$\text{net-flow}(i,j) = \text{flow}(i,j) - \text{flow}(j,i) = a_i p_{ij} - a_j p_{ji} = \pi_i p_{ij} v_i - \pi_j p_{ji} v_j$$

$$\text{(stationary def)} \qquad = \pi_j p_{ji}(v_i - v_j) \ge 0.$$

Thus, for any two nodes, there is nonnegative flow from heavier to lighter.

$\Rightarrow$ since $\ell \ge k$, flow from A to $\{k+1, k+2, \dots \ell\}$ minus flow from $\{k+1, \dots, \ell\}$ to A is non-negative.

For $i \le k$, $j > \ell$, $v_i \ge v_k$ and $v_j \le v_{\ell+1}$, so the net loss from A is at least

$$\sum_{\substack{i\le k\\ j>\ell}} \pi_j p_{ij}(v_i - v_j) \ge (v_k - v_{\ell+1})\sum_{\substack{i\le k\\ j>\ell}} \pi_j p_{ji}.$$

Thus $$(v_k - v_{\ell+1}) \sum_{\substack{i\le k \\ j>\ell}} \pi_j p_{ji} \le \frac{2}{t}$$ (By other computation)

But $$\sum_{i=1}^{k} \sum_{j=k+1}^{\ell} \pi_j p_{ji} \le \sum_{j=k+1}^{\ell} \pi_j \le \frac{\varepsilon \Phi \pi(A)}{4},$$ by def. of ℓ.

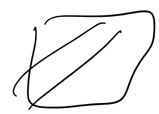
By def. of Φ and using $\min(\pi(A), \pi(\bar{A})) \ge \frac{\varepsilon}{2}\pi(A)$,

$$\sum_{i\le k<j} \pi_j p_{ji} \ge \Phi \min(\pi(A), \pi(\bar{A})) \ge \frac{\varepsilon \Phi \gamma_k}{2}$$

Then $$\sum_{\substack{i\le k \\ j>\ell}} \pi_j p_{ji} = \sum_{i\le k<j} \pi_j p_{ji} - \sum_{\substack{i\le k \\ j\le \ell}} \pi_j p_{ji} \ge \frac{\varepsilon \Phi \gamma_k}{4}.$$

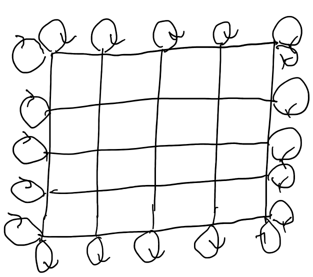
$$\Rightarrow v_k - v_{\ell+1} \le \frac{8}{t \varepsilon \Phi \gamma_k}.$$

(similar but easier for $k = i_0$)



Turns out we also have a lower bound $\Omega(1/\Phi)$ on mixing, but not proven here.

Let's use conductance to prove fast mixing.



$n \times n$ grid

$p_{ij} = \frac{1}{4}$ if i and j are adjacent

$f_i = \frac{1}{n^2}$ is the stationary distribution.

Consider any subset S, $|S| \le \frac{n^2}{2}$

If $|S| \ge \frac{n^2}{4}$, then at least n edges leave S.

(square $\frac{n}{2} \times \frac{n}{2}$ or entire columns + partial)

Thus, $$\Phi(S) = \sum_{i\in S} \frac{\pi_i}{\pi(S)} \sum_{j\in \bar{S}} p_{ij} \ge \frac{1}{|S|} \sum_{\substack{i\in S \\ j\in \bar{S}}} p_{ij} = \frac{n}{4|S|} = \Omega\left(\frac{1}{n}\right)$$

If $|S| < \frac{n^2}{4}$, then S is a square at a corner of the grid.

$\Rightarrow$ at least $2\sqrt{|S|}$ points adjacent to S.

Thus $$\Phi(S) = \sum_{i\in S} \frac{\pi_i}{\pi(S)} \sum_{j\in \bar{S}} p_{ij} \ge \frac{1}{|S|} \sum_{\substack{i\in S \\ j\in \bar{S}}} p_{ij} = \frac{2\sqrt{|S|}}{4|S|} = \Omega\left(\frac{1}{\sqrt{|S|}}\right) = \Omega\left(\frac{1}{n}\right)$$

$$\Rightarrow \Phi = \Omega\left(\frac{1}{n}\right) \Rightarrow \varepsilon\text{-mixing time} = O\left(\frac{n^2 \ln n}{\varepsilon^3}\right).$$

Similarly, can show for d-dim, $n \times n \times n \cdots \times n$ grid, $\Phi = \Omega\left(\frac{1}{dn}\right)$, $\pi_i = n^{-d}$

$$\Rightarrow \varepsilon\text{-mixing time} = O\left(\frac{d^3 n^2 \ln n}{\varepsilon^3}\right)$$

poly in n & d, not exponential.

Pagerank

Consider the WWW (world wide web) of webpages. If you are a search engine, you have two tasks: (1) find webpages containing a search term
(2) rank by importance.

Model a websurfer as a random walk on the hyperlinks.
Not strongly connected, so add random restarts to give stationary probabilities.
These stationary probabilities are the Pagerank, and correspond to the frequency with which a page will be randomly visited over a period of time.