Lecture 10 : Markov Chains & Mixing
Expanders & Eigenvalues

Last lecture we saw Markov Chains Monte Carlo.

Define an M.C. M s.t. its stationary distrib π^* is what you want to sample from.

sps. M is irreducible & aperiodic,

then $\pi^{(t)} \to \pi^*$. for any $\pi^{(0)}$ starting distrib.

So qn is: how fast does the chain "mix".

when is $\| \pi^{(t)} - \pi^* \|_{TV} \le \varepsilon$?

We saw: if coupling time is

then that gives UB on time for chain to mix.

(Need to finish proof that coupling time of Glauber dynamics on colorings (with $k \geq 4\Delta+1$ colors) is $O(kn \log(n/\varepsilon))$)

Today: another technique for showing mixing. → eigenvalue based.

- graphs which have good "eigenvalue gap" = expanders.
- another view of expanders (combinatorial)
- using expanders for probability amplification.

random walks.

More recap (from Lecture 6)

If chain is $\underbrace{\text{irreducible + aperiodic}}_{\text{ergodic}}$ then $\pi^{(t)} \to \pi^*$.

Btw: π^* satisfies $\pi^* = \pi^* P$ for transition matrix P for MC

π^* is (left) eigenvector

$(P_{ij} = Pr(X_{t+1} = j \mid X_t = i))$ $\forall ij \in n \times n$.

also $\pi^{(t)} = \pi^{(0)} \cdot P^t$.

so
$$\pi^{(t)} - \pi^* = \pi^{(0)} P^t - \pi^* P^t = (\pi^{(0)} - \pi^*) P^t.$$

So the question is: how does P^t change $(\pi^{(0)} - \pi^*)$?

Need to understand the eigenvalue structure of P.

What are its eigenvalues / eigenvectors? } "spectrum".

—— × ——

Theorem (uses the Perron-Frobenius theorem).

Let P be the transition matrix of an irreducible, aperiodic, reversible ~~matrix~~ markov chain. $(P_{ij} \geq 0)$

then ① all eigenvalues lie in $[-1, 1]$. $1 \geq \lambda_1 \geq \lambda_2 \cdots \geq \lambda_n \geq -1$

② there is a unique eigenvalue / eigenvector with $\lambda_1 = 1$. $\Rightarrow \lambda_2 < 1$. (by irred.)

③ $\lambda_n > -1$ so there is also unique e.v. of absolute value 1. (by aperiodicity)

④ all eigenvalues eigenvectors are real. (by reversibility).

(Time) Reversible Markov Chains

An important class of MCs.

Suppose $\exists \pi \in \mathbb{R}^n_{\geq 0}$ st $\pi_i P_{ij} = \pi_j P_{ji}$ $\forall i j$

then chain is reversible and π (suitably scaled so that $\|\pi\|_1 = \sum_i \pi_i = 1$)

is stationary distribution.

Subsub class:

symmetric chains where $P_{ij} = P_{ji}$ $\forall ij$

clearly time reversible (with $\pi_i = 1/n$ $\forall i$ being the "certificate").

- if chain symmetric then matrix P is also symmetric.

$\Rightarrow$ all eigenvalues real.

- if chain reversible then $\pi_i P_{ij} = \pi_j P_{ji}$ $\Leftrightarrow$ $\pi_i^{1/2} P_{ij} \pi_j^{-1/2} = \pi_j^{-1/2} P_{ji} \pi_i^{1/2}$

$\Rightarrow$ if $D^{1/2} = \text{diag}(\pi_i^{1/2})$ we have that

$D^{1/2} P D^{-1/2}$ is symmetric

technical term

$\Rightarrow$ P is "similar" to a symmetric matrix,

$\Rightarrow$ its eigenvalues are also real.

Thm: if $|\lambda_i| < 1$ then $\pi^{(t)} \Rightarrow \pi^*$ as $t \to \infty$.

Pf: Let's simplify our situation and consider case when P is symmetric.
(Else need to account for the similarity transformation we did above).

$\Rightarrow \pi^* =$ uniform $= \frac{1}{n} \cdot \mathbb{1}$.

Note: not unit vector in ℓ_2 $\Rightarrow$ $\sqrt{n} \cdot \pi^*$ is unit vector.

- P has eigen basis $v_1, v_2, \ldots v_n$. with $\underbrace{1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n > -1}_{\text{by facts above}}$
  - $v_1 = \sqrt{n} \cdot \pi^*$.
  - $v_2, v_3 \ldots v_n$ all orthogonal to v_1.

$\Rightarrow (\pi^0 - \pi^*) P^t = (\langle \pi^0, v_1 \rangle v_1 - v_1/\sqrt{n}) \cdot P^t + \sum_{i \geq 2} \langle \pi^0, v_i \rangle v_i \cdot P^t$

- $\langle \pi^0, v_1 \rangle = \sum_j \pi^0_j \cdot 1/\sqrt{n} = 1/\sqrt{n}$ $\Rightarrow$ first term $= 0$.
- $v_i P^t = \lambda_i^t \cdot v_i$.

say t is even and then since $|\lambda_i| < 1$
we have this term goes to 0 as $t \to \infty$.

$\blacksquare$

btw $\|(\pi^0 - \pi^*) P^t\|_2^2 \to 0$ as $t \to \infty$.

$$\| \quad \sum \underbrace{\langle \pi^0, v_i \rangle^2}_{} \cdot \lambda_i^{2t}$$

depends on how much mass you put in various dirs
but $\leq \sqrt{n}$ since $\|\pi^0\|_1 = 1$
$\Rightarrow \|\pi_0\|_2 \leq \sqrt{n}$.

Using Cauchy Schwarz one can prove a bound on the mixing time. (TV distance) too.

Thm: [see texts]

sps we start at any state x. then after t steps (with "nice" MC having above pp ties)

$$\|\pi^t - \pi^0\|_{TV} \leq O\left(\frac{\lambda_*^{\,t}}{\sqrt{\pi^*(x)}}\right)$$

where $\lambda_* = \max(|\lambda_2|, |\lambda_n|)$.

$\Rightarrow$ to get this down to ε, suffices to walk

$$O\left(\frac{\log 1/\pi(x) + \log 1/\varepsilon}{1-\lambda_*}\right) \text{ steps.}$$

Great: the mixing time depends on the gap between $\lambda_1 = 1$ and the other extreme evs.

For simplicity again, imagine that the chain is lazy

i.e. wp $\frac{1}{2}$ it stays at same place

$\frac{1}{2}$ it makes a move.

Changes transition matrix from P to $\frac{1}{2}(1+P)$

shifts the eigenvalues upwards.

makes $\lambda_n \geq 0$.

But keep general λ_* in mind (not for today, but for later)

$\Rightarrow \lambda_* = \lambda_2$.

typically can care about this case.

$\Rightarrow$ want large $\lambda_1 - \lambda_2$ "spectral gap".

Summary of material so far:

- If MC has large spectral gap ⇒ it mixes fast.
- Special case of MCs. is random walks on graphs.

(for simplicity do random walks on d-regular graphs)

every vertex has degree d.

(for simplicity make walk lazy

wp $\frac{1}{2}$ stay where you are

wp $\frac{1}{2}$ move to one of d randomly nbrs.)

$\Rightarrow P = \frac{1}{2}I + \frac{1}{2} \cdot \frac{1}{d} \cdot A$

Adjacency matrix of graph.

Can talk of the spectrum / eigen values of A. (~~similar~~ analogous theory)

$\boxed{\text{A graph where } P \text{ has large spectral gap (say constant) is called a "spectral expander".}}$

- Another "combinatorial" notion of expander graphs is.

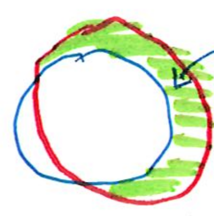
$\forall S$ not too large say $|S| \le n/2$

Look at how S "expands".

- $\dfrac{|N(S) \setminus S|}{|S|} \geq$ large.

"vertex expander".

$N(S) = \{ v : (u,v) \in E \text{ and } u \in S \}$



large compared to $|S|$.

or an "edge expansion" criterion :-

$$\boxed{\forall S \subseteq V \qquad \frac{|E(S,\bar{S})|}{|S|\cdot|\bar{S}|} \geq \phi. \qquad (**)}$$

(note that if (**) $\Rightarrow$ for any $|S| \leq n/2$ have $|\bar{S}| \in [n/2, n]$

for d-regular graph. $\Rightarrow \frac{|N(S)\setminus S|}{|S|} \geq \frac{\phi \cdot n/2}{d}$).

• <u>def</u>: $\phi(G) = \min_{S \subseteq V} \frac{|E(S,\bar{S})|}{|S|\cdot|\bar{S}|}$ as sparsity of G

—— x ——

$$\boxed{\begin{array}{l}\underline{\text{Fact Thm}}: \\ \text{Spectral expansion} \Leftrightarrow \text{combinatorial expansion}\end{array}}$$

(though need to be careful with the quantitative bounds).

—— x ——

<u>Many Questions</u>

① How to prove this equivalence
- Alon / Cheeger Inequality < Noga Alon at IAS (Princeton) / Jeff Cheeger on the 8th floor
- recent extensions ....

② Why do we care about this equivalence?
- two views of same thing
- can use both techniques to prove expansion ($\Rightarrow$ mixing)

③ Why care about expanders in the first place?
- randomness amplification
- expander decompositions (used in algos)
- expander codes (error correction) and many many more apps...

Aside

Constructions of Expanders (constant spectral gap).

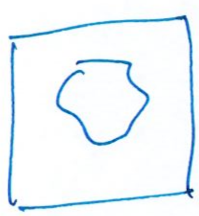
- A random constant-degree graph is an expander whp.
- Take 3 random perfect matchings on n vertices $\Rightarrow$ expander whp.
- Consider $\mathbb{Z}_p$ and for each $x \in \mathbb{Z}_p$, connect $(x, x+1)$, $(x, x-1)$

  $\forall x \in \mathbb{Z}_p^*$, connect $(x, 1/x)$. $\Rightarrow$ expander.

  [Hoory Linial Wigderson]

- Gabber Galil Expander

  Consider the maps $(x, y) \mapsto (x, x \pm y)$, $(x, y) \mapsto (x \pm y, y)$ $\forall (x,y) \in [0,1]$



  Intuition is that any region in $[0,1]^2$ expands.

  To discretize this, consider $[n] \times [n]$, $(x, y) \mapsto (x, x \pm y)$

  $(x, x \pm y + 1)$

  $(x \pm y, x)$

  $(x \pm y + 1, y)$

  [Margulis] definition & first proof

  [Gabber-Galil] This is a constant degree constant spectral gap expander. (via harmonic analysis)

- Newer results (from CS)
  - Zig Zag product. [Reingold Vadhan Wigderson]
  - Random Lifts [Bilu Linial]

    [Marcus Spielman Srivastava]

  combinatorial constructions

- Oh, one older result

  "Optimal" Ramanujan expanders — [Lubotzky Phillips Sarnak]. Number theoretic.

Randomness Amplification

Sps I have an algo that on each input is correct w.p. $99/100$.

$\Rightarrow$ Pr(Algo wrong on input x) $\leq 1/100$ $\forall x$.

Sps. want error prob to be $(1/100)^k$. (Uses r random bits)

① **Easy way**: run algo on ck independent r-bit strings

Take the majority vote.

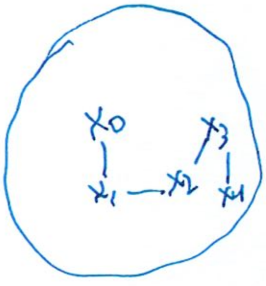
Pr(majority incorrect) $\leq (1/100)^k$ by Chernoff. (for large enough c)

$\boxed{O(rk) \text{ bits}}$

② **More interesting way**:

. Take a random string x_0 of length r, run algo on it.

. take an expander graph of $n = 2^r$ nodes. (constant Δ. degree)



View x_0 as label of some node in G.

Now pick random neighbor of x_0, say x_1.

Use that string as random string for Algo.

repeat $T = c \cdot k$ times: at step i pick random nbr of $x_{i-1} \rightarrow x_i$

Use it as random string for Algo.

. return majority vote of $A(x_0)$ $A(x_1)$ ... $A(x_T)$

$\boxed{\text{Uses } r + O(k \log \Delta) = r + O(k) \text{ random bits instead !}}$

Claim: if G is a good enough expander
(say $|\lambda_*| \le 1/10$) then Prob of error still $\le \left(\frac{1}{100}\right)^k$.

Pf: Suppose incorrect then must make mistake on $\ge T/2$ strings.

will show: P_r (make mistake on $\ge T/2$ strings) $\cdot\, 2^T \le$ tiny.

: Suppose π = some prob distrib.

P = transition matrix of expander random walk.

Z = diagonal matrix with $Z_{ii} = 1$ if i is a bad string
$Z_{ii} = 0$ if i is a good string.

$\Rightarrow$ πZ restricts the non-zeroes to entries that are bad

$\Rightarrow$ prob of error $= \|\pi Z\|_1$.

Main Lemma:

$$\|\pi P Z\|_2 \le \tfrac{1}{5}\|\pi\|_2.$$

Pf: Let π^* be the $1/n$ "1" vector, the stationary distribution.

Same arguments as before, can write π as

$$\underbrace{\pi}_{}= \underbrace{\pi^*}_{x} + \underbrace{\sum_{i \ge 2} w_i v_i}_{y}$$

for other vectors $v_2, v_3 \ldots v_n$ in eigenbasis.

$$\Rightarrow \|\pi P Z\|_2 = \|(x+y)PZ\|_2$$
$$\le \|xPZ\|_2 + \|yPZ\|_2 \quad \text{triangle inequality}$$

(a) $\|xPZ\|_2 = \|xZ\|_2$ since $xP = x$

$\leq \frac{1}{10}\|x\|_2$ since $\sqrt{\frac{I}{100}} = \frac{1}{10}$.

(b) $\|yPZ\|_2 \leq \|yP\|_2$

$\leq \frac{1}{10}\|y\|_2$ since the eigenvalues are all small $\in [-\frac{1}{10}, \frac{1}{10}]$ by assumption.

$\Rightarrow \|\pi PZ\|_2 \leq \frac{1}{10}\|x\| + \frac{1}{10}\|y\|$

$\leq \frac{1}{5}\|\pi\|$. ∎

"either we restrict to Z when π is close to uniform (and hence lose mass)

or we get closer to uniform and hence ℓ_2 norm drops."

Great: Same pf. for $\|\pi P^i Z\| \leq \frac{1}{5}\|\pi\|$ for all $i \geq 1$.

OK:

Look at $\underbrace{\pi^0 P\ PPP\ PP\ P \ldots\ P}_{T \text{ times}}$

Prob that we are incorrect at some $t = T/2$ specific steps is

$\|\pi^0 PZPPPZPPZ \ldots PZ\|_1$.

locations of t "tests" which we failed.

$\leq \sqrt{n} \cdot \|\text{same vector}\|_2 \leq \sqrt{n} \cdot (1/5)^t \cdot \|\pi^0\|_2$ $\frac{1}{\sqrt{n}}$

$= (1/5)^t = (1/5)^{T/2}$

$\Rightarrow Pr(\text{wrong on some } T/2 \text{ steps}) \leq 2^T (1/5)^{T/2} \leq \left(\frac{1}{100}\right)^k$ if $T = ck$ for large c.

Summary:

Used expander graph walk to reduce amount of randomness from $\Theta(r \cdot k)$ to $\Theta(r+k)$.

- Requires construction of explicit expander graphs.
  (But known — see the page giving constructions)
- Requirement that $\max(|\lambda_2|, |\lambda_n|) \le \frac{1}{10}$ strong.

  If not: instead walk some α steps between samples such that $\max(|\lambda_2|, |\lambda_n|)^\alpha \le 1/10$.

Easy!

—— x ——

A different application of expanders.

Error Correcting Codes via Bipartite expanders.

[Gallager]
[Tanner]
[Spielman Sipser]
+ followups.

Want codeword set $C \subseteq \{0,1\}^n$

st. ① C is large (has $2^{\Omega(n)}$ codewords)

② Hamming distance between codewords is large (so that error correction is possible)

③ Efficient encoding & Decoding.

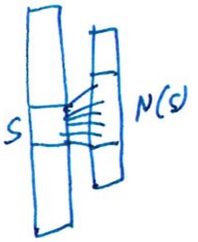
Here's how to use Expanders to get Good codes

Bipartite expanders:

graphs $G = (L, R, E)$ s.t. $|L| = n$

$|R| = m$

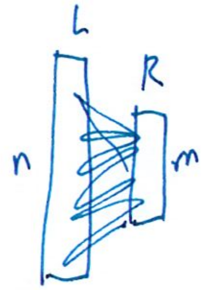
Every left vertex has d neighbors on right



Want it to be (α, β)-expanding

i.e. $\forall S \subseteq L$ st $|S| \leq \alpha n$

the neighborhood $|N(S)| \geq \beta . |S|$.



Claim: $\forall d$ constant,

$\exists$ graphs with $m = \Omega(n)$ so that $|R| = \Omega(|L|)$

and $\beta \geq \frac{3d}{4}$ and $\alpha = \Omega(1)$.

So $|RHS| = \Omega(LHS)$

and all sets of size some constant fraction of LHS

expand by $\geq \frac{3d}{4}$ of their size!

(note: cannot expand by more than $d|S|$).

— X —

Codes:

put $n+m$ bits on the vertices.

Is a valid codeword if

$\forall$ vertex $v \in R$, $\bigoplus_{u \in \partial v} b_u = 0$

(the parity of its neighbors on Left is zero)

Note: its a Linear code

i.e. if $x, y \in C \Rightarrow x+y \in C$
modulo 2.

Fact 1: for a linear code,

min distance = min wt of any codeword.

Thm 1: distance $\geq \alpha n \iff$ min wt of codeword $\geq \alpha n$.

Pf: sps not sps $\exists$ codeword x of wt $\|x\|_1 < \alpha n$

$\Rightarrow$ let S be the positions that x has 1's in

$$|N(S)| \geq \frac{3d}{4}|S|.$$

So avg # of edges from S that such a nbr sees $\leq \dfrac{|S|\,d}{\frac{3d}{4}|S|}$

$$= \frac{4}{3}$$

$\Rightarrow \exists$ nbr that sees only one edge from S.

But this vtx cannot have its nbrhood have parity 0.

$\Rightarrow$ min wt codeword has wt $\geq \alpha n$.

$\Rightarrow$ min distance $= \Omega(n)$. ∎

__Great__ : So the min distance of the code is large ($\geq \alpha n$, i.e. linear fraction of n).

But can we decode?

YES! Very nice algorithm.

Call a node v (right) __unhappy__ if $\sum_{u \in \partial v} b_u \neq 0$.

__Algo (Belief Propagation)__

[Find a vertex v s.t. majority of its neighbors (i.e. $> d/2$ nbrs) are unhappy

Flip it.

———— x ————

__Note__: if $U(x)$ is # of unhappy nbrs if bits on the Left are x.

$\Rightarrow$ BP causes $U(x)$ to decrease (by at least 1)

But

__Q1__: Why is there a valid move when $U(x) \neq 0$.

__Q2__: Even if we can correct some non-codeword y to some y', why is y' the closest codeword to y?

———— x ————

__Lemma1__: if x is not a codeword (so $U(x) \neq 0$) $\Rightarrow \exists$ a valid move (a vertex u on the Left with majority unhappy neighbor nodes)

(See notes)

Lemma 2: if x is at distance $\leq \frac{\alpha n}{4}$ from any codeword y

then can efficiently decode (via Belief Propagation) to y.

Very fast decoding since # unhappy nodes only decreases!

—— X ——

Expander based codes quite useful

- Tornado Codes

- Turbo Codes

—— X ——

In general Expanders very useful in

- theory — useful counterexamples
  - useful as algorithmic constructs (expander decomp)

- practice — ECCs.
  - routing networks

Explicit "random" like graphs!

—— X ——