Advanced Algorithms

Jingcheng Liu

Recap

<u>Previous lecture</u>:
<u>Electrical networks</u>

- Electrical flows
- Effective resistance
- Laplacian system
- Thompson’s principle

<u>What next?</u>
<u>Electrical networks</u>

- Hitting, commute & cover time of random walks

<u>Markov chain Monte Carlo method</u>

- Coupling
- Path coupling

<u>Rapid mixing & random walks on expanders</u>

- Expander graphs
- Expander mixing lemma

Commute Time

Theorem. For any two vertices s and t, $C_{s,t} = 2mR_{\text{eff}}(s,t)$, where $m = |E(G)|$

Proof:

Fix any node t, let $h_{u,t}$ be the hitting time from node u to node t, then $\forall u \neq t$

$$h_{u,t} = 1 + \frac{1}{d_u}\sum_{v\sim u} h_{v,t} \Rightarrow d_u h_{u,t} - \sum_{v\sim u} h_{v,t} = d_u$$

Consider the vector $\overrightarrow{h_{*,t}}$, it satisfies:

$$\begin{pmatrix} & D - A & \\ & & \end{pmatrix}\begin{pmatrix} h_{u,t} \\ h_{t,t} \end{pmatrix} = \begin{pmatrix} d_u \\ d_t - 2m \end{pmatrix}$$

Note that we have artificially added one row of equation on $h_{t,t}$

To ensure there is a solution, we have to make sure that the right hand side sum up to 0

(To be cont'd..)

Commute Time

Theorem. For any two vertices s and t, $C_{s,t} = 2mR_{\text{eff}}(s,t)$, where $m = |E(G)|$

Proof (cont'd):

Fix any node s, let $h_{u,s}$ be the hitting time from node u to node s, then $\forall u \neq s$

$$h_{u,s} = 1 + \frac{1}{d_u} \sum_{v \sim u} h_{v,s} \Rightarrow d_u h_{u,s} - \sum_{v \sim u} h_{v,s} = d_u$$

Consider the vector $\overrightarrow{h_{*,s}}$, it satisfies:

$$\begin{pmatrix} & & \\ & D - A & \\ & & \end{pmatrix} \begin{pmatrix} h_{s,s} \\ h_{u,s} \\ \\ h_{t,s} \end{pmatrix} = \begin{pmatrix} d_s - 2m \\ d_u \\ \\ d_t \end{pmatrix}$$

Again, we have artificially added one row of equation on $h_{s,s}$

(To be cont'd..)

Commute Time

Theorem. For any two vertices s and t, $C_{s,t} = 2mR_{\text{eff}}(s,t)$, where $m = |E(G)|$

Proof (cont'd):

$$L\left(\overrightarrow{h_{*,t}} - \overrightarrow{h_{*,s}}\right) = \begin{pmatrix} d_s \\ d_u \\ \vdots \\ d_t - 2m \end{pmatrix} - \begin{pmatrix} d_s - 2m \\ d_u \\ \vdots \\ d_t \end{pmatrix} = \begin{pmatrix} 2m \\ 0 \\ \vdots \\ -2m \end{pmatrix}$$

Thus, $\frac{L\left(\overrightarrow{h_{*,t}} - \overrightarrow{h_{*,s}}\right)}{2m} = b_{s,t}$

Recall that $L\phi = b_{st}$ has a solution that is unique up to translation

Let $\phi = \frac{\overrightarrow{h_{*,t}} - \overrightarrow{h_{*,s}}}{2m}$, we have

$$R_{\text{eff}}(s,t) = \phi(s) - \phi(t) = \frac{h_{s,t} - h_{s,s}}{2m} - \frac{h_{t,t} - h_{t,s}}{2m} = \frac{h_{s,t} + h_{t,s}}{2m} = \frac{C_{s,t}}{2m}$$

Cover Time

Corollary. $C_{u,v} \leq 2m$ for every edge $uv \in E$.

Proof: Notice that $R_{\text{eff}}(u, v) \leq 1$ for every edge $uv \in E$. Then it follows from $C_{u,v} = 2mR_{\text{eff}}(u, v) \leq 2m$

Theorem. The cover time of a connected graph is at most $2m(n-1)$.

Proof: Consider any spanning tree T.

Then the cover time is at most the time to commute along each tree edges of T.

Approximating Cover Time by Resistance Diameter

Theorem. Let $R(G) \coloneqq \max_{u,v} R_{\text{eff}}(u,v)$ be the resistance diameter. Then,

$$m \cdot R(G) \;\leq\; \text{cover}(G) \leq 6em \cdot R(G) \cdot \ln n + n$$

Proof: Firstly,

$$\text{cover}(G) \geq \max\{h_{uv}, h_{vu}\} \geq \frac{C_{uv}}{2} = mR_{uv}$$

For the upperbound, notice that the maximum (expected) commute time from any vertex is at most $2mR(G)$

If the random walk is run for $2em \cdot R(G)$, by Markov's inequality, the probability that a vertex is not visited is at most $1/e$

If we repeat this $3 \ln n$ times, the probability that a vertex is not visited is at most $1/n^3$

By a union bound, the probability that there exists a vertex not visited is at most $1/n^2$

In such cases, we can pay for another pessimistic cover time of n^3

Combined, we have $\text{cover}(G) \leq 6em \cdot R(G) \cdot \ln n + \frac{1}{n^2} n^3$

Ding, Lee and Peres showed a constant factor approximation of cover time, based on an effective resistance embedding of discrete Gaussian free field

Graph Connectivity

Theorem. There is an $O(n^3)$ time algorithm to solve s-t connectivity using only $O(\log n)$ space

Using random walk, the space requirement is $O(\log n)$ and expected running time is $O(|V||E|) = O(n^3)$

You may wonder, is randomness necessary for checking graph connectivity in log-space?

Definition. A sequence σ is (d, n)-universal if for every labeled connected d-regular graphs and every starting vertex s, the walk defined by σ started from s covers every vertices

Theorem. There exists (d, n)-universal sequence of length $O(n^3 d^2 \log nd)$ for undirected graphs

HINT: Cover time is at most $O(n^2 d)$ for d-regular graphs

Reingold's Theorem For undirected graphs, one can explicitly construct such a universal sequence in log-space

It is an open problem to derandomize log-space connectivity
Though likely not through "directed" universal sequences

Algorithms from random walk (so far)

Finding certain objects faster

- Hitting time / return time
- Ex: Finding bipartite matching, algorithmic Lovász local lemma, 2-SAT, random 3-SAT…

Exploring graphs in space bounded computations

- Cover time
- Ex: checking undirected s-t connectivity, cat and mouse game
- Time-space trade-off

Rapid mixing of random walks: Markov chain Monte Carlo method

- Mixing time
- Ex: Card shuffling, sampling random combinatorial objects, approximate counting
- Exponentially large graph, yet mixes in polynomial time $\approx O(\log N)$ where N is the size of the graph

Theorem. For any finite, irreducible, aperiodic Markov chain, there is a unique $\vec{\pi}$, and $p_t \to \vec{\pi}$ as $t \to \infty$.

Recap: Mixing Time

From the fundamental theorem of Markov chain, we know that $p_t \to \vec{\pi}$ as $t \to \infty$ regardless of p_0

We would like to understand how fast it converges to $\vec{\pi}$

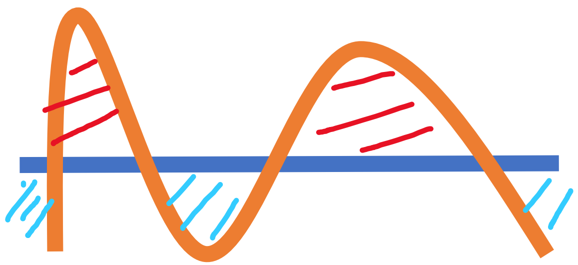
Recall how we measure closeness: $d_{TV}(p_t, \vec{\pi}) = \frac{1}{2}\|p_t - \vec{\pi}\|_1 = \frac{1}{2}\sum_{i=1}^{n}|p_t(i) - \pi(i)|$

Definition. The **ϵ-mixing time** of the random walk is defined as the smallest t such that

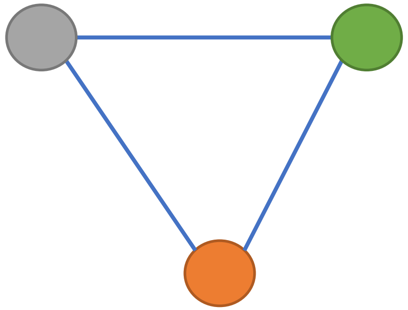
$$\|p_t - \vec{\pi}\|_1 \leq \epsilon \quad \forall\, p_0.$$

An observation: For any distributions p and q over $[n]$, let $p(S) = \sum_{i \in S} p(i)\,, q(S) = \sum_{i \in S} q(i)$, then

$$d_{TV}(p, q) = \frac{1}{2}\sum_{i=1}^{n}|p(i) - q(i)| = \max_{S \subseteq [n]}|p(S) - q(S)|$$



Recap: Graph coloring



Given an undirected graph with max. degree Δ and k colors
Goal: generate a k-coloring uniformly at random

This is presumably harder than deciding if there is a k-coloring
Nevertheless, the following random walk has a stationary distribution uniform over all k-colorings:

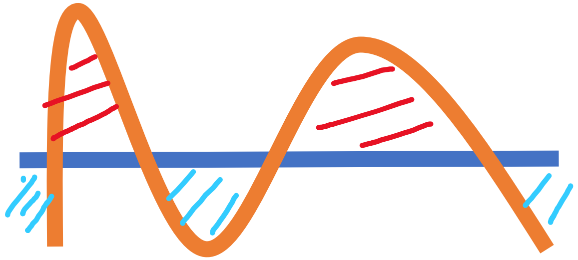
- Start with any k-coloring σ
- Pick a vertex v and a color c uniformly at random, recolor v with c if it is legal; otherwise do nothing;

This Markov chain is irreducible provided that $k \geq \Delta + 2$, and aperiodic

We prove rapid mixing assuming $k \geq 4\Delta + 1$, based on a coupling argument, and explain ideas for $k \geq 2\Delta + 1$
State of the art: $k \geq (\frac{11}{6} - \epsilon)\Delta$ for a small ϵ, or $k \geq \Delta + 3$ for sufficiently large girth graphs

This is known as the Metropolis chain
Other chains: Glauber dynamics, Wang–Swendsen–Kotecký chain, …

Coupling of two distributions



Given distributions p and q over $[n]$, a **coupling** between them is a joint distribution μ over $[n] \times [n]$ such that the marginals are p and q, respectively:

$$\sum_{j \in [n]} \mu(i,j) = p(i)$$

$$\sum_{i \in [n]} \mu(i,j) = q(j)$$

Independently joining p and q is obviously a coupling. More interesting are when they are not independent.

Theorem

For any distributions p and q, and any coupling μ between them, $d_{TV}(p,q) \le \Pr_{(X,Y)\sim\mu}[X \neq Y]$

Furthermore, there is a coupling μ such that $d_{TV}(p,q) = \Pr_{(X,Y)\sim\mu}[X \neq Y]$

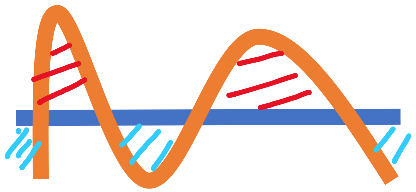
Intuitively, the best we can do is to make the random variables equal in the overlapping regions, that is, $\min\{p_i, q_i\}$; then with the remaining probability, they must be unequal.

Note that the region in red, and the region in light blue have the same area.

Coupling vs Indistinguishing game

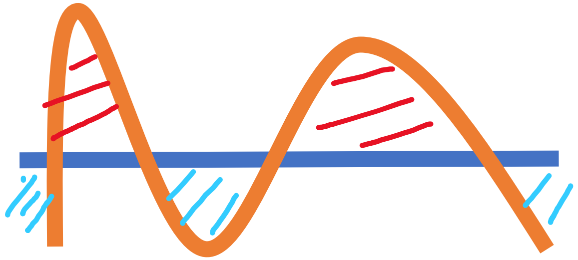
TV distance is also known as statistical distance

- A game to distinguish two distributions p and q over $[n]$
- Player A draw a sample $X \sim p$ and a sample $Y \sim q$
- Player A flips a fair coin to decide which sample to send to Player B
- Player B now needs to guess which distribution does it came from



Maximum likelihood!
Chance of error = TV distance!

Coupling of two random walks



Let (X_t) and (Y_t) be two copies of a Markov chain over $[n]$. A coupling between them is a joint **process** (X_t, Y_t) over $[n] \times [n]$ such that

1. Marginally, viewed in isolation, (X_t) and (Y_t) are both copies of the original chain
2. $X_t = Y_t \Rightarrow X_{t+1} = Y_{t+1}$

Basically, one can think of two random walkers on the same graph G

In isolation, they each behave faithfully as a random walk on G

But their moves could be dependent

The coupling technique is to design a joint moving process, such that

- The two random walkers meet quickly
- Once they meet, they make identical moves thereafter

Then by the coupling theorem, we know that the time they meet will roughly be an upperbound of mixing time

Random walk on the hypercube

- Start with $\sigma \in \{0,1\}^n$
- Pick a coordinate $i \in [n]$ u.a.r., and $b \in \{0, 1\}$ u.a.r.
- Update $\sigma_i = b$

To analyze its mixing time, we consider the following coupling
Say we have two arbitrary copies of the Markov chain, (X_t) and (Y_t)
At each step, we let them choose the same coordinate i and same b

Then, the time that they perfectly couple together is exactly the coupon collecting time!
Note that the probability of not collecting the i-th coupon after r rounds is at most $\left(1 - \frac{1}{n}\right)^r$
By a union bound, the probability of not collecting all the coupons after $n \ln \frac{n}{\epsilon}$ rounds is at most ϵ
So, the ϵ-mixing time for a random walk on the hypercube is $n \ln \frac{n}{\epsilon}$

Coupling for Graph Coloring

- Start with any k-coloring σ
- Pick a vertex v and a color c uniformly at random, recolor v with c if it is legal; otherwise do nothing

Say we have two arbitrary copies of the Markov chain, (X_t) and (Y_t)

At each step, we let them choose the same vertex v and same color c

Let $d_t = \text{number of vertices } X_t \text{ disagree with } Y_t$

Unlike the previous example, d_t can increase now

We need to consider Good Moves that decrease d_t, and balance them with Bad Moves that increase d_t

Start with any k-coloring σ
Pick a vertex v and a color c u.a.r.,
recolor v with c if legal

Coupling for Graph Coloring

Say we have two arbitrary copies of the Markov chain, (X_t) and (Y_t)

At each step, we let them choose the same vertex v and same color c

Let $d_t = \text{number of vertices } X_t \text{ disagree with } Y_t$

Good Moves that decrease d_t:

If we chose a disagreeing vertex v, and color c does not appear in the neighborhood of v in X_t or Y_t, this is a good move

Because we can safely recolor a disagreeing vertex v with color c, and they agree from then on

Let g_t be the number of good moves (among all possible kn choices)

There are d_t vertices to choose from, and each disagreeing vertex has a neighborhood of at most Δ colors in either process, so each disagreeing vertex has $k - 2\Delta$ “safe colors”

$$g_t \geq d_t(k - 2\Delta)$$

Start with any k-coloring σ
Pick a vertex v and a color c u.a.r.,
recolor v with c if legal

Coupling for Graph Coloring

Say we have two arbitrary copies of the Markov chain, (X_t) and (Y_t)

At each step, we let them choose the same vertex v and same color c

Let $d_t = \text{number of vertices } X_t \text{ disagree with } Y_t$

Bad Moves that increase d_t: a legal move in one process but not the other

This happens when (and only when) the chosen color c is already the color of some neighbor of v in one process but not the other

In other words, v must be a neighbor of some disagreeing vertex u, and c must be the color of u in either X_t or Y_t

Let b_t be the number of bad moves (among all possible kn choices)

There are d_t choices of disagreeing vertex u, then Δ choices for v, then 2 choices for

$$b_t \leq 2\Delta d_t$$

Start with any k-coloring σ
Pick a vertex v and a color c u.a.r.,
recolor v with c if legal

Coupling for Graph Coloring

Say we have two arbitrary copies of the Markov chain, (X_t) and (Y_t)
At each step, we let them choose the same vertex v and same color c
Let $d_t = \text{number of vertices } X_t \text{ disagree with } Y_t$

Combined: $\mathbb{E}[d_{t+1}|d_t] = d_t + \frac{b_t - g_t}{kn} \leq d_t + d_t \frac{4\Delta - k}{kn} \leq d_t\left(1 - \frac{1}{kn}\right)$

Since $d_0 \leq n$, we have $\mathbb{E}[d_t|d_0] \leq 1/e$ for $t = 2k\, n \ln n$. Thus,

$$\Pr[d_t > 0|X_0, Y_0] = \Pr[d_t \geq 1|X_0, Y_0] \leq \mathbb{E}[d_t|d_0] \leq 1/e$$

This concludes that the ϵ-mixing time is $\mathrm{O}\left(nk \log \frac{n}{\epsilon}\right)$

To improve this to $k \geq 2\Delta + 1$, one tries to pair bad moves in (X_t) but blocked in (Y_t), with bad moves in (Y_t) but blocked in (X_t)

Expander Graphs

- Combinatorial: graphs with good expansion
- Probabilistic: graphs in which random walks mix rapidly
- Algebraic: graphs with large spectral gap

Let G be a d-regular graph, and let $d = \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n \geq -d$ be the spectrum of its adjacency matrix.

We will be interested in the **spectral radius**, given by

$$\alpha := \max\{\alpha_2, |\alpha_n|\}$$

If α is much smaller than d, we have good **spectral expansion**.

There are many nice properties associated with expander graphs

Among others, say if we want more than one samples in MCMC, do we have to resample entirely?

Cauchy-Schwarz inequality:

$$\langle u, v\rangle \le \sqrt{\langle u, u\rangle} \cdot \sqrt{\langle v, v\rangle}$$

Expander Mixing lemma

Intuitively, an expander can be seen as an approximation to the complete graph, because edges are distributed evenly

Induced edges: $E(S,T) \coloneqq \{(u,v) : u \in S, v \in T, uv \in E\}$

We also allow non-disjoint S, T, in which case an edge can be counted twice.

Expander Mixing lemma

Let G be a d-regular graph with n vertices. If the spectral radius of G is α, then for every $S \subseteq [n], T \subseteq [n]$, we have

$$\left| E(S,T) - \frac{d|S||T|}{n} \right| \le \alpha\sqrt{|S||T|}.$$

Proof: Note that $E(S,T) = \chi_S^T A \chi_T$. Let $\chi_S = \sum_i a_i v_i, \chi_T = \sum_i b_i v_i$, where $\{v_i\}$ is an orthonormal basis for A, with eigenvalues $\{\alpha_i\}$.

$$E(S,T) = \frac{d|S||T|}{n} + \sum_{i \ge 2} \alpha_i a_i b_i .$$

By Cauchy-Schwarz,

$$\left| E(S,T) - \frac{d|S||T|}{n} \right| \le \alpha \|a\|_2 \|b\|_2 = \alpha \|\chi_S\|_2 \|\chi_T\|_2 = \alpha\sqrt{|S||T|}$$

Expander Mixing lemma

Intuition: Expander mixing lemma tells us that a spectral expander looks like a random graph.

Homework: Let G be a d-regular graph with spectral radius α. Show that the size of the maximum independent set of G is at most $\frac{\alpha n}{d}$.

Use this result to conclude that the chromatic number is at least $\frac{d}{\alpha}$.

Converse to Expander Mixing lemma

(By Bilu and Linial)

Suppose that for every $S \subseteq [n], T \subseteq [n]$ with $S \cap T = \emptyset$, we have

$$\left| E(S,T) - \frac{d|S||T|}{n} \right| \le \alpha \sqrt{|S||T|}.$$

Then all but the largest eigenvalue of A in absolute value is at most $O\left(\alpha\left(1 + \log\frac{d}{a}\right)\right)$.

- Proof is based on LP duality
- Would be nice to see an analog of Trevisan's Cheeger's rounding proof

Existence of expanders

- Complete graphs are obviously the best expanders in terms of “expansion” (in all three notions of “expansion”)
- What’s interesting is the existence of sparse expanders: e.g. d-regular expanders for constant d

- A random d-regular graph is a (combinatorial) expander with high probability
- However, deterministic and explicit construction of expanders seems to be much harder to come up with

Alon-Boppana Bound

- For d-regular graphs, how small can the spectral radius be?
- Ramanujan graphs: graphs whose spectral radius are at most $2\sqrt{d-1}$

Alon-Boppana Bound

Let G be a d-regular graph with n vertices, and α_2 be the second largest eigenvalue of its adjacency matrix. Then

$$\alpha_2 \geq 2\sqrt{d-1} - \frac{2\sqrt{d-1}-1}{\lfloor \mathrm{diam(G)}/2 \rfloor}$$

Alon-Boppana Bound

<u>An easy lower bound on spectral radius</u>

Let G be a d-regular graph with n vertices, and α be its spectral radius. Then $\alpha \geq \sqrt{d} \cdot \sqrt{\frac{n-d}{n-1}}$.

Proof: Consider $\mathrm{Tr}(A^2)$. Counting length-2 walks we have

$$\mathrm{Tr}(A^2) \geq nd$$

On the other hand, $\mathrm{Tr}(A^2) = \sum_i \alpha_i^2 \leq d^2 + (n-1)\alpha^2$.

Combined, we have $\alpha \geq \sqrt{d} \cdot \sqrt{\frac{n-d}{n-1}}$.

For the Alon-Boppana bound, one may consider $\mathrm{Tr}\left(A^{2k}\right)$.

Trace method/trick

Random walks in expanders

- We knew that it mixes rapidly, in time $O\left(\frac{\log n}{1-\epsilon}\right)$ for $\alpha = \epsilon d$.
- Perhaps surprisingly, not just the final vertex is close to the uniform distribution, but the entire sequence of walks looks like a sequence of independent samples for many applications.
- In fact, expander random walks can fool many test functions:

  ***Expander random walks: a Fourier-analytic approach*, by Cohen, Peri and Ta-Shma**

Hitting property of expander walks

Let G be a d-regular graph with n vertices, $\alpha = \epsilon d$ be its spectral radius and B be a set of size at most βn.

Then, starting from a uniformly random vertex, the probability that a t-step random walk has never escaped from B, denoted by $P(B, t)$, is at most $(\beta + \epsilon)^t$.

$$\Pr[X_0 \in B, X_1 \in B, X_2 \in B, \ldots, X_t \in B]$$

Remarks before a proof:

- Compare this to a sequence of independent samples.
- Expander mixing lemma is like $t = 2$: Note that $\varphi(S) = \Pr(X_2 \notin S \mid X_1 \sim \pi_S)$
- Bound can be strengthened → see Chapter 4 of ***Pseudorandomness*, by Vadhan**
- Applications to error reduction for randomized algorithms
  - Instead of using kt bits of randomness, only need $k + O(t \log d)$
  - for one-sided error, escaping the bad set of "random bits"
  - for two-sided error, a Chernoff type bound can also be shown → then take the majority of the answers

Hitting property of expander walks

$$\Pi_B = \begin{array}{cc} & \begin{array}{cc} B & V\backslash B \end{array} \\ \begin{array}{c} B \\ V\backslash B \end{array} & \begin{bmatrix} I_B & 0 \\ 0 & 0 \end{bmatrix} \end{array}$$

$$\Pi_B \Pi_B = \Pi_B$$

$W = \frac{1}{d} A$

$u = \frac{1}{n} \vec{1}$

Proof. Observe that $P(B,t) = \|(\Pi_B W)^t \Pi_B u\|_1$

To see this, notice that $\Pr[X_0 \in B] = \|\Pi_B u\|_1$

$$\Pr[X_0 \in B, X_1 \in B] = \|\Pi_B W \Pi_B u\|_1$$

And so on and so forth.

Suppose that we can show $\forall f: f$ is a probability distribution, we have

$$\|\Pi_B W \Pi_B f\|_2 \leq (\beta + \epsilon)\|f\|_2$$

Then,

$$\begin{aligned} \|(\Pi_B W)^t \Pi_B u\|_1 &\leq \sqrt{n}\|(\Pi_B W)^t \Pi_B u\|_2 \\ &= \sqrt{n}\|(\Pi_B W \Pi_B)^t u\|_2 \\ &\leq \sqrt{n}(\beta + \epsilon)^t \|u\|_2 \\ &= (\beta + \epsilon)^t \end{aligned}$$

Cauchy-Schwarz inequality:
$\langle u, v\rangle \leq \sqrt{\langle u, u\rangle} \cdot \sqrt{\langle v, v\rangle}$

$$\Pi_B = \begin{array}{cc} & \begin{array}{cc} B & V\backslash B \end{array} \\ \begin{array}{c} B \\ V\backslash B \end{array} & \begin{bmatrix} I_B & 0 \\ 0 & 0 \end{bmatrix} \end{array}$$

$W = \frac{1}{d}A$ has $\lambda_2(W^\top W) = \epsilon^2$

Hitting property of expander walks

Proof (cont'd):It remains to show $\forall f: f$ is a probability distribution,

$$\|\Pi_B W \Pi_B f\|_2 \le (\beta + \epsilon)\|f\|_2$$

Without loss of generality, we can assume f is supported only on B.

$$\|\Pi_B W \Pi_B f\|_2 = \|\Pi_B W f\|_2 = \|\Pi_B W(u+v)\|_2 \le \|\Pi_B u\|_2 + \|\Pi_B W v\|_2$$

$u = \frac{1}{n}\vec{1}$, so $\frac{\langle f,u\rangle}{\langle u,u\rangle}u = u$, then $v \perp \vec{1}$

Next, $\|\Pi_B W v\|_2 \le \|Wv\|_2 \le \epsilon\|v\|_2 \le \epsilon\|f\|_2$.

On the other hand, $\|\Pi_B u\|_2 = \sqrt{\frac{\beta}{n}} \le \beta\|f\|_2$,

where last inequality follows from Cauchy-Schwarz:

$$1 = \|f\|_1 = \langle 1_B, f\rangle \le \sqrt{\beta n}\|f\|_2$$

Combined together, we have $\|\Pi_B W \Pi_B f\|_2 \le (\beta+\epsilon)\|f\|_2$ as desired.

Cauchy-Schwarz inequality:
$\langle u,v\rangle \le \sqrt{\langle u,u\rangle}\cdot\sqrt{\langle v,v\rangle}$

Hitting property of expander

To get a tail bound, consider $P(S,t) = \left\|\Pi_{Z_t} W \Pi_{Z_{t-1}} W \ldots \Pi_{Z_1} u\right\|_1$
where $S = (Z_t, Z_{t-1}, \ldots, Z_1)$
indicates whether $Z_i \in \{B, \bar{B}\}$

Proof. Observe that $P(B,t) = \|(\Pi_B W)^t \Pi_B u\|_1$

Suppose that we can show $\forall f$: f is a probability distribution, we have $\|\Pi_B W \Pi_B f\|_2 \le (\beta + \epsilon)\|f\|_2$. Then,

$$\|(\Pi_B W)^t \Pi_B u\|_1 \le \sqrt{n}\|(\Pi_B W)^t \Pi_B u\|_2 = \sqrt{n}\|(\Pi_B W \Pi_B)^t u\|_2 \le \sqrt{n}(\beta + \epsilon)^t \|u\|_2 = (\beta + \epsilon)^t$$

It remains to show $\forall f$: f is a probability distribution,

$$\|\Pi_B W \Pi_B f\|_2 \le (\beta + \epsilon)\|f\|_2$$

Without loss of generality, we can assume f is supported only on B.

$$\|\Pi_B W \Pi_B f\|_2 = \|\Pi_B W f\|_2 = \|\Pi_B W (u + v)\|_2 \le \|\Pi_B u\|_2 + \|\Pi_B W v\|_2$$

Next, $\|\Pi_B W v\|_2 \le \|W v\|_2 \le \epsilon\|v\|_2 \le \epsilon\|f\|_2$.

On the other hand, $\|\Pi_B u\|_2 = \sqrt{\frac{\beta}{n}} \le \beta\|f\|_2$,

The last inequality follows from Cauchy-Schwarz:

$$1 = \|f\|_1 = \langle 1_B, f\rangle \le \sqrt{\beta n}\|f\|_2$$

Combined together, we have $\|\Pi_B W \Pi_B f\|_2 \le (\beta + \epsilon)\|f\|_2$ as desired.