

Lecture 20

Lecturer: Rong Ge

Scribe: Xixiang Chen

1 Overview

In this lecture, we talk about mixing time in Markov chains. In particular, we want to know how fast can mixing happen. We make certain assumptions on the Markov chains we work with in this lecture and define fast mixing, and illustrate graphs with either fast mixing or slow mixing. We then introduce the concept of expansion of a graph and show the relationship between expander graph and fast mixing.

2 Mixing

2.1 Definition

Recall that from previous lecture we talked about "good" Markov chains, and we define the mixing process of a Markov chain as follow:

Definition 1. *The state distribution X_t at time t , as t tends to infinity, converges to a unique stationary distribution π .*

Note that here in order for the distribution to converge to π , we need $t \rightarrow \infty$ which cannot happen in reality as we can only run definite steps on the Markov chain. Then the question is how fast can this mixing happen. More precisely, we want to know how large t need to be to make sure that

$$\|X_t - \pi\| \leq \varepsilon$$

Remark 1. Here we think about the norm as l_2 norm. It may appear counter-intuitive since X_t, π are both probability distributions and we often consider l_1 norm. But here we choose l_2 norm because they are not very different up to certain degree and l_2 norm makes it easier to prove.

2.2 What Does Fast Mean?

In this lecture we only consider a simpler case of Markov chains: undirected, unweighted regular graphs (What we talk about in this lecture can also be generalized to all undirected graphs, not necessarily regular, but not directed graphs). A regular graph is a graph whose nodes (or states in the world of Markov chains) all have the same degree d . Recall that $P_{ij} = \frac{C_{ij}}{\deg(i)}$, and in this case, since $\deg(i) = \deg(j)$, we have $P_{ij} = P_{ji}$ for all edge $(i, j) \in E$.

Then we define "fast mixing" as follow:

Definition 2. *Let X_t be the state distribution at time t , n the number of states, π the unique stationary distribution. Then the Markov chain is fast mixing if for all initial state X_0 , $\|X_t - \pi\| \leq \varepsilon$ happens in $O(\log \frac{n}{\varepsilon})$ time (This bound can be further relaxed to $O(\text{polylog} \frac{n}{\varepsilon})$).*

Remark 2. Note that here $\pi = \vec{1} \frac{1}{n}$, and $\pi_i = \frac{1}{n}$ for all i since all states have the same degree.

2.3 Examples

An example of a fast mixing Markov chain is a complete graph, since it is already uniformly distributed (or converges to uniform distribution very quickly if we also consider self loops on vertices).

An example of a slow mixing Markov chain is a graph with two complete subgraphs connected by just one edge. In this case, even at the vertex v on the connecting edge, there are more edges ($\frac{n}{2}$) on the side of v compared to only 1 edge leading to the other side. So the mixing cannot happen in log time.

Note that a graph with small cut does not necessarily implies slow mixing. What it takes to result in slow mixing is a graph with small cut that has many vertices on both sides of the cut (so that it does not go from one side to the other frequently). An example of this would be a cycle.

3 Expansion

3.1 Definitions

Definition 3. The *edge expansion* of a cut (S, \bar{S}) $h(S)$ is defined as:

$$h(S) = \frac{|E(S, \bar{S})|}{d \cdot \min\{|S|, |\bar{S}|\}}$$

Definition 4. The *sparsity* of a cut (S, \bar{S}) $\phi(S)$ is defined as:

$$\phi(S) = \frac{|E(S, \bar{S})|}{\frac{d}{n} \cdot |S| \cdot |\bar{S}|}$$

Remark 3. Assume $|S| \leq |\bar{S}|$. Since $\frac{1}{2} \leq \frac{|\bar{S}|}{n} \leq 1$, we know that $h(S) \leq \phi(S) \leq 2h(S)$.

Definition 5. The *edge expansion* of a graph G (or Cheeger constant) $h(G)$ is defined as:

$$h(G) = \min_{cut(S, \bar{S})} h(S)$$

Definition 6. Graph G is called an *edge expander* if $h(G) \geq \text{constant } h$.

Remark 4. Edge expansion of each graph in the family of graphs G belongs to is at least h , for an absolute constant h independent of number of vertices n .

3.2 Cheeger's inequality

Claim: edge expander \iff fast mixing.

Proof idea. Let P be the transition matrix. $1 = \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_n \geq -1$ be the eigenvalues of P , and respectively v_1, v_2, \dots, v_n where $Pv_i = \lambda_i v_i$ the eigenvectors of P . We know that from previous lecture $\pi = \pi P$ and $X_t = P^t X_0$. Let $X_0 = \sum_{i=1}^n \alpha_i v_i$, and $X_t = P^t X_0 = \sum_{i=1}^n \alpha_i \lambda_i^t v_i$.

Lemma 1. If $\max\{|\lambda_2|, |\lambda_n|\} \leq 1 - \delta$, then in $t = O(\log \frac{n}{\epsilon})$ time, $\|X_t - \pi\| \leq \epsilon$ for any initial state.

Proof. We know that

$$\pi = \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$$

$$v_1 = \left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right)$$

So

$$\begin{aligned} \|X_t - \pi\|^2 &= \sum_{i=1}^n (\alpha_i \lambda_i^t)^2 \\ &\leq (1 - \delta)^2 \sum_{i=2}^n (\alpha_i \lambda_i^{t-1})^2 \\ &= (1 - \delta)^2 \|X_{t-1} - \pi\|^2 \\ &\leq \dots \end{aligned}$$

And we know that $l_2 \text{norm} \leq l_1 \text{norm}$ and the distance between two probability distributions is ≤ 2 , we have

$$\|X_0 - \pi\| \leq \|X_0 - \pi\|_1 \leq 2$$

which completes the proof.

Claim: If we do a lazy walk, $\lambda_n \geq 0$.

Proof. Because $\lambda_n \geq -1$:

$$P_{\text{lazy}} = \frac{1}{2} \cdot I + \frac{1}{2} \cdot P$$

$$\lambda_{\text{lazy}, n} = \frac{1}{2} + \frac{1}{2} \cdot \lambda_n \geq 0$$

This implies that we only need to worry about λ_2 being close 1.

Cheeger's inequality: $\frac{1-\lambda_2}{2} \leq h(G) \leq \sqrt{2(1-\lambda_2)}$

The proof of Cheeger's inequality will be in next lecture.

4 Summary

In this lecture, we have introduced the definition of fast mixing Markov chains. We also talked about edge expansion of graph and graph expander. We claimed that a graph is an expander if and only if the Markov chain is fast mixing. And we introduce Cheeger's inequality which we will prove in next lecture.