Lecture 4

Lecturer: Nati Linial *Scribe: Yoad Lustig*

This lecture presents some applications of spectral properties of non-negative matrices. We first introduce the notion of Markov chain and its relation to non-negative matrices. We then associate between non-negative matrices in $M_n(\mathbb{R})$ and directed graphs on n vertices. Finally we discuss the relation between expander-graphs and rapidly mixing Markov chains.

1 Markov Chains

Markov chain is a random (or stochastic) process. We first introduce the notion of Markov chain and discuss some basic properties of Markov chains. Afterwards we use Perron's theorem to investigate the limiting distribution of Markov chains.

We deal with discrete random variable X_t with n values $\{1, ..., n\}$. A Markov chain is a sequence of random variables $X_1, ..., X_T$. Intuitively, we think on Markov chain as moving from a state i in time t to a state j in time $t+1$ with probability $P(X_{t+1} = j | X_t = i)$. Thus a Markov chain is completely determined by its initial distribution $\lambda_i = P(X_1 = i)$ and its transition probability $P_{ij} = P(X_{t+1} = j | X_t = i)$, we described it by (λ, P) where $\lambda \in \mathbb{R}^n$ is called *distribution vector* and $P \in \mathbb{R}^{n \times n}$ is called *stochastic matrix*[1].

We investigate the algebraic properties of the Markov process (λ, P). we state three simple properties:

1. Let λ_t be the distribution vector $P(X_t)$. Following the calculation $P(X_{t+1} = j) = \sum_{i=1}^{n} P(X_t = i) P(X_{t+1} = j | X_t = i)$ we induce that $\lambda_{t+1} = \lambda_0 P^t$.
2. The transition matrix P has the positive right eigenvector $\overline{1}$ with eigenvalue 1, and $\rho(P) = 1$ is the spectral radius of the matrix P, since for every $i \in \{1, ...n\}$: $\sum_j P_{ij} = \sum_j P(X_t = j | X_{t-1} = i) = 1$
3. The non-negative left eigenvector π of the matrix P corresponding to the eigenvalue $\rho(P) = 1$ is the stationary distribution of the Markov process, i.e., if $\lambda_t = \pi$ then $\lambda_{t+1} = \pi P = \pi$.

We will use Perron's theorem to show that for every Markov process with initial distribution λ and transition matrix P its limiting distribution $\lim_{m \to \infty} \lambda P^m$ approaches to the stationary distribution π. By Perron's theorem:

$$\lim_{m \to \infty} \left(\frac{P}{\rho(P)} \right)^m = \frac{\mathbf{x}_r \otimes \mathbf{x}_\ell}{< \mathbf{x}_r, \mathbf{x} >} \quad .$$

[1]Some define $P^\top$ as stochastic matrix. In this case the multiplication of stochastic matrix with distribution vector changes sides: $(\lambda P)^\top = P^\top \lambda^\top$

where x_r and x_ℓ are the right and left eigenvalues that correspond to $\rho(A)$. As we just mentioned, $x_r = \overline{1}$ and $\rho(P) = 1$. So,

$$\lim_{m\to\infty} P^m = \overline{1} \cdot x_\ell$$

It is easy to check that $\overline{1} \cdot x_\ell$ is a matrix each row of which us x_ℓ. Consequently, if u is any distribution vector, i.e $\sum_i u_i = 1, u \geq 0$, we get $\lim_{m\to\infty} uP^n = x_\ell$. This means that the process tends to a limit distribution.

For practical purposes, it is important to know the rate of convergence. The eigenvalues of a positive stochastic matrix are $1, \lambda_2, ..., \lambda_n$, where $1 > |\lambda_2| \geq |\lambda_3| \geq ... \geq |\lambda_n|$. The rate of convergence is determined by $|\lambda_2|$. We define the spectral gap to be $1 - |\lambda_2|$. A large spectral gap translates to a fast convergence rate. We will say more about this below.

2 Graphs

A graph $G = (V, E)$ has vertex set V and edge set E.

Definition 1 (regular graphs) *Let d be a natural number, a graph $G = (V, E)$ is* d regular *if every vertex has exactly d incident edges.*

Definition 2 (adjacency matrix) *The* adjacency matrix *associated with graph $G = (V, E)$ is a $A \in M_n(\mathbb{R})$ defined by*

$$A_{ij} = \begin{cases} 1 & (i,j) \in E \\ 0 & (i,j) \notin E \end{cases}$$

Definition 3 (Periodic graphs) *A directed graph $G = (V, E)$ is* Periodic with period k *if there exists a partition of V into k parts $S_1, \ldots, S_k$ such that:*

1. *The parts are disjoint (i.e. if $i \neq j$ then $S_i \cap S_j = \emptyset$).*
2. *The parts cover V (i.e. $V = \bigcup_{i=1}^{k} S_k$).*
3. *All the edges are between consecutive parts (mod k). That is, if $j \neq i + 1 (\text{mod } k)$ then $E \cap (S_i \times S_j) = \emptyset$.*

Definition 4 (bipartite graphs) *A graph $G = (V, E)$ is said to be* bipartite *if there is a decomposition $V = A \dot\cup B$ so that all edges in G connect a vertex in A and a vertex in B.*

Note that an undirected graph is periodic iff it is bipartite (since if there is an edge from S_i to S_{i+1} there is also an edge from S_{i+1} to S_i).

A Stochastic matrix (a non-negative matrix which its rows sum up to 1) can be seen as *transition matrix* of some Markov chain, in which the states are $\{1, \ldots, n\}$ and the probability of transition from state i to state j is given by A_{ij}.

For a graph $G = (V, E)$ and a set of vertices $S \subseteq V$ denote by $\overline{S}$ the complement of S (i.e. $\overline{S} = V \setminus S$). Denote by $e(S, \overline{S})$ the number of edges between S and $\overline{S}$ (i.e. $2|E \cap (S \times \overline{S})|$) and by $e(S)$ the number of edges within S (i.e. $|E \cap S \times S|$). Similarly, $e(\overline{S})$ denotes the number of edges within $\overline{S}$.

2.1 Random walks on graphs

Given a undirected graph $G = (V, E)$ with nonnegative weights w_{ij} to the edges, we construct a Markov chain with n states. The transition probability from state i to state j is defined to be $P_{i,j} = \frac{w_{ij}}{\sum_{k=1}^{n} w_{ik}}$. Let $w_i = \sum_{j=1}^{n} w_{ij1}$ and $w = \sum_{i=1}^{n} w_i$ then $\pi = (w_1, ..., w_n)/n$ is the left eigenvector of P:

$$(\pi P)_j = \sum_{i=1}^{n} \pi_i P_{i,j} = \sum_{i=1}^{n} \frac{w_i}{w} \frac{w_{ij}}{w_i} = \sum_{i=1}^{n} \frac{w_j}{w} \frac{w_{ji}}{w_j} = \frac{w_j}{w} \sum_{i=1}^{n} \frac{w_{ji}}{w_j} = \frac{w_j}{w} = \pi_j$$

Since we view a graph as Markov process we define a random walk on a graph as the running of the Markov process. Intuitively, from every node v the random walk continues to one of its neighbors according to the probability $P_{i,j}$.

we see that in a d-regular graph both the left and the right eigenvectors of ρ are $\overrightarrow{1}$. Therefore, by Perron-Frobenius theorem, the limit distribution is $\frac{1}{n}\overrightarrow{1}$ (namely, the uniform distribution):

Corollary 5 *For any $d > 0$, let G be a d-regular non-bipartite connected graph, and x an arbitrary probability distribution on its vertices. The probability to be in some vertex $v \in V$ after a random walk of length m with starting distribution x is $\frac{1}{n} + o(1)$. (Where $o(1)$ converges to 0 as m goes to ∞.)*

The question about the rate of converges arises naturally. We have seen earlier, that the rate of convergence depends on the second eigenvalue $|\lambda_2|$ and its distance from the first eigenvalue λ_1 (which is 1 for a Markov chain). In the next section we will see a family of graphs in which the random walk converges exponentially fast. That is, after m steps the probability to be in a vertex $v \in V$ is $\frac{1}{n} + e^{-\Omega(m)}$. (Note that the transition matrix P is $\frac{1}{d}A$ where A is the adjacency matrix. Therefore the second eigen value $|\lambda_2|$ of P is $\frac{1}{d}$ times the second eigenvalue of A.)

The difference between the first and second eigenvalues is the *spectral gap.*

3 Expanders

Intuitively, expander graphs (a.k.a. expanders) are graphs without "bottlenecks". That is, for any subset of vertices S and its complement $\overline{S}$, the number of edges between S and $\overline{S}$ is fairly large (when compared to the size of S).

Definition 6 (expander graphs) *Let $\delta \geq 0$, a graph $G = (V, E)$ is δ-*edge expanding *if for every set S of size at most $\frac{|V|}{2}$, it holds that $e(S, \overline{S}) \geq \delta|S|$.*

It turns out that expanders are graphs with large spectral graphs. There are theorems for both directions. First, if the spectral gap is large then the graph is an expander (we will prove such a theorem). Second, if the graph is an expander then the spectral gap is large. This is a harder theorem which we will not prove here.

Theorem 7 *Let $G = (V, E)$ be an undirected, d-regular non-bipartite graph with adjacency matrix A with no loops and no parallel edges. Let $\lambda_1 \geq \ldots, \lambda_n$ be the eigenvalues of A in descending order. Then G is δ-edge expanding for $\delta \geq \frac{d-\lambda_2}{2}$.*

Proof By Rayleigh-Ritz theorem $\lambda_2 = \max\limits_{x \perp \overrightarrow{1}} \frac{x^t A x}{||x||^2}$.

For an arbitrary set $S \subseteq V$ with $|S| \leq \frac{|V|}{2}$ we define a vector x representing S in the following way: $x_i = \begin{cases} |\overline{S}| & i \in S \\ -|S| & i \in \overline{S} \end{cases}$

It is easy to see that $x \perp \overrightarrow{1}$ since the sum of its coordinates is 0. Therefore $\lambda_2 \geq \frac{x^t A x}{||x||^2}$. Next we compute the value of $\frac{x^t A x}{||x||^2}$.

First, $||x||^2 = \sum\limits_{i=1}^{n} x_i^2 = |S||\overline{S}|^2 + |\overline{S}||S|^2 = |S||\overline{S}|(|S| + |\overline{S}|) = |S||\overline{S}|n$.

Next, $x^t A x = \sum_{i,j} x_i x_j A_{ij}$. However, A_{ij} is 0 unless there is an edge between i and j. Therefore, the sum can be split into three cases:

1. First, $(i,j) \in E$ and $i,j \in S$. In this case $x_i x_j = |\overline{S}|^2$ and these terms contribute to the sum a total of $2e(S)|\overline{S}|^2$. (Note that each edge is counted from both ends.)
2. Next, $(i,j) \in E$ and $i,j \in \overline{S}$. In this case $x_i x_j = |S|^2$ and this contributes to the sum a total of $2e(\overline{S})|S|^2$.
3. Finally, $(i,j) \in E$ and one of the indices is in S while the other in $\overline{S}$. In this case $x_i x_j = -|S||\overline{S}|$ and this contributes to the sum a total of $-2e(S,\overline{S})|\overline{S}||S|$.

Thus, $x^t A x = \sum\limits_{i,j} x_i x_j A_{ij} = 2(e(S)|\overline{S}|^2 + e(\overline{S})|S|^2 - e(S,\overline{S})|\overline{S}||S|)$.

Since the graph is d-regular we also know that the number of edges touching a vertex in S is $d|S|$. These edges can be partitioned into edges within S and edges touching a vertex in $\overline{S}$. Therefore $d|S| = 2e(S) + e(S,\overline{S})$ (note that the edges within S are counted twice). Similarly $d|\overline{S}| = 2e(\overline{S}) + e(S,\overline{S})$. Denoting $e(|S|,|\overline{S})$ by n_e we do the bookkeeping and get:

$$\begin{aligned} x^t A x &= 2(e(S)|\overline{S}|^2 + e(\overline{S})|S|^2 - n_e|\overline{S}||S|) \\ &= 2(\frac{1}{2}(d|S| - n_e)|\overline{S}|^2 + \frac{1}{2}(d|\overline{S}| - n_e)|S|^2 - n_e|\overline{S}||S|) \\ &= d|S||\overline{S}|(|\overline{S}| + |S|) - n_e(|\overline{S}|^2 + 2|\overline{S}||S| + |S|^2) \\ &= d|S||\overline{S}|n - n_e n^2 \end{aligned}$$

Therefore $\lambda_2 \geq \frac{x^t A x}{||x||} = d - n_e \frac{n}{|S||\overline{S}|}$. Since $|\overline{S}| \geq \frac{n}{2}$ we get $\lambda_2 \geq d - n_e \frac{2}{|S|}$ i.e. $\frac{n_e}{|S|} = \frac{e(S,\overline{S})}{|S|} \geq \frac{d-\lambda_2}{2}$ as claimed. ■