18.419 Random Walks and Polynomial-Time Algorithms Feb 21, 2002

Lecture 5:Isoperimetric coefficients and mixing times

Lecturer: Santosh Vempala *Scribe: Jan Vondrák*

In this lecture, we discuss the rate of convergence of a random walk and its relation to eigenvalues of the random walk. There are several differenent measures of convergence that we can consider. However, for our purposes it will not make much difference which one we choose.

Measures of convergence

Suppose the random walk starts at state i and after t steps, it is at state j with probability $P_t(j)$. We can measure the distance of this distribution from the stationary distribution $\pi(j)$ in the following ways:

- Pointwise distance: $\max_j |P_t(j) - \pi(j)|$
- χ^2-distance: $(\sum_j \frac{1}{\pi(j)} |P_t(j) - \pi(j)|^2)^{1/2}$
- Variation distance: $\sum_j |P_t(j) - \pi(j)|$

Clearly, the pointwise distance is smaller than the variation distance.

The variation distance is always less than or equal to the χ^2-distance. This can be shown using the Cauchy-Schwarz inequality: $\sum_j a_j b_j \le (\sum_j a_j^2)^{1/2} (\sum_j b_j^2)^{1/2}$. Here, we set

$$a_j = \frac{|P_t(j) - \pi(j)|}{\sqrt{\pi(j)}},$$

$$b_j = \sqrt{\pi(j)}$$

and we get

$$\sum_j |P_t(j) - \pi(j)| \le \left(\sum_j \frac{|P_t(j) - \pi(j)|^2}{\pi(j)} \right)^{1/2} \left(\sum_j \pi(j) \right)^{1/2} = \left(\sum_j \frac{|P_t(j) - \pi(j)|^2}{\pi(j)} \right)^{1/2}.$$

In the following, we will consider the pointwise distance.

Next, we define the mixing rate of a random walk:

$$\mu = \lim_{t \to \infty} \max_{i,j} |P_j(t) - \pi(j)|^{1/t}$$

The mixing time is defined as the time needed to halve the distance $\max_j |P_j(t) - \pi(j)|$. In this case, the mixing time would be $(\log_2 \frac{1}{\mu})^{-1}$.

It seems that mixing time should be related to cover time, but in some cases these two quantities can differ significantly. For a periodic d-dimensional lattice, with n^d vertices, the cover time is obviously at least n^d. However, since the random walk proceeds independently in different coordinates, the mixing time is roughly d times the mixing time for a cycle, i.e. $O(dn^2)$.

Convergence and eigenvalues

Let M be the transition matrix of an irreducible random walk (which means that any state can be reached from any other state). Then, by the Perron-Frobenius theorem, there exists a unique vector π such that

$$M^T \pi = \pi.$$

Moreover, all components of π are positive.

This implies that the stationary distribution is unique. In order to analyze the evolution of a probability distribution under the operation of M, it is convenient to symmetrize the matrix in the following way. Define

$$D = \text{diag}(\sqrt{\pi(1)}, \sqrt{\pi(2)}, \ldots, \sqrt{\pi(n)}),$$

$$Q = DMD^{-1}.$$

It can be verified that Q is symmetric, since

$$D^2 M = M^T D^2 = diag(\pi(1), \pi(2), \ldots, \pi(n)).$$

Moreover, Q and M have the same eigenvalues, and u is a (left) eigenvector of M if and only if $D^{-1}u$ is an eigenvector of Q:

$$u^T D^{-1} Q = u^T (M D^{-1}) = \lambda (u^T D^{-1}).$$

In particular, the eigenvector corresponding to eigenvalue 1 is

$$v_1 = D^{-1}\pi = (\sqrt{\pi(1)}, \sqrt{\pi(2)}, \ldots, \sqrt{\pi(n)}).$$

Since Q is symmetric, we can choose the eigenvectors so that they form an orthonormal basis $[v_1, v_2, \ldots, v_n]$, with corresponding eigenvalues $\lambda_1, \ldots \lambda_n$. Then Q can be decomposed into a sum of matrices of rank 1 like this:

$$Q = \sum_{k=1}^{n} \lambda_k v_k v_k^T.$$

Since the vectors are orthogonal,

$$Q^t = \sum_{k=1}^{n} \lambda_k^t v_k v_k^T,$$

$$M^t = D^{-1} Q^t D = \Pi + \sum_{k=2}^{n} \lambda_k^t D^{-1} v_k v_k^T D$$

where $\Pi_{ij} = \pi(j)$. This means that M^t converges to Π as quickly as the remaining terms with factors λ_i^t tend to zero. Finally, we estimate the mixing rate:

$$|P_t(j) - \pi(j)| = |(e_i M^t)_j - \pi(j)| = \left| \left(\sum_{k=2}^{n} \lambda_k^t v_k v_k^T \sqrt{\frac{d(j)}{d(i)}} \right)_{ij} \right| \leq |\lambda_2|^t \sqrt{\frac{d(j)}{d(i)}}$$

where λ_2 is the eigenvalue of second largest *absolute value.* Therefore the mixing rate is at most $|\lambda_2|$ (and this bound is tight). For non-bipartite and connected graphs, all eigenvalues except λ_1 are smaller than 1 in absolute value, so the random walk always converges to the stationary distribution.

Conductance

Conductance is a parameter defined in a combinatorial way which is closely related to the rate of convergence. Suppose that B is a transition matrix of a time-reversible random walk ($\pi_i b_{ij} = \pi_j b_{ji}$). For a subset of vertices S, we define

$$\Phi(S) = \frac{\sum_{i \in S, j \notin S} \pi_i b_{ij}}{\min(\pi(S), \pi(\overline{S}))}.$$

The conductance is then

$$\Phi = \min_{0 \neq S \subset V} \Phi(S).$$

Sometimes, $\Phi(S)$ is defined with $\pi(S)\pi(\overline{S})$ in the denominator. This can differ from our definition by a factor of at most 2. This definition would represent the ratio between the probability of crossing the cut between S and $\overline{S}$ when we are in the stationary distribution and make one random step, as opposed to the probability of crossing the cut when we choose two vertices randomly from the stationary distribution.

Next, we derive a relation between Φ and λ_2. (Here, λ_2 denotes the second largest positive eigenvalue.) If $u = D^{-1}\pi$ is the eigenvector for $\lambda_1 = 1$, we have

$$\lambda_2 = \max_{u^T x = 0} \frac{x^T Q x}{x^T x} = \max_{\pi^T y = 0} \frac{y^T D^2 B y}{y^T D^2 y}$$

(by substitution $y = D^{-1}x$).

$$1 - \lambda_2 = \max_{\pi^T y = 0} \frac{y^T D^2 (I - B) y}{y^T D^2 y}$$

$$y^T D^2 (I - B) y = \sum_{i,j} y_i \pi_i (\delta_{ij} - b_{ij}) y_j = \sum_i y_i^2 \pi_i (1 - b_{ii}) - \sum_{i \neq j} y_i y_j \pi_i b_{ij}$$

$$= \sum_i y_i^2 \pi_i \sum_{j \neq i} b_{ij} - \sum_{i \neq j} y_i y_j \pi_i b_{ij} = \frac{1}{2} \sum_{i \neq j} \pi_i b_{ij} (y_i - y_j)^2,$$

and

$$y^T D^2 y = \sum_i \pi_i y_i^2.$$

Therefore

$$1 - \lambda_2 = \min_{\pi^T y = 0} \frac{\sum_{i \neq j} \pi_i b_{ij} (y_i - y_j)^2}{2 \sum_i \pi_i y_i^2}.$$

Now suppose $S \subset V$ is such that

$$\Phi = \frac{\sum_{i \in S, j \notin S} \pi_i b_{ij}}{\min(\pi(S), \pi(\overline{S}))}.$$

We define a vector y orthogonal to π as

- $y_i = \pi(\overline{S})$ for $i \in S$
- $y_i = -\pi(S)$ for $i \in \overline{S}$

Then for $i \in S, j \notin S$, $(y_i - y_j)^2 = (\pi(S) + \pi(\overline{S}))^2 = 1$ and otherwise $(y_i - y_j)^2 = 0$. In the denominator, we get $\sum_i \pi_i y_i^2 = \pi(S)\pi(\overline{S})^2 + \pi(\overline{S})^2\pi(S) = \pi(S)\pi(\overline{S})$. Thus:

$$1 - \lambda_2 \leq \frac{\sum_{i \neq j} \pi_i b_{ij} (y_i - y_j)^2}{2 \sum_i \pi_i y_i^2} = \frac{\sum_{i \in S, j \notin S} \pi_i b_{ij}}{\pi(S)\pi(\overline{S})} \leq 2\ \Phi.$$

This is one part of a theorem which will be proved in the next lecture.

Theorem:

$$\frac{1}{2}\Phi^2 \leq 1 - \lambda_2 \leq 2\ \Phi.$$