The Eigenvalue Gap and Mixing Time

Nathan McNew

March 14, 2011

Abstract

The eigenvalue or spectral gap of a Markov chain is the difference between the two largest eigenvalues of the transition matrix of its underlying (state space) graph. In this paper we explore the intimate relationship between the spectral gap of a Markov chain and its mixing time, as well as another closely related structural property of a Markov chain known as conductance. The relationships among these properties can be used to put bounds on a chain's mixing time, and can be used to prove both rapid and slow mixing.

As the spectral gap and conductance of a Markov chain are often difficult to calculate, an additional tool, canonical paths, is introduced which can be used to put a lower bound on the spectral gap. Several theorems relating these properties to mixing time as well as an example of using these techniques to prove rapid mixing are given.

1 Introduction

Given any Markov chain, we can represent it as a random walk on some weighted directed graph G. Let P be the transition matrix for the directed graph associated to an ergodic Markov chain, with entries P_{ij} corresponding to the probability of transitioning from state i to state j in one iteration of the Markov chain. Let $\{\lambda_i\}$ be the eigenvalues of P, with $|\lambda_1| \geq |\lambda_2| \geq \cdots |\lambda_n|$. Since the Markov chain is ergodic, we know that the system has a stationary distribution π, and thus has an eigenvalue of 1 (corresponding to the eigenvector π.) By Perron Frobenius theory for nonnegative matrices [5], we can conclude that $\lambda_1 = 1$, and that $|\lambda_i| < 1$ for all $2 \leq i \leq n$. Further, if the Markov chain is reversible then we can conclude that all of its eigenvalues are real.

A Markov chain is *lazy* if the probability of staying in any state is at least 1/2. (And any Markov chain can be made lazy by adding self loops of weight one half to each state). If P is a reversible, lazy Markov chain, then all of its eigenvalues will be positive. This can be shown by writing $P = \frac{1}{2}(Q + I)$ for some transition matrix Q. Since Q also describes a valid reversible Markov chain, its eigenvalues $\mu_i = 2\lambda_i + 1$ must satisfy the Perron Frobenius condition that $\mu_i \geq -1$ and thus each $\lambda_i \geq 0$. As such, since any Markov chain can be made lazy without substantially increasing its mixing time, the negative eigenvalues of a Markov chain are never a significant concern. So we will only consider lazy Markov chains for the remainder of the paper.

Then, if the Markov chain is lazy, we can define its *eigenvalue* or *spectral gap*, the difference of its two largest eigenvalues, to be $\lambda_1 - \lambda_2 = 1 - \lambda_1$. (In greater generality, this would be $1 - \max\{|\lambda_2|, |\lambda_n|\}$.) The eigenvalue gap is useful because it can be used to put a bound on the mixing time for the Markov chain.

Specifically, let Ω be the state space of the Markov chain. Since the chain is ergodic, we know that the chain will converge to the stationary distribution π, and thus for every $i, j \in \Omega$, $\lim_{t\to\infty}(P^t)_{ij} = \pi_j$. We thus define the variation distance of a state i at time t from the stationary distribution to be

$$\Delta_i(t) = \frac{1}{2}\sum_{j\in\Omega}(P^t)_{ij}$$

and the mixing time to be

$$\tau(\epsilon) = \max_{i\in Omega} \min\{t \mid (\forall t' > t)\Delta_i(t') \leq \epsilon\}$$

Theorem 1.1. *For an ergodic, reversible, Markov chain and $\epsilon > 0$,*

$$\frac{1}{2\ln(2\epsilon)}\frac{\lambda_2}{1-\lambda_2} \leq \tau(\epsilon) \leq \frac{1}{1-\lambda_2}\ln\left(\frac{1}{\pi_*\epsilon}\right)$$

Where $\pi_* = \min_{j\in\Omega}\pi_j$.

Thus, if we can calculate or bound the size of the eigenvalue gap of our Markov chain we can determine the chains mixing time, and can determine whether our chain exhibits rapid (or slow) mixing depending on how the eigenvalue gap changes as we scale the size of our state space. The most commonly used method for bounding the eigenvalue gap of a Markov chain is via a property of the Markov chain known as conductance.

2 Conductance

The main idea of conductance is to formalize the idea that chains with fewer "bottlenecks" will mix faster. If a chain has a significant region in its state space graph that is difficult to enter or leave, then the Markov chain will necessarily take longer to reach the stationary distribution. We first define the *ergodic flow* between two subsets $S, T \subset \Omega$ to be

$$Q(S,T) = \sum_{i \in S, j \in T} \pi(i) P_{ij}$$

and the quantity

$$\Phi_S = \frac{Q(S, S^c)}{\pi(S)}$$

where $\pi(S) = \sum_{i \in S} \pi_i$. The quantity Φ_S can be thought of as the conditional probability that a Markov chain in the stationary distribution crosses the cut from S to S^c in a single iteration, given that it starts in S. [6] Finally, we can define the *conductance* (also known as the Cheeger constant) of the chain:

$$\Phi = \min_{\substack{S \subset \Omega \\ \pi(S) \leq \frac{1}{2}}} \Phi_S$$

We would expect the conductance of a Markov chain, which is effectively a measure of how well connected the state space graph is, to be directly correlated with the required mixing time for the chain, and thus to the eigenvalue gap. This relationship is given by an inequality known as the Cheeger inequality:

Theorem 2.1 (Cheeger Innequality)**.** *For a lazy, reversible markov chain, the eigenvalue gap* $1 - \lambda_2$ *satisfies:*

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

The Cheeger inequality was originally a result in Riemannian Manifolds [3] which was modified by Alon [1] and Alon and Milman [2] to a discrete case of the second eigenvalue of the adjacency matrix for simple unweighted graphs. In this case the analogue of the conductance was called the *magnification* of the graph, and was a generalization of the widely studied concept of *expansion* for bipartite graphs.

Combining this result with the previous result relating the eigenvalue gap to the mixing time we can relate the conductance directly to the mixing time as follows: [6]

$$\frac{1-\Phi}{2\Phi} \ln\left(\frac{1}{\epsilon}\right) \leq \tau(\epsilon) \leq \frac{2}{\Phi^2}\left(\ln\frac{1}{\pi_*} + \ln\frac{1}{\epsilon}\right)$$

Corollary 2.2. *A family of ergodic, reversible Markov chain with state space of size n and conductance conductance Φ_n is rapidly mixing if and only if*

$$\Phi_n \geq \frac{1}{p(n)}$$

for some polynomial p.

This result is commonly used to show rapid mixing of Markov chains, and almost all proofs of slow mixing rely on showing that the conductance is exponentially small in the problem size. Mihail [4] later showed that the upper bound holds even if the chain is non-reversible.

3 Canonical Paths

While the eigenvalue gap and conductance of a Markov chain can provide close bounds on the mixing time of the chain, these values are often prohibitively difficult to calculate. Since we are generally interested in proving rapid mixing, we are most interested in calculating a lower bound on the conductance of our graph. Another method, canonical paths and congestion can be useful in this regard, as they are much easier to calculate and can be used to bound the conductance of the graph from below.

For any pair $i, j \in \Omega$ we can define a canonical path $\gamma_{ij} = (i = z_0, z_1 \cdots z_k = j)$ running from from i to j through adjacent states in the state space graph of the Markov chain. Let $\Gamma = \{\gamma_{ij}\}$ be the family of canonical paths running between all pairs of states of the system. The *congestion* of the Markov chain is then defined to be

$$\rho = \rho(\Gamma) = \max_{(u,v)} \left\{ \frac{1}{\pi(u)P_{uv}} \sum_{\substack{i,j\in\Omega \\ \gamma_{ij} uses(u,v)}} \pi(i)\pi(j) \right\}$$

Where the maximum runs over all edges of the state space graph. The $\pi(u)P_{uv}$ term in the definition above can be thought of as being the natural "capacity" of the edge uv, or how much traffic it would normally experience in the stationary state. The sum above then counts the flow or "load" of the edge within this family of canonical paths, and the congestion, ρ is the maximum load of any edge of the state space graph as a fraction of its capacity.

As one might expect, high congestion in a graph corresponds to a lower conductance, as demonstrated by the following theorem. [7]

Theorem 3.1. *For any reversible Markov chain, and any choice of canonical paths*

$$\Phi \geq \frac{1}{2\rho}$$

Proof. Let $S \subset \Omega$ be the subset with $0 < \pi(S) < \frac{1}{2}$ which minimizes the quotient $Q(S, S^c)/\pi(S)$ and thus defines the conductance of the graph. For any choice of paths, the total flow crossing the cut from S to S^c is $\pi(S)\pi(S^c)$, while the total capacity of the cut edges (x, y), with $x \in S$ and $y \in S^c$, is $Q(S, S^c)$. Hence there must exist a cut edge $(u, v), u \in S, v \in S^c$ with

$$\frac{1}{\pi(u)P_{uv}} \sum_{\substack{x,y\in\Omega \\ \gamma_{xy} uses(u,v)}} \pi(x)\pi(y) \geq \frac{\pi(S)\pi(S^c)}{Q(S, S^c)} \geq \frac{\pi(S)}{2Q(S, S^c)} = \frac{1}{2\Phi}$$

□

Note that the above result applies to all possible choices of canonical paths, for example, no requirement was ever made that the shortest path between two states be chosen. In order to prove rapid mixing by this result, however, we would necessarily need to make a choice of canonical paths which does not excessively "overload" any edge relative to its capacity.

4 An example

Consider the lazy random walk on an odd cycle of length n. It can easily be seen that this chain is both ergodic and reversible, with stationary distribution $\pi(i) = \frac{1}{n}$ for every state i. We make the "natural" choice of canonical paths in this chain, namely the shortest route around the circle. Both the Markov chain and the choice of canonical paths is completely symmetric, so we need only count the number of times that each edge is used by one of these canonical paths.

Note that each edge is used once by a path of length 1, twice by paths of length 2 and so on, up to paths of length (n-1)/2. so the total number of times each path is used is

$$\sum_{i=1}^{(n-1)/2} i = \frac{\frac{n-1}{2} \cdot \frac{n+1}{2}}{2} = \frac{(n+1)(n-1)}{8}$$

We can then use this to calculate the congestion in the graph:

$$\rho = \frac{1}{\pi(u)P_{uv}} \sum_{\substack{x,y\in\Omega \\ \gamma_{xy} uses (u,v)}} \pi(x)\pi(y) = \frac{1}{\frac{1}{n}\frac{1}{4}} \left(\frac{(n+1)(n-1)}{8} \right) \frac{1}{n^2} \approx \frac{n}{2}$$

Thus by Theorem 3.1, we see that the conductance of this graph satisfies $\Phi > \frac{1}{n}$, and thus that the eigenvalue gap is bounded: $1 - \lambda_2 \geq \frac{1}{8\frac{n^2}{4}} = \frac{1}{2n^2}$. (The eigenvalue gap can be computed more precisely by other methods, in which case the actual gap is found to be asymptotically $\frac{\pi^2}{n^2}$, so our estimate is off only by a constant factor of $2\pi^2$.) We can then bound the mixing time by

$$\tau(\epsilon) < \frac{1}{1-\lambda_2} \log\left(\frac{1}{\pi_* \epsilon}\right) = 2n^2 \log\left(\frac{n}{\epsilon}\right)$$

thus demonstrating that this Markov chain exhibits rapid mixing.

References

[1] N. Alon (1986), Eigenvalues and Expanders. *Combinatorica* **6**, 83-96.

[2] N. Alon and V.D. Milman (1985), λ_1, isoperimetric inequalities for graphs and superconcentrators. *Journal of Combinatorial Theory Series B.* **38**, 73-88.

[3] J. Cheeger (1971), A lower bound for the smallest eigenvalue of the Laplacian *Problems in Analysis*, Princeton University Press, New Jersey, 195-199

[4] M. Mihail (1989) Conductance and convergence of markov chains-a combinatorial treatment of expanders. *30th Annual Symposium on Foundations of Computer Science*, 526-531.

[5] E. Seneta, (1981) *Non-negative matrices and Markov chains* (2nd ed.),Springer Verlag, New York.

[6] A. Sinclair, *Algorithms for Random Generation and Counting* Birkhauser, Boston.

[7] A. Sinclair (1992), Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Probability and Computing* **1**, 351-370.