Lecture 18: Eigenvalues, expansion, and rapid mixing CSE 525, Spring 2019
Instructor: James R. Lee

1 Eigenvalues, conductance, and flows

Let P be the transition kernel of a reversible, irreducible, aperiodic Markov chain on the state space Ω. Suppose that P has stationary measure π (this exists and is unique by the Fundamental Theorem of Markov Chains). Let us also assume that all the eigenvalues of P lie in $[0,1]$. In the last lecture, we proved that they must lie in $[-1,1]$. Now by replacing P with $P' = \frac{1}{2}I + \frac{1}{2}P$, we can ensure that all eigenvalues are nonnegative while only changing the mixing time by a factor of 2.

Suppose the eigenvalues of P are $1 = \lambda_1 > \lambda_2 \geqslant \cdots \geqslant \lambda_{|\Omega|} \geqslant 0$. In the last lecture, we defined τ_{mix} and showed that

$$\frac{1}{1-\lambda_2} - 1 \leqslant \tau_{\mathrm{mix}} \leqslant O(\log(1/\pi_{\min}))\frac{1}{1-\lambda_2}\,,$$

where $\pi_{\min} := \min\{\pi(x) : x \in \Omega\}$ is the minimum stationary probability. In other words, up to a factor of $O(\log(1/\pi_{\min}))$, the mixing time is controlled by the inverse spectral gap of P.

The Gibbs distribution on matchings. To understand the phrase "rapid mixing," let us consider sampling from a particular measure on an exponentially large state space. Fix an n-vertex graph $G = (V, E)$ and consider the set $\mathcal{M}(G)$ of all matchings in G; these are precisely subsets of the edges E in which every vertex has degree at most one. It is clear that $\mathcal{M}(G)$ can be very large; for instance, in the complete graph on $2n$ vertices, we have $\log|\mathcal{M}(G)| \asymp n \log n$.

For a parameter $\lambda \geqslant 1$, let π_λ denote the measure on $\mathcal{M}(G)$ where a matching m has probability proportional to $\lambda^{|m|}$. Here, $|m|$ denotes the number of edges in m. Thus $\pi_\lambda(m) = \lambda^{|m|}/Z$, where

$$Z = \sum_{m \in \mathcal{M}(G)} \lambda^{|m|}$$

is the corresponding *partition function*, which can itself be very difficult to compute. (In fact, the ability to approximate Z efficiently is essentially equivalent to the ability to sample efficiently from π_λ.)

Our goal is to produce a sample from a distribution that is very close to π_λ. To do this, we will define a Markov chain on $\mathcal{M}(G)$ whose stationary distribution is π_λ. We will then show that $\tau_{\mathrm{mix}} \leqslant n^{O(1)}$, implying that there is a polynomial-time algorithm to sample via simulating the chain for $n^{O(1)}$ steps. In general, for such an exponentially large state space indexed by objects of size n, we say that the chain is "rapidly mixing" if the mixing time is at most $n^{O(1)}$.

1.1 Conductance

For a pair of states $x, y \in \Omega$, define $Q(x,y) = \pi(x)P(x,y)$ and note that since P is reversible, the detailed balance conditions give us $Q(x,y) = Q(y,x)$. For two sets $S, T \subseteq \Omega$, define $Q(S,T) = \sum_{x\in S}\sum_{y\in T} Q(x,y)$. Finally, given a subset $A \subseteq \Omega$, we define its *conductance* as the quantity

$$\Phi(A) = \frac{Q(A,\bar{A})}{\pi(A)}\,.$$

Note that $Q(A,\bar{A})$ represents the "ergodic flow" from A to $\bar{A}$—this is the probability of a transition going between A and $\bar{A}$ at stationarity. This quantity has a straightforward operational interpretation:

It is precisely the probability that one step of the Markov chain leaves A when we start from the stationary measure restricted to A. Note that if $\Phi(A)$ is small, we expect that the chain might get "trapped" inside A, and thus perhaps such a "bottleneck" could be an obstruction to mixing. In fact, we will see momentarily that this is true, and moreover, these are the only obstructions to rapid mixing.

We define the *conductance of the chain P* to capture the conductance of the "worst" set

$$\Phi^* = \max_{\pi(A) \leqslant \frac{1}{2}} \Phi(A)\,.$$

Then we have the following probabilistic version of the discrete Cheeger inequality (proved independently by Jerrum-Sinclair and Lawler-Sokal in the context of Markov chains on discrete spaces).

Theorem 1.1. *It always holds that*

$$\frac{1}{2}(\Phi^*)^2 \leqslant 1 - \lambda_2 \leqslant 2\,\Phi^*\,.$$

This is a basic fact in spectral graph theory; we will not prove it here. Let us mention, though, that the right-hand side is straightforward—it verifies that indeed a low-conductance set is an obstruction to rapid mixing. The left-hand side, which claims that those are the only such obstructions, is more subtle.

The best way to prove the right-hand side is as follows: Recall the inner product

$$\langle u, v\rangle_{\ell_2(\pi)} = \sum_{x\in\Omega} \pi(x) u_x v_x$$

and the associated Euclidean norm $\|v\|_{\ell_2(\pi)} = \sqrt{\langle v, v\rangle_{\ell_2(\pi)}}$. Then using the variational principle for eigenvalues, we have

$$\lambda_2 = \max_{v:\langle v,\mathbf{1}\rangle_{\ell_2(\pi)}=0} \langle v, vP\rangle\,,$$

where $\mathbf{1}$ denotes the all-ones vector. Consider now any $A \subseteq \Omega$ with $\pi(A) \leqslant \frac{1}{2}$, and define

$$v_x = \begin{cases} \sqrt{\frac{1-\pi(A)}{\pi(A)}} & x \in A \\ -\sqrt{\frac{\pi(A)}{1-\pi(A)}} & x \notin A\,. \end{cases}$$

Note that $\langle v, \mathbf{1}\rangle_{\ell_2(\pi)} = \pi(A)\sqrt{\frac{1-\pi(A)}{\pi(A)}} - (1-\pi(A))\sqrt{\frac{\pi(A)}{1-\pi(A)}} = 0$, and

$$\|v\|^2_{\ell_2(\pi)} = 1 - \pi(A) + \pi(A) = 1\,.$$

Therefore

$$1 - \lambda_2 = \langle v, v(I-P)\rangle_{\ell_2(\pi)} = \frac{1}{2}\sum_{x,y} Q(x,y)(v_x - v_y)^2\,,$$

where the last equality is the usual one we have done with Laplacian matrices (like $I - P$) in preceding lectures. But the latter quantity is precisely

$$Q(A,\bar{A})\left(\sqrt{\frac{1-\pi(A)}{\pi(A)}} + \sqrt{\frac{\pi(A)}{1-\pi(A)}}\right)^2 \leqslant 2\frac{Q(A,\bar{A})}{\pi(A)} = 2\Phi(A)\,,$$

where the inequality uses the fact that $\pi(A) \leqslant \frac{1}{2}$.

1.2 Multi-commodity flows

Although Theorem 1.1 gives a nice characterization of rapid mixing in terms of conductance, the quantity Φ^* is NP-hard to compute, and can be difficult to get a handle on for explicit chains. Thus we now present another connection between conductance and multi-commodity flows.

We consider a multi-commodity flow instance on a graph with vertices corresponding to states Ω and edges $\{x, y\}$ with capacity $Q(x, y)$. The demand between x and y is $\pi(x)\pi(y)$. Let C^* be the optimal congestion that can be achieved by a multi-commodity flow satisfying all the demands (recalling that the congestion of an edge in a given flow is the ratio of the total flow over the edge to its capacity).

Theorem 1.2. *It holds that*

$$\frac{1}{2C^*} \leqslant \Phi^* \leqslant \frac{1}{C^*} O(\log |\Omega|) .$$

The right-hand side is due to Leighton and Rao (1988). We will only need the much simpler left-hand side inequality which can be proved as follows. Suppose there exists a flow achieving congestion C and consider some $A \subseteq \Omega$. Then

$$C \cdot Q(A, \bar{A}) \geqslant \pi(A)\pi(\bar{A}) .$$

This is because the left-hand side represents an upper bound on the total flow going across the cut—$Q(A, \bar{A})$ is the capacity across the cut $(A, \bar{A})$, and we have to rescale by C to account for the congestion. On the other hand, $\pi(A)\pi(\bar{A})$ represents the amount of flow that must be traveling across the cut to satisfy the demand. If $\pi(A) \leqslant \frac{1}{2}$, we conclude that

$$Q(A, \bar{A}) \geqslant \frac{\pi(A)\pi(\bar{A})}{C} \geqslant \frac{\pi(A)}{2C} ,$$

completing the proof.

Remark 1.3 (Proof sketch of RHS of Theorem 1.2). (This is related to HW#4(c), which would give the worse bound $O((\log |\Omega|)^2)$.) If we use linear programming duality to characterize C^*, it has the following dual representation:

$$\frac{1}{C^*} = \min_{d} \frac{\sum_{\{x,y\}} Q(x, y)d(x, y)}{\sum_{x,y\in\Omega} \pi(x)\pi(y)d(x, y)}, \tag{1.1}$$

where the minimum is over all symmetric distance functions $d(x, y)$ on $\Omega \times \Omega$ that satisfy the triangle inequality $d(x, y) \leqslant d(x, z) + d(z, y)$ for all $x, y, z \in \Omega$.

Recall that every finite metric space (X, d) admits a mapping $F : X \to \mathbb{R}^n$ with distortion $D \leqslant O(\log n)$, i.e.,

$$\frac{d(x, y)}{D} \leqslant \|F(x) - F(y)\|_2 \leqslant d(x, y) \qquad x, y \in X.$$

Now let us decompose the Euclidean distance on $\mathbb{R}^n$ into a convex combination over cuts. First, note that for any $a, b \in \mathbb{R}$, we have

$$|a - b| = \int_{-\infty}^{\infty} |\chi_s(a) - \chi_s(b)| \, ds,$$

where $\chi_s := \mathbf{1}_{(-\infty,s]}$. In other words, $\chi_s(a) = 1$ if $a \leqslant s$ and $\chi_s(a) = 0$ otherwise.

Let $\boldsymbol{g}$ denote a random n-dimensional Gaussian vector, i.e., $\boldsymbol{g} = (g_1, \ldots, g_n)$ where $\{g_i\}$ are i.i.d. $N(0, 1)$ random variables. Recall that for $u, v \in \mathbb{R}^n$, we have $\|u - v\|_2^2 = \mathbb{E}[\langle u - v, \boldsymbol{g}\rangle^2]$, because

$\langle u - v, \boldsymbol{g} \rangle$ is an $N(0, \|u - v\|_2^2)$ random variable (by the 2-stability property of normal random variables). One can also calculate: If g_0 is an arbitrary normal random variable with mean zero, then

$$\mathbb{E}[|g_0|] = \sqrt{\frac{2}{\pi}}\sqrt{\mathbb{E}[g_0^2]}.$$

Therefore:

$$\|u - v\|_2 = \sqrt{\mathbb{E}[\langle u - v, \boldsymbol{g} \rangle^2]} = \sqrt{\frac{\pi}{2}}\,\mathbb{E}\left[|\langle u - v, \boldsymbol{g} \rangle|\right]$$

We thus arrive at the following "cut decomposition" for all of $\mathbb{R}^n$:

$$\|u - v\|_2 = \mathop{\mathbb{E}}_{\boldsymbol{g}}\left[\int_{-\infty}^{\infty} |\chi_s(\langle u, \boldsymbol{g} \rangle) - \chi_s(\langle v, \boldsymbol{g} \rangle)|\; ds\right]$$

Suppose now that d is the optimal metric in (1.1) and let $F : \Omega \to \mathbb{R}^n$ denote a distortion $D \leqslant O(\log n)$ embedding. The distortion condition yields

$$\frac{1}{C^*} \geqslant \frac{1}{D}\frac{\sum_{\{x,y\}} Q(x,y)\|F(x) - F(y)\|_2}{\sum_{x,y} \pi(x)\pi(y)\|F(x) - F(y)\|_2} = \frac{\mathbb{E}_{\boldsymbol{g}}\left[\int_{-\infty}^{\infty} \sum_{\{x,y\}} Q(x,y)\left|\chi_s(\langle F(x), \boldsymbol{g} \rangle) - \chi_s(\langle F(y), \boldsymbol{g} \rangle)\right|\, ds\right]}{\mathbb{E}_{\boldsymbol{g}}\left[\int_{-\infty}^{\infty} \sum_{x,y} \pi(x)\pi(y)\left|\chi_s(\langle F(x), \boldsymbol{g} \rangle) - \chi_s(\langle F(y), \boldsymbol{g} \rangle)\right|\, ds\right]}$$

Finally, we observe that

$$\frac{\int f(x)\,dx}{\int g(x)\,dx} \geqslant \min_x \frac{f(x)}{g(x)}.$$

Thus there exists some choice of $g \in \mathbb{R}^n$ and $s \in \mathbb{R}$ such that

$$\frac{1}{C^*} \geqslant \frac{1}{D}\frac{\sum_{\{x,y\}} Q(x,y)|\chi_s(\langle F(x), g \rangle) - \chi_s(\langle F(y), g \rangle)|}{\sum_{x,y} \pi(x)\pi(y)|\chi_s(\langle F(x), g \rangle) - \chi_s(\langle F(y), g \rangle)|},$$

but the latter ratio is precisely $\frac{1}{D}\frac{Q(A,\bar{A})}{\pi(A)\pi(\bar{A})}$ for the set $A = \{x \in \Omega : \langle F(x), g \rangle \leqslant s\}$, hence

$$\frac{1}{C^*} \geqslant \frac{1}{D}\frac{Q(A,\bar{A})}{\pi(A)\pi(\bar{A})} \geqslant \frac{1}{2D}\frac{Q(A,\bar{A})}{\min(\pi(A), \pi(\bar{A}))} \geqslant \frac{\Phi^*}{2D},$$

verifying the RHS of Theorem 1.2.

1.3 The Markov chain

Recall now that our goal is to sample from the Gibbs measure π_λ introduced earlier. The following Markov chain is due to Jerrum and Sinclair. If we are currently at a matching $m \in \mathcal{M}(G)$, we define our local transition as follows.

1. With probability $1/2$, we stay at m.
2. Otherwise, choose an edge $e = \{u, v\} \in E(G)$ uniformly at random and:
   (a) If both u and v are unmatched in m, set $m := m \cup \{e\}$.
   (b) If $e \in m$, then with probability $1/\lambda$, put $m := m \setminus \{e\}$, and otherwise stay at m.
   (c) If exactly one of u or v is matched in m, then let e' be the unique edge that contains one of u or v and put $m := m \setminus \{e'\} \cup \{e\}$.

(d) If both u and v are matched, stay at m.

Exercise: Show that this chain is reversible with respect to the measure π_λ.

Now we would like to prove that this chain is rapid mixing by giving a low-congestion multi-commodity flow in the corresponding graph. In fact, we will give an "integral flow," i.e. we will specify for every pair of matchings $x, y \in \mathcal{M}(G)$, a path γ_{xy}.

To do this, consider the edges of x to be colored red and the edges of y to be covered blue. Then the colored union $x \cup y$ is a multi-graph where every node has degree at most 2. It is easy to see that every such graph breaks into a disjoint union of paths and even-length cycles. (Note also the trivial cycles of length two when x and y share an edge.)

The path γ_{xy} will "fix" each of these components one at a time (in some arbitrary order). The trivial cycles are already fine (we don't have to move those edges). To explain how to handle the path components, we look at a simple example. Suppose the path is $e_1, e_2, e_3, e_4, e_5, e_6$. Then we define a path from the red matching to blue the matching (in this component as follows):

$$e_1, e_3, e_5 \to e_3, e_5 \to e_2, e_5 \to e_2, e_4 \to e_2, e_4, e_6 \,.$$

Note that each transition is a valid step of the chain. We can do a similar thing for a cycle by first deleting a red edge so that it becomes a path.

Congestion analysis. So now we have given a path γ_{xy} between every pair of states $x, y \in \mathcal{M}(G)$. In the flow, this path should have flow value $\pi_\lambda(x)\pi_\lambda(y)$ so that is satisfies the corresponding demand. We are left to analyze the weight of paths that use a given "edge" (a transition) of the chain. The interested reader is referred to the beautiful argument at its original source [JS89].

References

[JS89] Mark Jerrum and Alistair Sinclair, *Approximating the permanent*, SIAM J. Comput. **18** (1989), no. 6, 1149–1178. MR 1025467 (91a:05075) 5