

Lecture 5

Instructor: Salil Vadhan

Scribe: Louis Golowich & Daniel Rodrigues

1 Recap / Corrections from Last Time

1.1 Eigenvalues on Cayley Graphs

Eigenvalues of undirected n -cycle = $w^{-r} + w^r = 2 \cos(\frac{2\pi r}{n})$, $r = 0, \dots, n - 1$.

1.2 Eigenvalues of Hypercube

$$G = \mathbb{Z}_2^d \text{ (}\{0, 1\}^d \text{ w/addition mod 2)}$$

$$S = \{e_1, e_2, \dots, e_d\} \text{ (i.e. } W = 1_s)$$

Fourier basis: for each $r \in \mathbb{Z}_2^d$,

$$\chi_r(x) = (-1)^{\langle r, x \rangle} = (-1)^{\langle r, x \rangle \pmod{2}}$$

$$\hat{W}_r = \sum_{s \in \mathbb{Z}_2^d} 1_s(s) \cdot (-1)^{\langle r, s \rangle}$$

$$= \sum_{i=1}^d (-1)^{\langle r, e_i \rangle} = \sum_{i=1}^d (-1)^{r_i}$$

$$= \#(0\text{'s in } r) - \#(1\text{'s in } r) = d - 2|r|$$

Table 1: Calculation of the adjacency matrix's eigenvalues of the hypercube.

$ r $	# of e-values	e-value of M	e-value of L	e-value of W	e-value of N
0	1	d	0	1	0
1	d	$d - 2$	2	$1 - 2/d$	$2/d$
2	$\binom{d}{2}$	$d - 4$	4	$1 - 4/d$	$4/d$
3	$\binom{d}{3}$	$d - 6$	6	$1 - 6/d$	$6/d$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$d/2$	$\binom{d}{d/2}$	0	d	0	1
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$d - 1$	d	$-(d - 2)$	$2d - 2$	$-1 + 2/d$	$2 - 2/d$
d	1	$-d$	$2d$	-1	2

1.3 Perron-Frobenius Theorem for *Symmetric* Matrices

Let M be a symmetric non-negative real matrix (e.g. adjacency or normalized adjacency matrix of **undirected graph** G) with the corresponding graph G (i.e. $E = \{(a, b) : M(a, b) > 0\}$) of which eigenvalues $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$. Then:

1. There exists a strictly positive eigenvector v_1 with the eigenvalue μ_1 (and the only non-negative eigenvectors are multiples of v_1).
2. $\mu_1 \geq \mu_2$ **with strict inequality if the graph is connected.**
3. $\mu_1 \geq -\mu_n$ **with strict inequality if the graph is connected and bipartite.** Furthermore, if the graph is bipartite then $\mu_{n-i} = -\mu_{i+1}$. for $i = 0, \dots, n - 1$.

2 Random walks on undirected graphs

Let G be undirected and weighted, with the random walk matrix defined as usual by $W = MD^{-1}$. For a random walk, the initial probability distribution on the vertices is denoted by p_0 , so that $p_t = W^t p_0$ gives the probability distribution after t steps of the walk. Our analysis of the random walk will use the (possibly non-orthogonal) eigendecomposition of $W = D^{1/2}AD^{-1/2}$ for the normalized adjacency matrix $A = D^{-1/2}MD^{-1/2}$. Specifically, because A is symmetric and therefore has an orthonormal eigenbasis ψ_1, \dots, ψ_n with eigenvalues $\omega_1 \geq \dots \geq \omega_n$, it follows that the vectors $\phi_i = D^{1/2}\psi_i$ form a (not necessarily orthogonal) eigenbasis for W with the same eigenvalues.

Now we write $p_0 = c_1\phi_1 + \dots + c_n\phi_n$, so $W^t p_0 = c_1\omega_1^t\phi_1 + \dots + c_n\omega_n^t\phi_n$. The eigenvalue of largest magnitude dominates. Here we have $\omega_1 = 1$ because $\psi_1 = d^{1/2}/\|d^{1/2}\|$ is a strictly positive eigenvector of A of eigenvalue 1. Thus by the Perron-Frobenius Theorem, $|\omega_i| \leq 1$ for all $i > 1$. So here $\phi_1 = D^{1/2}\psi_1 = d/\|d^{1/2}\|$, and

$$c_1 = (D^{-1/2}p_0)^T \psi_1 = p_0^T D^{-1/2} \frac{d^{1/2}}{\|d^{1/2}\|} = \frac{p_0^T \cdot 1}{\|d^{1/2}\|} = \frac{1}{\|d^{1/2}\|},$$

and therefore

$$c_1\phi_1 = \frac{d}{\|d^{1/2}\|^2} := \pi,$$

which is called the stationary distribution, and satisfies $W\pi = \pi$.

The following theorem works towards our goal of bounding the rate of convergence of the random walk to the stationary distribution.

Theorem 1. *Define*

$$\|x\|_{\pi^{-1}} := \sqrt{\sum_a \frac{x(a)^2}{\pi(a)}}.$$

Then

$$\max_{j>1} |\omega_j| = \max_{\text{prob. dist. } p} \frac{\|Wp - \pi\|_{\pi^{-1}}}{\|p - \pi\|_{\pi^{-1}}} := \omega_\pi.$$

Proof.

$$\begin{aligned} \max_{j>1} |\omega_j| &= \max_{x \perp \psi_1} \frac{\|Ax\|}{\|x\|} \\ &= \max_{y: D^{-1/2}y \perp \psi_1} \frac{\|D^{-1/2}Wy\|}{\|D^{-1/2}y\|} \\ &= \max_{y: \sum_a y_a = 0} \frac{\|Wy\|_{d^{-1}}}{\|y\|_{d^{-1}}} \\ &= \max_{\text{prob. dist. } p} \frac{\|W(p - \pi)\|_{d^{-1}}}{\|p - \pi\|_{d^{-1}}} \\ &= \max_{\text{prob. dist. } p} \frac{\|Wp - \pi\|_{\pi^{-1}}}{\|p - \pi\|_{\pi^{-1}}} \end{aligned}$$

The first equality above follows by Problem Set 1. The second follows by letting $x = D^{-1/2}y$. The third follows by the evaluation of ψ_1 above, and by our definition of the norm. The fourth follows by letting $p = \pi + cy$ for some sufficiently small constant c , as the third expression on the right hand side is invariant under scaling y . The fifth follows by the definition of π . \square

By definition, $\|W^t p_0 - \pi\|_{\pi^{-1}} \leq \omega_\pi^t \|p_0 - \pi\|_{\pi^{-1}}$. The initial distance satisfies

$$\|p_0 - \pi\|_{\pi^{-1}} \leq \|p_0\|_{\pi^{-1}} \leq \sqrt{1/\pi_{\min}}.$$

The final ℓ_1 distance (which equals double the total variation distance) satisfies

$$\|W^t p_0 - \pi\|_1 = \sum_a \frac{|W^t p_0(a) - \pi(a)|}{\pi(a)^{1/2}} \cdot \pi(a)^{1/2} \leq \sqrt{\sum_a \frac{(W^t p_0(a) - \pi(a))^2}{\pi(a)}} = \|W^t p_0 - \pi\|_{\pi^{-1}},$$

where the inequality above follows by the Cauchy-Schwartz inequality. Combining the above inequalities, we have shown the following result.

Theorem 2. *For every start distribution p_0 ,*

$$\|W^t p_0 - \pi\|_1 \leq \omega_\pi^t \|p_0\|_{\pi^{-1}} \leq \omega_\pi^t \cdot \frac{1}{\sqrt{\pi_{\min}}}$$

where $\pi_{\min} = \frac{d_{\min}}{\sum_a d(a)} \in [\frac{d_{\min}}{nd_{\max}}, \frac{1}{n}]$.

Corollary 3. *The “mixing time” to get within ℓ_1 distance ϵ of the stationary distribution is*

$$t = O\left(\frac{\log(nd_{\max}/\epsilon d_{\min})}{1 - \omega_\pi}\right).$$

The numerator here is fine, but we want to ensure that $\omega_\pi < 1$ to keep the denominator nonzero. Here $\omega_\pi = \max\{\omega_2, -\omega_n\}$, with $\omega_2 < 1$ iff G is connected and $\omega_n > -1$ iff G is nonbipartite.

If we instead switch to a lazy random walk with $\tilde{W} = \frac{1}{2}W + \frac{1}{2}I$, then $\tilde{\omega}_j = \frac{1+\omega_j}{2} \geq 0$, so $\tilde{\omega}_\pi = \tilde{\omega}_2 = \frac{1+\omega_2}{2}$ and $1 - \tilde{\omega}_\pi = \frac{1-\omega_2}{2} = \frac{\nu_2}{2}$, where ν_2 is the second smallest eigenvalue of the normalized Laplacian $N = I - A$.

Exercise 4. *Bound the mixing time of lazy and non-lazy random walks on the k -dimensional hypercube combinatorially and using eigenvalues, and compare.*

Proof. Recall the eigenvalues of the random walk matrix (computed with Cayley graphs) are $\omega_1 = 1$, $\omega_2 = 1 - 2/k$, and $\omega_n = -1$. So $\omega_\pi = 1$, which implies that there is no mixing in the non-lazy case. Combinatorially, this occurs because the hypercube is bipartite. In the lazy case, because $\tilde{\omega}_\pi = \frac{1+\omega_2}{2} = 1 - \frac{1}{k}$, our formula for the mixing time gives

$$t_{\text{mix}} = O\left(\frac{\log(n/\epsilon)}{1 - \tilde{\omega}_\pi}\right) = O\left(k^2 + k \log \frac{1}{\epsilon}\right),$$

as $n = 2^k$ and $d_{\max} = d_{\min} = k$. Combinatorially, the mixing time may be computed as follows. Consider a random walk starting at some length- k bit string. A step in the lazy random walk on the hypercube is given by choosing a uniformly random $1 \leq i \leq k$, and then flipping bit i with probability $\frac{1}{2}$, and leaving bit i unchanged with probability $\frac{1}{2}$, the latter case reflecting the laziness of the walk. Therefore for a given probability distribution p on the hypercube \mathbb{Z}_2^k , when a step of the random walk is executed and bit i is chosen, the resulting probability distribution $\tilde{W}(p)$ has bit i distributed uniformly on \mathbb{Z}_2 and independently of all other bits, while the joint distribution of all other bits is given by their marginal distribution in p . Thus once every bit is chosen as the bit i that is flipped with probability $\frac{1}{2}$, then the walk is perfectly mixed, as the distribution of each bit must be uniform on \mathbb{Z}_2 and independent of all other bits. To determine how many steps in the walk are needed to ensure each bit i is chosen with high probability, we refer to the solution of the well-known coupon collector’s problem (see [https://en.wikipedia.org/wiki/Coupon_collector%](https://en.wikipedia.org/wiki/Coupon_collector%27s_problem)

27s_problem), which says that the expected number of steps until each bit is chosen grows as $\Theta(k \log k)$. It follows by Markov's inequality that for any $\epsilon > 0$, there exists a sufficiently large constant C such that after $t = Ck \log k$ steps in the random walk, the probability that some bit has not been chosen is at most $\frac{\epsilon}{2}$. Thus with $t = Ck \log k$, for any initial distribution p_0 we may express the final distribution $\tilde{W}^t p_0$ as a linear combination of the stationary (uniform) distribution π and an error distribution p_{error} by conditioning on the event at each bit was chosen at least once, which gives $\tilde{W}^t p_0 = (1 - \frac{\epsilon}{2})\pi + \frac{\epsilon}{2}p_{\text{error}}$. Therefore by the triangle inequality $|\tilde{W}^t - \pi|_1 \leq \epsilon$, so we have shown a mixing time of $t_{\text{mix}} = O(k \log k)$. Note that the eigenvalue bound of $O(k^2)$ appears to be quadratically worse than the combinatorial bound, but because $k = \log n$, the gap is only logarithmic in n . \square

The converse to our mixing time result (may be on Problem Set 2) is stated below

Theorem 5. *For all undirected G , there exists a start vertex s such that for $p_0 = \delta_s$,*

$$t = \Omega\left(\frac{\omega_\pi}{1 - \omega_\pi} \cdot \log \frac{1}{\epsilon}\right)$$

steps are needed to get within ℓ_1 distance ϵ of π .

On Problem Set 1, it is shown that for every undirected, connected, nonbipartite G , it holds that $\omega_\pi \leq 1 - 1/\text{poly}(nd_{\text{max}})$. Therefore we have shown that mixing time is always at most $\text{poly}(nd_{\text{max}})$.

Next lecture we'll talk about random walks on directed graphs, and then discuss applications of random graphs, specifically Markov Chain Monte Carlo, and Pagerank.