Counting and Sampling **Fall 2017**

Lecture 7: Advanced Coupling & Mixing Time via Eigenvalues

Lecturer: Shayan Oveis Gharan *October 18*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

In this lecture first we discuss [HV05] to prove that the Glauber dynamics generates a random coloring of a graph G with maximum degree Δ using $q \geq 1.764\Delta$ colors in $O(n \log n)$. Our main motivation is to introduce additional more technical tools in coupling, beyond the path coupling technique. We prove the following theorem

Theorem 7.1. *Let $\alpha \approx 1.763\ldots$ satisfies $\alpha = e^{1/\alpha}$. If G is triangle free with max degree Δ, and $q \geq \Delta\alpha(1+\epsilon)$ and $\Delta \gg \frac{1}{\epsilon^2}\log(n)$, then the heat-bath chain mixes in $O(\frac{n}{\epsilon}\log(n))$ time.*

The main idea behind the proof is a *local uniformity property* which holds with high probability at the stationarity distribution. They show that for any triangle-free graph an easy lower bound on the number of available colors for all vertices of graph which holds with high probability at stationarity. We use the following notation throughout this section and the next one. For a coloring configuration X and a vertex v let $A(X, v)$ be the set of colors available to v. In other words, $A(X, v)$ is all colors that are not appeared on any neighbors of v in X.

Lemma 7.2. *Let G be a triangle-free graph with maximum degree Δ, $\epsilon > 0$ and $q > \Delta + 1$. For a uniformly random q-coloring of G, we have*

$$\mathbb{P}\left[\exists v : |A(X, v)| \leq q(e^{-\Delta/q} - \epsilon)\right] \leq ne^{-\epsilon^2 q}.$$

Proof. Fix a vertex v. We condition on the colors of all vertices which are not neighbors of v. We call the conditional information $\mathcal{F}$. Observe that, conditioned on $\mathcal{F}$, all neighbors of v are colored independently. This is because they form an independent set (by the traingle-free assumption).

Let $X_{c,u}$ be the random variable indicating that a neighbor u of v is colored with c.

$$|A(X, v)| = \sum_c \prod_{u \sim v} (1 - X_{c,u}).$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[|A(X, v)||\mathcal{F}\right] &= \sum_c \prod_{u \sim v} (1 - \mathbb{E}\left[X_{c,u}|\mathcal{F}\right]) \\
&= \sum_c \prod_{u \sim v : c \in A(X,u)} \left(1 - \frac{1}{|A(X, u)|}\right) \\
&\leq q \prod_c \prod_{u \sim v : c \in A(X,u)} \left(1 - \frac{1}{|A(X, u)|}\right)^{1/q} \\
&= q \prod_{u \sim v} \prod_{c \in A(X,u)} \left(1 - \frac{1}{|A(X, u)|}\right)^{1/q} \\
&\approx q \prod_{u \sim v} e^{-1/q} = qe^{-\Delta/q}.
\end{aligned}$$

where the first inequality follows by the AM-GM inequality. Note that since $X_{c,u}$'s are independent under $\mathcal{F}$, and $A(X, v)$ is a Lipschitz function of $X'_{c,u}s$ it follows that we have strong concentration on the above expected bound. Therefore, the claim simply follows by McDiarmid's inequality and a union bound.

Recall that McDiarmids inequality says that for any function $f(X_1, \ldots, X_n)$ which is a Lipshitz function with Lipshitz constant 1 of independent variables $X_1, \ldots, X_n$ we have

$$\mathbb{P}\left[|f(X_1, \ldots, X_n) - \mathbb{E}f(X_1, \ldots, X_n)| \geq \epsilon\right] \leq e^{-\epsilon^2/n}.$$

□

Equipped with the above lemma, let $\alpha \approx 1.763\ldots$ be such that $\alpha = e^{1/\alpha}$. Then, for $q = \Delta\alpha$, and for $\Delta \geq \Omega(\log n)$ we have

$$qe^{-\Delta/q} \geq qe^{-1/\alpha} = \frac{q}{\alpha} = \Delta$$

Therefore, if we take a slightly larger q, say $q = \Delta\alpha(1+\epsilon)$, for $\Delta \gg \frac{1}{\epsilon^2}\log n$, we have that with probability $1 - 1/n$ a random coloring X satisifies the following: For all v, $|A(X, v)| > \Delta(1 + \Omega(\epsilon))$.

7.1 Improved bound on Coupling

In this section step we show that if in a coloring X all vertices have strictly more than Δ available colors, then we can couple X with any proper coloring Y such that the distance decreases in expectation. Note that as a special case, this will also give us a proof of the $q \geq 2\Delta + 1$ bound from last lecture using ordinary coupling.

We work with a slightly different Markov chain: We choose a random vertex v and we assign a color that is not appeared in any of its neighbors uniformly at random, i.e., we work with the Heat-Bath chain as opposed to the Metropolis rule.

Lemma 7.3. *Let X be a proper coloring of G such that for some $0 < \beta < 1$ and all vertices v,*

$$|A(X, v)| \geq \frac{\Delta}{1-\beta}.$$

Then, for every $Y \in \Omega$, there exists a coupling $(X, Y) \to (X', Y')$ such that

$$\mathbb{E}\left[d(X', Y')|X, Y\right] \leq (1 - \beta/n)\, d(X, Y).$$

Note that in particular, if $q \geq 2\Delta + 1$, then, we can let $\beta \geq \Omega(1/\Delta)$, and the above lemma would simply imply the mixing.

Proof. We use a coupling strategy similar to the one we considered last time. First of all, both X, Y choose the same vertex v. For every color $c \in A(X, v) \cap A(Y, v)$ both chains choose c with probability $\min\{\frac{1}{|A(X,v)|}, \frac{1}{|A(Y,v)|}\}$. Note that in all other cases v will end up with different colors in X', Y'. So, we don't

need to specify how to match the rest of the events. Now, observe that

$$\begin{aligned}\mathbb{P}\left[c_{X'}(v) \neq c_{Y'}(v)|v\right] &= 1 - \frac{|A(X,v) \cap A(Y,v)|}{\max\{|A(X,v)|, |A(Y,v)|\}} \\ &= \frac{\max\{A(X,v), A(Y,v)\} - |A(X,v) \cap A(Y,v)|}{\max\{A(X,v), A(Y,v)\}} \\ &\leq \frac{|\{u \sim v : c_X(u) \neq c_Y(u)\}|}{\max\{A(X,v), A(Y,v)\}} \\ &\leq \frac{1-\beta}{\Delta}|\{u \sim v : c_X(u) \neq c_Y(u)\}|\end{aligned}$$

Note that the second to last inequality can be justified as follows: Say $|A(X,v)| > |A(Y,v)|$. Then, for every color in $c \in A(X,v) \setminus (A(X,v) \cap A(Y,v))$ there must be a vertex u that is a neighbor of v such that $c_Y(u) = c \neq c_X(u)$. The last inequality follows by the assumption of the lemma.

Therefore, for every vertex v,

$$\mathbb{P}\left[c_{X'}(v) \neq c_{Y'}(v)\right] \leq \frac{1-\beta}{n\Delta}|\{u \sim v : c_X(u) \neq c_Y(u)\}| + \left(1 - \frac{1}{n}\right)\mathbb{I}\left[c_X(v) \neq c_Y(v)\right].$$

Summing this over all v, we get

$$\mathbb{E}\left[d(X',Y')|X,Y\right] \leq \frac{1-\beta}{n\Delta}(\Delta d(X,Y)) + \left(1 - \frac{1}{n}\right)d(X,Y) = \left(1 - \frac{\beta}{n}\right)d(X,Y)$$

as desired. □

7.2 Coupling with stationarity

Using lemmas 7.2 and 7.3, to prove Theorem 7.1, it remains to show that for a coupling proof it is enough to show that for most states X the distance decreases:

Theorem 7.4. *Suppose that the distance function $d(.,.)$ is integral. For $\epsilon > 0$ and $S \subseteq \Omega$ suppose that for all $X \in \Omega$ and $Y \in S$ there is a coupling $(X,Y) \to (X',Y')$ such that*

$$\mathbb{E}\left[d(X',Y')|X,Y\right] \leq (1-\epsilon)d(X,Y).$$

if $\pi(S) \geq 1 - \frac{\epsilon}{2\,\mathrm{diam}(\Omega)}$, then

$$\tau_{mix} \leq O\left(\frac{\log(\mathrm{diam}(\Omega))}{\epsilon}\right).$$

Proof. The main insight of the proof is that in a coupling proof we can always assume $X_0 = x$ for some state $x \in \Omega$ and $Y_0 \sim \pi$. Therefore, for all t,

$$\mathbb{P}\left[Y_t \notin S\right] = 1 - \pi(S).$$

As usual if $X = Y$, then $X' = Y'$. Otherwise, if Y_t is not in S we just follow an independent coupling of X_t, Y_t. Otherwise, we use the coupling promised in the statement of the theorem.

Now, observe that on the event that $Y_t \notin S$, we have no control on the $d(X_{t+1}, Y_{t+1})$ but we know in the worst case it is at most $\mathrm{diam}(\Omega)$. Otherwise, we know the distance decreases. Therefore,

$$\begin{aligned}\mathbb{E}\left[d(X',Y')\right] &= \mathbb{E}\left[d(X',Y')|Y \in S\right]\mathbb{P}\left[Y \in S\right] + \mathbb{E}\left[d(X',Y')|Y \notin S\right]\mathbb{P}\left[Y \notin S\right] \\ &\leq (1-\epsilon)d(X,Y) + \mathrm{diam}(\Omega) \cdot \frac{\epsilon}{2\,\mathrm{diam}(\Omega)} \leq (1-\epsilon/2)d(X,Y).\end{aligned}$$

where we used that $d(X,Y) \geq 1$ because $X \neq Y$ and $d(.,.)$ is integer. So, the chain mixes in time $O(\frac{1}{\epsilon}\log(\text{diam}(\Omega)))$. □

This completes the proof of Theorem 7.1.

7.3 Mixing Time Using Eigenvalues

In this section we discuss a spectral theory of reversible Markov chains. Based on this theory in the next lecture we introduce the *path technology* which will be one of the strongest tools to analyze Markov chains.

Note that we particular focus on reversible chains because every chain that we will use in our counting tasks is reversible.

In this section we will treat the Markov kernel K as an operator. It turns out that K is not a symmetric matrix (operator). But, it has real eigenvalues. There are several ways to see this. Perhaps, one of the cleanest ways is to study K in a different linear product space, in which K is self-adjoint, i.e., symmetric. The inner product space $L^2(\pi)$ is defined as follows:

$$\langle f, g\rangle = \sum_x f(x)g(x)\pi(x).$$

In particular, $\|f\| = \sqrt{\sum_x f(x)^2\pi(x)}$.

Recall that for any vector/function f,

$$Kf(x) = \sum_y K(y,x)f(y).$$

Although K is not a symmetric matrix, it is self-adjoint with respect the above inner product, i.e., for any two functions f, g, $\langle Kf, g\rangle = \langle f, Kg\rangle$. This is because,

$$\begin{aligned}\langle Kf, g\rangle = \sum_x Kf(x)g(x)\pi(x) &= \sum_x \left(\sum_y K(x,y)f(y)\right) g(x)\pi(x)\\ &= \sum_{x,y} K(x,y)f(y)g(x)\pi(x)\\ &= \sum_{x,y} K(y,x)f(y)g(x)\pi(y) = \langle f, Kg\rangle.\end{aligned}$$

Now, we can apply the spectral theorem: K has orthonormal set of eigenvalues $\lambda_1 \geq \cdots \geq \lambda_{|\Omega|}$ with corresponding eigenvectors $\psi_1, \ldots, \psi_{|\Omega|}$ such that for all i,

$$K\psi_i = \lambda_i\psi_i,$$

and for all $i \neq j$,

$$\langle \psi_i, \psi_j\rangle = 0.$$

It is not hard to see that we always have $\lambda_1 = 1$ and $\psi_1 = \mathbf{1}$ because K is stochastic. Furthermore, also by stochasticity, $|\lambda_i| \leq 1$ for all $i \geq 2$.

We can translate the ergodicity properties of a Markov chain to this language. A (reversible) Markov chain is irreducible iff $\lambda_2 < 1$, and a (reversible) markov chain is aperiodic, i.e., it is not bipartite, iff $\lambda_{|\Omega|} > -1$. Let

$$\lambda^* = \max_{i>1} |\lambda_i| = \max\{\lambda_2, |\lambda_n|\}.$$

Next, we show that if $\lambda^* < 1$, then we can bound the mixing time by $\log(n)/(1-\lambda^*)$.

Firstly, it turns out that it is more convenient to measure the mixing time in L_2 distance as opposed to the total variation distance; this is because eigenvalues are "L_2 properties" of the chain.

Definition 7.5 (L_P distance)**.** *For a kernel K, the L_p distance of $K^t(x,.)$ to stationary is defined as*

$$\left\| \frac{K^t(x,.)}{\pi(.)} - \mathbf{1} \right\|_p = \left(\sum_y \pi(y) \left| \frac{K^t(x,y)}{\pi(y)} - 1 \right|^p \right)^{1/p}.$$

Here, we are thinking of $K^t(x,y)/\pi(y)$ as a function of y.

Observe that

$$\begin{aligned} \left\| K^t(x,.) - \pi \right\|_{TV} &= \frac{1}{2} \sum_y |K^t(x,y) - \pi(y)| = \frac{1}{2} \sum_y \pi(y) \left| \frac{K^t(x,y)}{\pi(y)} - 1 \right| \\ &\leq \frac{1}{2} \sqrt{\sum_y \pi(y) \left| \frac{K^t(x,y)}{\pi(y)} - 1 \right|^2} = \frac{1}{2} \left\| \frac{K^t(x,.)}{\pi} - \mathbf{1} \right\|_2. \end{aligned} \tag{7.1}$$

So, if we prove that the L_2 distance is small, it automatically implies that the total variation distance is small. Note that if $p = \infty$, then the distance of K^t/π from 1 is equal to $\max_y |K^t(x,y)/\pi(y) - 1|$.

Theorem 7.6. *Let $\lambda^* = \max_{i>1} |\lambda_i| = \max\{\lambda_2, |\lambda_n|\}$. Then,*

$$\left\| \frac{K_x^t}{\pi} - \mathbf{1} \right\|_2 \leq \frac{\lambda^{*2t}}{\pi(x)}.$$

Therefore, by (7.1),

$$\tau_x(\epsilon) \leq \frac{\log\left(\frac{1}{\epsilon \cdot \pi(x)}\right)}{1-\lambda^*}.$$

Note that the above bound this bound is ideal in applications where π is very far from a uniform distribution. In these scenarios we want to start the chain from a state with large $\pi(x)$. This idea is usually called a *warm start*.

Proof. First, observe that for any eigenvector ψ_i we have

$$\left\langle \frac{K_x^t}{\pi}, \psi_i \right\rangle = \sum_y \frac{K^t(x,y)}{\pi(y)} \psi_i(y)\pi(y) = \lambda_i^t \psi_i(x).$$

Note that by above, $\langle \frac{K_x^t}{\pi}, \mathbf{1} \rangle = 1$. Therefore,

$$\frac{K_x^t}{\pi} - \mathbf{1} = \sum_{i>1} \left\langle \frac{K_x^t}{\pi}, \psi_i \right\rangle \psi_i = \sum_{i>1} \lambda_i^t \psi_i(x) \psi_i.$$

And,

$$\left\| \frac{K_x^t}{\pi} - \mathbf{1} \right\|^2 = \sum_{i>1} \lambda_i^{2t} \psi_i(x)^2 \leq \lambda^{*2t} \sum_{i>1} \psi_i(x)^2 \leq \frac{\lambda^{*2t}}{\pi(x)}$$

The only nontrivial step is the last inequality which holds because

$$\sum_i \psi_i(x)^2 = \sum_i \langle \psi_i, \frac{\mathbf{1}_x}{\pi} \rangle^2 = \left\| \sum_i \langle \psi_i, \frac{\mathbf{1}_x}{\pi} \rangle \psi_i \right\|^2 = \left\| \frac{\mathbf{1}_x}{\pi} \right\|^2 = \frac{1}{\pi(x)}.$$

where $\mathbf{1}_x$ is the indicator function of x. □

References

[HV05] Tom Hayes and Eric Vigoda. Coupling with the stationary distribution and improved sampling for colorings and independent sets. In *SODA*, pages 971–979. Society for Industrial and Applied Mathematics, 2005. 7-1