[CS3958: Lecture 12] Reversible Chains, Spectral Decomposition

Instructor: Chihao Zhang, Scribed by Yulin Wang

December 20, 2022

1 Reversible Chains

A Markov chain P over state space $[n]$ is *(time) reversible* if there exists some distribution π satisfying

$$\forall i, j \in [n],\ \pi(i)P(i,j) = \pi(j)P(j,i).$$

This family of identities is called *detailed balance conditions*. Moreover, we have

Proposition 1 *If P is reversible w.r.t a distribution π, then π must be a stationary distribution of P.*

Proof. To see this, note that

$$\pi^\top P(j) = \sum_{i\in[n]} \pi(i)P(i,j) = \sum_{i\in[n]} \pi(j)P(j,i) = \pi(j).$$

□

The name *reversible chain* comes from the fact that for any sequence of variables $X_0, X_1, \ldots, X_t$ following the chain, the distribution of $(X_0, X_1, \ldots, X_{t-1}, X_t)$ is identical to the distribution of $(X_t, X_{t-1}, \ldots, X_1, X_0)$.

We will study reversible chains since their transition matrices are essentially *symmetric* in some sense, so many powerful tools in linear algebra apply. We will also see that reversible chains are general enough for most of our (algorithmic) applications.

You can verify that the the random walks on the hypercube studied in the last lecture are reversible Markov chains with respect to uniform distribution.

- P is aperiodic $\iff$ G is not bipartite.
- P is irreducible $\iff$ G is connected.
- P is reversible $\iff$ G is undirected.

2 Metropolis Algorithm

Given a distribution π over a state space Ω, how can we design a Markov chain P so that π is the stationary distribution of P? This is easy because we can simply let $P = \mathbf{1}\pi^\top$. However, what if we only allow a given set of entries of P to be nonzero?

This is equivalent to the problem of assigning transition probabilities in a transition graph G so that π is the stationary distribution of the random walk. The *Metropolis algorithm* provides a way to achieve the goal as long as G is connected and undirected.

Let Δ be the maximum degree of the transition graph except self-loops.[1] We describe the following process to construct a transition matrix P : Choose $k \in [\Delta + 1]$ uniformly at random. For any $i \in [n]$, let $\{j_1, j_2, \ldots, j_d\}$ be the d neighbours of i. We consider the transition at state i

[1] That is $\Delta \triangleq \max_{u\in[n]} \sum_{v\neq u\in[n]} \mathbb{1}[(u,v) \in E]\Big)$.

: - If $d+1 \le k \le \Delta+1$, do nothing. - If $k \le d$, - propose to move from i to j_k. - accept the proposal with probability $\min\left\{\frac{\pi(j_k)}{\pi(i)}, 1\right\}$. Then the transition matrix is, for every $i, j \in [n]$,

$$P(i,j) = \begin{cases} \frac{1}{\Delta}\min\left\{\frac{\pi(j)}{\pi(i)}, 1\right\}, & \text{if } i \neq j; \\ 1 - \sum_{k\neq i} p_{ik}, & \text{if } i = j. \end{cases}$$

We can verify that P is reversible with respect to π:

$$\forall i, j \in \Omega: \ \pi(i)P(i,j) = \pi(i)\cdot\frac{1}{\Delta}\min\left\{\frac{\pi(j)}{\pi(i)}, 1\right\} = \frac{\min\{\pi(i), \pi(j)\}}{\Delta} = \pi(j)P(j,i).$$

The advantage of the Metropolis algorithm is that we do not need to know π in order to implement the algorithm. We only need to know the quantity $\frac{\pi(j)}{\pi(i)}$, which is much easier to compute in many applications.

3 Sample Proper Coloring

Let's consider the problem of sampling proper colorings. Given a graph $G = (V, E)$, we want to color the vertices using q colors under the condition that no two adjacent vertices share the same color. More formally, a coloring of G is a mapping $c : V \mapsto [q]$, and we call it *proper* iff $\forall \{u, v\} \in E, c(u) \neq c(v)$. The proper coloring problem is **NP**-hard in general. However, for $q > \Delta$ there always exists a proper coloring that can be easily obtained by a greedy algorithm, where Δ is the maximum degree of the graph.

If we want to count the number of proper colorings, then the problem becomes harder. It is known that for every $q \ge \Delta$, the problem is #**P**-hard. On the other hand, we can use a uniform sampler to obtain an algorithm to approximately counting the number of proper coloring, at an arbitrarily low cost in the precision.

In fact, it is known that an approximate counting algorithm is equivalent to an uniform sampler in many cases (for example, sampling proper coloring). We only show one direction here: a sampler implies an algorithm for approximate counting. Given a graph $G = (V, E)$ with $V = [n]$, let C be the set of proper colorings and $Z = |C|$. Suppose we have an oracle that can uniformly generate a proper coloring from C. Fix a proper coloring σ. We have

$$\begin{aligned} \frac{1}{Z} =& \mathbf{Pr}_{x\sim C}[x = \sigma] \\ =& \mathbf{Pr}_{x\sim C}[x(1) = \sigma(1) \wedge x(2) = \sigma(2) \wedge \dots] \\ =& \prod_{i=1}^{n} \mathbf{Pr}\left[x(i) = \sigma(i) \,\middle|\, \bigcap_{j<i} x(j) = \sigma(j)\right]. \end{aligned}$$

The above probability can be estimated by taking a number of samples from the oracle, and computing the ratio between colorings such that $x(j) = \sigma(j)$ for $j \le i$ and ones that $x(j) = \sigma(j)$ for $j < i$. Moreover, the ratio we just estimated is bounded below by an inverse polynomial and therefore polynomial number of sample suffices to estimate ratio accurately. The

strategy works even if the sampler is an approximate one. Hence one can approximately compute Z. See [JVV86] for more details.

Now we use MCMC to do sampling. Consider the following Markov chain to sample proper colorings:

- Pick $v \in V$ and $c \in [q]$ uniformly at random.
- Recolor v with c if possible.

It is indeed a Metropolis algorithm. Let

$$\sigma^{v \leftarrow c}(u) = \begin{cases} \sigma(u) & u \neq v \\ c & u = v. \end{cases}$$

$\sigma^{v \leftarrow c}$ is a neighbor of σ on the transition graph, and we accept it if $\sigma^{v \leftarrow c}$ is a proper coloring, i.e. $\frac{\pi(\sigma^{v \leftarrow c})}{\pi(\sigma)} = 1$.

The chain is aperiodic since self-loops exist in the walk. For $q \geq \Delta + 2$, the chain is irreducible. The bound $q \geq \Delta + 2$ is tight for irreducibility since when $q = \Delta + 1$, each proper coloring of complete graph is frozen. It is still an open problem if the mixing time of the chain is polynomial in the size of the graph under the condition $q \geq \Delta + 2$. The best bound so far requires that $q \geq (\frac{11}{6} - \varepsilon)\Delta$ for a certain constant $\varepsilon > 0$. Here, we shall give a rapid mixing proof when $q > 4\Delta$ using the method of coupling.

The coupling we used is simple: Both players pick same v and c to move. However, we are not able to reduce the analyze the coupling to *coupon collector* as we did before. We introduce a more general method to analyze couplings. We define a certain distance $d(x, y)$ for any two configurations $x, y \in \Omega$. We can assume without loss of generality that if $x \neq y$ then $d(x, y) \geq 1$ since Ω is finite. Consider a coupling ω_t of μ_t, ν_v. Then for every $t \geq 0$ and $(X_t, Y_t) \sim \omega_t$, we try to establish

$$\mathbf{E}\left[d(X_{t+1}, Y_{t+1}) \mid (X_t, Y_t)\right] \leq (1 - \alpha) d(X_t, Y_t)$$

for some $\alpha \in (0, 1]$. In other words, $\{d(X_t, Y_t)\}_{t \geq 0}$ is a supermartingale. This implies that for every $t \geq 1$,

$$\mathbf{E}\left[d(X_t, Y_t)\right] \leq (1 - \alpha)\mathbf{E}\left[d(X_{t-1}, Y_{t-1})\right] \leq (1 - \alpha)^t d(X_0, Y_0).$$

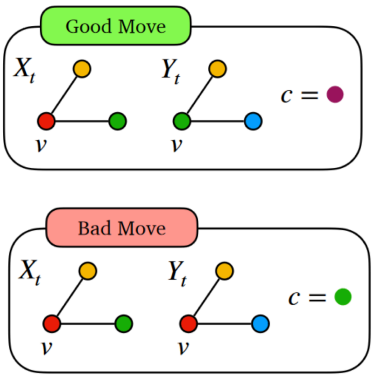
If we have a universal upper bound for $d(X_0, Y_0)$, say n, then by coupling lemma

$$\begin{aligned} D_{\mathrm{TV}}(\mu_t, \nu_t) &\leq \mathbf{Pr}_{(X_t, Y_t) \sim \omega_t}\left[X_t \neq Y_t\right] \\ &= \mathbf{Pr}\left[d(X_t, Y_t) \geq 1\right] \\ &\leq \mathbf{E}\left[d(X_t, Y_t)\right] \\ &\leq (1 - \alpha)^t \cdot n. \end{aligned}$$

Now come back to our problem of sampling proper colorings. Suppose X_t, Y_t are two proper colorings. We define the distance $d(X_t, Y_t)$ as their Hamming distance, i.e. the number of vertices colored differently in two colorings. Our coupling of two chains is that we always choose the same v, c in each step. The distance between two colorings can change at most 1 since only v is affected. The possible changes can be divided into two kinds:

- Good move: $X_t(v) \neq Y_t(v)$, and both change into c successfully. It will decrease distance by 1.

- Bad move: $X_t(v) = Y_t(v)$, one succeeds and one fails in the changing. It will increase distance by 1.



Consider the probabilities of two types of moves. For good moves, w.p. $\frac{d(X_t,Y_t)}{n}$, $X_t(v) \neq Y_t(v)$, and there are at least $q - 2\Delta$ choices of c to make it a good move. So

$$\begin{aligned}\mathbf{Pr}\left[d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) - 1\right] =& \mathbf{Pr}_{(v,c)\in V\times[q]}\left[(v,c) \text{ is a good move}\right] \\ \geq& \frac{d(X_t, Y_t)}{n} \cdot \frac{q - 2\Delta}{q}.\end{aligned}$$

For bad moves, there exists a neighbor w of v such that its color is different in two colorings, and in one coloring w is of color c. By a counting argument, we have

$$\mathbf{Pr}\left[d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) + 1\right] = \mathbf{Pr}_{(v,c)\in V\times[q]}\left[(v,c) \text{ is a bad move}\right] \leq \frac{\Delta d(X_t, Y_t)}{n} \cdot \frac{2}{q}.$$

Therefore,

$$\begin{aligned}\mathbf{E}\left[d(X_{t+1}, Y_{t+1})|(X_t, Y_t)\right] &= d(X_t, Y_t) + \mathbf{Pr}\left[d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) + 1\right] - \mathbf{Pr}\left[d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) - 1\right] \\ &\leq d(X_t, Y_t) + \frac{\Delta d(X_t, Y_t)}{n} \cdot \frac{2}{q} - \frac{d(X_t, Y_t)}{n} \cdot \frac{q - 2\Delta}{q} \\ &\leq d(X_t, Y_t)\left(1 - \frac{q - 4\Delta}{nq}\right).\end{aligned}$$

In the case $q > 4\Delta$, if we want

$$D_{\mathrm{TV}} \leq \left(1 - \frac{1}{nq}\right)^t n \leq \varepsilon,$$

we have the mixing time is bounded by

$$\tau_{\mathrm{mix}}(\varepsilon) \leq nq \log \frac{n}{\varepsilon}.$$

4 *Spectrum of Reversible Markov Chains*

Another advantage to use reversible chains is that their transition matrices have *real* eigenvalues. This follows from the fact that those matrices are essentially symmetric. As a result, we can apply tools in linear algebra to study them. We now develop the spectral decomposition theorem for *reversible* P. First, we have the following spectral decomposition theorem for *symmetric* P:

Theorem 2 (Spectral Decomposition Theorem) *If $P \in \mathbb{R}^{n\times n}$ is a symmetric matrix, then it has n real eigenvalues $\lambda_1, \ldots, \lambda_n$ with corresponding n orthonormal eigenvectors $v_1, \ldots, v_n$ satisfying*

$$P = \sum_{i=1}^{n} \lambda_i v_i v_i^\top.$$

If we let $V = \left[v_1, v_2, \ldots, v_n\right]$ *and* $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$*, then above can be written in the matrix form*

$$P = V\Lambda V^\top.$$

Now we prove a similar decomposition theorem for *reversible* P. Suppose P is reversible with respect to π. Let $\Pi = \operatorname{diag}(\pi)$ be the diagonal matrix with $\Pi(i,i) = \pi(i)$. Define $Q = \Pi^{\frac{1}{2}} P \Pi^{-\frac{1}{2}}$, then we can verify that Q is symmetric:

$$Q(i,j) = \pi(i)^{\frac{1}{2}} P(i,j) \pi(j)^{-\frac{1}{2}} = \pi(j)^{\frac{1}{2}} P(j,i) \pi(i)^{-\frac{1}{2}} = Q(j,i).$$

So we can apply the spectral decomposition theorem for Q, which yields

$$Q = \sum_{i=1}^{n} \lambda_i u_i u_i^\top,$$

where $\lambda_1, \ldots, \lambda_n$ are eigenvalues of Q with corresponding orthonormal eigenvectors $u_1, \ldots, u_n$. If we let $v_i \triangleq \Pi^{-\frac{1}{2}} u_i$, then the above is equivalent to

$$P = \sum_{i=1}^{n} \lambda_i \Pi^{-\frac{1}{2}} u_i u_i^\top \Pi^{\frac{1}{2}} = \sum_{i=1}^{n} \lambda_i v_i v_i^\top \Pi.$$

We claim that $\lambda_1, \ldots, \lambda_n$ are eigenvalues of P with corresponding eigenvectors $v_1, \ldots, v_n$. To see this, we have for any $j \in [n]$:

$$\begin{aligned} P v_j &= \sum_{i=1}^{n} \lambda_i \Pi^{-\frac{1}{2}} u_i u_i^\top \Pi^{\frac{1}{2}} v_j \\ &= \sum_{i=1}^{n} \lambda_i \Pi^{-\frac{1}{2}} u_i u_i^\top \Pi^{\frac{1}{2}} \Pi^{-\frac{1}{2}} u_j \\ &= \lambda_j v_j. \end{aligned}$$

Everything looks nice if we equip $\mathbb{R}^n$ with the inner product $\langle \cdot, \cdot \rangle_\Pi$ defined as $\langle x, y \rangle_\Pi = x^\top \Pi y = \sum_{i=1}^n \pi(i) x(i) y(i)$. It is clear that $v_1, \ldots, v_n$ are orthonormal with respect to the inner product:

$$\langle v_i, v_j \rangle_\Pi = \begin{cases} 0, & \text{if } i \neq j; \\ 1, & \text{if } i = j. \end{cases}$$

References

[JVV86] Mark R Jerrum, Leslie G Valiant, and Vijay V Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical computer science*, 43:169–188, 1986. 3