ADVANCED ALGORITHMS (XII)

CHIHAO ZHANG

1. Review of Graph Spectrum

We will study the relation between eigenvalues of the transition matrix of a Markov chain and its mixing time. Recall from Lecture 9 that we always work on reversible chains to which the spectral decomposition theorem applies. Let us first restate the theorem: Recall that for every $x, y \in \mathbb{R}^n$, we define the inner product $\langle x, y\rangle_\pi = \sum_{i\in[n]} \pi(i)x(i)y(i)$.

Theorem 1. *Let $P \in \mathbb{R}^{n\times n}$ be reversible with respect to π and $\mathcal{G}_P$ be its transition graph. Then the Hilbert space $(\mathbb{R}^n, \langle\cdot,\cdot\rangle_\pi)$ has an orthonormal basis $\{v_i\}_{i\in[n]}$ corresponding to real eigenvalues $\{\lambda_i\}_{i\in[n]}$. Moreover, assuming $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$, then it holds that*

1. *$\lambda_n = 1$;*
2. *$\lambda_1 \ge -1$ and $\lambda_1 = -1$ if and only if one of components of $\mathcal{G}_P$ is bipartite;*
3. *$\lambda_{n-1} = 1$ if and only if P is reducible.*

If we let $D_\pi \triangleq \mathrm{diag}(\pi(1), \dots, \pi(n))$, then we can write

$$P = \sum_{i=1}^{n} \lambda_i v_i v_i^T D_\pi.$$

We can also let $v_n = \mathbf{1}$ be the eigenvector corresponding to $\lambda_n = 1$ in the above theorem. Therefore, we have for every $t \ge 0$, it holds that

$$P^t = \Pi + \sum_{i=1}^{n-1} \lambda_i^t v_i v_i^T D_\pi,$$

where $\Pi = \begin{bmatrix} \pi^T \\ \vdots \\ \pi^T \end{bmatrix}$ is the matrix whose rows are all π^T. It follows from Theorem 1 that if P is irreducible and aperiodic, then $\lim_{t\to\infty} P^t = \Pi$, and the speed of the convergence can be controlled by $(1 - \lambda^*)^t$ where

$$\lambda^* \triangleq \max\{|\lambda_1|, |\lambda_{n-1}|\}.$$

We define the *relaxation time* of P as

$$\tau_{\mathrm{rel}} \triangleq \frac{1}{1-\lambda^*}.$$

Both the mixing time τ_{mix} and the relaxation time τ_{rel} measure how fast a chain converges to its stationary distribution. They are also related as in the following proposition:

Proposition 2. *Let P be a reversible irreducible aperiodic Markov chain with stationary distribution π. Then for every $\varepsilon > 0$,*

$$(\tau_{\mathrm{rel}} - 1)\log\left(\frac{1}{2\varepsilon}\right) \le \tau_{\mathrm{mix}}(\varepsilon) \le \tau_{\mathrm{rel}} \log\left(\frac{1}{\varepsilon\pi_{\min}}\right),$$

where $\pi_{\min} \triangleq \min_{x\in\Omega} \pi(x)$.

Recall the variational characterization of eigenvalues and the Rayleigh quotient we introduced before. We have similar notions in the Hilbert space $(\mathbb{R}^n, \langle\cdot,\cdot\rangle_\pi)$.

Definition 3. Let $f, g \in \mathbb{R}^n$ be two vectors (we also view them as functions $[n] \to \mathbb{R}$). The *Dirichlet form* of f and g is defined to be

$$\mathcal{E}(f, g) \triangleq \langle (I - P)f, g\rangle_\pi.$$

The matrix $I - P$ is the normalized Laplacian. The following proposition is a property of Laplacians (we also met it when studying graph Laplacians):

Proposition 4.

$$\mathcal{E}(f, f) = \frac{1}{2} \sum_{x,y \in \Omega} \pi(x)P(x, y)(f(x) - f(y))^2.$$

Proof.

$$\begin{aligned}
\text{RHS} &= \frac{1}{2} \sum_{x,y \in \Omega} f(x)^2 \pi(x)P(x, y) + \frac{1}{2} \sum_{x,y \in \Omega} f(y)^2 \pi(x)P(x, y) - \sum_{x,y \in \Omega} f(x)f(y)\pi(x)P(x, y) \\
&= \frac{1}{2} \sum_{x,y \in \Omega} f(x)^2 \pi(x)P(x, y) + \frac{1}{2} \sum_{x,y \in \Omega} f(y)^2 \pi(y)P(y, x) - \sum_{x \in \Omega} f(x)\pi(x) \sum_{y \in \Omega} f(y)P(x, y) \\
&= \sum_{x \in \Omega} f(x)^2 \pi(x) - \sum_{x \in \Omega} f(x)\pi(x)[Pf](x) \\
&= \langle f, f \rangle_\pi - \langle Pf, f \rangle_\pi \\
&= \langle (I - P)f, f \rangle_\pi,
\end{aligned}$$

where the second equality follows from the detailed balance condition of P. □

We can view the quantity $\pi(x)P(x, y)$ as the *weight* or *capacity* of the directed edge (x, y) and sometimes denote it by $Q(x, y)$.

If we let $\gamma \triangleq 1 - \lambda_{n-1}$, the variational characterization of λ_{n-1} is

$$\gamma = \min_{\substack{f \perp_\pi \mathbf{1} \\ f \neq 0}} \frac{\mathcal{E}(f, f)}{\langle f, f \rangle_\pi}.$$

For every $S \subseteq \Omega$, we define the expansion

$$\Phi(S) = \frac{\sum_{x \in S, y \in \Omega \setminus S} Q(x, y)}{\pi(S)} = \frac{\sum_{x \in S, y \in \Omega \setminus S} \pi(x)P(x, y)}{\sum_{x \in S} \pi(x)}.$$

Also the expansion of the chain P

$$\Phi_P \triangleq \min_{S \subseteq \Omega : \pi(S) \leq \frac{1}{2}} \Phi(S).$$

The Cheeger's inequality is then

$$\frac{\gamma}{2} \leq \Phi_P \leq \sqrt{2\gamma}.$$

2. Random Walk on Hypercubes

In this section, we study the eigenvalues and eigenvectors of the random walk on hypercubes. Last week, we introduced the following chain on $\{-1, 1\}^n$: each step when one is standing at a state $\mathbf{x} \in \{-1, 1\}^n$,

- choose an index $i \in [n]$ u.a.r. and $b \in \{-1, 1\}$ u.a.r.;
- change $x(i)$ to b.

We first look at the case when $n = 1$, so the transition matrix is $P_1 = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$. The two eigenvalues are 0 and 1 with corresponding eigenvectors $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\mathbf{1}$. Now we introduce the concept of *product chains*:

Let $P_1, P_2 \ldots P_t$ be t chains on spaces $\Omega_1, \Omega_2, \ldots, \Omega_t$ respectively. Consider the following chain defined on $\Omega \triangleq \Omega_1 \times \Omega_2 \cdots \times \Omega_t$: when one is standing at a point $\mathbf{x} \in \Omega$,

- choose an index $i \in [t]$ uniformly at random; and then
- perform a move in P_i.

If we use $P \in \mathbb{R}^{2^n \times 2^n}$ to denote the transition matrix of this product chain, then for every $\mathbf{x}, \mathbf{y} \in \{0,1\}^n$,

$$P(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \frac{1}{n} \cdot P(x_i, y_i) \cdot \prod_{j \neq i} \mathbf{1}[x_j = y_j].$$

Product chains enjoy the following properties

Proposition 5. *For every $1 \le i \le t$, if λ_i is a eigenvalue of P_i with corresponding eigenvector v_i, then*

- *$v_1 \otimes v_2 \otimes \cdots \otimes v_t$ is an eigenvector of P;*
- *$\frac{1}{n} \sum_{i=1}^{t} \lambda_i$ is an eigenvalue of P.*

We leave the proof of these properties as an exercise.

The random walk on an n-dim hypercube fits perfectly in the framework of product chains. Each P_i is the one-dim hypercube which we already understood. So the two eigenvectors of P_i, if viewed as eigenfunctions, are $f_1(x) = x$ and $f_2(x) = 1$. Therefore, if we use P to denote the transition matrix of this random walk, then its 2^n eigenfunctions are $f_S(\mathbf{x}) = \prod_{i \in S} x_i$, indexed by every $S \subseteq [n]$. Moreover, the eigenvalue of f_S is $\frac{n-|S|}{n}$. Therefore, the second largest eigenvalue of P is $\frac{n-1}{n}$ and its relaxation time is n.

Combining with Proposition 2, we have an $O(n^2)$ upper bound for the mixing time of the random walk on the n-dim hypercubes. This is worse than the bound we obtained using coupling argument last week. The reason is that Proposition 2 is not tight in this example.