

Rapid Mixing on Matroid Bases

Xiaoyu Chen

最近仔细看了一下 Nima Anari 在 Georia Tech 的讲座, Nima 在讲座上从人的正常思维的顺序给出了证明 rapid mixing 的思路, 感觉很受启发.

1 Preliminary

之前我学过的用于度量两个分布的距离的方法实际上可以被统一到同一个框架下:

Define 1 (f -divergence).

(1) $f : [0, +\infty) \rightarrow \mathbb{R}$ is convex

$$(2) D_f(\nu \parallel \mu) := \mathbb{E}_{S \sim \mu} \left[f\left(\frac{\nu(S)}{\mu(S)}\right) \right] - f\left(\mathbb{E}_{S \sim \mu} \left[\frac{\nu(S)}{\mu(S)} \right] \right)$$

Since f is convex, by Jensen's Inequality, we have

$$D_f(\nu \parallel \mu) \geq 0, \quad \forall \nu, \mu$$

定义好 f -divergence 之后, 立即就有:

Theorem 1 (Implied by Data Processing Inequality).

$$D_f(\nu P \parallel \mu P) \leq D_f(\nu \parallel \mu)$$

holds for any random walk operator P .

Proof. 个人感觉这个不等式是不是特别像 Data Processing Inequality, 反而直接通过 Jensen's Inequality 来证明比较简单. 首先发现不等式两边都有 $f(1)$, 我们可以直接去掉, 所以实际上我们要证明的东西是

$$\mathbb{E}_{S \sim \mu P} \left[f\left(\frac{\nu P(S)}{\mu P(S)}\right) \right] \leq \mathbb{E}_{S \sim \mu} \left[f\left(\frac{\nu(S)}{\mu(S)}\right) \right]$$

注意到当 S 固定时, $\frac{\nu P(S)}{\mu P(S)}$ 实际上可以看成在一个分布 π 上求 $\frac{\nu}{\mu}$ 的期望值

$$\begin{aligned} \frac{\nu P(S)}{\mu P(S)} &= \frac{\sum_x \nu(x) P(x, S)}{\mu P(S)} \\ &= \sum_x \frac{\mu(x) P(x, S)}{\mu P(S)} \frac{\nu(x)}{\mu(x)} \\ &= \sum_x \pi(x) \frac{\nu(x)}{\mu(x)} \end{aligned}$$

所以上面不等式的 LHS 可以写成:

$$\begin{aligned}
\mathbb{E}_{S \sim \mu P} \left[f\left(\frac{\nu P(S)}{\mu P(S)}\right) \right] &= \mathbb{E}_{S \sim \mu P} \left[f\left(\mathbb{E}_{x \sim \pi} \left[\frac{\nu(x)}{\mu(x)} \right] \right) \right] \\
&\leq \mathbb{E}_{S \sim \mu P} \left[\mathbb{E}_{x \sim \pi} \left[f\left(\frac{\nu(x)}{\mu(x)}\right) \right] \right], \quad \text{Jensen's Inequality} \\
&= \sum_S \mu P(S) \sum_x \frac{\mu(x) P(x, S)}{\mu P(S)} f\left(\frac{\nu(x)}{\mu(x)}\right) \\
&= \mathbb{E}_{x \sim \mu} \left[f\left(\frac{\nu(x)}{\mu(x)}\right) \right]
\end{aligned}$$

□

Fact 1.

- (1) $f(x) = \frac{1}{2}|x - 1|$ $D_f(\nu \parallel \mu)$ is the total variation distance of ν and μ .
- (2) $f(x) = x^2$ $D_f(\nu \parallel \mu)$ is the variance of $\frac{\nu}{\mu}$ w.r.t. μ .
- (3) $f(x) = x \log x$ $D_f(\nu \parallel \mu)$ is the KL-divergence of ν and μ .

Fact 2.

$$D_f(\nu \parallel \mu) = \mathbb{E}_{S \sim \mu} \left[f\left(\frac{\nu(S)}{\mu(S)}\right) \right] - f(1)$$

有些地方喜欢取满足 $f(1) = 0$ 的 f , 这样可以简化计算.

2 Proof of Rapid Mixing

记拟阵中大小为 i 的独立集的集合为 $\mathcal{M}(i)$. 记 $\mathcal{M}(i)$ 中的独立集的分布为 $\mu_i : \mathcal{M}(i) \rightarrow \mathbb{R}_{\geq 0}$. 记 r 为拟阵的秩, 则 μ_r 就是拟阵基的分布.

Fact 3 (Normalizing). $\forall x \in \mathcal{M}(k)$, $\mu_k(x) = \frac{1}{\binom{r}{k}} \sum_{\substack{y \in \mathcal{M}(r) \\ y \supseteq x}} \mu_r(y)$

通过 telescope sum, 我们可以将 $D_f(\nu_r \parallel \mu_r)$ 进行纵向的分解.

$$\begin{aligned}
D_f(\nu_r \parallel \mu_r) &= \mathbb{E}_{S \sim \mu_r} \left[f\left(\frac{\nu_r(S)}{\mu_r(S)}\right) \right] - f(1) \\
&= \sum_{i=1}^r \left(\mathbb{E}_{S \sim \mu_i} \left[f\left(\frac{\nu_i(S)}{\mu_i(S)}\right) \right] - \mathbb{E}_{S \sim \mu_{i-1}} \left[f\left(\frac{\nu_{i-1}(S)}{\mu_{i-1}(S)}\right) \right] \right)
\end{aligned}$$

方便起见, 不妨令

$$A_i = \underbrace{\mathbb{E}_{S \sim \mu_i} \left[f\left(\frac{\nu_i(S)}{\mu_i(S)}\right) \right]}_{a_i} - \underbrace{\mathbb{E}_{S \sim \mu_{i-1}} \left[f\left(\frac{\nu_{i-1}(S)}{\mu_{i-1}(S)}\right) \right]}_{a_{i-1}}$$

则上面的式子可以简写为

$$D_f(\nu_r \parallel \mu_r) = \sum_{i=1}^r A_i$$

Fact 4.

$$D_f(\nu_{r-1} \parallel \mu_{r-1}) = D_f(\nu_r P_r^\downarrow \parallel \mu_r P_r^\downarrow)$$

Observation 1. If $A_i \geq A_{i-1}, i = 2, \dots, r$, then

$$D_f(\nu_r P_r^\downarrow \parallel \mu_r P_r^\downarrow) \leq \left(1 - \frac{1}{r}\right) D_f(\nu_r \parallel \mu_r)$$

where P_r^\downarrow represent the random walk which remove 1 element u.a.r.

Proof. 显然, $D_f(\nu_r \parallel \mu_r) - D_f(\nu_r P^\downarrow \parallel \mu_r P^\downarrow) = A_r$. 然后, 因为 $A_i \geq A_{i-1}, i = 2, \dots, r$, 所以

$$A_r \geq \frac{1}{r} \sum_{i=1}^r A_i = \frac{1}{r} D_f(\nu_r \parallel \mu_r)$$

所以

$$\begin{aligned} D_f(\nu_r P^\downarrow \parallel \mu_r P^\downarrow) &= D_f(\nu_r \parallel \mu_r) - A_r \\ &\leq D_f(\nu_r \parallel \mu_r) - \frac{1}{r} D_f(\nu_r \parallel \mu_r) \\ &= (1 - \frac{1}{r}) D_f(\nu_r \parallel \mu_r) \end{aligned}$$

□

不难发现, 只要证明 $A_i \geq A_{i-1}$, 就能够证明 rapid mixing, 得到有意义的 mixing time.

3 Prove $A_i \geq A_{i-1}$ when $[f = x \log x]$

$$\begin{aligned} A_i &\geq A_{i-1} \\ a_i - a_{i-1} &\geq a_{i-1} - a_{i-2} \end{aligned}$$

所以, 我们只需证明 $a_i + a_{i-2} \geq 2a_{i-1}$ 即可. 故只需证明:

$$\mathbb{E}_{\mu_i} \left[f\left(\frac{\nu_i}{\mu_i}\right) \right] + \mathbb{E}_{\mu_{i-2}} \left[f\left(\frac{\nu_{i-2}}{\mu_{i-2}}\right) \right] \geq 2 \mathbb{E}_{\mu_{i-1}} \left[f\left(\frac{\nu_{i-1}}{\mu_{i-1}}\right) \right]$$

因为

$$\mathbb{E}_{\mu_{i-2}} \left[f\left(\frac{\nu_{i-2}}{\mu_{i-2}}\right) \right] \geq f(1) = 0$$

所以实际上只用证明

$$\mathbb{E}_{\mu_i} \left[f\left(\frac{\nu_i}{\mu_i}\right) \right] \geq 2 \mathbb{E}_{\mu_{i-1}} \left[f\left(\frac{\nu_{i-1}}{\mu_{i-1}}\right) \right]$$

证明这个式子通常使用的方法都是将其在水平方向上进行分解.

$$\begin{aligned} \sum_{J \in \mathcal{M}(i)} \mu_i(J) f\left(\frac{\nu_i(J)}{\mu_i(J)}\right) &\geq 2 \sum_{J \in \mathcal{M}(i-1)} \mu_{i-1}(J) f\left(\frac{\nu_{i-1}(J)}{\mu_{i-1}(J)}\right) \\ \underbrace{\sum_{J \in \mathcal{M}(i)} \nu_i(J) \log \frac{\nu_i(J)}{\mu_i(J)}}_{(1)} &\geq 2 \underbrace{\sum_{J \in \mathcal{M}(i-1)} \nu_{i-1}(J) \log \frac{\nu_{i-1}(J)}{\mu_{i-1}(J)}}_{(2)} \end{aligned}$$

$$\begin{aligned}
(1) &= \sum_{J \in \mathcal{M}(i)} \frac{1}{\binom{i}{2}} \sum_{\substack{I \in \mathcal{M}(i-2) \\ I \subseteq J}} \nu_i(J) \log \frac{\nu_i(J)}{\mu_i(J)}, \quad I \subseteq J \wedge I \in \mathcal{M}(i-2) \text{ 这样的 } I \text{ 有 } \binom{i}{2} \text{ 个} \\
&= \sum_{I \in \mathcal{M}(i-2)} \sum_{\substack{J \in \mathcal{M}(i) \\ J \supseteq I}} \frac{1}{\binom{i}{2}} \nu_i(J) \log \frac{\nu_i(J)}{\mu_i(J)} \\
&= 2 \sum_{I \in \mathcal{M}(i-2)} \nu_{i-2}(I) \sum_{\substack{\{a,b\} \in \mathcal{M}_I(2) \\ J=I \cup \{a,b\}}} \nu_{I,2}(\{a,b\}) \log \frac{\nu_i(J)}{\mu_i(J)}, \quad \nu_{i-2}(I) = \frac{(i-2)!}{i!} \sum_{\substack{J \in \mathcal{M}(i) \\ J \supseteq I}} \nu_J(J) \\
&= 2 \sum_{I \in \mathcal{M}(i-2)} \nu_{i-2}(I) \sum_{\substack{\{a,b\} \in \mathcal{M}_I(2) \\ J=I \cup \{a,b\}}} \nu_{I,2}(\{a,b\}) \left[\log \frac{\nu_{I,2}(\{a,b\})}{\mu_{I,2}(\{a,b\})} + \log \frac{\nu_{i-2}(I)}{\mu_{i-2}(I)} \right] \\
&= 2 \sum_{I \in \mathcal{M}(i-2)} \nu_{i-2}(I) \sum_{\substack{\{a,b\} \in \mathcal{M}_I(2) \\ J=I \cup \{a,b\}}} \nu_{I,2}(\{a,b\}) \log \frac{\nu_{I,2}(\{a,b\})}{\mu_{I,2}(\{a,b\})} + 2 \sum_{I \in \mathcal{M}(i-2)} \nu_{i-2}(I) \log \frac{\nu_{i-2}(I)}{\mu_{i-2}(I)} \\
(2) &= 2 \sum_{J \in \mathcal{M}(i-1)} \frac{1}{i-1} \sum_{\substack{I \in \mathcal{M}(i-2) \\ I \subseteq J}} \nu_{i-1}(J) \log \frac{\nu_{i-1}(J)}{\mu_{i-1}(J)}, \quad I \subseteq J \wedge I \in \mathcal{M}(i-2) \text{ 这样的 } I \text{ 有 } i-1 \text{ 个} \\
&= 2 \sum_{I \in \mathcal{M}(i-2)} \sum_{\substack{J \in \mathcal{M}(i-1) \\ J \supseteq I}} \frac{1}{i-1} \nu_{i-1}(J) \log \frac{\nu_{i-1}(J)}{\mu_{i-1}(J)} \\
&= 2 \sum_{I \in \mathcal{M}(i-2)} \nu_{i-2}(I) \sum_{\substack{a \in \mathcal{M}_I(1) \\ J=I \cup \{a\}}} \nu_{I,1}(\{a\}) \log \frac{\nu_{i-1}(J)}{\mu_{i-1}(J)}, \quad \nu_{i-2}(I) = \frac{(i-2)!}{(i-1)!} \sum_{\substack{J \in \mathcal{M}(i-1) \\ J \supseteq I}} \nu_{i-1}(J) \\
&= 2 \sum_{I \in \mathcal{M}(i-2)} \nu_{i-2}(I) \sum_{\substack{a \in \mathcal{M}_I(1) \\ J=I \cup \{a\}}} \nu_{I,1}(\{a\}) \left[\log \frac{\nu_{I,1}(\{a\})}{\mu_{I,1}(\{a\})} + \log \frac{\nu_{i-2}(I)}{\mu_{i-2}(I)} \right] \\
&= 2 \sum_{I \in \mathcal{M}(i-2)} \nu_{i-2}(I) \sum_{\substack{a \in \mathcal{M}_I(1) \\ J=I \cup \{a\}}} \nu_{I,1}(\{a\}) \log \frac{\nu_{I,1}(\{a\})}{\mu_{I,1}(\{a\})} + 2 \sum_{I \in \mathcal{M}(i-2)} \nu_{i-2}(I) \log \frac{\nu_{i-2}(I)}{\mu_{i-2}(I)}
\end{aligned}$$

比较 (1), (2), 实际上, 最终我们只需要证明 $\forall I \in \mathcal{M}(i-2)$ 都有:

$$\sum_{\{a,b\} \in \mathcal{M}_I(2)} \nu_{I,2}(\{a,b\}) \log \frac{\nu_{I,2}(\{a,b\})}{\mu_{I,2}(\{a,b\})} \geq \sum_{a \in \mathcal{M}_I(1)} \nu_{I,1}(\{a\}) \log \frac{\nu_{I,1}(\{a\})}{\mu_{I,1}(\{a\})}$$

Intuitively, 我们将原来要证明的期望不等式分解到了每一个 $I \in \mathcal{M}(i-2)$ 上 (\mathcal{M}_I 的 1, 2 层上). 这样分解的原因是:

$$\begin{aligned}
\mu_{I,2}(\{a,b\}) &\propto \nabla^2 \partial_I g_\mu(a,b) \\
\mu_{I,1}(\{a\}) &\propto \nabla \partial_I g_\mu(a)
\end{aligned}$$

这样分解可以方便我们将 $\partial_I g_\mu$ 是 log-concave 的这个性质拿过来用.

下面继续来证明上面的式子, 将右侧放到 2 倍可以得到一个更紧的不等式, 这个不等式要好证明一些.

$$\sum_{\{a,b\} \in \mathcal{M}_I(2)} \nu_{I,2}(\{a,b\}) \log \frac{\nu_{I,2}(\{a,b\})}{\mu_{I,2}(\{a,b\})} \geq 2 \sum_{a \in \mathcal{M}_I(1)} \nu_{I,1}(\{a\}) \log \frac{\nu_{I,1}(\{a\})}{\mu_{I,1}(\{a\})}$$

观察可以发现:

$$\nu_{I,1}(\{a\}) = \frac{1}{2} \sum_{\{a,b\} \in \mathcal{M}_I(2)} \nu_{I,2}(\{a,b\})$$

所以上面的式子可以继续变形:

$$\begin{aligned}
\sum_{\{a,b\} \in \mathcal{M}_I(2)} \nu_{I,2}(\{a,b\}) \log \frac{\nu_{I,2}(\{a,b\})}{\mu_{I,2}(\{a,b\})} &\geq \sum_{a \in \mathcal{M}_I(1)} \sum_{\{a,b\} \in \mathcal{M}_I(2)} \nu_{I,2}(\{a,b\}) \log \frac{\nu_{I,1}(\{a\})}{\mu_{I,1}(\{a\})} \\
&\geq \sum_{\{a,b\} \in \mathcal{M}_I(2)} \nu_{I,2}(\{a,b\}) \sum_{a \in \{a,b\}} \log \frac{\nu_{I,1}(\{a\})}{\mu_{I,1}(\{a\})} \\
&\geq \sum_{\{a,b\} \in \mathcal{M}_I(2)} \nu_{I,2}(\{a,b\}) \log \left[\frac{\nu_{I,1}(\{a\})}{\mu_{I,1}(\{a\})} \cdot \frac{\nu_{I,1}(\{b\})}{\mu_{I,1}(\{b\})} \right]
\end{aligned}$$

所以, 我们只需要证明:

$$\begin{aligned}
&\sum_{\{a,b\} \in \mathcal{M}_I(2)} \nu_{I,2}(\{a,b\}) \left(\log \frac{\nu_{I,2}(\{a,b\})}{\mu_{I,2}(\{a,b\})} - \log \left[\frac{\nu_{I,1}(\{a\})}{\mu_{I,1}(\{a\})} \cdot \frac{\nu_{I,1}(\{b\})}{\mu_{I,1}(\{b\})} \right] \right) \geq 0 \\
&\sum_{\{a,b\} \in \mathcal{M}_I(2)} \mu_{I,2}(\{a,b\}) \frac{\nu_{I,2}(\{a,b\})}{\mu_{I,2}(\{a,b\})} \left(\log \frac{\nu_{I,2}(\{a,b\})}{\mu_{I,2}(\{a,b\})} - \log \left[\frac{\nu_{I,1}(\{a\})}{\mu_{I,1}(\{a\})} \cdot \frac{\nu_{I,1}(\{b\})}{\mu_{I,1}(\{b\})} \right] \right) \geq 0
\end{aligned}$$

因为 $a \log \frac{a}{b} \geq a - b$, 所以上面的式子可以化成

$$\begin{aligned}
&\sum_{\{a,b\} \in \mathcal{M}_I(2)} \mu_{I,2}(\{a,b\}) \left(\frac{\nu_{I,2}(\{a,b\})}{\mu_{I,2}(\{a,b\})} - \frac{\nu_{I,1}(\{a\})}{\mu_{I,1}(\{a\})} \cdot \frac{\nu_{I,1}(\{b\})}{\mu_{I,1}(\{b\})} \right) \geq 0 \\
&\sum_{\{a,b\} \in \mathcal{M}_I(2)} \mu_{I,2}(\{a,b\}) \frac{\nu_{I,1}(\{a\})}{\mu_{I,1}(\{a\})} \cdot \frac{\nu_{I,1}(\{b\})}{\mu_{I,1}(\{b\})} \leq 1
\end{aligned}$$

令 $W_I(a,b) = \frac{1}{2} \mu_{I,2}(\{a,b\})$, 我们可以发现 $W_I \propto \nabla^2 \partial_I g_\mu$. 不妨令 $W_I = \gamma \nabla^2 \partial_I g_\mu$.

Fact 5. if g_μ is log-concave, then W_I has at most one positive eigenvalue.

Proof. 首先不难发现 $\nabla^2 \partial_I g_\mu$ 最多只有一个正特征值. 又因为

$$\begin{aligned}
\text{Det}(\nabla^2 \partial_I g_\mu - \lambda I) &= 0 \\
\text{Det}(\gamma \nabla^2 \partial_I g_\mu - \gamma \lambda I) &= 0 \\
\text{Det}(W_I - \gamma \lambda I) &= 0
\end{aligned}$$

所以其实 W_I 的特征方程和 $\nabla^2 \partial_I g_\mu$ 的特征方程之间只差一个常数, 所以 W_I 当然也最多只有一个正特征值. □

所以上面的式子可以进一步写成

$$x^T W_I x \leq 1, \quad \text{where, } x = \frac{v_{I,1}}{\mu_{I,1}}.$$

不妨假设 $W_I \in \mathbb{R}_{\geq 0}^{n \times n}$ 是一个 $n \times n$ 的矩阵 [这个矩阵的大小不影响后面的逻辑]. 不难发现:

$$\sum_{j=1}^n W_I(i,j) = \mu_{I,1}(\{i\}) = \sum_{j=1}^n W_I(j,i)$$

Lemma 1. 如果 $M \in \mathbb{R}^{n \times n}$ 是一个对称矩阵, 且 M 最多只有一个正特征值. 则对于任意的 $x, v \in \mathbb{R}^n$, 都有

$$(x^T M x)(v^T M v) \leq (x^T M v)^2$$

Proof. 令 $S = [x|v]$ 则

$$\begin{aligned}
A = S^T M S &= \begin{bmatrix} x^T \\ v^T \end{bmatrix} M \begin{bmatrix} x \\ v \end{bmatrix} \\
&= \begin{bmatrix} x^T M x & x^T M v \\ v^T M x & v^T M v \end{bmatrix}
\end{aligned}$$

因为 M 最多有一个正特征值, 所以 A 最多有一个正特征值. 又因为 A 是一个正项矩阵, 所以 A 一定有一个正特征值. 所以 A 的另外一个特征值 ≤ 0 . 所以

$$\lambda_1 \lambda_2 = \text{Det}(A) \leq 0$$

$$(x^T M x)(v^T M v) \leq (x^T M v)^2$$

□

所以

$$(x^T W_I x)(\mathbf{1}^T W_I \mathbf{1}) \leq (x^T M \mathbf{1})^2$$

$$x^T W_I x \leq 1$$