Refined Understanding: Spectrum, Coupling and ℓ^p-distance

Xiaoyu Chen

1 Preliminaries

Let P be a time reversible Markov chain defined on the state space Ω with its stationary distribution π.

Definition 1.1. *For any $f, g \in \mathbb{R}^\Omega$, let:*

$$\langle f, g\rangle_\pi \triangleq \sum_{x\in\Omega} \pi(x) f(x) g(x)$$

Definition 1.2 (ℓ_p-norm)**.** *For any $f : \Omega \to \mathbb{R}$, we have:*

$$\| f \|_{\pi,p} \triangleq \begin{cases} \left(\sum_{x\in\Omega} \pi(x)|f(x)|^p\right)^{1/p} & 1 \le p < \infty \\ \max_x |f(x)| & o.w. \end{cases}$$

Fact 1.1. *For any $f : \Omega \to \mathbb{R}$ we have*

$$\| f \|_{\pi,1} \le \| f \|_{\pi,2} \le \cdots \le \| f \|_{\pi,\infty}$$

Proof. Recall that we have Jenson's Inequality for concave function g, such that

$$\mathbb{E}[g(x)] \le g(\mathbb{E}[x])$$

Then for any $p < r$, note that $x \mapsto x^{\frac{p}{r}}$ is a concave function, we have

$$\| f \|_{\pi,p} = (\mathbb{E}_{x\sim\pi}[|f(x)|^p])^{1/p} = (\mathbb{E}_{x\sim\pi}[(|f(x)|^r)^{\frac{p}{r}}])^{1/p} \le ((\mathbb{E}_{x\sim\pi}[|f(x)|^r])^{\frac{p}{r}})^{1/p} = (\mathbb{E}_{x\sim\pi}[|f(x)|^r])^{1/r} = \| f \|_{\pi,r}$$

□

Fact 1.2. *For $c \ge 0$, we have $\| c \cdot f \|_{\pi,p} = c \| f \|_{\pi,p}$*

Definition 1.3 (self-adjoint operator)**.** *We say P is a self-adjoint operator over $\langle\cdot,\cdot\rangle_\pi$ if for $\forall f, g : \Omega \to \mathbb{R}$, we have $\langle f, Pg\rangle_\pi = \langle Pf, g\rangle_\pi$.*

Definition 1.4.

$$D \triangleq \mathrm{diag}(\pi)$$

Definition 1.5. *For any $x \in \Omega$ we have its indicator function $\delta_x : \Omega \to \{0,1\}$ defined as:*

$$\delta_x(y) \triangleq \begin{cases} 1 & y = x \\ 0 & y \ne x \end{cases}$$

Definition 1.6. *$q_t \in \mathbb{R}^{\Omega\times\Omega}$ is a matrix that is widely used in measure the distance between the current distribution with the stationary distribution, which is defined as $q_t \triangleq P^t D^{-1}$ and thus:*

$$q_t(x,y) = \frac{P^t(x,y)}{\pi(y)}$$

Moreover, let $q_t^x \triangleq q_t(x,\cdot)$ be the x-th row of q_t.

2 Spectral Decomposition

Fact 2.1. *P is a self-adjoint operator of $\langle\cdot,\cdot\rangle_\pi$ iff P is time reversible.*

Proof. $\Rightarrow$: If P is a self-adjoint operator of $\langle\cdot,\cdot\rangle_\pi$, then for any $x,y\in\Omega$, we have

$$\langle\delta_x, P\delta_y\rangle_\pi = \langle P\delta_x,\delta_y\rangle_\pi = \langle\delta_y, P\delta_x\rangle_\pi$$

. So, we have $\pi(x)P(x,y)=\pi(y)P(y,x)$.

$\Leftarrow$: Could be verified by simple calculation. □

Theorem 2.1. *P is a self-adjoint operator of $\langle\cdot,\cdot\rangle_\pi$ iff P has eigenbasis $\{f_i\}_{i=1}^{|\Omega|}$ on the inner product space $(\mathbb{R}^\Omega,\langle\cdot,\cdot\rangle_\pi)$.*

Proof. $\Rightarrow$: Since $\pi(x)P(x,y)=\pi(y)P(y,x)$, we have $\pi^{1/2}(x)P(x,y)\pi^{-1/2}(y)=\pi^{1/2}P(y,x)\pi^{-1/2}(x)$. And it turns out that if $D\triangleq\mathrm{diag}(\pi)$ then:

$$D^{1/2}PD^{-1/2}\text{ is a symmetric matrix}$$

So, $D^{1/2}PD^{-1/2}$ have eigenbasis $\{g_i\}_{i=1}^{|\Omega|}$ over $(\mathbb{R}^\Omega,\langle\cdot,\cdot\rangle)$. And it is suffice to show that $\{D^{-1/2}g_i\}_{i=1}^{|\Omega|}$ are eigenfunctions of P and they are orthogonal according to $\langle\cdot,\cdot\rangle_\pi$. Just note that

- (eigenfunciton) $D^{1/2}PD^{-1/2}g_i=\lambda_i g_i\Rightarrow pD^{-1/2}g_i=\lambda_i D^{-1/2}g_i$.
- (orghognoal) for $i\neq j$, $\langle D^{-1/2}g_i, D^{-1/2}g_j\rangle_\pi=\langle g_i,g_j\rangle=0$

($\Leftarrow$: Omit, may be added in the future.) □

Theorem 2.2. *If P has eigenbasis $\{f_i\}_{i=1}^{|\Omega|}$ on the inner product space $(\mathbb{R}^\Omega,\langle\cdot,\cdot\rangle_\pi)$, then $P=\sum_{i=1}^{|\Omega|}\lambda_i f_i f_i^T D$*

Proof. Its suffice to prove that $P(x,y)=\sum_{i=1}^{|\Omega|}\lambda_i f_i(x)f_i(y)\pi(y)$. Note that $P\delta_y(x)=P(x,y)$. And we have $\delta_y=\sum_{i=1}^{|\Omega|}\langle f_i,\delta_y\rangle_\pi f_i=\sum_{i=1}^{|\Omega|}\pi(y)f_i(y)f_i$. So, $P\delta_y=\sum_{i=1}^{|\Omega|}\lambda_i\pi(y)f_i(y)f_i$. So we have

$$P(x,y)=P\delta_y(x)=\sum_{i=1}^{|\Omega|}\lambda_i f_i(x)f_i(y)\pi(y)$$ □

Colorllary 2.1. *If P has eigenbasis $\{f_i\}_{i=1}^{|\Omega|}$ on the inner product space $(\mathbb{R}^\Omega,\langle\cdot,\cdot\rangle_\pi)$, then $P^t=\sum_{i=1}^{|\Omega|}\lambda_i^t f_i f_i^T D$ and thus $q_t=P^tD^{-1}=\sum_{i=1}^{|\Omega|}\lambda_i^t f_i f_i^T$.*

3 Distances

Definition 3.1.

$$d(t)\triangleq\max_{x\in\Omega}\parallel P^t(x,\cdot)-\pi\parallel_{TV}$$

Definition 3.2 (ℓ^p-distances)**.**

$$d^{(p)}(t)\triangleq\max_{x\in\Omega}\parallel q_t^x-1\parallel_{\pi,p}$$

Colorllary 3.1 (from Fact 1.1)**.** $2d(t)=d^{(1)}(t)\le d^{(2)}(t)\le\cdots\le d^{(\infty)}(t)$

Fact 3.1. $d^{(\infty)}(2t)=\left[d^{(2)}(t)\right]^2=\max_{x\in\Omega}q_{2t}(x,x)-1=\max_{x\in\Omega}\langle q_t^x-1,q_t^x-1\rangle_\pi$

Proof. Background: $\langle q_t^x,q_t^y\rangle_\pi=\sum_{z\in\Omega}\pi(z)\frac{P^t(x,z)}{\pi(z)}\frac{P^t(y,z)}{\pi(z)}=\sum_{z\in\Omega}\frac{P^t(x,z)P^t(z,y)}{\pi(y)}=\frac{P^{2t}(x,y)}{\pi(y)}=q_{2t}(x,y)$ And thus we have $\langle q_t^x-1,q_t^y-1\rangle_\pi=q_{2t}(x,y)-1$.

($d^{(\infty)}(2t)=\max_{x\in\Omega}|q_{2t}(x,x)-1|$): By Cauchy-Schwarz Inequality, we have

$$\langle q_t^x-1,q_t^y-1\rangle_\pi\le\sqrt{\langle q_t^x-1,q_t^x-1\rangle_\pi\langle q_t^y-1,q_t^y-1\rangle_\pi}\le\max_{\theta\in\{x,y\}}\langle q_t^\theta-1,q_t^\theta-1\rangle_\pi$$

Thus, $d^{(\infty)}(2t)=\max_{x,y\in\Omega}q_{2t}(x,y)-1=\max_{x,y\in\Omega}\langle q_t^x-1,q_t^y-1\rangle_\pi=\max_{x\in\Omega}\langle q_t^x-1,q_t^x-1\rangle_\pi$

($\left[d^{(2)}(t)\right]^2=\max_{x\in\Omega}|q_{2t}(x,x)-1|$): $[d^{(2)}(t)]^2=\max_{x\in\Omega}\parallel q_t^x-1\parallel_{\pi,2}^2=\max_{x\in\Omega}\langle q_t^x-1,q_t^x-1\rangle_\pi$ □

Fact 3.2 (submultiplicative)**.** $d^{(p)}(s+t)\le d^{(p)}(s)d^{(p)}(t)$

4 The Relationship Between Distances and Spectrum

Fact 4.1. *Since $\lambda_1 = 1$ and $f_1 = \mathbb{1}$, we have:*

$$d^{(p)}(t) = \max_{x\in\Omega} \| q_t^x - \mathbb{1}^T \|_{\pi,p} = \max_{x\in\Omega} \| \sum_{i=1}^{|\Omega|} \lambda_i^t f_i(x) f_i^T - \mathbb{1}^T \|_{\pi,p} = \max_{x\in\Omega} \| \sum_{i=2}^{|\Omega|} \lambda_i^t f_i(x) f_i^T \|_{\pi,p}$$

Definition 4.1. *Let $\lambda_\star \triangleq \{|\lambda| : \lambda$ is an eigenvalue of P but $\lambda \neq 1\}$. **absolute spectral gap** is defined as $\gamma_\star \triangleq 1-\lambda_\star$. If we sort all the eigenvalues in the order $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{|\Omega|} \geq -1$. **spectral gap** is defined as $\gamma \triangleq 1 - \gamma_2$. Moreover, the **relaxation time** is defined as $\frac{1}{\gamma_\star}$*

Fact 4.2. $d^{(p)}(t) \leq \lambda_\star^{t-1} d^{(p)}(1)$

Proof. $d^{(p)}(t) = \max_{x\in\Omega} \| \sum_{i=2}^{|\Omega|} \lambda_i^t f_i(x) f_i^T \|_{\pi,p} \leq \lambda_\star^{t-1} \max_{x\in\Omega} \| \sum_{i=2}^{|\Omega|} \lambda_i f_i(x) f_i^T \|_{\pi,p} = \lambda_\star^{t-1} d^{(p)}(1)$ □

Fact 4.3. $\lambda_\star^t \leq d^{(1)}(t)$

Proof. Let f be any eigenfunctions of P that satisfies $f \perp_\pi \mathbb{1}$, and λ be its corresponding eigenvalue and thus we have $\langle f, \mathbb{1}\rangle_\pi = \mathbb{E}[f] = 0$ Then for $\forall x \in \Omega$:

$$|\lambda^t f(x)| = |P^t f(x)| = |\sum_{y\in\Omega} P(x,y) f(y)| = |\sum_{y\in\Omega} P(x,y)f(y) - \pi(y)f(y)| \leq \| f \|_\infty d^{(1)}(t)$$

Let $x_\star \triangleq \text{argmax}_x |f(x)|$, then we have

$$|\lambda^t||f(x_\star)| \leq |f(x_\star)| d^{(1)}(t) \Rightarrow |\lambda|^t \leq d^{(1)}(t) \Rightarrow \lambda_\star^t \leq d^{(1)}(t)$$ □

Theorem 4.1. $\lim_{t\to\infty} \left[d^{(p)}(t)\right]^{1/t} = \lambda_\star$

Proof.

$$\begin{aligned} \lambda_\star^t &\leq d^{(p)}(t) \leq \lambda_\star^{t-1} d^{(p)}(1) \\ \lambda_\star &\leq d^{(p)}(t)^{1/t} \leq \lambda_\star^{\frac{t-1}{t}} d^{(p)}(1)^{1/t} \\ \lim_{t\to\infty} \lambda_\star &\leq \lim_{t\to\infty} d^{(p)}(t)^{1/t} \leq \lim_{t\to\infty} \lambda_\star^{\frac{t-1}{t}} \cdot \lim_{t\to\infty} d^{(p)}(1)^{1/t} \\ \lambda_\star &\leq \lim_{t\to\infty} d^{(p)}(t)^{1/t} \leq \lambda_\star \end{aligned}$$ □

5 The Relationship Between Coupling and Spectrum

TODO: [M. F. Chen (98)] and its extension.