Mixing time of Markov chains

Jiwoon Park

January 2019

Will follow the lecture notes by Nathanaël Berestycki, Ch 1 - 4. (use pdf from from the lecturer's website)
Primary reference would be from Liven-Perres, "Markov chains and mixing times"
Also useful : Montenagro-Tetali, "mathematical aspects of mixing times". This text is very analytical.

(17th January, Thursday)

1 Preliminary

Mixing times is at the cross road of probability theory, analyis, geometry, statistical mechanics with ties to other fields such as representations theory and theoretical computer science. We will not focus on the application side of the theory, but on the theoeretical side.

$P = (p(x,y))_{x,y}$ is **reversible** with respect to the sationary distribution π if $\forall x, y$, $\pi(x)p(x,y) = \pi(y)p(x,y)$, i.e. satisfies the detailed balance equation. This condition is equivalent to

(i) $P = P^*$ with respect to $\langle f, g\rangle_\pi = \pi(fg)$, where $\pi(h) = \sum_x \pi(x)h(x)$, i.e. $\langle Pf, g\rangle_\pi = \langle f, Pg\rangle_\pi$ for all f, g.

(ii) P is a weighted random walk of $G = (V, E)$ for some graph : assign to each $xy \in E$ a symmetric edge weight $c(x,y) = c(y,x)$. Then $p(x,y) = c(x,y)/\sum_z c(x,z)$. Then for $\pi(x) = c(x)/\sum_y c(y)$, has $\pi(x)p(x,y) = \pi(y)p(y,x)$.

If reversibility is already satisfied, let $c(x,y) = \pi(x)p(x,y) = c(y,x)$. Then $c(x,y)/\sum_z c(x,z) = \pi(x)p(x,y)/\pi(x) = p(x,y)$, so we have the equivalence.

Definition 1.1.) The **total-variation distance of** μ, ν, distributions on the state space S, is defined by

$$\left\|\mu - \nu\right\|_{TV} = \max_{A\subset S} \mu(A) - \nu(A)$$

Lemma 1.1) $\left\|\mu - \nu\right\|_{TV} = \frac{1}{2}\sum_x |\mu(x) - \nu(x)| = \sum_{x:\mu(x)>\nu(x)} |\mu(x) - \nu(x)|$.

proof) Let $A\mu = \{x : \mu(x) > \nu(x)\}$ and $A\nu$ be similar. Then

$$\mu(A) - \nu(A) \le \mu(A \cap A\mu) - \nu(A \cap A\mu) \le \mu(A\mu) - \nu(A\mu)$$

with equality if $A = A\mu$ and because $\mu(\{\mu(x) = \nu(x)\}) = \nu(\{\mu(x) = \nu(x)\})$,

$$\mu(A\mu) - \nu(A\mu) = \nu(A\nu) - \mu(A\nu)$$

so

$$\begin{aligned}\|\mu - \nu\|_{TV} &= \frac{1}{2}(\mu(A\mu) - \nu(A\mu) + \nu(A\nu) - \mu(A\nu)) \\ &= \frac{1}{2}\sum_x |\mu(x) - \nu(x)|\end{aligned}$$

(End of proof) $\square$

Under irreducibility and aperiodicity condition, we have

$$p^t(x,y) \to \pi(y) \quad t \to \infty, \ \forall (x,y)$$

(Given $|S|$ is finite, such π exists and is unique)

- **ϵ-TV mixing time** is defined by

  $$\text{tmix}(\epsilon) = \inf\{t : d(t) \leq \epsilon\}$$

  where $d(t) = \max_x \|p^t(x,\cdot) - \pi\|_{TV}$. Also let tmix = tmix(1/4). Soon, we will show that

  $$\text{tmix}(\epsilon^k) \leq \text{tmix}(\epsilon/2) \cdot k$$

- **Claim :** $d(t)$ is non-increasig in t.

  proof)

  $$\begin{aligned}\frac{1}{2}\sum_y |p^{t+s}(x,y) - \pi(y)| &= \frac{1}{2}\sum_y |\sum_z p^t(x,z)p^s(z,y) - \pi(z)p^s(z,y)| \\ &\leq \frac{1}{2}\sum_z |p^t(x,z) - \pi(z)|\end{aligned}$$

Definition) Let $(X^n)_n$ be a family of Markov Chains. Write $d_n(t)$ for $d(t)$ with respect to X^n. We say that **cutoff** occurs at time t_n if

$$\forall \epsilon > 0, d_n((1-\epsilon)t_n) \to 1, d_n((1+\epsilon)t_n) \to 0 \quad \text{as } x \to \infty$$

Equivalently, $\text{tmix}^n(\epsilon) \sim t_n$, where $a_n \sim b_n$ (reversibility)iff $a_n/b_n \to 1$.

Definition 1.4) A **coupling** of probability measures μ and ν is a realization (X,Y) on the same probability space such that $X \sim \mu$ and $Y \sim \nu$.

Example : Consider $\mu = \nu$. Then may choose (i) $X = Y$ or (ii) X, Y independent. So we have complete freedom to of choice whether X and Y are dependent or not.

Theorem 1.1) $\|\mu - \nu\|_{TV} = \inf\{\mathbb{P}(X \neq Y) : (X,Y) \text{ a coupling of } \mu, \nu\}$. The coupling with this infimum is called the **optimal coupling.**

proof) For any coupling (X, Y) of μ and ν, we have

$$\mu(A) - \nu(A) = \mathbb{P}(X \in A) - \mathbb{P}(Y \in A) \leq \mathbb{P}(X \neq Y)$$

For the converse direction, take Z, X', Y', W independently with

$$\begin{aligned}
&Z \sim \text{Ber}(\|\mu - \nu\|_{TV}) \\
&W \text{ takes value } x \text{ with probability } \mu(x) \wedge \nu(x)/p \\
&X' \text{ takes value } x \text{ with probability } (\mu(x) - \nu(x)) + \|\mu - \nu\|_{TV} \\
&Y' \text{ takes value } y \text{ with probability } (\nu(y) - \mu(y)) + \|\mu - \nu\|_{TV}
\end{aligned}$$

where

$$\begin{aligned}
p &= \sum_z \mu(z) \wedge \nu(z) = \sum_y \nu(y) - \sum_y (\nu(y) - \mu(y))_+ \\
&= 1 - \|\mu - \nu\|_{TV}
\end{aligned}$$

Now set $X = ZX' + (1 - Z)W$, $Y = ZY' + (1 - Z)W$. Then X' and Y' have disjoint supports, so $\mathbb{P}(X' = Y') = 0$ and

$$\begin{aligned}
\mathbb{P}(X \neq Y) &= \mathbb{P}(Z = 1) = \|\mu - \nu\|_{TV} \\
\mathbb{P}(X = x) &= \mathbb{P}(Z = 1)\mathbb{P}(X' = x) + \mathbb{P}(Z = 0)\mathbb{P}(W = x) \\
&= \mathbb{P}(Z = 1)\frac{(\mu(x) - \nu(x))_+}{\mathbb{P}(Z = 1)} + \mathbb{P}(Z = 0)\frac{\mu(x) \wedge \nu(x)}{\mathbb{P}(Z = 0)} = \mu(x) \\
\mathbb{P}(Y = y) &= \nu(y)
\end{aligned}$$

so X, Y satisfy the desired eqaulity.

(End of proof) $\square$

Proposition 1.2) Let $p(t) = \max_{x,y \in S}\|p^t(x, \cdot) - p^t(y, \cdot)\|_{TV}$. Then

$$\begin{aligned}
&d(t) \leq p(t) \leq 2d(t) \\
&p(t + s) \leq p(s)p(t) \qquad \forall t, s \geq 0
\end{aligned}$$

Idea : let $A = \{f : \pi(f) = 0\}$, $p^* : A \to A$ the adjoint of p. Check $p(t) = \max_{f \in A}\|(p^*)^t f\|_1$, $\|f\|_1 = 2$, $\|h\|_1 = \pi(|h|)$.

(22nd Jaunary, Tuesday)

Proposition 1.2) Let $p(t) = \max_{x,y \in S}\|p^t(x, \cdot) - p^t(y, \cdot)\|_{TV}$. Then

$$\begin{aligned}
&d(t) \leq p(t) \leq 2d(t) \\
&p(t + s) \leq p(s)p(t) \qquad \forall t, s \geq 0
\end{aligned}$$

proof)

1. Let $A = A(x) \subset S$ be such that $\left\|p^t(x,\cdot) - \pi\right\|_{TV}$. Then $d(t) \leq p(t)$ comes from

$$\begin{aligned}\left\|p^t(x,\cdot) - \pi\right\|_{TV} &= p^t(x,A) - \pi(A) = \sum_z \pi(z)(p^t(x,A) - p^t(z,A)) \\ &\leq \max_z p^t(x,A) - p^t(z,A) \leq p(t)\end{aligned}$$

   The inequality $p(t) \leq 2d(t)$ is obtained using simple application of triangular inequality,

$$\begin{aligned}p(t) &= \max_{x,y}\left\|p^t(x,\cdot) - p^t(y,\cdot)\right\|_{TV} \\ &\leq \max_x\left\|p^t(x,\cdot) - \pi\right\|_{TV} + \max_y\left\|p^t(y,\cdot) - \pi\right\|_{TV} = 2d(t)\end{aligned}$$

2. Couple (X_s, Y_s) such that $\mathbb{P}(X_s = \hat{x}) = p^s(x,\hat{x})$, $p(Y_s = \hat{y}) = p^s(y,\hat{y})$ they are optimally coupled, as given in **Theorem 1.1**, *i.e.* $\mathbb{P}(X_s \neq Y_s) = \left\|p^s(x,\cdot) - p^s(y,\cdot)\right\|_{TV} \leq p(s)$ (the inequality comes from the previous part of the proposition). Conditionally on the event $X_s = z = Y_s$, take $X_{s+t} = Y_{s,t} \sim p^t(z,\cdot)$. Otherwise, if $X_s = x'$, $Y_s = y'$, with $x' \neq y'$ then take (X_{s+t}, Y_{s+t}) such that

$$\begin{aligned}\mathbb{P}[X_{s+t} = \hat{x}|X_s = x', Y_s = y'] = p^t(x',\hat{x}) \\ \mathbb{P}[Y_{s+t} = \hat{y}|X_s = x', Y_s = y'] = p^t(y',\hat{y})\end{aligned}$$

   again optimally coupled. Then by **Theorem 1.1**,

$$\begin{aligned}p(t+s) &\leq \mathbb{P}[X_{t+s} \neq Y_{t+s}] = \mathbb{P}[X_{t+s} \neq Y_{t+s}|X_s \neq Y_s]\mathbb{P}[X_s = Y_s] \\ &= p(s)\mathbb{P}[X_{t+s} \neq Y_{t+s}|X_s \neq Y_s]\end{aligned}$$

   Let $\nu(x,y)$ be the law of (X_s, Y_s) given $X_s \neq Y_s$, we have

$$\begin{aligned}\mathbb{P}[X_{t+s} \neq Y_{t+s}|X_s \neq Y_s] = \sum_{z\neq z'} \nu(z,z')\mathbb{P}[X_{t+s} \neq Y_{t+s}|X_s = z, Y_s = z'] \\ p(t)\sum_{z\neq z'}\nu(z,z') = p(t)\end{aligned}$$

   and we have the result.

(End of proof) □

Example : (reversibility)(Random to top shuffle) We have n cards labelled $1, \cdots, n$. Pick one at random and move to top.

Theorem 1.2) There cutoff around time is $n \log n$.

proof) Let X_0 have a fixed arbitrary distribution and Y_0 have the equilibrium distribution. Couple these two objects as the following : given X_t and Y_t, choose a randomly a number $i \in \{1, \cdots, n\}$ and put ith card of both X_t and Y_t on the top.

Let τ =(first time every card was picked). Then we may observe that $X_t = Y_t$ if $t \geq \tau$. Then $p(t) \leq \mathbb{P}(\tau > t)$ - so we make bound on probabilities on τ to bound $p(t)$.

: Let z_i =the time at which i difference cards were picked, and $T_i = z_i - z_{i-1}$. Then they are independent, $\tau = \sum_{i=1}^n T_i$, $T_i \sim \text{Geo}(\frac{n-i+1}{n})$, $\mathbb{E}[\tau] = \sum_{i=1}^n \frac{n}{n-i+1} \sim n\log n$, and $\text{Var}[\tau] = \sum_{i=1}^n \text{Var}[T_i] = \sum_{i=1}^n (n/i)^2 \leq n^2\pi^2/6 \ll (E[\tau])^2$. By Chebyshev,

$$\mathbb{P}[\tau > (1+\epsilon)n\log n] \xrightarrow{n\to\infty} 0$$

Therefore we also have $p((1+\epsilon)n\log n) \le \mathbb{P}[\tau > (1+\epsilon)n\log n] \to 0$ as $n \to \infty$.

For the other direction, let x_i =the card at position i, and A_j be the event of having $A_j = \{x : x_n > x_{n-1} > \cdots > x_{n-j}\}$, where x_j is the j-th card of the deck. Then $\pi(A_j) = 1/(j+1)!$ (π is the uniform distribution). If exactly s cards were not shuffled then they will be the bottm s cards at the original relative order. Start from $x = \begin{pmatrix} 1 \\ \vdots \\ n \end{pmatrix}$.

Then for $t = (1-\epsilon)n\log n$, $j = \lceil \log n \rceil$, again using similar estimate for $\sum T_i$ as above,

$$\mathbb{P}_x(X_t \in A_j) \ge \mathbb{P}[\sum_{i=1}^{n-j} T_i > t] = 1 - o(t)$$

Hence by definition of the distance function d,

$$d(t) \ge \mathbb{P}_x(X_t \in A_j) - \pi(A_j) \ge 1 - o(1)$$

and we have the desired result *(End of proof)* □

L_p norms

For $f \in R^S$, let $\|f\|_p = (\pi|f|^p)^{1/p}$, $\pi(h) = \sum \pi(x)h(x)$ and $\|f\|_\infty = \max_x |f(x)|$. For a signed measure σ, let

$$\|\sigma\|_{p,\pi} = \|\sigma/\pi\|_p = \Big(\sum_x \pi(x)\Big|\frac{\sigma(x)}{\pi(x)}\Big|^p\Big)^{1/p}$$
$$\|\sigma\|_{\infty,\pi} = \max_x \Big|\frac{\sigma(x)}{\pi(x)}\Big|$$

If μ, ν are distributions on X, $\|\mu - \nu\|_{1,\pi} = \sum_x \pi(x)|\frac{\mu(x)}{\pi(x)} - \frac{\nu(x)}{\pi(x)}| = 2\|\mu - \nu\|_{TV}$.

By Jensen's inequality, $\|\nu - \mu\|_{p,\pi}$ is non-decreasing in p.

Lemma 1) For reversible chains,

$$\|p^s(x,\cdot) - \pi\|_{2,\pi}^2 = \frac{p^{2s}(x,x)}{\pi(x)} - 1$$

proof)

$$\|p^s(x,\cdot) - \pi\|_{2,\pi}^2 = \sum_y \pi(y)\Big(\frac{p^s(x,y)}{\pi(y)} - 1\Big)^2 = -1 + \sum_y \frac{p^s(x,y)^2}{\pi(y)}$$
$$= -1 + \sum_y \frac{p^s(x,y)p^s(y,x)}{\pi(x)} = -1 + \frac{p^{2s}(x,x)}{\pi(x)}$$

(End of proof) □

Lemma 2) Under reversibility,

$$\Big|\frac{p^{s+t}(x,y)}{\pi(y)} - 1\Big|^2 \le \Big(\frac{p^{2s}(x,x)}{\pi(x)} - 1\Big)\Big(\frac{p^{2t}(y,y)}{\pi(y)} - 1\Big)$$

proof)

$$(LHS) = \Big|\frac{\sum_z (p^s(x,z)-\pi(z))(p^t(z,y)-\pi(y))}{\pi(y)}\Big|^2$$

and by reversibility, $p^t(z,y)\pi(z) = p^t(y,z)\pi(y)$ so

$$= \Big|\sum_z \pi(z)\frac{p^s(x,z)-\pi(z)}{\pi(z)}\cdot\frac{p^t(y,z)-\pi(z)}{\pi(z)}\Big|^2$$

$$\le \sum_z \pi(z)\Big(\frac{p^s(x,z)-\pi(z)}{\pi(z)}\Big)^2 \sum_z \pi(z)\Big(\frac{p^t(y,z)-\pi(z)}{\pi(z)}\Big)^2$$

and using **Lemma 1**, we have the desired result.

The most interesting case of the lemma is obtained when $t = s$.

Corollary)

$$\max_{x,y}\Big|\frac{p^{2s}(x,y)}{\pi(y)} - 1\Big| = \max_x \frac{p^{2s}(x,x)}{\pi(x)} - 1$$

In other words, if we write $d_p(t) = \max_x \big\|p^t(x,\cdot)-\pi\big\|_{p,\pi}$ then $d_\infty(t)^2 = d_\infty(wt)$

Example : The hypercube (n-dimensional) is $G = (\{0,1\}^n, E)$, $E = \{\{x,y\} : x, y$ differ on exactly 1 corrdinate$\}$. Consider 'lazy' random walk

$$p(x,y) = \begin{cases} 1/2 & \text{if } x = y \\ 1/2n & \text{if } x \sim y \end{cases}$$

(24th January, Thursday)

(no lecture on Tuesda Jan 29th)

(Example continues) The transition matrix of lazy simple random walk on hypercube is $P(x,y) = \frac{1}{1}_{x=y} + \frac{1}{2n}1_{(x,y)\in E}$.

Theorem) Cutoof around time is $\frac{1}{2}n\log n$.

At each step, pick a random co-oordinate and "refresh" it, *i.e.* flip with probability $\frac{1}{2}$. Let

$$\tau = \text{first time every co-ordinate is refreshed}$$

Then τ is a coupon-collector time(as before), so is concentrated around time $n\log n$. Then $X_\tau \sim \pi$, the equilibrium distribution, and (X_τ, τ) are independent.

$$d(t) \le \mathbb{P}[\tau > t] = o(1) \quad \text{if } t = (1+\epsilon)n\log n$$

: let $\text{sep}(\mu,\nu) = \max_x(1 - \frac{\mu(x)}{\nu(x)})$. Then

$$\big\|\mu-\nu\big\|_{TV} = \sum_x (\mu(x)-\nu(x))_+ = \sum_x \mu(x)(1-\frac{\nu(x)}{\mu(x)})_+$$

$$\le \max_x 1 - \frac{\nu(x)}{\mu(x)} = \text{sep}(\mu,\nu)$$

Before we complete the proof, we have to prove some results.

Definition) T is a **stationary time** if for some filtrations $\mathcal{F}_t \supset \sigma(X_s : s \le t)$ such that $1_{\tau > t} \in \mathcal{F}_t$ and $X_\tau \sim \pi$.

It is **strong stationary time (SST)** if alaso X_τ is independent form τ.

Lemma) If τ is a SST, then

$$\max_x \operatorname{sep}(p^t(x,\cdot),\pi) \le \max_x \mathbb{P}_x(\tau > t)$$

proof)

$$\begin{aligned} p^t(x,y) \ge \sum_{s=0}^t \mathbb{P}_x[X_t = y, \tau = s] &= \sum_{s=0}^t \mathbb{P}[\tau = s]\mathbb{P}_\pi(X_{t-s} = y) \\ &= \pi(y)\mathbb{P}[\tau \le t] \end{aligned}$$

(End of proof) □

This lemma justifies the inequality $d(t) \le \mathbb{P}[T > t]$ above.

Now let us work in the continuous time setting - "Refresh" each coordinate independently at rate $\frac{1}{n}$. By symmetry, $p^{2t}(x,x)$ is independent of x,

$$d_2^2(t) = \frac{p^{2t}(0,0)}{\pi(0)} - 1 = 2^n Q_{2t}(0,0) - 1 = (\star)$$

where $Q = \begin{pmatrix} -1/2n & 1/2n \\ 1/2n & -1/2n \end{pmatrix}$. Then $Q_{2t}(0,0) = \frac{1}{2} + \frac{1}{2}e^{-2t/n}$ so

$$(\star) = (1 + e^{-2t/n})^n - 1 \le \exp(ne^{-wt/n})) - 1 = \exp(n^{-\epsilon}) - 1 \le 2n^{-\epsilon} \xrightarrow{n\to\infty} 0$$

for $t = \frac{1+\epsilon}{2} nlogn$, where we used $e^x \le 1 + x + x^2$ for $x \in [-1,1]$.

2 Spectral Methods

(we follow Levin Chapter 12)

Proposition 2.7) Let P be a transition matrix. Then

1. If λ is an eigenvector of P, then $|\lambda| \le 1$.
2. If P is irreducible and $Pf = f$, then $f = (c, \cdots, c)^T$.
3. If P is irreducible aperiodic, and $\lambda \ne 1$ is an eigenvalue, then $|\lambda| < 1$.

   If P has period, then $e^{-2\pi i/n}$ is a eigenvalue.

proof)

2. Let $x \in S$ be such that $f(x) = \max_y |f(y)|$ then $Pf(x) = f(x) = \sum P(x,y)f(y)$ hence $f(y) = f(x)$ for all y such that $P(x,y) > 0$.

   Now apply this argument iterativel to find that f is constant.

3. (later part) We can partition S to s sets $A_1, \cdots, A_s$ such that $P(x, A_{i+1}) = 1$ for all i and $x \in A_i$, Then $f(x) = e^{\frac{2\pi i}{s} j}$ for $x \in A_j$, so $Pf = f \cdot e^{\frac{2\pi i}{s}}$.

(reversibility)Recall : $P^*(x, y) = \frac{\pi(y)}{\pi(x)} P(y, x)$. Then we have

$$\rho(t) = \max_{f:\pi(f)=0, f\neq 0} \frac{\left\|(P^*)^T f\right\|_1}{\left\|f\right\|_1}$$

(reversibility)[Hint : try $f = \frac{1_x}{\pi(x)} - \frac{1_y}{\pi(y)}$ and express general $f \neq 0$ with $\pi f = 0$ as $f = \sum_{xy} c_{xy} \left(\frac{1_x}{\pi(x)} - \frac{1_y}{\pi(y)}\right)$ with $\sum_{xy} |c_{xy}| = \frac{1}{2}\left\|f\right\|_1$.]

Lemma)

1. $\sum_y P^*(x, y) = 1$.
2. P is irreducbie (reversibility)iff P^* is
3. $\pi(x_1)P(x_1, x_2) \cdots P(x_{s-1}, x_s) = \pi(x_s)p^*(x_s, x_{s-1}, \cdots p^*(x_2, x_1))$ so $\pi(x)(P^*)^s(x, y) = \pi(y)P^s(y, x)$, i.e. $(P^*)^s = (P^s)^*$. In particular, $(P^*)^s(x, x) = P^s(x, x)$.
4. P is aperiodic (reversibility)iff P^* is.
5. $\pi P^* = \pi$.
6. $\langle Pf, g\rangle_\pi = \langle f, P^* g\rangle_\pi$

proof of 6)

6.
$$\begin{aligned}\sum_x \pi(x)g(x)\sum_y P(x,y)f(y) &= \sum_x g(x)\sum_y \pi(y)P^*(y,x)f(y) \\ &= \sum_y f(y)\sum_x \pi(y)P^*(y,x)g(x) = \langle f, P^* g\rangle_\pi\end{aligned}$$

Lemma) If $Pf = \lambda f$, $\lambda \neq 1$, then $\pi f = 0$.

proof)

$$\lambda\pi(f) = \langle Pf, 1\rangle_\pi = \langle, f, P^*\rangle_\pi = \pi(f)$$

(End of proof) □

proof of 1,3) Has $\rho_*(t) \geq |\lambda|^t$ if λ is an eigenvalue of P and ρ_* is *rho* w.r.t. P^*. But P is irreducuble aperiodic then so is P^*. Therefore $P_*(t) \xrightarrow{t\to\infty} 0$, so $|\lambda| < 1$.

(End of proof) □

Theorem 2.1) Assume P is irducible and reversible w.r.t. π.

1. $\exists$ an orthonormal basis $f_1, \cdots, f_n$ w.r.t $\langle\cdot,\cdot\rangle_\pi$ such that $f_1 = 1$ and $Pf_i = \lambda_i f_i$, $\lambda_1 = 1 > \lambda_2 \geq \cdots \geq \lambda_n \geq -1$.

2. $\frac{P^t(x,y)}{\pi(y)} - 1 = \sum_{j=2}^n f_j(x) f_j(y) \lambda_j^t$.

proof) Note that $A(x,y) = \sqrt{\frac{\pi(x)}{\pi(y)}} P(x,y)$ is symmetric, so $A = D_\pi P D_\pi^{-1}$, $D_\pi = \mathrm{Diag}(\sqrt{\pi(x)})$. Then $\exists \phi_1, \cdots, \phi_n$ such that $\langle \phi_i, \phi_j \rangle = 1_{i=j}$ and $A\phi_j = \lambda\phi)j$. Let $f_j = D_\pi^{-1}\phi_j$, then

$$\begin{aligned} Pf_j &= D_\pi^{-1} A\phi_j = \lambda_j D_\pi^{-1}\phi_j = \lambda_j f_j \\ \langle f_j, f_i \rangle_\pi &= \langle \phi_j, \phi_i \rangle = 1_{i=j} \end{aligned}$$

2.

$$P^T 1_y = \sum_j P^T f_j \langle f_j, 1_y \rangle_\pi = \sum_j \lambda_j^T f_j \pi(y) f_j(y) \quad P^T(x,y) = P^t 1_y(x) = \sum_{j=1}^n f_j(x) f_j(y) \pi_y \lambda_j^t$$

Corollary) Let $\lambda_* = |\lambda_2| \vee |\lambda_n|$. Then

$$\frac{P^{2t}(x,x)}{\pi(x)} - 1 \le \lambda_*^2 (\frac{p^{2t-2}(x,x)}{\pi(x)} - 1) \le \lambda_*^{2t} (\frac{1-\pi(x)}{\pi(x)})$$