Markov Chains Mixing Time, Random Walk on Groups and the 15 Puzzle Problem

Yang Chu, Robert Hough(Advisor)

May 24, 2020

Abstract

In this thesis we study Markov chains and representation theory, which lead us to the research of random walk on 15 puzzle problem.

In the first section, we study the basics of Markov chains. Several probabilistic methods such as coupling, stationary time and distinguishing statistics are discussed.

In the second section, we study representation theory.

In the third section, we investigate Green's function on 2-dimensional lattice. This eventually provides a way to calculate return probability on 2D lattice.

In the fourth section, we illustrate several examples and analyze their mixing time by the methods we develop in previous sections.

In the fifth section, we develop comparison techniques, which can give bounds of spectrum of unknown chains by comparing with known chains on the same state space.

In the sixth section, we give an example of counting problem, which is an application of Markov Chain mixing times.

In the seventh section, we discuss some results in our research of random walk on the $n^2 - 1$ puzzle.

In the eighth section, two concentration inequalities are presented, which are used in section 7.

Acknowledgements

I'm very fortune and grateful to be advised and helped by many people at Stony Brook University during writing this thesis. In particular, I would like to thank my advisor Professor Robert Hough for his supervision and thorough support, Professor Scott Sutherland for general advice, and Summer Math Foundation for their generous financial support.

Chapter 1

Markov chains

Most of this chapter is based on [2]

1.1 Basics

Definition 1.1.1. *A Markov chain is a process which moves among the elements of a set $\mathcal{X}$ in the following manner: when at $x \in \mathscr{X}$, the next position is chosen according to a fixed probability distribution $P(x, \cdot)$ depending only on x. More precisely, a sequence of random variables $(X_0, X_1, \ldots)$ is a* ***Markov chain with state space*** $\mathscr{X}$ ***and transition matrix*** P *if for all $x, y \in \mathcal{X}$, all $t \geqslant 1$, and all events $H_{t-1} = \bigcap_{s=0}^{t-1} \{X_s = x_s\}$ satisfying $\mathbf{P}(H_{t-1} \cap \{X_t = x\}) > 0$, we have*

$$\mathbf{P}\{X_{t+1} = y | H_{t-1} \cap \{X_t = x\}\} = \mathbf{P}\{X_{t+1} = y | X_t = x\} = P(x, y) \tag{1.1}$$

1.1 is called ***Markov property***

The x -th row of P is the distribution $P(x, \cdot)$. Thus P is stochastic, that is, its entries are all non-negative and

$$\sum_{y \in \mathcal{X}} P(x, y) = 1 \quad \text{for all } x \in \mathscr{X}$$

The distribution at time t can be found by matrix multiplication. Let $(X_0, X_1, \ldots)$ be a finite Markov chain with state space $\mathscr{X}$ and transition matrix P, and let the row vector μ_t be the distribution of X_t :

$$\mu_t(x) = \mathbf{P}\{X_t = x\} \quad \text{for all } x \in \mathscr{X}$$

By conditioning on the possible predecessors of the $(t+1)$ -st state, we see that

$$\mu_{t+1}(y)=\sum_{x \in \mathcal{X}} \mathbf{P}\left\{X_t=x\right\} P(x, y)=\sum_{x \in \mathcal{X}} \mu_t(x) P(x, y) \quad \text{for all } y \in \mathcal{X}$$

Rewriting this in vector form gives

$$\mu_{t+1}=\mu_t P \quad \text{for } t \geqslant 0,$$

and hence

$$\mu_t=\mu_0 P^t \quad \text{for } t \geqslant 0.$$

Since we will often consider Markov chains with the same transition matrix but different starting distributions, we introduce the notation $\mathbf{P}_\mu$ and $\mathbf{E}_\mu$ for probabilities and expectations given that $\mu_0=\mu$.

Quite often, the initial distribution will be concentrated at a single definite starting state x (start at x). We denote this distribution by δ_x (Dirac measure):

$$\delta_x(y)=\begin{cases}1 & \text{if } y=x \\ 0 & \text{if } y \neq x\end{cases}$$

We write simply $\mathbf{P}_x$ and $\mathbf{E}_x$ for $\mathbf{P}_{\delta_x}$ and $\mathbf{E}_{\delta_x}$, respectively. These definitions and together imply that

$$\mathbf{P}_x\left\{X_t=y\right\}=\left(\delta_x P^t\right)(y)=P^t(x, y)$$

That is, the probability of moving in t steps from x to y is given by the (x, y) -th entry of P^t. We call these entries **the t -step transition probabilities**.

Definition 1.1.2. *For a function (a column vector) f on the state space $\mathscr{X}$. Consider multiplying f by P from the left and the x -th entry of the resulting vector:*

$$P f(x)=\sum_y P(x, y) f(y)=\sum_y f(y) \mathbf{P}_x\left\{X_1=y\right\}=\mathbf{E}_x\left(f\left(X_1\right)\right).$$

We say f is ***harmonic*** *at $x \in \mathscr{X}$ if $P f(x)=f(x)$ (that is, f has stationary expectation at x.) f is* ***harmonic*** *on $A \subset \mathscr{X}$ if $P f(x)=f(x)$ for all $x \in A$*

Definition 1.1.3. *For $i, j \in \mathscr{X}$, we say j is* ***reachable*** *from i if there is a positive integer such that $P^t(i, j)>0$.*

We say i and j ***communicate*** *if i is reachable from j and j is reachable from i*

It's easy to see that communicate is a equivalent relation on $\mathscr{X}$, so it has corresponding equivalent classes, which we call **communicating classes**.

Definition 1.1.4. *A chain P is called* ***irreducible*** *if for any two states $x, y \in \mathscr{X}$ there exists an integer t such that $P^t(x, y) > 0$.*

We see the above definition is equivalent to that $\mathscr{X}$ has only 1 communicating class.

Definition 1.1.5. *Let $\mathcal{T}(x) := \{t \geqslant 1 : P^t(x, x) > 0\}$ be the set of times when it is possible for the chain to return to starting position x. The* ***period*** *of state x is defined to be the greatest common divisor of $\mathcal{T}(x)$*

Lemma 1.1.1. *If P is irreducible, then $\gcd \mathcal{T}(x) = \gcd \mathcal{T}(y)$ for all $x, y \in \mathscr{X}$*

Proof. Fix two states x and y. There exist non-negative integers r and ℓ such that $P^r(x, y) > 0$ and $P^l(y, x) > 0$.

Then $P^{r+l}(x, x) \geqslant P^r(x, y)P^l(y, x) > 0$. So $r+l$ is a multiple of $\gcd \mathcal{T}(x)$.

Also, given $m \in \mathcal{T}(y)$, we have

$$P^{r+m+l}(x, x) \geqslant P^r(x, y)P^m(y, y)P^l(y, x) > 0$$

. Then $r + m + l$ is a multiple of $\gcd \mathcal{T}(x)$, and so is m.

Hence $\gcd \mathcal{T}(y) \geqslant \gcd \mathcal{T}(x)$.

Exchange x and y we get $\gcd \mathcal{T}(x) \geqslant \gcd \mathcal{T}(y)$.

Therefore $\gcd \mathcal{T}(y) = \gcd \mathcal{T}(x)$.

□

Definition 1.1.6. *For an irreducible chain, the period of the chain is defined to be the period which is common to all states. The chain will be called* ***aperiodic*** *if all states have period 1. If a chain is not aperiodic, we call it* ***periodic****.*

Lemma 1.1.2. *If P is aperiodic and irreducible, then there is an integer r_0 such that $P^r(x, y) > 0$ for all $x, y \in \mathcal{X}$ and $r \geqslant r_0$*

Definition 1.1.7. *A distribution π on $\mathcal{X}$ satisfying*

$$\pi = \pi P$$

can have another interesting property: in that case, π was the long-term limiting distribution of the chain. We call such probability π satisfying a stationary distribution of the Markov chain.

Definition 1.1.8. ***Hitting and first return times.*** *Assume that the Markov chain $(X_0, X_1, \ldots)$ under discussion has finite state space $\mathscr{X}$ and transition matrix P. For $x \in \mathcal{X}$, define the hitting time for x to be*

$$\tau_x := \min\{t \geqslant 0 : X_t = x\}$$

the first time at which the chain visits state x. For situations where only a visit to x at a positive time will do, we also define

$$\tau_x^+ := \min\{t \geqslant 1 : X_t = x\}$$

When $X_0 = x$, we call τ_x^+ the first return time.

Lemma 1.1.3. *For any states x and y of an irreducible chain, $\mathbf{E}_x\left(\tau_y^+\right) < \infty$.*

Proof. The definition of irreducibility implies that there exist an integer $r > 0$ and a real $\varepsilon > 0$ with the following property: for any states $z, w \in \mathscr{X}$, there exists a $j \leqslant r$ with $P^j(z, w) > \varepsilon$. Thus for any value of X_t, the probability of hitting state y at a time between t and $t + r$ is at least ε. Hence for $k > 0$ we have

$$\begin{aligned}
\mathbf{P}_x\left\{\tau_y^+ > kr\right\} &= \mathbf{P}_{X_{(k-1)r}}\left\{\tau_y^+ > r\right\}\mathbf{P}_x\left\{\tau_y^+ > (k-1)r\right\} \\
&\leqslant (1-\varepsilon)\mathbf{P}_x\left\{\tau_y^+ > (k-1)r\right\}
\end{aligned}$$

Repeated above yields

$$\mathbf{P}_x\left\{\tau_y^+ > kr\right\} \leqslant (1-\varepsilon)^k$$

Recall that when Y is a non-negative integer-valued random variable, we have

$$\mathbf{E}(Y) = \sum_{t \geqslant 0} \mathbf{P}\{Y > t\}.$$

Since $\mathbf{P}_x\left\{\tau_y^+ > t\right\}$ is a decreasing function of t, it suffices to bound all terms of the corresponding expression for $\mathbf{E}_x\left(\tau_y^+\right)$:

$$\mathbf{E}_x\left(\tau_y^+\right) = \sum_{t \geqslant 0} \mathbf{P}_x\left\{\tau_y^+ > t\right\} \leqslant \sum_{k \geqslant 0} r\mathbf{P}_x\left\{\tau_y^+ > kr\right\} \leqslant r\sum_{k \geqslant 0}(1-\varepsilon)^k < \infty$$

$\square$

Next, we study the existence of the stationary distribution of finite aperiod irreducible chains, where the distribution is given by

$$\pi(x) = \frac{1}{\mathbf{E}_x[\tau_x^+]}. \tag{1.2}$$

Let z be an arbitrary initial state. To construct the stationary distribution π, we consider the expected time the chain reaching a given state y.

Thus we define

$$\begin{aligned}
\tilde{\pi}(y) &:= \mathbf{E}_z(\text{ number of visits to } y \text{ before returning to } z) \\
&= \mathbf{E}_z \sum_{t=0}^{\tau_z^+ - 1} \mathbf{1}_{X_t = y} \\
&= \sum_{n=1}^{\infty} \sum_{t=0}^{n-1} \mathbf{P}_z \left\{ X_t = y | \tau_z^+ = n \right\} \mathbf{P}_z\{\tau_z^+ = n\} \\
&= \sum_{n=1}^{\infty} \sum_{t=0}^{n-1} \mathbf{P}_z \left\{ X_t = y, \tau_z^+ = n \right\} \\
&= \sum_{t=0}^{\infty} \mathbf{P}_z \left\{ X_t = y, \tau_z^+ > t \right\},
\end{aligned}$$

where the last equality is by changing order of summation (sum by columns is equal to sum by rows, since

$$\tilde{\pi}(y) = \sum_{t=0}^{\infty} \mathbf{P}_z \left\{ X_t = y, \tau_z^+ > t \right\} \leqslant \sum_{t=0}^{\infty} \mathbf{P}_z\{\tau_z^+ > t\} = \mathbf{E}_z \tau_z^+ < \infty,$$

Proposition 1.1.4. *Let $\tilde{\pi}$ be the measure on $\mathcal{X}$ defined above*

(i) If $\mathbf{P}_z \{\tau_z^+ < \infty\} = 1$, then $\tilde{\pi}$ satisfies $\tilde{\pi} P = \tilde{\pi}$

(ii) If $\mathbf{E}_z(\tau_z^+) < \infty$, then $\pi := \frac{\tilde{\pi}}{\mathbf{E}_z(\tau_z^+)}$ is a stationary distribution.

Proof. We have

$$\tilde{\pi}(y) = \sum_{t=0}^{\infty} \mathbf{P}_z \left\{ X_t = y, \tau_z^+ > t \right\} \leqslant \sum_{t=0}^{\infty} \mathbf{P}_z\{\tau_z^+ > t\} = \mathbf{E}_z \tau_z^+ < \infty,$$

To check $\tilde{\pi}$ is stationary:

$$\tilde{\pi} P(y) = \sum_{x \in \mathcal{X}} \tilde{\pi}(x) P(x,y) = \sum_{x \in \mathcal{X}} \sum_{t=0}^{\infty} \mathbf{P}_z \left\{ X_t = x, \tau_z^+ > t \right\} P(x,y) \tag{1.3}$$

We know that $\{\tau_z^+ \geqslant t+1\} = \{\tau_z^+ > t\}$. So

$$\mathbf{P}_z \left\{ X_t = x, X_{t+1} = y, \tau_z^+ \geqslant t+1 \right\} = \mathbf{P}_z \left\{ X_t = x, \tau_z^+ \geqslant t+1 \right\} P(x,y)$$

Change order of 1.3 and use above identity, we get

$$\begin{aligned}\sum_{x\in\mathcal{X}} \tilde{\pi}(x)P(x,y) &= \sum_{t=0}^{\infty} \mathbf{P}_z\left\{X_{t+1}=y, \tau_z^+ \geqslant t+1\right\} \\ &= \sum_{t=1}^{\infty} \mathbf{P}_z\left\{X_t=y, \tau_z^+ \geqslant t\right\}\end{aligned}$$

Thus,

$$\begin{aligned}&\sum_{t=1}^{\infty} \mathbf{P}_z\left\{X_t=y, \tau_z^+ \geqslant t\right\} \\ &= \tilde{\pi}(y) - \mathbf{P}_z\left\{X_0=y, \tau_z^+ > 0\right\} + \sum_{t=1}^{\infty} \mathbf{P}_z\left\{X_t=y, \tau_z^+ = t\right\} \\ &= \tilde{\pi}(y) - \mathbf{P}_z\left\{X_0=y\right\} + \mathbf{P}_z\left\{X_{\tau_z^+}=y\right\} \quad (1.4) \\ &= \tilde{\pi}(y). \quad (1.5)\end{aligned}$$

The last equality follows by considering two cases:

$y=z$: since $X_0=z$ and $X_{\tau^+}=z$, the last two terms of 1.4 are both 1, and they cancel each other out.

$y\neq z$: Here the last two terms of 1.4 are both 0.

Therefore,$\tilde{\pi}=\tilde{\pi}P$.

Normalize the measure by $\sum_x \tilde{\pi}(x)=\mathbf{E}_z\left(\tau_z^+\right)$, we get

$$\pi(x)=\frac{\tilde{\pi}(x)}{\mathbf{E}_z\left(\tau_z^+\right)} \quad \text{satisfies } \pi=\pi P \tag{1.6}$$

□

Next, we show that the stationary distribution of an irreducible Markov chain is unique, which implies 1.2. To see why, by 1.6,

$$\pi(z)=\frac{\tilde{\pi}(z)}{\mathbf{E}_z\left(\tau_z^+\right)}=\frac{1}{\mathbf{E}_z\left(\tau_z^+\right)}.$$

Since z is arbitrary and π is unique, we conclude 1.2 is true.

We need the following lemma to prove the stationary distribution is unique:

Lemma 1.1.5. *Suppose that P is irreducible. A function h which is harmonic at every point of $\mathcal{X}$ is constant.*

Proof. Since $\mathcal{X}$ is finite, there must be a state x_0 such that $h(x_0) = M$ is maximal. If for some state z such that $P(x_0, z) > 0$ we have $h(z) < M$, then

$$h(x_0) = P(x_0, z)\, h(z) + \sum_{y \neq z} P(x_0, y)\, h(y) < M, \tag{1.7}$$

contradiction. So $h(z) = M$ for all states z such that $P(x_0, z) > 0$

For any $y \in \mathcal{X}$, irreducibility implies that there is a sequence

$x_0, x_1, \ldots, x_n = y$ with $P(x_i, x_{i+1}) > 0$. Repeating the argument above tells us that $h(y) = h(x_{n-1}) = \cdot = h(x_0) = M$. Thus h is constant. □

Theorem 1.1.6. *For an irreducible Markov chain, there is an unique stationary distribution.*

Proof. By 1.1.4, there is at least one stationary distribution. 1.1.5 implies that the kernel of $P - I$ has dimension 1 (vectors with the same value in each coordinate). Thus the rank of $P - I$ is $|\mathcal{X}| - 1$. Since the column rank is equal to the row rank, the space of solutions of row vector equation $\nu = \nu P$ has dimension 1. The space contains only one vector whose entries sum to 1. □

Next, we discuss reversibility and time reversals of Markov chains. For our interests here, the mixing time for a random walk on groups is the same as the reversed one, which will be shown later.

Suppose a probability distribution π on $\mathcal{X}$ satisfies

$$\sum_{y \in \mathcal{X}} \pi(y) P(y, x) = \sum_{y \in \mathcal{X}} \pi(x) P(x, y) = \pi(x) \tag{1.8}$$

1.8 is called **detailed balance equation**.

Proposition 1.1.7. *Let P be the transition matrix of a Markov chain with state space $\mathcal{X}$. Any distribution π satisfying the detailed balance equations is stationary for P*

Proof.

$$\sum_{y \in \mathcal{X}} \pi(y) P(y, x) = \sum_{y \in \mathcal{X}} \pi(x) P(x, y) = \pi(x)$$

□

If 1.8 holds, then

$$\pi\left(x_0\right) P\left(x_0, x_1\right) \cdots P\left(x_{n-1}, x_n\right)=\pi\left(x_n\right) P\left(x_n, x_{n-1}\right) \cdots P\left(x_1, x_0\right) \tag{1.9}$$

, which is same as

$$\mathbf{P}_\pi\left\{X_0=x_0, \ldots, X_n=x_n\right\}=\mathbf{P}_\pi\left\{X_0=x_n, X_1=x_{n-1}, \ldots, X_n=x_0\right\}. \tag{1.10}$$

In other words, if a chain (X_t) satisfies (1.29) and has stationary initial distribution, then the distribution of $(X_0, X_1, \ldots, X_n)$ is the same as the distribution of‘$(X_n, X_{n-1}, \ldots, X_0)$. For this reason, a chain satisfying 1.8 is called **reversible**.

The time reversal of an irreducible Markov chain with transition matrix P and stationary distribution π is the chain with matrix

$$\widehat{P}(x, y):=\frac{\pi(y) P(y, x)}{\pi(x)} \tag{1.11}$$

Proposition 1.1.8. *Let (X_t) be an irreducible Markov chain with transition matrix P and stationary distribution π. Write $\left(\widehat{X}_t\right)$ for the time-reversed chain with transition matrix $\widehat{P}$. Then π is stationary for $\widehat{P}$, and for any $x_0, \ldots, x_t \in \mathcal{X}$ we have*

$$\mathbf{P}_\pi\left\{X_0=x_0, \ldots, X_t=x_t\right\}=\mathbf{P}_\pi\left\{\widehat{X}_0=x_t, \ldots, \widehat{X}_t=x_0\right\}$$

Proof. To check that π is stationary for $\widehat{P}$, we simply compute

$$\sum_{y \in \mathcal{X}} \pi(y) \widehat{P}(y, x)=\sum_{y \in \mathcal{X}} \pi(y) \frac{\pi(x) P(x, y)}{\pi(y)}=\pi(x)$$

To show the probabilities of the two trajectories are equal, note that

$$\begin{aligned}
\mathbf{P}_\pi\left\{X_0=x_0, \ldots, X_n=x_n\right\} & =\pi\left(x_0\right) P\left(x_0, x_1\right) P\left(x_1, x_2\right) \cdots P\left(x_{n-1}, x_n\right) \\
& =\pi\left(x_n\right) \widehat{P}\left(x_n, x_{n-1}\right) \cdots \widehat{P}\left(x_2, x_1\right) \widehat{P}\left(x_1, x_0\right) \\
& =\mathbf{P}_\pi\left\{\hat{X}_0=x_n, \ldots, \hat{X}_n=x_0\right\},
\end{aligned}$$

since $P\left(x_{i-1}, x_i\right)=\pi\left(x_i\right) \widehat{P}\left(x_i, x_{i-1}\right) / \pi\left(x_{i-1}\right)$ for each i □

Now, we introduce random walks on groups as an example of Markov chains.

Definition 1.1.9. *Given a probability distribution μ on a group $(G, \cdot)$,* ***the left random walk on*** G ***with increment distribution*** μ *is an Markov chain with state space G, initial distribution μ. The transition probability is given by:*

$$P(g, hg) = \mu(h), (\textit{equivalently}, P(g, h) = \mu(hg^{-1})).$$

The ***right random walk*** *is the Markov chain on the same state space and initial distribution, with transition probability:*

$$P(g, gh) = \mu(h), (\textit{equivalently}, P(g, h) = \mu(g^{-1}h)).$$

Definition 1.1.10. *Convolution*

Suppose P and Q are probabilities on finite group G. We define convolution of P and Q by

$$P * Q(s) := \sum_t P(st^{-1})Q(t)$$

For above definition, we see for a random walk on a group G with driven distribution μ, the distribution after 1 (step 2) move is $\mu * \mu = \mu^{*2}$. The distribution after n moves is $\mu^{*(n+1)}$

Due to the symmetry of groups, for any random walk on groups, the uniform distribution is unique:

Proposition 1.1.9. *Let P be the transition matrix of a random walk on a finite group G and let U be the uniform probability distribution on G. Then U is a stationary distribution for P*

Proof.

$$\sum_{h \in G} U(h)P(h, g) = \frac{1}{|G|} \sum_{k \in G} P\left(k^{-1}g, g\right) = \frac{1}{|G|} \sum_{k \in G} \mu(k) = \frac{1}{|G|} = U(g)$$

□

We call a probability distribution μ on a group G **symmetric** if $\mu(g) = \mu(g^{-1})$ for every $g \in G$.

Proposition 1.1.10. *The random walk on a finite group G with increment distribution μ is reversible if μ is symmetric.*

Proof. Let U be the uniform probability distribution on G. For any $g, h \in G$ we have that

$$U(g)P(g, h) = \frac{\mu\left(hg^{-1}\right)}{|G|} \quad \text{and} \quad U(h)P(h, g) = \frac{\mu\left(gh^{-1}\right)}{|G|}$$

are equal if and only if $\mu\left(hg^{-1}\right) = \mu\left(\left(hg^{-1}\right)^{-1}\right)$ □

Now we discuss the mixing time, the central property of Markov chains we are interested in this thesis.

1.2 Total variation distance and Mixing time

Firstly, we study **total variation distance** of two probability measures, which is a metric of the difference between two probability measures on the same state space.

Definition 1.2.1. *The total variation distance between two probability distributions μ and ν on $\mathcal{X}$ is defined by*

$$\|\mu - \nu\|_{\mathrm{TV}} = \max_{A \subset \mathcal{X}} |\mu(A) - \nu(A)|$$

Proposition 1.2.1. *Let μ and ν be two probability distributions on $\mathcal{X}$. Then*

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)| \tag{1.12}$$

Proof. Without loss of generality, assume

$$\|\mu - \nu\|_{\mathrm{TV}} = \max_{A \subset \mathcal{X}} |\mu(A) - \nu(A)| = \mu(X) - \nu(X)$$

We claim that, except for a set where μ and ν agrees, $X = \{x \in \mathcal{X} : \mu(x) - \nu(x) \geqslant 0\}$

Suppose $\exists x \in X$, such that $\mu(x) - \nu(x) < 0$. Then clearly $\mu(X \backslash x) - \nu(X \backslash x) > \mu(X) - \nu(X)$, contradiction. Then by the definition of total variation distance, X is the largest such set.

Similarly, if

$$\|\mu - \nu\|_{\mathrm{TV}} = \max_{A \subset \mathcal{X}} |\mu(A) - \nu(A)| = -(\mu(Y) - \nu(Y)),$$

$Y = \{y \in \mathcal{X} : \mu(x) - \nu(x) \leqslant 0\}$.

We see that $X \cup Y = \mathcal{X}$, and $X \cap Y$ is the set where μ and ν agrees.

We know $\sum_{a \in \mathcal{X}} \mu(a) = 1$ and $\sum_{a \in \mathcal{X}} \nu(a) = 1$. So $-(\mu(Y) - \nu(Y)) = -(1 - \mu(X) - (1 - \nu(Y))) = \mu(X) - \mu(Y)$. Indeed, the two expression of total variation distance agree.

So

$$\frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)| = \frac{1}{2} \left(\mu(A) - \nu(A) - (\mu(Y) - \nu(Y))\right) = \|\mu - \nu\|_{\mathrm{TV}}$$

□

Also, from the above proof we see

$$\|\mu - \nu\|_{\text{TV}} = \sum_{x:\mu(x)>\nu(x)} \mu(x) - \nu(x). \tag{1.13}$$

By 1.12 and triangle inequality in $\mathbb{R}$, we see that the total variation distance satisfying the triangle inequality: for probability distributions μ, ν and η

$$\|\mu - \nu\|_{\text{TV}} \leqslant \|\mu - \eta\|_{\text{TV}} + \|\eta - \nu\|_{\text{TV}}$$

The next theorem shows that for an irreducible aperiodic Markov chain, the chain will eventually converges to uniform distribution.

Theorem 1.2.2. *(Convergence Theorem). Suppose that P is irreducible and aperiodic, with stationary distribution π. Then there exist constants $\alpha \in (0,1)$ and $C > 0$ such that*

$$\max_{x\in\mathcal{X}} \left\|P^t(x,\cdot) - \pi\right\|_{\text{TV}} \leqslant C\alpha^t$$

The proof decomposes the chain into a mixture of repeated independent sampling from the stationary distribution and another Markov chain.

Proof. Since P is irreducible and aperiodic, there exists an r such that P^r has strictly positive entries. Let Π be the matrix with $|\mathcal{X}|$ rows, each of which is the row vector π. For sufficiently small $\delta > 0$, we have

$$P^r(x,y) \geqslant \delta\pi(y)$$

for all $x, y \in \mathcal{X}$. Let $\theta = 1 - \delta$. The equation

$$P^r = (1-\theta)\Pi + \theta Q \tag{1.14}$$

defines a stochastic matrix Q. It is a straightforward computation to check that $M\Pi = \Pi$ for any stochastic matrix M and that $\Pi M = \Pi$ for any matrix M such that $\pi M = \pi$ Next, we use induction to demonstrate that

$$P^{rk} = \left(1-\theta^k\right)\Pi + \theta^k Q^k \tag{1.15}$$

for $k \geqslant 1$. If $k = 1$, this holds by 1.14. Assuming that 1.15 holds for $k = n$

$$P^{r(n+1)} = P^{rn}P^r = \left[(1-\theta^n)\Pi + \theta^n Q^n\right]P^r$$

Distributing and expanding P^r in the second term (using 1.14) gives

$$P^{r(n+1)} = [1-\theta^n]\,\Pi P^r + (1-\theta)\theta^n Q^n \Pi + \theta^{n+1} Q^n Q$$

Using that $\Pi P^{r}=\Pi$ and $Q^{n}\Pi=\Pi$ shows that

$$P^{r(n+1)}=\left[1-\theta^{n+1}\right]\Pi+\theta^{n+1}Q^{n+1}$$

Using that $\Pi P^{r}=\Pi$ and $Q^{n}\Pi=\Pi$ shows that

$$P^{r(n+1)}=\left[1-\theta^{n+1}\right]\Pi+\theta^{n+1}Q^{n+1}$$

This establishes 1.15 for $k=n+1$ (assuming it holds for $k=n$), and hence it holds for all k.

Multiplying by P^{j} and rearranging terms now yields

$$P^{rk+j}-\Pi=\theta^{k}\left(Q^{k}P^{j}-\Pi\right)$$

To complete the proof, sum the absolute values of the elements in row x_0 on both sides and divide by 2. On the right, the second factor is at most the largest possible total variation distance between distributions, which is 1. Hence for any x_0 we have

$$\left\|P^{rk+j}\left(x_{0},\cdot\right)-\pi\right\|_{\mathrm{TV}}\leqslant\theta^{k}.$$

Taking $\alpha=\theta^{1/r}$ and $C=1/\theta$ finishes the proof.

□

Bounding the maximal distance (over $x_0\in\mathcal{X}$) between $P^{t}\left(x_{0},\cdot\right)$ and π is among our primary objectives. It is therefore convenient to define

$$d(t):=\max_{x\in\mathcal{X}}\left\|P^{t}(x,\cdot)-\pi\right\|_{\mathrm{TV}}\tag{1.16}$$

Later we will show it is often possible to bound $\|P^{t}(x,\cdot)-P^{t}(y,\cdot)\|_{\mathrm{TV}}$ uniformly over all pairs of states (x,y) by coupling. We therefore make the definition

$$\bar{d}(t):=\max_{x,y\in\mathcal{X}}\left\|P^{t}(x,\cdot)-P^{t}(y,\cdot)\right\|_{\mathrm{TV}}\tag{1.17}$$

Lemma 1.2.3. *If $d(t)$ and $\bar{d}(t)$ are as defined in 1.16 and 1.17 respectively, then*

$$d(t)\leqslant\bar{d}(t)\leqslant 2d(t)$$

Proof. It is immediate from the triangle inequality for the total variation distance that $\bar{d}(t)\leqslant 2d(t)$ To show that $d(t)\leqslant\bar{d}(t)$, note first that since π is stationary, we have $\pi(A)=\sum_{y\in\mathcal{X}}\pi(y)P^{t}(y,A)$ for any set A. (This is the definition of stationarity if A is a singleton $\{x\}$. To get this for arbitrary A, just sum over the elements in A.) Using this shows that

$$\begin{aligned}\left|P^t(x,A) - \pi(A)\right| &= \left|\sum_{y\in\mathcal{X}} \pi(y)\left[P^t(x,A) - P^t(y,A)\right]\right| \\ &\leqslant \sum_{y\in\mathcal{X}} \pi(y)\left\|P^t(x,\cdot) - P^t(y,\cdot)\right\|_{\mathrm{TV}} \leqslant \bar{d}(t)\end{aligned}$$

by the triangle inequality and the definition of total variation. Maximizing the left-hand side over x and A yields $d(t) \leqslant \bar{d}(t)$ ☐

Lemma 1.2.4. *The function $\bar{d}$ is submultiplicative: $\bar{d}(s+t) \leqslant \bar{d}(s)\bar{d}(t)$*

Proof. Fix $x, y \in \mathcal{X}$, and let (X_s, Y_s) be the optimal coupling of $P^s(x,\cdot)$ and $P^s(y,\cdot)$ whose existence is guaranteed by 1.20 (which we shall prove later). Hence

$$\|P^s(x,\cdot) - P^s(y,\cdot)\|_{\mathrm{TV}} = \mathbf{P}\{X_s \neq Y_s\}$$

We have

$$P^{s+t}(x,w) = \sum_z \mathbf{P}\{X_s = z\} P^t(z,w) = \mathbf{E}\left(P^t(X_s,w)\right)$$

For a set A, summing over $w \in A$ shows that

$$\begin{aligned}P^{s+t}(x,A) - P^{s+t}(y,A) &= \mathbf{E}\left(P^t(X_s,A) - P^t(Y_s,A)\right) \\ &\leqslant \mathbf{E}\left(\bar{d}(t)\mathbf{1}_{\{X_s\neq Y_s\}}\right) = \mathbf{P}\{X_s \neq Y_s\}\bar{d}(t)\end{aligned}$$

By 1.20, the right-hand side is at most $\bar{d}(s)\bar{d}(t)$

☐

Lemma 1.2.5. *$\bar{d}(t)$ is non-increasing in t. If c and t are positive integers, then*

$$d(ct) \leqslant \bar{d}(ct) \leqslant \bar{d}(t)^c$$

Definition 1.2.2. *The **mixing time** is defined by*

$$t_{\mathrm{mix}}(\varepsilon) := \min\{t : d(t) \leqslant \varepsilon\}$$

and

$$t_{\mathrm{mix}} := t_{\mathrm{mix}}(\frac{1}{4})$$

For random walk on groups, we set

$$t_{\mathrm{mix}} := t_{\mathrm{mix}}(\frac{1}{e})$$

1.2.3 and 1.2.5 show that when ℓ is a positive integer,

$$d\left(\ell t_{\text{mix}}(\varepsilon)\right) \leqslant \bar{d}\left(t_{\text{mix}}(\varepsilon)\right)^{\ell} \leqslant (2\varepsilon)^{\ell}$$

In particular, taking $\varepsilon = 1/4$ above yields

$$d\left(\ell t_{\text{mix}}\right) \leqslant 2^{-\ell}$$

and

$$t_{\text{mix}}(\varepsilon) \leqslant \lceil \log_2 \varepsilon^{-1} \rceil t_{\text{mix}}$$

Other distances between distributions are useful. Given a distribution π on $\mathcal{X}$ and $1 \leqslant p \leqslant \infty$, the $\ell^p(\pi)$ norm of a function $f : \mathcal{X} \to \mathbb{R}$ is defined as

$$\|f\|_p := \begin{cases} \left[\sum_{y\in\mathcal{X}} |f(y)|^p \pi(y)\right]^{1/p} & 1 \leqslant p < \infty \\ \max_{y\in\mathcal{X}} |f(y)| & p = \infty \end{cases}$$

The d_2 distance is a scaled version of the ℓ^2 norm,

$$\|\mu - \nu\|_{d_2}^2 = |\mathscr{X}| \sum_{x\in\mathscr{X}} (\mu(x) - \nu(x))^2. \tag{1.18}$$

For $0 < \epsilon < 1$, the $\frac{\epsilon}{|\mathscr{X}|}$-$\ell^\infty$ distance between μ and ν is

$$\|\mu - \nu\|_{\epsilon,\infty} = \frac{|\mathscr{X}|}{\epsilon} \sup_{x\in\mathscr{X}} |\mu(x) - \nu(x)|. \tag{1.19}$$

The distance and mixing times corresponding to these metric are defined in similar way as for total variation distance.

Given any of these metrics, the mixing time of the chain $\{e_x^t P^N\}$ to uniformity ν is the first steps N such that $\|e_x^t P^N - \nu\| < \frac{1}{e}$. For symmetric random walk on a group, the total variation mixing time and $\frac{\epsilon}{|G|} - \ell^\infty$ mixing time are bounded up to constants by the d_2 mixing time.

1.3 Coupling

Next, we discuss **coupling of Markov chains**, which is a powerful probabilistic method. We firstly introduce coupling of two distributions.

Definition 1.3.1. *A* ***coupling*** *of two probability distributions μ and ν is a pair of random variables (X, Y) defined on a single probability space such that the marginal distribution of X is μ and the marginal distribution of Y is ν. That is, a coupling (X, Y) satisfies $\mathbf{P}\{X = x\} = \mu(x)$ and $\mathbf{P}\{Y = y\} = \nu(y)$*

We first point out a relation between coupling and total variation distance.

Proposition 1.3.1. *Let μ and ν be two probability distributions on $\mathcal{X}$. Then*

$$\|\mu - \nu\|_{\mathrm{TY}} = \inf\{\mathbf{P}\{X \neq Y\} : (X, Y) \textit{ is a coupling of } \mu \textit{ and } \nu\} \tag{1.20}$$

*The coupling (X, Y) attaches the infimum is called the **optimal**.*

Proof. For any coupling (X, Y) and $A \subset \Omega$

$$\begin{aligned} \mu(A) - \nu(A) &= \mathbb{P}(X \in A) - \mathbb{P}(Y \in A) \\ &\leqslant \mathbb{P}(X \in A, Y \notin A) \\ &\leqslant \mathbb{P}(X \neq Y) \end{aligned}$$

Similarly, $\nu(A) - \mu(A) \leqslant \mathbb{P}(X \neq Y)$ Therefore, $\|\mu - \nu\|_{tv} \leqslant \mathbb{P}(X \neq Y)$

We construct a coupling for which $\mathbf{P}\{X \neq Y\}$ is exactly $\|\mu - \nu\|_{\mathrm{TV}}$.

We use the following procedure to generate X and Y. Let

$$p = \sum_{x \in \mathcal{X}} \mu(x) \wedge \nu(x).$$

Write

$$\sum_{x \in \mathcal{X}} \mu(x) \wedge \nu(x) = \sum_{\substack{x \in \mathcal{X}, \\ \mu(x) \leqslant \nu(x)}} \mu(x) + \sum_{\substack{x \in \mathcal{X}, \\ \mu(x) > \nu(x)}} \nu(x)$$

Adding and subtracting $\sum_{x : \mu(x) > \nu(x)} \mu(x)$ to the right-hand side above shows that

$$\sum_{x \in \mathcal{X}} \mu(x) \wedge \nu(x) = 1 - \sum_{\substack{x \in \mathcal{X}, \\ \mu(x) > \nu(x)}} [\mu(x) - \nu(x)] = 1 - \|\mu - \nu\|_{\mathrm{TV}} = p.$$

The coupling is constructed as following: Flip a coin with probability of heads equal to p (i) If the coin comes up heads, then choose a value Z according to the probability distribution

$$\gamma_{\mathrm{III}}(x) = \frac{\mu(x) \wedge \nu(x)}{p}$$

and set $X = Y = Z$

(ii) If the coin comes up tails, choose X according to the probability distribution

$$\gamma_{\mathrm{I}}(x) = \begin{cases} \frac{\mu(x) - \nu(x)}{\|\mu - \nu\|_{\mathrm{TV}}} & \text{if } \mu(x) > \nu(x) \\ 0 & \text{otherwise} \end{cases}$$

and independently choose Y according to the probability distribution

$$\gamma_{\mathrm{II}}(x) = \begin{cases} \frac{\nu(x)-\mu(x)}{\|\mu-\nu\|_{\mathrm{TV}}} & \text{if } \nu(x) > \mu(x) \\ 0 & \text{otherwise} \end{cases}$$

Clearly,

$$\begin{aligned} p\gamma_{\mathrm{III}} + (1-p)\gamma_{\mathrm{I}} &= \mu \\ p\gamma_{\mathrm{III}} + (1-p)\gamma_{\mathrm{II}} &= \nu \end{aligned}$$

so that the distribution of X is μ and the distribution of Y is ν. Note that in the case that the coin lands tails up, $X \neq Y$ since γ_{I} and γ_{II} are positive on disjoint subsets of $\mathcal{X}$. Thus $X = Y$ if and only if the coin toss is heads. We conclude that

$$\mathbf{P}\{X \neq Y\} = \|\mu - \nu\|_{\mathrm{TV}}$$

□

Definition 1.3.2. *Given a Markov chain on $\mathcal{X}$ with transition matrix P,* ***a Markovian coupling*** *of two P -chains is a Markov chain $\{(X_t, Y_t)\}_{t \geqslant 0}$ with state space $\mathcal{X} \times \mathcal{X}$ which satisfies, for all x, y, x', y'*

$$\begin{aligned} \mathbf{P}\left\{X_{t+1} = x' | X_t = x, Y_t = y\right\} &= P\left(x, x'\right) \\ \mathbf{P}\left\{Y_{t+1} = y' | X_t = x, Y_t = y\right\} &= P\left(y, y'\right) \end{aligned}$$

Any Markovian coupling of Markov chains with transition matrix P can be modified so that the two chains stay together at all times after their first simultaneous visit to a single state more precisely, so that if $X_s = Y_s$, then $X_t = Y_t$ for $t \geqslant s$ To construct such a coupling, simply run the chains according to the original coupling until they meet, then run them together.

NOTATION: If (X_t) and (Y_t) are coupled Markov chains with $X_0 = x$ and $Y_0 = y$, then we will often write $\mathbf{P}_{x,y}$ for the probability on the space where (X_t) and (Y_t) are both defined.

The next theorem is a powerful tool to bound total variation distance.

Theorem 1.3.2. *Let $\{(X_t, Y_t)\}$ be a coupling satisfying (5.2) for which $X_0 = x$ and $Y_0 = y$. Let τ_{couple} be the coalescence time of the chains:*

$$\tau_{couple} := \min\{t : X_s = Y_s \textit{ for all } s \geqslant t\} \tag{1.21}$$

Then

$$\left\|P^t(x, \cdot) - P^t(y, \cdot)\right\|_{\mathrm{TV}} \leqslant \mathbf{P}_{x,y}\left\{\tau_{\text{couple}} > t\right\} \tag{1.22}$$

Proof. Notice that $P^t(x, z) = \mathbf{P}_{x,y}\{X_t = z\}$ and $P^t(y, z) = \mathbf{P}_{x,y}\{Y_t = z\}$ Consequently, (X_t, Y_t) is a coupling of $P^t(x, \cdot)$ with $P^t(y, \cdot)$, whence 1.20 implies that

$$\left\|P^t(x, \cdot) - P^t(y, \cdot)\right\|_{\mathrm{TV}} \leqslant \mathbf{P}_{x,y}\{X_t \neq Y_t\}$$

By construction, $\mathbf{P}_{x,y}\{X_t \neq Y_t\} = \mathbf{P}_{x,y}\{\tau_{\text{couple}} > t\}$, which establishes 1.22 □

Combine above theorems we get the following proposition.

Proposition 1.3.3. *Suppose that for each pair of states $x, y \in \mathcal{X}$ there is a coupling (X_t, Y_t) with $X_0 = x$ and $Y_0 = y$. For each such coupling, let τ_{couple} be the coalescence time of the chains, as defined in 1.21. Then*

$$d(t) \leqslant \max_{x,y \in \mathcal{X}} \mathbf{P}_{x,y}\{\tau_{couple} > t\}$$

and therefore $t_{\text{mix}} \leqslant 4 \max_{x,y} \mathbf{E}_{x,y}(\tau_{couple})$.

The last inequality is by Markov inequality.

1.4 Stationary Times

This chapter we study stationary times of Markov chains, which gives a method to bound mixing times.

Definition 1.4.1. *A stopping time τ for the filtration $\{\mathcal{F}_t\}$ is a $\{0, 1, 2, ...\} \cup \{\infty\}$ valued random variable satisfying $\{\tau = t\} \in \mathcal{F}_t$. For the Markov chains (X_t), we consider the natural filtration generated by $X_1, X_2, \ldots, X_t$.*

An important property in this work is the *strong Markov property*

$$\mathbf{Prob}_{x_0}\{(X_{\tau+1}, X_{\tau+2}, ..., X_{\tau+\ell} \in A | \tau = k \wedge (X_1, ..., X_k) = (x_1, ..., x_k)\} \tag{1.23}$$
$$= \mathbf{Prob}_{x_k}\{(X_1, ..., X_\ell) \in A\}.$$

Definition 1.4.2. *Let (X_t) be a Markov chain with respect to the filtration $\{\mathcal{F}_t\}$, with stationary distribution π. A strong stationary time for (X_t) and starting position x is an $\{\mathcal{F}_t\}$ -stopping time τ, such that for all times t and all y*

$$\mathbf{P}_x\{\tau = t, X_\tau = y\} = \mathbf{P}_x\{\tau = t\}\pi(y). \tag{1.24}$$

In other words, X_τ has distribution π and is independent of τ.

Proposition 1.4.1. *If τ is a strong stationary time for starting state x, then*

$$\left\| P^t(x,\cdot) - \pi \right\|_{\mathrm{TV}} \leqslant \mathbf{P}_x\{\tau > t\} \tag{1.25}$$

To prove this proposition, we introduce the **separation distance** and it suffices to prove two related lemmas.

Definition 1.4.3. *The separation distance of a Markov chain is defined by*

$$s_x(t) := \max_{y \in \mathcal{X}} \left[1 - \frac{P^t(x,y)}{\pi(y)} \right]. \tag{1.26}$$

We also define

$$s(t) := \max_{x \in \mathcal{X}} s_x(t)$$

Lemma 1.4.2. *If τ is a strong stationary time for starting state x, then*

$$s_x(t) \leqslant \mathbf{P}_x\{\tau > t\}$$

Proof.

$$\begin{aligned}
\mathbf{P}_x\left\{\tau \leqslant t, X_t = y\right\} &= \sum_{s \leqslant t} \sum_{z} \mathbf{P}_x\left\{\tau = s, X_s = z, X_t = y\right\} \\
&= \sum_{s \leqslant t} \sum_{z} P^{t-s}(z,y) \mathbf{P}_x\{\tau = s\} \pi(z) \\
&= \mathbf{P}_x\{\tau \leqslant t\} \pi(y)
\end{aligned}$$

So for $x \in \mathcal{X}$. Observe that for every $y \in \mathcal{X}$,

$$\begin{aligned}
1 - \frac{P^t(x,y)}{\pi(y)} = 1 - \frac{\mathbf{P}_x\left\{X_t = y\right\}}{\pi(y)} &\leqslant 1 - \frac{\mathbf{P}_x\left\{X_t = y, \tau \leqslant t\right\}}{\pi(y)} \\
= 1 - \frac{\pi(y)\mathbf{P}_x\{\tau \leqslant t\}}{\pi(y)} &= \mathbf{P}_x\{\tau > t\}
\end{aligned}$$

□

Lemma 1.4.3. *The separation distance $s_x(t)$ satisfies*

$$\left\| P^t(x,\cdot) - \pi \right\|_{\mathrm{TV}} \leqslant s_x(t),$$

and therefore $d(t) \leqslant s(t)$.

Proof.

$$\begin{aligned}
\left\| P^t(x,\cdot) - \pi \right\|_{\mathrm{TV}} = \sum_{P^t(x,y) < \pi(y)} \left[\pi(y) - P^t(x,y)\right] &= \sum_{P^t(x,y) < \pi(y)} \pi(y) \left[1 - \frac{P^t(x,y)}{\pi(y)} \right] \\
&\leqslant \max_{y} \left[1 - \frac{P^t(x,y)}{\pi(y)} \right] = s_x(t)
\end{aligned}$$

□

1.5 Eigenvalues

In this section, we present several important properties of the eigenvalues of transition matrices, which turns out to be very important to understand the asymptotic behavior of Markov chains.

Lemma 1.5.1. *Let P be the transition matrix of a finite Markov chain.*

(i) If λ is an eigenvalue of P, then $|\lambda| \leqslant 1$

(ii) If P is irreducible, the vector space of eigenfunctions corresponding to the eigenvalue 1 is the one-dimensional space generated by the column vector $1 := (1, 1, \ldots, 1)^T$

(iii) If P is irreducible and aperiodic, then -1 is not an eigenvalue of P

Proof. (a) Let $\|f\|_\infty := \max_{x\in\mathcal{X}} |f(x)|$. We have $\|Pf(x)\| = \|\sum P(x,y)f(y)\| \leqslant \|\sum P(x,y)\|f\|_\infty\| = \|f\|_\infty$ for every $x \in \mathcal{X}$. Therefore $\|Pf\|_\infty \leqslant \|f\|_\infty$.

When f is an eigenfunction, we have $\|\lambda\|_\infty \leqslant \|f\|_\infty$. Therefore $|\lambda| \leqslant 1$.

(b) is proved in 1.1.6.

(c) is guaranteed by the Convergence Theorem: convergence of P^n implies convergence of $P^n f$. But if f is an eigenfunction with eigenvalue -1. $P^n f = (-1)^n f$, which is not convergent.

□

Denote by $\langle \cdot, \cdot \rangle$ the usual inner product on $\mathbb{R}^{\mathcal{X}}$, given by $\langle f, g \rangle = \sum_{x\in\mathcal{X}} f(x)g(x)$ We will also need another inner product, denoted by $\langle \cdot, \cdot \rangle_\pi$ and defined by

$$\langle f, g \rangle_\pi := \sum_{x\in\mathcal{X}} f(x)g(x)\pi(x) \tag{1.27}$$

We write $\ell^2(\pi)$ for the vector space $\mathbb{R}^{\mathcal{X}}$ equipped with the inner product (12.1) Recall that the transition matrix P is reversible with respect to the stationary distribution π if $\pi(x)P(x,y) = \pi(y)P(y,x)$ for all $x, y \in \mathcal{X}$. The reason for introducing the inner product 1.27 is given by the following lemma:

Lemma 1.5.2. *Let P be reversible with respect to π*

(i) The inner product space $\left(\mathbb{R}^{\mathcal{X}}, \langle \cdot, \cdot \rangle_\pi\right)$ has an orthonormal basis of real-valued eigenfunctions $\{f_j\}_{j=1}^{|\mathcal{X}|}$ corresponding to real eigenvalues $\{\lambda_j\}$

(ii) The matrix P can be decomposed as

$$\frac{P^t(x,y)}{\pi(y)} = \sum_{j=1}^{|\mathcal{X}|} f_j(x)f_j(y)\lambda_j^t$$

(iii) The eigenfunction f_1 corresponding to the eigenvalue 1 can be taken to be the constant vector $\mathbf{1}$, *in which case*

$$\frac{P^t(x,y)}{\pi(y)} = 1 + \sum_{j=2}^{|\mathcal{X}|} f_j(x)f_j(y)\lambda_j^t \tag{1.28}$$

Proof. Define $A(x,y) := \pi(x)^{1/2}\pi(y)^{-1/2}P(x,y)$. Reversibility of P implies that A is symmetric. The spectral theorem for symmetric matrices guarantees that the inner product space $\left(\mathbb{R}^{\mathcal{X}}, \langle\cdot,\cdot\rangle\right)$ has an orthonormal basis $\{\varphi_j\}_{j=1}^{|\mathcal{X}|}$ such that φ_j is an eigenfunction with real eigenvalue λ_j

We define $\varphi_1 = \sqrt{\pi}$ be a function which takes values of the square root of the stationary distribution. Then

$$A\varphi_1(x) = \sum_y A(x,y)\phi_(y) = \sum_y \pi(x)^{1/2}P(x,y) = \pi(x)^{1/2} = \varphi_1(x).$$

Thus φ_1 is an eigenfunction of A with corresponding eigenvalue $\lambda_1 = 1$.

If D_π denotes the diagonal matrix with diagonal entries $D_\pi(x,x) = \pi(x)$, then $A = D_\pi^{\frac{1}{2}} P D_\pi^{-\frac{1}{2}}$. If $f_j := D_\pi^{-\frac{1}{2}}\varphi_j$, then f_j is an eigenfunction of P with eigenvalue λ_j

$$Pf_j = PD_\pi^{-\frac{1}{2}}\varphi_j = D_\pi^{-\frac{1}{2}}\left(D_\pi^{\frac{1}{2}}PD_\pi^{-\frac{1}{2}}\right)\varphi_j = D_\pi^{-\frac{1}{2}}A\varphi_j = D_\pi^{-\frac{1}{2}}\lambda_j\varphi_j = \lambda_j f_j$$

Although the eigenfunctions $\{f_j\}$ are not necessarily orthonormal with respect to the usual inner product, they are orthonormal with respect to the inner product $\langle\cdot,\cdot\rangle_\pi$ defined in 1.27.

$$\delta_{ij} = \langle\varphi_i,\varphi_j\rangle = \left\langle D_\pi^{\frac{1}{2}}f_i, D_\pi^{\frac{1}{2}}f_j\right\rangle = \langle f_i, f_j\rangle_\pi$$

Considering $\left(\mathbb{R}^{\mathcal{X}}, \langle\cdot,\cdot\rangle_\pi\right)$ with its orthonormal basis of eigenfunctions $\{f_j\}_{j=1}^{|\mathcal{X}|}$, the function δ_y can be written via basis decomposition as

$$\delta_y = \sum_{j=1}^{|\mathcal{X}|}\langle\delta_y, f_j\rangle_\pi f_j = \sum_{j=1}^{|\mathcal{X}|} f_j(y)\pi(y)f_j$$

(1.29)

since $P^t f_j = \lambda_j^t f_j$ and $P^t(x,y) = \left(P^t\delta_y\right)(x)$

$$P^t(x,y) = \sum_{j=1}^{|\mathcal{X}|} f_j(y)\pi(y)\lambda_j^t f_j(x)$$

Dividing by $\pi(y)$ completes the proof of (ii), and (iii) follows from observations above.

$\square$

Lemma 1.5.3. *If φ is an eigenfunction of the transition matrix P with eigenvalue $\lambda \neq 1$, then $E_\pi(\varphi) = 0$*

Proof. Multiplying the equation $P\varphi = \lambda\varphi$ on the left by the stationary distribution π shows that

$$E_\pi(\varphi) = \pi P\varphi = \lambda E_\pi(\varphi).$$

Then $E_\pi(\varphi) = 0$ when $\lambda \neq 1$. $\square$

1.6 The Relaxation Time and Spectral Gap

Definition 1.6.1. $\lambda_\star := \max\{|\lambda| : \lambda$ *is an eigenvalue of* $P, \lambda \neq 1\}$ *The difference* $\gamma_\star := 1 - \lambda_\star$ *is called the* ***absolute spectral gap****.* *1.5.1 implies that if P is aperiodic and irreducible, then $\gamma_\star > 0$*

For a reversible transition matrix P, we label the eigenvalues of P in decreasing order:

$$1 = \lambda_1 > \lambda_2 \geqslant \cdots \geqslant \lambda_{|\mathcal{X}|} \geqslant -1$$

The ***spectral gap*** *of a reversible chain is defined by $\gamma := 1 - \lambda_2$.*

Lemma 1.6.1. *Let $P_L = (P + I)/2$ be the transition matrix of the lazy version of the chain with transition matrix P. Show that all the eigenvalues of P_L are nonnegative. Therefore, for $\frac{1}{2}$-lazy markov chains, $\gamma_\star = \gamma$.*

Proof. $P_L f = \lambda f$ implies $Pf = (2\lambda - 1)f$.

Then from 1.5.1 we have $-1 \leqslant 2\lambda - 1 \leqslant 1$. Thus $0 \leqslant \lambda \leqslant 1$. $\square$

Similarly, for $\frac{1}{n}$-th lazy Markov chains we have

$$-1 + \frac{2}{n} \leqslant \lambda \leqslant 1. \tag{1.30}$$

Definition 1.6.2. *The relaxation time t_{rel} of a reversible Markov chain with absolute spectral* gap *$\gamma_\star$ is defined to be*

$$t_{\mathrm{rel}} := \frac{1}{\gamma_\star}$$

We prove upper and lower bounds on the mixing time in terms of the relaxation time and the stationary distribution of the chain.

Theorem 1.6.2. *Let P be the transition matrix of a reversible, irreducible Markov chain with state space $\mathcal{X}$, and let $\pi_{\min} := \min_{x \in \mathcal{X}} \pi(x)$. Then*

$$t_{\text{mix}}(\varepsilon) \leqslant \left\lceil t_{\text{rel}} \left(\frac{1}{2} \log \left(\frac{1}{\pi_{\min}} \right) + \log \left(\frac{1}{2\varepsilon} \right) \right) \right\rceil \leqslant t_{\text{rel}} \log \left(\frac{1}{\varepsilon \pi_{\min}} \right) \tag{1.31}$$

$$t_{\text{mix}}^{(\infty)}(\varepsilon) \leqslant \left\lceil t_{\text{rel}} \log \left(\frac{1}{\varepsilon \pi_{\min}} \right) \right\rceil \tag{1.32}$$

Proof. Using 1.28 and applying the Cauchy-Schwarz inequality yields

$$\left| \frac{P^t(x,y)}{\pi(y)} - 1 \right| \leqslant \sum_{j=2}^{|\mathcal{X}|} |f_j(x) f_j(y)| \, \lambda_\star^t \leqslant \lambda_\star^t \left[\sum_{j=2}^{|\mathcal{X}|} f_j^2(x) \sum_{j=2}^{|\mathcal{X}|} f_j^2(y) \right]^{1/2} \tag{1.33}$$

Using 1.5 and the orthonormality of $\{f_j\}$ shows that

$$\pi(x) = \langle \delta_x, \delta_x \rangle_\pi = \left\langle \sum_{j=1}^{|\mathcal{X}|} f_j(x) \pi(x) f_j, \sum_{j=1}^{|\mathcal{X}|} f_j(x) \pi(x) f_j \right\rangle_\pi = \pi(x)^2 \sum_{j=1}^{|\mathcal{X}|} f_j(x)^2$$

Consequently, $\sum_{j=2}^{|\mathcal{X}|} f_j(x)^2 \leqslant \pi(x)^{-1}$. This bound and 1.33 imply that

$$\left| \frac{P^t(x,y)}{\pi(y)} - 1 \right| \leqslant \frac{\lambda_\star^t}{\sqrt{\pi(x)\pi(y)}} \leqslant \frac{\lambda_\star^t}{\pi_{\min}} = \frac{(1-\gamma_\star)^t}{\pi_{\min}} \leqslant \frac{e^{-\gamma_\star t}}{\pi_{\min}} \tag{1.34}$$

The bound on $t_{\text{mix}}^{(\infty)}(\varepsilon)$ follows from its definition and the above inequality. □

Theorem 1.6.3. *Suppose that $\lambda \neq 1$ is an eigenvalue for the transition matrix P of an irreducible and aperiodic Markov chain. Then*

$$t_{\text{mix}}(\varepsilon) \geqslant \left(\frac{1}{1-|\lambda|} - 1 \right) \log \left(\frac{1}{2\varepsilon} \right).$$

In particular, for reversible chains,

$$t_{\text{mix}}(\varepsilon) \geqslant (t_{\text{rel}} - 1) \log \left(\frac{1}{2\varepsilon} \right). \tag{1.35}$$

Proof. We may assume that $\lambda \neq 0$. Suppose that $Pf = \lambda f$ with $\lambda \neq 1$. By 1.5.3 $E_\pi(f) = 0$. It follows that

$$\left| \lambda^t f(x) \right| = \left| P^t f(x) \right| = \left| \sum_{y \in \mathcal{X}} \left[P^t(x,y) f(y) - \pi(y) f(y) \right] \right| \leqslant \|f\|_\infty 2d(t)$$

With this inequality, we can obtain a lower bound on the mixing time. Taking x with $|f(x)| = \|f\|_\infty$ yields

$$|\lambda|^t \leqslant 2d(t) \tag{1.36}$$

Therefore, $|\lambda|^{t_{\mathrm{mix}}}(\varepsilon) \leqslant 2\varepsilon$, whence

$$t_{\mathrm{mix}}(\varepsilon)\left(\frac{1}{|\lambda|}-1\right) \geqslant t_{\mathrm{mix}}(\varepsilon)\log\left(\frac{1}{|\lambda|}\right) \geqslant \log\left(\frac{1}{2\varepsilon}\right)$$

Minimizing the left-hand side over eigenvalues different from 1 and rearranging finishes the proof. □

Lemma 1.6.4. *For a reversible, irreducible, and aperiodic Markov chain,*

$$\lim_{t\to\infty} d(t)^{1/t} = \lambda_\star$$

Proof. The proof is directly from previous theorems. □

1.7 Distinguishing Statistics

One way to produce a lower bound on the mixing time t_{mix} is to find a statistic f(a real-valued function) on $\mathcal{X}$ such that the distance between the distribution of $f\left(X_t\right)$ and the distribution of f under the stationary distribution π can be bounded from below.

We firstly provide a useful lemma. When μ is a probability distribution on $\mathcal{X}$ and $f : \mathcal{X} \to \Lambda$, The distribution of f is given by μf^{-1}:

$$\left(\mu f^{-1}\right)(A) := \mu\left(f^{-1}(A)\right)$$

Lemma 1.7.1. *Let μ and ν be probability distributions on $\mathcal{X}$, and let $f : \mathcal{X} \to \Lambda$ be a function on $\mathcal{X}$, where Λ is a finite set. Then*

$$\|\mu - \nu\|_{\mathrm{TV}} \geqslant \left\|\mu f^{-1} - \nu f^{-1}\right\|_{\mathrm{TV}}$$

Proof. $|\mu f^{-1}(B) - \nu f^{-1}(B)| = |\mu\left(f^{-1}(B)\right) - \nu\left(f^{-1}(B)\right)|$. Since $f^{-1}(B) \subset \mathcal{X}$, we have

$$\max_{B\subset\Lambda}\left|\mu f^{-1}(B) - \nu f^{-1}(B)\right| \leqslant \max_{A\subset\mathcal{X}}|\mu(A) - \nu(A)|.$$

□

Proposition 1.7.2. *For $f : \mathcal{X} \to \mathbb{R}$, define $\sigma_\star^2 := \max\{\mathrm{Var}_\mu(f), \mathrm{Var}_\nu(f)\}$. If*

$$|E_\nu(f) - E_\mu(f)| \geqslant r\sigma_\star.$$

Then,

$$\|\mu - \nu\|_{\mathrm{TV}} \geqslant 1 - \frac{8}{r^2}.$$

In particular, if for a Markov chain (X_t) with transition matrix P the function f satisfies

$$|\mathbf{E}_x[f(X_t)] - E_\pi(f)| \geqslant r\sigma_\star$$

Then,

$$\left\|P^t(x,\cdot) - \pi\right\|_{\mathrm{TV}} \geqslant 1 - \frac{8}{r^2}.$$

Proof. Suppose without loss of generality that $E_\mu(f) \leqslant E_\nu(f)$. If $A = (E_\mu(f) + r\sigma_\star/2, \infty)$, then Chebyshev's inequality yields that

$$\mu f^{-1}(A) = \mu\{f - E_\mu(f) > r\sigma_\star/2\} \leqslant \frac{4}{r^2} \quad \text{and thus} \quad \nu f^{-1}(A) \geqslant 1 - \frac{4}{r^2}$$

Then the result follows by the previous lemma. □

The following gives a better constant in the lower bound.

Proposition 1.7.3. *Let μ and ν be two probability distributions on $\mathcal{X}$, and let f be a real-valued function on $\mathcal{X}$. If*

$$|E_\mu(f) - E_\nu(f)| \geqslant r\sigma$$

where $\sigma^2 = [\mathrm{Var}_\mu(f) + \mathrm{Var}_\nu(f)]/2$, then

$$\|\mu - \nu\|_{\mathrm{TV}} \geqslant 1 - \frac{4}{4 + r^2}$$

1.8 Wilson's Method

A general method due to David Wilson [11] for obtaining a lower bound on mixing time uses an eigenfunction Φ to construct a distinguishing statistic.

For an example of Wilson's Method, see 4.2.2.

Theorem 1.8.1. *(Wilson's method). Let (X_t) be an irreducible aperiodic Markov chain with state space $\mathcal{X}$ and transition matrix P. Let Φ be an eigenfunction of P with real eigenvalue λ satisfying $1/2 < \lambda < 1$. Fix $0 < \varepsilon < 1$ and let $R > 0$ satisfy*

$$\mathbf{E}_x\left(|\Phi(X_1) - \Phi(x)|^2\right) \leqslant R \tag{1.37}$$

for all $x \in \mathcal{X}$. Then for any $x \in \mathcal{X}$

$$t_{\text{mix}}(\varepsilon) \geqslant \frac{1}{2\log(1/\lambda)}\left[\log\left(\frac{(1-\lambda)\Phi(x)^2}{2R}\right) + \log\left(\frac{1-\varepsilon}{\varepsilon}\right)\right] \tag{1.38}$$

Proof. Since

$$\mathbf{E}\left(\Phi\left(X_{t+1}\right)|X_t = z\right) = \lambda\Phi(z) \tag{1.39}$$

for all $t \geqslant 0$ and $z \in \mathcal{X}$, we have

$$\mathbf{E}_x\Phi\left(X_t\right) = \lambda^t\Phi(x) \quad \text{for } t \geqslant 0 \tag{1.40}$$

by induction. Fix a value t, let $z = X_t$, and define $D_t = \Phi\left(X_{t+1}\right) - \Phi(z)$. By 1.39 and 1.37 respectively, we have

$$\mathbf{E}_x\left(D_t|X_t = z\right) = (\lambda - 1)\Phi(z) \tag{1.41}$$

and

$$\mathbf{E}_x\left(D_t^2|X_t = z\right) \leqslant R \tag{1.42}$$

Hence

$$\begin{aligned}
\mathbf{E}_x\left(\Phi\left(X_{t+1}\right)^2|X_t = z\right) &= \mathbf{E}_x\left(\left(\Phi(z) + D_t\right)^2|X_t = z\right) && (1.43)\\
&= \Phi(z)^2 + 2\mathbf{E}_x\left(D_t\Phi(z)|X_t = z\right) + \mathbf{E}_x\left(D_t^2|X_t = z\right) && (1.44)\\
&\leqslant (2\lambda - 1)\Phi(z)^2 + R && (1.45)
\end{aligned}$$

Averaging over the possible values of $z \in \mathcal{X}$ with weights $P^t(x,z) = \mathbf{P}_x\left\{X_t = z\right\}$ gives

$$\mathbf{E}_x\Phi\left(X_{t+1}\right)^2 \leqslant (2\lambda - 1)\mathbf{E}_x\Phi\left(X_t\right)^2 + R$$

Averaging over the possible values of $z \in \mathcal{X}$ with weights $P^t(x,z) = \mathbf{P}_x\left\{X_t = z\right\}$ gives

$$\mathbf{E}_x\Phi\left(X_{t+1}\right)^2 \leqslant (2\lambda - 1)\mathbf{E}_x\Phi\left(X_t\right)^2 + R$$

At this point, we could apply this estimate inductively, then sum the resulting geometric series. It is equivalent (and neater) to subtract $R/(2(1-\lambda))$ from both sides, obtaining

$$\mathbf{E}_x\Phi\left(X_{t+1}\right)^2 - \frac{R}{2(1-\lambda)} \leqslant (2\lambda - 1)\left(\mathbf{E}_x\Phi\left(X_t\right)^2 - \frac{R}{2(1-\lambda)}\right)$$

Iterating the above inequality shows that

$$\mathbf{E}_x \Phi\left(X_t\right)^2 - \frac{R}{2(1-\lambda)} \leqslant (2\lambda - 1)^t \left[\Phi(x)^2 - \frac{R}{2(1-\lambda)}\right]$$

Leaving off the non-positive term $-(2\lambda-1)^t R/[2(1-\lambda)]$ on the right-hand side above shows that

$$\mathbf{E}_x \Phi\left(X_t\right)^2 \leqslant (2\lambda - 1)^t \Phi(x)^2 + \frac{R}{2(1-\lambda)}$$

Combining 1.40 and 1.43 gives

$$\operatorname{Var}_x \Phi\left(X_t\right) \leqslant \left[(2\lambda-1)^t - \lambda^{2t}\right] \Phi(x)^2 + \frac{R}{2(1-\lambda)} < \frac{R}{2(1-\lambda)}. \tag{1.46}$$

Since $2\lambda - 1 < \lambda^2$ ensures the first term is negative. 1.5.3 implies that $E_\pi(\Phi) = 0$. Letting $t \to \infty$ in 1.46 the Convergence Theorem implies that

$$\operatorname{Var}_\pi(\Phi) \leqslant \frac{R}{2(1-\lambda)}$$

Applying Proposition 1.7.3 with $r^2 = \frac{2(1-\lambda)\lambda^{2t}\Phi(x)^2}{R}$ gives

$$\left\|P^t(x,\cdot) - \pi\right\|_{\mathrm{TV}} \geqslant \frac{r^2}{4+r^2} = \frac{(1-\lambda)\lambda^{2t}\Phi(x)^2}{2R + (1-\lambda)\lambda^{2t}\Phi(x)^2} \tag{1.47}$$

If t satisfies

$$(1-\lambda)\lambda^{2t}\Phi(x)^2 > \frac{\varepsilon}{1-\varepsilon}(2R) \tag{1.48}$$

then the right-hand side of 1.47 is strictly greater than ε, whence, $d(t) > \varepsilon$. For any

$$t < \frac{1}{2\log(1/\lambda)}\left[\log\left(\frac{(1-\lambda)\Phi(x)^2}{2R}\right) + \log\left(\frac{1-\varepsilon}{\varepsilon}\right)\right] \tag{1.49}$$

the inequality 1.48 holds, so $t_{\mathrm{mix}}(\varepsilon) > t$. Thus $t_{\mathrm{mix}}(\varepsilon)$ is at least the right-hand side of 1.49.

□

1.9 Heat Kernel

We first introduce **continuous time Markov chains**.

Definition 1.9.1. *Given a transition matrix P, $(X_t)_{t\in[0,\infty)}$ is the continuous-time chain with transition matrix P, if the occurrence of transitions is a Poisson Process. More precisely, let $S_1, S_2, \ldots$ be the transition times that change-of-state occurs. Then $T_i = S_i - S_{i-1}$ is i.i.d. exponential random variables with rate r. At these transition times moves are made according to P.*

Define $N_t := \max\{k : S_k \leqslant t\}$ to be the number of transition times up to and including time t, which is a Poisson Process. Observe that $N_t = k$ if and only if $S_k \leqslant t < S_{k+1}$. From the definition,

$$\mathbf{P}_x\{X_t = y|N_t = k\} = \mathbf{P}_x\{\Phi_k = y\} = P^k(x,y).$$

Definition 1.9.2. *The time t heat kernel associated to P is the transition probabilities from initial states at time 0 to states at time t, i.e.*

$$\begin{aligned} H_t(x,y) &:= \mathbf{P}_x\{X_t = y\} = \sum_{k=0}^{\infty} \mathbf{P}_x\{X_t = y|N_t = k\}\mathbf{P}_x\{N_t = k\} \\ &= \sum_{k=0}^{\infty} \frac{e^{-rt}(rt)^k}{k!} P^k(x,y) \end{aligned}$$

The time t heat kernel associated to P with rate 1 is

$$H_t(P) = e^{-t}\sum_{k=0}^{\infty} \frac{t^k P^k}{k!}. \tag{1.50}$$

Write $\sigma(P)$ for the spectrum, including multiplicity, of P. If $P = \sum_{\lambda\in\sigma(P)} \lambda v_\lambda v_\lambda^t$ is a diagonalization of P in an orthonormal eigenbasis $\{v_\lambda\}_{\lambda\in\sigma(P)}$ then

$$H_t(P) = \sum_{\lambda\in\sigma(P)} e^{(\lambda-1)t} v_\lambda v_\lambda^t. \tag{1.51}$$

Chapter 2

Representation Theory and Fourier Analysis

2.1 Basics of Representation Theory

Definition 2.1.1. *Group representation*

A representation ρ of a group G is a homomorphism from G to $GL(V)$ where V is a vector space. The dimension of V is called the degree of ρ.

If W is a subspace of V and W is stable under G (i.e., $\rho(g)W \subset W$ for $\forall g \in G$), then ρ restricted to W gives a **subrepresentation**. If the representation ρ admits no non-trivial subrepresentation, then ρ is called **irreducible**.

Definition 2.1.2. *For a group G, a homomorphism f from a group representation ρ on V to a group representation σ on W is a function $f : V \to W$ such that $f(\rho_g \cdot v) = \sigma_g f(v)$ for all $g \in G$.*

f is an isomorphism if f is a bijection.

$$\begin{array}{ccc} V & \xrightarrow{\rho} & V \\ {\scriptstyle f}\big\downarrow & & \big\downarrow{\scriptstyle f} \\ W & \xrightarrow{\sigma} & W \end{array}$$

Definition 2.1.3. *Given $W \subset V$ is a subrepresentation, we define the quotient representation*

$$\rho_{V/U} : G \to GL(V/U), \ \textit{with } \rho_U(g)(v + U) = \rho(g)(v) + U$$

Theorem 2.1.1. ***Isomorphism Theorem***. *For a homomorphism $f : V \to W$,*

$$V/ker(f) \cong Im(f)$$

Definition 2.1.4. *Fourier transform*

Suppose P is a probability measure (or, generally any function from $G \to \mathbb{C}$) on finite group G. We define Fourier transform of P by

$$\hat{P}(s) := \sum_s P(s)\rho(t)$$

Lemma 2.1.2. *For any representation ρ, $\widehat{P * Q}(\rho) = \hat{P}(\rho)\hat{Q}(\rho)$*

Proof.

$$\begin{aligned}
\hat{P}(\rho)\hat{Q}(\rho) &= \sum_s \sum_t P(s)Q(t)\rho(st) \\
&= \sum_t Q(t)(\sum_s P(s)\rho(st)) \\
&= \sum_t Q(t)(\sum_s P(st^{-1})\rho(s)) \\
&= \sum_t (\sum_s P(st^{-1})Q(t)\rho(s)) \\
&= \sum_s P * Q(s)\rho(s) \\
&= \widehat{P * Q}(\rho)
\end{aligned}$$

□

Lemma 2.1.3. *Schur's Lemma*

If f is a homomorphism : $V_1 \to V_2$, such that V_1, V_2 are irreducible representations of a group G. Then f is either 0 or invertible.

Proof. If f is not invertible, there are 2 cases:

1. $ker(f) \neq 0$, then $ker(f) = V_1$ since $ker(f)$ is a subrepresentation and V_1 is irreducible. Thus $f = 0$.

2. $Im(f) \neq V_2$, then $Im(f) = 0$, since $Im(f)$ is a subrepresentation of V_2 and V_2 is irreducible. Thus $f = 0$

□

Lemma 2.1.4. *For an irreducible representation V over $\mathbb{C}$ and $f : V \to V$ is a homomorphism, then $f = \lambda I$ for $\lambda \in \mathbb{C}$*

Proof. Let λ be an eigenvalue of f, then $f - \lambda I$ is not invertible. By the previous lemma, $f - \lambda I = 0$ $\square$

Lemma 2.1.5. *For uniform distribution U on G, we have $\hat{U}(\rho) = I$ for the trivial representation ρ, $\hat{U}(\rho) = 0$ for any nontrivial irreducible representation ρ.*

Proof. Notice that $\hat{U(\rho)}$ is an homomorphism from V to V:

$$\hat{U}(\rho)\rho(s) = \frac{1}{|G|}\sum_t U(t)\rho(ts) = \hat{U}(\rho)$$

$$\rho(s)\hat{U}(\rho) = \frac{1}{|G|}\sum_t U(t)\rho(st) = \hat{U}(\rho)$$

Then by Schur's lemma $\hat{U}(\rho) = \lambda I$.

When ρ is trivial, clearly $\lambda = 1$.

When ρ is not trivial, clearly $\lambda = 0$. $\square$

Theorem 2.1.6. *Let $\rho : G \to GL(V)$ be a linear representation of G in V and let W be a subspace of V stable under G. Then there exists a complement $W^\perp$ of W in V which is stable under G.*

Proof. Let $\langle , \rangle_1$ be a scalar inner product on V. Using the average trick, define a new inner product by

$$\langle u, v\rangle = \sum_s \langle \rho(s)u, \rho(s)v\rangle_1.$$

Then $\langle , \rangle$ is invariant: $\langle \rho(s)u, \rho(s)v\rangle = \langle u, v\rangle$. The orthogonal complement of W in V serves as $W^\perp$.

$\square$

Theorem 2.1.7. *Every representation of a finite group G on a complex vector space V is completely reducible (a direct sum of irreducible representations).*

Proof. This is a direct result by applying the above theorem inductively on V.

$\square$

Next, we present an alternative version of Schur's Lemma.

Lemma 2.1.8. *Let h be any linear map of V_1 into V_2. Let*

$$h^0 = \frac{1}{|G|} \sum_t \left(\rho_t^2\right)^{-1} h \rho_t^1$$

Then

(1) If ρ^1 and ρ^2 are not equivalent, $h^0 = 0$

(2) If $V_1 = V_2$ and $\rho^1 = \rho^2$, then h^0 is a constant times the identity, the constant being $\operatorname{tr} h / d_\rho$.

Proof. For any $s, \rho^2_{s^{-1}} h^0 \rho^1_s = \frac{1}{|G|} \Sigma \rho^2_{s^{-1}t^{-1}} h \rho^1_{ts} = \frac{1}{|G|} \Sigma \left(\rho^2_{ts}\right)^{-1} h \rho^1_{ts} = h^0$. If ρ^1 and ρ^2 are not isomorphic then $h^0 = 0$ by Schur's lemma. If $V_1 = V_2, \rho_1 = \rho_2 = \rho$, then h^0 is an homomorphism from V_1 to V_2.

By Schur's lemma $h^0 = cI$. Take the trace of both sides and solve for c. □

Suppose ρ^1 and ρ^2 are given in matrix form

$$\rho_t^1 = \left(r_{i_1 j_1}(t)\right), \quad \rho_t^2 = \left(r_{i_2 j_2}(t)\right)$$

The linear maps h and h^0 are defined by matrices $x_{i_2 i_1}$ and $x^0_{i_2 i_1}$. We have

$$x^0_{i_2 i_1} = \frac{1}{|G|} \sum_{t, j_1, j_2} r_{i_2 j_2}\left(t^{-1}\right) x_{j_2 j_1} r_{j_1 i_1}(t)$$

In case (1), $h^0 \equiv 0$ for all choices of h. This can only happen if the coefficients of $x_{j_2 j_1}$ are all zero (we view $x^0_{j_2 j_1}$ as a linear combination of $x_{j_2 j_1}$). This gives

$$\frac{1}{|G|} \sum_{t \in G} r_{i_2 j_2}\left(t^{-1}\right) r_{j_1 i_1}(t) = 0 \text{ for all } i_1, i_2, j_1, j_2 \tag{2.1}$$

In case (2),

$$x^0_{i_2 i_1} = \frac{\operatorname{tr} h}{d_\rho} = \frac{\sum_{i_1} x_{i_1 i_1}}{d_\rho}$$

, so

$$\frac{1}{|G|} \sum_{t \in G} r_{i_2 j_2}\left(t^{-1}\right) r_{j_1 i_1}(t) = \begin{cases} \frac{1}{d_\rho} & \text{if } i_1 = i_2 \text{ and } j_1 = j_2 \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

Definition 2.1.5. *Characters*

Given a representation φ of a finite group G, the character χ of $\varphi(g)$ is

$$\chi_\varphi(g) = \operatorname{tr}(\varphi(g))$$

By recalling that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$, *we see that* $\chi_\varphi\left(hgh^{-1}\right) = \chi_\varphi\left(gh^{-1}h\right) = \chi_\varphi(g)$. *Therefore, χ is constant on conjugacy class. Hence, χ is a class function.*

Let J be the space of complex-valued functions on group G. Then it has inner product:

$$(f_1, f_2) := \frac{1}{|G|} \sum_{a \in G} f_1(a) \overline{f_2(a)}$$

Lemma 2.1.9. *We have*

(a).

$$\chi_{V \oplus W} = \chi_V + \chi_W$$

(b).If V_1, V_2 are irreducible non-isomorphic representations, then

$$(\chi_{V_1}, \chi_{V_2}) = 0$$

(c).If V is irreducible,

$$(\chi_V, \chi_V) = 1$$

Proof. (a) Notice that If we choose a basis of V_1 and a basis of V_2, together it forms a basis for $V_1 \oplus V_2$. Then the matrix of a representation respect to this basis is the block diagonal matrix. The result follows immediately.

(b)

$$(\chi_{V_1}, \chi_{V_2}) := \frac{1}{|G|} \sum_{a \in G} \operatorname{tr}(\rho(a)) \operatorname{tr}(\sigma(a^{-1})) = 0$$

. The last equality is by 2.1.

(c)

$$(\chi_V, \chi_V) := \frac{1}{|G|} \sum_{a \in G} \operatorname{tr}(\rho(a)) \operatorname{tr}(\rho(a^{-1})) = \frac{1}{|G|} \sum_{a \in G} \sum_{i,j} r_{ii}(a) r_{jj}(a^{-1}) = 1.$$

The last equality is by 2.2.

□

Lemma 2.1.10. *For $V = n_1 V_1 \oplus n_2 V_2 \oplus ... \oplus n_k V_k$, We have*

a.$(\chi_V, \chi_{V_i}) = (n_i \chi_{V_i}, \chi_{V_i})$.

b.$(\chi_V, \chi_V) = \sum_i n_i^2$.

c. There are only finitely many irreducible representations.

Proof. These are direct results by previous lemmas. □

Let the irreducible characters be labelled χ_i. Suppose their degrees are d_i The regular representation is based on a vector space with basis $\{e_s\}, s \in G$ Define $\rho_s(e_t) = e_{st}$. Observe that the underlying vector space can be identified with the set of all functions on G

Proposition 2.1.11. *The character r_G of the regular representation is given by*

$$r_G(1) = |G|$$
$$r_G(s) = 0, \quad s \neq 1$$

Proof. $\rho_1(e_s) = e_s$ so $\operatorname{Tr}\rho_1 = |G|$. For $s \neq 1, \rho_s e_t = e_{st} \neq e_t$ so all diagonal entries of the matrix for ρ_s are zero. i.e., group automorphism doesn't have any fixed points, except the identity map. □

Proposition 2.1.12. *Every irreducible representation W_i is contained in the regular representation with multiplicity equal to its degree.*

Proof.

$$(r_G|\chi_i) = \frac{1}{|G|}\sum_{s\in G} r_G(s)\chi_i^*(s) = \chi_i^*(1) = d_i.$$

□

Proposition 2.1.13. *(a) The degrees d_i satisfy $\sum d_i^2 = |G|$*

(b) If $s \in G$ is different from $1, \Sigma d_i\chi_i(s) = 0$

Proof. The result is immediate by the previous proposition. $r_G(s) = \Sigma d_i\chi_i(s)$. For (a) take $s = 1$, for (b) take any $s \neq 1$. □

Proposition 2.1.14. *(a) Fourier Inversion Theorem. Let f be a function on G, then*

$$f(s) = \frac{1}{|G|}\sum d_i \operatorname{tr}\left(\rho_i\left(s^{-1}\right)\hat{f}\left(\rho_i\right)\right)$$

(b) Plancherel Formula. Let f and h be functions on G, then

$$\Sigma f\left(s^{-1}\right)h(s) = \frac{1}{|G|}\Sigma d_i \operatorname{tr}\left(\hat{f}\left(\rho_i\right)\hat{h}\left(\rho_i\right)\right)$$

Proof. (a). Both sides are linear in f so it is sufficient to check the formula for $f(s) = \delta_t(s)$. Then $\hat{f}(\rho_i) = \rho_i(t)$, and the right side equals

$$\frac{1}{|G|}\sum d_i\chi_i\left(s^{-1}t\right).$$

The result follows by 2.1.13.

(b)Both sides are linear in f; taking $f(s) = \delta_t(s)$, the equation thus is reduced to

$$h\left(t^{-1}\right) = \frac{1}{|G|}\sum d_i \operatorname{tr}\left(\rho_i(t)\hat{h}\left(\rho_i\right)\right)$$

This was proved in part (a).

□

Remark. *For real valued functions , then the Plancherel formula is given by*

$$\sum f(s)h(s) = \frac{1}{|G|}\sum_i \operatorname{tr}\left(\hat{f}(\rho_i)\hat{h}(\rho_i)^*\right). \tag{2.3}$$

Definition 2.1.6. *$f : G \to \mathbb{C}$ is a **class function** on G if it is invariant under conjugation. i.e. $f(h^{-1}gh) = f(g)$.*

Proposition 2.1.15. *Let f be a class function on G. Let $\rho : G \to GL(V)$ be an irreducible representation of G. Then $\hat{f}(\rho) = \lambda I$ with*

$$\lambda = \frac{1}{d_\rho}\sum f(t)\chi_\rho(t) = \frac{|G|}{d_\rho}\left(f|\chi_\rho^*\right)$$

Proof.

$$\rho_s\hat{f}(\rho)\rho_s^{-1} = \sum f(t)\rho(s)\rho(t)\rho\left(s^{-1}\right) = \sum f(t)\rho\left(\mathrm{sts}^{-1}\right) = \hat{f}(\rho).$$

So, by Schur's lemma $\hat{f}(\rho) = \lambda I$. Take traces of both sides and solve for λ □

Proposition 2.1.16. *The characters of the irreducible representations $\{\chi_i\}_{i=1}^k$ (there are only finitely many) form an orthonormal basis for the class functions.*

Proof. We have shown that the characters of irreducible representations are orthogonal. It remains to show they are enough.

Suppose $(f|\chi_i^*) = 0$, for f a class function. Then 2.1.15 gives $\hat{f}(\rho) = 0$ for every irreducible ρ and the inversion theorem gives $f = 0$. □

Lemma 2.1.17. ***(Upper bound lemma.)*** *Let Q be a probability on the finite group G. Then*

$$\|Q - U\|^2 \leqslant \frac{1}{4}\sum^* d_\rho \operatorname{Tr}\left(\hat{Q}(\rho)\overline{Q}(\rho)^*\right)$$

where the sum is over all non-trivial irreducible representations. Then,

$$\|Q - U\|^2 \leqslant \frac{1}{4}\sum^* d_\rho \operatorname{Tr}\left(\hat{Q}(\rho)\hat{Q}(\rho)^*\right)$$

where the sum $\sum^*$ is over all non-trivial irreducible representations.

Proof.

$$4\|Q-U\|^2 = \left\{\sum_s |Q(s)-U(s)|\right\}^2 \leqslant |G| \sum_s |Q(s)-U(s)|^2$$

$$= \sum^* d_\rho \operatorname{Tr}\left(\hat{Q}(\rho)\hat{Q}(\rho)^*\right)$$

The inequality is by Cauchy-Schwarz. The final equality is by 2.3, and $\hat{U}(\rho) = 1$ for ρ trivial, $\hat{U}(\rho) = 0$ for ρ non-trivial by 2.1.5. □

2.2 Connections with Markov chains

As shown in [3] chapter 3E, we present the connection between representation theory and random walk on groups as defined in 1.1.9, where the eigenvalues of the transition matrix are precisely the eigenvalues of the Fourier transform of the probability measure with respect to the regular representation, each appearing with multiplicity d_ρ.

Let finite group $G = \{s_1, \ldots, s_N\}, N = |G|$ Given a probability measure Q on a group G. Recall that the transition probability is given by $Q(s,t) = Q(ts^{-1})$. We denote $Q(i,j) = Q\left(s_j s_i^{-1}\right)$.

Suppose irreducible representations are numbered as $\rho_1, \ldots, \rho_K$. Define

$$M_k = \begin{pmatrix} \hat{Q}(\rho_k) & & 0 \\ & \ddots & \\ 0 & & \hat{Q}(\rho_k) \end{pmatrix},$$

a $d_k^2 \times d_k^2$ block matrix with $\hat{Q}(\rho_k)$ the Fourier transform of Q at ρ_k.

Let M be the $N \times N$ block diagonal matrix $\begin{pmatrix} M_1 & & 0 \\ & \ddots & \\ 0 & & M_K \end{pmatrix}$ Choose a basis such that each irreducible representation is given by a unitary matrix. Define

$$\psi_k(s) = \sqrt{\frac{d_k}{N}} \left(\rho_k(s)_{11}, \rho_k(s)_{21}, \ldots, \rho_k(s)_{d_k 1}, \rho_k(s)_{12}, \ldots, \rho_k(s)_{d_k d_k}\right)^T$$

a column vector of length d_k^2. Let $\phi(s) = \left(\psi_1(s)^T, \psi_2(s)^T, \ldots, \psi_K(s)^T\right)^T$ be a column vector of length N obtained by concatenating the $\psi_k(s)$ vectors.

Let ϕ be the $N \times N$ matrix $(\phi(s_1), \ldots, \phi(s_N))$ and ϕ^* its conjugate transpose.

Theorem 2.2.1. *Then transition matrix $Q(i,j)$ satisfies*

$$Q = \phi^* M^* \phi.$$

Remark. *The Schur orthogonality relations show that ϕ is a unitary matrix. It implies that each eigenvalue of $\hat{Q}(\rho)$ is an eigenvalue of $Q(i,j)$ with multiplicity d_ρ. Together these are all the eigenvalues of $Q(i,j)$. If M is diagonal, then (5) is the spectral decomposition of Q with respect to an orthonormal basis of eigenvectors.*

Also, $\operatorname{tr} Q = \operatorname{tr} \phi^* M^* \phi = \operatorname{tr} M^* \phi \phi^* = \operatorname{tr} M^*$.

Proof.

$$Q(i,j) = \frac{1}{N} \sum_{k=1}^{K} d_k \operatorname{Tr}\left[\hat{Q}\left(\rho_k\right) \rho_k\left(s_i\right) \rho_k\left(s_j^{-1}\right)\right] = \frac{1}{N} \sum_{k=1}^{K} d_k \operatorname{Tr}\left[\rho_k\left(s_j^{-1}\right) \hat{Q}\left(\rho_k\right) \rho_k\left(s_i\right)\right]$$

Expanding the trace, this equals

$$\sum_{k=1}^{K} \psi_k\left(s_j\right)^* M_k \psi_k\left(s_i\right).$$

$\square$

Chapter 3

Green's function on lattice points and Harmonic Extension.

This chapter combines the idea of harmonic extension, as illustrated in [2], and harmonic functions discussed in [7].

The return time of a Markov chain at a vertex is the first positive time after leaving the vertex at which the Markov chain again is at the vertex. The hitting time of a Markov chain to a set B is the first non-negative time at which the chain reaches the set. We require some estimates for return times and hitting times of simple random walk on the torus $(\mathbb{Z}/n\mathbb{Z})^2$, as well as the hitting probabilities regarding the likelihood of first reaching individual vertices in sets.

The following proposition can be used to estimate return probabilities.

Proposition 3.0.1. *Let (X_t) be a Markov chain on a finite state space $\mathscr{X}$ with irreducible transition matrix P, let $B \subset \mathscr{X}$, and let $h_B : B \to \mathbb{R}$ be a function defined on B. The function $h : \mathscr{X} \to \mathbb{R}$ defined by*

$$h(x) := \mathbf{E}_x h_B(X_{\tau_B}) \tag{3.1}$$

is the unique extension $h : \mathscr{X} \to \mathbb{R}$ of h_B such that $h(x) = h_B(x)$ for all $x \in B$ and h is harmonic for P at all $x \in \mathscr{X} \backslash B$.

The function h is called the 'harmonic extension' of h_B to $\mathscr{X} \backslash B$.

The following lemma can be deduced from Proposition 3.0.1. Let X_t be $\frac{1}{5}$-lazy simple random walk on $(\mathbb{Z}/n\mathbb{Z})^2$ started from $(1, 0)$, and let $p_{y,n}, y \in \{(1, 0), (-1, 0), (0, 1), (0, -1)\}$ be the probability that X_t first reaches $(0, 0)$ from y.

Lemma 3.0.2. *There are limiting probabilities $p_{(1,0)}, p_{(-1,0)}, p_{(0,1)} = p_{(0,-1)} > 0$ such that $p_{y,n} \to p_y$ as $n \to \infty$.*

Proof. Since simple random walk on $\mathbb{Z}^2$ is recurrent with probability 1, the probability of the first return to 0 taking more than n steps tends to 0 as $n \to \infty$. Those return paths which take fewer than n steps are the same on $(\mathbb{Z}/n\mathbb{Z})^2$ as on $\mathbb{Z}^2$, which proves the limit. $\square$

Next we calculate the limiting return probabilities p_x. Firstly we introduce the Green's functions on $\mathbb{Z}^2$, which could be used to represent harmonic functions on $\mathbb{Z}^2$.

Definition 3.0.1. *We define Green's function on $\mathbb{Z}^2$ as following: for $x \in \mathbb{Z}^2$,*

$$G_{\mathbb{Z}^2}(x) = \frac{1}{4}\sum_{n=0}^{\infty}[v^{*n}(x) - v^{*n}(0,0)]$$

, where $v = \frac{1}{4}[\delta_{(0,1)} + \delta_{(0,-1)} + \delta_{(1,0)} + \delta_{(-1,0)}]$.

It's easy to verify $G_{\mathbb{Z}^2}(x)$ is harmonic modulo 1 on $\mathbb{Z}^2$.

To calculate $G_{\mathbb{Z}^2}(x)$, consider the Fourier Transformation:

$$\hat{v}(n) = \int_{\mathbb{R}^d/\mathbb{Z}^d} f(x)e(n \cdot x)dx. \tag{3.2}$$

,where $e(x) = e^{2\pi i x}$.

For v:

$$\hat{v}(\zeta_1, \zeta_2) = \frac{\cos(2\pi\zeta_1) + \cos(2\pi\zeta_2)}{2}$$

So

$$\widehat{G_{\mathbb{Z}^2}}(\zeta) = \frac{1}{4}\sum_{n=0}^{\infty}\widehat{v^{*n}}(\zeta_1,\zeta_2) = \frac{1}{4}\sum_{n=0}^{\infty}\hat{v}^n = \frac{1}{4 - 2\cos(2\pi\zeta_1) - 2\cos(2\pi\zeta_2)}$$

Then by Fourier inversion formula, we can compute $G_{\mathbb{Z}^2}(x)$ by calculating the integral:

$$G_{\mathbb{Z}^2}(x) = \int_{\mathbb{R}^2/\mathbb{Z}^2} e(-\zeta \cdot x)\widehat{G_{\mathbb{Z}^2}}(\zeta)d\zeta \tag{3.3}$$

We give the return probability to the origin in the following lemmas.

Lemma 3.0.3. *Started at $(1,0)$, the return probability to the origin is given by*

$$p_{(1,0)} = \frac{1}{2}, \qquad p_{(0,\pm 1)} = \frac{1}{2} - \frac{1}{\pi}, \qquad p_{(-1,0)} = \frac{2}{\pi} - \frac{1}{2}. \tag{3.4}$$

Proof. We work first on $(\mathbb{Z}/n\mathbb{Z})^2$ and then take the limit as $n \to \infty$. Denote $G_{x,n}$ the Green's function on $(\mathbb{Z}/n\mathbb{Z})^2$ started at 0 and evaluated at x and G_x the Green's function started at 0 and evaluated at x on $\mathbb{Z}^2$. We define function h_B on set B, where $B = \{(0,0), (-1,0)\}$, and extend h_B on $(\mathbb{Z}/n\mathbb{Z})^2$ by Lemma 3.0.1.

By [7], any harmonic modulo 1 function on $\mathbb{Z}^2$ is a sum of discrete derivatives of the Green's function.

Let $h((0,0)) = -G_{(1,0),n}$, $h((-1,0)) = G_{(1,0),n}$, then

$$h(x) = G_{x,n} - G_{x-(-1,0),n} \tag{3.5}$$

By conditioning on the state which is first hit in B, we get

$$h(x) = p_{B_{(0,0)},n}(x)h((0,0)) + p_{B_{(-1,0)},n}(x)h((-1,0)), \tag{3.6}$$

where $p_{B_{(0,0)},n}(x)$ is, starting at point x the probability of hitting the origin when first hitting set B and $p_{B_{(-1,0)},n}(x)$ is the probability of hitting $(-1,0)$ when first hitting set B. For convenience, let $p_{B_{(0,0)},n}$ and $p_{B_{(-1,0)},n}$ denote $p_{B_{(0,0)},n}(1,0)$ and $p_{B_{(-1,0)},n}(1,0)$.

Since we know $p_{B_{(0,0)},n} + p_{B_{(-1,0)},n} = 1$, plug in $(1,0)$ in (3.5) and we get

$$p_{B_{(0,0)},n} = \frac{G_{(2,0),n}}{2G_{(1,0),n}}, \qquad p_{B_{(-1,0)},n} = 1 - \frac{G_{(2,0),n}}{2G_{(1,0),n}}. \tag{3.7}$$

Plug in $(0,1)$ in (3.5), we get

$$p_{B_{(-1,0)},n}((0,1)) = 1 - \frac{G_{(1,1),n}}{2G_{(1,0),n}}. \tag{3.8}$$

Due to symmetry, we have

$$p_{(0,1),n} = p_{B_{(-1,0)},n}((0,1))p_{(1,0),n}$$

i.e. start at $(0,1)$, calculate $p_{(-1,0),n}$ by conditioning on the state of hitting $(-1,0)$ before hitting the origin.

Let $p'_{(1,0),n}$, $p'_{(0,1),n}$, $p'_{(0,-1),n}$ denote the probability of returning to origin through $(1,0), (0,1), (0,-1)$, without passing through point $(-1,0)$. We have

$$\begin{aligned}
p_{B_{(0,0)},n} &= p'_{(0,1),n} + p'_{(0,-1),n} + p'_{(1,0),n} = 2p'_{(0,1),n} + p'_{(1,0),n} \qquad (3.9)\\
p_{(-1,0),n} &= p_{B_{(-1,0)},n}p_{(1,0),n}\\
p_{(1,0),n} &= p'_{(1,0),n} + p_{B_{(-1,0)},n}p_{(-1,0),n} = \frac{p'_{(1,0),n}}{1 - p^2_{B_{(-1,0)},n}}\\
p_{(0,1),n} &= p'_{(0,1),n} + p_{B_{(-1,0)},n}p_{(0,1),n} = \frac{p'_{(0,1),n}}{1 - p_{B_{(-1,0)},n}}.
\end{aligned}$$

By solving the above linear system, we get the desired quantity of return probabilities,

$$\begin{aligned} p_{(1,0),n} &= \frac{2G_{(1,0),n}}{8G_{(1,0),n} - 2G_{(1,1),n} - G_{(2,0),n}} \\ p_{(0,\pm1),n} &= \frac{2G_{(1,0),n} - G_{(1,1),n}}{8G_{(1,0),n} - 2G_{(1,1),n} - G_{(2,0),n}} \\ p_{(-1,0),n} &= \frac{2G_{(1,0),n} - G_{(2,0),n}}{8G_{(1,0),n} - 2G_{(1,1),n} - G_{(2,0),n}}. \end{aligned} \tag{3.10}$$

Letting $n \to \infty$, $G_{x,n} \to G_x$. The exact values were calculated in Mathematica.

□

Use similar ideas, we can get the return probability of any given start point:

Lemma 3.0.4. *Started at (a,b), where $(a,b) \neq (\pm1,0)$ and $(0,\pm1)$, the return probability to the origin is given by*

$$\begin{aligned} p_{(-1,0)}(a,b) &= \frac{G_{(a,b)} - G_{(a+1,b)}}{4G_{(1,0)}} + \frac{1}{4} \\ p_{(1,0)}(a,b) &= \frac{G_{(a,b)} - G_{(a-1,b)}}{4G_{(1,0)}} + \frac{1}{4} \\ p_{(0,1)}(a,b) &= \frac{G_{(a,b)} - G_{(a,b-1)}}{4G_{(1,0)}} + \frac{1}{4} \\ p_{(0,-1)}(a,b) &= \frac{G_{(a,b)} - G_{(a,b+1)}}{4G_{(1,0)}} + \frac{1}{4}. \end{aligned} \tag{3.11}$$

Proof. Using the same set B and function h in lemma 3.0.3, plug in (a,b) we get:

$$\begin{aligned} p_{B_{(0,0)},n}(a,b) &= -\frac{G_{(a,b),n} - G_{(a+1,b),n}}{2G_{(1,0),n}} + \frac{1}{2} \\ p_{B_{(-1,0)},n}(a,b) &= \frac{G_{(a,b),n} - G_{(a+1,b),n}}{2G_{(1,0),n}} + \frac{1}{2}. \end{aligned} \tag{3.12}$$

By similar ideas in lemma 3.0.3, conditioning on the state hitting point $(-1,0)$, we get:

$$p_{(-1,0),n}(a,b) = p_{B_{(-1,0)},n}(a,b)p_{(1,0),n} = p_{(1,0),n}\left(\frac{G_{(a,b),n} - G_{(a+1,b),n}}{2G_{(1,0),n}} + \frac{1}{2}\right)$$

By symmetry (i.e. reflection by line $x = 0$ and $y = x$, rotation 90 degree clockwise), we observe that

$$p_{(0,-1),n}(a, b) = p_{(-1,0),n}(b, a)$$

$$p_{(1,0),n}(a, b) = p_{(-1,0),n}(-a, b)$$
$$p_{(0,1),n}(a, b) = p_{(-1,0),n}(-b, a)$$

The right hand sides are immediate from the above identity we get.

Letting $n \to \infty$, $G_{x,n} \to G_x$, $p_{(1,0),n} \to \frac{1}{2}$, the desired quantity is shown in the lemma.

□

Lemma 3.0.5. *Given point $P = (a, b)$, started at a neighbor of P, the probability of hitting the origin without passing through P is given by:*

$$\begin{aligned}
p_{B_{(0,0),n}}(a-1, b) &= -\frac{G_{(a-1,b),n} - G_{(1,0),n}}{2G_{(a,b),n}} + \frac{1}{2} \\
p_{B_{(0,0),n}}(a+1, b) &= -\frac{G_{(a+1,b),n} - G_{(1,0),n}}{2G_{(a,b),n}} + \frac{1}{2} \\
p_{B_{(0,0),n}}(a, b-1) &= -\frac{G_{(a,b-1),n} - G_{(1,0),n}}{2G_{(a,b),n}} + \frac{1}{2} \\
p_{B_{(0,0),n}}(a, b+1) &= -\frac{G_{(a,b+1),n} - G_{(1,0),n}}{2G_{(a,b),n}} + \frac{1}{2}.
\end{aligned} \tag{3.13}$$

Proof. Let $B = \{(0, 0), (a, b)\}$ and $h_B(0, 0) = -G_{(a,b),n}$, $h_B(a, b) = G_{(a,b),n}$. Use the harmonic extension and the calculation is similar as we did in lemma 3.0.3 and 3.0.4. □

Chapter 4

Examples

In this chapter, we give several examples of Markov chains and random walk on groups. Also, we analyze their mixing time based on methods we have developed so far. To guarantee convergence of chains, we always want to make sure the chain is irreducible and aperiodic.

This chapter is mostly based on [2] chapter 6 and 8, and [11].

The first example is a classic Markov chain which shall be used later.

4.1 Coupon collecting

Question: A company issues n different types of coupons. A collector desires a complete set. We suppose each coupon he acquires is equally likely to be each of the n types. How many coupons must he obtain so that his collection contains all n types?

Let X_t denote the number of different types represented among the collector's first t coupons. Clearly $X_0 = 0$. When the collector has coupons of k different types, there are nk types missing. Of the n possibilities for his next coupon, only nk will expand his collection. Hence

$$\mathbf{P}\left\{X_{t+1} = k+1 | X_t = k\right\} = \frac{n-k}{n}$$

and

$$\mathbf{P}\left\{X_{t+1} = k | X_t = k\right\} = \frac{k}{n}.$$

Thus we see this is indeed a Markov chain. Once the chain arrives at state n (corresponding to a complete collection), it is absorbed there. We are interested in the number of steps required to reach the absorbing state.

Proposition 4.1.1. *Consider a collector attempting to collect a complete set of coupons. Assume that each new coupon is chosen uniformly and independently from the set of n possible types, and let τ be the (random) number of coupons collected when the set first contains every type. Then*

$$\mathbf{E}(\tau) = n \sum_{k=1}^{n} \frac{1}{k}$$

Proof. The expectation $\mathbf{E}(\tau)$ can be computed by writing τ as a sum of geometric random variables. Let τ_k be the total number of coupons accumulated when the collection first contains k distinct coupons. Then

$$\tau = \tau_n = \tau_1 + (\tau_2 - \tau_1) + \cdots + (\tau_n - \tau_{n-1})$$

Furthermore, $\tau_k - \tau_{k-1}$ is a geometric random variable with success probability $(n-k+1)/n$: after collecting τ_{k-1} coupons, there are $n-k+1$ types missing from the collection. Each subsequent coupon drawn has the same probability $(n-k+1)/n$ of being a type not already collected, until a new type is finally drawn. Thus $\mathbf{E}(\tau_k - \tau_{k-1}) = n/(n-k+1)$ and

$$\mathbf{E}(\tau) = \sum_{k=1}^{n} \mathbf{E}(\tau_k - \tau_{k-1}) = n \sum_{k=1}^{n} \frac{1}{n-k+1} = n \sum_{k=1}^{n} \frac{1}{k}$$

□

Proposition 4.1.2. *Let τ be a coupon collector random variable, as defined above. For any $c > 0$*

$$\mathbf{P}\{\tau > \lceil n \log n + cn \rceil\} \leqslant e^{-c}$$

Proof. Let A_i be the event that the i-th type does not appear among the first $\lceil n \log n + cn \rceil$ coupons drawn. Observe first that

$$\mathbf{P}\{\tau > \lceil n \log n + cn \rceil\} = \mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{i=1}^{n} \mathbf{P}(A_i)$$

since each trial has probability $1 - n^{-1}$ of not drawing coupon i and the trials are independent, the right-hand side above is equal to

$$\sum_{i=1}^{n} \left(1 - \frac{1}{n}\right)^{\lceil n \log n + cn \rceil} \leqslant n \exp\left(-\frac{n \log n + cn}{n}\right) = e^{-c}.$$

The first inequality is by $1 + x < e^x$. □

Proposition 4.1.3. *Consider the coupon collecting problem with n distinct coupon types, and let $I_j(t)$ be the indicator of the event that the j -th coupon has not been collected by time t. Let $R_t = \sum_{j=1}^{n} I_j(t)$ be the number of coupon types not collected by time t. The random variables $I_j(t)$ are negatively correlated, and letting $p = \left(1 - \frac{1}{n}\right)^t$, we have for $t \geqslant 0$*

$$\mathbf{E}\left(R_t\right) = np$$
$$\mathrm{Var}\left(R_t\right) \leqslant np(1-p) \leqslant \frac{n}{4}$$

Proof. Since $I_j(t) = 1$ if and only if the first t coupons are not of type j, it follows that

$$\mathbf{E}\left(I_j(t)\right) = \left(1 - \frac{1}{n}\right)^t = p \quad \text{and} \quad \mathrm{Var}\left(I_j(t)\right) = p(1-p)$$

Similarly, for $j \neq k$

$$\mathbf{E}\left(I_j(t) I_k(t)\right) = \left(1 - \frac{2}{n}\right)^t$$

whence

$$\mathrm{Cov}\left(I_j(t), I_k(t)\right) = \left(1 - \frac{2}{n}\right)^t - \left(1 - \frac{1}{n}\right)^{2t} \leqslant 0$$

$\square$

The next two examples emphasize the method of coupling.

4.2 Random walk on the hypercube

The n**-dimensional hypercube** is a graph whose vertices are the binary n tuples $\{0,1\}^n$. Two vertices are connected by an edge when they differ in exactly one coordinate. The simple random walk on the hypercube moves from a vertex $(x^1, x^2, \ldots, x^n)$ by choosing a coordinate $j \in \{1, 2, \ldots, n\}$ uniformly at random and setting the new state equal to $(x^1, \ldots, x^{j-1}, 1 - x^j, x^{j+1}, \ldots, x^n)$. That is, the bit at the walk's chosen coordinate is flipped.

It's easy to see that the simple random walk on the hypercube is periodic. To avoid the periodicity, we consider **the lazy random walk**, which does not have this problem. It remains at its current position with probability 1/2 and moves as above with probability 1/2.

A convenient way to generate the lazy walk is as follows: pick one of the n coordinates uniformly at random, and refresh the bit at this coordinate with a random fair bit (one which equals 0 or 1 each with probability 1/2).

This leads to the following coupling of two walks with possibly different starting positions: first, pick among the n coordinates uniformly at random; suppose that coordinate i is selected. In both walks, replace the bit at coordinate i with the same random fair bit.

If τ is the first time when all of the coordinates have been selected at least once, then the two walkers agree with each other from time τ onwards. (If the initial states agree in some coordinates, the first time the walkers agree could be strictly before τ.) The distribution of τ is exactly the same as the coupon collector random variable as discussed in the previous example.

Thus by 1.3.3 and 4.1.2,

$$d(n \log n + cn) \leqslant \mathbf{P}\{\tau > n \log n + cn\} \leqslant e^{-c}$$

It is immediate from the above that

$$t_{\mathrm{mix}}(\varepsilon) \leqslant n \log n + \log(1/\varepsilon)n.$$

To prove the mixing time lower bound, we use the method of distinguished statistics.

Proposition 4.2.1 ([2] Proposition 7.14)**.** *For the lazy random walk on the n-dimensional hypercube*

$$d\left(\frac{1}{2}n \log n - \alpha n\right) \geqslant 1 - 8e^{2-2\alpha}$$

Proof. Let $\mathbf{1}$ denote the vector of ones $(1, 1, \ldots, 1)$, and let $W(\boldsymbol{x}) = \sum_{i=1}^{n} x^i$ be the Hamming weight of $\boldsymbol{x} = (x^1, \ldots, x^n) \in \{0, 1\}^n$. We will apply 1.7.2 with $f = W$. The position of the walker at time t, started at $\mathbf{1}$ is denoted by $\boldsymbol{X}_t = (X_t^1, \ldots, X_t^n)$

As π is uniform on $\{0, 1\}^n$, the distribution of the random variable W under π is binomial with parameters n and $p = 1/2$. In particular

$$E_\pi(W) = \frac{n}{2}, \quad \mathrm{Var}_\pi(W) = \frac{n}{4}$$

Let R_t be the number of coordinates not updated by time t. When starting from 1, the conditional distribution of $W(\boldsymbol{X}_t)$ given $R_t = r$ is the same as that of $r + B$, where B is a binomial random variable with parameters $n - r$ and $1/2$ Consequently,

$$\mathbf{E}_{\mathbf{1}}\left(W\left(\boldsymbol{X}_t\right) | R_t\right) = R_t + \frac{(n - R_t)}{2} = \frac{1}{2}\left(R_t + n\right).$$

By 4.1.3,

$$\mathbf{E_1}\left(W\left(\boldsymbol{X}_t\right)\right)=\frac{n}{2}\left[1+\left(1-\frac{1}{n}\right)^t\right]$$

Using the identity

$$\operatorname{Var}_1\left(W\left(X_t\right)\right)=\operatorname{Var}_1\left(\mathbf{E}\left(W\left(X_t\right)\mid R_t\right)\right)+\mathbf{E}_1\left(\operatorname{Var}_1\left(W\left(X_t\right)\mid R_t\right)\right),$$

$$\operatorname{Var}_1\left(W\left(\boldsymbol{X}_t\right)\right)=\frac{1}{4}\operatorname{Var}_{\mathbf{1}}\left(R_t\right)+\frac{1}{4}\left[n-\mathbf{E_1}\left(R_t\right)\right]$$

By 4.1.3, R_t is the sum of negatively correlated indicators, and consequently $\operatorname{Var}_{\mathbf{1}}\left(R_t\right)\leqslant\mathbf{E_1}\left(R_t\right)$. We conclude that

$$\operatorname{Var}_{\mathbf{1}}\left(W\left(\boldsymbol{X}_t\right)\right)\leqslant\frac{n}{4}$$

Setting

$$\sigma=\sqrt{\max\left\{\operatorname{Var}_\pi(W),\operatorname{Var}_1\left(W\left(\boldsymbol{X}_t\right)\right)\right\}}=\frac{\sqrt{n}}{2}$$

we have

$$\left|E_\pi(W)-\mathbf{E}_1\left(W\left(\boldsymbol{X}_t\right)\right)\right|=\frac{n}{2}\left(1-\frac{1}{n}\right)^t=\sigma\sqrt{n}\left(1-\frac{1}{n}\right)^t$$

Setting

$$t_n:=\frac{1}{2}(n-1)\log n-(\alpha-1)n>\frac{1}{2}n\log n-\alpha n$$

and using that $(1-1/n)^{n-1}>e^{-1}>(1-1/n)^n$, gives

$$\left|E_\pi(W)-\mathbf{E_1}\left(W\left(\boldsymbol{X}_{t_n}\right)\right)\right|>e^{\alpha-1}\sigma$$

and applying 1.7.2 yields

$$d\left(\frac{1}{2}n\log n-\alpha n\right)\geqslant\left\|P^{t_n}(\mathbf{1},\cdot)-\pi\right\|_{\mathrm{TV}}\geqslant 1-8e^{2-2\alpha}$$

□

In addition, we give a proof based on Wilson's Method discussed in chapter 1.

Proposition 4.2.2. *For the random walk on the n-dimensional hypercube, we have*

$$t_{\mathrm{mix}}(\varepsilon)=\frac{1}{2}n\log n+O(n)$$

Proof. To use Wilson's method, we need to find an eigenfunction on the hypercube. To find the eigenfunction easily, we denote the n-dimensional hypercube as $\{-1,1\}^n$. Notice that, for $n=1$, $f(x) = x(f(-1) = -1, f(1) = 1)$ is the only eigenfunction for the non-lazy walk. For the lazy walk, we define, for $J \subset \{1,\ldots,n\}$,

$$f_J(x) = \prod_{j\in J} x_j.$$

$$Pf_J(x) = \sum_y P(x,y) f_J(y) = \sum_y P(x,y) \prod_{j\in J} y_j \tag{4.1}$$

$P(x,y) > 0$ only when y is a neighbor of x or $y = x$. Uniformly choose a coordinate from 1 to n, then update the coordinate. We see that the two outcomes x and y, both have probability $\frac{1}{2n}$. Notice that if we choose a coordinate that is in J, then $f(x) = -f(y)$. Thus, such x, y contributes 0 to the sum.

If the coordinate is not in J, we have $f(x) = f(y) = \prod_{j\in J} x_j$. Therefore we have

$$4.1 = \frac{2n - 2|J|}{2n} \prod_{j\in J} x_j.$$

Therefore, these are all eigenfunctions on $\{-1,1\}^n$ (since the number is 2^n, same as the cardinality of the state space). Each eigenfunction has associated eigenvalue

$$\lambda_J = \frac{n - |J|}{n}.$$

This gives us all the eigenfunctions and hence

$$\gamma^* = \frac{1}{n} \text{ and hence } t_{\text{rel}} = n.$$

Let $W(x)$ be the Hamming weight of the vector x, i.e. the number of 1's in x. Define $\Phi(x) = W(x) - \frac{n}{2}$.

We see that

$$P\Phi(x) = \sum_y P(x,y)\Phi(y) = \frac{1}{2}[\Phi(x)] + \frac{W(x)}{2n}[\Phi(x) - 1] + \frac{n - W(x)}{2n}[\Phi(x) + 1]$$

$$= W(x) - \frac{n}{2} + \frac{1}{2} - \frac{W(x)}{n} = \left(1 - \frac{1}{n}\right)\Phi(x).$$

Therefore, we see Φ is an eigenfunction with eigenvalue $1 - \frac{1}{n}$.

We apply Φ to the Wilson's Method:

$\mathbb{E}_x\left((\Phi\left(X_1\right)-\Phi(x))^2\right)=\frac{1}{2}$ for all x since Φ changes by exactly 1 whenever the chain moves (i.e., with probability $1/2$).

Hence if we take $R=\frac{1}{2}$ and the initial state to be the all 1 's vector, by 1.8.1:

$$
\begin{aligned}
t_{\mathrm{mix}}(\varepsilon) & \geqslant \frac{1}{-2 \log \left(1-n^{-1}\right)}\left[\log \left\{n^{-1}\left(\frac{n}{2}\right)^2\right\}+\log \{(1-\varepsilon) / \varepsilon\}\right] \qquad (4.2)\\
& \geqslant \frac{1}{2 n^{-1}}\left[\log \frac{n}{4}+\log \{(1-\varepsilon) / \varepsilon\}\right] \\
& =\frac{n}{2} \log n+\frac{n}{2}[\log \{(1-\varepsilon) / \varepsilon\}-\log 4] .
\end{aligned}
$$

□

4.3 Random walk on the torus

The d -dimensional torus is the graph whose vertex set is the Cartesian product

$$
\mathbb{Z}_n^d=\underbrace{\mathbb{Z}_n \times \cdots \times \mathbb{Z}_n}_{d \text { times }}
$$

Vertices $\boldsymbol{x}=\left(x^1, \ldots, x^d\right)$ and $\boldsymbol{y}=\left(y^1, y^2, \ldots, y^d\right)$ are neighbors in $\mathbb{Z}_n^d$ if for some $j \in\{1,2, \ldots, d\}$, we have $x^i=y^i$ for all $i \neq j$ and $x^j \equiv y^j \pm 1 \bmod n$.

When n is even, the graph $\mathbb{Z}_n^d$ is bipartite and the associated random walk is periodic. Again we consider the lazy random walk on $\mathbb{Z}_n^d$ to avoid this complication.

Theorem 4.3.1. *[[2] Theorem 5.6] For the lazy random walk on the d-dimension torus $\mathbb{Z}_n^d$, if $\varepsilon<\frac{1}{2}$ then*

$$
t_{\operatorname{mix}}(\varepsilon) \leqslant d^2 n^2\left\lceil\log _4(d / \varepsilon)\right\rceil
$$

Proof. We use coupling to prove this theorem. To couple together a random walk $\left(\boldsymbol{X}_t\right)$ started at $\boldsymbol{x}$ with a random walk $\left(\boldsymbol{Y}_t\right)$ started at $\boldsymbol{y}$, first pick one of the d coordinates at random. If the positions of the two walks agree in the chosen coordinate, we move both of the walks by +1,-1 or 0 in that coordinate, with probabilities $1/4, 1/4$ and $1/2$, respectively. If the positions of the two walks differ in the chosen coordinate, we randomly choose one of the chains to move, leaving the other fixed. We then move the selected walk by +1 or -1 in the chosen coordinate, with the sign determined by a fair coin toss.

Let $\boldsymbol{X}_t=\left(X_t^1, \ldots, X_t^d\right)$ and $\boldsymbol{Y}_t=\left(Y_t^1, \ldots, Y_t^d\right)$, and let

$$
\tau_i:=\min \left\{t \geqslant 0: X_t^i=Y_t^i\right\}
$$

be the time required for the chains to agree in coordinate i. The clockwise difference between X_t^i and Y_t^i, viewed at the times when coordinate i is selected, behaves just as the coupling of the lazy walk on the cycle $\mathbb{Z}_n$ discussed above. Thus, the expected number of moves in coordinate i needed to make the two chains agree on that coordinate is not more than $n^2/4$ since coordinate i is selected with probability $1/d$ at each move, there is a geometric waiting time between moves with expectation d. It follows that

$$\mathbf{E}_{\boldsymbol{x},\boldsymbol{y}}\left(\tau_i\right) \leqslant \frac{dn^2}{4}$$

The coupling time we are interested in is $\tau_{\text{couple}} = \max_{1\leqslant i\leqslant d}\tau_i$, and we can bound the maximum by a sum to get

$$\mathbf{E}_{\boldsymbol{x},\boldsymbol{y}}\left(\tau_{\text{couple}}\right) \leqslant \frac{d^2n^2}{4},$$

which is true for any x, y. Then by Markov's inequality,

$$\mathbf{P}_{\boldsymbol{x},\boldsymbol{y}}\left\{\tau_{\text{couple}} > t\right\} \leqslant \frac{\mathbf{E}_{\boldsymbol{x},\boldsymbol{y}}\left(\tau_{\text{couple}}\right)}{t} \leqslant \frac{1}{t}\frac{d^2n^2}{4}$$

Taking $t_0 = d^2n^2$ shows that $d\left(t_0\right) \leqslant 1/4$, and so $t_{\text{mix}} \leqslant d^2n^2$. □

4.4 Top-to-Random Shuffle

Consider the following (slow) method of shuffling a deck of n cards: take the top card and insert it uniformly at random in the deck. This process will eventually mix up the deck—the successive arrangements of the deck are a random walk on the group S_n of $n!$ possible permutations of the cards.

Let τ_{top} be the time one move after the first occasion when the original bottom card has moved to the top of the deck. We show now that the arrangement of cards at time τ_{top} is distributed uniformly on the set S_n of all permutations of $\{1, \ldots, n\}$ and moreover this random element of S_n is independent of the time $\tau_{\text{top.}}$

Proposition 4.4.1. *Let (X_t) be the random walk on S_n corresponding to the top-to-random shuffle on n cards. Given at time that there are k cards under the original bottom card, each of the $k!$ possible orderings of these cards are equally likely. Therefore, if τ_{top} is one shuffle after the first time that the original bottom card moves to the top of the deck, then the distribution of $X_{\tau_{top}}$ is uniform over S_n and the time τ_{top} is independent of $X_{\tau_{top}}$. Thus τ_{top} is a strong stationary time for (X_t)*

Proof. When $t = 0$, there are no cards under the original bottom card, and the claim is trivially valid. Now suppose that the claim holds at time t. There are two possibilities at time $t+1$: either a card is placed under the original bottom card, or not. In the second case, the cards under the original bottom card remain in random order. In the first case, given that the card is placed under the original bottom card, each of the $k+1$ possible locations for the card is equally likely, and so each of the $(k+1)!$ orderings are equal likely. □

Proposition 4.4.2. *Let (X_t) be the random walk on S_n corresponding to the top-to-random shuffle on n cards. The corresponding mixing time satisfies*

$$t_{\text{mix}}(\varepsilon) \leqslant n \log n + \log\left(\varepsilon^{-1}\right) n.$$

Proof. Consider the motion of the original bottom card. When there are k cards beneath it, the chance that it rises one card remains $(k+1)/n$ until a shuffle puts the top card underneath it. Thus, the distribution of τ_{top} is the same as the coupon collector's time. Then by 4.1.2 and 1.4.1,

$$d(n \log n + \alpha n) \leqslant e^{-\alpha} \quad \text{for all } n.$$

Therefore,

$$t_{\text{mix}}(\varepsilon) \leqslant n \log n + \log\left(\varepsilon^{-1}\right) n.$$

□

Proposition 4.4.3 ([2] Proposition 7.15)**.** *Let (X_t) be the top-to-random chain on n cards. For any $\varepsilon > 0$, there exists a constant $\alpha(\varepsilon)$ such that $\alpha > \alpha(\varepsilon)$ implies that for all sufficiently large n*

$$d_n(n \log n - \alpha n) \geqslant 1 - \varepsilon$$

That is

$$t_{\text{mix}}(1-\varepsilon) \geqslant n \log n - \alpha n$$

Proof. The bound is based on the following events:

$$A_j = \{ \text{ the original bottom } j \text{ cards are in their original relative order } \} \tag{4.3}$$

Let τ_j be the time required for the card initially j -th from the bottom to reach the top. Then

$$\tau_j = \sum_{i=j}^{n-1} \tau_{j,i}$$

where $\tau_{j,i}$ is the time it takes the card initially j -th from the bottom to ascend from position i (from the bottom) to position $i+1$. The variables $\{\tau_{j,i}\}_{i=j}^{n-1}$ are independent and $\tau_{j,i}$ has a geometric distribution with parameter $p = i/n$, whence $\mathbf{E}(\tau_{j,i}) = n/i$ and $\operatorname{Var}(\tau_{j,i}) < n^2/i^2$. We obtain the bounds.

$$\mathbf{E}(\tau_j) = \sum_{i=j}^{n-1} \frac{n}{i} \geqslant n \int_j^n \frac{dx}{x} = n(\log n - \log j). \tag{4.4}$$

and

$$\operatorname{Var}(\tau_j) \leqslant n^2 \sum_{i=j}^{\infty} \frac{1}{i(i-1)} \leqslant \frac{n^2}{j-1}. \tag{4.5}$$

Using the bounds 4.4 and 4.5 together with Chebyshev's inequality, yields

$$\begin{aligned}\mathbf{P}\{\tau_j < n\log n - \alpha n\} &\leqslant \mathbf{P}\{\tau_j - \mathbf{E}(\tau_j) < -n(\alpha - \log j)\} \\ &\leqslant \frac{1}{j-1}\end{aligned}$$

provided that $\alpha \geqslant \log j + 1$. Define $t_n(\alpha) = n\log n - \alpha n$. If $\tau_j \geqslant t_n(\alpha)$, then the original j bottom cards remain in their original relative order at time $t_n(\alpha)$, so

$$P^{t_n(\alpha)}(\mathrm{id}, A_j) \geqslant \mathbf{P}\{\tau_j \geqslant t_n(\alpha)\} \geqslant 1 - \frac{1}{j-1}$$

for $\alpha \geqslant \log j + 1$. On the other hand, for the uniform stationary distribution

$$\pi(A_j) = 1/(j!) \leqslant (j-1)^{-1}$$

whence, for $\alpha \geqslant \log j + 1$

$$d_n(t_n(\alpha)) \geqslant \left\|P^{t_n(\alpha)}(\mathrm{id}, \cdot) - \pi\right\|_{\mathrm{TV}} \geqslant P^{t_n(\alpha)}(\mathrm{id}, A_j) - \pi(A_j) > 1 - \frac{2}{j-1}$$

Taking $j = \lceil e^{\alpha-1} \rceil$, provided $n \geqslant e^{\alpha-1}$, we have

$$d_n(t_n(\alpha)) > g(\alpha) := 1 - \frac{2}{\lceil e^{\alpha-1} \rceil - 1}$$

Therefore

$$\liminf_{n\to\infty} d_n(t_n(\alpha)) \geqslant g(\alpha)$$

where $g(\alpha) \to 1$ as $\alpha \to \infty$.

$\square$

4.5 Random Transpositions

The Random transposition shuffle is a random walk on the symmetric group S_n, where the driven probability is given at transpositions. We give several different models and analysis, based on methods of coupling, stationary time as described in [2].

The first model is constructed as follows: Choose card X_t and an independent position Y_t uniformly. Exchange X_t with $\sigma_t(Y_t)$ (the card at Y_t).

Coupling of σ_t, σ_t' " Choose **card** X_t and independent **position** Y_t uniformly. "Use X_t and Y_t to update both σ_t and σ_t' Let M_t = number of cards at the same position in σ and σ'

Case 1: X_t in same position, $M_{t+1} = M_t$.

Case 2: X_t in different positions. $\sigma(Y_t) = \sigma'(Y_t)$. $M_{t+1} = M_t$

Case 3: X_t in different positions. $\sigma(Y_t) \neq \sigma'(Y_t)$ $M_{t+1} > M_t$

Proposition 4.5.1. *Let τ^* be the first time $M_t = n$, for any x, y:*

$$\mathbb{E}_{x,y}(\tau^*) < \frac{\pi^2}{6}n^2, \quad t_{mix} = O\left(n^2\right) \tag{4.6}$$

Proof. Let τ_i = steps to increase M_t from $i-1$ to i so

$$\tau^* = \tau_1 + \tau_2 + \cdots + \tau_n$$

As the case analysis discussed above, only when X_t in different positions and $\sigma(Y_t) \neq \sigma'(Y_t)$, M_t shall increase. Both probabilities are $\frac{n-i}{n}$. So

$$\mathbb{P}(M_{t+1} > M_t | M_t = i) = \frac{(n-i)^2}{n^2} \Rightarrow \mathbb{E}(\tau_{i+1} | M_t = i) = \frac{n^2}{(n-i)^2}$$

Therefore, for any x, y

$$\mathbb{E}_{x,y}(\tau^*) \leqslant n^2 \sum_{i=0}^{n-1} \frac{1}{(n-i)^2} < \frac{\pi^2}{6}n^2$$

□

Another different model is the following:

At time t, choose two cards, labelled L_t and R_t, independently and uniformly at random. If L_t and R_t are different, transpose them. Otherwise, do nothing. The resulting distribution μ satisfies

$$\mu(\sigma) = \begin{cases} 1/n & \text{if } \sigma = \text{id} \\ 2/n^2 & \text{if } \sigma = (ij) \\ 0 & \text{otherwise} \end{cases}$$

Proposition 4.5.2 ([2] Proposition 8.6). *In the random transposition shuffle, let R_t and L_t be the cards chosen by the right and left hands, respectively, at time t. Assume that when $t = 0$, no cards have been marked. At time t, mark card R_t if both of the following are true:*

R_t is unmarked.

$-$Either L_t is a marked card or $L_t = R_t$.

Let τ be the time when every card has been marked. Then τ is a strong stationary time for this chain.

Remark. *One way to generate a uniform random permutation is to build a stack of cards, one at a time, inserting each card into a uniformly random position relative to the cards already in the stack. For the stopping time described above, the marked cards are carrying out such a process.*

Lemma 4.5.3. *The stopping time τ defined above satisfies*

$$\mathbf{E}(\tau) = 2n(\log n + O(1))$$

and

$$\operatorname{Var}(\tau) = O\left(n^2\right)$$

Proof. The proof is based on decompose the coupon collector time $\tau = \sum_{i=0}^{n-1} \tau_i$. Then calculate for each τ_i. □

Lemma 4.5.4 ([2]Corollay 8.10). *For the random transposition chain on an n -card deck.*

$$t_{\text{mix}} \leqslant (2 + o(1))n \log n$$

Proof. Let τ be the stopping time defined above and let $t_0 = \mathbf{E}(\tau) + 2\sqrt{\operatorname{Var}(\tau)}$. By Chebyshev's inequality,

$$\mathbf{P}\left\{\tau > t_0\right\} \leqslant \frac{1}{4}$$

Then the result follows by 1.4.1 and the above lemma. □

We present a lower bound as following:

Proposition 4.5.5 ([2] Proposition 8.4). *Let $0 < \varepsilon < 1$. For the random transposition chain on an $n-$ card deck*

$$t_{\text{mix}}(\varepsilon) \geqslant \frac{n-1}{2} \log\left(\frac{1-\varepsilon}{6} n\right)$$

Proof. Let $F(\sigma)$ denote the number of fixed points of the permutation σ. If σ is obtained from the identity by applying t random transpositions, then $F(\sigma)$ is at least as large as the number of cards that were touched by none of the transpositions - no such card has moved, and some moved cards may have returned to their original positions.

Our shuffle chain determines transpositions by choosing pairs of cards independently and uniformly at random. Hence, after t shuffles, the number of untouched cards has the same distribution as the number R_{2t} of uncollected coupon types after $2t$ steps of the coupon collector chain. By 4.1.3

$$\mu := \mathbf{E}\left(R_{2t}\right) = n\left(1-\frac{1}{n}\right)^{2t}$$

and $\operatorname{Var}\left(R_{2t}\right) \leqslant \mu$. Let $A = \{\sigma : F(\sigma) \geqslant \mu/2\}$. We will compare the probabilities of A under the uniform distribution π and $P^t(\text{ id}, \cdot)$. First

$$\pi(A) \leqslant \frac{2}{\mu}$$

by Markov's inequality. By Chebyshev's inequality,

$$P^t\left(\text{ id }, A^c\right) \leqslant \mathbf{P}\left\{R_{2t} \leqslant \mu/2\right\} \leqslant \frac{\mu}{(\mu/2)^2} = \frac{4}{\mu}$$

Then we have, by definition of total variation distance,

$$\left\|P^t(\mathrm{id}, \cdot) - \pi\right\|_{\mathrm{TV}} \geqslant 1 - \frac{6}{\mu}$$

We want to find how small t must be so that $1 - 6/\mu > \varepsilon$, or equivalently

$$n\left(1-\frac{1}{n}\right)^{2t} = \mu > \frac{6}{1-\varepsilon}$$

The above holds if and only if

$$\log\left(\frac{n(1-\varepsilon)}{6}\right) > 2t\log\left(\frac{n}{n-1}\right) \tag{4.7}$$

Using the inequality $\log(1+x) < x$, we have $\log\left(\frac{n}{n-1}\right) < \frac{1}{n-1}$, so the inequality 4.7 holds provided that

$$\log\left(\frac{n(1-\varepsilon)}{6}\right) \geqslant \frac{2t}{n-1}$$

That is, if $t \leqslant \frac{n-1}{2}\log\left(\frac{n(1-\varepsilon)}{6}\right)$, then $d(t) \geqslant 1 - 6/\mu > \varepsilon$.

□

Chapter 5

Dirichlet form, minimax characterization and Comparison Techniques

In this chapter, we consider all Markov Chains to be reversible. We develop methods for getting upper and lower bounds of eigenvalues β_i of the transition kernel by comparison with a second reversible chain on the same state space.:

Let x be a finite set. Let $P(x,y)$ be an irreducible Markov kernel on $\mathscr{X}$ with stationary probability $\pi(x)$. Recall that P, π is reversible:

$$\pi(x)P(x,y) = \pi(y)P(y,x)$$

By symmetry, P has eigenvalues Let $l^2(\mathscr{X})$ have scalar product as we showed in 1.27

$$\langle f, g\rangle = \sum_{x \in X} f(x)g(x)\pi(x).$$

Because of reversibility, the operator $f \mapsto Pf$, with $Pf(x) = \sum f(y)P(x,y)$, is self-adjoint on l^2 with eigenvalues $1 = \beta_0 > \beta_1 \geqslant \cdots \geqslant \beta_{|\mathscr{X}|-1} \geqslant -1$. These eigenvalues can be characterized by the Dirichlet form:

Definition 5.0.1. *We define the* ***Dirichlet Form*** *on f by*

$$\mathscr{E}(f,f) = \langle (I-P)f, f\rangle = \frac{1}{2}\sum_{x,y}(f(x)-f(y))^2\pi(x)P(x,y)$$

Recall the minimax characterization of the eigenvalues:
For a subspace W of $\mathbb{R}^n$, define

$$\begin{aligned} m(W) &= \min\{\langle Pf, f\rangle / \langle f, f\rangle : f \in W\backslash\{0\}\} \\ M(W) &= \max\{\langle Pf, f\rangle / \langle f, f\rangle : f \in W\backslash\{0\}\} \end{aligned}$$

The minimax characterization of eigenvalues gives

$$\lambda_i = \max\left\{m(W) : \dim\left(W^{\perp}\right) = i\right\} = \min\{M(W) : \dim(W) = i+1\}$$

Given a subspace W of $L^2(X)$, set

$$\begin{aligned} M_g(W) &= \max\left\{\mathscr{E}(f,f); \|f\|_2 = 1, f \in W\right\}, \\ m_g(W) &= \min\left\{\mathscr{E}(f,f); \|f\|_2 = 1, f \in W\right\}. \end{aligned}$$

We have

$$\mathbf{1} - \beta_i = \min\left\{M_g(W); \dim W = i+1\right\} = \max\left\{m_g(W); \dim W^{\perp} = i\right\}$$

If $\tilde{P}(x,y), \tilde{\pi}$ is a second reversible Markov chain on X, the minimax characterization yields, for $1 \leqslant i \leqslant |X|-1$

$$\beta_i \leqslant 1 - \frac{a}{A}\left(1 - \tilde{\beta}_i\right), \qquad \text{if } \tilde{\mathscr{E}} \leqslant A\mathscr{E}, \tilde{\pi} \geqslant a\pi$$

In the applications, P, π is the chain of interest and $\tilde{P}, \tilde{\pi}$ is a chain with known eigenvalues. Both π and $\tilde{\pi}$ are assumed to be supported on X. For each pair $x \neq y$ with $\tilde{P}(x,y) > 0$, fix a sequence of steps $x_0 = x, x_1, x_2, \ldots, x_k = y$ with $P\left(x_i, x_{i+1}\right) > 0$. This sequence of steps will be called a path γ_{xy} of length $|\gamma_{xy}| = k$. Set $E = \{(x,y); P(x,y) > 0\}$ $\tilde{E} = \{(x,y); \tilde{P}(x,y) > 0\}$ and $\tilde{E}(e) = \left\{(x,y) \in \tilde{E}; e \in \gamma_{xy}\right\}$, where $e \in E$. In other words, E is the set of "edges" for P and $E(e)$ is the set of paths that contain e. For convention, in this section all graphs are undirected graphs. However, we describe such a graph as a set of vertices X and a symmetric set of directed edges $E \subset X \times X$

Theorem 5.0.1. *Let $\tilde{P}, \tilde{\pi}$ and P, π be reversible Markov chains on a finite set X. The Dirichlet forms $\tilde{\mathscr{E}} \leqslant A\mathscr{E}$*
, with

$$A = \max_{(z,w)\in E}\left\{\frac{1}{\pi(z)P(z,w)} \sum_{\tilde{E}(z,w)} |\gamma_{xy}|\, \tilde{\pi}(x)\tilde{P}(x,y)\right\}. \tag{5.1}$$

Proof. We may assume that none of the paths γ_{xy} contains loops. For an edge $e = (z,w) \in E$, let $f(e) = f(z) - f(w)$. Then

$$\tilde{\mathscr{E}} = \frac{1}{2}\sum_{x,y\in X} (f(x) - f(y))^2 \tilde{\pi}(x)\tilde{P}(x,y)$$

$$= \frac{1}{2} \sum_{x,y} \left\{ \sum_{e \in \gamma_{xy}} f(e) \right\}^2 \tilde{\pi}(x) \tilde{P}(x,y)$$

$$\leqslant \frac{1}{2} \sum_{x,y \in X} |\gamma_{xy}| \, \tilde{\pi}(x) \tilde{P}(x,y) \sum_{e \in \gamma_{xy}} |f(e)|^2$$

$$\leqslant \frac{1}{2} \sum_{e=(z,w)} |f(e)|^2 \frac{\pi(z)P(z,w)}{\pi(z)P(z,w)} \sum_{e \in \gamma_{xy}} |\gamma_{xy}| \, \tilde{\pi}(x) \tilde{P}(x,y)$$

$$\leqslant A\mathscr{E}(f,f),$$

where the first inequality is by Cauchy-Schwarz.

□

In the case that the Markov chains are symmetric random walks on a group, we have the following simplified estimate as shown in [4]. Let E be a symmetric set of generators of a finite group G. For $y \in G$, let $y = z_1 z_2 \cdots z_k$ with $z_i \in E$. Denote the least such k, $|y|$. Let $N(z,y)$ denote the number of times which z appears in the chosen representation of y.

Theorem 5.0.2. *Let $\tilde{p}$ and p be symmetric probabilities on a finite group G. Let E be a symmetric set of generators. Suppose that the support of p contains E. Then the Dirichlet forms satisfy (48)*

$$\tilde{\mathscr{E}} \leqslant A\mathscr{E}$$

with

$$A = \max_{z \in E} \frac{1}{p(z)} \sum_{y \in G} |y| N(z,y) \tilde{p}(y) \tag{5.2}$$

Proof. The proof is similar as we did in the previous Theorem. □

Chapter 6

Counting Problem: an application of mixing time.

In theoretical computer science, there are several counting problems which are hard to solve in polynomial time. For example, counting the number of perfect matches in a graph is P-complete, as illustrated in [12]. However, there is an approximate counting based on random sampling. We can represent the desired number as a expected value of a random variable. We get the approximate counting by calculating the expectation with enough many samples. The complexity of such algorithm, depends on how long we can generate a random sample, i.e. the mixing time of the Markov chain since we use Markov Chains Monte Carlo to generate random samples.

Definition 6.0.1. *Given an undirected graph $G = (V, E)$, a matching $M \subseteq E$ is a set of vertex disjoint edges. A matching is perfect if $|M| = n/2$ where n = number of vertices (and m = number of edges).*

Let e be an arbitrary edge. Use sampling to determine the fraction of matchings that do not use e. And we define the indicator random variable X on matches of G.

If the match contains e, $X = 1$. Otherwise $X = 0$.

We see that

$$E[X] = \frac{\#\text{ matchings without } e}{\#\text{ matchings}}$$

We have $1/2 \leqslant E[X] \leqslant 1$, since for any match containing e we can drop e to get a new match.

Let $X_1 = X$ $G_0 = G$, $G_1 = G \backslash e$, we see also

$$E[X] = \frac{\#\text{ matchings in } G_1}{\#\text{ matchings in } G_0}.$$

Define inductively for X_2 on G_1, we observe that

$$\prod_{i=1}^{m} E[X_i] = \frac{1}{\#\text{ matchings in } G_0}.$$

Chapter 7

The 15 puzzle problem

7.1 Description of the problem

A '15 puzzle' consists of a 4×4 board with 15 numbered unit tiles and one empty square. A move in the puzzle consists of sliding a numbered tile into the empty square. The 15 puzzle gained notoriety in the United States in the 1870's when an article in the American Journal of Math [9] asked whether the board with positions 14 and 15 exchanged and an empty tile in the lower right corner could be shifted into sorted order, again with the empty tile in its initial position, see Figure 7.1 (it cannot, the group of permutations generated is A_{15}). In general, an $n^2 - 1$ puzzle consists of an $n \times n$ board with $n^2 - 1$ numbered tiles and one empty square. In the book [3], Diaconis considers the problem of randomizing an $n^2 - 1$ puzzle given periodic boundary conditions by, at each step, shifting a uniform random neighbor of the open square into its place. He conjectures that the total variation mixing time to randomize the position of a single numbered piece is order n^3, and that the mixing time to stationarity for the whole puzzle is order $n^3 \log n$. The main results in [1] solve Diaconis' '15 puzzle' problem in corrected form.

Theorem 7.1.1. *The $n^2 - 1$ puzzle Markov Chain can be identified with random walk on the group $G_n = S_{n^2-1} \times (\mathbb{Z}/n\mathbb{Z})^2$ driven with the measure $\mu = \frac{1}{5}\left(\delta_{\mathrm{id}} + \delta_R + \delta_L + \delta_U + \delta_D\right)$,*

$$\text{where } R = \begin{bmatrix} (n, n-1, \cdots, 1) \\ (2n, 2n-1, \cdots, n+1) \\ \vdots \\ (n^2-n, n^2-n-1, \cdots, n^2-2n+1) \\ (n^2-1, n^2-2, \cdots, n^2-n+1) \end{bmatrix} \times (1, 0),$$

1	2	3	4
5	6	7	8
9	10	11	12
13	15	14	

Figure 7.1: A 15 puzzle. A move in the puzzle slides a numbered tile into the empty space.

$$U = \begin{bmatrix} (1, n+1, \cdots, n^2-n+1) \\ (2, n+2, \cdots, n^2-n+2) \\ \vdots \\ (n-1, 2n-1, \cdots, n^2-1) \\ (n, 2n, \cdots, n^2-n) \end{bmatrix} \times (0,1) \textit{ and } L = R^{-1}, D = U^{-1}.$$

Proof. While moving up, we consider the empty piece is still on the right down corner by wrapping around the broad into a torus. Thus it's a product of n-cycles. □

In [1], including the above one, several different random walks are discussed. Here we introduce the random walk where we only track the location of the empty square and one piece of labeled square, and we forget the locations of all other pieces.

7.2 Mixing of a single piece

One of the main theorems proved in [1] is as following:

Theorem 7.2.1 ([1], Theorem 1)**.** *Let $d_{\mathrm{Br}}(t)$ be the total variation distance to uniformity at time $t > 0$ of standard Brownian motion started from $(0,0)$ on $(\mathbb{R}/\mathbb{Z})^2$. Let $c_{\mathrm{puz}} = \frac{5}{2}(\pi - 1)$. As $n \to \infty$, the total variation distance to uniformity of a single piece in the $n^2 - 1$ puzzle at time $c_{\mathrm{puz}} n^4 t$ converges to $d_{\mathrm{Br}}(t)$ uniformly for t in compact subsets of $(0, \infty)$.*

The proof is by tracking one or several marked pieces on the board as they move. The pieces move at the times of a renewal process when the empty square moves next to one of them and then the piece is shifted into it. A local limit theorem is proved which demonstrates the approximate independence of the piece's location and the number of moves of the empty square after the marked piece has moved approximately n^2 times. To prove this, we prove several estimates of the characteristic functions.

Since the expected time of a renewal is of order n^2, this takes time order n^2. The lower order fluctuations from the sum of the renewal process times are then absorbed as an error term.

We track the location of a single numbered piece $\mathscr{P}$, along with the empty piece $\mathscr{P}_e$.

Consider stopping times $\{t_i\}_0^\infty$: every time $\mathscr{P}_e$ swapping positions with $\mathscr{P}$ alternatively from vertical and horizontal directions. Here t_0 is time of the first vertical swap. Then t_1 is time of the first horizontal swap after t_0, and so on.

Let H_0 be the number of positions (left is negative, right is positive) that $\mathscr{P}$ moves prior to t_1 and V_0 the number of positions (up is positive, down is negative) that $\mathscr{P}$ moves prior to t_1.

For $i \geqslant 1$, let H_i be the number of horizontal moves (right is positive and left is negative) of $\mathscr{P}$ in $[t_{2i-1}, t_{2i})$ and let V_i be the number of vertical moves (up is positive and down is negative) of $\mathscr{P}$ in $[t_{2i}, t_{2i+1})$.

Lemma 7.2.2. *The collection of random variables $\{H_i, V_i\}_{i=1}^\infty$ are i.i.d. symmetric, mean 0, and have exponentially decaying tails. They are independent of H_0, V_0, and these variables have exponentially decaying tails.*

Proof. By symmetry and strong Markov property, each inter-arrival time r_i, s_i is independent identically distributed. The collection of random variables

$\{H_i, V_i\}_{i=1}^\infty$ are i.i.d. symmetric, mean 0. Also, $\{H_i, V_i\}_{i=1}^\infty$ have exponentially decaying tails.

To see this, recall that by 3.0.2, 3.0.3, we have the return probability $p_{x,n} \to p_x > 0$ as $n \to \infty$ for $x = (1,0), (-1,0), (0,-1), (0,1)$. Then the H_1 is a sum of geometric random variables with parameter less than 1. □

For $i \geqslant 1$, define $r_i = t_{2i} - t_{2i-1}, \quad s_i = t_{2i+1} - t_{2i}$.

The collection $\{(H_i, r_i), (V_i, s_i)\}_{i=1}^\infty$ are also i.i.d. and are independent of $\{H_0, V_0, t_0, t_1\}$. Set

Set $s_n^2 = \mathbf{E}\left[H_1^2\right], \mu_n = \mathbf{E}\left[r_1\right], v_n^2 = \mathbf{Var}\left[r_1\right]$.

Let m_i be the number of times $\mathscr{P}$ moves either left or right between t_{2i-1} and t_{2i} and n_i the number of times $\mathscr{P}$ moves either up or down between t_{2i} and t_{2i+1}. Call a type I return of $\mathscr{P}_e$ a sequence of moves in which $\mathscr{P}_e$ begins adjacent to $\mathscr{P}$, ends at the next time $\mathscr{P}_e$ swaps position with $\mathscr{P}$ from the same direction (horizontal or vertical), and does not swap positions with $\mathscr{P}_e$

from the opposite direction in between. Call a type II return a sequence of moves of $\mathscr{P}_e$ in which it begins adjacent to $\mathscr{P}$, ends adjacent to $\mathscr{P}_e$ from the opposite direction, and does not swap position with $\mathscr{P}_e$ in between, but will swap positions with $\mathscr{P}_e$ from the opposite direction on the next move. Thus r_i is the sum of the length of m_i independent type I returns and one type II return, and s_i is the sum of the lengths of n_i independent type I returns and one type II return.

We have

Proposition 7.2.3.

$$\lim_{n\to\infty} s_n^2 = s^2, \quad \lim_{n\to\infty} \frac{\mu_n}{n^2} = \mu, \textit{ with}$$

$$s^2 = \frac{1}{2p_{(0,\pm1)}} \frac{1 - p_{(1,0)} + p_{(-1,0)}}{1 + p_{(1,0)} - p_{(-1,0)}}, \quad \mu = \frac{5}{4}\left(\frac{1}{2p_{(0,\pm1)}}\right).$$

Also, $v_n = O\left(n^2 \log n\right)$.

Proof. Consider the process in H_1^2. Due to symmetry, without loss of generality we assume the initial move in H_1 is -1, so that afterwards, $\mathscr{P}_e$ is at position $\mathscr{P} + (1,0)$. Then it makes $k \geqslant 0$ right or left returns followed by an up or down return.

Condition on k and let Z_k be the conditional displacement of the moves following the first one. Thus $Z_0 = 0$ and, for $k \geqslant 1$

$$Z_k = \begin{cases} 1 - Z_{k-1} & \text{prob. } \frac{p_{(1,0),n}}{1-2p_{(0,\pm1),n}} \\ -1 + Z_{k-1} & \text{prob. } \frac{p_{(-1,0),n)}}{1-2p_{(0,\pm1),n}} \end{cases}$$

The recurrent relation is based on conditioning the first move: notice that initially $\mathscr{P}_e$ is on the right of $\mathscr{P}$. If the first move is $+1$, afterwards $\mathscr{P}_e$ is on the left of $\mathscr{P}$, which is the opposite from initial thus $Z_t = 1 - Z_{t-1}$. While if the first move is -1, afterward $\mathscr{P}_e$ is on the right of $\mathscr{P}$, thus $Z_t = -1 + Z_{t-1}$.

Hence

$$\mathbf{E}[Z_k] = \frac{p_{(1,0),n} - p_{(-1,0),n}}{1 - 2p_{(0,\pm1),n}}(1 - \mathbf{E}[Z_{k-1}]). \tag{7.1}$$

Solving the recurrence, it follows that

$$\mathbf{E}[Z_k] = \sum_{i=1}^{k}(-1)^{i-1}\left(\frac{p_{(1,0),n} - p_{(-1,0),n}}{1 - 2p_{(0,\pm1),n}}\right)^i. \tag{7.2}$$

Similarly,

$$Z_k^2 = (1 - Z_{k-1})^2 = 1 - 2Z_{k-1} + Z_{k-1}^2.$$

$$\mathbf{E}[Z_k^2] = 1 - \mathbf{E}[2Z_{k-1}] + \mathbf{E}[Z_{k-1}^2] \tag{7.3}$$

from which it follows that

$$\begin{aligned} \mathbf{E}[Z_k^2] &= k - 2\sum_{j=1}^{k-1}\sum_{i=1}^{j}(-1)^{i-1}\left(\frac{p_{(1,0),n} - p_{(-1,0),n}}{1-2p_{(0,\pm1),n}}\right)^i \\ &= k - 2\sum_{i=1}^{k-1}(-1)^{i-1}(k-i)\left(\frac{p_{(1,0),n} - p_{(-1,0),n}}{1-2p_{(0,\pm1),n}}\right)^i . \end{aligned} \tag{7.4}$$

The second equality is because the sum by columns is equal to the sum by rows.

Conditioning on the number k of type I returns,

$$\begin{aligned} \mathbf{E}[H_1^2] &= \sum_{k=0}^{\infty}\mathbf{E}[(-1+Z_k)^2]2p_{(0,\pm1),n}(1-2p_{(0,\pm1),n})^k \\ &= \sum_{k=0}^{\infty}2p_{(0,\pm1),n}(1-2p_{(0,\pm1),n})^k - 2\sum_{k=1}^{\infty}\mathbf{E}[Z_k]2p_{(0,\pm1),n}(1-2p_{(0,\pm1),n})^k \\ &+ \sum_{k=1}^{\infty}\mathbf{E}[Z_k^2]2p_{(0,\pm1),n}(1-2p_{(0,\pm1),n})^k. \end{aligned} \tag{7.5}$$

We have

$$\sum_{k=0}^{\infty}2p_{(0,\pm1),n}(1-2p_{(0,\pm1),n})^k = 1 \tag{7.6}$$

and

$$\begin{aligned} &\sum_{k=1}^{\infty}\mathbf{E}[Z_k]2p_{(0,\pm1),n}(1-2p_{(0,\pm1),n})^k \\ &= 2p_{(0,\pm1),n}\sum_{k=1}^{\infty}\sum_{i=1}^{k}(-1)^{i-1}(p_{(1,0),n} - p_{(-1,0),n})^i(1-2p_{(0,\pm1),n})^{k-i} \\ &= \sum_{i=1}^{\infty}(-1)^{i-1}(p_{(1,0),n} - p_{(-1,0),n})^i \\ &= \frac{p_{(1,0),n} - p_{(-1,0),n}}{1+p_{(1,0),n} - p_{(-1,0),n}}. \end{aligned} \tag{7.7}$$

Also,

$$\sum_{k=1}^{\infty} \mathbf{E}[Z_k^2] 2p_{(0,\pm1),n}(1-2p_{(0,\pm1),n})^k \tag{7.8}$$

$$= 2p_{(0,\pm1),n} \sum_{k=1}^{\infty} k(1-2p_{(0,\pm1),n})^k$$

$$- 4p_{(0,\pm1)} \sum_{k=1}^{\infty} \left(1-2p_{(0,\pm1),n}\right)^k \sum_{i=1}^{k-1} (-1)^{i-1}(k-i) \left(\frac{p_{(1,0),n} - p_{(-1,0),n}}{1-2p_{(0,\pm1),n}}\right)^i$$

$$= 2p_{(0,\pm1),n} \sum_{k=1}^{\infty} k(1-2p_{(0,\pm1),n})^k$$

$$- 4p_{(0,\pm1)} \sum_{k=1}^{\infty} \sum_{i=1}^{k-1} \left(1-2p_{(0,\pm1),n}\right)^{k-i} (-1)^{i-1}(k-i) \left(p_{(1,0),n} - p_{(-1,0),n}\right)^i$$

$$= 2p_{(0,\pm1),n} \sum_{k=1}^{\infty} k(1-2p_{(0,\pm1),n})^k$$

$$- 4p_{(0,\pm1),n} \sum_{i=1}^{\infty} (-1)^{i-1} (p_{(1,0),n} - p_{(-1,0),n})^i \sum_{k=1}^{\infty} k(1-2p_{(0,\pm1),n})^k$$

$$= \frac{1-2p_{(0,\pm1),n}}{2p_{(0,\pm1),n}} - 2\frac{1-2p_{(0,\pm1),n}}{2p_{(0,\pm1),n}} \frac{p_{(1,0),n} - p_{(-1,0),n}}{1+p_{(1,0),n} - p_{(-1,0),n}}.$$

Combining the above obtains

$$\mathbf{E}[H_1^2] = \frac{1}{2p_{(0,\pm1),n}} \frac{1-p_{(1,0),n} + p_{(-1,0),n}}{1+p_{(1,0),n} - p_{(-1,0),n}}. \tag{7.9}$$

The number of times that the piece moves between t_{2i} and t_{2i+1} is $1+m_i$. We have, by the law of large numbers,

$$\frac{1}{N}\sum_{i=1}^{N} m_i \to \mathbf{E}[m_1] = \sum_{k=0}^{\infty} k(2p_{(0,\pm1),n})(1-2p_{(0,\pm1),n})^k = \frac{1}{2p_{(0,\pm1),n}} - 1. \tag{7.10}$$

Also, $\frac{1}{N}\sum_{i=1}^{N} r_i \to \mathbf{E}[r_1]$. Similarly, the averages converge for n_i and s_i, which have the same distribution. Since, on average, the piece moves once every $\frac{5}{4}(n^2-1)$ steps of the walk, $\frac{\mathbf{E}[r_1]}{1+\mathbf{E}[m_1]} = \frac{5}{4}(n^2-1)$, or

$$\mathbf{E}[r_1] = \frac{5}{4}(n^2-1)\left(\frac{1}{2p_{(0,\pm1),n}}\right). \tag{7.11}$$

□

The primary step in establishing the mixing of the piece $\mathscr{P}$ is establishing the asymptotic independence of the coordinates of the sum

$$S_N = \left(\sum_{i=1}^{N} H_i, \sum_{i=1}^{N} V_i, \sum_{i=1}^{N} (r_i + s_i)\right),$$

which is demonstrated by considering the **characteristic function**:

$$\chi(\xi_1, \xi_2) = \mathbf{E}\left[e^{2\pi i \frac{\xi_1}{n} H_1 + 2\pi i \xi_2 r_1}\right], \quad \xi_1 \in \mathbb{Z}/n\mathbb{Z}, \quad \xi_2 \in \mathbb{R}/\mathbb{Z}.$$

Firstly, we develop several methods to get estimates of hitting times of random walk on $(\mathbb{Z}/n\mathbb{Z})^2$: Let P be the transition kernel of $\frac{1}{5}$-lazy simple random walk on $(\mathbb{Z}/n\mathbb{Z})^2$, and let P' be P with row and column corresponding to $(0,0)$ deleted. Let

$$R(z) = (I - zP')^{-1} = I + zP' + (zP')^2 + \ldots \tag{7.12}$$

be the resolvent. I.e., the coefficients in z^n in each entry is the n-th transition probability.

Lemma 7.2.4. *The characteristic function of the hitting time from $(1,0)$ to $(0,0)$ under $\frac{1}{5}$-lazy simple random walk at $z = e^{2\pi i \xi}$ is*

$$\chi(z) = \frac{z}{5} e^t_{(1,0)} R(z) \left(e_{(1,0)} + e_{(-1,0)} + e_{(0,1)} + e_{(0,-1)}\right). \tag{7.13}$$

The expected hitting time is

$$1 + \frac{1}{5} e^t_{(1,0)} R'(1) \left(e_{(1,0)} + e_{(-1,0)} + e_{(0,1)} + e_{(0,-1)}\right) \tag{7.14}$$

which is of order n^2.

Proof. $(P')^n$ enumerates the transition probabilities that result from length n paths which do not visit $(0,0)$. To obtain the characteristic function formula, condition on the number of steps taken under P', and then use that the probability of transitioning from one of the neighbors of $(0,0)$ to $(0,0)$ is $\frac{1}{5}$.

We have $\chi(1) = 1$, since the walk hits $(0,0)$ in finite time with probability 1.

The formula for the expected hitting time holds since the expectation is $\chi'(1)$. $\square$

Since P' is symmetric, it can be diagonalized using an orthonormal set of eigenvectors. Let the corresponding eigenvalues be $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{n^2-1}$ with eigenvectors $v_1, v_2, ..., v_{n^2-1}$. We have v_1 is non-negative and $1 > \lambda_1$. Write

$$R(z) = \sum_{i=1}^{n^2-1} \frac{v_i v_i^t}{1 - z\lambda_i}. \tag{7.15}$$

Let $c_{i,x} = \langle v_i, e_x \rangle$.

Lemma 7.2.5. *The largest eigenvalue λ_1 satisfies $\frac{1}{n^2 \log n} \ll 1 - \lambda_1 \ll \frac{1}{n^2}$. Also,*

$$\frac{5}{4} \leqslant \sum_{i=1}^{n^2-1} \frac{c_{i,(1,0)}^2}{1 - \lambda_i} < 5 \tag{7.16}$$

and

$$\sum_{i=1}^{n^2-1} \frac{c_{i,(1,0)}^2}{(1 - \lambda_i)^2} \asymp n^2. \tag{7.17}$$

Furthermore, there is a constant $c > 0$ such that

$$\sum_{i:(1-\lambda_i) > \frac{c}{n}} \frac{c_{i,(1,0)}^2}{(1 - \lambda_i)} \gg \frac{1}{\log n}. \tag{7.18}$$

Proof. Let v_1 be in the top eigen-space, $1^t v_1 = 1$ so that v_1 is a probability vector. Since $(P')^m v_1 = \lambda_1^m v_1$, $1^t (P')^m v_1 = \lambda_1^m$ is the probability of not reaching $(0,0)$ in m steps, started from a distribution proportional to v_1. Since the expected hitting time to $(0,0)$ is $O(n^2 \log n)$ uniformly in the starting point, it follows that for some $c > 0$, $\lambda_1 \leqslant 1 - \frac{c}{n^2 \log n}$.

The bound $\lambda_1 > 1 - \frac{c}{n^2}$ will follow after establishing (7.17), since $\sum_i c_{i,(1,0)}^2 = 1$ by orthogonality. To prove (7.16) note that $\chi(1) = 1$ may be written

$$1 = \frac{1}{5} \sum_{i=1}^{n^2-1} \frac{c_{i,(1,0)}(c_{i,(1,0)} + c_{i,(-1,0)} + c_{i,(0,1)} + c_{i,(0,-1)})}{1 - \lambda_i}. \tag{7.19}$$

Furthermore, by symmetry, $\sum_i \frac{c_{i,x}^2}{1-\lambda_i}$ is independent of $x \in \{(\pm 1, 0), (0, \pm 1)\}$, so that

$$\begin{aligned} \sum_{i=1}^{n^2-1} \frac{c_{i,(1,0)}^2}{1 - \lambda_i} &\leqslant \sum_{i=1}^{n^2-1} \frac{c_{i,(1,0)}(c_{i,(1,0)} + c_{i,(-1,0)} + c_{i,(0,1)} + c_{i,(0,-1)})}{1 - \lambda_i} \\ &\leqslant 4 \sum_{i=1}^{n^2-1} \frac{c_{i,(1,0)}^2}{1 - \lambda_i}. \end{aligned} \tag{7.20}$$

Similarly, the expected hitting time formula may be written

$$\sum_{i=1}^{n^2-1} \frac{c_{i,(1,0)}\left(c_{i,(1,0)} + c_{i,(-1,0)} + c_{i,(0,1)} + c_{i,(0,-1)}\right)\lambda_i}{(1-\lambda_i)^2} \asymp n^2. \tag{7.21}$$

Those terms with $\frac{1}{1-\lambda_i}$ bounded contribute $O(1)$ to this sum, so that the negative terms may be dropped, and hence

$$\sum_{i=1}^{n^2-1} \frac{c_{i,(1,0)}\left(c_{i,(1,0)} + c_{i,(-1,0)} + c_{i,(0,1)} + c_{i,(0,-1)}\right)}{(1-\lambda_i)^2} \asymp n^2. \tag{7.22}$$

Again by symmetry, and Cauchy-Schwarz,

$$\sum_{i=1}^{n^2-1} \frac{c_{i,(1,0)}^2}{(1-\lambda_i)^2} \asymp n^2. \tag{7.23}$$

To prove (7.18), note that if $c > 0$ is sufficiently small, since $\sum_i c_{i,(1,0)}^2 = 1$,

$$\sum_{i:(1-\lambda_i)>\frac{c}{n}} \frac{c_{i,(1,0)}^2}{(1-\lambda_i)^2} \gg n^2. \tag{7.24}$$

The claim now follows from $\frac{1}{1-\lambda_i} \ll n^2 \log n$. $\square$

The following bounds are useful in bounding the characteristic function of the hitting time.

Lemma 7.2.6. *Let ϑ be maximal such that*

$$\sum_{i:(1-\lambda_i)>\vartheta} \frac{c_{i,(1,0)}^2}{1-\lambda_i} \geqslant \left(1 - \frac{1}{(\log n)^3}\right) \sum_{i=1}^{n^2-1} \frac{c_{i,(1,0)}^2}{1-\lambda_i}. \tag{7.25}$$

Then for $\vartheta < \xi \leqslant \frac{1}{2}$,

$$\sum_{i:1-\lambda_i<\xi} c_{i,(1,0)}^2 \left(\frac{1}{1-\lambda_i} - \frac{1}{|1-\lambda_i e^{2\pi i\xi}|}\right) \gg \frac{1}{(\log n)^3}. \tag{7.26}$$

For $0 < \xi \leqslant \vartheta$,

$$\sum_{i:(1-\lambda_i)>\vartheta} c_{i,(1,0)}^2 \left(\frac{1}{1-\lambda_i} - \frac{1}{|1-\lambda_i e^{2\pi i\xi}|}\right) \gg \xi^2 n^4. \tag{7.27}$$

Proof. By the previous lemma, $\vartheta \ll \frac{1}{n}$. We have

$$\frac{1}{|1-\lambda_i e^{2\pi i \xi}|} \leqslant \frac{1}{|(1-\lambda_i)+i\lambda_i \sin(2\pi\xi)|} = \frac{1}{1-\lambda_i}\frac{1}{\left|1+\frac{i\lambda_i \sin(2\pi\xi)}{1-\lambda_i}\right|}. \tag{7.28}$$

and, thus,

$$\begin{aligned} \frac{1}{1-\lambda_i} - \frac{1}{|1-\lambda_i e^{2\pi i\xi}|} &\gg \frac{1}{1-\lambda_i}\left(1 - \frac{1}{\sqrt{1+\frac{\xi^2}{(1-\lambda_i)^2}}}\right) \\ &\gg \frac{1}{1-\lambda_i}\min\left(1, \frac{\xi^2}{(1-\lambda_i)^2}\right). \end{aligned} \tag{7.29}$$

We now conclude, when $\vartheta < \xi \leqslant \frac{1}{2}$,

$$\sum_{i:(1-\lambda_i)<\xi} c_{i,(1,0)}^2 \left(\frac{1}{1-\lambda_i} - \frac{1}{|1-\lambda_i e^{2\pi i\xi}|}\right) \gg \sum_{i:(1-\lambda_i)<\xi} \frac{c_{i,(1,0)}^2}{1-\lambda_i} \gg \frac{1}{(\log n)^3}. \tag{7.30}$$

For $0 < \xi \leqslant \vartheta$, note that

$$\sum_{i:(1-\lambda_i)<\vartheta} \frac{c_{i,(1,0)}^2}{(1-\lambda_i)^3} \ll \frac{1}{(1-\lambda_1)^2}\sum_{i:(1-\lambda_i)<\vartheta} \frac{c_{i,(1,0)}^2}{1-\lambda_i} \ll \frac{n^4}{\log n} \tag{7.31}$$

while, by Hölder,

$$\sum_{i=1}^{n^2-1} \frac{c_{i,(1,0)}^2}{(1-\lambda_i)^3} \gg n^4. \tag{7.32}$$

Hence,

$$\sum_{i:1-\lambda_i>\vartheta} c_{i,(1,0)}^2 \left(\frac{1}{1-\lambda_i} - \frac{1}{|1-\lambda_i e^{2\pi i\xi}|}\right) \gg \xi^2 \sum_{i:1-\lambda_i>\vartheta} \frac{c_{i,(1,0)}^2}{(1-\lambda_i)^3} \gg \xi^2 n^4. \tag{7.33}$$

□

Now we are ready to prove bounds of $\chi(\xi_1, \xi_2)$.

Lemma 7.2.7. *There is a constant $c > 0$ such that, uniformly in n and uniformly in $\xi_1 \in \left(-\frac{n}{2}, \frac{n}{2}\right]$, and $\xi_2 \in \left(-\frac{1}{2}, \frac{1}{2}\right]$,*

$$|\chi(\xi_1, \xi_2)| \leqslant 1 - c\max\left(\frac{\xi_1^2}{n^2}, \xi_2^2\right). \tag{7.34}$$

Proof. Consider paths of either fixed, bounded length, or fixed displacement of $\mathscr{P}$ but bounded varying length. All of these have positive probability, which is not dependent on n for all n sufficiently large. The variation in their phase is of the given magnitude. $\square$

We use generating function to get a matrix formula for $\chi(\xi_1, \xi_2)$. Recall that we let P be the transition matrix of $\frac{1}{5}$-lazy simple random walk on $(\mathbb{Z}/n\mathbb{Z})^2$, and that P' is the minor excluding the row and column of $(0,0)$, $R(z) = (I - zP')^{-1}$ Let $M(z_1, z_2)$ be the transition matrix on $(\mathbb{Z}/n\mathbb{Z})^2 \backslash \{(0,0)\}$ with $\frac{1}{5} z_1 z_2$ in the transition from $(1,0)$ to $(-1,0)$ and $\frac{1}{5}\frac{z_2}{z_1}$ in the transition from $(-1,0)$ to $(1,0)$, and zeros elsewhere. Let $w(z_1, z_2)$ be the vector with entries $\frac{z_1 z_2}{2}$ at $(-1,0)$ and $\frac{z_2}{2z_1}$ at $(1,0)$, with zeros elsewhere, and let v be the vector with value $\frac{1}{5}$ at $(0, \pm 1)$ and zeros elsewhere.

Lemma 7.2.8. *Let $\xi_1 \in \mathbb{Z}/n\mathbb{Z}$, $\xi_2 \in \mathbb{R}/\mathbb{Z}$, and set $z_1 = e^{\frac{2\pi i \xi_1}{n}}, z_2 = e^{2\pi i \xi_2}$. Then*

$$\chi(\xi_1, \xi_2) = w(z_1, z_2)^t (I - R(z_2) M(z_1, z_2))^{-1} R(z_2) v. \tag{7.35}$$

Proof. The sequence of moves described in phase space by $\chi(\xi_1, \xi_2)$ are as follows. An initial move, which involves one move of the empty square and one update, right or left of $\mathscr{P}$ occurs. This is recorded by $w(z_1, z_2)^t$ in which if the empty square swaps places from the right, it now has the position to the left $(-1, 0)$ of $\mathscr{P}$ and makes one move, hence contributes $z_1 z_2$ to the phase. If instead the empty square swaps places from the left then it now occupies $(1,0)$ relative to $\mathscr{P}$ and contributes $\frac{z_2}{z_1}$ to the phase.

Now there are 0 or more excursions of the empty square followed by a right or left move of $\mathscr{P}$. A right or left move of $\mathscr{P}$ entails moving the empty square from the position on the right of $\mathscr{P}$ to the position on the left, or vice versa. This is captured by $M(z_1, z_2)$. Finally there is a final excursion, captured by R, which is finished by moving onto $\mathscr{P}$ from above or below, captured by v. $\square$

The previous lemma implies the following bound.

Lemma 7.2.9. *Uniformly in n, $\xi_1 \in (-\frac{n}{2}, \frac{n}{2}]$ and $\xi_2 \in (-\frac{1}{2}, \frac{1}{2}]$,*

$$1 - |\chi(\xi_1, \xi_2)| \gg e_{(1,0)}^t R(1) e_{(1,0)} - \left| e_{(1,0)}^t R(e^{2\pi i \xi_2}) e_{(1,0)} \right|. \tag{7.36}$$

Proof. In the term $(I - R(z_2)M(z_1, z_2))^{-1} = \sum_{k=0}^{\infty} (R(z_2) M(z_1, z_2))^k$, there is a probability, bounded uniformly away from 0 that the $k = 1$ term is taken, and a probability bounded uniformly from 0 that the return is of type $e_{(1,0)}$ to $e_{(1,0)}$. With no cancellation, the sum of path probabilities making up the return is $e_{(1,0)}^t R(1) e_{(1,0)}$, while with the phase, the sum has size $\left| e_{(1,0)}^t R(e^{2\pi i \xi_2}) e_{(1,0)} \right|$. $\square$

Lemma 7.2.10. *There is a constant $c > 0$ such that, for all $\xi_1 \in \mathbb{Z}/n\mathbb{Z}$ and $-\frac{1}{2} < \xi_2 \leqslant \frac{1}{2}$,*

$$|\chi(\xi_1, \xi_2)| \leqslant 1 - c \min\left(\frac{1}{(\log n)^3}, \xi_2^2 n^4\right). \tag{7.37}$$

Proof. By the previous lemma,

$$\begin{aligned} 1 - |\chi(\xi_1, \xi_2)| &\gg \sum_{i=1}^{n^2-1} \frac{c_{i,(1,0)}^2}{1-\lambda_i} - \left| \sum_{i=1}^{n^2-1} \frac{c_{i,(1,0)}^2}{1-\lambda_i e^{2\pi i \xi_2}} \right| \\ &\geqslant \sum_{i=1}^{n^2-1} c_{i,(1,0)}^2 \left(\frac{1}{1-\lambda_i} - \frac{1}{|1-\lambda_i e^{2\pi i \xi_2}|}\right). \end{aligned} \tag{7.38}$$

By Lemma 7.2.6, it follows that

$$1 - |\chi(\xi_1, \xi_2)| \gg \min\left(\frac{1}{(\log n)^3}, \xi_2^2 n^4\right). \tag{7.39}$$

□

At small frequencies, the characteristic function may be estimated by Taylor expansion. Recall $\mathbf{E}[H_1^2] = s_n^2$, $\mathbf{E}[r_1] = \mu_n$ and $\mathbf{Var}[r_1] = v_n^2$.

Lemma 7.2.11. *For $\xi_1 \in \mathbb{Z}/n\mathbb{Z}$, $|\xi_1| \leqslant \frac{n}{2}$ and for complex ξ_2, $|\xi_2| \ll \frac{1}{n^2(\log n)^2}$,*

$$\begin{aligned} \chi(\xi_1, \xi_2) = \exp&\left(2\pi i \xi_2 \mu_n - \frac{2\pi^2 \xi_1^2}{n^2} s_n^2 - 2\pi^2 \xi_2^2 v_n^2\right) \\ &\times \left(1 + O\left(\frac{\xi_1^3}{n^3} + \xi_2^3 (n^2 \log n)^3\right)\right). \end{aligned} \tag{7.40}$$

Proof. To obtain this estimate, write

$$\chi(\xi_1, \xi_2) = e^{2\pi i \xi_2 \mu_n} \mathbf{E}\left[e^{2\pi i \left(\frac{\xi_1}{n} H_1 + \xi_2 (r_1 - \mu_n)\right)}\right]. \tag{7.41}$$

Now Taylor expand $e^{2\pi i x} = 1 + 2\pi i x - 2\pi^2 x^2 + O(x^3)$ and use the moments

$$\mathbf{E}[H_1] = 0, \qquad \mathbf{E}[H_1^2] = s_n, \qquad \mathbf{E}[|H_1|^3] = O(1) \tag{7.42}$$

and

$$\begin{aligned} \mathbf{E}[r_1 - \mu_n] &= 0, \\ \mathbf{E}[(r_1 - \mu_n)^2] &= v_n, \\ \mathbf{E}[H_1(r_1 - \mu_n)] &= 0, \\ \mathbf{E}[|r_1 - \mu_n|^3] &= O((n^2 \log n)^3). \end{aligned} \tag{7.43}$$

□

Combine above we get **a local limit theorem:**

Theorem 7.2.12 ([1] Theorem 29)**.** *Let* $n \geqslant 2, (\log n)^{19} \leqslant N \leqslant n^3, |t-2N\mu_n| < \sqrt{N}v_n \log n$ *for any* $A > 0$,

$$\text{Prob}\left(S_N = (i,j,t)\right) = O\left(\frac{(\log n)^6}{N^2 v_n}\exp\left(-\frac{(t-2N\mu_n)^2}{4Nv_n^2}\right)\right) + O_A\left(n^{-A}\right)$$
$$+\frac{\exp\left(-\frac{(t-2N\mu_n)^2}{4Nv_n^2}\right)}{\sqrt{4\pi N}v_n}\left(\frac{1}{n^2}\sum_{\substack{\xi\in\mathbb{Z}/n\mathbb{Z})^2\\ \|\xi\|_2\ll\frac{n}{\sqrt{N}}\log n}} e^{-2\pi i\frac{\xi\cdot(i,j)}{n}}\exp\left(-\frac{2\pi^2\|\xi\|_2^2 s_n^2 N}{n^2}\right)\right).$$

Proof. The characteristic function of S_N at frequencies $\frac{\xi_1}{n}, \frac{\xi_2}{n}, \xi_3$ is given by $\chi(\xi_1,\xi_3)^N\chi(\xi_2,\xi_3)^N$. Hence, by Fourier inversion,

$$\begin{aligned}&\mathbf{Prob}(S_N = (i,j,t)) \\ &= \frac{1}{n^2}\sum_{\xi=(\xi_1,\xi_2)\in(\mathbb{Z}/n\mathbb{Z})^2}\int_{\xi_3\in\mathbb{R}/\mathbb{Z}} e^{-2\pi i\left(\frac{\xi_1 i+\xi_2 j}{n}+\xi_3 t\right)}\chi(\xi_1,\xi_3)^N\chi(\xi_2,\xi_3)^N d\xi_3.\end{aligned} \tag{7.44}$$

Using the bound $|\chi(\xi_1,\xi_3)| \leqslant 1 - c\max\left(\frac{\xi_1^2}{n^2}, \xi_3^2\right)$, truncate the torus variables to $\|\xi\|_2 \ll \frac{n}{\sqrt{N}}\log n$ with error, for any $A > 0$, $O_A(n^{-A})$. Next using the bound $|\chi(\xi_1,\xi_3)| \leqslant 1 - c\min\left(\frac{1}{(\log n)^3}, \xi_3^2 n^4\right)$, truncate the ξ_3 integral to $|\xi_3| \ll \frac{(\log n)^2}{n^2\sqrt{N}}$ with the same error. Inserting the Taylor expansion for $\chi(\xi_1,\xi_3)$ and $\chi(\xi_2,\xi_3)$ at low frequencies,

$$\begin{aligned}\mathbf{Prob}(S_N = (i,j,t)) &= \frac{1}{n^2}\sum_{\|\xi\|_2\ll\frac{n}{\sqrt{N}}\log n} e^{-2\pi i\left(\frac{\xi_1 i+\xi_2 j}{n}\right)} \\ &\times \int_{|\xi_3|\ll\frac{(\log n)^2}{n^2\sqrt{N}}} e^{-2\pi i\xi_3(t-2N\mu_n)}\exp\left(-N\left(\frac{2\pi^2(\xi_1^2+\xi_2^2)s_n^2}{n^2}+4\pi^2\xi_3^2 v_n^2\right)\right) \\ &\times\left(1+O\left(\frac{\|\xi\|_2^3}{n^3}+|\xi_3|^3(n^2\log n)^3\right)\right)^{2N} d\xi_3 + O_A(n^{-A}).\end{aligned} \tag{7.45}$$

In the integral over ξ_3 substitute $\xi_3' = 2\pi N^{\frac{1}{2}}v_n\xi_3$ to obtain

$$\begin{aligned}\mathbf{Prob}\left(S_N = (i,j,t)\right) &= \frac{1}{2\pi N^{\frac{1}{2}}v_n n^2}\sum_{\|\xi\|_2\ll\frac{n}{\sqrt{N}}\log n} e^{-2\pi i\left(\frac{\xi_1 i+\xi_2 j}{n}\right)} \\ &\times\int_{|\xi_3'|\ll\frac{v_n(\log n)^2}{n^2}}\exp\left(\frac{-i\xi_3'(t-2N\mu_n)}{N^{\frac{1}{2}}v_n} - {\xi_3'}^2 - N\left(\frac{2\pi^2(\xi_1^2+\xi_2^2)s_n^2}{n^2}\right)\right) \\ &\times\left(1+O\left(\frac{\|\xi\|_2^3}{n^3}+|\xi_3'|^3\left(\frac{n^2\log n}{v_n\sqrt{N}}\right)^3\right)\right)^{2N} d\xi_3' + O_A(n^{-A}).\end{aligned} \tag{7.46}$$

$$\tilde{\xi}_3 = \xi_3' + \frac{i(t-2N\mu_n)}{2N^{\frac{1}{2}} v_n} \tag{7.47}$$

then shifting the $\tilde{\xi}_3$ integral to be on the real axis obtains a horizontal integral bounded by $O_A(n^{-A})$ together with a shifted integral

$$\begin{aligned}\mathbf{Prob}\left(S_N = (i,j,t)\right) &= \frac{\exp\left(-\frac{(t-2N\mu_n)^2}{4Nv_n^2}\right)}{2\pi N^{\frac{1}{2}} v_n n^2} \sum_{\|\xi\|_2 \ll \frac{n}{\sqrt{N}}\log n} e^{-2\pi i\left(\frac{\xi_1 i + \xi_2 j}{n}\right)} \\ &\times \int_{|\tilde{\xi}_3| \ll \frac{v_n(\log n)^2}{n^2}} \exp\left(-\tilde{\xi}_3^2 - N\left(\frac{2\pi^2(\xi_1^2+\xi_2^2)s_n^2}{n^2}\right)\right) \\ &\times \left(1 + O\left(\frac{\|\xi\|_2^3}{n^3} + \left(|\tilde{\xi}_3|^3 + \frac{|t-2N\mu_n|^3}{N^{\frac{3}{2}} v_n^3}\right)\left(\frac{n^2\log n}{v_n\sqrt{N}}\right)^3\right)\right)^{2N} d\tilde{\xi}_3 \\ &+ O_A(n^{-A}).\end{aligned} \tag{7.48}$$

The main term is obtained by dropping the big O terms and extending the $\tilde{\xi}_3$ integral to $\mathbb{R}$ with acceptable error.

To bound the error, note that in the region of integration, using $n^2 \ll v_n \ll n^2 \log n$,

$$O\left(\frac{\|\xi\|_2^3}{n^3} + \left(|\tilde{\xi}_3|^3 + \frac{|t-2N\mu_n|^3}{N^{\frac{3}{2}} v_n^3}\right)\left(\frac{n^2\log n}{v_n\sqrt{N}}\right)^3\right) = o\left(\frac{1}{N}\right), \tag{7.49}$$

so that the exponential may be bounded linearly. Bound integration over $\tilde{\xi}_3$ by a constant. This obtains an error of

$$\begin{aligned}&\ll \frac{\exp\left(-\frac{(t-2N\mu_n)^2}{4Nv_n^2}\right)}{2\pi N^{\frac{1}{2}} v_n n^2} \sum_{\|\xi\|_2 \ll \frac{n}{\sqrt{N}}\log n} \exp\left(\frac{-2\pi^2 N\|\xi\|_2^2 s_n^2}{n^2}\right) \\ &\times N\left(\frac{\|\xi\|_2^3}{n^3} + \left(1 + \frac{|t-2N\mu_n|^3}{N^{\frac{3}{2}} v_n^3}\right)\left(\frac{n^2\log n}{v_n\sqrt{N}}\right)^3\right).\end{aligned} \tag{7.50}$$

Use $v_n \gg n^2$, and use $|t-2N\mu_n| \ll N^{\frac{1}{2}} v_n \log n$, and approximate the sum over ξ with an integral over $\mathbb{R}^2$ to estimate the error by

$$\ll \frac{\exp\left(-\frac{(t-2N\mu_n)^2}{4Nv_n^2}\right) N^{\frac{1}{2}}}{2\pi v_n} \int_{\mathbb{R}^2} \exp\left(-2\pi^2 N s_n^2 x^2\right)\left(\|x\|_2^3 + \frac{(\log n)^6}{N^{\frac{3}{2}}}\right) dx. \tag{7.51}$$

This obtains the claimed error bound.

□

Due to symmetry of the walk, the distribution of the marked piece $\mathscr{P}$ is determined after N steps of the renewal process by the above local limit theorem.

We can now prove Theorem 7.2.1.

Proof of Theorem 7.2.1. Recall that the time t Brownian motion $B(t)$ on $(\mathbb{R}/\mathbb{Z})^2$ has a distribution which is a θ function $\theta_t(x)$. The convergence in Theorem 7.2.1 consists of a lower bound and an upper bound approximating the distance to uniformity of the single piece with

$$\begin{aligned} \left\|\theta_t - \mathbb{U}_{(\mathbb{R}/\mathbb{Z})^2}\right\|_{\mathrm{TV}} &= \frac{1}{2}\int_{(\mathbb{R}/\mathbb{Z})^2} |\theta_t(x) - 1| dx \\ &= \int_{(\mathbb{R}/\mathbb{Z})^2} \mathbf{1}(\theta_t(x) > 1)(\theta_t(x) - 1)dx. \end{aligned} \tag{7.52}$$

Let $\rho : \mathbb{R} \to [0,1]$ be a smooth approximation identity (cut-off function) which satisfies $\rho(x) = 0$ if $x \leqslant -1$ and $\rho(x) = 1$ if $x \geqslant 1$. Let, for $\varepsilon > 0$, $\rho_\varepsilon(x) = \rho\left(\frac{x}{\varepsilon}\right)$ and define $\psi_\varepsilon(x,t) = \rho_\varepsilon(\theta_t(x) - 1)$. For fixed ε and for $t \in K$ with $K \subset \mathbb{R}^+$ compact, $\psi_\varepsilon(x,t)$ is uniformly C^1 since $\theta_t(x)$ is uniformly C^j for every j. Also,

$$\left| \left\|\theta_t(x) - \mathbb{U}_{(\mathbb{R}/\mathbb{Z})^2}\right\|_{\mathrm{TV}} - \int_{(\mathbb{R}/\mathbb{Z})^2} \psi_\varepsilon(x,t)(\theta_t(x) - 1)dx \right| \leqslant \varepsilon, \tag{7.53}$$

since $|\theta_t(x) - 1| \leqslant \varepsilon$ whereever $\psi_\varepsilon(x,t)$ and $\mathbf{1}(\theta(x,t) > 1)$ differ.

Define $c_{\mathrm{puz}} = \frac{2\mu}{\sigma^2} = \frac{5}{2}(\pi - 1)$. Let (i_T, j_T) be the displacement from its initial position of the piece $\mathscr{P}$ after $T = \lfloor c_{\mathrm{puz}} n^4 t \rfloor$ steps of the Markov chain P. We show that for each fixed $\varepsilon > 0$, uniformly for $t \in K$,

$$\lim_{n\to\infty} \mathbf{E}_{\mathbb{U}_{(\mathbb{Z}/n\mathbb{Z})^2}} \left[\psi_\varepsilon\left(\left(\frac{i}{n}, \frac{j}{n}\right), t\right)\right] = \int_{(\mathbb{R}/\mathbb{Z})^2} \psi_\varepsilon(x,t)dx \tag{7.54}$$

and

$$\lim_{n\to\infty} \mathbf{E}_{P^T} \left[\psi_\varepsilon\left(\left(\frac{i_T}{n}, \frac{j_T}{n}\right), t\right)\right] = \int_{(\mathbb{R}/\mathbb{Z})^2} \psi_\varepsilon(x,t)\theta_t(x)\, dx, \tag{7.55}$$

which proves that the total variation distance of the single piece process is bounded below in the limit by that of Brownian motion. Note that (7.54) holds since the expectation is a Riemann sum for the integral, so that the convergence holds by uniform convergence.

The distribution of $\mathscr{P}$ is determined after N steps of the renewal process in Theorem 7.2.12, so we now remove the stopping time implicit in the renewal process and include the moves H_0, V_0 prior to the renewal process beginning.

Let, for $\epsilon > 0$,

$$N = \left\lfloor \frac{T}{2\mu_n} - n^{1+\epsilon} \right\rfloor. \tag{7.56}$$

Since μ_n is of order n^2 by Lemma 7.2.3, N is of order n^2. Let

$$M = \left\lfloor \frac{T - S_{N,3}}{2\mu_n} - n^{\frac{1}{2}+2\epsilon} \right\rfloor. \tag{7.57}$$

Then outside a set S_{bad}

$$(i_T, j_T) = (H_0, V_0) + (S_{N,1}, S_{N,2}) + (S_{M,1}, S_{M,2}) + (E_1, E_2) \tag{7.58}$$

where E_1, E_2 are the set of moves of the piece after time $t_1 + S_{N,3} + S_{M,3}$. The condition for membership in S_{bad} is that either $|S_{N,3} - 2N\mu_n| \geqslant \sqrt{N} v_n \log n$ or $|S_{M,3} - 2M\mu_n| \geqslant \sqrt{M} v_n \log n$. Since $\frac{r_i}{v_n}$ and $\frac{s_i}{v_n}$ have exponentially decaying tails, by the variant of Chernoff's inequality, Lemma 8.0.2, S_{bad} has probability, for any $A > 0$, $O_A(n^{-A})$, so can be ignored.

Outside S_{bad}, M is of order $n^{1+\epsilon}$, so that $S_{M,1}, S_{M,2} \ll n^{\frac{1}{2}+\epsilon}$ with probability $1 - O_A(n^{-A})$ by Lemma 8.0.2 this time applied to H_i and V_i, which have exponentially decaying tails. Similarly, by excluding S_{bad}, $T - S_{N,3} - S_{M,3} - t_1 = O(n^{\frac{5}{2}+2\epsilon})$. It then follows by Chernoff's inequality for the sum of r_i and s_i, that $\mathscr{P}$ moves $O(n^{\frac{1}{2}+3\epsilon})$ times in (E_1, E_2). Since H_0 and V_0 are bounded $\ll \log n$ w.o.p., it follows that $(i_T, j_T) = (S_{N,1}, S_{N,2}) + O(n^{\frac{1}{2}+3\epsilon})$. Since $\psi_\varepsilon(x, t)$ is uniformly C^1, it suffices to prove (7.55) for (i_T, j_T) replaced by $(S_{N,1}, S_{N,2})$.

For any fixed i, j, the error term in applying Theorem 7.2.12 to S_N summed in t is $O(n^{-3+\epsilon})$, and hence may be ignored. Also, the sum over ξ may be extended to all of $\mathbb{Z}^2$ with negligible error. This gives a main term of

$$\frac{1}{n^2} \sum_{\xi \in \mathbb{Z}^2} e^{-2\pi i \frac{\xi \cdot (i,j)}{n}} \exp\left(-\frac{2\pi^2 \|\xi\|_2^2 s_n^2 N}{n^2} \right) = \frac{1}{n^2} \theta_{\frac{s_n^2 N}{n^2}} \left(\frac{i}{n}, \frac{j}{n} \right). \tag{7.59}$$

It follows that uniformly in t, as $n \to \infty$,

$$\mathbf{E}_{P^T} \left[\psi_\varepsilon \left(\left(\frac{i_T}{n}, \frac{j_T}{n} \right), t \right) \right] \sim \frac{1}{n^2} \sum_{(i,j) \in (\mathbb{Z}/n\mathbb{Z})^2} \psi_\varepsilon \left(\frac{i}{n}, \frac{j}{n} \right) \theta_{\frac{s_n^2 N}{n^2}} \left(\frac{i}{n}, \frac{j}{n} \right). \tag{7.60}$$

Since $\theta_{\frac{s_n^2 N}{n^2}} \to \theta_t$ uniformly as $n \to \infty$, the claim now follows by uniform convergence. This completes the proof of the lower bound.

To prove the upper bound, note that, conditioned on $S_{N,3}$, $S_{N,1}$ and $S_{N,2}$ are independent of (H_0, V_0), $(S_{M,1}, S_{M,2})$ and (E_1, E_2). As in the lower bound, drop the error terms from the limit in Theorem 7.2.12, since these contribute measure $o(1)$. Denote by $\tilde{S}_N$ the main term. Also, we may restrict attention to $\tilde{S}_{N,3}$ such that $|\tilde{S}_{N,3} - 2N\mu_n| \leqslant Av_n$ for a constant A, since the remaining part has measure $o(1)$ as $A \to \infty$. Since convolution with the remaining distributions can only decrease the total variation distance, as can removing the conditioning, it suffices to prove that, conditioned on any $\tilde{S}_{N,3}$ which differs from its mean by a bounded multiple of its variance, the distribution of $(\tilde{S}_{N,1}, \tilde{S}_{N,2})$ has distance from uniform bounded by the total variation distance of $\theta_t(x)$ to uniform. This in fact follows from the convergence of the Fourier series in Theorem 7.2.12.

□

7.3 The upper bound

We show an upper bound of the mixing time of this random walk, by using the comparison techniques on the walk generated by 3-cycle, which is analysed in [10].

We firstly introduce the lemma from [10].

Lemma 7.3.1. *The spectral gap in the regular representation of* $\mathrm{Alt}(n)$ *for the measure supported uniformly on 3-cycles is* $\frac{3}{n-1}$, *and the* $\frac{\epsilon}{|G|} - \ell^\infty$ *mixing time is of order* $n \log n$.

Proof. See [10], Appendix A. □

The proof of the upper bound is motivated by the observation that A_n is generated by elementary 3-cycle $(i, i+1, i+2)$. We can move the empty piece to a desired position by a path (which is $O(n)$)on the board, then do a $ULDR$ and return back to the right-down corner by the same path. This generates any 3-cycle. and thus gives the constant of the comparison theorem.

Let $G_n = S_{n^2-1} \times (\mathbb{Z}/n\mathbb{Z})^2 (n$ odd) or $G_n = A_{n^2-1} \times (\mathbb{Z}/n\mathbb{Z})^2$ even) be the $n^2 - 1$ group. Consider the symmetric set

$$S = \{Rc, Lc, Uc, Dc, c : c = (c_3, \mathrm{id}), c_3 \text{ a } 3 \text{ -cycle } \}$$

and let μ_S be its uniform probability measure.

Lemma 7.3.2. *The measure* μ_S *has* d_2 *mixing time on* G_n *of order* $O\left(n^2 \log n\right)$.

Proof. Since this is a symmetric random walk on a group, by Plancherel,

$$\|\mu_S^{*N} - \mathbb{U}_{G_n}\|_{d_2}^2 = \sum_{1\neq\lambda\in\sigma(P_S)} \lambda^{2N}, \tag{7.61}$$

where P_S is the transition kernel of the random walk.

Let $\rho = \rho_1 \otimes \rho_2$ be an irreducible representation of G_n, so that ρ_1 is an irreducible representation of S_n or A_n and ρ_2 is a character of $(\mathbb{Z}/n\mathbb{Z})^2$. Thus $\dim \rho_1 \otimes \rho_2 = \dim \rho_1$. We have

$$\begin{aligned}\hat{\mu}_S(\rho_1 \otimes \rho_2) &= \mathbf{E}_{x\in\{R,L,U,D,\mathrm{id}\}} [\rho_1 \otimes \rho_2(x)] \, \mathbf{E}_{c_3 \text{ 3 cycle}} [\rho_1(c_3)] \\ &= \frac{\chi_{\rho_1}(c)}{d_{\rho_1}} \mathbf{E}_{x\in R,L,U,D,\mathrm{id}}[\rho_1 \otimes \rho_2(x)].\end{aligned} \tag{7.62}$$

where $\chi_{\rho_1}(c)$ is the character of ρ_1 at a 3-cycle, and d_{ρ_1} is the dimension.

When $\rho_1 \otimes \rho_2$ is one dimensional, but not trivial, ρ_1 is the identity on 3-cycles, so that in this case, $|\hat{\mu}_S(\rho_1 \otimes \rho_2)| \leqslant 1 - \frac{c}{n^2}$. When ρ_2 is not the trivial representation, this follows by bounding the spectrum of simple random walk on the torus, while when ρ_2 is trivial, since ρ_1 is not, use that $\rho_1(\mathrm{id}) = 1$ while $\rho_1(R) = -1$. Since there are $O(n^2)$ one dimensional representations, this part of the spectrum is mixed in $O(n^2 \log n)$ steps.

When ρ_1 has dimension $d_{\rho_1} > 1$, $\frac{\chi_{\rho_1}(c)}{d_{\rho_1}}$ is the eigenvalue, with multiplicity d_{ρ_1} of the 3-cycle walk in A_{n^2-1} in this representation. There are now n^2 representations having the same ρ_1 factor corresponding to the choices for ρ_2, each having their spectrum bounded in size by $\left|\frac{\chi_{\rho_1}(c)}{d_{\rho_1}}\right|$. Since the spectral gap of the 3-cycle walk is of order $\frac{1}{n^2}$, an arbitrary factor of n^2 in the multiplicity can be saved by increasing the constant in the mixing time of order $n^2 \log n$ for the 3-cycle walk.

□

Theorem 7.3.3 ([1], Theorem 4)**.** ***The mixing time upper bound:*** *The total variation and* $\left(\frac{\epsilon}{|G|}, \ell^\infty\right)$ *mixing time of an* $n^2 - 1$ *puzzle is* $O\left(n^4 \log n\right)$.

Proof. By Cauchy-Schwarz, the total variation distance is bounded by half the d_2 distance. Also, since the $n^2 - 1$ puzzle is a symmetric random walk on a group, the $\frac{\epsilon}{|G|} - \ell^\infty$ mixing time is bounded by a constant times the d_2 mixing time. Thus we only estimate the d_2 mixing time.

By Plancherel, we have

$$\|e_{\mathrm{id}}^t P_{n^2-1}^N - \mathbb{U}_{G_n}\|_{d_2}^2 = \sum_{1\neq\lambda\in\sigma(P_{n^2-1})} \lambda^{2N}. \tag{7.63}$$

Let $1 = \lambda_{0,n^2-1} > \lambda_{1,n^2-1} \geqslant \cdots$ be the eigenvalues of P_{n^2-1}, and let $1 = \lambda_{0,S} > \lambda_{1,S} \geqslant \cdots$ be the eigenvalues of the transition kernel associated to the symmetric generating set μ_S above.

Note that the commutator $URDL$ is a 3-cycle which leaves the empty space fixed. Any other 3-cycle may be obtained by finding a word w of length $O(n)$ which shifts any 3 pieces i, j, k into the positions cycled by $URDL$ and performing $w^{-1}URDLw$, which again leaves the empty square fixed, and cycles i, j, k. It follows that each element of S can be obtained as a word in $O(n)$ letters on generators, so A in the comparison theorem may be taken $O(n^2)$. In particular, $1 - \lambda_{i,n^2-1} \gg \frac{1}{n^2}(1 - \lambda_{i,S})$.

As P_{n^2-1} is $\frac{1}{5}$-lazy, by 1.30 the negative eigenvalues are bounded below by $-\frac{3}{5}$, and since the purported mixing time $O(n^4 \log n)$ is large compared to $\log |G_n| = O(n^2 \log n)$, the negative eigenvalues may be ignored when bounding the d_2 mixing time. Thus by comparison, the d_2 mixing time is bounded by a constant times A times the d_2 mixing time for S, and hence is $O(n^4 \log n)$, as wanted. □

Chapter 8

Concentration Inequality

The following inequalities are used to prove mixing of a single piece .

Lemma 8.0.1 (Chernoff's inequality)**.** *Let $X_1, X_2, ..., X_n$ be i.i.d. random variables satisfying $|X_i - \mathbf{E}[X_i]| \leqslant 1$ for all i. Set $X := X_1 + \cdots + X_n$ and let $\sigma := \sqrt{\mathbf{Var}(X)}$. For any $\lambda > 0$,*

$$\mathbf{Prob}\left(X - \mathbf{E}[X] \geqslant \lambda\sigma\right) \leqslant \max\left(e^{-\frac{\lambda^2}{4}}, e^{\frac{-\lambda\sigma}{2}}\right). \tag{8.1}$$

The following variant handles random variables which have exponentially decaying tails.

Lemma 8.0.2. *Let $X_1, X_2, ..., X_n$ be i.i.d. non-negative random variables of variance σ^2, $\sigma > 0$, satisfying the tail bound, for some $c > 0$ and for all $Z > 0$, $\mathbf{Prob}(X_1 > Z) \ll e^{-cZ}$. Let $X = X_1 + X_2 + \cdots + X_n$. Then for any $\lambda > 1$, for $c_1 = \frac{\sqrt{c\sigma}}{2}$,*

$$\mathbf{Prob}\left(|X - \mathbf{E}[X]| \geqslant \lambda\sigma\sqrt{n}\right) \ll e^{-\frac{\lambda^2}{16}} + ne^{-c_1\lambda^{\frac{1}{2}}n^{\frac{1}{4}}}. \tag{8.2}$$

Proof. Let Z be a parameter, $Z \gg n^{\frac{1}{4}}$. Let X_i' be X_i conditioned on $X_i \leqslant Z$. Let $\mu' = \mathbf{E}[X_i']$. Let $X_i'' = X_i \cdot \mathbf{1}(X_i \leqslant Z) + \mu' \cdot \mathbf{1}(X_i > Z)$ and $X'' = X_1'' + X_2'' + \cdots + X_n''$. We have

$$\begin{aligned}
\mathbf{E}[X_i \cdot \mathbf{1}(X_i \geqslant Z)] &= -\int_Z^\infty x d\mathbf{Prob}(X_i \geqslant x) \qquad (8.3)\\
&= Z\mathbf{Prob}(X_i \geqslant Z) + \int_Z^\infty \mathbf{Prob}(X_i \geqslant x)dx \\
&\ll Ze^{-cZ} + \int_Z^\infty e^{-cx}dx \leqslant \left(Z + \frac{1}{c}\right)e^{-cZ}.
\end{aligned}$$

Thus, for some $c' > 0$, $\mathbf{E}[X''] = \mathbf{E}[X] + O(ne^{-c'Z})$. Also,

$$\begin{aligned}\mathbf{Var}(X_i) &= \mathbf{E}[(X_i - \mathbf{E}[X_i])^2] \\ &\geqslant \mathbf{E}[(X_i - \mathbf{E}[X_i])^2 \mathbf{1}(X_i \leqslant Z)] \\ &\geqslant \mathbf{E}[(X_i - \mu')^2 \mathbf{1}(X_i \leqslant Z)] \\ &= \mathbf{Var}(X_i'').\end{aligned} \tag{8.4}$$

Since $|X_i''| \leqslant Z$, for all n sufficiently large, applying Chernoff's inequality,

$$\begin{aligned}\mathbf{Prob}(|X - \mathbf{E}[X]| > \lambda\sigma\sqrt{n}) &\leqslant \sum_{i=1}^{n} \mathbf{Prob}(X_i'' \neq X_i) \\ &\quad + \mathbf{Prob}\left(|X'' - \mathbf{E}[X'']| > \frac{\lambda}{2}\sigma\sqrt{n}\right) \\ &\ll ne^{-cZ} + 2\max\left(e^{-\frac{\lambda^2}{16}}, e^{-\frac{\lambda\sigma\sqrt{n}}{4Z}}\right).\end{aligned} \tag{8.5}$$

To optimize the exponents, choose $Z^2 = \frac{\lambda\sigma\sqrt{n}}{4c}$ to obtain the claim. □

Bibliography

[1] Chu, Yang, and Hough, Robert *Solution of the 15 puzzle problem.* arXiv:1908.07106., 2019.

[2] Levin, D.A., Y. Peres, and E.L. Wilmer. *Markov chains and mixing times.* American Mathematical Soc., 2009.

[3] Diaconis, Persi. *Group representations in probability and statistics.* Lecture notes-monograph series 11 (1988): i-192.

[4] Diaconis, Persi, and Laurent Saloff-Coste. "Comparison techniques for random walk on finite groups." *The Annals of Probability* (1993): 2131-2156.

[5] Diaconis, Persi, and Laurent Saloff-Coste. "Comparison theorems for reversible Markov chains." *The Annals of Applied Probability* (1993): 696-730.

[6] Durrett, Rick. *Probability: theory and examples.* Vol. 49. Cambridge University Press, 2019.

[7] Hough, Robert D., Daniel C. Jerison, and Lionel Levine. "Sandpiles on the square lattice." *Communications in Mathematical Physics* 367.1 (2019): 33-87.

[8] Diaconis, Persi, and Mehrdad Shahshahani. "Generating a random permutation with random transpositions." *Probability Theory and Related Fields* 57.2 (1981): 159-179.

[9] Johnson, Wm Woolsey, and William Edward Story. "Notes on the '15' puzzle." *American Journal of Mathematics* 2.4 (1879): 397-404.

[10] Helfgott, Harald A., Ákos Seress, and Andrzej Zuk. "Random generators of the symmetric group: diameter, mixing time and spectral gap." *Journal of Algebra* 421 (2015): 349-368.

[11] David Bruce Wilson "Mixing Times of Lozenge Tiling and Card Shuffling Markov Chains" *The Annals of Applied Probability* , 14.1(2004), 274–325

[12] L. G. Valiant. "The complexity of computing the permanent." *Theoretical Computer Science*, 8(2): 189-201 (1979)