Bounding Mixing Times of Lattice Path Markov Chains and Related Models: Utilizing Contraction Properties, Coupling, and Wilson's Method

MATH GR 6153 Probability II Final Essay

Heyuan Yao [1*]

1 School of General Studies, Columbia University, New York, NY, United States

* hy2704@columbia.edu

Abstract and Summary

This report essay focuses on the paper "Mixing times of lozenge tiling and card shuffling Markov chains" by Wilson (2004) [1]. The paper presents methods for bounding the mixing time of some Markov chains that are connected (i.e. irreducible), aperiodic, reversible, and have a stationary uniform distribution. The two main ideas in the paper are (1) using the contraction property and coupling methods to establish a lower bound on the absolute spectral gap and then constructing an upper bound, and (2) using distinguished statistics and Wilson's method to establish a lower bound on the mixing time. The analysis employs a subtle approach of assigning a function to each state, to measure the "displacement" of a state and the "gap" between two partially ordered states. The goal of this essay is to provide a clear introduction to the some important Markov chains discussed in Wilson's article. It will also carefully explain the connections between these chains and the subtle techniques used to bound the mixing time. While the computational results will be presented, they will not be shown in detail.

This article is organized as follows. Section 1 provides the preliminaries, which were mainly covered in the course MATH GR 6153 Probability II, Spring 2023, and in the book "Markov chains and mixing times" by Levin and Peres (2017) [2]. In section 2, we discuss the lattice path Markov chain and the main methods employed in the paper. Relatively detailed proofs will be presented in this section as its results will be utilized in the subsequent chapters. We then present similar arguments for constructing upper and lower bounds for the random adjacent transportation (section 3), lozenge-tiling Markov chain (section 4), and Karzanov-Khachiyan Markov chain (section 5). Finally, section 6 provides conclusions and discusses the challenges in future research.

1 Preliminaries

In this section, we will introduce several definitions and theorems, assuming that the readers are already familiar with them. While we will not provide formal proofs, we will provide explanations to help readers understand or recall these concepts, as we will use these theories extensively throughout the remainder of

the text.

1.1 Variation distance and mixing times

An irreducible and aperiodic Markov chain is known to converge to its unique stationary distribution, regardless of its initial measure. However, describing and quantifying this convergence is important, and we use the (total) variation distance to do so.

Definition 1. *Given the finite state space χ, the total variation between two probability measures μ and ν on χ is defined by*

$$\|\mu - \nu\|_{TV} = max_{A \subseteq \chi} |\mu(A) - \nu(A)| = \frac{1}{2}\|\mu - \nu\|_1. \tag{1}$$

The first equation provides a probabilistic interpretation, namely, the largest difference in probabilities of an event occurring. The second equation offers an equivalent idea using L^1-norm. Using this metric, we can explain the convergence theorem as follows: we denote χ as the finite state space, P_x^t as the probability measure from state $x \in \chi$ after t transitions, and μ as the unique stationary measure.

Theorem 1. *With the above setting, there exists constants $\alpha \in (0, 1)$ and $C > 0$ such that*

$$max_{x \in \chi} \|P_x^t - \mu\|_{TV} \leq C\alpha^t. \tag{2}$$

By utilizing the theorem and the non-increasing properties of d(t) and $\bar{d}(t)$ [1], it becomes evident that the variation distance approaches zero, where

$$d(t) = max_{x \in \chi} \|P_x^t - \mu\|_{TV}, \tag{3}$$

and

$$\bar{d}(t) = max_{x,y \in \chi} \|P_x^t - P_y^t\|_{TV}. \tag{4}$$

Here one can notice that $d(t) \leq \bar{d}(t)$. An intriguing question that arises from this theorem is the speed of convergence, i.e., the amount of time required to attain equilibrium, which indicates that the chain is mixing. It is worth noting that the mixing time quantifies the duration for which the chain needs to be sufficiently close to the stationary distribution.

Definition 2. *The mixing time is defined by*

$$t_{mix}(\epsilon) = min\{t : d(t) \leq \epsilon\} \tag{5}$$

and

$$t_{mix} = min\{t : d(t) \leq \frac{1}{4}\}. \tag{6}$$

[1] the proof can be found in Exercise 4.2, [2]

1.2 Coupling

A widely employed technique to bound the mixing time of a given Markov chain is known as coupling, which is defined as follows.

Definition 3. *A coupling of two probability distribution μ and ν is a pair of random variables (X,Y) defined on a single probability space such that the marginal distribution of X and Y is respectively μ and ν.*

It can be demonstrated with ease that, given μ and ν are two probability measure on χ, the total variation distance can be defined equivalently as

$$\|\mu - \nu\|_{TV} = inf\{\mathbb{P}(X \neq Y) : (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\}. \tag{7}$$

Coupling can be employed to a Markov chain, which is a process $\{(X_t, Y_t)\}_{t\geq 0}$ on $\chi \times \chi$. Marginally both $\{(X_t)\}_{t\geq 0}$ and $\{(Y_t)\}_{t\geq 0}$ are Markov chains with the same transition rule, starting at x_0 and y_0. Usually one utilize Markovian coupling.

Definition 4. *Given the above setting and denote P the transition matrix, a Markovian coupling of two P-chains satisfies, for all x, y, x', y' $\in$ χ,*

$$\mathbb{P}\{X_{t+1} = x' | X_t = x,\ Y_t = y\} = P(x, x'),\ \text{and } \mathbb{P}\{Y_{t+1} = y' | X_t = x,\ Y_t = y\} = P(y, y'). \tag{8}$$

It's worth noting that once the two chains coalesce, they will move together. This leads to the following results.

Theorem 2. *Let $\{(X_t, Y_t)\}_{t\geq 0}$ be a Markovian coupling for which $X_0 = x$ and $Y_0 = y$. Let τ_{couple} be the coalescence time of the chains:*

$$\tau_{couple} = min\{t :\ X_s = Y_s \text{ for all } s\}. \tag{9}$$

Then

$$\|P_y^t - P_x^t\|_{TV} \leq \mathbb{P}_{x,y}\{\tau_{couple} > t\}. \tag{10}$$

Therefore a corollary can be derived.

Corollary 1. *Suppose that for each pair of states $x, y \in \chi$, $\exists$ a coupling (X_t, Y_t) with initial state X_0, Y_0 = (x,y). For each coupling,*

$$d(t) \leq max_{x,y\in\chi}\mathbb{P}_{x,y}\{\tau_{couple} > t\}. \tag{11}$$

With this corollary, it is possible to specify a lower bound for the mixing time. Specifically, one can set the right-hand side to be less than a given ϵ to obtain the upper bound. Moreover, if a quickly coalescent coupling is used, the upper bound will be even smaller.

1.3 Absolute Spectral Gap

We note that Wilson's paper focuses primarily on defining a function on the state space naturally, instead of highlight the method of specifying a lower bound using eigenfunctions and eigenvalues. However, including these terms can provide valuable intuition for studying mixing time. Nonetheless, readers will not encounter any difficulties understanding the main text of this essay, even if they choose to skip this subsection.

In the context of a transition matrix P, the function $f : \chi \to C$ is an eigenfunction with a corresponding eigenvalue λ if $Pf = \lambda f$. We claim that in this essay, the eigenfunction and eigenvalue are both real, considering that each Markov chain discussed is irreducible, aperiodic and reversible[2]. Furthermore, we can observe that $-1 < \lambda \leq 1$, where 1 is always an eigenvalue with a corresponding constant eigenfunction.[3]

This property provides a powerful method to determine a lower bound using the absolute spectral gap.

Definition 5. *The absolute spectral gap $\gamma_*$is defined as the difference between 1 and the second-largest absolute value of the eigenvalue, i.e., $\gamma_* = 1 - \lambda_*$, where $\lambda_* := max\{|\lambda| : \lambda \text{ is an eigenvalue of } P, \lambda \neq 1\}$.*

And relaxation time t_{rel}, is then defined as $t_{rel} := \frac{1}{\gamma_*}$. And this term further leads to the following theorem:

Theorem 3. *With the setting above and let μ the stationary measure, and $\mu_{min} := \min_{xin\chi} \mu(x)$. Then*

$$t_{mix}(\epsilon) \leq \lceil t_{rel}[\frac{1}{2} log(\frac{1}{\mu_{min}}) + log(\frac{1}{2\epsilon})]\rceil \leq t_{rel} log(\frac{1}{\epsilon \mu_{min}}). \tag{12}$$

The proof is approachable and can be found in Chapter 12.2 of [2]. However, determining the absolute spectral gap is not always straightforward, and one may try to approximate this upper bound of the mixing time by using a lower bound of δ_*. One method is given by Chen in 1998:

Theorem 4. *If χ is a metric space with metric ρ. Suppose there exists a constant $\theta < 1$ such that $\forall x,\ y \in \chi$, there exists a coupling (X_1, Y_1) of $P(x, \cdot)$ and $P(y, \cdot)$ satisfying*

$$\mathbb{E}_{x,y}(\rho(X_1, Y_1)) \leq \theta \rho(x, y), \tag{13}$$

then $\lambda_* \leq \theta$.

And we will introduce a related theorem from Wilson's paper.

Another theorem utilizing spectral theory is exactly developed by Wilson in this paper we will discuss, called Wilson's method, which have already got familiar to readers in Chapter 13.5 in [2].

Lemma 1. *Let Φ be an eigenfunction of P with eigenvalue $\frac{1}{2} < \lambda < 1$, For $0 < \epsilon < 1$ and let $R > 0$ satisfy*

$$\mathbb{E}_x(|\Phi(X_1) - \Phi(x)|^2) \leq R \tag{14}$$

[2]From now, in this subsection, we demonstrate everything with this presumed condition.
[3]Interested readers can refer to their solutions to exercise 12.1 in [2].

for all $x \in \chi$. Then for any x,

$$t_{mix}(\epsilon) \geq \frac{1}{2log(\frac{1}{\lambda})}[log(\frac{(1-\lambda\Phi^2(x))}{2R} + log(\frac{1-\epsilon}{\epsilon}))]. \tag{15}$$

In the next section, we will present the original version of this lemma.

2 Lattice Path Markov Chain

We highlight this section with detailed proofs and explanations, because not only the main methods and lemmas will be provided here, but lattice paths Markov chain will be further utilized throughout this essay.

2.1 Background: Problem Setting and Motivation

In this article, a lattice path is defined as a traversal from the leftmost corner to the rightmost corner of an $a \times b$ rectangle, as illustrated in Figure 1 for the case where $a = 4$ and $b = 5$. The lattice path Markov chain is a chain with state space χ, which is the set of all possible lattice paths with the given a and b. Letting $n = a + b$, we have $|\chi| = \frac{n!}{a!b!}$. The transition rule for this chain is as follows: for any path in χ, one of the $n-1$ internal columns (points) is randomly chosen and refreshed by pushing it up or down with probability $\frac{1}{2}$, provided that the resulting path is valid. If the resulting path is not valid, we let the path remain idle. Note that a point can only be refreshed if it is a local maximum or local minimum. When we encode each upward move as 1 and each downward move as 0, we obtain a permutation of a 0s and b 1s. This transition rule is equivalent to performing a random adjacent shift on this binary sequence, as we will discuss in the next section. One can notice that this chain is irreducible, aperiodic and reversible, with the stationary distribution Unif(χ).

Since the lattice path Markov chain can be extended to other types of Markov chains, such as the random adjacent transportation and lozenge-tiling Markov chain, it is important to understand the mixing technique of this chain and to find upper and lower bounds for its mixing time.

To analyze this chain, Wilson devised two functions that measure the displacement of a single path and the gap between two ordered paths (where each point of one path is not below the other). The height function of a point in a path describes the deviation of this point from the diagonal line. This can be defined by the following iterative process: the rectangle is equipped with a horizontal coordinate ranging from $-\frac{n}{2}$ (leftmost) to $\frac{n}{2}$ (rightmost) with a step size of $\frac{1}{2}$. Then,

$$h(\frac{n}{2}) = h(-\frac{n}{2}) = 0, \ and \ h(x) = \begin{cases} h(x-\frac{1}{2}) + \frac{a}{n}, & if \ there \ is \ an \ up \ move \ from \ x-\frac{1}{2} \ to \ x, \\ h(x-\frac{1}{2}) - \frac{b}{n}, & otherwise \end{cases} \tag{16}$$

A probabilistic explanation of this function is that h(x) is the difference between the number of upper moves

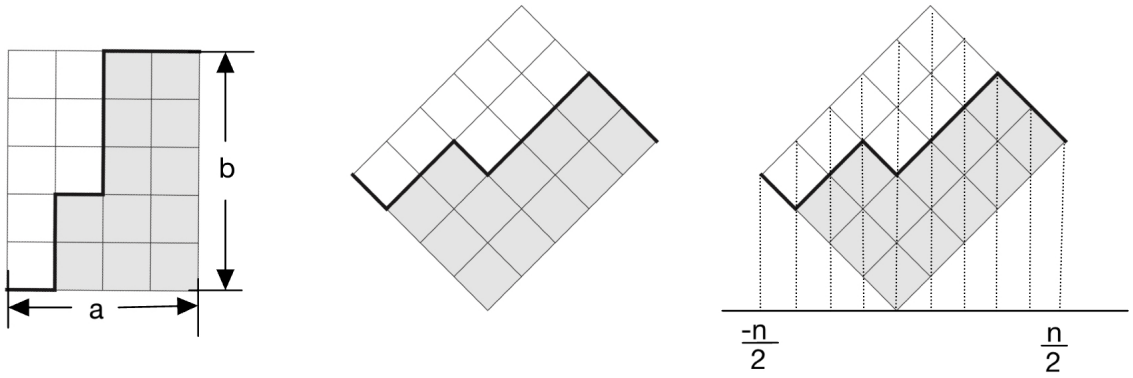

Figure 1: Three lattice paths with $a = 4$ and $b = 5$ (i.e., $n = 9$). The middle one is obtained by rotating the left graph $45°$, and the right one is the middle graph with horizontal coordinates equipped.

to x and its expectation. And based on this height function, the displacement function Φ of h is

$$\Phi(h) = \sum_{x=-\frac{n}{x}}^{\frac{n}{2}} h(x) cos\frac{\beta x}{n},\ \beta \in [0, \pi]. \tag{17}$$

Note that each internal point x corresponds to a positive weight of $cos\frac{\beta x}{n}$. When $\beta = 0$, equal emphasis is placed on each position, and as β increases, the deviation in the middle of the path is highlighted. When comparing two comparable states (i.e., every point in one path is not lower than the corresponding point in another) characterized by $\check{h}$ and $\hat{h}$, we can assume without loss of generality that $\check{h}(x) \leq \hat{h}(x)$ for all x. The gap function between these states is

$$\Phi(\check{h}, \hat{h}) = \Phi(\hat{h} - \check{h}) = \Phi(\hat{h}) - \Phi(\check{h}), \tag{18}$$

because Φ is linear. As mentioned above, Wilson's paper talked less about spectral decomposition of the height function. However, Wilson controlled the weights of the points to obtain the optimal upper bounds.

2.2 Contraction Property and Upper Bound

Using the definitions and interpretations presented in the previous subsection, we will now explain how to obtain an upper bound for this chain. It is important to note that this chain exhibits the contraction property, which can be described by the following lemma.

Lemma 2. *Suppose that either*

$$h \text{ is a height function and } \beta = \pi, \text{ or},$$
$$h = \hat{h} - \check{h} \text{ is a gap function and } \beta \in [0, \pi],$$

Let h' be the height function or gap function after one step of transition. Then

$$\mathbb{E}[\Phi(h') - \Phi(h)|h] \leq \frac{-1 + cos\frac{\beta}{n}}{n-1}\Phi(h) \tag{19}$$

with equality holds when $\beta = \pi$[4]*.*

We point that

$$-\frac{\beta^2}{2n^2(n-1)} \leq \frac{-1 + cos\frac{\beta}{n}}{n-1} \leq -\frac{\beta}{2n^3}, \tag{20}$$

and hence we can control the bound further using a more readable version. Although we do not provide the proof of the inequality (19) due to the intensive computation and length limit of this paper, it is important to emphasize that with the contraction property and coupling, one can obtain a desired upper bound.

Theorem 5. *When n is large, after $\frac{2+o(1)}{\pi^2}n^3 log\frac{ab}{\epsilon}$ steps, $d(t) \leq \epsilon$ and the two extreme paths have coalesced with probability greater than $1 - \epsilon$.*

Proof. Recall that Corollary 1 provides an approach to upper bound the mixing time. In our coupling method, we consider two lattice paths and randomly select a common point to determine whether to sort or unsort them simultaneously. This coupling preserves the partial order, as coalesced points move together. By induction, we can verify that the topmost and bottommost paths coalesce with the lowest probability after a fixed time. Therefore, we can focus on the pair $(\hat{h}_t, \check{h}_t)$ with the initial state $(\hat{h}_0, \check{h}_0)$ mentioned earlier. Let $\Phi_t = \hat{h}_t - \check{h}_t$, and we have that $\Phi_t = 0$ only if $\hat{h}_t = \check{h}_t$, which occurs when the two chains coalesce.

From Lemma 2, we have that

$$\mathbb{E}[\Phi_t] \leq \Phi_0[1 - \frac{1 - cos\frac{\beta}{n}}{n-1}]^t \leq \Phi_0[1 - \frac{\beta^2}{2n^3}]^t \leq \Phi_0 e^{-\frac{\beta^2}{2n^3}t} \tag{21}$$

Notice that $\mathbb{E}[\Phi_t] \geq \mathbb{P}(\Phi_t > 0)\Phi_{min} = \mathbb{P}(\hat{h}_t \neq \check{h}_t)\Phi_{min}$, where $\Phi_{min} = cos[\beta\frac{\frac{n}{2}-1}{n}]$ the lowest positive gap, according to Definition (17). We can further notice that $\Phi_{min} > cos\frac{\beta}{2} \approx \frac{\pi-\beta}{2}$ and $\Phi_0 \leq ab$ (because each weight is not greater than 1). And with inequality (21), we have that when $t \geq \frac{2}{\beta}n^3 log(\frac{\Phi_0}{\Phi_{min}\epsilon})$, $\mathbb{P}[\hat{h}_t \neq \check{h}_t] \leq \epsilon$. The optimal value of β to control this bound is let $\beta = \pi - \Theta(\frac{1}{log(n)})$, such that $\pi - \beta \to 0$ whereas $log\frac{1}{\pi-\beta} \ll log(ab)$ and finally we have that

$$t \geq (\frac{2}{\pi^2} + \Theta(\frac{log(log(n))}{log(n)}))n^3 log(\frac{ab}{\epsilon}), \tag{22}$$

which is exactly the upper bound. □

The subtlety in Wilson's idea here is to replace the traditional integer-valued metric (e.g. Hamming distance) with the gap function, which not only preserves the property of the metric of the partially ordered

[4]This equality is extremely important, as it will be further used to obtain an eigendecomposition and to construct a lower bound of mixing time.

paths, but also can be controlled by adjusting the weight. A similar reasoning using absolute absolute spectral gap is Theorem 4.

2.3 Wilson's Method and Lower Bound

In Section 1.3, we briefly discussed Wilson's method (lemma 1) in the context of eigendecomposition. And in Wilson's original paper, this lemma states as follows:

Lemma 3. *If a function Φ on the state space of a Markov chain satisfies*

$$\mathbb{E}(\Phi(X_{t+1})|X_t) = (1-\gamma)\Phi(X_t), \tag{23}$$

and $\mathbb{E}[(\Delta\Phi)^2|X_0] \leq R$ for any state X_0, where $\Delta\Phi = \Phi(X_1) - \Phi(X_0)$, then when the number of move t is bounded by

$$\frac{log(\Phi_{max}) + \frac{1}{2}log(\frac{\gamma\epsilon}{4R})}{-log(1-\gamma)}, \tag{24}$$

and

$$either\ 0 \leq \gamma \leq 2-\sqrt{2},\ or\ 0 < \gamma \leq 1\ with\ odd\ t, \tag{25}$$

we have $d(t) \geq 1-\epsilon$.

We remark that although this lemma bears similarity to Lemma 1, it is not as sharp as the version presented by Levin and Peres after ensuring consistency in symbols. Nonetheless, it still performs well for this lattice path Markov chain. The proof of this main result is as follows:

Proof. We denote $\Phi_t = \Phi(X_t)$ for simplicity. As at equilibrium, each states occurs with the same probability, $\mathbb{E}(\Phi) = 0$. Also, $\Phi_{t+1}^2 = \Phi_t^2 + 2\Phi_t\Delta\Phi + (\Delta\Phi)^2$, and thus $\mathbb{E}[\Phi_{t+1}] \leq (1-2\gamma)\Phi_t^2 + \mathbb{E}[(\Delta\Phi)^2|\Phi_t] \leq (1-2\gamma)\Phi_t^2 + R$. by subtracting $\frac{R}{2\gamma}$ in both sides and taking iteration, we have that

$$\mathbb{E}[\Phi_t^2] \leq \Phi_0^2(1-2\gamma)^t + \frac{R}{2\gamma}. \tag{26}$$

By induction we further have that

$$\mathbb{E}[\Phi_t] = \Phi_0(1-\gamma)^t. \tag{27}$$

Letting 26-27[2], the upper bound of the variance is given by

$$Var[\Phi_t] \leq \Phi_0^2[(1-2\gamma)^t(1-\gamma)^{2t}] + \frac{R}{2\gamma} \leq \frac{R}{2\gamma}. \tag{28}$$

We point that the the condition 25 makes the last inequality hold. Now we apply Chebychev's inequality

and have that $\mathbb{P}[|\Phi_t - \mathbb{E}(\Phi_t)| \geq \sqrt{\frac{R}{2\gamma\epsilon}}] \leq \epsilon$. When $\mathbb{E}(\Phi_t) \geq \frac{4R}{\gamma\epsilon}$,

$$\mathbb{P}[|\Phi_t| \leq \sqrt{\frac{R}{\gamma\epsilon}}] \leq \mathbb{P}[\Phi_t \leq \sqrt{\frac{R}{\gamma\epsilon}}] \leq \mathbb{P}[\Phi_t \leq \mathbb{E}(\Phi_t) - \sqrt{\frac{R}{\gamma\epsilon}}] \leq \mathbb{P}[|\Phi_t - \mathbb{E}(\Phi_t)| \geq \sqrt{\frac{R}{\gamma\epsilon}}] \leq \frac{\epsilon}{2}. \quad (29)$$

And when Φ approaches the equilibrium, $\mathbb{E}(\Phi) = 0$ and $\mathbb{P}[|\Phi| \geq \sqrt{\frac{R}{\gamma\epsilon}}] \leq \frac{\epsilon}{2}$. Thus the variation distance

$$d(t) \geq |P^t - \mu|\{X : |\Phi(X)| \leq \sqrt{\frac{R}{\gamma\epsilon}}\} \geq \mathbb{P}(|\Phi| \leq \sqrt{\frac{R}{\gamma\epsilon}}) - \mathbb{P}(|\Phi_t| \leq \sqrt{\frac{R}{\gamma\epsilon}}) \geq 1 - \frac{\epsilon}{2} - \frac{\epsilon}{2} = 1 - \epsilon. \quad (30)$$

When the initial state in χ maximizes $\Phi_0 = \Phi_{max}$, after t steps such that $\mathbb{E}(\Phi_t) = \Phi_{max}(1-\gamma)^t \geq \frac{4R}{\gamma\epsilon}$, i.e.,

$$t \leq \frac{log[\frac{\Phi_{max}}{\sqrt{\frac{4R}{\gamma\epsilon}}}]}{-log(1-\gamma)}, \quad (31)$$

we have that $d(t) \geq 1 - \epsilon$. □

We point out that the Wilson's method is used to derive the lower bound involves evaluating the deviation of the displacement functions of the states at a given time t and at equilibrium. If the range of values in their "confidence intervals" do not overlap significantly, it will indicate that mixing has not yet occurred. This concept can also be observed in the works of Diaconis and Shahshahni, as well as Lee and Yau, where they applied it to establish lower bounds for certain Markov chains.

Utilizing Lemma (3), we obtain the lower bound of lattice path Markov chain.

Theorem 6. *If $min(a,b) \gg 1$, then after*

$$\frac{1 - o(1)}{\pi^2} n^3 log[min(a,b)] \quad (32)$$

steps, the variation distance from stationarity is 1-o(1).

Proof. When $\beta = \pi$ and $min(a,b) \gg 1$, we have that (23) is satisfied with

$$\gamma = \frac{1 - cos\frac{\pi}{n}}{n-1} \sim \frac{\pi^2}{2n^3}, \quad (33)$$

We remark that condition (25) holds when $n \geq 3$. The bound of R is $\frac{mix(a,b)}{n-1}$, which happens when there is $mix(a,b)$ local extreme point, each with $\Delta\Phi \leq 1$. Taking γ, Φ_{max} to (24) and with the fact that $-log(1 - \frac{\pi^2}{n^3}) \geq \frac{\pi^2}{n^3}$, we can obtain the lower bound. □

3 Random Adjacent Transportation

From this section, we will introduce several well-known Markov chains whose mixing time bounds can be determined through the use of their contraction properties and Wilson's method. Our focus will be on explaining the methodology rather than providing detailed calculations.

3.1 Background: Problem Setting and Previous Work

One well-known Markov chain is the card shuffling by random adjacent transposition (RAT), which is familiar to most readers. The state space of this chain is the symmetric group S_n, with the transition rule that picking one of the n-1 adjacent pair and refresh this pair by sorting or reverse-sorting it with probability a half. Aldous (1983) demonstrated that to achieve equilibrium, $O(n^3 \log n)$ shuffles are sufficient, while $\Omega(n^3)$ shuffles are necessary. Aldous (1997) and Disconis (1997) presented heuristic arguments suggesting that $\Omega(n^3 \log n)$ is a lower bound, although no rigorous proof was provided. The readers were given the upper bound and lower bound in Section 16.1 in [2], where the upper bounds via comparison and coupling are respectively $O(n^4(logn)^2$ and $2n^3log_2n$, while the lower bounds via tracing a single card and Wilson's method (absolute spectral gap) are respectively $\frac{n^2(n-1)}{16}$ and $\frac{1-o(1)}{\pi^2}n^3logn$ – which is precisely the bound we will prove. In this section, we will demonstrate how this problem can be transformed into a lattice path Markov chain, and how the lower bounds are within a constant factor of the upper bounds. The refined result of its mixing time can be further applied to analysis of the mixing time of the card shuffling Markov chain via move-ahead-one update rule.

3.2 Upper and Lower Bounds

As mentioned earlier, while the lattice path can be represented as a binary sequence with a certain number of 0s and 1s, permutations in S_n cannot be directly represented in this way. However, we can characterize a permutation using n+1 lattice paths. We first represent a permutation using an n+1 by n matrix, where the i-th column corresponds to the number at the i-th position of the permutation (denoted by j). This column is an (n+1)-dimensional vector with 1s at each of the first j entries and 0s otherwise. In this way, the k-th row corresponds to n+1-k 1s and then k-1 0s. We can then represent each row as a lattice path, where the top path is strictly increasing and the bottom path is strictly decreasing. Each path is, by definition, dominated by the path above it and dominates the path below it without any point coalescing. A clear visualization of an example can be found in Figure 2.

We point out that the equilibrium of the $\lfloor \frac{n}{2} \rfloor + 1$-th lattice path is necessary condition for the equilibrium of RAT chain. For this path, we have that a = $\lfloor \frac{n}{2} \rfloor \gg 1$ and b = n - $\lfloor \frac{n}{2} \rfloor \gg 1$, when n is large. Without using absolute spectral gap but instead referring Theorem 6, we can conveniently show that the after $\frac{1-o(1)}{\pi^2}n^3logn$ steps, the variation distance is still $1 - o(1)$.

Now we apply Theorem 5 to establish an upper bound. We let $\epsilon = \frac{\delta}{n}$ and have that $\frac{ab}{\epsilon} \leq \frac{n^3}{\delta}$ for any

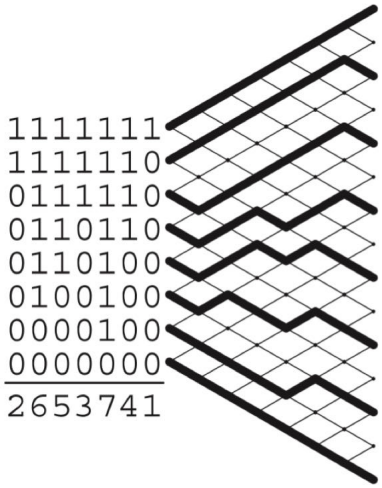

Figure 2: The n+1 (n=7) lattice paths which characterise the permutation 2653741

path. After $(\frac{2}{\pi^2} + o(1))n^3 log(\frac{n^3}{\delta})$ steps, each chain approaches equilibrium with probability $1 - \frac{\delta}{n}$ and thus the variation distance is at most δ. Similarly, we control δ such that $\delta \ll 1$ and $\log \frac{1}{\delta} \ll logn$. And we show that after $(\frac{6}{\pi^2} + o(1))n^3 log(n)$, $d(t) \leq o(1)$.

It's worth noting that though we've ensured the lower bounds are within a constant factor of the upper bounds, both bounds can still be refined. Wilson demonstrated that for a lattice path Markov chain, $t_{mix}(\epsilon) \leq (\frac{2}{\pi^2} + o(1))n^3 log(\frac{10n}{\epsilon})$. Specifically, when the path lies within a $\frac{n}{2} \times \frac{n}{2}$ box, $d(t) = 1 - o(1)$ when $t \leq (\frac{2}{\pi^2} - o(1))n^3 log(n)$. However, these bounds were obtained using different methods and their computation is not straightforward. We therefore refrain from delving deeper into this topic and suggest that readers who are interested in these methods refer to Section 8 in Wilson's paper.

4 Lozenge-tiling Markov Chain

4.1 Background: Problem Setting and Previous Work

Before delving into the history of this fascinating model, let us first explain the framework of random lozenge-tiling. Literally, lozenge-tiling is a way for tiling a region, such as a regular hexagon, using rhombuses with angles of 60° and 120°. This tiling can be visualized through non-intersecting lozenge lattice paths. An example of this tiling with side length $l = 20$, as well as an example of the corresponding characterization, can be seen in Figure 3, where the shadowy lozenges form seven lattice paths from the leftmost edge to the rightmost edge.

Wilson's paper analyzes the Markov chain introduced by Luby, Randall, and Sinclair, where each state can be represented by lattice paths. The tiled region is assumed to be simply connected to ensure that the lattice path can cross the interior without passing through a hole. The state space is the set of all workable tiling ways, which is equivalent to a group of lattice paths. The transition rule is nearly the same, randomly

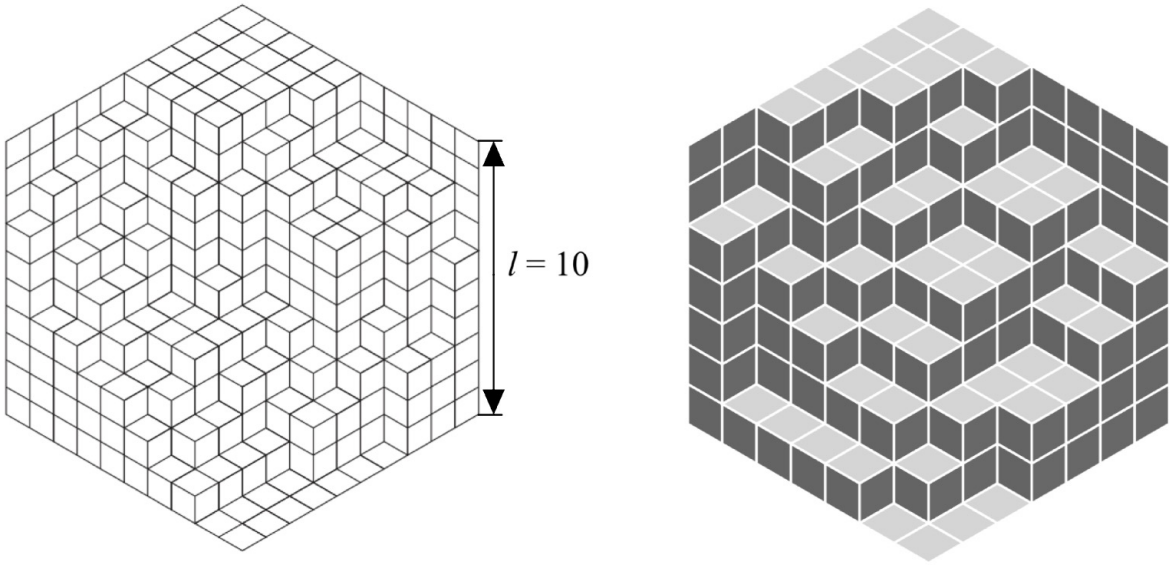

Figure 3: The left one shows an example of a lozenge-tiling in a hexagon, where l=10. The right figure visualizes an expression of a lozenge-tiling using l=7 lattice paths.

picking a point on a lattice path and deciding whether to push it up (if it is a local minimum) or down (if it is a local maximum) with a probability of one-half, except when two conditions arise: 1) refresh is not allowed because the point is bounded by the border of the tiled region and 2) refresh is not allowed because the operation will make the new path overlap the path lying above (below). If 1) occurs, no action is taken, and if 2) occurs, a "nonlocal move" is adopted. If there are k chains bounded above (below) the local minimum (maximum), and we decide to move the point up (down), then with a probability of $\frac{1}{k+1}$, we do so, and with a probability of $\frac{k}{k+1}$, we do nothing. An example can be seen below for this special situation in Figure 4. This operation can not only accelerate the mixing, but also preserve the expected change of the height, which will be mentioned later.

It is worth noting that random tiling encompasses various types of models, as it has been extensively studied in physics as a model of dimer systems. Among these models, random lozenge tiling stands out as a popular one, where monomers are assumed to be two regular triangles contained in a lozenge. Since the late 1990s, numerous studies have investigated the asymptotic properties of this tiling model. Several articles describe techniques for randomly generating lozenge tilings, such as Markov chain and linear algebra approaches[5]. A classical model of the former is the Luby-Randall-Sinclair Markov chain, which is of great importance in studying the mixing time of this chain. This is because we need to ensure that the chain efficiently converges, and to establish a physical interpretation for it that connects its mixing properties to dimer systems. Moreover, these properties may be extended to other types of random tiling, such as tiling a whole plane.

[5]Wilson also developed the monotone coupling from the past (monotone-CFPT) method for generating random structures such as lozenge tilings. For further information, readers can consult Chapter 25 in [2].

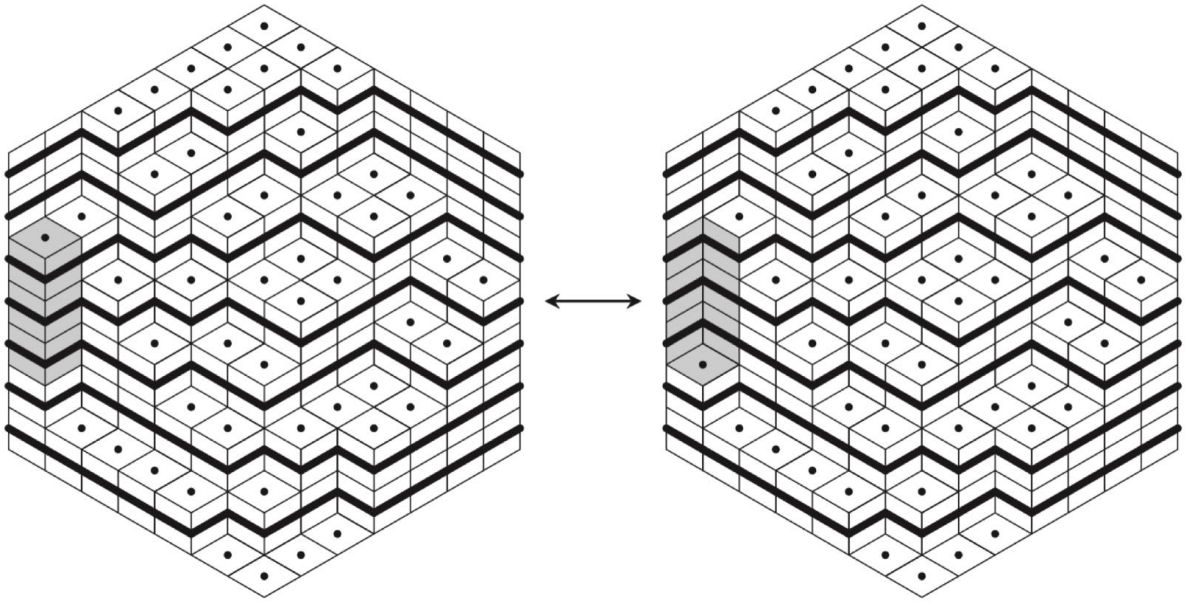

Figure 4: An example of a nonlocal move in Luby-Randall-Sinclair Markov chain, where from the left one to the right one, if we decide to move up the lowest shadowy point, we move all three w.p. $\frac{1}{3}$, and vice versa.

Luby, Randall, and Sinclair analyzed this Markov chain and proved that its mixing time is of the order $O(n^4)$, which they further reduced to $O(n^{3.5})$. For a regular hexagon tiling with side length l, the bound is $O(l^7)$. Wilson's paper further improves this upper bound and provides a lower bound for the hexagon tiling.

4.2 Upper and Lower Bounds

It is easy to observe that the Luby-Randall-Sinclair Markov chain is irreducible, aperiodic, reversible, and has a stationary uniform distribution (as it is symmetric). The region where the lattice paths reside, which is equivalent to the tiled region, is assumed to have a width of n (as arranged in section 2), with m local moves separating the top and bottom configurations, and p points may be moved.

Constructing an upper bound is similar to the lattice path Markov chain, as it involves using the contraction property of the gap function. In this context, we define the height function of path i with a minor adjustment:

$$h_i(-\frac{n}{2}) = 0,\ and h_i(x) = \begin{cases} h_i(x-\frac{1}{2})+\frac{1}{2}, & if\ there\ is\ an\ up\ move\ from\ x-\frac{1}{2}\ to\ x, \\ h_i(x-\frac{1}{2})-\frac{1}{2}, & otherwise \end{cases} \tag{34}$$

Hence each local move results a height change by 1. And the displacement function of a tiling is modified to

$$\Phi(h) = \sum_i \sum_{x=-\frac{n}{2}}^{\frac{n}{2}} h_i(x) cos\frac{\beta x}{n},\ \beta \in [0,\pi]. \tag{35}$$

We further introduce the gap function between two ordered tilings as $g = \hat{h} - \check{h}$, where $\hat{h}$ and $\check{h}$ represent the height functions of the tilings, respectively. We define $\Phi(g) = \Phi(\hat{h}) - \Phi(\check{h})$. It is important to note that the path may not have a length of n, and for any point x where path i does not exist, we set $h_i(x)$ to an arbitrary constant. This ensures that the gap at this location on the path is zero. Additionally, we observe that $\Phi(g) = 0$ if and only if $\hat{h}$ and $\check{h}$ completely overlap.

With this setting, given the condition that the location x at path i is picked, we have that

$$\mathbb{E}[\Delta\Phi(g)|x,i] \leq [\frac{g_i(x-\frac{1}{2}) + g_i(x+\frac{1}{2})}{2} - g_i(x)]cos(\frac{\beta x}{n}) := B(g,x,i). \tag{36}$$

Readers can verify that the equality holds even when a nonlocal move happens, if there are no "border effects", meaning that the border does not obstruct lattice movement. In this explanation, we mainly consider situations where border effects exist. If $\mathbb{E}(\hat{h}(x)) < B(\hat{h},x,i)$ and $\mathbb{E}(\check{h}(x)) > B(\hat{h},x,i)$,we directly have that $\mathbb{E}[\Delta\Phi(g)|x,i] < B(g,x,i)$. If $\mathbb{E}(\hat{h}(x)) > B(\hat{h},x,i)$, then the point hits its minimum value (as do $\hat{h}_i(x-\frac{1}{2})$ and $\hat{h}_i(x+\frac{1}{2})$) with a local maximum. Since $\hat{h}$ dominates $\check{h}$, we can see that $\check{h}$ is in the same situation as $\hat{h}$. Therefore, $\mathbb{E}[\Delta\Phi|x,i] = 0$, and this holds when $\mathbb{E}(\check{h}(x)) < B(\hat{h},x,i)$ due to symmetry.

With some algebra and 36, we can construct the contraction property in this situation, which is

$$\mathbb{E}[\Delta\Phi] \leq \frac{-1 + cos(\frac{\beta}{n})}{p}\Phi, \tag{37}$$

and the equality holds when $\beta = \pi$. Using this property and following the computation in the proof of Theorem 5, we obtain that

$$t_{mix}(\epsilon) \leq \frac{2+o(1)}{\beta^2} pn^2 log(\frac{m}{\Phi_{min}\epsilon}). \tag{38}$$

Wilson further argued that the optimal way to lower the upper bound is to let $\beta = \pi - \Theta(\frac{1}{logn})$, which leads that

$$t_{mix}(\epsilon) \leq \frac{2+o(1)}{\pi^2} pn^2 log(\frac{m}{\epsilon}). \tag{39}$$

When the tiling region is a regular hexagon with side length l, we have that $n = 2l$, $m = l^3$ and $p = l(2l-1)$ [6]. And the upper bound for $t_{mix}(\epsilon)$ is $\frac{48+o(1)}{\pi^2} l^4 log(l)$, which is much better than the polynomial bound given by Luby, Randall and Sinclair.

Since this is a regular hexagon, there are no "border effects." As l becomes larger, Theorem 3 is satisfied, where $\gamma \approx \frac{\beta^2}{2pn^2} \approx \frac{\pi^2}{16l^4}$ and $R < l$ due to the nonlocal move[7]. Therefore, according to Wilson's method, the lower bound is $\frac{8-o(1)}{\pi^2} l^4 \log(l)$.

[6] In the original paper, Wilson had that $p = 2l(l-1)$ but we do not think it is correct. However, this error does not influence the bound.

[7] The "worst" case is that each path have $2l-1$ local maxima and minima at the same positions. Then $R = \frac{l-1}{2p}\sum_{k=1}^{l}\frac{k^2}{k} = \frac{l+1}{4} < l$

5 Karzanov-Khachiyan Markov Chain

In this section, we introduce the Karzanov-Khachiyan Markov chain, which somehow differs from traditional Markov chains in terms of its state space. Through this model cannot be converted to or characterized by lattice paths, studying methods to optimize its bound mixing time is still valuable. This involves controlling the weight of each displacement function and adjusting the gap function.

5.1 Background: Problem Setting and Previous Work

The Karzanov-Khachiyan Markov chain was originally designed not to address physics problems or sample variables from stationary distributions, but rather to approximate the number of linear extensions of a partially ordered set. Counting these extensions is a difficult problem, proven to be P-complete. In 1991, Matthews introduced a geometric Markov chain for generating random linear extensions, which achieved a mixing time of $O(550n^8(logn)^3log(1/\epsilon))$. That same year, Karzanov and Khachiyan introduced the combinatorial Karzanov-Khachiyan chain, which achieved a mixing time bound of $O(n^6logn)$. Specifically, they showed that the mixing time was upper bounded by $8n^5log(|\chi|/\epsilon)$, where χ is the state space of all possible linear extensions. This fruitful year had also seen the improvement of this upper bound given by Dyer and Frieze, which is $O(n^4log(\frac{|\chi|}{\epsilon})) \leq O(n^5logn)$. Furthermore, in 1998, Bubley and Dyer demonstrated that a related chain mixed in $O(n^3logn)$, while the original chain mixed in $O(n^3lognlog(\frac{|\chi|}{\epsilon})) \leq O(n^4(logn)^2)$.

Bubley and Dyer defined this revised Markov chain in which the state space consists of all possible linear extensions of a poset with n elements. Unlike the original model, where pairs of adjacent points were chosen uniformly, this chain selects items at positions i and i+1 with a probability $f(i) \propto i(n-i)$. Bubley and Dyer compared both chains and demonstrated that the mixing time of the original chain is upper bounded by $O(n^4(logn)^2)$, whereas the revised chain mixes in time $O(\frac{1+o(1)}{3})n^3logn$. Wilson used the contraction property of the original Markov chain to show that its mixing time is no greater than $O(\frac{4+o(1)}{\pi^2})n^3logn$ [8].

5.2 Upper Bound

We will focus solely on the original Karzanov-Khachiyan chain in this discussion, as the argument for the upper bound of the revised chain is nearly identical, which can be found in Wilson's paper. The width or weight of transporting the positions i and i+1 is defined by the function $w(i)$. Given two partially ordered extensions X and Y, the distance function $\delta(X,Y)$ is defined as the minimum sum of transportation weights required to transform X into Y, i.e., $\delta(X,Y) = \min\sum_{i=1}^{r-1}\delta(Z_i, Z_{i+1})$, where $X = Z_0, Z_1, ..., Z_r = Y$ is a sequence of valid extensions, and Z_i and Z_{i+1} differ by a single transportation. While Bubley and Dyer showed that a constant weight can optimize the upper bound, Wilson argued that the best choice in a situation with uniform frequency is a sinusoidal weight. By letting $w(i) = \cos[\beta(\frac{i}{n} - \frac{1}{2})]$, where $\beta \in [0, \pi]$

[8] Although this upper bound is not better than that of the revised chain, the original chain is still preferred because it is more convenient to pick pairs uniformly.

can be controlled, one can obtain the optimal weight. For any two permutations A and B that differ by a single transportation (i,j), one can transform them into A' and B' using Markovian coupling,

$$\mathbb{E}(\delta(A',B') = \delta(A,B) + \frac{1}{2}[f(i-1)w(i-1) - f(i)w(i) - f(j-1)w(j-1) + f(j)w(j)] = \delta(A,B)(1-\gamma_{i,j}), \quad (40)$$

where

$$\gamma_{i,j} = \frac{-1}{2}\frac{[f(i-1)w(i-1) - f(i)w(i) - f(j-1)w(j-1) + f(j)w(j)]}{\delta(A,B)}. \quad (41)$$

Wilson noted that to determine the smallest value of γ such that $\mathbb{E}(\delta(A',B') = \delta(A,B)(1-\gamma)$, it suffices to find $min_{i,j}\gamma_{i,j}$, which is equivalent to finding $min_{i,i+1}\gamma_{i,i+1}$. This is because $\gamma_{i,j} \geq c$ if and only if $\gamma_{i,j} \geq c$ for any i. (A heuristic way to understand this is that the supremum of $|f'|$ always dominates the absolute average change.) Therefore, we only need to minimize the expression:

$$\gamma_{i,i+1} = \frac{1}{2(n-1)cos[\beta(\frac{i}{n}-\frac{1}{2})]}(-cos[\beta(\frac{i-1}{n}-\frac{1}{2})+2cos[\beta(\frac{i}{n}-\frac{1}{2})]-cos[\beta(\frac{i+1}{n}-\frac{1}{2})) = \frac{1-cos(\beta/n)}{n-1} \geq \frac{\beta^2}{2n^3}. \quad (42)$$

We now simplify the problem by using a familiar version of it. By setting $\beta = \pi - \Theta(\frac{1}{\log n})$, we can ensure that γ is sufficiently large, and that the ratio of the largest distance to the smallest distance, denoted by D, is not too large ($O(n^2)$). With these restrictions, we can bound the upper limit of $t_{mix}(\epsilon)$:

$$\frac{2n^3}{\beta^2}\log\frac{D}{\epsilon} = \frac{4+o(1)}{\pi^2}n^3\log n. \quad (43)$$

6 Discussion

Considering the lack of space, and that readers are familiar with related knowledge in the class, we will not include the application of Wilson's method to lower bound the mixing times of certain Markov chains, such as card shuffling on a hypercube or a grid, nearest-neighbor random walk on product graphs, and the Markov chain introduced by Diaconis and Saloff-Coste. Interested readers may refer to chapter 9 of the original article for more information. In this section, we provide a brief summary and highlight several open problems and potential directions for future research in this report essay.

6.1 Conclusion

For Markov chains with complex structures, one may define a distance or displacement function on the state space with adjustable weights. By examining the contraction property of the gap function, which measures the difference between partially ordered states, one can obtain a lower bound on the absolute spectral gap and an upper bound on the mixing time. Moreover, by controlling the weight, an optimal bound

can be selected. Alternatively, if the expected one-step change of the displacement function is proportional to the displacement of the initial state[9] and this conditional change has a bounded second moment, Wilson's method can be used to construct a lower bound.

6.2 Potential Avenues for Future Research

We will briefly discuss two key challenges that merit further exploration: verifying the existence of cutoff and designing fast-mixing chains.

6.2.1 The Cutoff Phenomenon

One of the most interesting directions in Markov chain mixing time is the study of cutoff phenomena. In previous examples, the upper and lower bounds of mixing time for a chain often differ by only a small constant factor, indicating the possibility of a cutoff. Proving the existence of a cutoff in a Markov chain allows for a more precise determination of the mixing time threshold within a negligible window.

In his article, Wilson provided some heuristic arguments based on several assumptions about the multiplicity of the second-largest eigenvalue, the geometric decrease of variation and separation distances, and other factors. However, there is a need for rigorous analysis to determine whether a cutoff exists in the aforementioned chains.

6.2.2 Designing a Markov Chain

In section 5, we discussed two methods for updating a linear extension of a poset, which differ only in the probability of selecting an adjacent pair. This suggests that in addition to controlling weights to obtain optimal upper bounds, adjusting the frequency function may be a viable approach to accelerate mixing. Wilson noted that this could be an interesting challenge for future research.

Acknowledgement

I would like to express my gratitude to Professor Ivan Corwin and Dr. Weitao Zhu for their outstanding lectures this semester. Their teachings have greatly contributed to my understanding and engagement in the study of Markov chain mixing time. In particular, I am thankful to Professor Corwin for suggesting the paper that inspired this report, and to Dr. Zhu for her explanations for me of the model in Chapter 8, though it was not directly covered in this essay.

[9] Actually this function is an eigenfunction, and we should determine if the proportion (or equivalently, the eigenvalue) is in the valid range.

References

[1] Wilson, David Bruce. "Mixing times of lozenge tiling and card shuffling Markov chains". In: *The Annals of Applied Probability* 14.1 (2004), pp. 274–325.

[2] Levin, David A and Peres, Yuval. *Markov chains and mixing times*. Vol. 107. American Mathematical Soc., 2017.