PERMUTATIONS WITH INTERVAL RESTRICTIONS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF MATHEMATICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Olena Blumberg
January 2012

Abstract

This thesis studies the problem of the random transposition walk on permutations with interval restrictions. The mixing time of this Markov chain is explored, and a number of different cases are considered. For the case of bounded interval restrictions, an $O(n^4)$ bound for the mixing time is achieved. For a specific example of bounded interval restrictions called Fibonacci permutations, it is shown that a sped-up walk using only adjacent transpositions mixes in order $n \log n$ time. An example of a family of interval restriction matrices for which the random walk mixes in exponential time is provided, showing that the walk in general does not mix in polynomial time. The case of one-sided interval restrictions is also studied, and cut-off is shown for a large class of one-sided interval restriction matrices. Furthermore, examples are provided in which chi-squared cut-off occurs, while total variation mixing occurs significantly earlier without cut-off. Finally, a coupling argument showing that the random transposition walk on the whole symmetric group mixes in $O(n \log n)$ time is presented. This is achieved via projection to conjugacy classes and then a path coupling argument.

Acknowledgements

First of all, I would like to thank my advisor, Persi Diaconis, for introducing to a beautiful area of mathematics, guiding my journey through it, and reading my thesis with scrupulous care. I am immensely grateful.

Thank you to my family, who raised me with an enduring love for math and who always encouraged and supported my interest in it. A particular thank you to my grandmother Luda, a wonderful teacher who introduced me to the subject and nurtured my love for it for many years. I wish you were around to see this.

Thank you to all the professors who have shown interest in my work over the years, including but not limited to Nathanael Berestycki, Amir Dembo, Ben Morris, Yuval Peres, and Ofer Zeitouni.

A big thanks to all my Stanford friends – you've made the last five years an experience I'll cherish. And an extra big thank you to Sukhada Fadnavis and Aaron Smith, who I've talked shop with over the years. Thanks for listening, guys; I hope I occasionally returned the favor!

And finally, a gigantic thank you to my husband Andrew for being eternally patient, loving, and supportive. The days of thesis complaints are now over; unfortunately for you, I'll probably find something else to worry about. Thanks, love!

Contents

Chapter 1

Introduction

Permutations with restricted positions are elements σ of the symmetric group S_n such that for each i, $\sigma(i)$ is only allowed to be in a particular subset of $\{1, 2, \ldots, n\}$. These are specified by a $\{0, 1\}$-matrix M, where $\sigma(i)$ is allowed to be equal to j precisely if $M(i, j) = 1$. Let S_M be the set of permutations corresponding to M. More concisely,

$$S_M = \{\sigma \mid M(i, \sigma(i)) = 1 \text{ for all } i\} \tag{1.1}$$

If $S(i) = \{j \mid M(i, j) = 1\}$ then $S(i)$ is the set of allowable values for $\sigma(i)$. Call $\{0, 1\}$-matrices M *restriction matrices.* If $S(i)$ is an interval for each i, then call M an *interval* restriction matrix.

There are many familiar examples of sets of permutations with restricted positions. The simplest case is the one where M is the $n \times n$ matrix of all 1s. In this case, each set $S(i)$ is $\{1, 2, \ldots, n\}$ and there are no restrictions; therefore, $S_M = S_n$. This M is clearly an interval restriction matrix. A slightly trickier example is one where M has 0s on the diagonal and 1s everywhere else. The following is the example for $n = 4$:

$$\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

In this case, the set $S(i)$ is precisely $\{1, 2, \ldots, n\}/\{i\}$. Here, the restriction on $\sigma \in S_M$ is that $\sigma(i)$ can be anything other than i. Therefore, S_M is the set of derangements of n. In this case, M is not a interval restriction matrix.

While the above examples are easy to analyse, enumerating permutations with a given set of restrictions is usually very tricky. Even if each $S(i)$ is an interval with a small number of elements, the size of S_M is not at all apparent. This problem is equivalent to evaluating the *permanent* of a $\{0,1\}$-matrix, where the permanent is defined as

$$\mathrm{Per}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} A(i, \sigma(i)).$$

The permanent is reminiscent of the determinant without the minus signs. However, unlike the determinant, calculating the permanent is a provably difficult problem. The developments in this area are summarized in Section 1.1.

The permanent also has connections to combinatorics, in particular to matching theory. Recall that a graph G consists of a set of vertices V and edges E, where an edge is an undirected pair of distinct vertices. A *matching* in G is a subset S of E such that each vertex in V appears in at most one edge in S. A perfect matching is one that uses every vertex in V. Given a $n \times n$ $\{0,1\}$-matrix M, construct a graph G_M on the vertex set $\{1, 2, \ldots, n\} \cup \{1', 2', \ldots, n'\}$ such that i is connected to j' if and only if $M(i, j) = 1$. Then G_M is a bipartite graph on $2n$ vertices, and M is the adjacency graph of G_M. Furthermore, the perfect matchings in G_M are in bijection with the permutations in S_M. Hence, calculating the size of S_M is equivalent to calculating the number of matchings in a bipartite graph. Matching theory is a thriving area, and bipartite matchings are the simplest case. An excellent reference on this subject is the book of Lovász and Plummer [37].

An algebraic approach to bipartite matchings is rook theory, which visualizes the permutations with restricted positions as placements of non-attacking rooks on a chess-board. Using the terminology of chess, let a board B be a subset of the $n \times n$ board $\{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$, and let $r(B, k)$ be the number of non-attacking rook placements of k rooks on a board B. Since rooks can move any number of squares either vertically or horizontally, such a configuration shouldn't have any rooks in the same row or column. Then, following the notation of Stanley [49], define the *rook polynomial* of B by

$$R_B(x) = \sum_{k=1}^{\infty} r(B, k) x^k$$

Since for $k > n$, $r(B, k) = 0$, this is indeed a polynomial. If M is a $n \times n$ restriction matrix, it can be used as a board by treating the matrix M as a $n \times n$ grid, where the 1s are included in the board and the 0s are not. In this case, $r\left(M, n\right)$ is equal to $|S_M|$, while in general, $r\left(M, k\right)$ is the number of matchings with k edges in the bipartite graph G_M defined above. Rook polynomials satisfy useful recurrences and have a number of interesting properties. Chapter 2 of Stanley [49] gives a useful treatment of the essentials. There's also an extensive review of the classical literature in the book by Riordan [45], while Lovász and Plummer [37] extend the theory to general graphs (following Heilmann-Lieb.)

Permutations with restricted positions also make an appearance in statistics, in particular, in testing for independence based on paired data (X_1, Y_1), $(X_2, Y_2), \ldots, (X_n, Y_n)$. Such data sets often come with natural truncations: for each x, there exists a set $S(x)$ such that a pair (X, Y) can be observed precisely if Y is in $S(X)$. For an easy example, if X measures the velocity of the fastest red car on the street and Y measures the velocity of the fastest car of any color, then it is self-evident that $X \leq Y$. Such truncations necessarily induce dependence. However, it is possible to test for independence given the constraints: in this case, independence would imply that for any $\sigma \in S_n$ such that $Y_{\sigma(i)} \in S(X_i)$ for all i, the paired data set $\left(X_1, Y_{\sigma(1)}\right), \left(X_2, Y_{\sigma(2)}\right), \ldots, \left(X_n, Y_{\sigma(n)}\right)$ is equally likely. Define the $\{0,1\}$-matrix M by

$$M(i,j) = \begin{cases} 1 & \text{if } \; Y_i \in S(X_j) \\ 0 & \text{otherwise} \end{cases}$$

The σ above are precisely the σ in S_M. Therefore, sampling from S_M allows standard tests for independence to be used. There is further discussion of this in Diaconis, Graham, and Holmes [11]; the topic will also be elaborated on in Chapter 3.

For the remainder of this document, it is assumed that M is an interval restriction matrix. The main objects of study are slight variants of the following random tranposition walk: pick a tranposition (i, j) uniformly at random from the $1/\binom{n}{2}$ possibilities, and attempt to transpose i and j. If this tranposition takes us to an element of S_M, move there; if this transposition takes us outside S_M, the random walk stays in place. It is shown in Diaconis, Graham, and Holmes [11] that S_M is connected under this walk; furthemore, it is an appealingly simple walk. In particular, it is an example of a *Gibbs sampler*: a walk which proceeds by fixing most of the variables, and moving by using the low-dimensional distribution of the remaining variables. These reasons combine into making the random transposition walk a natural object of study.

An irreducible aperiodic Markov chain (X_t) on a finite state space will converge to its stationary distribution as $t \to \infty$ (see Chapter 15 of Feller [18]). There are a number of ways to measure convergence to stationarity; for probability measures π and σ on a finite set Ω, a common metric is the *total variation distance*, which is defined as follows:

$$\|\pi - \sigma\|_{TV} = \max_{A \subseteq X} |\pi(A) - \sigma(A)| = \frac{1}{2} \sum_{s \in S} |\pi(s) - \sigma(s)| \tag{1.2}$$

The *mixing time* of a Markov chain is defined to be the time it takes for the chain to be within a certain prescribed distance of the stationary distribution. Letting P be the transition matrix of the Markov chain, and letting π be its stationary distribution, define

$$d(t) = \max_{x} \left\| P^t(x, \cdot) - \pi \right\|_{TV} \tag{1.3}$$

and

$$\tau_{\text{mix}}(\epsilon) = \min \{ t \mid d(t) \leq \epsilon \} \tag{1.4}$$

Define τ_{mix} to be $\tau_{\text{mix}}(1/4)$. The questions tackled in this thesis are the standard questions for the field of Markov chain mixing times: namely, given a family of Markov chains $(X_t^n)_{t \geq 0}$, what is the behavior of $\tau_{\text{mix}}^{(n)}(\epsilon)$ as $n \to \infty$.

1.1 Motivation and Contributions

This thesis is concerned with sampling from sets of permutations with interval restrictions. It is well known that getting an almost random sample from a set is equivalent to approximating its size [30]. Therefore, the goal of this document is to to approximate $\text{Per}(M)$, where M is a interval restriction matrix.

Determinining the permanent of a matrix is an old and celebrated problem. The fastest known exact algorithm is due to Ryser [46] and takes $O(n2^n)$ time. As this algorithm takes exponentially long to run, it is still far too slow for most applications. This state of affairs was explained by Valiant in 1979 [51], who used calculating the permanent of a $\{0,1\}$-matrix as the first example of a #P-complete problem: a counting problem which is not expected to have a polynomial time algorithm. The complexity class #P is similar to the class NP, except that #P problems are concerned with counting the number of solutions, while NP problems are only required to check whether a solution exists. As noted above, calculating the permanent of a $\{0,1\}$-matrix M is equivalent to calculating the number of perfect matchings in a bipartite graph with adjacency matrix M. Interestingly enough, checking the existence of a matching in any graph has an $O(n^2)$ algorithm [9], and therefore this related question is in the class P.

Valiant's discovery redirected research to approximation algorithms for the permanent. The goal was now a *fully polynomial randomized approximation scheme* (FPRAS) which would be polynomial in both the input size and the magnitude of error. In 1986, Broder [7] proposed using a Markov chain with an enlarged state space: this random walk used both perfect matchings and "near-perfect" matchings – that is, matchings missing precisely one edge. This led to a paper of Jerrum and Sinclair [28] showing that the aforementioned random walk was a FPRAS as long as the ratio between perfect and near-perfect matchings is bounded, and culminated in a seminal paper by Jerrum, Sinclair and Vigoda [29] demonstrating a FPRAS for the permanent of any matrix with non-negative entries. (As noted in [29], an algorithm for matrices with negative entries could be used to calculate the permanent precisely and hence is assumed to be intractable.) This approximation scheme generates clever weights for the random walk, enabling perfect matchings not to get overwhelmed by near-perfect matchings even when the ratio between them isn't bounded. Currently, the best results for this FPRAS can be found in a paper by Bezáková, Štefankovič, Vazirani, and Vigoda [5], which achieves an $O(n^7 \log n^4)$ bound.

This thesis takes a different approach to the problem. By focusing attention on permutations with interval restrictions, a more tractable subclass of permutations with restricted positions, it was hoped that better bounds could be achieved using the simpler random transposition walk. A further incentive to study interval restrictions came from statistics, as the truncated paired data often comes with with natural interval structure (this will be further discussed in the introduction to Chapter 3.) Diaconis, Graham and Holmes conjectured in [11] that in general, the random transposition walk on S_M mixes in $n^2 \log n$ time. In Section 3.5, the conjecture is resolved in the negative; indeed, an example is provided (courtesy of my colleague John Jiang) that mixes in exponential time. However, the case of *bounded* interval restriction matrices proves to be more tractable. In this case, the random transposition walk mixes in $O(n^4)$ time, which is a substantial improvement on the best bounds for the general permanent algorithm discussed above.

This thesis also gets more precise answers for subclasses of interval restriction matrices (these will be defined precisely in Section 1.2 below.) The mixing time for the sped-up random transposition walk on sets called Fibonacci permutations in [11] is derived to within a constant. Furthermore, a phenomenon called cut-off is observed for a subclass of one-sided interval restriction matrices. In the process of analysis, methodological contributions are made: depending on the specific set up, different tools are required to obtain optimal bounds. Adapting the techniques elucidates the strengths and weaknesses of each one. This will be discussed in the summary of the results below.

1.2 Results

This thesis is split into four chapters after this introductory chapter. Chapter 2 gives a short overview of the techniques used. Chapter 3 presents the results for matrices with bounded interval restrictions. Chapter 4 deals with matrices with one-sided interval restrictions. Finally, Chapter 5 presents a coupling argument for the random transposition walk on S_n. The results of each chapter are now described in more detail.

1.2.1 Bounded Interval Restrictions

Recall that if the restriction matrix is M, then $S(i)$ is defined to be $\{j \mid M(i,j) = 1\}$, which is the set of allowed values for $\sigma(i)$ for $\sigma \in S_M$. Chapter 3 studies the case where each interval $S(i)$ is of size at most k for a fixed k. Such matrices are called "bounded interval restriction matrices." The following theorem is proved (Theorem 3.2):

Theorem 1.1. *If M is an $n \times n$ restriction matrix such that $|S(i)| \le k$ for each i, then*

$$\tau_{\text{mix}} \le C(k)n^4$$

for a constant $C(k)$ depending only on k.

This result will be proven by using the path method of Diaconis and Stroock [15] to bound the spectral gap, and then using that to bound the mixing time. Like the other techniques mentioned below, this will be described in more detail in Chapter 2. In order to use the method, a clever bijective trick (used earlier by Jerrum and Sinclair in [28] and rediscovered here) is needed. This is the most general approach utilized in this thesis: it can be applied even when there's limited symmetry. However, it rarely gives the right answer.

In special cases, the results of Theorem 1.1 can be strengthened: consider the problem of the random transposition walk on the Fibonacci permutations. To define Fibonacci permutations, let M_n be the $n \times n$ tridiagonal matrix; that is,

$$M_n(i,j) = \begin{cases} 1 & |i-j| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

In Section 2.4 of the paper "Statistical problems involving permutations with restricted positions" by Diaconis, Graham, and Holmes [11], it was shown that $|S_{M_n}|$ is the $(n+1)$st Fibonacci number, explaining the name for this set. For this set, it turns out that the only necessary transpositions are the adjacent transpositions; hence, only using those speeds things up. The following theorem is proved (Theorem 3.15):

Theorem 1.2. *The mixing time for the random adjacent transposition walk on the set of Fibonacci permutations is of order $n \log n$.*

The upper bound is shown using a tricky coupling argument, and the lower bound follows from the paper "A general lower bound for mixing of single-site dynamics on graphs" by Hayes and Sinclair [23]. Unlike the path method, coupling arguments are difficult to implement without a substantial amount of structure. However, they are considerably more successful at getting the right answer for special cases.

Finally, a counterexample to the conjecture of Diaconis, Graham and Holmes which states that the random transposition walk on S_M mixes in order $n^2 \log n$ time is given. Unfortunately, not only is this not true, but there isn't even a general polynomial bound for this walk. A family of $(2n-1) \times (2n-1)$ matrices A_n is provided for which the random transposition walk needs exponential time to mix. This counterexample comes from my colleague John Jiang and uses conductance for the lower bound.

1.2.2 One-Sided Interval Restrictions

Chapter 4 studies the case of permutations with one-sided interval restrictions: that is, $n \times n$ restriction matrices M such that $S(i) = [a_i, n]$ for each i. This chapter proves very sharp results on the mixing time of the chain under further restrictions to 'two-step' one-sided restriction matrices, where each $S(i)$ is either equal to $[1,n]$ or $[a,n]$ for some fixed a. This makes heavy results of Hanlon, who diagonalized this walk in the wonderful paper "A random walk on the rook placements on a Ferrers board" [22]. As will be seen, a lot of further work is needed to make use of the diagonalization.

Two definitions are needed to state the results. First is the definition of a phenomenon called *cut-off*, which occurs when a chain goes from being very far from mixed to being thoroughly mixed in a window of order smaller than τ_{mix}. To be more precise, say that a family of chains $(X_t^n)_{t\geq 0}$ with mixing times $\tau_{\text{mix}}^{(n)}(\epsilon)$ experiences a cut-off if for all $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{\tau_{\text{mix}}^{(n)}(\epsilon)}{\tau_{\text{mix}}^{(n)}(1-\epsilon)} = 1 \tag{1.5}$$

This section will be largely using the tools of Fourier analysis, so an L^2 distance instead of an L^1 distance is needed. Accordingly, define the *chi-squared distance* between distributions μ and π with respect to π to be

$$\|\mu - \pi\|_{2,\pi} = \sqrt{\sum_x \left(\frac{\mu(x)}{\pi(x)} - 1\right)^2 \pi(x)} \tag{1.6}$$

Further background is in Chapter 2. For now, it suffices to define it, and to state that it is an upper bound on two times the total varation distance. Chi-squared mixing time and chi-squared cut-off are defined analogously to Equations (1.4) and (1.5).

One of the first times this distance was used to obtain sharp bounds on mixing times was by Diaconis and Shahshahani, in their paper “Generating a random permutations with random transpositions" [14]. Theorems 1.3-1.5 make much use of the techniques of this proof, especially the more streamlined version presented by Diaconis in the book “Group representations in probability and statistics" [10]. The following is the first theorem of this chapter (Theorem 4.9.)

Theorem 1.3. *Under certain conditions, the random transposition walk on S_M, where M is a two-step interval restriction matrix, undergoes a chi-squared cutoff.*

Theorem 1.3 and some combinatorial tools combine to show the following theorem (Theorem 4.10):

Theorem 1.4. *Under certain conditions that are stronger than the ones in Theorem 1.3, the random transposition walk on S_M, where M is a two-step restriction matrix, undergoes a total variation cutoff.*

In addition, the following theorem (Theorem 4.11) shows that the above result does not hold for all the two-step restriction matrices in Theorem 1.3.

Theorem 1.5. *There exists a class of two-step interval restriction matrices M, such that the total variation mixing time of the random transposition walk on S_M is of lower order than the chi-squared mixing time. Furthermore, the walk experiences a chi-squared cutoff but does not experience a total variation cutoff.*

This chapter serves as a reflection on Fourier analysis as a method for bounding total variation mixing times. First of all, it presents a large class of examples for which diagonalizing the random walk provides very sharp bounds. As such, it generalizes the results of Diaconis and Shahshahani in [14]. However, it also demonstrates the difficulties and limitations of the approach. The combinatorial calculations needed to use the eigenvalues require considerable ingenuity, even in the (inherently simpler) cases where the walk is vertex transitive. Furthermore, Theorem 1.5 presents a large class of examples for which the spectral analysis does not give the right answer for the total variation mixing, showing a clear limitation of the method.

1.2.3 Coupling for the Random Transposition Walk

Chapter 5 studies the random transposition walk on the entire symmetric group: that is, the case where the matrix M is all 1s. It was shown by Diaconis and Shahshahani [14] that this walk experiences a cut-off, as defined above in Equation (1.5), around time $\frac{1}{2}n \log n$. There is also an argument by Matthews in [39] proving the same result using a purely probablistic technique called a strong stationary time. However, there did not previously exist a coupling argument that showed mixing after time of $O(n \log n)$; furthermore, it can be shown that the best possible bound resulting from a direct Markovian coupling argument is of order n^2. The main result in this chapter is the following theorem (Theorem 5.1):

Theorem 1.6. *There exists a coupling argument which shows that the random transposition walk on S_n mixes in $O(n \log n)$ time.*

This theorem is proven by first projecting the walk to the conjugacy classes of S_n, which are well known to be indexed by partitions of n [26]. This projection is also a Markov chain, called a split-merge random walk [47]. It is easy to show that bounding the mixing time of the walk on conjugacy classes suffices to bound the original walk on S_n; accordingly, a coupling argument for the walk on conjugacy classes is presented.

In order to define this coupling, the path coupling method introduced by Bubley and Dyer in [8] is used; that is, a coupling is defined only for pairs of partitions (σ, τ) that are one step away in the split-merge random walk. The trick is to define the coupling so that at each step, the coupled pair is no more than one step away from each other, and then to find a lower bound on the probability of colliding at each step. For a straightforward $O(n \log n)$ bound, a probability of at least $\frac{1}{n}$ of coupling at each step is required. This is not that case; however, it is shown that this is true on average, using the techniques developed by Schramm in his paper "Compositions of random transpositions" [47]. This chapter again demonstrates the utility of coupling for obtaining good answers for problems with sufficient structure, as well as the convenience of the path coupling method. Furthermore, by adapting the techniques of Lemma 2.3 from [47], it shows the applicability of the ideas therein. In particular, the clever probabilistic bounds used to control the growth of a key quantity in this lemma were successfully adapted in Lemmas 5.35 and 5.23 below.

Chapter 2

Background and Techniques

2.1 Introduction

This chapter defines a number of important concepts in the field of Markov chain mixing times, and introduces the techniques that will be used in the rest of thesis. An excellent reference for much of this chapter is the book “Markov chains and Mixing Times" by Levin, Peres, and Wilmer [35].

For the rest of this document, let $(X_t)_{t\geq 0}$ denote a time-homogeneous Markov chain on a finite state space Ω, with transition matrix P: that is,

$$\mathbb{P}\{X_{t+1} = y \mid X_t = x, X_{t-1} = z_{t-1}, \ldots, X_0 = z_0\} = \mathbb{P}\{X_{t+1} = y \mid X_t = x\} = P(x, y)$$

for all $x, y \in \Omega$ and any choice of $z_0, \ldots, z_{t-1} \in \Omega$. There are a number of properties that enter the definitions needed to talk about the convergence of Markov chains to their stationary distributions.

Definition 2.1. *A Markov chain is said to be irreducible if for every $x, y \in \Omega$, there exists an integer t such that $P^t(x, y) > 0$.*

Definition 2.2. *The period of a state x in a Markov chain is defined to be*

$$d_x = \gcd\{i \mid P^i(x, x) > 0\}$$

We say that a state is aperiodic if $d_x = 1$, and that a Markov chain is aperiodic if $d_x = 1$ for all states $x \in \Omega$.

As is well known, an aperiodic irreducible Markov chain on a finite state spaces converges to its stationary distribution, where a distribution π on Ω is called *stationary* if

$$\pi(y) = \sum_{x \in \Omega} \pi(x) P(x, y) \tag{2.1}$$

for all y. The following theorem can be found in Chapter 15 of Feller [18].

Theorem 2.3. *An irreducible, aperiodic Markov chain on a finite state space Ω has a stationary distribution π to which it converges as $t \to \infty$. To be more precise, if $d(t)$ is defined as in Equation* (1.3)*, then $d(t) \to 0$ as $t \to \infty$.*

Recall that the mixing time of a Markov chain is

$$\tau_{\text{mix}}(\epsilon) = \min\{t \mid d(t) \leq \epsilon\}$$

where τ_{mix} is conventionally defined to be $\tau_{\text{mix}}(1/4)$ and $d(t)$ is defined as in Equation (1.3). Given a family of Markov chains $(X_t^n)_{t \geq 0}$ with mixing times $\tau_{\text{mix}}^{(n)}(\epsilon)$, we will be interested in the rate of growth of $\tau_{\text{mix}}^{(n)}$.

There are many techniques which can be used to bound this rate of growth, many of which are outlined in [35]. This chapter focuses on the ones which will be directly used in this document. One of the methods described is a probabilistic technique called coupling, as well as a useful variant called path coupling. Also included are a number of spectral techniques, which use the eigenvalues of the matrix P to find bounds on the mixing time of the walk. A final subsection shows how to find lower bounds for mixing times using the bottleneck ratio.

2.2 Coupling

Coupling is a purely probabilistic technique usually traced back to Doeblin [16]. Two good reference books which illustrate its many uses are Lindvall's “Lectures on the coupling method" [36] and Thorisson's “Coupling, stationarity, and regeneration" [50]. The concept of coupling applies both to pairs of distributions and to pairs of Markov chains. In either case, the idea is to define objects on the same probability space, and to use this to study their properties. The following definitions are needed.

Definition 2.4. *A coupling of a pair of probability measures μ and ν is a pair of random variables (X, Y) defined on the same probability space, such that the marginal distriubtion of X is μ and the marginal distribution of Y is ν.*

The definition of a coupling of Markov chains is entirely analogous, except with processes instead of random variables.

Definition 2.5. *A coupling of a pair of Markov chains both with transition matrix P is a process $(X_t, Y_t)_{t\geq 0}$ such that both $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ are Markov chains with transition matrix P, but which possibly have different starting distributions.*

Remark 2.6. If the joint process $(X_t, Y_t)_{t\geq 0}$ is itself a Markov chain, then the coupling is called *Markovian*. In practice, most couplings are Markovian; a good example of a non-Markovian coupling is the paper “A non-Markovian coupling for sampling colorings" by Hayes and Vigoda [25].

Coupling can be used to bound Markov chain mixing times. The idea is the following: define a coupling $(X_t, Y_t)_{t\geq 0}$ of two Markov chains, both with transition matrix P, one starting at x and the other starting at y. The goal is to have the two chains meet. By the time that this has happened with high probability for every choice of $x, y \in \Omega$, the coupling inequality shows that the chain has mixed. One of the first times this technique was used to bound mixing times was by Aldous to upper bound the mixing time for the random walk on the n-dimensional hypercube [1]. The following is a precise version – it's a restatement of Theorem 5.2 from [35].

Theorem 2.7. *Let (X_t, Y_t) be a coupling of a pair of Markov chains, both with transition matrix P, such that $X_0 = x$ and $Y_0 = y$, and let T be a random time at which the chains have met – that is, $X_T = Y_T$. Then, the following inequality holds:*

$$\left\|P^t(x, \cdot) - P^t(y, \cdot)\right\|_{TV} \leq \mathbb{P}\{T > t\}$$

Note: *A stopping time T such that $X_T = Y_T$ is called a coupling time.*

Remark 2.8. It turns out that the above theorem is sharp: there always exists a coupling that achieves equality. This was first shown by Griffeath in [21], and Pitman presented a simpler proof in [44]. However, the proofs are non-constructive, and maximal couplings are usually non-Markovian; hence this is not often useful in practice.

Now, using Theorem 2.7 and the following well-known inequality (Lemma 4.11 in [35]),

$$\max_x \left\| P^t(x,\cdot) - \pi \right\| \leq \max_{x,y} \left\| P^t(x,\cdot) - P^t(y,\cdot) \right\|_{TV}$$

easily gives a way to go from a coupling to a bound on mixing times.

Corollary 2.9. *Suppose that for all pairs of states $x, y \in \Omega$, there exists a coupling (X_t, Y_t) such that $(X_0, Y_0) = (x, y)$, with a coupling time $T_{x,y}$. Then, the following inequality holds:*

$$d(t) \leq \max_{x,y} \mathbb{P}\{T_{x,y} > t\}$$

2.2.1 Path Coupling

The coupling inequality in Corollary 2.9 requires a coupling (X_t, Y_t) starting at any pair (x, y). It is not always clear how to do this, making this a difficult task (albeit one that has been carried out in a large number of examples.) An approach called *path coupling*, introduced by Bubley and Dyer in [8], can make defining couplings considerably easier.

The idea of path coupling is the following: instead of having to define a coupling starting at each pair (x, y), it is only necessary to define couplings for certain pairs. To be more precise, if $(X_t)_{t\geq 0}$ is a Markov chain on Ω, endow Ω with a connected graph structure: that is, select a set of edges E' between elements of Ω, such that for any $u, v \in \Omega$, there exists a path between u and v only using the edges in E'. It is now only necessary to define a coupling for $(x, y) \in E'$.

Remark 2.10. Note that E' can be entirely unrelated to the Markov chain (X_t), which allows considerable latitude when using path coupling.

Continuing the definitions, assign lengths $l(x, y) \geq 1$ to each edge $(x, y) \in E'$, and define a path metric ρ on Ω by

$$\begin{aligned}\rho(z, w) &= \text{Minimal length of a path between } z \text{ and } w \\ &= \min \left\{ \sum_{i=0}^{n-1} l(x_i, x_{i+1}) \mid x_0 = z, x_n = w, (x_i, x_{i+1}) \in E' \text{ for all } i \right\} \end{aligned} \tag{2.2}$$

Furthermore, define the diameter of the set Ω as usual as

$$\text{diam}(\Omega) = \max_{u,v\in\Omega} \rho(u, v)$$

The following theorem is the basic path coupling bound.

Theorem 2.11. *Let $(X_t)_{t\geq 0}$ be a Markov chain on a set Ω, and let E', l and ρ be defined as above. Let (X_1, Y_1) be the first step of a coupling started at $(x, y) \in E'$. Then, if there is a $\kappa < 1$ such that for every $(x, y) \in E'$,*

$$\mathbb{E}\left[\rho(X_1, Y_1)\right] \leq \kappa \rho(x, y) \tag{2.3}$$

then for all $t \geq 1$,

$$d(t) \leq \mathrm{diam}(S)\kappa^t$$

Path coupling is a powerful technique that has been used to tackle a number of problems. In particular, it has been very useful for studying the mixing time of the graph coloring Markov chain. One of the first places this was done was by Jerrum in [27]; another good reference is Vigoda's paper "Improved bounds for sampling colorings" [52]. An interesting refinement of the method is the idea of variable-length coupling, in which contraction is allowed to occur after a random number of steps (instead of just one step, as in standard path coupling.) This was developed by Hayes and Vigoda in [24].

Another recent development has utilized the idea of curvature, which requires contraction of the transportation metric (also known as the Wasserstein distance.) In the case of finite state spaces, this reduces to Theorem 2.11 above; when X is a manifold, it is equivalent to having positive (Ricci) curvature in the ordinary geometric sense. A good reference is "Curvatures, Concentration and Error estimates for Markov Chain Monte Carlo" by Joulin and Ollivier [32]. An earlier paper with a number of instructive examples is "Ricci curvature of Markov chains on metric spaces" by Ollivier [42].

2.3 Spectral Methods

Turn next to a more analytical and less probabilistic technique. It turns out that in some cases, it is possible to use the eigenvalues of the transition matrix P to bound the mixing time of a Markov chain. Before describing the history of this approach, here is a definition.

Definition 2.12. *A Markov chain on Ω with a transition matrix P and a stationary distribution π is reversible if for every $x, y \in \Omega$,*

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

Furthermore, if a probability distribution π satisfies the above equations, it is a stationary distribution of our Markov chain.

Remark 2.13. Many natural Markov chains are reversible; in particular, it is easy to check that Markov chains with symmetric transition matrices are reversible with respect to the uniform distribution.

Spectral techniques are most often used for reversible Markov chains and they exploit the fact that such chains are orthogonally diagonalizable. This method of analyzing Markov chains is classical – a good reference is Chapter 16 of Feller [18]. One of the first uses of spectral methods in the context of analyzing families of chains is the paper of Diaconis and Shahshahani examining the random transpositon walk on S_n [14].

Definition 2.14. *Let P be a transition matrix of a reversible Markov chain on a set Ω with stationary distribution π. Define the inner product $\langle \cdot, \cdot \rangle_\pi$ by*

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

Furthermore, define the L^2 norm as

$$\|f\|_2 = \sqrt{\langle f, f \rangle_\pi}$$

in the usual way.

The fact that P is orthogonally diagonalizable under $\langle \cdot, \cdot \rangle_\pi$ means that P has an amenable spectral profile. The following is a consequence of Lemma 12.1 in “Markov chains and mixing times" [35].

Lemma 2.15. *If P is the transition matrix of a reversible Markov chain then the following hold:*

1. *All of the eigenvalues of P are real and in the interval $[-1, 1]$.*
2. *If the Markov chain is irreducible, then P has exactly one copy of the eigenvalue 1.*
3. *If the Markov chain is aperiodic, then -1 is not an eigenvalue of P.*

For the remainder of this section, assume a reversible, irreducible Markov chain with the following eigenvalues:

$$1 = \beta_0 > \beta_1 \geq \cdots \geq \beta_{|\Omega|-1} \geq -1 \tag{2.4}$$

and let $\beta_{\min} = \beta_{|\Omega|-1}$.

2.3.1 The Spectral Gap and the Path Method

Often, the full spectral profile of the transition matrix P is not accessible (although when it is, it can be an extremely powerful tool.) However, there are theorems that only require a limited amount of spectral information. If P has eigenvalues as in Equation (2.4), define

$$\beta_* = \max(\beta_1, |\beta_{\min}|) \tag{2.5}$$

Thus, β_* is the eigenvalue with the highest absolute value. The quantity $1 - \beta_*$ is referred to as the *spectral gap*, and it can be used to analyze the mixing time of the Markov chain. The following theorem (Theorem 12.3 in [35]) is often used for this purpose.

Theorem 2.16. *Let* $\pi_{\min} = \min_x \pi(x)$. *Then, using the definitions above, the following inequality holds:*

$$\tau_{\text{mix}}(\epsilon) \leq \log\left(\frac{1}{\epsilon\pi_{\min}}\right)\frac{1}{1-\beta_*}$$

The above theorem raises an obvious question: how do we find bounds on the spectral gap of a transition matrix? A powerful approach to this problem is the *path method*, which was developed by Diaconis and Stroock in the paper "Geometric bounds for eigenvalues of Markov chains" [15], and which uses the geometric structure of the underlying Markov chain to find bounds on β_*. Before proceeding, here is a useful definition.

Definition 2.17. *If* $(X_t)_{t\geq 0}$ *is a Markov chain with a transition matrix* P, *define the edge set* E *induced by* $(X_t)_{t\geq 0}$ *to be all pairs* $\{x, y\}$ *in* Ω *such that* $P(x, y) > 0$ *or* $P(y, x) > 0$.

The path method requires us to define a family of paths $\{\gamma_{xy}\}$ using the edges in E between every pair of elements $x, y \in \Omega$. In order to get a good bound, we need to have control on both the lengths of the paths and the number of paths going through each edge. After the paths are defined, an upper bound on β_* is derived using Cauchy-Schwarz. Since there are choices about how to apply the inequality, a number of similar theorems can be derived. Various bounds have been proved in the aforementioned paper by Diaconis and

Stroock [15] and by Sinclair in [48]. For our purposes, the specific version of the theorem does not matter. The following is Proposition 1 in Diaconis and Stroock.

Theorem 2.18. *Let $\{\gamma_{xy}\}$ be a family of paths using the edges of E between elements of Ω, such that γ_{xy} starts at x and ends at y. Define $Q(z,w) = \pi(z)P(z,w)$, and let*

$$|\gamma_{xy}|_Q = \sum_{e \in \gamma_{xy}} Q(e)^{-1} \tag{2.6}$$

where $e \in \gamma_{xy}$ means that edge e is used by the path γ_{xy}. Now, let

$$K = \max_{e \in E} \sum_{x,y \text{ s.t. } e \in \gamma_{xy}} |\gamma_{xy}|_Q \pi(x)\pi(y)$$

Then we have that

$$\beta_1 \leq 1 - \frac{1}{K}$$

There is a similar theorem to bound the smallest eigenvalue of P. This is Proposition 2 from Diaconis and Stroock. Here, instead of making paths between distinct pairs of elements x and y, a path of odd length from x to x is needed for each $x \in \Omega$.

Theorem 2.19. *Let $\{\gamma_x\}$ be a family of paths using the edges of E such that γ_x both starts and ends at x and has an odd number of edges. Define $|\gamma_x|_Q$ like we did in Equation (2.6). Then, if*

$$L = \max_{e} \sum_{x \ s.t. \ e \in \gamma_x} |\gamma_x|_Q \, \pi(x),$$

then

$$\beta_{\min} \geq -1 + \frac{2}{L}$$

2.3.2 Fourier Analysis

As noted above, Theorem 2.16 only requires bounds on β_*. However, good bounds on all the eigenvalues and eigenvectors of P give considerably more information about the mixing time. This technique was pioneered by Diaconis and Shahshahani in their paper "Generating a random permutation with random transpositions" [14], in which they used Fourier analysis to find tight bounds on the mixing time of the random transposition walk on S_n. Since this method requires a lot of information, it has primarily been used for random walks on groups, although the techniques apply in general. A good reference is the book "Group representations in probability and stastistics" by Diaconis [10].

The effectiveness of Fourier analysis stems from the tractability of $L^2(\pi)$. This uses the chi-squared distance, as defined above in Equation (1.6). This distance is closely related to the inner product $\langle \cdot, \cdot \rangle_\pi$.

Definition 2.20. *For a distribution μ on S, define the chi-squared distance between μ and π as*

$$\|\mu - \pi\|_{2,\pi} = \left\| \frac{\mu}{\pi} - 1 \right\|_2 = \sqrt{\sum_x \left(\frac{\mu(x)}{\pi(x)} - 1 \right)^2 \pi(x)}$$

where $\|\cdot\|_2$ is the standard norm on $L^2(\pi)$, as defined as above in Definition 2.14.

Note that it is easy to see that $\|\mu - \pi\|_{TV}$ is precisely half the $L^1(\pi)$ distance between $\frac{\mu}{\pi}$ and 1. This means that chi-squared distance is an upper bound for twice the total variation distance. Furthermore, chi-squared distance turns out to precisely computable if all the eigenvalues and eigenvectors of the matrix P are known. The following two results are standard, and follow from Lemma 12.16 in [35].

Theorem 2.21. *Let P be the transition matrix of an irreducible reversible Markov chain on Ω with eigenvalues $1 = \beta_0 > \beta_1 \geq \cdots \geq \beta_{|\Omega|-1} \geq -1$, and with a corresponding basis of orthonormal eigenvectors $\{v_j\}$. Then,*

$$\left\|P^t(x,\cdot) - \pi\right\|_{TV} \leq \frac{1}{2} \left\|P^t(x,\cdot) - \pi\right\|_{2,\pi} = \frac{1}{2} \sqrt{\sum_{i=1}^{|\Omega|-1} v_i(x)^2 \beta_i^{2t}}$$

The above theorem requires all the information about the eigenvalues and eigenvectors to compute the chi-squared distance. However, given a sufficient degree of symmetry, eigenvalues suffice. Say that a Markov chain on a state space Ω is *vertex transitive* if for every $x, y \in \Omega$, there exists a bijec tion $f : \Omega \to \Omega$ such that $f(x) = y$, and f preserves the random walk. In this case, the following easy corollary can be proved.

Corollary 2.22. *If the assumptions from Theorem 2.21 hold, and the Markov chain is also vertex transitive, then*

$$\left\|P^t(x,\cdot) - \pi\right\|_{TV} \leq \frac{1}{2} \left\|P^t(x,\cdot) - \pi\right\|_{2,\pi} = \frac{1}{2} \sqrt{\sum_{i=1}^{|\Omega|-1} \beta_i^{2t}}$$

Note that in both Theorem 2.21 and Corollary 2.22, there is an equality between the chi-squared distance and the expression on the right hand side. This fact will be useful in Chapter 4.

2.4 Lower Bounds Using the Bottleneck Ratio

The bottleneck ratio has a number of different names in the literature, including the Cheeger constant and conductance. The relationship between it and the spectral gap was first noted by Alon and Milman [2] in 1985. The lower bound for nonreversible chains was discovered later by Mihail [41] in 1989. For the sake of completeness, this section presents the results for both cases. First define

$$Q(x,y) = \pi(x)P(x,y), \quad Q(A,B) = \sum_{x\in A, y\in B} Q(x,y)$$

Hence, $Q(A,B)$ is the probability of moving from A to B in one step, starting from the stationary distribution. Then, define the *bottleneck ratio* of the Markov chain to be

$$\Phi_* = \min_{S:\pi(S)\leq\frac{1}{2}} \frac{Q(S,S^c)}{\pi(S)} \tag{2.7}$$

Then, the following is Theorem 7.3 from “Markov chains and mixing times" [35]:

Theorem 2.23. *For any Markov chain $(X_t)_{t\geq 0}$ and Φ_* as defined above,*

$$\tau_{\text{mix}} = \tau_{\text{mix}}(1/4) \geq \frac{1}{4\Phi_*}$$

It is possible to obtain a slight improvement on this result in the reversible case. First of all, if β_* is defined as above in Equation (2.5), then it is follows from Theorem 13.14 in [35] that $1-\beta_* \leq 2\Phi_*$. Combining this fact with a well-known theorem (Theorem 12.4 in [35], originally from [28] and [34]) gives a slightly better lower bound than Theorem 2.23 above:

Theorem 2.24. *For a reversible, irreducible, and aperiodic Markov chain*

$$\tau_{\text{mix}}(\epsilon) \geq \left(\frac{1}{1-\beta_*} - 1\right)\log\left(\frac{1}{2\epsilon}\right) \geq \left(\frac{1}{2\Phi_*} - 1\right)\log\left(\frac{1}{2\epsilon}\right)$$

Chapter 3

Bounded Interval Restrictions

3.1 Introduction

As discussed in Chapter 1, one motivation for studying permutations with restricted positions is statistical testing for independence. Recall that in this set-up there is a paired data set $(X_1, Y_1), \ldots, (X_n, Y_n)$ with truncations: for each x there is a set $S(x)$ such that the pair (X, Y) is observable if and only if $Y \in S(X)$. In real world applications, the sets $S(x)$ often have a lot of structure; in particular, a common type of truncation is one where for each x, $S(x)$ is an interval. As noted above, independence testing in this case requires sampling from $\sigma \in S_M$, where M is defined by

$$M(i,j) = \begin{cases} 1 & \text{if } \ Y_i \in S(X_j) \\ 0 & \text{otherwise} \end{cases}$$

It is easy to see that if all the $S(x)$ are intervals, then M is an interval restriction matrix. This motivates the current discussion.

Diaconis, Graham and Holmes provide a fascinating real-world example of such a truncation in [11], which originally came from a paper of Efron and Petrosian [17]. This example concerned red-shift data: the observation X_i was the red-shift of a quasar, while the observation Y_i was the log-luminosity. The data included $n = 210$ quasars, and the goal was to check a hypothesis of luminosity evolution: were the quasars brighter in the early universe and then evolve to be dimmer? Testing this hypothesis required removing the spurious correlations resulting from truncation; to this end, an efficient sampling algorithm from S_M for interval restriction matrices M would be highly desirable. Further examples of real-world doubly truncated data can for instance be seen in [40]. This chapter examines the random transposition algorithm introduced in Chapter 1.

3.2 Preliminaries

Recall from Equation (1.1) that for a $n \times n$ $\{0,1\}$-matrix M,

$$S_M = \{\sigma \mid M(i, \sigma(i)) = 1 \text{ for all } i\}$$

and for $i \leq n$, $S(i) = \{j \mid M(i,j) = 1\}$. Clearly, σ is in S_M if $\sigma(i)$ is in $S(i)$ for each i.

This section explains how to visualize S_M, as well as the random transposition walk on this set. It also proves that the order of rows of M doesn't matter. Recall that attention is restricted to interval restriction matrices M – that is, matrices for which each row $S(i)$ is an interval for each i. Here's an example of such a matrix:

$$M = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

Note that $\alpha \in S_M$ precisely if $M(i, \alpha(i)) = 1$ for all i. This means that α can be represented by marking the 1s in positions $(i, \alpha(i))$. For example, for the matrix M above, $\alpha = 1432$ is represented by boxing 1s in the following way:

$$\begin{bmatrix} \boxed{1} & 1 & 1 & 0 \\ 0 & 1 & 1 & \boxed{1} \\ 1 & 1 & \boxed{1} & 0 \\ 0 & \boxed{1} & 1 & 0 \end{bmatrix}$$

Thus, the elements of S_M are in bijection with the ways of marking 1s in M, such that there is exactly one 1 in each row and column; clearly, this characterization is equivalent to representing S_M as the set of placements of n rooks on the board induced by M. The random transposition walk can be visualized as follows: to attempt to transpose the elements labelled i and j, consider the marked 1s in columns i and j and construct a rectangle with those 1s at its corners. If the two unmarked corners of the rectangle are both 1s, unmark the currently marked corners and mark the other two; otherwise stay in place. For example, transposing the 2 and 3 in the above α looks like this:

$$\begin{bmatrix} \boxed{1} & 1 & 1 & 0 \\ 0 & 1 & 1 & \boxed{1} \\ 1 & \widehat{1} & \boxed{1} & 0 \\ 0 & \boxed{1} & \widehat{1} & 0 \end{bmatrix} \Longrightarrow \begin{bmatrix} \boxed{1} & 1 & 1 & 0 \\ 0 & 1 & 1 & \boxed{1} \\ 1 & \boxed{1} & 1 & 0 \\ 0 & 1 & \boxed{1} & 0 \end{bmatrix}$$

Similarly, attempting to transpose the elements in positions i and j involves an identical procedure with the 1s in rows i and j. This visualization will be useful for proving Theorem 3.2 below, which states that the random transposition walk on bounded interval restriction matrices mixes in $O(n^4)$ time. In the language of the above description, transposing the elements labeled i and j amounts to 'transposing columns i and j,' and transposing the elements in positions i and j amounts to 'transposing the rows i and j.'

The following result both helps with the intuition about the set S_M, and is useful for some of the later chapters: it shows that the rows of M can be permuted without changing the structure of the random transposition walk on S_M. Note that this result doesn't require M to be an interval restriction matrix (although the walk may not be connected for other matrices.)

Lemma 3.1. *Let M be a restriction matrix, and let N be M with its rows permuted. Then there is an isomorphism between the random transposition walk on N and the random transposition walk on M.*

Proof: Assume that N is M with its rows permuted by a permutation α (so row i of M become row $\alpha(i)$ of N.) Then, define a map $\psi : S_N \to S_M$ by

$$\psi(\sigma) = \sigma\alpha$$

First, show that ψ is a bijection from S_N to S_M. Since N is M with its rows permuted by

α, $M(i,j) = N(\alpha(i), j)$. Therefore,

$$\begin{aligned} S_N = \{\sigma \mid N(i,\sigma(i)) = 1 \text{ for all } i\} &= \{\sigma \mid N(\alpha(i),\sigma(\alpha(i))) = 1 \text{ for all } i\} \\ &= \{\sigma \mid M(i,\sigma\alpha(i)) = 1 \text{ for all } i\} \\ &= \{\sigma \mid \sigma\alpha \in S_M\} \end{aligned}$$

This suffices to show that ψ is a bijection. Therefore, it just remains to prove that ψ preserves the random walk. By the definition, it must be demonstrated that $\sigma\tau^{-1}$ is a transposition if and only if $\psi(\sigma)\psi(\tau)^{-1}$ is one as well. But

$$\psi(\sigma)\psi(\tau)^{-1} = \sigma\alpha\alpha^{-1}\tau^{-1} = \sigma\tau^{-1}$$

so the above statement follows trivially. This completes the proof. $\square$

3.3 Bounded Interval Restrictions

This section proves the following theorem:

Theorem 3.2. *Fix a positive integer k, and let M be an $n \times n$ interval restriction matrix such that each interval $S(i)$ is of size at most k. Then,*

$$\tau_{\text{mix}} \leq C(k)n^4$$

for a constant $C(k)$ independent of M and n.

Note that the random walk is symmetric and therefore reversible, and hence the spectral theorems in Section 2.3 apply. Therefore, this problem can be approached using the path method for bounding the spectral gap (Theorems 2.18 and 2.19). As usual, bounding β_1 will be the far more difficult task. Some lemmas simplifying these theorems for the particular case of our random walk are useful.

Lemma 3.3. *Let $\{\gamma_{\sigma\tau}\}$ be a family of paths between every pair of elements $\sigma, \tau \in S_M$, and assume that none of the edges are loops (that is, go from α to α). Define $c(e)$ to be the number of paths that go through edge e (the congestion of e), and let $|\gamma_{\sigma\tau}|$ be the length (that* *is, the number of edges) of the path $\gamma_{\sigma\tau}$. If*

$$K_1 = \frac{n(n-1)}{2|S_M|} \max_e c(e) \max_{\sigma,\tau} |, \gamma_{\sigma\tau}|$$

then

$$\beta_1 \leq 1 - \frac{1}{K_1}$$

Proof: It clearly suffices to show that for K as defined in Theorem 2.18,

$$K \leq K_1 = \frac{n(n-1)}{2|S_M|} \max_e c(e) \max_{\sigma,\tau} |\gamma_{\sigma\tau}|$$

By definition,

$$K = \max_e \sum_{\sigma,\tau \text{ s.t. } e\in\gamma_{\sigma\tau}} |\gamma_{\sigma\tau}|_Q$$

Now, $Q(z,w) = \pi(z)P(z,w) = \frac{1}{|S_M|}\binom{n}{2}^{-1}$, since the stationary distribution is uniform, and the only non-loop edges in correspond to uniformly picked random tranpositions. Thus,

$$\begin{aligned} |\gamma_{\sigma\tau}|_Q = \sum_{e\in\gamma_{\sigma\tau}} Q(e)^{-1} &= \sum_{e\in\gamma_{\sigma\tau}} |S_M|\frac{n(n-1)}{2} \\ &= |\gamma_{\sigma\tau}||S_M|\frac{n(n-1)}{2} \end{aligned}$$

Thus,

$$\begin{aligned} K = \max_e \sum_{e\in\gamma_{\sigma\tau}} |\gamma_{\sigma\tau}|_Q \pi(\sigma)\pi(\tau) &= \max_e \sum_{e\in\gamma_{\sigma\tau}} |\gamma_{\sigma\tau}|\frac{n(n-1)}{2|S_M|} \\ &\leq \frac{n(n-1)}{2|S_M|} \max_e c(e) \max_{\sigma,\tau} |\gamma_{\sigma\tau}| \end{aligned}$$

as required. □

Theorem 2.19 now gives a lower bound on $\beta_{\min}$.

Lemma 3.4. *For sufficiently large n, the random transposition walk on S_M described above satisfies*

$$\beta_{\min} \geq -1 + \frac{2}{n^2}$$

Proof: Theorem 2.19 requires a family $\{\gamma_\sigma\}$ of paths of odd length starting and ending at σ for each $\sigma \in S_M$. Since the number of 1s in each row of M is at most k, it is easy to see that for sufficiently large n, there will be at least one transposition for each element of S_M that moves the walk out of S_M. This means that for all $\sigma \in S_M$,

$$P(\sigma, \sigma) \geq \frac{1}{\binom{n}{2}} = \frac{2}{n(n-1)}$$

Let γ_σ be the single edge (σ, σ) for each σ. Then

$$Q(\sigma, \sigma) = \pi(\sigma)P(\sigma, \sigma) \geq \frac{1}{|S_M|n^2}$$

and thus $|\gamma_\sigma|_Q \leq n^2|S_M|$. Since each edge e is used in at most 1 path, and $\pi(\sigma) = |S_M|^{-1}$,

$$L = \max_e \sum_{\sigma \text{ s.t. } e \in \gamma_\sigma} |\gamma_\sigma|_Q \, \pi(\sigma) \leq n^2$$

and this gives

$$\beta_{\min} \geq -1 + \frac{2}{L} \geq -1 + \frac{2}{n^2}$$

as required. □

The main technical result of this section will be the following proposition, which establishes the existence of a 'good' family of paths between pairs σ and τ in S_M. This is proved in Section 3.3.1 below.

Proposition 3.5. *There exists a family of paths $\{\gamma_{\sigma\tau}\}$ between pairs of elements σ, τ in S_M such that each $\gamma_{\sigma\tau}$ is at most of length kn, and such that every edge $e \in E$, as defined in Definition 2.17, satisfies*

$$c(e) \leq C_1(k)|S_M|$$

for some constant $C_1(k)$ independent of n.

The above results can be used to prove Theorem 3.2:

Proof of Theorem 3.2. From Lemma 3.4,

$$\beta_{\min} \geq -1 + \frac{2}{n^2}$$

Furthermore, using the family of paths defined above in Proposition 3.5, and combining it with the result in Lemma 3.3,

$$\beta_1 \leq 1 - \frac{1}{K_1}$$

where

$$\begin{aligned} K_1 &= \frac{n(n-1)}{2|S_M|} \max_e c(e) \max_{\sigma,\tau} |\gamma_{\sigma\tau}| \\ &\leq \frac{n(n-1)}{2|S_M|} C_1(k)|S_M|kn \\ &\leq C_2(k)n^3 \end{aligned}$$

for some constant $C_2(k)$. Thus, $\beta_1 \leq 1 - \frac{1}{C_2(k)n^3}$, so for sufficiently large n,

$$\beta_* = \max(\beta_1, |\beta_{\min}|) \leq 1 - \frac{1}{C_2(k)n^3}$$

Therefore, from Theorem 2.16,

$$\tau_{\text{mix}} \leq \log\left(\frac{4}{\pi_{\min}}\right) \frac{1}{1-\beta_*} \leq \log\left(4|S_M|\right) C_2(k)n^3$$

It remains to bound $|S_M|$. Note that to pick $\sigma \in S_M$, it is necessary to pick $\sigma(i)$ for each i; since there are at most k choices for each i, this means that $|S_M| \leq k^n$. Therefore,

$$\begin{aligned} \tau_{\text{mix}} &\leq \log(4k^n)C_2(k)n^3 \leq (n\log k + \log 4)C_2(k)n^3 \\ &\leq C(k)n^4 \end{aligned}$$

for some constant $C(k)$ independent of n, as required. □

3.3.1 Information-Preserving Paths, and Bounding β_1

This section proves Proposition 3.5: it finds a family of paths $\{\gamma_{\sigma\tau}\}$ which are of length at most kn and satisfy

$$c(e) \leq C_1(k)|S_M|$$

for each edge e, such that the constant $C_1(k)$ is independent of n and M. It turns out to be easy to find paths between σ and τ that aren't too long. On the other hand, the problem of bounding the maximal congestion at an edge turns out to be a trickier one, since there do not exist good bounds on the size of S_M; therefore, the best direct bounds for $c(e)/|S_M|$ are worse than polynomial in n. This suggests a different approach – one of creating a correspondence between edges $e \in E$ and elements $\sigma \in S_M$, and bounding the number of possible edges that can be matched to a particular element. The following lemma is used:

Lemma 3.6. *Let $\{\gamma_{\sigma\tau}\}$ be a family of paths between pairs of elements in S_M. Assume that for every pair of elements σ, τ there is a map $\phi_{\sigma\tau}$ such that*

$$\phi_{\sigma\tau} : \{e \in \gamma_{\sigma\tau}\} \to S_M$$

and such that for every $\alpha \in S_M$,

$$|\{(\sigma, \tau) \mid e \in \gamma_{\sigma\tau}, \phi_{\sigma\tau}(e) = \alpha\}| \leq C_1(k) \tag{3.1}$$

Then,

$$c(e) \leq C_1(k)|S_M|$$

for every edge e, as required.

Proof: The above maps $\phi_{\sigma\tau}$ bound $c(e)$ for an arbitrary edge e, since

$$\begin{aligned} c(e) = |\{(\sigma, \tau) \mid e \in \gamma_{\sigma\tau}\}| &= \sum_{\alpha \in S_M} |\{(\sigma, \tau) \mid e \in \gamma_{\sigma\tau}, \phi_{\sigma\tau}(e) = \alpha\}| \\ &\leq \sum_{\alpha \in S_M} C_1(k) = C_1(k)|S_M| \end{aligned}$$

as required. □

Now, of course, a family of paths $\{\gamma_{\sigma\tau}\}$ which makes it possible to define appropriate maps $\phi_{\sigma\tau}$ is needed. A good way to do this is to define a set of paths such that an edge on the path $\gamma_{\sigma\tau}$ taken together with a corresponding vertex on the backwards path $\gamma_{\tau\sigma}$ preserve almost all the information about the pair (σ, τ). Then $\phi_{\sigma\tau}(e)$ is defined to be a vertex α in $\gamma_{\tau\sigma}$, and the combination of the edge and the vertex restricts the number of possible (σ, τ) to a small number. Note that it is also possible to take two different families of paths $\{\gamma_{\sigma\tau}\}$ and $\{\tilde{\gamma}_{\sigma\tau}\}$ such that an edge on $\gamma_{\sigma\tau}$ together with a vertex on $\tilde{\gamma}_{\sigma\tau}$ preserve almost all the information about the pair (σ, τ) – there's no reason to restrict attention to the backwards path $\gamma_{\tau\sigma}$.

Call all paths for which the above construction is possible "information-preserving." I invented this approach in the process of working on this problem. It had also been used earlier by Jerrum and Sinclair in [28] to show that their approximation scheme for the permanent was polynomial for a wide class of $\{0,1\}$-matrices (including dense matrices). This method also made an appearance in the celebrated paper by Jerrum, Sinclair and Vigoda [29] which established the existence of a FPRAS for the permanent of any matrix with non-negative entries. It is a very useful technique when the actual size of the state space is difficult to approximate.

A convenient type of information-preserving path is the 'left-to-right' path. The approximate idea here is that at the ith step, $\gamma_{\sigma\tau}$ contains information about $\sigma^{-1}(j)$ for $j \leq i$, and contains information about $\tau^{-1}(j)$ for $j > i$. In this case knowing the ith step of both $\gamma_{\sigma\tau}$ and $\gamma_{\tau\sigma}$ uniquely specifies the pair (σ, τ). Alas, things are not quite this simple.

A recursive definition is used for paths: in order to create a path $\sigma = \sigma_0, \sigma_1, \ldots, \sigma_m = \tau$ between σ and τ, first define the first step σ_1 using σ and τ, and then define σ_{i+1} to be the first step in a path between σ_i and τ. The following lemma is needed:

Lemma 3.7. *Let $\sigma, \tau \in S_M$ and assume $\sigma \neq \tau$. Pick the smallest j such that $\sigma^{-1}(j) \neq \tau^{-1}(j)$ – considerating permutations as marked 1s in the matrix M, j is the smallest (that is, leftmost) column on which σ and τ disagree. Let $i = \tau^{-1}(j)$. Then, the following things hold*

$$j = \tau(i) < \sigma(i) \tag{3.2}$$

and there exists an r such that

$$\tau(i) \leq \sigma(r) < \sigma(i) \leq \tau(r) \tag{3.3}$$

Example: Below, σ is represented by the boxed 1s, and τ is represented by the 1s with hats above them.

$$\sigma \text{ shown by } \boxed{1}, \tau \text{ shown by } \widehat{1}: \quad \begin{bmatrix} \boxed{\widehat{1}} & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & \boxed{1} & 1 & \widehat{1} & 0 \\ 0 & \boxed{\widehat{1}} & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & \boxed{\widehat{1}} & 1 & 0 \\ 0 & 0 & \widehat{1} & 1 & 1 & \boxed{1} \\ 1 & 1 & 1 & 1 & \boxed{1} & \widehat{1} \end{bmatrix} \tag{3.4}$$

Here, $\sigma = 132465$ and $\tau = 152436$. Clearly, the first column that σ and τ disagree on is the third column. Thus, $j = 3$, and $i = \tau^{-1}(3) = 5$. Now,

$$3 = \tau(5) < \sigma(5) = 6$$

so Equation (3.2) holds. Next, an r must be found such that Equation (3.3) holds, which in this case means that

$$3 \leq \sigma(r) < 6 \leq \tau(r)$$

Since $\sigma(6) = 5$ and $\tau(6) = 6$, $r = 6$ works. Visually, we're looking for a row r with the following kind of arrangement:

$$\begin{array}{llllll} \text{Row } i: & \widehat{1} & \cdots & \boxed{1} & & \\ \text{Row } r: & & & \boxed{1} & \cdots & \widehat{1} \end{array}$$

allowing $\tau(i) = \sigma(r)$ and $\sigma(i) = \tau(r)$. Finding such an r gives first step on the path between σ and τ.

Proof of Lemma 3.7. Begin by showing that Equation (3.2) holds. Proceed by contradiction: assume that $\sigma(i) \leq \tau(i)$. First, note that that if $\sigma(i) = \tau(i) = j$, then $\sigma^{-1}(j) = \tau^{-1}(j) = i$, which contradicts the choice of j as the first column on which σ and τ disagree. Thus, it can be assumed that $\sigma(i) < \tau(i) = j$. But by definition, if $l < j$, $\sigma^{-1}(l) = \tau^{-1}(l)$, and thus

$$\begin{aligned} \sigma^{-1}(\sigma(i)) &= \tau^{-1}(\sigma(i)) \\ \Rightarrow \sigma(i) &= \tau(i) = j \end{aligned}$$

This contradicts the fact that $\sigma^{-1}(j) \neq \tau^{-1}(j)$. Thus, $j = \tau(i) < \sigma(i)$, as required.

Next, find a row r such that Equation (3.3) holds. Define the set A as follows:

$$A = \{s \mid 1 \leq \sigma(s) \leq \sigma(i)\} = \sigma^{-1}\left([1, \sigma(i)]\right)$$

Since σ is a permutation, $|A| = \sigma(i)$. Since τ is a bijection between $\{1, 2, \ldots, n\}$ and itself, there must exist an $r \in A$ such that $\tau(r) \geq \sigma(i)$. This r satisfies

$$\tau(i) \leq \sigma(r) < \sigma(i) \leq \tau(r)$$

as desired, via the following argument.

By choice of r, $\sigma(i) \leq \tau(r)$. Thus, it suffices to show that $\tau(i) \leq \sigma(r) < \sigma(i)$. Since $r \in A$,

$$1 \leq \sigma(r) \leq \sigma(i)$$

Furthermore, since $\tau(r) \geq \sigma(i)$, and $\tau(i) < \sigma(i)$, $r = i$ is impossible. This yields that

$$1 \leq \sigma(r) < \sigma(i) \leq \tau(r)$$

Thus, $\sigma(r) \neq \tau(r)$, and hence σ and τ disagree on the column $\sigma(r)$. Since σ and τ agree on the columns $1, 2, \ldots, j-1$, $\sigma(r) \geq j = \tau(i)$. Thus,

$$\tau(i) \leq \sigma(r) < \sigma(i) \leq \tau(r)$$

as required. □

This preparation allows the following definition (no path is needed for $\sigma = \tau$.)

Definition 3.8. *Let σ and τ be two distinct elements in S_M, and let $\sigma_0 = \sigma$. Using the same notation as in Lemma 3.7, pick the smallest j such that $\sigma^{-1}(j) \neq \tau^{-1}(j)$, and let $i = \tau^{-1}(j)$. Then, from Equation* (3.3) *in Lemma 3.7, there exists an r such that*

$$\tau(i) \leq \sigma(r) < \sigma(i) \leq \tau(r)$$

Pick the minimal such r, and define $f(\sigma, \tau)$ to be σ with the elements in positions r and i tranposed. Furthermore, define $\sigma_1, \sigma_2, \ldots$ recursively by letting

$$\sigma_r = f(\sigma_{r-1}, \tau)$$

Note: *There's no particular reason to pick the minimal r – there just needs to be an arbitrary canonical way of choosing it.*

The above definition leads to a path $\sigma = \sigma_0, \sigma_1, \ldots, \sigma_m = \tau$ in S_M between σ and τ. To be more precise, it must be shown that $f(\sigma, \tau)$ is in S_M for any pair σ and τ, and that there exists some finite m for which $\sigma_m = \tau$. The easiest way to check the latter is to define a metric on the set S_M which decreases with each step of our proposed path.

Definition 3.9. *Let σ and τ be a pair of elements of S_M. Define*

$$d_{\text{row}}(\sigma, \tau) = \sum_{s=1}^{n} |\sigma(s) - \tau(s)|$$

Visually, the above definition just sums the absolute value of the differences between σ and τ on each row. For example, if σ and τ are as in (3.4) above, then

$$d_{\text{row}}(\sigma, \tau) = 0 + 2 + 0 + 0 + 3 + 1 = 6$$

Lemma 3.10. *For $f(\sigma, \tau)$ as defined above in Definition 3.8, $f(\sigma, \tau) \in S_M$ and*

$$d_{\text{row}}(f(\sigma, \tau), \tau) < d_{\text{row}}(\sigma, \tau)$$

Thus, Definition 3.8 specifies a path $\gamma_{\sigma\tau}$ which is $\sigma = \sigma_0, \sigma_1, \ldots, \sigma_m = \tau$ between σ and τ, and furthermore, that this path satisfies

$$|\gamma_{\sigma\tau}| \leq kn$$

for every pair σ, τ.

Proof: First, $f(\sigma, \tau) = \sigma_1$ is shown to be in S_M. Let i and r be defined as in Definition 3.8, and let σ be represented by boxed 1s, and τ be represented by hatted 1s, as above. By choice of r, the following relationship between rows i and r holds:

$$\begin{array}{lcccc} \text{Row } i: & \widehat{1} & \cdots & \boxed{1} & & \\ \text{Row } r: & & & \boxed{1} & \cdots & \widehat{1} \end{array}$$

σ_1 is defined to be σ with the elements in positions i and r in transposed, so $\sigma_1(i) = \sigma(r)$ and $\sigma_1(r) = i$. The above picture should make it clear that the interval structure of the matrix M means that this transposition keeps the walk in S_M. To be more precise, by choice of r

$$\tau(i) \leq \sigma(r) < \sigma(i) \leq \tau(r)$$

By definition, $\sigma(i), \tau(i) \in S(i)$, and since $S(i)$ is an interval, the same holds for every number between $\sigma(i)$ and $\tau(i)$. Thus, $\sigma(r) \in S(i)$. Similarly, we get that $\sigma(i) \in S(r)$, and hence $\sigma_1 \in S_M$.

The next step is to show that $d_{\text{row}}(\sigma_1, \tau) < d_{\text{row}}(\sigma, \tau)$. Since σ_1 and σ agree on every element but i and r, it suffices to show that

$$|\sigma_1(i) - \tau(i)| + |\sigma_1(r) - \tau(r)| < |\sigma(i) - \tau(i)| + |\sigma_1(r) - \tau(r)| \tag{3.5}$$

But

$$\begin{aligned} |\sigma_1(i) - \tau(i)| &= |\sigma(r) - \tau(i)| = \sigma(r) - \tau(i) < \sigma(i) - \tau(i) \\ &= |\sigma(i) - \tau(i)|. \end{aligned}$$

Similarly,

$$\begin{aligned} |\sigma_1(r) - \tau(r)| &= |\sigma(i) - \tau(r)| = \tau(r) - \sigma(i) < \tau(r) - \sigma(r) \\ &= |\sigma(r) - \tau(r)|. \end{aligned}$$

Therefore the inequality (3.5) follows immediately.

The fact that $d_{\text{row}}(\sigma_1, \tau) < d_{\text{row}}(\sigma, \tau)$ shows that the sequence $\sigma_0, \sigma_1, \ldots$ will eventually lead to τ, and that this path will take no more than kn steps: clearly,

$$d_{\text{row}}(\sigma, \tau) = \sum_{r=1}^{n} |\sigma(r) - \tau(r)|$$

is an integer-valued metric. Furthermore, since $\sigma(r)$ and $\tau(r)$ are both in $S(r)$, and $S(r)$ is an interval with at most k elements, $|\sigma(r) - \tau(r)| \leq k$ for all r. Thus,

$$d_{\text{row}}(\sigma, \tau) \leq kn$$

Since

$$d_{\text{row}}(\sigma_s, \tau) < d_{\text{row}}(\sigma_{s-1}, \tau) < \cdots < d_{\text{row}}(\sigma, \tau)$$

and d_{row} is integer-valued, the path can't have more than kn steps. Hence, $|\gamma_{\sigma\tau}| \leq kn$, and the conclusion follows. $\square$

Example: Consider $\sigma = 4231, \tau = 1243$ as shown below:

$$\sigma \text{ shown by } \boxed{1}, \tau \text{ shown by } \widehat{1} : \begin{bmatrix} \widehat{1} & 1 & 1 & \boxed{1} \\ 1 & \boxed{\widehat{1}} & 1 & 0 \\ 0 & 1 & \boxed{1} & \widehat{1} \\ \boxed{1} & 1 & \widehat{1} & 1 \end{bmatrix}$$

The first column on which σ and τ don't agree is clearly 1. Thus, $j = 1$ and hence $i = \tau^{-1}(j) = 1$. Now, find a row r such that

$$\tau(i) \leq \sigma(r) < \sigma(i) \leq \tau(r)$$

which here means that $1 \leq \sigma(r) < 4 \leq \tau(r)$. This implies that $\tau(r)$ must be 4, and therefore that $r = 3$. Hence, the elements in positions 1 and 3 must be transposed. This means that $\sigma_1 = 3241$, giving the following arrangement:

$$\sigma_1 \text{ shown by } \boxed{1}, \tau \text{ shown by } \widehat{1} : \begin{bmatrix} \widehat{1} & 1 & \boxed{1} & 1 \\ 1 & \boxed{\widehat{1}} & 1 & 0 \\ 0 & 1 & 1 & \boxed{\widehat{1}} \\ \boxed{1} & 1 & \widehat{1} & 1 \end{bmatrix}$$

Again, σ_1 and τ disagree on the first column, so $i = \tau^{-1}(i) = 1$. Since $\sigma_1(1) = 3$, find an r such that

$$1 \leq \sigma_1(r) < 3 \leq \tau(r)$$

Now, the only choice is $r = 4$. Transposing the elements in positions 1 and 4 gives $\sigma_2 = 1243$, as desired.

The above example should demonsrate the idea of the paths $\gamma_{\sigma\tau}$ – find the first column j on which σ_l and τ disagree, and then keep transposing the element in position $i = \tau^{-1}(j)$ with some other element in a position r until we get the permutations to match on column j. Note that while the mismatch is at column j, the same row i is reused for all the transpositions. It should be clear that these paths were chosen to be a type of left-to-right path discussed before. The following lemma makes this claim precise.

Lemma 3.11. *Let σ and τ be elements of S_M, and let $\gamma_{\sigma\tau}$ be the path $\sigma = \sigma_0, \sigma_1, \ldots, \sigma_m = \tau$. Now, fix a σ_s in this path, and let the first column on which σ_s disagrees with τ be j.*

Then, σ_s agrees with σ on all columns that are at least $j+k$; that is,

$$\sigma^{-1}(c) = \sigma_s^{-1}(c) \text{ for all } c \geq j+k.$$

Proof: All that must be shown is that at no point in the path between σ and σ_s was there an alteration in any columns which are at least $j+k$. Let $q < s$, and let j_1 be the first column on which σ_q disagrees with τ: since σ_q appears in the path $\gamma_{\sigma\tau}$ before σ_s, by definition of the path, $j_1 \leq j$. Now, letting $i_1 = \tau^{-1}(j_1)$ and copying the notation above, to get to σ_{q+1} transpose the elements in positions i_1 and r_1, where

$$\tau(i_1) \leq \sigma_q(r_1) < \sigma_q(i_1) \leq \tau(r_1)$$

Equivalently, σ_{q+1} is σ_q with the elements in columns $\sigma_q(r_1)$ and $\sigma_q(i_1)$ transposed. Now, $j_1 = \tau(i_1) \in S(i_1)$, and $S(i_1)$ is an interval containing at most k elements. Thus, for all $a \in S(i_1)$, $a \leq j_1 + k - 1 \leq j + k - 1$, and hence

$$\sigma_q(i_1) \leq j + k - 1$$

Since by definition, $\sigma_q(r_1) < \sigma_q(i_1)$, the step from σ_q to σ_{q+1} transposes two elements in columns which are both at most $j+k-1$. This doesn't change $\sigma^{-1}(c)$ for any $c \geq j+k$ before getting to σ_s, leading to the desired conclusion. $\square$

The above 'left-to-right' property defines maps $\phi_{\sigma\tau}$ which are required for Lemma 3.6. As described above, edges of $\gamma_{\sigma\tau}$ are mapped to vertices of the path $\gamma_{\tau\sigma}$.

Lemma 3.12. *Let the family of paths $\{\gamma_{\sigma\tau}\}$ be defined as above. Now, fix σ and τ, and let $\gamma_{\tau\sigma}$ be the backwards path $\tau = \tau_0, \tau_1, \ldots, \tau_l = \sigma$. Consider an edge $e = (\sigma_s, \sigma_{s+1})$ of the path $\gamma_{\sigma\tau}$, and assume σ_s agrees with τ on precisely the first $j-1$ columns, and disagrees with τ on the jth column. Then, let τ_p be the first vertex on the path $\gamma_{\tau\sigma}$ such that τ_p matches with σ on the first j columns, and define*

$$\phi_{\sigma\tau}(e) = \tau_p$$

Then, the conditions of Lemma 3.6 are satisfied; that is, for every choice of edge e and $\alpha \in S_M$,

$$|\{(\sigma,\tau) \mid e \in \gamma_{\sigma\tau}, \phi_{\sigma\tau}(e) = \alpha\}| \leq C_1(k) = 2(k!)^2$$

Proof: Fix α and the edge $e = (\rho_1, \rho_2)$. An upper bound on the number of (σ, τ) such that $\phi_{\sigma\tau}(e) = \alpha$ must be found. For convenience of notation, define

$$A = \{(\sigma, \tau) \mid e \in \gamma_{\sigma\tau}, \phi_{\sigma\tau}(e) = \alpha\}$$

Divide the above set according the order in which the edge (ρ_1, ρ_2) appears in the path $\gamma_{\sigma\tau}$. Thus, define

$$A_1 = \{(\sigma, \tau) \mid (\rho_1, \rho_2) \in \gamma_{\sigma\pi} \text{ with } \rho_1 = \sigma_s, \rho_2 = \sigma_{s+1} \text{ for some } s, \phi_{\sigma\tau}(\rho_1, \rho_2) = \alpha\}$$
$$A_2 = \{(\sigma, \tau) \mid (\rho_1, \rho_2) \in \gamma_{\sigma\pi} \text{ with } \rho_2 = \sigma_s, \rho_1 = \sigma_{s+1} \text{ for some } s, \phi_{\sigma\tau}(\rho_1, \rho_2) = \alpha\}$$

Clearly, A is the disjoint union of A_1 and A_2. An upper bound on $|A_1|$ will be given. Since A_2 is a completely analogous set, the same bound will apply.

In picking the edge (σ_s, σ_{s+1}) in the path $\gamma_{\sigma\tau}$, the first column j on which σ_s and τ disagree is chosen. Break down the set A_1 by choice of j; that is, fix j and consider the set

$$A_1^j = \{(\sigma, \tau) \in A_1 \mid j \text{ is the first column on which } \rho_1 \text{ and } \tau \text{ disagree}\}$$

Clearly, if $(\sigma, \tau) \in A_1^j$, then

$$\tau^{-1}(c) = \rho_1^{-1}(c) \text{ for all } c \leq j - 1 \tag{3.6}$$

Furthermore, from Lemma 3.11 above,

$$\sigma^{-1}(c) = \rho_1^{-1}(c) \text{ for all } c \geq j + k \tag{3.7}$$

Thus, ρ_1 determines τ on the 'small' columns and σ on the 'large' columns. Now the information provided by α is used; by definition, $\phi_{\sigma\tau}(\rho_1, \rho_2)$ is chosen to be the first vertex on the path $\gamma_{\tau\sigma}$ that agrees with σ on the first j columns. Thus,

$$\sigma^{-1}(c) = \alpha^{-1}(c) \text{ for all } c \leq j \tag{3.8}$$

Now α is used to determine $\tau^{-1}(c)$ for 'large' c. Note that α was chosen to be the first vertex τ_p on the path $\gamma_{\tau\sigma}$ to satisfy the property in Equation (3.8) above. There are two possibilities: the first is that $\alpha = \tau$, in which case clearly

$$\tau^{-1}(c) = \alpha^{-1}(c) \text{ for all } c \geq j + k. \tag{3.9}$$

The other, less trivial, possibility is that α is not the first vertex on the path $\gamma_{\tau\sigma}$, in which case there is a vertex τ_{p-1} preceding it that disagrees with σ on one of the first j columns. Using Lemma 3.11 above,

$$\tau^{-1}(c) = \tau_{p-1}^{-1}(c) \text{ for all } c \geq j + k$$

But τ_{p-1} and $\alpha = \tau_p$ differ only by a transposition; furthermore, since α agrees with σ on all columns that are at most j, this transposition must involve a column $q \leq j$. In addition, since each interval $S(i)$ is of size at most k, it is easy to see that such a column can only be transposed with columns up to $j+k-1$. Therefore, α and τ_{p-1} agree on columns $c \geq j+k$, so Equation (3.9) holds.

Thus, Equations (3.6), (3.7), (3.8), and (3.9) above yield that if $(\sigma, \tau) \in A_1^j$, then $\sigma^{-1}(c)$ is determined for all $c \leq j$ and $c \geq j + k$, while $\tau^{-1}(c)$ is determined for all $c \leq j - 1$ and $c \geq j+k$. Thus σ can be specified by choosing its values on columns $\{j+1, j+2, \ldots, j+k-1\}$ out of the $k-1$ available values. This gives $(k-1)!$ choices for σ; similarly, to pick τ, choose its values on $\{j, j+1, \ldots, j+k-1\}$ giving $k!$ choices for τ. This gives

$$|A_1^j| \leq (k-1)!k!$$

Since

$$|A_1| = \sum_j |A_1^j|$$

the number of possible j using α and (ρ_1, ρ_2) must be bounded. Let ρ_2 be ρ_1 with columns x and y transposed, and assume that $x < y$. If the first column on which ρ_1 differs from σ is j, then as usual let $i = \tau^{-1}(j)$, and find an r such that

$$\tau(i) \leq \rho_1(r) < \rho_1(i) \leq \tau(r).$$

Then transpose columns $\rho_1(r)$ and $\rho_1(i)$. Since $\rho_1(i)$ is the larger of the two, it must be that $\rho_1(i) = y$. Since $S(i)$ has at most k elements and contains both $\tau(i)$ and $\rho_1(i) = y$,

$\tau(i) \geq y - k$. From above, $\tau(i) < \rho_1(i) = y$. Therefore,

$$y - k \leq \tau(i) < y$$

Since $j = \tau(i)$, this means that given y, the possible values for j are in the interval $[y - k, y - 1]$. Therefore,

$$|A_1| = \sum_{j=y-k}^{y-1} |A_1^j| \leq k(k-1)!k! = (k!)^2$$

Since $|A| = |A_1| + |A_2|$, and since $|A_2| \leq (k!)^2$ by an identical argument,

$$|\{(\sigma, \tau) \mid e \in \gamma_{\sigma\tau}, \phi_{\sigma\tau}(e) = \alpha\}| = |A| \leq 2(k!)^2 = C_1(k)$$

as required. □

This gives the following two-line proof for Proposition 3.5.

Proof of Proposition 3.5: From Lemma 3.12, for paths $\gamma_{\sigma\tau}$ and maps $\phi_{\sigma\tau}$ as defined above,

$$|\{(\sigma, \tau) \mid e \in \gamma_{\sigma\tau}, \phi_{\sigma\tau}(e) = \alpha\}| \leq C_1(k)$$

for some constant $C_1(k)$. Furthermore, from Lemma 3.10, we have that these paths satisfy $|\gamma_{\sigma\tau}| \leq kn$ for each pair σ, τ. This completes the proof. □

Remark 3.13. Unfortunately, the above derivation makes it clear that the constant $C(k)$ in Theorem 3.2 is super-exponential in k, as a factor of $(k!)^2$ makes an appearance. While this could probably be improved, the constant could never be smaller than exponential in k, since Section 3.5 below provides a family of $n \times n$ restriction matrices for which the random transposition walk takes exponentially long to mix.

3.4 Fibonacci permutations

This section studies the mixing time for a particular example of resriction matrices with bounded intervals. These were introduced by Diaconis, Graham, and Holmes in [11] as 'Fibonacci permutations.' For reasons that will become apparent shortly, a sped-up version of the random transposition walk is introduced. The bound uses coupling, and is significantly smaller than the bound in Theorem 3.2.

Let M_n be the $n \times n$ tridiagonal matrix. That is,

$$M_n(i,j) = \begin{cases} 1 & |i-j| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

For example,

$$M_4 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Call S_{M_n} the set of *Fibonacci permutations.* This terminology was chosen because

$$|S_{M_n}| = F_{n+1} \tag{3.10}$$

where F_n is the nth Fibonacci number, a result derived in Section 2.4 of [11]. It is not difficult to prove using a recursion on the sizes of S_{M_n}.

From Proposition 2.1 in [11], the cycle representations of the permutations in S_{M_n} only contain fixed points and adjacent transpositions. Furthermore, the set of Fibonacci permutations is in bijective correspondence with the set of matchings on the n-path: that is, with the number of ways of putting edges between adjacent points in a path such that no vertex is used by more than one edge. The mapping between the two sets is simple: any pairwise adjacent transposition that appears in the cycle form of σ is mapped to an edge. For example, if $\sigma = 2134$, which in cycle form is $(12)(3)(4)$, then σ is mapped to

where the points in the above diagram are labelled in order.

It is easy to check that for $\sigma \in S_{M_n}$, the only allowed moves are either transposing a pair of adjacent fixed points, or using one of the transpositions which currently appear in the cycle form of σ. For example, for the above $\sigma = (12)(3)(4)$, the possible transpositions are (12) and (34). This means that the standard random transposition walk is extremely inefficient, since none but adjacent transpositions can ever succeed. We will therefore speed up the walk by only ever attempting to transpose adjacent elements – at each step, pick one of the adjacent transpositions $\{(1,2),(2,3),\ldots,(n-1,n)\}$ uniformly at random, and attempt to transpose it. Translating this into the language of matchings on the n-path, this leads to the following description of the walk.

Description 3.14. The adjacent transposition walk on the S_{M_n} is isomorphic to the following walk on the matchings of the n-path. At each step, pick an adjacent pair $(i, i+1)$, each with probability $\frac{1}{n-1}$. Then try to 'flip' the edge corresponding to that pair: that is, if the edge $(i, i+1)$ already appears in the matching, remove it, and otherwise attempt to add it to the matching. (Note that if either i or $i+1$ are already used by an edge in the matching, then we will fail to add $(i, i+1)$ and the walk will remain where it was.)

From now on, $\sigma \in S_{M_n}$ will be represented as a matching on the n-path instead of a permutation, and the walk will be the one described above. In this section, the following theorem is proved:

Theorem 3.15. *The random walk on the Fibonacci permutation sdescribed in Description 3.14 mixes in $\Theta(n \log n)$ time in total variation – that is, there exist constants c_1 and c_2 such that the walk has not mixed at time $c_1 n \log n$, and has mixed at time $c_2 n \log n$.*

The lower bound uses a paper of Hayes and Sinclair (see [23]) that applies in a more general setting. The upper bound uses the path coupling technique outlined in Section 2.2.1.

Remark 3.16. Theorem 3.15 is a considerable improvement on the direct use of the second eigenvalue. For the original random transposition walk, the arguments in the proof of Theorem 3.2 bound the second eigenvalue below by $1 - \frac{C}{n^3}$. Since the walk on the Fibonacci permutations in this section is sped up by a factor of n, the lower bound here is $1 - \frac{C}{n^2}$. This leads to an upper bound on the mixing time of $O(n^3)$, which is much bigger than the correct answer of $\Theta(n \log n)$.

3.4.1 Upper Bound

This subsection proves the upper bound in Theorem 3.15 using path coupling (Theorem 2.11 from Chapter 2.) For this, an edge set E' on S_{M_n} and a contracting one-step coupling for pairs (x, y) in E' must be specified. The following theorem will be proved:

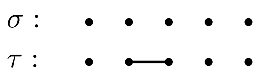
Theorem 3.17. *Consider the random transposition walk on S_{M_n} as described in Description 3.14. Then, using the notation of Theorem 2.11, there is an edge set E' and lengths $l(\alpha, \beta) \geq$*

1 *for pairs* (α, β) *in* E'*, such that for any* $(\sigma, \tau) \in E'$,

$$\mathbb{E}\left[\rho(X_1, Y_1)\right] \leq \left(1 - \frac{1}{4(n-1)}\right) \rho(\sigma, \tau)$$

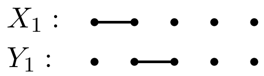
where (X_1, Y_1) *is the first step of a coupling started at* σ *and* τ*, and* ρ *is as usual the distance induced by the edge lengths* $l(\alpha, \beta)$.

Here is an intuitive description of the coupling. A technique that often succeeds is letting E' simply be the pairs that are one step away from each other in the random walk. This will not precisely work for this walk: here, pairs that are one step away from each other are pairs σ and τ that differ by one edge. Here's an example of such a pair:



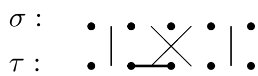
Recall that the steps of the walk consist of simply picking an edge and attempting to flip it. Hence, a reasonable coupling could be run as follows: create a bijection ϕ between pairs $(i, i+1)$ in the two n-paths corresponding to σ and τ, try to flip these edges simultaneously.

Example 3.18. An obvious example of a bijection ϕ is the one that maps each edge to itself. This leads to attempting to flip the same edge in σ and τ. For example, for σ and τ above, try to flip the edge $(1, 2)$ in both walks. This would result in the following pair (X_1, Y_1) as the next step of the coupling,



since the edge can be flipped in σ but not in τ.

A good visual way to represent the bijection ϕ is to draw lines between pairs of corresponding edges $(e, \phi(e))$. For example, again for the σ and τ above, here's a possible bijection:

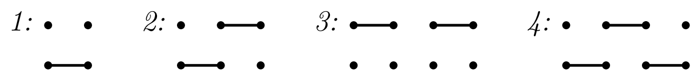


Thus, here ϕ maps edges of σ to edge of τ in the following way:

$$\phi(1,2) = (1,2),\ \phi(2,3) = (3,4),\ \phi(3,4) = (2,3),\ \phi(4,5) = (4,5)$$

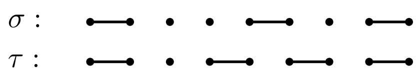
The above example helps illustrate why the edge set E' can't be pairs that differ by an edge: whatever the attempted coupling, the probability of the distance increasing to 2 at the next step is at least as big as the probability of the distance decreasing to 0. This leads to a non-contracting coupling. To overcome this problem, let the set of edges E' be a bit more complicated, and define the lengths $l(\sigma, \tau)$ in a specific way to ensure contraction after one step. Here is the desired definition.

Definition 3.19. *An edge (σ, τ) is in E' if σ and τ disagree on at most 3 consecutive edges $(i, i+1)$, and furthermore, if these disagreeing edges fall into one of the following four configurations*



where the order of σ and τ does not matter. Furthermore, define $l(\sigma, \tau)$ to be the number of edges σ and τ disagree on, plus 1. For example, if σ and τ are in the configuration labelled 1 above, then $l(\sigma, \tau) = 2$.

Example 3.20. For $n = 9$, let



It should be clear that σ and τ agree on all edges but $(4, 5)$, $(5, 6)$, and $(6, 7)$, and that these three edges are in configuration 4. Thus, $(\sigma, \tau) \in E'$, and $l(\sigma, \tau) = 4$.

Recall that by definition,

$$\rho(\sigma, \tau) = \min \left\{ \sum_{i=0}^{n-1} l(x_i, x_{i+1}) \mid x_0 = \sigma, x_n = \tau, (x_i, x_{i+1}) \in E' \text{ for all } i \right\}$$

which is the minimal length of a path between σ and τ in E'. The next result proves ρ and l match on pairs (σ, τ) in E' – this is generally expected, since otherwise some of the edges in E' would be redundant.

Proposition 3.21. *For E' as defined above in Definition 3.19,*

$$\rho(\sigma, \tau) = l(\sigma, \tau).$$

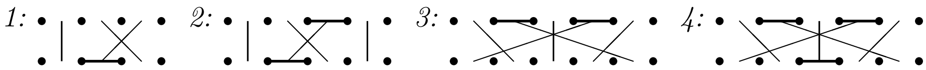
Proof: Let $(\sigma, \tau) \in E'$. Note that by definition, if $(x, y) \in E'$, then $l(x, y) \geq 2$. Thus, a path with at least two edges between σ and τ would have length at least 4, which is already at least $l(\sigma, \tau)$. This means that $\rho(\sigma, \tau) = l(\sigma, \tau)$, as required. □

To define the coupling, as discussed above, a bijection ϕ between edges of σ and τ is created. The coupling then consists of trying to flip the paired edges concurrently. This bijection will be described visually, as it was in Example 3.18.

Definition 3.22. *A coupling for pairs $(\sigma, \tau) \in E'$ is defined via a bijection ϕ from edges of σ to edges of τ, which consists of flipping the edge e of σ concurrently with flipping the edge $\phi(e)$ of τ.*

1. *If $(i, i+1)$ is an edge on which σ and τ match, and which does not neighbor one of the edges of disagreement, then let $\phi(i, i+1) = (i, i+1)$. That is, flip such edges simultaneously in both σ and τ.*

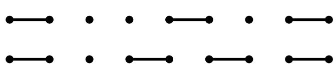
2. *Otherwise, the edge must either be in one of the configurations in Definition 3.19, or a neighboring edge. The following is a visual description of the bijections, corresponding to the various pairs in E':*



Note: *It is possible that a configuration is at the 'end' of the n-path (that is, includes the edge $(1, 2)$ or $(n-1, n)$) and hence has no neighboring edges on one side. In that case, it's possible that there exists an edge e in σ such that $\phi(e)$ is non-existent. If this is the case, there also exists an edge f in τ such that $\phi^{-1}(f)$ is non-existent – this problem is easy to resolve by setting $\phi(e) = f$.*

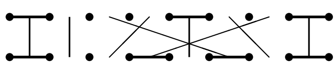
Example 3.23. Here is an example of the one-step coupling for an actual pair $(\sigma, \tau) \in E'$. Consider the following pair, which has already appeared in the example above:

σ :
τ :



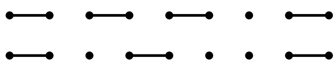
This pair is clearly in configuration 4. Accordingly, the bijection is

σ :
τ :



For example, trying to flip edge $(3,4)$ in σ leads to flipping the edge $(6,7)$ in τ, and the following pair (X_1, Y_1):

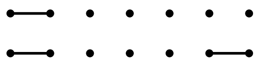
X_1 :
Y_1 :



Note that this is also in configuration 4.

Here is an example of a pair where the configuration is at the end of the n-path. Consider the pair

σ :
τ :

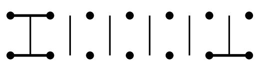


This is clearly in configuration 1. Thus, the prescribed bijection is:

σ :
τ :

which 'uses' non-existent edges. As noted in the definition of the coupling, in this case the two edges that are currently in bijection with the non-existent edges are connected, resulting in the following coupling:

σ :
τ :



Return to the proof that the above coupling indeed contracts, and hence fulfills the requirements of Theorem 3.17. This is a little tedious, as each of the four configurations is treated separately.

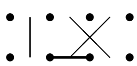
Proof of Theorem 3.17: It must be shown

$$\mathbb{E}\left[\rho(X_1, Y_1)\right] \leq \left(1 - \frac{1}{4(n-1)}\right) \rho(\sigma, \tau)$$

for pairs $(\sigma, \tau) \in E'$. This is treated case by case.

Configuration 1: Here, σ and τ are in the following configuration:

σ:
τ:



While the neighboring edges are drawn, this won't matter in the calculation. Here, clearly $\rho(\sigma, \tau) =$ number of edges of disagreement $+ 1 = 2$.

To increase the distance at the next step, an edge must be added for them to disagree on. This could only be accomplished by attempting to flip $(1, 2)$ in σ. This would put things in configuration 2, yielding

$$\mathbb{P}(\rho(X_1, Y_1) = 3) \leq \mathbb{P}(\text{flipping edge } (1, 2) \text{ in } \sigma) = \frac{1}{n-1} \tag{3.11}$$

Note that the inequality stems from the fact that $(1, 2)$ is a neighboring edge and hence might not exist, in which case it clearly can't be added.

A way to decrease distance is to decrease the number of disagreements. This can be accomplished by flipping edge $(2, 3)$ in σ – since it is mapped to the unflippable edge $(3, 4)$, this choice leads to $X_1 = Y_1$. Thus,

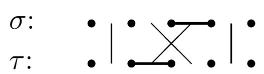
$$\mathbb{P}(\rho(X_1, Y_1) = 0) \geq \mathbb{P}(\text{flipping edge } (2, 3) \text{ in } \sigma) = \frac{1}{n-1}$$

Note that the edge $(2, 3)$ is an integral part of the configuration and hence clearly actually exists. Since we have already shown that the only way distance increases corresponds to Equation 3.11,

$$\begin{aligned}\mathbb{E}\left[\rho(X_1, Y_1) - \rho(\sigma, \tau)\right] &\leq (3-2) \cdot \frac{1}{n-1} + (0-2) \cdot \frac{1}{n-1} \\ &= -\frac{2}{n-1} = -\frac{1}{n-1}\rho(\sigma, \tau) \end{aligned} \tag{3.12}$$

which leads to the desired inequality for configuration 1.

Configuration 2: Here, σ and τ are in the following configuration:



The analysis is similar to the one above. To increase distance, attempt flipping either edge $(1, 2)$ or edge $(4, 5)$ in σ. Both of those lead to configuration 3. Thus,

$$\mathbb{P}(\rho(X_1, Y_1) = 4) \leq \frac{2}{n-1}$$

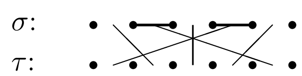
Similarly, to decrease distance flip edge $(3,4)$ in σ (and of course, concurrently edge $(2,3)$ in τ), leading to $X_1 = Y_1$. Thus,

$$\mathbb{P}(\rho(X_1, Y_1) = 0) = \frac{1}{n-1}$$

Having accounted for all the ways distance could increase, because $\rho(\sigma,\tau) = 3$,

$$\begin{aligned}\mathbb{E}\left[\rho(X_1,Y_1) - \rho(\sigma,\tau)\right] &\leq (4-3)\cdot\frac{2}{n-1} + (0-3)\cdot\frac{1}{n-1} \\ &= -\frac{1}{n-1} = -\frac{1}{3(n-1)}\rho(\sigma,\tau)\end{aligned} \tag{3.13}$$

Configuration 3: Here, σ and τ are in the following configuration:



Now, distance increases by trying to flip edge $(3,4)$ in σ, going to configuration 4. Hence

$$\mathbb{P}(\rho(X_1, Y_1) = 4) = \frac{1}{n-1}.$$

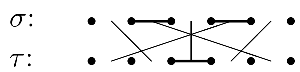
Similarly, distance decreases via edges $(2,3)$ and $(4,5)$ in τ (talking about edges of τ since there's a chance that the edges of σ they are in bijection with don't exist), both of which lead either to configuration 1 if the neighboring edges exist, or to distance 0 if some of them don't. Hence,

$$\mathbb{P}(\rho(X_1, Y_1) \leq 2) \geq \frac{2}{n-1}.$$

Edges $(2,3)$ and $(4,5)$ in σ change the configuration, but clearly don't add a disagreement. Therefore,

$$\begin{aligned}\mathbb{E}\left[\rho(X_1,Y_1) - \rho(\sigma,\tau)\right] &\leq (4-3)\cdot\frac{1}{n-1} + (2-3)\cdot\frac{2}{n-1} \\ &= -\frac{1}{n-1} = -\frac{1}{3(n-1)}\rho(\sigma,\tau)\end{aligned} \tag{3.14}$$

Configuration 4: Here, σ and τ are in the following configuration:



This one is very simple: there's no way to increase distance here. However, edge $(3, 4)$ in σ (and in τ given the bijection) decreases the distance by 1. Hence,

$$\mathbb{E}\left[\rho(X_1, Y_1) - \rho(\sigma, \tau)\right] \le (3-4) \cdot \frac{1}{n-1} \tag{3.15}$$

$$= -\frac{1}{n-1} = -\frac{1}{4(n-1)}\rho(\sigma, \tau) \tag{3.16}$$

Finally, combining Equations 3.12, 3.13, 3.14, and 3.15 above,

$$\mathbb{E}\left[\rho(X_1, Y_1)\right] \le \left(1 - \frac{1}{4(n-1)}\right)\rho(\sigma, \tau)$$

as required. □

Before continuing the proof of Theorem 3.15, a simple lemma concerning the diameter of the graph S_{M_n} under the metric ρ is needed.

Lemma 3.24. *The diameter* $\operatorname{diam}(S_{M_n})$ *is at most* $2n$*.*

Proof: By definition of configuration 1 above, if σ and τ differ by one edge, they are distance 2 apart. A path from any σ to any τ can be created by removing all the edges of σ, getting to the empty matching, then adding all the edges of τ. Since edges aren't allowed to be adjacent, σ and τ both have at most $\frac{n}{2}$ edges, and hence this path takes n steps, each of length 2. Thus, $\operatorname{diam}(S_{M_n}) \le 2n$, as required. □

Proof of Upper Bound in Theorem 3.15: By Theorem 3.17, there is a path coupling that satisfies

$$\mathbb{E}\left[\rho(X_1, Y_1)\right] \le \left(1 - \frac{1}{4(n-1)}\right)\rho(\sigma, \tau)$$

Furthermore, $\operatorname{diam}(S_{M_n}) \le 2n$. Thus, applying Theorem 2.11 with $\kappa = 1 - \frac{1}{4(n-1)}$ gives

$$d(t) \le \operatorname{diam}(S)\kappa^t = 2n\left(1 - \frac{1}{4(n-1)}\right)^t$$

Hence, if $t = 8(n-1)\log n$,

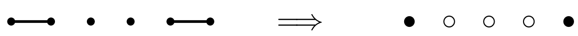
$$d(t) \le 2n\left(1 - \frac{1}{4(n-1)}\right)^t \le 2ne^{-2\log n} = \frac{2}{n}$$

so by this time the walk has thoroughly mixed. Since this t is of order $n \log n$, the argument is complete. $\square$

3.4.2 Lower Bound

The lower bound in Theorem 3.15 follows from a general lower bound theorem of Hayes and Sinclair [23]. Before stating the theorem, a number of observations are useful.

First of all, Fibonacci matchings on the n-path can be thought of as colorings on the $(n-1)$-path: map the edge $(i, i+1)$ to the vertex i, coloring the vertex black if the edge is present in the matching and coloring it white if it's not. Here's an example of how the map works:



Clearly, the restriction on colorings is that two black vertices cannot be next to each other. This is called the *hardcore model.* The walk in Description 3.14 can be translated to the hardcore model in the following way: at each step, pick a vertex in $\{1, 2, \ldots, n-1\}$ uniformly at random, and attempt to recolor it. If this is possible without violating the restriction about neighboring vertices, do so; if it's not, stay in place.

Following the definition in [23], a Glauber dynamics chain is a Markov chain on the colorings of a graph G, which is reversible (as defined in Definition 2.12) and which updates one site at a time. It's clear that the above walk on the hardcore model satisfies both those constraints. Theorem 1.1 from [23] can now be stated.

Theorem 3.25. *Let $\Delta \geq 2$ be a fixed positive integer, and let G be any graph on n vertices of degree at most Δ. Any nonredundant Glauber dynamics on G has mixing time $\Omega(n \log n)$, where the constant in $\Omega(\cdot)$ depends only on Δ.*

Here, nonredundant dynamics means simply that each vertex has at least two viable states, which clearly holds for the hardcore model. Furthermore, for Glauber dynamics on the $(n-1)$-path, the maximum degree of the graphs is $\Delta = 2$. This means that the above theorem applies to the Fibonacci walk, and hence the lower bound part of Theorem 3.15 follows.

3.5 Counterexample

The examples in this section show that, in general, the above walk on permutations with two sided interval restrictions does not mix in polynomial time. This disproves a conjecture of Diaconis, Graham and Holmes, who guessed that such a walk must in general mix in order $n^2 \log n$ [11]. I'm grateful for John Jiang for providing this counterexample.

Define a $(2n-1) \times (2n-1)$ matrix A_n in the following way:

$$A_n(i,j) = \begin{cases} 1 & \text{if } j-i \leq n \text{ and } i \leq n \leq j \\ 1 & \text{if } i-j \leq n \text{ and } j \leq n \leq i \\ 0 & \text{otherwise} \end{cases}$$

For example, here's the matrix for $n = 4$:

$$A_4 = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

The following technical lemma is needed:

Lemma 3.26. *Let A_n be the $(2n-1) \times (2n-1)$ matrix defined above. Then, S_{A_n} can be written as $H_1 \cup H_2$, where $|H_1| = |H_2| = 2^{n-1}$, and H_1 and H_2 intersect in one vertex v. Furthermore, the only edges between H_1 and H_2 are at v, where v has precisely $n-1$ neighbors in H_1 and precisely $n-1$ neighbors in H_2.*

The above lemma yields the following theorem.

Theorem 3.27. *Let Φ_* be the bottleneck ratio for S_{A_n} as defined in Equation (2.7). Then, $\Phi_* \leq \frac{2}{n(2^{n-1}-1)}$, and therefore*

$$\tau_{\text{mix}} \geq \frac{n(2^{n-1}-1)}{8}$$

Proof: Decompose S_{A_n} as $H_1 \cup H_2$ as in the above lemma, and let v be the vertex in the intersection of H_1 and H_2. Define $S = H_1/\{v\}$. Then,

$$Q(S, S^c) = \sum_{x \in H_1/\{v\}, y \in H_2} \pi(x)P(x, y) = \sum_{\substack{(x,y) \in E(S_M) \\ x \in H_1/\{v\}, y \in H_2}} \frac{2}{n(n-1)|S_M|}$$

From the above lemma, the only edges between H_1 and H_2 are at v. Thus, the only pairs (x, y) such that $x \in H_1/\{v\}$, $y \in H_2$ and $(x, y) \in E(S_M)$ are pairs (x, v), where x is a neighbor of v in H_1. Now, v has precisely $n - 1$ neighbors in H_1; thus,

$$Q(S, S^c) = \frac{2}{n|S_M|}$$

Furthermore,

$$\pi(S) = \frac{|H_1/\{v\}|}{|S_M|} = \frac{2^{n-1} - 1}{|S_M|} = \frac{2^{n-1} - 1}{2^n - 1} < \frac{1}{2}$$

Therefore,

$$\Phi_* \leq \frac{Q(S, S^c)}{\pi(S)} = \frac{2}{n(2^{n-1} - 1)}$$

Finally, Theorem 2.23 states that $\tau_{\text{mix}} \geq \frac{1}{4\Phi_*}$ and therefore

$$\tau_{\text{mix}} \geq \frac{n(2^{n-1} - 1)}{8}$$

as required. □

Remark 3.28. Since this chain is reversible, Theorem 2.24 gives a slightly better constant; however, in this case this is immaterial.

The decomposition $S_{A_n} = H_1 \cup H_2$ partitions S into a union of two hypercubes that overlap at $(1, 1, \ldots, 1)$. In order to show this, the following lemma is needed.

Lemma 3.29. *If $\sigma \in S_{A_n}$, and $\sigma(n) \geq n$, then $\sigma(n+i) = i$ for each i such that $1 \leq i \leq n-1$. Similarly, if $\sigma(n) \leq n$, then $\sigma(i) = n + i$ for each i such that $1 \leq i \leq n - 1$.*

Proof: To prove the first half of the above statement, assume that $\sigma(n) \geq n$. Proceed by induction. Note that there are exactly two choices for $\sigma^{-1}(1)$, and these are n and $n + 1$. Since $\sigma(n) \geq n$, $\sigma^{-1}(1)$ must be $n + 1$. This gives the expression for $i = 1$.

For the inductive step, assume that $\sigma(n + i) = i$ for all $i < j \leq n - 1$. Now, the choices for $\sigma^{-1}(j)$ are by definition $\{n, n+1, n+2, \cdots, n+j\}$. Since $\sigma(n+i) = i$ for all $i < j \leq n-1$

by the inductive hypothesis, and $\sigma(n) \geq n$, the only remaining choice for $\sigma^{-1}(j)$ is $n + j$. Thus, $\sigma(n + j) = j$ and so by induction, we're done. The second half of the statement follows identically. $\square$

Example 3.30. Suppose that $n = 4$ and $\sigma(4) = 6$. Why does $\sigma(n+i) = i$ for each i? Using the visual representation for permutations, σ has a boxed 1 at $(4, 6)$. Now, 'crossing out' the row and column of the boxed 1 gives

$$\sigma : \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & \boxed{1} & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Looking at the first column, there's now only one 1 that isn't crossed out. This means that $\sigma(5) = 1$. Boxing that 1, and crossing out its row and column gives

$$\sigma : \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & \boxed{1} & 1 \\ \boxed{1} & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

By looking at the second column, the only 1 that isn't crossed out is in row 6. Thus, $\sigma(6) = 2$. Continuing in this way show that $\sigma(7) = 3$, and hence $\sigma(n + i) = i$ for each i.

Proof of Lemma 3.26: Now, define

$$H_1 = \{\sigma \in S_{A_n} \mid \sigma(n) \geq n\}$$
$$H_2 = \{\sigma \in S_{A_n} \mid \sigma(n) \leq n\}$$

From the above lemma, if $\sigma \in H_1$, then $\sigma(n + i) = i$ for all i such that $1 \leq i \leq n - 1$.

This means that H_1 is isomorphic to S_N, where N is the $n \times n$ submatrix in the upper right cornder of A_n. Now, it was shown in [22] that S_N is isomorphic to the $(n-1)$-dimensional hypercube, and hence it is of size 2^{n-1}. Similarly, H_2 is isomorphic to $S_{N'}$, where N' is the $n \times n$ submatrix in the bottom left corner of A_n, and hence H_2 is also isomorphic to the $(n-1)$-dimensional hypercube.

Now, $H_1 \cap H_2 = \{\sigma \in S_{A_n} \mid \pi(n) = n\}$. The above lemma implies that if $\sigma(n) = n$, then both $\sigma(n+i) = i$ for each i such that $1 \leq i \leq n-1$, and $\sigma(i) = n+i$ for each i such that $1 \leq i \leq n-1$. This means that there is exactly one vertex in the intersection of H_1 and H_2. Furthermore, note that if $\sigma(n) > n$, then any neighbor of σ must also be in H_1 – since $\sigma(i) \leq n$ for all $i \geq n+1$, the condition $\sigma(n) \geq n$ prevents swapping row n with any row $i \geq n+1$. A similar statement holds for H_2.

Thus, $|H_1| = |H_2| = 2^{n-1}$, there is precisely one vertex v lying in the intersection of H_1 and H_2, and the only edges between H_1 and H_2 are at v. Finally, since v is just a vertex of each of the $(n-1)$-dimensional hypercubes, v has precisely $n-1$ neighbors in H_1, and the same number of neighbors in H_2. This leads to the desired conclusion. $\square$

Chapter 4

One-Sided Interval Restrictions

4.1 Introduction

This chapter works with one-sided restriction matrices M, as defined in Chapter 1. For such matrices M, the permutations in S_M correspond to rook placements on a Ferrers board – that is, a board obtained by removing a Ferrers diagram from one of the corners of an $n \times n$ matrix. These objects have an elegant combinatorial structure, originally studied by Foata, and Schützenberger [19], and then later by Goldman, Joichi and White [20]. In particular, it is possible to perfectly sample elements of S_M.

Studying the random transposition walk on S_M gives insight to more general restrictions. The additional structure of the set of permutations with one-sided restrictions allows sharp estimates of convergence. Indeed, in this case a proof can be given of *cut-off*, a phenomenon which occurs when the walk transitions from being barely mixed to being thoroughly mixed in a window much smaller than the mixing time. This definition can be applied to various distances; accordingly, this chapter will examine both total variation and chi-squared distance.

The sharp estimates on rates of convergence use the spectral information for the walk. This walk has been diagonalized by Hanlon [22], who derived all the eigenvalues and eigenvectors in terms of combinatorial objects called $\vec{b}$-partitions, defined later in Definition 4.20. This amazing result is an extension of the diagonalization of the random transposition walk on S_n, which used representation theory (see James and Kerber [26] for a treatment of the standard theory.) Here, the diagonalization was possible despite the absence of a group structure.

The present approach is inspired by the results of Diaconis and Shahshahani [14], who used the eigenvalues of the random transposition walk on S_n to show total variation cut-off of the walk at time $\frac{1}{2}n\log n$ with a window of size n. The argument proceeds as follows: attention is restricted to a subclass of one-sided restriction matrices, for which the set of allowed positions $S(i)$ is either the whole interval $[1, n]$ or $[a, n]$ for some fixed a. Colloquially, these are called 'two-step' restriction matrices. For this class of restriction matrices, conditions are specified such that the random transposition walk experiences chi-squared cut-off. However, it turns out that total variation mixing is more complicated: an extra condition will be needed to ensure total variation cut-off. Furthermore, a class of examples is given for which total variation mixing occurs considerably earlier, and for which cut-off does not occur.

These examples of the random transposition walk on S_M, where M is a one-sided restriction matrix, turns out to be instructive in two ways. In some subset of cases, they show the utility of using spectral information to obtain cutoff with respect to the total variation distance; on the other hand, a large class of examples is produced in which the total variation mixing time and the chi-squared mixing time do not match. Here, the spectral information does not give the right answer for total variation distance.

4.2 Definitions and Setup

If M is an $n \times n$ restriction matrix, since it is one-sided, $S(i)$ must be equal to $[b_i, n]$ for some b_i. To simplify notation, make the following definition.

Definition 4.1. *Let n be an integer, and let $\vec{b} = (b_1, b_2, \cdots, b_n)$ be a vector of elements of $\{1, 2, \ldots, n\}$. Define $M(\vec{b})$ to be the $n \times n$ interval restriction matrix satisfying $S(i) = [b_i, n]$ for all i.*

Here is an example: if $n = 4$ and $\vec{b} = (1, 1, 2, 3)$, then

$$M(\vec{b}) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Remark 4.2. The notation here is different from Hanlon's – he doesn't explicitly talk about restriction matrices and uses the notation $R_n(\vec{b})$ for the set we have called $S_{M(\vec{b})}$. The new notation is consistent with the rest of the thesis.

Lemma 3.1 shows that permuting the rows of M doesn't change the underlying graph. Therefore, from now on assume that the vector $\vec{b}$ satisfies

$$b_1 \le b_2 \le \cdots \le b_n \tag{4.1}$$

With the above assumption, Proposition 2.1 (d) in Hanlon [22] shows that:

Lemma 4.3. *If $b_i \le i$ for all i then the graph on $S_{M(\vec{b})}$ induced by the random transposition walk is regular, with degree*

$$\Delta = \sum_{i=1}^{n}(i - b_i) \tag{4.2}$$

If $b_i > i$ for some i, the set $S_{M(\vec{b})}$ is empty.

The basic walk proceeds as follows. First, pick a number i from $\{1, 2, \ldots, n\}$ with your left hand, and an independent number j from $\{1, 2, \ldots, n\}$ with your right hand. Note that it is is possible to chose the same number. Now, if the cards in positions i and j can be transposed while remaining in the set $S_{M(\vec{b})}$, do so. If this is not possible, go back and pick a different pair i' and j' using the same procedure. Continue selecting pairs until arriving at one that can be exchanged without leaving $S_{M(\vec{b})}$ – transposing this pair will then be the next step in the chain. Note that in this variant of the walk, picking a pair i and j that cannot be transposed does not get counted as a holding step, unlike in Chapter 3. Since the graph induced on $S_{M(\vec{b})}$ by the random transposition walk is regular, this is just a scaled version of that walk. This walk was chosen to match the one described by Hanlon [22].

Definition 4.4. *Given a vector $\vec{b} = (b_1, b_2, \ldots, b_n)$, define $U(\vec{b})$ to be the adjacency matrix of $S_{M(\vec{b})}$ under the random transposition walk.*

With Δ from Equation 4.2, if P is the transition matrix for the random walk, then for τ and σ that differ by a transposition (i, j),

$$P(\sigma, \tau) = \frac{2}{n + 2\Delta}$$

since i can be chosen with the right hand and j with the left hand, or vice versa. Furthermore,

$$P(\sigma, \sigma) = \frac{n}{n + 2\Delta}$$

since the same i can be chosen with both hands. This shows that

$$P = \frac{1}{n+2\Delta}(nI + 2U) \tag{4.3}$$

where I is the identity matrix. This demonstrates that finding the spectral information for P reduces to diagonalizing U. Correspondingly, many of the theorems in this chapter will be stated in terms of $U(\vec{b})$.

As noted earlier, attention here is restricted to a subclass of one-sided restriction matrices:

Definition 4.5. *Let $f(n), g(n)$ be two functions from $\mathbb{Z}^+$ to $\mathbb{Z}^+$. Define $\vec{b}_n(f,g)$ to be the vector $(b_1, b_2, \ldots, b_n)$ such that*

$$b_i = \begin{cases} 1 & i \le f(n) \\ g(n)+1 & f(n) < i \le n \end{cases}$$

That is, $\vec{b}_n(f,g) = (1, 1, \ldots, 1, g(n)+1, \ldots, g(n)+1)$ where the number of 1s in the beginning of the vector is $f(n)$.

Remark 4.6. From Lemma 4.3, $b_i \le i$ for all i is necessary for $S_{M(\vec{b})}$ not to be empty. This condition at $i = g(n)+1$ above forces

$$f(n) \ge g(n) \tag{4.4}$$

to a have a non-empty walk. Thus, from here on assume the above inequality for functions f and g for all n.

Now, let $\vec{b} = \vec{b}_n(f,g)$ as defined above, and consider the matrix $M(\vec{b})$. By the above definitions, $S[i]$ is $[1,n]$ if i is between 1 and $f(n)$, while $S[i]$ is $[g(n)+1, n]$ for i between $f(n)+1$ and n. This means that $\sigma \in S_{M(\vec{b})}$ has no restrictions on the first $f(n)$ rows, and must be at least $g(n)+1$ on rows that are at least $f(n)+1$. Alternatively, on the first $g(n)$ columns, σ is only allowed to take values up to $f(n)$. For example, if $n = 5$, $f(5) = 3$ and $g(5) = 2$, then

$$\vec{b} = \vec{b}_5(f,g) = (1,1,1,3,3) \tag{4.5}$$

and

$$M(\vec{b}) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

The next lemma calculates the degree Δ of a vertex in S_M for above matrices M.

Lemma 4.7. *Let $\vec{b}_n(f,g)$ be defined as in Definition 4.5 above, and let $M_n = M(\vec{b}_n(f,g))$. Then, if Δ is defined as in Equation* (4.2) *above,*

$$\Delta = \frac{n^2 - n - 2ng(n) + f(n)g(n)}{2}$$

Proof: From the definition of b_i,

$$\begin{aligned} \Delta = \sum_{i=1}^{n}(i - b_i) &= \frac{n(n+1)}{2} - \sum_{i=1}^{n} b_i \\ &= \frac{n(n+1)}{2} - f(n)\cdot 1 - (n - f(n))\cdot(g(n)+1) \\ &= \frac{n^2 - n - 2ng(n) + f(n)g(n)}{2} \end{aligned}$$

as desired. □

Notation 4.8. From now on, when f and g are implied, the convention

$$M_n = M(\vec{b}_n(f,g)) \tag{4.6}$$

is used to simplify notation.

The first theorem below shows that, given certain assumptions on $f(n)$ and $g(n)$, the walk always achieve cut-off in chi-squared distance, where chi-squared distance is defined as in Definition 2.20: that is,

$$\|\mu - \pi\|_{2,\pi} = \left\|\frac{\mu}{\pi} - 1\right\|_2 = \sqrt{\sum_x \left(\frac{\mu(x)}{\pi(x)} - 1\right)^2 \pi(x)} \tag{4.7}$$

This uses Corollary 2.22 from Chapter 2, which states that if P is the transition matrix of a vertex transitive Markov chain on a set S with eigenvalues $1 = \beta_0 > \beta_1 \geq \cdots \geq \beta_{|S|-1} \geq -1$, then

$$\left\|P^t(x,\cdot) - \pi\right\|_{TV} \leq \frac{1}{2}\left\|P^t(x,\cdot) - \pi\right\|_{2,\pi} = \frac{1}{2}\sqrt{\sum_{i=1}^{|S|-1} \beta_i^{2t}} \tag{4.8}$$

Theorem 4.9. *Let $\vec{b}_n = \vec{b}_n(f,g)$ for some functions f and g which satisfy $f(n) \geq g(n)$ for all n, and which also satisfy*

$$\lim_{n\to\infty} \frac{f(n)}{n} = 0 \text{ and } \lim_{n\to\infty} f(n) = \infty$$

and consider the random transposition walk on S_{M_n}. Let Δ denote the degree of the graph induced on S_M; hence, from Equation 4.7 above, $n + 2\Delta = n^2 - 2ng(n) + f(n)g(n)$. Then, there is chi-squared cut-off around

$$\frac{(n+2\Delta)(\log f(n) + \log g(n))}{4f(n)}$$

with a window of size $\frac{n+2\Delta}{4f(n)}$. More precisely,

1. *If*

$$t = \frac{(n+2\Delta)(\log f(n) + \log g(n))}{4f(n)} + c\frac{n+2\Delta}{4f(n)} \tag{4.9}$$

then

$$\left\|P^t(x,\cdot) - \pi\right\|_{2,\pi} \leq 4e^{-\frac{c}{2}}$$

for $c > 10$ and n sufficiently large.

2. *Furthermore, if $c > 0$ and*

$$t = \frac{(n+2\Delta)(\log f(n) + \log g(n))}{4f(n)} - c\frac{n+2\Delta}{4f(n)} \tag{4.10}$$

then

$$\left\|P^t(x,\cdot) - \pi\right\|_{2,\pi} \geq \frac{1}{2}e^{c/2}$$

for n sufficiently large.

As noted earlier, there are also interesting results for the more intuitive total variation distance. The following theorem states assumptions on f and g which imply total variation cut-off for the walk, while the theorem after it presents a class of examples for which total variation mixing occurs significantly before chi-squared mixing, and in which total variation cut-off does not occur.

Theorem 4.10. *Assume that f and g satisfy the hypotheses for Theorem 4.9 above, and furthermore, that* $\liminf_{n\to\infty} \frac{g(n)}{f(n)} > r > 0$. *Let* $M_n = M(\vec{b}_n(f,g))$ *as in Equation* (4.6) *above. Then, if*

$$t = \frac{(n+2\Delta)(\log g(n) + \log f(n))}{4f(n)} - c\frac{n+2\Delta}{4f(n)} \tag{4.11}$$

then the random transposition walk on S_{M_n} satisfies

$$\liminf_{n\to\infty} \left\| P^t(x,\cdot) - \pi \right\|_{TV} \geq \frac{1}{e} - e^{-re^c}$$

and thus the walk hasn't mixed by time t for sufficiently large c.

Theorem 4.11. *Let $g(n) = 1$, and assume that*

$$\lim_{n\to\infty} f(n) = \infty \text{ and } f(n) \leq \frac{n}{5\log n} \text{ for sufficiently large } n.$$

Then, the random transposition walk on S_{M_n} mixes in total variation distance in order $\frac{n+2\Delta}{f(n)}$ time. Furthermore, the walk does not have total variation cut-off.

The next section gives heuristics and proves Theorem 4.10, while Section 4.4 proves Theorem 4.11, and also proves that the walk is vertex transitive. The chi-squared results are then proved. Section 4.5 reviews the necessary background material, restates Hanlon's eigenvalue results and provides examples of how they apply for the specific matrices $M_n = M(\vec{b}_n(f,g))$ under consideration. Section 4.6 provides bounds on the eigenvalues of the transition matrix P, which will be used in the following section to do lead term analysis and prove the lower bound in Theorem 4.9. Finally, Sections 4.8 and 4.9 prove the upper bound in Theorem 4.9.

4.3 Heuristics and Total Variation Lower Bound

This section gives heuristics and proves Theorem 4.10. First, here is an argument explaining why a total variation cut-off is expected around

$$t = \frac{(n+2\Delta)(\log f(n) + \log g(n))}{4f(n)}$$

if $f(n)$ and $g(n)$ are of the same order, as the conjunction of Theorems 4.9 and 4.10 implies.

Consider the instructive case $f(n) = g(n)$. For example, if $n = 5$ and $f(n) = g(n) = 2$, the restriction matrix is

$$M = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

In this case, $\sigma \in S_M$ can take values up to $f(n)$ on the first $f(n)$ columns. This makes it easy to see that $\sigma \in S_m$ can be represented as a pair of permutations, one on $\{1, 2, \ldots, f(n)\}$ and the other on $\{f(n)+1, f(n)+2, \ldots, n\}$. Thus,

$$S_M \cong S_{f(n)} \times S_{n-f(n)}$$

Now consider how fast the walk mixes. It has been shown by Diaconis and Shahshahani that the transposition walk on S_k has cut-off at time $\frac{1}{2}k \log k$ with a window of size k [14]. At each step, the present walk uses either a transposition in $S_{f(n)}$ (with probability $\frac{f(n)^2}{n+2\Delta}$) or a transposition in $S_{n-f(n)}$ (with probability $\frac{(n-f(n))^2}{n+2\Delta}$). This is equivalent to two separate walks: the walk on $S_{f(n)}$, slowed down by a factor of $\frac{n+2\Delta}{f(n)^2}$, and the walk on $S_{n-f(n)}$, slowed down by a factor of $\frac{n+2\Delta}{(n-f(n))^2}$. Thus, the walk on $S_{f(n)}$ has cut-off around time

$$t_1 = \frac{1}{2} f(n) \log f(n) \cdot \frac{n+2\Delta}{f(n)^2} = \frac{(n+2\Delta) \log f(n)}{f(n)}$$

with a window of size $f(n) \cdot \frac{n+2\Delta}{f(n)^2} = \frac{n+2\Delta}{f(n)}$, while the walk on $S_{n-f(n)}$ analogously has cut-off around time

$$t_2 = \frac{(n+2\Delta) \log(n - f(n))}{n - f(n)}$$

with a window of size $\frac{n+2\Delta}{n-f(n)}$. It is easy to see (and is proved below in Lemma 4.36), that if $f(n)$ is comparitively small with respect to n, then t_1 is the larger of the two times. Since in this case, $f(n) = g(n)$, t_1 gives precisely the answer of Theorems 4.9 and 4.10. The above analysis shows that the first $g(n)$ columns are the limiting component of the walk. Theorem 4.9 assumes that $\frac{f(n)}{n} \to 0$ to simplify computations.

Remark 4.12. This discussion shows why $f(n) \to \infty$ is assumed for the main theorems –

in order to reasonably talk about asymptotics, the limiting term should at least go to ∞.

Before proving Theorem 4.10, a supporting lemma calculating the size of $S_{M(\vec{a})}$ is needed.

Lemma 4.13. *If $\vec{a} = (1, 1, \ldots, 1, y+1, \ldots, y+1)$ is a vector of length n, where the number of initial 1s is precisely x, and $x \geq y$, then*

$$\left|S_{M(\vec{a})}\right| = \frac{x!(n-y)!}{(x-y)!}$$

Proof: Count the number of elements in $S_{M(\vec{a})}$ column by column. How many choices are there for $\alpha^{-1}(1)$? Since $\alpha(i)$ is only allowed to be 1 for $i \leq x$, there are x choices for the first column. Similarly, there are $(x-1)$ choices for $\alpha^{-1}(2)$, and continuing, there are $(x-i+1)$ choices for $\alpha^{-1}(i)$ for $i \leq y$. Thus, the number of choices for the tuple $(\alpha^{-1}(1), \alpha^{-1}(2), \ldots, \alpha^{-1}(y))$ is precisely

$$x \cdot (x-1) \cdots (x-y+1) = \frac{x!}{(x-y)!}$$

Now, consider the number of choices for $\alpha^{-1}(y+1)$. Since $\alpha(i)$ is allowed to be equal to $y+1$ for every single $i \leq n$, and furthermore, $\alpha^{-1}(1), \cdots, \alpha^{-1}(y)$ are already specifiied, there are precisely $n-y$ choices for $\alpha^{-1}(y+1)$, $n-y-1$ choices for $\alpha^{-1}(y+2)$, etc. Thus, the total number of choices for these is precisely $(n-y) \cdot (n-y-1) \cdots 2 \cdot 1 = (n-y)!$. Multiplying these together gives

$$\left|S_{M(\vec{a})}\right| = \frac{x!}{(x-y)!}(n-y)! = \frac{x!(n-y)!}{(x-y)!}$$

as required. □

Theorem 4.10 gives a lower bound on the mixing time. Here, an explicit set A is found such that for t in the above theorem, $|P^t(id, A) - \pi(A)|$ is large. In their paper, Diaconis and Shahshahani use the set of permutations with at least 1 fixed point [14]. The heuristics above suggest trying a modification: clearly, if $f(n) = g(n)$, the fixed points in the top $f(n) \times f(n)$ square of M_n could be used. In general, the number of fixed points in the first $g(n)$ columns works.

Before the proof some notation mainly used for the next two sections is needed. Denote the Markov chain by $(X_k)_{k=0}^{\infty}$, where $X_k = (X_k(1), X_k(2), \ldots, X_k(n))$ and let the random row transposition used at time k be r_k. Clearly, $X_k = X_0 r_1 r_2 \ldots r_k$.

Proof of Theorem 4.10: It must be shown that at time

$$t = \frac{(n+2\Delta)(\log g(n) + \log f(n))}{4f(n)} - c\frac{n+2\Delta}{4f(n)} \tag{4.12}$$

the walk on S_{M_n} has not yet mixed, where $M_n = M(\vec{b}_n(f,g))$. By assumption,

$$\liminf_{n\to\infty} \frac{g(n)}{f(n)} = r > 0.$$

This implies that $\limsup_{n\to\infty}(\log f(n) - \log g(n)) < -\log r$. Thus, for sufficiently large n,

$$t < \frac{(n+2\Delta)\log g(n)}{2f(n)} - (c + \log r)\frac{n+2\Delta}{4f(n)}.$$

Hence it suffices to show that the walk has not mixed at the time on the right-hand side above. Call this time t'.

Define A_n as

$$A_n = \{\sigma \in S_{M_n} \text{ such that } \sigma(i) = i \text{ for at least one } i \leq g(n)\}$$

Call $i \leq g(n)$ such that $\sigma(i) = i$ a *small fixed point.* The first step calculates an upper bound for $\pi(A_n)$ for the uniform distribution π on S_{M_n}. This is done by a standard inclusion-exclusion argument. Consider the number of $\sigma \in S_{M_n}$ such that $\sigma(i) = i$ for a specified value of $i \leq g(n)$. In the language of permutation matrices, to pick such a σ, mark the 1 in position (i,i), cross out the ith row and ith column, and select the values of σ on the remaining rows and columns. It is easy to check that this corresponds to picking an element in $S_{M(\vec{a})}$, for $\vec{a} = (1, 1, \ldots, 1, g(n), \ldots, g(n))$ where $\vec{a}$ starts with $f(n)-1$ ones and the vector is of length $n-1$. Thus, from Lemma 4.13 above,

$$|\{\sigma \in S_{M_n} \mid \sigma(i) = i\}| = \left|S_{M(\vec{a})}\right| = \frac{(f(n)-1)!(n-g(n))!}{(f(n)-g(n))!}$$

From the same lemma, $S_{M_n} = \frac{f(n)!(n-g(n))!}{(f(n)-g(n))!}$. Thus,

$$|\{\sigma \in S_{M_n} \mid \sigma(i) = i\}| = \frac{|S_{M_n}|}{f(n)}.$$

In a similar way, given k distinct values of $i_1, i_2, \ldots, i_k$ in $[1, g(n)]$, choose σ satisfying

$\sigma(i_j) = i_j$ for all j by crossing out all the rows i_j and columns i_j. This would correspond to the vector $\vec{a}' = (1, 1, \ldots, 1, g(n)+1-k, \ldots, g(n)+1-k)$ of length $n-j$, and thus

$$\begin{aligned} |\{\sigma \in S_{M_n} \mid \sigma(i_1) = i_1, \ldots, \sigma(i_k) = i_k\}| &= \frac{(f(n)-k)!(n-g(n))!}{(f(n)-g(n))!} \\ &= |S_{M_n}| \frac{f(n)!}{(f(n)-k)!} \end{aligned}$$

Picking an element σ that has precisely k small fixed points involves choosing the values $i_1, i_2, \ldots, i_k$ above. Thus,

$$\begin{aligned} |\{\sigma \in S_{M_n} \mid \sigma \text{ has } k \text{ small fixed points}\}| &= \binom{g(n)}{k} |S_{M_n}| \frac{f(n)!}{(f(n)-k)!} \\ &= |S_{M_n}| \frac{g(n)(g(n)-1)\cdots(g(n)-k+1)}{k! f(n)(f(n)-1)\cdots(f(n)-k+1)} \end{aligned}$$

Now, the standard inclusion-exclusion argument gives

$$|A_n| = |S_{M_n}| \left(\frac{g(n)}{f(n)} - \frac{g(n)(g(n)-1)}{2f(n)(f(n)-1)} + \frac{g(n)(g(n)-1)(g(n)-2)}{6f(n)(f(n)-1)(f(n)-2)} - \ldots \right)$$

From $f(n) \to \infty$, and $\liminf \frac{g(n)}{f(n)} > 0$, $g(n) \to \infty$. Thus, the above sum becomes arbitrarily well-approximated by

$$|S_{M_n}| \left(\frac{g(n)}{f(n)} - \frac{g(n)^2}{2! f(n)^2} + \frac{g(n)^3}{3! f(n)^3} - \ldots \right) = |S_{M_n}| (1 - e^{-g(n)/f(n)})$$

and thus $\pi(A_n)$ is arbitrarily well-approximated by $1 - e^{-g(n)/f(n)}$. Therefore, noting that $1 - e^{-x}$ is an increasing function, and that $\frac{g(n)}{f(n)} \le 1$,

$$\limsup \pi(A_n) \le 1 - \frac{1}{e}.$$

Now, consider the probability of A_n after t' steps of the random walk, started at the identity. Define F_k to be the number of small fixed points of our random walk at time k: that is, the number of rows $i \le g(n)$ such that $X_k(i) = i$. Consider the distribution of F_k. Note that since the walk starts at the identity, any $i \le g(n)$ that has not yet been tranposed with anything is in F_t. Furthermore, if $X_k(i) \le g(n)$, row i can only be transposed with row j if $j \le f(n)$ – that is, a row corresponding to one of the first $g(n)$ columns can only be transposed with one of the first $f(n)$ rows. Otherwise, $X_{k+1}(j) = X_k(i) \leq g(n)$, which is not allowed for $j > f(n)$.

From the arguments above, $F_0 = \{1, 2, \ldots, g(n)\}$, and furthermore, row i can leave the set F_{k-1} only by being transposed with a row that is at most $f(n)$. Consider two cases – first, the next step could transpose two rows which are both at most $g(n)$. This has probability

$$\mathbb{P}\left(r_k = (ij) \mid i, j \leq g(n)\right) = \frac{g(n)^2}{n + 2\Delta}.$$

Secondly, the next step could transpose a pair of rows one of which is below $g(n)$ and the other one is between $g(n)$ and $f(n)$. This has probability

$$\mathbb{P}\left(r_k = (ij) \mid i \leq g(n) < j \leq f(n) \text{ or vice versa}\right) = \frac{2g(n)(f(n) - g(n))}{n + 2\Delta}.$$

The argument proceeds by estimating the probability of having transposed each element from $\{1, 2, \ldots, g(n)\}$ at time t'. Note that if two rows below $g(n)$ are transposed at time k, then they both leave the set F_{k-1}, whereas the second case above corresponds to only one row leaving the set. This rephrases the question as the following coupon collectors problem: at each step, collect two coupons with probability $\frac{g(n)^2}{n+2\Delta}$ and one coupon with probability $\frac{2g(n)(f(n)-g(n))}{n+2\Delta}$. What is the probability of not collecting every coupon from $\{1, 2, \ldots, g(n)\}$ by time t'?

Begin by calculating the mean number of coupons collected at each step, counting each coupon however many times it's collected. This number is clearly

$$\begin{aligned}
\mathbb{E}\left(\text{coupons collected in one step}\right) &= 2 \cdot \frac{g(n)^2}{n + 2\Delta} + 1 \cdot \frac{2g(n)(f(n) - g(n))}{n + 2\Delta} \\
&= \frac{2f(n)g(n)}{n + 2\Delta}
\end{aligned}$$

Similarly, the variance of the number of coupons collected in one step is bounded above by $\frac{4f(n)g(n)}{n+2\Delta}$. Thus, after t' total steps of the walk, the total number of coupons collected is concentrated around

$$\begin{aligned}
t' \cdot \frac{2f(n)g(n)}{n + 2\Delta} &= \left(\frac{(n + 2\Delta)\log g(n)}{2f(n)} - (c + \log r)\frac{n + 2\Delta}{2f(n)}\right) \cdot \frac{2f(n)g(n)}{n + 2\Delta} \\
&= g(n)\log g(n) - (c + \log r)g(n)
\end{aligned}$$

with a window of at most $\sqrt{2g(n)\log g(n)}$. Since $g(n) \to \infty$, and since $\sqrt{2g(n)\log g(n)}$ is $o(g(n))$, using standard coupon collector arguments [18] we can conclude that

$$P^{t'}(id, A_n) \to 1 - e^{-e^{c+\log r}} = 1 - e^{-re^c}$$

as $n \to \infty$.

Therefore,

$$\begin{aligned}\liminf_{n\to\infty} \left\| P^{t'}(id,\cdot) - \pi \right\|_{TV} &\geq \liminf_{n\to\infty} \left(P^{t'}(id, A_n) - \pi(A_n) \right) \\ &\geq \left(1 - e^{-re^c}\right) - \left(1 - \frac{1}{e}\right) \\ &= \frac{1}{e} - e^{-re^c},\end{aligned}$$

the desired inequality for t'. As noted earlier, since for sufficiently large n, $t \leq t'$, and since distance to stationarity is non-decreasing,

$$\liminf_{n\to\infty} \left\| P^{t}(id,\cdot) - \pi \right\|_{TV} \geq \frac{1}{e} - e^{-re^c}$$

and so the walk hasn't mixed by time t, as required. $\square$

4.4 Vertex Transitivity and Example of Fast Mixing

This section contains the proof of some of the total variation results, in particular Theorem 4.11. Here, $g(n) = 1$, which means that $\vec{b}_n(f,g) = (1,\ldots,1,2,\ldots,2)$. Hence $\sigma \in S_{M_n}$ has the following restrictions: it's allowed to be at most $f(n)$ on column 1, and has no restrictions at all on the remaining columns. It is easy to calculate that the probability of the first column being tranposed in a particular step is

$$P(\text{Column 1 tranposed at step } k) = \frac{2f(n)-1}{n+2\Delta} \approx \frac{2f(n)}{n^2}$$

using the value for Δ derived in Lemma 4.7.

Now, since $f(n) \leq \frac{n}{5\log n}$ for sufficiently large n, the first column gets tranposed at most every $\frac{5}{2}n\log n$ steps or so. Furthermore, the remaining $n-1$ columns have no restrictions, and hence the walk without the first column is just the walk on S_{n-1}. Since this walk has cutoff at time $\frac{1}{2}(n-1)\log(n-1)$, these $n-1$ columns should be thoroughly mixed by the time the first column is used at all. Therefore, the walk should be mixed as soon as the first column is used. Furthermore, since mixing is driven by one column, a cut-off is not expected.

In order to simplify calculations, it is first shown that for any $\vec{b} = \vec{b}_n(f,g)$, $S_{M(\vec{b})}$ is vertex transitive. This will simplify the proof of Theorem 4.11, as the walk may be started at the identity. Furthermore, this will also be useful later for spectral analysis, as Equation (4.8) will be used.

Lemma 4.14. *For $M_n = M(\vec{b}_n(f,g))$, the graph induced on S_{M_n} by the random transposition walk is vertex transitive.*

Note: This chapter uses a different convention from Hanlon for permutation multiplication: if α and β are permutations, multiplication is treated as function composition. That is,

$$(\alpha\beta)(i) = \alpha(\beta(i))$$

Proof: Let π and σ be elements of S_{M_n}. A graph isomorphism $\phi : S_{M_n} \to S_{M_n}$ is found such that

$$\phi(\pi) = \sigma.$$

Only graph isomorphisms of the form

$$\phi(\tau) = \alpha\tau\beta \tag{4.13}$$

are considered, where α is a permutation in S_n that only acts non-trivially on the elements $\{g(n)+1, g(n)+2, \ldots, n\}$ and β is an elements of S_n that only acts non-trivially on the elements $\{1, 2, \ldots, f(n)\}$. The proof consists of the following steps. First, it is shown that any ϕ as defined in (4.13) maps S_{M_n} to itself. Secondly, it is shown that any such ϕ is an isomorphism. Finally, α and β are found such that the ϕ defined above satisfies

$$\phi(\pi) = \sigma$$

Step 1. Note that visually, α rotates the last $n - g(n)$ columns of the restriction matrix, while β rotates the first $f(n)$ rows of the restriction matrix; this formulation makes it clear that ϕ must map S_{M_n} to itself. To prove this formally, use the definition

$$S_{M_n} = \{\tau \in S_n \mid \tau(i) \geq g(n) + 1 \text{ for all } i \geq f(n) + 1\}.$$

Thus, it must be shown that for $i \geq f(n) + 1$,

$$\alpha\tau\beta(i) \geq g(n) + 1.$$

But since β only acts on $\{1, 2, \ldots, f(n)\}$, for $i \geq f(n) + 1$, $\beta(i) = i$. Thus, $\alpha\tau\beta(i) = \alpha\tau(i)$. Since $\tau \in S_{M_n}$, and $i \geq f(n) + 1$, $\tau(i) \geq g(n) + 1$. Since α only acts on $\{g(n) + 1, g(n) + 2, \ldots, n\}$, $\alpha\tau(i) = \alpha(\tau(i)) \geq g(n) + 1$. Thus,

$$\alpha\tau\beta(i) = \alpha\tau(i) \geq g(n) + 1$$

as desired.

Step 2. It now must be shown that any ϕ defined by (4.13) is an isomorphism. By Step 1, ϕ maps into S_{M_n}, and it's obviously invertible, so it suffices to show that it preserves edges. That is, for any transposition (i, j) and $\tau \in S_{M_n}$, such that $(i, j)\tau \in S_{M_n}$, $\phi\left((i, j)\tau\right)$ is a neighbor of $\phi(\tau)$. But

$$\begin{aligned} \phi((i, j)\tau) &= \alpha(i, j)\tau\beta = (\alpha(i), \alpha(j))\alpha\tau\beta \\ &= (\alpha(i), \alpha(j))\phi(\tau) \end{aligned}$$

which is clearly a transposition away from $\phi(\tau)$. Thus, the map ϕ preserves edges, and hence is an isomorphism of S_{M_n}.

Step 3. Finally, find α and β such that

$$\phi(\pi) = \sigma$$

This entails $\alpha\pi\beta = \sigma$. Since β acts only on $\{1, 2, \ldots, f(n)\}$, this means that for $i > f(n)$,

$$\begin{aligned} \alpha\pi\beta(i) &= \sigma(i) \\ \Rightarrow \alpha\pi(i) &= \sigma(i) \end{aligned}$$

Thus, α must satisfy

$$\alpha\pi(i) = \sigma(i) \text{ for } i > f(n) \tag{4.14}$$

This defines α on $\pi(S)$, where $S = \{i > f(n)\}$. Since $\pi \in S_{M_n}$, for all $i > f(n)$, $\pi(i) > g(n)$. Thus, $\pi(S) \subset \{g(n)+1, g(n)+2, \ldots, n\}$. Furthermore, for $i > f(n), \sigma(i) > g(n)$: therefore, Equation (4.14) says that α must send a subset of $\{g(n)+1, g(n)+2, \ldots, n\}$ to some other subset of $\{g(n)+1, g(n)+2, \ldots, n\}$. Therefore, it is possible to pick a permutation that acts only on $\{g(n)+1, g(n)+2, \ldots, n\}$ and satisfies Equation (4.14): call this permutation α_0.

Now, define $\beta_0 = \pi^{-1}\alpha_0^{-1}\sigma$. Clearly, with this definition,

$$\phi(\pi) = \alpha_0\pi\beta_0 = \sigma$$

To check that this β_0 acts only on $\{1, 2, \ldots, f(n)\}$, it suffices to show this for β_0^{-1}. Let $i > f(n)$:

$$\beta_0^{-1}(i) = \sigma^{-1}\alpha_0\pi(i) = \sigma^{-1}(\sigma(i)) = i$$

where the second equality follows by Equation (4.14), which was used to define α_0. Thus, β_0^{-1}, and hence β_0, fixes $\{f(n)+1, f(n)+2, \ldots, n\}$, so β_0 acts only on $\{1, 2, \ldots, f(n)\}$. Therefore, for this choice of α_0 and β_0, $\phi(\tau) = \alpha_0\tau\beta_0$ is an isomorphism of S_{M_n} which maps π to σ, as required. □

Remark 4.15. Unfortunately, S_M is not vertex transitive for all one-sided interval restriction matrices. While I have not discovered an easy characterization for vertex transitivity, it is easy to provide counterexamples. For example, let

$$M = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

Both $\sigma = 12345$ and $\tau = 45123$ are in S_M. Furthermore, a simple computer calculation shows that

$$P^6(\sigma,\sigma) = \frac{5207}{117649} \neq P^6(\tau,\tau) = \frac{5287}{117649}$$

implying that the walk is not vertex transitive. I conjecture that S_M is not vertex transitive for almost all one-sided restriction matrices that are not 'two-step.'

Return to proving Theorem 4.11, following the outline at the beginning of this section.

Lemma 4.16. *Let T be the first time the first column is used in the walk. Furthermore, fix ϵ and assume that i is chosen such that at time $i-1$, the random tranposition walk on S_{n-1} is within ϵ in total variation distance from uniformity. Then, if π is the uniform distribution on S_{M_n} for $t \geq i$,*

$$\frac{1}{2}\sum_{x\in S_{M_n}} |\mathbb{P}\left(X_t = x \mid T = i\right) - \pi(x)| < \frac{4}{2f(n)-1} + \epsilon.$$

Thus, the conditional distribution of X_t given $T = i$ is very close in total variation to the uniform distribution.

Proof: From Lemma 4.14, the walk is vertex transitive. Without loss of generality, the walk starts at the identity. Now, at times $t \geq i$, the walk simply evolves as usual, since the event conditioned on happened already. Since the total variation distance of a Markov chain to stationarity is non-decreasing, for $t \geq i$,

$$\frac{1}{2}\sum_{x\in S_{M_n}} |\mathbb{P}\left(X_t = x \mid T = i\right) - \pi(x)| \leq \frac{1}{2}\sum_{x\in S_{M_n}} |\mathbb{P}\left(X_i = x \mid T = i\right) - \pi(x)|$$

Clearly, the expression on the right-hand side above is the just total variation distance between X_i conditioned on $T = i$ and stationarity. Thus, it is equal to

$$\sup_{B \subseteq S_{M_n}} |\mathbb{P}\left(X_i \in B \mid T = i\right) - \pi(B)|$$

and that is precisely what will be bounded. Accordingly, fix a subset B of S_{M_n}.

By definition, until time i the walk is precisely the random transposition walk on the permutations of $\{2, 3, \cdots, n\}$, and hence it is identical to the random transposition walk on S_{n-1}. Recall that i is chosen such that this walk is within ϵ of stationarity at time $i - 1$. Now, let

$$S = \{\sigma \in S_{M_n} \mid \sigma(1) = 1\} \cong S_{n-1} \tag{4.15}$$

and let $\tilde{\pi}$ is the uniform distribution on S. For any $B \subseteq S$,

$$|P(X_{i-1} \in B \mid T = i) - \tilde{\pi}(B)| < \epsilon.$$

Now consider the transposition r_i which takes X_{i-1} to X_i. Conditioning on $T = i$, this transposition is uniformly distributed between all the transpositions involving the first row (as the walk starts at the identity, this is equivalent to using the first column.) Note that among these, the transposition $(1, 1)$ appears once, while the remaining transpositions $(1, k)$ appear twice. This means that there are precisely $2f(n) - 1$ possible choices for r_i. Now, r_i determines the value of X_i on column 1. Accordingly, write $B = B_1 \cup B_2 \cup \cdots \cup B_{f(n)}$, where

$$B_k = \{\sigma \in B \mid \sigma(k) = 1\}$$

This decomposes B into equivalence classes that depend on the value of $\sigma \in B$ on column 1. Since $X_{i-1}(1) = 1$, if $r_i = (1, j)$, then $X_i(j) = 1$. Therefore, X_i is in B_k if and only if r_i is $(1, k)$ – that is, if at time i rows 1 and k are transposed. Hence,

$$\begin{aligned}
\mathbb{P}(X_i \in B_k \mid T = i) &= \mathbb{P}(X_{i-1}(1, k) \in B_k \mid T = i)\mathbb{P}(r_i = (1, k)) \\
&= \mathbb{P}(X_{i-1} \in B_k(1, k) \mid T = i)\mathbb{P}(r_i = (1, k))
\end{aligned}$$

As noted earlier, there are precisely $2f(n) - 1$ possibilities for r_i, and the cases $k = 1$ and

$k \neq 1$ are slightly different. Consider those separately. For $k = 1$,

$$\begin{aligned}\mathbb{P}(X_i \in B_1 \,|\, T = i) &= \mathbb{P}(X_{i-1} \in B_1 \,|\, T = i)\mathbb{P}(r_i = (1,1)) \\ &= \frac{\mathbb{P}(X_{i-1} \in B_1 \,|\, T = i)}{2f(n) - 1} \leq \frac{1}{2f(n) - 1}\end{aligned} \tag{4.16}$$

and for $k \neq 1$,

$$\begin{aligned}\mathbb{P}(X_i \in B_k \,|\, T = i) &= \mathbb{P}(X_{i-1} \in B_k(1,k) \,|\, T = i)\mathbb{P}(r_i = (1,k)) \\ &= \frac{2\mathbb{P}(X_{i-1} \in B_k(1,k) \,|\, T = i)}{2f(n) - 1}\end{aligned} \tag{4.17}$$

By choice of i, $\mathbb{P}(X_{i-1} \in B_k(1,k) \,|\, T = i)$ is well-approximated by $\tilde{\pi}(B_k(1,k))$, where $\tilde{\pi}$ is the uniform distribution on S as defined above in Equation (4.15). It is easy to see that $\tilde{\pi}(B_k(1,k)) = f(n)\pi(B_k(1,k)) = f(n)\pi(B_k)$, and thus that

$$|\mathbb{P}(X_{i-1} \in B_k(1,k) \,|\, T = i) - f(n)\pi(B_k)| < \epsilon$$

for each k. Therefore, for $k \neq 1$, using Equation (4.17),

$$\begin{aligned}|\mathbb{P}(X_i \in B_k \,|\, T = i) - \pi(B_k)| &= \left| \frac{2\mathbb{P}(X_{i-1} \in B_k(1,k) \,|\, T = i)}{2f(n) - 1} - \pi(B_k) \right| \\ &\leq \frac{2\,|\mathbb{P}(X_{i-1} \in B_k(1,k) \,|\, T = i) - f(n)\pi(B_k)| + \pi(B_k)}{2f(n) - 1} \\ &\leq \frac{2\epsilon + \pi(B_k)}{2f(n) - 1}\end{aligned}$$

Combining the above with Equation (4.16),

$$\begin{aligned}|\mathbb{P}(X_i \in B \,|\, T = i) - \pi(B)| &= \left| \sum_{k=1}^{f(n)} (\mathbb{P}(X_i \in B_k \,|\, T = i) - \pi(B_k)) \right| \\ &\leq \mathbb{P}(X_i \in B_1 \,|\, T = i) + \pi(B_1) + \sum_{k=2}^{f(n)} \left(\frac{2\epsilon + \pi(B_k)}{2f(n) - 1} \right) \\ &\leq \frac{1}{2f(n) - 1} + \pi(B_1) + \epsilon + \frac{1}{2f(n) - 1} \sum_{k=2}^{f(n)} \pi(B_k)\end{aligned} \tag{4.18}$$

Now, since $B_1 \subseteq S$, and $\pi(S)$ is clearly $\frac{1}{f(n)}$, $\pi(B_1) < \frac{2}{2f(n)-1}$. Also,

$$\sum_{k=2}^{f(n)} \pi(B_k) \leq \pi(B) \leq 1$$

Plugging these back into Equation (4.18),

$$|\mathbb{P}(X_i \in B \mid T = i) - \pi(B)| \leq \frac{4}{2f(n) - 1} + \epsilon$$

and hence

$$\sup_{B \subseteq S_{M_n}} |\mathbb{P}\left(X_i \in B \mid T = i\right) - \pi(B)| < \frac{4}{2f(n) - 1} + \epsilon$$

as required. □

Turn now to the proof of Theorem 4.11. As above, condition on the first time the first column is transposed.

Proof of Theorem 4.11. Without loss of generality, assume that the walk starts at the identity. Define

$$t = \frac{3(n + 2\Delta)}{2f(n) - 1}. \tag{4.19}$$

It must be shown that the walk is sufficiently mixed by time t.

Let T is the first time that column 1 is used in a transposition. Then,

$$\begin{aligned}
\left\|P^t(id, \cdot) - \pi\right\|_{TV} &= \frac{1}{2} \sum_x \left|P^t(id, x) - \pi(x)\right| \\
&= \frac{1}{2} \sum_x \left| \sum_{i=1}^{\infty} \mathbb{P}(X_t = x \mid T = i)\mathbb{P}(T = i) - \pi(x) \right| \\
&= \frac{1}{2} \sum_x \left| \sum_{i=1}^{\infty} \mathbb{P}(X_t = x \mid T = i)\mathbb{P}(T = i) - \sum_{i=1}^{\infty} \pi(x)\mathbb{P}(T = i) \right| \\
&\leq \sum_{i=1}^{\infty} \mathbb{P}(T = i) \frac{1}{2} \sum_x |\mathbb{P}(X_t = x \mid T = i) - \pi(x)| \qquad (4.20)
\end{aligned}$$

Now, from the results of Diaconis and Shahshahani, the random transposition walk on S_{n-1} has cut-off at time $\frac{1}{2}(n-1)\log(n-1)$ with a window of size $n-1$. Fix $\epsilon > 0$. If $i \geq n \log n$, then for sufficiently large n, the total variation distance between the random transposition walk on S_{n-1} at time i and the uniform distribution is less than ϵ. Thus, by Lemma 4.16 above, if $i \geq n \log n$ and n is sufficiently large, then for $t \geq i$,

$$\frac{1}{2} \sum_{x \in S_{M_n}} \left| \mathbb{P}\left(X_t = x \mid T = i\right) - \pi(x) \right| < \frac{4}{2f(n)-1} + \epsilon \tag{4.21}$$

Now, using Equation (4.20), since the total variation distance between distributions is always at most 1, $\left\| P^t(id, \cdot) - \pi \right\|$ is bounded above by

$$\sum_{i=n\log n}^{t} \mathbb{P}(T=i) \frac{1}{2} \sum_{x} |\mathbb{P}(X_t = x \mid T = i) - \pi(x)| + \mathbb{P}(T \notin [n \log n, t]) \tag{4.22}$$

The probability of transposing the first column in one step is precisely $\frac{2f(n)-1}{n+2\Delta}$. Thus,

$$\begin{aligned} \mathbb{P}(T \leq n \log n) &= 1 - \mathbb{P}(T > n \log n) \\ &= 1 - \left(1 - \frac{2f(n)-1}{n+2\Delta}\right)^{n \log n} \end{aligned}$$

Since $f(n) \leq \frac{n}{5 \log n}$ for sufficiently large n, and since $n + 2\Delta \approx n^2$,

$$1 - \left(1 - \frac{2f(n)-1}{n+2\Delta}\right)^{n \log n} \leq 1 - e^{-1/2}$$

for sufficiently large n. Similarly, the probability that $T > t$ is just

$$\left(1 - \frac{2f(n)-1}{n+2\Delta}\right)^{t} \leq \exp\left(-\frac{t(2f(n)-1)}{n+2\Delta}\right) = e^{-3}$$

and thus $\mathbb{P}\left(T \notin [n \log n, t]\right) \leq 1 - e^{-1/2} + e^{-3} < 0.45$. Next, by Equations (4.22) and (4.21), for sufficiently large n,

$$\begin{aligned} \left\| P^t(id, \cdot) - \pi \right\|_{TV} &\leq \sum_{i=n\log n}^{t} \mathbb{P}(T=i) \frac{1}{2} \sum_{x} |\mathbb{P}(\alpha_t = x \mid T = i) - \pi(x)| + 0.45 \\ &\leq \sum_{i=n\log n}^{t} \mathbb{P}(T=i) \left(\frac{4}{2f(n)-1} + \epsilon\right) + 0.45 \\ &\leq \frac{4}{2f(n)-1} + \epsilon + 0.45 \end{aligned}$$

Since ϵ can be chosen to be anything, and $f(n) \to \infty$, the above is clearly less than $\frac{1}{2}$ for sufficiently large n. Thus, the walk has mixed by time $t = \frac{3(n+2\Delta)}{2f(n)-1}$, which is clearly of order $\frac{n+2\Delta}{f(n)}$.

Finally, the walk does not have cut-off: starting at the identity,

$$\mathbb{P}\left(X_k(1) = 1\right) \geq \mathbb{P}(T > k) = \left(1 - \frac{2f(n)-1}{n+2\Delta}\right)^k$$

Under the uniform distribution π, the probability of the set $S = \{\sigma \in S_{M_n} \mid \sigma(1) = 1\}$ is precisely $\frac{1}{f(n)}$. This means that

$$\left\|P^k(id, \cdot) - \pi\right\|_{TV} \geq \left(1 - \frac{2f(n)-1}{n+2\Delta}\right)^k - \frac{1}{f(n)}$$

It is easy to see that the above function falls off smoothly as opposed to exhibiting cut-off. For example, at time $2t = \frac{6(n+2\Delta)}{2f(n)-1}$, it is approximately e^{-6}, which does not approach 0 as $n \to \infty$. Thus, the walk does not have cut-off, completing the proof. $\square$

4.5 Hanlon's Results and Other Preliminaries

The remainder of this chapter focuses on proving Theorem 4.9, which is concerned with chi-squared cut-off. This section contains a review of Hanlon's work and other background material, which shows how to diagonalize the adjacency matrix for $S_{M(\vec{b})}$. The following is Definition 4.2 from his paper [22]:

Definition 4.17. *Let $\vec{b} = (b_1, \ldots, b_n)$ be a sequence chosen from $\{1, 2, \ldots, n\}$ satisfying $b_1 \leq b_2 \leq \cdots \leq b_n$. Call u and v left-equivalent if $b_u = b_v$. Let $L_1, L_2, \ldots, L_s$ denote the left-equivalence classes of $\{1, 2, \ldots, n\}$, where the ordering is chosen so that whenever $i < j$, the elements of L_i are less than the elements of L_j.*

For notational simplicity, define $b_{n+1} = n + 1$, and call u and v right-equivalent if there exist i and $i + 1$ such that $b_i \leq u, v \leq b_{i+1}$. It is straightforward to check that the number of left-equivalence classes is equal to the number of right-equivalence classes. Let $R_1, \ldots, R_s$ denote the right-equivalence classes of $\{1, 2, \ldots, n\}$, where the R_i are ordered in the same way as the L_i above.

Example 4.18. Left-equivalence and right-equivalence classes are very easy to visualize via the restriction matrix $M(\vec{b})$. For example, let $\vec{b} = (1, 1, 1, 2, 4)$. Then, the corresponding

$M(\vec{b})$ is below:

$$M(\vec{b}) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Now, imagine the restriction matrix as an $n \times n$ chessboard, where each square contains either a 0 or a 1. Shade in the squares that contain 1s, and look at the southwest boundary of the shaded area. This has a 'step pattern': in the above example, go down 3 steps, right 1 step, down 1 step, right 2 steps, down 1 step, right 2 steps. It is easy to check that the left-equivalence classes correspond to the down stretches of the step pattern, and the right-equivalence classes correspond to the stretches pointing right.

Using this, for $\vec{b}$ as defined above, the left-equivalence classes are $\{1, 2, 3\}, \{4\}, \{5\}$, and the right-equivalence classes are $\{1\}, \{2, 3\}, \{4, 5\}$. (This visualization also makes it clear why the numbers of left-equivalence and right-equivalence classes match.)

Remark 4.19. It should be clear that if $\vec{b}_n(f, g)$ is defined as in Definition 4.5, then M_n is a 'two-step' restriction matrix: that is, the southwest boundary described above will go down for $f(n)$ steps, then will go right for $g(n)$ steps, then will go down for $n - f(n)$ steps, then will go right for $n - g(n)$ steps. Thus,

$$L_1 = \{1, 2, \ldots, f(n)\}, L_2 = \{f(n) + 1, \ldots, n\} \tag{4.23}$$

and

$$R_1 = \{1, 2, \ldots, g(n)\}, L_2 = \{g(n) + 1, \ldots, n\} \tag{4.24}$$

For example, if $n = 5$, $f(5) = 3$ and $g(5) = 2$, then $\vec{b} = (1, 1, 1, 3, 3)$ as in Equation (4.5) above. From the $M(\vec{b})$ below it, it's clear that the left-equivalence classes are $\{1, 2, 3\}$ and $\{4, 5\}$, while the right-equivalence classes are $\{1, 2\}$ and $\{3, 4, 5\}$ matching the expressions above.

The following is Definition 4.3 from Hanlon [22].

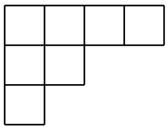
Definition 4.20. *A $\vec{b}$-partition $\alpha = (\lambda_1, \mu_1, \lambda_2, \ldots, \mu_{s-1}, \lambda_s)$ is a sequence of partitions such that*

1. $\lambda_i \supseteq \mu_i \subseteq \lambda_{i+1}$ *for all* $i = 1, 2, \ldots, s - 1$

2. $|\lambda_i \setminus \mu_i| = |R_i|$ *for all* $i = 1, 2, \ldots, s$ *(defining* $\mu_s = \emptyset$*)*

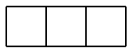
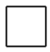
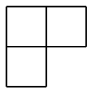
3. $|\lambda_{i+1} \setminus \mu_i| = |L_{i+1}|$ *for all* $i = 0, 1, \ldots, s-1$ *(defining* $\mu_0 = \emptyset$*)*

It is helpful to have some standard terminology (a good reference for this is Stanley [49]). Recall that a *Ferrers diagram* is a finite collection of boxes, arranged in left-justified rows, such that each row has at least as many boxes as the row directly below it. If $(a_1, a_2, \ldots, a_m)$ is a partition, then the Ferrers diagram associated to it has precisely a_i boxes in row i. For example, below is the Ferrers diagram of the partition $(4, 2, 1)$ of 7:



The *transpose* λ^T of a partition λ is a partition whose Ferrers diagram is a reflection of the Ferrers diagram of λ along the main diagonal $y = -x$. For example, for $\lambda = (4, 2, 1)$, the tranpose partition λ^T is $(3, 2, 1, 1)$.

Ferrers diagrams give a good way of visualizing $\vec{b}$-partitions. If the left-equivalence classes are $L_1, L_2, \ldots, L_s$ and the right-equivalence classes are $R_1, R_2, \ldots, R_s$, then to get a $\vec{b}$-partition start with $\mu_0 = \emptyset$, add $|L_1|$ squares to get to λ_1, delete $|R_1|$ squares to get μ_1, add $|L_2|$ squares to get λ_2, etc. For example, if $\vec{b} = (1, 1, 1, 3, 3)$, with left-equivalence classes $\{1, 2, 3\}$ and $\{4, 5\}$, and right-equivalence classes $\{1, 2\}$ and $\{3, 4, 5\}$, then the following is a $\vec{b}$-partition:

$\mu_0 = \emptyset$, $\lambda_1 =$ , $\mu_1 =$ , $\lambda_2 =$ , $\mu_2 = \emptyset$

A labeling of a Ferrers diagram is a Ferrers diagram with numbers filled into the boxes. A *standard Young tableau* is a labeling in which the entries of each row are strictly increasing from left to right, while the entries of each column are strictly increasing from top to bottom. A *semi-standard Young tableau* has strictly increasing columns, but weakly increasing (non-decreasing) rows. The following is a semi-standard Young tableau for the partition $(4, 2, 1)$ of 7:

1	1	2	3
2	4		
3			

Definition 4.21. *Let α and β be partitions satisfying $\alpha_i \geq \beta_i$ for all i. Let $n = |\alpha| - |\beta|$. Then, the skew-shape α/β is well-defined, and there exists a representation of S_n corresponding to α/β. Denote the degree of this representation by $|X_{\alpha/\beta}|$.*

Remark 4.22. It is well-known that $|X_{\alpha/\beta}|$ is the number of standard Young tableaux of shape α/β (a good source for this and results like it is Macdonald's book [38]).

Definition 4.23. *Let $\alpha = (\lambda_1, \mu_1, \lambda_2, \ldots, \lambda_s)$ be a $\vec{b}$-partition. Define the indicator tableau $T(\alpha)$ to be the tableau whose entry in a square x is the number of skew shapes λ_i/μ_{i-1} that contain x. Denote this entry by $T_x(\alpha)$.*

Now, recall that the *content* of a square x in a Ferrers diagram is denoted by c_x, and is defined to be $j - i$, where x lies in column j from the left, and row i from the top. The following theorem restates a main result of Hanlon's paper, Theorem 4.15, which derives all the eigenvalues of the adjacency matrix U using the language indicated above.

Theorem 4.24. *For every $\vec{b}$-partition $\alpha = (\lambda_1, \mu_1, \lambda_2, \ldots, \lambda_s)$, there is an eigenspace $R_n(\alpha)$ of $U(\vec{b})$ such that the vector space $\mathbb{C}S_{M(\vec{b})}$ decomposes as a direct sum of the spaces $R_n(\alpha)$; that is,*

$$\mathbb{C}S_{M(\vec{b})} = \bigoplus_{\alpha} R_n(\alpha)$$

where the sum runs over all $\vec{b}$-partitions α. The dimension of $R_n(\alpha)$ is equal to

$$\prod_{i=0}^{s} |X_{\lambda_i/\mu_i}|\,|X_{\lambda_{i+1}/\mu_i}| \tag{4.25}$$

where $\mu_0 = \mu_s = \emptyset$.

Furthermore, letting the eigenvalue of U corresponding to $R_n(\alpha)$ be $\Lambda(\alpha)$,

$$\Lambda(\alpha) = \sum_{x} T_x(\alpha) c_x \tag{4.26}$$

where the sum is over all the squares x in the indicator tableau $T(\alpha)$.

The following lemma finds another expression for the eigenvalue $\Lambda(\alpha)$. It requires the following definition:

Definition 4.25. *If λ is a partition, define $C(\lambda)$ to be the sum of the contents of all the squares of λ.*

Lemma 4.26. *If* $\alpha = (\lambda_1, \mu_1, \ldots, \mu_{s-1}, \lambda_s)$ *is a* $\vec{b}$*-partition,*

$$\begin{aligned}\Lambda(\alpha) &= C(\lambda_1) - C(\mu_1) + \cdots - C(\mu_{s-1}) + C(\lambda_2) \\ &= \sum_{i=1}^{s} C(\lambda_i) - \sum_{i=1}^{s-1} C(\mu_i) \end{aligned} \tag{4.27}$$

Proof: From Equation (4.26),

$$\Lambda(\alpha) = \sum_x T_x(\alpha) c_x$$

where the sum is over all the squares x in the indicator tableau $T(\alpha)$. Recall that $T_x(\alpha)$ is the number of skew shapes λ_i/μ_{i-1} that contain x. Thus, since $C(\mu_0) = 0$,

$$\begin{aligned}\sum_{i=1}^{s} C(\lambda_i) - \sum_{i=1}^{s-1} C(\mu_i) &= \sum_{i=1}^{s} \sum_{x \in \lambda_i} c_x - \sum_{i=0}^{s-1} \sum_{x \in \mu_i} c_x \\ &= \sum_{i=1}^{s} \left(\sum_{x \in \lambda_i} c_x - \sum_{x \in \mu_{i-1}} c_x \right) \\ &= \sum_{i=1}^{s} \sum_{x \in \lambda_i/\mu_{i-1}} c_x = \sum_{i=1}^{s} \sum_{x \in T(\beta)} c_x \mathbf{1}_{\{x \in \lambda_i/\mu_{i-1}\}} \\ &= \sum_{x \in T(\beta)} c_x \sum_{i=1}^{s} \mathbf{1}_{\{x \in \lambda_i/\mu_{i-1}\}} = \sum_{x \in T(\beta)} c_x T_x(\alpha) \\ &= \Lambda(\alpha)\end{aligned}$$

as required. □

To diagonalize the transition matrix P, make the following simple definition.

Definition 4.27. *Define*

$$\Lambda_1(\alpha) = \frac{n + 2\Lambda(\alpha)}{n + 2\Delta} \tag{4.28}$$

From Equation (4.3)*,*

$$P = \frac{1}{n + 2\Delta}(nI + 2U),$$

so $\Lambda_1(\alpha)$ *are the eigenvalues of* P*.*

4.6 Bounding the Eigenvalues

Lemma 4.14 above shows that the graph on $S_{M(\vec{b})}$ induced by the walk is vertex transitive. Thus, Theorem 4.24 and Equation (4.8) combine to show that

$$\left\|P^t(x,\cdot)-\pi\right\|_{2,\pi} = \sqrt{\sum_{\alpha} \dim(R_n(\alpha))\Lambda_1(\alpha)^{2t}} \tag{4.29}$$

where the sum is over $\vec{b}$-partitions α such that $\Lambda_1(\alpha) \neq 1$. Using the above expression calls for good bounds on the eigenvalues $\Lambda_1(\alpha)$ and the dimensions of the eigennspaces $R_n(\alpha)$. This section concentrates on the eigenvalues. It is first shown that it suffices to consider non-negative eigenvalues $\Lambda_1(\alpha)$ by showing that the eigenvalues come in pairs. Since this is always true, it is proved for a general $\vec{b}$:

Lemma 4.28. *Let $\alpha = (\lambda_1, \mu_1, \lambda_2, \ldots, \lambda_s)$ be a $\vec{b}$-partition. Letting*

$$\alpha^T = (\lambda_1^T, \mu_1^T, \lambda_2^T, \ldots, \lambda_s^T)$$

it can be concluded that

$$\dim(R_n(\alpha^T)) = \dim(R_n(\alpha)) \text{ and } \Lambda(\alpha^T) = -\Lambda(\alpha) \tag{4.30}$$

Proof: From Theorem 4.24,

$$\dim(R_n(\alpha^T)) = \prod_{i=0}^{s} \left|X_{\lambda_i^T/\mu_i^T}\right| \left|X_{\lambda_{i+1}^T/\mu_i^T}\right|$$

It is well-known that for general $\alpha \supseteq \beta$, X_{α^T/β^T} and $X_{\alpha/\beta}$ are conjugate representations, and hence their dimensions are equal. Thus,

$$\dim(R_n(\alpha^T)) = \prod_{i=0}^{s} \left|X_{\lambda_i/\mu_i}\right| \left|X_{\lambda_{i+1}/\mu_i}\right| = \dim(R_n(\alpha))$$

Furthermore, it is clear from the definition that $C(\alpha^T) = -C(\alpha)$. Lemma 4.27 gives

$$\begin{aligned}\Lambda(\alpha^T) &= \sum_{i=1}^{s} C(\lambda_i^T) - \sum_{i=1}^{s-1} C(\mu_i^T) = -\sum_{i=1}^{s} C(\lambda_i) + \sum_{i=1}^{s-1} C(\mu_i) \\ &= -\Lambda(\alpha)\end{aligned}$$

as required. □

Now, assume that $\Lambda_1(\alpha) < 0$. Then,

$$\begin{aligned}|\Lambda_1(\alpha)| &= -\frac{n+2\Lambda(\alpha)}{n+2\Delta} = \frac{-n+2\Lambda(\alpha^T)}{n+2\Delta} \leq \frac{n+2\Lambda(\alpha^T)}{n+2\Delta} \\ &= \Lambda_1(\alpha^T)\end{aligned}$$

Therefore,

$$\left\| P^t(x,\cdot) - \pi \right\|_{2,\pi} \leq \left(2 \sum_{1 \neq \Lambda_1(\alpha) \geq 0} \dim(R_n(\alpha)) \Lambda_1(\alpha)^{2t} \right)^{1/2} \tag{4.31}$$

Hence, it will suffice to provide upper bounds for the eigenvalues. The argument follows the approach laid out by Diaconis and Shahshahani in their analysis for S_n [14]. For their walk, the eigenvalues were functions of partitions λ, and it turned out that a good bound for these eigenvalues could be derived using only the largest part of λ. A similar trick is used to bound $\Lambda_1(\alpha)$.

As noted earlier in Remark 4.19, for $\vec{b} = \vec{b}_n(f,g)$, there are only two left-equivalence and two right-equivalence classes. Furthermore, by Equations (4.23) and (4.24), $|L_1| = f(n), |L_2| = n - f(n)$, and $|R_1| = g(n), |R_2| = n - g(n)$. Thus, by Definition 4.20, a $\vec{b}$-partition α can be written as

$$\alpha = (\lambda_1, \mu_1, \lambda_2), \quad \text{where } |\lambda_1| = f(n), |\mu_1| = f(n) - g(n), |\lambda_2| = n - g(n), \quad \text{and } \lambda_1 \supseteq \mu_1 \subseteq \lambda_2 \tag{4.32}$$

Hence, by Lemma 4.26, for $\alpha = (\lambda_1, \mu_1, \lambda_2)$,

$$\Lambda(\alpha) = C(\lambda_1) - C(\mu_1) + C(\lambda_2) \tag{4.33}$$

Since by Equation (4.28), $\Lambda_1(\alpha)$ is just $(n + 2\Lambda(\alpha))/(n + 2\Delta)$, it suffices to find bounds on the above quantity.

The largest parts of the partitions will be used to bound $\Lambda(\alpha)$. For sequences of partitions $\alpha = (\lambda_1, \mu_1, \lambda_2)$, the bounds will depend on the sequence of largest parts of λ_1, μ_1 and λ_2. In order to slightly simplify calculations, instead of working directly with the largest parts of α, it is convenient to work with what remains when the largest part has been taken away. The following definitions and lemmas carry out these ideas.

Definition 4.29. *Let λ be a partition of n. Then, write λ as*

$$\lambda = (\lambda^1, \lambda^2, \lambda^3, \dots)$$

where $\lambda^1 \geq \lambda^2 \geq \lambda^3 \dots$; in particular, the largest part of λ will be denoted by λ^1. Futhermore, note that $(\lambda^2, \lambda^3, \dots)$ is a partition of $n - \lambda^1$. Call this partition the remainder of λ, and denote it by λ^{Re}.

Now, let α be a $\vec{b}$-partition, where $\alpha = (\lambda_1, \mu_1, \lambda_2)$. Then, define

$$\begin{aligned} \alpha^1 &= \left(\lambda_1^1, \mu_1^1, \lambda_2^1\right) \\ \alpha^{Re} &= \left(\lambda_1^{Re}, \mu_1^{Re}, \lambda_2^{Re}\right) \end{aligned}$$

analogously to above. Furthermore, define

$$\left|\alpha^{Re}\right| = \left(f(n) - \lambda_1^1, f(n) - g(n) - \mu_1^1, n - g(n) - \lambda_2^1\right)$$

Note that

$$\left|\alpha^{Re}\right| = \left(\left|\lambda_1^{Re}\right|, \left|\mu_1^{Re}\right|, \left|\lambda_2^{Re}\right|\right)$$

The eigenvalue $\Lambda(\alpha)$ is bounded by finding a function of $\left|\alpha^{Re}\right|$ which is greater than it. The following lemmas carry this out.

Lemma 4.30. *Let λ be a partition of l and μ be a partition of m such that $\lambda \supseteq \mu$. Assume that*

$$\left|\lambda^{Re}\right| = i \text{ and } \left|\mu^{Re}\right| = j$$

Then, for $C(\lambda)$ as defined in Defintion 4.25, the following inequality holds:

$$C(\lambda) - C(\mu) \leq \frac{l^2 - l}{2} - \frac{m^2 - m}{2} - i(l - i + 1) + j(m - j + 1)$$

Furthermore, if $j \leq m/2$ and $i \leq l/2$ then equality is achieved at $\lambda = (l-i, i)$ and $\mu = (m-j, j)$.

Proof: Let $sq(a,b)$ denote the square of the Ferrers diagram in column a and row b, with the square in the upper left corner denoted by $sq(1,1)$. The content of a square $x = sq(a,b)$ is $a-b$. This increases as the column a increases, and decreases as the row b increases.

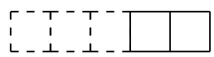
Note that $C(\lambda) - C(\mu) = C(\lambda/\mu)$, since $\lambda \supseteq \mu$. Since $\left|\lambda^{Re}\right| = i$ and $\left|\mu^{Re}\right| = j$, $\lambda_1 = l-i$ and $\mu_1 = m-j$. Thus, the first row of λ/μ contains the squares $sq(m-j+1,1), sq(m-j+2,1), \ldots, sq(l-i,1)$. Furthermore, λ/μ contains another $i-j$ squares in other rows.

Now, start with the Ferrers diagram that consists only of the first row squares $sq(m-j+1,1), sq(m-j+2,1), \ldots, sq(l-i,1)$ and add squares until reaching a skew partition that contains $l-m$ squares. Since $\left|\mu^{Re}\right| = j$, the first square added must be at most in column $j+1$ and at least in row 2. By similar logic, the rth square added must be at least in column $j+r$ and at least in row 2. Thus, the content of the rth square we add is at most $j+r-2$, so

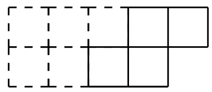
$$\begin{aligned}
C(\lambda/\mu) &= C(\text{squares in first row}) + C(\text{remaining squares}) \\
&\leq ((m-j) + \cdots + (l-i-1)) + ((j-1) + \cdots + (i-2)) \\
&= \binom{l-i}{2} - \binom{m-j}{2} + \binom{i-1}{2} - \binom{j-1}{2} \\
&= \frac{l^2-l}{2} - \frac{m^2-m}{2} - i(l-i+1) + j(m-j+1)
\end{aligned}$$

as required. Furthermore, if $j \leq m/2$ and $i \leq l/2$, then is straightforward to check that $\lambda = (l-i, i)$ and $\mu = (m-j, j)$ are partitions, satisfy $\lambda \subseteq \mu$, and achieve equality. $\square$

Example 4.31. Let $l = 9$ and $i = 4$, and let $m = 5$ and $j = 2$. Then, λ has 5 squares in the first row, and μ has 3. This λ/μ has 2 squares in the first row, and 2 more squares somewhere below the first row. Clearly, the first row of λ/μ looks like



where the dotted lines correspond to the squares of μ. Now, the remaining 2 squares of λ/μ need to placed somewhere. By definition of content, $C(\lambda/\mu)$ is maximized by placing these squares as far up and as far to the right as possible. Clearly the best thing to do is to put all the remaining squares in the second row – it is easy to see that otherwise, a square can be moved up and to the right, increasing content. Furthermore, since μ has 2 squares below the first row, all of these should be put in the second row as well, to make sure that the squares of λ/μ are as far to the right as possible. Thus, the optimal arrangement is



which is precisely the case $\lambda = (5, 4)$ and $\mu = (3, 2)$. Note that this arrangement would have been impossible if i or j was too large, but the same argument would have still provided an upper bound.

Corollary 4.32. *Let λ be a partition of l and μ be a partition of m such that $\lambda \supseteq \mu$. Then,*

$$C(\lambda) - C(\mu) \leq \frac{l^2 - l}{2} - \frac{m^2 - m}{2}$$

Furthermore, equality is achieved at $\lambda = (l)$ and $\mu = (m)$.

Proof: A proof identical to the one for Lemma 4.30 above works, but instead the lemma itself is used. Assume that $\left|\lambda^{Re}\right| = i$ and $\left|\mu^{Re}\right| = j$. Since $\lambda \supseteq \mu$, it follows that $i \geq j$ and $\lambda^1 \geq \mu^1$, so $l - i \geq m - j$. Combining, $i(l - i + 1) \geq j(m - j + 1)$. Thus,

$$\begin{aligned} C(\lambda) - C(\mu) &\leq \frac{l^2 - l}{2} - \frac{m^2 - m}{2} - i(l - i + 1) + j(m - j + 1) \\ &\leq \frac{l^2 - l}{2} - \frac{m^2 - m}{2} \end{aligned}$$

as required. It is easy to check that $\lambda = (l)$ and $\mu = (m)$ achieve equality. $\square$

The following result appears in Diaconis (Lemma 2, Chapter 3) [10]. A proof is given.

Lemma 4.33. *Let λ be a partition of l, and let $C(\lambda)$ be the sum of the contents of the squares in λ, as defined above. Let $\left|\lambda^{Re}\right| = i$. Then,*

$$C(\lambda) \leq \frac{l^2 - l}{2} - i(l - i + 1)$$

for all i, with equality achieved for $i \leq \frac{l}{2}$ at $\lambda = (l - i, i)$.

Furthermore, if $i \geq l/2$,

$$C(\lambda) \leq \frac{l^2 - l}{2} - \frac{il}{2}$$

Proof: The first inequality follows easily from Lemma 4.30. Let $\mu = \emptyset$. Then, $m = j = 0$, and

$$C(\lambda) \leq \frac{l^2 - l}{2} - i(l - i + 1)$$

as required. From the same lemma, if $i \leq l/2$, equality is achieved at $(l - i, i)$.

To prove the second inequality, note that if $\lambda = (\lambda^1, \lambda^2, \ldots, \lambda^r)$, then

$$C(\lambda) = \sum_{s=1}^{r} \frac{\lambda^s(\lambda^s - s - 1)}{2}$$

Clearly, $\lambda^s \leq \lambda^1$ for each s. Thus,

$$\begin{aligned} C(\lambda) &\leq \sum_{s=1}^{r} \frac{\lambda^s(\lambda^1 - s - 1)}{2} \leq \sum_{s=1}^{r} \frac{\lambda^s(\lambda^1 - 1)}{2} \\ &= \frac{\lambda^1 - 1}{2} \sum_{s=1}^{r} \lambda^s \end{aligned}$$

Now, $\lambda^1 - 1 = l - i - 1$, while $\sum \lambda^s$ is the total number of squares in the partition λ, which is equal to l. This gives

$$C(\lambda) \leq \frac{(l - i - 1)l}{2} = \frac{l^2 - l}{2} - \frac{il}{2}$$

as required. □

The next two lemmas bound the eigenvalues $\Lambda_1(\alpha)$. The first lemma will be used when $\left|\lambda_2^{Re}\right|$ is sufficiently large, and the second lemma will be used the rest of the time.

Lemma 4.34. *Let $\alpha = (\lambda_1, \mu_1, \lambda_2)$ be a $\vec{b}$-partition, where $\vec{b} = \vec{b}_n(f, g)$, such that $\left|\lambda_2^{Re}\right| = k \geq (n - g(n))/5$. Then, for sufficiently large n,*

$$\Lambda_1(\alpha) \leq \frac{9}{10}$$

Proof: From Equation (4.33),

$$\Lambda(\alpha) = C(\lambda_1) - C(\mu_1) + C(\lambda_2)$$

From Equation (4.28),

$$\Lambda_1(\alpha) = \frac{n + 2\Lambda(\alpha)}{n + 2\Delta}$$

and from Lemma 4.7,

$$n + 2\Delta = n^2 - 2ng(n) + 2g(n)f(n)$$

Since λ_2 is a partition of $n - g(n)$, and $\left|\lambda_2^{Re}\right| = k \geq (n - g(n))/5$, Lemma 4.33 gives

$$C(\lambda_2) \leq \frac{(n - g(n))^2 - (n - g(n))}{2} - \frac{k(n - g(n))}{2} \leq \frac{2(n - g(n))^2}{5}$$

Further, from Corollary 4.32,

$$C(\lambda_1) - C(\mu_1) \leq \frac{f(n)^2 - f(n)}{2} - \frac{(f(n) - g(n))^2 - (f(n) - g(n))}{2}$$

Since $f(n)/n$ approaches 0, $g(n) \leq f(n)$, and $n + 2\Delta = n^2 - 2ng(n) + 2g(n)f(n)$, the above expression for $C(\lambda_1) - C(\mu_1)$ is negligible compared to $n + 2\Delta$. Similarly, $n/(n + 2\Delta)$ approaches 0 as $n \to \infty$,

$$\begin{aligned}
\lim_{n\to\infty} \Lambda_1(\alpha) &= \lim_{n\to\infty} \frac{n + 2C(\lambda_1) - 2C(\mu_1) + 2C(\lambda_2)}{n + 2\Delta} \\
&= \lim_{n\to\infty} \frac{4}{5}\left(\frac{(n - g(n))^2}{n^2 - 2ng(n) + 2f(n)g(n)}\right) \\
&= \frac{4}{5}
\end{aligned}$$

again using the fact that $f(n)$ and $g(n)$ are negligible compared to n. Thus, for sufficiently large n, $\Lambda_1(\alpha) \leq 9/10$, completing the proof. □

Lemma 4.35. *Let $\alpha = (\lambda_1, \mu_1, \lambda_2)$ be a $\vec{b}$-partition, where $\vec{b} = \vec{b}_n(f, g)$, such that $\left|\beta^{Re}\right| = (i, j, k)$. Then,*

$$\Lambda_1(\alpha) \leq 1 - s_1(i) + s_2(j) - s_3(k)$$

where

$$\begin{aligned}
s_1(i) &= \begin{cases} \frac{2i(f(n)-i+1)}{n+2\Delta} & i < \frac{f(n)}{2} \\ \frac{if(n)}{n+2\Delta} & i \geq \frac{f(n)}{2} \end{cases} \\
s_2(j) &= \frac{2j(f(n) - g(n) - j + 1)}{n + 2\Delta} \\
s_3(k) &= \frac{2k(n - g(n) - k + 1)}{n + 2\Delta}
\end{aligned}$$

Proof: From Equation (4.33),

$$\Lambda(\alpha) = C(\lambda_1) - C(\mu_1) + C(\lambda_2)$$

Furthemore, from Lemma 4.33,

$$C(\lambda_1) \leq \begin{cases} \frac{f(n)^2 - f(n)}{2} - i(f(n) - i + 1) & i < \frac{f(n)}{2} \\ \frac{f(n)^2 - f(n)}{2} - \frac{if(n)}{2} & i \geq \frac{f(n)}{2} \end{cases}$$

Thus, from the definition of $s_1(i)$,

$$C(\lambda_1) \leq \frac{f(n)^2 - f(n)}{2} - \frac{n + 2\Delta}{2} s_1(i)$$

Using Lemma 4.30 and simplifying,

$$C(\lambda_2) - C(\mu_1) \leq \Delta - \frac{f(n)^2 - f(n)}{2} - \frac{n + 2\Delta}{2} s_2(j) + \frac{n + 2\Delta}{2} s_3(k)$$

using the fact that $2\Delta = n^2 - 2ng(n) + 2f(n)g(n) - n$.

Thus, combining the two,

$$\Lambda(\alpha) = C(\lambda_1) - C(\mu_1) + C(\lambda_2) \leq \Delta - \frac{n + 2\Delta}{2} (s_1(i) - s_2(j) + s_3(k))$$

Since $\Lambda_1(\alpha) = \frac{n + 2\Lambda(\alpha)}{n + 2\Delta}$,

$$\Lambda_1(\alpha) \leq 1 - s_1(i) + s_2(j) - s_3(j)$$

as required. $\square$

4.7 Lead Term Analysis and Chi-squared Lower Bound

The eigenvalue bounds derived above will now be used to provide some lead term analysis and also to prove the lower bound of Theorem 2.20. This will use Lemmas 4.32 and 4.33 – Lemma 4.35 could also be used, but the previous lemmas are more hands on.

Let $\alpha = (\lambda_1, \mu_1, \lambda_2)$ be a $\vec{b}$-partition, where $\vec{b} = \vec{b}_n(f, g)$. Then,

$$\Lambda(\alpha) = C(\lambda_1) - C(\mu_1) + C(\lambda_2),$$

Lemmas 4.32 and 4.33 show that $C(\lambda_1) - C(\mu_1)$ is maximized at $\lambda_1 = (f(n))$ and $\mu_1 = (f(n) - g(n))$, whereas $C(\lambda_2)$ is maximized at $(n - g(n))$. Thus, letting

$$\alpha_0 = ((f(n)), (f(n) - g(n)), (n - g(n))),$$

$\Lambda(\alpha_0)$ must be the maximal eigenvalue of U. Indeed, doing the calculation gives

$$\Lambda(\alpha_0) = \frac{n^2 - n - 2ng(n) + f(n)g(n)}{2} = \Delta$$

which is precisely the expected maximal eigenvalue of an adjacency matrix of a regular graph with degree Δ. Intuitively, the next highest eigenvalues should be $\Lambda(\alpha)$ for α close to α_0. If $\alpha = (\lambda_1, \mu_1, \lambda_2) \neq \alpha_0$, clearly either $\lambda_1 \neq (f(n))$ or $\lambda_2 \neq (n - g(n))$. Thus, the obvious candidates for next highest eigenvalue are

$$\begin{aligned}
\alpha_1 &= ((f(n) - 1, 1), (f(n) - g(n)), (n - g(n))) \text{ and} \\
\alpha_2 &= ((f(n)), (f(n) - g(n)), (n - g(n) - 1, 1))
\end{aligned}$$

Indeed, if $\alpha = (\lambda_1, \mu_1, \lambda_2)$ and $\lambda_1 \neq (f(n))$, then from Lemmas 4.33 and 4.32 it is easy to conclude that

$$\begin{aligned}
\Lambda(\alpha) &= C(\lambda_1) - C(\mu_1) + C(\lambda_2) \\
&\leq \frac{f(n)^2 - f(n)}{2} - f(n) + (C(\lambda_2) - C(\mu_1)) \\
&\leq C(f(n)) - f(n) - C(f(n) - g(n)) + C(n - g(n)) \\
&\leq \Lambda(\alpha_1)
\end{aligned}$$

and similarly, if $\lambda_2 \neq (n - g(n))$,

$$\begin{aligned}
\Lambda(\alpha) &\leq C(f(n)) - C(f(n) - g(n)) + C(n - g(n)) - (n - g(n)) \\
&\leq \Lambda(\alpha_2)
\end{aligned}$$

The above arguments (or a straighforward calculation) should make it clear that

$$\Lambda(\alpha_1) = \Delta - f(n) \quad \text{and} \quad \Lambda(\alpha_2) = \Delta - (n - g(n)) \tag{4.34}$$

The following steps derive the bounds corresponding to α_1 and α_2 for the lead-term analysis.

Bound corresponding to α_1:
Here, the term $\dim(R_n(\alpha_1))\Lambda_1(\alpha_1)^{2t}$ is used, where

$$\alpha_1 = ((f(n)-1,1),(f(n)-g(n)),(n-g(n))) \tag{4.35}$$

as defined above. Now, from Equation 4.25,

$$\dim(R_n(\alpha_1)) = |X_{\lambda_1}|\,|X_{\lambda_1/\mu_1}|\,|X_{\lambda_2/\mu_1}|\,|X_{\lambda_2}|$$

As noted in Remark 4.22, the degree of a skew representation is just the number of standard Young tableaux of that skew shape. Thus,

$$|X_{\lambda_1}| = |X_{(f(n)-1,1)}| = (f(n)-1)$$

since choosing a standard Young tableau for $(f(n)-1,1)$ just requires picking a number other than 1 for the second row – the numbers in the first row have to be ordered, and and picking 1 to go in the second row would result in a contradiction in the first column.

Now, $\lambda_1/\mu_1 = (f(n)-1,1)/(f(n)-g(n))$, which consists of precisely $g(n)-1$ squares starting at $f(n)-g(n)+1$ in the first row, and one square in the second row. Since the square in the second row is not directly below any square in the first row, choosing a standard Young tableau just requires picking any number for the second row. Thus,

$$|X_{\lambda_1/\mu_1}| = g(n)$$

Since λ_2/μ_1 and λ_2 are both just single rows,

$$|X_{\lambda_2/\mu_1}| = |X_{\lambda_2}| = 1$$

Combining,

$$\dim(R_n(\alpha_1)) = (f(n)-1)g(n).$$

From Equation (4.34),

$$\Lambda_1(\alpha_1) = \frac{n+2\Lambda(\alpha_1)}{n+2\Delta} = 1 - \frac{2f(n)}{n+2\Delta}$$

and hence

$$\begin{aligned}\dim(R_n(\alpha_1))\Lambda_1(\alpha_1)^{2t} &= (f(n)-1)g(n)\left(1-\frac{2f(n)}{n+2\Delta}\right)^{2t} \\ &\approx (f(n)-1)g(n)\exp\left(-\frac{4tf(n)}{n+2\Delta}\right)\end{aligned} \tag{4.36}$$

Thus, making this lead term at most e^{-c} for some constant c requires

$$\frac{4tf(n)}{n+2\Delta} \geq \log(f(n)-1)+\log g(n)+c$$

so it clearly suffices to have

$$t \geq \frac{(n+2\Delta)(\log f(n)+\log g(n))}{4f(n)} + c\frac{n+2\Delta}{4f(n)} \tag{4.37}$$

Thus, the above bound is the contribution of α_1. Turn next to α_2.

Bound corresponding to α_2:

The case of

$$\alpha_2 = ((f(n)),(f(n)-g(n)),(n-g(n),1))$$

is entirely analogous. Identical calculations show that

$$\dim(R_n(\alpha_2)) = (n-g(n)-1)(n-f(n))$$

and that

$$\Lambda_1(\alpha_2) = 1-\frac{2(n-g(n))}{n+2\Delta}$$

Hence,

$$\dim(R_n(\alpha_2))\Lambda_2(\alpha_1)^{2t} \approx (n-g(n)-1)(n-f(n))\exp\left(-\frac{4t(n-g(n))}{n+2\Delta}\right)$$

giving the bound

$$t \geq \frac{(n+2\Delta)(\log(n-f(n))+\log(n-g(n)))}{4(n-g(n))} + c\frac{n+2\Delta}{4(n-g(n))} \tag{4.38}$$

To compare the bounds, first prove the following simple lemma:

Lemma 4.36. *If (x_n) and (y_n) are positive sequences such that $\lim_{n\to\infty} \frac{x_n}{y_n} = 0$, and $\lim_{n\to\infty} x_n = \infty$, then for an arbitrarily large constant c,*

$$\frac{\log x_n}{x_n} \geq c\frac{\log y_n}{y_n}$$

for sufficiently large n.

Proof: Rewriting,

$$\begin{aligned}\frac{\log y_n}{y_n} &= \frac{\log(y_n/x_n) + \log x_n}{x_n} \cdot \frac{x_n}{y_n} \\ &= \frac{1}{x_n}\left(-\frac{x_n}{y_n}\log\left(\frac{x_n}{y_n}\right) + \frac{x_n}{y_n} \cdot \log x_n\right)\end{aligned}$$

Since $\frac{x_n}{y_n}$ approaches 0, and x_n approaches ∞, $-\frac{x_n}{y_n}\log\left(\frac{x_n}{y_n}\right)$ approaches 0 and thus is eventually less than $\frac{1}{2c}\log x_n$. Similarly, $\frac{x_n}{y_n}\log x_n$ is eventually less than $\frac{1}{2c}\log x_n$. Combining, for sufficiently large n,

$$\frac{\log y_n}{y_n} \leq \frac{1}{c}\frac{\log x_n}{x_n}$$

as required. □

One of the assumptions in Theorem 4.9 is that $f(n)/n$ approaches 0. Since $g(n) \leq f(n)$ by definition, the bound in Equation (4.38) is of order $(n+2\Delta)\log n/n$, whereas the bound in Equation (4.37) is of order $(n+2\Delta)\log f(n)/f(n)$. The above lemma with $x_n = f(n)$ and $y_n = n$ makes it clear that Equation (4.37) is the stronger bound. This lead term analysis suggests chi-squared cut-off around

$$t = \frac{(n+2\Delta)(\log f(n) + \log g(n))}{4f(n)} \tag{4.39}$$

with a window of order $\frac{n+2\Delta}{4f(n)}$.

Having derived the second highest eigenvalue, the lower bound part of Theorem 4.9 can be proved. If t is defined as

$$t = \frac{(n+2\Delta)(\log f(n) + \log g(n))}{4f(n)} - c\frac{n+2\Delta}{4f(n)} \tag{4.40}$$

it is shown that

$$\left\|P^t(x,\cdot) - \pi\right\|_{2,\pi} \geq \frac{1}{2}e^{c/2}$$

for sufficiently large n.

Proof of Lower Bound in Theorem 4.9: Let $\alpha_1 = ((f(n)-1,1),(f(n)-g(n)),(n-g(n)))$ as in Equation (4.35). From Equation (4.29),

$$\begin{aligned}\left\|P^t(x,\cdot)-\pi\right\|_{2,\pi} &= \sqrt{\sum_{\alpha}\dim(R_n(\alpha))\Lambda_1(\alpha)^{2t}} \\ &\geq \sqrt{\dim(R_n(\alpha_1))\Lambda_1(\alpha_1)^{2t}}\end{aligned}$$

Furthermore, from Equation (4.36),

$$\dim(R_n(\alpha_1))\Lambda_1(\alpha_1)^{2t} = (f(n)-1)g(n)\left(1-\frac{2f(n)}{n+2\Delta}\right)^{2t}$$

Since $\frac{f(n)}{n}\to 0$, $\frac{f(n)}{n+2\Delta}\to 0$ as n approaches ∞. Thus, for t as defined in Part 2 of Theorem 4.9 and as restated above in Equation (4.40),

$$\begin{aligned}\lim_{n\to\infty}(f(n)-1)g(n)\left(1-\frac{2f(n)}{n+2\Delta}\right)^{2t} &= \lim_{n\to\infty}(f(n)-1)g(n)\exp\left(-\frac{4tf(n)}{n+2\Delta}\right) \\ &= \lim_{n\to\infty}(f(n)-1)g(n)e^{-\log f(n)-\log g(n)+c} \\ &= \lim_{n\to\infty}\frac{f(n)-1}{f(n)}e^c = e^c\end{aligned}$$

since $f(n)\to\infty$. Combining the equations above,

$$\left\|P^t(x,\cdot)-\pi\right\|_{2,\pi} \geq \sqrt{\dim(R_n(\alpha_1))\Lambda_1(\alpha_1)^{2t}} \longrightarrow e^{c/2} \text{ as } n\to\infty$$

and therefore, for sufficiently large n, the chi-squared distance is at least $\frac{e^{c/2}}{2}$. This proves the lower bound. □

Remark 4.37. Note that while the assumptions of Theorem 4.9 are used in the above proof to simplify computation, their full strength is not needed. Indeed, a similar lower bound can be derived for almost any functions f and g. However, if $f(n)\geq n/2$ then this lower bound will not correspond to the second-highest eigenvalue, as that will be associated to the α_2 defined above.

4.8 Dimensions of Eigenspaces

This section continues with the proof of the the upper bound in Theorem 4.9. Here is a recap of the bounds in Section 4.6.

Definition 4.38. *Let (i, j, k) be a triple of integers. Then, define*

$$s(i,j,k) = \begin{cases} \frac{9}{10} & k \geq (n - g(n))/5 \\ 1 - s_1(i) + s_2(j) - s_3(k) & \textit{otherwise} \end{cases} \tag{4.41}$$

where s_1, s_2 and s_3 are defined as in Lemma 4.35 – that is,

$$s_1(i) = \begin{cases} \frac{2i(f(n)-i+1)}{n+2\Delta} & i < \frac{f(n)}{2} \\ \frac{if(n)}{n+2\Delta} & i \geq \frac{f(n)}{2} \end{cases}$$
$$s_2(j) = \frac{2j(f(n) - g(n) - j + 1)}{n + 2\Delta}$$
$$s_3(k) = \frac{2k(n - g(n) - k + 1)}{n + 2\Delta}$$

Remark 4.39. The above $s(i, j, k)$ are chosen to simplify notation. Combining Lemmas 4.34 and 4.35,

$$\Lambda_1(\alpha) \leq s\left(\left|\alpha^{Re}\right|\right) \tag{4.42}$$

which will clearly be useful.

Here is a sketch out the rest of the proof. By Equation (4.31), the quantity to be bounded is

$$\sum_{1 \neq \Lambda_1(\alpha) \geq 0} \dim(R_n(\alpha))\Lambda_1(\alpha)^{2t}$$

Since the upper bound above depends purely on $\left|\alpha^{Re}\right|$, rearrange the above quantity by the value of $\left|\alpha^{Re}\right|$. From the previous section, it is clear that $\left|\alpha^{Re}\right| = (0, 0, 0)$ corresponds to $\Lambda_1(\alpha) = 1$. Thus,

$$\begin{aligned} \sum_{1 \neq \Lambda_1(\alpha) \geq 0} \dim(R_n(\alpha))\Lambda_1(\alpha)^{2t} &\leq \sum_{\left|\alpha^{Re}\right| \neq (0,0,0)} \dim(R_n(\alpha)) s\left(\left|\alpha^{Re}\right|\right)^{2t} \\ &= \sum_{(i,j,k) \neq (0,0,0)} s(i,j,k)^{2t} \sum_{\left|\alpha^{Re}\right| = (i,j,k)} \dim(R_n(\alpha)) \end{aligned} \tag{4.43}$$

The proof below will be organized as follows. The current section finds an expression for the sum

$$\sum_{|\alpha^{Re}|=(i,j,k)} \dim(R_n(\alpha))$$

in terms of i, j, and k. This leaves the (rather unwieldy) sum on the right-hand side of Equation (4.43) in terms of the three indices i, j, and k. In Section 4.9, this sum is split into a number of pieces that depend on the precise values of the indices, and supporting lemmas are proved for the size of each piece. All the bounds are then combined into a proof of the upper bound part of Theorem 4.9.

The current section is devoted to proving the following lemma:

Lemma 4.40. *Assume that $i \geq j \leq k$. Then,*

$$\sum_{|\alpha^{Re}|=(i,j,k)} \dim(R_n(\alpha)) \leq \binom{f(n)}{i}\binom{g(n)}{i-j}\binom{n-f(n)}{k-j}\binom{n-g(n)}{k}\frac{i!k!}{j!}$$

Remark 4.41. Note that $\left|\alpha^{Re}\right|$ is defined to be $\left(\left|\lambda_1^{Re}\right|, \left|\mu_1^{Re}\right|, \left|\lambda_2^{Re}\right|\right)$, where $\lambda_1 \supseteq \mu_1 \subseteq \lambda_2$. Thus, if

$$\left|\alpha^{Re}\right| = (i, j, k)$$

then $i \geq j \leq k$. Therefore, the the condition in Lemma 4.40 is the natural one.

Definition 4.42. *To simplify notation from now on, define*

$$a(i,j,k) = \binom{f(n)}{i}\binom{g(n)}{i-j}\binom{n-f(n)}{k-j}\binom{n-g(n)}{k}$$

The n in the above expression will always be implied.

A number of supporting lemmas will eventually yield Lemma 4.40 above.

Lemma 4.43. *Let $\alpha = (\lambda_1, \mu_1, \lambda_2)$ be a $\vec{b}$-partition, where $\vec{b} = \vec{b}_n(f, g)$, satisfying $\left|\alpha^{Re}\right| = (i, j, k)$. Then,*

$$\dim(R_n(\alpha)) \leq a(i,j,k)\left|X_{\lambda_1^{Re}}\right|\left|X_{\lambda_1^{Re}/\mu_1^{Re}}\right|\left|X_{\lambda_2^{Re}/\mu_1^{Re}}\right|\left|X_{\lambda_2^{Re}}\right| \tag{4.44}$$

Proof: From Equation (4.25) in Theorem 4.24,

$$\dim(R_n(\alpha)) = |X_{\lambda_1}|\left|X_{\lambda_1/\mu_1}\right||X_{\lambda_2}|\left|X_{\lambda_2/\mu_1}\right|$$

Furthermore, as noted in Remark 4.22, $\left|X_{\alpha/\beta}\right|$ is the number of standard Young tableaux of shape α/β. Use this to examine the expression above. Clearly, $|X_{\lambda_1}|$ is the number of standard Young tableaux of shape λ_1. A naive way of trying to construct such a tableau would be to pick the $f(n) - i$ elements of $\{1, 2, \ldots, f(n)\}$ that will go in the first row of the standard Young tableau of shape λ_1, and then use the remaining i elements to construct a 'shifted' standard Young tableau of the remainder of λ_1, called λ_1^{Re} above. Note that once the elements of the first row are chosen, they must be arranged in exactly increasing order: thus, each choice of subset and of remaining 'shifted' tableau results in exactly one Young tableau. While this will vastly overcount the number of standard Young tableaux of shape λ_1, every single one can be constructed in this way. Thus,

$$|X_{\lambda_1}| \leq \binom{f(n)}{f(n)-i} \left|X_{\lambda_1^{Re}}\right| = \binom{f(n)}{i} \left|X_{\lambda_1^{Re}}\right|$$

Similarly, to construct a skew tableau of shape λ_1/μ_1, pick the $g(n) - (i-j)$ elements for the first row, and construct a 'shifted' tableau of shape $\lambda_1^{Re}/\mu_1^{Re}$ with the remaining numbers. Thus,

$$\left|X_{\lambda_1/\mu_1}\right| \leq \binom{g(n)}{i-j} \left|X_{\lambda_1^{Re}/\mu_1^{Re}}\right|$$

Proceeding in this way,

$$\left|X_{\lambda_2/\mu_1}\right| \leq \binom{n-f(n)}{k-j} \left|X_{\lambda_2^{Re}/\mu_1^{Re}}\right|$$
$$|X_{\lambda_2}| \leq \binom{n-g(n)}{k} \left|X_{\lambda_2^{Re}}\right|.$$

Multiplying the above inequalitites together gives precisely

$$\dim(R_n(\alpha)) \leq a(i,j,k) \left|X_{\lambda_1^{Re}}\right| \left|X_{\lambda_1^{Re}/\mu_1^{Re}}\right| \left|X_{\lambda_2^{Re}/\mu_1^{Re}}\right| \left|X_{\lambda_2^{Re}}\right|$$

as desired. □

Using the lemma above,

$$\sum_{|\alpha^{Re}|=(i,j,k)} \dim(R_n(\alpha)) \leq a(i,j,k) \sum_{|\alpha^{Re}|=(i,j,k)} \left|X_{\lambda_1^{Re}}\right| \left|X_{\lambda_1^{Re}/\mu_1^{Re}}\right| \left|X_{\lambda_2^{Re}/\mu_1^{Re}}\right| \left|X_{\lambda_2^{Re}}\right|$$

In order to simplify things slightly, the following simple lemma is useful:

Lemma 4.44. *For any i, j, k,*

$$\sum_{|\alpha^{Re}|=(i,j,k)} \left|X_{\lambda_1^{Re}}\right| \left|X_{\lambda_1^{Re}/\mu_1^{Re}}\right| \left|X_{\lambda_2^{Re}/\mu_1^{Re}}\right| \left|X_{\lambda_2^{Re}}\right| \leq \sum_{(\lambda_1',\mu_1',\lambda_2')} \left|X_{\lambda_1'}\right| \left|X_{\lambda_1'/\mu_1'}\right| \left|X_{\lambda_2'/\mu_1'}\right| \left|X_{\lambda_2'}\right|$$

where the right-hand sum is over triples of partitions such that $|\lambda_1'| = i, |\mu_1'| = j, |\lambda_2'| = k$, and $\lambda_1' \supseteq \mu_1' \subseteq \lambda_2'$.

Proof: Clearly, if $\left|\alpha^{Re}\right| = (i, j, k)$, then

$$\left|\lambda_1^{Re}\right| = i, \left|\mu_1^{Re}\right| = j, \left|\lambda_2^{Re}\right| = k, \text{ and } \lambda_1^{Re} \supseteq \mu_1^{Re} \subseteq \lambda_2^{Re}$$

Furthermore, given $\vec{b} = \vec{b}_n(f, g)$, α^{Re} uniquely determines α (although not every choice of $(\lambda_1', \mu_1', \lambda_2')$ will actually produce an α). Hence, the lemma follows trivially. □

Now, note that the sum on the right-hand side of the above lemma looks a lot like a sum of $\dim(R_n(\alpha'))$ over $\vec{a}$-partitions α' for some $\vec{a}$. Using this heuristic, consider rewriting the above as

$$\sum_{\alpha'} \dim(R_n(\alpha')) = \left|S_{M(\vec{a})}\right|$$

where the equality follows since the sum of the dimensions of the eigenspaces is just the dimension of the whole space.

However, the above heuristic has the following trouble: if either $i = j$ or $j = k$, then the $(\lambda_1', \mu_1', \lambda_2')$ above are not in fact $\vec{a}$-partitions for any $\vec{a}$, since $\lambda_1' = \mu_1'$ or $\mu_1' = \lambda_2'$ are forbidden. However, this turns out to be a non-essential part of the definition of $\vec{a}$-partitions. The following supporting lemma overcomes the issue.

Lemma 4.45. *Let $n \geq m$, and let λ be a partition of n. Then, the following equality holds*

$$|X_\lambda| = \sum_{\mu \leq \lambda, |\mu|=m} \left|X_{\lambda/\mu}\right| |X_\mu|$$

Proof: This may well be a well-known formula. However, it has a simple combinatorial proof presented below. As noted above, $|X_\lambda|$ is the number of standard Young tableaux of shape λ, and similarly for μ and λ/μ. Let $Tab(s)$ denote the set of standard Young tableaux of shape s, whether s is a partition or a skew-partition. Then, what is needed is a bijection

g such that

$$g : \bigcup_{|\mu|=m} Tab(\mu) \times Tab(\lambda/\mu) \longrightarrow Tab(\lambda) \tag{4.45}$$

where the union is over partitions μ.

To define the bijection g, note that if Y_λ is a standard Young tableau of shape λ and $Y_{\lambda/\mu}$ is a Young tableau of shape λ/μ, a Young tableau of shape λ can be created by adding m to every entry of the $Y_{\lambda/\mu}$ and then sticking the two tableaux together. Call this new tableau $g(Y_\mu, Y_{\lambda/\mu})$.

For example, if $\lambda = (4, 2, 1)$, and $\mu = (3, 1)$, then the following Young tableaux

$$Y_\mu = \begin{array}{|c|c|c|} \hline 1 & 3 & 4 \\ \hline 2 \\ \cline{1-1} \end{array} \qquad\qquad Y_{\lambda/\mu} = \begin{array}{cccc} & & & \boxed{1} \\ & \boxed{3} & & \\ \boxed{2} & & & \end{array}$$

can be combined into $g(Y_\mu, Y_{\lambda/\mu})$, which is

$$Y_\lambda = \begin{array}{|c|c|c|c|} \hline 1 & 3 & 4 & 5 \\ \hline 2 & 7 \\ \cline{1-2} 6 \\ \cline{1-1} \end{array}$$

Clearly, for any $Y_\mu \in Tab(\mu)$ and $Y_{\lambda/\mu} \in Tab(\lambda/\mu)$, the above bijection results in a tableau of shape λ which is filled with the numbers $\{1, 2, \ldots, n\}$. Thus, it remains to check that the rows and columns in $g(Y_\mu, Y_{\lambda/\mu})$ are in increasing order from left to right and from top to bottom.

Proceed by contradiction: assume there are i and j such that $i < j$, but j is strictly to the left of i, or strictly above i in $g(Y_\mu, Y_{\lambda/\mu})$. If i and j are both at most m, then they both appeared in Y_μ, and this is impossible; similarly, if both i and j are greater than m, then $i - m$ and $j - m$ both appeared in $Y_{\lambda/\mu}$ and this is similarly impossible. Thus, assume that $i \leq m < j$. But then j must be in a square belonging to λ/μ, and i must be in a square belonging to μ, and therefore it is impossible for j to be strictly to the left or strictly above i. Thus, $g(Y_\mu, Y_{\lambda/\mu})$ is a standard Young tableau, so the map g is well-defined.

Now for the inverse f of the map g. It is easy to see from above that μ is defined precisely by the set of squares which contain the numbers $\{1, 2, \ldots, m\}$. Using arguments identical to the above, in any standard Young tableau Y_λ the set of squares containing the elements $\{1, 2, \ldots, m\}$ is a partition of m. Thus, the inverse map f must map Y_λ to $(Y_\mu, Y_{\lambda/\mu})$, where Y_μ is simply the standard Young tableaux induced by the squares containing $\{1, 2, \ldots, m\}$, and $Y_{\lambda/\mu}$ is obtained by deleting the squares in μ from Y_λ, and subtracting m from the remaining squares. For example, if

$$Y_\lambda = \begin{array}{|c|c|c|c|}\hline 1 & 2 & 5 & 6 \\ \hline 3 & 4 \\ \cline{1-2} 7 \\ \cline{1-1}\end{array}$$

then the inverse map f maps it to the pair

$$Y_\mu = \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & 4 \\ \hline\end{array} \qquad\qquad Y_{\lambda/\mu} = \begin{array}{cccc} & & \boxed{1} & \boxed{2} \\ \\ \boxed{3} & & & \end{array}$$

and hence the bijection has an explicit inverse. This shows that the sizes of the sets in Equation (4.45) are equal, and hence

$$|X_\lambda| = \sum_{\mu \leq \lambda, |\mu| = m} \left|X_{\lambda/\mu}\right| |X_\mu|$$

as required. □

The next lemmas tackle the expression on the right-hand side of Lemma 4.44.

Lemma 4.46. *Let (i, j, k) be a triple of positive integers that satisifes $i \geq j \leq k$. Define*

$$A = \sum_{(\lambda_1', \mu_1', \lambda_2')} \left|X_{\lambda_1'}\right| \left|X_{\lambda_1'/\mu_1'}\right| \left|X_{\lambda_2'/\mu_1'}\right| \left|X_{\lambda_2'}\right| \tag{4.46}$$

where the sum is over triples of partitions $(\lambda_1', \mu_1', \lambda_2')$ such that $|\lambda_1'| = i, |\mu_1'| = j, |\lambda_2'| = k$ and $\lambda_1' \supseteq \mu_1' \subseteq \lambda_2'$. Then,

$$A = \frac{i!k!}{j!}$$

Proof: If $i > j < k$, define the vector $\vec{a}$ to be

$$\vec{a} = (1, 1, \ldots, 1, i - j + 1, \ldots, i - j + 1)$$

where the number of initial 1s is i, and the total number of terms in the vector is $i - j + k$. Then it is easy to check that the set of all $(\lambda_1', \mu_1', \lambda_2')$ such that $|\lambda_1'| = i, |\mu_1'| = j, |\lambda_2'| = k$

and $\lambda_1' \supseteq \mu_1' \subseteq \lambda_2'$ is the set of $\vec{a}$-partitions, as defined in Definition 4.20. Thus,

$$\begin{aligned} A &= \sum_{(\lambda_1',\mu_1',\lambda_2')\ \vec{a}\text{-partition}} \left|X_{\lambda_1'}\right| \left|X_{\lambda_1'/\mu_1'}\right| \left|X_{\lambda_2'/\mu_1'}\right| \left|X_{\lambda_2'}\right| \\ &= \sum_{(\lambda_1',\mu_1',\lambda_2')\ \vec{a}\text{-partition}} \dim(R_n(\lambda_1',\mu_1',\lambda_2')) \end{aligned}$$

where the second equality follows from Equation (4.25) in Theorem 4.24. But from the same theorem,

$$\mathbb{C}S_{M(\vec{a})} = \bigoplus_{\alpha\ \vec{a}\text{-partition}} R_n(\alpha)$$

Thus,

$$\sum_{\alpha\ \vec{a}\text{-partition}} \dim(R_n(\alpha)) = \dim \mathbb{C}S_{M(\vec{a})} = \left|S_{M(\vec{a})}\right|$$

But for the choice of $\vec{a}$, from Lemma 4.13,

$$\left|S_{M(\vec{a})}\right| = \frac{i!k!}{j!}$$

finally yielding that

$$A = \sum_{\alpha\ \vec{a}\text{-partition}} \dim(R_n(\alpha)) = \left|S_{M(\vec{a})}\right| = \frac{i!k!}{j!}$$

as required.

It remains to explain how to handle the case where $i = j$, or $j = k$, or both. Consider the case $i = j < k$ (the same method will apply to all the other cases.) As before,

$$A = \sum_{(\lambda_1',\mu_1',\lambda_2')} \left|X_{\lambda_1'}\right| \left|X_{\lambda_1'/\mu_1'}\right| \left|X_{\lambda_2'/\mu_1'}\right| \left|X_{\lambda_2'}\right|$$

where the sum is over $|\lambda_1'| = i, |\mu_1'| = j, |\lambda_2'| = k$ and $\lambda_1' \supseteq \mu_1' \subseteq' \lambda_2$. Since $i = j$, and $\lambda_1' \supseteq \mu_1'$, it must be that $\lambda_1' = \mu_1'$. Thus,

$$A = \sum_{(\lambda_1',\lambda_2')} \left|X_{\lambda_1'}\right| \left|X_{\lambda_2'/\lambda_1'}\right| \left|X_{\lambda_2'}\right|$$

where the sum is over $|\lambda_1'| = i$, $|\lambda_2'| = k$, and $\lambda_1' \subseteq \lambda_2'$. But from Lemma 4.45, for a fixed λ_2',

$$\sum_{\lambda_1' \subseteq \lambda_2', |\lambda_1'|=i} \left| X_{\lambda_2'/\lambda_1'} \right| \left| X_{\lambda_1'} \right| = \left| X_{\lambda_2'} \right|$$

Thus,

$$\begin{aligned} A &= \sum_{(\lambda_1', \lambda_2')} \left| X_{\lambda_1'} \right| \left| X_{\lambda_2'/\lambda_1'} \right| \left| X_{\lambda_2'} \right| \\ &= \sum_{|\lambda_2'|=k} \left| X_{\lambda_2'} \right| \sum_{\lambda_1' \subseteq \lambda_2', |\lambda_1'|=i} \left| X_{\lambda_2'/\lambda_1'} \right| \left| X_{\lambda_1'} \right| \\ &= \sum_{|\lambda_2'|=k} \left| X_{\lambda_2'} \right|^2 \end{aligned}$$

and the sum on the right is well-known to be $k!$. Therefore, in this case,

$$A = k! = \frac{i!k!}{j!}$$

since $i = j$. The other cases with equality can be done similarly, completing the proof. $\square$

With all these preliminaries, the lemma from the beginning of the section can now be proved.

Proof of Lemma 4.40. From Lemma 4.43,

$$\sum_{|\alpha^{Re}|=(i,j,k)} \dim(R_n(\alpha)) \leq a(i,j,k) \sum_{|\alpha^{Re}|=(i,j,k)} \left| X_{\lambda_1^{Re}} \right| \left| X_{\lambda_1^{Re}/\mu_1^{Re}} \right| \left| X_{\lambda_2^{Re}/\mu_1^{Re}} \right| \left| X_{\lambda_2^{Re}} \right|$$

Combining this with Lemma 4.44,

$$\sum_{|\alpha^{Re}|=(i,j,k)} \dim(R_n(\alpha)) \leq a(i,j,k) \sum_{(\lambda_1', \mu_1', \lambda_2')} \left| X_{\lambda_1'} \right| \left| X_{\lambda_1'/\mu_1'} \right| \left| X_{\lambda_2'/\mu_1'} \right| \left| X_{\lambda_2'} \right|$$

where the right-hand sum is over triples of partitions such that $|\lambda_1'| = i, |\mu_1'| = j, |\lambda_2'| = k$, and $\lambda_1' \supseteq \mu_1' \subseteq \lambda_2'$. However, from Lemma 4.46,

$$\sum_{(\lambda_1', \mu_1', \lambda_2')} \left| X_{\lambda_1'} \right| \left| X_{\lambda_1'/\mu_1'} \right| \left| X_{\lambda_2'/\mu_1'} \right| \left| X_{\lambda_2'} \right| = \frac{i!k!}{j!}$$

and thus

$$\sum_{|\alpha^{Re}|=(i,j,k)} \dim(R_n(\alpha)) \leq a(i,j,k)\frac{i!k!}{j!}$$

as required. □

4.9 Chi-Squared Upper Bound

This final section puts all the quantities together to prove the upper bound part of Theorem 4.9. For the remainder of this section, let t be defined the way it is for the upper bound; that is,

$$t = \frac{(n+2\Delta)(\log f(n) + \log g(n))}{4f(n)} + c\frac{n+2\Delta}{4f(n)} \tag{4.47}$$

From Equation (4.43),

$$\sum_{1\neq\Lambda_1(\alpha)\geq 0} \dim(R_n(\alpha))\Lambda_1(\alpha)^{2t} \leq \sum_{(i,j,k)\neq(0,0,0)} s(i,j,k)^{2t} \sum_{|\alpha^{Re}|=(i,j,k)} \dim(R_n(\alpha)) \tag{4.48}$$

where $s(i,j,k)$ is defined as in Definition 4.38. From the lead-term analysis in Section 4.7, the 'limiting' eigenvalue corresponds to the case $(i,j,k) = (1,0,0)$. This suggests that term is the largest.

The above sum will be broken up into various pieces and bounds will be proved for each piece. There are three zones:

1. The first zone is $k \geq (n - g(n))/5$. As should be clear from the definition of $s(i,j,k)$, this is the zone with a constant upper bound for the eigenvalues $\Lambda_1(\alpha)$. This case is fairly straightforward, and will be done in Lemma 4.48 below.

2. The second zone is $i = 0$ and $k < (n - g(n))/5$. For this case, Lemma 4.40 is used for the upper bound. From the heuristics in the lead term analysis, the expected limiting term for this piece is $(0,0,1)$ – since it was noted that this term imposes lower order restrictions than $(1,0,0)$, this case should also be fairly simple. This will be done in Lemma 4.50.

3. The final zone is $i > 0$ and $k < (n - g(n))/5$, for which Lemma 4.40 is again used. This is the case that contains the limiting term $(1,0,0)$, and as such should provide the biggest bound. This will be done in Lemma 4.51.

To simplify notation, make the following definition:

Definition 4.47. *Define*

$$\begin{aligned}
A_1 &= \left\{(i,j,k) \mid k \geq \frac{n-g(n)}{5}\right\} \\
A_2 &= \left\{(i,j,k) \neq (0,0,0) \mid k < \frac{n-g(n)}{5}, i = 0\right\} \\
A_3 &= \left\{(i,j,k) \mid k < \frac{n-g(n)}{5}, i \neq 0\right\}
\end{aligned}$$

This corresponds to the pieces being bounded.

Lemma 4.48. *Using the definition of A_1 above, let*

$$Q_1 = \sum_{(i,j,k)\in A_1} s(i,j,k)^{2t} \sum_{|\alpha^{Re}|=(i,j,k)} \dim(R_n(\alpha))$$

Then, for sufficiently large n, $Q_1 \leq n^{-n}$.

Proof: By Definition 4.38, for $(i,j,k) \in A_1$, $s(i,j,k) = 9/10$. Thus,

$$\begin{aligned}
Q_1 &\leq \sum_{(i,j,k)\in A_1} \left(\frac{9}{10}\right)^{2t} \sum_{|\alpha^{Re}|=(i,j,k)} \dim(R_n(\alpha)) \\
&\leq \left(\frac{9}{10}\right)^{2t} \sum_{\alpha} \dim(R_n(\alpha)) \\
&\leq \left(\frac{9}{10}\right)^{2t} n!
\end{aligned}$$

since $\sum_\alpha \dim(R_n(\alpha)) = \left|S_M(\vec{b})\right| \leq n!$. Then, using t as defined as in Equation (4.47) above,

$$\begin{aligned}
Q_1 &\leq \left(\frac{9}{10}\right)^{n^2(\log f(n)+\log g(n))/2f(n)} n! \\
&\leq \left(\frac{9}{10}\right)^{n^2(\log f(n)+\log g(n))/2f(n)} e^{n\log n} \\
&\leq \exp\left(n^2\left(\frac{\log n}{n} - \log(10/9)\frac{\log(f(n))}{2f(n)}\right)\right)
\end{aligned}$$

From Lemma 4.36, for sufficiently large n, the above is at most $e^{-n\log n} = n^{-n}$, completing the proof. $\square$

Before continuing, a simple supporting lemma is needed.

Lemma 4.49. *With notation as above,*

$$\frac{i!k!}{j!}a(i,j,k) \leq \frac{f(n)^i g(n)^{i-j}(n-f(n))^{k-j}(n-g(n))^k}{(i-j)!(k-j)!j!}$$

Proof: This follows easily from the definition of $a(i,j,k)$ in Definition 4.42 and the fact that $\binom{x}{y} \leq \frac{x^y}{y!}$. □

Lemma 4.50. *For*

$$A_2 = \left\{(i,j,k) \neq (0,0,0) \mid k < \frac{n-g(n)}{5}, i = 0\right\}$$

let

$$Q_2 = \sum_{(i,j,k)\in A_2} s(i,j,k)^{2t} \sum_{|\alpha^{Re}|=(i,j,k)} \dim(R_n(\alpha)).$$

Then, for sufficiently large n, $Q_2 \leq 2(n-g(n))^{-6}$.

Proof: By Definition 4.38,

$$\begin{aligned} s(i,j,k) &= 1 - s_1(i) + s_2(j) - s_3(k) \\ &\leq e^{-s_1(i)+s_2(j)-s_3(k)} \end{aligned}$$

since $1 - x \leq e^{-x}$ for all x. Furthermore, from Lemma 4.40,

$$\sum_{|\alpha^{Re}|=(i,j,k)} \dim(R_n(\alpha)) \leq a(i,j,k)\frac{i!k!}{j!}$$

for $a(i,j,k)$ as in Definition 4.42. Combining the above,

$$Q_2 \leq \sum_{(i,j,k)\in A_2} e^{-2ts_1(i)+2ts_2(j)-2ts_3(k)} a(i,j,k)\frac{i!k!}{j!} \tag{4.49}$$

Now, for $(i,j,k) \in A_2$, $i = 0$. Since $i \geq j$, this also means that $j = 0$. Using Definition 4.38 and Lemma 4.49,

$$s_1(i) = 0, s_2(j) = 0 \text{ and } \frac{i!k!}{j!}a(i,j,k) \leq \frac{(n-f(n))^k(n-g(n))^k}{k!}$$

Thus, the above inequality simplifies to

$$Q_2 \le \sum_{1\le k\le \frac{n-g(n)}{5}} \exp\left(-\frac{4tk(n-g(n)-k+1)}{n+2\Delta}\right)\frac{(n-f(n))^k(n-g(n))^k}{k!}$$

and hence, using the fact that $g(n) \le f(n)$,

$$\begin{aligned} Q_2 &\le \sum_{1\le k< \frac{n-g(n)}{5}} \exp\left(-\frac{4tk(n-g(n)-k+1)}{n+2\Delta}\right)\frac{(n-g(n))^{2k}}{k!} \\ &\le \sum_{1\le k< \frac{n-g(n)}{5}} \exp\left(-\frac{16tk(n-g(n))}{5(n+2\Delta)}\right)\frac{(n-g(n))^{2k}}{k!} \end{aligned}$$

Now, using Lemma 4.36 with $x_n = f(n)$ and $y_n = n - g(n)$, for sufficiently large n, $\frac{t}{n+2\Delta} \ge \frac{f(n)}{4f(n)} \ge 10\frac{\log(n-g(n))}{n-g(n)}$. Thus,

$$Q_2 \le \sum_{1\le k< \frac{n-g(n)}{5}} \exp\left(-8k\log(n-g(n))\right)\frac{(n-g(n))^{2k}}{k!}$$

and hence the above simplifies to

$$Q_2 \le \sum_{1\le k< \frac{n-g(n)}{5}} \frac{(n-g(n))^{-6k}}{k!} \le 2(n-g(n))^{-6}$$

as required. □

It is easy to see that the lemmas above can be manipulated to provide arbitrarily good bounds. As noted previously, this is because the 'limiting' term $(1, 0, 0)$ does not make an appearance here. The case $i \ne 0$ will be considerably more tricky (or at least more tedious.)

Lemma 4.51. *For*

$$A_3 = \left\{(i,j,k) \mid k < \frac{n-g(n)}{5}, i \ne 0\right\}$$

let

$$Q_3 = \sum_{(i,j,k)\in A_2} s(i,j,k)^{2t} \sum_{|\alpha^{Re}|=(i,j,k)} \dim(R_n(\alpha))$$

Then, for sufficiently large n,

$$Q_3 \leq 12e^{-c} + 4e^{-\frac{cf(n)}{10}+f(n)}$$

Proof: Manipulating as in Lemma 4.50 gives an equation analogous to Equation (4.49) above:

$$Q_3 \leq \sum_{(i,j,k)\in A_3} e^{-2ts_1(i)+2ts_2(j)-2ts_3(k)} a(i,j,k)\frac{i!k!}{j!}$$

Do the above sum over k, then over j, then finally over i. Using Lemma 4.49,

$$\begin{aligned} Q_3 &\leq \sum_{(i,j,k)\in A_3} e^{-2ts_1(i)+2ts_2(j)-2ts_3(k)} \frac{f(n)^i g(n)^{i-j}(n-f(n))^{k-j}(n-g(n))^k}{(i-j)!(k-j)!j!} \\ &\leq \sum_{(i,j)} e^{-2ts_1(i)+2ts_2(j)} \frac{f(n)^i g(n)^{i-j}}{(i-j)!j!} \sum_{k:(i,j,k)\in A_3} e^{-2ts_3(k)} \frac{(n-f(n))^{k-j}(n-g(n))^k}{(k-j)!} \end{aligned} \tag{4.50}$$

Fix i and j and do the sum over k:

Summing over k: If $(i,j,k) \in A_3$, then by definition $k \leq (n-g(n))/5$. Also, $k \geq j$, giving

$$\sum_{k=j}^{(n-g(n))/5} e^{-2ts_3(k)} \frac{(n-f(n))^{k-j}(n-g(n))^k}{(k-j)!}$$

Denote the kth term of the above sum by r_k. The idea will be to show that r_{k+1}/r_k is less than $1/2$, and thus to bound the sum by $2r_j$. Explicitly,

$$\begin{aligned} \frac{r_{k+1}}{r_k} &= e^{-2ts_3(k+1)+2ts_3(k)} \frac{(n-f(n))(n-g(n))}{k-j+1} \\ &\leq \exp\left(\frac{4t(2k-n+g(n))}{n+2\Delta}\right) \frac{(n-g(n))^2}{k-j+1} \\ &\leq \exp\left(\frac{-12t(n-g(n))}{5(n+2\Delta)}\right)(n-g(n))^2 \end{aligned}$$

and again using Lemma 4.36 with $x_n = f(n)$ and $y_n = n - g(n)$, for sufficiently large n, $\frac{t}{n+2\Delta} \geq \frac{f(n)}{4f(n)} \geq 5\frac{\log(n-g(n))}{n-g(n)}$, and thus

$$\begin{aligned}\frac{r_{k+1}}{r_k} &\leq \exp\left(-12\log(n-g(n))\right)(n-g(n))^2 \\ &= (n-g(n))^{-10} \leq \frac{1}{2}\end{aligned}$$

for sufficiently large n. Therefore,

$$\begin{aligned}\sum_{k=j}^{(n-g(n))/5} e^{-2ts_3(k)}\frac{(n-f(n))^{k-j}(n-g(n))^k}{(k-j)!} &= \sum_{k=j}^{(n-g(n))/5} r_k \leq 2r_j \\ &= 2e^{-2ts_3(j)}(n-g(n))^j\end{aligned}$$

Plugging this back into Equation (4.50),

$$\begin{aligned}Q_3 &\leq \sum_{(i,j)} e^{-2ts_1(i)+2ts_2(j)}\frac{f(n)^i g(n)^{i-j}}{(i-j)!j!}\left(2e^{-2ts_3(j)}(n-g(n))^j\right) \\ &\leq 2\sum_{i=1}^{f(n)} e^{-2ts_1(i)}f(n)^i \sum_{j=0}^{i} e^{2ts_2(j)-2ts_3(j)}\frac{g(n)^{i-j}(n-g(n))^j}{(i-j)!j!} \qquad (4.51)\end{aligned}$$

again using the fact that if $\left|\alpha^{Re}\right| = (i,j,k)$, then $i \geq j$, and also $i \leq f(n)$. Now fix i, and do the sum over j.

Summing over j: The sum to be bounded is

$$\sum_{j=0}^{i} e^{2ts_2(j)-2ts_3(j)}\frac{g(n)^{i-j}(n-g(n))^j}{(i-j)!j!}$$

From Definition 4.38,

$$s_2(j) - s_3(j) = -\frac{2j(n-f(n))}{n+2\Delta}$$

Hence the sum simplifies to

$$\sum_{j=0}^{i} \exp\left(-\frac{4tj(n-f(n))}{n+2\Delta}\right)\frac{g(n)^{i-j}(n-g(n))^j}{(i-j)!j!}$$

By a slight abuse of notation, let r_j again be the jth summand of the above sum. As above, bound the ratio between r_{j+1} and r_j to bound the sum. Here,

$$\begin{aligned}\frac{r_{j+1}}{r_j} &= \exp\left(-\frac{4t(n-f(n))}{n+2\Delta}\right)\frac{(n-g(n))(i-j)}{g(n)(j+1)} \\ &\leq \exp\left(-\frac{4t(n-f(n))}{n+2\Delta}\right)(n-g(n))f(n)\end{aligned}$$

Now, using Lemma 4.36 with $x_n = f(n)$ and $y_n = n - f(n)$, for sufficiently large n, $\frac{t}{n+2\Delta} \geq \frac{\log f(n)}{4f(n)} \geq \frac{\log(n-f(n))}{n-f(n)}$. Thus,

$$\begin{aligned}\frac{r_{j+1}}{r_j} &= \exp\left(-4\log(n-f(n))\right)(n-g(n))f(n) \\ &\leq \frac{(n-g(n))f(n)}{(n-f(n))^4} < \frac{1}{2}\end{aligned}$$

for sufficiently large n, and as before,

$$\sum_{j=0}^{i} e^{2ts_2(j)-2ts_3(j)}\frac{g(n)^{i-j}(n-g(n))^j}{(i-j)!j!} = \sum_{j=0}^{i} r_j \leq 2r_0 = \frac{2g(n)^i}{i!}$$

Plugging this back into Equation (4.51),

$$Q_3 \leq 4\sum_{i=1}^{f(n)} e^{-2ts_1(i)}\frac{f(n)^i g(n)^i}{i!}. \tag{4.52}$$

Summing over i: Break up the above sum into two pieces: $i \leq \frac{f(n)}{5}$ and $i > \frac{f(n)}{5}$. Bound the first case first. Consider the summation

$$\sum_{i=1}^{f(n)/5} e^{-2ts_1(i)}\frac{f(n)^i g(n)^i}{i!}$$

Recall that

$$s_1(i) = \begin{cases} \frac{2i(f(n)-i+1)}{n+2\Delta} & i < \frac{f(n)}{2} \\ \frac{if(n)}{n+2\Delta} & i \geq \frac{f(n)}{2}\end{cases} \tag{4.53}$$

Thus, for $i < \frac{f(n)}{5}$,

$$\begin{aligned}2ts_1(i) &= 2\left(\frac{(n+2\Delta)(\log f(n) + \log g(n))}{4f(n)} + c\frac{n+2\Delta}{4f(n)}\right)\frac{2i(f(n)-i+1)}{n+2\Delta} \\ &= (\log f(n) + \log g(n) + c)\left(i - \frac{i^2-i}{f(n)}\right)\end{aligned}$$

Now, $i - \frac{i^2-i}{f(n)}$ is a quadratic function in i which corresponds to an upside down parabola, and as such is minimized at the endpoints of an interval. Plugging in $i = 1$ and $i = \frac{f(n)}{5}$, for $1 \leq i \leq \frac{f(n)}{5}$, $i - \frac{i^2-i}{f(n)} \geq 1$. This gives that

$$2ts_1(i) \geq (\log f(n) + \log g(n))\left(i - \frac{i^2-i}{f(n)}\right) + c$$

Thus,

$$\begin{aligned}e^{-2ts_1(i)} &\leq \exp\left(-(\log f(n) + \log g(n))\left(i - \frac{i^2-i}{f(n)}\right) - c\right) \\ &= f(n)^{-i}g(n)^{-i}f(n)^{\frac{i^2-i}{f(n)}}g(n)^{\frac{i^2-i}{f(n)}}e^{-c}\end{aligned}$$

Hence,

$$\sum_{i=1}^{f(n)/5} e^{-2ts_1(i)}\frac{f(n)^i g(n)^i}{i!} \leq e^{-c}\sum_{i=1}^{f(n)/5} f(n)^{\frac{i^2-i}{f(n)}}g(n)^{\frac{i^2-i}{f(n)}}\frac{1}{i!}$$

By yet another slight abuse of notation, let r_i be the ith summand of the above right-hand sum – bound the ratio between consecutive terms to find bounds on the sum. Then, since $g(n) \leq f(n)$,

$$\frac{r_{i+1}}{r_i} = \frac{1}{i+1}f(n)^{\frac{2i}{f(n)}}g(n)^{\frac{2i}{f(n)}} \leq \frac{1}{i+1}f(n)^{\frac{4i}{f(n)}}$$

Now, consider the above expression on the right-hand side as a function of i. By differentiating, it is easy to check that it is increasing for $i+1 > \frac{f(n)}{4\log f(n)}$ and decreasing for $i+1 < \frac{f(n)}{4\log f(n)}$. Thus, in order to find its maximum, just check the endpoints. Plugging $i = 1$, the result is $\frac{1}{2}f(n)^{4/f(n)}$. Since $f(n) \to \infty$, this approaches $1/2$ as $n \to \infty$. Thus, for sufficiently large n, this is at most $2/3$. Plugging in $i = f(n)/5$,

$$\frac{1}{f(n)/5+1}f(n)^{\frac{4f(n)/5}{f(n)}} \leq 5f(n)^{-1/5}$$

which clearly goes to 0, and hence is less than 2/3 for sufficiently large n. Therefore, r_{i+1}/r_i is less than 2/3 for sufficiently large n, and hence

$$\sum_{i=1}^{f(n)/5} f(n)^{\frac{i^2-i}{f(n)}} g(n)^{\frac{i^2-i}{f(n)}} \frac{1}{i!} = \sum_{i=1}^{f(n)/5} r_i \leq 3r_1 = 3$$

and therefore

$$\sum_{i=1}^{f(n)/5} e^{-2ts_1(i)} \frac{f(n)^i g(n)^i}{i!} \leq 3e^{-c} \tag{4.54}$$

It remains to bound the sum for $i \geq f(n)/5$. This is

$$\sum_{i=f(n)/5+1}^{f(n)} e^{-2ts_1(i)} \frac{f(n)^i g(n)^i}{i!}$$

It is easy to check from the restated definition in Equation (4.53) that for $i > \frac{f(n)}{5}$,

$$s_1(i) \geq \frac{if(n)}{n+2\Delta}$$

Thus,

$$\begin{aligned}
e^{-2ts_1(i)} &\leq \exp\left(-2\left(\frac{(n+2\Delta)(\log f(n) + \log g(n))}{4f(n)} + c\frac{(n+2\Delta)}{4f(n)}\right)\frac{if(n)}{n+2\Delta}\right) \\
&= \exp\left(-\frac{i(\log f(n) + \log g(n) + c)}{2}\right) \\
&= e^{-\frac{ic}{2}} f(n)^{-\frac{i}{2}} g(n)^{-\frac{i}{2}}
\end{aligned}$$

Therefore, since $g(n) \leq f(n)$,

$$\begin{aligned}
\sum_{i=f(n)/5+1}^{f(n)} e^{-2ts_1(i)} \frac{f(n)^i g(n)^i}{i!} &\leq \sum_{i=f(n)/5+1}^{f(n)} e^{-\frac{ic}{2}} \frac{f(n)^{\frac{i}{2}} g(n)^{\frac{i}{2}}}{i!} \\
&\leq e^{-\frac{cf(n)}{10}} \sum_{i=f(n)/5+1}^{f(n)} \frac{f(n)^i}{i!} \\
&\leq e^{-\frac{cf(n)}{10}+f(n)}
\end{aligned}$$

Combining the above with Equation (4.54),

$$\sum_{i=1}^{f(n)} e^{-2ms_1(i)} \frac{f(n)^i g(n)^i}{i!} \leq 3e^{-c} + e^{-\frac{cf(n)}{10}+f(n)}$$

and thus from Equation (4.52),

$$Q_3 \leq 12e^{-c} + 4e^{-\frac{cf(n)}{10}+f(n)}$$

as required. □

Proof of Upper Bound in Theorem 4.9. Combining Equation (4.48) with Lemmas 4.48, 4.50 and 4.51,

$$\begin{aligned}\sum_{1\neq\Lambda_1(\alpha)\geq 0} \dim(R_n(\alpha))\Lambda_1(\alpha)^{2t} &\leq \sum_{(i,j,k)\neq(0,0,0)} s(i,j,k)^{2t} \sum_{|\alpha^{Re}|=(i,j,k)} \dim(R_n(\alpha)) \\ &\leq n^{-n} + 2(n-g(n))^{-6} + 12e^{-c} + 4e^{-\frac{cf(n)}{10}+f(n)}\end{aligned}$$

Thus, for $c > 10$ and n sufficiently large,

$$\sum_{1\neq\Lambda_1(\alpha)\geq 0} \dim(R_n(\alpha))\Lambda_1(\alpha)^{2t} \leq 16e^{-c}$$

Now, from Equation (4.29),

$$\left\|P^t(x,\cdot) - \pi\right\|_{2,\pi} = \sqrt{\sum_{\alpha} \dim(R_n(\alpha))\Lambda_1(\alpha)^{2t}}$$

and thus, for $c > 10$ and for n sufficiently large,

$$\left\|P^t(x,\cdot) - \pi\right\|_{2,\pi} \leq 4e^{-\frac{c}{2}}$$

as required. □

Chapter 5

Coupling for the Random Transposition Walk

5.1 Introduction

This chapter provides a coupling argument that shows that the random transposition walk mixes in $O(n \log n)$ time. Before proceeding with the history and motivation, here is a description of the walk. Lay n cards out in a row, and pick one card uniformly with your right hand, and another card independently uniformly with your left hand (note that you may have picked the same card.) Then, swap the cards – this is an extremely simple shuffling scheme for n cards.

The existence of a coupling argument showing an $O(n \log n)$ mixing time is a long-standing open problem. Due to its simplicity and symmetry, the random transposition walk was one of the first ones considered in burgeoning field of Markov chains mixing times. As noted in [14], the mixing time of this walk was first bounded by Aldous in 1980, who showed that it must be between order n and n^2 and conjectured that it must be of order $n \log n$. As noted earlier, this was proved in 1981 in "Generating a random permutation with random transpositions" by Diaconis and Shahshahani [14]. This paper uses Fourier analysis to show that the walk experiences a cut-off, as defined in Equation (1.5), mixing in a window of order n around time $\frac{1}{2} n \log n$.

The beautiful proof in [14] uses the tools of representation theory and Fourier analysis, and hence is non-probabilistic. While a purely probabilistic strong stationary time proof for an $O(n \log n)$ mixing time was discovered by Broder in 1985 [6], a coupling argument proved to be more elusive. The main difficulty is due to the fact that a Markovian coupling cannot succeed; indeed, Lemma 5.8 shows that such an approach can never prove a bound of order better than n^2. As noted in Remark 2.8, it has been shown by Griffeath [21] and then Pitman [44] that a maximal coupling must exist, but it evidently has to be non-Markovian. There has been continued interest in finding such a coupling – for example, Peres named it as an interesting open problem in [43]. This chapter resolves this problem. (Another approach for finding such a non-Markovian coupling can be seen in the preprint “Mixing times via super-fast coupling" [33].)

This question is approached here by first projecting the random transposition walk to conjugacy classes. As noted in Chapter 1, this projection is also a Markov chain, called a split-merge random walk [47]. Using the fact that the random transposition walk started from the identity is constant on conjugacy classes, it suffices to find the mixing time of the split-merge random walk. The path coupling technique of Bubley and Dyer [8] is used to examine the split-merge random walk. However, this is not straightforward, since in the worst case scenario, the split-merge random walk contracts by only $1 - \frac{1}{n^2}$.

It is shown here that on average, the split-merge random walk does indeed contract by $1 - \frac{1}{n}$, enabling the use of path coupling to conclude that the walk mixes in $O(n \log n)$ time. This argument does not, however, show cut-off: indeed, as noted in Remark 5.36 below, the constant in front of the $n \log n$ is very large. To show that the contraction coefficient is of the right order, the techniques of Schramm from his paper “Compositions of random transpositions" [47] are used. He shows that large cycles in the random transposition walk emerge after time $\frac{n}{2}$, and then proves the law for the scaled cycles. Methods from “Compositions of random transpositions" have given rise to the wonderful paper “Mixing times for random k-cycles and coalescence-fragmentation chains" by Berestycki, Schramm, and Zeitouni [4], which uses probabilistic techniques to get the right answer for a generalization of the random transposition walk.

5.2 Background and Definitions

The main result of this chapter is the following theorem:

Theorem 5.1. *There exists a coupling argument that shows that the random transposition walk on S_n mixes in time of order $n \log n$.*

Before launching into the proof, it is instructive to consider the many ways an $O(n \log n)$

mixing time has been obtained for this walk, as well as the uses of the result. This bound was first obtained by Diaconis and Shahshahani in [14]; the techniques of their proof are discussed in some detail in Chapter 4. This result is beautiful and extremely precise; however, the scope of the technique is limited as it requires fully diagonalizing the random walk. As seen in the previous chapter, this is possible for a number of walks, including walks that are not random walks on groups. Nonetheless, this is a drawback to the method. This result is also extremely useful for comparison theory. As shown by Diaconis and Saloff-Coste in [13], the Dirichlet form can be used to compare all the eigenvalues of the chain, resulting in good bounds for a variety of walks. For example, Jonasson uses this result in [31] to show that the overlapping cycle shuffle mixes in $O(n^3 \log n)$ time.

As noted above, the first probabilistic proof of the result was by Broder [6] and used *strong stationary times*: stopping times T such that the conditional distribution of X_T given T is stationary. Since the stationary distribution for the random transposition walk is uniform, this is equivalent to stating that for all $\sigma \in S_n$ and all positive integers k,

$$\mathbb{P}(X_T = \sigma \mid T = k) = \frac{1}{n!}$$

The following is Broder's strong stationary time argument, as summarised in Chapter 9 of [35]. Let R_t and L_t be the cards chosen by the right and left hand, respectively. Start the process with no marked cards, and use the following marking scheme: at each step, mark a card R_t if R_t is unmarked, and either (a) L_t is marked or (b) $R_t = L_t$. Define the stopping time T to be the first time all n cards are marked. It is easy to show that this is indeed a strong stationary time, and that T is around $2n \log n$. This argument provides an $O(n \log n)$ mixing time, but not the correct constant. It was improved by Matthews [39] in 1988 by creating a more complicated rule for marking the cards. This argument showed a cut-off for the walk at time $\frac{1}{2} n \log n$. These arguments are probabilistic and intuitive, and eluciate the reasons for the mixing time in a way that Fourier analysis does not. However, they are heavily reliant on the symmetry of the random transposition walk and as such are difficult to generalise.

The recent paper by Berestycki, Schramm and Zeitouni [4] uses a different approach. Their technique provides the correct answer for the following generalization of the Markov chain: instead of using a uniformly chosen random transposition at each step, a random k-cycle is used. This paper obtains the correct $\frac{1}{k} n \log n$ answer for any fixed k. Like this chapter, this result begins by projecting the walk to conjugacy classes and makes use of the results of Schramm in [47]. The tools of both this result and Schramm's original paper are graph theoretic: for example, a transposition is considered to be an edge in a random graph process on n vertices. Unfortunately, this exciting method again requires considerable symmetry, since the projection to conjugacy classes has to be a Markov chain. This is also a drawback of the coupling approach of this chapter.

Another intriguing technique by Burton and Kovchegov [33] uses non-Markovian coupling. While the probability community is yet to fully understand the argument, the idea is that the standard coupling argument by Aldous which results in in $O(n^2)$ bound can be improved by 'looking into the future.' A non-Markovian argument with a somewhat similar flavor has previously been implemented for the coloring chain by Hayes and Vigoda [25]. Here's a very approximate sketch of the idea for random transpositions: say that a pair (σ, τ) in S_n currently differs in the cards labeled i and j. The standard coupling for this pair transposes the cards with the same labels in both σ and τ, unless the next tranposition is (i, j). However, it is possible to do something different: if the next step transposes cards labeled i and k in σ, the next step in τ can transpose either cards labeled i and k or cards labeled j and k. If the coupling is Markovian, then the choice makes no difference; however, 'looking into the future' can substantially improve the bounds. In work stemming from an unrelated project, I hope to show this for a number of different walks in an upcoming paper.

The argument in this chapter proceeds by projecting the walk to conjugacy classes. It is a well-known result that the conjugacy classes of S_n are indexed by partitions of n. Recall that a partition of n is an m-tuple $(a_1, a_2, \ldots, a_m)$ of positive integers that sum to n, where m can be any integer, and $a_1 \geq a_2 \geq \cdots \geq a_m$. Let $\mathcal{P}_n$ be the set of partitions of n. The projection of the random transposition walk on S_n to conjugacy classes is also a Markov chain, called a *split-merge* random walk. It is often referred to as a coagulation-fragmentation chain, and it has been extensively studied – see [12] for some references.

Definition 5.2. *Assume the random walk is currently at the partition* $(a_1, a_2, \ldots, a_m)$. *Then, there are three possibilities for the next move: either merge a pair of parts, split a part into two pieces, or stay in place. (All of these moves are followed by rearranging the new parts to be in non-decreasing order.)*

- **Split:** *A pair* a_i *can be replaced by the pair* $(r, a_i - r)$. *For each* r *between* 1 *and* $a_i - 1$, *the probability of this particular split is* $\frac{a_i}{n^2}$.

Note that this phrasing takes the order into account: here, a more convenient phrasing is the following: for each $r < \frac{a_i}{2}$*, split* a_i *into* $\{r, a_i - r\}$ *with probability* $\frac{2a_i}{n^2}$*. If* a_i *is even and* $r = \frac{a_i}{2}$*, split* a_i *into* $\{r, a_i - r\}$ *with probability* $\frac{a_i}{n^2}$*.*

- **Merge:** *Replace the parts* a_i *and* a_j *by* $a_i + a_j$*. This is done with probability* $\frac{2a_ia_j}{n^2}$*.*
- **Stay in Place:** *Stay at the partition* $(a_1, a_2, \ldots, a_m)$ *with probability* $\frac{1}{n}$*.*

Example 5.3. Here is an example of the split-merge random walk. Let $n = 5$, and assume the walk is currently at $(4, 1)$. Then, the next step X_1 is distributed as follows:

$$X_1 = \begin{cases} (5) & \text{with probability } \frac{8}{25} \\ (4,1) & \text{with probability } \frac{1}{5} \\ (3,1,1) & \text{with probability } \frac{8}{25} \\ (2,2,1) & \text{with probability } \frac{4}{25} \end{cases}$$

The primary walk under consideration is the split-merge random walk, but for some of the proofs, the original transposition walk is needed. With that in mind, make the following two definitions:

Definition 5.4. *For* $\alpha \in S_n$*, define* $\mathrm{Cyc}(\alpha)$ *to be the partition corresponding to the cycle type of* α*. For* $\sigma \in \mathcal{P}_n$*, let*

$$\mathrm{Perm}(\sigma) = \{\alpha \in S_n \mid \mathrm{Cyc}(\alpha) = \sigma\}$$

be the set of all permutations with cycle type σ*.*

Definition 5.5. *Let* $(X_t)_{t\geq 0}$ *denote the split-merge random walk, and let* $(\bar{X}_t)_{t\geq 0}$ *denote the random transposition walk, so that for all* t*,*

$$X_t = \mathrm{Cyc}(\bar{X}_t)$$

Furthermore, let P *and* π *be the transition matrix and stationary distribution for* $(X_t)_{t\geq 0}$*, respectively, and define* $\bar{P}$ *and* $\bar{\pi}$ *analagously for* $(\bar{X}_t)_{t\geq 0}$*.*

The next argument shows it suffices to consider the split-merge random walk. The following proof take a little bit of space to write down, but is actually very simple – the key idea is that the random transpositon walk started at the identity is always uniformly distributed over each conjugacy class.

Lemma 5.6. *Let $P, \bar{P}, \pi$ and $\bar{\pi}$ be defined as in Definition 5.5 above. Then,*

$$\max_{\alpha \in S_n} \left\| \bar{P}^t(\alpha, \cdot) - \bar{\pi} \right\|_{TV} \leq \max_{\sigma \in \mathcal{P}_n} \left\| P^t(\sigma, \cdot) - \pi \right\|_{TV}$$

Proof: Since the random transposition walk is a random walk on a group, it's vertex transitive. Therefore, for all $\alpha \in S_n$,

$$\left\| \bar{P}^t(\alpha, \cdot) - \bar{\pi} \right\|_{TV} = \left\| \bar{P}^t(id, \cdot) - \bar{\pi} \right\|_{TV}$$

where id is the identity permutation. Thus, it suffices to show that

$$\left\| \bar{P}^t(id, \cdot) - \bar{\pi} \right\|_{TV} \leq \max_{\sigma \in \mathcal{P}_n} \left\| P^t(\sigma, \cdot) - \pi \right\|_{TV}$$

Now, let $\sigma_0 = \mathrm{Cyc}(id) = (1, 1, \ldots, 1)$. It suffices to show that

$$\left\| \bar{P}^t(id, \cdot) - \bar{\pi} \right\|_{TV} = \left\| P^t(\sigma_0, \cdot) - \pi \right\|_{TV} \tag{5.1}$$

Since the split-merge random walk is a projection of the random transposition walk, for $\sigma \in \mathcal{P}_n$,

$$\pi(\sigma) = \sum_{\alpha \in \mathrm{Perm}(\sigma)} \bar{\pi}(\alpha) = \frac{|\mathrm{Perm}(\sigma)|}{n!} \tag{5.2}$$

since $\bar{\pi}$ is the uniform distribution on S_n. Similarly,

$$P^t(\sigma_0, \sigma) = \sum_{\alpha \in \mathrm{Perm}(\sigma)} \bar{P}^t(id, \alpha)$$

Furthermore, note that both the identity permutation and the the random transposition walk are symmetric with respect to $\{1, 2, \ldots, n\}$. Hence for any α_1, α_2 with the same cycle structure, $\bar{P}^t(id, \alpha_1) = \bar{P}^t(id, \alpha_2)$ for all t. Combining this with the equation above shows that for $\alpha \in \mathrm{Perm}(\sigma)$,

$$P^t(\sigma_0, \sigma) = |\mathrm{Perm}(\sigma)| \, \bar{P}^t(id, \alpha) \tag{5.3}$$

Using Equations (5.2) and (5.3),

$$\sum_{\alpha \in \mathrm{Perm}(\sigma)} \left| \bar{P}^t(id, \alpha) - \frac{1}{n!} \right| = \left| P^t(\sigma_0, \sigma) - \pi(\sigma) \right|$$

Finally, putting all this together,

$$\begin{aligned} \left\| \bar{P}^t(id, \cdot) - \bar{\pi} \right\|_{TV} &= \frac{1}{2} \sum_{\alpha \in S_n} \left| \bar{P}^t(id, \alpha) - \frac{1}{n!} \right| = \frac{1}{2} \sum_{\sigma \in \mathcal{P}_n} \sum_{\alpha \in \mathrm{Perm}(\sigma)} \left| \bar{P}^t(id, \alpha) - \frac{1}{n!} \right| \\ &= \frac{1}{2} \sum_{\sigma \in \mathcal{P}_n} \left| P^t(\sigma_0, \sigma) - \pi(\sigma) \right| = \left\| P^t(\sigma_0, \cdot) - \pi \right\|_{TV} \end{aligned}$$

which proves Equation (5.1), as desired. □

Remark 5.7. Although it is not needed, it is very easy to use the triangle inequality to prove the opposite inequality to the one in Lemma 5.1. Hence, the two quantities are actually equal.

Before proceeding to sketch the upcoming proof, it is shown that a Markovian coupling for the random transposition walk cannot hope to give an $O(n \log n)$ mixing time.

Lemma 5.8. *A Markovian coupling $(\bar{X}_t, \bar{Y}_t)$ of the random transpositon walk takes at least $\Omega(n^2)$ time to meet.*

Proof: It easy to check that wherever the two random transposition walks currently are, if $\bar{X}_t \neq \bar{Y}_t$, then

$$\mathbb{P}\left(\bar{X}_{t+1} = \bar{Y}_{t+1}\right) \leq \frac{6}{n^2}$$

To verify this, note that if $\bar{X}_t$ and $\bar{Y}_t$ differ only in the transposition (i, j), then the only way to meet is to transpose i and j in one of them, and to stay in place in the other one; similar arguments hold if $\bar{X}_t$ and $\bar{Y}_t$ are two transpositions apart, and in all other cases, the probability of meeting at the next step is 0. Combining the above inequality with the Markov property leads to the desired result. □

Turn next to an explanation of the idea behind the coupling. The argument uses path coupling – that is, coupling a pair of split-merge random walks started at a neighboring pair of partitions (σ, τ). Define neighboring pairs to be precisely the pairs which are one step away in the split-merge random walk. Then, define a coupled process (X_t, Y_t) such that

$X_0 = \sigma$ and $Y_0 = \tau$, making sure that the distance between X_t and Y_t at each step is at most 1. Here are some useful definitions.

Definition 5.9. *For σ and τ partitions of n, define $\rho(\sigma, \tau)$ to be the distance between σ and τ induced by the split-merge random walk; that is, $\rho(\sigma, \tau)$ is the number of split-merge steps it takes to get from σ to τ.*

The next definiton is useful for finding a lower bound on the probability of coupling at each step given the current location of the two walks.

Definition 5.10. *Let σ and τ be partitions of n such that $\rho(\sigma, \tau) = 1$. Then σ and τ are exactly one merge away, so rearranging parts appropriately and without loss of generality letting σ be the partition with more parts,*

$$\begin{aligned} \sigma &= (a_1, a_2, \ldots, a_m, b, c) \\ \tau &= (a_1, a_2, \ldots, a_m, b + c) \end{aligned} \tag{5.4}$$

where $b \leq c$. Then, define

$$s(\tau, \sigma) = s(\sigma, \tau) = b \quad \textit{and} \quad m(\tau, \sigma) = m(\sigma, \tau) = c \tag{5.5}$$

That is, since σ and τ differ in the parts b, c and $b + c$, $s(\sigma, \tau)$ is the smallest part in which they differ, and $m(\sigma, \tau)$ is the medium part in which they differ.

For later use, define $m(\sigma, \sigma) = n$ and $s(\sigma, \sigma) = \frac{n}{2}$.

In the next section, the coupling is given along with the following lemma:

Lemma 5.11. *Assume that $(X_t, Y_t) = (\sigma, \tau)$, where $\rho(\sigma, \tau) = 1$. Then, $\rho(X_{t+1}, Y_{t+1}) \leq 1$, and*

$$\mathbb{P}(X_{t+1} = Y_{t+1}) \geq \frac{4s(\sigma, \tau)}{n^2}$$

That is, the chain stays at most distance 1 apart, and gives the above lower bound for the probability of coupling.

After proving the above lemma, it is shown below that after order n steps, $s(X_t, Y_t)$ is on average of order n. The lemma then implies that the probability of coupling at each step is of order $\frac{1}{n}$, which will show that there is a high probability of coupling after order n steps. Using the fact that the diameter of the set of partitions is no greater than n, the path coupling theorem (Theorem 2.11) shows that the random transposition walk mixes in $O(n \log n)$ time.

5.3 The Coupling

This section defines the coupling for neighboring pairs for the split-merge random walk, and proves Lemma 5.11. The coupling is defined in such a way that the distance between X_t and Y_t at each step is at most 1 for all t. As usual, once the two chains meet, they are run together.

Definition 5.12. *Consider the next step (X_1, Y_1) of a coupling which is currently at $(X_0, Y_0) = (\sigma, \tau)$, where $\rho(\sigma, \tau) = 1$ and*

$$\sigma = (a_1, a_2, \dots, a_m, b, c)$$
$$\tau = (a_1, a_2, \dots, a_m, b + c)$$

where $b \le c$. There are a number of cases, considered in the following order: go through the possible moves in σ, then provide corresponding moves in τ.

- **Operations only using the a_i:** *If a_i and a_j are merged in σ for any i and j, perform the same operation in τ. Similarly, if a_i is split in σ into $\{r, a_i - r\}$, do the same for a_i in τ. Then,*

$$X_1 = (a'_1, \dots, a'_k, b, c)$$
$$Y_1 = (a'_1, \dots, a'_k, b + c)$$

for the appropriate $\{a'_1, a'_2, \dots, a'_k\}$.

- **Merging b or c and a_i:** *If b and a_i are merged in σ, merge $b + c$ and a_i in τ. If c and a_i are merged in σ, also merge $b + c$ and a_i in τ. In the first case,*

$$X_1 = (a'_1, \dots, a'_{m-1}, b + a_i, c)$$
$$Y_1 = (a'_1, \dots, a'_{m-1}, b + c + a_i)$$

where $\{a'_1, a'_2, \dots, a'_{m-1}\} = \{a_1, a_2, \dots, a_m\}/\{a_i\}$. The case where c and a_i are merged in σ is analogous.

- **Splitting b or c:** *If b is split in σ into $\{r, b-r\}$ where $r \leq \frac{b}{2}$, then split $b+c$ in τ into $\{r, b+c-r\}$. Similarly, if c is split in σ into $\{r, c-r\}$ where $r \leq \frac{c}{2}$, then split $b+c$ in τ into $\{r, b+c-r\}$. The first case results in*

$$X_1 = (a_1, \ldots, a_m, r, b-r, c)$$
$$Y_1 = (a_1, \ldots, a_m, r, b+c-r)$$

  The second case, where c is split into $\{r, c-r\}$, is analogous.

- **Staying in place:** *If the walk stays in place in σ, it is coupled with either staying in place in τ or with splitting $b+c$ in τ into $\{b, c\}$. Since splitting $b+c$ into $\{b, c\}$ may have already been coupled with splitting c into $\{b, c-b\}$, let p be the remaining probability of splitting $b+c$ into $\{b, c\}$. Then, couple staying in place in σ with splitting $b+c$ into $\{b, c\}$ in τ with probability $\min\left(p, \frac{1}{n}\right)$. This results in*

$$X_1 = (a_1, \ldots, a_m, b, c)$$
$$Y_1 = (a_1, \ldots, a_m, b, c)$$

  That is, the chains will couple.

  Couple staying in place in σ to staying in place in τ the rest of the time – that is, with probability $\frac{1}{n} - \min\left(p, \frac{1}{n}\right)$.

- **Merging b and c:** *Couple merging b and c in σ to any remaining possibilities in τ. It is easy to check that these are either staying in place or splitting $b+c$ into $\{r, b+c-r\}$. The first leads to the chains coupling; the second leads to*

$$X_1 = (a_1, \ldots, a_m, b+c)$$
$$Y_1 = (a_1, \ldots, a_m, r, b+c-r)$$

  for some r.

Example 5.13. As this coupling looks fairly complicated, here are a couple of examples. The possible pairs for (X_1, Y_1) are listed, as well as the probability of each pair.

1. Let $(X_0, Y_0) = (\sigma, \tau) = ((2,3),(5))$. Here, there are no a_i, $b = 2$, $c = 3$, and $b + c = 5$.

A description is provided for each pair of moves.

$$(X_1, Y_1) = \begin{cases} ((1,1,3),(1,4)), p = \frac{2}{25} & \text{split 2 into } \{1,1\} \text{ in } \sigma\text{, split 5 into } \{1,4\} \text{ in } \tau \\ ((1,2,2),(1,4)), p = \frac{6}{25} & \text{split 3 into } \{1,2\} \text{ in } \sigma\text{, split 5 into } \{1,4\} \text{ in } \tau \\ ((2,3),(2,3)), p = \frac{5}{25} & \text{stay at } \sigma\text{, split 5 into } \{2,3\} \text{ in } \tau \\ ((5),(1,4)), p = \frac{2}{25} & \text{merge 2 and 3 in } \sigma\text{, split 5 into } \{1,4\} \text{ in } \tau \\ ((5),(2,3)), p = \frac{5}{25} & \text{merge 2 and 3 in } \sigma\text{, split 5 into } \{2,3\} \text{ in } \tau \\ ((5),(5)), p = \frac{5}{25} & \text{merge 2 and 3 in } \sigma\text{, stay at } \tau \end{cases}$$

2. Let $(X_0, Y_0) = (\sigma, \tau) = ((2,1,3),(2,4))$, written with the above convention that the parts σ and τ disagree on are written last. Here, $a_1 = 2$, $b = 1$, $c = 3$, and $b + c = 4$.

$$(X_1, Y_1) = \begin{cases} ((1,1,1,3),(1,1,4)), p = \frac{2}{36} & \text{split 2 into } \{1,1\} \text{ in } \sigma \text{ and } \tau \\ ((3,3),(6)), p = \frac{4}{36} & \text{merge 2 and 1 in } \sigma\text{, merge 2 and 4 in } \tau \\ ((1,5),(6)), p = \frac{12}{36} & \text{merge 2 and 3 in } \sigma\text{, merge 2 and 4 in } \tau \\ ((2,1,1,2),(2,1,3)), p = \frac{6}{36} & \text{split 3 as } \{1,2\} \text{ in } \sigma\text{, split 4 as } \{1,3\} \text{ in } \tau \\ ((2,1,3),(2,1,3)), p = \frac{2}{36} & \text{stay at } \sigma\text{, split 4 into } \{1,3\} \text{ in } \tau \\ ((2,4),(2,2,2)), p = \frac{4}{36} & \text{merge 1 and 3 in } \sigma\text{, split 4 into } \{2,2\} \text{ in } \tau \\ ((2,4),(2,4)), p = \frac{2}{36} & \text{merge 1 and 3 in } \sigma\text{, stay at } \tau \\ ((2,1,3),(2,4)), p = \frac{4}{36} & \text{stay at } \sigma \text{ and } \tau \end{cases}$$

Going back to the general case, here is a check that the above definition provides the correct marginal distribution for Y_1. Note that given the way that the coupling was defined, it clearly provides the correct distribution for X_1.

Lemma 5.14. *The coupling in Definition 5.12 has the correct marginal distribution for Y_1.*

Proof: Since σ and τ share the parts a_i, the operations only using the a_i are distributed identically in both and hence pose no problem. Furthermore,

$$\begin{aligned} \mathbb{P}(\text{Merging } b \text{ and } a_i \text{ in } \sigma) + \mathbb{P}(\text{Merging } c \text{ and } a_i \text{ in } \sigma) &= \frac{2ba_i}{n^2} + \frac{2ca_i}{n^2} = \frac{2(b+c)a_i}{n^2} \\ &= \mathbb{P}(\text{Merging } b + c \text{ and } a_i \text{ in } \tau) \end{aligned}$$

Thus, all the operations involving any a_i work properly.

Consider next operations that only involve b and c in σ. Splitting b into $\{r, b-r\}$ where $r \leq \frac{b}{2}$ in σ is coupled with splitting $b+c$ into $\{r, b+c-r\}$ in τ, and similarly for c. It needs to be checked that this is possible – that is, that the probability of splitting $b+c$ into $\{r, b+c-r\}$ in τ is sufficiently large to accommodate all these moves in σ.

There are a number of possibilities. First of all, if $r \leq \frac{b}{2}$, then clearly $r < \frac{b+c}{2}$, and hence according to Definition 5.2,

$$\begin{aligned}\mathbb{P}(\text{Splitting } b+c \text{ into } \{r, b+c-r\} \text{ in } \tau) = \frac{2(b+c)}{n^2} &= \frac{2b}{n^2} + \frac{2c}{n^2} \\ &\geq \mathbb{P}(\text{Splitting } b \text{ into } \{r, b-r\} \text{ in } \sigma) + \\ &\qquad \mathbb{P}(\text{Splitting } c \text{ into } \{r, c-r\} \text{ in } \sigma)\end{aligned}$$

In this case, the probability of splitting $b+c$ into $\{r, b+c-r\}$ in τ is sufficiently large.

Now, if $\frac{b}{2} < r \leq \frac{c}{2}$, the procedure couples splitting $b+c$ into $\{r, b+c-r\}$ with splitting c into $\{r, c-r\}$. Thus, since in this case r is still less than $\frac{b+c}{2}$,

$$\begin{aligned}\mathbb{P}(\text{Splitting } b+c \text{ into } \{r, b+c-r\} \text{ in } \tau) = \frac{2(b+c)}{n^2} &\geq \frac{2c}{n^2} \\ &\geq \mathbb{P}(\text{Splitting } c \text{ into } \{r, c-r\} \text{ in } \sigma)\end{aligned}$$

which again works.

Finally, if $r > \frac{c}{2}$, splitting $b+c$ into $\{r, b+c-r\}$ is not coupled to splitting either b or c in σ, which obviously does not pose a problem. None of the other moves considered in Definition 5.12 could be an issue, and hence the marginal distribution of Y_1 under this definition is correct. $\square$

The next step proves Lemma 5.11. This states that the coupled chains stay at most one step apart, and that

$$\mathbb{P}(X_{t+1} = Y_{t+1}) \geq \frac{4s(X_t, Y_t)}{n^2}$$

Proof of Lemma 5.11: It should be clear from Definition 5.12 that the coupling stays at most one step apart for all t. To show that if $(X_t, Y_t) = (\sigma, \tau)$, where $\rho(\sigma, \tau) = 1$, then

$$\mathbb{P}(X_{t+1} = Y_{t+1}) \geq \frac{4s(\sigma, \tau)}{n^2}$$

let

$$\sigma = (a_1, \ldots, a_m, b, c)$$
$$\tau = (a_1, \ldots, a_m, b + c)$$

where $b \leq c$. Then by Definition 5.10, $s(\sigma, \tau) = b$.

From Definition 5.12, the chains can couple either if σ stays in place, or if b and c are merged in σ. Consider those two cases separately.

Staying in place in σ: The chains will couple if σ stays in place and $b+c$ is split in τ into $\{b, c\}$. As noted in the definition, these are coupled together with probability $\min\left(p, \frac{1}{n}\right)$, where p is the remaining probability of splitting $b+c$ into $\{b, c\}$ in τ – the probability that this split hasn't already been coupled to something else. To find a lower bound on p, first note that splitting $b+c$ into $\{b, c\}$ in τ couldn't have been coupled with any splits of b in σ. However, it might have been coupled with a split of c in σ. Consider two cases: $c < 2b$ and $c \geq 2b$.

If $c < 2b$, then splitting c into $\{c - b, b\}$ in σ is coupled to splitting $b + c$ into $\{c - b, 2b\}$ in τ since $c - b < b$. This means that nothing is coupled to splitting $b + c$ into $\{b, c\}$, and therefore

$$p = \mathbb{P}(\text{Splitting } b + c \text{ into } \{b, c\} \text{ in } \tau) \geq \frac{b + c}{n^2} \geq \frac{2b}{n^2} \tag{5.6}$$

If $c \geq 2b$, then splitting $b + c$ into $\{b, c\}$ in τ is indeed coupled with splitting c into $\{b, c - b\}$ in σ. In this case, clearly $b \neq c$, and hence

$$\mathbb{P}(\text{Splitting } b + c \text{ into } \{b, c\} \text{ in } \tau) = \frac{2(b + c)}{n^2}$$

Therefore,

$$\begin{aligned} p &\geq \mathbb{P}(\text{Splitting } b + c \text{ into } \{b, c\} \text{ in } \tau) - \mathbb{P}(\text{Splitting } c \text{ into } \{b, c - b\} \text{ in } \sigma) \\ &\geq \frac{2(b + c)}{n^2} - \frac{2c}{n^2} = \frac{2b}{n^2} \end{aligned} \tag{5.7}$$

Equations (5.6) and (5.7) give $p \geq \frac{2b}{n^2}$. Furthermore, note that $b \leq c$, and $b + c \leq n$, and hence $b \leq \frac{n}{2}$. Therefore,

$$\min\left(p, \frac{1}{n}\right) \geq \min\left(\frac{2b}{n^2}, \frac{1}{n}\right) \geq \frac{2b}{n^2} \tag{5.8}$$

Hence,

$$\mathbb{P}(\text{Coupling if staying in place in } \sigma) = \min\left(p, \frac{1}{n}\right) \geq \frac{2b}{n^2} \tag{5.9}$$

Merging b and c in σ: Next, consider the probability of coupling if b and c are merged in σ. Clearly, this would need to be coupled with staying in place in τ. The only other thing that staying in place in τ could have been coupled with so far is staying in place in σ. As noted in Definition 5.12,

$$\mathbb{P}(\text{Both } \sigma \text{ and } \tau \text{ stay in place}) = \frac{1}{n} - \min\left(p, \frac{1}{n}\right)$$

for the same p used above. Thus,

$$\begin{aligned}
\mathbb{P}(\text{Coupling if merging } b \text{ and } c \text{ in } \sigma) &= \mathbb{P}(b \text{ and } c \text{ merged in } \sigma, \tau \text{ stays in place}) \\
&= \frac{1}{n} - \mathbb{P}(\text{Both } \sigma \text{ and } \tau \text{ stay in place}) \\
&= \min\left(p, \frac{1}{n}\right) \geq \frac{2b}{n^2} \qquad (5.10)
\end{aligned}$$

using Equation (5.8) above.

Finally, combining Equations (5.9) and (5.10),

$$\mathbb{P}(X_{t+1} = Y_{t+1}) \geq \frac{4b}{n^2} = \frac{4s(\sigma, \tau)}{n^2} = \frac{4s(X_t, Y_t)}{n^2}$$

as required. □

Continuing with the proof, as sketched out earlier, the rest of this chapter will be concerned with showing that $s(X_t, Y_t)$ is of order n after $O(n)$ time. The next section shows how that proves Theorem 5.1, and provides a summary of the proof.

5.4 Proof of Main Theorem Using $\mathbb{E}\left[s(X_t, Y_t)\right]$

As described above, one of the main tools of this paper is the following theorem:

Theorem 5.15. *There exist constants α and β such that for all $t \geq \alpha n$,*

$$\mathbb{E}\left[s(X_t, Y_t)\right] \geq \beta n$$

This section uses the above result to prove Theorem 5.1. To start, prove the following easy lemma:

Lemma 5.16. *Let (X_t, Y_t) be defined as in Definition 5.12, where as usual $\rho(X_0, Y_0) = 1$. Let α and β be the constants in Theorem 5.15 above. Then,*

$$\mathbb{P}\left(X_{\alpha n+\frac{n}{2}} = Y_{\alpha n+\frac{n}{2}}\right) \geq \beta$$

Proof: Since by Lemma 5.11, $\mathbb{P}(X_t = Y_t)$ is non-decreasing, if $\mathbb{P}(X_t = Y_t) \geq \beta$ for any $t \leq \alpha n + \frac{n}{2}$, the argument is complete. Thus, assume that

$$\mathbb{P}(X_t = Y_t) \leq \beta \tag{5.11}$$

for all $t \leq \alpha n + \frac{n}{2}$.

Clearly,

$$\mathbb{P}(X_{t+1} = Y_{t+1}) = \mathbb{P}(X_t = Y_t) + \mathbb{P}(X_{t+1} = Y_{t+1} \mid X_t \neq Y_t)\mathbb{P}(X_t \neq Y_t)$$

Rearranging, and using Lemma 5.11,

$$\begin{aligned}\mathbb{P}(X_{t+1} = Y_{t+1}) - \mathbb{P}(X_t = Y_t) &\geq \mathbb{E}\left[\frac{4s(X_t,Y_t)}{n^2}\middle| X_t \neq Y_t\right]\mathbb{P}(X_t \neq Y_t) \\ &= \frac{4}{n^2}\mathbb{E}\left[s(X_t,Y_t)\middle| X_t \neq Y_t\right]\mathbb{P}(X_t \neq Y_t) \qquad (5.12)\end{aligned}$$

A lower bound is now needed for the right-hand side. Assume that $t \geq \alpha n$, and hence that $\mathbb{E}\left[s(X_t, Y_t)\right] \geq \beta n$ by Lemma 5.15. Then,

$$\begin{aligned}\mathbb{E}\left[s(X_t,Y_t)\middle| X_t \neq Y_t\right]\mathbb{P}(X_t \neq Y_t) &= \mathbb{E}\left[s(X_t,Y_t)\right] - \mathbb{E}\left[s(X_t,Y_t)\middle| X_t = Y_t\right]\mathbb{P}(X_t = Y_t) \\ &\geq \beta n - \frac{n}{2}\mathbb{P}\left(X_t = Y_t\right)\end{aligned}$$

since if $X_t = Y_t$, $s(X_t, Y_t) = \frac{n}{2}$. Furthermore, using Equation (5.11),

$$\mathbb{E}\left[s(X_t,Y_t)\middle| X_t \neq Y_t\right]\mathbb{P}(X_t \neq Y_t) \geq \frac{\beta n}{2}$$

Combining this with Equation (5.12),

$$\mathbb{P}(X_{t+1} = Y_{t+1}) - \mathbb{P}(X_t = Y_t) \geq \frac{2\beta}{n}$$

for all $\alpha n \leq t \leq \alpha n + \frac{n}{2}$. Adding up these inequalities for all t in $[\alpha n, \alpha n + \frac{n}{2}]$,

$$\mathbb{P}\left(X_{\alpha n + \frac{n}{2}} = Y_{\alpha n + \frac{n}{2}}\right) \geq \beta$$

as required. □

For path coupling, a lemma about the diameter of P_n under the split-merge random walk is needed.

Lemma 5.17. *The diameter of P_n under the split-merge random walk is at most n.*

Proof: Proceed by induction on n. This statement is clearly true for $n = 1$. Now, assume it's true for all $m \leq n - 1$, and show it for n. Let $\sigma = (a_1, \ldots, a_k)$ and $\tau = (b_1, \ldots, b_l)$ be two partitions of n. Without loss of generality, assume that $a_1 \geq b_1$.

If $a_1 = b_1$, create a path from σ to τ by just changing the parts $(a_2, \ldots, a_k)$ to $(b_2, \ldots, b_l)$. Since $(a_2, \ldots, a_k)$ is a partition of $n - a_1$, by the inductive hypothesis,

$$\rho(\sigma, \tau) \leq n - a_1 \leq n - 1.$$

so this case follows.

Otherwise, $a_1 > b_1$. Let σ_1 be σ with a_1 split into $(b_1, a_1 - b_1)$. Then, σ_1 and τ match on the part b_1, and hence by the argument above,

$$\rho(\sigma_1, \tau) \leq n - 1$$

Since σ is a neighbor of σ_1, this implies that $\rho(\sigma, \tau) \leq n$, completing the proof. □

Theorem 5.1 is now proved using path coupling. It shows an $O(n \log n)$ bound on the split-merge random walk, and hence on the random transposition walk.

Proof of Theorem 5.1. Let $t_1 = \alpha n + \frac{n}{2}$. Consider the walk $(\tilde{X}_k)_{k \geq 1}$, where each step consists of making t_1 steps of the split-merge random walk. Let $(\tilde{X}_k, \tilde{Y}_k)$ be the coupling on this new walk induced by the current coupling (X_t, Y_t). Now, Proposition 5.16 shows that if $(\tilde{X}_0, \tilde{Y}_0) = (\sigma, \tau)$, where $\rho(\sigma, \tau) = 1$, then

$$\begin{aligned} \mathbb{E}\left[\rho(\tilde{X}_1, \tilde{Y}_1)\right] &= \mathbb{E}\left[\rho(X_{t_1}, Y_{t_1})\right] = \mathbb{P}(X_{t_1} \neq Y_{t_1}) \\ &\leq (1-\beta)\rho(\sigma, \tau) \end{aligned}$$

using the fact that $\rho(X_t, Y_t)$ is always either 0 or 1. Therefore, if $\tilde{d}(k)$ is defined to be the distance from stationarity of $(\tilde{X}_k, \tilde{Y}_k)$, then from Theorem 2.11,

$$\tilde{d}(k) \leq \operatorname{diam}(P_n)\,(1-\beta)^k$$

Since neighboring pairs are pairs that are one step apart in the split-merge random walk, Proposition 5.17 implies that $\operatorname{diam}(P_n) \leq n$. Also using the fact that $1 - x \leq e^{-x}$,

$$\tilde{d}(k) \leq n e^{-\beta k}$$

Thus, if $k = \frac{\log n}{2\beta}$, then $\tilde{d}(k) \leq e^{-2} < \frac{1}{4}$. But it's clear from the definition of the new walk that

$$d(kt_1) = \tilde{d}(k)$$

Thus,

$$d\left(\left(\frac{\alpha}{2\beta} + \frac{1}{4\beta}\right) n \log n\right) = d(kt_1) < \frac{1}{4}$$

which means that the walk has mixed by time $\left(\frac{\alpha}{2\beta} + \frac{1}{4\beta}\right) n \log n$, completing the proof. □

5.5 Proving $\mathbb{E}\left[s(X_t, Y_t)\right]$ is large

Let us now summarize the rest of the proof. The remainder of this chapter will be devoted to proving Theorem 5.15, which states that after $O(n)$ time, the expected value of $s(X_t, Y_t)$ is of order n.

The proof will be structured as follows: it is shown that in $O(n)$ time, $s(X_t, Y_t)$ will have a high probability of being at least order $n^{1/3}$. Then it is shown that it takes another $o(n)$ time for $s(X_t, Y_t)$ to have a high probability of being of order n. This will clearly suffice to show that that after $O(n)$ time, $\mathbb{E}\left[s(X_t, Y_t)\right]$ is of order n. Section 5.6 below will be concerned with growing $s(X_t, Y_t)$ to order $n^{1/3}$, while Section 5.7 will be concerned with growing it to order n.

Before stating the theorems and sketching their proofs, a number of useful definitions are needed. Note that some of these definitions are asymmetrical: they are defined in terms of $\bar{X}_t$ and not $\bar{Y}_t$. This is an arbitrary choice; since the pair (X_t, Y_t) is only a step apart, it doesn't make any difference.

Definition 5.18. *For $v \in \{1, 2, \ldots, n\}$, define $C_t(v)$ to be the cycle of $\bar{X}_t$ containing v. Furthermore, for a number x, define*

$$V_t(x) = \{v \in \{1, 2, \ldots, n\} \mid |C_t(v)| \geq x\}$$

Thus, $V_t(x)$ is the union of all cycles of size at least x.

Remark 5.19. Note that if $X_t = (a_1, a_2, \ldots, a_m)$, then

$$|V_t(x)| = \sum_{a_i \geq x} a_i$$

Thus, the size of $V_t(x)$ is a function of X_t.

The first proposition that grows $s(X_t, Y_t)$ to order $n^{1/3}$ is now stated.

Proposition 5.20. *Let (X_t, Y_t) be the usual coupling started at $(X_0, Y_0) = (\sigma, \tau)$, where $\rho(\sigma, \tau) \leq 1$. Then, for n sufficiently large and $t \geq 9n$,*

$$\mathbb{P}\left\{s(X_t, Y_t) \geq n^{1/3}, \left|V_t\left(n^{1/3}\right)\right| \geq \frac{n}{2}\right\} \geq \frac{1}{2}$$

Remark 5.21. Here, the choice of $n^{1/3}$ is in some sense arbitrary – any n^{α}, where $\alpha < \frac{1}{2}$, would have done just as well.

A few other definitions which are needed for the statement of the theorem growing $s(X_t, Y_t)$ from order $n^{1/3}$ to order n. Indeed, a more general theorem is proved. Fix constants ϵ and δ: then, if $s(X_t, Y_t)$ starts by being of size 2^{j+1} (where j can be a function of n), after a certain amount of time q, $s(X_{t+q}, Y_{t+q})$ has a high probability of being least $\epsilon\delta n$. The following definition introduces some notation necessary for stating the theorem; it currently looks completely inexplicable, but will be justified in Section 5.7.

Definition 5.22. *Assume ϵ and δ are fixed constants, and j is a number (possibly a function of n). Then, define*

$$K = \lceil \log_2(\epsilon\delta n) \rceil \tag{5.13}$$

Furthermore, for r between j and K define

$$a_r = \lceil 2\delta^{-1} 2^{-r} n(\log_2 n - r) \rceil \quad \textit{and} \quad \tau_r = \sum_{i=j}^{r-1} a_i \tag{5.14}$$

where as usual, $\lceil \cdot \rceil$ stands for the ceiling function.

The following proposition proves that $s(X_t, Y_t)$ grows to order n.

Proposition 5.23. *Let (X_t, Y_t) be the usual coupling started at $(X_0, Y_0) = (\sigma, \tau)$, where $\rho(\sigma, \tau) \leq 1$. Let j be a number and let $\delta \in (0, 1]$ be a constant such that $\left|V_0(2^{j+1})\right| \geq \delta n$ and $s(\sigma, \tau) \geq 2^{j+1}$. If K and τ_K are defined as in Definition 5.22 and $\epsilon \in (0, 1/32)$, then*

$$\mathbb{P}\{s(X_{\tau_K}, Y_{\tau_K}) < \epsilon\delta n\} \leq O(1)\delta^{-1}\epsilon \left|\log(\epsilon\delta)\right| \tag{5.15}$$

where the constant implied in the $O(1)$ notation is universal.

Proof of Theorem 5.15. Propositions 5.20 and 5.23 can be used to prove Theorem 5.15: let $t_1 \geq 9n$, and condition on $(X_{t_1}, Y_{t_1}) \in Q_{t_1}$, where

$$Q_{t_1} = \left\{(X_{t_1}, Y_{t_1}) \text{ such that } s(X_{t_1}, Y_{t_1}) \geq n^{1/3}, \left|V_{t_1}^{\pi}\left(n^{1/3}\right)\right| \geq \frac{n}{2}\right\} \tag{5.16}$$

Letting $2^{j+1} = n^{1/3}$ and $\delta = \frac{1}{2}$, if $(X_{t_1}, Y_{t_1}) \in Q_{t_1}$, then

$$s(X_{t_1}, Y_{t_1}) \geq 2^{j+1} \text{ and } \left|V_{t_1}(2^{j+1})\right| \geq \delta n$$

Since $\rho(X_{t_1}, Y_{t_1}) \leq 1$, Proposition 5.23 applies to pairs (X_{t_1}, Y_{t_1}) in Q_{t_1}. Therefore, averaging over $(X_{t_1}, Y_{t_1}) \in Q_{t_1}$,

$$\mathbb{P}\left\{s(X_{t_1+\tau_K}, Y_{t_1+\tau_K}) < \epsilon\delta n \mid (X_{t_1}, Y_{t_1}) \in Q_{t_1}\right\} \leq O(1)\delta^{-1}\epsilon \left|\log(\epsilon\delta)\right|$$

for any $\epsilon \in (0, 1/32)$. Now, pick ϵ such that the right hand side of the above inequality is at most 1/2. Then,

$$\mathbb{P}\left\{s(X_{t_1+\tau_K}, Y_{t_1+\tau_K}) \geq \epsilon\delta n \mid (X_{t_1}, Y_{t_1}) \in Q_{t_1}\right\} \geq \frac{1}{2}$$

and therefore, for sufficiently large n,

$$\mathbb{P}\left\{s(X_{t_1+\tau_K}, Y_{t_1+\tau_K}) \geq \epsilon\delta n\right\} \geq \frac{\mathbb{P}(Q_{t_1})}{2} \geq \frac{1}{4}$$

using Lemma 5.20. Therefore,

$$\mathbb{E}\left[s(X_{t_1+\tau_K}, Y_{t_1+\tau_K})\right] \geq \frac{\epsilon\delta n}{4} \tag{5.17}$$

It now just remains to show that is that $t_1 + \tau_K$ can be of order n. Since $\delta = \frac{1}{2}$ and $2^{j+1} = n^{1/3}$, by Equation (5.14)

$$\begin{aligned}\tau_K &= \sum_{i=j}^{K-1} \lceil 2\delta^{-1}2^{-i}n(\log_2 n - i)\rceil = O\left(n \log n \sum_{r=j}^{K-1} 2^{-i}\right)\\ &= O\left(n \log n \cdot 2^{-j+1}\right) = O(n^{2/3}\log n)\end{aligned}$$

Since $t_1 \geq 9n$ is arbitrary and τ_K is $o(n)$, Equation (5.17) implies that

$$\mathbb{E}\left[s(X_t, Y_t)\right] \geq \frac{\epsilon\delta n}{4}$$

for all $t \geq 10n$, which is precisely what is needed. □

Before the next two sections, in which Propositions 5.20 and 5.23 are proved, some technical results are needed. These are proved in Section 5.8 below, and are instrumental for controlling the probabilities in the next two sections.

Lemma 5.24. *Let σ be in S_n, and let $(\bar{X}_t)_{t\geq 1}$ be the random transposition walk starting at σ. Then, the expected number of v such that $|C_1(v)| < |C_0(v)|$ and $|C_1(v)| < x$ is no greater than $\frac{x^2}{n}$.*

For the next four lemmas, let (X_t, Y_t) be the usual coupling starting at (σ, τ), where $\rho(\sigma, \tau) = 1$, $s(\sigma, \tau) = b$ and $m(\sigma, \tau) = c$.

Lemma 5.25. *If $x \leq c$, then*

$$\mathbb{P}\{m(X_1, Y_1) < x\} \leq \frac{2x^2}{n^2}.$$

Lemma 5.26. *If $x \leq c$, and if $|V_0(y)| \geq R$, then*

$$\mathbb{P}\{m(X_1, Y_1) \geq x + y\} \geq \frac{2c(R - 2c)}{n^2}$$

Lemma 5.27. *If $x \leq b$, then*

$$\mathbb{P}\{s(X_1, Y_1) < x\} \leq \frac{4x^2}{n^2}$$

Lemma 5.28. *If x and y satisfy $x \leq b < x + y \leq c$, and $|V_0(y)| \geq R$, then*

$$\mathbb{P}\{s(X_1, Y_1) \geq x + y\} \geq \frac{2b(R - 3x - 3y)}{n^2}$$

5.6 Growing to $\Theta\left(n^{1/3}\right)$

This section proves Proposition 5.20. It makes a lot of use of the results of Schramm in “Compositions of random transpositions" [47]. A number of definitions are needed to state his main result.

Definition 5.29. *If $(\bar{X}_t)_{t\geq 0}$ is the random transposition walk, define G_t to be the graph on $\{1, 2, \ldots, n\}$ such that $\{u, v\}$ is an edge in G_t if and only if the random transposition (u, v) has appeared in the first t steps of our walk. Furthermore, let W_t denote the set of vertices of the largest component of G_t.*

Note that the behavior of the W_t defined above is well-understood; indeed,by an Erdős-Rényi theorem (see for example [3]), if $t = cn$, then

$$\frac{|W_t|}{n} \to z(2c) \tag{5.18}$$

in probability as $n \to \infty$, where $z(s)$ is the positive solution of $1 - z = e^{-zs}$.

Definition 5.30. *The Poisson-Dirichlet ($PD(1)$) distribution is a probability measure on the infinite dimensional simplex $\Omega = \{(x_1, x_2, \ldots) \mid \sum_{i=0}^{\infty} x_i = 1\}$. Sample from this simplex* *as follows: let $U_1, U_2, \ldots$ be an i.i.d sequence of random variables uniform on $[0,1]$. Then, set $x_1 = U_1$, and recursively,*

$$x_j = U_j \left(1 - \sum_{i=1}^{j-1} x_i\right)$$

Let (y_i) be the (x_i) sorted in nonincreasing order; then, the $PD(1)$ distribution is defined as the law of (y_i).

The main theorem (Theorem 1.1) of Schramm's paper [47] can now be stated. This remarkable result was proved using the tools of graph theory and coupling. A clever lemma showing that vertices that start in 'sufficiently large' cycles are likely to end up in cycles of order n also played a pivotal role (Lemma 5.34 below is an almost exact reproduction of the result.) The full strength of the result is not needed: while Schramm determines the law of the large parts of X_t, the only fact necessary for this chapter is that after a sufficiently long time, these cycles are of order n. For this theorem, treat X_t as an infinite vector by adding infinitely many 0s at the end of it.

Theorem 5.31 (Schramm)**.** *Let $c > 1/2$, and take $t = cn$. As $n \to \infty$, the law of $\frac{X_t}{|W_t|}$ converges weakly to the $PD(1)$ distribution; that is, for every $\epsilon > 0$, if n is sufficiently large and $t \geq cn$, then there is a coupling of X_t and a $PD(1)$ sample Y such that*

$$P\left\{\left\|Y - \frac{X_t}{|W_t|}\right\|_\infty < \epsilon\right\} > 1 - \epsilon \tag{5.19}$$

where $\|\cdot\|_\infty$ is the standard l^∞ distance.

The proof that follows uses Theorem 5.31 to show that at time $t = n$, more than half the vertices are in cycles of order n with high probability. This is used to 'grow' $m(X_t, Y_t)$ to order $n^{1/3}$, after which the same is done for $s(X_t, Y_t)$. The results for $m(X_t, Y_t)$ are needed before the results for $s(X_t, Y_t)$: since $s(X_t, Y_t) \leq m(X_t, Y_t)$, $m(X_t, Y_t)$ constrains the growth of $s(X_t, Y_t)$ from above. Good control on m is needed before tackling s.

Lemma 5.32. *Let k be a natural number not dependent on n. For sufficiently large n, that is, for $n > N = N(k)$,*

$$\mathbb{P}\left\{|V_n(n/k)| > n/2\right\} \geq 1 - \frac{6}{k}$$

Proof: For convenience of notation, let $X = (x_1, x_2, \ldots)$ be X_n, let $Q = (q_1, q_2, \ldots)$ be $\frac{X_n}{|W_n|}$, and let $Y = (y_1, y_2, \ldots)$ be a $PD(1)$ sample which is coupled with Q to satisfy Theorem

5.31 above. With current notation,

$$|V_n(n/k)| = \sum_{x_i \geq \frac{n}{k}} x_i \tag{5.20}$$

For the rest of the proof, fix $\epsilon = \frac{1}{9k}$. First note that Equation (5.18) implies that

$$\frac{|W_n|}{n} \to z(2) \approx 0.797$$

in probability, which means that $\lim_{n\to\infty} \mathbb{P}\{|W_n|/n < 3/4\} = 0$. Since $Q = X_n/|W_n|$, for sufficiently large n,

$$\mathbb{P}\left\{x_i \geq \frac{3n}{4} q_i \text{ for all } i\right\} > 1 - \epsilon$$

Furthermore, Theorem 5.31 implies that for sufficiently large n,

$$\mathbb{P}\{q_i \geq y_i - \epsilon \text{ for all } i\} > 1 - \epsilon$$

Combining the above two equations,

$$\mathbb{P}\left\{x_i \geq \frac{3n}{4}(y_i - \epsilon) \text{ for all } i\right\} > 1 - 2\epsilon \tag{5.21}$$

for sufficiently large n.

Thus, to estimate $|V_n(n/k)|$ it suffices to consider the large parts of the $PD(1)$ sample Y. To that end, define the random variable

$$G_Y(x) = \sum_{y_i \geq x} y_i$$

It is easy to check that $\mathbb{E}[G_Y(x)] = 1 - x$, and therefore $\mathbb{E}[1 - G_Y(x)] = x$. Thus, Markov's inequality implies that

$$\mathbb{P}\{G_Y(x) \leq 3/4\} = \mathbb{P}\{1 - G_Y(x) \geq 1/4\} \leq 4x$$

Recall that $\epsilon = \frac{1}{9k}$. Then, combining the above with Equation (5.21),

$$\mathbb{P}\left\{x_i \geq \frac{3n}{4}(y_i - \epsilon) \text{ for all } i, G_Y\left(\frac{13}{9k}\right) \geq \frac{3}{4}\right\} \geq 1 - \frac{6}{k} \tag{5.22}$$

Finally, assume that $x_i \geq \frac{3n}{4}(y_i - \epsilon)$ for each i, and that $G_Y\left(\frac{13}{9k}\right) \geq \frac{3}{4}$. Then, Equation (5.20) implies that

$$\begin{aligned}
|V_n(n/k)| &\geq \sum_{\frac{3n}{4}(y_i-\epsilon)\geq\frac{n}{k}} \frac{3n}{4}(y_i - \epsilon) = \frac{3n}{4}\left(\sum_{y_i\geq 13/9k} y_i - \sum_{y_i \geq 13/9k} \frac{1}{9k}\right) \\
&\geq \frac{3n}{4}\left(G_Y\left(\frac{13}{9k}\right) - \frac{1}{13}\right) \geq \frac{n}{2} \qquad (5.23)
\end{aligned}$$

using the fact that there can be at most $\frac{9k}{13}$ values of y_i that are greater than $\frac{13}{9k}$, since the y_i are positive and sum to 1. Therefore, using Equation (5.22), for sufficiently large n

$$\mathbb{P}\left\{|V_n(n/k)| \geq \frac{n}{2}\right\} \geq 1 - \frac{6}{k}$$

as required. □

The above lemma is now applied to find a t of order n such that the probability of having $m(X_t, Y_t) \geq n^{1/3}$ is sufficiently high. Lemmas 5.25 and 5.26 give control of $m(X_t, Y_t)$.

Lemma 5.33. *If n is sufficiently large and $t \geq 5n$, then*

$$\mathbb{P}\left\{m(X_t, Y_t) \geq n^{1/3}, \left|V_t\left(n^{1/3}\right)\right| \geq \frac{n}{2}\right\} \geq \frac{4}{5}$$

Proof: From Lemma 5.32, at time $t = n$,

$$\mathbb{P}\left\{|V_t(n/k)| \geq \frac{n}{2}\right\} \geq 1 - \frac{6}{k} \qquad (5.24)$$

Average over the possible values of X_{t-n} to conclude that Equation (5.24) also holds for any time $t \geq n$. Now, for convenience of notation, define

$$S_t = \left\{(X_t, Y_t) \text{ s.t. } \left|V_t\left(n^{1/3}\right)\right| \geq \frac{n}{2}\right\} \qquad (5.25)$$

For sufficiently large n, $n^{1/3} \leq \frac{n}{k}$ for any fixed value of k. Fix $\epsilon > 0$. Then, for $t \geq n$ and sufficiently large n, Equation (5.24) implies that $\mathbb{P}(S_t) \geq 1 - \epsilon$. Furthermore, define

$$A_t = \left\{(X_t, Y_t) \mid m(X_t, Y_t) \geq n^{1/3}\right\} \qquad (5.26)$$

To find a lower bound for $\mathbb{P}(A_t \cap S_t)$ for $t \geq 10n$, note that

$$\mathbb{P}(A_t \cap S_t) \geq \mathbb{P}(A_t) - \mathbb{P}(S_t^c) \geq \mathbb{P}(A_t) - \epsilon \tag{5.27}$$

and hence it suffices to bound $\mathbb{P}(A_t)$. This is done using a recursive argument: at each step t, calculate the probability that $m(X_t, Y_t)$ was too small, but $m(X_{t+1}, Y_{t+1})$ is large enough, and vice versa. The probability of A_t is shown to grow sufficiently quickly with t.

Start by bounding the probability that $m(X_t, Y_t) \geq n^{1/3}$, while $m(X_{t+1}, Y_{t+1}) < n^{1/3}$. By Lemma 5.25 with $x = n^{1/3}$,

$$\mathbb{P}\{(X_{t+1}, Y_{t+1}) \notin A_{t+1} \mid (X_t, Y_t) \in A_t\} \leq \frac{2x^2}{n^2} = \frac{2}{n^{4/3}}$$

and therefore

$$\mathbb{P}\{(X_{t+1}, Y_{t+1}) \notin A_{t+1}, (X_t, Y_t) \in A_t\} \leq \frac{2}{n^{4/3}}\mathbb{P}(A_t) \tag{5.28}$$

Now bound the probability that $m(X_t, Y_t) < n^{1/3}$, while $m(X_{t+1}, Y_{t+1}) \geq n^{1/3}$. In order to bound this in a satisfactory way, enough parts of size $n^{1/3}$are needed; accordingly, work with $(X_t, Y_t) \in A_t^c \cap S_t$. If $m(X_t, Y_t) < n^{1/3}$ and $V_t\left(n^{1/3}\right) \geq \frac{n}{2}$, then using Lemma 5.26 with $x = 0, y = n^{1/3}$, and $R = \frac{n}{2}$,

$$\mathbb{P}\{(X_{t+1}, Y_{t+1}) \in A_{t+1} \mid (X_t, Y_t) \in A_t^c \cap S_t\} \geq \frac{2\left(n/2 - 2n^{1/3}\right)}{n^2} \geq \frac{1-\epsilon}{n}$$

for sufficiently large n. Thus, for $t \geq n$, using the fact that $\mathbb{P}(S_t) \geq 1 - \epsilon$,

$$\begin{aligned}\mathbb{P}\{(X_{t+1}, Y_{t+1}) \in A_{t+1}, (X_t, Y_t) \notin A_t\} &\geq \left(\frac{1-\epsilon}{n}\right)\mathbb{P}(A_t^c \cap S_t) \\ &\geq \left(\frac{1-\epsilon}{n}\right)(1 - \mathbb{P}(A_t) - \epsilon) \end{aligned} \tag{5.29}$$

for sufficiently large n. Combining Equations (5.28) and (5.29),

$$\begin{aligned}\mathbb{P}(A_{t+1}) - \mathbb{P}(A_t) &\geq -\frac{2}{n^{4/3}}\mathbb{P}(A_t) + \left(\frac{1-\epsilon}{n}\right)(1 - \mathbb{P}(A_t) - \epsilon) \\ &\geq \frac{1 - \mathbb{P}(A_t) - 3\epsilon}{n}\end{aligned}$$

for sufficiently large n and $t \geq n$. Rearranging the above,

$$(1 - 3\epsilon - \mathbb{P}(A_{t+1})) \leq \left(1 - \frac{1}{n}\right)(1 - 3\epsilon - \mathbb{P}(A_t)) \tag{5.30}$$

and hence using recursion and the lower bound in Equation (5.27),

$$\begin{aligned}(1 - 3\epsilon - \mathbb{P}(A_t)) \leq \left(1 - \frac{1}{n}\right)^{t-n} \leq e^{-(t-n)/n} \\ \Rightarrow \mathbb{P}(A_t \cap S_t) \geq 1 - 4\epsilon - e^{-(t-n)/n}\end{aligned}$$

Thus, for $t \geq 5n$, $\mathbb{P}(A_t \cap S_t) \geq 1 - 4\epsilon - e^{-4} \approx 1 - 4\epsilon - 0.018$, and picking ϵ appropriately completes the proof. □

Proposition 5.20 (Restatement). *For sufficiently large n, and $t \geq 9n$,*

$$\mathbb{P}\left\{s(X_t, Y_t) \geq n^{1/3}, \left|V_t\left(n^{1/3}\right)\right| \geq \frac{n}{2}\right\} \geq \frac{1}{2}$$

Proof: This proof is very similar to the one above. Let $t \geq 5n$, and define

$$R_t = \left\{m(X_t, Y_t) \geq n^{1/3}, \left|V_t\left(n^{1/3}\right)\right| \geq \frac{n}{2}\right\}$$

From the above lemma, $\mathbb{P}(R_t) \geq \frac{4}{5}$. Now, define

$$C_t = \left\{(X_t, Y_t) \mid s(X_t, Y_t) \geq n^{1/3}\right\}$$

It is shown below that $\mathbb{P}(C_t \cap R_t) \geq \frac{1}{2}$, which will clearly suffice. Note that for $t \geq 5n$,

$$\mathbb{P}(C_t \cap R_t) \geq \mathbb{P}(C_t) - \frac{1}{5} \tag{5.31}$$

and hence it suffices to find a lower bound on $\mathbb{P}(C_t)$. As above, this is done by finding recursive bounds on the probability of C_{t+1} given the probability of C_t. By Lemma 5.27 with $x = n^{1/3}$,

$$\mathbb{P}\{(X_{t+1}, Y_{t+1}) \notin C_{t+1} \mid (X_t, Y_t) \in C_t\} \leq \frac{4x^2}{n^2} = \frac{4}{n^{4/3}}$$

and therefore

$$\mathbb{P}\{(X_{t+1}, Y_{t+1}) \notin C_{t+1}, (X_t, Y_t) \in C_t\} \leq \frac{4}{n^{4/3}}\mathbb{P}(C_t) \tag{5.32}$$

Now, assume that $(X_t, Y_t) \in C_t^c \cap R_t$. Then $m(X_t, Y_t) \geq n^{1/3} > s(X_t, Y_t)$ and $V_t\left(n^{1/3}\right) \geq \frac{n}{2}$. Therefore, using Lemma 5.28 with $x = 0, y = n^{1/3}$, and $R = \frac{n}{2}$,

$$\mathbb{P}\{(X_{t+1}, Y_{t+1}) \in C_{t+1} \mid (X_t, Y_t) \in C_t^c \cap R_t\} \geq \frac{2\left(n/2 - 3n^{1/3}\right)}{n^2} = \frac{1}{n} - \frac{6}{n^{5/3}}$$

Thus, for $t \geq 5n$, using the fact that $\mathbb{P}(R_t) \geq \frac{4}{5}$,

$$\begin{aligned}\mathbb{P}\{(X_{t+1}, Y_{t+1}) \in C_{t+1}, (X_t, Y_t) \notin C_t\} &\geq \left(\frac{1}{n} - \frac{6}{n^{5/3}}\right)\mathbb{P}(C_t^c \cap R_t) \\ &\geq \left(\frac{1}{n} - \frac{6}{n^{5/3}}\right)\left(\frac{4}{5} - \mathbb{P}(C_t)\right) \qquad (5.33)\end{aligned}$$

for sufficiently large n. Therefore, combining Equations (5.32) and (5.33) and picking n sufficiently large,

$$\begin{aligned}\mathbb{P}(C_{t+1}) - \mathbb{P}(C_t) &\geq -\frac{2}{n^{4/3}}\mathbb{P}(C_t) + \left(\frac{1}{n} - \frac{6}{n^{5/3}}\right)\left(\frac{4}{5} - \mathbb{P}(C_t)\right) \\ &\geq \frac{3/4 - \mathbb{P}(C_t)}{n} \qquad (5.34)\end{aligned}$$

for $t \geq 5n$. Rearranging analogously to Equation (5.30),

$$\left(\frac{3}{4} - \mathbb{P}(C_{t+1})\right) \leq \left(1 - \frac{1}{n}\right)\left(\frac{3}{4} - \mathbb{P}(C_t)\right)$$

As before, for $t \geq 9n$, $\mathbb{P}(C_t) \geq \frac{7}{10}$. Combining this with Equation (5.31),

$$\mathbb{P}(C_t \cap R_t) \geq \frac{1}{2}$$

for $t \geq 9n$ and n sufficiently large, as required. □

5.7 Growing to $\Theta(n)$

This section proves Proposition 5.23, which shows that $s(X_t, Y_t)$ can be grown to order n. This section is structured similarly to the previous one: proving a lemma about overall cycle sizes, then a lemma about $m(X_t, Y_t)$, and then finally Proposition 5.23. Again, use is made of the technical results in Lemmas 5.24 through 5.28.

The idea behind the proof is largely based on Lemma 2.3 from “Compositions of random transpositions" [47]. Let ϵ, δ and j be chosen as in Proposition 5.23. Recall that Definition 5.22 defines $K = \lceil \log_2(\epsilon\delta n) \rceil$ and

$$a_r = \lceil 2\delta^{-1} 2^{-r} n(\log_2 n - r) \rceil \quad \text{and} \quad \tau_r = \sum_{i=j}^{r-1} a_i$$

for r between j and K, with $\tau_j = 0$. Then, define

$$I_r = [\tau_r, \tau_{r+1} - 1] \tag{5.35}$$

and for convenience of notation, define $I_K = \{\tau_K\}$.

As should be clear from the statement of Proposition 5.23, the argument starts with $s(\pi, \sigma) \geq 2^{j+1}$ and $V_0(2^{j+1}) \geq \delta n$, and shows that at time τ_K, the probability that $s(X_{\tau_K}, Y_{\tau_K})$ is less than $\epsilon\delta n$ is appropriately bounded above. In fact, something stronger is shown: for the intervals I_r as defined above, one 'expects' to have

$$V_t(2^{r+1}) \geq \frac{\delta n}{2}, s(X_t; Y_t) \geq 2^r, \text{ and } m(X_t, Y_t) \geq 2^{r+1}$$

for all r between j and K. This would clearly suffice to prove the result.

The first lemma is almost identical to Lemma 2.3 from [47] – it is reproven here for completeness, and to illustrate the technique. This lemma starts with $\sigma \in S_n$, and $\left|V_0(2^{j+1})\right| \geq \delta n$. It gives an upper bound for the expected number of vertices that start in cycles of size at least 2^{j+1}, and that are not in cycles of size $\epsilon\delta n$ at time τ_K. This shows that 'most' vertices that start in cycles of size 2^{j+1} are in cycles of order n at time τ_K.

Lemma 5.34. *Let $\sigma \in S_n$. Let $\delta \in (0,1)$ be a constant such that $\left|V_0(2^{j+1})\right| \geq \delta n$, and let K and τ_K be defined as they are above and in Definition 5.22. Fix $\epsilon \in (0, 1/32)$. For the random transposition walk $\left(\bar{X}_t\right)_{t\geq 0}$,*

$$\mathbb{E}\left|V_0(2^{j+1}) \setminus V_{\tau_K}(2\epsilon\delta n)\right| \leq O(1)\delta^{-1}\epsilon \left|\log(\epsilon\delta)\right| n \tag{5.36}$$

where the constant implied in the $O(1)$ notation is universal.

Proof: Before beginning the proof, consider what is being shown. Starting with a σ such that $\left|V_0(2^{j+1})\right| > \delta n$ means that at least δn of the vertices in σ are in cycles of size at least 2^{j+1}. An upper bound on the expected size of $V_0(2^{j+1}) \setminus V_{\tau_K}(2\epsilon\delta n)$ is needed: that is, an upper bound on the expected number of vertices that started off in cycles of size at least 2^{j+1} in σ, and ended up in cycles of size less than $2\epsilon\delta n$ at time τ_K.

Something stronger is shown: conditioned on $v \in V_0(2^{j+1})$,

$$\mathbb{E}\left|v \text{ s.t. } C_t(v) < 2^{r+1} \text{ for any } t \in I_r, \text{ for } r \in [j, K]\right| \le O(1)\delta^{-1}\epsilon\left|\log(\epsilon\delta)\right| n \tag{5.37}$$

This requires an upper bound on the expected number of vertices that for any time $t \in I_r$ are 'too small' for I_r: they are of size less than 2^{r+1}. Note that the above set includes all vertices such that $C_{\tau_K}(v) < 2\epsilon\delta n \le 2^{K+1}$, and hence the above bound would suffice.

Three different possibilities are considered. First of all, an upper bound is needed on the expected number of vertices v such that at any point, the cycle containing v is split, and becomes too small. Secondly, all vertices that appear in permutations with an insufficient number of large parts are rejected. And thirdly, it is necessary to bound the possibility that the cycle containing v does not grow sufficiently during I_r. Call the vertices that fall into any of these undesirable categories 'failed.'

In the next three sections, condition on $v \in V_0(2^{j+1})$: that is, assume that v is in a cycle of size 2^{j+1} in σ. This means that v has not failed at time 0.

The cycle containing v becomes too small Let $r \in [j, K-1]$, and let $t \in I_r + 1 = [\tau_r + 1, \tau_{r+1}]$. For $C_t(v)$ to be of size 2^{r+2} by time τ_{r+1}, calculate the probability that for any $t \in I_r + 1$, the cycle containing v is split, and v is then contained in a cycle of size less than 2^{r+2}. To be precise, define F_t to be the set of vertices at time t such that $|C_t(v)| < |C_{t-1}(v)|$ and $|C_t(v)| < 2^{r+2}$. Find the expected size of F_t: by definition, this is the expected number of vertices v, whose cycle is split from time $t-1$ to time t, and which are in cycles of size less than 2^{r+2} at time t. By Lemma 5.24,

$$\mathbb{E}\left|F_t\right| \le \frac{2\left(2^{r+2}\right)^2}{n} = \frac{2^{2r+5}}{n}$$

Now, define the cumulative set $\tilde{F}_t = \bigcup_{x=1}^{t} F_x$. This is the set of all vertices up to time

t, whose cycles have at any time $x \le t$ been split into ones that are 'too small.' Clearly,

$$\begin{aligned}
\mathbb{E}\left|\tilde{F}_{\tau_K}\right| &\le \sum_{x=1}^{\tau_K} \mathbb{E}\left|F_x\right| \le \sum_{r=j}^{K-1} a_r \frac{2^{2r+5}}{n} \\
&\le \sum_{r=j}^{K-1} \lceil 2\delta^{-1}2^{-r}n(\log_2 n - r)\rceil \frac{2^{2r+5}}{n} \\
&\le \sum_{r=j}^{K-1} \left(2\delta^{-1}2^{-r}n(\log_2 n - r) + 1\right) \frac{2^{2r+5}}{n} \\
&\le \sum_{r=j}^{K-1} 2^6\delta^{-1}2^r(\log_2 n - r) + \sum_{r=j}^{K-1} \frac{2^{2r+5}}{n} \qquad (5.38)
\end{aligned}$$

Now,

$$\begin{aligned}
\sum_{r=j}^{K-1} r2^r &= (K-2)2^K - (j-2)2^j \\
\sum_{r=j}^{K-1} 2^r &= 2^K - 2^j
\end{aligned}$$

shows that

$$\mathbb{E}\left|\tilde{F}_{\tau_K}\right| \le 2^8\left|\log_2(\epsilon\delta)\right|\epsilon n \qquad (5.39)$$

Permutations with insufficiently many large parts It is also necessary to rule out vertices in permutations for which the union of the 'large parts' isn't sufficiently high. This will be useful for the next part of the proof. To be more precise, let $t \in I_r$: if $\left|V_t(2^{r+1})\right| < \delta n/2$, and this is the first t for which the inequality holds, then consider all vertices in X_t to have failed, and set $H_t = \{1, \dots, n\}$. Otherwise, set $H_t = \emptyset$.

Again, define the cumulative set $\tilde{H}_t = \bigcup_{x=0}^{t} H_x$. This is the union of all vertices that up to time t have been in a permutation with insufficiently many large parts, by the above definition. It is clear that this set is either empty, or contains all the vertices. There is no current available upper bound on the expectation for $\tilde{H}_t$; one will be derived after the next section of the proof.

The cycle containing v doesn't grow sufficiently Next, consider how a vertex v might fail at time t, if it does not fall into $\tilde{F}_t$ or $\tilde{H}_{t-1}$. Assume t is the minimal time for which v

fails: since failed vertices include all vertices contained in cycles that are 'too small', if $s < t$ and $s \in I_k$ then $|C_s(v)| \geq 2^{k+1}$. Now, assume that t, the first time at which v fails, is in I_r: thus, $t-1$ is either in I_{r-1} or in I_r. Either way, since it was assumed that v is not in F_t^π, it can't be that $|C_t(v)| < |C_{t-1}(v)|$ and $|C_t(v)| < 2^{r+1}$. Since the vertex v fails at time t, $C_t(v)$ must contain fewer than 2^{r+1} vertices. Combine this with the preceding statement to conclude that $C_{t-1}(v)$ also contains fewer than 2^{r+1} vertices. However, by definition the vertex v did not fail at time $t-1$. This implies $t-1$ must have been in I_{r-1}. Thus, the only remaining times at which vertices could fail are $t = \tau_r$, for $r \in \{j, j+1, \ldots, K\}$. Having conditioned on $v \in V_0(2^{j+1})$, it may be concluded that v can't fail at time $\tau_j = 0$.

Now, define B_r to be the set of vertices at time τ_r that are not in $\tilde{F}^{\tau_r} \cup \tilde{H}^{\tau_r - 1}$, such that $|C_{\tau_r}(v)| < 2^{r+1}$, and that have not failed previously. As before, define $\tilde{B}_r = \bigcup_{x=j}^r B_x$ and estimate the expected size of B_r.

Condition on $v \notin \tilde{F}_{\tau_r} \cup \tilde{H}_{\tau_r - 1}$ and calculate the probability that v fails at τ_r, given that it has not failed up to that time. First, for $t \in I_{r-1}$, $|C_t(v)| \geq 2^r$. Furthermore, since $v \notin \tilde{F}_{\tau_r}$, there was no time between τ_{r-1} and τ_r at which the cycle containing v was split to contain fewer than 2^{r+1} vertices. This means that if v failed at time τ_r, then $C_t(v)$ must have been of size less than 2^{r+1} for all $t \in [\tau_{r-1}, \tau_r - 1]$. Therefore, for $t \in I_{r-1}$,

$$2^r \leq |C_t(v)| < 2^{r+1} \tag{5.40}$$

Furthermore, since v is not in $\tilde{H}_t$ for any $t \in I_{r-1}$, for every $t \in I_{r-1}$, $|V_t(2^r)| \geq \delta n/2$. Consider the probability that from time t to time $t+1$, the cycle containing v is merged with a cycle of size at least 2^r. By (5.40) above, the size of $C_t(v)$ is at least 2^r, so such a merge would result in $C_{t+1}(v) \geq 2^{r+1}$. Using the above reasoning implies that $|C_{\tau_r}(v)| \geq 2^{r+1}$, and therefore v does not fail at time τ_r. Now, again by (5.40), the cycle containing v is of size at most 2^{r+1}. Since $|V_t(2^r)| \geq \delta n/2$, this means the union of the cycles disjoint from $C_t(v)$ of size at least 2^r contains at least $\delta n/2 - 2^{r+1}$ vertices. Now, since $r \leq K = \lceil \log_2(\epsilon \delta n) \rceil$, $2^{r+1} \leq 2^{K+1} \leq 4\epsilon \delta n$, and since $\epsilon < 1/32$,

$$\frac{\delta n}{2} - 2^{r+1} \geq \frac{\delta n}{4}$$

Thus, the union of the cycles of size at least 2^r disjoint from $C_t(v)$ is of size at least $\delta n/4$, and therefore

$$\mathbb{P}\{C_t(v) \text{ merges with a cycle of size } \geq 2^r\} \geq 2 \cdot 2^r \frac{\delta n/4}{n^2} = 2^{r-1}\delta n^{-1}$$

Clearly, for v to be in in B_r, it cannot be that $C_t(v)$ merges with a cycle of size $\geq 2^r$ for any $t \in I_{r-1}$. Therefore,

$$\begin{aligned} \mathbb{P}\{v \in B_r\} &\leq \left(1 - 2^{r-1}\delta n^{-1}\right)^{a_{r-1}} \\ &\leq \exp(-2^{r-1}\delta n^{-1} a_{r-1}) \end{aligned} \tag{5.41}$$

and since $a_{r-1} \geq 2\delta^{-1}2^{-r+1}n(\log_2 n - r + 1)$,

$$\mathbb{P}\{v \in B_r\} \leq e^{2(r-1-\log_2 n)}$$

Now, $r - 1 \leq K - 1 \leq \log_2(\epsilon\delta n)$, and therefore, $\log_2 n - r + 1 \leq 0$. Thus,

$$\mathbb{P}\{v \in B_r\} \leq e^{2(\log_2 n - r + 1)} \leq 2^{\log_2 n - r + 1} = \frac{2^{r-1}}{n}$$

This yields

$$\mathbb{E}\,|B_r| \leq n\mathbb{P}\{v \in B_r\} = 2^{r-1}$$

and therefore,

$$\begin{aligned} \mathbb{E}\left|\tilde{B}_K\right| &\leq \sum_{r=j}^{K} \mathbb{E}\,|B_r| \leq \sum_{r=j}^{K} 2^{r-1} \\ &\leq 2^K \leq \epsilon\delta n + 1 \end{aligned} \tag{5.42}$$

Finally, bound the expected size of H_t, the set of vertices in permutations with insufficiently many large parts. Recall that for $t \in I_r$, if $\left|V_t(2^{r+1})\right| < \delta n/2$ and t was the first time this inequality held, H_t was defined to be the set of all vertices, and it was otherwise defined to be the empty set. If H_t is non-empty, then the set of vertices in X_t that are in cycles of size less than 2^{r+1} has size at least $n - \delta n/2 \geq \delta n/2$. Now, consider v in H_t such that $|C_t(v)| < 2^{r+1}$. By definition, v has failed by time t, and v is not in H_s for any $s < t$.

Therefore, each such vertex is in $\tilde{F}_t \cup \tilde{B}_r$. Thus,

$$\mathbb{E}\left|\tilde{H}_{\tau_K}\right| \leq n\mathbb{P}\left\{\left|\tilde{F}_{\tau_K} \cup \tilde{B}_K\right| \geq \delta n/2\right\} \leq 2\delta^{-1}\mathbb{E}\left|\tilde{F}_{\tau_K} \cup \tilde{B}_K\right|$$

so using (5.39) and (5.42) above,

$$\begin{aligned}\mathbb{E}\left|\tilde{H}_{\tau_K}\right| &\leq 2\delta^{-1}(\epsilon\delta n + 1 + 2^8 \left|\log(\epsilon\delta)\right| \epsilon n) \\ &\leq (2^9+1)\epsilon\delta^{-1}\left|\log_2(\epsilon\delta)\right| n \qquad (5.43)\end{aligned}$$

as desired. Finally, adding up the expectations for H_{τ_K}, B_K and F_{τ_K} in (5.43), (5.42) and (5.39) completes the proof. □

The next lemma is similar. It shows that $m(X_t, Y_t)$ becomes sufficiently large at time τ_K. The proof is almost entirely analogous; the only substantial difference is in the bound for the probability of X_t having insufficiently many 'large parts.' For this bound, Equation (5.43) above has to be used. Lemmas 5.25 and 5.26 will also be used.

Lemma 5.35. *Assume $\rho(\sigma,\tau) = 1$. Let j be a natural number such that $m(\sigma,\tau) \geq 2^{j+1}$, and let $\delta \in (0,1]$ be a constant such that $\left|V_0(2^{j+1})\right| \geq \delta n$ and $m(\pi,\sigma) \geq 2^{j+1}$. Let K and τ_K be defined as above, and let $\epsilon \in (0, 1/16)$. Then,*

$$\mathbb{P}\{m(X_{\tau_K}, Y_{\tau_K}) < 2\epsilon\delta n\} \leq O(1)\delta^{-1}\epsilon\left|\log(\epsilon\delta)\right| \qquad (5.44)$$

where the constant implied in the $O(1)$ notation is universal.

Proof: This proof is almost exactly analogous to the previous one, except that instead of keeping track of failed vertices, failed pairs of partitions will be tracked. Something stronger is shown:

$$\mathbb{P}\{m(X_t, Y_t) < 2^{r+1} \text{ for any } t \in I_r, \text{ for } j \in [j,K]\} \leq O(1)\delta^{-1}\epsilon\left|\log(\epsilon\delta)\right| \qquad (5.45)$$

Again, the argument requires upper bounds on three different cases: the one where $m(X_t, Y_t)$ shrinks to become too small at time t, the one where X_t doesn't have sufficiently many large parts, and the one where $m(X_t, Y_t)$ fails to grow sufficiently during I_r. The only major difference in the proof is use of the bound from Lemma 5.34 to bound the probability of X_t having insufficiently many large parts.

Since the quantities specified are precisely analogous, use the names $\mathcal{F}_t, \mathcal{B}_t$ and $\mathcal{H}_t$.

Probability $m(X_t, Y_t)$ gets too small during I_r For $t \in I_r + 1$, define $\mathcal{F}_t$ to be the set of pairs (X_t, Y_t) such that $m(X_t, Y_t) < m(X_{t-1}, X_{t-1})$ and $m(X_t, Y_t) < 2^{r+2}$. Apply Lemma 5.25 above. Let $x = \min(2^{r+2}, m(X_{t-1}, Y_{t-1}))$. Then, $x \leq m(X_{t-1}, Y_{t-1})$, and therefore from Lemma 5.25, the probability that $m(X_t, Y_t)$ is less than x is bounded above by $\frac{2x^2}{n^2}$. By definition of $\mathcal{F}_t$, this means that

$$\mathbb{P}\{\mathcal{F}_t\} \leq \frac{2x^2}{n^2} \leq 2\frac{\left(2^{r+2}\right)^2}{n^2} = \frac{2^{2r+5}}{n^2}$$

Define the cumulative set $\tilde{\mathcal{F}}_t = \bigcup_{x=1}^{t} \mathcal{F}_x$. Therefore,

$$\mathbb{P}\{\tilde{\mathcal{F}}_{\tau_K}\} \leq \sum_{x=1}^{\tau_K} \mathbb{P}\{\mathcal{F}_x\} \leq \sum_{r=j}^{K-1} a_r \frac{2^{2r+5}}{n^2}$$

and doing a calculation almost identical to (5.39),

$$\mathbb{P}\{\tilde{\mathcal{F}}_{\tau_K}\} \leq 2^8 \epsilon \left|\log_2(\epsilon\delta)\right| \tag{5.46}$$

Note that the only difference in the calculation was an extra factor of n in the denominator.

Probability X_t doesn't have enough large parts Define $\mathcal{H}_t$ almost exactly as H_t in the last lemma, except that instead of making it a set of vertices, let it be a set of pairs (X_t, Y_t). (X_t, Y_t) is included in $\mathcal{H}_t$ precisely when X_t doesn't have enough large parts: that is, if $t \in I_r$, then (X_t, Y_t) is in $\mathcal{H}_t$ if $\left|V_t(2^{r+1})\right| < \delta n/2$, and t is the first time for which this inequality holds. Define $\tilde{\mathcal{H}}_t$ as usual to be the cumulative set.

Clearly, if $(X_t, Y_t) \in \mathcal{H}_t$, then H_t contains n vertices, and otherwise H_t is empty. Since $\left|V_0(2^{j+1})\right| \geq \delta n$, the results derived in Lemma 5.34 can be used. Therefore,

$$\mathbb{P}\{\mathcal{H}_t\} = \frac{1}{n}\mathbb{E}\left|H_t\right|$$

and thus, from (5.43) above,

$$\mathbb{P}\{\mathcal{H}_t\} \leq (2^9 + 1)\epsilon\delta^{-1} \left|\log_2(\epsilon\delta)\right| \tag{5.47}$$

Probability $m(X_t, Y_t)$ doesn't grow sufficiently during I_r As before, the only remaining times that (X_t, Y_t) can fail is at times τ_r. Accordingly, define $\mathcal{B}_r$ to be those pairs (X_{τ_r}, Y_{τ_r}) that are not in $\tilde{\mathcal{F}}_{\tau_r}$ or $\tilde{\mathcal{H}}_{\tau_r - 1}$, such that $m(X_{\tau_r}, Y_{\tau_r}) < 2^{r+1}$ and that have not failed previously. As before, if (X_{τ_r}, Y_{τ_r}) is in $\mathcal{B}_r$, then it had not failed in I_{r-1}, and therefore, for $t \in I_{r-1}$, $m(X_t, Y_t) \geq 2^r$. Furthermore, since (X_{τ_r}, Y_{τ_r}) is not in $\tilde{\mathcal{F}}_{\tau_r}$, it must be that $m(X_t, Y_t)$ is less than 2^{r+1} for $t \in I_{r-1}$. Thus, for $t \in I_{r-1}$,

$$2^r \leq m(X_t, Y_t) < 2^{r+1} \tag{5.48}$$

Furthermore, since $\mathcal{B}_r$ is disjoint from $\mathcal{H}_{\tau_r - 1}$, for every $t \in I_{r-1}$, $|V_t(2^r)| \geq \delta n/2$. Since $m(X_t, Y_t) \geq 2^r$, Lemma 5.26 holds with $R = \delta n/2$ and $x = y = 2^r$. Let $c = m(X_t, Y_t)$. Thus, for any $t \in I_r$,

$$\mathbb{P}\{m(X_{t+1}, Y_{t+1}) \geq 2^{r+1}\} \geq \frac{2c(\delta n/2 - 2c)}{n^2} \geq \frac{2^{r+1}(\delta n/2 - 2^{r+2})}{n^2}$$

and since $\epsilon < \frac{1}{32}$, and $r \leq K \leq \log_2(\epsilon\delta n) + 1$, $\delta n/2 - 2^{r+2} \geq \delta n/4$. Thus,

$$\mathbb{P}\{m(X_{t+1}, Y_{t+1}) \geq 2^{r+1}\} \geq 2^{r-1}\delta n^{-1}$$

Finally, the probability of $\mathcal{B}_r$ is the probability that $m(X_{t+1}, Y_{t+1})$ isn't at least 2^{r+1} for any $t \in I_r$, and therefore,

$$\mathbb{P}\{\mathcal{B}_r\} \leq \left(1 - 2^{r-1}\delta n^{-1}\right)^{a_{r-1}}$$

and since this is precisely the same inequality as in (5.41),

$$\mathbb{P}\{\mathcal{B}_r\} \leq \frac{2^{r-1}}{n}$$

and hence

$$\mathbb{P}\{\tilde{\mathcal{B}}_K\} \leq \sum_{r=j}^{K} \mathbb{P}\{\mathcal{B}_r\} \leq \sum_{r=j}^{K} \frac{2^{r-1}}{n} \leq \epsilon\delta + \frac{1}{n} \tag{5.49}$$

Thus, adding (5.46), (5.47), and (5.49),

$$\mathbb{P}\{m(X_t, Y_t) < 2^{r+1} \text{ for any } t \in I_r, \text{ for } r \in [j, K]\} \leq (2^9 + 2^8 + 1)\epsilon\delta^{-1} |\log_2(\epsilon\delta)| \tag{5.50}$$

which is what is needed. □

The stage is almost set to prove an analogous result for $s(X_t, Y_t)$. As above, the two technical Lemma 5.27 and 5.28 are used. As in the previous section, $m(X_t, Y_t)$ must be 'sufficiently large' to allow $s(X_t, Y_t)$ to grow. This is the reason for proving the lemma concerning $m(X_t, Y_t)$ first.

Proposition 5.23 (Restatement). *Let (X_t, Y_t) be the usual coupling started at $(X_0, Y_0) = (\sigma, \tau)$, where $\rho(\sigma, \tau) \leq 1$. Let j be a number and let $\delta \in (0, 1]$ be a constant such that $\left|V_0(2^{j+1})\right| \geq \delta n$ and $s(\sigma, \tau) \geq 2^{j+1}$. If K and τ_K are defined as in Definition 5.22 and $\epsilon \in (0, 1/32)$, then*

$$\mathbb{P}\{s(X_{\tau_K}, Y_{\tau_K}) < \epsilon \delta n\} \leq O(1)\delta^{-1}\epsilon \left|\log(\epsilon\delta)\right|$$

where the constant implied in the $O(1)$ notation is universal.

Proof of Lemma 5.23. This proof is analogous to the proof of Lemma 5.34 and 5.35, except that the previous two lemmas are used to bound the probability that $s(X_t, Y_t)$ shrinks or grows. As before, a stronger statement is proved:

$$\mathbb{P}\{s(X_t, Y_t) < 2^r \text{ for any } t \in I_r, \text{ for } r \in [j, K]\} \leq O(1)\delta^{-1}\epsilon \left|\log(\epsilon\delta)\right| \tag{5.51}$$

Again, bounds are needed for a number of different cases: for the probability that $s(X_t, Y_t)$ shrinks to become too small during I_r, the probability that X_t doesn't have enough large parts, and that the probability that $s(X_t, Y_t)$ doesn't grow sufficiently on I_r. Furthermore, note that Lemma 5.28 requires the assumption that $m(\sigma, \tau) \geq 2x$ to lower bound on the probability that $s(X_1, Y_1) \geq 2x$. Since $s(X_t, Y_t)$ must grow during I_r to be at least 2^{r+1} by τ_{r+1}, $m(X_t, Y_t)$ must be at least 2^{r+1} on I_r. Lemma 5.35 is used to bound the probability that $m(X_t, Y_t)$ is too small.

The quantities are precisely analogous to the ones in the two similar previous lemmas. Accordingly, name them $\mathscr{F}_t$, $\mathscr{H}_t$, and $\mathscr{B}_t$, using the same letters but yet another font. The new quantity $\mathscr{M}_t$ is added, as discussed above.

Probability $s(X_t, Y_t)$ gets too small during I_r For $t \in I_r + 1 = [\tau_r + 1, \tau_{r+1}]$, define $\mathscr{F}_t$ to be the set of pairs (X_t, Y_t) such that $s(X_t, Y_t) < s(X_{t-1}, Y_{t-1})$ and $s(X_t, Y_t) < 2^{r+1}$. Apply Lemma 5.27 above. Define $x = \min(2^{r+1}, s(X_{t-1}, Y_{t-1}))$. Then, $x \leq s(X_{t-1}, Y_{t-1})$, and therefore Lemma 5.27 applies. Plugging it in, the probability that $s(X_t, Y_t)$ is less than

x is at most $\frac{4x^2}{n^2}$. Thus,

$$\mathbb{P}\{\mathscr{F}_t\} \leq \frac{4x^2}{n^2} \leq \frac{4(2^{r+1})^2}{n^2} = \frac{2^{2r+4}}{n^2}$$

Now, define the cumulative set $\tilde{\mathscr{F}}_t = \bigcup_{x=1}^{t} \mathscr{F}_x$. Then,

$$\mathbb{P}\{\tilde{\mathscr{F}}_{\tau_K}\} \leq \sum_{x=1}^{\tau_K} \mathbb{P}\{\mathscr{F}_x\} \leq \sum_{r=j}^{K-1} a_r \frac{2^{2r+4}}{n^2}$$

Doing a calculation identical to the one in (5.39) and (5.46),

$$\mathbb{P}\{\tilde{\mathscr{F}}_{\tau_K}\} \leq 2^7 \epsilon \left|\log_2(\epsilon\delta)\right| \tag{5.52}$$

Probability X_t doesn't have enough large parts For $t \in I_r$, define $\mathscr{H}_t$ very similarly to before, to be the set of (X_t, Y_t) such that $|V_t(2^r)| < \delta n/2$, whenever this is the first t for which this inequality holds. Define $\tilde{\mathscr{H}_t}$ to be the usual cumulative set. Now, from Lemma 5.35,

$$\mathcal{H}_t = \{(X_t, Y_t) \mid \left|V_t(2^{r+1})\right| < \delta n/2\}$$

Since $V_t(2^r) \supseteq V_t(2^{r+1})$, clearly $\mathscr{H}_t \supseteq \mathcal{H}_t$, and therefore, using (5.47)

$$\mathbb{P}\{\tilde{\mathscr{H}_t}\} \leq \mathbb{P}\{\tilde{\mathcal{H}}_t\} \leq (2^9 + 1)\epsilon\delta^{-1} \left|\log_2(\epsilon\delta)\right| \tag{5.53}$$

Probability $m(X_t, Y_t)$ is too small For $t \in I_r$, define $\mathscr{M}_t$ to be the set of all (X_t, Y_t) such that $m(X_t, Y_t) < 2^{r+1}$. As usual, define $\tilde{\mathscr{M}_t}$ to be the cumulative set. Since at the start $s(\sigma, \tau) \geq 2^{j+1}$, $m(\sigma, \tau) \geq 2^{j+1}$ is forced. By assumption, $V_0(2^{j+1}) \geq \delta n$, so any inequalities derived in Lemma 5.35 are in force. Thus, from Equation (5.50),

$$\mathbb{P}\{m(X_t, Y_t) < 2^{r+1} \text{ for any } t \in I_r, \text{ for } r \in [j, K]\} \leq (2^9 + 2^8 + 1)\epsilon\delta^{-1} \left|\log_2(\epsilon\delta)\right|$$

and clearly, from the definition of $\tilde{\mathscr{M}}^t$,

$$\mathbb{P}\{\tilde{\mathscr{M}}^t\} \leq (2^9 + 2^8 + 1)\epsilon\delta^{-1} \left|\log_2(\epsilon\delta)\right| \tag{5.54}$$

Probability $s(X_t, Y_t)$ doesn't grow sufficiently during I_r As before, the only remaining times that $s(X_t, Y_t)$ can fail is at time τ_r. Therefore, define $\mathscr{B}_r$ to be the set of (X_{τ_r}, Y_{τ_r})

that are not in $\tilde{\mathscr{F}}_{\tau_r}$, $\tilde{\mathscr{H}}_{\tau_r-1}$ or $\tilde{\mathscr{M}}_{\tau_r}$, such that $s(X_{\tau_r}, Y_{\tau_r}) < 2^r$ and that have not failed previously. If (X_{τ_r}, Y_{τ_r}) is in $\mathscr{B}_r$, then it had not failed in I_{r-1}, and therefore for $t \in I_{r-1}$, $s(X_t, Y_t) \geq 2^{r-1}$. Furthermore, since (X_{τ_r}, Y_{τ_r}) is not in $\tilde{\mathscr{F}}_{\tau_r}$, for $t \in I_{r-1}$, $s(X_t, Y_t) < 2^r$. Thus, for $t \in I_{r-1}$,

$$2^{r-1} \leq s(X_t, Y_t) < 2^r \tag{5.55}$$

Furthermore, since $(X^{\pi}_{\tau_r}, X^{\sigma}_{\tau_r})$ is not in $\tilde{\mathscr{M}}_{\tau_r}$, for $t \in I_{r-1}$

$$m(X_t, Y_t) \geq 2^r$$

Finally, since $\mathscr{B}_r$ is disjoint from $\mathscr{H}_{\tau_r-1}$, for every $t \in I_{r-1}$, $\left|V_t(2^{r-1})\right| \geq \delta n/2$. Now apply Lemma 5.28 with $R = \delta n/2$ and $x = y = 2^{r-1}$. For any $t \in I_r$,

$$\mathbb{P}\{s(X_{t+1}, Y_{t+1}) \geq 2^r\} \geq \frac{2x(R - 3x - 3y)}{n^2} = \frac{2^r(\delta n/2 - 3 \cdot 2^r)}{n^2}$$

Since $r \leq K = \lceil \log_2(\epsilon \delta n) \rceil$, and since $\epsilon < \frac{1}{32}$, $3 \cdot 2^r \leq 6\epsilon\delta n \leq \frac{\delta n}{4}$. Thus,

$$\mathbb{P}\{s(X_{t+1}, Y_{t+1}) \geq 2^r\} \geq 2^{r-2}\delta n^{-1}$$

Finally, the probability of $\mathscr{B}_r$ is the probability that $s(X_{t+1}, Y_{t+1})$ isn't at least 2^{r+1} for any $t \in I_r$, and therefore,

$$\mathbb{P}\{\mathscr{B}_r\} \leq \left(1 - 2^{r-2}\delta n^{-1}\right)^{a_{r-1}} \leq \exp(-2^{r-2}\delta n^{-1} a_{r-1})$$

Now, since $a_{r-1} = \lceil 2\delta^{-1} 2^{-r+1} n(\log_2 n - r + 1) \rceil$,

$$\mathbb{P}\{\mathscr{B}_r\} \leq e^{r-1-\log_2 n} \leq 2^{r-1-\log_2 n} = \frac{2^{r-1}}{n}$$

using the fact that $r \leq K \leq \log_2 n + 1$, and hence $r - 1 - \log_2 n \leq 0$. Therefore,

$$\mathbb{P}\{\tilde{\mathscr{B}}_r\} \leq \sum_{r=j}^{K} \mathbb{P}\{\mathscr{B}_r\} \leq \sum_{r=j}^{K} \frac{2^{r-1}}{n} \leq \frac{2^r}{n} \leq \epsilon\delta + \frac{1}{n} \tag{5.56}$$

Now, adding (5.52), (5.53), (5.54) and (5.56),

$$\mathbb{P}\{s(X_t, Y_t) < 2^r \text{ for any } t \in I_r, \text{ for } r \in [j, K]\} \leq 2^{11}\delta^{-1}\epsilon \left|\log(\epsilon\delta)\right| \tag{5.57}$$

as required. □

Remark 5.36. Assiduously tracking down all the constants in the above argument shows that the mixing time was bounded above by $2^{25} n \log n$ or so. This, of course, is very far from the correct answer of $\frac{1}{2} n \log n$. While this argument can almost certainly be mildly tweaked to give a less intimidating answer such as $10 n \log n$, it is unlikely that it could be manipulated to give the right constant.

5.8 Technical Lemmas

In this section, the technical results in Lemmas 5.24 through 5.28 are proved. For the convenience of the reader, the results are restated.

Lemma 5.24 (Restatement). *Let σ be in S_n, and let $(\bar{X}_t)_{t \geq 1}$ be the random transposition walk starting at σ. Then, the expected number of v such that $|C_1(v)| < |C_0(v)|$ and $|C_1(v)| < x$ is no greater than $\frac{x^2}{n}$.*

Proof: Let $\sigma = (a_1, \ldots, a_m)$. Clearly, the only way that $|C_1(v)| < |C_0(v)|$ is if the cycle containing v is split; furthermore, the only way that $|C_1(v)| < x$ is if v winds up in a piece of size less than x. The 'ordered' splitting formula shows that the probability of splitting a_i into $(r, a_i - r)$ is $\frac{a_i}{n^2}$. Consider the cases where either $r < x$ or $a_i - r < x$. Thus, summing over the possible a_i,

$$\mathbb{E}\left|\{v \text{ s.t. } |C_1(v)| < |C_0(v)|, |C_1(v)| < x\}\right| \leq \sum_{i=1}^{m} \left(\sum_{r=1}^{x-1} r \cdot \frac{a_i}{n^2} + \sum_{r=a_i-x+1}^{a_i-1} (a_i - r) \cdot \frac{a_i}{n^2} \right)$$

It's clear that

$$\sum_{r=a_i-x+1}^{a_i-1} (a_i - r) \frac{a_i}{n^2} = \sum_{r=1}^{x-1} r \cdot \frac{a_i}{n^2} = \frac{a_i}{n^2} \sum_{c=1}^{x-1} r \leq \frac{a_i x^2}{2n^2}$$

Therefore,

$$\begin{aligned} \mathbb{E}\left|\{v \text{ s.t. } |C_1(v)| < |C_0(v)|, |C_1(v)| < x\}\right| &\leq \sum_{i=1}^{m} \frac{a_i x^2}{n^2} = \frac{x^2}{n^2} \sum_{i=1}^{m} a_i \\ &= \frac{x^2}{n^2} \cdot n = \frac{x^2}{n} \end{aligned}$$

as required. □

For the next four lemmas, let (X_t, Y_t) be our usual coupling starting at (σ, τ), where $\rho(\sigma, \tau) = 1$, $s(\sigma, \tau) = b$ and $m(\sigma, \tau) = c$. For these proofs, it will be useful to reference the original definition of the coupling and the possible pairs (X_1, Y_1) in Definition 5.12.

Lemma 5.25 (Restatement). *If $x \leq c$, then*

$$\mathbb{P}\left\{m(X_1, Y_1) < x\right\} \leq \frac{2x^2}{n^2}.$$

Proof: Let us assume without loss of generality that

$$\begin{aligned} \sigma &= (a_1, \ldots, a_n, b, c) \\ \tau &= (a_1, \ldots, a_n, b + c) \end{aligned} \tag{5.58}$$

Consider how $m(X_1, Y_1)$ could be smaller than c. Note that performing an operation involving only the a_i on σ and τ, then X_1 and Y_1 will still differ in b, c and $b + c$, so $m(X_1, Y_1) = c$. Furthermore, merging a_i with b in σ and a_i with $b + c$ in τ, then X_1 and Y_1 will differ in the parts $(b + a_i, c, b + c + a_i)$, which are greater, respectively, than $(b, c, b + c)$. This means that $m(X_1, Y_1) \geq c$. Similar reasoning holds for merging a_i with c in σ, and hence these cases do not contribute to $\mathbb{P}\{m(X_1, Y_1) < x\}$.

Also, note that if b is split into $\{r, b - r\}$ for $r \leq \frac{b}{2}$, then

$$\begin{aligned} X_1 &= (a_1, \ldots, a_m, r, b - r, c) \\ Y_1 &= (a_1, \ldots, a_m, r, b + c - r) \end{aligned}$$

Clearly, $c \geq b \geq b - r$, and therefore $m(X_1, Y_1) = c$. Thus, m cannot decrease if b is split in σ. This gives cases: splitting c in σ, and merging b and c in σ. The cases in which the coupling meets can be ignored, since $m(\alpha, \alpha) = n \geq c$, and hence these cases do not contribute to $\mathbb{P}\{m(X_1, Y_1) < x\}$.

Splitting c in σ: If c is split into $\{r, c - r\}$ for $r \leq \frac{c}{2}$, then

$$\begin{aligned} X_1 &= (a_1, \ldots, a_m, r, b, c - r) \\ Y_1 &= (a_1, \ldots, a_m, r, b + c - r) \end{aligned}$$

Clearly, $m(X_1, Y_1) \geq c - r$. Thus, to have $m(X_1, Y_1) < x$, it must be that $c - r < x$, and thus $r > c - x$. By definition, $r \leq \frac{c}{2}$, and hence

$$c - x < r \leq \frac{c}{2}$$

If $2x < c$, this set contains no elements, so assume for now that $2x \geq c$. Then the number of possible r is at most $\frac{c}{2} - (c - x) = \frac{2x-c}{2}$. Since the probability of splitting c into $\{r, c-r\}$ is at most $\frac{2c}{n^2}$ for each r,

$$\mathbb{P}\left(m(X_1, Y_1) < x,\ c \text{ split in } \sigma\right) \leq \frac{c(2x-c)}{n^2} \leq \frac{x^2}{n^2} \tag{5.59}$$

using the AM-GM inequality and the assumption that $2x - c \geq 0$. Furthermore, the above inequality also holds when $2x < c$, since in that case, the left-hand side is 0.

Merging b and c in σ: If b and c are merged in σ,

$$X_1 = (a_1, \ldots, a_m, b + c)$$
$$Y_1 = (a_1, \ldots, a_m, s, b + c - s)$$

for some $s \leq \frac{b+c}{2}$. Hence, $m(X_1, Y_1) = b + c - r$. Again, to have $b + c - r < x$, it must be that $s > b + c - x$, and the probability of each split is at most $\frac{2(b+c)}{n^2}$. Thus, analogously to above, consider

$$b + c - x < r \leq \frac{b+c}{2}$$

and hence the total number of such s is at most $\frac{2x-(b+c)}{2}$ if $2x \geq b + c$, and 0 otherwise. Therefore, if $2x \geq b + c$,

$$P\{m(X_1, Y_1) < x,\ b \text{ and } c \text{ merged in } \pi\} \leq \frac{(b+c)(2x-(b+c))}{n^2} \leq \frac{x^2}{n^2} \tag{5.60}$$

again using AM-GM. This clearly also holds for $2x < b + c$.

Finally, adding (5.59) and (5.60),

$$P\{m(X_1, Y_1) < x\} \leq \frac{2x^2}{n^2}$$

as required. □

Lemma 5.26 (Restatement). *If $x \leq c$, and $|V_0(y)| \geq R$, then*

$$\mathbb{P}\{m(X_1, Y_1) \geq x + y\} \geq \frac{2c(R-2c)}{n^2}$$

Proof: Consider both the possibilities that

$$\begin{aligned} \pi &= (a_1, \ldots, a_m, b, c) \\ \sigma &= (a_1, \ldots, a_m, b + c) \end{aligned} \tag{5.61}$$

and that

$$\begin{aligned} \pi &= (a_1, \ldots, a_m, b + c) \\ \sigma &= (a_1, \ldots, a_m, b, c) \end{aligned} \tag{5.62}$$

with $b \leq c$, since $V_t(y)$ is defined for X_t and not Y_t, and therefore the symmetry breaks down. Merging c with an $a_i \geq y$ will result in $m(X_1, Y_1) = c + a_i \geq x + y$. To calculate the probability of such a merge, the sum of these a_i is needed.

In both cases (5.61) and (5.62), since σ and τ agree on the a_i,

$$\sum_{a_i \geq y} a_i \geq |V_0(y)| - (b + c) \geq R - 2c \tag{5.63}$$

using Remark 5.19. For case (5.61), merging c and some $a_i \geq y$ in σ gives

$$\begin{aligned} X_1 &= (a'_1, \ldots, a'_{m-1}, b, c + a_i) \\ Y_1 &= (a'_1, \ldots, a'_{m-1}, b + c + a_i) \end{aligned}$$

where $\{a'_1, \ldots, a'_{m-1}\} = \{a_1, \ldots, a_m\}/\{a_i\}$. Clearly, $c + a_i \geq b$, and therefore $m(X_1, Y_1) = c + a_i \geq x + y$. The probability of merging c with a_i in σ is $\frac{2ca_i}{n^2}$, and thus

$$\mathbb{P}\{m(X_1^\pi, X_1^\sigma) \geq x + y\} \geq \sum_{a_i \geq y} \frac{2ca_i}{n^2} = \frac{2c}{n^2} \sum_{a_i \geq y} a_i \geq \frac{2c(R-2c)}{n^2}$$

using Equation (5.63) for the last inequality. Thus, in case (5.61) the proof is finished. Furthermore, since Equation (5.63) is symmetric for the cases (5.61) and (5.62), the second case is completely analogous. $\square$

Lemma 5.27 (Restatement). *If $x \leq b$, then*

$$\mathbb{P}\{s(X_1, Y_1) < x\} \leq \frac{4x^2}{n^2}$$

Proof: For simplicity, assume without loss of generality that π and σ satisfy (5.58) above. In the same way as in Lemma 5.25 above, any operations involving a_i cannot make $s(X_1, Y_1)$ smaller than b. Thus, the operations that might produce $s(X_1, Y_1) < x$ involve either splitting b in σ, splitting c in σ, or merging b and c in σ. Consider these cases separately. In the same way as before, the cases where the coupling meets can be ignored.

Splitting b in σ: Recall that if b is split into $\{r, b-r\}$ for $r \leq \frac{b}{2}$, then

$$X_1 = (a_1, \ldots, a_m, r, b-r, c)$$
$$Y_1 = (a_1, \ldots, a_m, r, b+c-r)$$

Thus, $s(X_1, Y_1) = \min(b-r, c) = b-r$. To have $s(X_1, Y_1) < x$, $b-r < x$ is needed. Hence, consider r such that

$$b - x < r \leq \frac{b}{2}$$

If $2x < b$, this set contains no elements, so assume $2x \geq b$. Clearly, the above set is of size at most $x - \frac{b}{2}$. The probability of splitting b into $(s, b-s)$ is at most $\frac{2b}{n^2}$ for each $s \leq \frac{b}{2}$, and therefore

$$\mathbb{P}\{s(X_1, Y_1) < x,\ b \text{ split in } \sigma\} \leq \frac{2b}{n^2}\left(x - \frac{b}{2}\right) = \frac{b(2x-b)}{n^2} \leq \frac{x^2}{n^2} \tag{5.64}$$

using AM-GM and the assumption that $2x \geq b$ for the last inequality. This clearly also holds if $2x < b$, since in that case the left-hand side is 0.

Splitting c in σ: This calculation is very similar to the above. The probability that c is split into $\{r, c-r\}$, where $r \leq \frac{c}{2}$ and $c - r < x$ is needed. Again, consider

$$c - x < r \leq \frac{c}{2}$$

and since the probability of a particular split is at most $\frac{2c}{n^2}$, assuming that $2x \geq c$, the total probability of all these cases is at most

$$\mathbb{P}\{s(X_1, Y_1) < x,\ c \text{ split in } \sigma\} \leq \frac{c(2x-c)}{n^2} \leq \frac{x^2}{n^2} \tag{5.65}$$

which again holds trivially when $2x < c$.

Merging b and c in σ: Recall that merging b and c in σ is coupled with splitting $b+c$ into $\{r, b+c-r\}$ in τ, where each split in τ occurs with the probability that it has not already been coupled with a split of b or c in σ. Thus, in this case,

$$X_1 = (a_1, \ldots, a_m, b+c)$$
$$Y_1 = (a_1, \ldots, a_m, r, b+c-r)$$

Assuming as usual that $r \leq \frac{b+c}{2}$, $s(X_1, Y_1) = r$. Now calculate the probability that $r < x$. Define

$$P_r = \mathbb{P}\{b \text{ and } c \text{ merge in } \pi,\ b+c \text{ splits into } \{r, b+c-r\} \text{ in } \sigma\}$$

and bound P_r for various values of r. Consider three different cases:

- $r < \frac{b}{2}$: In this case, splitting $b+c$ into $\{r, b+c-r\}$ in τ is coupled with both spliting b into $\{r, b-r\}$ in σ and with splitting c into $\{r, c-r\}$ in σ. Thus,

$$P_r = \frac{2(b+c)}{n^2} - \frac{2b}{n^2} - \frac{2c}{n^2} = 0 \tag{5.66}$$

- $\frac{b}{2} \leq r < \frac{c}{2}$: In this case, splitting $b+c$ into $\{r, b+c-r\}$ in τ is coupled with splitting c into $\{r, c-r\}$ in σ. Thus,

$$P_r \leq \frac{2(b+c)}{n^2} - \frac{2c}{n^2} = \frac{2b}{n^2}$$

- $\frac{c}{2} \leq r$: In this case, splitting $b+c$ into $\{r, b+c-r\}$ in τ isn't coupled with any splits in σ. Hence,

$$P_r \leq \frac{2(b+c)}{n^2}$$

Therefore, the reasoning above shows

$$P_r \le \begin{cases} 0 & r < \frac{b}{2} \\ \frac{2b}{n^2} & \frac{b}{2} \le r < \frac{c}{2} \\ \frac{2b+2c}{n^2} & \frac{c}{2} \le r \le \frac{b+c}{2} \end{cases} \implies P_r \le \frac{4r}{n^2} \tag{5.67}$$

where the right-hand inequality uses the fact that $b \le c$. Therefore,

$$\begin{aligned} \mathbb{P}\{s(X_1, Y_1) < x,\, b \text{ and } c \text{ merge in } \sigma\} &= \sum_{r=1}^{x-1} P_r \le \sum_{r=1}^{x-1} \frac{4r}{n^2} = 4\frac{x(x-1)}{2n^2} \\ &\le \frac{2x^2}{n^2} \end{aligned} \tag{5.68}$$

Adding Equations (5.64), (5.65) and (5.68) gives

$$\mathbb{P}\{s(X_1, Y_1) < x\} \le \frac{4x^2}{n^2}$$

as required. □

Lemma 5.28 (Restatement). *If x and y satisfy $x \le b < x + y \le c$, and $|V_0(y)| \ge R$,*

$$\mathbb{P}\{s(X_1, Y_1) \ge x + y\} \ge \frac{2b(R - 3x - 3y)}{n^2}$$

Proof: Just like in Lemma 5.26, consider the two possibilities that

$$\begin{aligned} \sigma &= (a_1, \ldots, a_m, b, c) \\ \tau &= (a_1, \ldots, a_m, b + c) \end{aligned} \tag{5.69}$$

and that

$$\begin{aligned} \sigma &= (a_1, \ldots, a_m, b + c) \\ \tau &= (a_1, \ldots, a_m, b, c) \end{aligned} \tag{5.70}$$

since $V_t(y)$ depends on X_t and not on Y_t. As in the previous lemma, in both cases (5.69) and (5.70),

$$\sum_{a_i \geq y} a_i \geq |V_0(y)| - (b+c) \geq R - (b+c) \tag{5.71}$$

so case (5.69) may be assumed. Identical arguments will apply for (5.70).

There are two possible ways to have $s(X_1, Y_1) \geq x + y$: either b can merge with an $a_i \geq y$ in σ, or b and c can merge in σ, while $b+c$ can be split into $\{r, b+c-r\}$ in τ, where $r \geq x + y$. Consider those cases separately.

Merging b and $a_i \geq y$ in σ: Note that if b and a_i are merged in σ, then

$$\begin{aligned} X_1 &= (a'_1, \ldots, a'_{m-1}, b + a_i, c) \\ Y_1 &= (a'_1, \ldots, a'_{m-1}, b + a_i + c) \end{aligned}$$

where $\{a'_1, \ldots, a'_{m-1}\} = \{a_1, \ldots, a_m\}/\{a_i\}$. Therefore, $s(X_1, Y_1) = \min(b + a_i, c)$. Since $b \geq x$ and $a_i \geq y$, $b + a_i \geq x + y$. By assumption, $c \geq x + y$, and so $s(X_1, Y_1) \geq x + y$.

The probability of b merging with a particular a_i is $\frac{2ba_i}{n^2}$, and using the bound in Equation (5.71),

$$\begin{aligned} \mathbb{P}\{s(X_1, Y_1) \geq x + y, b \text{ merges with some } a_i \text{ in } \sigma\} &= \sum_{a_i \geq y} \frac{2ba_i}{n^2} = \frac{2b}{n^2} \sum_{a_i \geq y} a_i \\ &\geq \frac{2b(R - (b+c))}{n^2} \end{aligned} \tag{5.72}$$

Merging b and c in σ: If $c < 2x + 2y$, it will later show that the above bound in Equation (5.72) suffices. Therefore, for this case, assume that $c \geq 2x + 2y$. Consider the probability of merging b and c in σ, while splitting $b + c$ in τ into $\{r, b + c - r\}$, where $r \geq x + y$.

Let P_r be defined as in Equation (5.66). Now a lower bound on

$$\sum_{r \geq x+y} P_r = \mathbb{P}\{\text{merge } b \text{ and } c \text{ in } \sigma\} - \mathbb{P}\{\text{merge } b \text{ and } c \text{ in } \sigma, \text{ stay at } \tau\} - \sum_{r < x+y} P_r$$

is needed. The above equality follows because merging b and c in σ is always either coupled with splitting $b + c$ into $\{r, b + c - r\}$ in τ, or staying at τ. Here is a lower bound for the right-hand side.

To start, $c \geq 2x + 2y > 2b$. By Equation (5.10),

$$\mathbb{P}\left\{\text{merge } b \text{ and } c \text{ in } \sigma, \text{ stay at } \tau\right\} = \min\left(p, \frac{1}{n}\right)$$

where $p = \mathbb{P}(b + c \text{ split into } \{b, c\} \text{ in } \tau, \text{ staying at } \sigma)$. Now, since $c > 2b$,

$$\begin{aligned} p &= \mathbb{P}(b + c \text{ split into } \{b, c\} \text{ in } \tau) - \mathbb{P}(c \text{ split into } \{b, c - b\} \text{ in } \sigma) \\ &= \frac{2(b + c)}{n^2} - \frac{2c}{n^2} = \frac{2b}{n^2} \end{aligned}$$

and hence, since $b \leq \frac{n}{2}$,

$$\mathbb{P}\left\{\text{merge } b \text{ and } c \text{ in } \sigma, \text{ stay at } \tau\right\} = \min\left(\frac{2b}{n^2}, \frac{1}{n}\right) = \frac{2b}{n^2}$$

Furthermore, since $x + y \leq \frac{c}{2}$, Equation (5.67) above implies that if $r < x + y$ then $P_r \leq \frac{2b}{n^2}$, and therefore

$$\mathbb{P}\{\text{merge } b \text{ and } c \text{ in } \sigma, \text{ stay at } \tau\} + \sum_{r < x+y} P_r \leq \frac{2b(x + y)}{n^2}$$

Thus, since the probability of merging b and c is $\frac{2bc}{n^2}$,

$$\begin{aligned} \mathbb{P}\{s(X_1, Y_1) \geq x + y, b \text{ and } c \text{ merge in } \sigma\} &= \sum_{r \geq x+y} P_r \geq \frac{2bc}{n^2} - \frac{2b(x + y)}{n^2} \\ &= \frac{2b(c - x - y)}{n^2} \end{aligned} \tag{5.73}$$

Combining all this information, if $c < 2x + 2y$, then Equation (5.72) shows that

$$\mathbb{P}\{s(X_1, Y_1) \geq x + y\} \geq \frac{2b(R - (b + c))}{n^2} \geq \frac{2b(R - 3x - 3y)}{n^2}$$

using the fact that $b \leq x + y$. Furthermore, if $c \geq 2x + 2y$, then combining Equation (5.72) and (5.73),

$$\begin{aligned} \mathbb{P}\{s(X_1, Y_1) \geq x + y\} &\geq \frac{2b\left(R - (b + c)\right)}{n^2} + \frac{2b\left(c - x - y\right)}{n^2} \\ &\geq \frac{2b(R - 2x - 2y)}{n^2} \end{aligned}$$

Hence, in either case $\mathbb{P}\{s(X_1, Y_1) \geq x + y\} \geq \frac{2b(R-2x-2y)}{n^2}$, completing the proof. $\square$

Bibliography

[1] David Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on probability, XVII*, volume 986 of *Lecture Notes in Math.*, pages 243–297. Springer, Berlin, 1983.

[2] N. Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Combin. Theory Ser. B*, 38(1):73–88, 1985.

[3] Noga Alon and Joel H. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., New York, 1992. With an appendix by Paul Erdős, A Wiley-Interscience Publication.

[4] Nathanael Berestycki, Oded Schramm, and Ofer Zeitouni. Mixing times for random k-cycles and coalescence-fragmentation chains. `http://arxiv.org/abs/1001.1894`. [Online; accessed 4-July-2011].

[5] Ivona Bezáková, Daniel Štefankovič, Vijay V. Vazirani, and Eric Vigoda. Accelerating simulated annealing for the permanent and combinatorial counting problems. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 900–907, New York, 2006. ACM.

[6] Andrei Z. Broder. Unpublished manuscript. 1985.

[7] Andrei Z. Broder. How hard is it to marry at random? (on the approximation of the permanent). In *Proceedings of the eighteenth annual ACM symposium on Theory of computing*, STOC '86, pages 50–58, New York, NY, USA, 1986. ACM.

[8] R. Bubley and M. Dyer. Path coupling: A technique for proving rapid mixing in markov chains. In *Proceedings of the 38th Annual Symposium on Foundations of Computer Science*, pages 223–, Washington, DC, USA, 1997. IEEE Computer Society.

[9] William J. Cook, William H. Cunningham, William R. Pulleyblank, and Alexander Schrijver. *Combinatorial optimization*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., New York, 1998. A Wiley-Interscience Publication.

[10] Persi Diaconis. *Group representations in probability and statistics*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988.

[11] Persi Diaconis, Ronald Graham, and Susan P. Holmes. Statistical problems involving permutations with restricted positions. In *IMS Lecture Notes Monogr. Ser*, pages 195–222, 1999.

[12] Persi Diaconis, Eddy Mayer-Wolf, Ofer Zeitouni, and Martin P. W. Zerner. The Poisson-Dirichlet law is the unique invariant distribution for uniform split-merge transformations. *Ann. Probab.*, 32(1B):915–938, 2004.

[13] Persi Diaconis and Laurent Saloff-Coste. Comparison techniques for random walk on finite groups. *Ann. Probab.*, 21(4):2131–2156, 1993.

[14] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981.

[15] Persi Diaconis and Daniel Stroock. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1(1):36–61, 1991.

[16] W. Doeblin. Esposé de la théorie des chaînes simple constantes de Markov à un nombre fini d'états. *Rev. Math. Union Interbalkan*, 2:77–105, 1938.

[17] Bradley Efron and Vahe Petrosian. Nonparametric methods for doubly truncated data. *J. Amer. Statist. Assoc.*, 94(447):824–834, 1999.

[18] William Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.

[19] D. Foata and M. P. Schützenberger. On the rook polynomials of Ferrers relations. In *Combinatorial theory and its applications, II (Proc. Colloq., Balatonfüred, 1969)*, pages 413–436. North-Holland, Amsterdam, 1970.

[20] Jay R. Goldman, J. T. Joichi, and Dennis E. White. Rook theory. I. Rook equivalence of Ferrers boards. *Proc. Amer. Math. Soc.*, 52:485–492, 1975.

[21] David Griffeath. A maximal coupling for Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31:95–106, 1974/75.

[22] Phil Hanlon. A random walk on the rook placements on a Ferrers board. *Electron. J. Combin.*, 3(2):Research Paper 26, approx. 24 pp. (electronic), 1996. The Foata Festschrift.

[23] Thomas P. Hayes and Alistair Sinclair. A general lower bound for mixing of single-site dynamics on graphs. *Ann. Appl. Probab.*, 17(3):931–952, 2007.

[24] Thomas P. Hayes and Eric Vigoda. Variable length path coupling. *Random Structures Algorithms*, 31(3):251–272, 2007.

[25] T.P. Hayes and E. Vigoda. A non-markovian coupling for randomly sampling colorings. In *Foundations of Computer Science, 2003. Proceedings. 44th Annual IEEE Symposium on*, pages 618–627, oct. 2003.

[26] Gordon James and Adalbert Kerber. *The representation theory of the symmetric group*, volume 16 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley Publishing Co., Reading, Mass., 1981. With a foreword by P. M. Cohn, With an introduction by Gilbert de B. Robinson.

[27] Mark Jerrum. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Structures Algorithms*, 7(2):157–165, 1995.

[28] Mark Jerrum and Alistair Sinclair. Approximating the permanent. *SIAM J. Comput.*, 18(6):1149–1178, 1989.

[29] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. In *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing*, pages 712–721 (electronic), New York, 2001. ACM.

[30] Mark R. Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoret. Comput. Sci.*, 43(2-3):169–188, 1986.

[31] J. Jonasson. Mixing time bounds for overlapping cycles shuffles. *Electronic Journal of Probability*, 16:1281–1295, 2011.

[32] Aldéric Joulin and Yann Ollivier. Curvature, concentration and error estimates for Markov chain Monte Carlo. *Ann. Probab.*, 38(6):2418–2442, 2010.

[33] Yevgeniy Kovchegov and Robert Burton. Mixing times via super-fast coupling. `arXiv:math/0609568v2`. [Online; accessed 4-July-2011].

[34] Gregory F. Lawler and Alan D. Sokal. Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.*, 309(2):557–580, 1988.

[35] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.

[36] Torgny Lindvall. *Lectures on the coupling method*. Dover Publications Inc., Mineola, NY, 2002. Corrected reprint of the 1992 original.

[37] László Lovász and Michael D. Plummer. *Matching theory*. AMS Chelsea Publishing, Providence, RI, 2009. Corrected reprint of the 1986 original [MR0859549].

[38] I. G. Macdonald. *Symmetric functions and Hall polynomials*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, second edition, 1995. With contributions by A. Zelevinsky, Oxford Science Publications.

[39] Peter Matthews. A strong uniform time for random transpositions. *J. Theoret. Probab.*, 1(4):411–423, 1988.

[40] C. E. McLaren, M. Wagstaff, G. M. Brittenham, and A. Jacobs. Detection of two-component mixtures of lognormal distributions in grouped, doubly truncated data: Analysis of red blood cell volume distributions. *Biometrics*, 47(2):pp. 607–622, 1991.

[41] M. Mihail. Conductance and convergence of Markov chains – a combinatorial treatment of expanders. In *Foundations of Computer Science, 1989., 30th Annual Symposium on*, pages 526 –531, oct-1 nov 1989.

[42] Yann Ollivier. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.*, 256(3):810–864, 2009.

[43] Yuval Peres. Mixing for markov chains and spin systems. `http://www.stat.berkeley.edu/users/peres/ubc.pdf`. [Online; accessed 4-July-2011].

[44] J. W. Pitman. On coupling of Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 35(4):315–322, 1976.

[45] John Riordan. *An introduction to combinatorial analysis.* Dover Publications Inc., Mineola, NY, 2002. Reprint of the 1958 original [Wiley, New York; MR0096594 (20 #3077)].

[46] Herbert John Ryser. *Combinatorial mathematics.* The Carus Mathematical Monographs, No. 14. Published by The Mathematical Association of America, 1963.

[47] Oded Schramm. Compositions of random transpositions. *Israel Journal of Mathematics*, vol. 147:pp.221–243, 2005.

[48] Alistair Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combin. Probab. Comput.*, 1(4):351–370, 1992.

[49] Richard P. Stanley. *Enumerative combinatorics. Vol. 1*, volume 49 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 1997. With a foreword by Gian-Carlo Rota, Corrected reprint of the 1986 original.

[50] Hermann Thorisson. *Coupling, stationarity, and regeneration.* Probability and its Applications (New York). Springer-Verlag, New York, 2000.

[51] L. G. Valiant. The complexity of computing the permanent. *Theoret. Comput. Sci.*, 8(2):189–201, 1979.

[52] Eric Vigoda. Improved bounds for sampling colorings. *J. Math. Phys.*, 41(3):1555–1569, 2000. Probabilistic techniques in equilibrium and nonequilibrium statistical physics.