Paths, Sampling, and Markov Chain Decomposition

A Thesis
Presented to
The Faculty of the Division of Graduate Studies

by

Russell A. Martin

In Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy in Mathematics

Georgia Institute of Technology
December 2001

Paths, Sampling, and Markov Chain Decomposition

Approved:

Dana Randall

Milena Mihail

Thomas Morley

Prasad Tetali

Vijay Vazirani

Date Approved by Chairman _____

Summary

For many decades sampling has been a valuable algorithmic tool in many scientific and engineering disciplines. Recent advances have allowed a rigorous analysis for some of these algorithms, as well as the development of new methods. We extend these methods and provide the first rigorous sampling algorithms for problems of interest in combinatorics and statistical mechanics.

The common underlying connection among the models we examine is representing them using sets of non–intersecting lattice paths, or routings, in directed graphs. Routings have been used to model structures from statistical mechanics and combinatorics such as domino tilings in the plane, triangulations, and polymers. We describe an exact sampling method that uses enumerative combinatorics and self–reducibility to generate random free routings, where the number of paths and the location of sources and sinks can vary. This technique relies on the Gessel-Viennot method for counting fixed routings in digraphs, where the number of paths is fixed, and that of Stembridge for counting free routings. The running time of this new algorithm is $O(l^2 n)$ where l is the maximum number of paths, and n is the number of vertices in the graph. Using this sampling algorithm, we describe how to generate domino tilings of reduced Aztec diamonds and lozenge tilings of hexagonal regions in the triangular lattice, both with free boundary conditions.

We also examine staircase walks, which are paths joining $(0, 0)$ to $(2n, 0)$ that use $(1, 1)$ and $(1, -1)$ steps and do not go below the x-axis. A path hitting the x-axis k times is assigned a weight of λ^k where $\lambda > 0$. A natural Markov chain that connects the state space and converges to the Gibbs measure (which normalizes these weights)

is known to be rapidly mixing when $\lambda \leq 1$. We give the first proof that this Markov chain is rapidly mixing in the case $\lambda > 1$, known in the statistical physics community as adsorbing staircase walks. The main new tool is a new decomposition method where we partition the state space into a number of smaller pieces. By analyzing the mixing time of each restricted Markov chain, together with that of a projection chain that captures the ergodic flow amongst the sets, we can bound the mixing time of the original Markov chain on the entire state space. This new disjoint decomposition method is quite general and is one of the main contributions of this thesis.

Contents

List of Figures

Acknowledgements

As many people have said in similar circumstances, no work like this is the result of one person (or even two).

First, I must thank my family for their love and support. Without my parents, of course, I wouldn't be here. They've always been a huge support, and have taught me how to be a better person just by example. My sister, Annette, whom I miss a lot, taught me something about persistence and patience. My brother, Eric, inspires me with his writing, which, in my opinion, shows perspicacity of someone with more wisdom than his age might suggest. I like to think that one day I can pick up a bit of this.

Secondly, it has been a great pleasure to work with my advisor Dana Randall. From her I have learned a great many things. Her patience with me has been generous and her enthusiasm inspiring. Her comments have certainly helped make my writing so much better than it otherwise would have been. I hope I've picked up some clarity of thought from her, as well as insight into how to think and communicate in the "computer science world." I have the distinction of being her first student. I hope those that follow are all smarter, more talented, and quicker on the uptake than me (no comments, please). It's hard to find proper words to express my thanks. Dana, let this be the start of more and even better collaborative work to come.

My friends deserve most of the accolades. Without them I would be lost. Two people in particular merit special attention, Erik Boczko and Juan Gerardo González Ibarra. I met Juan while at Clemson University working towards my Master's degree. He proved to be an intelligent man with a generous spirit and a good insight as to the

things that are most important in life. Above all, his friendship and support, with that of family and other friends, helped carry me through one very bad day and the weeks that followed. To him and the rest, I literally owe my life. I've been fortunate to visit Juan in his native México, allowing me to meet many interesting people, his family, and enjoy the hospitality of his wife Mayra and her family in El Roble, once again emphasizing the truly important aspects of life. Juan seems to have some deep hidden bit of wisdom that if I can only convince him to share it, I would be much happier for it.

Erik, it's hard to sum up what your friendship means to me in a few words. We survived the first year trial–by–fire of endless real analysis and, later, the comprehensive exams. Right from the start, it's been a pleasure to talk math with you. I often learn something every time we chat. On the other hand, you're also one of the few people I can talk with about anything, as neither of us are easily offended (except by the willful ignorance of others). Some people are taken aback by your (and, to a lesser extent, my) propensity to say exactly what's on your mind, but it's probably something that we all need to do a bit more often. I miss being able to blow off an afternoon with you and your fabulous wife, Debbie Murdock, who also deserves tremendous kudos. She's demonstrated her kindness and patience with me, letting her husband "go out and play" with his tall, gangly, goofy friend (and joining us on many occasions). She also resisted knocking me flat on my back during the (hopefully) few occasions when I crossed her generous bounds of decency.

For the past six months or so the "gamers of Tech" have helped me hold onto the one marble I have left rolling around in my head. For that I'm grateful. I hope you all had as much fun as I did, and that we can get together again sometime real soon.

Finally, here's some more of those who've made a significant difference to me and helped make things a lot more interesting. May you continue to do so for me and others. I hope I have done (and can keep doing) the same for you too: Anthony

Baker, George Cain, Sean and Kyle Curran, Sarah “Dyna-girl” Day, Shuli Fu, Luis González, Rajneesh Hegde, Luis Hernández, Rafal Komendarczyk, Mark and Heather Lockwood, Hamed Maroofi, Lisa McShine, Todd Moeller, Victor Morales Duarte and Lucrecia Vizcaino Cabarrus, Tom Morley, Shobhana Murali, Bryan and Amanda Rasmussen, Jose Miguel Renom and Claudia Antonini, Chris and Melaine Rottkamp, Laura “Shep” Sheppardson, Brian Schulte, Jacob Taylor, Prasad Tetali, and Marion Weederman.

For my family and friends.

Without you, I would be a shadow of the person
I am today and none of this would have been possible.

Chapter 1

Introduction

For many decades sampling has been a valuable algorithmic tool used in several scientific and engineering disciplines. Recently, there have been new advances in theoretical computer science which allow for a rigorous analysis of some of these algorithms, as well as the development of new methods. In this thesis we extend these methods and provide the first rigorous sampling algorithms for problems of interest in both combinatorics and statistical physics. Moreover, the methods utilized are quite general and can be applied to problems in computer science as well.

Statistical mechanics is the study of systems of matter to explain and predict the macroscopic (large–scale) properties of these systems based on the microscopic (small–scale) elements that comprise it. One outstanding difficulty is that these systems contain a large number of component particles. This often means that exact computations are intractable because of the huge number of possible configurations. One way to investigate such systems is to sample typical configurations, measure or compute the thermodynamic properties of interest, then find an average over a large number of samples. A common method for doing this is using Markov Chain Monte Carlo (MCMC) to perform a random walk on the set of all possible configurations Ω, making small local changes to move from one configuration to another. In order for this scheme to be successful, it is desired that after performing a "small" number of these changes, the resulting sample really does approximate a "typical" configuration. In other words, there is a probability distribution π on Ω that is determined by the

interactions between the component elements and the thermodynamic properties that result from these interactions. Samples obtained using MCMC, or any other method, should appear with probabilities that are close (if not equal) to this distribution.

The underlying principle of MCMC is to define a Markov chain with Ω as the state space. We can view this Markov chain as a random walk on a finite (possibly directed) graph, where the vertices represent elements of Ω. An edge joining a pair of vertices indicates it is possible to move, in one step, from one state to another. This graph representing transitions between states in Ω is called the *Markov kernel*. A common method of defining the transition probabilities is to use the *Metropolis algorithm*. With this approach, the Markov kernel is an undirected graph. Letting Δ denote the largest degree of a vertex in the Markov kernel, the Metropolis transition probabilities, also called the *acceptance ratios*, are

$$P(x, y) = \frac{1}{2\Delta} \min\{1, \frac{\pi(y)}{\pi(x)}\}$$

for any pair $x, y \in \Omega$ that are joined by an edge. These transition probabilities define a *reversible* Markov chain and, assuming the Markov kernel is a connected graph, by starting at any state and performing a random walk we will in fact converge to the desired distribution π.

Physicists have long used Monte Carlo methods for generating samples. Only recently have some of these methods been placed on a firm footing, by rigorously showing that they do produce samples distributed (approximately) according to π in a small number of steps of the random walk. The goal of this monograph is to present new methods that aid in analyzing such sampling methods. In particular, we present techniques that will allow us to sample configurations of some physical systems. While we consider specific models, we believe our methods will extend to a much wider variety of problems including not only models from physics, but many types of sampling problems from theoretical computer science and mathematics. Many such

questions may be cast in a combinatorial setting allowing us to examine them free from consideration of their physical interpretations. In some cases, we avail ourselves of enumerative methods from combinatorics for exact solutions, or to demonstrate auxiliary properties of some systems useful in counting or approximately counting the number of configurations, sampling configurations, or otherwise examining these systems.

We first describe some statistical physics models and related combinatorial models, some of which we explore in this thesis.

1.1 Physical systems and related models

To motivate the versatility of the Markov Chain Monte Carlo method, we present some examples to show why efficient sampling methods are desirable. A classical example from statistical mechanics is the *Ising model*, a model for ferromagnets. The essential idea is the magnet is composed of a large number of dipoles caused by the spins of atoms and molecules, and the large scale properties of the material are determined by the interactions of these component spins. In such a model we consider a finite lattice G and let σ denote an assignment of spins at each vertex of the lattice, each spin either $+1$ or -1. The *interaction energy* $E(\sigma)$ of a spin configuration σ is defined as

$$E(\sigma) = -J \sum_{i \sim j} \sigma(i)\sigma(j) - H \sum_{i} \sigma(i)$$

where the first summation is over all vertices i, j that are neighbors in the lattice, H is the *external magnetic field*, and J is the *coupling constant*.

Gibbs [18] showed that for a given temperature T, the probability of observing the system in the state σ is given by

$$\pi(\sigma) = \frac{\exp(-\beta E(\sigma))}{Z(G)}$$

where β is inversely proportional to the temperature. This probability distribution is often referred to as the *Gibbs distribution* on the set of spin configurations and is also sometimes called the *Boltzmann distribution* after another pioneer of statistical mechanics. The normalizing constant $Z(G)$, called the *partition function*, is

$$Z(G) = \sum_{\sigma} \exp(-\beta E(\sigma))$$

where the sum is over all 2^n possible spin configurations, with n denoting the number of vertices of G. The Gibbs distribution assigns more probability to those configurations having low energy, agreeing with the thermodynamics convention that systems will naturally tend towards a low energy state. Note also that taking $J > 0$ favors neighbors to have the same spin, and for fixed $J > 0$ and H, at very high temperatures (very small values of β) the assignment of individual spins to the vertices is essentially independent, as the Gibbs distribution is approximately uniform. At low temperatures (large values of β) the Gibbs distribution is bimodal, where configurations that are predominantly $+1$ or predominantly -1 have much higher probability than those with an even mixture of the two states.

One fundamental interest when examining the Ising model is to determine $Z(G)$. More than simply a normalizing constant for the Gibbs distribution, the partition function captures many of the thermodynamic properties of the system such as the free energy, internal energy, entropy, and magnetic susceptibility; these quantities may be determined from computations involving $Z(G)$, $\log Z(G)$, various derivatives of $Z(G)$, etc. For most applications physicists are concerned with the *thermodynamic limit* of $Z(G)$, that is, the limit as the size of G increases through a series of lattices to the infinite lattice. The only interesting case for which an exact solution has been obtained is that of Onsager for the 2-dimensional square lattice $\mathbb{Z}^2$ with no external field ($H = 0$) [3, 49]. As yet, more than fifty years later, no extension for an exact solution into higher dimensions is known. It is clear that, even for small lattices,

exact computation of the partition function by enumeration of all configurations is infeasible; a $5 \times 5 \times 5$ cube in $\mathbb{Z}^3$ has $2^{125} \approx 4 \cdot 10^{37}$ possible assignments of signs.

Kasteleyn demonstrated a connection between the 2-dimensional Ising model and another statistical physics problem involving dimers [29, 46]. A *dimer*, or diatomic molecule, is the edge between two nearest neighbors in a graph. A *perfect matching* is a collection of dimers (edges) such that every vertex is adjacent to exactly one of the dimers. The *dimer problem* is that of counting the number of perfect matchings in the lattice. Fisher, Kasteleyn, and Temperley independently showed how to exactly count the number of perfect matchings for regions of $\mathbb{Z}^2$ [14, 28, 60]. Kasteleyn also demonstrated that counting the number of perfect matchings in general planar graphs can be done by computing a Pfaffian of a matrix (which can be performed efficiently by computing a determinant) [30]. However, counting the number of perfect matchings in a general graph is $\#P$-complete [61] meaning that, under current complexity theory views, an exact solution to this problem is believed to be computationally intractable. For a precise description of the correspondence between spin configurations of the 2-dimensional Ising model and dimer configurations on a related lattice (and the solution of finding the partition function using this relationship) the interested reader is referred to [29, 46]. Barring exact solutions, we seek methods to approximate the number of matchings, or ways to estimate the partition function of the Ising model. In some cases we are able to do this efficiently, and we describe one way to do so for the dimer (matching) problem.

Counting perfect matchings in bipartite graphs is equivalent to computing the permanent of the adjacency matrix of the graph. For an $n \times n$ matrix A, the *permanent* is defined as

$$per(A) = \sum_{\tau} \prod_{i=1}^{n} a_{i,\tau(i)}$$

where the sum is over all permutations τ of $\{1, \ldots, n\}$. Despite the similarity to

the determinant, Valiant showed that exact computation of the permanent is $\#P$-complete [61]. A general approximation algorithm for the permanent of non–negative matrices has only recently been given by Jerrum, Sinclair, and Vigoda [25]. This Monte Carlo method is an extension of a previous result by Jerrum and Sinclair for sampling perfect matchings in dense bipartite graphs [24, 57]. An obvious generalization of the matching problem is the *monomer–dimer* problem where monomers correspond to unmatched vertices in the lattice, the goal being to count all such matchings, matchings with a specified number of edges, or, perhaps, to select a matching at random. Exactly counting the number of matchings of all sizes is a $\#P$-complete problem, even when restricted to planar graphs [21, 62].

One further interesting problem involving matchings is obtained when we give weights to each matching. Let $\lambda > 0$ denote a fixed parameter. For a matching M with k edges, we assign the weight $wt(M) = \lambda^k$ to M. Normalizing these weights, we obtain a probability distribution on the set of all matchings where $\pi(M) = wt(M)/Z$ with $Z = \sum_{M'} wt(M')$ as the partition function in this case, the sum being over all possible matchings M'. If we treat λ as a formal variable, the partition function is the generating function for the number of matchings in the graph, i.e. the coefficient of λ^k in Z is equal to the number of matchings with exactly k edges. However, as stated above, exact computation of the partition function for a general graph is (believed to be) computationally intractable since counting the total number of matchings is $\#P$-complete.

Another way to interpret dimer problems on a lattice are as tilings of regions in the plane. We describe this interpretation because it will be relevant in chapter 2 for the first set of results in this thesis. Let R denote a $2n \times 2n$ region in the plane. A *domino tiling* of R is a cover of R by non–overlapping 2×1 and 1×2 rectangles. While the results of Fisher, Kasteleyn, and Temperley [14, 30, 60] let us count the total number of tilings (dimer arrangements), some physical properties of these systems may be

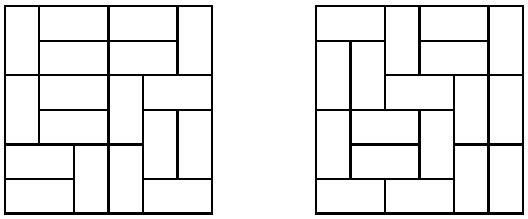


Figure 1.1: Domino tilings of a 6×6 region

expressed as the expected value, over the uniform distribution, of a function defined on the set of configurations. Such a function might measure the correlation between the orientation of dominoes covering two squares, or count the number of horizontal dominoes in the whole tiling. Because of this, much interest lies in being able to sample random domino tilings in order to estimate these physical properties using an average over a large number of samples. Using Markov Chain Monte Carlo Luby, Randall, and Sinclair [36] have given a method for sampling domino tilings from a distribution that is approximately uniform over the set of possible tilings. The basic method is to iteratively choose a random 2×2 window in the region R and, if the four squares are covered by a pair of dominoes, rotating the dominoes by 90° in the plane. See figure 1.2 for an example. By repeating this iterative step $\lceil 512en^8 \ln(\varepsilon^{-1}) \rceil$

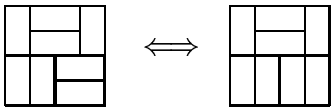


Figure 1.2: Domino rotation

times (for $0 < \varepsilon < 1$), we obtain a domino tiling of R from a distribution that is within a distance of ε from the uniform distribution. Remarkably, this means that by performing only a *polynomial* number of steps, we may obtain a (nearly) random tiling from an *exponentially* sized set; the number of domino tilings of this $2n \times 2n$ region is at least 2^{n^2}, which can be seen by dividing R into n^2 small 2×2 squares and tiling each square by a pair of vertical or horizontal dominoes.

Much attention also has been given to tiling a differently shaped region. The *Aztec diamond of order n* is a region consisting of $2n$ centered rows of squares, where the kth row has $\min\{2k, 4n - 2k + 2\}$ squares. We tile this region with dominoes (see figure 1.3, a tiling generated using an applet by Jason Woolever [50]). For this

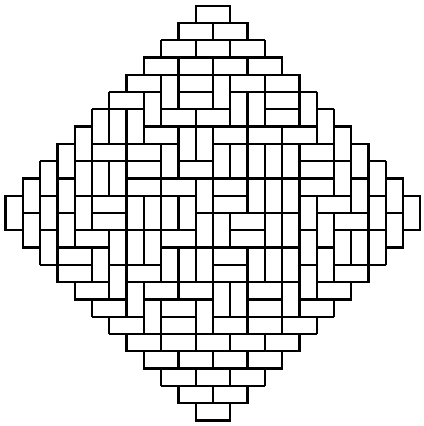


Figure 1.3: A random domino tiling of an Aztec diamond of order 12

specially shaped region, in addition to the domino rotation algorithm described for $\mathbb{Z}^2$, there is also a "shuffling algorithm" that may be used to generate perfectly random tilings by starting with a tiling of single square (an Aztec diamond of order one, with two possible tilings) and successively generating random tilings of larger Aztec diamonds until the proper size is obtained. This algorithm was used to prove these regions have exactly $2^{n(n+1)/2}$ distinct domino tilings [12].

Some of these tiling problems point to an interesting phenomena when we consider the thermodynamic limit. The *entropy* is defined as $h(\Lambda) = \lim_{n\to\infty} \frac{\log \#(\lambda_n)}{\mathrm{Area}(\lambda_n)}$ where $\Lambda = \{\lambda_n\}$ is a family of (tileable) regions that tend to the infinite lattice and $\#(\lambda_n)$ is the number of tilings of λ_n. The family of square regions has been shown to have maximal entropy over all families of finite regions, while that of Aztec diamonds is strictly smaller. This is related to the so–called arctic circle phenomena whereby

frozen regions of the Aztec diamond emerge near the boundary having completely predictable local tilings [27]. So, along with being able to compute sample averages of functions defined on the set of tilings previously mentioned, it is desired to have fast methods to generate random tilings to obtain empirical evidence for conjectures about properties of the collection of tilings (or other general combinatorial structures).

The tilings described and pictured so far are those having *fixed boundary conditions* where the tiles are forced to lie entirely within the region of interest. Alternatively, we may consider *free boundary conditions* where we allow tilings that can be seen within a *window* in the shape of the region. This means that tiles are allowed to overlap the boundary of the region, provided that the tiling can be extended to one of the plane. A third possibility is *periodic boundary conditions* where the lattice "wraps around" so that it is embedded on a torus, and we might even consider hybrid lattices that have fixed boundaries in one direction and periodic in another, etc. See figure 1.4 for examples. Much of the work to this point has focused on counting (or sampling from)

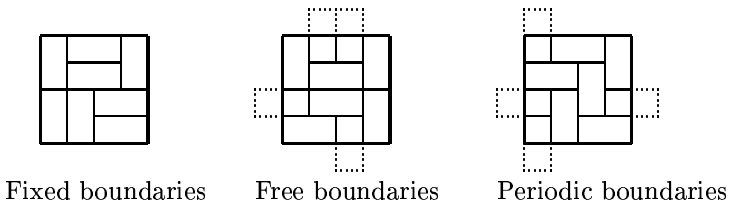


Figure 1.4: Domino tilings on regions with different boundary conditions

the collection of fixed boundary tilings, although recent work of Kenyon, Randall and Sinclair allows approximation of the number of tilings of rectangular lattices (in any dimension) with periodic boundary conditions [31]. Their method extends to approximating the number of monomer–dimer arrangements (matchings with some unmatched vertices) for any lattice with periodic boundary conditions or any finite vertex–transitive graph.

One reason for studying free boundary conditions is to eliminate the effect of the

shape of the region when considering the thermodynamic limit. For example, we have already mentioned that with fixed boundary conditions, the entropy of the family of Aztec diamonds is strictly less than that of square planar regions. With free boundary conditions, for any family of regions where the ratio of the perimeter to the area tends to zero, it has been shown that the entropy converges to the same (maximal) value. In other words, statistics of tilings of square regions with free boundary conditions agree with statistics of tilings of Aztec diamonds with free boundary conditions. Part of this thesis describes a method that allows us to count the number of free boundary tilings of some regions. Using this counting technique, we give an algorithm that allows us to sample uniformly at random from the set of all such free boundary tilings. This algorithm relies on a bijection between tilings and routings in a related graph, where a *routing* is a collection of non–intersecting paths between sources and sinks in a directed graph. In next section we show how many statistical mechanics problems share this common underlying combinatorial interpretation, describe some of the methods and tools available to help investigate them, and how this thesis addresses further avenues that will aid in the analysis of these combinatorial and physical models. While we examine some specific applications, our overall intent is to present general methods that can be be used to probe a wide variety of questions, those arising from statistical mechanics and combinatorics, along with more elaborate ones that arise in theoretical computer science.

1.2 Paths and Routings

A common theme in this work will be exploiting the correspondence between statistical mechanics problems, combinatorial models, and various path structures. For example, we have noted that Kasteleyn demonstrated the connection between the 2-dimensional Ising model and dimer arrangements [29, 46]. We have also seen how

dimer arrangements correspond to tilings of regions. Several algorithms for sampling tilings on regions with fixed boundary conditions rely on a bijection between tilings and sets of non–intersecting paths. These are *fixed routings*. The first is a Markov chain approach of Luby, Randall, and Sinclair [36] which samples routings uniformly at random (and can be extended to the case where the paths are edge disjoint, but not necessarily vertex disjoint). A second approach is the determinant based algorithm Wilson has used [63].

In this thesis we explore a variant called *free routings*, where the number and location of sources and sinks are not fixed (but lie in specified sets). Using them we can gain insight into free boundary tilings of some regions. Let us examine the correspondence between domino tilings of regions in $\mathbb{Z}^2$ and routings in a digraph in the case of fixed boundaries. First, color the lower left square of the region black, then extend the coloring of the region with alternating black and white squares, like a checkerboard. Mark the midpoint of the left edge of each black square (doing so on the right boundary of the region as if the coloring extended beyond the boundary). These points form the vertex set of the digraph. Join the vertex (x, y) to vertices $(x+1, y+1), (x+1, y-1)$, and $(x+2, y)$. Directing these edges from left to right gives the digraph we want. Superimpose the tiling of the region on top of this graph. The edges we take for the corresponding routing in the digraph are those that lie entirely inside of dominoes in the tiling. For any tiling of the 6×6 square region in figure 1.5 there will be exactly three paths in the corresponding routing. The analysis in [36] shows there is an efficient Markov chain for sampling fixed routings in some planar graphs. Therefore, we have an efficient method for sampling fixed boundary tilings using the bijection between the set of tilings and fixed routings. In chapter 2 we use a similar correspondence when we present our algorithm that can be used to sample free boundary tilings of some regions.

The second approach to sampling, due to Wilson [63], uses the Gessel-Viennot

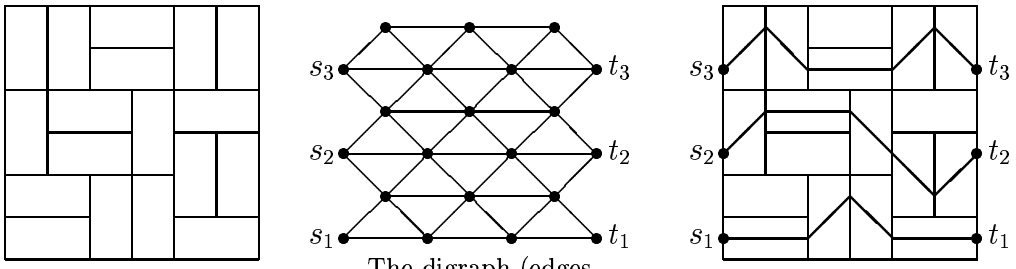


Figure 1.5: Domino tilings and routings

method for enumerating fixed routings (and the close relationship between counting and sampling formalized by Jerrum, Valiant and Vazirani [26]). The remarkable result by Gessel and Viennot [15, 16] lets us compute the number of routings in a finite, directed, acyclic graph by finding the determinant of a certain matrix.

The relationship between counting and sampling will be utilized in chapter 2 when we generalize the approach for sampling free routings, so we outline the idea here. Jerrum, Valiant and Vazirani [26] characterized the notion of *self–reducibility*. Informally, a problem is self–reducible if the solution may be found by examining a few similar problems of smaller size. For example, to build a path in a digraph starting at a specified source vertex, first choose an outgoing edge from that vertex. Then, delete the source (and all edges adjacent to it) to get a smaller graph, and iterate this procedure, starting again with the chosen neighbor of the first vertex as the new source. The algorithm of Wilson builds routings edge by edge in this fashion, utilizing other techniques to speed up computation of the probabilities of selecting an edge at each step. Problems that are self–reducible are also important in that if we are able to (approximately) count the size of the solution set, then we are able to efficiently sample elements from that set, and vice–versa. This relationship has been exploited for many combinatorial problems.

Our new result combines self–reducibility with a method of Stembridge [59] for

counting the number of free routings in a digraph. In this approach, the Gessel-Viennot matrix is replaced by a special skew–symmetric matrix (where $A^T = -A$), and instead of the determinant we use the Pfaffian of the matrix. We further extend the result of Stembridge by showing how we may compute the number of routings where certain of the sources are specified to be used or unused in the routings. With this further extension, we give an algorithm for sampling free routings. Using this algorithm we show how to generate random tilings of reduced Aztec diamonds and lozenge tilings of hexagons with free boundary conditions, relying on a similar bijection between these tilings and free routings in related digraphs.

The second part of this thesis examines a different, more direct, role of paths in combinatorics and physics problems. One example physicists study that is modeled by paths is polymers. A linear polymer (in dilute solution) is typically modeled as a self–avoiding random walk in a lattice, i.e. a path in the lattice with no self intersections. There are many variations on these types of models, where one or both ends of the polymer are attached to specific sites, the polymer has interactions (attraction or repulsion) with itself or other surfaces, restrictions where the polymer may not penetrate surfaces, etc. See [40] for an extensive treatise on self–avoiding walks and [52, 53, 54] for details of some of the statistical physics models. One specific example we examine is staircase walks. A *staircase walk* is a path joining $(0, 0)$ to $(2n, 0)$ that uses $(1, 1)$ and $(1, -1)$ steps, and does not go below the x-axis. It is well known (see Stanley [58]) that there are $C(n)$ such paths where $C(n) = \frac{1}{n+1}\binom{2n}{n}$ are the Catalan numbers.

Let $\mathcal{S}^n$ denote the set of staircase walks joining $(0, 0)$ to $(2n, 0)$. Similar to matchings, we define a probability distribution on $\mathcal{S}^n$ by assigning weights to them. Let $\lambda > 0$. Assign the weight λ^k to a staircase walk $\sigma \in \mathcal{S}^n$ that hits the x-axis k times. Normalize the weights to obtain a probability distribution on $\mathcal{S}^n$. Sampling staircase walks is one example where the Monte Carlo Markov Chain method has been shown

to be useful. A natural Markov chain on $\mathcal{S}^n$ consists of changing local maxima to local minima by changing a pair of consecutive "up,down" edges to a "down,up" pair, or vice–versa. This Markov chain has proven to be efficient for sampling from $\mathcal{S}^n$ in the uniform case ($\lambda = 1$), as well as the case when $\lambda < 1$ [36, 47, 64]. We provide the first proof that this chain is also efficient in the case $\lambda > 1$, demonstrating our new disjoint decomposition method for analyzing Markov chains. This new decomposition theorem is presented in chapter 5, and the analysis of the chain for staircase walks when $\lambda > 1$ is shown in chapter 6.

As we can see from the examples above, many models can be represented as a path or routing in a (directed) graph. Because of this, one over–riding goal is to find efficient ways to count and/or sample from the collection of all paths (or routings) in graphs, since then many seemingly different types of problems can then be examined at the same time.

This thesis is uses two approaches to these counting and sampling problems. The first, based on the enumerative method of Stembridge previously mentioned, uses counting in order to sample. The second part of the thesis is based on Monte Carlo methods to sample directly. We now give an overview of the Markov Chain Monte Carlo method for sampling from a large combinatorial set.

1.3 Markov Chain Monte Carlo

Physicists and other scientist use Markov Chain Monte Carlo (MCMC) methods extensively, as they are often the only available approach to sampling. The main idea is that we have a set Ω, from which we desire to sample elements according to some probability distribution π. To do so, we define a graph G whose vertex set is $V(G) = \Omega$. The edges of G represent a way of moving from one element of Ω to another, typically by performing some small change in the combinatorial structure of

the elements. G is called the *Markov kernel*. For example, recall the Ising model is a representation of a magnetic material, given by an assignment of a spin of $+1$ or -1 to each site in a lattice L. The typical Markov kernel in this case is given by using the set of all spin configurations as the vertex set of G, joining two spin configurations by an edge if they differ solely at a single site in L.

Having defined the Markov kernel G, we perform a random walk on this graph. We want to do it in such a way that the distribution converges to π, the desired distribution on Ω. As previously mentioned, a common way to perform the random walk is to use the Metropolis algorithm. If $x, y \in \Omega$, where $x \neq y$, are joined by an edge in the Markov kernel, then

$$P(x,y) = \frac{1}{2\Delta} \min\{1, \frac{\pi(y)}{\pi(x)}\}$$

where Δ is the maximum degree of a vertex in G. We also let $P(x,x) = 1 - \sum_{y \neq x} P(x,y)$. If the Markov kernel is a connected graph, these transition probabilities define a Markov chain $\mathcal{M}$ whose stationary distribution is π.

The main task, then, is to determine the number of steps of this random walk that are necessary until we are "close" to the distribution π. More precisely, for $\varepsilon > 0$ we want to show that there is a polynomial q in n and $\log \varepsilon^{-1}$, where n is the size of a typical element in Ω, such that after $q(n, \log \varepsilon^{-1})$ steps of the random walk, we are within distance ε of the distribution π on Ω. A Markov chain having this property is said to be *rapidly mixing*. We define this more formally in chapter 3.

The main tool we use to investigate these chains is coupling. Informally, a *coupling* of the Markov chain $\mathcal{M}$ is a new Markov chain on $\Omega \times \Omega$ which we view as pair of random walks on Ω. The transitions of the pair may be correlated, but, each random walk, viewed in isolation, looks like the original chain $\mathcal{M}$. Bounding the expected time until the random walks coalesce, maximized over all possible pairs of starting states of each walk, allows us to infer a bound on the mixing time of the chain.

The salient requirement to use this method is to show the coupling is contracting on the state space so that pairs of states in Ω tend to move togther in expectation under some distance measure on the state space. Path coupling is a more specialized version where we need only examine a smaller subset of pairs in Ω and show that the coupling is contracting on that smaller set in order to show rapid mixing of the chain. Couplings without this contraction property are of no use to us in helping bound mixing times.

One impetus for much of the work in this thesis came initially from a problem where coupling fails to show rapid mixing, namely sampling adsorbing staircase walks. Recall that a staircase walk is a path from $(0,0)$ to $(2n,0)$ using $(1,1)$ and $(1,-1)$ steps, not going below the x-axis. McShine and Tetali [47] have used coupling to show the natural Markov chain is rapidly mixing when sampling uniformly from the set $\mathcal{S}_n$ of all staircase walks of length $2n$. Also recall that we want to consider the weighted probability distribution obtained by assigning the weight λ^k to a staircase walk hitting the x-axis k times. If $\lambda = 1$ we have the uniform distribution on $\mathcal{S}^n$, and when $\lambda \leq 1$ an application of path coupling proves rapid mixing as pairs of paths will tend to move together. However, when $\lambda > 1$ (called *adsorbing* staircase walks), coupling no longer applies as pairs of paths will tend to move apart in expectation. We need something else in order to prove rapid mixing in this case.

In order to circumvent the difficulties in applying coupling to this and other problems, indirect methods such as comparison of Markov chains and decomposition are proving to be indispensable tools. One of the main results in this thesis is a new decomposition method that allows analysis of a complicated Markov chain by breaking the state space into smaller pieces and analyzing simpler, derivative Markov chains. If the restricted Markov chains on each of the pieces (the *restrictions*) are rapidly mixing, and if a *projection chain* (both defined later) is also rapidly mixing, suggesting that it is easy to move from one set to another, we may conclude that the original

Markov chain on the entire state space is rapidly mixing. This method is similar in spirit to an earlier decomposition method of Madras and Randall [38, 39]. However, in our method the decomposition is a partition of Ω, whereas in the Madras/Randall theorem the decomposition must be a cover of Ω where overlapping sets are necessary. Our method is typically more straightforward to apply as often there is a natural partition of Ω based on some combinatorial property of the elements. With our method, we can avoid the requirement of converting the partition to a cover. This new decomposition technique provides another general tool in the arsenal of techniques for analyzing Markov chains.

We demonstrate the application of this new theorem by examining the Markov chain on adsorbing staircase walks (when $\lambda > 1$). In doing so we utilize couplings, the comparison method, and analyze the auxiliary projection Markov chains that arise from decomposition. Through this analysis, we gain insight into the relevant difficulty that prevents a simpler, direct argument for adsorbing staircase walks. In chapter 6, we use the decomposition method twice to establish rapid mixing of the staircase walk chain. One of the projection chains that arises is an interesting exclusion process, a Markov chain with state space consisting of arrangements of k particles on a linear set of n sites, which is of independent combinatorial interest. Each configuration is weighted by a product of Catalan numbers, which gives rise to the probability distribution on the set of particle configurations by normalizing these weight.

The remainder of this thesis is organized as follows: Chapter 2 outlines the exact sampling method mentioned previously that relies on enumerative methods to count lattice paths and recursive probabilistic branching to construct such paths. In chapter 3 we review some background and known results about Markov chains. We apply these results in chapter 4 when we examine exclusion processes. Chapter 5 includes the statement and proof of our new disjoint decomposition theorem. In chapter 6, we apply disjoint decomposition to show the staircase walk Markov chain is rapidly

mixing. Some of this work has already appeared in published form [43, 44].

Chapter 2

Sampling Free Routings

In this chapter we present the first rigorous technique for sampling tilings with free boundary conditions using methods from enumerative combinatorics. Recall that a routing is a collection of non–intersecting paths in a directed graph. Free routings are those where the sources and sinks lie in specified sets, but they need not be utilized as endpoints of paths in the routing. As we outlined in the introduction, the sampling scheme for generating tilings with free boundary conditions is based on a method of counting free routings in digraphs. We then use the idea of self–reducibility to convert the counting scheme to a sampling method by generating routings one edge at a time. The running time of our algorithm is $O(l^2 n)$ where n is the size of the region and l is the maximal number of paths in a routing. For graphs arising from tiling problems we typically have $l = O(\sqrt{n})$, yielding an $O(n^2)$ algorithm. We apply this sampling method to provide the first method for generating random domino tilings of reduced Aztec diamonds and lozenge tilings of hexagonal regions, both with free boundary conditions.

The counting technique of Stembridge we use extends the method of Gessel and Viennot for counting fixed routings in graphs. Fixed routings are those where the number and location of all of the sources and sinks of the paths in the routing are specified. Wilson [63] uses the Gessel-Viennot counting method to generate tilings of regions with fixed boundary conditions. Counting free routings entails evaluating a Pfaffian of a certain matrix, which can be done efficiently by instead computing a

determinant.

Our first contribution is a straightforward extension of the method of Stembridge. In particular, we can require that certain of the sources are used or unused in the routings. This enables us to build the paths edge by edge in a fast way, re-using much of the information from one step to the next as we construct the routing.

Our second contribution is to apply some results from linear algebra to speed up the algorithm. We adapt the method of Colbourn, Myrvold and Neufeld [7] to show how ratios of Pfaffians can be evaluated quickly, and also make use of the Sherman-Morrison formulas [7, 63] for updating the inverse of a matrix after changing one of its rows or columns.

We conclude this chapter by demonstrating how, through a bijection between tilings and routings, the algorithm allows us to sample free boundary tilings of reduced Aztec diamonds and lozenge tilings of hexagons in the triangular lattice.

2.1 Background: Counting routings

First we begin with an overview of the methods of Gessel and Viennot for counting fixed routings and Stembridge for counting free routings. Wilson has shown how to sample fixed routings using self–reducibility and iterative applications of the Gessel-Viennot method. We will see later that similar methods can be used to sample free boundary tilings of some regions by extending the technique of Stembridge for counting free routings.

2.1.1 The Gessel-Viennot method for fixed routings

Gessel and Viennot [15, 16], and Lindström [35] introduce a method for finding the number of non–intersecting paths, with specified sources and sinks, in certain directed graphs by computing a determinant of a matrix. For their technique to work, the

graph must be directed and acyclic. Furthermore, the sources and sinks must satisfy a condition known as *compatibility*. In this definition, we require that both the set of sources $\mathcal{S}$ and (separately) the set of sinks $\mathcal{T}$ be ordered.

Definition. Let D be a directed acyclic graph. The ordered sets $\mathcal{S}$ and $\mathcal{T}$ are said to be *compatible* if $s < s'$ in $\mathcal{S}$ and $t < t'$ in $\mathcal{T}$ implies that every $s - t'$ path intersects every $s' - t$ path.

Thus, if there is a set of l non–intersecting paths using sources $s_1 < s_2 < \cdots < s_l$ and sinks $t_1 < t_2 < \cdots < t_l$, then it must be the case that s_i is joined to t_i for all i. We call such a set of l non–intersecting paths, where all of the sources and sinks appear on paths, a *fixed routing* of D. We also sometimes use the term *(fixed) l-routing* to specify the number of paths in the routing.

We have the following remarkable theorem [15, 59] stating that, in this setting, the number of fixed routings can be computed with a single determinant:

Theorem 2.1 *Let D be an acyclic digraph with compatible sets of sources $\mathcal{S} = \{s_1, \ldots, s_n\}$ and sinks $\mathcal{T} = \{t_1, \ldots, t_n\}$. Let P denote the matrix where p_{ij} is the number of directed paths with source s_i and sink t_j. The number of l-routings in D is equal to* $\det(P)$.

For example, consider the digraph in figure 2.1 where the edges are directed from

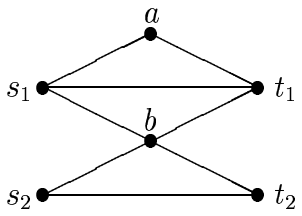


Figure 2.1: An example for the Gessel-Viennot method

left to right. Note that the sources and sinks of this digraph are compatible. In this

example, the matrix P of theorem 2.1 is $P = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$. We see that $\det(P) = 5$ agreeing with the fact that this digraph has five fixed 2-routings, namely (specifying the vertices on the paths of each routing)

1. s_1, a, t_1 and s_2, b, t_2
2. s_1, a, t_1 and s_2, t_2
3. s_1, t_1 and s_2, b, t_2
4. s_1, t_1 and s_2, t_2
5. s_1, b, t_1 and s_2, t_2.

Note, however, that if D is not acyclic, or if $\mathcal{S}$ and $\mathcal{T}$ are not compatible, then the preceding theorem fails, as the simple example in figure 2.2 illustrates. This graph (with edges directed from left to right) has two fixed routings, the pair of edges labeled

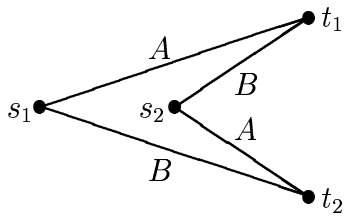


Figure 2.2: An example where compatibility fails

"A" and the pair labeled "B" (the routing that violates the compatibility condition). On the other hand, the "Gessel-Viennot matrix" for this graph is $P = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ which has a determinant of 0.

Wilson shows that the entries of P can be computed very efficiently, and uses repeated applications of the Gessel-Viennot method to sample routings.

Theorem 2.2 (Wilson [63]) *Let D be a planar, acyclic digraph with n vertices, having compatible sources and sinks. Fixed routings of D can be uniformly sampled in $O(l^{1.688}n)$ time.*

2.1.2 Stembridge's extension for free routings

Stembridge [59] provides an interesting extension of the Gessel-Viennot method to count free routings. In the case of free routings, the number of paths is no longer fixed, so S is now a the set of *potential* sources and T is a set of *potential* sinks. Also, it is no longer always true that s_i will be joined to t_i, as was true in the case of fixed routings. Again, we let D denote the underlying directed, acyclic graph.

If $s_i \in \mathcal{S}$ is a source in a free routing, we say that s_i is *used* in the routing; otherwise s_i is *unused*. We assume that $|\mathcal{S}| = |\mathcal{T}| = l$, but in the case of free routings it is not necessary that these sets have the same size.

The main idea behind Stembridge's method is to find a generating function for the number of routings. The coefficient of t^k in this generating function is the number of k-routings (routings with exactly k paths) in the digraph. This method implicitly allows sources and sinks to be used or unused in routings. First, we need a bit of linear algebra.

Definition. Let B be a $2n \times 2n$ skew–symmetric matrix (i.e. $B^T = -B$), and let

$$\pi = \{\{i_1, j_1\}, \{i_2, j_2\}, \ldots, \{i_n, j_n\}\}$$

be a partition of the set $\{1, \ldots, 2n\}$ into pairs. Let

$$b_\pi = \operatorname{sgn} \begin{pmatrix} 1 & 2 & 3 & 4 & \ldots & 2n-1 & 2n \\ i_1 & j_1 & i_2 & j_2 & \ldots & i_n & j_n \end{pmatrix} b_{i_1,j_1} b_{i_2,j_2} \cdots b_{i_n,j_n}.$$

The *Pfaffian* of B, denoted $\mathrm{Pf}(B)$, is defined by

$$\mathrm{Pf}(B) = \sum_\pi b_\pi.$$

Theorem 2.3 *If B is a skew–symmetric matrix of even size, then* $\det(B) = \mathrm{Pf}(B)^2$.

A skew–symmetric matrix, Q, will take the role of the matrix P in theorem 2.1, but instead of the determinant of Q, we look at its Pfaffian. For $1 \leq i < j \leq l$ and $1 \leq h < k \leq l$, let $\alpha_{ij}(h,k)$ denote the number of non–intersecting paths in D with sources s_i, s_j, and sinks t_h, t_k. From theorem 2.1 we find $\alpha_{ij}(h,k) = \det \begin{pmatrix} p_{ih} & p_{ik} \\ p_{jh} & p_{jk} \end{pmatrix}$, where, recall, p_{ih} is the number of paths from s_i to t_h.

Let $q_{ij} = \sum_{h<k} \alpha_{ij}(h,k)$. Then q_{ij} is the number of pairs of non–intersecting paths with sources s_i and s_j, where the sinks range over all pairs where t_h precedes t_k in the ordering of $\mathcal{T}$. Finally, let q_i denote the number of paths with source s_i to any sink in $\mathcal{T}$.

We assume that l is odd; if not, we can add an additional isolated vertex s_{l+1} to $\mathcal{S}$. The following is due to Stembridge [59]:

Theorem 2.4 *Let $\mathcal{S} = (s_1, \ldots, s_l)$ be an l-tuple of vertices in an acyclic digraph D, with l odd. Let $\mathcal{T}$ be an ordered subset of vertices that is compatible with $\mathcal{S}$. Let Q be the $(l+1) \times (l+1)$ skew–symmetric matrix where the upper triangular entries are given by*

$$[Q]_{ij} = (-1)^{i+j-1} + q^*_{ij}$$

for $1 \leq i < j \leq l+1$, where $q^*_{ij} = q_{ij}$ for $j \leq l$ and $q^*_{i,l+1} = q_i$. Then $\Phi = \mathrm{Pf}(Q)$ is the total number of free routings of D.

The matrix Q is:

$$Q = \begin{pmatrix} 0 & 1+q_{12} & \cdots & -1+q_{1l} & 1+q_1 \\ -1-q_{12} & 0 & \cdots & 1+q_{2l} & -1+q_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1-q_{1l} & -1-q_{2l} & \cdots & 0 & 1+q_l \\ -1-q_1 & 1-q_2 & \cdots & -1-q_l & 0 \end{pmatrix}.$$

Stembridge uses this theorem to study shifted tableaux, plane partitions, and Schur's Q-functions. As we will see in section 2.3 it can be used to count and generate tilings with free boundary conditions. We first give an extension of Stembridge's result.

2.1.3 Fixing sources

For our applications we will need to count routings where some of the sources are fixed and some are free. So we need a hybrid form of the Gessel-Viennot and Stembridge methods. Here we provide a technical, but straightforward, generalization of theorem 2.4 that allows us to count the number of such routings. This extension relies on carefully choosing the entries of the matrix so that the Pfaffian of this matrix counts the routings of interest to us.

Informally, if s_i is used as a source then we replace terms $\pm 1 + q_{ij}$ in Q by q_{ij}, and if not used, by ± 1. The following theorem formalizes this idea:

Theorem 2.5 *Let $\mathcal{S} = (s_1, \ldots, s_l)$ be an l-tuple of vertices in an acyclic digraph D, with l odd. Let $\mathcal{T}$ be an ordered subset of vertices that is compatible with $\mathcal{S}$. Suppose $\mathcal{S}_{in}, \mathcal{S}_{out} \subseteq \mathcal{S}$ with $\mathcal{S}_{in} \cap \mathcal{S}_{out} = \emptyset$. Let Q be the skew–symmetric matrix where the upper triangular entries are given by*

$$[Q]_{ij} = \begin{cases} 0 & \text{if } s_i \in \mathcal{S}_{in} \text{ and } s_j \in \mathcal{S}_{out}, \text{ (or vice-versa)} \\ q_{ij} & \text{if } j \leq l \text{ and } (s_i \text{ and/or } s_j \text{ in } \mathcal{S}_{in}, \text{neither in } \mathcal{S}_{out}) \\ (-1)^{i+j-1} & \text{if } j \leq l \text{ and } (s_i \text{ and/or } s_j \text{ in } \mathcal{S}_{out}, \text{neither in } \mathcal{S}_{in}) \\ q_i & \text{if } j = l+1 \text{ and } s_i \in \mathcal{S}_{in} \\ (-1)^{i+l} & \text{if } j = l+1 \text{ and } s_i \in \mathcal{S}_{out} \\ (-1)^{i+j-1} + q^*_{ij} & \text{otherwise, where } q^*_{ij} \text{ is as in theorem 2.4} \end{cases}$$

for $1 \leq i < j \leq l+1$. Then $\Phi = \mathrm{Pf}(Q)$ is the number of free routings of D with $\mathcal{S}_{in}$ included in the set of used sources, and $\mathcal{S}_{out}$ in the set of unused sources.

Proof. For $J \subseteq \{1, 2, \ldots, n\}$, let A_J denote the square submatrix of A obtained by selecting the rows and columns indexed by J. We use the result (from [59, lemma 4.2]) that (for n even, A and B $n \times n$ in size) we can write

$$\mathrm{Pf}(A+B) = \sum_J (-1)^{\sigma(J)-\frac{|J|}{2}} \mathrm{Pf}[A_J]\mathrm{Pf}[B_{\overline{J}}] \tag{2.1}$$

where $\sigma(J) = \sum_{j \in J} j$, and the sum is taken over all partitions $J, \overline{J}$ of $\{1, \ldots, n\}$ with $|J|$ even.

Decompose Q into a sum of two matrices, A and B, where $[A]_{ij} \in \{0, 1, -1\}$ and $[B]_{ij} \in \{0, q_{ij}^*, -q_{ij}^*\}$. Now apply the above result for the Pfaffian of a sum. We have $\mathrm{Pf}[A_J] = 0$ if $J \cap \mathcal{S}_{in} \neq \emptyset$ since A will contain a row of zeros. Similarly, $\mathrm{Pf}[B_{\overline{J}}] = 0$ if $\overline{J} \cap \mathcal{S}_{out} \neq \emptyset$. So the only terms that survive in the sum (2.1) are those with $\mathcal{S}_{in} \subseteq \overline{J}$ and $\mathcal{S}_{out} \subseteq J$. Note that if q_{ij} or q_i appears in one of the terms of Φ, that term corresponds to a set of paths (not necessarily non–intersecting) that uses s_i. Thus, if $s_i \in \mathcal{S}_{in}$, by the choice of the entries of Q we ensure that s_i is used in every routing of D, as one of the q_{ij}'s or q_i will appear in each term of Φ. Similarly, if $s_i \in \mathcal{S}_{out}$ then none of the q_{ij}'s or q_i appears in Φ, so that s_i is unused as a source in every routing of D, as required. □

2.2 Generating random routings

We are now in a position to define an algorithm to uniformly generate a free routing of a planar acyclic digraph D with compatible sources and sinks. The general approach is now obvious and is similar to the determinant algorithm of [63] for generating fixed routings. The main idea is to generate the routings using self–reducibility. Paths are constructed single edges at a time, where first we decide whether a particular candidate source will appear in the routing. If so, we then select one of its outgoing edges, move the source forward to a new vertex, and repeat this procedure. Extensive

use of theorem 2.5 is made during this algorithm to ensure, for example, once a path is begun from a source in the set $\mathcal{S}$ it will continue and terminate at a sink in $\mathcal{T}$.

In our algorithm we will need to compute ratios of Pfaffians to determine the probability that a candidate source or edge is included in the routing. We will see that these two Pfaffians are of very similar matrices. The following linear algebra result will help us significantly improve the time required to compute these ratios.

Theorem 2.6 *Let A be an invertible, skew–symmetric matrix and let B be a skew–symmetric matrix which differs from A by only the ith row and column. Then*

$$\frac{\mathrm{Pf}(B)}{\mathrm{Pf}(A)} = [BA^{-1}]_{ii}.$$

Proof. The proof relies on a closely related fact that if A is an invertible matrix and C differs from A by only the ith row, then

$$\frac{\det(C)}{\det(A)} = [CA^{-1}]_{ii}.$$

This follows from Cramer's rule. Now given A, an invertible, skew–symmetric matrix and B, a skew–symmetric matrix which differs from A by only the ith row and column, let C be the matrix formed by replacing the ith row of A with the ith row of B (so C differs from A by the ith row and from B by the ith column). Assume first that C is invertible. Then we have that $\det(C)/\det(A) = [CA^{-1}]_{ii}$ and $\det(B)/\det(C) = \det(B^T)/\det(C^T) = [B^T C^{-1\ T}]_{ii} = [C^{-1}B]_{ii}$. Finally, we use the fact that since A and C differ in only the ith row and $A_{ii} = C_{ii} = 0$ since A and B are skew–symmetric, then the ith rows of A^{-1} and C^{-1} must agree. Hence,

$$\begin{aligned}\left(\frac{\mathrm{Pf}(B)}{\mathrm{Pf}(A)}\right)^2 &= \frac{\det(B)}{\det(A)} = \frac{\det(B)}{\det(C)} \cdot \frac{\det(C)}{\det(A)} = [C^{-1}B]_{ii}[CA^{-1}]_{ii} \\ &= [A^{-1}B]_{ii}[BA^{-1}]_{ii} = \left([BA^{-1}]_{ii}\right)^2. \end{aligned} \tag{2.2}$$

If C is not invertible, let B' be obtained from B by perturbing the ith row of B by $\varepsilon \cdot$(random vector) and the ith column so that B' is skew-symmetric (and differs from

A only in the ith row and ith column). Letting C' be the matrix formed by replacing the ith row of A by the ith row of B', we proceed as before (as C' is invertible), then let $\varepsilon \to 0$ (so $B' \to B$) to get the same result as in (2.2).

Taking square roots, and recalling that $A^{-1} = \operatorname{adj}(A)/\det(A)$, where $\operatorname{adj}(A)$ is the (classical) adjoint of A, we can write (2.2) as

$$\operatorname{Pf}(B)\det(A) = \pm \operatorname{Pf}(A)[B\operatorname{adj}(A)]_{ii}. \tag{2.3}$$

Choosing an invertible skew–symmetric matrix A and letting $B = A$, we see in this case the choice of sign in (2.3) is $+$. By continuity, the sign for a whole neighborhood of the parameter values is also $+$. Taking partial derivatives and evaluating at 0, we see that the coefficients of the polynomials must be equal, so that the sign in (2.3) is everywhere $+$. This means the correct form of (2.3) is actually

$$\operatorname{Pf}(B)\det(A) = \operatorname{Pf}(A)[B\operatorname{adj}(A)]_{ii}. \tag{2.4}$$

Rewriting (2.4), we have

$$\frac{\operatorname{Pf}(B)}{\operatorname{Pf}(A)} = [BA^{-1}]_{ii}$$

as we desired (again using $A^{-1} = \operatorname{adj}(A)/\det(A)$). □

For our algorithm, the input is a planar digraph D with compatible sources $\mathcal{S}$ and sinks $\mathcal{T}$, where $|\mathcal{S}| = |\mathcal{T}| = l$, with l odd. Note that D need not be planar, however the planarity of D gives the running time $O(l^2 n)$ specified below in theorem 2.7, using the fact that $|E(D)| = O(|V(D)|)$.

We use the Sherman-Morrison formula for updating A^{-1} after changing a single row or column of the $l \times l$ matrix A. In our case we will be changing both a row and a column, but we can update the inverse by applying the Sherman-Morrison formula twice. Updating A^{-1} can be done in $\Theta(l^2)$ time using this method. We give a description that is useful for our purposes. For a more general result see [7]. For

the procedure outlined below, we assume that B differs from the matrix A only in row i.

Updating the matrix inverse after changing a single row

1. Set $C = A^{-1}$. (C will eventually equal B^{-1}.)
2. Let $w =$ (row i of A) $-$ (row i of B).
3. For $k = 1$ to l do

   Set $z_k = \sum_{j=1}^{l} w_j A^{-1}_{jk}$. ($z_k$ is the dot product of w with column k of A^{-1}.)
4. Set $\alpha = \frac{1}{1-z_i}$ and $z = (z_1, z_2, \ldots, z_l)$. (Note that if $z_i = 1$, B is not invertible.)
5. For $k = 1$ to l do

   Add $\alpha\, A^{-1}_{ki}\, z$ to row k of C.
6. C now equals B^{-1}. (This assumes α is defined in step 4, so that B^{-1} exists.)

Remark. The Sherman-Morrison formula for updating an inverse has shown some numerical instability in practice; we may achieve greater numerical stability by using other schemes for updating A^{-1} at a small cost in the running time of the FreeRoute algorithm given below. □

Now we are ready to describe the algorithm. The input to the algorithm is the planar digraph, D, having n vertices and m edges, sources $\mathcal{S}$ and sinks $\mathcal{T}$. The variable x_i records the current position of source i, and the array R records the routing as it is constructed. We maintain a matrix Q, initially equal to the matrix Q of theorem 2.4, and a matrix U, initially the inverse of Q, which we use to compute probabilities of using sources or edges in the routing. Q and U are updated as we move through the digraph. We compute $P[v, i]$, which will be the number of paths from vertex v to sink t_i, and $\widehat{P}[v, i]$, the number of paths from v to any of the sinks $t_i, t_{i+1}, \ldots, t_l$. (We use

the $\widehat{P}[v,i]$'s to help initialize the matrix Q in time $O(l^3)$ instead of $O(l^4)$, and later for updating entries of Q as we move through D.) With "$v \to w$" denoting that there is a directed edge from v to w the algorithm is:

The FreeRoute algorithm

1. Do a topological sort on D, numbering the vertices 1 through n, so that $v \to w$ implies $v < w$. ($O(n)$ time.)

2. For $v = 1$ to n, set $q_v = 0$. ($O(n)$ time.)

3. For $i = l$ down to 1 (Dynamic programming step) ($O(ln)$ time.)

   (a) Set $x_i = s_i$.

   (b) For $v = n$ down to 1

      i. If $v = t_i$, set $P[v,i] = 1$, else set $P[v,i] = \sum_{w:w\to v} P[w,i]$.
      ($P[v,i]$ now contains the number of paths from v to t_i.)

      ii. Set $q_v = q_v + P[v,i]$.

      iii. If $i = l$, set $\widehat{P}[v,i] = P[v,i]$, else set $\widehat{P}[v,i] = \widehat{P}[v,i+1] + P[v,i]$.

4. For $i = 1$ to l ($O(l^3)$ time.)

   (a) Find v such that $v = s_i$.

   (b) Set $q_i = q_v$. (Initialize q_i's.)

   (c) For $j = i + 1$ to l

      i. Find w such that $w = s_j$.

      ii. Set $q_{ij} = \sum_{k=1}^{l-1} \det \begin{pmatrix} P[v,k] & \widehat{P}[v,k+1] \\ P[w,k] & \widehat{P}[w,k+1] \end{pmatrix}$. (Initialize q_{ij}'s.)

5. Initialize the matrix Q as in theorem 2.4 using the q_i's and q_{ij}'s, find $U = Q^{-1}$, and set $\mathcal{S}_{in} = \emptyset$ and $\mathcal{S}_{out} = \emptyset$. ($O(l^3)$ time.)

6. For $v = 1$ to n, if $v = x_i$ for some i then

   (a) If $v \in \mathcal{S} \backslash (\mathcal{S}_{in} \cup \mathcal{S}_{out})$ then decide if v is used as a source (see details below). If it is, add v to $\mathcal{S}_{in}$. If not, add v to $\mathcal{S}_{out}$ and set $R[v] = 0$.
   In either case, update row and column i of Q and U. ($O(n + l^3)$ time.)

   (b) If $v \in \mathcal{S}_{in}$ then decide which edge leaving v to include in the path of the routing (see details and remarks). Let w be the other endpoint of this edge.
   If $w = s_k$ for some k, see remark 1 below.
   Set $R[v] = w$, $x_i = w$, and add w to $\mathcal{S}_{in}$.
   Set $q_i = q_w$. Update the ith row and column of Q and update U. ($O(l^2 n)$ time.)

Remark. In step 6(b), we may try to push $v = x_i$ forward to an (as yet) unused source $w = s_k \in \mathcal{S}$. In this case, we want to add w to $\mathcal{S}_{out}$ so that it is not used during some later step to begin a different path. However, we also want to add w to $\mathcal{S}_{in}$ so that in later iterations of step 6(b) we push w forward to complete a full path into $\mathcal{T}$ that started from s_i. This conflicts with the condition of theorem 2.5 that $\mathcal{S}_{in} \cap \mathcal{S}_{out} = \emptyset$. We get around this difficulty as follows: Remove x_i from $\mathcal{S}_{in}$ and add it to $\mathcal{S}_{out}$, then add $w = s_k$ to S_{in} so that it is pushed forward in later steps of the algorithm. Update row and column i of Q to reflect that x_i is unused, then row and column k so that s_k is used, and update U accordingly with successive applications of the Sherman-Morrison formula. Finally, set $R[v] = w$ to join the path between s_i and w to the path from w into $\mathcal{T}$. We will see examples of digraphs in which this situation might arise in section 2.3 where we consider tilings of reduced Aztec diamonds. □

Remark. During step 6(b) it is possible that $x_i \in \mathcal{T}$ but we might still push x_i forward. This could occur if x_i has out–neighbors that are also in $\mathcal{T}$, or there is some path that eventually ends in $\mathcal{T}$. Informally, in this situation we may consider that x_i

is joined to a phantom sink by a single (phantom) edge. Pushing x_i forward to this phantom sink corresponds to terminating the path at x_i and not continuing to any of x_i's neighbors. In practice, we need not handle this situation as a special case, since we can examine all of the out–neighbors of x_i in turn and if we reject using any of them then terminate the path, i.e., x_i is not pushed. □

Details for step 6(a): In this step, we determine if the source s_i is used in a routing The probability that s_i is used is given by $\frac{\mathrm{Pf}(Q')}{\mathrm{Pf}(Q)}$, where Q' is a skew–symmetric matrix differing from Q in the ith row and ith column. In particular, the ith row of Q' can be found using theorem 2.5, where we apply the theorem with s_i used in the set of current potential sources (the x_j's, restricting $\mathcal{S}_{in}$ and $\mathcal{S}_{out}$ to that set). We use theorem 2.6 to compute this probability as the dot product of the new ith row of Q' with column i of U. If v is used, we replace the ith row of Q by the ith row of Q' to reflect this (and then update the ith column of Q so that it remains skew–symmetric), and add v to $\mathcal{S}_{in}$. If v is not used, we update row and column i of Q as appropriate in theorem 2.5, where v is now in $\mathcal{S}_{out}$. In either case, we update U (so that it is still equal to Q^{-1}), using two successive applications of the Sherman-Morrison formula, once for changing row i of Q, and again for changing column i. Updating U takes time $\Theta(l^2)$, and hence the total time spent in step 6(a) is $O(n + l^3)$.

Details for step 6(b): Moving the source x_i forward in step 6(b) changes the ith row and column of Q. As before, the probability that the edge $v \to w$ is used is $\frac{\mathrm{Pf}(Q')}{\mathrm{Pf}(Q)}$, where Q' is the matrix with w used as a source in place of x_i. If this edge is taken, we update Q (and U) by replacing the ith row and column of Q with those of Q'. In the special case that $w \in \mathcal{S}$, we proceed as outlined in remark 1. The time to update U (at any instance when Q is updated) is $\Theta(l^2)$, so the total time spent in step 6(b) is $O(l^2 n)$.

We have demonstrated the following theorem:

Theorem 2.7 *Let D be a planar acyclic digraph with n vertices, having compatible sources and sinks. The FreeRoute algorithm uniformly samples a free routing of D in time $O(l^2 n)$.*

2.3 Lattice paths and tilings

We demonstrate applications of the FreeRoute algorithm from the previous section, showing how to generate random domino tilings of the reduced Aztec diamond with free boundary conditions and lozenge tilings of the hexagon with free boundary conditions. The key idea is the existence of a bijection between the set of tilings of this region and the set of free routings in a related digraph. The bijections described below have previously appeared in [36] in the context of tilings with fixed boundary conditions. The proof of theorem 2.8 has been adapted from there.

2.3.1 Domino tilings of the reduced Aztec diamond

The *reduced Aztec diamond of order n*, denoted Γ_n, is a region composed of $2n^2$ unit squares arranged as $2n$ centered rows of squares, where the kth row has $\min\{2k-1, 4n-2k+1\}$ squares in it. A domino tiling is a cover of Γ_n using non–overlapping dominoes, where a domino covers two adjacent squares. A domino tiling with *free boundary conditions* is a tiling in which all the squares of Γ_n are covered, but the dominoes are allowed to "stick out" of (or overlap) the boundary of the region. We assume that we know the orientation of a domino that overlaps the boundary, i.e., a single square (or half-domino) is designated as the bottom, top, left or right half of a domino.

Given a tiling of Γ_n with free boundary conditions (or simply, a free tiling), we define a routing of a related digraph, D_n. Figure 2.3 gives an example of the digraph that arises from a reduced Aztec diamond. To get D_n, first color the left square

of row n of Γ_n black, then extend the coloring to Γ_n using alternating black and white squares (as on the underlying infinite chessboard). Mark the midpoint of each vertical edge that has a black square to its right. These marked points will comprise most of the vertices of D_n. To define the rest of the vertex set, fix $(0,0)$ as the coordinates of the point on the left edge of row n. Add $n+1$ additional points at coordinates $(-1,-1),(0,-2),(1,-3),\ldots,(n-1,-n-1)$, and another $n+1$ points at $(n,n),(n+1,n-1),(n+2,n-2),\ldots,(2n,0)$. To define the edges of D_n, join a point with coordinates (x,y) to the points $(x+1,y+1),(x+1,y-1)$, and $(x+2,y)$. Finally, delete edges that lie completely outside the boundary of Γ_n.

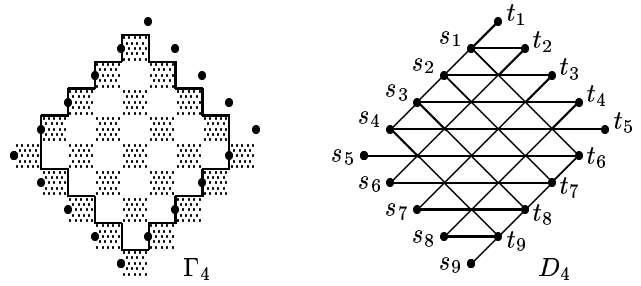


Figure 2.3: The reduced Aztec diamond and the corresponding digraph

The edges of D_n are those that remain between points after the deletion step. Direct edges from left to right. Starting at the source in the top square, label the sources $s_1, s_2, \ldots, s_{2n+1}$ in the counterclockwise direction, and then label each sink t_i where s_i is the last unmatched source. The left picture of figure 2.3 is Γ_4, the reduced Aztec diamond of order 4, along with the sources and sinks of D_4. The right picture is the digraph D_4. It should be noted that the sources and sinks are compatible, taking the natural orderings $s_1 < s_2 < \cdots < s_9$ and $t_1 < t_2 < \cdots < t_9$.

Theorem 2.8 *There is a bijection between free boundary tilings of* Γ_n *and free routings of* D_n.

Proof. Given a free tiling of Γ_n we map it to a free routing of D_n. This procedure

of constructing the routing corresponds to a tiling of the Aztec diamond using a set

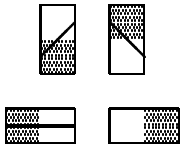
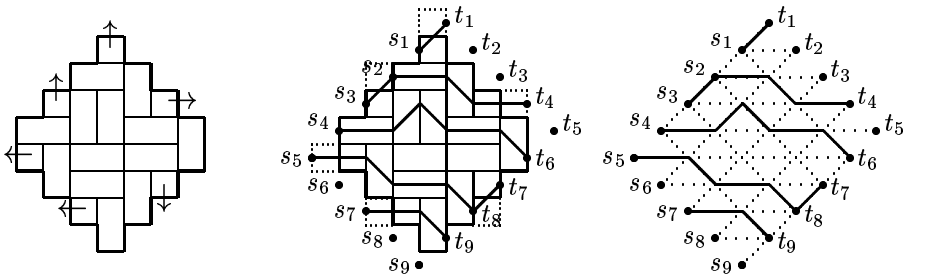


Figure 2.4: Paths through dominoes

of four types of tiles, three of which are marked, as shown in figure 2.4. The shading shows the relative positions of the black and white squares in the underlying region that are covered by the tile.

Examine the sources in this order: $s_n, s_{n-1}, \ldots, s_1, s_{n+1}, s_{n+2}, \ldots, s_{2n+1}$. It is possible that no source lies on the edge of a domino, in which case the routing is empty. Otherwise, the routing consists of the paths constructed as follows: If s_n lies on the edge of a domino, this determines the first edge in a path starting at s_n (otherwise move onto s_{n-1}). Connect s_n to the unique vertex in D_n that lies on the right side of the domino. This new vertex must lie on the left side of another domino, so repeat this process. Stop when we reach a vertex in $\mathcal{T}$ that does not have a domino to its right. Choose the next source, in the prescribed order, that is not on a path already constructed, and repeat this procedure. The paths are non–intersecting since dominoes cannot overlap and because of the order in which the sources were examined. See figure 2.5 for an example of a free boundary tiling of Γ_4 and the corresponding routing. (An arrow in the tiling points to the location of the other half of a domino that overlaps the boundary.)

To demonstrate the above map is a bijection, we construct the inverse map from routings to tilings. Each path starts at a sink s_i and follows edges in D_n to some sink t_j. As we follow the path, we tile the black squares on the paths in the underlying lattice. Each such edge in the path corresponds to using one of the marked dominoes in the tiling. Since the paths are non–intersecting the dominoes don't overlap, and we



A free domino tiling | Marked tiles | The corresponding routing

Figure 2.5: A domino tiling with free boundary conditions and its free routing

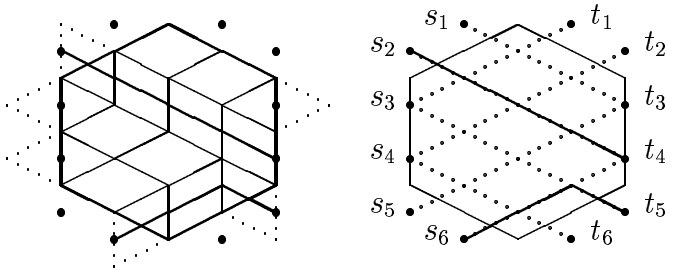
have a partial (free boundary) tiling of Γ_n. The remaining portions may be tiled in a unique manner using only horizontal unmarked dominoes. To see this, consider an untiled black square. The white square to the left of it must also be untiled, otherwise there would be a path leaving its right boundary so the black square would be tiled. We cover this pair of squares with a horizontal domino. Tiling each sub–region in this manner completes the tiling. Uniqueness follows since the leftmost square of each untiled region in the interior of the Aztec diamond must be white, so these portions cannot be completed using any vertical dominoes. In addition, an untiled black square appearing on the leftmost boundary of Γ_n has an untiled white square to its immediate left (that lies outside the boundary of the Aztec diamond). This pair of squares must also be covered by a horizontal unmarked domino. □

From this connection between tilings of Γ_n and routings of D_n, we can generate free boundary tilings of Γ_n by using the FreeRoute algorithm given in section 2.2 for generating free routings of D_n.

2.3.2 Lozenge tilings of the hexagon

We use a similar approach as in the previous section to generate lozenge tilings of a hexagonal region of the triangular lattice with free boundary conditions. There is a bijection between the collection of free boundary tilings, and the set of free routings

of a related digraph.



A free lozenge tiling The corresponding routing

Figure 2.6: A lozenge tiling with free boundary conditions and its free routing

Let H_n denote a hexagonal region on the triangular lattice with n edges on each side. A *lozenge tiling* of H_n is a covering of the region with lozenges, where a lozenge covers two adjacent triangles, and lozenges do not overlap. As in the previous section, a lozenge tiling of H_n with *free boundary conditions* is a tiling in which lozenges may overlap the boundary of the region. We describe a digraph, G_n, associated with H_n, in which free routings correspond to free boundary tilings of H_n. First, augment H_n to get a region $\hat{H}_n$ by adding the triangles in the underlying lattice that share an edge with the boundary of H_n. Mark the midpoint of each vertical edge in $\hat{H}_n$. These marked points form the vertex set of G_n. Join two points if they lie on adjacent triangles. These are the edges of G_n. Direct these edges from left to right. Labeling the sources and sinks from top to bottom in the graph, and ordering them in the natural manner shows they are compatible. See figure 2.6 for an example of this labeling. A free boundary lozenge tiling of H_n corresponds to a free routing of G_n. Similar to domino tilings, we may consider that a tiling of the region is done using a set of marked lozenges where the markings on the tiles line up to give the free routing in the digraph G_n that corresponds with the region. The proof of this bijection follows analogously to the proof given in [36] for establishing a bijection between fixed lozenge tilings and fixed routings. Figure 2.6 provides a pictorial illustration of this correspondence.

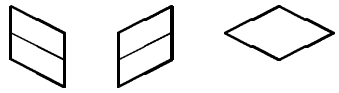


Figure 2.7: Paths through lozenges

Applying the FreeRoute algorithm of section 2.2 allows us to uniformly generate free routings of G_n, which we may then map to their corresponding free boundary tilings of H_n.

Chapter 3

Mixing Machinery

In this chapter, we shift our focus to the Markov Chain Monte Carlo (MCMC) methods which will be the essential tool for the remainder of this thesis. This method lends itself to a wider range of applications than the methods of the previous chapter. We sacrifice achieving perfect sampling, satisfying ourselves with approximate sampling, in return for a much richer theory and broader range of applications.

We start by reviewing some of the fundamentals underlying the theory of Markov chains. Recall that we may think of a Markov chain as a random walk on a graph or directed graph G. In this monograph we only consider finite graphs. At each unit of time in this random walk we are at some vertex v. We select one of the neighbors of v according to some fixed probability distribution and move to that neighbor, or we can choose to remain at v until the next time step. We repeat this step many times. The set of vertices of G is called the *state space* of the Markov chain, typically denoted Ω. Mathematically, we describe the random walk by labeling the vertices using the integers $\{1, \ldots, N\}$ where $N = |\Omega|$. Let p_{ij} denote the probability that, given we are at vertex i, we select vertex j to move to during the current step. These probabilities define the *transition matrix* P of the chain. The transition matrix satisfies two properties:

1. $p_{ij} \geq 0$ for all $i, j \in \Omega$;

2. P is *row stochastic*, i.e. for each $i \in \Omega$, $\sum_{j \in \Omega} p_{ij} = 1$.

Let the row vector π_0 denote the probability distribution on Ω that describes where we begin the random walk. It is well–known that the distribution on Ω after t steps of the chain is given by the product $\pi_0 P^t$.

A typical example of a Markov chain arises when we consider shuffling a deck of cards. Suppose we shuffle the (standard 52 card) deck in the following manner:

A card shuffling procedure:

1. Select an integer $i \in \{1, 2, \ldots, 52\}$ uniformly at random.

2. Remove the top card and reinsert it so that it is now the ith card in the deck (counting, say, from the top).

3. Repeat this procedure (steps 1 and 2) a large number of times.

A natural question is to determine what is meant by "large" in step 3. In other words, if, after we finish shuffling, we want to have an equal chance of obtaining one of the $52! \approx 8 \cdot 10^{67}$ possible orderings of the cards, how many times must we iterate steps 1 and 2? A standard argument (see, e.g. [2]) shows that we must repeat the "top in at random" steps about $52 \ln 52 \approx 206$ times to do this. This means that even though the state space is exponentially sized ($52! \approx 52^{40}$), we can select an element (nearly) at random in only a polynomially bounded number of steps ($52 \ln 52$).

We should note that the transition matrix for the "top in at random" Markov chain is $52! \times 52!$ in size (recall that $52! \approx 8 \cdot 10^{67}$). Therefore we see that even *writing down* the transition matrix is not practical. We need methods to examine Markov chains defined on such large state spaces. Before we formalize the mathematical framework, we consider another example taken from [55].

Imagine 1000 lily pads arranged in a circle. A frog starts on pad 1. Once a minute she either jumps one pad to her immediate left, one to her immediate right, or stays where she is, each with equal 1/3 probability. It is clear that after a "long time" the frog will be at any of the lily pads with approximately equal probability, and as the

amount of time increases to infinity the probability will converge to 1/1000. How long does it take until this probability is "close to" 1/1000? The result from [55] is that we will have to wait more than 120,000 minutes for the frog to have an approximately equally like chance to be at any of the lily pads. We will see that the eigenvalues of the matrix determine the convergence rate of the Markov chain. At the lowest level, finding the eigenvalues for this matrix amounts to determining the roots of a polynomial of degree 1000, not an attractive task. Even worse would be to determine the roots of the characteristic polynomial of the transition matrix for the card shuffling chain (having degree $\approx 8 \cdot 10^{67}$). In the general case, there is often a parameter n that measures the size of an instance of the problem. Often the state space Ω has size that is exponential in n. Therefore, we seek other ways to analyze convergence rates and similar questions without resorting to computing the eigenvalues of the large transition matrices for these chains.

To lay the groundwork for things to come, we formalize the definition of total variation distance, which measures how close we are to the limiting distribution. We also state the definition of mixing time, and relevant results used later in our analysis of Markov chains.

3.1 Definitions

In what follows, we assume that $\mathcal{M}$ is a Markov chain with finite state space Ω and transition matrix P. For $x, y \in \Omega$ and $t \geq 1$, $P^t(x, y)$ denotes the t-step probability of going from x to y.

A Markov chain is *ergodic* if there exists a probability distribution π on Ω such that

$$\lim_{t \to \infty} P^t(x, y) = \pi(y), \quad \forall x, y \in \Omega.$$

In this case, π is the unique vector satisfying

$$\pi P = \pi, \quad \pi(x) > 0 \ \ \forall x \in \Omega, \quad \text{and} \ \sum_{x \in \Omega} \pi(x) = 1,$$

so π is a left eigenvector (with eigenvalue 1) of the transition matrix P. The vector π is called the *stationary distribution* or *limiting distribution* of the chain. For both of the examples above the stationary distribution is uniform on Ω, i.e. $\pi(x) = \frac{1}{|\Omega|}, \ \forall x \in \Omega$, but this need not always be the case. We will see limiting distributions that are not uniform in chapters 4 and 6.

The statement $\pi P = \pi$ is equivalent to the collection of equations

$$\sum_{x \in \Omega} \pi(x) P(x, y) = \pi(y), \quad \text{for each } y \in \Omega. \tag{3.1}$$

These equations are commonly called the *global balance* equations. We will consider *reversible* Markov chains that satisfy the stronger *detailed balance* equations, namely

$$\pi(x) P(x, y) = \pi(y) P(y, x), \ \forall x, y \in \Omega. \tag{3.2}$$

In the two examples above, the "top in at random" card shuffling chain is not reversible since we always move the card on top of the deck, while the travels of the frog do describe a reversible chain.

Necessary and sufficient conditions for $\mathcal{M}$ to be ergodic are that $\mathcal{M}$ is *irreducible* and *aperiodic*. Irreducibility means that for every pair $x, y \in \Omega$ there is a sequence of states $x = z_0, z_1, \ldots, z_{k-1}, z_k = y$ such that $P(z_{i-1}, z_i) > 0$ for $i = 1, \ldots, k$. As we have already pointed out, it is often convenient to represent a Markov chain as a random walk on a graph or directed graph, where Ω is the set of vertices of the graph, and there is an edge from x to y if $P(x, y) > 0$. This graph is the *Markov kernel* of the chain. Irreducibility then means that the Markov kernel is strongly connected, so there is a path (or directed path in the case of digraphs) from x to y for every pair of vertices (states) x and y.

A state $x \in \Omega$ is *periodic* with period $\kappa \geq 1$ if $\kappa = \gcd\{t : P^t(x,x) > 0\}$, i.e. starting from state x, if x may be revisited in t steps, then κ divides t and is the largest such integer satisfying this property. The Markov chain $\mathcal{M}$ is said to be *aperiodic* if $\gcd\{\kappa(x) : x \in \Omega\} = 1$ where $\kappa(x)$ is the period of state x.

Suppose we have a finite set Ω and some probability distribution π on that set, with $\pi(x) > 0 \ \ \forall x \in \Omega$. One method of defining a Markov chain whose limiting distribution is π is to use the *Metropolis algorithm*. Recall this algorithm uses *acceptance ratios* to define the transition probabilities. What is often done is to first define the *Markov kernel*, the graph on which the random walk is performed. The vertices of the Markov kernel are the elements of Ω. The edges are typically specified by describing small local changes in the elements of Ω that allow us to move around the state space, or otherwise defining the edges of this graph. For example, in the Markov kernel for the "top in at random" chain there is a directed edge from a deck configuration x to another configuration y if y is obtained by moving the top card of x to obtain y. The Markov kernel of the "frog chain" is a cycle with 1000 vertices and loops at each vertex. In the Ising model from the introduction, we might change a single $+1$ to -1 or vice–versa to move between possible states, and in the case of domino tilings the rotations of pairs as in figure 1.2 define the Markov kernel. Provided the Markov kernel is strongly connected, the chain is irreducible. Let Δ denote the largest degree of a vertex in the Markov kernel and $x \sim y$ means that there is an edge between x and y in the Markov kernel. One common variant of the Metropolis algorithm has transition probabilities:

$$P(x,y) = \begin{cases} \frac{1}{2\Delta}\min\{1, \frac{\pi(y)}{\pi(x)}\} & \text{for } x \sim y, x \neq y \text{ in the Markov kernel} \\ 1 - \sum_{y \neq x} P(x,y) & \text{if } x = y. \end{cases}$$

These transitions ensure the chain is aperiodic as each vertex has a self–loop with probability at least $1/2$, i.e. $P(x,x) \geq \frac{1}{2}, \ \forall x \in \Omega$. It is easy to check that these

transition probabilities satisfy the detailed balance equations (3.2) so that π is the stationary distribution of this Markov chain.

The time a Markov chain takes to converge to its stationary distribution, the *mixing time* of the chain, is measured in terms of the distance between the distribution at time t and the stationary distribution π. Recall $P^t(x, y)$ denotes the t-step probability of going from x to y.

Definition. The *total variation distance* at time t from the distribution π is

$$\|P^t, \pi\|_{tv} = \max_{x \in \Omega} \frac{1}{2} \sum_{y \in \Omega} |P^t(x, y) - \pi(y)|.$$

The mixing time tells how many iterations of the chain must be performed until the total variation distance is small.

Definition. For $\varepsilon > 0$, the *mixing time* $\tau(\varepsilon)$ of the Markov chain $\mathcal{M}$ is

$$\tau(\varepsilon) = \min\{t : \|P^{t'}, \pi\|_{tv} \leq \varepsilon, \quad \forall t' \geq t\}.$$

We say a Markov chain is *rapidly mixing* if the mixing time is bounded above by a polynomial in n and $\log \frac{1}{\varepsilon}$, where n is the size of each configuration in the state space. One of the main goals of Markov Chain Monte Carlo is to design Markov chains that are provably rapidly mixing. Such chains can then be used to generate samples from Ω in a small (polynomial) number of iterations of the chain and the probability of generating the state $x \in \Omega$ is provably close to $\pi(x)$.

It is well known that the mixing rate is related to the *spectral gap* of the transition matrix. For the transition matrix P, we let $Gap(P) = \lambda_0 - |\lambda_1|$ denote its spectral gap, where $\lambda_0, \lambda_1, \ldots, \lambda_{|\Omega|-1}$ are the eigenvalues of P and $1 = \lambda_0 > |\lambda_1| \geq |\lambda_i|$ for all $i \geq 2$. The following result relates the spectral gap with the mixing time of the chain (see, e.g., [57]):

Theorem 3.1 *Let $\pi_* = \min_{x \in \Omega} \pi(x)$. For all $\varepsilon > 0$ we have*

(a) $\tau(\varepsilon) \leq \frac{1}{Gap(P)} \log(\frac{1}{\pi_* \varepsilon})$

(b) $\tau(\varepsilon) \geq \frac{|\lambda_1|}{2Gap(P)} \log(\frac{1}{2\varepsilon})$.

We give a brief review of some of the techniques that are used to bound the mixing time (or spectral gap) of a Markov chain. Our first application of these methods is to exclusion processes in chapter 4. Our new decomposition method follows in chapter 5.

3.2 Path coupling

Couplings have proven to be a useful source for bounding the mixing times of some Markov chains. For example, Jerrum [22] uses a coupling to examine a Markov chain for graph colorings, and Luby and Vigoda [37] use couplings to approximately count independent sets in low degree graphs. Luby, Randall, and Sinclair use this method to prove rapid mixing of Markov chains on the set of domino tilings of regions in $\mathbb{Z}^2$, eulerian orientations, and lozenge tilings of regions in the triangular lattice [36].

A *coupling* is a Markov chain on $\Omega \times \Omega$ with the property that each coordinate process, viewed in isolation, is just performing transitions of the original Markov chain $\mathcal{M}$. Also, if the two coordinates are equal at some iteration, then they will also agree after the next iteration of the coupled chain. Typically, instead of updating the coordinates independently, the coupling updates them so that the two processes will tend to coalesce, or "move together" under some measure of distance. The mixing time of $\mathcal{M}$ can be bounded by the expected time for configurations to coalesce under any valid coupling. We will consider a *complete coupling*. Informally, a complete coupling specifies the evolution of the entire state space during one step of the chain.

Path coupling lets us bound the mixing time by analyzing a subset of $\Omega \times \Omega$. The method of path coupling is described in the next theorem, adapted from [11].

Theorem 3.2 (Dyer and Greenhill [11]) *Let Φ be an integer valued metric defined on $\Omega \times \Omega$ which takes values in $\{0, \ldots, B\}$. Let U be a subset of $\Omega \times \Omega$ such that for all $(x_t, y_t) \in \Omega \times \Omega$ there exists a path $x_t = z_0, z_1, \ldots, z_r = y_t$ between x_t and y_t such that $(z_i, z_{i+1}) \in U$ for $0 \leq i < r$ and*

$$\sum_{i=0}^{r-1} \Phi(z_i, z_{i+1}) = \Phi(x_t, y_t).$$

Let $\mathcal{M}$ be a Markov chain on Ω with transition matrix P. Consider any random function $f : \Omega \to \Omega$ such that $Pr[f(x) = y] = P(x, y)$ for all $x, y \in \Omega$, and define a coupling of the Markov chain by $(x_t, y_t) \to (x_{t+1}, y_{t+1}) = (f_t(x_t), f_t(y_t))$, where $\{f_t\}_{t=0}^{\infty}$ are independent copies of f. Suppose that $\mathrm{E}(\Phi(x_{t+1}, y_{t+1})) \leq \eta\, \Phi(x_t, y_t)$ for all $(x_t, y_t) \in U$. Then the mixing time of $\mathcal{M}$ satisfies:

(a) If $\eta < 1$, then

$$\tau(\varepsilon) \leq \frac{\log(B\varepsilon^{-1})}{1 - \eta}.$$

(b) If $\eta = 1$ and there exists $\alpha > 0$ such that $Pr[\Phi(x_{t+1}, y_{t+1}) \neq \Phi(x_t, y_t)] \geq \alpha$ whenever $x_t \neq y_t$, then

$$\tau(\varepsilon) \leq \left\lceil \frac{eB^2}{\alpha} \right\rceil \lceil \log \varepsilon^{-1} \rceil.$$

Typically, the subset U in the theorem is taken as the set $\{(x, y) \in \Omega \times \Omega : \Phi(x, y) = 1\}$. In order to apply this theorem it is critical to use a coupling that has the contraction property where the distance between pairs x, y (with $(x, y) \in U$) tends to decrease, or, at least, not increase.

Remark. For all of the Markov chains discussed in this thesis, we ensure that there are self–loop probabilities of at least $\frac{1}{2}$, i.e. $P(x, x) \geq \frac{1}{2}$, $\forall x \in \Omega$. In particular, this allows us to guarantee the existence of $\alpha > 0$ in part (b) of theorem 3.2. One way to implement this in practice is to first define the coupling on all pairs $(x, y) \in U$. Then before each step of the coupling, flip a fair coin for each state in Ω where each coin

flip is independent of those at the other states. If the coin shows "heads" stay where you are; if it shows "tails" perform the transition as usual. □

3.3 The comparison method

In some cases, indirect methods of analyzing a Markov chain are necessary in combination with methods like coupling. Here we describe the technique of comparing Markov chains. In chapter 5 we describe the indirect method of decomposing the state space Ω to derive bounds on the mixing time of the chain.

The comparison method of Diaconis and Saloff-Coste [9] is handy for relating the mixing rates of two similar Markov chains. It is most useful in cases where it is easier to analyze a Markov chain if some auxiliary moves are added, but then bounds on the mixing time of the original Markov chain are desired.

Let $\widetilde{P}$ and P denote the transition matrices of two reversible Markov chains on the state space Ω with the same stationary distribution π. The comparison method allows us to relate the mixing times of these two chains. The idea is that the mixing time, $\tau_{\widetilde{P}}(\varepsilon)$, of $\widetilde{P}$ is known (or a suitable bound) and we desire to obtain a bound for the mixing time, $\tau_P(\varepsilon)$, of P.

Let $E(P) = \{(x,y) : P(x,y) > 0\}$ and $E(\widetilde{P}) = \{(x,y) : \widetilde{P}(x,y) > 0\}$ denote the sets of edges of the two chains, viewed as directed graphs. For each x, y with $\widetilde{P}(x,y) > 0$, define a *path* γ_{xy} using a sequence of states $x = x_0, x_1, \ldots, x_k = y$ with $P(x_i, x_{i+1}) > 0$, and let $|\gamma_{xy}|$ denote the length of the path. Let $\Gamma(z,w) = \{(x,y) \in E(\widetilde{P}) : (z,w) \in \gamma_{xy}\}$ be the set of paths that use the transition (z,w) of P. Finally, define

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

The following result is a combination of theorem 2.1 in [9] and theorem 4 in [51]:

Theorem 3.3 *With the above notation, we have*

(a) $Gap(P) \geq \frac{1}{A} Gap(\widetilde{P})$

(b) For $0 < \varepsilon < 1$, *we have that*

$$\tau_P(\varepsilon) \leq \frac{4\log(1/(\varepsilon\pi_*))}{\log(1/(2\varepsilon))} A\tau_{\widetilde{P}}(\varepsilon)$$

where $\pi_* = \min_{x\in\Omega} \pi(x)$.

Additionally, we will also use another comparison theorem of a different flavor. If the transition probabilities of two chains are related by simple inequalities, then so are their spectral gaps. This lemma follows immediately from the "functional definition" of the spectral gap (see [39, Eq. (7)]):

Lemma 3.4 *Suppose* P *and* $\widetilde{P}$ *are Markov chains on the same state space, each reversible with respect to the distribution* π. *Suppose there are constants* c_1 *and* c_2 *such that*

$$c_1 P(x,y) \leq \widetilde{P}(x,y) \leq c_2 P(x,y)$$

for all $x \neq y$. *Then*

$$c_1 Gap(P) \leq Gap(\widetilde{P}) \leq c_2 Gap(P).$$

In the next chapter we apply some of these results to a simple Markov chain involving an exclusion process. In addition to illustrating these techniques, this will be one of the key ingredients in chapter 6.

Chapter 4

Exclusion Processes

An *exclusion process* is a stochastic process whose configurations are arrangements of particles in distinct vertices of a graph G. The particles move along edges between neighboring vertices in G, maintaining the condition that two or more particles cannot occupy the same vertex simultaneously.

Exclusion processes are one class of particle systems. Such systems have been used to model the spread of an infectious disease or a forest fire, how voters change their views by interacting with people they know, and they have provided an alternative way of analyzing or interpreting some problems from statistical mechanics like the Ising model. See Durrett [10] and Liggett [34] for more on these examples. Some of these systems are also closely related to percolation, or so–called "oriented percolation." We do not define percolation here, but again refer the interested reader to Durrett [10], the detailed book by Grimmett [19], and the references therein.

For some of these systems, the number of particles is allowed to vary, but in an exclusion processes, the number of particles is fixed. In our case, the transitions of the exclusion process are governed by a Markov chain. We will consider the exclusion process that will be most useful to us in chapter 6 when we examine the problem of sampling staircase walks in $\mathbb{Z}^2$. The graph G is the interval [n] in $\mathbb{Z}$, where nearest neighbors are joined by edges, i.e. i is joined to $i-1$ and $i+1$ (1 is joined only to 2 and n only to $n-1$). Each configuration of k particles has a weight associated with it, and these weights are normalized to give a probability distribution, ρ, on the set of

particle configurations. The Markov chain consists of choosing a particle and moving it one step to the left or right, provided the destination is unoccupied. The transition probabilities are chosen so that ρ is the stationary distribution of this Markov chain.

In general, exclusion processes should arise naturally when using the decomposition method of chapter 5. In this method, we need to examine a projection chain that captures the interaction among the component sets that partition the state space.

4.1 A weighted exclusion process

As mentioned above, we focus on the exclusion process in the case when G is the interval $[n]$ in $\mathbb{Z}$. Let Ω_k denote the set of all $\binom{n}{k}$ ways to arrange k particles in distinct vertices of G. To formalize the weighted probability distribution on Ω_k, let f_i denote some sequence of positive numbers for $i \geq 0$. The collection of k particles divides the interval $[n]$ into $k+1$ pieces. The weight of a configuration $\sigma \in \Omega_k$ is given by $wt(\sigma) = f_{x_1} f_{x_2} \cdots f_{x_{k+1}}$, where x_i is the number of unoccupied sites in the ith interval. Therefore, we have $0 \leq x_i \leq n-k$ for all i, and $\sum_i^{k+1} x_i = n-k$. Normalizing, we have a probability distribution ρ on Ω_k, with $\rho(\sigma) = \frac{wt(\sigma)}{Z_k}$ where $Z_k = \sum_{\sigma' \in \Omega_k} wt(\sigma')$. Note that taking $f_i = 1$ for all i gives us the uniform distribution on Ω_k, so our formulation includes the uniform distribution as a special case.

In our analysis that follows, it will be convenient to augment each configuration in Ω_k with sites at locations 0 and $n+1$, each containing a fixed particle which never moves. We refer to this additional pair as particle 0 and particle $k+1$, respectively. With this addition, the weight of the (augmented) configuration σ is $wt(\sigma) = f_{x_1} f_{x_2} \cdots f_{x_{k+1}}$ where $x_i \in \{0, \ldots, n-k\}$ is the number of unoccupied sites between particles i and $i+1$ for $i = 0, \ldots, k$.

The remainder of this chapter is spent showing how, for some general classes of sequences f_i, we can sample from Ω_k. The procedure is based on a simple Markov

chain.

4.2 Markov chains for exclusion processes

4.2.1 A particle process with nearest–neighbor dynamics

A natural Markov chain $\mathcal{M}_{Gl}$ on Ω_k, called *nearest–neighbor (or Glauber) dynamics*, consists of moving some particle one space to the left or right, provided the destination is unoccupied. We recall that the two particles we added at sites 0 and $n+1$ are fixed and cannot move. For $\sigma \in \Omega_k$, we use $\sigma(i)$ to denote the position of particle i. Define the distance between pairs of configurations σ_1, σ_2 as $\delta(\sigma_1, \sigma_2) = \sum_i |\sigma_1(i) - \sigma_2(i)|$, i.e. the sum of distances between corresponding particles. We let P_k denote the transition matrix of the nearest–neighbor Markov chain $\mathcal{M}_{Gl}$ on Ω_k, and use Metropolis transitions for this chain. Note that we use the phrase "select $a \in_u A$" to mean that we select an element a from the set A uniformly at random.

One step of the nearest–neighbor Markov chain $\mathcal{M}_{Gl}$:

1. Select a particle uniformly at random (excluding the two fixed particles), and a direction $d \in_u \{\ell, r\}$, where ℓ means "left" and r "right".

2. With probability $\frac{1}{2}\min\{1, \frac{\rho(\sigma_2)}{\rho(\sigma_1)}\}$ move the particle one unit in the direction d if the destination is unoccupied.

3. In all other cases, do nothing.

With the Markov chain defined above, the transition probabilities for the Glauber

chain $\mathcal{M}_{Gl}$ are given by

$$P_k(\sigma_1, \sigma_2) = \begin{cases} \frac{1}{4k} \min\{1, \frac{\rho(\sigma_2)}{\rho(\sigma_1)}\} & \text{if } \delta(\sigma_1, \sigma_2) = 1 \\ 0 \quad \text{if } \delta(\sigma_1, \sigma_2) > 1 \\ 1 - \sum_{\sigma' \neq \sigma_1} P(\sigma_1, \sigma') & \text{if } \sigma_1 = \sigma_2 \end{cases} \tag{4.1}$$

These transitions ensure that $\mathcal{M}_{Gl}$ is aperiodic since $P_k(\sigma, \sigma) \geq \frac{1}{2}$ for all $\sigma \in \Omega_k$.

Our approach to establishing a bound for the mixing time of $\mathcal{M}_{Gl}$ is indirect. First we define a new Markov chain that includes some non-local moves and show this chain is rapidly mixing. Using the comparison method of section 3.3 we then establish rapid mixing of the simpler Glauber chain $\mathcal{M}_{Gl}$. In subsequent sections we demonstrate this when the sequence f_i is log–concave, or log–convex and satisfies an additional condition.

4.2.2 A particle process with heat bath dynamics

A second natural Markov chain on Ω_k is *heat bath dynamics*. The transitions differ from nearest–neighbor dynamics in that particles may move more than one space at a time, but can never pass over another particle in doing so. Recall that $\sigma(j)$ denotes the location of particle j in the configuration σ. We let σ denote the current configuration and σ' the new configuration obtained after one step of the heat bath dynamics. The new Markov chain $\mathcal{M}_{HB}$ on Ω_K has transition matrix $\widetilde{P}_k$.

One step of the heat bath Markov chain $\mathcal{M}_{HB}$:

1. With probability $\frac{3}{4}$, set $\sigma' = \sigma$.

2. Otherwise,

   - Select a particle uniformly at random (excluding the two fixed particles). Suppose we select particle j. Let $m = \sigma(j+1) - \sigma(j-1) - 1$. Then m is the number of sites between particles $j-1$ and $j+1$.

- Remove particle j.
- Replace particle j in location $\sigma(j-1)+t$ to obtain σ' where t is selected according to the probability distribution

$$\mathcal{P}r(t=i) = \frac{f_{i-1}f_{m-i}}{Z} \qquad \forall i \in \{1,\ldots,m\}$$

and $Z = \sum_{i=1}^{m} f_{i-1}f_{m-i}$ is the normalizing constant.

Step 1 is introduced for technical reasons, and only slows down the chain by a factor of four. We see that in one step of $\mathcal{M}_{HB}$ we remove a particle and reinsert it between its neighbors according to the conditional distribution. That is, the transition probabilities are

$$\widetilde{P}_k(\sigma_1,\sigma_2) = \begin{cases} \frac{1}{4k}\frac{f_{x_{j-1}}f_{x_j}}{Z} & \text{if } \sigma_1,\sigma_2 \text{ differ solely at particle } j \\ 0 \quad \text{if } \sigma_1,\sigma_2 \text{ differ by two or more particles} \\ 1-\sum_{\sigma'\neq\sigma_1}\widetilde{P}_k(\sigma_1,\sigma') \quad \text{if } \sigma_1=\sigma_2 \end{cases} \tag{4.2}$$

where x_j is the number of empty sites between particles j and $j+1$ in σ_2 and Z is the normalizing constant defined above in step 2 of this chain. It is important to note again that the ordering of the particles in the interval does not change during the transitions. Any particle that moves must remain between its neighbors and cannot move over them.

Finally, note this Markov chain is reversible and the stationary probabilities are those defined earlier, namely $\rho(\sigma) = f_{x_1}f_{x_2}\cdots f_{x_{k+1}}/Z_k$, where $Z_k = \sum f_{y_1}f_{y_2}\cdots f_{y_{k+1}}$ is the normalizing constant and the sum is over all non–negative solutions to $y_1 + y_2 + \ldots + y_{k+1} = n-k$. Having defined the heat bath Markov chain, we prove it is rapidly mixing for some natural classes of sequences.

4.3 Analysis of the Markov chain

The general approach we use is to apply path coupling to show the heat bath Markov chain $\mathcal{M}_{HB}$ is rapidly mixing. Recall that we need to define a coupling that is contracting on the state space Ω_k in order to use theorem 3.2 for this method to work. We consider some types of sequences for which we are able to prove the heat bath chain is indeed rapidly mixing.

4.3.1 Log–concave sequences

A log–concave sequence is one satisfying $f_{i-1}f_{i+1} \leq f_i^2$ for all i. For these sequences, the configurations having the largest weight (stationary probability) are those where the particles are distributed as evenly as possible in the range of sites. Intuitively, this suggests that regardless of the initial arrangement of the particles, they will tend to spread out over time.

Recall we have already defined a distance metric on Ω_k, namely $\delta(\sigma_1, \sigma_2) = \sum_i |\sigma_1(i) - \sigma_2(i)|$. Note that $0 \leq \delta(\sigma_1, \sigma_2) \leq k(n-k)$ for any pair $\sigma_1, \sigma_2 \in \Omega_k$. We use path coupling to examine the heat bath Markov chain $\mathcal{M}_{HB}$. In the notation of theorem 3.2 we take $U \subseteq \Omega_k \times \Omega_k$ to be pairs (σ_1, σ_2) with $\delta(\sigma_1, \sigma_2) = 1$, so the pair differs by a single (nearest–neighbor) transition.

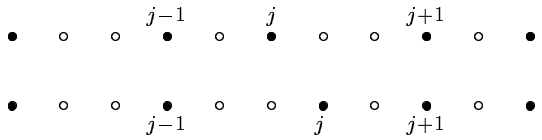


Figure 4.1: Typical situation for path coupling

Figure 4.1 shows a portion of two configurations with k particles, differing only at particle j. In our coupling, if particle j is chosen in the first step of the move, we can reinsert it at the same position in each configuration, decreasing the distance by one. Also, choosing any other particle except $j-1$ or $j+1$ allows us to reinsert it at the

same position in each configuration with identical probabilities, leaving the distance unchanged. If we choose particle $j-1$ or $j+1$ we need to define how to couple the chains to minimize the expected change in distance. For example, consider figure 4.2 which shows the configuration of figure 4.1 with particle $j-1$ removed. When we couple the chains together, it could be very bad to replace particle $j-1$ at the first position in σ_1 and at the fifth position in σ_2 since the new distance between the configurations is now four where it was previously one.

The dashed lines in figure 4.2 indicate how we would like to couple the moves. Replacing particle $j-1$ between its neighboring particles in the manner indicated either keeps the distance between the pair of configurations unchanged or increases

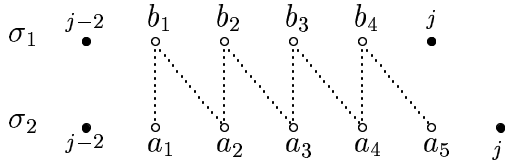


Figure 4.2: The coupling for the particle system

it by only one. We show that such a coupling is possible by demonstrating majorization inequalities for the transition probabilities a_i and b_i for the pair of particle configurations.

Referring to figure 4.2 as a representative case, let m denote the number of sites between particles $j-2$ and j in σ_1, so $m = \sigma_1(j) - \sigma_1(j-2) - 1$. For $i \in [m]$ let $b_i = f_{i-1}f_{m-i}/Z_b$ be the probability with which we insert particle $j-1$ at position i between $j-2$ and j in the upper configuration; for $i \in [m+1]$, let $a_i = f_{i-1}f_{m-i+1}/Z_a$ be the probability to insert particle $j-1$ in the lower configuration σ_2, where $Z_b = \sum_{i=1}^{m} f_{i-1}f_{m-i}$ is the normalizing constant for σ_1 and $Z_a = \sum_{i=1}^{m+1} f_{i-1}f_{m-i+1}$ is the normalizing constant for σ_2. We have the following surprising combinatorial lemmas:

Lemma 4.1 *Suppose f is a log–concave sequence. With a_i, b_i as above, these probabilities satisfy $a_i \leq b_i$ for $i \in [m]$.*

Proof. We have $a_i \leq b_i$ if and only if

$$\begin{aligned}
&\frac{f_{i-1}f_{m-i+1}}{Z_a} \leq \frac{f_{i-1}f_{m-i}}{Z_b} \\
&\quad \Leftrightarrow f_{m-i+1} \cdot Z_b \leq f_{m-i} \cdot Z_a \\
&\quad \Leftrightarrow f_{m-i+1} \cdot [f_0 f_{m-1} + f_1 f_{m-2} + \cdots + f_{m-1} f_0] \leq \\
&\qquad\qquad\qquad\qquad f_{m-i} \cdot [f_0 f_m + f_1 f_{m-1} + \cdots + f_m f_0].
\end{aligned}$$

This last inequality follows by comparing corresponding terms, e.g. $f_{m-i+1}f_{m-1} \leq f_{m-i}f_m$ by log–concavity of f. Hence $a_i \leq b_i$. □

Lemma 4.2 *Suppose f is a log–concave sequence. With a_i, b_i as above, for all $r \in [m]$, we have*

$$\sum_{i=1}^{r} a_i \leq \sum_{i=1}^{r} b_i \leq \sum_{i=1}^{r+1} a_i.$$

Proof. Half of these inequalities come directly from lemma 4.1, i.e. $a_1 \leq b_1$ and $a_2 \leq b_2$ imply $a_1 + a_2 \leq b_1 + b_2$. Similarly, $a_3 \leq b_3$ then implies that $a_1 + a_2 + a_3 \leq b_1 + b_2 + b_3$ and so forth. This establishes the inequalities

$$\sum_{i=1}^{r} a_i \leq \sum_{i=1}^{r} b_i \quad \forall r \in [m].$$

For the complementary set of inequalities, we first note from the symmetric definition of a_i and b_i that $a_i = a_{m-i+2}$ (for $i = 1, \ldots, m+1$) and $b_i = b_{m-i+1}$ (for $i = 1, \ldots, m$).

Start with the equality $b_1 + \cdots + b_m = a_1 + \cdots + a_{m+1}$. Using the symmetry of the transition probabilities, from lemma 4.1 we see $a_{m+1} = a_1 \leq b_1 = b_m$ or $-b_m \leq -a_{m+1}$. Adding this to the first equality, we find $b_1 + \cdots + b_{m-1} \leq a_1 + \cdots + a_m$. Since $a_m = a_2 \leq b_2 = b_{m-1}$ we see that $b_1 + \cdots + b_{m-2} \leq a_1 + \cdots + a_{m-1}$. Continuing in this manner we get the set of inequalities

$$\sum_{i=1}^{r} b_i \leq \sum_{i=1}^{r+1} a_i \quad \forall r \in [m]$$

completing the proof. □

Lemmas 4.1 and 4.2 allow us to couple moves in the "zig–zag" manner shown in figure 4.2. We describe the coupling for pairs σ_1, σ_2 with $\delta(\sigma_1, \sigma_2) = 1$. As before, we suppose that they differ at particle j. We use the notation (σ_1', σ_2') for the configuration obtained from the pair (σ_1, σ_2) after one step of the coupling. Let $\Delta\delta(\sigma_1, \sigma_2) = \delta(\sigma_1', \sigma_2') - \delta(\sigma_1, \sigma_2)$ denote the change in distance after one step of the coupled chain.

Coupling for the heat bath chain $\mathcal{M}_{HB}$:

1. Pick a particle, say particle i, uniformly at random.

2. If $i \notin \{j-1, j, j+1\}$ we remove particle i from σ_1 and σ_2 and reinsert it in the same location in both, with the exact same transition probabilities (according to equation (4.2)). In this case we have $\Delta\delta(\sigma_1, \sigma_2) = 0$.

3. If $i = j$ we can replace particle j in the same position in each configuration so that they now agree. Conditioned on $i = j$ we have $\Delta\delta(\sigma_1, \sigma_2) = -1$.

4. If $i \in \{j-1, j+1\}$ pick an ordered pair (u, v) according to the following probability distribution:

$$\mathcal{P}r\{(u,v) = (t,t)\} = \sum_{r=1}^{t} a_r - \sum_{r=1}^{t-1} b_r \quad \text{for } t = 1, \ldots, m$$
$$\mathcal{P}r\{(u,v) = (t,t+1)\} = \sum_{r=1}^{t} b_r - \sum_{r=1}^{t} a_r \quad \text{for } t = 1, \ldots, m$$

   Here we assume that σ_1 has the shorter interval between particles $i-1$ and $i+1$ and $m = \sigma_1(i+1) - \sigma_1(i-1) - 1$ denotes the number of sites between the neighbors of particle i, and a_i, b_i are defined as before. Having picked the ordered pair (u, v), we place particle i at position u in σ_1 and at position v in σ_2. In this case we have $\Delta\delta(\sigma_1, \sigma_2) \in \{0, +1\}$.

Lemma 4.3 *Suppose f is a log–concave sequence. Let $\sigma_1, \sigma_2 \in \Omega_k$ with $\delta(\sigma_1, \sigma_2) = 1$. After one step of the coupled Markov chain we have $E(\Delta\delta(\sigma_1, \sigma_2)) \leq 0$.*

Proof. We have noted that selecting a particle other than $j-1$, j, or $j+1$ does not change the distance, and that by choosing particle j the distance decreases by one. The "bad moves" that increase the distance are those in which we insert $j-1$ (or $j+1$) at different positions in σ_1 and σ_2. If we select either of these two particles, say $j-1$, the expected change in distance is

$$\begin{aligned} \beta &= \sum_{i=1}^{m} \left(\sum_{r=1}^{i} b_r - \sum_{r=1}^{i} a_r \right) \\ &= m\, b_1 + (m-1) b_2 + \cdots + b_m - m\, a_1 - (m-1) a_2 - \cdots - a_m. \end{aligned}$$

Recalling that $b_i = b_{m-i+1}$ and $a_i = a_{m-i+2}$, we can also write

$$\beta = m\, b_m + (m-1) b_{m-1} + \cdots + b_1 - m\, a_{m+1} - (m-1) a_m - \cdots - a_2.$$

Summing these equations and simplifying yields $\beta = \frac{1}{2}$. By symmetry this also represents the expected change in distance for particle $j+1$.

Putting these pieces together to determine the overall expected change, we find that $E(\Delta\delta(\sigma_1, \sigma_2)) \leq \frac{1}{2k}(-1 + \frac{1}{2} + \frac{1}{2}) = 0$. (We write this as an inequality, as it is possible, say, that $j = 1$ so there at most one bad move for the neighboring particle 2. Recall the fixed particle 0 cannot move.) □

All the preceding lemmas result in the following theorem:

Theorem 4.4 *The heat bath Markov chain $\widetilde{P}_k$ is rapidly mixing for any log–concave sequence f and:*

(a) $Gap(\widetilde{P}_k) \geq \frac{c}{n^5}$ for some constant c.

(b) $\tau_{\widetilde{P}_k}(\varepsilon) = \mathrm{O}(n^5 \log \varepsilon^{-1})$

Proof. Whenever $\delta(\sigma_1, \sigma_2) > 0$, the probability that the distance changes in one step is at least $1/2k$ (by choosing the particle where σ_1 and σ_2 disagree). The path coupling theorem gives a bound on the mixing time for the heat bath Markov chain $\widetilde{P}_k$ on Ω_k:

$$\tau_{\widetilde{P}_k}(\varepsilon) \quad \leq \quad \lceil \frac{e(k(n-k-1))^2}{\frac{1}{2k}} \log \varepsilon^{-1} \rceil = \mathrm{O}(n^5 \log \varepsilon^{-1}).$$

For part (a), first note that since we have ensured self–loop probabilities of at least $\frac{3}{4}$, this implies that $\lambda_1(\widetilde{P}_k) \geq \frac{1}{2}$. Using theorem 3.1(b) we can lower bound $Gap(\widetilde{P}_k)$ (using that $\lambda_1(\widetilde{P}_k) \geq \frac{1}{2}$):

$$Gap(\widetilde{P}_k) \geq \frac{1}{4\tau_{\widetilde{P}_k}(\varepsilon)} \log(\frac{1}{2\varepsilon}) = c \frac{\log(\frac{1}{2\varepsilon})}{n^5 \log(\frac{1}{\varepsilon})}$$

for some constant c. As this inequality must hold for every $0 < \varepsilon < 1$, letting $\varepsilon \to 0$ (so that $\frac{\log(\frac{1}{2\varepsilon})}{\log(\frac{1}{\varepsilon})} \to 1$) we have $Gap(\widetilde{P}_k) \geq \frac{c}{n^5}$. Part (b) follows from part (a) and theorem 3.1. □

Remark. Recall the original Markov chain $\mathcal{M}_{Gl}$ on Ω_k that has nearest–neighbor dynamics. The comparison theorem may be used in a straightforward manner to give a bound on the mixing time (or, equivalently, the spectral gap) of P_k. Because the parameter A of theorem 3.3 involves the transition probabilities of both the heat bath and nearest–nearest chains, we will apply that theorem once a particular sequence has been specified. □

4.3.2 Log–convex sequences

Another natural class of sequence to consider are log–convex sequences, those satisfying $f_i^2 \leq f_{i-1} f_{i+1}$ for all i. Unfortunately, the analysis we used in the previous section does not hold for an arbitrary log–convex sequence.

The difficulty arises in that log–convex sequences can grow "too fast" so the majorization lemmas 4.1 and 4.2 do not generalize to any log–convex sequence. To see this, consider a sequence of numbers beginning $1, 1, 1, 1, \alpha$, where $\alpha \gg 1$ which is a log–convex sequence. Suppose that when we remove particle j the interval between its neighbors in σ_2 has five sites in it as in figure 4.3 below. To use this same coupling, we require, for example, that $\frac{1\cdot\alpha}{1\cdot\alpha+1\cdot1+1\cdot1+1\cdot1+\alpha\cdot1} \leq \frac{1\cdot1}{1\cdot1+1\cdot1+1\cdot1+1\cdot1} = \frac{1}{4}$ (showing $a_1 \leq b_1$). Letting $\alpha \to \infty$, we see this clearly cannot hold since the fraction involving α approaches one. It is clear that no coupling can overcome this difficulty.

For our purposes in chapter 6, we need to consider an exclusion process whose weights are defined by the Catalan numbers $C(n) = \frac{1}{n+1}\binom{2n}{n}$. This sequence is log–convex. We shall see that, unlike the previous example, log–convex sequences that satisfy a suitable restriction do permit results analogous to lemmas 4.1 and 4.2. Hence, for sequences satisfying this restriction we also have analogues to lemma 4.3 and theorem 4.4, meaning the heat bath chain is rapidly mixing for such sequences.

One restriction for which the path coupling argument in the previous subsection will work in the log–convex case is one that rules out sequences that grow "too fast." As before, to use path coupling we need to consider two configurations σ_1 and σ_2 which differ only in the location of particle j and $\delta(\sigma_1, \sigma_2) = 1$. Referring

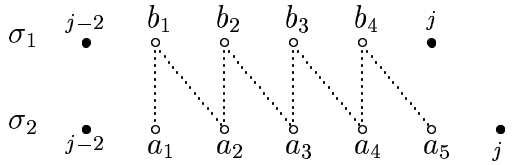


Figure 4.3: The coupling for the particle system

to figure 4.3 (a reproduction of figure 4.2) as a representative case, we again let $m = \sigma_1(j) - \sigma_1(j-2) - 1$. For $i \in [m]$ $b_i = f_{i-1}f_{m-i}/Z_b$ is the probability to insert at position i in σ_1, and for $i \in [m+1]$ the probability to insert at position i in σ_2 is $a_i = f_{i-1}f_{m-i+1}/Z_a$ where Z_b and Z_a are the respective normalizing constants. Let

us define $g_t = \sum_{i=1}^{t} f_{i-1} f_{t-i}$. With this definition, we have $Z_b = g_m$ and $Z_a = g_{m+1}$. We first show a result analogous to lemma 4.1.

Lemma 4.5 *Suppose f is a log–convex sequence and $\frac{f_t}{f_{t-1}} \leq \frac{g_{t+1}}{g_t}$ for all $t \geq 1$, where $g_t = \sum_{i=1}^{t} f_{i-1} f_{t-i}$. Then $a_i \leq b_i$ for all $i \in [m]$.*

Proof. The first inequality comes directly from the assumption relating f and g, i.e. $\frac{f_m}{f_{m-1}} \leq \frac{g_{m+1}}{g_m}$ implies that $a_1 = \frac{f_0 f_m}{g_{m+1}} \leq \frac{f_0 f_{m-1}}{g_m} = b_1$. Using log–convexity of f we have $\frac{f_{m-1}}{f_{m-2}} \leq \frac{f_m}{f_{m-1}} \leq \frac{g_{m+1}}{g_m}$ so by rearranging this inequality and multiplying by f_1 we get $a_2 = \frac{f_1 f_{m-1}}{g_{m+1}} \leq \frac{f_1 f_{m-2}}{g_m} = b_2$. Similarly, using log–convexity of f, we have $\frac{f_{m-2}}{f_{m-3}} \leq \frac{g_{m+1}}{g_m}$, so $a_3 = \frac{f_2 f_{m-2}}{g_{m+1}} \leq \frac{f_2 f_{m-3}}{g_m} = b_3$. Continuing in this manner, we obtain the inequalities $a_i \leq b_i$ for all $i \in [m]$. □

Using this lemma we can prove results analogous to lemmas 4.2 and 4.3 for log–convex sequences satisfying the hypotheses of lemma 4.5. We will not repeat those statements here. The proof of the following theorem is identical to that of theorem 4.4.

Theorem 4.6 *Suppose f is a log–convex sequence and $\frac{f_t}{f_{t-1}} \leq \frac{g_{t+1}}{g_t}$ for all $t \geq 1$. Then the heat bath Markov chain $\mathcal{M}_{HB}$ is rapidly mixing and:*

(a) $Gap(\widetilde{P}_k) \geq \frac{c}{n^5}$ for some constant c.

(b) $\tau_{\widetilde{P}_k}(\varepsilon) = \mathrm{O}(n^5 \log \varepsilon^{-1})$

Remark. As in the previous case, we may again use the comparison theorem to derive a bound on the mixing time of the simpler nearest–neighbor chain. We will discuss this further in section 6.3.7 once we have specified a particular log–convex sequence for our application. □

Chapter 5

Markov Chain Decomposition

We have already seen in the previous chapter that studying auxiliary chains can be a powerful tool in the analysis of Markov chains. In this chapter our goal is to present a method that allows us to analyze Markov chains by considering a collection of intermediate, often simpler, Markov chains. We apply this technique in chapter 6, but we feel the method is very general, and is one of the main contributions of this thesis.

In [39] Madras and Randall describe a *state factorization* or *state decomposition* method. To use this method we restrict the Markov chain to smaller pieces of the state space. Intuitively, if the *restricted* Markov chains on each piece are all rapidly mixing, and a related *projection chain* (defined below) is rapidly mixing, suggesting good flow among the smaller pieces, then we may conclude the original Markov chain on all of Ω is rapidly mixing. One salient feature of this method is the necessity to find a collection of subsets that cover Ω and have significant overlaps.

The main idea of our new theorem is similar to theirs, but eliminates this last condition. We examine several Markov chains, each restricted to a smaller subset of Ω, as well as a projection chain that captures flow among the subsets. We infer a bound on the spectral gap of the original chain from bounds on the spectral gaps of the restricted chains and the projection. The significant difference between their state decomposition theorem and the new one here is that we use a collection of sets that *partition* Ω, rather than requiring a family of overlapping sets. Often a

natural partition of Ω suggests itself based on some combinatorial feature of the state space elements. To apply the Madras-Randall theorem we are then left with the nettlesome task of contriving a family of overlapping sets from this partition. The new decomposition theorem sidesteps this requirement. This new result, theorem 5.3 and corollary 5.4, first appeared in [44].

One significant factor to bear in mind is that while decomposition allows us to show rapid mixing of some Markov chains for which path coupling or other methods do not suffice, this method can potentially add several orders of magnitude to the bounds on the mixing time. This is because we are indirectly analyzing the Markov chain through the use of the projection chain and the restrictions on the smaller subsets of the state space, possibly even using other indirect analysis like the comparison method on the component chains. More direct analysis is desirable if possible, but decomposition can be useful to show the mixing time is polynomially bounded if other methods fail.

We first review the approach of Madras and Randall, more background provided by Caracciolo, Pelissetto, and Sokal, then follow with our disjoint decomposition theorem.

5.1 Previous results

5.1.1 The Madras-Randall decomposition method

Let $\mathcal{M}$ be a Markov chain we are interested in analyzing, with state space Ω and transition matrix P. Following [39], a *cover* of Ω is a collection of subsets $\Omega_1, \ldots, \Omega_m$ such that $\cup_i \Omega_i = \Omega$. We are interested in two classes of induced Markov chains. The first is a set of *restricted* Markov chains, obtained by confining $\mathcal{M}$ to each subset Ω_i, i.e., any move of $\mathcal{M}$ that would take us from an element $x \in \Omega_i$ to some $y \notin \Omega_i$

is rejected. In particular, the restriction to Ω_i is a Markov chain, $\mathcal{M}_i$, where the transition matrix P_{Ω_i} is defined as follows: If $x \neq y$ and $x, y \in \Omega_i$ then $P_{\Omega_i}(x, y) = P(x, y)$; if $x \in \Omega_i$ then $P_{\Omega_i}(x, x) = 1 - \sum_{y \in \Omega_i, y \neq x} P_{\Omega_i}(x, y)$.

The second Markov chain is the *projection* $\mathcal{M}_H$ of the cover $\{\Omega_1, \ldots, \Omega_m\}$, defined on the set $[m] = \{1, \ldots, m\}$, where each point i is associated with the set Ω_i. Let $\Theta = \max_{x \in \Omega} |\{i : x \in \Omega_i\}|$. The transition matrix P_H for Markov chain $\mathcal{M}_H$ is defined by letting $P_H(i, j) = \frac{\pi(\Omega_i \cap \Omega_j)}{\Theta \pi(\Omega_i)}$ for $i \neq j$, and $P_H(i, i) = 1 - \sum_{j \neq i} P_H(i, j)$. The limiting distribution ρ of this chain is $\rho(i) = \pi(\Omega_i)/\widehat{Z}$. The normalizing constant $\widehat{Z}$ is $\widehat{Z} = \sum_i \sum_{x \in \Omega_i} \pi(x)$. Note that $\widehat{Z} \leq \Theta$. Also, note that the projection chain is reversible.

Recall that the spectral gap of the transition matrix P is $Gap(P) = \lambda_0 - |\lambda_1|$, where $1 = \lambda_0 > |\lambda_1| \geq |\lambda_i|$ for all $i \geq 2$. The result of Madras and Randall relates the spectral gap of P with that of P_H and the spectral gap of the restricted chains. Using theorem 3.1 we can then bound the mixing time of $\mathcal{M}$.

Theorem 5.1 (Madras and Randall) [39] *In the preceding framework,*

$$Gap(P) \geq \frac{1}{\Theta^2} Gap(P_H) \min_{i=1,\ldots,m} Gap(P_{\Omega_i}).$$

5.1.2 The CPS tempering method

Caracciolo, Pelissetto, and Sokal (CPS) gave a result that relates the spectral gap of a "composite" Markov chain to the gaps of other intermediate chains. Their work was in the context of simulated tempering, which was proposed independently by Marinari and Parisi [42], and by Geyer and Thompson [17].

To formalize their result, as usual let P denote a transition matrix of a Markov chain $\mathcal{M}$ on Ω that is reversible with respect to the probability distribution π. Suppose that the state space is partitioned into m *disjoint* pieces $\Omega_1, \ldots, \Omega_m$. For each $i = 1, \ldots, m$, define P_{Ω_i}, the restriction of P to Ω_i, by rejecting jumps that leave Ω_i (as

in section 5.1.1). Let π_i be the normalized restriction of π to Ω_i, i.e., $\pi_i(A) = \frac{\pi(A \cap \Omega_i)}{b_i}$ where $b_i = \pi(\Omega_i)$. Let Q be another transition matrix that is also reversible with respect to π. Define $\overline{Q}$ to be the following aggregated transition matrix on the state space $[m]$:

$$\overline{Q}(i,j) = \frac{1}{b_i} \sum_{\substack{x \in \Omega_i, \\ y \in \Omega_j}} \pi(x) Q(x,y) \qquad \forall i,j \in [m]. \tag{5.1}$$

We note that

$$b_i \overline{Q}(i,j) = b_j \overline{Q}(j,i)$$

so $\overline{Q}$ is reversible with respect to the probability measure $b = (b_1, \ldots, b_m)$ on $[m]$.

Caracciolo, Pelissetto, and Sokal gave a bound on the spectral gap of the composite chain QPQ in terms of the restrictions P_{Ω_i} and the aggregate chain $\overline{Q}$.

Theorem 5.2 ([39], Thm A.1) *Assume Q is positive semi–definite. Let $Q^{1/2}$ denote the nonnegative square root of Q. Then*

$$Gap(Q^{1/2} P Q^{1/2}) \geq Gap(\overline{Q}) \min_{i=1,\ldots,m} Gap(P_{\Omega_i}).$$

5.2 A new decomposition result

Our goal is to give a method, similar to that of theorem 5.1 in section 5.1.1, but using a partition of Ω into *disjoint* pieces. We relate the spectral gap of the original chain to the spectral gap of the restriction to each set in the partition, and that of a new projection of this partition. The main ingredient in the new theorem we present is theorem 5.2 of the previous section.

5.2.1 Disjoint decomposition

In the framework of Caracciolo, Pelissetto, and Sokal, the matrices P and Q were very different. However, there is no reason why we should not take $Q = P$ in theorem 5.2.

Doing so allows us to derive a bound on the spectral gap of P. We assume the eigenvalues of P are all non–negative. Recall we may replace P by the transition matrix $\frac{1}{2}(I+P)$ that does have non–negative eigenvalues, and the same stationary distribution as P. The proofs of the next theorem, and the corollary that follows, are expanded versions of those in [44]. As before, we assume that $\Omega_1, \Omega_2, \ldots, \Omega_m$ is a partition of the state space Ω.

Theorem 5.3 *Let P_{Ω_i} denote the restriction of P to Ω_i for $i \in [m]$. Let $\overline{P}$ be defined as above in equation (5.1) with P in place of Q. Then*

$$Gap(P) \geq \frac{1}{2} Gap(\overline{P}) \min_{i=1,\ldots,m} Gap(P_{\Omega_i}).$$

Proof. Take $Q = P$ in theorem 5.2 above. This gives the inequality

$$Gap(P^2) \geq Gap(\overline{P}) \min_{i=1,\ldots,m} Gap(P_{\Omega_i}). \tag{5.2}$$

Note that the eigenvalues of P^2 are the squares of the eigenvalues of P. Since $Gap(P) = 1 - \lambda_1$ and $Gap(P^2) = 1 - \lambda_1^2$, we find that $Gap(P) = 1 - \sqrt{1 - Gap(P^2)}$.

To complete the proof, consider the function $1 - \sqrt{1-x}$. The Taylor series of this function is

$$1 - \sqrt{1-x} \;=\; \frac{1}{2}x + \frac{1}{8}x^2 + \frac{1}{16}x^3 + \frac{5}{128}x^4 + \cdots. \tag{5.3}$$

From (5.3) we see that $1 - \sqrt{1-x} \geq \frac{1}{2}x$ for $x \geq 0$. Taking $x = Gap(P^2)$ gives us the inequality $Gap(P) = 1 - \sqrt{1 - Gap(P^2)} \geq \frac{1}{2} Gap(P^2)$. Multiplying (5.2) by $\frac{1}{2}$, we obtain the result. □

Remark. The interested reader may be curious how we justify taking $Q = P$ in theorem 5.2 since the hypothesis requires that Q is positive semi–definite. Consider the function space $\ell^2(\pi)$, the set of all square summable sequences defined on Ω where for two functions f and g on Ω we define the inner product:

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

Since P is the transition matrix of a reversible Markov chain, in this inner product space it defines a self–adjoint operator, i.e.

$$\langle Pf, g\rangle = \langle f, Pg\rangle \qquad \forall f, g \in \ell^2(\pi).$$

Therefore we see P is positive semi–definite in this inner product space so $P^{1/2}$ is well–defined, theorem 5.2 applies and all the machinery of self–adjoint operators is available to us. For example, the so–called "variational characterization" of the eigenvalues applies to P. The eigenvalue gap of P can be found by solving an optimization problem:

$$Gap(P) = \min_f \frac{\langle f, (I-P)f\rangle}{\langle f, f\rangle}$$

where the minimum is over all non–constant functions f in $\ell^2(\pi)$. See [8] and [20] for general discussion about self–adjoint operators and [39] for details on the proof of theorem 5.2. □

One difficulty of directly applying theorem 5.3 is that we must find (or bound) the eigenvalue gap of the aggregated transition matrix $\overline{P}$. Even computing the transition probabilities of $\overline{P}$ might be difficult for a large state space. Therefore, we also derive a corollary to theorem 5.3. We replace $\overline{P}$ with a new matrix for which it is typically easier to compute the transition probabilities, but is also still useful to provide a bound on the spectral gap of the original chain P. Keeping the same Markov kernel of $\overline{P}$, we replace $\overline{P}$ with the transition matrix P_M on the set $\{1, \ldots, m\}$ that uses Metropolis transitions, i.e.,

$$P_M(i,j) = \frac{1}{2\Delta} \min\{1, \frac{\pi(\Omega_j)}{\pi(\Omega_i)}\}$$

where, as usual, Δ is the maximum degree of a vertex in the Markov kernel of $\overline{P}$. In the Madras-Randall decomposition method a cover of sets having large overlaps is necessary. In this corollary, the idea is that the sets that partition Ω have large

boundaries so that we can cross from one set to another efficiently. To formalize this, let

$$\partial_i(\Omega_j) = \{y \in \Omega_j : \exists x \in \Omega_i \text{ with } P(x,y) > 0\} \quad \forall i,j \in [m].$$

We see that $\partial_i(\Omega_j)$ denotes the set of elements in Ω_j that may be obtained from some element of Ω_i by one step of the Markov chain $\mathcal{M}$.

Corollary 5.4 *With P_M as above, suppose there exists $\beta > 0$ and $\gamma > 0$ such that*

(a) $P(x,y) \geq \beta$ for all $x \sim y$ in P;

(b) $\pi(\partial_i(\Omega_j)) \geq \gamma\pi(\Omega_j)$ for all pairs $i \sim j$ in the Markov kernel of $\overline{P}$.

Then

$$Gap(P) \geq \frac{1}{2}\beta\gamma\, Gap(P_M) \min_{i=1,\ldots,m} Gap(P_{\Omega_i}).$$

Proof. Note that

$$\begin{aligned}
\sum_{\substack{x\in\Omega_i,\\ y\in\Omega_j}} \pi(x)P(x,y) &= \sum_{\substack{x\in\partial_j(\Omega_i),\\ y\in\partial_i(\Omega_j)}} \pi(x)P(x,y) \\
&= \sum_{\substack{x\in\partial_j(\Omega_i),\\ y\in\partial_i(\Omega_j)}} \pi(y)P(y,x) \\
&\geq \sum_{y\in\partial_i(\Omega_j)} \pi(y)\beta \\
&\geq \beta\gamma\pi(\Omega_j),
\end{aligned}$$

where the second equality follows from reversibility and the inequalities follow from conditions (a) and (b). Multiplying by $\frac{1}{\pi(\Omega_i)}$, we see $\overline{P}(i,j) \geq \beta\gamma P_M(i,j)$, so $Gap(\overline{P}) \geq \beta\gamma\, Gap(P_M)$ by lemma 3.4. □

In the next chapter we will see applications of this new decomposition theorem and its corollary.

Chapter 6

Sampling Staircase Walks Using Decomposition

For the remainder of the thesis we examine a sampling problem that has connections to combinatorics and statistical mechanics. Polymers, i.e. long chains of molecules, in a solution are often modeled as *self-avoiding* random walks in a lattice, or those walks that do not cross themselves. We consider a special case where the ends of the polymer are attached to an impenetrable surface. One mathematical interpretation of such polymers is staircase walks. A *staircase walk* is a path joining $(0,0)$ to (n,n) that use $(1,0)$ and $(0,1)$ steps, and does not go below the line $y = x$. Rotating by 45°, we can view them as paths joining $(0,0)$ to $(2n,0)$ that use $(1,1)$ and $(1,-1)$ steps, and does not go below the x-axis. This is the representation we will consider. See figure 6.1 for an example. For obvious reasons, these walks are also referred

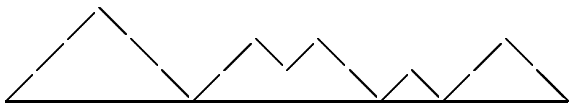


Figure 6.1: A staircase walk ($n = 9$)

to as mountain/valley diagrams. It is well–known (see Stanley [58]) that there are $C(n) = \frac{1}{n+1}\binom{2n}{n}$ staircase walks joining $(0,0)$ to $(2n,0)$. These are the *Catalan numbers*. (See [58] for a multitude of other combinatorial interpretations of the Catalan numbers.)

A natural Markov chain for sampling staircase walks has appeared in [36] and [64] in the context of tilings and triangulations. The transitions replace local maxima with local minima, or vice-versa. Letting U denote a $(1,1)$ step and D a $(1,-1)$ step, we can change a UD pair of consecutive edges to a DU pair, or vice–versa.

In order to represent attraction or repulsion between the components of the polymer and the wall, we assign a weight to each staircase walk. The weights are determined by the number of times the walk hits the x-axis. Let $\lambda > 0$ be a fixed parameter. Assign a weight of λ^k to a staircase walk that hits the x-axis k times. The Gibbs distribution normalizes these weights, defining a probability distribution on the set of staircase walks with $2n$ edges. Taking $\lambda = 1$ gives the uniform distribution. If $\lambda < 1$, the Gibbs distribution favors walks that hit the x-axis a small number of times. These are *repelling* staircase walks. The case $\lambda > 1$ are called *adsorbing* staircase walks, and the Gibbs distribution favors walks that hit the x-axis many times.

The mountain/valley Markov chain may be adapted to incorporate the weights, so that the stationary distribution is now the Gibbs distribution. We present this chain in section 6.1. The analysis in [36] and [64] shows that the mountain/valley chain is rapidly mixing when $\lambda = 1$ and similar analysis using path coupling, demonstrated in section 6.2, will show it is also rapidly mixing when $\lambda < 1$. However, the natural coupling does not extend to the case when $\lambda > 1$, since the coupling is not contracting on the state space, meaning that theorem 3.2 does not apply in this case. We apply the disjoint decomposition method of the previous chapter to examine the Markov chain in the case of adsorbing staircase walks. Before doing so, we present one more connection between staircase walks and a problem from combinatorics.

Consider an infinite d-regular tree. Using the set $A = \{a_1, a_2, \ldots, a_d\}$, label this tree so that each vertex is incident to exactly one edge with each of the labels in A. Fix a vertex as the root. We can label walks in the tree by specifying the sequence of

edges in the walk. Sampling random walks with a fixed number of edges is easy to do. We just sequentially generate elements uniformly at random from A to describe the sequence of edges in the walk. We want to consider the more interesting problem of sampling *closed* walks, or *circuits*, of a fixed length. As we shall see, this is a special case of sampling adsorbing staircase walks.

The key observation we use to sample circuits is a many-to-one mapping between the set of circuits on this tree to the set of staircase walks. A circuit of length $2n$, starting at the root, has n edges leading away from the root and n edges leading back towards the root. When we are at the root there are d choices of labeled edges out of the root; whenever we are away from the root there are $d-1$ edges which move us farther away and a unique edge which brings us closer to the root. Hence, for $1 \leq k \leq n$ there are $d^k(d-1)^{n-k}$ walks of length $2n$ that exit the root k times.

Using adsorbing staircase walks, sampling circuits in these trees (with a fixed length) is easy:

A sampling scheme for circuits in infinite d-regular trees:

1. Select a staircase walk of length $2n$ according to the Gibbs measure with $\lambda = \frac{d}{d-1}$. A staircase walk hitting the x-axis k times appears with probability proportional to $(\frac{d}{d-1})^k = \frac{d^k(d-1)^{(n-k)}}{(d-1)^n}$. The up edges in this walk correspond to steps in the tree that move away from the root, and the down edges are those that move back towards the root.

2. Assign labels to the up edges uniformly at random (from the set of d labels for edges starting from the x-axis, and from a suitable set of $d-1$ labels for edges above the x-axis), assigning labels to the down edges that equal the label of the most recent unpaired up edge preceding it. This gives a sequence of labeled edges corresponding to a labeled walk of length $2n$ in the d-regular tree.

Using this scheme, sampling circuits in the tree reduces to the problem of sampling staircase walks with the appropriate Gibbs measure, then labeling a staircase walk as outlined above. We see this method to requires us to sample staircase walks with $\lambda = \frac{d}{d-1} > 1$, the adsorbing case.

We should note that there are exact methods for sampling staircase walks; for instance dynamic programming may be used to generate them. However, our main goal is to demonstrate the use of the disjoint decomposition method in section 5.2. Furthermore, we gain insight into why a standard coupling argument fails to directly show mixing of the simple mountain/valley Markov chain when $\lambda > 1$, and how decomposition helps to overcome this difficulty.

First, we describe a simple Markov chain in section 6.1. In section 6.2 we use path coupling to show this chain is rapidly mixing for small values of λ. The analysis of adsorbing staircase walks ($\lambda > 1$) begins in section 6.3. Much of the analysis that follows first appeared in [44] with some elaboration here.

6.1 The Markov chain on staircase walks

We define a natural Markov chain $\mathcal{M}$ on the set of staircase walks $\mathcal{S}$. This mountain/valley Markov chain is variant of that used by Luby, Randall and Sinclair [36], McShine and Tetali [47], and Wilson [64] for uniform sampling. We adapt their chain to incorporate the weights so that detailed balance is satisfied. The transitions of the chain are *inversions* which replace local maxima with local minima, or vice–versa, by interchanging two edges along the walk. If the cth point on the path is $v_c = (x_c, y_c)$, we call it a *mountain* if $y_{c-1} = y_c - 1 = y_{c+1}$ and inverting it consists of setting $y'_c = y_c - 2$. Likewise, inverting a *valley* where $y_{c-1} = y_c + 1 = y_{c+1}$ consists of setting

$y'_c = y_c + 2$. The Markov chain $\mathcal{M}$ iterates the following steps.

One step of the mountain/valley Markov chain $\mathcal{M}$:

1. Pick c uniformly at random from $\{1, \ldots, 2n\}$, and let v denote the point on the walk whose x-coordinate is c. Also pick $d \in_u \{up, down\}$.
2. (a) If v is the bottom of a valley lying on the x-axis and $d = up$, with probability $\frac{1}{(1+\lambda)}$ set σ_{t+1} equal to σ_t inverted at v. Otherwise, set $\sigma_{t+1} = \sigma_t$.

   (b) If v is the top of a mountain, inverting it will put it on the x-axis, and $d = down$, with probability $\frac{\lambda}{(1+\lambda)}$ set σ_{t+1} equal to σ_t inverted at v.

   (c) If v is the bottom of a valley not lying on the x-axis, and $d = up$, with probability $\frac{1}{4}$ set σ_{t+1} equal to σ_t inverted at v.

   (d) If v is the top of a mountain, inverting at v does not put it on the x-axis, and $d = down$, with probability $\frac{1}{4}$ set σ_{t+1} equal to σ_t inverted at v.

   (e) In all other cases, set $\sigma_{t+1} = \sigma_t$.

This procedure shows the transition probabilities for a pair $\sigma_1 \neq \sigma_2$ differing by a single mountain/valley move are:

$$P(\sigma_1, \sigma_2) = \begin{cases} \frac{1}{4n(1+\lambda)} & \text{(pushing a valley off the } x\text{-axis)} \\ \frac{\lambda}{4n(1+\lambda)} & \text{(pushing a mountain down onto the } x\text{-axis)} \\ \frac{1}{8n} & \text{(inverting a mountain/valley away from the } x\text{-axis)} \end{cases}$$

Note that this mountain/valley Markov chain is aperiodic and reversible. We may check that detailed balance (equations (3.2)) is satisfied to see that the stationary distribution is the Gibbs distribution, namely $\pi(\sigma) = \frac{\lambda^k}{Z}$, where k is the number of times the walk σ touches the x-axis. For example, consider two staircase walks σ_1, σ_2 that differ by an inversion that involves pushing a valley off the x-axis. The

walk with the extra x-axis hit, say σ_1, has $\pi(\sigma_1) = \lambda^{k+1}/Z$ and $P(\sigma_1, \sigma_2) = \frac{1}{4n(1+\lambda)}$. For the other staircase walk σ_2, we have $\pi(\sigma_2) = \lambda^k/Z$ and $P(\sigma_2, \sigma_1) = \frac{\lambda}{4n(1+\lambda)}$. In this case we see $\pi(\sigma_1)P(\sigma_1, \sigma_2) = \pi(\sigma_2)P(\sigma_2, \sigma_1)$ so that detailed balance is satisfied. Other cases are similar. First we consider what is possible to demonstrate using path coupling in the next section.

6.2 Repelling staircase walks ($\lambda \leq 1$)

In the case of repelling staircase walks ($0 < \lambda \leq 1$) we can use path coupling to prove a polynomial bound on the mixing rate. This is a straightforward generalization of the uniform case ($\lambda = 1$).

First we define our distance measure Φ on the set $\mathcal{S}$ to be one–half of the area between the configurations, i.e., drawing a pair of staircase walks on the same set of axis bounds rectangular regions between the pair of walks. The distance between the two walks is one–half of the sum of the areas of these rectangular regions. In this case, to define the coupling, we take the point c and the direction d in step 1 of $\mathcal{M}$, and attempt to perform the same transition in each walk. To use path coupling, we take $U \subseteq \mathcal{S} \times \mathcal{S}$ to be the pairs of configurations that differ by a single transition of the chain (a single square). We show the coupled chain is contracting on U.

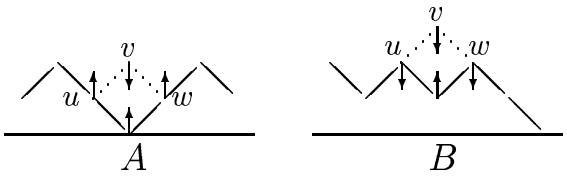


Figure 6.2: Typical situations for path coupling

Lemma 6.1 *Let $\sigma_1, \sigma_2 \in \mathcal{S}$ with $\Phi(\sigma_1, \sigma_2) = 1$. After one step of $\mathcal{M}$ we have $E(\Delta\Phi(\sigma_1, \sigma_2)) \leq 0$, provided $\lambda \leq 1$.*

Proof. Consider the configurations in figure 6.2, which show pieces of walks that agree everywhere except at a single square, and let σ_2 denote the "upper" walk. If the square is adjacent to the x-axis as in figure 6.2.A, then there are two transitions that decrease the distance by one, inverting v in one of the walks so they now agree everywhere. Inverting u or w in σ_2 increases the distance by one. Every other transition not involving u, v, or w does not change the distance between σ_1 and σ_2. Therefore, in this case we find $E(\Delta\Phi(\sigma_1, \sigma_2)) \leq \frac{1}{2n}(\frac{1}{4} + \frac{1}{4} - \frac{1}{2(1+\lambda)} - \frac{\lambda}{2(1+\lambda)}) = 0$. (This is an inequality since one or both of the moves which increase the distance might not be valid moves.) The second case is if the differences between the two walks occur a unit distance from the x-axis, as in figure 6.2.B. There are also two inversions at v that decrease the distance, each occurring with probability $\frac{1}{4}$. The moves at u and w increase the distance between σ_1 and σ_2 by one; all other moves preserve the distance between the pair. In this case, we have $E(\Delta\Phi(\sigma_1, \sigma_2)) \leq \frac{1}{2n}(\frac{\lambda}{2(1+\lambda)} + \frac{\lambda}{2(1+\lambda)} - \frac{1}{4} - \frac{1}{4})$. This last expression is non–positive if $\lambda \leq 1$. Other situations where σ_1 and σ_2 differ by a square that is far away from the x-axis are neutral; two good moves decrease the distance by one, and (at most) two bad moves increase the distance by one. Each of these moves occurs with equal probability, so in these cases we also have $E(\Delta\Phi(\sigma_1, \sigma_2)) \leq 0$. □

An application of theorem 3.2 gives a polynomial bound on the mixing rate. The distance between any pair of staircase walks is at most $\frac{1}{2}(n^2 - n)$. Since we consider the case $0 < \lambda \leq 1$, we see that the minimum probability that the distance changes is $\frac{\lambda}{4n(1+\lambda)}$. Applying theorem 3.2 with $B = \frac{1}{2}(n^2 - n)$ and $\alpha = \frac{\lambda}{4n(1+\lambda)}$ we find that

$$\tau_{\mathcal{M}}(\varepsilon) = \mathrm{O}(n^5 \frac{1+\lambda}{\lambda} \log \varepsilon^{-1}).$$

Using a different argument in the uniform case ($\lambda = 1$), Wilson [64] has shown a tight bound of $\mathrm{O}(n^3(\log n + \log \varepsilon^{-1}))$ for the mixing time, better than the bound we have shown here by this simple argument.

Note that in the case demonstrated in figure 6.2.B the distance will increase in expectation if $\lambda > 1$. This means theorem 3.2 does not apply in this case. The effect of the adsorption of the staircase walks onto the x-axis means that the coupling is no longer contracting on all of $\mathcal{S}$.

6.3 Adsorbing staircase walks ($\lambda > 1$)

Showing rapid mixing for sampling staircase walks in the case $\lambda > 1$ is much harder, precisely because the simple path coupling approach fails. We outline the direction of the remainder of this chapter for the case of adsorbing staircase walks. We treat this problem by applying the decomposition theorem, in fact twice. A natural choice is to partition $\mathcal{S}$ into sets $\mathcal{S}_k$, where each walk in $\mathcal{S}_k$ hits the x-axis exactly k times between the endpoints. We first show the cardinalities $|\mathcal{S}_0|, |\mathcal{S}_1|, \ldots, |\mathcal{S}_{n-1}|$ form a log–concave sequence (section 6.3.2). In section 6.3.3 we explain why this is enough to show rapid mixing for the projection chain. Hence, it suffices to show the restriction to each $\mathcal{S}_k$ is rapidly mixing.

To examine the restricted chain on $\mathcal{S}_k$, we apply decomposition a second time. We partition $\mathcal{S}_k$ into $\binom{n-1}{k}$ sets. Within each set of this second partition, every walk hits the x-axis at exactly the same locations. Note that the distribution on the set $\mathcal{S}_k$ is uniform and, hence, in each set after further partitioning $\mathcal{S}_k$ the distribution is still uniform. We will see that rapid mixing of the second restrictions to the subsets of $\mathcal{S}_k$ is now readily established using path coupling, and is very much like the case when $\lambda \leq 1$.

The projection of the partition of $\mathcal{S}_k$ turns out to be basically the exclusion processes of section 4.1. The location of the particles correspond to the places where the walks hit the x-axis. The stationary probability of a particle configuration is proportional to a product of Catalan numbers. This allows us to use the analysis

in section 4.3.2 as the Catalan numbers are a log–convex sequence, and satisfy the special condition given in theorem 4.6. These pieces together imply rapid mixing for each chain on $\mathcal{S}_k$. Having established mixing for the projection chain for the partition of $\mathcal{S}$, we can finally establish a bound on the mixing time for the Markov chain $\mathcal{M}$ on all of $\mathcal{S}$. We formalize the analysis in what follows.

6.3.1 Decomposition of $\mathcal{S}$

We apply the decomposition theorem to the partition $\mathcal{S} = \dot{\cup}\mathcal{S}_k$ to analyze $\mathcal{M}$. An unfortunate technicality is that the restricted chains on $\mathcal{S}_k$ are no longer ergodic. To get around this difficulty, we will define a new Markov chain $\widehat{\mathcal{M}}$ by augmenting $\mathcal{M}$ with new moves; $\widehat{\mathcal{M}}$ is very similar to $\mathcal{M}$, and the rapid mixing of $\widehat{\mathcal{M}}$ will imply rapid mixing of $\mathcal{M}$ via an application of the comparison theorem.

In this new Markov chain $\widehat{\mathcal{M}}$ there are two basic types of moves. The first type of moves are inversions that change a mountain to a valley or vice–versa. The second type of move consists of changing one "propeller-like" structure into its mirror image. Letting D denote a "down" edge and U an "up" edge, if there is a sequence of four edges $DUUD$, we can change it to the sequence $UDDU$, or vice-versa. These moves are only allowed when one point of the propeller touches the boundary. See figure 6.3 for a pictorial depiction of this move. We call such a change a propeller move (centered) at v.

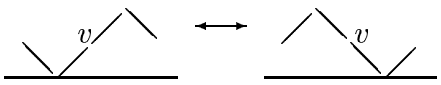


Figure 6.3: The propeller move

More formally, $\widehat{\mathcal{M}}$ iterates these steps:

One step of the augmented mountain/valley Markov chain $\widehat{\mathcal{M}}$:

1. Pick c uniformly at random from $\{1, \ldots, 2n\}$, and let v denote the point on the walk whose x-coordinate is c. Also pick $d \in_u \{up, down\}$.
2. (a) If $c = 2$, v is the bottom of a valley on the x-axis, and $d = up$, with probability $\frac{1}{(1+\lambda)}$ set X_{t+1} equal to X_t inverted at v.

   (b) If $c = 2$, v is the top of a mountain and inverting it will put it on the x-axis, and $d = down$, with probability $\frac{\lambda}{(1+\lambda)}$ set X_{t+1} equal to X_t inverted at v.

   (c) If v is the bottom of a valley not lying on the x-axis, and $d = up$, with probability $\frac{1}{4}$ set X_{t+1} equal to X_t inverted at v.

   (d) If v is the top of a mountain and inverting at v does not push it onto the x-axis, and $d = down$, with probability $\frac{1}{4}$ set X_{t+1} equal to X_t inverted at v.

   (e) If v is the central vertex of a propeller structure where the lowest point lies on the x-axis, with probability $\frac{1}{4}$ set X_{t+1} equal to X_t after performing a propeller move at v.

   (f) In all other cases, set $X_{t+1} = X_t$.

First note that $\widehat{\mathcal{M}}$ is aperiodic ($X_{t+1} = X_t$ with probability at least $\frac{1}{2}$). Second, the only time a transition is possible from a path hitting the boundary k times to one hitting the boundary $k+1$ (respectively $k-1$) times is when there is a mountain (respectively valley) at the beginning of the walk, and we select that vertex in step 1 of the chain. All other moves of $\widehat{\mathcal{M}}$ preserve the weight of the walk.

The transition probabilities $\widehat{P}(\sigma_1, \sigma_2)$ for this chain $\widehat{\mathcal{M}}$ for walks $\sigma_1 \neq \sigma_2$ differing

by a single transition are

$$\widehat{P}(\sigma_1,\sigma_2) = \begin{cases} \frac{1}{4n(1+\lambda)} & \text{(pushing a valley off the } x\text{-axis when } c{=}2) \\ \frac{\lambda}{4n(1+\lambda)} & \text{(pushing a mountain down onto the } x\text{-axis when } c=2) \\ \frac{1}{8n} & \text{(inverting a mountain/valley away from the } x\text{-axis)} \\ \frac{1}{8n} & \text{(if } \sigma_1 \text{ differs from } \sigma_2 \text{ by a propeller move)} \end{cases}$$

Having described the Markov chain $\widehat{\mathcal{M}}$, we use it to define a new metric $\widehat{\Phi}$ on $\mathcal{S}$. For any pair of states $\sigma_1, \sigma_2 \in \mathcal{S}$, if $\widehat{P}(\sigma_1,\sigma_2) > 0$ (so σ_1 and σ_2 are nearest neighbors), we define $\widehat{\Phi}(\sigma_1,\sigma_2)$ to equal one–half of the area of the symmetric difference of the two staircase walks. If $\widehat{P}(\sigma_1,\sigma_2) = 0$ (i.e., moving from σ_1 to σ_2 requires more than one move), first consider a *path of states* $\sigma_1 = Z_0, Z_1, \ldots, Z_r = \sigma_2$ between σ_1 and σ_2, where $P(Z_i, Z_{i+1}) > 0$ for each $i = 0, \ldots r-1$; then define $\widehat{\Phi}(\sigma_1,\sigma_2) = \min \sum_{i=0}^{r-1} \widehat{\Phi}(Z_i, Z_{i+1})$ where the minimum is taken over all paths joining σ_1 and σ_2. We call $\widehat{\Phi}$ the *transition metric.*

We are now prepared to start analyzing the derivative Markov chains necessary for the decomposition theorem. We will start with the projection chain. In the next section, we prove some combinatorial lemmas that will be the key to a simple argument that this projection is rapidly mixing.

6.3.2 A combinatorial look at $\mathcal{S}$

For this subsection, we let $\mathcal{S}^n$ denote the set of staircase walks with $2n$ edges and set $s^n = |\mathcal{S}^n|$. We use $\mathcal{S}^n_k$ to denote the subset of $\mathcal{S}^n$ containing those walks with k internal x-axis hits and let $s^n_k = |\mathcal{S}^n_k|$. These cardinalities can be shown to be log–concave, i.e., the sequence of numbers $s^n_0, s^n_1, s^n_2, \ldots, s^n_{n-1}$ is a log–concave sequence. This follows from two simple lemmas.

Lemma 6.2 *For $n \geq 3$ and $1 \leq k \leq n-2$, $s^n_k = s^{n-1}_{k-1} + s^n_{k+1}$.*

Proof. We partition $\mathcal{S}_k^n$ into two sets A_k^n, the subset of walks that begin with two edges UD, and B_k^n, the remaining walks in $\mathcal{S}_k^n$. We define two bijections. The first maps $\sigma \in A_k^n$ to an element in $\mathcal{S}_{k-1}^{n-1}$ by deleting the first two edges (the initial "bump" in the walk), and sliding the remaining piece of σ left two spaces. The second bijection is slightly more complicated; for $\sigma \in B_k^n$ let u denote the first point on the walk with y-coordinate equal to one (this is the point with x-coordinate equal to one). Let v be the first point to the right of u having y-coordinate one. Map σ to a walk in $\mathcal{S}_{k+1}^n$ by removing the first edge of σ, sliding the piece of σ between u and v to the left and down by one unit, adding an upwards edge at the end of this piece, and continuing the walk with the remainder of σ. See figure 6.4 for an example of this mapping. It is easy to verify that each map is a bijection. □

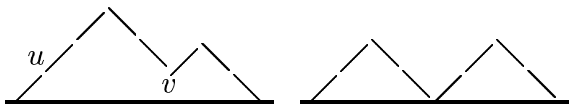


Figure 6.4: An example of the map from B_0^4 to $\mathcal{S}_1^4$

Lemma 6.3 *For $n \geq 3$ and $1 \leq k \leq n-2$,*

$$s_k^n = s_{k-1}^{n-1} + s_k^{n-1} + \cdots + s_{n-2}^{n-1}.$$

Also, for $n \geq 2$,

$$s_0^n = s^{n-1} = s_0^{n-1} + \cdots + s_{n-2}^{n-1}.$$

Proof. For the first part, we use lemma 6.2 iteratively, so

$$\begin{aligned} s_k^n &= s_{k-1}^{n-1} + s_{k+1}^n \\ &= s_{k-1}^{n-1} + s_k^{n-1} + s_{k+2}^n = \cdots \end{aligned}$$

For the second, there is a bijection between $\mathcal{S}_0^n$ and $\mathcal{S}^{n-1}$ by taking $\sigma \in \mathcal{S}_0^n$, deleting the initial and terminal edges of the walk, and shifting the walk down and to the left by one to obtain a staircase walk joining $(0,0)$ to $(2n-2,0)$. □

Theorem 6.4 *For a fixed $n \geq 3$, s_k^n is log–concave. In particular, for $1 \leq k \leq n-2$,*

$$s_{k-1}^n \cdot s_{k+1}^n \leq (s_k^n)^2. \tag{6.1}$$

Proof. We use induction on n.

For $n = 3$, by a simple enumeration of the possibilities, we find that $s_0^3 = s_1^3 = 2$ and $s_2^3 = 1$, so that $s_0^3 \cdot s_2^3 \leq (s_1^3)^2$.

Now assume for some $n-1$ that s_k^{n-1} is log–concave. Also, assume first that $k \geq 2$. We want to show that (6.1) holds. To do this, it suffices to show the inequality

$$s_{k-2}^{n-1} \cdot [s_k^{n-1} + s_{k+1}^{n-1} + \cdots + s_{n-2}^{n-1}] \;\leq\; s_{k-1}^{n-1} \cdot [s_{k-1}^{n-1} + s_k^{n-1} + \cdots + s_{n-2}^{n-1}] \tag{6.2}$$

since lemma 6.3 implies that (6.2) is equivalent to

$$s_{k-2}^{n-1} \cdot s_{k+1}^n \;\leq\; s_{k-1}^{n-1} \cdot s_k^n. \tag{6.3}$$

By adding $s_k^n \cdot s_{k+1}^n$ to both sides of (6.3), factoring, and applying lemma 6.2, we get (6.1).

To show (6.2), it suffices to show the set of inequalities

$$s_{k-2}^{n-1} \cdot s_{k-1+i}^{n-1} \;\;\leq\;\; s_{k-1}^{n-1} \cdot s_{k-2+i}^{n-1}$$

for all $i \in [n-k-1]$. These inequalities all hold by our induction hypothesis that s_k^{n-1} is log–concave. Adding them, and the extra term $s_{k-1}^{n-1} \cdot s_{n-2}^{n-1}$ to the right hand side, gives us (6.2).

All that remains is the case $s_0^n \cdot s_2^n \leq (s_1^n)^2$ (when $k = 1$). We use that $s_0^n = s_1^n$ and, from lemma 6.2, we see $s_2^n \leq s_1^n$. Therefore, $s_0^n \cdot s_2^n = s_1^n \cdot s_2^n \leq s_1^n \cdot s_1^n$. □

6.3.3 Projection 1 using the partition $\mathcal{S} = \dot{\cup}\mathcal{S}_k$

The state space of this projection is the set $\{0, 1, \ldots, n-1\}$ corresponding to the number of internal x-axis hits for the walks in $\mathcal{S}_k$. The stationary probability $\rho(k)$ of the point k is defined by the Gibbs measure on the set $\mathcal{S}_k$. Therefore, the stationary probability is $\rho(k) = \frac{\lambda^k |\mathcal{S}_k|}{Z_M}$, where $Z_M = \sum_{k=0}^{n-1} \lambda^k |\mathcal{S}_k|$. Because the cardinalities of these sets are log–concave, implying the stationary distribution is unimodal, we would expect a random walk to converge quickly.

Since we use corollary 5.4 to apply decomposition, we want to consider the projection chain P_M that uses Metropolis transitions. The transition probabilities are given by

$$P_M(k,j) = \begin{cases} \frac{1}{4}\min\{1, \frac{\rho(j)}{\rho(k)}\} & \text{for } j \in \{k-1, k+1\} \\ 1 - P_M(k-1) - P_M(k+1) & \text{if } k = j \\ 0 & \text{for } j \notin \{k-1, k, k+1\}. \end{cases}$$

As usual, these transitions give at least $\frac{1}{2}$ holding probabilities, and the Metropolis acceptance ratios ensure that ρ is the stationary distribution of P_M. We bound the mixing rate of the projection by appealing to theorem 6.4 and using path coupling.

Lemma 6.5 *The mixing time of P_M satisfies*

$$\tau_{P_M}(\varepsilon) = O(n^2 \log \varepsilon^{-1}).$$

Proof. To use path coupling, we need a metric on the state space $\{0, \ldots, n-1\}$. In this case, the distance between two configurations is simply the distance between the location of the particles in this interval. Using path coupling, we take U to be the pairs of states that differ by unit distance as in figure 6.5, where the particle is at position k in the upper configuration, σ_1, and $k+1$ in the lower one σ_2.

We couple the moves so that the particle moves in only one of the configurations during each time step of the coupled process, i.e. if we move the particle left or

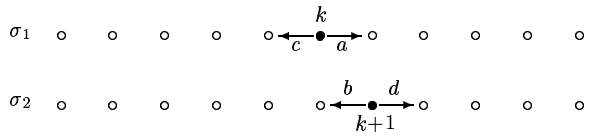


Figure 6.5: Typical situation for path coupling in P_M

right in σ_1, then we hold the particle fixed in σ_2. The 1/2 holding probabilities allow us to couple the moves in this fashion. Letting a, b, c, d denote the transition probabilities as labeled in figure 6.5, the expected change in distance after one move in the coupled chain is $(-1)\cdot(a+b)+(1)\cdot(c+d)$. To apply theorem 3.2 we need to show this quantity is less than or equal to zero, or, equivalently, show $a+b \geq c+d$. Consider first the case when $\lambda = \frac{|\mathcal{S}_k|}{|\mathcal{S}_{k+1}|}$. With this value, we have that $a = b = \frac{1}{4}$, $c = \frac{1}{4}\min\{1, \frac{|\mathcal{S}_{k+1}|\cdot|\mathcal{S}_{k-1}|}{|\mathcal{S}_k|^2}\} \leq \frac{1}{4}$, and $d = \frac{1}{4}\min\{1, \frac{|\mathcal{S}_k|\cdot|\mathcal{S}_{k+2}|}{|\mathcal{S}_{k+1}|^2}\} \leq \frac{1}{4}$. So $a+b \geq c+d$ as desired.

If $0 < \lambda < \frac{|\mathcal{S}_k|}{|\mathcal{S}_{k+1}|}$ we note that $a = \frac{1}{4}\frac{\lambda|\mathcal{S}_{k+1}|}{|\mathcal{S}_k|} \geq \frac{1}{4}\frac{\lambda|\mathcal{S}_{k+2}|}{|\mathcal{S}_{k+1}|} = d$, where the inequality follows using log–concavity of the sequence $|\mathcal{S}_k|$. Since $b = \frac{1}{4}\min\{1, \frac{|\mathcal{S}_k|}{\lambda|\mathcal{S}_{k+1}|}\} = \frac{1}{4} \geq c$ we again have $a+b \geq c+d$. The last case where $\lambda > \frac{|\mathcal{S}_k|}{|\mathcal{S}_{k+1}|}$ is similar, so that $a+b \geq c+d$ for all $\lambda > 0$.

Appealing to theorem 3.2, noting that the maximum distance between two particles is at most n and $\alpha = \frac{1}{4}$ suffices in the hypothesis of the theorem (either $a = \frac{1}{4}$ or $b = \frac{1}{4}$ for any value of λ) we have

$$\tau_{P_M}(\varepsilon) \leq \lceil \frac{en^2}{\frac{1}{4}} \rceil \lceil \log \varepsilon^{-1} \rceil = O(n^2 \log \varepsilon^{-1}).$$

□

6.3.4 Restriction 1: Mixing on S_k

By the disjoint decomposition theorem, it now suffices to show that the restricted Markov chains $\widehat{\mathcal{M}}_k$ ($\widehat{\mathcal{M}}$ restricted to S_k) are rapidly mixing in order to conclude that

$\widehat{\mathcal{M}}$ is mixing on the whole state space $\mathcal{S}$. This is somewhat trickier and we establish this indirectly by applying decomposition a second time and using the comparison method in the next several sections.

6.3.5 The second decomposition: $S_k = \dot{\cup}\mathcal{S}_{k,T}$

In order to show rapid mixing on $\mathcal{S}$, we partition this set based on the location of the x-axis hits. Let T denote a subset of $\{2, 4, \ldots, 2n-2\}$ where $|T| = k$, and let $\mathcal{S}_{k,T}$ denote the subset of $\mathcal{S}_k$ having walks that touch the x-axis in the same k locations specified by T. For example, (in the case that $n \geq 6$) we can take $T = \{2, 6, 10\}$ and consider the set of walks that hit the x-axis at the points with x-coordinates $2, 6$, and 10 in the interior between the two endpoints. There are $\binom{n-1}{k}$ ways to specify the location of k internal hits, as the x-coordinate of each hit must be an even number. We write $\mathcal{S}_k = \dot{\cup}_T \mathcal{S}_{k,T}$, where this union is over all $\binom{n-1}{k}$ ways of specifying the hits on the x-axis.

The distribution is uniform on each of the subsets $\mathcal{S}_{k,T}$. Showing rapid mixing of the restriction of $\widehat{\mathcal{M}}_k$ to each subset $\mathcal{S}_{k,T}$ proceeds as it did in section 6.2 in the case $\lambda \leq 1$. We show this in the next section. Following that, in section 6.3.7 we examine the projection of the second partition $\mathcal{S}_k = \dot{\cup} S_{k,T}$. Analysis of this projection makes use of the previous results on particle processes from chapter 4.

6.3.6 Restriction 2: Mixing of $\mathcal{S}_{k,T}$

Let $\widehat{\mathcal{M}}_{k,T}$ denote the restriction of $\widehat{\mathcal{M}}_k$ to the set $\mathcal{S}_{k,T}$. We have the following result, whose proof is a simple application of path coupling:

Lemma 6.6 *Let $\mathcal{S}_{k,T}$ be a subset of $\mathcal{S}_k$ as above, and let $\sigma_1, \sigma_2 \in \mathcal{S}_{k,T}$ with $\widehat{\Phi}(\sigma_1, \sigma_2) = 1$. After one step of the Markov chain $\widehat{\mathcal{M}}_{k,T}$ on $\mathcal{S}_{k,T}$ we have $E(\Delta\widehat{\Phi}(\sigma_1, \sigma_2)) \leq 0$.*

Proof. This proof is similar to lemma 6.1, except that we need only consider the situation when the paths differ by a square that is at least distance one from the x-axis as in figure 6.6. There are two good inversions at v that decrease the distance by one, and at most two inversions increasing the distance by one. Each of these inversions happens with equal probability, so $E(\Delta\widehat{\Phi}(\sigma_1,\sigma_2)) \leq 0$. □

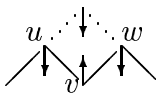


Figure 6.6: Possible transitions inside $\mathcal{S}_{k,T}$

Lemma 6.6 gives the first piece for the path coupling theorem. If $\widehat{\Phi}(\sigma_1,\sigma_2) > 0$, the probability of the distance changing in one step of $\widehat{\mathcal{M}}_{k,T}$ is at least $\frac{1}{4(n-k)}$, since when we select a vertex on the walk we may avoid choosing one that lies on the x-axis, and its immediate neighbor to the right, as these vertices will never move.

For pairs of walks in $\mathcal{S}_k$, we have $0 \leq \widehat{\Phi}(\sigma_1,\sigma_2) \leq (n-k-1)^2 - (n-k-1)$. For $\sigma_1,\sigma_2 \in \mathcal{S}_{k,T}$ there is a sequence of $\widehat{\Phi}(\sigma_1,\sigma_2)$ inversions that will transform one walk into the other.

By a straightforward application of path coupling (theorem 3.2(b)), we have

Lemma 6.7 *The mixing time of $\widehat{\mathcal{M}}_{k,T}$ satisfies*

$$\begin{aligned}\tau_{\widehat{\mathcal{M}}_{k,T}}(\varepsilon) &\leq \lceil \frac{e((n-k-1)(n-k-2))^2}{\frac{1}{4(n-k)}} \log \varepsilon^{-1} \rceil \\ &= O(n^5 \log \varepsilon^{-1}).\end{aligned}$$

6.3.7 Projection 2 using $\mathcal{S}_k = \dot{\cup}_T \mathcal{S}_{k,T}$

The projection P_k of the partition $\mathcal{S}_k = \dot{\cup}_T \mathcal{S}_{k,T}$ can be viewed as a particle process on $[n-1]$. The k particles represent the places that a path hits the x-axis. Because of the propeller moves in the Markov chain, the projection of the chain $\widehat{\mathcal{M}}_k$ corresponds

to nearest–neighbor dynamics on the set of particles which moves one particle to the left or right one space in each step. As we have seen in chapter 4 it is difficult to analyze the mixing rate of this particle process using a simple path coupling argument, which seems to isolate the difficulty with using path coupling on the original mountain/valley chain. However, the more intricate analysis in that chapter can be used to help bound the mixing time of this projection.

Recalling the framework in section 4.1, we have a linear arrangement of $n-1$ sites, k of which contain particles. (In that section we had n sites, but the result is the same.) Denoting the collection of all $\binom{n-1}{k}$ arrangements by Ω_k, an element $\sigma \in \Omega_k$ corresponds to the set $\mathcal{S}_{k,T(\sigma)}$ of all staircase walks in $\mathcal{S}_k$ that have x-axis hits at the locations determined by the particles in σ. For example, if $k=3$ and σ is the configuration with particles at sites 2, 3, and 8, then $\mathcal{S}_{3,T(\sigma)}$ consists of all walks in $\mathcal{S}_3$ that hit the x-axis at coordinates 4, 6, and 16 (noting that walks only hit the x-axis at even coordinates). The fixed particles added to each configuration at sites 0 and n correspond to the fixed endpoints of each staircase walk.

The Catalan numbers are used to assign weights to the particle configurations in order to count the number of paths in each set of the partition of $\mathcal{S}_k$. That is, the stationary probability of $\sigma \in \Omega_k$ is proportional to $C(x_1)C(x_2)\cdots C(x_{k+1})$, where x_j is the number of empty sites between particles j and $j+1$ in σ. The transitions of the projection are the single site dynamics described in section 4.2.1 but, as mentioned, path coupling fails to directly show rapid mixing. However, taking $f_t = C(t)$, we check that the Catalan numbers satisfy the two conditions of theorem 4.6, namely

1. $C(t)$ is a log–convex sequence, i.e. $C(t)^2 \leq C(t-1)C(t+1)$
2. $\frac{C(t)}{C(t-1)} \leq \frac{g_{t+1}}{g_t}$ (for $t \geq 1$) where $g_t = \sum_{i=1}^{t} C(i-1)C(t-i)$.

We recall the recurrence relation for the Catalan numbers is $C(n+1) = C(0)C(n) + C(1)C(n-1) + C(2)C(n-2) + \cdots + C(n)C(0)$ for $n \geq 0$, with the initial condition

$C(0) = 1$. Using that $C(t) = \frac{1}{t+1}\binom{2t}{t}$ (for example, see [58]), it is easy to verify the Catalan numbers are log–convex and the recurrence relation implies that $g_t = \sum_{i=1}^{t} C(i-1)C(t-i) = C(t)$. In fact, the inequality $\frac{C(t)}{C(t-1)} \leq \frac{g_{t+1}}{g_t} = \frac{C(t+1)}{C(t)}$ that is needed in theorem 4.6 follows directly from the log–convexity of the Catalan numbers. Therefore, we see that theorem 4.6 applies in this case and the gap of the heat bath chain $\widetilde{P}_k$ on Ω_k (where particles may move more than a single step left or right) has a lower bound of $\frac{c}{n^5}$ for a constant c. Lower bounding the gap of the nearest–neighbor projection chain P_k is done using the comparison theorem.

Lemma 6.8 *The Markov chain P_k on Ω_k with nearest–neighbor dynamics is mixing in polynomial time and $Gap(P_k) \geq \frac{C_1}{n^8}$, where C_1 is a constant independent of n.*

Proof. By part (a) of theorem 3.3 it suffices to bound the parameter A for the chains P_k and $\widetilde{P}_k$ to find a lower bound on $Gap(P_k)$. In the notation of that theorem, for any transition (z, w) of P_k, we have that $\Gamma(z, w) \leq n - k$ and $|\gamma_{xy}| \leq n - k$ for any pair x, y with $(x, y) \in E(\widetilde{P}_k)$. We also have $\frac{\pi(x)}{\pi(z)P(z,w)} \leq \frac{k}{2}$. Using the trivial bound $\widetilde{P}_k(x, y) \leq 1$, we conclude that $A \leq k(n-k)^2 = \mathrm{O}(n^3)$, so $Gap(P_k) \geq \frac{C_1}{n^8}$. □

We now have shown in lemmas 6.7 and 6.8 that the restrictions defined by the decomposition $\mathcal{S}_k = \dot{\cup}_T \mathcal{S}_{k,T}$, as well as the projection, are all mixing in polynomial time. Appealing to corollary 5.4 with $\beta = \frac{1}{8n}$ and $\gamma = \frac{1}{4}$, we find:

Lemma 6.9 *The Markov chain $\widehat{\mathcal{M}}_k$ on $\mathcal{S}_k$ is rapidly mixing and $Gap(\widehat{\mathcal{M}}_k) \geq \frac{C_2}{n^{14}}$.*

Proof. Lemma 6.7 (with theorem 3.1(b)) tells us that

$$Gap(\widehat{\mathcal{M}}_{k,T}) \geq \frac{C_2}{n^5}$$

for some constant C_2 that is independent of n. Lemma 6.8 bounds the gap of the projection of the partition of $\mathcal{S}_k$. We have $\beta = \frac{1}{8n}$ because the transitions of the

Markov chain $\widehat{\mathcal{M}}_k$ on $\mathcal{S}_k$ are either simple inversions or propeller moves. These happen with probability $\frac{1}{4n}$ and $\frac{1}{8n}$, respectively; β is the minimum of these transition probabilities. Recall from corollary 5.4 that γ satisfies

$$\pi(\partial_i(\Omega_j)) \geq \gamma\pi(\Omega_j)$$

for all pairs $i \sim j$ in the Markov kernel of the projection chain, where

$$\partial_i(\Omega_j) = \{y \in \Omega_j : \exists x \in \Omega_i \text{ with } P(x,y) > 0\}.$$

To see that $\gamma = \frac{1}{4}$ suffices, consider the following argument: A propeller move that changes the staircase walk x into y corresponds to moving a particle one space to the left or right in the projection chain P_k on Ω_k. The reverse statement is not necessarily true, so we need to determine the fraction of staircase walks for which a particle movement does actually correspond to a propeller movein $\widehat{\mathcal{M}}$. Suppose that $\sigma_1, \sigma_2 \in \Omega_k$ are two particle configurations where, without loss of generality, σ_2 is obtained from σ_1 by moving particle i to the right one space in σ_1. The weight of σ_1 is $C(x_1)\cdots C(x_i)C(x_{i+1})\cdots C(x_k+1)$, where, as in chapter 4, x_i is the number of empty sites between particles i and $i+1$ in σ_1. Therefore, the weight of σ_2 is $wt(\sigma_2) = C(x_1)\cdots C(x_i+1)C(x_{i+1}-1)\cdots C(x_k+1)$. For each walk having σ_2 as its configuration in the projection chain invert the mountain in the part of the propellor to get a new staircase walk in $\mathcal{S}_{k+1}$. This defines a map into $\mathcal{S}_{k+1}$ and, given that we know which mountain we inverted to obtain the new walk, is a bijection (i.e we may invert the $i+1$st valley that lies on the x-axis to get back the original walk in $\mathcal{S}_k$). This new walk corresponds to a particle configuration σ_3 in Ω_{k+1} that has weight $wt(\sigma_3) = C(y_1)\cdots C(y_i)C(y_{i+1})\cdots C(y_{k+2})$. Note that in this case we have $y_j = x_j$ for all $j < i$, $y_i = 0, y_{i+1} = x_i - 1$, and $y_{j+1} = x_j$ for all $j > i$. Therefore, the fraction of walks having the configuration σ_2 in the projection chain for which the particle move corresponds to a real propeller move is $wt(\sigma_3)/wt(\sigma_2)$. In the notation

of corollary 5.4 this means

$$\frac{\pi(\partial_{\sigma_1}(\mathcal{S}_{k,\sigma_2}))}{\pi(\mathcal{S}_{k,\sigma_2})} \geq \frac{C(y_1)\cdots C(y_i)C(y_{i+1})\cdots C(y_{k+2})}{C(x_1)\cdots C(x_i)C(x_{i+1})\cdots C(x_k+1)} = \frac{C(y_{i+1})}{C(x_i)} = \frac{C(x_i-1)}{C(x_i)} \geq \tfrac{1}{4}.$$

This implies that $\gamma \geq \frac{1}{4}$.

Using the bounds on the restrictions, projection, and the values of β and γ specified, the result follows from corollary 5.4. □

6.3.8 Mixing for $\mathcal{S}$: The final word

A polynomial bound on the mixing time for $\widehat{\mathcal{M}}$ now follows from all of our previous work. By lemma 6.9 the restrictions to each set $\mathcal{S}_k$ are all rapidly mixing, and the mixing time of the projection followed from the log–concavity of the sets $\mathcal{S}_k$ (lemma 6.5). Using corollary 5.4, we can bound the mixing time of the chain $\widehat{\mathcal{M}}$ on all of $\mathcal{S}$. We first prove a preliminary result that enables us to find a γ that suffices for our purposes. Recall our notation from section 6.3.2 where $s_k^n = |\mathcal{S}_k^n|$.

Lemma 6.10 *Let $n \geq 3$. For all $0 \leq k \leq n-2$ we have*

$$\frac{s_k^{n-1}}{s_k^n} \geq \frac{1}{n}.$$

Proof. This follows from an easy induction on n. First consider the base case when $n = 3$. In this case,

$$\frac{s_0^2}{s_0^3} = \frac{1}{2} \geq \frac{1}{3} \quad \text{and} \quad \frac{s_1^2}{s_1^3} = \frac{1}{2} \geq \frac{1}{3}.$$

Now suppose the statement holds for some $n \geq 3$, and also suppose first that $k \geq 1$. We want to show

$$\frac{s_k^n}{s_k^{n+1}} \geq \frac{1}{n+1} \tag{6.4}$$

or that

$$(n+1)\cdot s_k^n \geq s_k^{n+1}. \tag{6.5}$$

From lemma 6.3, equation (6.5) is equivalent to

$$n(s_{k-1}^{n-1}+s_k^{n-1}+\cdots+s_{n-2}^{n-1})+(s_{k-1}^{n-1}+s_k^{n-1}+\cdots+s_{n-2}^{n-1}) \geq s_{k-1}^n+s_k^n+\cdots+s_{n-2}^n+s_{n-1}^n \quad (6.6)$$

By induction, we have $n \cdot s_i^{n-1} \geq s_i^n$ for $i \in \{k-1, k, \cdots, n-2\}$. Therefore, (6.6) holds if and only if

$$s_{k-1}^{n-1} + s_k^{n-1} + \cdots + s_{n-2}^{n-1} \geq s_{n-1}^n.$$

But this last statement is trivial since, $s_{n-2}^{n-1} = s_{n-1}^n = 1$.

The last case to consider is when $k = 0$. Then we want to show that

$$\frac{s_0^n}{s_0^{n+1}} \geq \frac{1}{n+1}.$$

As before, using lemma 6.3 this is true if and only if

$$(n+1)(s_0^{n-1} + s_1^{n-1} + \cdots + s_{n-2}^{n-1}) \geq s_0^n + s_1^n + \cdots + s_{n-1}^n.$$

Using our inductive hypothesis, this reduces to the statement

$$s_0^{n-1} + s_1^{n-1} + \cdots s_{n-2}^{n-1} \geq s_{n-1}^n$$

which is again trivially true, since $s_{n-2}^{n-1} = s_{n-1}^n = 1$. □

We are now in a position to derive a bound on the spectral gap of $\widehat{\mathcal{M}}$ on $\mathcal{S}$. Note that in this case we have $\beta = \frac{1}{4n(\lambda+1)}$ and $\gamma = \frac{1}{n}$.

Theorem 6.11 *The Markov chain $\widehat{\mathcal{M}}$ on $\mathcal{S}$ is rapidly mixing and $Gap(\widehat{\mathcal{M}}) \geq \frac{C_3}{\lambda n^{18}}$.*

Proof. Theorem 6.9 gives us a bound on the spectral gap of the restricted chains. Recall the projection chain P_M for the first decomposition is a random walk on the interval [n-1]. From lemma 6.5 we have $P_M = \mathrm{O}(n^2 \log \varepsilon^{-1})$. Theorem 3.1(b) implies that $Gap(P_M) \geq \frac{c'}{n^2}$ for some constant c'.

For this application of corollary 5.4 we see that $\beta = \frac{1}{4n(\lambda+1)}$ is the minimum probability of any transition of $\widehat{\mathcal{M}}$ (using that $\lambda > 1$).

For the bound on γ recall that the only way to move from a staircase walk hitting the x-axis k times to a walk hitting the x-axis $k-1$ times is to invert a valley at the very beginning of the walk. Let us also recall the notation from lemma 6.2. We partition the set $\mathcal{S}_k$ into two sets A_k^n, the set of walks that begin UDU, and B_k^n, the remaining walks in $\mathcal{S}_k$. Any walk $\sigma \in A_k^n$ can be mapped into a walk in $\mathcal{S}_{k-1}$ by inverting the valley that occurs at the beginning of the walk. Therefore, using our notation that $s_k^n = |\mathcal{S}_k|$, we see

$$\frac{\pi(\partial_{k-1}(\mathcal{S}_k))}{\pi(\mathcal{S}_k)} = \frac{|\partial_{k-1}(\mathcal{S}_k)|}{|\mathcal{S}_k|} \geq \frac{|A_k^n|}{|\mathcal{S}_k|} = \frac{s_{k-1}^{n-1}}{s_k^n} \tag{6.7}$$

since $|A_k^n| = s_{k-1}^{n-1}$ by lemma 6.2. Lemma 6.3 tells us that

$$s_k^n = s_{k-1}^{n-1} + s_k^{n-1} + s_{k+1}^{n-1} + \cdots + s_{n-2}^{n-1}\,. \tag{6.8}$$

Using lemma 6.2 again, we note

$$s_{k-1}^{n-1} \geq s_k^{n-1} \geq s_{k+1}^{n-1} \geq \cdots \geq s_{n-2}^{n-1}. \tag{6.9}$$

Therefore, (6.8) and (6.9) together imply $s_k^n \leq n \cdot s_{k-1}^{n-1}$. This inequality, with (6.7), shows

$$\frac{\pi(\partial_{k-1}(\mathcal{S}_k))}{\pi(\mathcal{S}_k)} \geq \frac{s_{k-1}^{n-1}}{s_k^n} \geq \frac{s_{k-1}^{n-1}}{n \cdot s_{k-1}^{n-1}} \geq \frac{1}{n}.$$

In a similar manner, consider the set $\mathcal{S}_k$. We want to find (a bound for) the fraction of walks in $\mathcal{S}_k$ that have a valley at the beginning of the walk that may be inverted to give a walk in $\mathcal{S}_{k-1}$. From lemma 6.2 once more, we see that this fraction is

$$\frac{|\partial_k(\mathcal{S}_{k-1})|}{|\mathcal{S}_{k-1}|} \geq \frac{s_{k-1}^{n-1}}{s_{k-1}^n}.$$

Lemma 6.10 shows that this fraction is also at least $\frac{1}{n}$. Therefore, $\gamma = \frac{1}{n}$ suffices for us in corollary 5.4. The bound on the spectral gap of $\widehat{\mathcal{M}}$ then follows using our information about β, γ, the restrictions (theorem 6.9), and the projection chain (where $Gap(P_M) \geq \frac{c'}{n^2}$). $\square$

Finally, a simple application of the comparison theorem establishes a polynomial bound for the mixing time of the original simpler mountain/valley Markov chain $\mathcal{M}$ (from section 6.1) on $\mathcal{S}$.

Theorem 6.12 *The mountain/valley Markov chain $\mathcal{M}$ on $\mathcal{S}$ is rapidly mixing and* $\tau_{\mathcal{M}}(\varepsilon) = \mathrm{O}(\lambda^2 n^{19} \log(\frac{\lambda}{\varepsilon}))$.

Proof. The only interesting case to consider is a propeller move in $\widehat{\mathcal{M}}$. This move can be broken into a pair of inversions in $\mathcal{M}$. The propeller move occurs with probability $\frac{1}{4n}$, while each inversion happens with probability $\frac{\lambda}{(2n)2(1+\lambda)}$ or $\frac{1}{(2n)2(1+\lambda)}$. Since we consider the case with $\lambda > 1$ we have $A = 2(1+\lambda)$. From theorem 3.3(a) we have $Gap(\mathcal{M}) \geq \frac{C_4}{\lambda^3 n^{18}}$. Using theorem 3.1(a), observing that $\pi_* = \min_{x \in \mathcal{S}} \pi(x) \geq \frac{1}{4^n \lambda^n}$ so $\log(\frac{1}{\pi_* \varepsilon}) \leq n \log(\frac{4\lambda}{\varepsilon})$, we derive the stated bound on $\tau_{\mathcal{M}}(\varepsilon)$. □

Chapter 7

Concluding remarks

The FreeRoute algorithm described in section 2.2 allows us to sample free boundary tilings of some regions using a bijection between routings and tilings. One problem left to consider is finding a method for sampling free boundary domino tilings of square regions in $\mathbb{Z}^2$. Coming up with a digraph having compatible sources and sinks, and for which free routings correspond to free boundary tilings in a bijective manner, has proven an elusive task. This naturally leads to the more general open problem of finding exact methods for counting and sampling (fixed or free) routings for digraphs with sources and sinks that are not compatible. One possible avenue to explore is to determine if a sum and difference of determinants (Pfaffians) of several matrices can be used to count the number of fixed routings (respectively, free routings) in digraphs. Determining methods to approximately count or sample routings in these cases would also be appropriate if exact methods prove too difficult to find.

We have seen how Markov chain decomposition offers a method for proving rapid mixing for some Markov chains. We believe the new approach presented here that utilizes a partition of the state space is more natural and lends itself more easily to application than does the previous state decomposition theorem of Madras and Randall.

As we have seen, the essential feature is that instead of analyzing a complicated state space, we consider restrictions to smaller pieces of the state space on which typically the probability distribution is uniform. Showing these restricted chains (or

slight variants of them) are rapidly mixing can be more straightforward. For example, since the effect of differing stationary probabilities can be dampened (assuming the distribution is uniform on a subset), path coupling might now suffice to show rapid mixing of the restricted chains. We saw this was true when we considered the restricted chains on each of the small component sets $\mathcal{S}_{k,T}$ in section 6.3.6 where we were able to use path coupling in that case.

Moreover, we feel that the projection chains that arise from decomposition will often be random walks on an interval $[n]$ or an integer simplex. These types of Markov processes are typically well–studied or, at the very least, are problems that have a more well–defined structure on the state space that might help to analyze these chains more easily than chains on complicated state space. Chapter 4 on exclusion processes presents only a pair of examples that might arise in projections chains; we believe the methods presented there should readily extend to some other types of exclusion processes, for example one where n sites are arranged in a circular pattern.

A further example where the disjoint decomposition method has proven useful is sampling circuits in $\mathbb{Z}^d$. We may represent a circuit as a string using the elements from the set $\{a_1, a_2, \ldots, a_d\} \cup \{a_1^{-1}, a_2^{-1}, \ldots, a_d^{-1}\}$, where a_i represents a positive step in the ith coordinate direction and a_i^{-1} a corresponding negative step. In [45] the authors define a Markov chain whose transitions involve transposing adjacent elements in the string, and occasionally changing an adjacent pair $a_i a_i^{-1}$ to an $a_j a_j^{-1}$ pair. The state space is the set of all circuits of length $2n$, and is decomposed into sets where the number of each symbol is fixed. This means the restricted Markov chains involve only transpositions of elements in the string. Interestingly enough, for this partition, the projection chain is a random walk on an integer simplex. Using this disjoint decomposition of the state space, they are able to show an upper bound of $\mathrm{O}(dn^{12} \log^3 n \log \varepsilon^{-1})$ for the mixing time of the Markov chain for sampling circuits.

As we stated previously, using decomposition will typically not provide optimal

bounds for the mixing time of Markov chains. In our application to staircase walks, we did not attempt to optimize the mixing time for the restricted chains or the projections. A finer analysis of some of these chains could produce a better overall result for the mountain/valley Markov chain $\mathcal{M}$ of section 6.1. Directly proving the nearest–neighbor Markov chain for the second projection in section 6.3.7 is rapidly mixing should also improve the bound on the mixing time of $\mathcal{M}$. More generally, another open question pertains to examining the exclusion processes in chapter 4, and determining other classes of sequences for which there will be similar results of rapid mixing.

As a final remark, we want to note that Madras and Zheng [41] have recently presented two more applications of decomposition, some of that work along similar lines to that given here. Using the theorem of Caracciolo, Pelissetto, and Sokal (theorem 5.2 in this paper), results on product Markov chains, and decomposition of the state space, they show that the so–called swapping algorithm is rapidly mixing when sampling from a bimodal distribution on an interval $[n]$. They also examine the low temperature mean–field Ising model and prove the swapping algorithm is rapidly mixing in this case as well. The interested reader is referred to [41] for further details on this interesting work.

Bibliography

[1] D. Aldous. Random walks of finite groups and rapidly mixing Markov chains. *Séminaire de Probabilités XVII. Lecture Notes in Mathematics* **986** (1983), Springer, New York, pp. 243–297.

[2] D. Aldous and P. Diaconis. Shuffling cards and stopping times. *American Mathematical Monthly* **93** (1986), pp. 333–348.

[3] R.J. Baxter. *Exactly Solved Models in Statistical Mechanics.* Academic Press, New York, 1982.

[4] D. Bayer and P. Diaconis. Trailing the dovetail shuffle to its lair. *Annals of Applied Probability* **2** (1992), pp. 294–313.

[5] R. Bubley and M. Dyer. Faster random generation of linear extensions. *Discrete Mathematics* **201** (1999), pp. 81–88.

[6] R. Bubley and M. Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. *Proc. 38th Annual IEEE Symposium on Foundations of Computer Science* (1997), pp. 223–231.

[7] C.J. Colbourn, W.J. Myrvold, and E. Neufeld. Two algorithms for unranking arborescences. *J. Algorithms* **20** (1996), pp. 268–281.

[8] J.B. Conway. *A Course in Functional Analysis*, 2nd ed. Springer-Verlag, New York, 1990.

[9] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov chains. *Annals of Applied Probability* **3** (1993), pp. 696–730.

[10] R. Durrett. *Lecture Notes on Particle Systems and Percolation.* Wadsworth & Brooks, Belmont, 1988.

[11] M. Dyer and C. Greenhill. A more rapidly mixing Markov chain for graph colorings. *Random Structures and Algorithms* **13** (1998), pp. 285–317.

[12] N. Elkies, G. Kuperberg, M. Larsen, and J. Propp. Alternating sign matrices and domino tilings. *J. Algebraic Combinatorics* **1** (1992), pp. 111–132, 219–234.

[13] R. Epstein. *The Theory of Gambling and Statistical Logic*, revised ed. Academic Press, New York, 1977.

[14] M.E. Fisher. Statistical mechanics of dimers on a plane lattice. *Physics Review* **124** (1961), pp. 1664–1672.

[15] I.M. Gessel and X.G. Viennot. Binomial determinants, paths, and hook length formulae. *Advances in Mathematics* **58** (1985), pp. 300–321.

[16] I.M. Gessel and X.G. Viennot. Determinants, paths, and plane partitions. Preprint (1989).

[17] C.J. Geyer and E.A. Thompson. Annealing Markov Chain Monte Carlo with applications to ancestral inference. *J. American Statistical Association* **90** (1995), pp. 909–920.

[18] J.G. Gibbs. *Elementary Principles in Statistical Mechanics.* Yale University Press, New Haven, 1902.

[19] G. Grimmett. *Percolation, 2nd edition.* Springer-Verlag, New York, 1999.

[20] R.A. Horn and C.R. Johnson. *Matrix Analysis.* Cambridge University Press, New York, 1985.

[21] M.R. Jerrum. 2–dimensional monomer–dimer systems are computationally intractable. *J. Statistical Physics* **48** (1987), pp. 121–134.

[22] M.R. Jerrum. A very simple algorithm for estimating the number of k-colorings of a low degree graph. *Random Structures and Algorithms* **7** (1995), pp. 157–165.

[23] M.R. Jerrum and A.J. Sinclair. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation* **82** (1989), pp. 93–133.

[24] M.R. Jerrum and A.J. Sinclair. Approximating the permanent. *SIAM J. Computing* **18** (1989), pp. 1149–1178.

[25] M.R. Jerrum, A.J. Sinclair, and E. Vigoda. A polynomial–time approximation algorithm for the permanent of a matrix with non–negative entries. *Electronic Colloquium on Computational Complexity*, **http://www.eccc.uni-trier.de/eccc/**, Report TR00–079 (2000), ISSN 1433–8092, 7th year, 79th report.

[26] M.R. Jerrum, L.G. Valiant, and V.V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science* **43** (1986), pp. 169–188.

[27] W. Jockusch, J. Propp, and P. Shor. Random domino tilings and the arctic circle theorem. Preprint (1995).

[28] P.W. Kasteleyn. The statistics of dimers on a lattice I. The number of dimer arrangements on a quadratic lattice. *Physica* **27** (1961), pp. 1209–1225.

[29] P.W. Kasteleyn. Dimer statistics and phase transitions. *J. Mathematical Physics* **4** (1963), pp. 287–293.

[30] P.W. Kasteleyn. Graph theory and crystal physics. In *Graph Theory and Theoretical Physics* (F. Harary, ed.), Academic Press, London, 1967, pp. 43–110.

[31] C. Kenyon, D. Randall, and A.J. Sinclair. Approximating the number of monomer–dimer coverings of a lattice. *J. Statistical Physics* **83** (1996), pp. 637–659.

[32] R. Kenyon. Local statistics of dimers on a lattice. *Annales de l'Institut Henri Poincaré – Probabilités et Statistiques* **33** (1997), pp. 591–618.

[33] V.G. Kulkarni. *Modeling and Analysis of Stochastic Systems.* Chapman & Hall, New York, 1995.

[34] T.M. Liggett. *Interacting Particle Systems.* Springer-Verlag, New York, 1985.

[35] B. Lindström. On the vector representations of induced matroids. *Bulletin of the London Mathematical Society* **5** (1973), pp. 85–90.

[36] M. Luby, D. Randall, and A.J. Sinclair. Markov chain algorithms for planar lattice structures. *Proc. 36th Annual IEEE Symposium on Foundations of Computer Science* (1995), pp. 150–159.

[37] M. Luby and E. Vigoda. Approximately counting up to four. *Proc. 29th Annual IEEE Symposium on Theory of Computation (STOC)* (1997), ACM Press, pp. 682–687.

[38] N. Madras and D. Randall. Factoring graphs to bound mixing rates. *Proc. 37th Annual IEEE Symposium on Foundations of Computer Science* (1996), pp. 194–203.

[39] N. Madras and D. Randall. Markov chain decomposition for convergence rate analysis. *Annals of Applied Probability*, (to appear).

[40] N. Madras and G. Slade. *The Self-Avoiding Walk.* Birkhäuser, Boston, 1996.

[41] N. Madras and Z. Zheng. On the Swapping Algorithm. Preprint (2001).

[42] E. Marinari and G. Parisi. Simulated tempering: a new Monte Carlo scheme. *Europhysics Letters* **19** (1992), pp. 451–458.

[43] R.A. Martin and D. Randall. Pfaffian algorithms for sampling routings on regions with free boundary conditions. *3rd International Workshop on Randomization and Approximation Techniques in Computer Science* in *Lecture Notes in Computer Science* **1671** (1999), pp. 257–268.

[44] R.A. Martin and D. Randall. Sampling adsorbing staircase walks using a new Markov chain decomposition method. *Proc. 41st Annual IEEE Symposium on Foundations of Computer Science* (2000), pp. 492–502.

[45] R.A. Martin and D. Randall. Disjoint decomposition and sampling circuits in Cayley graphs. Preprint (2001).

[46] B.M. McCoy and T.T. Wu. *The Two-Dimensional Ising Model.* Harvard University Press, Cambridge, 1973.

[47] L. McShine and P. Tetali. On the mixing time of the triangulation walk and other Catalan structures. *Randomization methods in algorithm design*, *DIMACS Series on Discrete Mathematics and Theoretical Computer Science* **43** (1999), pp. 147–160.

[48] M.E.J. Newman and G.T. Barkema. *Monte Carlo Methods in Statistical Physics.* Oxford University Press, New York, 1999.

[49] L. Onsager. Crystal statistics. I.A two-dimensional model with an order–disorder transition. *Physical Review (2)* **65** (1944), pp. 117–149.

[50] J. Propp and J. Woolever.
http://www.math.wisc.edu/∼propp/software.html

[51] D. Randall and P. Tetali. Analyzing Glauber dynamics by comparison of Markov chains. *Journal of Mathematical Physics* **41** (2000), pp. 1598–1615.

[52] B. van Rensburg. Adsorbing staircase walks and staircase polygons. *Annals of Combinatorics* **3** (1999), pp. 451–473.

[53] E.J.J. van Rensburg. Collapsing and adsorbing polygons. *Journal of Physics A: Mathematical and General* **31** (1998), pp. 8295–8306.

[54] E.J.J. van Rensburg. *The Statistical Mechanics of Interacting Walks, Polygons, Animals, and Vesicles.* Oxford University Press, New York, 2000.

[55] J.S. Rosenthal. Convergence rates for Markov chains. *SIAM Review* **37** (1995), pp. 387–405.

[56] J. Sherman and W. J. Morrison. Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix. *The Annals of Mathematical Statistics* **20** (1949), p. 621.

[57] A.J. Sinclair. *Algorithms for Random Generation & Counting: a Markov Chain Approach.* Birkhäuser, Boston, 1993.

[58] R. Stanley. *Enumerative Combinatorics*, Vol. II. Cambridge University Press, New York, 1999.

[59] J.R. Stembridge. Non-intersecting paths, Pfaffians, and plane partitions. *Advances in Mathematics* **83** (1990), pp. 96–131.

[60] H.N.V. Temperley and M.E. Fisher. Dimer problem in statistical mechanics–an exact reult. *Philosophical Magazine* **6** (1961), pp. 1061–1063.

[61] L.G. Valiant. The complexity of computing the permanent. *Theoretical Computing Science* **8** (1979), pp. 189–201.

[62] L.G. Valiant. The complexity of enumeration and reliability problems. *SIAM J. Computing* **8** (1979), pp. 410–421.

[63] D.B. Wilson. Determinant algorithms for random planar structures. *Proc. 8th Annual ACM-SIAM Symposium on Discrete Algorithms* (1997), pp. 258–267.

[64] D.B. Wilson. Mixing times of lozenge tiling and card shuffling Markov chains. Preprint (1997).

Vita

Russell Martin thrust himself upon the world in July 1970 in Chattanooga, Tennessee. Growing up there, and later in Wappingers Falls, New York, his youth was somewhat uneventful, characterized by his disregard for social cliques and general cluelessness (which, some argue, continues to this day).

Russell obtained his Bachelor of Science degree in mathematics from Syracuse University in May 1992, and his Master of Science degree from Clemson University in August 1994. After a brief period in the "real world", he started attending the Georgia Institute of Technology to work towards his PhD. Surviving the upheaval of Atlanta due to the Centennial Olympic Games, after many opportunities to make a fool of himself (many of them accomplished successfully), Russell graduated in December 2001 with his Doctor of Philosophy degree in mathematics. He quietly continues the quest to become Master of Time and Space, and his search for the perfect donut.