Eigenfunctions for Random Walks on Hyperplane Arrangements

by

John Pike

A Dissertation Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(APPLIED MATHEMATICS)

August 2013

Acknowledgements

I would like to extend my sincerest thanks to my advisor, Jason Fulman, who introduced me to many of my favorite areas of mathematics and has been a constant source of guidance and encouragement over these past several years. His knowledge and taste in mathematics has had a profound influence on my academic development and my work has benefited greatly from our many conversations. I could not have been more fortunate in my choice of advisor and it has been a true pleasure working with him.

I would also like to thank Larry Goldstein for all that he has taught me about the content and practice of mathematics. He has been extremely generous with his time and knowledge, and I always walk away from our discussions with a deeper understanding and appreciation of fundamental issues.

I have had nothing but positive experiences with all of the professors at USC, and have learned a great deal during my time here. I do not have the space to thank everyone properly, but I am especially grateful to Peter Baxendale for his help, kindness, and many informative conversations, to Paul Newton and Nicolai Haydn for taking the time to absorb a subject outside of their primary research interests in order to offer valuable contributions as members of my dissertation and oral exam committees, and to Cymra Haskell for her tireless efforts to help all of the math TAs excel in their teaching duties. Finally, I have had a great time discussing mathematics of all sorts with Sergey Lototsky in the student probability seminar and will miss our lively discourses.

Thanks also to my mathematical compatriots Haining Ren, Umit Islak, and Radoslav Marinov. I have profited immensely from all the hours spent learning new subjects together, and have thoroughly enjoyed getting to know each of them. I look forward to further collaborations in the future.

My parents have been wonderful role models throughout my life and, in addition to teaching me to be a decent person in general, helped to instill the importance of creativity, curiosity, open-mindedness, critical thought, and a demand for logical and empirical rigor. I thank them deeply for helping me become the person I am today and for their unwavering patience, love, and support.

Finally, I cannot sufficiently express my appreciation for my girlfriend, Arie Armstrong, who has been my constant companion for nearly a decade and counting. She is a perpetual source of happiness in my life and her love and devotion have helped make all of this possible. This dissertation is for her.

Table of Contents

Acknowledgements ii

Abstract vi

Chapter 0 Introduction 1

Chapter 1 Markov Chains 5
- 1.1 Basic Concepts . . . 5
- 1.2 Mixing . . . 10
- 1.3 Products and Projections of Markov Chains . . . 21
  - 1.3.1 Projections . . . 22
  - 1.3.2 Products . . . 30

Chapter 2 Random Walks on Hyperplane Arrangements 40
- 2.1 Hyperplane Arrangements . . . 40
- 2.2 Hyperplane Walks . . . 48
- 2.3 Examples . . . 64
  - 2.3.1 Boolean Arrangement . . . 65
  - 2.3.2 Braid Arrangement . . . 73
  - 2.3.3 Dihedral Arrangement . . . 83
- 2.4 Extensions . . . 86
  - 2.4.1 Rows of Colored Tiles . . . 86
  - 2.4.2 Oriented Matroids and Left-Regular Bands . . . 90

Chapter 3 Eigenfunctions 94
- 3.1 Eigenfunctions for Hyperplane Walks . . . 94
- 3.2 Examples . . . 103
- 3.3 Applications . . . 119
  - 3.3.1 Mixing Time Bounds . . . 122
    - 3.3.1.1 Wilson's Method and Lower Bounds . . . 122
    - 3.3.1.2 Stochastic Monotonicity and Upper Bounds . . . 134
  - 3.3.2 Connections with Stein's Method . . . 142
    - 3.3.2.1 Using Stein's Method to Study Eigenfunctions . . . 143
    - 3.3.2.2 Eigenfunctions for Random Walks on S_n . . . 148

Chapter 4 Concluding Remarks 161

Bibliography 169

Abstract

A large class of seemingly disparate Markov chains can be modeled as random walks on the chambers of hyperplane arrangements. Examples include models from computer science, statistical mechanics, military campaigns, and card shuffling, as well as many natural random walks on finite reflection groups. A remarkable feature of these chains is that the general theory is fairly complete. For instance, it has been established that the corresponding transition matrices are diagonalizable over the reals with explicitly computable nonnegative eigenvalues of known multiplicity. Moreover, a description of the stationary distributions and a criterion for ergodicity are known, and there are simple upper bounds on the rates of convergence.

The present work continues these investigations by providing an explicit construction of the right eigenfunctions corresponding to the largest eigenvalues and a general prescription for finding others. For certain important classes of chamber walks, we are able to provide a basis for the eigenspace corresponding to the subdominant eigenvalue and we show that several interesting statistics arise in this fashion. In addition, we demonstrate how these eigenfunctions can be used to obtain upper and lower bounds on the variation distance to stationarity. We also discuss connections with Stein's method, including as an aside a derivation of some of the eigenfunctions for certain random walks on the symmetric group which do not fit into the hyperplane paradigm. Along the way, we give generalizations and alternate proofs of some known results and introduce a novel combinatorial description of hyperplane walks which we have found to be quite useful.

Chapter 0

Introduction

This thesis has been written for a very general audience and thus contains a good deal of background material which can be skipped by those already familiar with various aspects of the subject.

We begin in Chapter 1 with an overview of (finite state space, discrete time) Markov chain theory and much of the first chapter will be routine to practitioners of the subject. The first section reviews basic concepts such as ergodicity and reversibility, and the second addresses ideas related to the study of mixing times. No new results appear in either of these sections and they can be skimmed over quickly by anyone with a background in probability. The third section is concerned with lumped chains and product chains, with many of the basic ideas coming from [48]. Markov chain lumping plays a central role in the derivation of eigenfunctions for hyperplane chamber walks, and ideas involving product chains are also used in several examples. Though the most important results about lumped Markov chains for purposes of the present work are classical, we also introduce a new construction which allows one to recover left eigenfunctions from associated chains on smaller state spaces in this section. These results do not apply to hyperplane walks

in general, but we show that they can be used to study random walks on finite groups. We also extend some known results concerning product chains to allow for more general transition dynamics.

In Chapter 2 we move on to random walks on hyperplane arrangements. The first section mostly covers standard facts about hyperplane arrangements, but we also introduce a new combinatorial description of faces and their products which provides a nice way of visualizing hyperplane chamber walks and generalizes to an intermediate position between random walks on oriented matroids and left-regular bands. Briefly, we associate each sign sequence corresponding to a face in the arrangement with a row of colored tiles and explain how the product of two faces can be visualized in terms of stacking. Once one knows the rows of tiles corresponding to the initial chamber and the support of the face measure (which depend on the underlying arrangement), the hyperplane description can be abstracted away entirely. In addition to providing a convenient visual aid and a generalization of hyperplane chamber walks, this perspective captures many of the essential features of these chains with regard to mixing behavior and simplifies some of the arguments contained herein. In the second section, we survey the known results concerning hyperplane chamber walks and include some original proofs as well. We also derive an upper bound on the variation distance which is equivalent to that obtained in [8] by Möbius inversion, but is more computationally tractable. The third section contains examples of various chamber walks to motivate the subject and provide concrete examples to illustrate the results presented so far. Most of these examples can be found in [8, 12], though we also present some original models of conquering territories, provide an alternative analysis of random-to-top shuffles in terms of product chains, and suggest

some other models related to coupon collecting, voting, and random walks on Young tableaux. We conclude the second chapter with a brief discussion of generalizations to random walks on rows of colored tiles, oriented matroids, and left-regular bands.

Chapter 3 forms the core of this thesis and readers who are already acquainted with the material in [12, 8, 7] and are primarily interested in results concerning eigenfunctions could probably begin here without missing very much. We begin in the first section by showing that chamber walks are lumpable with respect to equivalence relations that induce random walks on the chambers of subarrangements. This is equivalent to a result in [5] concerning functions of such Markov chains. It then follows that we can recover eigenfunctions for the original chain from those of the lumped chain, which enables us in particular to compute the eigenfunctions corresponding to eigenvalues indexed by hyperplanes. Under standard assumptions on the face measure, we show that the eigenfunctions so constructed are linearly independent. In the second section, we provide explicit computations for a variety of hyperplane chamber walks and show that a large class of walks on the chambers of the braid arrangement have a common eigenspace for the subdominant eigenvalue for which we can compute a convenient basis. The third section begins with a discussion of possible uses of Markov chain eigenfunctions and some examples which include the expectation of various statistics after any number of steps in several standard Markov chains. We then recall the framework using Wilson's method which allows one to compute lower bounds on the variation distance to stationarity in terms of eigenfunctions, and we provide a general variance bound for eigenfunctions for walks on the braid arrangement which is useful for determining good choices of distinguishing statistics both in terms of optimality and computation simplicity. To exemplify the technique, we establish a lower

bound for shuffles which proceed by choosing m cards at random and then moving them to the top of the deck while retaining their original relative order. These shuffles were studied previously in [24] and [8], and our lower bound matches the upper bound up to a factor of 4. To the best of the author's knowledge, this is the first lower bound given for these shuffles, though we conjecture that the known upper bound gives the correct mixing time. Next, we establish that hyperplane chamber walks are stochastically monotone with respect to a variety of partial orders related to the weak Bruhat order and use techniques from [31, 47] to find upper bounds on variation distance in certain cases. These upper bounds only provide a technical improvement on the known bounds, but serve as a nice demonstration of the applicability of this relatively novel method in Markov chain theory. We conclude the third chapter with a discussion of connections between Stein's method and Markov chain eigenfunctions, and as an aside, we use ideas from representation theory and the combinatorial central limit theorem to construct an explicit eigenbasis for a typically large eigenvalue of any random walk on the symmetric group which is driven by a measure that is constant on conjugacy classes.

In the final chapter, we provide a brief overview of our results and their significance, as well as a discussion of possible directions for future research.

Chapter 1

Markov Chains

1.1 Basic Concepts

Given a measurable space $(S, \mathcal{S})$, a sequence of S-valued random variables $\{X_k\}_{k=0}^{\infty}$ is said to be a Markov chain on S with respect to a filtration $\{\mathcal{F}_k\}$ if $X_k \in \mathcal{F}_k$ and

$$\mathbb{P}\{X_{k+1} \in B | \mathcal{F}_k\} = \mathbb{P}\{X_{k+1} \in B | X_k\}$$

for all $n \in \mathbb{N}_0$, $B \in \mathcal{S}$ [34]. S is called the state space of the Markov chain and in all examples considered here we assume that $|S| = N < \infty$, $\mathcal{S} = 2^S$. Also, we will typically take $\mathcal{F}_k = \sigma(X_0, X_1, ..., X_k)$, so the interpretation is that $\{X_k\}_{k=0}^{\infty}$ is a random S-valued process which depends on the past only through the present. We will restrict out attention to temporally homogeneous Markov chains so that the transition mechanism can be expressed as the $N \times N$ stochastic matrix P, indexed by the elements of S, given by

$$P(s,t) = \mathbb{P}\{X_{k+1} = t | X_k = s\},$$

independently of n. Taking rth powers gives the r-step transition matrix $P^r = P \cdot P^{r-1}$ with entries $P^r(s,t) = \mathbb{P}\{X_{k+r} = t | X_k = s\}$. Thus the distribution of X_r given that the chain began at $X_0 = s$ is given by $P_s^r(\cdot) = P^r(s, \cdot)$. If the chain is initially distributed as $X_0 \sim \mu$, then the distribution after r steps is written as $X_r \sim P_\mu^r$. (By a slight abuse of notation, P_s^r represents the case when the initial distribution is the point mass at s.) We will often find it useful to think of P as an operator which acts on functions $f : S \to \mathbb{C}$ by

$$Pf(s) = \sum_{t \in S} P(s,t) f(t)$$

and acts on probability measures by

$$\mu P(t) = \sum_{s \in S} \mu(s) P(s,t).$$

A useful construct in the study of temporally homogeneous, finite state space Markov chains is that of a random mapping representation of the transition operator. This is a function $f : S \times \Lambda \to S$, along with a Λ-valued random variable Z, satisfying

$$\mathbb{P}\{f(s,Z) = t\} = P(s,t).$$

The corresponding Markov chain is obtained by successive applications of the random mapping: $X_{k+1} = f(X_k, Z_{k+1})$ where $Z, Z_1, Z_2, ...$ is an i.i.d. sequence. If P is a transition matrix on a finite state space $S = \{s_1, s_2, ..., s_N\}$, then we can define

$$F_s(0) = 0, \; F_s(m) = \sum_{i=1}^{m} P(s, s_i)$$

for each $s \in S$, $m \in [N]$. Letting $Z \sim U(0,1)$, and setting

$$f(s,z) = \sum_{i=1}^{N} s_i 1_{(F_s(i-1),F_s(i)]}(z) \text{ where } 1_{(a,b]}(z) = \begin{cases} 1, & z \in (a,b] \\ 0, & else \end{cases},$$

we see that

$$\mathbb{P}\{f(s,Z) = s_i\} = \mathbb{P}\{F_s(i-1) < Z \leq F_s(i)\} = P(s,s_i),$$

hence every transition matrix on a finite state space has a random mapping representation [48]. It should be noted that random mapping representations are not unique and in many cases there are much more natural representations than the one used above to demonstrate existence. For example, simple random walk on a p-cycle has transition probabilities

$$\mathbb{P}\{X_{k+1} = j | X_k = i\} = \begin{cases} \frac{1}{2}, & j = i \pm 1 (\text{mod} p) \\ 0, & else \end{cases}$$

and random mapping representation (f, Z) where $Z = 2W - 1$, $W \sim Bernoulli(\frac{1}{2})$, and $f(s,z) = s + z(\text{mod}p)$.

One of the more interesting features of many Markov chains is that the k-step distributions tend towards an equilibrium as k grows. We say that a probability measure π is a stationary distribution for P if it satisfies

$$\pi(t) = \sum_{s \in S} \pi(s) P(s,t)$$

for all $t \in S$ - that is, if π is a left eigenvector of P with corresponding eigenvalue 1. For general finite state space Markov chains, convergence to a stationary distribution depends on the properties of irreducibility and aperiodicity. (When dealing with countably infinite state space Markov chains, one needs the additional assumption of positive recurrence.)

A Markov chain is said to be irreducible if for all $s, t \in S$, there exists some $k = k(s,t)$ such that $P^k(s,t) > 0$. The period of a state s is defined as

$$d_s = \gcd\{i \in \mathbb{N} : P^i(s,s) > 0\},$$

and a Markov chain is called aperiodic if $d_s = 1$ for all $s \in S$. If the chain is irreducible, then it is aperiodic if $d_s = 1$ for any $s \in S$. A Markov chain is both irreducible and aperiodic if and only if there is some $K \in \mathbb{N}$ such that $P^k(s,t) > 0$ for all $s, t \in S$, $k \geq K$, so it follows from the Perron-Frobenius theorem for positive matrices and the fact that P is stochastic that the eigenvalues satisfy $1 = \lambda_1 > |\lambda_i|$ for $i = 2, ..., N$. (The Perron-Frobenius Theorem for nonnegative matrices shows that if P is irreducible, then 1 is a simple eigenvalue and there are corresponding left and right eigenvectors with strictly positive entries.) These observations imply the fundamental theorem of finite state space Markov chains: An irreducible Markov chain has a unique and strictly positive stationary distribution π, and if the chain is aperiodic as well, then

$$\lim_{n \to \infty} P^n(s,t) = \pi(t)$$

for all $s, t \in S$. If a Markov chain is both irreducible and aperiodic, we will call it regular. If the chain converges to a unique stationary distribution, independent of the initial state, then we say that is ergodic. The above result can thus be stated as "Every regular finite state space Markov chain is ergodic."

Also, we say that P is reversible with respect to a probability measure π if it satisfies the detailed balance equations

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

for all $x, y \in S$. Reversibility is a sufficient, but not necessary, condition for the existence of a stationary distribution. If we let

$$\langle f, g \rangle = \sum_{s \in S} f(s)g(s)$$

be the usual inner product on $\mathbb{R}^S$ and define the new inner product

$$\langle f, g \rangle_\pi = \sum_{s \in S} f(s)g(s)\pi(s),$$

then reversibility implies that P is a self-adjoint operator with respect to the inner product $\langle \cdot, \cdot \rangle_\pi$, so it follows from the spectral theorem that P has real eigenvalues and a basis of eigenvectors which are orthonormal with respect to $\langle \cdot, \cdot \rangle_\pi$. This fact gives rise to many useful theorems concerning the convergence rates of reversible Markov chains. Even if an irreducible Markov chain is not reversible, we can define its time reversal in terms of the transition matrix

$$P^*(s, t) = \frac{\pi(t)}{\pi(s)} P(t, s).$$

Thus P^* is the adjoint of P with respect to the inner product $\langle \cdot, \cdot \rangle_\pi$, and it is not difficult to verify that if $\{X_k\}$ is a Markov chain with transition matrix P and positive stationary distribution π, then, P^* is the transition matrix of a Markov chain $\{Y_k\}$ which has stationary distribution π and satisfies

$$\mathbb{P}_\pi\{X_0 = x_0, X_1 = x_1, ..., X_T = x_T\} = \mathbb{P}_\pi\{Y_0 = x_T, Y_1 = x_{T-1}, ..., Y_T = x_0\}$$

for all $T \in \mathbb{N}$, $x_0, ...x_T \in S$ [48]. In many cases one can use facts about reversible chains to obtain results for a nonreversible chain by considering the multiplicative reversibilization of P given by the product PP^*, which is a stochastic matrix with stationary distribution

π [36]. (A similar construct which is useful for continuous time Markov processes is the additive reversibilization $\frac{P+P^*}{2}$ [50].) Though most chains considered here will not be reversible, it is good to keep this definition in mind for the sake of comparison as we will be dealing primarily with chains having real eigenvalues and linearly independent eigenvectors. Also, we will be considering time reversals in several examples.

1.2 Mixing

The most common question about ergodic Markov chains is how long will it take a given chain with initial distribution μ to get close to its equilibrium distribution π. In order to make this definition rigorous, we must introduce a measure of distance between probability measures. We will primarily be concerned with the total variation metric defined by

$$\|\mu - \nu\|_{TV} = \max_{A \subseteq S} |\mu(A) - \nu(A)| .$$

Thus the total variation distance measures the largest disagreement between two probability measures concerning the likelihood of an event. One can show that equivalent (and often useful) definitions are [21]

$$\frac{1}{2} \sum_{s \in S} |\mu(s) - \nu(S)| = \|\mu - \nu\|_{TV} = \frac{1}{2} \sup_{\|f\|_\infty \leq 1} |\mu(f) - \nu(f)|$$

where $\mu(f) = \sum_{s \in S} \mu(s) f(s)$.

Another useful characterization of total variation distance involves coupling. A coupling of two probability measures P_1 and P_2 defined on a finite set S is a probability Q

defined on $S \times S$ with margins P_1 and P_2. Denoting the diagonal by $\Delta = \{(s,s) : s \in S\}$, we can define

$$\|P_1 - P_2\|_{TV} = \inf\{Q(\Delta^C) : Q \text{ is a coupling of } P_1 \text{ and } P_2\}.$$

Equivalently, a pair of random variables (X, Y) on $S \times S$ is said to be a coupling of P_1 and P_2 if $X \sim P_1$ and $Y \sim P_2$, and we can define

$$\|P_1 - P_2\|_{TV} = \inf\{\mathbb{P}\{X \neq Y\} : (X, Y) \text{ is a coupling of } P_1 \text{ and } P_2\}.$$

The infima are achieved in both cases. Now suppose that P is the transition matrix for a Markov chain on S with stationary distribution π. Let $\{X_k^1\}_{k=0}^{\infty}$ be a Markov chain with transition matrix P and initial distribution μ, and let $\{X_k^2\}_{k=0}^{\infty}$ be a Markov chain with transition matrix P and initial distribution π. Define the coupling time by $T = \min\{k \geq 0 : X_k^1 = X_k^2\}$. Then, setting

$$X_k^3 = \begin{cases} X_k^2, & T \leq k \\ X_k^1, & T > k \end{cases},$$

the pair (X_k^1, X_k^3) is a coupling of P_μ^k and π, so the coupling definition of total variation distance implies that

$$\left\| P_\mu^k - \pi \right\|_{TV} \leq \mathbb{P}\{T > k\}.$$

The general idea is that we start one copy of the chain in a specified distribution, let another copy begin in stationarity, and then make them evolve according to the same transition mechanism until they meet and proceed simultaneously forever after. As the

second chain was stationary to begin with, it remains so for all time, hence the first chain must have equilibriated by the time that they have coupled [21, 48].

A related (but distinct) construction is that of strong stationary times introduced in [3]: If T is a randomized stopping time for $\{X_n\}_{n=0}^{\infty}$ such that

$$\mathbb{P}\{X_k = s | T = k\} = \pi(s) \text{ for all } s \in S,\ 0 \leq k < \infty,$$

then T is called a strong stationary time. Equivalently, a strong stationary time T is a randomized stopping time such that $X_T \sim \pi$ and T is independent of X_T. (A randomized stopping time allows the decision of whether to stop at time k to depend not only on $\{X_0, ..., X_k\}$, but also on the values of other random quantities independent of $\{X_{k+1}, X_{k+2}, ...\}$.) If we drop the independence assumption and only require that T is a randomized stopping time with $\mathbb{P}\{X_T = s\} = \pi(s)$, then we say that T is a stationary time. It can be shown that if T is a strong stationary time, then

$$\left\| P_x^k - \pi \right\|_{TV} \leq \text{sep}(P_x^k, \pi) \leq P(T > k),$$

where sep denotes the separation distance defined in the following paragraph. Moreover, there is an optimal strong stationary time such that the second inequality is an equality, and this provides one way to define separation distance. Note that coupling times are not necessarily strong stationary times. However, Aldous and Diaconis have shown that coupling subsumes strong stationary times in the sense that if T is a strong stationary time for a Markov chain, then there exists a coupling with coupling time T [4]. Chapter 4 in [21] contains more information on coupling and strong stationary times, and Igor

Pak's PhD thesis [53] includes numerous clever applications of strong stationary times to the study of Markov chain convergence.

There are many other standard measures of distance between Markov chains, including the l^p distances, separation distance, Hellinger distance, and Kullback-Leibler divergence. We will only define the first two as they are the most relevant to the ensuing discussion. Given $1 \leq p < \infty$ and a probability measure π, we can define the $l^p(\pi)$ norm on $\mathbb{R}^S$ by

$$\|f\|_{\pi,p} = \left(\sum_{s \in S} |f(s)|^p \, \pi(s) \right)^{\frac{1}{p}}.$$

We define the l^∞ norm by

$$\|f\|_{\pi,\infty} = \|f\|_\infty = \max_{s \in S} |f(s)| \, .$$

When P_μ^k is the k-step distribution of an ergodic Markov chain with initial distribution μ and positive stationary measure π, these norms gives rise to the l^p distances between P_μ^k and π by setting

$$d_p(P_\mu^k, \pi) = \left\| \frac{P_\mu^k}{\pi} - 1 \right\|_{\pi,p}.$$

As expected, we have the inequality $d_p(P_\mu^k, \pi) \leq d_q(P_\mu^k, \pi)$ for all $1 \leq p \leq q \leq \infty$. The case $p = 1$ gives (twice) the total variation distance, and the case $p = \infty$ is related to the more commonly used separation distance,

$$\text{sep}(\mu, \nu) = d_\infty(\mu, \nu) = \max_{s \in S} \left\{ 1 - \frac{\mu(s)}{\nu(s)} \right\},$$

which satisfies

$$\|\mu - \nu\|_{TV} \leq \text{sep}(\mu, \nu).$$

Note that the l^p and separation distances are not actually metrics because of asymmetry. However, the separation distance, like the total variation metric, always takes values in $[0, 1]$ [50, 21].

Because we often want a measure of convergence which is independent of the initial distribution, we also define the maximal variation

$$\left\|P^k - \pi\right\|_{TV*} = \sup\{\left\|P_\mu^k - \pi\right\|_{TV} : \mu \text{ is a probability measure}\}.$$

It can be shown that this supremum is realized by a point mass, so we may simply write

$$\left\|P^k - \pi\right\|_{TV*} = \max_{s\in S} \left\|P_s^k - \pi\right\|_{TV}.$$

Similarly, we write

$$d_p^*(P^k, \pi) = \max_{s\in S} d_p(P_s^k, \pi), \quad \text{sep}^*(P^k, \pi) = \max_{s\in S} \text{sep}(P_s^k, \pi).$$

It is worth observing that $\text{sep}^*(P^k, \pi)$ and $d_p^*(P^k, \pi)$ are submultiplicative as functions of k in the sense that

$$d_p^*(P^{k+l}, \pi) \leq d_p^*(P^k, \pi) d_p^*(P^l, \pi).$$

Total variation is not submultiplicative, but we do have the inequality

$$\left\|P^t - \pi\right\|_{TV*} \leq 2^{m-1} \prod_{i=1}^{m} \left\|P^{t_i} - \pi\right\|_{TV*} \text{ for } t = \sum_{i=1}^{m} t_i.$$

Finally, it should be noted that if P is the kernel of an ergodic Markov chain, then all of the above distances to stationarity are nonincreasing in k [51, 50].

With these definitions at hand, we can ask quantitative questions about convergence rates. The quantity of interest here is the (total variation) mixing time defined as follows: Given $\epsilon > 0$, set

$$t_{mix}(\epsilon) = \min\{k \in \mathbb{N}_0 : \left\| P_s^k - \pi \right\|_{TV} < \epsilon\},$$

$$t^*_{mix}(\epsilon) = \min\{k \in \mathbb{N}_0 : \left\| P^k - \pi \right\|_{TV^*} < \epsilon\}.$$

We define the mixing time as $t_{mix} = t_{mix}(\frac{1}{4})$ and similarly for t^*_{mix}. Of course, these quantities can also be defined with respect to other distances in an analogous manner. Much of the current research in Markov chain theory is focused on estimating mixing times and the literature is replete with different techniques for doing so. However, one typically has to have a very good understanding of the particular chain in question in order to get accurate results, so the general techniques will only get you so far. We have already seen how to upper bound total variation distance, and thus total variation mixing time, by constructing couplings and strong stationary times. The bound from couplings used the fact that total variation can be defined in terms of an infimum over certain couplings, and the strong stationary bound arose because separation may be defined as an infimum over strong stationary times and separation upper-bounds total variation. Again, though these methods work in theory, one must usually have a detailed knowledge of the chain in question in order to construct couplings or strong stationary times which yield sharp estimates.

As with the strong stationary time bound, other inequalities between various distances allow one to bound a given distance in terms of bounds involving another, and different

definitions of distances involving maxima and minima may also be used to get bounds. For example, since total variation is defined as a maximum over subsets of the state space, this definition can be used to lower-bound the total variation distance in the sense that

$$\left|P_s^k(A) - \pi(A)\right| \leq \left\|P_s^k - \pi\right\|_{TV} \leq \left\|P^k - \pi\right\|_{TV^*}$$

for all $A \subseteq S$. Often one chooses $A = \{s \in S : f(s) > \alpha\}$ for some $\alpha \in \mathbb{R}$, $f \in \mathbb{R}^S$. This is known as the method of distinguishing statistics and f is generally chosen to be an eigenfunction corresponding to the largest nontrivial eigenvalue. The term $\left|P_s^k(A) - \pi(A)\right|$ may be estimated using Chebychev's inequality or more refined arguments such as appears in Proposition 7.8 in [48]. David Wilson's lower bound technique is related to this idea and provides a computationally simpler method of obtaining lower bounds via eigenfunctions [70]. These ideas are discussed further in Subsection 3.3.1. Other useful lower bounds involve representing the transition matrix as a weighted directed graph with vertices corresponding to the states and edge set $\{(s,t) \in S^2 : Q(s,t) = \pi(s)P(s,t) > 0\}$ where the edge (x,y) has weight $Q(x,y)$. One can then obtain lower bounds in terms of the degrees, diameters, and bottleneck ratios, respectively, of these transition graphs. The reader is referred to chapter 7 in [48] for more details.

Another common class of bounds involve the spectrum of the operator P. For instance, we have seen that the eigenvalues of the transition operator of a regular Markov chain satisfy $1 = \lambda_1 > |\lambda_i|$ for $i = 2, ..., N$. By considering the Jordan normal form of P, one

sees that the exponential rate of convergence is governed by second largest eigenvalue (in modulus)

$$\lambda_* = \max_{2 \le i \le N} \{|\lambda_i|\}.$$

Indeed, for any regular Markov chain, we have the bound [50]

$$\frac{1}{2}\lambda_*^k \le \left\| P^k - \pi \right\|_{TV^*}.$$

An easy proof of this fact follows from the characterization of total variation $\|\mu - \nu\|_{TV} = \frac{1}{2}\sup_{\|f\|_\infty \le 1} |\mu(f) - \nu(f)|$. Namely, letting λ_i be an eigenvalue with $|\lambda_i| = \lambda_*$ and φ an associated right eigenfunction normalized so that $\|\varphi\|_\infty = 1$, and observing that $\pi(\varphi) = 0$ since π is a left eigenfunction corresponding to $1 \neq \lambda_i$, we have

$$\begin{aligned} \left\| P_s^k - \pi \right\|_{TV} &= \frac{1}{2}\sup_{\|f\|_\infty \le 1} \left| P_s^k(f) - \pi(f) \right| \\ &\ge \frac{1}{2}\left| P_s^k(\varphi) - \pi(\varphi) \right| = \frac{1}{2}\left| \lambda_i^k \varphi(s) - 0 \right| = \frac{1}{2}\lambda_*^k \left| \varphi(s) \right|, \end{aligned}$$

and the result follows upon maximizing over $s \in S$.

If P is reversible as well, then it has an orthonormal set of eigenfunctions $\{\phi_i\}_{i=1}^N$ corresponding to real eigenvalues $1 = \lambda_1 > \lambda_2 \ge ... \ge \lambda_N > -1$ and a relatively straightforward computation gives the upper bound

$$4\left\| P_s^k - \pi \right\|_{TV}^2 \le \sum_{i=2}^N \lambda_1^{2k} \phi_i(s)^2 \le \lambda_*^{2k} \frac{1 - \pi(s)}{\pi(s)}.$$

If P is nonreversible, then we have the bound

$$4\left\| P_s^k - \pi \right\|_{TV}^2 \le \frac{\lambda_{PP^*}^k}{\pi(s)}$$

where λ_{PP*} is the largest nontrivial eigenvalue of PP^*, the multiplicative reversibilization of P [36]. Even when λ_* is not known, there are a whole host of fascinating techniques for estimating it, such as the canonical path bounds discussed in [23]. Nonetheless, though the first of the above upper bounds for reversible chains is often sharp, estimates involving only the subdominant eigenvalue do not typically suffice for accurate non-asymptotic results.

There are numerous other methods, drawing from a wide range of disciplines, for obtaining bounds on Markov chain convergence rates that have been successfully applied in several examples. Among the more prominent omitted from the above discussion are the isoperimetric bounds (including the method of evolving sets) and the Nash, log-Sobolev, and other functional analytic bounds. There is also an extensive literature on random walks on finite groups which includes comparison techniques and several remarkable representation theoretic arguments for obtaining bounds. For the sake of brevity, we refer the interested reader to [50, 63, 21, 2]. (For more connections between Markov chains and other areas of mathematics, the author also highly recommends the survey [28].)

Before moving on, however, a few words about random walks on groups are in order since some of the examples we consider can be modeled as such. Given a finite group G and a probability measure p with support $K \subseteq G$, we can construct a random walk on G by beginning at some initial state $X_0 = h$ and proceeding by choosing g_k from p and forming the product $X_k = g_k X_{k-1}$. The transition matrix is thus given by $P(x, y) = p(yx^{-1})$. It is elementary to show that the Markov chain so constructed is irreducible as long as K generates G and is ergodic if and only if K is not contained in a coset of a proper subgroup of G, in which case the stationary distribution is the uniform measure on G. Also, the

distance to stationarity is independent of the initial state [63]. Such random walks are reversible precisely when $p(g) = p(g^{-1})$ for all $g \in G$. When the walk is not reversible, we can construct the time-reversal by setting $p^*(g) = p(g^{-1})$ (so that $P^*(x, y) = p^*(yx^{-1})$). Letting $\{X_k^*\}$ denote the walk associated with P^*, a few simple calculations show that

$$P(X_k^* = g | X_0^* = h) = P(X_k = h | X_0 = g),$$

$$P(X_k^* = g^{-1} | X_0^* = id) = P(X_k = g | X_0 = id),$$

$$\left\| (P^*)^k - U \right\|_{TV^*} = \left\| P^k - U \right\|_{TV^*}.$$

These facts will be of interest later because certain hyperplane walks can be viewed as time-reversals of well-known random walks on the symmetric group.

One of the most intriguing discoveries in Markov chain theory is that many families of random walks exhibit a sharp phase transition in their convergence to stationarity known as the cutoff phenomenon. In the words of Persi Diaconis [25], "The distance $\left\| P_x^k - \pi \right\|_{TV}$ stays close to its maximum value at 1 for a while, then suddenly drops to a quite small value and then tends to zero exponentially fast." More formally, suppose that for a sequence of Markov chains $\{X_k^{(1)}\}, \{X_k^{(2)}\}, ...$ (with transition matrices $P^{(1)}, P^{(2)}, ...$ and stationary distributions $\pi^{(1)}, \pi^{(2)}, ...$), $\{X_k^{(n)}\}$ has $\epsilon-$mixing time (with respect to some distance)

$$t_{mix}^{(n)}(\epsilon) = \min\{k \in \mathbb{N}_0 : d^* \left((P^{(n)})^k, \pi^{(n)} \right) < \epsilon\}.$$

We say that this sequence exhibits cutoff (with respect to that distance) if

$$\lim_{n \to \infty} \frac{t_{mix}^{(n)}(\epsilon)}{t_{mix}^{(n)}(1 - \epsilon)} = 1$$

for all $\epsilon \in (0, \frac{1}{2})$. Of course, we always have

$$\frac{t_{mix}^{(n)}(\epsilon)}{t_{mix}^{(n)}(1-\epsilon)} \geq 1,$$

and if we only know that

$$\sup_{0<\epsilon<\frac{1}{2}} \limsup_{n\to\infty} \frac{t_{mix}^{(n)}(\epsilon)}{t_{mix}^{(n)}(1-\epsilon)} < \infty,$$

then we say that the sequence exhibits a pre-cutoff. Alternatively, if $\{t_n\}_{n=1}^{\infty}$ and $\{w_n\}_{n=1}^{\infty}$ are two sequences of nonnegative numbers such that $w_n \in_{n\to\infty} o(t_n))$ and

$$\lim_{c\to-\infty} \liminf_{n\to\infty} d^*\left((P^{(n)})^{\lfloor t_n+cw_n\rfloor}, \pi^{(n)}\right) = 1,$$

$$\lim_{c\to\infty} \limsup_{n\to\infty} d^*\left((P^{(n)})^{\lfloor t_n+cw_n\rfloor}, \pi^{(n)}\right) = 0,$$

then we say that the sequence exhibits a (t_n, w_n) cutoff. Typically, $t_n = t_{mix}^{(n)}$ and we call $\{w_n\}$ the cutoff window [33, 48]. The first example in which cutoff was observed was the random transposition walk analyzed by Diaconis and Shahshahani using the representation theory of the symmetric group [20], and the notion of cutoff first appeared explicitly in [3], though a rigorous definition was not introduced until [25].

Many natural Markov chains have been shown to exhibit cutoff, but the general phenomenon is still somewhat of a mystery. Persi Diaconis conjectures that in many cases cutoff results from high multiplicity of the subdominant eigenvalue [25]. At an AIM workshop held in Palo Alto in 2004, Yuval Peres noted that a necessary condition for a sequence of reversible Markov chains to exhibit total variation cutoff is that $\lim_{n\to\infty} t_{mix}^{(n)} \cdot \gamma_n = \infty$ where $\gamma_n = 1 - \lambda_*^{(n)}$ is the absolute spectral gap of $P^{(n)}$, and a slight modification of Peres' argument shows that this is true even for nonreversible chains. Peres conjectured

that this condition is also sufficient, but counterexamples have since been constructed, showing that this is not always the case [33, 15]. Still, it is likely that Peres' conjecture does hold under certain conditions. For example, it has been shown that if the transition operators of such a sequence of Markov chains are normal (in the sense that they commute with their adjoints), then in both discrete and continuous time, the sequence presents a cutoff with respect to the distance d_p^*, $1 < p < \infty$, if and only if $\lim_{n\to\infty} t_{mix}^{(n)} \cdot \gamma_n = \infty$ [15, 16]. Peres' conjecture has also been verified for continuous-time birth-and-death chains started at an endpoint with respect to separation distance [29] and for all continuous-time and lazy discrete-time birth-and-death chains with respect to total variation [33]. The cutoff phenomenon has been identified in certain realizations of hyperplane walks [25], but no general criterion for cutoff in hyperplane walks is known at present.

1.3 Products and Projections of Markov Chains

Two final concepts in Markov chain theory which we will find useful in our discussion of hyperplane walks are projections and products of Markov chains. In the case of projections, one studies induced Markov chains on equivalence classes of the state space of a given Markov chain, essentially restricting one's attention to the evolution of particular aspects of the original process. With products, one aggregates individual Markov chains to obtain a description of a larger system in terms of component processes. Both concepts involve a "big" chain and "little" chain(s), and the interest is in inferring properties about the "big" chain from those of the "little" chain(s) and vice versa. We will see that

random walks on hyperplane arrangements are particularly amenable to study in terms of product and quotient constructions.

1.3.1 Projections

We begin with projections or "lumpings" of Markov chains. The setup here is a Markov chain $\{X_k\}$ with finite state space S and transition matrix P, and an equivalence relation $\sim$ on S. Letting $[s]$ denote the equivalence class containing $s \in S$, we are interested in the behavior of $\{[X_k]\}$ regarded as a random process on $S^* = S/\sim$. The main results relating $\{X_k\}$ and $\{[X_k]\}$ are summarized in the following theorem.

Theorem 1.3.1. *Let P be the transition matrix of a Markov chain $\{X_k\}$ having finite state space S, and let $\sim$ be an equivalence relation on S. Set $P(r,[s]) = \sum_{t\sim s} P(r,t)$. If $P(r,[s]) = P(q,[s])$ for all $[s] \in S^*$ and all $r,q \in S$ with $r \sim q$, then $\{[X_k]\}$ is a Markov chain on S^* with transition matrix $P^{\#}([r],[s]) := P(r,[s])$. The k-step transition probabilities are given by $(P^{\#})^k([r],[s]) = \sum_{t\sim s} P^k(r,t)$ for all $k \in \mathbb{N}$. Moreover,*

1. *If ϕ is a right eigenfunction of P corresponding to the eigenvalue λ and $\phi(q) = \phi(r)$ for all $r,q \in S$ with $r \sim q$, then $\phi^{\#}([r]) := \phi(r)$ is a right eigenfunction of $P^{\#}$ with associated eigenvalue λ.*

2. *If φ is a right eigenfunction of $P^{\#}$ with eigenvalue β, then $\varphi^{\flat}(r) := \varphi([r])$ is a right eigenfunction of P with eigenvalue β.*

3. *If ψ is a left eigenfunction of P with eigenvalue α, then $\psi^{\#}([s]) := \sum_{t\sim s} \psi(t)$ is a left eigenfunction for $P^{\#}$ with eigenvalue α. In particular, if π is stationary for P, then $\pi^{\#}$ is stationary for $P^{\#}$.*

4. *$\{[X_k]\}$ is ergodic (respectively, regular, reversible) whenever $\{X_k\}$ is, and we have the total variation bound $\left\|P^k - \pi\right\|_{TV^*} \geq \left\|(P^{\#})^k - \pi^{\#}\right\|_{TV^*}$ for all $k \in \mathbb{N}$.*

5. *If $\phi_1, ..., \phi_k : S \to \mathbb{C}$ are linearly independent with $\phi_i(q) = \phi_i(r)$ for all $q \sim r$, $i = 1, ..., k$, then $\phi_1^{\#}, ..., \phi_k^{\#} : S^* \to \mathbb{C}$ are linearly independent as well. Similarly, if $\varphi_1, ..., \varphi_k : S^* \to \mathbb{C}$ are linearly independent, then so are $\varphi_1^{\flat}, ..., \varphi_k^{\flat} : S \to \mathbb{C}$.*

Most of Theorem 1.3.1 is classical and proofs may be found in [48]. The distance bounds follow from the extreme event characterization of total variation and the k-step transition probabilities can be deduced from obvious induction arguments. Inheritance of ergodicity/regularity/reversibility may be checked directly from the definitions. For property 5, note that if $\alpha_1\phi_1^{\#} + ... + \alpha_k\phi_k^{\#} \equiv 0$ is a nontrivial dependence relation, then so is $\alpha_1\phi_1 + ... + \alpha_k\phi_k \equiv 0$ and conversely. The remaining assertions are verified by straightforward computations.

Much of Theorem 1.3.1 can also be interpreted in terms of linear algebra. For example, writing $S = \{x_1, ..., x_n\}$, $S^* = \{[x_{s_1}], ..., [x_{s_m}]\}$, we can define R to be the $n \times m$ matrix with entries $R_{ij} = 1_{[x_{s_j}]}(x_i)$ and define Q to be the $m \times n$ matrix with entries $Q_{ij} = |[x_{s_i}]|^{-1} 1_{[x_{s_i}]}(x_j)$. A bit of matrix multiplication shows that QR is the $m \times m$ identity matrix and $P^{\#} = QPR$. The condition $P(r, [s]) = P(q, [s])$ for all $r, s \in S$, $q \sim r$ is equivalent to the statement $PR = RP^{\#} = RQPR$, which gives another inductive derivation of the k-step transitions for the lumped chain since

$$(P^{\#})^k = (QPR)^{k-1}QPR = [QP^{k-1}R]QPR$$

$$= QP^{k-1}(RQPR) = QP^{k-1}PR = QP^kR.$$

The preceding equations, along with the Cayley-Hamilton theorem, also provide an alternative proof that every eigenvalue of $P^{\#}$ is an eigenvalue of P. This is because if p is the characteristic polynomial of P, then $p(P^{\#}) = Qp(P)R = 0$. Thus if q is the minimal polynomial of $P^{\#}$, it must be the case that q divides p, so the set of roots of q (eigenvalues of $P^{\#}$) is contained in the set of roots of p (eigenvalues of P) [72].

The primary utility of the preceding theorem is to deduce spectral properties of the original chain from those of the lumped chains. The latter are typically easier to work with since the corresponding state spaces are smaller. Indeed, item 2 in Theorem 1.3.1 is the crux of our derivation of right eigenfunctions for random walks on the chambers of hyperplane arrangements (where the equivalence classes are defined in terms of projections onto various subarrangements).

One problem with this theorem is that one cannot recover left eigenfunctions of the original chain from those of the lumped chain. However, under certain assumptions on P and $\sim$, one can construct a Markov chain $\{Y_k\}$ on S^* (which is not equal to $\{[X_k]\}$ in general) from which this information may be gleaned. This is made explicit in the following theorem, which is in some sense dual to Theorem 1.3.1.

Theorem 1.3.2. *Let P be the transition matrix of a Markov chain $\{X_k\}$ having finite state space S, and let $\sim$ be an equivalence relation on S. Set $P([r], s) = \sum_{q \sim r} P(q, s)$. If $P([r], s) = P([r], t)$ for all $[r] \in S^*$ and all $s, t \in S$ with $s \sim t$, then $P_{\#}([r], [s]) := \frac{|[s]|}{|[r]|} P([r], s)$ is the transition matrix for a Markov chain $\{Y_k\}$ on S^*. The k-step transition probabilities are given by $(P^{\#})^k([r], [s]) = \frac{|[s]|}{|[r]|} \sum_{q \sim r} P^k(q, s)$ for all $k \in \mathbb{N}$. Furthermore,*

1. *If ϕ is a left eigenfunction of P with eigenvalue λ such that ϕ is constant on equivalence classes of S, then $\phi_{\#}([s]) := |[s]|\,\phi(s)$ is a left eigenfunction of $P_{\#}$ with eigenvalue λ.*

2. *If φ is a left eigenfunction of $P_{\#}$ with eigenvalue β, then $\varphi_{\flat}(s) := \frac{1}{|[s]|}\varphi([s])$ is a left eigenfunction of P with eigenvalue β.*

3. *If χ is a right eigenfunction of P with eigenvalue α, then $\chi_{\#}([r]) := \frac{1}{|[r]|}\sum_{q\sim r}\chi(q)$ is a right eigenfunction for $P_{\#}$ with eigenvalue α.*

4. *Assume that $\{X_k\}$ is ergodic with stationary distribution π. Then $\pi(s) = \pi(t)$ for all $s, t \in S$ with $s \sim t$ and $\{Y_k\}$ is ergodic with stationary distribution $\pi_{\#}$. Also, if P is reversible with respect to π, then $P_{\#}$ is reversible with respect to $\pi_{\#}$. Finally, we have the total variation bound $\left\|P_{\#}^k - \pi_{\#}\right\|_{TV^*} \le \left\|P^k - \pi\right\|_{TV^*}$.*

Proof. To see that $P_{\#}$ is the transition matrix for a Markov chain on S^*, we first note that the assumptions imply that $P_{\#}([r],[s])$ does not depend on the choice of equivalence class representatives and $P_{\#}([r],[s]) = \frac{|[s]|}{|[r]|}\sum_{q\sim r} P(q,s) \ge 0$ for all $[r],[s] \in S^*$. Moreover, for any $[r] \in S^*$, we have

$$\begin{aligned}\sum_{[s]\in S^*} P_{\#}([r],[s]) &= \sum_{[s]\in S^*}\frac{|[s]|}{|[r]|}P([r],s) = \sum_{[s]\in S^*}\frac{1}{|[r]|}\sum_{t\sim s}P([r],t)\\ &= \sum_{[s]\in S^*}\sum_{t\sim s}\frac{1}{|[r]|}\sum_{q\sim r}P(q,t) = \frac{1}{|[r]|}\sum_{q\sim r}\sum_{t\in S}P(q,t)\\ &= \frac{1}{|[r]|}\sum_{q\sim r}1 = \frac{|[r]|}{|[r]|} = 1,\end{aligned}$$

hence $P_{\#}$ is a well-defined stochastic matrix.

The k-step transition probabilities may be computed inductively because if we assume that $(P^{\#})^k([r],[s]) = \frac{|[s]|}{|[r]|}\sum_{q\sim r} P^k(q,s)$ for all $[r],[s] \in S^*$, then

$$\begin{aligned}(P^{\#})^{k+1}([r],[s]) &= \sum_{[u]\in S^*} P^{\#}([r],[u])(P^{\#})^k([u],[s]) \\ &= \sum_{[u]\in S^*} \frac{|[u]|}{|[r]|} P([r],u)\left(\frac{|[s]|}{|[u]|}\sum_{v\sim u} P^k(v,s)\right) \\ &= \frac{|[s]|}{|[r]|}\sum_{[u]\in S^*}\sum_{v\sim u} P([r],v)P^k(v,s) = \frac{|[s]|}{|[r]|}\sum_{v\in S} P([r],v)P^k(v,s) \\ &= \frac{|[s]|}{|[r]|}\sum_{q\sim r}\sum_{v\in S} P(q,v)P^k(v,s) = \frac{|[s]|}{|[r]|}\sum_{q\sim r} P^{k+1}(q,s).\end{aligned}$$

Since $(P^{\#})^{k+1}([r],[s]) = \sum_{[u]\in S^*} P^{\#}([r],[u])(P^{\#})^k([u],[s])$ and $P^{\#}([r],[s])$ does not depend on the choice of equivalence class representatives, it follows by induction that the k-step transitions do not depend on the choice of representatives.

Now assume that ϕ satisfies $\phi P = \lambda\phi$ and $\phi(s) = \phi(t)$ whenever $s \sim t$. Then

$$\begin{aligned}(\phi_{\#}P_{\#})([s]) &= \sum_{[r]\in S^*} \phi_{\#}([r])P_{\#}([r],[s]) = \sum_{[r]\in S^*} |[r]|\,\phi(r)\frac{|[s]|}{|[r]|}\sum_{q\sim r} P(q,s) \\ &= |[s]|\sum_{[r]\in S}\sum_{q\sim r}\phi(q)P(q,s) = |[s]|\sum_{q\in S}\phi(q)P(q,s) \\ &= |[s]|\,\lambda\phi(s) = \lambda\phi_{\#}([s]).\end{aligned}$$

Similarly, if φ satisfies $\varphi P_{\#} = \lambda\varphi$, then

$$\begin{aligned}(\varphi_{\flat}P)(s) &= \sum_{r\in S}\varphi_{\flat}(r)P(r,s) = \sum_{[r]\in S^*}\sum_{q\sim r}\varphi_{\flat}(q)P(q,s) \\ &= \sum_{[r]\in S^*}\varphi_{\flat}(r)\sum_{q\sim r} P(q,s) = \sum_{[r]\in S^*}\frac{1}{|[r]|}\varphi([r])P([r],s) \\ &= \frac{1}{|[s]|}\sum_{[r]\in S^*}\varphi([r])\frac{|[s]|}{|[r]|}P([r],s) = \frac{1}{|[s]|}\sum_{[r]\in S^*}\varphi([r])P_{\#}([r],[s])\end{aligned}$$

$$= \frac{1}{|[s]|}(\varphi P_{\#})([s]) = \frac{1}{|[s]|}\beta\varphi([s]) = \beta\varphi_{\flat}(s),$$

and if $P\chi = \alpha\chi$, then

$$\begin{aligned}(P_{\#}\chi_{\#})([r]) &= \sum_{[s]\in S^*} P_{\#}([r],[s])\chi_{\#}([s]) = \sum_{[s]\in S^*} \frac{|[s]|}{|[r]|}P([r],s)\frac{1}{|[s]|}\sum_{t\sim s}\chi(t) \\ &= \frac{1}{|[r]|}\sum_{[s]\in S^*}\sum_{t\sim s} P([r],t)\chi(t) = \frac{1}{|[r]|}\sum_{t\in S}\sum_{q\sim r} P(q,t)\chi(t) \\ &= \frac{1}{|[r]|}\sum_{q\sim r}(P\chi)(q) = \frac{1}{|[r]|}\sum_{q\sim r}\alpha\chi(q) = \alpha\chi_{\#}([r]).\end{aligned}$$

If $\{X_k\}$ is ergodic with stationary distribution π, then

$$\begin{aligned}\lim_{k\to\infty}(P^{\#})^k([r],[s]) &= \lim_{k\to\infty}\frac{|[s]|}{|[r]|}\sum_{q\sim r}P^k(q,s) \\ &= \frac{|[s]|}{|[r]|}\sum_{q\sim r}\pi(s) = |[s]|\,\pi(s) = \pi_{\#}([s]).\end{aligned}$$

Since $P^k([r],[s])$ does not depend on the equivalence class representatives, this means that π must be constant on the equivalence classes. If it is also the case that $\pi(r)P(r,s) = \pi(s)P(s,r)$ for all $r,s\in S$, then

$$\begin{aligned}\pi_{\#}([r])P_{\#}([r],[s]) &= |[r]|\,\pi(r)\frac{|[s]|}{|[r]|}P([r],s) = \sum_{t\sim s}\pi(r)P([r],t) \\ &= \sum_{t\sim s}\pi(r)\sum_{q\sim r}P(q,t) = \sum_{t\sim s}\sum_{q\sim r}\pi(q)P(q,t) = \sum_{t\sim s}\sum_{q\sim r}\pi(t)P(t,q) \\ &= \sum_{q\sim r}\pi(s)\sum_{t\sim s}P(t,q) = \sum_{q\sim r}\pi(s)P([s],q) = |[r]|\,\pi(s)P([s],r) \\ &= |[s]|\,\pi(s)\frac{|[r]|}{|[s]|}P([s],r) = \pi_{\#}([s])P_{\#}([s],[r]).\end{aligned}$$

Finally,

$$\begin{aligned}
\left\| P^k_{\#}([r],\cdot) - \pi_{\#}(\cdot) \right\|_{TV} &= \frac{1}{2} \sum_{[s]\in S^*} \left| P^k_{\#}([r],[s]) - \pi_{\#}([s]) \right| \\
&= \frac{1}{2} \sum_{[s]\in S^*} \left| \frac{|[s]|}{|[r]|} \sum_{q\sim r} P^k(q,s) - |[s]|\,\pi(s) \right| \\
&= \frac{1}{2} \sum_{[s]\in S^*} |[s]| \left| \frac{1}{|[r]|} \sum_{q\sim r} \Big(P^k(q,s) - \pi(s) \Big) \right| \\
&= \frac{1}{2} \sum_{[s]\in S^*} \sum_{t\sim s} \left| \frac{1}{|[r]|} \sum_{q\sim r} \Big(P^k(q,t) - \pi(t) \Big) \right| \\
&\leq \frac{1}{2} \sum_{[s]\in S^*} \sum_{t\sim s} \frac{1}{|[r]|} \sum_{q\sim r} \left| P^k(q,t) - \pi(t) \right| \\
&\leq \frac{1}{2} \sum_{[s]\in S^*} \sum_{t\sim s} \max_{q\sim r} \left| P^k(q,t) - \pi(t) \right| \\
&= \max_{q\sim r} \frac{1}{2} \sum_{t\in S} \left| P^k(q,t) - \pi(t) \right| \\
&= \max_{q\sim r} \left\| P^k(q,\cdot) - \pi(\cdot) \right\|.
\end{aligned}$$

Maximizing over $[r] \in S^*$ yields $\left\| P^k_{\#} - \pi_{\#} \right\|_{TV^*} \leq \left\| P^k - \pi \right\|_{TV^*}$. □

To the best of the author's knowledge, Theorem 1.3.2 is original, albeit quite similar in spirit to the Theorem 1.3.1. Since we are assuming that $P([r],s) = P([r],t)$ whenever $s \sim t$, $P_{\#}([r],[s]) = \frac{|[s]|}{|[r]|} P([r],s) = \frac{1}{|[r]|} \sum_{q\sim r} \sum_{t\sim s} P(q,t)$, so given that $Y_k = [r]$, the probability that $Y_{k+1} = [s]$ is equal to the probability of transitioning from a state chosen uniformly at random from the equivalence class $[r]$ to any state in the equivalence class $[s]$ under the original dynamics. When viewed in this light, the condition that $P([r],\cdot)$ is constant on equivalence classes just means that beginning at a random state in $[r]$, the original chain is equally likely to end up at each state in $[s]$. Though the preceding

theorem does not seem to be applicable to hyperplane walks in general, it is useful for the study of random walks on groups. For example, suppose that p is a probability measure on a finite group G. Then p gives rise to a Markov chain on G with transition probabilities $P(x,y) = p(yx^{-1})$. For any $H < G$, we can define an equivalence relation on G by $x \sim_H y$ if x and y belong to a common left coset of H. Then for any $g_1, g_2 \in G$, if $x \in g_1 H$ and $y, z \in g_2 H$ (say $y = g_2 h_1$ and $z = g_2 h_2$ for some $h_1, h_2 \in H$), we have

$$\begin{aligned} P([x], y) &= \sum_{h \in H} P(g_1 h, y) = \sum_{h \in H} p(y h^{-1} g_1^{-1}) = \sum_{h \in H} p(yhg_1^{-1}) \\ &= \sum_{h \in H} p(g_2 h_1 h g_1^{-1}) = \sum_{h \in H} p(g_2 h g_1^{-1}) \\ &= \sum_{h \in H} p(g_2 h_2 h g_1^{-1}) = \sum_{h \in H} p(z h g_1^{-1}) = \sum_{h \in H} p(z h^{-1} g_1^{-1}) \\ &= \sum_{h \in H} P(g_1 h, z) = P([x], z). \end{aligned}$$

Because the conditions Theorem 1.3.2 are satisfied, we can find left eigenfunctions for random walks on groups by examining induced Markov chains on various coset spaces. Observe that in this case we have $|[x]| = |H| = |[y]|$ for all $[x], [y] \in G^*$, so we can drop the scaling factors to obtain $(P_\#)^k([x],[y]) = \sum_{w \sim_H x} P^k(w,y)$, $\phi_\#([x]) = \phi(x)$, $\varphi_\flat(y) = \varphi([y])$, and $\chi_\#([x]) = \sum_{w \sim_H x} \chi(w)$.

Additionally, in the setting of the preceding paragraph, if $x = g_1 h_3$ and $w = g_1 h_4$ we also have that

$$\begin{aligned} P(x, [y]) &= \sum_{h \in H} P(g_1 h_3, g_2 h) = \sum_{h \in H} p(g_2 h h_3^{-1} g_1^{-1}) = \sum_{h \in H} p(g_2 h g_1^{-1}) \\ &= \sum_{h \in H} p(g_2 h h_4^{-1} g_1^{-1}) = \sum_{h \in H} p(g_1 h_4, g_2 h) = \sum_{h \in H} p(w, g_2 h) \end{aligned}$$

$$= P(w, [y]),$$

so random walks on groups are also lumpable with respect to $\sim_H$ in the sense of Theorem 1.3.1 as well. Note that in both of these cases, the only requirement is that G is a finite group. In particular, we are not assuming that p is constant on conjugacy classes of G or that the support of p generates G.

1.3.2 Products

A kind of converse method for constructing new chains from old ones involves taking products. Here we are assuming that we have Markov chains on $\Omega_1, ..., \Omega_n$ with transition matrices $P_1, ..., P_n$, and we wish to use these to build a chain on the product space $\Omega := \Omega_1 \times \cdots \times \Omega_n$. One way to do so is to start with a probability measure μ on $[n] = \{1, 2, ..., n\}$ and to proceed by choosing i from μ, then updating the ith coordinate according to P_i. The transition probabilities are thus

$$P(x, y) = \sum_{i=1}^{n} \mu(i) P_i(x_i, y_i) \prod_{j \neq i} 1\{y_j = x_j\}$$

for $x = (x_1, ..., x_n), y = (y_1, ..., y_n) \in \Omega$. Without loss of generality, we will suppose henceforth that $\mu(i) > 0$ for all $i \in [n]$. One easily verifies that if every P_i is irreducible, then P is irreducible as well. The same also holds for aperiodicity. Now, if $f^{(1)}, ..., f^{(n)}$ are functions on $\Omega_1, ..., \Omega_n$, then their tensor product is given by $(f^{(1)} \otimes f^{(2)} \cdots \otimes f^{(n)})(x_1, x_2, ..., x_n) := f^{(1)}(x_1) f^{(2)}(x_2) \cdots f^{(n)}(x_n)$. One may check that if, for each $i \in [n]$, $\varphi^{(i)}$ is a right (respectively, left) eigenfunction for P_i corresponding

to the eigenvalue λ_i, then $\varphi = \varphi^{(1)} \otimes \cdots \otimes \varphi^{(n)}$ is a right (respectively, left) eigenfunction for P corresponding to the eigenvalue $\sum_{i=1}^n \mu(i)\lambda_i$. This follows by writing $P(x,y) = \sum_{i=1}^n \mu(i)\widetilde{P}_i(x,y)$ where $\widetilde{P}_i(x,y) = P_i(x_i,y_i)\prod_{j\neq i} 1\{y_j = x_j\}$ is the chain on Ω which always moves in the ith coordinate according to P_i, and then verifying that φ is an eigenfunction for $\widetilde{P}_i$ with eigenvalue λ_i. In particular, we see that if $\pi^{(1)}, ..., \pi^{(n)}$ are stationary for $P_1, ..., P_n$, then $\pi := \pi^{(1)} \otimes \cdots \otimes \pi^{(n)}$ is stationary for P.

Now, keeping the above notation, suppose that $\mathcal{B}_1 = \{f_i\}_{i=1}^{N_1}$ is a basis of eigenfunctions for P_1 and $\mathcal{B}_2 = \{g_j\}_{j=1}^{N_2}$ is a basis of eigenfunctions for P_2. We claim that $\mathcal{B} = \{f_i \otimes g_j : f_i \in \mathcal{B}_1, g_j \in \mathcal{B}_2\}$ is a basis of eigenfunctions for the product chain P under any weighting μ such that $\mu(1), \mu(2) > 0$. We already know that each $f_i \otimes g_j \in \mathcal{B}$ is an eigenfunction for the product chain and that $|\mathcal{B}| = N_1 N_2 = |\Omega_1| \cdot |\Omega_2| = |\Omega_1 \times \Omega_2|$, so we need only show that the $f_i \otimes g_j$ are linearly independent. To this end, suppose that there exist $\{a_{i,j}\}_{i\in[N_1], j\in[N_2]}$ such that

$$0 = \sum_{i,j} a_{i,j} f_i \otimes g_j(x) = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} a_{i,j} f_i(x_1) g_j(x_2) = \sum_{i=1}^{N_1} f_i(x_1)\left(\sum_{j=1}^{N_2} a_{i,j} g_j(x_2)\right)$$

for all $x \in \Omega_1 \times \Omega_2$. Then, since the f_i are linearly independent, we must have that $\sum_{j=1}^{N_2} a_{i,j} g_j(x_2) = 0$ for all $x_2 \in \Omega_2$, $i \in [N_1]$. Of course, since the g_j are linearly independent, this means that $a_{i,j} = 0$ for all $i \in [N_1]$, $j \in [N_2]$. Consequently, the $f_i \otimes g_j$ are linearly independent, so $\mathcal{B}$ is indeed a basis. It follows by induction that if, for $i = 1, ..., n$, $\mathcal{B}_i = \{f_j^{(i)}\}_{j=1}^{|\Omega_i|}$ is a basis of eigenfunctions for P_i, then $\mathcal{B} = \{f_{j_1}^{(1)} \otimes \cdots f_{j_n}^{(n)} : f_{j_i}^{(i)} \in \mathcal{B}_i\}$ is a basis of eigenfunctions for P. This shows that P is diagonalizable whenever all of the P_i are diagonalizable. Moreover, in this case, the set of eigenvalues of P is precisely $\{\sum_{i=1}^n \mu(i)\lambda_i : \lambda_i \text{ is an eigenvalue of } P_i\}$. Thus if λ_{i*} is the second largest (in

modulus) eigenvalue of P_i, setting $J = \text{argmax}_{i\in[n]}\mu(i)\lambda_{i*}$, we see that the subdominant eigenvalue of P is given by

$$\lambda_* = \sum_{i\neq J}\mu(i) + \mu(J)\lambda_{J*} = 1 - \mu(J) + \mu(J)\lambda_{J*} = 1 - \mu(J)\gamma_{J*}$$

where $\gamma_{i*} = 1 - \lambda_{i*}$ is the absolute spectral gap of P_i. A quick computation shows that λ_* is also the subdominant eigenvalue of the chain $Q_J = \mu(J)P_J + (1-\mu(J))I$ obtained from P_J by adding a holding probability of $1 - \mu(J)$. Thus the exponential mixing rate of the product chain is the maximum of the exponential mixing rates of the composite chains endowed with the appropriate holding probabilities.

If, in addition, each of the P_i is reversible with respect to π_i, then there exist bases of eigenfunctions for each P_i, $\mathcal{B}_i = \{f_j^{(i)}\}_{j=1}^{|\Omega_i|}$, such that the $f_j^{(i)}$ are orthonormal in $L^2(\pi_i)$. Consequently, if $\varphi = \varphi^{(1)} \otimes \cdots \otimes \varphi^{(n)}$ and $\psi = \psi^{(1)} \otimes \cdots \otimes \psi^{(n)}$ are eigenfunctions in $\mathcal{B} = \{f_{j_1}^{(1)} \otimes \cdots f_{j_n}^{(n)} : f_{j_i}^{(i)} \in \mathcal{B}_i\}$, then

$$\langle \varphi, \psi \rangle_\pi = \prod_{i=1}^n \left\langle \varphi^{(i)}, \psi^{(i)} \right\rangle_{\pi_i} = \begin{cases} 1, & \varphi = \psi \\ 0, & \varphi \neq \psi \end{cases},$$

so $\mathcal{B}$ is a basis of eigenfunctions for P which is orthonormal in $L^2(\pi)$, hence P is reversible with respect to π. It turns out that the spectrum of P is given by $\sigma(P) = \{\sum_{i=1}^n \mu(i)\lambda_i : \lambda_i \text{ is an eigenvalue of } P_i\}$ regardless of whether the P_i's are all diagonalizable, but we will postpone the proof temporarily as it will follow from a more general result.

Another way to form a Markov chain on the product space is to move every coordinate at each step where P_i is the transition mechanism for the ith coordinate. This is referred to as the tensor product chain $P^{\otimes} = P_1 \otimes \cdots \otimes P_n$. Its transition probabilities are given by $P^{\otimes}(x,y) = \prod_{i=1}^n P_i(x_i, y_i)$. One readily verifies that if, for each $i \in [n]$, ϕ_i is an

eigenfunction for P_i with eigenvalue λ_i, then $\phi_1 \otimes \cdots \otimes \phi_n$ is an eigenfunction for $P^{\otimes}$ with eigenvalue $\prod_{i=1}^{n} \lambda_i$. To see that this is the case, note that if ϕ_1, ϕ_2 are right eigenfunctions for P_1, P_2, then

$$\begin{aligned}
\left[(P_1 \otimes P_2)(\phi_1 \otimes \phi_2)\right](w,x) &= \sum_{(y,z)\in\Omega_1\times\Omega_2} (\phi_1 \otimes \phi_2)(y,z)\,(P_1 \otimes P_2)((w,x),(y,z)) \\
&= \sum_{y\in\Omega_1} \phi_1(y) P_1(w,y) \sum_{z\in\Omega_2} \phi_2(z) P_2(x,z) \\
&= \lambda_1\phi_2(w)\lambda_2\phi_2(x) = \lambda_1\lambda_2\phi_1 \otimes \phi_2(w,x).
\end{aligned}$$

The same argument applies to left eigenfunctions and the result follows by induction.

In fact, the above argument applies to generalized eigenvectors as well and this can be used to establish eigenvalue multiplicities. Namely, suppose that λ_1 is an eigenvalue of P_1 with algebraic multiplicity m_1. Then it follows from the theory of Jordan forms that there exist linearly independent vectors $u_1, ..., u_{m_1}$ such that for all $1 \leq k \leq m_1$, $P_1 u_k = \lambda_1 u_k + \alpha_{k,k-1} u_{k-1} + ... + \alpha_{k,1} u_1$ for some constants $\alpha_{i,j} \in \mathbb{C}$, $1 \leq j < i \leq k$. Similarly, if λ_2 is an eigenvalue of P_2 with algebraic multiplicity m_2 there exist linearly independent vectors $v_1, ..., v_{m_2}$ and constants $\{\beta_{i,j}\}_{1\leq j<i\leq m_2} \subseteq \mathbb{C}$ with $P_2 v_k = \lambda_2 v_k + \beta_{k,k-1} v_{k-1} + ... + \beta_{k,1} v_1$ for all $1 \leq k \leq m_2$. Now let $k \in [m_1]$, $l \in [m_2]$ be given. Then, as with the ordinary eigenvectors, we have

$$\begin{aligned}
\left[(P_1 \otimes P_2)(u_k \otimes v_l)\right](w,x) &= \sum_{(y,z)\in\Omega_1\times\Omega_2} (u_k \otimes v_l)(y,z)\,(P_1 \otimes P_2)((w,x),(y,z)) \\
&= \sum_{y\in\Omega_1} u_k(y) P_1(w,y) \sum_{z\in\Omega_2} v_l(z) P_2(x,z) \\
&= (P_1 u_k)(w)(P_2 v_l)(x)
\end{aligned}$$

$$
\begin{aligned}
&= (\lambda_1 u_k(w) + ... + \alpha_{k,1} u_1(w)) (\lambda_2 v_l(x) + ... + \beta_{l,1} v_1(x)) \\
&= \lambda_1\lambda_2 (u_k \otimes v_l)(w,x) + \sum_{i=1}^{k-1} \lambda_2 \alpha_{k,i} (u_i \otimes v_l)(w,x) \\
&\quad + \sum_{j=1}^{l-1} \lambda_1 \beta_{l,j} (u_k \otimes v_j)(w,x) \\
&\quad + \sum_{i=1}^{k-1}\sum_{j=1}^{l-1} \alpha_{k,i}\beta_{l,j} (u_i \otimes v_j)(w,x) \\
&= \lambda_1\lambda_2 (u_k \otimes v_l)(w,x) + \sum_{r=1}^{kl-1} \gamma_r (u \otimes v)_r (w,x).
\end{aligned}
$$

Therefore, since $k \in [m_1]$, $l \in [m_2]$ were arbitrary and tensor products of linearly independent vectors are linearly independent, $\{u_k \otimes v_l\}_{k\in[m_1],l\in[m_2]}$ is a collection of $m_1 m_2$ linearly independent generalized eigenvectors for $P_1 \otimes P_2$ corresponding to the eigenvalue $\lambda_1\lambda_2$. Moreover, if $(P_1 \otimes P_2)(u_k \otimes v_l) = \lambda_1\lambda_2 (u_k \otimes v_l)$, then the above equation implies that $\sum_{r=1}^{kl-1} \gamma_r (u \otimes v)_r = 0$, so, since the $(u\otimes v)_r$'s are linearly independent, we must have that $\gamma_r = 0$ for all $r = 1, ..., kl-1$. In particular $\lambda_2\alpha_{k,i} = \lambda_1\beta_{l,j} = 0$ for all $i = 1, ..., k-1$, $j = 1, ..., l-1$. If $\lambda_1, \lambda_2 \neq 0$, then this means that u_k, u_l are eigenvectors. Therefore, the geometric multiplicity of any nonzero eigenvalue $\lambda_1\lambda_2$ of $P_1 \otimes P_2$ is equal to the product of the corresponding geometric multiplicities. Repeating the above argument and invoking associativity shows that if for $i = 1, ..., n$, ϕ_i is a generalized eigenvector for P_i with eigenvalue λ_i, then $\phi_1 \otimes \cdots \otimes \phi_n$ is a generalized eigenvector for $P_1 \otimes \cdots \otimes P_n$ with eigenvalue $\prod_{i=1}^n \lambda_i$ and that there are $\prod_{i=1}^n m_i$ such generalized eigenvectors for $\prod_{i=1}^n \lambda_i$. Because

$$
|\Omega| = \prod_{i=1}^n |\Omega_i| = \prod_{i=1}^n (m_1^{(i)} + ... + m_{k(i)}^{(i)}) = \sum_{i_1=1}^{k(1)} \cdots \sum_{i_n=1}^{k(n)} \prod_{j=1}^n m_{i_j}^{(j)},
$$

this gives a full basis of generalized eigenvectors, hence the spectrum of $P^{\otimes}$ is given by $\sigma(P^{\otimes}) = \{\prod_{i=1}^{n} \lambda_i : \lambda_i \in \sigma(P_i)\}$. If $0 \notin \sigma(P^{\otimes})$ or each P_i is diagonalizable, the eigenfunctions of $P^{\otimes}$ are precisely the tensor products of the component eigenfunctions. (The preceding can also be inferred by letting $V_i J_i V_i^{-1}$ be the Jordan decomposition for each P_i and observing that

$$(V_1 \otimes \cdots \otimes V_n)^{-1}(P_1 \otimes \cdots \otimes P_n)(V_1 \otimes \cdots \otimes V_n)$$
$$= (V_1^{-1} P_1 V_1) \otimes \cdots \otimes (V_n^{-1} P_n V_n) = J_1 \otimes \cdots \otimes J_n$$

where $J_1 \otimes \cdots \otimes J_n$ is upper triangular with the desired diagonal entries.) Finally, one may directly check that the tensor product chain is irreducible/aperiodic whenever each of the component chains is, and arguing as in the preceding paragraphs shows that diagonalizability and reversibility are inherited as well.

Having examined these product and tensor product constructions, it is natural to consider the following generalization. Namely, let Q be a probability measure on $2^{[n]}$, the collection of subsets of $[n] = \{1, ..., n\}$. Then one can construct a Markov chain $\{X_k\} = \{(X_k^{(1)}, ..., X_k^{(n)})\}$ on Ω by first picking $S \subseteq [n]$ from Q and then moving each coordinate $i \in S$ according to P_i. We may assume that for each $i \in [n]$, there is a set $S \subseteq [n]$ such that $i \in S$ and $Q(S) \neq 0$ because if, we have $X_k^{(j)} = X_0^{(j)}$ for all $k = \mathbb{N}_0$, then there is no point in including Ω_j in the product to begin with. In this case, the transition probabilities are given by

$$P_Q(x, y) = \sum_{S \subseteq [n]} Q(S) \prod_{i \in S} P_i(x_i, y_i) \prod_{j \notin S} 1\{y_j = x_j\}.$$

By construction, P_Q is a stochastic matrix. When Q is supported on the singletons of $[n]$, this reduces to the product chain we first considered, and when $Q([n]) = 1$, we obtain the tensor product chain. If we define $P^{\otimes S} := \bigotimes_{i=1}^{n} \left(I_{|\Omega_i|} + 1_S(i) \left[P_i - I_{|\Omega_i|} \right] \right)$ for each $S \subseteq [n]$ (so that, for example, $P^{\otimes \{1,3\}} = P_1 \otimes I_{|\Omega_2|} \otimes P_3 \otimes I_{|\Omega_4|} \otimes \cdots \otimes I_{|\Omega_n|}$), then we can write $P_Q = \sum_{S \subseteq [n]} Q(S) P^{\otimes S}$, hence P_Q is a convex combination of tensor product chains composed of the P_i's and appropriate identity matrices.

Now suppose that $\phi_1, ..., \phi_n$ are eigenfunctions for $P_1, ..., P_n$ with eigenvalues $\lambda_1, ..., \lambda_n$. Then for each $S \subseteq [n]$, $\phi_1 \otimes \cdots \otimes \phi_n$ is an eigenfunction for $P^{\otimes S}$ with eigenvalue $\prod_{i \in S} \lambda_i$, and thus, by linearity, $\phi_1 \otimes \cdots \otimes \phi_n$ is an eigenfunction for P_Q with eigenvalue $\sum_{S \subseteq [n]} Q(S) \prod_{i \in S} \lambda_i$. Applying the same argument using the results for generalized eigenvectors of tensor product chains shows that this completely accounts for the spectrum of P_Q. Moreover, if P_Q is nonsingular or each P_i is diagonalizable, then its eigenfunctions are precisely the tensor products of the component eigenfunctions.

For ease of reference, we record these observations as

Theorem 1.3.3. *Let $P_1, ..., P_n$ be transition kernels for Markov chains on $\Omega_1, ..., \Omega_n$ and let Q be a probability measure on $2^{[n]}$ such that for each $i \in [n]$, there is a set $S \subseteq [n]$ such that $i \in S$ and $Q(S) \neq 0$. Let $\Omega = \Omega_1 \times \cdots \times \Omega_n$. Then the matrix given by $P_Q(x, y) = \sum_{S \subseteq [n]} Q(S) \prod_{i \in S} P_i(x_i, y_i) \prod_{j \notin S} 1\{y_j = x_j\}$ for $x, y \in \Omega$ is the transition kernel for a Markov chain on Ω. Moreover,*

1. *If ϕ_i is an eigenfunction of P_i with eigenvalue λ_i for $i = 1, ..., n$, then $\phi = \phi_1 \otimes \cdots \otimes \phi_n$ is an eigenfunction for P_Q with eigenvalue $\sum_{S \subseteq [n]} Q(S) \prod_{i \in S} \lambda_i$.*

2. *The spectrum of $P^{\otimes}$ is given by*

$$\sigma(P^{\otimes}) = \{ \sum_{S \subseteq [n]} Q(S) \prod_{i \in S} \lambda_i : \lambda_i \in \sigma(P_i) \text{ for } i = 1, ..., n \}.$$

3. *Every eigenfunction of P_Q corresponding to a nonzero eigenvalue may be expressed as a tensor product of the eigenfunctions of the component chains. If each P_i is diagonalizable, then all eigenfunctions of P_Q can be expressed in such a form.*

4. *If each of the component chains is irreducible, aperiodic, diagonalizable, or reversible, respectively, then so is P_Q.*

The above set up can be further generalized by letting the component transition probabilities depend on which coordinates have been selected for updating. Specifically, suppose that for each Ω_i, we have a family of Markov kernels $\{P_i^{(S)}\}_{S \subseteq [n]}$. Then a step in this product chain corresponds to first choosing a set $S \subseteq [n]$ and then moving each coordinate $i \in S$ according to $P_i^{(S)}$, so that the product kernel is given by

$$P(x, y) = \sum_{S \subseteq [n]} Q(S) \prod_{i \in S} P_i^{(S)}(x_i, y_i) \prod_{j \notin S} 1\{y_j = x_j\}.$$

On the whole, it seems difficult to derive universal results at this level of generality. However, if we suppose that for each $i \in [n]$, the matrices $\{P_i^{(S)}\}_{S \subseteq [n]}$ have the same set of eigenfunctions, then the preceding analysis carries over directly. In particular, if $\{P_i^{(S)}\}_{S \subseteq [n]}$ is a commuting family of diagonalizable matrices for each $i \in [n]$, then P is diagonalizable and each of its eigenfunctions may be represented as a tensor product of the component eigenfunctions. In a similar vein, we can define chains on the product space as linear combinations of tensor product chains where the families of component kernels have index set different from $2^{[n]}$. That is, if for each P_i, we have a family $\{P_i^{(S)}\}_{S \in I}$ of

transition operators and Q is a probability measure on I, then $P = \sum_{S \in I} Q(S) P_1^{(S)} \otimes \cdots \otimes P_n^{(S)}$ is the transition operator for a Markov chain on $\Omega_1 \times \cdots \times \Omega_n$ and much of the same analysis applies. We will find this latter construction to be useful in computing eigenfunctions for the random-to-top chain on S_n.

Finally, we observe that, as one would expect, product chains are lumpable with respect to the equivalence relation induced by projection onto a subset of coordinates. As before, let $P_1, ..., P_n$ be transition matrices for Markov chains on $\Omega_1, ..., \Omega_n$, let Q be a probability on the subsets of $[n]$, and consider the product chain on $\Omega = \Omega_1 \times \cdots \times \Omega_n$ given by

$$P_Q(x, y) = \sum_{S \subseteq [n]} Q(S) \prod_{i \in S} P_i(x_i, y_i) \prod_{j \notin S} 1\{y_j = x_j\}.$$

Observe that for any $A \subseteq [n]$, we can define an equivalence relation on Ω by $x \sim_A y$ if and only if $x_i = y_i$ for all $i \in A$. Moreover, for every $[y]_A \in \Omega^* = \Omega / \sim_A$ and every $x \in \Omega$, if we set $Q_A(B) = \sum_{T \subseteq [n] \setminus A} Q(B \cup T)$ for $B \subseteq A$, we have

$$\begin{aligned}
P_Q(x, [y]_A) &= \sum_{z \sim_A y} \sum_{B \subseteq A} \sum_{C \subseteq [n] \setminus A} \Bigg[Q(B \cup C) \prod_{i \in B} P_i(x_i, z_i) \prod_{j \in C} P_j(x_j, z_j) \\
&\qquad \times \prod_{k \in A \cap B^C} 1\{z_k = x_k\} \prod_{l \in A^C \cap C^C} 1\{z_l = x_l\} \Bigg] \\
&= \sum_{B \subseteq A} \Bigg[\prod_{i \in B} P_i(x_i, y_i) \prod_{k \in A \cap B^C} 1\{y_k = x_k\} \\
&\qquad \times \sum_{z \sim_A y} \sum_{C \subseteq [n] \setminus A} \left(Q(B \cup C) \prod_{j \in C} P_j(x_j, z_j) \prod_{l \in A^C \cap C^C} 1\{z_l = x_l\} \right) \Bigg] \\
&= \sum_{\substack{B \subseteq A: \\ Q_A(B) \neq 0}} \Bigg[\prod_{i \in B} P_i(x_i, y_i) \prod_{k \in A \cap B^C} 1\{y_k = x_k\}
\end{aligned}$$

$$\times \sum_{z\sim_A y} \sum_{C\subseteq[n]\setminus A} \left(Q(B\cup C) \prod_{j\in C} P_j(x_j, z_j) \prod_{l\in A^C\cap C^C} 1\{z_l = x_l\} \right) \Bigg]$$

$$= \sum_{\substack{B\subseteq A:\\ Q_A(B)\neq 0}} \Bigg[Q_A(B) \prod_{i\in B} P_i(x_i, y_i) \prod_{k\in A\cap B^C} 1\{y_k = x_k\}$$

$$\times \sum_{z\sim_A y} \sum_{C\subseteq[n]\setminus A} \left(\frac{Q(B\cup C)}{Q_A(B)} \prod_{j\in C} P_j(x_j, z_j) \prod_{l\in A^C\cap C^C} 1\{z_l = x_l\} \right) \Bigg]$$

$$= \sum_{\substack{B\subseteq A:\\ Q_A(B)\neq 0}} Q_A(B) \prod_{i\in B} P_i(x_i, y_i) \prod_{k\in A\cap B^C} 1\{y_k = x_k\}$$

$$= \sum_{B\subseteq A} Q_A(B) \prod_{i\in B} P_i(x_i, y_i) \prod_{k\in A\cap B^C} 1\{y_k = x_k\}.$$

The penultimate equality is due to the fact that for each $B \subseteq A$ with $Q_A(B) \neq 0$,

$$P_{Q_{A,B}}(x_{[n]\setminus A}, z_{[n]\setminus A}) := \sum_{C\subseteq[n]\setminus A} \frac{Q(B\cup C)}{Q_A(B)} \prod_{j\in C} P_j(x_j, z_j) \prod_{l\in A^C\cap C^C} 1\{z_l = x_l\}$$

is a transition probability for the product chain on $\Omega_{[n]\setminus A} := \prod_{i\in[n]\setminus A} \Omega_i$ generated by the probability measure $Q_{A,B}(C) := \frac{Q(B\cup C)}{Q_A(B)}$ on $2^{[n]\setminus A}$. Since $P_{Q_{A,B}}(x_{[n]\setminus A}, z_{[n]\setminus A})$ only involves the coordinates in $[n] \setminus A$, it follows from the definition of $\sim_A$ that

$$\sum_{z\sim_A y} P_{Q_{A,B}}(x_{[n]\setminus A}, z_{[n]\setminus A}) = \sum_{z\in\Omega_{[n]\setminus A}} P_{Q_{A,B}}(x_{[n]\setminus A}, z) = 1.$$

Thus we see that P_Q is lumpable with respect to $\sim_A$ in the sense of Theorem 1.3.1. A completely analogous argument shows that the same applies to the more general chain

$$P(x, y) = \sum_{S\subseteq[n]} Q(S) \prod_{i\in S} P_i^{(S)}(x_i, y_i) \prod_{j\notin S} 1\{y_j = x_j\}$$

where the coordinate transition matrices are allowed to depend on the subset chosen from Q.

Chapter 2

Random Walks on Hyperplane Arrangements

2.1 Hyperplane Arrangements

The main class of Markov chains we will be considering in this thesis arise as random walks on the chambers or faces of (finite, real, central) hyperplane arrangements. The basic idea is that a collection of hyperplanes carves the underlying space into a bunch of pieces called faces, and there is a natural way in which these faces can be multiplied by one another. One can construct a Markov chain by repeatedly choosing a face according to some probability distribution and then moving from the present face to its product with the randomly chosen face. Before examining these chains in detail, we need to establish some background and terminology concerning hyperplane arrangements. The standard reference for hyperplane arrangements is the text [52] by Peter Orlik and Hiroaki Terao, and Richard Stanley's lecture notes [64] provide a nice, readable introduction. Our discussion and notation is primarily based upon the paper [12] by Brown and Diaconis.

A finite hyperplane arrangement $\mathcal{A} = \{H_i\}_{i=1}^m$ is a finite collection of hyperplanes in a linear, affine, or projective space V. We will assume throughout that $V = \mathbb{R}^n$, though

it can also be of interest to consider vector spaces over other fields. (For example, Anders Björner has an interesting paper which generalizes the Tsetlin library by considering complex arrangements [10].) We will impose the additional requirement that $\cap_{i=1}^m H_i \neq \emptyset$, in which case we may assume that $0 \in \cap_{i=1}^m H_i$, so that the H_i are just codimension 1 subspaces of $\mathbb{R}^n$. Such an arrangement is called central. If $\cap_{i=1}^m H_i = \{0\}$, then we say that the arrangement is essential. This assumption is occasionally invoked in discussions of hyperplane walks, but we will not need it here. (One can always work with essential arrangements by taking the quotient of V by the intersection $\cap_{i=1}^m H_i$, and one can pass from noncentral arrangements to central arrangements by a process called coning, which is essentially an embedding of the arrangement in $V \times \mathbb{R}$ defined so that the intersection is nonempty. Neither of these processes will affect the combinatorial constructs which are of interest in the context of hyperplane walks. See [12, 64, 52] for more on passing to central and essential arrangements.)

Now if $\mathcal{A} = \{H_i\}_{i=1}^m$ is a collection of hyperplanes in $V = \mathbb{R}^n$ and $\gamma \in V \setminus \mathcal{A}$, then each H_i partitions V into three subsets: $H_i^+ = \{\text{open half-space containing } \gamma\}$, $H_i^0 = H_i$, and $H_i^- = \{\text{open half-space not containing } \gamma\}$. (Equivalently, each hyperplane can be identified with the solutions to a linear equation $H_i = H_i^0 = \{x \in V : L_i x = c_i\}$, and we can define the half-spaces by $H_i^+ = \{x \in V : L_i x > c_i\}$, $H_i^- = \{x \in V : L_i x < c_i\}$. For central arrangements, $c_i = 0$ for all $i \in [m]$.) The set of faces of the arrangement, $\mathcal{F} = \mathcal{F}(\mathcal{A})$, consists of all nonempty intersections of the form

$$F = \cap_{i=1}^m H_i^{\sigma_i(F)},\ \sigma_i(F) \in \{-, 0, +\}.$$

Thus each face may be identified with its sign sequence

$$F \sim \sigma(F) = (\sigma_1(F), \sigma_2(F), ..., \sigma_m(F)).$$

The n-dimensional faces are called chambers, and the set of chambers is denoted by $\mathcal{C} = \mathcal{C}(\mathcal{A})$. These are the (convex) connected components of $V \setminus \mathcal{A}$, and their sign sequences satisfy $\sigma_i(C) \neq 0$ for all $i \in [m]$. Note that our method of orienting the hyperplanes always guarantees the existence of a chamber with sign sequence $(+, +, ..., +)$ - namely, the chamber containing γ. We observe that except in the Boolean case (to be discussed later), not all of the 3^m possible sign sequences are realized by the faces of a given arrangement.

The intersection poset associated with $\mathcal{A}$, $\mathcal{L} = \mathcal{L}(\mathcal{A})$, is defined as the collection of all intersections of the form $\cap_{i \in A} H_i$, $A \subseteq [m]$, ordered by reverse inclusion. (In the noncentral case, we only consider nonempty intersections. Also, beware that some authors order $\mathcal{L}$ by inclusion.) For $U, W \in \mathcal{L}$, the join of U and W is $U \vee W = U \cap W$ and the meet of U and W is $U \wedge W = U + W$, the smallest subspace containing U and W. The elements of the intersection poset are referred to as flats. To avoid ambiguity in representation, we will define the support set of a flat $W \in \mathcal{L}$ to be the set $A_W = \{i \in [m] : W \subseteq H_i\}$, so that $W = \cap_{i \in A_W} H_i$. We include the following proposition as motivation for this definition.

Proposition 2.1.1. *For any $W, U \in \mathcal{L}$, $W \leq U$ if and only if $A_W \subseteq A_U$.*

Proof. If $A_W \subseteq A_U$, then $U = \cap_{i \in A_U} H_i \subseteq \cap_{i \in A_W} H_i = W$, hence $W \leq U$. On the other hand, if $W \leq U$, then $U \subseteq W$, so for every $H_i \in A_W$, we have $U \subseteq W \subseteq H_i$, so $H_i \in A_U$. Therefore, $W \leq U$ implies that $A_W \subseteq A_U$ as well. □

Thus we see that ordering the flats by reverse inclusion is equivalent to ordering their support sets by inclusion.

As $\mathcal{A}$ is assumed to be central, it can be shown that $\mathcal{L}$ is actually a (geometric) lattice with top element $\hat{1} = \cap_{i=1}^{m} H_i$ and bottom element $\hat{0} = V$, the empty intersection. The Möbius function of $\mathcal{L}$ is defined recursively by

$$\mu(W, W) = 1,$$

$$\mu(W, W') = -\sum_{W \leq U < W'} \mu(W, U) \text{ for } W < W',$$

$$\mu(W, W') = 0 \text{ for } W \not\leq W'.$$

(See chapter 5 in [1].) The Möbius function is involved in the formula for the multiplicities of the eigenvalues of hyperplane walks and also in an upper bound on the distance to stationarity after k steps. Moreover, we have the following famous theorem due to Thomas Zaslavsky [71]:

Theorem 2.1.1 (Zaslavsky)**.** *Let $\mathcal{A}$ be a finite arrangement of hyperplanes in $V = \mathbb{R}^n$ with chamber set $\mathcal{C}$ and intersection lattice $\mathcal{L}$. Let μ be the Möbius function of $\mathcal{L}$. Then* $|\mathcal{C}| = \sum_{W \in \mathcal{L}} |\mu(V, W)|$.

In addition to the partial order on flats defined above, there is a related partial order on faces. For $F, G \in \mathcal{F}$, we write $G \leq F$ if for each $i \in [m]$, either $\sigma_i(F) = 0$ or $\sigma_i(F) = \sigma_i(G)$, in which case we say that F is a face of G. (As with the intersection lattice, some authors reverse the inequalities in the definition of the face poset.) We define the support set of a face $F \in \mathcal{F}$ by

$$A_F = \{i \in [m] : F \subseteq H_i\} = \{i \in [m] : \sigma_i(F) = 0\}$$

and define the support of F by $\text{supp}(F) = \cap_{i \in A_F} H_i \in \mathcal{L}$. Thus the support of a face is the maximal element of the intersection lattice in which that face is contained. Note that $G \leq F$ in $\mathcal{F}$ implies that $\text{supp}(G) \leq \text{supp}(F)$ in $\mathcal{L}$, but the converse is not true in general. Similarly, if $F \in \mathcal{F}$ and $W \in \mathcal{L}$, then $F \subseteq W$ if and only if $\text{supp}(F) \geq W$. In terms of support sets, we have $A_{\text{supp}(F)} = A_F$. Also, since the chambers are precisely those faces which are not contained in any hyperplane, we see that $F \in \mathcal{C}$ if and only if $\text{supp}(F) = V$.

Two chambers are said to be adjacent if they share a common codimension 1 face, and we can define the chamber graph to have vertices indexed by the chambers and edges given by the adjacency relation. The geodesic distance on the chamber graph provides a metric on $\mathcal{C}$, and the distance between two chambers given by this metric is equal to the minimum number of hyperplanes one crosses when traveling from one chamber to the other [12].

There is a natural product on the faces of a hyperplane arrangement defined in terms of sign sequences by

$$\sigma_i(FG) = \begin{cases} \sigma_i(F), & \sigma_i(F) \neq 0 \\ \sigma_i(G), & \sigma_i(F) = 0 \end{cases}.$$

The face product may be interpreted geometrically as follows: FG is the first face encountered after traveling a positive distance in a straight line from a point in F to a point in G. It is routine to verify that this product makes $\mathcal{F}$ a semigroup. (Since we are dealing with central arrangements, the face $O = \cap_{i=1}^{m} H_i$, which has sign sequence $(0, ..., 0)$, is a two sided identity with respect to this product, so the faces have a monoidal structure.) In addition, it is clear that the chambers form a two-sided ideal in the face semigroup,

so, in particular, $FC \in \mathcal{C}$ for all $F \in \mathcal{F}$, $C \in \mathcal{C}$. In fact, FC is the unique chamber having F as a face which is closest to C in terms of the metric defined above [12]. It is worth pointing out that, by construction, F is a face of FG for all $G \in \mathcal{F}$, thus successive (right) face multiplications tend to move one downwards in the face poset (i.e. towards the chambers). Indeed, one may define the partial order on faces in terms of the face product by saying that F is a face of G if and only if $FG = G$. The definition of the face product also shows that $\mathcal{F}$ is idempotent ($FF = F$ for all $F \in \mathcal{F}$) and has the left regular property: $FGF = FG$ for all $F, G \in \mathcal{F}$. An idempotent semigroup with the left regular property is known as a left-regular band, and much of the analysis of hyperplane arrangements and the random walks thereon can be extended easily to this more general setting (see [14, 13, 62]).

Because much of the foregoing is difficult to visualize, even in dimensions 2 and 3, we introduce an alternative description of the face semigroup: Recalling that each face is uniquely determined by its sign sequence, we associate the faces in $\mathcal{F}$ with rows of m colored tiles where the ith tile in the row corresponding to a face F is red if $\sigma_i(F) = +$, green if $\sigma_i(F) = -$, and clear if $\sigma_i(F) = 0$. The product FG corresponds to the row of tiles one observes when stacking the row corresponding to F on top of the row corresponding to G, keeping in mind that one can see through the clear tiles while the red and green tiles are opaque. We will henceforth refer to this as the RCT (rows of colored tiles) description of hyperplane walks. This interpretation is illustrated in the following diagram.

A graphical representation for multiplying $G \sim (-,+,+,-,0,0)$ by $F \sim (+,0,-,0,0,+)$ to obtain $FG \sim (+,+,-,-,0,+)$

One final concept that we will need when describing the eigenvectors for hyperplane walks is that of a subarrangement. Given a hyperplane arrangement $\mathcal{A} = \{H_i\}_{i=1}^m$ and a set $B \subseteq [m]$, we can define a subarrangement $\mathcal{B} = \{H_i\}_{i \in B}$ - that is, $\mathcal{B}$ is obtained from $\mathcal{A}$ by restricting attention to a subset of the hyperplanes in $\mathcal{A}$. In terms of the face semigroup, this is equivalent to projecting the sign sequences of faces in $\mathcal{A}$ onto $\{-,0,+\}^{|B|}$ by only keeping track of the coordinates in B. Thus in the RCT description, we are just ignoring the tiles corresponding to hyperplanes which are not in our subarrangement, perhaps by painting all such tiles black in every row. In the case where $B = A_W$ for some $W \in \mathcal{L}$, we denote the subarrangement by $\mathcal{B}_W = \{H_i : W \subseteq H_i\}$. Similarly, when $B = A_F = A_{\text{supp}(F)}$ for some $F \in \mathcal{F}$, we denote the subarrangement by $\mathcal{B}_F = \{H_i : F \subseteq H_i\}$. Observe that for $U, U' \in \mathcal{L}$, $U \leq U'$ if and only if $\mathcal{B}_U$ is a subarrangement of $\mathcal{B}_{U'}$. The following proposition shows that the Möbius function of the intersection lattice corresponding to such a subarrangement agrees with the Möbius function of $\mathcal{L}$ on the flats of the subarrangement.

Lemma 2.1.1. *Given a hyperplane arrangement $\mathcal{A} = \{H_i\}_{i=1}^m$ in $V = \mathbb{R}^n$ and a flat $W \in \mathcal{L}$, let $\mathcal{B}_W$ be the subarrangement $\{H_i\}_{H_i \supseteq W}$. Then $\mathcal{L}_W$, the intersection lattice of $\mathcal{B}_W$, is isomorphic to the interval $[V, W]_{\mathcal{L}} \subseteq \mathcal{L}$. Thus for all $U, U' \in \mathcal{L}_W$, $\mu_W(U, U') = \mu(U, U')$ where μ_W and μ are the Möbius functions of $\mathcal{L}$ and $\mathcal{L}_W$, respectively.*

Proof. We note first that $\mathcal{L}_W \subseteq \mathcal{L}$ by construction, so the definition of the intersection lattices implies that $U \leq U'$ in $\mathcal{L}_W$ if and only if $U' \subseteq U$ if and only if $U \leq U'$ in $\mathcal{L}$. Because $\mathcal{L}_W$ inherits the partial order on $\mathcal{L}$, we will have $\mathcal{L}_W \cong [V, W]_{\mathcal{L}}$ as long as $\mathcal{L}_W$ and $[V, W]_{\mathcal{L}}$ are equal as sets. To see that this is indeed the case, we observe that if $U \in \mathcal{L}_W$, then $U = \cap_{i \in S} H_i$ for some $S \subseteq B_W \subseteq [m]$, hence $W = \cap_{i \in B_W} H_i \subseteq U \subseteq V$, so $\mathcal{L}_W \subseteq [V, W]_{\mathcal{L}}$. Conversely, if $U' \in [V, W]_{\mathcal{L}}$, then $W \subseteq U'$, so $U' \leq W$, hence $A_{U'} \subseteq A_W$ by Proposition 2.1.1, thus $U' \in \mathcal{L}_W$. Therefore, since the Möbius function is defined recursively on intervals and $\mathcal{L}_W \cong [V, W]_{\mathcal{L}}$, the Möbius function of $\mathcal{L}_W$ is just the restriction of the Möbius function of $\mathcal{L}$ to the interval $[V, W]_{\mathcal{L}}$, hence the two agree on $\mathcal{L}_W$. □

It is worth noting that for an arbitrary subarrangement $\mathcal{B} = \{H_i\}_{i \in B}$, it is not necessarily true that the intersection lattice of $\mathcal{B}$ is isomorphic to an interval in $\mathcal{L}$ since one may have flats in $[V, \cap_{i \in B} H_i]$ which cannot be expressed as intersections over a subset of B. (If B is a proper subset of the support set of $\cap_{i \in B} H_i$, then there is a hyperplane $H_j \notin B$ containing $\cap_{i \in B} H_i$ which is in $[V, \cap_{i \in B} H_i]$, but not in the intersection lattice of $\mathcal{B}$.)

Subarrangements are formed by deleting hyperplanes from a given arrangement. A related operation is that of restriction. Formally, given a flat $W \in \mathcal{L}(\mathcal{A})$, we can define

the restricted arrangement $\mathcal{B}^W = \{H \cap W : H \in \mathcal{A} \setminus \mathcal{B}_W\}$ in the ambient vector space W. One can show that for any $W \in \mathcal{L}$, the intersection lattice for $\mathcal{B}^W$ is isomorphic to the interval $[W, \bigcap_{H \in \mathcal{A}} H]_{\mathcal{L}}$ in $\mathcal{L}(\mathcal{A})$ just as $\mathcal{B}_W \cong [V, W]_{\mathcal{L}}$. The proof of Zaslavsky's theorem is essentially a recursive argument based on the identity $|\mathcal{C}(\mathcal{A})| = |\mathcal{C}(\mathcal{B}_{H_1})| + \left|\mathcal{C}(\mathcal{B}^{H_1})\right|$ - the number of chambers in $\mathcal{A}$ is equal to the number of chambers in the arrangement $\mathcal{B}_{H_1}$ formed by deleting H_1 plus the number of chambers in $\mathcal{B}_{H_1}$ which are cut into 2 pieces by H_1, the latter being equal to the number of chambers in the restriction to H_1 [64]. Though subarrangements are central to our analysis of the eigenfunctions of hyperplane chamber walks in chapter 3, we do not see a way to use the dual notion of restricted arrangements to obtain analogous results.

2.2 Hyperplane Walks

In light of the semigroup structure on the faces of a hyperplane arrangement $\mathcal{A} = \{H_i\}_{i=1}^m$, it is obvious how to construct a random walk on $\mathcal{F}$. Namely, given a probability measure w on $\mathcal{F}$ and some initial face F_0, we define $X_0 = F_0$ and $X_{k+1} = F_{k+1}X_k$ for all $k \in \mathbb{N}_0$ where $F_1, F_2, ...$ are chosen independently from w. Then $\{X_k\}_{k=0}^{\infty}$ is clearly a Markov chain and the transition probabilities are given by

$$P(G, H) = \sum_{\substack{F \in \mathcal{F}: \\ FG = H}} w(F)$$

for all $G, H \in \mathcal{F}$. Because $\mathcal{C}$ is a left ideal in the face semigroup, we see that if $X_0 = F_0 \in \mathcal{C}$, then $X_k \in \mathcal{C}$ for all k. Consequently, we may restrict the state space to $\mathcal{C}$ without changing the underlying dynamics. For various reasons which will be discussed later, we will focus on random walks on the chambers of $\mathcal{A}$ in what follows.

These hyperplane chamber walks were introduced by Pat Bidigare in his PhD thesis [7] and subsequent investigations were soon carried out by Persi Diaconis and Ken Brown [12] and by Bidigare, Phil Hanlon, and Dan Rockmore [8]. They are often referred to as BHR walks after the latter authors. The main results for these Markov chains are summarized in the following theorems [12, 8].

Theorem 2.2.1. *Let $\mathcal{A}$ be an arrangement of hyperplanes in $V = \mathbb{R}^n$ with face poset $\mathcal{F}$ and intersection poset $\mathcal{L}$, and let w be a probability measure on $\mathcal{F}$. Then the matrix $P(C, C') = \sum_{FC=C'} w(F)$, $C, C' \in \mathcal{C}$, is diagonalizable over $\mathbb{R}$, and for each $W \in \mathcal{L}$, there is an eigenvalue*

$$\lambda_W = \sum_{\substack{F \in \mathcal{F}: \\ F \subseteq W}} w(F)$$

occurring with multiplicity

$$m_W = |\mu(V, W)| = (-1)^{\operatorname{codim}(W,V)} \mu(V, W)$$

where μ is the Möbius function of $\mathcal{L}$ and $\operatorname{codim}(W, V)$ denotes the codimension of W in V.

Note that Theorem 2.2.1 gives an alternate proof of Zaslavsky's Theorem. Also, despite the fact that hyperplane walks are not generally reversible, we see that all of the eigenvalues of P are nonnegative real numbers. However, one should be aware that it may be the case that $\lambda_{W_1} = ... = \lambda_{W_l} = \lambda$ for distinct $W_1, ..., W_l \in \mathcal{L}$, in which case the multiplicity of λ is $m_\lambda = m_{W_1} + ... + m_{W_l}$.

In addition to establishing diagonalizability and computing eigenvalues, the above authors were able to determine a criterion for ergodicity, describe the stationary distribution, and upper-bound the rate of convergence (with respect to the total variation metric) in terms of the eigenvalues. To state the results as succinctly as possible, we adopt the following terminology: A probability measure w on $\mathcal{F}$ is called separating if it is not concentrated on the faces of any $H_i \in \mathcal{A}$. That is, for every $i \in [m]$, there is some $F \in \mathcal{F}$ such that $\sigma_i(F) \neq 0$ and $w(F) > 0$.

Theorem 2.2.2. *Let $\mathcal{A}$, $\mathcal{F}$, $\mathcal{C}$, w, and P be as in Theorem 2.2.1. Then*

1. *P has a unique stationary distribution π if and only if w is separating.*

2. *Assume that w is separating. Sample without replacement from w to obtain an ordering $F_1, F_2, ..., F_r$ of $\{F \in \mathcal{F} : w(F) > 0\}$. Then the product $F_1 F_2 \cdots F_r$ is a chamber distributed according to π.*

3. *Still assuming that w is separating, and letting P_C^k be the distribution of the walk started at $C \in \mathcal{C}$ after k steps, the total variation distance between P_C^k and π satisfies*

$$\left\| P_C^k - \pi \right\|_{TV} \leq - \sum_{W > V} \mu(V, W) \lambda_W^k \leq \sum_{H \in \mathcal{A}} \lambda_H^k.$$

Theorem 2.2.2 is the easier of the two to prove, so we will begin there. Brown and Diaconis [12] establish the theorem by considering the stationary $\mathcal{F}$-valued process $..., F_{-1}, F_0, F_1, ...$ where each F_i is distributed according to w. They then argue that the condition that w is separating guarantees that the infinite product $F_1^\infty = \prod_{i=1}^\infty F_i = \lim_{k \to \infty} F_1 \cdots F_k$ is almost surely constant and take π to be the distribution of F_1^∞. Next,

they show that for any $C_0 \in \mathcal{C}$, letting π_k denote the distribution of $C_k = F_1^k C_0$ where $F_1^k = F_1 \cdots F_k$, one has

$$\|\pi_k - \pi\|_{TV} \leq \mathbb{P}\{F_1^k \neq F_1^\infty\} = \mathbb{P}\{F_1 \cdots F_k \notin \mathcal{C}\} \to 0 \text{ as } k \to \infty.$$

Since $C_k = F_1 \cdots F_k C_0 =_d F_{-k} \cdots F_{-1} C_0$ by shift-invariance of the stationary process, this establishes the first two parts of the theorem up to some fairly trivial observations. The third part follows by using Möbius inversion to write

$$\mathbb{P}\{F_1 \cdots F_k \notin \mathcal{C}\} = \mathbb{P}\{F_1, ..., F_k \in H_i \text{ for some } i \in [m]\} = -\sum_{W > V} \mu(V, W) \lambda_W^k.$$

The first two parts of the theorem can be deduced in a similar fashion by applying the machinery from [27] to the infinite composite of the random mapping representation of the chain. For the sake of variety and to highlight some of the more salient aspects of the result, we now provide a different proof of Theorem 2.2.2 along with an equivalent upper bound on distance to stationarity which is more computationally tractable.

To begin with, let w be a separating probability measure on $\mathcal{F}$ and consider the Markov chain on $\mathcal{C}$ with transitions given by $X_k = F_k X_{k-1}$ where F_k is chosen from w independent of $F_1, ..., F_{k-1}$. Let $\mathcal{W} = \{F_{[1]}, ..., F_{[r]}\} \subseteq \mathcal{F}$ be an enumeration of the support of w and write $\mathcal{C}' = \{F_{[\sigma(r)]} \cdots F_{[\sigma(1)]} : \sigma \in S_r\}$. We first observe that since w is separating, $\mathcal{C}' \subseteq \mathcal{C}$. This is easy to see in terms of the RCT description of the faces since in this context the requirement that w is separating is equivalent to the statement that for each tile position, there is some $F \in \mathcal{W}$ such that F has an opaque tile in that position. As such, once all rows of tiles corresponding to the faces in $\mathcal{W}$ have been stacked on top of one another, the entire row will be opaque regardless of the order in which they were

stacked. Notice also that if $F_k \cdots F_1 \in \mathcal{C}$ for some $k \geq 1$, then $X_k = F_k \cdots F_1 X_0 \in \mathcal{C}'$. This is because we can use the left-regularity property of $\mathcal{F}$ ($FGF = FG$) to delete all but the last instance of any face occurring with multiplicity in the sequence $F_1, ..., F_k$ to obtain $F_{i_l} \cdots F_{i_1} = F_k \cdots F_1 = X_k$ with the $F'_{i_j}s$ distinct, and then use the fact that $CF = C$ for all $C \in \mathcal{C}$, $F \in \mathcal{F}$ to append the remaining faces (in any order) at the beginning, yielding $X_k = F_{i_l} \cdots F_{i_1} F_{j_1} \cdots F_{j_{r-l}} \in \mathcal{C}'$.

Now for any $C \in \mathcal{C}'$, $F \in \mathcal{W}$, we have $C = F_{[\sigma(r)]} \cdots F_{[\sigma(1)]}$ for some $\sigma \in S_r$ and $F = F_{[\sigma(i)]}$ for some $i \in \{1, ..., r\}$. If $i = r$, then

$$FC = F_{[\sigma(r)]} F_{[\sigma(r)]} \cdots F_{[\sigma(1)]} = F_{[\sigma(r)]} \cdots F_{[\sigma(1)]} \in \mathcal{C}'$$

by idempotence, and if $i \neq r$, then

$$FC = F_{[\sigma(i)]} F_{[\sigma(r)]} \cdots F_{[\sigma(i)]} \cdots F_{[\sigma(1)]} = F_{[\sigma(i)]} F_{[\sigma(r)]} \cdots F_{[\sigma(i+1)]} F_{[\sigma(i-1)]} \cdots F_{[\sigma(1)]} \in \mathcal{C}'$$

by left-regularity. Thus, if $X_k \in \mathcal{C}'$, then $X_n \in \mathcal{C}'$ for all $n \geq k$. Also, for any $C, D \in \mathcal{C}'$ and any $n \in \mathbb{N}_0$, $k \geq r$, if $D = F_{[\tau(r)]} \cdots F_{[\tau(1)]}$, then

$$\mathbb{P}\{X_{n+k} = D \,|X_n = C\} \geq \mathbb{P}\{F_{n+k} = F_{[\tau(r)]}, ..., F_{n+k-r+1} = F_{[\tau(1)]}\} > 0$$

because $F_{n+k} = F_{[\tau(r)]}, ..., F_{n+k-r+1} = F_{[\tau(1)]}$ implies

$$F_{n+k} \cdots F_{n+k-r+1} F_{n+k-r} \cdots F_{n+1} C = D F_{n+k-r} \cdots F_{n+1} C = D$$

as $D \in \mathcal{C}' \subseteq \mathcal{C}$ and $BF = B$ for all $B \in \mathcal{C}$, $F \in \mathcal{F}$). Consequently, the Markov chain on $\mathcal{C}'$ is irreducible and aperiodic, so there exists a unique and strictly positive stationary distribution $\widetilde{\pi}$ such that $\lim_{k\to\infty} \mathbb{P}\{X_{n+k} = D \,|X_n = C\} = \widetilde{\pi}(D)$ for all $C, D \in \mathcal{C}'$.

We can extend $\widetilde{\pi}$ to all of $\mathcal{C}$ by defining $\pi(C) = \widetilde{\pi}(C)1\{C \in \mathcal{C}'\}$, and we claim that π is the unique stationary distribution for the random walk on $\mathcal{C}$. To see that this is the case, we construct the following coupling: Let $X_0 \sim \mu$, $Y_0 \sim \pi$ and recursively define $X_k = F_k X_{k-1}$, $Y_k = F_k Y_{k-1}$ where $F_1, F_2, ...$ are i.i.d. picks from w. Then for all $k \in \mathbb{N}$, $X_k \sim P_\mu^k$ by definition and $Y_k \sim \pi$ since $Y_0 \in \mathcal{C}'$ a.s. and $\widetilde{\pi} = \pi|_{\mathcal{C}'}$ is stationary. Consequently, it follows from the coupling characterization of the total variation distance that $\left\|P_\mu^k - \pi\right\|_{TV} \leq \mathbb{P}\{X_k \neq Y_k\}$. Setting $T = \min\{k \in \mathbb{N} : F_k \cdots F_1 \in \mathcal{C}\}$, we have

$$X_k = F_k \cdots F_1 X_0 = F_k \cdots F_1 = F_k \cdots F_1 Y_0 = Y_k$$

whenever $k \geq T$, hence $\left\|P_\mu^k - \pi\right\|_{TV} \leq \mathbb{P}\{T > k\}$.

Now

$$\begin{aligned}\mathbb{P}\{T > k\} &= \mathbb{P}\{F_k \cdots F_1 \notin \mathcal{C}\} = \mathbb{P}\{F_1, ..., F_k \in H_i \text{ for some } i \in [m]\} \\ &\leq \sum_{i=1}^{m} \mathbb{P}\{F_1, ..., F_k \in H_i\},\end{aligned}$$

and for each $i \in [m]$,

$$\mathbb{P}\{F_1, ..., F_k \in H_i\} = \mathbb{P}\{F_1 \in H_i\} \cdots \mathbb{P}\{F_k \in H_i\} = \mathbb{P}\{F_1 \in H_i\}^k$$

where

$$\mathbb{P}\{F_1 \in H_i\} = \sum_{\substack{F \in \mathcal{W}: \\ F \subseteq H_i}} w(F) \leq 1 - \min_{F \in \mathcal{W}} w(F)$$

since w is separating. Therefore,

$$\left\|P_\mu^k - \pi\right\|_{TV} \leq \mathbb{P}\{T > k\} \leq \sum_{i=1}^{m} \mathbb{P}\{F_1, ..., F_k \in H_i\} \leq m\left(1 - \min_{F \in \mathcal{W}} w(F)\right)^k \to 0$$

as $k \to \infty$, so the random walk has a unique stationary distribution whenever w is separating. If w is not separating, then there exists some H_i such that $F \subseteq H_i$ for all $F \in \mathcal{W}$. Thus when $X_0 = (+, ..., +)$, we have $\sigma_i(X_k) = +$ for all k, and when $X_0 = (-, ..., -)$,we have $\sigma_i(X_k) = -$ for all k so the condition that w is separating is necessary as well. (Both of the preceding sign sequences correspond to faces in any arrangement by virtue of our method of orienting the hyperplanes.)

Also, note that the proof of the first part of Theorem 2.2.2 shows that $F_T \cdots F_1 = F_T \cdots F_1 X_0 \sim \pi$. The interpretation is thus "Sample without replacement from w until the product $F_T \cdots F_1$ is a chamber. This chamber is distributed according to π." Moreover, the RCT description shows that $F_T \cdots F_1 \in \mathcal{C}$ if and only if $F_{\sigma(T)} \cdots F_{\sigma(1)} \in \mathcal{C}$ for all $\sigma \in S_T$ (since the opacity of the stack does not depend on the ordering of the rows), so it is equivalent say "Sample faces from w until $F_1 \cdots F_T \in \mathcal{C}$. Then $F_T \cdots F_1$ is distributed according to π." Because the left regular property allows us to delete all but the last occurrences of any faces in the sequence $F_1, ..., F_T$ without changing the value of the product, it is also equivalent to sample without replacement as this is the same as sampling with replacement and then deleting repeats. At the other extreme, rather than just eliminating the face just sampled at each stage, we may, upon having already chosen $F_1, ..., F_i$, remove all faces contained in the support of $F_1 \cdots F_i$ before picking F_{i+1} since $F \subseteq \text{supp}(F_1 \cdots F_i)$ implies that $\text{supp}(F) \subseteq \text{supp}(F_1 \cdots F_i)$ and thus $F_1 \cdots F_i = F_1 \cdots F_i F$. This follows from the same argument as before, and both of these alternate descriptions of π are given in [12]. These observations combine to give the second part of Theorem 3 (and several equivalent formulations as well).

In particular, the description of π given in Theorem 2.2.2 shows that we may write

$$\pi(C) = \sum_{\substack{\sigma \in S_r: \\ F_{[\sigma(1)]} \cdots F_{[\sigma(r)]} = C}} \prod_{1 \le j \le r} \frac{w(F_{[\sigma(j)]})}{1 - \sum_{i<j} w(F_{[\sigma(i)]})}.$$

Now observe that

$$\begin{aligned}
\sum_{C \in \mathcal{C}} \pi(C) P(C, D) &= \sum_{C \in \mathcal{C}} \left(\sum_{\substack{\sigma \in S_r: \\ F_{[\sigma(1)]} \cdots F_{[\sigma(r)]} = C}} \prod_{1 \le j \le r} \frac{w(F_{[\sigma(j)]})}{1 - \sum_{i<j} w(F_{[\sigma(i)]})} \right) \left(\sum_{\substack{F \in \mathcal{W}: \\ FC = D}} w(F) \right) \\
&= \sum_{C \in \mathcal{C}} \sum_{\substack{\sigma \in S_r: \\ F_{[\sigma(1)]} \cdots F_{[\sigma(r)]} = C}} \sum_{\substack{F \in \mathcal{W}: \\ F F_{[\sigma(1)]} \cdots F_{[\sigma(r)]} = D}} w(F) \prod_{1 \le j \le r} \frac{w(F_{[\sigma(j)]})}{1 - \sum_{i<j} w(F_{[\sigma(i)]})} \\
&= \sum_{\sigma \in S_r} \sum_{\substack{F \in \mathcal{W}: \\ F F_{[\sigma(1)]} \cdots F_{[\sigma(r)]} = D}} w(F) \prod_{1 \le j \le r} \frac{w(F_{[\sigma(j)]})}{1 - \sum_{i<j} w(F_{[\sigma(i)]})} \\
&= \sum_{F \in \mathcal{W}} w(F) \sum_{\substack{\sigma \in S_r: \\ F F_{[\sigma(1)]} \cdots F_{[\sigma(r)]} = D}} \prod_{1 \le j \le r} \frac{w(F_{[\sigma(j)]})}{1 - \sum_{i<j} w(F_{[\sigma(i)]})},
\end{aligned}$$

and for each $\sigma \in S_r$, $F \in \mathcal{W}$, we have that $F F_{[\sigma(1)]} \cdots F_{[\sigma(r)]} = F_{[\tau(1)]} \cdots F_{[\tau(r)]}$ for some $\tau \in S_r$ uniquely determined by F and σ (by the same argument that showed $X_0 \in \mathcal{C}'$ implies $X_k \in \mathcal{C}'$ for all $k \in \mathbb{N}$). Consequently,

$$\begin{aligned}
\sum_{C \in \mathcal{C}} \pi(C) P(C, D) &= \sum_{F \in \mathcal{W}} w(F) \sum_{\substack{\sigma \in S_r: \\ F F_{[\sigma(1)]} \cdots F_{[\sigma(r)]} = D}} \prod_{1 \le j \le r} \frac{w(F_{[\sigma(j)]})}{1 - \sum_{i<j} w(F_{[\sigma(i)]})} \\
&= \left(\sum_{F \in \mathcal{W}} w(F) \right) \left(\sum_{\substack{\tau \in S_r: \\ F_{[\tau(1)]} \cdots F_{[\tau(r)]} = D}} \prod_{1 \le j \le r} \frac{w(F_{[\sigma(j)]})}{1 - \sum_{i<j} w(F_{[\sigma(i)]})} \right) \\
&= \sum_{\substack{\tau \in S_r: \\ F_{[\tau(1)]} \cdots F_{[\tau(r)]} = D}} \prod_{1 \le j \le r} \frac{w(F_{[\sigma(j)]})}{1 - \sum_{i<j} w(F_{[\sigma(i)]})} = \pi(D),
\end{aligned}$$

providing another proof that π is stationary.

At this point, it remains only to establish the third part of Theorem 2.2.2. To this end, recall that

$$\left\|P_\mu^k - \pi\right\|_{TV} \leq \mathbb{P}\{T > k\} = \mathbb{P}\{F_k \cdots F_1 \notin \mathcal{C}\} = \mathbb{P}\{\text{supp}(F_k \cdots F_1) \neq V\}$$

$$= \mathbb{P}\{\text{supp}(F_k \cdots F_1) > V\} = \sum_{W > V} \mathbb{P}\{\text{supp}(F_k \cdots F_1) = W\}.$$

Now for each $W \in \mathcal{L}$, set

$$\lambda_W = \sum_{\substack{F \in \mathcal{F}: \\ F \subseteq W}} w(F).$$

These are the purported eigenvalues from Theorem 2.2.1. By definition, we have

$$\lambda_W = \sum_{\substack{F \in \mathcal{F}: \\ F \subseteq W}} w(F) = \mathbb{P}\{\text{supp}(F_i) \geq W\},$$

so

$$\lambda_W^k = \mathbb{P}\{\text{supp}(F_k) \geq W\} \cdots \mathbb{P}\{\text{supp}(F_1) \geq W\} = \mathbb{P}\{\text{supp}(F_k), ..., \text{supp}(F_1) \geq W\}$$

$$= \mathbb{P}\{\text{supp}(F_k \cdots F_1) \geq W\} = \sum_{U \geq W} \mathbb{P}\{\text{supp}(F_k \cdots F_1) = U\}.$$

As such, Möbius inversion (see [1]) gives

$$\mathbb{P}\{\text{supp}(F_k \cdots F_1) = W\} = \sum_{U \geq W} \lambda_U^k \mu(W, U),$$

hence

$$\left\|P_\mu^k - \pi\right\|_{TV} \leq \sum_{W > V} \mathbb{P}\{\text{supp}(F_k \cdots F_1) = W\} = \sum_{W > V} \sum_{U \geq W} \lambda_U^k \mu(W, U)$$

$$= \sum_{U > V} \lambda_U^k \sum_{U \geq W > V} \mu(W, U) = \sum_{U > V} \lambda_U^k \left(-\mu(V, U)\right)$$

$$= -\sum_{U > V} \lambda_U^k \mu(V, U),$$

which is the third part of Theorem 2.2.2. Note that since the argument involves inversion from above, it is important that we are working with a central arrangement so that $\mathcal{L}$ has top element $\hat{1} = \cap_{i=1}^{m} H_i$.

Though it is nice to have an upper bound on variation distance in terms of the eigenvalues of the transition matrix, the Möbius function for the intersection lattice of an arbitrary arrangement is not easy to compute, making the bound fairly unwieldy. Of course, we could use the simpler bound

$$\begin{aligned}\left\|P_{\mu}^{k} - \pi\right\|_{TV} &\leq \mathbb{P}\{F_k \cdots F_1 \notin \mathcal{C}\} = \mathbb{P}\{F_1, ..., F_k \in H_i \text{ for some } i \in [m]\} \\ &\leq \sum_{i=1}^{m} \mathbb{P}\{F_1, ..., F_k \in H_i\} = \sum_{i=1}^{m} \mathbb{P}\{F_1 \in H_i\}^k \\ &= \sum_{i=1}^{m} \lambda_{H_i}^{k}\end{aligned}$$

which is equivalent to truncating the alternating sum $-\sum_{U>V} \lambda_U^k \mu(V,U)$ after the atoms of $\mathcal{L}$ (i.e. the hyperplanes), but then we lose the advantage of all the cancellation. One way to remedy this problem is to employ the same basic reasoning in a more familiar lattice. To wit, for each $S \subseteq [m]$, define

$$\beta_s = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = 0 \, \forall i \in S}} w(F)$$

and for each $i \in [m]$, set $B_i^k = \{F_1, ..., F_k \in H_i\}$. Then inclusion-exclusion yields

$$\begin{aligned}\left\|P_{\mu}^{k} - \pi\right\|_{TV} &\leq \mathbb{P}\{F_1, ..., F_k \in H_i \text{ for some } i \in [m]\} = \mathbb{P}\left\{\cup_{i=1}^{m} B_i^k\right\} \\ &= \sum_{i=1}^{m} \mathbb{P}\{B_i^k\} - \sum_{1 \leq i < j \leq m} \mathbb{P}\{B_i^k \cap B_j^k\} + ... - (-1)^m \mathbb{P}\left\{\cap_{i=1}^{m} B_i^k\right\}.\end{aligned}$$

Since

$$\begin{aligned}\mathbb{P}\{B_{i_1}^k \cap \cdots \cap B_{i_r}^k\} &= \mathbb{P}\{F_1, ..., F_k \subseteq \cap_{i\in\{i_1,...,i_r\}} H_i\} \\ &= \mathbb{P}\{F_1 \subseteq \cap_{i\in\{i_1,...,i_r\}} H_i\} \cdots \mathbb{P}\{F_k \subseteq \cap_{i\in\{i_1,...,i_r\}} H_i\} \\ &= \mathbb{P}\{F_1 \subseteq \cap_{i\in\{i_1,...,i_r\}} H_i\}^k = \beta_{\{i_1,...,i_r\}}^k,\end{aligned}$$

we have

Theorem 2.2.3. *In the setting of Theorem 2.2.2, we have the equivalent bound*

$$\left\| P_\mu^k - \pi \right\|_{TV} \leq - \sum_{\substack{S\subseteq[m]: \\ S\neq\emptyset}} (-1)^{|s|} \beta_S^k$$

where

$$\beta_S = \sum_{\substack{F\in\mathcal{F}: \\ \sigma_i(F)=0\ \forall i\in S}} w(F).$$

Of course, inclusion-exclusion is Möbius inversion on the lattice of subsets, so both arguments have the same underlying structure. In fact, the two formulations are equivalent when $\mathcal{A}$ is a Boolean arrangement (to be discussed in the next section). In some sense, the latter bound can be seen as embedding the hyperplane walk into a Boolean arrangement and carrying out the computations in this simpler setting. It is also worth pointing out that the $\lambda_W' s$ are a subset of the $\beta_S' s$, and the two are equivalent if one ignores multiplicity. When S is a singleton, we have $\beta_{\{i\}} = \lambda_{H_i}$, and when $S = A_W$, we have $\beta_{A_W} = \lambda_W$.

Finally, observe that the proof of Theorem 2.2.2 carries through unchanged if we allow the state space to consist of all of $\mathcal{F}$ rather than just $\mathcal{C}$. In particular, it is informative to consider the random walk on $\mathcal{F}$ with initial state $X_0 = \mathcal{O} = \cap_{i=1}^m H_i$. Recalling that

the conditional distribution of $X_k = F_k \cdots F_1$ given that $F_k \cdots F_1 \in \mathcal{C}$ is π, if we let $P_{\mathcal{O}}^k$ denote the distribution of X_k where $X_0 = \mathcal{O}$, the for all $C \in \mathcal{C}$,

$$P_0^k(C) = \mathbb{P}\{X_k = C\} = \mathbb{P}\{X_k = C, X_k \in \mathcal{C}\} = \mathbb{P}\{X_k = C \,|X_k \in \mathcal{C}\}\mathbb{P}\{X_k \in \mathcal{C}\}$$

$$= \pi(C)\mathbb{P}\{X_k \in \mathcal{C}\} \leq \pi(C).$$

(We have tacitly assumed that $\mathbb{P}\{X_k \in \mathcal{C}\} > 0$, but the conclusion holds trivially if this is not the case.) Since π is supported on $\mathcal{C}' \subseteq \mathcal{C}$, it follows from the extreme event characterization of total variation that

$$\left\|P_0^k - \pi\right\|_{TV} = \sum_{\substack{F \in \mathcal{F}: \\ \pi(F) \geq P_0^k(F)}} \left[\pi(F) - P_0^k(F)\right] = \sum_{C \in \mathcal{C}} \left[\pi(C) - P_0^k(C)\right]$$

$$= 1 - \sum_{C \in \mathcal{C}} P_0^k(C) = 1 - P_0^k(\mathcal{C}) = P_0^k(\mathcal{C}^C) = \mathbb{P}\{F_k \cdots F_1 \notin \mathcal{C}\},$$

thus the upper bound $\left\|P_\mu^k - \pi\right\|_{TV} \leq \mathbb{P}\{T > k\} = \mathbb{P}\{F_k \cdots F_1 \notin \mathcal{C}\}$ is tight in the case of random walk on the faces of a hyperplane arrangement with initial state $X_0 = \mathcal{O}$. (This also implies that the maximal total variation distance $\left\|P^k - \pi\right\|_{TV^*}$ is given by the upper bounds in Theorems 2.2.2 and 2.2.3 in the case of hyperplane face walks.) Moreover, since $\left\|P_{\mathcal{O}}^k - \pi\right\|_{TV} = \mathbb{P}\{F_k \cdots F_1 \notin \mathcal{C}\}$, the Bonferroni inequalities give

$$\sum_{i=1}^m \beta_{\{i\}}^k - \sum_{1 \leq i < j \leq m} \beta_{\{i,j\}}^k \leq \left\|P_{\mathcal{O}}^k - \pi\right\|_{TV} \leq \sum_{i=1}^m \beta_{\{i\}}^k.$$

For walks started at any other face, we can still use Theorem 2.2.3 and the Bonferroni inequalities to get upper bounds on total variation distance which are better than the bound

in terms only of the eigenvalues corresponding to hyperplanes and are not significantly harder to compute in principle. For example, we have

$$\left\|P_\mu^r - \pi\right\|_{TV} \leq \sum_{i=1}^{m} \beta_{\{i\}}^r - \sum_{1\leq i<j\leq m} \beta_{\{i,j\}}^r + \sum_{1\leq i<j<k\leq m} \beta_{\{i,j,k\}}^r.$$

Though it is nice to have an exact formula for the total variation distance, it is not generally that useful to consider walks on the faces. This is largely because the chambers typically have a different combinatorial description than the rest of the faces and the stationary distribution will be concentrated on $\mathcal{C}$ whenever the chain is ergodic. Also, though Theorems 2.2.2 and 2.2.3 hold in the case of face walks, it turns out that the proofs of Theorem 2.3.1 do not carry over since $\mathcal{F}$ is not a meet semilattice.

Having dispensed with Theorem 2.2.2, we now turn our attention to Theorem 2.2.1. The idea here is to consider the vector space $\mathbb{R}\mathcal{F}$ of formal linear combinations of elements in $\mathcal{F}$ - that is, the vectors are of the form $\sum_{F\in\mathcal{F}} a_F F$, $a_F \in \mathbb{R}$. Extending the semigroup product on $\mathcal{F}$ yields the bilinear product

$$\left(\sum_{F\in\mathcal{F}} a_F F\right)\left(\sum_{F\in\mathcal{F}} b_F F\right) = \left(\sum_{F\in\mathcal{F}} c_F F\right)$$

where

$$c_F = \sum_{GH=F} a_G b_H$$

is the standard convolution product. Thus we can view $\mathbb{R}\mathcal{F}$ as an $\mathbb{R}$-algebra. If $\mathcal{I}$ is an ideal of $\mathcal{F}$, then the action of $\mathcal{F}$ on $\mathcal{I}$ makes the free vector space $\mathbb{R}\mathcal{I}$ an $\mathbb{R}\mathcal{F}$-module. (This holds for any ideal, including $\mathcal{I} = \mathcal{F}$, but we will primarily be interested the case

$\mathcal{I} = \mathcal{C}$, and so will work in this setting henceforth.) Given a probability measure w on $\mathcal{F}$, we can consider the element

$$T_w = \sum_{F \in \mathcal{F}} w(F) F \in \mathbb{R}\mathcal{F}$$

which acts as an operator on $\mathbb{R}\mathcal{C}$. That is, given $\alpha = \sum_{C \in \mathcal{C}} \alpha_C C \in \mathbb{R}\mathcal{C}$, we have

$$T_w(\alpha) = \left(\sum_{F \in \mathcal{F}} w(F) F \right) \left(\sum_{C \in \mathcal{C}} \alpha_C C \right) = \sum_{C \in \mathcal{C}} \sum_{F \in \mathcal{F}} w(F) \alpha_C FC = \sum_{B \in \mathcal{C}} \gamma_B B$$

where

$$\gamma_B = \sum_{C \in \mathcal{C}} \alpha_C \sum_{\substack{F \in \mathcal{F}: \\ FC = B}} w(F) = \sum_{C \in \mathcal{C}} \alpha_C P(C, B).$$

Consequently, if we think of the elements of $\mathbb{R}\mathcal{C}$ as row vectors, then T_w acts as right multiplication by the transition matrix P. Thus P is diagonalizable if T_w is diagonalizable and the eigenvectors of T_w on $\mathbb{R}\mathcal{C}$ correspond to the left eigenvectors of P, giving the spectrum of P provided that the spectrum of T_w is known [12, 13].

Brown and Diaconis use tools from topology to show that T_w is diagonalizable and to compute its eigenvalues and their multiplicities [12]. In the more general setting of left-regular bands, Brown gives a purely algebraic derivation of the eigenvalues using semigroup representation theory and shows that the subalgebra generated by T_w is isomorphic to a direct product of copies of $\mathbb{R}$ so that $\mathbb{R}[T_w]$ is semisimple, and thus the action of T_w is diagonalizable in every $\mathbb{R}\mathcal{F}$-module [13]. The paper [14] gives a nice, readable overview of this argument. Using ideas from [12] and [13], Graham Denham was able to use combinatorial Heaviside functions to describe the left eigenspaces of the transition matrices of hyperplane chamber walks in terms of flags in the intersection lattices [19],

but it seems difficult to use these ideas to explicitly write down the eigenvectors in general. In the setting of left-regular bands, Franco Saliola was able to express the analog of T_w as a linear combination of primitive orthogonal idempotents in the semigroup algebra to obtain similar results for left eigenspaces of random walks on LRBs and he used this to deduce diagonalizability [61, 62].

These algebraic analyses of the face semigroup algebra are fascinating in their own right (see [61] for further remarkable properties of $k\mathcal{F}$), but as the author has nothing substantial to add to these results and the focus of this thesis is probabilistic, the reader is referred to the above references for more details. Before moving on though, we provide a purely combinatorial derivation of the eigenvalues and their multiplicities due to Christos Athanasiadis and Persi Diaconis in an article which foreshadows some of our results concerning the right eigenvectors of hyperplane chamber walks [5].

To begin, we note that if $A \in M_n(\mathbb{C})$ has eigenvalues $\lambda_1, ..., \lambda_n$, then $\mathrm{Tr}(A^k) = \lambda_1^k + ... + \lambda_n^k$ for all $k \in \mathbb{N}$ - this is clearly true for diagonalizable matrices and since the set of $n \times n$ diagonalizable matrices is dense in $M_n(\mathbb{C})$, the continuity of the trace function implies the result. Conversely, if $\mathrm{Tr}(A^k) = \lambda_1^k + ... + \lambda_n^k$ for some $A \in M_n(\mathbb{C})$, then it follows from Newton's identities (see [42]) that $\lambda_1, ..., \lambda_n$ are the eigenvalues of A. Consequently, it suffices to prove that

$$\mathrm{Tr}(P^k) = \sum_{W \in \mathcal{L}} |\mu(V, W)| \, \lambda_W^k$$

for all $k \in \mathbb{N}$.

Now observe that for any $F \in \mathcal{F}$, the set $\{C \in \mathcal{C} : FC = C\}$ is in one-to-one correspondence with $\mathcal{C}(\mathcal{B}_F)$, the set of chambers in the subarrangement $\mathcal{B}_F = \{H_i : F \subseteq$

$H_i\}$. This is easy to see in terms of the RCT description of the faces: Given a collection of rows of colored tiles representing $\mathcal{F}$, a representation of the faces of the subarrangement $\mathcal{B}_F$ can be obtained by painting the ith tile in each row black if $\sigma_i(F) \neq 0$ and then throwing out any duplicates. The chambers of $\mathcal{B}_F$ are represented by the remaining rows which have no clear tiles. The chambers in $\{C \in \mathcal{C} : FC = C\}$ correspond precisely to the rows of tiles obtained by painting the black tiles in these rows the corresponding colors of F - that is, if $\sigma_i(F) = +$, change the ith tiles from black to red, and if $\sigma_j(F) = -$, change the jth tiles from black to green. The associated map from $\mathcal{C}(\mathcal{B}_F)$ to $\{C \in \mathcal{C} : FC = C\}$ is well-defined since every row of tiles in $\mathcal{C}(\mathcal{B}_F)$ is sent to a unique row in $\mathcal{C}$ which has the same color as F in the positions for which the F row is opaque, it is injective since we threw out duplicates, and it is surjective since every row in $\{C \in \mathcal{C} : FC = C\}$ is the image of the row in $\mathcal{C}(\mathcal{B}_F)$ obtained by painting the ith tile black whenever $\sigma_i(F) \neq 0$. Consequently, it follows from Theorem 2.1.1 and Lemma 2.1.1 that

$$|\{C \in \mathcal{C} : FC = C\}| = |\mathcal{C}(\mathcal{B}_F)| = \sum_{W \in \mathcal{L}_F} |\mu_{\mathcal{L}_F}(V, W)| = \sum_{\substack{W \in \mathcal{L}: \\ F \subseteq W}} |\mu(V, W)| .$$

Therefore, by definition of the random walk, we have

$$\begin{aligned}
\mathrm{Tr}(P^k) &= \sum_{C \in \mathcal{C}} \sum_{\substack{(F_1, \ldots, F_k) \in \mathcal{F}^k: \\ F_k \cdots F_1 C = C}} w(F_1) \cdots w(F_k) = \sum_{C \in \mathcal{C}} \sum_{\substack{F \in \mathcal{F}: \\ FC = C}} \sum_{\substack{(F_1, \ldots, F_k) \in \mathcal{F}^k: \\ F_k \cdots F_1 = F}} w(F_1) \cdots w(F_k) \\
&= \sum_{F \in \mathcal{F}} |\{C \in \mathcal{C} : FC = C\}| \sum_{\substack{(F_1, \ldots, F_k) \in \mathcal{F}^k: \\ F_k \cdots F_1 = F}} w(F_1) \cdots w(F_k) \\
&= \sum_{F \in \mathcal{F}} \sum_{\substack{W \in \mathcal{L}: \\ F \subseteq W}} |\mu(V, W)| \sum_{\substack{(F_1, \ldots, F_k) \in \mathcal{F}^k: \\ F_k \cdots F_1 = F}} w(F_1) \cdots w(F_k)
\end{aligned}$$

$$= \sum_{W \in \mathcal{L}} |\mu(V,W)| \sum_{\substack{(F_1,\ldots,F_k) \in \mathcal{F}^k: \\ F_k \cdots F_1 \subseteq W}} w(F_1) \cdots w(F_k)$$

$$= \sum_{W \in \mathcal{L}} |\mu(V,W)| \sum_{\substack{(F_1,\ldots,F_k) \in \mathcal{F}^k: \\ F_1,\ldots,F_k \subseteq W}} w(F_1) \cdots w(F_k)$$

$$= \sum_{W \in \mathcal{L}} |\mu(V,W)| \left(\sum_{\substack{F \in \mathcal{F}: \\ F \subseteq W}} w(F) \right)^k = \sum_{W \in \mathcal{L}} |\mu(V,W)| \, \lambda_W^k,$$

which establishes the result.

2.3 Examples

In order to motivate the study of hyperplane walks and illustrate the concepts mentioned above, we will examine several interesting Markov chains which can be modeled in terms of random walks on the Boolean, braid, and dihedral arrangements, respectively. Most of the examples considered here (or analogues thereof) can be found in [12].

We remark that the three arrangements considered in this section are each examples of what are known as reflection arrangements. An element $s \in GL(V)$ is called a reflection if it has finite order and its fixed point set is a hyperplane H_s (called the reflecting hyperplane of s), and a finite group $G \leq GL(V)$ is called a finite reflection group (or finite Coxeter group) if it is generated by reflections. The set of reflecting hyperplanes of a finite reflection group G is known as the reflection arrangement of G [52]. (We are considering isomorphic groups to be equivalent, so a reflection group is a group which is isomorphic to a subgroup of some general linear group which is generated by reflections.) If $\mathcal{A}$ is the reflection arrangement of a finite reflection group G, then the chambers of $\mathcal{A}$ correspond to the elements of G and the faces of $\mathcal{A}$ correspond to left cosets of parabolic

subgroups of G [38, 43]. As such, there is a natural action of G on $\mathcal{C}(\mathcal{A})$ and $\mathcal{F}(\mathcal{A})$, respectively, and G acts simply transitively on $\mathcal{C}$. Consequently, we can view hyperplane chamber walks as Markov chains on G, though they do not represent random walks on G in the standard sense of the term. However, if the face measure w is invariant under the action of G on $\mathcal{F}$, then the chamber walk corresponds to the random walk on G generated by the measure

$$Q(g) = \sum_{\substack{F \in \mathcal{F}: \\ F \cdot id = g}} w(F)$$

(where g is regarded as a group element on the left-hand side and as the corresponding chamber on the right-hand side, and similarly for id) [12]. The Boolean, braid, and dihedral arrangements are reflection arrangements corresponding to the Boolean group $(\mathbb{Z}/2\mathbb{Z})^n$, the symmetric group S_n, and the dihedral group D_{2n}, respectively.

2.3.1 Boolean Arrangement

The Boolean arrangement in $\mathbb{R}^n$ is the collection of coordinate hyperplanes $\mathcal{B}_n = \{H_i\}_{i=1}^n$ where $H_i = \{x \in \mathbb{R}^n : x_i = 0\}$. The chamber set of this arrangement consists of the 2^n orthants in $\mathbb{R}^n$ (which may be identified with the set $\{+,-\}^n$ or the vertices of the n-dimensional hypercube), and the set of faces correspond to the set of all 3^n possible sign sequences $\{-,0,+\}^n$. The fact that one need not worry whether a given sign sequence corresponds to a face and that the chambers may be identified with such familiar objects makes the Boolean arrangement one of the simplest arrangements to think about.

It is also in some sense the most general class of hyperplane arrangements since every hyperplane walk can be embedded into a walk on a Boolean arrangement. Specifically,

given a hyperplane arrangement $\mathcal{A} = \{H_i\}_{i=1}^m$ in $\mathbb{R}^n$, n arbitrary, and a measure w on the face semigroup $\mathcal{F}(\mathcal{A})$, one can construct a walk on the arrangement $\mathcal{B}_m = \{x \in \mathbb{R}^m : x_i = 0\}_{i=1}^m$ by defining $\widetilde{w}(\widetilde{F}) = w(F)$ if $\widetilde{F} \in \mathcal{F}(\mathcal{B}_m)$ has the same sign sequence as $F \in \mathcal{F}(\mathcal{A})$ and $\widetilde{w}(\widetilde{F}) = 0$ if there is no $F \in \mathcal{F}(\mathcal{A})$ with the same sign sequence as $\widetilde{F}$. Because $\mathcal{F}(\mathcal{A})$ is a semigroup, we see that this induced walk on $\mathcal{B}_m$ is identical to the walk on $\mathcal{A}$ provided that we begin at some chamber in $\mathcal{C}(\mathcal{B}_m)$ whose sign sequence corresponds to an element of $\mathcal{C}(\mathcal{A})$. If w is separating as a measure on $\mathcal{F}(\mathcal{A})$, then $\mathcal{C}(\mathcal{A}) \subseteq \mathcal{C}(\mathcal{B}_m)$ is an absorbing set regardless of the initial state. This means that the transition matrix for the walk on the larger state space $\mathcal{C}(\mathcal{B}_m)$ can be written as

$$\widetilde{P} = \begin{bmatrix} P & 0 \\ Q & R \end{bmatrix}$$

where P is the transition matrix for the original walk. One advantage of this perspective is that the spectrum of P is contained in the spectrum of $\widetilde{P}$ and it turns out that one can analyze walks on a Boolean arrangement using much more elementary machinery than is required for a walk on an arbitrary arrangement. Moreover, this procedure applies even when the original arrangement is noncentral, so one can answer many questions about walks on noncentral (or nonessential) arrangements by embedding them in to the appropriate Boolean arrangement. This notion of enlarging the state space also comes into play when considering walks on the faces of an arrangement and involves many of the same ideas.

A concrete example of a Markov chain which can be modeled in terms of a random walk on the Boolean arrangement $\mathcal{B}_n = \{x \in \mathbb{R}^n : x_i = 0\}_{i=1}^n$ is given by defining $w(F) = \frac{1}{2n}$ if the sign sequence of F contains exactly $n-1$ zeros and $w(F) = 0$ otherwise.

Thus the walk evolves by picking a coordinate of $x \in \{\pm 1\}^n$ at random and then flipping a fair coin to decide whether to set that coordinate to 1 or -1. This is just the lazy nearest-neighbor walk on the hypercube, which is closely related to the Ehrenfest urn model [48]. When we sample without replacement from the support of w to obtain $F_1, ..., F_{2n}$, we see that the chamber $F_1 \cdots F_{2n}$ is distributed according to the uniform distribution on $\{\pm 1\}^n$. Because the intersection lattice $\mathcal{L}(\mathcal{B}_n)$ is isomorphic to the lattice of subsets of $[n]$ with $W \in \mathcal{L}(\mathcal{B}_n)$ corresponding to the set $s_W = \{i \in [n] : x_i = 0 \text{ on } W\}$, it follows from Theorem 2.2.1 that each subset $s \subseteq [n]$ with cardinality $|s| = k$ contributes an eigenvalue

$$\lambda_W = \sum_{F \subseteq W} w(F) = 2(n-k)\frac{1}{2n} = 1 - \frac{k}{n}$$

of multiplicity $m_W = |\mu(V, W)| = \left|(-1)^k\right| = 1$ (as the Möbius function on the intersection lattice agrees with the Möbius function on the subset lattice evaluated at the corresponding subset). Because there are $\binom{n}{k} = \binom{n}{n-k}$ subsets of $[n]$ of size k, this agrees with the classical eigenvalue result due to Mark Kac. The upper bound from Theorem 2.2.2 gives

$$\left\| P_C^l - \pi \right\|_{TV} \leq \sum_{H \in \mathcal{A}} \lambda_H^l = n(1 - \frac{1}{n})^l,$$

hence $n\log(n) + cn$ steps suffice to make the total variation distance less than e^{-c}. It is known that the correct bound is $\frac{1}{2}n\log(n) + cn$, so this bound is good but not perfect [12].

Another way to conceive of random walks on the chambers of Boolean arrangements is in terms of conquering territories [12]. The idea here is that each $C \in \mathcal{C} \sim x \in \{+, -\}^n$ may be regarded as a landscape consisting of n sites, each of which can be in one of two states, "+" or "−". The action of a face F on C corresponds to changing the state of site

i to "+" if $\sigma_i(F) = +$, to "−" if $\sigma_i(F) = -$, and leaving the state as is if $\sigma_i(F) = 0$. Thus we can think of the random walk as modeling two opposing forces engaged in successive campaigns for control of the various sites. For example, a random walk on $\mathcal{B}_5$ corresponds to an ongoing war over a region divided into five sites. If the current state of affairs is $X_i \sim (+, +, -, +, -)$ (so that "+" controls sites 1, 2, and 4, and "−" controls sites 3 and 5) and the result of campaign $i + 1$ is $F_{i+1} \sim (-, +, 0, 0, +)$ (in which battles occur at sites 1, 2, and 5 with "+" the victor in sites 2 and 5 and "−" winning at site 1), then the new division of territories is $X_{i+1} = F_{i+1}X_i \sim (-, +, -, +, +)$. Example 3 in section 3B of [12] presents a simplified model of this setup in which the respective forces attack only from the left or right so that the face measure is supported on those $F \in \mathcal{F}$ such that for some $1 \leq j \leq n$, $\sigma_i(F) = +$ for $i = 1, ..., j$ and $\sigma_i(F) = -$ for $i = j + 1, ..., n$. (A similar scenario arises if the face measure is supported on those $F \in \mathcal{F}$ such that there exist $1 \leq j \leq k \leq n$ with $\sigma_i(F) = +$ for $i = 1, ..., j$, $\sigma_i(F) = -$ for $i = j+1, ..., k-1$, and $\sigma_i(F) = -$ for $i = k + 1, ..., n$.)

We will consider the following generalization of this militaristic model. Consider two forces battling over a region which is divided into territories $T_1, ..., T_n$. Each territory has independent battle propensity $b_i = \mathbb{P}\{T_i \text{ is involved in a battle}\}$, and outcome probabilities

$$p_S(i) = \mathbb{P}\{+ \text{ wins a battle at } T_i | \text{battles occur at sites in } S\},$$

$$q_S(i) = \mathbb{P}\{- \text{ wins a battle at } T_i | \text{battles occur at sites in } S\} = 1 - p_S(i)$$

for $S \subseteq [n]$. The corresponding hyperplane walk is given by assigning face probabilities

$$w(F) = \prod_{i \in A_F} (1 - b_i) \prod_{j \notin A_F} b_i \left[p_{A_F^C}(j) 1_{\{+\}}(\sigma_j(F)) + q_{A_F^C}(j) 1_{\{-\}}(\sigma_j(F)) \right]$$

where $A_F = \{i \in [n] : \sigma_i(F) = 0\}$ is the support set of F. Thus, when $n = 5$, the probability of a campaign resulting in $F \sim (0, +, -, 0, +)$ is given by

$$w(F) = (1 - b_1) b_2 p_{\{2,3,5\}}(2) b_3 q_{\{2,3,5\}}(3) (1 - b_4) b_5 p_{\{2,3,5\}}(5).$$

Now each $W \in \mathcal{L}(\mathcal{B}_n)$ corresponds to its support set $A_W \subseteq [n]$ and thus, by Theorem 2.2.1, contributes an eigenvalue

$$\lambda_W = \sum_{F \subseteq W} w(F) = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = 0 \; \forall \; i \in A_W}} w(F) = \prod_{i \in A_W} (1 - b_i)$$

of multiplicity $m_W = |\mu(V, W)| = 1$. In particular, the eigenvalues, and thus the upper bounds on distance to stationarity from Theorem 2.2.2, do not depend on the outcome probabilities $p_S(1), ..., p_S(n)$. Letting $b_* = \min_{i \in [n]} b_i$, the crude upper bound in Theorem 2.2.2 gives

$$\left\| P_C^l - \pi \right\|_{TV} \leq \sum_{i=1}^{n} (1 - b_i)^l \leq n(1 - b_*)^l \leq n e^{-l b_*},$$

so that $l \geq b_*^{-1}(\log(n) + c)$ implies $\left\| P_C^l - \pi \right\|_{TV} \leq e^{-c}$ for all $c > 0$.

Observe that this example may also be modeled as a generalized product chain on $\{+, -\}^n$. Specifically, for each $i \in [n]$, $S \subseteq [n]$, define

$$P_i^{(S)} = \begin{bmatrix} p_S(i) & q_S(i) \\ \\ p_S(i) & q_S(i) \end{bmatrix}, \quad Q(S) = \prod_{i \in S} b_i \prod_{j \notin S} (1 - b_j).$$

Then the transition probabilities are

$$P(x,y) = \sum_{S\subseteq[n]} Q(S) \prod_{i\in S} P_i^{(S)}(x_i, y_i) \prod_{j\notin S} 1\{y_j = x_j\}.$$

However, unless the parameters $p_S(i)$ are independent of S for each i (that is, the outcome probabilities do not depend on where the battles are being fought), the collections $\{P_i^{(S)}\}_{S\subseteq[n]}$ do not have the same eigenvectors, so the methods of subsection 1.3.2 do not apply (though we can still recover some of the eigenfunctions using techniques to be introduced in the following chapter). If $p_S(i) = p_i$ for all $i \in [n]$, $S \subseteq [n]$, then each $P_i^{(S)} = P_i = \begin{bmatrix} p_i & 1-p_i \\ p_i & 1-p_i \end{bmatrix}$ has eigenvalues 1 and 0 corresponding to right eigenvectors $\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}p_i-1\\p_i\end{smallmatrix}\right]$ and it follows from Theorem 1.3.3 that P has eigenvalues

$$\sum_{S\subseteq[n]} \prod_{i\in S} b_i \prod_{j\notin S} (1-b_j) \prod_{i\in S} z_i$$

as z ranges over $\mathbb{Z}_2^n$. Thus for each $z \in \mathbb{Z}_2^n$, letting $S_z = \{i \in [n] : z_i = 1\}$, we see that the corresponding eigenvalue is

$$\sum_{S\subseteq S_z} \prod_{i\in S} b_i \prod_{j\notin S} (1-b_j) = \prod_{j\notin S_z} (1-b_j).$$

This agrees with the derivation in terms of the theory of random walks on hyperplane arrangements, but Theorem 1.3.3 also gives every eigenfunction (in terms of tensor products of the component eigenfunctions) as well. Specifically, the eigenvalue $\lambda_W = \prod_{i\in A_W}(1-b_i)$ corresponds to the right eigenfunction defined by $\phi_W(C) = \prod_{i\in A_W}[p_i - 1_{\{+\}}(\sigma_i(C))]$. The left eigenfunctions can be computed similarly.

A variation on the above model would be to allow for battles to end in a draw, in which case the state of the corresponding territory remains unchanged. This is equivalent

to saying that the current occupant of a site has an advantage in battles over that site. To make this concrete, we set

$$\begin{aligned} p_S(i) &= \mathbb{P}\{+ \text{ wins a battle at } T_i | \text{battles occur at sites in } S\}, \\ q_S(i) &= \mathbb{P}\{- \text{ wins a battle at } T_i | \text{battles occur at sites in } S\}, \\ r_S(i) &= \mathbb{P}\{\text{The battle at } T_i \text{ ends in a draw} | \text{battles occur at sites in } S\} \\ &= 1 - p_S(i) - q_S(i) \end{aligned}$$

for each $i \in [n]$, $S \subseteq [n]$ and define the face measure w by

$$\begin{aligned} w(F) = \sum_{S \subseteq A_F} \prod_{i \in S} (1 - b_i) \prod_{j \in A_F \setminus S} b_j r_{S^C}(j) \\ \times \prod_{k \in A_F^C} b_k \left[p_{S^C}(k) 1_{\{+\}}(\sigma_k(F)) + q_{S^C}(k) 1_{\{-\}}(\sigma_k(F)) \right]. \end{aligned}$$

In this case, Theorem 2.2.1 implies that each $W \in \mathcal{L}(\mathcal{B}_n)$ contributes an eigenvalue

$$\lambda_W = \sum_{F \subseteq W} w(F) = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = 0 \, \forall \, i \in A_W}} w(F) = \sum_{S \subseteq A_W} \prod_{i \in S} (1 - b_i) \prod_{j \in A_W \setminus S} b_j r_{S^C}(j)$$

of multiplicity $m_W = |\mu(V, W)| = 1$. We can also represent this as a product chain by taking

$$P_i^{(S)} = \begin{bmatrix} 1 - q_S(i) & q_S(i) \\ p_S(i) & 1 - p_S(i) \end{bmatrix}, \quad Q(S) = \prod_{i \in S} b_i \prod_{j \notin S} (1 - b_j)$$

for each $i \in [n]$, $S \subseteq [n]$. Then the transition probabilities are

$$P(x, y) = \sum_{S \subseteq [n]} Q(S) \prod_{i \in S} P_i^{(S)}(x_i, y_i) \prod_{j \notin S} 1\{y_j = x_j\}.$$

If we suppose that the two sides are evenly matched in the sense that $p_S(i) = q_S(i)$ for all $i \in [n]$, $S \subseteq [n]$, then the component kernels are given by

$$P_i^{(S)} = \begin{bmatrix} 1 - p_S(i) & p_S(i) \\ p_S(i) & 1 - p_S(i) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 - 2p_S(i) \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & r_S(i) \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^{-1}.$$

Because $\{P_i^{(S)}\}_{S \subseteq [n]}$ is simultaneously diagonalizable, the argument from Theorem 1.3.3 shows that the eigenfunctions of the product chain are given by the tensor products of the component eigenfunctions, $\phi_i^{(0)}(\pm) = 1$ and $\phi_i^{(1)}(\pm) = \pm 1$.

These military models are primarily interesting because they show that one can sometimes model hyperplane walks as product chains and vice versa. When viewed as product chains, one may sometimes be able to extract all of the eigenfunctions using Theorem 1.3.3. Conversely, while it may be difficult to compute the spectrum of some generalized product chains using Theorem 1.3.3, if they can be modeled as hyperplane walks, then the task may be much simpler.

Before moving on to other types of arrangements, we remark that a wide variety of coupon collecting problems can be modeled as random walks on Boolean arrangements. The setup here is that we are trying to obtain a complete set of coupons $\{c_1, ..., c_N\}$ by successively choosing books of coupons with various probabilities. Suppose that there are

m different types of books $B_k = \{c_{i_1}, ..., c_{i_{n_k}}\}$, $k \in [m]$, and the probability of choosing book B_k at any stage is p_k. We may associate each book B_k with the face F_k defined by

$$\sigma_j(F_k) = \begin{cases} +, & c_j \in B_k \\ 0, & c_j \notin B_k \end{cases}$$

and define the measure w on $\mathcal{F}(\mathcal{B}_n)$ by $w(F_k) = p_k$ for $k = 1, ..., n$. Letting $\{X_k\}$ denote the BHR walk on $\mathcal{B}_n$ driven by w and having initial state $X_0 \sim (-, ..., -)$, the coupons collected stage k are precisely those c_i with of $\sigma_i(X_k) = +$. Note that w is separating if and only if every coupon is contained in some book. In this case, the stationary distribution is the point mass at $C \sim (+, ..., +)$. Also, this process is equivalent to the walk on $\mathcal{F}(\mathcal{B}_n)$ started at $X_0 = \mathcal{O}$, so our previous analysis shows that the total variation distance to stationarity is equal to the bounds in Theorems 2.2.2 and 2.2.3. In many ways, this example encapsulates the main ideas underlying convergence of hyperplane chamber walks. In particular, for the chain to equilibriate, one needs to "collect" faces with nonzero sign-sequence coordinates for each i.

2.3.2 Braid Arrangement

Another important example is the braid arrangement on $\mathbb{R}^n$. This is the set of all $\binom{n}{2}$ hyperplanes of the form $\{x \in \mathbb{R}^n : x_i - x_j = 0\}$. The chambers of this arrangement can be indexed by the symmetric group on n symbols with $\pi \in S_n$ corresponding to the chamber $C_\pi = \{x \in \mathbb{R}^n : x_{\pi(1)} > x_{\pi(2)} > ... > x_{\pi(n)}\}$. This is because the points in a given chamber are not contained in any hyperplane $\{x_i = x_j\}$, thus all coordinates of such points must differ. The relative ordering of the coordinates of two points in the same chamber are identical since one must pass through a hyperplane to change that

ordering. Similarly, the faces of the braid arrangement are in bijective correspondence with the ordered set partitions of $[n]$, the partition $(B_1, ..., B_l)$ corresponding to the face consisting of all points x such that $x_i = x_j$ if $i, j \in B_r$ for some $r = 1, ..., l$ and $x_i > x_j$ if $x_i \in B_r$, $x_j \in B_s$ for some $1 \leq r < s \leq l$. (The chambers are the faces corresponding to partitions where all blocks have size 1.) In terms of sign sequences, if $F \sim (B_1, ..., B_l)$, then the $(i,j)th$ coordinate of $\sigma(F)$ - that is, the coordinate indicating which side of the hyperplane $\{x_i = x_j\}$ F belongs to - is given by

$$\sigma_{(i,j)}(F) = \begin{cases} -, & i \in B_s,\ j \in B_r \text{ for some } 1 \leq r < s \leq l \\ 0, & i, j \in B_r \text{ for some } 1 \leq r \leq l \\ +, & i \in B_r,\ j \in B_s \text{ for some } 1 \leq r < s \leq l \end{cases}.$$

(Unless specifically stated otherwise, we will take it as implicit when indexing objects corresponding to the braid arrangement with the ordered pair (i, j) that $i < j$.)

When we identify the chambers with S_n so that $\pi \in S_n$ corresponds to the chamber $C_\pi = \{x_{\pi(1)} > ... > x_{\pi(n)}\}$, then $\sigma(C_\pi)$ is related to the inversion set of π^{-1}, where the inversion set of $\tau \in S_n$ is defined as $\text{Inv}(\tau) = \{(i,j) : i < j,\ \tau(i) > \tau(j)\}$. Specifically, since for all $i < j$, $\sigma_{(i,j)} = -$ if and only if $j = \pi(r)$ and $i = \pi(s)$ with $r < s$ if and only if $(i, j) \in \text{Inv}(\pi^{-1})$, we have

$$\sigma_{(i,j)}(\pi) := \sigma_{(i,j)}(C_\pi) = \begin{cases} -, & (i,j) \in \text{Inv}(\pi^{-1}) \\ +, & (i,j) \notin \text{Inv}(\pi^{-1}) \end{cases}.$$

In particular, we see that the identity permutation has sign sequence $\sigma(id) = (+, ..., +)$ and the permutation π_{rev} defined by $\pi_{rev}(k) = n - k + 1$ has sign sequence $\sigma(\pi_{rev}) = (-, ..., -)$.

The face $F \sim (B_1, ..., B_k)$ acts on the face $G \sim (C_1, ..., C_l)$ by refinement so that $FG \sim (B_1 \cap C_1, ..., B_1 \cap C_l, ..., B_k \cap C_1, ..., B_k \cap C_l)^\wedge$ where the hat means "delete empty intersections." If you think of the chamber C_τ as a deck of n cards ordered with the card labeled $\tau(i)$ in the ith position from the top, then the face $F \sim (B_1, ..., B_k)$ acts on C_τ by removing all cards with labels in B_1 and placing them on top, retaining their relative order, then removing all cards with labels in B_2 and placing them next, and so on. For example, take $n = 7$ and consider the faces $F \sim (\{1, 2\}, \{5\}, \{3, 4, 7\}, \{6\})$, $G \sim (\{5, 7\}, \{1, 2, 6\}, \{3, 4\})$ and the chamber $C \sim (3625174)$. Then we have $FG \sim (\{1, 2\}, \{5\}, \{7\}, \{3, 4\}, \{6\})$ and $FC \sim (2153746)$.

As a first example of random walks on braid arrangements, consider the probability measure that assigns weight $w_i \geq 0$ to the face F_i corresponding to the ordered set partition $(\{i\}, [n] \setminus \{i\})$ for $i = 1, ..., n$ where $\sum_{i=1}^{n} w_i = 1$. If we think of chambers as orderings of a deck of cards, then the walk proceeds by picking the card labeled i with probability w_i and placing it on top of the deck. This model appears under the name of the Tsetlin library in the theory of dynamic file management. If we picture the chambers as stacks of files with file i being used with propensity w_i, then the walk corresponds to placing a file on the top of the stack every time it is used. Over time, the most used files will tend to be near the top. If the weights are all equal, $(w_1 = ... = w_n = \frac{1}{n})$, then the chamber walk is equivalent to the random-to-top shuffle in which at each stage a card is chosen uniformly at random and placed on top of the deck. This is the inverse of the more commonly discussed top-to-random shuffle and it follows from the basic theory of random walks on groups that both shuffles mix at the same rate. Since a face F is contained in the hyperplane $H_{(i,j)} = \{x_i - x_j = 0\}$ if and only if i and j

are contained in the same block in the partition corresponding to F, it follows from Theorem 2.2.1 that, for the random-to-top walk, the eigenvalues corresponding to the hyperplanes are given by $\lambda_{H_{(i,j)}} = \sum_{F \subseteq H_{(i,j)}} w(F) = 1 - \frac{2}{n}$ for all $i \neq j$ (as every face in the support of w is contained in $H_{(i,j)}$ except the ones corresponding to the partitions $(\{i\}, [n] \setminus \{i\})$ and $(\{j\}, [n] \setminus \{j\})$). Consequently, we have the upper bound $\left\| P_C^l - \pi \right\|_{TV} \leq \sum_{H \in \mathcal{A}} \lambda_H^l = \binom{n}{2}(1 - \frac{2}{n})^l$, so the total variation distance is less than $\frac{1}{2}e^{-c}$ after $n \log(n) + cn$ shuffles. Previous work of Diaconis and Aldous [3] and Diaconis, Fill, and Pitman [24] shows that this bound is sharp. In the case of the Tsetlin library, the upper bound is $\left\| P_C^l - \pi \right\|_{TV} \leq \sum_{1 \leq i < j \leq n} (1 - w_i - w_j)^l$. The largest terms in this sum correspond to the smallest values of $w_i + w_j$, or the least frequently used files. This also makes sense from the coupling perspective since the chain equilibriates once all files have been used at least once, so it should take longer when some of the files are used very rarely.

Before continuing our analysis of walks on the braid arrangement, we observe that random-to-top shuffles also admit a product chain description. The idea is inspired by the standard correspondence between permutations and the chambers in the braid arrangement. Recall that we have been identifying the permutation π with the vector $\sigma(\pi) \in \{-, +\}^{\{(i,j) \in \mathbb{N}^2 : 1 \leq i < j \leq n\}}$ defined by

$$\sigma_{(i,j)}(\pi) = \begin{cases} -, & (i,j) \in \mathrm{Inv}(\pi^{-1}) \\ +, & (i,j) \notin \mathrm{Inv}(\pi^{-1}) \end{cases}.$$

Now consider the chain on $\Omega = \{-, +\}^{\{(i,j) \in \mathbb{N}^2 : 1 \leq i < j \leq n\}}$ (which is a strictly larger state space than the set of sign sequences of chambers in the braid arrangement) which proceeds

by picking a number k uniformly from $[n]$ and then setting $\sigma_{(i,k)} = -$ and $\sigma_{(k,l)} = +$ for all $i < k < l$. If the chain begins at a state $x = \sigma(\pi)$ for some $\pi \in S_n$, then this is equivalent to the random-to-top shuffle since we are just declaring that k precedes all numbers in the list $\pi(1), \pi(2), ..., \pi(n)$ and the relative positions of all other numbers remain unchanged. To better visualize this, think of π as representing a deck of cards labeled 1 through n with $\pi(r)$ the label of the card in position r. Then $\pi^{-1}(s)$ gives the position of the card labeled s. Thus the card labeled k is on the top of the deck if and only if $\pi^{-1}(k) < \pi^{-1}(r)$ for all $r \in [n] \setminus \{k\}$. This is equivalent to requiring that $(i, k) \in \text{Inv}(\pi^{-1})$ for all $i < k$ and $(k, j) \notin \text{Inv}(\pi^{-1})$ for all $j > k$. Because all coordinates of $\sigma(\pi)$ which do not involve k stay the same, the effect of the proposed transition is to place card k at the top of the deck originally ordered as π and leave all other cards in the same relative order. This shows that the chain restricted to $\mathcal{S} = \{x \in \Omega : x = \sigma(\pi) \text{ for some } \pi \in S_n\}$ is equivalent to the random-to-top shuffle and that $\mathcal{S}$ is a closed and irreducible subset of Ω. Consequently, we may choose a basis so that this Markov chain defined on Ω has transition matrix

$$P = \begin{bmatrix} R & 0 \\ M & N \end{bmatrix}$$

where R is the transition matrix for the random-to-top chain on S_n. Accordingly, if we can find eigenfunctions for P, then they will restrict to an eigenfunctions of R.

Now we can represent P as a linear combination of tensor product chains as in subsection 1.3.2. To do so, let Q be uniform on $[n]$ and for each $k \in [n]$, $1 \leq i < j \leq n$, define transition matrices on $\Omega_{(i,j)} = \{-, +\}$ by

$$P^{(k)}_{(k,j)} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad P^{(k)}_{(i,k)} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \quad P^{(k)}_{(i,j)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

for all $k \neq i, j$. It follows from the definition of the chain and the results of subsection 1.3.2 that

$$P = \sum_{k=1}^{n} Q(k) \bigotimes_{(i,j)} P^{(k)}_{(i,j)} = \frac{1}{n} \sum_{k=1}^{n} \bigotimes_{(i,j)} P^{(k)}_{(i,j)}$$

where the product is taken in lexicographic order. The family $\left\{P^{(k)}_{(i,j)}\right\}_{k=1}^{n}$ is not simultaneously diagonalizable, so Theorem 1.3.3 and its generalizations are not directly applicable. Nonetheless, we may guess that certain eigenfunctions can be represented as tensor products of functions. Since $\phi_0 \equiv 1$ is an eigenfunction for each $P^{(k)}_{(i,j)}$, if we assume that most of the components are copies of ϕ_0, then our analysis will be greatly simplified. Specifically, given any $i < j$, let $\varphi_{(i,j)} = \bigotimes_{(k,l)} \phi_{(k,l)}$ where $\phi_{(k,l)} \equiv 1$ when $(k,l) \neq (i,j)$ and $\phi_{(i,j)} : \Omega_{(i,j)} \to \mathbb{R}$ is yet to be determined. Then for any $x \in \Omega = \bigotimes_{(i,j)} \Omega_{(i,j)}$, we have

$$\begin{aligned} P\varphi_{(i,j)}(x) &= \sum_{y} P(x,y)\varphi_{(i,j)}(y) = \sum_{y} \varphi_{(i,j)}(y) \left(\frac{1}{n} \sum_{r=1}^{n} \bigotimes_{(k,l)} P^{(r)}_{(k,l)}(x,y) \right) \\ &= \frac{1}{n} \sum_{r=1}^{n} \left(\sum_{y} \varphi_{(i,j)}(y) \bigotimes_{(k,l)} P^{(r)}_{(k,l)}(x,y) \right) \\ &= \frac{1}{n} \sum_{r=1}^{n} \prod_{(k,l)} \left(\sum_{y_{(k,l)} \in \Omega_{(k,l)}} \phi_{(k,l)}(y_{(k,l)}) P^{(r)}_{(k,l)}(x_{(k,l)}, y_{(k,l)}) \right) \end{aligned}$$

$$= \frac{1}{n}\sum_{r=1}^{n}\sum_{y_{(i,j)}\in\Omega_{(i,j)}} \phi_{(i,j)}(y_{(i,j)})P^{(r)}_{(i,j)}(x_{(i,j)}, y_{(i,j)})$$

$$= \sum_{y_{(i,j)}\in\Omega_{(i,j)}} \phi_{(i,j)}(y_{(i,j)}) \left(\frac{1}{n}\sum_{r=1}^{n} P^{(r)}_{(i,j)}(x_{(i,j)}, y_{(i,j)})\right)$$

Thus in order to have

$$\begin{aligned} P\varphi_{(i,j)}(x) &= \lambda\varphi_{(i,j)}(x) \\ &= \lambda\left(\prod_{(k,l)<(i,j)} \phi_{(k,l)}(x_{(k,l)})\right)\phi_{(i,j)}(x_{i,j})\left(\prod_{(k,l)<(i,j)} \phi_{(k,l)}(x_{(k,l)})\right) \\ &= \lambda\phi_{(i,j)}(x_{i,j}), \end{aligned}$$

it is necessary and sufficient that

$$\sum_{y_{(i,j)}\in\Omega_{(i,j)}} \phi_{(i,j)}(y_{(i,j)}) \left(\frac{1}{n}\sum_{r=1}^{n} P^{(r)}_{(i,j)}(x_{(i,j)}, y_{(i,j)})\right) = \lambda\phi_{(i,j)}(x_{(i,j)}).$$

It follows from the definition of $\left\{P^{(k)}_{(i,j)}\right\}_{k=1}^{n}$ that the left-hand side of the above equation may be more compactly represented as $P_{(i,j)}\phi_{(i,j)}(x_{(i,j)})$ where

$$\begin{aligned} P_{(i,j)} &= \frac{1}{n}\left(\sum_{k\neq i,j} I + P^{(k)}_{(i,k)} + P^{(k)}_{(k,j)}\right) \\ &= \frac{1}{n}\left(\begin{bmatrix} n-2 & 0 \\ 0 & n-2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}\right) \\ &= \begin{bmatrix} \frac{n-1}{n} & \frac{1}{n} \\ \frac{1}{n} & \frac{n-1}{n} \end{bmatrix}, \end{aligned}$$

hence $\varphi_{(i,j)}(y)$ will be an eigenfunction of P with eigenvalue λ if and only if $\phi_{(i,j)}$ is an eigenfunction of $P_{(i,j)}$ with eigenvalue λ. One readily checks that $P_{(i,j)}$ has eigenvalues

$\lambda = 1$ and $\lambda = 1 - \frac{2}{n}$ and corresponding eigenvectors $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right]$. Thus a nontrivial eigenfunction is obtained by defining

$$\phi_{(i,j)}(x_{(i,j)}) = \begin{cases} -1, & x_{(i,j)} = - \\ 1, & x_{(i,j)} = + \end{cases},$$

so that

$$\varphi_{(i,j)}(x) = \begin{cases} -1, & x_{(i,j)} = - \\ 1, & x_{(i,j)} = + \end{cases}$$

is an eigenfunction of P with eigenvalue $1 - \frac{2}{n}$. Accordingly, for each $1 \leq i < j \leq n$, restricting $\varphi_{(i,j)}$ to $\mathcal{S}$ shows that

$$\varphi^{\#}_{(i,j)}(\pi) = \begin{cases} -1, & (i,j) \in \mathrm{Inv}(\pi^{-1}) \\ 1, & (i,j) \notin \mathrm{Inv}(\pi^{-1}) \end{cases}$$

is an eigenfunction for the random-to-top shuffle with eigenvalue $\lambda_{(i,j)} = 1 - \frac{2}{n}$. We will obtain the same result with less work in chapter 3, but it is interesting to note that some information can be gained from the product chain perspective even when the conditions of Theorem 1.3.3 are not met. Observe that the above procedure is applicable in the analysis of other generalized product chains as well and essentially amounts to lumping the product chain according to the equivalence relation defined by agreement in a particular (subset of) coordinate(s).

Another example involving the braid arrangement is the inverse of the famous riffle shuffle introduced by Gilbert, Shannon, and Reeds. In this case, the distribution of w is uniform over all two-block partitions $(S, [n] \setminus S)$, $S \subseteq [n]$ (including the degenerate cases $S = \emptyset$ and $S = [n]$). In terms of card shuffling, multiplication by the face $F \sim (S, [n] \setminus S)$

corresponds to "unriffling" the deck by moving all cards indexed by S to the front of the deck while retaining their original order. Since all two-block partitions are equally likely, we are essentially flipping a fair coin n times, marking the ith card if the ith flip results in a heads, and then moving all of the marked cards to the front of the deck. This is precisely the inverse of the GSR shuffle as explained in [6]. The bound from Theorem 2.2.2 works out to be $\left\|P_C^l - \pi\right\|_{TV} \leq \binom{n}{2}\frac{1}{2^l}$, hence $2\log_2(n)+c-1$ steps make the distance to uniformity less than 2^{-c}. The correct mixing time is known to be about $\frac{3}{2}\log_2(n)$, so, as with the Ehrenfest urn, the bound is good but not optimal [12]. More generally, if we assign weight a^{-n} to each of the a^n ordered partitions of $[n]$ into a blocks (some of which may be empty), then the corresponding hyperplane walk is equivalent to inverse a-shuffles. (An a-shuffle results from successively cutting the deck into a piles and then dropping cards from each pile with probability proportional to the pile's size [6].) Then every face $F_{B_1,...,B_a} \sim (B_1, ..., B_a)$ has sign sequence

$$\sigma_{(i,j)}(F_{B_1,...,B_a}) = \begin{cases} -, & i \in B_l, j \in B_k, 1 \leq k < l \leq a \\ 0, & i, j \in B_k, 1 \leq k \leq a \\ +, & i \in B_k, j \in B_l, 1 \leq k < l \leq a \end{cases},$$

so each of the $\binom{n}{2}$ hyperplanes contributes the eigenvalue

$$\lambda_{H_{(i,j)}} = \frac{1}{a^n}\left|\{(B_1, ..., B_a) : i, j \in B_k \text{ for some } k \in [a]\}\right| = \frac{1}{a^n}a^{n-2} \cdot a = \frac{1}{a}$$

as the number of ordered partitions of $[n]$ into a blocks such that i and j are contained in a common block can be enumerated by first picking one of the a^{n-2} ordered partitions of $[n] \setminus \{i, j\}$ into a blocks and then choosing which of the a blocks to add $\{i, j\}$ to.

We can also think of the faces of the braid arrangement as representing non-strict preferences where the face $F \sim (B_1, ..., B_k)$ corresponds to a ranking of n alternatives with the alternatives indexed by B_1 being the most preferred (but deemed equal to one another), those indexed by B_2 being the second most preferred, and so on. The chambers then correspond to strict linear rankings. If k judges rank n items, then we can define the face measure w by letting $w(F)$ be equal to the number of judges with preferences corresponding to F divided by k. This measure is separating as long as no two alternatives are tied in each judge's ranking. A step in this walk corresponds to picking a judge at random and then letting them update the current ranking by moving their favorite alternatives to the top of the list (with tied alternatives retaining their original order), then moving their second favorite alternatives to below those, and so on. This provides a means of passing from a distribution on non-strict rankings, w, to a distribution on strict rankings, π, the stationary measure of the hyperplane walk. Hiroaki Terao has demonstrated the applicability of the hyperplane perspective to voting theory by establishing a version of Arrow's impossibility theorem as a corollary of a result concerning *admissible mappings* on the chambers of hyperplane arrangements [68], and the foregoing suggests other possible connections.

Finally, the faces $F \sim (B_1, ..., B_k)$ with $|B_1| \geq ... \geq |B_k|$ are in an obvious bijective correspondence with Young tabloids (Young tableaux modulo row equivalence). The action of permutations on tabloids is fundamental to the representation theory of S_n, and it might be interesting to study how tabloids act on permutations in terms of the face product. One natural measure would be $w(F) = \frac{f_\lambda}{n!}$ if $F \sim (B_1, ..., B_k)$ where B_i consists of the numbers in the ith row of a standard tableau of shape λ and f_λ denotes

the number of such tableaux. This is a probability measure since $\sum_{\lambda \vdash n} \frac{f_\lambda^2}{n!} = 1$ by RSK correspondence, and there are clear connections with the Plancherel measure and the representation theory of S_n. It may also be interesting to examine the dynamics of the random walk induced by letting w be uniformly distributed over the $\binom{n}{\lambda_1,...,\lambda_k}$ $F \in \mathcal{F}$ such that $F \sim (B_1, ..., B_k)$ with $|B_i| = \lambda_i$ for various choices of $\lambda = (\lambda_1, ..., \lambda_k) \vdash n$. Analyzing these walks seems to be quite a formidable task, but it would almost certainly yield some fascinating mathematics.

2.3.3 Dihedral Arrangement

Let $\mathcal{A}$ consist of m lines through the origin in $\mathbb{R}^2$. We will assume throughout that the lines are equally spaced so that $\mathcal{A}$ is the reflection arrangement of the dihedral group of order $2m$. The hyperplanes may be written in polar coordinates as $H_i = \{(r, \theta) : r \in \mathbb{R}, \theta = \frac{\pi i}{m}\}$. The $2m$ "wedges" $C_j = \left\{(r, \theta) : r > 0, \theta \in \left(\frac{\pi(j-1)}{m}, \frac{\pi j}{m}\right)\right\}$ are the chambers of this arrangement, the $2m$ rays $R_j = \left\{(r, \theta) : r > 0, \theta = \frac{\pi(j-1)}{m}\right\}$ are the one-dimensional faces, and the origin is a zero-dimensional face, hence $|\mathcal{F}| = 4m + 1$. If we orient the hyperplanes so that $\sigma(C_1) = (+, ..., +)$, then the chambers have sign sequences given by

$$\sigma_i(C_j) = \begin{cases} +, & j \leq i \text{ or } m + i < j \\ -, & \text{else} \end{cases},$$

the 1-dimensional faces have sign sequences

$$\sigma_i(R_j) = \begin{cases} +, & j < i \text{ or } m + i < j \\ 0, & i = j \text{ or } i + m = j \\ -, & \text{else} \end{cases},$$

and, of course, $\sigma(\{0\}) = (0, ..., 0)$. The intersection lattice has bottom element $V = \mathbb{R}^2$, which is covered by each of the m hyperplanes H_i, which are in turn each covered by the origin. Thus one has $\mu(V, V) = 1$, $\mu(V, H_i) = 1$, and $\mu(V, \{0\}) = m - 1$.

Define a face measure w by $w(R_j) = p_j$ for $j = 1, ..., 2m$ where $p_j \geq 0$, $\sum_{j=1}^{2m} p_j = 1$, and $p_i + p_{i+m} < 1$ for $i = 1, ..., m$. The last condition ensures that w is separating. Brown and Diaconis describe the resulting chamber walk in terms of a circular house containing $2m$ rooms (the chambers) separated by $2m$ walls (the 1-dimensional faces). The walls of the house are inhabited by a mouse and the rooms by a cat. At each time step, the mouse travels to wall R_j with probability p_j and the cat moves from its present room to the nearest room adjacent to that wall. The state of the chamber walk at time n represents the room the cat is then occupying. Alternatively, the chambers can be identified with the edges of a regular $2m$-gon and the 1-dimensional faces with its vertices. One can imagine a queuing system with service points at the vertices and a server which moves around on the edges. Requests arrive at the service point with propensities $p_1, ..., p_{2m}$ and the server moves to the nearest edge [12].

It follows from Theorem 2.2.1 that $\lambda_V = 1$ is a simple eigenvalue of the associated chamber walk, $\lambda_{\{0\}} = 0$ is an eigenvalue of multiplicity $m - 1$, and for $i = 1, ..., m$, each hyperplane contributes a single eigenvalue $\lambda_{H_i} = p_i + p_{i+m}$. In this case, the bound from Theorem 2.2.2 is $\left\| P_C^k - \pi \right\|_{TV} \leq \sum_{i=1}^{m} (p_i + p_{i+m})^k$. Now recall the description of the stationary distribution in terms of sampling faces from w in which all faces contained in the support of the product $F_1 \cdots F_k$ are discarded at the kth step and we cease sampling

once the product is a chamber. In this case, we need only sample two faces and one readily computes the stationary distribution as

$$\pi(C_j) = w(F_j)\frac{\sum_{i=j+1}^{j+m-1} w(F_i)}{1 - w(F_j) - w(F_{j+m})} + w(F_{j-1})\frac{\sum_{i=j}^{j+m} w(F_i)}{1 - w(F_{j-1}) - w(F_{j+m-1})}$$

where the indices are taken modulo $2m$.

One can construct slightly more general walks on the chambers of a dihedral arrangement by enlarging the support of the face measure to include the origin and/or some of the chambers, and indeed this may be useful for some models. However, for hyperplane walks in general, putting mass on the face $\mathcal{O} \sim (0, ..., 0)$ is equivalent to adding holding probability $w(\mathcal{O})$ to the walk driven by $\widetilde{w}(F) = (1 - w(\mathcal{O}))^{-1} w(F)1\{F \neq \mathcal{O}\}$ and so adds nothing substantially new to the analysis: If P is the walk driven by w and $\widetilde{P}$ is the walk driven by $\widetilde{w}$, then the eigenfunctions (and thus the stationary distribution) of P and $\widetilde{P}$ are the same, the eigenvalues of P are given by $w(\mathcal{O})(1 - \lambda) + \lambda$ where λ is an eigenvalue of $\widetilde{P}$, and P mixes more slowly than $\widetilde{P}$ by a factor of $w(\mathcal{O})$. Adding mass to the chambers has more dramatic effects. In particular, since the chain will be stationary as soon as a chamber has been drawn, it will mix faster than the walk with $w(\mathcal{C})$ distributed evenly over the rest of the faces and the stationary distribution will be more concentrated on and around those chambers C with $w(C) > 0$. However, one would study the behavior of such a chain at any given time by conditioning on whether or not a chamber had been sampled, so the problem essentially reduces to the case where the face measure is supported on faces of codimension at least one.

2.4 Extensions

2.4.1 Rows of Colored Tiles

We have seen that it is sometimes convenient to represent the face semigroup of a hyperplane arrangement in terms of rows of colored tiles where a face F in the arrangement $\mathcal{A} = \{H_i\}_{i=1}^m$ corresponds to a row of m tiles where the color of the ith tile corresponds to the ith coordinate of $\sigma(F)$. Multiplication is defined in terms of stacking the rows, keeping in mind that one can see through the clear tiles while the other tiles are opaque. A natural generalization of this idea is to allow for more than two colors of opaque tiles. Specifically, given a set of "colors" C, we can consider the set $RCT(C, m) = (C \cup \{0\})^m$. For $S \in RCT(C, m)$, we define $\tau_i(S) \in C \cup \{0\}$ to be the ith coordinate of S and we endow $RCT(C, m)$ with a product structure by

$$\tau_i(ST) = \begin{cases} \tau_i(S), & \tau_i(S) \neq 0 \\ \tau_i(T), & \tau_i(S) = 0 \end{cases}.$$

(Here 0 represents the clear tiles.) This makes $RCT(C, m)$ into a monoid with identity element I satisfying $\tau_i(I) = 0$ for all $i \in [m]$. It is also clear that for any $S, T \in RCT(C, m)$, we have $S^2 = S$ and $STS = ST$, thus $RCT(C, m)$ is an example of a left-regular band (which we will discuss further in the next subsection). Moreover, $\mathcal{R} = \{S \in RCT(C, m) : \tau_i(S) \neq 0 \text{ for all } i\}$ is a two-sided ideal (analogous to $\mathcal{C} \subseteq \mathcal{F}$) and many of the results for BHR walks carry over to this setting if we restrict the state space to $\mathcal{R}$.

The preceding also applies to any subsemigroup of $RCT(C,m)$. Indeed, for any arrangement $\mathcal{A} = \{H_i\}_{i=1}^m$, the face semigroup $\mathcal{F}(\mathcal{A})$ is isomorphic to the subsemigroup of $RCT(\{-,+\},m) < RCT(\{-,+,*\},m)$ given by

$$\begin{aligned} K(\mathcal{A}) &= K_{[m]}(\mathcal{A}) \\ &= \{S \in RCT(\{-,+\},m) : \exists\, F \in \mathcal{F}(\mathcal{A}) \text{ with } \tau_i(S) = \sigma_i(F)\ \forall\, i \in [m]\}. \end{aligned}$$

More generally, for any $B \subseteq [m]$, the face set of the subarrangement $\mathcal{B} = \{H_i\}_{i\in B}$ is isomorphic to the subsemigroup of $RCT(\{-,+,*\},m)$ given by

$$\begin{aligned} K(\mathcal{B}) &= K_B(\mathcal{A}) \\ &= \{S : \tau_i(S) = *\ \forall\, i \in B^C \text{ and } \exists\, F \in \mathcal{F}(\mathcal{A}) \text{ with } \tau_i(S) = \sigma_i(F)\ \forall\, i \in B\}. \end{aligned}$$

The isomorphism is given by $\phi_B(S) = F$ such that $\sigma_i(F) = \tau_i(S)$ for all $i \in B$. (We omit the subscript when $\mathcal{B} = \mathcal{A}$.)

Adopting this framework provides an interesting interpretation of projections onto subarrangements which makes the idea of "blacking out" tiles more rigorous. Suppose that $B \subseteq [m]$, and let $K(\mathcal{A})$, $K(\mathcal{B})$ be as above. Define the blackout map $g : K(\mathcal{A}) \to K(\mathcal{B})$ by

$$\tau_j(g(S)) = \begin{cases} *, & j \in B^C \\ \tau_j(S), & j \in B \end{cases}.$$

If $F \in \mathcal{F}(\mathcal{A})$ corresponds to $S = \phi^{-1}(F) \in K(\mathcal{A})$, then $g(S)$ corresponds to the face $\overline{F} = \phi_B(g(S)) \in \mathcal{F}(\mathcal{B})$, and $F \subseteq \overline{F}$ since $\sigma_i(\overline{F}) = \tau_i(g(S)) = \tau_i(S) = \sigma_i(F)$ for all $i \in B$.

For any $S, T \in K(\mathcal{A})$,

$$\tau_j(g(ST)) = \begin{cases} *, & j \in B^C \\ \tau_j(ST), & j \in B \end{cases} = \begin{cases} *, & j \in B^C \\ \tau_j(S), & j \in B, \tau_j(S) \neq 0 \\ \tau_j(T), & j \in B, \tau_j(S) = 0 \end{cases}$$

$$= \begin{cases} \tau_j(g(S)), & \tau_j(f(S)) \neq 0 \\ \tau_j(g(T)), & \tau_j(f(S)) = 0 \end{cases} = \tau_j(g(S)g(T)),$$

hence g is a semigroup homomorphism. Moreover, for any $R \in K(\mathcal{B})$, there is an $F \in \mathcal{F}(\mathcal{A})$ with $\tau_j(R) = \sigma_j(F)$ for all $j \in B$, so $R = g(S)$ where $S = \phi^{-1}(F)$, hence g is surjective as well. Therefore, by the first isomorphism theorem for semigroups, we have $K(\mathcal{B}) \cong K(\mathcal{A})/\ker(g)$. Now for any $S, T \in K(\mathcal{A})$, $g(S) = g(T)$ if and only if $\sigma_j(\phi(S)) = \tau_j(S) = \tau_j(T) = \sigma_j(\phi(T))$ for all $j \in B$. Thus the equivalence relation on $\mathcal{F}(\mathcal{A})$ given by $F \sim_B G$ if $\sigma_j(F) = \sigma_j(G)$ for all $j \in B$ is a semigroup congruence with canonical surjection $\pi_B = \phi_B \circ g \circ \phi^{-1}$, and we have $\mathcal{F}(\mathcal{A})/\sim_B \cong K(\mathcal{A})/\ker(g) \cong K(\mathcal{B}) \cong \mathcal{F}(\mathcal{B})$. Since the equivalence relation $F \sim G$ if $F, G \in \mathcal{C}(\mathcal{A})$ or $F, G \notin \mathcal{C}(\mathcal{A})$ is also a semigroup congruence, the second isomorphism theorem in universal algebra implies that if we restrict $\sim_B$ to the chambers of $\mathcal{B}$, then $\mathcal{C}(\mathcal{A})/\sim_B \cong \mathcal{C}(\mathcal{B})$. (Alternatively, one could just run through the above argument replacing $K(\mathcal{A})$ and $K(\mathcal{B})$ with their subsemigroups consisting of elements with no zero coordinates to conclude that $\mathcal{C}(\mathcal{A})/\sim_B \cong \mathcal{C}(\mathcal{B})$.)

The preceding paragraph says that we can project the faces of a hyperplane arrangement onto those of a subarrangement by sending each face $F \in \mathcal{F}(\mathcal{A})$ to the unique face in $\overline{F} \in \mathcal{F}(\mathcal{B})$ containing F in a manner which preserves the semigroup structure by introducing the equivalence relation on $\mathcal{F}(\mathcal{A})$ given by $F \sim_B G$ if and only if $\sigma_i(F) = \sigma_i(G)$

for all $i \in B$. This projection maps $\mathcal{C}(\mathcal{A})$ homomorphically onto $\mathcal{C}(\mathcal{B})$. We will be interested in lumping BHR chains using $\sim_B$ in order to find eigenfunctions via Theorem 1.3.1 and the foregoing provides a nice interpretation of the lumped chains in terms of both faces of subarrangements and rows of colored tiles. This could be accomplished without appealing to the various RCT semigroups, and indeed this is the approach taken in [5] to verify that the assumptions of an equivalent formulation of Theorem 1.3.1 are satisfied in this setting. The perspective adopted there is to view induced walks on subarrangements as functions of the original walk defined in terms of the maps which send $F \in \mathcal{F}(\mathcal{A})$ to $\overline{F} \in \mathcal{F}(\mathcal{B})$ such that $F \subseteq \overline{F}$ and verification of the properties of these maps is left to the reader. In section 3.1, we show directly that hyperplane walks are lumpable with respect to $\sim_B$ and the proof is a little more streamlined from this point of view, but the relation between $\mathcal{F}(\mathcal{A})/\sim_B$ and $\mathcal{F}(\mathcal{B})$ is not quite as transparent. The above is intended merely to clarify and make rigorous the relations between the face semigroups, congruence classes, and RCT interpretations.

Finally, we observe that in addition to providing a visual aid for hyperplane walks and a generalization thereof in which many of the proofs carry over directly, these RCT walks may be useful for modeling product chains in which the component state spaces have cardinality greater than two (in a manner analogous to the conquering territories examples) by taking $|C| = \max_i |\Omega_i|$. They also allow for generalizations of some of the models we have previously considered. For example, one could consider the case where there are more than two competing forces in the conquering territories models or more than two urns in the Ehrenfest urn example.

2.4.2 Oriented Matroids and Left-Regular Bands

Matroids are combinatorial objects introduced by Hassler Whitney in 1935 as a means of abstracting the notion of linear independence in finite dimensional vectors spaces over arbitrary fields [69]. Oriented matroids specialize to the case of dependence structures in vector spaces over ordered fields and generalize properties of directed graphs and face semigroups of hyperplane arrangements [55]. There are many ways of axiomatizing the structure of oriented matroids which are not obviously equivalent. We will adhere to the construction in [12] as it relates most clearly to faces in a central arrangement of hyperplanes. According to this definition, a set $\mathcal{X} \in \{0, \pm\}^m$ is an oriented matroid if

1. $0 = (0, ..., 0) \in \mathcal{X}$

2. $x = (x_1, ..., x_n) \in \mathcal{X}$ implies that $-x = (-x_1, ..., -x_n) \in \mathcal{X}$

3. $x, y \in \mathcal{X}$ implies $x \cdot y \in \mathcal{X}$ where

$$(x \cdot y)_i = \begin{cases} x_i, & x_i \neq 0 \\ y_i, & x_i = 0 \end{cases}$$

4. For $x, y \in \mathcal{X}$, define $S(x, y) = \{i : x_i = -y_i \neq 0\}$. Then for every $i \in S(x, y)$, there is a $z \in \mathcal{X}$ with $z_i = 0$ and $z_j = (x \cdot y)_j = (y \cdot x)_j$ for all $j \notin S(x, y)$.

By inspection, the sign sequences of faces in $\mathcal{F}(\mathcal{A})$ satisfy the above properties for any central arrangement $\mathcal{A}$. An oriented matroid which arises from the faces of central hyperplane arrangements is said to be realizable. There are oriented matroids which are not realizable, but the Folkman-Lawrence topological representation theorem says that all oriented matroids can be represented in terms of arrangements of pseudospheres (which

are analogues of hyperplane arrangements in which the "pseudo-hyperplanes" are not necessarily flat) [12, 55]. An element $x \in \mathcal{X}$ such that $x_i \neq 0$ for all i is called a chamber of $\mathcal{X}$. Also, for every face $z \in \mathcal{X}$, we can define the support of z in terms of the support set $A_z = \{i \in [m] : z_i = 0\}$ and the set of supports forms an intersection lattice just as in the case of hyperplane arrangements. If $\mathcal{X}$ is an oriented matroid with chamber set $\mathcal{C}(\mathcal{X}) \neq \emptyset$ and w is any probability measure on $\mathcal{X}$, then we may define a random walk on $\mathcal{C}(\mathcal{X})$ by $P(x, y) = \sum_{z \cdot x = y} w(z)$. Brown and Diaconis show that Theorems 2.2.1 and 2.2.1 carry over to oriented matroids [12]. One can also define correlates of hyperplane subarrangements for oriented matroids by restricting attention to the coordinates in support sets, and the argument that one can recover eigenfunctions via projections given here in chapter 3 can be easily adapted to the case of random walks on oriented matroids thereby. Moreover, analogously to embedding chamber walks on arbitrary arrangements into walks on Boolean arrangements as described in subsection 2.3.1, all random walks on oriented matroids can be realized as random walks on Boolean arrangements by defining the measure $\widetilde{w}$ on $\mathcal{F}(\mathcal{B}_m)$ by $\widetilde{w}(F) = w((x_1, ..., x_n))$ if $\sigma_i(F) = x_i$ for $i = 1, ..., n$ and $\widetilde{w}(F) = 0$ if there is no $(x_1, ..., x_n) \in \mathcal{X}$ such that $\sigma_i(F) = x_i$ for all $i \in [n]$. Similarly, oriented matroids can be seen as special cases of the RCT framework in which there are only two opaque colors. Thus many features of random walks on oriented matroids can be deduced from results concerning random walks on hyperplane arrangements and rows of colored tiles.

A further generalization of random walks on hyperplane arrangements, oriented matroids, and rows of colored tiles is given by the notion of a (finite) left-regular band. A band is an idempotent semigroup S and it is said to be left-regular if it also satisfies

$xyx = xy$ for all $x, y \in S$ [13]. Equivalently, a finite semigroup S is a LRB if there is a lattice $\mathcal{L}$ and a surjective map $\text{supp} : S \to \mathcal{L}$ which satisfies $\text{supp}(xy) = \text{supp}(x) \vee \text{supp}(y)$ and $xy = x$ if and only if $\text{supp}(x) \leq \text{supp}(y)$ [62]. (This characterization applies to finite bands in general if one only requires $\mathcal{L}$ to be an upper semilattice and replaces the second condition with $\text{supp}(x) \leq \text{supp}(y)$ if and only if $x = xyx$ [13].) The first of these relations says that supp is a semigroup homomorphism when $\mathcal{L}$ is regarded as a (commutative) semigroup under the join operation and the second says that supp is an order-preserving poset surjection. It is clear from the first definition that oriented matroids and face semigroups of hyperplane arrangements are both examples of left-regular bands. When applying the second definition to hyperplane walks, the lattice $\mathcal{L}$ should be viewed as the lattice of support sets of flats ordered by inclusion rather than the intersection lattice as we defined it because we chose to order the flats by reverse inclusion. The elements $x \in S$ with $\text{supp}(x) = \hat{1}$ are called chambers.

If S is a finite semigroup, $I \subseteq S$ is a left ideal, and w is a probability measure on S, then one can define a Markov chain on I by $P(s,t) = \sum_{xs=t} w(x)$. When S is a band, we will say that the measure w is separating if for each $H \in \mathcal{L}$ with $H \lessdot \hat{1}$, there is an $x \in S$ with $\text{supp}(x) \not\leq H$ and $w(x) > 0$. If S is a left-regular band, then upon identifying S with $\mathcal{F}(\mathcal{A})$, $\mathcal{C} = \{x \in S : \text{supp}(x) = \hat{1}\}$ with $\mathcal{C}(\mathcal{A})$, and $\mathcal{L}$ with $\mathcal{L}(\mathcal{A})$, the proof of Theorem 2.2.2 given here carries over directly to random walks on LRBs. (Alternatively, if we $\mathcal{W} = \{F_{[1]}, ..., F_{[r]}\} \subseteq S$ be an enumeration of the support of w and write $\mathcal{C}' = \{F_{[\sigma(r)]} \cdots F_{[\sigma(1)]} : \sigma \in S_r\}$ as in the proof of Theorem 2.2.2, then the argument given there shows that a necessary and sufficient condition for the existence of a unique stationary distribution is that $\mathcal{C}'$ is a right ideal in the semigroup $\mathcal{C}$.) Theorem 2.2.1 also

generalizes, but the approach is a little more delicate and relies on a careful analysis of the semigroup algebra.

In [14], Ken Brown established that the transition matrices for random walks on the chambers of left-regular bands are diagonalizable by appealing to a criterion for the semisimplicity of an algebra generated by a single element, $a = T_w$, in terms of the poles of the generating function for the powers of a. Using semigroup representation theory, Brown was also able to determine the eigenvalues, with multiplicity, of the transition matrix of the random walk on the chambers of any finite band [13]. These results were further generalized to a larger class of finite semigroups known as semilattices of combinatorial archimedian semigroups by Benjamin Steinberg [67]. Building on Brown's results, Franco Saliola gave a nice description of the eigenspaces of LRB walks in terms of the primitive orthogonal idempotents of the semigroup algebra [62].

All of these results involve identifying right multiplication by P with the element $T_w = \sum_{x \in S} w(x)x \in kS$ acting on the vector space $k\mathcal{C}$ by left multiplication as described in section 2.2. Though this approach is elegant, enlightening, and gives the eigenvalues of P directly, it is not particularly easy to recover the eigenvectors of P from the eigenspace decomposition in terms of orthogonal idempotents in the semigroup algebra. We provide a partial remedy in the case of hyperplane chamber walks in the following chapter where it is shown that many of the right eigenvectors arise from projecting the chains onto subarrangements. By reducing the state space, we can compute the eigenvectors of the lumped chains explicitly and then lift them to the original chain to obtain concrete expressions for the eigenvectors corresponding to the largest eigenvalues.

Chapter 3

Eigenfunctions

3.1 Eigenfunctions for Hyperplane Walks

The present section addresses one of the primary contributions of this dissertation, finding explicit right eigenfunctions for random walks on the chambers of central hyperplane arrangements in $\mathbb{R}^n$. The key insight here is that these Markov chains are lumpable (in the sense of Theorem 1.3.1) with respect to equivalence relations on the set of chambers induced by projections onto subarrangements, thus one can compute the eigenfunctions of the lumped chains and lift them to obtain eigenfunctions for the original chain. We begin with the following lemma, an equivalent formulation of which can be found in [5].

Lemma 3.1.1. *For any hyperplane arrangement $\mathcal{A} = \{H_i\}_{i=1}^m$, any set $B \subseteq [m]$, and any probability measure w on $\mathcal{F}(\mathcal{A})$, the Markov chain on $\mathcal{C}(\mathcal{A})$ defined by $P(C, D) = \sum_{\substack{F \in \mathcal{F}: \\ FC = D}} w(F)$ is lumpable in the sense of Theorem 1.3.1 with respect to the equivalence relation defined by $C \sim_B C'$ if and only if $\sigma_i(C) = \sigma_i(C')$ for all $i \in B$.*

Proof. For any $C \in \mathcal{C}(\mathcal{A})$, $[D] \in \mathcal{C}(\mathcal{A})/\sim_B$,

$$P(C,[D]) = \sum_{D' \sim_B D} P(C,D') = \sum_{D' \sim_B D} \sum_{\substack{F \in \mathcal{F}(\mathcal{A}): \\ FC = D}} w(F) = \sum_{\substack{F \in \mathcal{F}(\mathcal{A}): \\ FC \sim_B D}} w(F).$$

Now in order that $FC \sim_B D$, it is necessary and sufficient that

$$\sigma_i(D) = \sigma_i(FC) = \begin{cases} \sigma_i(F), & \sigma_i(F) \neq 0 \\ \sigma_i(C), & \sigma_i(F) = 0 \end{cases}$$

for all $i \in B$. Suppose that $C' \sim_B C$ so that $\sigma_i(C') = \sigma_i(C)$ for all $i \in B$. Then for any $F \in \mathcal{F}$, $i \in B$, we have that

$$\sigma_i(FC) = \begin{cases} \sigma_i(F), & \sigma_i(F) \neq 0 \\ \sigma_i(C), & \sigma_i(F) = 0 \end{cases} = \begin{cases} \sigma_i(F), & \sigma_i(F) \neq 0 \\ \sigma_i(C'), & \sigma_i(F) = 0 \end{cases} = \sigma_i(FC').$$

Accordingly, for all $C' \sim_B C$, $FC \sim_B D$ if and only if $FC' \sim_B D$, and thus

$$P(C,[D]) = \sum_{\substack{F \in \mathcal{F}(\mathcal{A}): \\ FC \sim_B D}} w(F) = \sum_{\substack{F \in \mathcal{F}(\mathcal{A}): \\ FC' \sim_B D}} w(F) = P(C',[D]).$$

□

As detailed in subsection 2.4.1, the lumped chain may be interpreted as a random walk on the chambers of the subarrangement $\mathcal{B} = \{H_i\}_{i \in B}$. In fact, using the notation of 2.4.1, it follows from the definition of the lumped chain $P^{\#}([C],[D]) = P(C,[D])$ that the induced walk is driven by the probability measure w_B on $\mathcal{C}(\mathcal{B})$ defined by $w_B(\overline{F}) = \sum_{\substack{G \in \mathcal{F}(\mathcal{A}): \\ \overline{G} = \overline{F}}} w(G)$. In terms of the RCT description, the lumped chain corresponds to painting the tiles indexed by B^C black in each row of tiles corresponding to the faces in $\mathcal{F}(\mathcal{A})$ and proceeding according to the original dynamics - that is, choosing a row of tiles corresponding to F with probability $w(F)$ and placing it on top of the stack. Of

course, Lemma 3.1.1 carries over directly to random walks on the faces of a hyperplane arrangements, oriented matroids, and $RCT(C, m)$.

We begin by considering the lumped chain corresponding to subarrangements consisting of a single hyperplane $\mathcal{B}_i = \mathcal{B}_{H_i}$. Projecting onto a subarrangement consisting of a single hyperplane is equivalent to restricting attention to a single coordinate of the chambers' sign sequences (or a single tile position in the RCT description). Now for any $U, W \in \mathcal{L}$, if $U \leq W$, then $W \subseteq U$, so $\lambda_W = \sum_{\substack{F \in \mathcal{F}: \\ F \subseteq W}} w(F) \leq \sum_{\substack{F \in \mathcal{F}: \\ F \subseteq U}} w(F) = \lambda_U$. As such, the largest nontrivial eigenvalue must correspond to some H_i as the atoms of $\mathcal{L}$ are precisely the hyperplanes. For $i \in [m]$, we define the equivalence relation $C \sim_i C'$ if and only if $\sigma_i(C) = \sigma_i(C')$. By Lemma 3.1.1 and Theorem 1.3.1, this gives rise to a random walk on $\mathcal{C}(\mathcal{A})/\sim_i \cong \mathcal{C}(\mathcal{B}_i)$ with transition probabilities

$$P^{\#}([C]_{H_i}, [D]_{H_i}) = \sum_{D' \sim_i D} P(C, D') = \sum_{D' \sim_i D} \sum_{\substack{F \in \mathcal{F}: \\ FC = D'}} w(F) = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(FC) = \sigma_i(D)}} w(F).$$

Because $\mathcal{C}(\mathcal{B}_i)$ consists of the two states $+$ and $-$, if we define

$$p_i = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = +}} w(F), \quad q_i = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = -}} w(F), \quad r_i = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = 0}} w(F) = \lambda_i,$$

then the we can write

$$P^{\#} = \begin{bmatrix} q_i + r_i & p_i \\ q_i & p_i + r_i \end{bmatrix}.$$

We assume throughout that w is separating so that p_i and q_i are not both zero. One easily checks that the eigenvalues of $P^{\#}$ are $\lambda = 1$ and $\lambda = r_i$ with corresponding right eigenvectors $\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}-p_i\\q_i\end{smallmatrix}\right]$. It follows from Theorem 1.3.1 that

$$\phi_i(C) = \varphi_i^{\flat}(C) = \begin{cases} -p_i, & \sigma_i(C) = - \\ q_i, & \sigma_i(C) = + \end{cases}$$

is a right eigenfunction for the original transition operator P with eigenvalue λ_i. (Lemma 3.1.1 still applies if w is concentrated on some H_i, but then projection gives 1 as an eigenvalue of multiplicity 2 and corresponding eigenvectors $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$.)

For the sake of consistency, note that if we take $B = \emptyset$, so that there is a single equivalence class, then the lumped transition matrix is given by $P^{\#} = 1$, which has eigenvalue 1 and eigenfunction $\varphi_0 \equiv 1$, so the trivial eigenfunction $\phi_0 \equiv 1$ also corresponds to projection onto a subarrangement. Also, since w is separating, for every $i \in [m]$, there is some $F \in \mathcal{F}$ such that $\sigma_i(F) \neq 0$ and $w(F) > 0$, hence $r_i = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = 0}} w(F) = \lambda_i < 1$. As such, $p_i + q_i = 1 - r_i > 0$, so at least one of p_i, q_i is positive (and, by construction, both p_i and q_i are nonnegative), hence $\phi_i(C) \neq \phi_i(D)$ if $\sigma_i(C) \neq \sigma_i(D)$. These facts enable us to conclude that $\{\phi_0, \phi_1, ..., \phi_n\}$ is a linearly independent set. To see that this is the case, suppose that $\alpha_{i_1}\phi_{i_1} + ... + \alpha_{i_k}\phi_{i_k} \equiv 0$ is a minimal dependence relation with $i_1 < ... < i_k$. Let $C, C' \in \mathcal{C}$ be such that $\sigma_{i_k}(C) = +$, $\sigma_{i_k}(C') = -$, and $\sigma_j(C) = \sigma_j(C')$ for all $j \in [n] \setminus \{i_k\}$. (By Zaslavsky's theorem, the arrangement obtained by deleting H_{i_k} has at least $|\mu(V, H_{i_k})| = 1$ less chamber than the original arrangement, so there is at least one chamber in the deleted arrangement which is divided into two by the addition

of H_{i_k}. These two chambers may be taken as C, C'.) Then, adopting the above notation, we have

$$\begin{aligned}\alpha_{i_1}\phi_{i_1}(C) + ... + \alpha_{i_{k-1}}\phi_{i_{k-1}}(C) + \alpha_{i_k}q_{i_k} &= \alpha_{i_1}\phi_{i_1}(C) + ... + \alpha_{i_k}\phi_{i_k}(C) = 0 \\ &= \alpha_{i_1}\phi_{i_1}(C') + ... + \alpha_{i_k}\phi_{i_k}(C') \\ &= \alpha_{i_1}\phi_{i_1}(C) + ... + \alpha_{i_{k-1}}\phi_{i_{k-1}}(C) - \alpha_{i_k}p_{i_k},\end{aligned}$$

hence $\alpha_{i_k}q_{i_k} = -\alpha_{i_k}p_{i_k}$ and thus $\alpha_{i_k}(q_{i_k} + p_{i_k}) = 0$. Because $q_{i_k} + p_{i_k} = 1 - r_{i_k} > 0$, this means that $\alpha_{i_k} = 0$, contradicting the minimality of the dependence relation.

We record these facts as

Theorem 3.1.1. *For each $i = 1, ..., m$, the Markov chain on $\mathcal{C}(\mathcal{A})$ defined by $P(C, D) = \sum_{\substack{F\in\mathcal{F}:\\ FC=D}} w(F)$ with w a separating probability measure on $\mathcal{F}$ has right eigenfunction*

$$\phi_i(C) = \begin{cases} -\sum_{\substack{F\in\mathcal{F}:\\ \sigma_i(F)=+}} w(F), & \sigma_i(C) = - \\ \sum_{\substack{F\in\mathcal{F}:\\ \sigma_i(F)=-}} w(F), & \sigma_i(C) = + \end{cases}$$

corresponding to the eigenvalue $\lambda_i = \sum_{\substack{F\in\mathcal{F}:\\ \sigma_i(F)=0}} w(F)$. P also has right eigenfunction $\phi_0 \equiv 1$ with eigenvalue $\lambda_0 = 1$, and $\phi_0, \phi_1, ..., \phi_m$ are linearly independent.

Of course, Lemma 3.1.1 applies to projections onto any subarrangement. For example, given any hyperplanes $H_i, H_j \in \mathcal{A}$, one can consider the random walk induced on the subarrangement $\{H_i, H_j\}$. Since this is a hyperplane arrangement in its own right and $\{H_i\}$, $\{H_j\}$, and the empty arrangement are all subarrangements, we already know three of the eigenfunctions - namely, the lifts of the trivial eigenfunction φ_0, and the eigenfunctions φ_i and φ_j corresponding to the eigenvalues λ_i and λ_j, and these in turn

lift to eigenfunctions for the random walk on the chambers of $\mathcal{A}$. The reader is spared the computations, but if we set

$$\rho_{x,y} = \rho_{x,y}^{(i,j)} = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F)=x \\ \sigma_j(F)=y}} w(F)$$

for $x, y \in \{-, 0, +\}$, then analysis of the resulting 4×4 matrix shows that the remaining eigenfunction lifts to

$$\begin{aligned}
\phi_{H_i \cap H_j}(C) &= \rho_{-1,0}\rho_{0,1}\rho_{1,-1} + \rho_{-1,1}\rho_{0,-1}\rho_{1,0} - \rho_{-1,0}\rho_{0,-1}\rho_{1,1} \\
&\qquad + \rho_{0,1}\rho_{1,0}(1 - \rho_{-1,-1} - \rho_{0,0}), \quad \sigma_i(C) = -, \sigma_j(C) = -; \\
\phi_{H_i \cap H_j}(C) &= \rho_{-1,0}\rho_{0,1}\rho_{1,-1} - \rho_{-1,-1}\rho_{0,1}\rho_{1,0} - \rho_{-1,0}\rho_{0,-1}\rho_{1,1} \\
&\qquad - \rho_{0,-1}\rho_{1,0}(1 - \rho_{-1,1} - \rho_{0,0}), \quad \sigma_i(C) = -, \sigma_j(C) = +; \\
\phi_{H_i \cap H_j}(C) &= \rho_{1,0}\rho_{0,-1}\rho_{-1,1} - \rho_{-1,-1}\rho_{0,1}\rho_{1,0} - \rho_{-1,0}\rho_{0,-1}\rho_{1,1} \\
&\qquad - \rho_{0,1}\rho_{-1,0}(1 - \rho_{1,-1} - \rho_{0,0}), \quad \sigma_i(C) = +, \sigma_j(C) = -; \\
\phi_{H_i \cap H_j}(C) &= \rho_{-1,0}\rho_{0,1}\rho_{1,-1} + \rho_{-1,1}\rho_{0,-1}\rho_{1,0} - \rho_{1,0}\rho_{0,1}\rho_{-1,-1} \\
&\qquad + \rho_{0,-1}\rho_{-1,0}(1 - \rho_{1,1} - \rho_{0,0}), \quad \sigma_i(C) = +, \sigma_j(C) = +.
\end{aligned}$$

As we are tacitly assuming that $H_i \neq H_j$, and both H_i and H_j are codimension 1 subspaces of V, we must have that $H_i + H_j = V$. Thus, since

$$\begin{aligned}
\dim(V) + \dim(H_i \cap H_j) &= \dim(H_i + H_j) + \dim(H_i \cap H_j) \\
&= \dim(H_i) + \dim(H_j) = 2\dim(V) - 2,
\end{aligned}$$

we see that $\dim(H_i \cap H_j) = \dim(V) - 2$. Quotienting out by $H_i \cap H_j$ shows that the subarrangement $\{H_i, H_j\}$ is combinatorially equivalent to an arrangement of two nonparallel lines in the plane. Consequently, for each $(x, y) \in \{+, -\}^2$, there is some $C \in \mathcal{C}$ with $\sigma_i(C) = x$ and $\sigma_j(C) = y$. In general, the state space of chamber walks on subarrangements of $k \geq 3$ hyperplanes may have cardinality less than 2^k, so one must take care when applying this procedure to arrangements other than the Boolean arrangement. Of course, computing eigenfunctions by hand once the size of the state space exceeds four is generally impractical, though using projections is still more numerically stable and often faster than dealing with the original state space.

It is perhaps worth mentioning that this construction could be guessed at without directly appealing to arguments involving projections. To see how this would work, let φ_W denote an eigenfunction corresponding to λ_W. Then for every $C \in \mathcal{C}$, we would have

$$\begin{aligned}
\lambda_W \varphi_W(C) = (P\varphi_W)(C) &= \sum_{D \in \mathcal{C}} P(C, D)\varphi_W(D) = \sum_{D \in \mathcal{C}} \sum_{\substack{F \in \mathcal{F}: \\ FC = D}} w(F)\varphi_W(D) \\
&= \sum_{F \in \mathcal{F}} w(F)\varphi_W(FC) = \sum_{r \in \{-,0,+\}^{k_W}} \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = r_i \, \forall i \in A_W}} w(F)\varphi_W(FC) \\
&= \sum_{r \in \{-,0,+\}^{k_W} \setminus \{0\}} \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = r_i \, \forall i \in A_W}} w(F)\varphi_W(FC) + \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = 0 \, \forall i \in A_W}} w(F)\varphi_W(FC).
\end{aligned}$$

If one were to suppose for the sake of convenience that φ_W was constant over all chambers with proscribed sign sequence coordinates in the positions indexed by the support set of W, then the preceding becomes

$$\lambda_W \varphi_W(C) = \sum_{r \in \{-,0,+\}^{k_W} \setminus \{0\}} \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = r_i \, \forall i \in A_W}} w(F)\varphi_W(FC) + \varphi_W(C) \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = 0 \, \forall i \in A_W}} w(F)$$

$$= \sum_{r \in \{-,0,+\}^{k_W} \setminus \{0\}} \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = r_i \ \forall i \in A_W}} w(F)\varphi_W(FC) + \lambda_W \varphi_W(C),$$

so φ_W must satisfy

$$\sum_{r \in \{-,0,+\}^{k_W} \setminus \{0\}} \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = r_i \ \forall i \in A_W}} w(F)\varphi_W(FC) = 0$$

for all chambers C. Solving this system of equations is equivalent to computing the eigenfunctions of the projected chain.

To illustrate this approach, consider the case $W = H_i$. We compute

$$\begin{aligned}
\lambda_i \varphi_i(C) &= (P\varphi_i)(C) \\
&= \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = +}} w(F)\varphi_i(FC) + \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = -}} w(F)\varphi_i(FC) + \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = 0}} w(F)\varphi_i(FC) \\
&= \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = +}} w(F)\varphi_i(FC) + \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = -}} w(F)\varphi_i(FC) + \lambda_i \varphi_i(C)
\end{aligned}$$

so that

$$\sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = +}} w(F)\varphi_i(FC) + \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = -}} w(F)\varphi_i(FC) = 0.$$

If we let $p_i = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = +}} w(F)$, $q_i = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = -}} w(F)$, and assume that

$$\varphi_i(C) = \begin{cases} a, & \sigma_i(C) = + \\ b, & \sigma_i(C) = - \end{cases},$$

then we see that

$$0 = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = +}} w(F)\varphi_i(FC) + \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = -}} w(F)\varphi_i(FC)$$

$$= a \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F)=+}} w(F) + b \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F)=-}} w(F) = ap_i + bq_i,$$

so, up to scaling, the solution is $a = q_i$, $b = -p_i$ as previously deduced by examining the projected chain.

Before concluding this section, we note that Lemma 3.1.1 and the accompanying proof shows that the state of the Markov chain is completely determined by its evaluation at eigenfunctions corresponding to the hyperplanes whenever the face measure is separating. This is because ϕ_i takes distinct values depending on whether its argument is on the positive or negative side of H_i and a chamber is uniquely determined by specifying which side of each hyperplane it lies in. Thus one expects that analysis of the hyperplane eigenfunctions should shed a good deal of light on the behavior of the associated chamber walk.

We also mention that it is tempting to conjecture that in many cases all of the eigenfunctions arise in terms of projection onto subarrangements corresponding to support sets of flats. Since such eigenfunctions are constant on equivalence classes, this would provide valuable information concerning the eigenspace decomposition. The general idea is that given any $W \in \mathcal{L}$, the lumped chain with respect to $\sim_{A_W}$ (where $A_W = \{i \in [m] : \sigma_i(W) = 0\}$) corresponds to a walk on the chambers of the subarrangement $\mathcal{B}_W = \{H_i\}_{i \in A_W} = \{H_i \in \mathcal{A} : W \subseteq H_i\}$. As this is a hyperplane chamber walk in its own right, Theorem 2.2.1 implies that the transition matrix for the projected chain is diagonalizable,

and Lemma 2.1.1 and Zaslavsky's Theorem show that the state space of the projected chain has size

$$|\mathcal{C}_W| = \sum_{U \in \mathcal{L}_W} |\mu_W(V,U)| = \sum_{U \in [V,W]} |\mu(V,U)| = \sum_{\substack{U \in \mathcal{L}: \\ U < W}} |\mu(V,U)| + |\mu(V,W)| \,.$$

Naively, projection onto further subarrangements corresponding to $U < W$ accounts for $\sum_{\substack{U \in \mathcal{L}: \\ U < W}} |\mu(V,U)|$ of these eigenvalues/eigenfunctions, so the remaining $|\mu(V,W)|$ can be said to arise from projection onto $\mathcal{B}_W$. The author has tried several approaches using well-founded induction on the intersection lattice and ordinary induction on the dimension or the number of hyperplanes (and appealing to Theorem 3.1.1 for the base case) to make this rigorous, but has been unsuccessful thus far. The main problems involve the fact that one may have $\lambda_U \neq \lambda_{U'}$ when $U \neq U'$ and the difficulty in establishing linear independence of the eigenfunctions corresponding to incomparable flats $U, U' < W$. Still it seems likely that something like this is true in many cases, and ideally one would like to be able to use the lumping framework to recover all or part of Theorem 2.2.1, perhaps with additional assumptions. Hopefully, further investigation investigation will yield concrete results in this direction.

3.2 Examples

In order to illustrate the utility of the results in the preceding section, we will now explicitly compute some of the top eigenfunctions for the random walks encountered in section 2.3. We begin with lazy random walk on the hypercube, which corresponds to a random walk on the Boolean arrangement $\mathcal{B}_n$ with face weights $w(F) = \frac{1}{2n}$ if the $\sigma(F)$ has exactly one nonzero coordinate and $w(F) = 0$ otherwise. We saw in subsection 2.3.1

that each $S \subseteq [n]$ contributes the eigenvalue $\lambda_S = 1 - \frac{|S|}{n}$ with multiplicity one. Thus the largest nontrivial eigenvalue is $\lambda = 1 - \frac{1}{n}$ with multiplicity n. By Theorem 3.1.1, a basis for the corresponding eigenspace is given by $\{\varphi_i\}_{i=1}^n$ where

$$\phi_i(C) = \begin{cases} -\sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = +}} w(F), & \sigma_i(C) = - \\ \\ \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = -}} w(F), & \sigma_i(C) = + \end{cases}.$$

For each $i \in [n]$, we have $\sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = +}} w(F) = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = -}} w(F) = \frac{1}{2n}$ since there is exactly one face in the support of w with a "+" in the ith sign sequence coordinate and exactly one face with a "−" in the ith coordinate.. Scaling by $2n$, we see that the functions

$$\varphi_i(C) = \begin{cases} 1, & \sigma_i(C) = + \\ \\ -1, & \sigma_i(C) = - \end{cases}, \ i = 1, ..., n$$

form a basis for the $1 - \frac{1}{n}$ eigenspace.

It should be mentioned that one can use the spectral decomposition of the transition matrix for the chain viewed as a random walk on $\mathbb{Z}_2^n$ to determine all of the eigenvalues/eigenfunctions in this particular case. The general technique was discovered by Peter Matthews and is discussed at greater length in [21] as well at the end of subsection 3.3.2 in the present work. The basic idea is that for any random walk on a group which is driven by a measure that is constant on conjugacy classes, the eigenvalues are given by the diagonal entries of the Fourier transforms of the irreducible representations and the eigenfunctions are given by the matrix entries of the corresponding representations. Every random walk on $\mathbb{Z}_2^n$ can be so decomposed since it is an abelian, group and the computations are quite simple for the same reason. As an n-fold Cartesian product of

the cyclic group of order 2, the irreducible representations of $\mathbb{Z}_2^n$ are given by all possible n-fold products of ± 1 and thus may be parameterized by $\rho_x(y) = (-1)^{x \cdot y}$ as x ranges over $\mathbb{Z}_2^n$. The lazy nearest-neighbor walk is driven by the measure $Q(0) = \frac{1}{2}$, $Q\left(y^{(1)}\right) = Q\left(y^{(n)}\right) = \frac{1}{2n}$ where $y^{(i)} \in \mathbb{Z}_2^n$ has ith coordinate 1 and all other coordinates 0, so, letting $\omega(x)$ denote the number of coordinates in x which are equal to 1, we compute the Fourier transform of Q at the representation ρ_x as

$$\begin{aligned}\widehat{Q}(\rho_x) &= \sum_{y \in \mathbb{Z}_2^n} (-1)^{x \cdot y} Q(y) = \frac{1}{2} + \frac{1}{2n} \sum_{i=1}^{n} (-1)^{x \cdot y^{(i)}} \\ &= \frac{1}{2} + \frac{1}{2n} \left[(n - \omega(x)) - \omega(x) \right] = 1 - \frac{\omega(x)}{n},\end{aligned}$$

in agreement with the eigenvalues deduced from Theorem 2.2.1. The corresponding eigenfunctions are the representations $\rho_x(y) = (-1)^{x \cdot y}$. In particular, the representations corresponding to the eigenvalue $1 - \frac{1}{n}$ are parametrized by those elements of $\mathbb{Z}_2^n$ with exactly one coordinate equal to 1, so a basis of for the $(1 - \frac{1}{n})$-eigenspace is given by

$$\rho_{y^{(i)}}(x) = (-1)^{y^{(i)} \cdot x} = \begin{cases} -1, & x_i = 1 \\ 1, & x_i = 0 \end{cases}, \; i = 1, ..., n.$$

Observe that $\rho_{y^{(i)}} = -\varphi_i$ under the obvious correspondence. Notice also that if $y, y' \in \mathbb{Z}_2^n$ agree in all coordinates i with $x_i = 1$, then $\rho_x(y) = (-1)^{x \cdot y} = (-1)^{x \cdot y'} = \rho_x(y')$, and if there is some set $A \subseteq [n]$ such that $y_i = y'_i$ for all $i \in A$ implies $\rho_x(y) = \rho_x(y')$, then it must be the case that $A \supseteq \{i \in [n] : x_i = 1\}$ - otherwise, there would exist some $j \in [n] \setminus A$ with $x_j = 1$, so, if y and y' disagreed only in coordinate j, then we would have $y_i = y'_i$ for all $i \in A$, but $\rho_x(y) = -\rho_x(y')$. Finally, it is worth mentioning that these facts

could also be deduced from Theorem 1.3.3 by taking Q to be uniform on the singletons and taking $P_i = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ for $i = 1, ..., n$.

Each of these three techniques can also be used to analyze biased random walks on the hypercube. For example, letting $F_i^{\pm}$ denote the face with $\sigma_i(F_i^{\pm}) = \pm$ and $\sigma_j(F_i^{\pm}) = 0$ for $j \neq i$, letting $p_i^{\pm} \geq 0$ be such that $p_i^+ + p_i^- > 0$ and $\sum_{i=1}^n \left(p_i^+ + p_i^-\right) = 1$, and defining w by $w(F_i^{\pm}) = p_i^{\pm}$, we see that for each $i \in [n]$, the associated random walk on the Boolean arrangement has eigenfunction

$$\phi_i(C) = \begin{cases} -p_i^+, & \sigma_i(C) = - \\ \\ p_i^-, & \sigma_i(C) = + \end{cases}$$

corresponding to the eigenvalue $\lambda_i = 1 - p_i^+ - p_i^-$.

As we have already obtained eigenfunctions for the militaristic models using the theory of product chains, and the analysis via projections is not substantially different from the the hypercube walks, we turn now to examples involving the braid arrangement. Before doing so, however, it is worth noting that the eigenfunctions arising from single hyperplane subarrangements can also be expressed in terms of indicators: Recall that for any random walk on a central arrangement of m hyperplanes in $\mathbb{R}^n$ driven by a separating face measure w, if we set $p_i = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = +}} w(F)$, $q_i = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = -}} w(F)$, $\lambda_i = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = 0}} w(F)$, then

$$\phi_i(C) = \begin{cases} -p_i, & \sigma_i(C) = - \\ \\ q_i, & \sigma_i(C) = + \end{cases}$$

is an eigenfunction of the transition operator with eigenvalue λ_i, $i = 1, ..., m$. It is often more illuminating to write

$$\phi_i(C) = (q_i + p_i)1\{\sigma_i(C) = +\} - p_i$$

and divide by $p_i + q_i = (1 - \lambda_i) > 0$ to obtain the eigenfunction

$$\varphi_i(C) = 1\{\sigma_i(C) = +\} - \frac{p_i}{p_i + q_i}.$$

In the case of the braid arrangement corresponding to S_n (the set of all $\binom{n}{2}$ hyperplanes of the form $\{x \in \mathbb{R}^n : x_i - x_j = 0\}$), we saw in subsection 2.3.2 that the chambers are in bijective correspondence with S_n via $\pi \leftrightarrow C_\pi = \{x \in \mathbb{R}^n : x_{\pi(1)} > x_{\pi(2)} > ... > x_{\pi(n)}\}$. Moreover, the sign sequence of C_π is related to the inversion set of π^{-1} under this correspondence:

$$\sigma_{(i,j)}(\pi) := \sigma_{(i,j)}(C_\pi) = \begin{cases} -, & (i,j) \in \text{Inv}(\pi^{-1}) \\ \\ +, & (i,j) \notin \text{Inv}(\pi^{-1}) \end{cases}.$$

Similarly, the faces of the braid arrangement are in bijective correspondence with the ordered set partitions of $[n]$ with the partition $(B_1, ..., B_l)$ corresponding to the face consisting of all points x such that $x_i = x_j$ if $i, j \in B_r$ for some $r = 1, ..., l$ and $x_i > x_j$ if $i \in B_r$, $j \in B_s$ for some $1 \leq r < s \leq l$. The sign sequences, of the faces can thus be represented as

$$\begin{aligned} \sigma_{(i,j)}\left((B_1, ..., B_l)\right) &:= \sigma_{(i,j)}(F_{(B_1,...,B_l)}) \\ &= \begin{cases} -, & i \in B_s,\ j \in B_r \text{ for some } 1 \leq r < s \leq l \\ \\ 0, & i, j \in B_r \text{ for some } 1 \leq r \leq l \\ \\ +, & i \in B_r,\ j \in B_s \text{ for some } 1 \leq r < s \leq l \end{cases}. \end{aligned}$$

The set partitions act on permutations by $(B_1, ..., B_l)\pi = \tau$ where $\tau(i) < \tau(j)$ if and only if $i \in B_r$, $j \in B_s$ for some $1 \leq r < s \leq l$ or $i, j \in B_r$ ($r \in [l]$) and $\pi(i) < \pi(j)$. Thus, if we identify the state space with S_n, then for each $1 \leq i < j \leq n$, we have eigenfunctions

$$\begin{aligned}\varphi_{(i,j)}(\pi) &= 1\left\{(i,j) \notin \mathrm{Inv}(\pi^{-1})\right\} - \frac{p_{(i,j)}}{p_{(i,j)} + q_{(i,j)}} \\ &= 1\left\{\pi^{-1}(i) < \pi^{-1}(j)\right\} - \frac{p_{(i,j)}}{p_{(i,j)} + q_{(i,j)}}\end{aligned}$$

corresponding to the eigenvalue $\lambda_{(i,j)}$ where

$$q_{(i,j)} = \sum_{\substack{B \vdash [n]: \\ j \text{ is in a block preceding } i}} w(B)$$

$$p_{(i,j)} = \sum_{\substack{B \vdash [n]: \\ i \text{ is in a block preceding } j}} w(B)$$

$$\lambda_{(i,j)} = \sum_{\substack{B \vdash [n]: \\ i \text{ and } j \text{ are in a common block}}} w(B).$$

The fact that the top eigenfunctions of BHR walks on braid arrangements can be represented as shifted indicators of inversions is not only interesting from a theoretical viewpoint, but has practical applications in the study of various permutation statistics using Stein's method techniques as we will see in subsection 3.3.2. Recalling the deck of cards interpretation of random walks on the braid arrangement (where $\pi(i)$ denotes the card which is in the ith position from the top of the deck), we see that $\left\{(i,j) \notin \mathrm{Inv}(\pi^{-1})\}\right\}$ is the event that card i is above card j. Thus, from a card-shuffling standpoint, the top eigenfunctions for these walks carry information about the pairwise relative ordering of the cards. In a similar vein, if we conceive of the chambers as representing rankings of alternatives (as alluded to at the end of subsection 2.3.2), then the top eigenfunctions are

related to outcomes in head-to-head competitions, and thus could be used as a measure of the degree to which the voting scheme we mentioned satisfies the Condorcet criterion.

Let us now consider examine some of the examples encountered in subsection 2.3.2. We begin with the Tsetlin library, a model for dynamic file management. Here we are given numbers $w_1, ..., w_n \geq 0$ with $\sum_{k=1}^{n} w_k = 1$ and we define a measure w on $\mathcal{F}$ by $w(F_k) = w_k$ for $F_k \sim (\{k\}, [n] \setminus \{k\})$. The interpretation is that we have files $1, ..., n$ with file k being accessed with probability w_k (where w_k is unknown to the user). We wish to sort the files so that those used most frequently are arranged at the top of the files, so we adopt the policy that every time a file is used it gets placed on top. The state of the random walk after r steps corresponds to the order of the files after r files have been accessed. We saw in subsection 2.3.2 that the eigenvalue corresponding to the hyperplane $H_{(i,j)}$ is given by $\lambda_{(i,j)} = 1 - w_i - w_j$. Because $\sigma_{(i,j)}(F) = +$ if and only the ordered set partition corresponding to F is such that i and j belong to separate blocks and the block containing i appears before the block containing j, we see that $\sum_{\substack{F \in \mathcal{F}: \\ \sigma_{(i,j)}(F) = +}} w(F) = w_i$. By the exact same reasoning, $\sum_{\substack{F \in \mathcal{F}: \\ \sigma_{(i,j)}(F) = -}} w(F) = w_j$. Thus if we identify the state space with S_n, then it follows from Theorem 3.1.1 that the corresponding eigenfunction is given by

$$\phi_{(i,j)}(\pi) = \begin{cases} -w_i, & (i,j) \in \mathrm{Inv}(\pi^{-1}) \\ w_j, & (i,j) \notin \mathrm{Inv}(\pi^{-1}) \end{cases}.$$

In the case of (uniform) random-to-top shuffles, we have $w_k = \frac{1}{n}$ for all $k = 1, ..., n$, so after rescaling, we see that the functions $\{\varphi_{(i,j)}\}_{i<j}$ defined as

$$\varphi_{(i,j)}(\pi) = \begin{cases} -1, & (i,j) \in \mathrm{Inv}(\pi^{-1}) \\ 1, & (i,j) \notin \mathrm{Inv}(\pi^{-1}) \end{cases}$$

are eigenfunctions for the random-to-top shuffle with eigenvalue $1 - \frac{2}{n}$. It turns out that the $\varphi_{(i,j)}$'s are also eigenfunctions for the subdominant eigenvalue of inverse a-shuffles (including ordinary riffle shuffles as the $a = 2$ case). Recall that a-shuffles can be realized as walks on the braid arrangement driven by a face measure which is uniform over all a-block partitions of $[n]$, including those containing empty blocks. Accordingly, we have

$$\sum_{\substack{F \in \mathcal{F}: \\ \sigma_{(i,j)}(F) = +}} w(F) = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_{(i,j)}(F) = +}} \frac{1}{a^n} = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_{(i,j)}(F) = -}} \frac{1}{a^n} = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_{(i,j)}(F) = -}} w(F)$$

(since interchanging i and j gives a bijection between partitions with i in a block preceding that containing j and those where j appears first), hence $\varphi_{(i,j)}$ is an eigenfunction corresponding to $\lambda_{(i,j)} = \frac{1}{a}$ for each $1 \leq i < j \leq n$. Note that for both the inverse a-shuffles and the uniform random-to-top shuffles, the face measures are invariant under the action of S_n, and thus correspond to random walks on the symmetric group in the usual sense. As a random walk on a group has an inverse whose transition matrix is just the transpose of the original transition matrix, the right eigenvectors of these inverse-shuffle hyperplane walks are the left eigenvectors of the corresponding shuffles. Thus a-shuffles and top-to-random shuffles have subdominant left eigenspaces spanned by the above shifted indicators of inversions. As discussed in [32], this suggests the possibility of using these eigenfunctions to study the associated quasi-stationary distributions of these shuffles.

It should be noted that Diaconis, Pang, and Ram were able to represent inverse a-shuffles in terms of the Hopf square (coproduct-then-product) map on the free associative algebra with respect to its word basis, and then use Hopf algebra techniques to find the left eigenfunctions of the transition matrix. They showed that a basis for the $\frac{1}{a}$ left-eigenspace is given by $\{f_{ij}\}_{1\leq i<j\leq n}$ where

$$f_{ij}(\pi) = \begin{cases} 1, & ij \text{ is a subword of } \pi \\ -1, & ji \text{ is a subword of } \pi \\ 0, & \text{else} \end{cases}$$

and then deduce that if $d(\pi) = |\{i \in [n-1] : \pi(i) > \pi(i+1)\}|$, the number of descents in π, then $f(\pi) = n-1-2d(\pi)$ is a left eigenfunction for the inverse a-shuffle with eigenvalue $\frac{1}{a}$. See [32] for more on this fascinating approach.

Our results for random-to-top shuffles and inverse a-shuffles show that $g(\pi) = f(\pi^{-1})$ is a right eigenfunction for the subdominant eigenvalues of these chains. More generally, for each $1 \leq d < n$, we can define the number of d-descents of a permutation $\pi \in S_n$ by

$$\mathrm{Des}_d(\pi) = \left|\{(i,j) \in [n]^2 : i < j \leq i+d,\, \pi(i) > \pi(j)\}\right|.$$

These permutation statistics first appeared in [18] where they were related to the Betti numbers of Hessenberg varieties, and generalize the notions of descents and inversions of a permutation (corresponding to $d = 1$ and $d = n$ respectively). We claim that the function $\mathrm{Des}_d(\pi^{-1})$ (when normalized to have mean zero under the uniform distribution on S_n) is an eigenfunction corresponding to the second largest eigenvalues of each of these

chains. To see that this is the case, we first observe that for $n \geq 2$, $1 \leq d < n$, the number of $(i,j) \in [n]^2$ with $i < j \leq i + d$ is

$$N_{n,d} = d(n-d) + (d-1) + (d-2) + ... + 2 + 1 = \frac{2nd - d^2 - d}{2}.$$

Note that under the uniform distribution on S_n, each pair each pair (i,j) with $i < j$ has probability $\frac{1}{2}$ of being an inversion, so the expected number of d-descents in a random permutation is $\frac{1}{2}N_{n,d}$. Now, writing

$$\mathrm{Asc}_d(\pi) = \left|\{(i,j) \in [n]^2 : i < j \leq i + d,\, \pi(i) < \pi(j)\}\right| = N_{n,d} - \mathrm{Des}_d(\pi)$$

for the number of d-ascents of π, we have

$$\sum_{1 \leq i < j \leq \min(n, i+d)} \varphi_{(i,j)}(\pi) = \mathrm{Asc}_d(\pi^{-1}) - \mathrm{Des}_d(\pi^{-1}) = N_{n,d} - 2\mathrm{Des}_d(\pi^{-1}),$$

hence

$$\mathrm{Des}_d(\pi^{-1}) - \frac{1}{2}N_{n,d} = -\frac{1}{2} \sum_{1 \leq i < j \leq \min(n, i+d)} \varphi_{(i,j)}(\pi)$$

is an eigenfunction for both the the random-to-top and inverse a-shuffle chains. Specializing to the case $d = 1$ and using the fact that a permutation has r rising sequences if and only if its inverse has has $r - 1$ descents, so that the number of rising sequences in π is $R(\pi) = \mathrm{Des}_1(\pi^{-1}) + 1$, we have the corollary that

$$R(\pi) - \frac{n+1}{2} = R(\pi) - 1 - \frac{2n-2}{4} = \mathrm{Des}_1(\pi^{-1}) - \frac{1}{2}N_{n,1}$$

is also an eigenfunction. Similarly, since the number of inversions in a permutation is the same as the number of inversions in its inverse, taking $d = n - 1$ shows that

$$\mathrm{Inv}(\pi) - \frac{n^2 - n}{4} = \mathrm{Inv}(\pi^{-1}) - \frac{n^2 - n}{4} = \mathrm{Des}_{n-1}(\pi^{-1}) - \frac{1}{2}N_{n,n-1}$$

is an eigenfunction of these inverse shuffles as well.

Before moving on, we note that there are many other walks on the braid arrangement which have eigenfunctions $\varphi_{(i,j)}$. For example, we could take the face measure to be uniform over all two-block partitions $(S, [n] \setminus S)$ such that $|S| = m$. The interpretation is that we choose m cards at random and then move them to the front of the deck, retaining their original order. We will call these *random-m-set-to-top* shuffles. For each hyperplane $H_{(i,j)} = \{x_i = x_j\}$, we have

$$\lambda_{(i,j)} = \sum_{\substack{F \in \mathcal{F}: \\ F \subseteq H_i}} w(F) = 1 - \sum_{\substack{F \in \mathcal{F}: \\ F \nsubseteq H_i}} w(F) = 1 - \frac{2\binom{n-2}{m-1}}{\binom{n}{m}} = 1 - 2\frac{m(n-m)}{n(n-1)}$$

since there are $\binom{n}{m}$ faces in the support of F (one corresponding to each $S \subset [n]$ with $|S| = m$) and the number of such faces in which i and j belong to different blocks is twice the number of faces where i's block proceeds j's (as seen by the map which interchanges i and j), of which there are $\binom{n-2}{m-1}$, the number of ways of choosing the other $m-1$ elements in the first block. Arguing as in the inverse a-shuffle case, we see that the functions $\{\varphi_{(i,j)}\}_{1 \leq i < j \leq n}$ are eigenfunctions for this eigenvalue as well, and so the statements about d-descents are also applicable in this case. Similarly, we could consider the inverse of the *top-m-to-random shuffles* studied in [24]. These are driven by probability measures which are constant over all $(m+1)$-block partitions where the first m blocks are singleton sets. For all $1 \leq i < j \leq n$, we have the eigenvalue

$$\lambda_{(i,j)} = \frac{\binom{n-2}{m}}{\binom{n}{m}} = \frac{(n-m)(n-m-1)}{n(n-1)}$$

as there are a total of $m!\binom{n}{m}$ equiprobable faces, $m!\binom{n-2}{m}$ of which have i and j in a common block. Again, the interchange i and j argument shows that corresponding

eigenfunctions are given by $\{\varphi_{(i,j)}\}_{i<j}$. Yet another example in which the $\varphi_{(i,j)}$'s arise as eigenfunctions for the subdominant eigenvalue is given by taking the face measure to be uniform over all two-block partitions in which one of the blocks has size 1. The inverse of this shuffle corresponds to picking a card at random and then flipping a fair coin to decide whether to put it on the top or bottom of the deck. Of course this reasoning applies to all walks on the braid arrangement driven by a probability measure which is constant over set partitions of a given shape.

In fact, it is not even necessary that the face measure is uniform over partitions with the same shape; it needs only to be invariant under pairwise transposition of elements. For example, consider the inverse of the biased a-shuffles. Here we let $p_1, ..., p_a > 0$ be such that $\sum_{j=1}^{a} p_j = 1$ and define a measure on the faces corresponding to a-block partitions by

$$w\left((B_1, ..., B_a)\right) = \prod_{j=1}^{a} p_j^{|B_j|}$$

As this measure satisfies $w(F) = w(\tau F)$ for all transpositions $\tau \in S_n$, reasoning as before shows that $\varphi_{(i,j)}$ is an eigenfunction for the eigenvalue $\lambda_{(i,j)}$ for each $i < j$. To compute this eigenvalue, we observe that for each ordered partition of $[n] \setminus \{i, j\}$ into a blocks we can add $\{i, j\}$ to any of the blocks to obtain an ordered partition of $[n]$ into a blocks where i and j lie in a common block. Thus it follows from the multinomial theorem that

$$\begin{aligned}
\lambda_{(i,j)} &= \sum_{(B_1,...,B_a)\vdash[n]\setminus\{i,j\}} \left(\sum_{l=1}^{a} p_l^2 \prod_{k=1}^{a} p_k^{|B_k|} \right) \\
&= \sum_{l=1}^{a} p_l^2 \left(\sum_{(B_1,...,B_a)\vdash[n]\setminus\{i,j\}} \prod_{k=1}^{a} p_k^{|B_k|} \right)
\end{aligned}$$

$$= \sum_{l=1}^{a} p_l^2 \left(\sum_{b_1+\ldots+b_a=n-2} \binom{n-2}{b_1,\ldots,b_a} \prod_{k=1}^{a} p_k^{|B_k|} \right) = \sum_{l=1}^{a} p_l^2.$$

Our next task is to show that in each of the preceding examples, $\{\varphi_{(i,j)}\}_{i<j}$ actually forms a basis for the eigenspace corresponding to the subdominant eigenvalue. We first note that since transpositions generate S_n, invariance of w under the action of transpositions on faces is equivalent to invariance of w under the action of S_n - the former is merely easier to check. We have seen that this condition guarantees that $\varphi_{(i,j)}$ is an eigenfunction corresponding to $\lambda_{(i,j)}$, and multiplying by $-\frac{1}{2}$ shows that

$$\phi_{(i,j)}(\pi) = 1\{(i,j) \notin \mathrm{Inv}(\pi^{-1})\} - \frac{1}{2} = 1\{\pi^{-1}(i) < \pi^{-1}(j)\} - \frac{1}{2}$$

is also such an eigenfunction. By the remarks at the beginning of section 2.3, all such hyperplane walks actually correspond to random walks on S_n in the standard sense. Moreover, invariance also implies that for all $(i,j),(k,l) \in \{(a,b) \in [n]^2 : a < b\}$, we have

$$\lambda_{(i,j)} = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_{(i,j)}(F)=0}} w(F) = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_{(i,j)}(F)=0}} w(\tau_{j,l}\tau_{i,k}F) = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_{(k,l)}(F)=0}} w(F) = \lambda_{(k,l)}$$

where $\tau_{a,b}$ is the permutation which transposes a and b. In addition, if w is not the point mass at the origin, then it must be the case that w is separating. This is because the origin corresponds to the one block partition $0 \sim ([n])$, so if $w(0) < 1$, then there is some face $F \sim (B_1,\ldots,B_l)$ with $l > 1$ and $w(F) > 0$, hence there is a pair $(i,j) \in [n]^2$ with $i \in B_1, j \in B_2$ and thus $\sigma_{(i,j)}(F) \neq 0$. By invariance, for any $(k,l) \in [n]^2$, $F' = \tau_{i,k}\tau_{j,l}F$ satisfies $\sigma_{(k,l)}(F') \neq 0$ and $w(F') = w(F) > 0$. Finally, under the preceding assumptions, we claim

Lemma 3.2.1. *If w is not concentrated on $\mathcal{C}$ and $n \neq 4$, then it cannot be the case that there exist $(i,j) \neq (k,l)$ such that $\sigma_{(k,l)}(F) = 0$ for all $F \in \mathcal{F}$ with $\sigma_{(i,j)}(F) = 0$ and $w(F) > 0$.*

Proof. This is vacuously true if $n = 2$. If $n = 3$, then the fact that w is not supported on $\mathcal{C} \cup \{0\}$ implies that there is a face $F \sim (B_1, B_2)$ with $w(F) > 0$ and either B_1 or B_2 has 2 elements. By invariance, we can take $\{i,j\}$ as the two-element block. We may thus assume that $n > 4$. Now the assumption that the support of w is not contained in $\mathcal{C} \cup \{0\}$ implies that there is some $F \sim (B_1, ..., B_l)$ with $w(F) > 0$, $l > 1$, and $\max_{r \in [l]} |B_r| > 1$. Let $s = \text{argmax}_{r \in [l]} |B_r|$. By invariance, we may assume that $i, j \in B_s$. Also, we may suppose without loss of generality that $k \neq i, j$ (as it cannot be the case that both k and l are in $\{i,j\}$ and the ensuing argument does not depend on the relative ordering of k and l). If $|B_s| = 2$, then there are at least two other blocks and either $l \in \{i,j\}$, so $\sigma_{(k,l)}(F) \neq 0$, or $l \notin \{i,j\}$ and we can use invariance to transpose k with an element which is not in the block containing l to obtain a partition with positive measure in which k and l lie in different blocks. If $|B_s| > 2$, then arguing as before, we can obtain a partition having positive measure with $i, j, l \in B_s$ (possibly because $l \in \{i,j\}$) and $k \notin B_s$. □

It will follow that the multiplicity of the eigenvalue corresponding to a single hyperplane is $\binom{n}{2}$ and thus the functions $\{\varphi_{(i,j)}\}_{1 \leq i < j \leq n}$ are a basis for the eigenspace corresponding to the subdominant eigenvalue (because we have already established that $\{\varphi_{(i,j)}\}_{1 \leq i < j \leq n}$ is a linearly independent set of such eigenfunctions). Clearly the multiplicity is at least $\binom{n}{2}$ since each of the hyperplanes contributes such an eigenvalue. To see that the multiplicity of the subdominant eigenvalue is no more than $\binom{n}{2}$, we observe that

if it were, then there would exist a flat $W > V$ which strictly contains some hyperplane $H_{(i,j)}$ and satisfies $\lambda_W = \lambda_{(i,j)}$. But it follows from Lemma 3.2.1 that for all $(k,l) \neq (i,j)$, there is a face F with $\sigma_{(i,j)}(F) = 0$, $\sigma_{(k,l)}(F) \neq 0$, and $w(F) > 0$. Accordingly, letting $(k,l) \in A_W \setminus \{(i,j)\} \neq \emptyset$, there is a $G \in \mathcal{F}$ with $\sigma_{(i,j)}(G) = 0$, $\sigma_{(k,l)}(G) \neq 0$, and $w(G) > 0$, so that

$$\begin{aligned}\lambda_{(i,j)} = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_{(i,j)}(F)=0}} w(F) &> \sum_{\substack{F \in \mathcal{F}: \\ \sigma_{(i,j)}(F)=0}} w(F) - w(G) \\ &\geq \sum_{\substack{F \in \mathcal{F}: \\ \sigma_{(r,s)}(F)=0 \, \forall (r,s) \in A_W}} w(F) = \lambda_W.\end{aligned}$$

Accordingly, the preceding examples are all special cases of

Theorem 3.2.1. *If w is a probability measure on the faces of the braid arrangement which is invariant under multiplication by transpositions, then the hyperplane chamber walk driven by w corresponds to a random walk on S_n. If $w(0) < 1$, then the functions*

$$\varphi_{(i,j)}(\pi) = \begin{cases} -1, & (i,j) \in \mathit{Inv}(\pi^{-1}) \\ 1, & (i,j) \notin \mathit{Inv}(\pi^{-1}) \end{cases}, \; 1 \leq i < j \leq n$$

are linearly independent eigenfunctions corresponding to the subdominant eigenvalue. If w is not supported on $\mathcal{C} \cup \{0\}$ and $n \neq 4$, then $\{\varphi_{(i,j)}\}_{1 \leq i < j \leq n}$ forms a basis for the eigenspace corresponding to the subdominant eigenvalue. An equivalent basis is given by $\{\phi_{(i,j)}\}_{1 \leq i < j \leq n}$ with

$$\phi_{(i,j)}(\pi) = 1\{\pi^{-1}(i) < \pi^{-1}(j)\} - \frac{1}{2}.$$

It is interesting to note that this eigenspace contains all of the information about the state of the chain in the sense that if one knows the value of $\phi_{(i,j)}(\pi_k)$ for all $1 \leq i < j \leq n$,

then one knows π_k. In the case of the braid arrangement, this is because a permutation is uniquely determined by its inversion set. More generally, the simultaneous values of the eigenfunctions corresponding to hyperplanes specify the sign sequence coordinates of the chamber at which they are evaluated.

We conclude this subsection with a computation of the eigenfunctions for the random walk on the dihedral arrangement discussed in subsection 2.3.3. This arrangement consists of $2m$ equally spaced lines through the origin in $\mathbb{R}^2$ where we have chosen to label the m hyperplanes by $H_i = \{(r,\theta) : r \in \mathbb{R}, \theta = \frac{\pi i}{m}\}$, the $2m$ chambers by $C_j = \left\{(r,\theta) : r > 0, \theta \in \left(\frac{\pi(j-1)}{m}, \frac{\pi j}{m}\right)\right\}$, and the $2m$ one-dimensional faces by $R_j = \left\{(r,\theta) : r > 0, \theta = \frac{\pi(j-1)}{m}\right\}$, so that the sign sequences of the faces are given by

$$\sigma_i(C_j) = \begin{cases} +, & j \leq i \text{ or } m+i < j \\ -, & \text{else} \end{cases},$$

$$\sigma_i(R_j) = \begin{cases} +, & j < i \text{ or } m+i < j \\ 0, & i = j \text{ or } i+m = j \\ -, & \text{else} \end{cases},$$

$$\sigma(\{0\}) = (0, ..., 0).$$

We restrict our attention to walks driven by measures supported on the one-dimensional faces which are not concentrated on any single hyperplane - that is, w is defined by $w(R_j) = p_j$ for $j = 1, ..., 2m$ where $p_j \geq 0$, $\sum_{j=1}^{2m} p_j = 1$, and $p_i + p_{i+m} < 1$ for $i = 1, ..., m$. We have already determined that the eigenvalues are given by $\lambda_V = 1$ with multiplicity 1, $\lambda_{\{0\}} = 0$ with multiplicity $m-1$, $\lambda_{H_i} = p_i + p_{i+m}$ with multiplicity 1 for $i = 1, ..., m$.

By Theorem 3.1.1 (using $-\phi_i$) and the above characterization of the sign sequences, we see that for each $j \in [2m]$, $i \in [m]$, the eigenfunction corresponding to λ_{H_i} is given by

$$\phi_i(C_j) = \begin{cases} \sum_{k<i} p_k + \sum_{k>m+i} p_k, & j < i \leq m+i \\ -\sum_{i<k<m+i} p_k, & \text{otherwise} \end{cases}.$$

3.3 Applications

In section 2.1 of [32], the authors mention several uses for Markov chain eigenfunctions. The first involves writing functions defined on the state space as linear combinations of eigenfunctions, $f = \sum_i \alpha_i \varphi_i$, so that the expectation of f under P_x^k can be easily computed as $E[f(X_k)|X_0 = x] = \sum_i \alpha_i \lambda_i^k \varphi_i(x)$. When P is diagonalizable, every function on the state space can be so expressed. Moreover, if P is reversible with respect to the stationary distribution π, then the eigenfunctions may be chosen to form an orthonormal basis for $L^2(\pi)$ - the space $\{f : S \to \mathbb{R}\}$ with inner product $\langle f, g \rangle_\pi = \sum_{x \in S} f(x)g(x)\pi(x)$ - so that $\alpha_i = \langle f, \varphi_i \rangle_\pi$.

As an example, recall that $\psi(\pi) = \mathrm{Des}_d(\pi^{-1}) - \frac{2nd-d^2-d}{4}$ is an eigenfunction for both the inverse a-shuffle and the random-to-top shuffle with eigenvalues $\lambda = \frac{1}{a}$ and $\lambda = 1 - \frac{2}{n}$, respectively. Accordingly, letting $\{X_k\}$ denote a generic representative of these chains, we have

$$\begin{aligned} E[\psi(X_k) + \frac{2nd-d^2-d}{4} | X_0 = id] &= \lambda^k \psi(id) + \frac{2nd-d^2-d}{4} \\ &= \lambda^k \left(\mathrm{Des}_d(id) - \frac{2nd-d^2-d}{4} \right) + \frac{2nd-d^2-d}{4} \\ &= (1-\lambda^k)\frac{2nd-d^2-d}{4}. \end{aligned}$$

Thus if P_*^k denotes the distribution of a permutation π after k inverse shuffles of an ordered deck of cards, and P^k denotes the distribution after k forward shuffles of an ordered deck, then

$$\begin{aligned}
(1-\lambda^k)\frac{2nd-d^2-d}{4} &= E_{P_*^k}[\mathrm{Des}_d(\pi^{-1})] = \sum_{j=0}^{N_{n,d}} jP_*^k\left(\{\mathrm{Des}_d(\pi^{-1})=j\}\right) \\
&= \sum_{j=0}^{N_{n,d}} j \sum_{\sigma\in S_n} 1\{\mathrm{Des}_d(\sigma^{-1})=j\}P_*^k(\sigma) \\
&= \sum_{j=0}^{N_{n,d}} j \sum_{\sigma\in S_n} 1\{\mathrm{Des}_d(\sigma^{-1})=j\}P^k\left(\sigma^{-1}\right) \\
&= \sum_{j=0}^{N_{n,d}} j \sum_{\tau\in S_n} 1\{\mathrm{Des}_d(\tau)=j\}P^k(\tau) \\
&= \sum_{j=0}^{N_{n,d}} jP^k\left(\{\mathrm{Des}_d(\pi)=j\}\right) = E_{P^k}[\mathrm{Des}_d(\pi)].
\end{aligned}$$

Accordingly, we have the following result.

Proposition 3.3.1. *Beginning with an ordered deck of n cards*

- *The expected number of d-descents after k a-shuffles is $(1-a^{-k})\frac{2nd-d^2-d}{4}$.*
- *The expected number of d-descents after k top-to-random shuffles is $\left[1-\left(1-\frac{2}{n}\right)^k\right]\frac{2nd-d^2-d}{4}$.*

Similar arguments can be used to compute the expectation of other permutation statistics of these shuffles (and many of their relatives) which can be expressed as linear combinations of indicators of inversions of inverses. We can also use this technique to compute the expected number of rising sequences and inversions in the corresponding inverse shuffles.

Proposition 3.3.2. *Beginning with an ordered deck of n cards*

- *The expected number of rising sequences after k inverse a-shuffles is* $\frac{n+1}{2} - a^{-k}\left(\frac{n-1}{2}\right)$.

- *The expected number of rising sequences after k random-to-top shuffles is* $\frac{n+1}{2} - \left(1 - \frac{2}{n}\right)^k \left(\frac{n-1}{2}\right)$.

- *The expected number of inversions after k inverse a-shuffles is* $(1 - a^{-k})\frac{n^2-n}{4}$.

- *The expected number of inversions after k random-to-top shuffles is* $\left[1 - \left(1 - \frac{2}{n}\right)^k\right] \frac{n^2-n}{4}$.

The above should be contrasted with Brad Mann's (apparently laborious) calculation of the expected number of rising sequences after k a-shuffles [49]:

$$E_{P_a^k}[R] = a^k - \frac{n+1}{a^{nk}} \sum_{r=0}^{a^k-1} r^n.$$

Descents and inversions after riffle shuffles have been studied previously in [37, 30, 44], and it is perhaps enlightening to compare the various approaches taken to establish results along the lines of Proposition 3.3.1. Note also that Propositions 3.3.1 and 3.3.2 demonstrate the relation of the subdominant eigenvalue to the rate of convergence of these shuffles as measured by such statistics.

Other uses for eigenfunctions mentioned in [32] include expressing k-step transition probabilities in terms of left and right eigenfunctions and their corresponding eigenvalues; using martingale methods to study the chain via the sequence $Y_k = \frac{1}{\lambda_i^k}\varphi_i(X_k)$; examining k-step self-correlations of eigenfunctions for the chain started in equilibrium; using left

eigenfunctions to study time-reversed chains and quasi-stationary distributions; and exploiting the fact that left and right eigenfunctions corresponding to different eigenvalues are orthogonal. (This latter fact can be used to show that a necessary condition for a sequence of ergodic Markov chains $\{X_k^{(n)}\}$ with total variation mixing times $t_{mix}^{(n)}$ and absolute spectral gaps γ_n to exhibit the cutoff phenomenon is that $\lim_{n\to\infty} \gamma_n t_{mix}^{(n)} = \infty$. Proofs which assume that the chains are reversible can be found in [33, 48], but if one replaces the term $\langle f, 1\rangle_\pi$ with $\langle f, \pi\rangle$ in these proofs and uses the orthogonality of left and right eigenfunctions, then the reversibility assumption can be dropped.) In the rest of this section, we consider two other uses for Markov chain eigenfunctions: bounding mixing times and applications to Stein's method computations.

3.3.1 Mixing Time Bounds

3.3.1.1 Wilson's Method and Lower Bounds

We have already mentioned that eigenfunctions can be used to obtain lower bounds on the k-step distance to stationarity via the method of distinguishing statistics and the use of Wilson's technique. The general idea behind this approach is to use the fact that for any subset B of the state space, one has $\left|P_x^k(B) - \pi(B)\right| \leq \max_{A\subseteq S} \left|P_x^k(A) - \pi(A)\right| = \left\|P_x^k - \pi\right\|_{TV}$. The method of distinguishing statistics employs this bound with the event $B = \{x : \phi(x) \leq \alpha\}$ for some constant α and some function ϕ. Typically, ϕ is taken to be an eigenfunction of P so that one can estimate the first two moments of $\phi(X)$ for $X \sim P_x^k$ and $X \sim \pi$, and bound the probability of B under P_x^k and π using Chebychev's inequality. Examples of direct computations using this idea can be found in [21, 40].

In general, if ϕ is a real-valued right eigenfunction of P with eigenvalue λ, then

$$E_{P_x^k}[\phi(X)] = \sum_{y \in S} P^k(x,y)\phi(y) = \left(P^k\phi\right)(x) = \lambda^k\phi(x),$$

and if $\lambda \neq 1$, then $E_\pi[\phi(X)] = \sum_{y\in S}\pi(y)\phi(y) = \langle \pi, \phi\rangle = 0$ since π is a left eigenfunction of P with eigenvalue $1 \neq \lambda$. To estimate the variance of $\phi(X)$ under P_x^k and π, respectively, we can often appeal to the following argument due to David Wilson [70].

Suppose that ϕ is a right eigenfunction of P with eigenvalue $(1-\gamma)$ where $0 < \gamma < 1$, and let

$$R \geq E\left[\left(\varphi_A(X_1) - \varphi_A(y)\right)^2 \middle| X_0 = y\right]$$

for all $y \in S$. Let $X_0 = x$ denote the initial state of the chain. By induction, we have that $E[\phi(X_t)] = (1-\gamma)^t\phi(x)$. Also,

$$\begin{aligned}
E[\phi(X_{t+1})^2|X_t] &= E\left[(\phi(X_{t+1}) - \phi(X_t))^2 + 2\phi(X_{t+1})\phi(X_t) - \phi(X_t)^2|X_t\right] \\
&\leq R + 2\phi(X_t)E[\phi(X_{t+1})|X_t] - \phi(X_t)^2 \\
&= R + 2\phi(X_t)(1-\gamma)\phi(X_t) - \phi(X_t)^2 \\
&= R + (1-2\gamma)\phi(X_t)^2,
\end{aligned}$$

hence

$$E[\phi(X_{t+1})^2] = E\left[E[\phi(X_{t+1})^2|X_t]\right] \leq R + (1-2\gamma)E\left[\phi(X_t)^2\right].$$

Solving this recursive inequality yields

$$E[\phi(X_t)^2] \leq (1-2\gamma)^t\phi(x)^2 + R\sum_{i=0}^{t-1}(1-2\gamma)^i$$

$$= (1-2\gamma)^t \phi(x)^2 + R\frac{1-(1-2\gamma)^t}{2\gamma},$$

so that

$$\begin{aligned}\mathrm{Var}\left(\phi(X_t)\right) &= E[\phi(X_t)^2] - E[\phi(X_t)]^2 \\ &\leq \phi(x)^2\left[(1-2\gamma)^t - (1-\gamma)^{2t}\right] + \frac{R}{2\gamma}\left[1-(1-2\gamma)^t\right].\end{aligned}$$

Since $\gamma \in (0,1)$ and $P_x^k \Rightarrow \pi$, letting $t \to \infty$ gives

$$\mathrm{Var}_\pi(\phi(X)) = \lim_{t\to\infty} \mathrm{Var}\left(\phi(X_t)\right) \leq \frac{R}{2\gamma}.$$

If we impose the additional constraint $0 < \gamma \leq \frac{1}{2}$, then a bit of algebra shows that $(1-2\gamma)^t - (1-\gamma)^{2t} \geq 0$ and thus $\mathrm{Var}\left(\phi(X_t)\right) \leq \frac{R}{2\gamma}$ for all t. (The inequality $(1-2\gamma)^t - (1-\gamma)^{2t} \geq 0$ actually holds for all t whenever $\gamma \leq 2-\sqrt{2}$, but one also has to worry about the $(1-2\gamma)^t$ term if t is odd and $\gamma > \frac{1}{2}$.)

This method of bounding the variance is known as Wilson's technique and is recorded in the following lemma[70].

Lemma 3.3.1 (Wilson)**.** *Suppose that $\{X_k\}_{k=0}^\infty$ is an ergodic Markov chain with state space S, transition matrix P, and stationary distribution π. If ϕ is a right eigenfunction for P with eigenvalue $(1-\gamma)$ for some $\gamma \in (0, \frac{1}{2}]$ and there is some number $R < \infty$ such that*

$$E\left[\left(\phi(X_1) - \phi(x)\right)^2 \middle| X_0 = x\right] \leq R$$

for all $x \in S$, then

$$Var_\pi\left(\phi(X)\right), Var_{P_x^k}\left(\phi(X)\right) \leq \frac{R}{2\gamma}$$

for every $x \in S$, $k \in \mathbb{N}$.

By taking the modulus of various terms when appropriate, the above argument carries over directly to complex valued eigenfunctions/eigenvalues. There are also various extensions of Wilson's theorem which allow for test functions which are not quite eigenfunctions [45] and for linear combinations of eigenfunctions under certain nonpositive covariance assumptions [46]. One uses these results to obtain lower bounds on the variation distance by bounding the probability of events of the form $B = \{x : \phi(x) \leq \alpha\}$ under the respective probabilities using Chebychev's inequality. Another possibility is to use a linear combination of eigenfunctions as a test statistic to obtain bounds using Wilson's technique with concentration inequalities rather than Chebychev, though the issue of dependence makes this difficult in general. Also, one may be able to get improved bounds using Wilson's technique and the following variation bound from [48] in place of Chebychev's inequality.

Lemma 3.3.2. *Let μ and ν be probability distributions on Ω and let f be a real valued function on Ω. If $|E_\mu(f) - E_\nu(f)| \geq r\sigma$ where $\sigma^2 = \frac{1}{2}[Var_\mu(f) + Var_\nu(f)]$, then $\|\mu - \nu\|_{TV} \geq 1 - \frac{4}{4+r^2}$.*

With the preceding results in hand, we are in a position to find lower bounds on the convergence rates of some hyperplane chamber walks. We will concentrate on random walks on the braid arrangement which satisfy the conditions of Theorem 3.2.1. For these chains, an eigenbasis for the subdominant eigenvalue is given by

$$\phi_{(i,j)}(\pi) = 1\{\pi^{-1}(i) < \pi^{-1}(j)\} - \frac{1}{2},\ 1 \leq i < j \leq n,$$

thus for any $A \subseteq \{(i,j) \in [n]^2 : i < j\}$, the function

$$\varphi_A(\pi) = \sum_{(i,j)\in A} \phi_{(i,j)}(\pi)$$

is also such an eigenfunction. In order to lower-bound the mixing time of these chains in terms of φ_A, it is necessary to upper-bound

$$\max_{x\in S} E\left[\left(\varphi_A(X_1) - \varphi_A(x)\right)^2 \middle| X_0 = x \right].$$

Now for any $\pi_0 \in S_n$, let $\pi \sim P_{\pi_0}$ (the distribution after a single step in the chain starting at π_0) and set

$$X_{(i,j)} = X_{(i,j)}(\pi) = 1\{\pi^{-1}(i) < \pi^{-1}(j)\} - 1\{\pi_0^{-1}(i) < \pi_0^{-1}(j)\}.$$

Then

$$\begin{aligned} &E\left[\left(\varphi_A(\pi) - \varphi_A(\pi_0)\right)^2\right] \\ &\qquad = E\left[\left(\sum_{(i,j)\in A} \left(1\{\pi^{-1}(i) < \pi^{-1}(j)\} - 1\{\pi_0^{-1}(i) < \pi_0^{-1}(j)\}\right)\right)^2\right] \\ &\qquad\qquad = \sum_{(i,j)\in A} E\left[X_{(i,j)}^2\right] + 2 \sum_{\substack{(i,j),(k,l)\in A: \\ (i,j)<(k,l)}} E\left[X_{(i,j)} X_{(k,l)}\right] \end{aligned}$$

where the pairs are ordered lexicographically. We will assume henceforth that $n > 4$.

Using the card shuffling interpretation of the chain, we see that $X_{(i,j)} = -1$ if card i precedes card j originally and card j is moved above card i after a shuffle, $X_{(i,j)} = 1$ if card i follows card j originally and is card i is moved above card j after a shuffle, and $X_{(i,j)} = 0$ if cards i and j maintain their relative order after a shuffle. Thus $X_{(i,j)}^2 = 1$ if the relative ordering of cards i and j changes after a shuffle and $X_{(i,j)}^2 = 0$ if cards i

and j maintain their relative order after a shuffle. Since the face measure is invariant under transpositions, the probability that i and j maintain their original relative orders is independent of the initial state π and is the same for all pairs (i,j). Thus for all $(i,j) \in A$, $E\left[X_{(i,j)}^2\right] = P\left(X_{(i,j)}^2 = 1\right) = P\left(E_{(1,2)}\right)$ where $E_{(1,2)}$ is the event that the relative orders of cards 1 and 2 change after a shuffle. Now consider the terms $E\left[X_{(i,j)}X_{(k,l)}\right]$ with $(i,j) < (k,l)$. If either cards i and j or cards k and l retain their original order after a shuffles, then $X_{(i,j)}X_{(k,l)} = 0$, and the probability of this occurring does not depend on the initial configuration of the deck since, for example, the probability that i is moved in front of j is the same as the probability that j is moved in front of i by invariance. If both pairs of cards change their relative order, then $X_{(i,j)}X_{(k,l)} = \pm 1$. Moreover, in order for $X_{(i,j)}X_{(k,l)} = -1$, it must be the case that either card i initially preceded card j or card k initially preceded card l. The initial state maximizing $E\left[\left(\varphi_A(X_1) - \varphi_A(x)\right)^2 \middle| X_0 = x\right]$ is thus $\pi_0 = \pi_{rev}$ defined by $\pi_{rev}(k) = n - k + 1$. In this case, $E_{(1,2)}$ is the event that a shuffle moves card 1 in front of card 2. By invariance, this is the same as the probability that card 2 is moved in front of card 1, so, since the sum of these two probabilities is $1 - \lambda_{(1,2)} = 1 - \lambda$, $E\left[X_{(i,j)}^2\right] = P\left(E_{(1,2)}\right) = \frac{1-\lambda}{2}$. Thus, taking $\pi \sim P_{\pi_{rev}}$, and letting $E_{(i,j),(k,l)} = \{\pi^{-1}(i) < \pi^{-1}(j), \pi^{-1}(k) < \pi^{-1}(l)\}$ be the event that the relative orders of cards i and j and cards k and l both change after a single shuffle of the reverse-ordered deck, we see that

$$R = E\left[\left(\varphi_A(\pi) - \varphi_A(\pi_{rev})\right)^2\right] = \sum_{(i,j)\in A} E\left[X_{(i,j)}^2\right] + 2 \sum_{\substack{(i,j),(k,l)\in A: \\ (i,j)<(k,l)}} E\left[X_{(i,j)}X_{(k,l)}\right]$$

$$= |A|\,\frac{1-\lambda}{2} + 2 \sum_{\substack{(i,j),(k,l)\in A: \\ (i,j)<(k,l)}} P\left(E_{(i,j),(k,l)}\right).$$

At this point, we note that the additional constraints $i < j$, $k < l$ allow only for the following orderings of the indices: 1) $i < j < k < l$; 2) $i < k < j < l$; 3) $i < k < l < j$; 4) $i < k < j = l$; 5) $i < j = k < l$; 6) $i = k < j < l$. In the first three cases (where the indices are all distinct), it follows from invariance that

$$P\left(E_{(i,j),(k,l)}\right) = P\left(E_{(1,2),(3,4)}\right) = \sum_{\substack{F\in\mathcal{F}:\\ \sigma_{(1,2)}(F)=\sigma_{(3,4)}(F)=+}} w(F).$$

In case 4, invariance implies that

$$P\left(E_{(i,j),(k,j)}\right) = P\left(E_{(1,3),(2,3)}\right) = \sum_{\substack{F\in\mathcal{F}:\\ \sigma_{(1,3)}(F)=\sigma_{(2,3)}(F)=+}} w(F).$$

Similarly, in case 5 we have

$$P\left(E_{(i,j),(j,l)}\right) = P\left(E_{(1,2),(2,3)}\right) = \sum_{\substack{F\in\mathcal{F}:\\ \sigma_{(1,2)}(F)=\sigma_{(2,3)}(F)=+}} w(F),$$

and in case 6 we have

$$P\left(E_{(i,j),(i,l)}\right) = P\left(E_{(1,2),(1,3)}\right) = \sum_{\substack{F\in\mathcal{F}:\\ \sigma_{(1,2)}(F)=\sigma_{(1,3)}(F)=+}} w(F)$$

Therefore, writing

$$A_1 = \left\{((i,j),(k,l)) \in A^2 : i,j,k,l \text{ are distinct}\right\},$$

$$A_2 = \left\{((i,j),(k,l)) \in A^2 : i < k < j = l\right\},$$

$$A_3 = \left\{((i,j),(k,l)) \in A^2 : i < j = k < l\right\},$$

$$A_4 = \left\{((i,j),(k,l)) \in A^2 : i = k < j < l\right\},$$

we have

$$R = |A| \frac{1-\lambda}{2} + 2 \sum_{\substack{(i,j),(k,l)\in A: \\ (i,j)<(k,l)}} P\left(E_{(i,j),(k,l)}\right)$$

$$= |A| \frac{1-\lambda}{2} + 2\,|A_1| \sum_{\substack{F\in\mathcal{F}: \\ \sigma_{(1,2)}(F)=\sigma_{(3,4)}(F)=+}} w(F) + 2\,|A_2| \sum_{\substack{F\in\mathcal{F}: \\ \sigma_{(1,3)}(F)=\sigma_{(2,3)}(F)=+}} w(F)$$

$$+ 2\,|A_3| \sum_{\substack{F\in\mathcal{F}: \\ \sigma_{(1,2)}(F)=\sigma_{(2,3)}(F)=+}} w(F) + 2\,|A_4| \sum_{\substack{F\in\mathcal{F}: \\ \sigma_{(1,2)}(F)=\sigma_{(1,3)}(F)=+}} w(F).$$

One can show that $\sum_{\substack{F\in\mathcal{F}: \\ \sigma_{(1,2)}(F)=\sigma_{(3,4)}(F)=+}} w(F) = \frac{1-2\lambda+\lambda_{H_{(1,2)}\cap H_{(3,4)}}}{4}$ using invariance and inclusion-exclusion, but since this does not generally simplify computations and the arguments do not carry over to the cases where the indices are not all distinct, we omit this fact in our description of R. Before moving on to specific examples, we note that the above decomposition can be quite helpful in deciding how to choose A both in terms of achieving optimal bounds and in reducing the amount of computations needed.

At this point, we turn our attention to the random-to-top shuffle. We have already seen that the subdominant eigenvalue is given by $\lambda = 1 - \frac{2}{n}$. Because the face measure is supported on two-block partitions where the first block is a singleton, it is clear that there are no faces F with $w(F) > 0$ and $\sigma_{(1,2)}(F) = \sigma_{(3,4)}(F) = +$, $\sigma_{(1,3)}(F) = \sigma_{(2,3)}(F) = +$ or $\sigma_{(1,2)}(F) = \sigma_{(2,3)}(F) = +$. Moreover, the only face in the support of w with $\sigma_{(1,2)}(F) = \sigma_{(1,3)}(F) = +$ is $F \sim (\{1\}, [n] \setminus \{1\})$, which has weight $w(F) = \frac{1}{n}$. Consequently, for any choice of A, we have

$$R = \frac{|A|}{n} + 2\,|A_4| \sum_{\substack{F\in\mathcal{F}: \\ \sigma_{(1,2)}(F)=\sigma_{(1,3)}(F)=+}} w(F) = \frac{|A| + 2\,|A_4|}{n}.$$

If we take $A = \{(i,j) \in [n]^2 : i < j\}$, then $|A| = \binom{n}{2}$ and $|B_4| = \binom{n}{3}$, hence $R = \frac{2n^2-3n+1}{6}$, so, since $\gamma = 1 - \lambda = \frac{2}{n}$, Wilson's method gives

$$\mathrm{Var}_\pi(\varphi_A), \mathrm{Var}_{P^k}(\varphi_A) \leq \frac{R}{2\gamma} = \frac{2n^3 - 3n^2 + n}{24}.$$

For this choice of A, $-\varphi_A$ is the number of inversions normalized to have mean 0 under the uniform distribution on S_n, and it is easy to show that the number of inversions in a random permutation has variance $\frac{2n^3+3n^2-5n}{72}$, so our estimate has the correct order. As this bound is independent of the initial state and so is the mixing time, it suffices to assume that the chain starts at the identity. In this case, using the fact that $\log(1-x) \geq -x - x^2$ for $0 \leq x \leq \frac{1}{2}$, we have

$$\begin{aligned}
\left| E_{P^k_{id}}[\varphi_A] - E_\pi[\varphi_A] \right| &= \left| \left(1 - \frac{2}{n}\right)^k \varphi_A(id) \right| = \left(1 - \frac{2}{n}\right)^k \frac{n(n-1)}{4} \\
&= \exp\left[k \log\left(1 - \frac{2}{n}\right)\right] \frac{n(n-1)}{4} \\
&\geq \exp\left[-\frac{2k}{n}\left(\frac{n+2}{n}\right)\right] \frac{n(n-1)}{4}
\end{aligned}$$

and, since we are assuming that $n > 4$,

$$\begin{aligned}
\sigma^2 &= \frac{1}{2}\left(\mathrm{Var}_{P^k_{id}}(\varphi_A) + \mathrm{Var}_\pi(\varphi_A)\right) \leq \frac{1}{2}\left(\frac{2n^3 - 3n^2 + n}{24} + \frac{2n^3 + 3n^2 - 5n}{72}\right) \\
&= \frac{4n^3 - 3n^2 - n}{72} \leq \frac{6n^3 - 12n^2 + 6n}{72} = \frac{n(n-1)^2}{12},
\end{aligned}$$

so we can take

$$\begin{aligned}
r &= \frac{\exp\left[-\frac{2k}{n}\left(\frac{n+2}{n}\right)\right] \frac{n(n-1)}{4}}{\sqrt{\frac{n(n-1)^2}{12}}} = \frac{\sqrt{3}}{2} \exp\left[-\frac{2k}{n}\left(\frac{n+2}{n}\right)\right] \sqrt{n} \\
&= \frac{\sqrt{3}}{2} \exp\left[-\frac{2k}{n}\left(\frac{n+2}{n}\right) + \frac{\log(n)}{2}\right]
\end{aligned}$$

in the statement of Lemma 3.3.2, to obtain

$$\left\|P^k - \pi\right\|_{TV^*} = \left\|P^k_{id} - \pi\right\|_{TV} \geq 1 - \frac{4}{4+r^2} \geq 1 - 4r^{-2}$$
$$\geq 1 - \frac{16}{3}\exp\left[\frac{4k}{n}\left(\frac{n+2}{n}\right) - \log(n)\right].$$

Taking $k = \frac{1}{4}n\log(n) - cn$ for any $c > 0$ yields $\left\|P^k - \pi\right\|_{TV^*} \geq 1 - \frac{16}{3}e^{-4c+1}$, so the mixing time for the random-to-top (and thus for the top-to-random) shuffle is at least $\frac{1}{4}n\log(n)$. It is known that the true mixing time for the random-to-top shuffle is $n\log(n)$, so the bound from Wilson's method is off by a factor of 4.

The same general analysis also applies for the random-m-set-to-top shuffles where m is some fixed number which is small relative to n. It was shown in section 3.2 that the subdominant eigenvalue for these chains is $\lambda = 1 - 2\frac{m(n-m)}{n(n-1)}$. We first observe that since the face measure is supported on two-block partitions, there is no F in the support of w with $\sigma_{(1,2)}(F) = \sigma_{(2,3)}(F) = +$. Also, in order to avoid dealing with the sums over A_2 and A_4, we will take $A = \{(i, i+1), i = 1, ..., n-1\}$, so that $-\varphi_A(\pi)$ is the number of rising sequences in π standardized to have mean zero under the uniform distribution. Finally, simple counting arguments show that

$$\sum_{\substack{F\in\mathcal{F}:\\ \sigma_{(1,2)}(F)=\sigma_{(3,4)}(F)=+}} w(F) = \frac{\binom{n-4}{m-2}}{\binom{n}{m}} = \frac{m(m-1)(n-m)(n-m-1)}{n(n-1)(n-2)(n-3)}$$

and

$$|A_1| = \sum_{i=1}^{n-2}(n-i-2) = \frac{(n-2)(n-3)}{2}.$$

Accordingly, we have the bound

$$R = \frac{m(n-m)}{n} + 2\frac{m(m-1)(n-m)(n-m-1)}{2n(n-1)} = \frac{m^2(n-m)^2}{n(n-1)}$$

so that

$$\mathrm{Var}_\pi(\varphi_A), \mathrm{Var}_{P^k}(\varphi_A) \leq \frac{R}{2\gamma} \leq \frac{m(n-m)}{4}.$$

Under the uniform distribution, φ_A has variance $\frac{n+1}{12}$, so this bound is of the correct order when m is a small fixed number. Arguing as in the random-to-top case, we have

$$\begin{aligned}\left|E_{P^k_{id}}[\varphi_A] - E_\pi[\varphi_A]\right| &= \left|\left(1 - \frac{2m(n-m)}{n(n-1)}\right)^k \varphi_A(id)\right| = \frac{n-1}{2}\left(1 - \frac{2m(n-m)}{n(n-1)}\right)^k \\ &= \exp\left[k\log\left(1 - \frac{2m(n-m)}{n(n-1)}\right)\right]\frac{n-1}{2} \\ &\geq \exp\left[-\frac{2km(n-m)}{n(n-1)}\left(1 + \frac{2m(n-m)}{n(n-1)}\right)\right]\frac{n-m}{2},\end{aligned}$$

so we can take

$$\begin{aligned}r &= \frac{\exp\left[-\frac{2km(n-m)}{n(n-1)}\left(1 + \frac{2m(n-m)}{n(n-1)}\right)\right]\frac{n-m}{2}}{\sqrt{\frac{m(n-m)}{4}}} \\ &= \exp\left[-\frac{2km(n-m)}{n(n-1)}\left(1 + \frac{2m(n-m)}{n(n-1)}\right) + \frac{1}{2}\log\left(\frac{n-m}{m}\right)\right]\end{aligned}$$

to obtain the bound

$$\begin{aligned}\left\|P^k - \pi\right\|_{TV^*} &\geq 1 - 4r^{-2} \\ &= 1 - 4\exp\left[\frac{4km(n-m)}{n(n-1)}\left(1 + \frac{2m(n-m)}{n(n-1)}\right) - \log\left(\frac{n-m}{m}\right)\right].\end{aligned}$$

When $k = \frac{n(n-1)}{4m(n-m)}\log(n) - c\frac{n(n-1)}{m(n-m)}$, we have

$$\begin{aligned}&\exp\left[\frac{4km(n-m)}{n(n-1)}\left(1 + \frac{2m(n-m)}{n(n-1)}\right) - \log\left(\frac{n-m}{m}\right)\right] \\ &= \exp\left[\log(n)\frac{2m(n-m)}{n(n-1)} - 4c\left(1 + \frac{2m(n-m)}{n(n-1)}\right) + \log\left(\frac{n}{n-m}\right) + \log(m)\right] \\ &\leq \exp\left[\frac{2m(n-m)}{n} - 4c + \log(2) + m - 1\right] \leq \frac{3}{4}\exp\left[3m - 4c\right],\end{aligned}$$

hence $\left\|P^k - \pi\right\|_{TV^*} \geq 1 - 3e^{3m-4c}$, which goes to 1 as $c \to \infty$. In fact, the preceding argument shows that the mixing time is at least $\frac{n(n-1)}{4m(n-m)}\log(n)$ when m is not held constant provided that $m = m(n) \in_{n\to\infty} o(\log(n))$. Using the bound from Theorem 2.2.2, we see that if $k = \frac{n(n-1)}{m(n-m)}\log(n) + c\frac{n(n-1)}{m(n-m)}$, then

$$\begin{aligned}\left\|P^k - \pi\right\|_{TV^*} &\leq \binom{n}{2}\left(1 - 2\frac{m(n-m)}{n(n-1)}\right)^k \\ &\leq \frac{n^2}{2}\exp\left(-\frac{2km(n-m)}{n(n-1)}\right) = \frac{1}{2}e^{-2c}.\end{aligned}$$

Thus our lower bound matches the upper bound up to a factor of 4. To the best of the author's knowledge, this is the first nontrivial lower bound which has been obtained for random-m-set-to-top shuffles - the lower bound trick for top-to random shuffles does not appear to be applicable in this case. However, based on the analogy with random-to-top shuffles, the mixing time is probably given by the upper bound.

The same general argument can be carried out for other walks on the braid arrangement satisfying the conditions of Theorem 3.2.1, and the representation of R given above usually makes the calculations quite manageable. Moreover, this representation often provides a good deal of insight about how to choose A both in terms of optimality and computational simplicity. Also, the general idea of using distinguishing statistics and Wilson's variance bound applies to all hyperplane chamber walks, and candidate eigenfunctions can be found with the aid of Theorem 3.1.1. The foregoing was intended primarily as an example of the utility of eigenfunctions for obtaining lower bounds, but it is likely that one can find novel bounds of the correct order by applying these arguments to other chamber walks. However, it should be mentioned that this method works better for some chains than others. For example, it fails spectacularly for lower-bounding the

mixing time of riffle shuffles, while yielding the correct bound for random walk on the hypercube.

3.3.1.2 Stochastic Monotonicity and Upper Bounds

We now show that eigenfunctions can also be used to upper-bound the mixing time of hyperplane chamber walks by appealing to stochastic monotonicity of the Markov kernel and the fact that the top eigenfunctions are also monotone with respect to certain partial orders on the chambers. A Markov kernel P is said to be stochastically monotone with respect to a partial order $\preceq$ on its state space S if for all $x, y \in S$ with $x \preceq y$, one has

$$\sum_{z' \preceq z} P(x, z') \geq \sum_{z' \preceq z} P(y, z') \text{ for all } z \in S.$$

(If S is infinite, then one replaces $\sum_{z' \preceq z} P(x, z')$ with $\int_{z' \preceq z} P(x, dz')$ in the above expression, and similarly for y.)

Building on work of David Wilson [70], Diaconis, Khare, and Saloff-Coste provided some general results for upper-bounding the mixing time of stochastically monotone Markov chains with real-valued state spaces in terms of strictly monotone eigenfunctions [31]. This technique was subsequently generalized to chains having partially ordered state spaces by Khare and Mukherjee in order to deal with multivariate Markov chains. Before presenting a version of these results, we first establish that hyperplane chamber walks are indeed stochastically monotone with respect to several partial orders. We begin with the following lemma, which is probably known but has not been found in the literature by the author.

Lemma 3.3.3. *Suppose that P is the transition matrix for a Markov chain on a partially ordered state space $(S, \preceq)$ which has a random mapping representation (f, Z) where $f : S \times \Lambda \to S$ is monotone nondecreasing in its first argument. Then P is stochastically monotone with respect to $\preceq$.*

Proof. Let $w, x, y \in S$ with $x \preceq y$ be given. Since f is nondecreasing in its first argument, we have that $f(x, z) \preceq f(y, z)$ for all $z \in \Lambda$, thus $\{z \in \Lambda : f(y, z) \preceq w\} \subseteq \{z \in \Lambda : f(x, z) \preceq w\}$. It follows that

$$\begin{aligned}\sum_{w' \preceq w} P(y, w') &= \sum_{w' \preceq w} \mathbb{P}\{f(y, Z) = w'\} = \sum_{w' \preceq w} \sum_{\substack{z \in \Lambda: \\ f(y,z) = w'}} \mathbb{P}\{Z = z\} \\ &= \sum_{\substack{z \in \Lambda: \\ f(y,z) \preceq w}} \mathbb{P}\{Z = z\} \leq \sum_{\substack{z \in \Lambda: \\ f(x,z) \preceq w}} \mathbb{P}\{Z = z\} \\ &= \sum_{w' \preceq w} \mathbb{P}\{f(x, Z) = w'\} = \sum_{w' \preceq w} P(x, w').\end{aligned}$$

If S is infinite, the sums are replaced with integrals. □

Now the partial order on the set of faces introduced in section 2.1 does not distinguish between chambers. However, since chamber walks have random mapping representation (f, Z) where $f : \mathcal{C} \times \mathcal{F} \to \mathcal{C}$ is defined by $f(C, F) = FC$ and Z is distributed according to w, it follows from Lemma 3.3.3 that the associated kernel is stochastically monotone with respect to any partial order on $\mathcal{C}$ satisfying $C \preceq D$ implies $FC \preceq FD$ for all $F \in \mathcal{F}$. In particular, for any fixed chamber $D \in \mathcal{C}$, consider the partial order $C \preceq_D C'$ if $\{i \in [m] : \sigma_i(C) \neq \sigma_i(D)\} \subseteq \{i \in [m] : \sigma_i(C') \neq \sigma_i(D)\}$. For any $F \in \mathcal{F}$ and any $C, C' \in \mathcal{C}$ with $C \preceq_D C'$, if $\sigma_i(FC) \neq \sigma_i(D)$, then either $\sigma_i(F) \neq \sigma_i(D)$ and thus $\sigma_i(FC') \neq \sigma_i(D)$, or $\sigma_i(F) = 0$ and $\sigma_i(C) \neq \sigma_i(D)$ so that $\sigma_i(C') \neq \sigma_i(D)$ and thus $\sigma_i(FC') \neq \sigma_i(D)$.

Therefore, the BHR kernel is stochastically monotone with respect to $\preceq_D$. In the case of the braid arrangement, where the chambers are indexed by permutations, the set of negative sign sequence coordinates of a chamber $C \sim \pi$ is equal to the inversion set of π^{-1}. Thus taking $D \sim id$ so that $\sigma(D) = (+, ..., +)$, one has $\pi_1 \preceq_D \pi_2$ if and only if $\text{Inv}\left(\pi_1^{-1}\right) \subseteq \text{Inv}\left(\pi_2^{-1}\right)$ if and only if $\pi_1^{-1} \preceq_{rwb} \pi_2^{-1}$ if and only if $\pi_1 \preceq_{lwb} \pi_2$ where $\preceq_{lwb}, \preceq_{rwb}$ denote the left and right weak Bruhat orders, respectively. As the left and right weak Bruhat orders are isomorphic via $\pi \mapsto \pi^{-1}$, we will omit this distinction henceforth.

Paul Edelman showed that (an equivalent definition of) the order $\preceq_D$ with $D \sim id$ is isomorphic to the weak Bruhat order on any finite Coxeter group when the group elements are identified with the chambers of the corresponding hyperplane arrangement [35], and it is shown in [9] that this definition of the weak Bruhat order extends naturally to an ordering on the topes of oriented matroids in general, of which the chambers of a real hyperplane arrangement are a special case. For hyperplane arrangements, the latter definition is given in terms of inclusion of the sets of hyperplanes separating the chambers from some distinguished chamber. Our construction of $\preceq_D$ is equivalent to taking the distinguished chamber to be D. Accordingly, we will henceforth refer to $\preceq_D$ as the D-weak Bruhat order on the chambers, and we record this fact as

Theorem 3.3.1. *Define the D-weak Bruhat order on the chambers of a hyperplane arrangement $\mathcal{A} = \{H_i\}_{i=1}^m$ by*

$$C \preceq_D C' \text{ if } \{i \in [m] : \sigma_i(C) \neq \sigma_i(D)\} \subseteq \left\{i \in [m] : \sigma_i(C') \neq \sigma_i(D)\right\}.$$

. Then the BHR walks are stochastically monotone with respect to $\preceq_D$.

It was noted in [12] that hyperplane chamber walks are stochastically monotone with respect to the weak Bruhat order and it is likely that the authors had something like the preceding argument in mind with $D \sim id$. They further observed that since there are least and greatest elements for the Bruhat order ($\hat{0} = D$ and $\hat{1} = -D$, the chamber with all sign sequence coordinates opposite to those of D, according to the definition given here), stochastic monotonicity implies that one can use monotone coupling from the past to obtain exact samples from the stationary distributions of these chains. Using very similar ideas, one can get upper bounds on the variation distance using the methods developed by Diaconis, Khare, and Saloff-Coste in [31]. We will make use of the following generalization of their result to partially ordered state spaces due to Khare and Mukherjee [47].

Theorem 3.3.2 (Khare and Mukherjee)**.** *Let P be the transition density for an ergodic Markov chain having finite state space S equipped with a partial order $\preceq$. Let π denote the stationary distribution and suppose that P is stochastically monotone with respect to $\preceq$. Suppose further that for all $x, y \in S$, there is some $z(x, y) \in S$ such that $z(x, y) \preceq x, y$ or $x, y \preceq z(x, y)$. If $\lambda \in (0, 1)$ is an eigenvalue of P with strictly increasing eigenfunction f, then for all $x \in S$*

$$\left\| P_x^k - \pi \right\|_{TV} \leq \frac{\lambda^k}{c} E\left[|f(Y) + f(x) - 2f\left(z(x, Y)\right)| \right]$$

where $Y \sim \pi$ and

$$c = \inf\{f(t) - f(s) : s \preceq t, s \neq t\}.$$

We remark that Khare and Mukherjee were also able to use stochastic monotonicity and increasing eigenfunctions to obtain the lower bound $\left\|P_x^k - \pi\right\|_{TV} \geq \frac{\lambda^k}{2C} |f(x)|$ where $C = \sup\{|f(z)| : z \in S\}$. However, this bound is equivalent to the universal lower bound presented here in section 1.2 when $|\lambda| = \lambda_*$ and is worse otherwise.

Now in order to apply Theorem 3.3.2 to hyperplane chamber walks we need only to find eigenfunctions which are strictly increasing with respect to the D weak Bruhat order - Theorem 3.3.1 establishes stochastic monotonicity and pairwise dominance is guaranteed since we have maximal and minimal elements. One case in which such eigenfunctions necessarily exist is when $\lambda_{H_1} = ... = \lambda_{H_n}$, as is the case when the conditions of Theorem 3.2.1 are satisfied. This is because Theorem 3.1.1 shows that each λ_{H_i} has eigenfunction

$$\varphi_i(C) = \begin{cases} -\frac{p_i,}{p_i+q_i} & \sigma_i(C) = - \\ \frac{q_i}{p_i+q_i}, & \sigma_i(C) = + \end{cases} \begin{aligned} &= 1\{\sigma_i(C) = +\} - \frac{p_i}{p_i+q_i} \\ &= \frac{q_i}{p_i+q_i} - 1\{\sigma_i(C) = -\} \end{aligned}$$

where $p_i = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = +}} w(F)$ and $q_i = \sum_{\substack{F \in \mathcal{F}: \\ \sigma_i(F) = -}} w(F)$. Thus, setting

$$D^+ = \{i \in [m] : \sigma_i(D) = +\}, \quad D^- = \{i \in [m] : \sigma_i(D) = -\},$$

and

$$K_D = \sum_{i \in D^-} \frac{p_i}{p_i + q_i} + \sum_{i \in D^+} \frac{q_i}{p_i + q_i},$$

we see that

$$\begin{aligned} \varphi(C) &= |\{i \in [m] : \sigma_i(C) \neq \sigma_i(D)\}| - K_D \\ &= \sum_{i \in D^-} \left(1\{\sigma_i(C) = +\} - \frac{p_i}{p_i + q_i}\right) + \sum_{i \in D^+} \left(1\{\sigma_i(C) = -\} - \frac{q_i}{p_i + q_i}\right) \end{aligned}$$

$$= \sum_{i \in D^-} \varphi_i(C) - \sum_{i \in D^+} \varphi_i(C)$$

is an eigenfunction corresponding to $\lambda = \lambda_{H_1} = ... = \lambda_{H_m}$. If $C \preceq C'$ with $C \neq C'$, then $\{i \in [m] : \sigma_i(C) \neq \sigma_i(D)\} \subset \{i \in [m] : \sigma_i(C') \neq \sigma_i(D)\}$, so

$$\begin{aligned}\varphi(C) &= |\{i \in [m] : \sigma_i(C) \neq \sigma_i(D)\}| - K_D \\ &< \left|\{i \in [m] : \sigma_i(C') \neq \sigma_i(D)\}\right| - K_D = \varphi(C'),\end{aligned}$$

hence φ is strictly increasing. Also, it is clear in this case that

$$c = \inf\{\varphi(D) - \varphi(C) : C \preceq D, C \neq D\} = 1,$$

the infimum being achieved when

$$\left\{i \in [m] : \sigma_i(C') \neq \sigma_i(D)\right\} = \{i \in [m] : \sigma_i(C) \neq \sigma_i(D)\} \cup \{j\}$$

with $j \in \{i \in [m] : \sigma_i(C) = \sigma_i(D)\}$.

Now if P is the transition kernel of a hyperplane chamber walk with all hyperplane eigenvalues equal to λ, then given any initial state C_0, P is stochastically monotone with respect to $\preceq_{C_0}$ and φ_{C_0} is a strictly increasing eigenfunction for λ. Moreover, for any $C \in \mathcal{C}$, we have that $z(C, C_0) = C_0$ is dominated by both C and C_0. Thus for any $\mathcal{C}$-valued random variable Y

$$\begin{aligned}&E\left[|\varphi_{C_0}(Y) + \varphi_{C_0}(C_0) - 2\varphi_{C_0}\left(z(C_0, Y)\right)|\right] \\ &\qquad = E\left[|\varphi_{C_0}(Y) - \varphi_{C_0}(C_0)|\right] \\ &\qquad = E\left[||\{i \in [m] : \sigma_i(Y) \neq \sigma_i(C_0)\}| - K_{C_0} - (-K_{C_0})|\right] \\ &\qquad = E\left[|\{i \in [m] : \sigma_i(Y) \neq \sigma_i(C_0)\}|\right]\end{aligned}$$

$$= E\left[\sum_{i=1}^{m} 1\{\sigma_i(Y) \neq \sigma_i(C_0)\}\right]$$
$$= \sum_{i=1}^{m} \mathbb{P}\{\sigma_i(Y) \neq \sigma_i(C_0)\}.$$

Therefore, Theorem 3.3.2 implies

Theorem 3.3.3. *Let P be the transition density for an ergodic hyperplane chamber walk with stationary distribution π. If the hyperplane eigenvalues satisfy $\lambda_{H_1} = ... = \lambda_{H_m} = \lambda \in (0,1)$, then for any $k \in \mathbb{N}$ and any initial state $C_0 \in \mathcal{C}$, we have*

$$\left\| P^k_{C_0} - \pi \right\|_{TV} \leq \lambda^k \sum_{i=1}^{m} \mathbb{P}\{\sigma_i(Y) \neq \sigma_i(C_0)\}$$

where $Y \sim \pi$.

Specializing to random walks on the chambers of the braid arrangement in $\mathbb{R}^n$ with translation invariant face measure, we have $m = \binom{n}{2}$ and π uniform over $\mathcal{C}$ so that $\mathbb{P}\{\sigma_{(i,j)}(Y) \neq \sigma_{(i,j)}(C_0)\} = \frac{1}{2}$ for all $i < j$, $C_0 \in \mathcal{C}$, hence

Corollary 3.3.1. *Suppose that P is the transition density of a random walk on the braid arrangement in $\mathbb{R}^n$ satisfying the conditions of Theorem 3.2.1. Then, letting λ denote the subdominant eigenvalue of P, we have the variation bound*

$$\left\| P^k - \pi \right\|_{TV^*} \leq \frac{1}{2}\lambda^k \binom{n}{2}.$$

Note that the crude bound from Theorem 2.2.2 gives $\left\| P^k - \pi \right\|_{TV^*} \leq \lambda^k \binom{n}{2}$ in cases where the corollary is applicable. Though the extra factor of $\frac{1}{2}$ will not affect bounds on the mixing time, it is often an improvement in concrete applications. For example, if one wanted to know how many times a deck of 52 cards had to be shuffled using the top-to-random method until the total variation distance to stationarity was less than $\frac{1}{10}$,

then the bound from Theorem 2.2.2 would recommend 243 shuffles whereas the Corollary 3.3.1 shows that 225 shuffles suffice. If one is playing with an eight-deck shoe, as is the case in some casinos, Corollary 3.3.1 would save the dealer over 330 shuffles.

In addition, this example serves to further motivate the method of monotone eigenfunctions for upper-bounding variation distance from [31], and to demonstrate its applicability to partially ordered state spaces as in [47]. Also, it is illuminating to be able to compare the bounds obtained by different techniques in specific examples of Markov chains in order to understand the types of chains for which the various methods are applicable. When contrasting the bounds on shuffles from Theorem 2.2.2 and Corollary 3.3.1, it is not surprising that one obtains effectively the same mixing time estimates since both results ultimately involve couplings with coupling times given by the number of iterations required for all pairwise comparisons of cards to be made. Still, in order to address more fundamental questions about Markov chain mixing times, such as "Under which conditions does the cut-off phenomenon occur?" or "What are the most important features of the transition rule for computing the mixing time?", it is useful to assemble a variety of specific examples and general techniques so as to obtain a more panoramic view of the subject. In many senses, this is the underlying motivation for studying hyperplane walks and one of the main contributions of this thesis.

In the spirit of investigating various features of the transition mechanism and their broader applications, we now shift gears and explore some connections between Markov chain eigenfunctions and Stein's method.

3.3.2 Connections with Stein's Method

Over the past four decades, probabilists and statisticians have developed a wide range of techniques, inspired by Charles Stein's seminal paper [65], for bounding the distance between the distribution of a random variable X and and that of a random variable Z having some specified target distribution. The metrics for which these techniques are applicable are of the form $d_{\mathcal{H}}(\mathscr{L}(X), \mathscr{L}(Z)) = \sup_{h \in \mathcal{H}} |E[h(X)] - E[h(Z)]|$ for some suitable class of functions $\mathcal{H}$, and include as special cases the Wasserstein, Kolmogorov, and total variation distances. (The Kolmogorov distance gives the L^∞ distance between the associated distribution functions, so $\mathcal{H} = \{1_{(-\infty,a]}(x) : a \in \mathbb{R}\}$. The total variation and Wasserstein distances correspond to letting $\mathcal{H}$ consist of indicators of Borel sets and 1-Lipschitz functions, respectively.) The basic idea is to find an operator $\mathcal{A}$ such that $E[(\mathcal{A}f)(X)] = 0$ for all f belonging to some sufficiently large class of functions $\mathcal{F}$ if and only if $\mathscr{L}(X) = \mathscr{L}(Z)$. For example, Charles Stein showed that $E[(\mathcal{A}_N f)(Z)] = E[Zf(Z) - \sigma^2 f'(Z)] = 0$ for all absolutely continuous functions f such that the expectations exist if and only if $Z \sim N(0, \sigma^2)$, and his student Louis Chen showed shortly thereafter that $Z \sim \text{Poisson}(\lambda)$ if and only if $E[(\mathcal{A}_P f)(Z)] = E[Zf(Z) - \lambda f(Z+1)] = 0$ for all functions f for which the expectations exist [17]. Similar characterizing operators have since been worked out for several other distributions (see [59] for some examples). Given such an operator $\mathcal{A}$ for $\mathscr{L}(Z)$, one can consider the solution $f_h \in \mathcal{F}$ to the equation $(\mathcal{A}f)(x) = h(x) - E[h(Z)]$ for $h \in \mathcal{H}$. Taking expectations, absolute values, and suprema gives

$$d_{\mathcal{H}}(\mathscr{L}(X), \mathscr{L}(Z)) = \sup_{h \in \mathcal{H}} |E[h(X)] - E[h(Z)]| = \sup_{h \in \mathcal{H}} |E[(\mathcal{A}f_h)(X)]| \,.$$

The intuition is that since $E[(\mathcal{A}f)(Z)] = 0$ for $f \in \mathcal{F}$, the distribution of X should be close to that of Z when $E[(\mathcal{A}f)(X)]$ is close to zero. Remarkably, it is often easier to work with the right-hand side of the above equation, and the tools for analyzing distances between distributions in this manner are collectively known as Stein's method. For more on this rich and fascinating subject, the author highly recommends [66, 59, 17, 26].

3.3.2.1 Using Stein's Method to Study Eigenfunctions

The connection between Stein's method and eigenvectors of Markov chains arises from one of the earliest and most common techniques within this framework for bounding the distance to normality, namely the use of Stein pairs introduced in [66]. A pair of random variables (W, W') is called a γ-Stein pair if it is exchangeable - that is, $(W, W') =_d (W', W)$ - and satisfies the linearity condition $E[W'|W] = (1-\gamma)W$ for some $\gamma \in (0, 1)$. A typical result involving Stein pairs is given by the following theorem due to Yosef Rinott and Vladimir Rotar [56]

Theorem 3.3.4 (Rinott and Rotar)**.** *Let W, W' be an exchangeable pair of real-valued random variables such that $E[W'|W] = (1-\gamma)W$ for some $\gamma \in (0, 1)$. If there exists a constant A such that $|W' - W| \leq A$ almost surely, then for all $x \in \mathbb{R}$,*

$$|P(W \leq x) - \Phi(x)| \leq \frac{12}{\gamma}\sqrt{\mathrm{Var}(E[(W'-W)^2|W]} + 48\frac{A^3}{\gamma} + 8\frac{A^2}{\sqrt{\gamma}}$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$ is the standard normal c.d.f.

It is often the case that Stein pairs are constructed by letting W and W' be functions of successive steps in a reversible Markov chain started in equilibrium [17, 59]. To see how this works, suppose that W is a random variable defined on a discrete probability space

$(\Omega, \mathcal{F}, \pi)$ and that (f, Z) is a random mapping representation of the transition matrix, P, of a Markov chain with state space Ω and stationary distribution π. Define a new random variable W' on $(\Omega, \mathcal{F}, \pi)$ by $W'(\omega) = W(f(\omega, Z))$. If P is reversible with respect to π, then this construction shows that for all $a, b \in W(\Omega) = \{W(\omega) : \omega \in \Omega\}$, we have

$$\begin{aligned}
\mathbb{P}\{W = a, W' = b\} &= \sum_{x \in W^{-1}(a)} \left(\pi(x) \sum_{y \in W^{-1}(b)} P(x, y) \right) \\
&= \sum_{x \in W^{-1}(a)} \sum_{y \in W^{-1}(b)} \pi(x) P(x, y) = \sum_{x \in W^{-1}(a)} \sum_{y \in W^{-1}(b)} \pi(y) P(y, x) \\
&= \sum_{y \in W^{-1}(b)} \left(\pi(y) \sum_{x \in W^{-1}(a)} P(y, x) \right) = \mathbb{P}\{W = b, W' = a\},
\end{aligned}$$

hence (W, W') is exchangeable. Note that we do not require that $W_n = W(\omega_n)$ is a Markov chain for the exchangeability condition to be satisfied. Indeed, this will only be the case if the underlying chain satisfies the lumping criterion from Theorem 1.3.1 with respect to the equivalence relation induced by W.

The linearity condition is typically verified by showing that $E[W'|\omega] = (1 - \gamma)W$ and then using the fact that $\sigma(W) \subseteq \sigma(\omega)$ to conclude that

$$E[W'|W] = E[E[W'|\omega]|W] = E[(1 - \gamma)W|W] = (1 - \gamma)W.$$

Because

$$E[W'|\omega] = E[W(f(\omega, Z))] = \sum_{y \in \Omega} W(y) \mathbb{P}\{f(\omega, Z) = y\} = \sum_{y \in \Omega} P(\omega, y) W(y),$$

the condition $E[W'|\omega] = (1 - \gamma)W$ is equivalent to the statement that W is a right eigenfunction for P with eigenvalue $(1 - \gamma)$, and this fact does not depend on reversibility.

The above discussion shows that one way to obtain a Stein pair involving a random variable W on $(\Omega, \mathcal{F}, \pi)$ is to construct a reversible Markov chain on Ω with stationary distribution π for which W is a (right) eigenfunction. (It also shows that when (W, W') is a γ-Stein pair constructed in terms of a Markov chain on the underlying space and $E[W'|\omega] = (1-\gamma)W$, then W is actually an eigenfunction of that chain.) Observe that since P has real entries, γ will be real provided that W is, and the spectral theorem shows that this always holds for reversible chains. In addition, as long as the chain is irreducible and W is not constant, we know from section 1.1 that $1-\gamma < 1$, so W has expectation 0 under π (because π is a left eigenfunction with eigenvalue $1 \neq 1-\gamma$ and thus is orthogonal to W), and we can always normalize so that $\mathrm{Var}(W) = E[W^2] = 1$. Also, this approach guarantees that W and W' have common law since ω and $f(\omega, Z)$ both have distribution π.

The following (special case of a) theorem due to Adrian Röllin shows that one can get bounds which are similar to those coming from Stein pairs without requiring (W, W') to be exchangeable, provided that they are equidistributed with mean zero and variance one, so this approach is actually quite general [57].

Theorem 3.3.5 (Röllin)**.** *Suppose that W and W' are a pair of real-valued random variables having common law such that $E[W] = 0$, $\mathrm{Var}(W) = 1$, and $E[W'|W] = (1-\gamma)W$ for some $\gamma \in (0,1)$. Suppose moreover that there is a constant A such that $|W' - W| \leq A$ almost surely. Then for all $x \in \mathbb{R}$,*

$$|P(W \leq x) - \Phi(x)| \leq \frac{12}{\gamma}\sqrt{\mathrm{Var}(E[(W'-W)^2|W])} + 32\frac{A^3}{\gamma} + 6\frac{A^2}{\sqrt{\gamma}}$$

where Φ is the standard normal c.d.f.

In light of Röllin's theorem, we see that one can use the above construction to study the asymptotic normality of Markov chain eigenfunctions even if the chain is not reversible.

One of the earliest uses of a nonreversible Markov chain to construct Stein pairs was in [39] to study the asymptotics of the number of descents and inversions in a random permutation. Using a similar construction, the author was able to extend this result to obtain rates of normal convergence for the number of generalized descents of a random permutation [54]. (Recall from subsection 3.2 that for $\pi \in S_n$ and $1 \leq d < n$, a pair (i, j) with $i < j \leq i + d$ and $\pi(i) > \pi(j)$ is called a d-descent of π and $\mathrm{Des}_d(\pi)$ denotes the number of these generalized descents.) In both cases, the pairs are constructed by taking the real, skew-symmetric $n \times n$ matrix $M(n, d) = [M_{i,j}(n, d)]_{i,j=1}^n$ given by

$$M_{i,j}(n,d) = \begin{cases} -1, & i < j \leq i + d \\ 1, & j < i \leq i + d \\ 0, & \text{otherwise} \end{cases}$$

and defining the random variable $Z_{n,d}$ by $Z_{n,d}(\pi) = \sum_{i<j} M_{\pi(i),\pi(j)}(n, d)$. One then computes that $Z_{n,d}(\pi) = 2\mathrm{Des}_d(\pi^{-1}) - \frac{2nd-d^2-d}{2}$ where $\frac{2nd-d^2-d}{2}$ is the expected number of descents. Next, one defines the random variable $W_{n,d}(\pi) = \frac{Z_{n,d}(\pi)}{\sqrt{\mathrm{Var}(Z_{n,d})}}$ which is distributed as the normalized number of d-descents of π. The complementary random variable $W'_{n,d}$ is defined by choosing I uniformly from $[n] = \{1, 2, ..., n\}$ and setting $W'_{n,d}(\pi) = W_{n,d}(\pi\sigma_I)$ where σ_I is the cycle $(I, I + 1, ..., n)$. Thus W' is obtained from W by performing a "random-to-bottom" shuffle on the input. After some careful calculations, one finds that $E[W'_{n,d}|W_{n,d}] = (1 - \frac{2}{n})W_{n,d}$. It follows that, for $1 \leq d \leq n - 1$, $W_{n,d}$ is a right eigenfunction of the random-to-end shuffle on S_n with eigenvalue $(1 - \frac{2}{n})$ [54]. Fulman's work

on the $d = 1$ and $d = n - 1$ cases predated Röllin's theorem and required additional arguments to establish that (W, W') was exchangeable, but one can circumvent this step by arguing as in the preceding paragraphs. After much work estimating the terms in Theorem 3.3.2 (guided largely by Fulman's analysis for descents and inversions), and deriving the variance of $\mathrm{Des}_d(\pi)$ under the uniform distribution on S_n, it was eventually established that [54]

Theorem 3.3.6. *The number of d-descents in a random permutation of length n satisfies*

$$\left| P\left(\frac{\mathrm{Des}_d - \mu_{n,d}}{\sigma_{n,d}} \leq x \right) - \Phi(x) \right| \leq \begin{cases} M_1 d^{\frac{3}{2}} n^{-\frac{1}{2}}, & d(n) \leq \sqrt{(n)} \\ M_2 n d^{-\frac{3}{2}}, & \sqrt{(n)} < d(n) \leq \frac{n}{2} \\ M_3 n^{-\frac{1}{2}}, & 2d(n) > n \\ M(d) n^{-\frac{1}{2}}, & d \text{ fixed} \end{cases}$$

where Φ is the standard normal c.d.f., M_1, M_2, M_3, and $M(d)$ are constants which do not depend on n, $\mu_{n,d} = \frac{2nd - d^2 - d}{4}$, and $\sigma_{n,d} = \sqrt{\mathrm{Var}(Des_d)}$.

Theorem 3.3.3 was deduced before the present work on eigenfunctions for hyperplane walks, but the construction of the pair (W, W') is made clear in retrospect: A random-to-bottom shuffle is combinatorially equivalent to random-to-top shuffle, and the results of section 3.2 show that

$$-Z_{n,d}(\pi) = \frac{2nd - d^2 - d}{2} - 2\mathrm{Des}_d(\pi^{-1})$$

is an eigenfunction corresponding to $1 - \frac{2}{n}$.

The main point of the foregoing is that Markov chain eigenfunctions are often interesting in their own right and their asymptotic properties can be studied using Stein's

method techniques. Many other questions about random permutations can be addressed by looking at statistics which can be represented in terms of linear combinations of indicators of inversions, and the fact that such statistics arise as eigenfunctions of Markov chains with nice combinatorial interpretations suggests the possibility of studying them via Stein's method. It is shown in [32] that other permutation statistics which can be expressed as linear combinations of products of indicators of inversions are also eigenfunctions for a-shuffles, and it is likely the case that this is true for top-to-random shuffles as well. (Inspection of Theorems 3.3.1 and 3.3.2 show that one typically gets better bounds by using eigenfunctions corresponding to larger eigenvalues, so the top-to-random shuffles should yield better results than a-shuffles. See [58] for a more extensive discussion.) This could be verified by using the lumping techniques and symmetry arguments from section 3.1 to study the eigenfunctions corresponding to the third and fourth largest eigenvalues and would open up a wide variety of permutation statistics whose asymptotic properties could likely be deduced using the aforementioned methodology. Another interesting line of research would be to carry out a similar program using eigenfunctions for the Tsetlin library. The stationary distribution of this chain is known as the Luce model in cognitive science and the above may offer a tractable approach to study its properties. See chapter 9 in [21] for more on the Luce model and its statistical applications.

3.3.2.2 Eigenfunctions for Random Walks on S_n

Having seen how one can exploit the fact that some statistics of interest are eigenfunctions of Markov chains to facilitate Stein's method computations, we now turn to an example of a sort of methodological converse: It is sometimes the case that calculations carried out

in Stein's method arguments produce Markov chain eigenfunctions as a byproduct. This turns out to be the case in the derivation of convergence rates in Hoeffding's combinatorial central limit theorem, originally worked out by Bolthausen [11]. The present argument is based on the discussion in [17] which establishes an L^1 bound on the convergence rate. Before stating the result, we need to introduce the following notation. Define

$$a_{\bullet\bullet} = \frac{1}{n^2}\sum_{i,j=1}^{n} a_{i,j},\ a_{i\bullet} = \frac{1}{n}\sum_{j=1}^{n} a_{i,j},\ \text{and } a_{\bullet j} = \frac{1}{n}\sum_{i=1}^{n} a_{i,j},$$

and let

$$\mu = n a_{\bullet\bullet},$$

$$\sigma^2 = \frac{1}{n-1}\sum_{i,j=1}^{n}(a_{i,j} - a_{i\bullet} - a_{\bullet j} + a_{\bullet\bullet})^2,$$

$$\gamma = \sum_{i,j=1}^{n} \left|a_{i,j} - a_{i\bullet} - a_{\bullet j} + a_{\bullet\bullet}\right|^3.$$

Then one has

Theorem 3.3.7. *For $n \geq 3$, let $[a_{i,j}]_{i,j=1}^n$ be the components of a matrix $A \in M_n(\mathbb{R})$, let π be a random permutation uniformly over S_n, and let $Y = \sum_{i=1}^{n} a_{i,\pi(i)}$. If F denotes the distribution function of $W = \frac{Y-\mu}{\sigma}$, and Φ denotes the standard normal distribution function, then the L^1 distance between F and Φ satisfies*

$$\|F - \Phi\|_1 \leq \frac{\gamma}{(n-1)\sigma^3}\left(16 + \frac{56}{n-1} + \frac{8}{(n-1)^2}\right).$$

The proof of Theorem 3.3.4 is based on the construction of a zero bias variable using an exchangeable pair constructed in terms of a random transposition walk on S_n. Briefly,

by replacing $a_{i,j}$ with $a_{i,j} - a_{i\bullet} - a_{\bullet j} + a_{\bullet\bullet}$, it can be assumed without loss of generality that $a_{i\bullet} = a_{\bullet j} = a_{\bullet\bullet} = 0$. Then taking $Y(\pi) = \sum_{i=1}^{n} a_{i,\pi(i)}$, one defines the random variable $Y'(\pi) = Y(\pi\tau)$ where τ is uniformly distributed over the $\binom{n}{2}$ transpositions in S_n. The Stein condition is established by computing $E[Y' - Y|\pi] = -\frac{2}{n-1}Y$. As detailed in the previous sub-subsection, this shows that Y is an eigenfunction (corresponding to $\lambda = 1 - \frac{2}{n-1}$) for the random walk on S_n driven by the measure which is uniform over the conjugacy class of 2-cycles. Examination of the proof shows that it suffices to assume that $a_{i\bullet} = 0$ for all i. We take this observation as our starting point.

Let $A = [a_{i,j}]_{i,j=1}^{n}$ be any matrix with real entries which satisfies

$$\sum_{j=1}^{n} a_{i,j} = 0$$

for all $i \in [n]$, and define the map, $\varphi_A : S_n \to \mathbb{R}$ by

$$\varphi_A(\pi) = \sum_{i=1}^{n} a_{i,\pi(i)}.$$

Note that $\varphi_A(\pi)$ gives the result of right-multiplying A by the permutation matrix corresponding to π (that is, the defining representation of S_n evaluated at π) and taking the trace.

Let $\{X_k\}_{k=1}^{\infty}$ denote the random transposition walk on S_n having transition matrix

$$P(\pi, \sigma) = \mathbb{P}\{X_{k+1} = \sigma | X_k = \pi) = \begin{cases} \binom{n}{2}^{-1}, & \sigma\pi^{-1} \in \mathcal{T}_n \\ 0, & \text{else} \end{cases}$$

where $\mathcal{T}_n = \{\tau \in S_n : \tau = (i,j), 1 \leq i < j \leq n\}$ denotes the set of transpositions. (Typical treatments of the random transposition walk add a holding probability to deal

with periodicity problems. Note however that the measure driving the random transposition walk with holding is still constant on conjugacy classes, the eigenfunctions are the same with or without holding, and the eigenvalues transform linearly.) The proof of the combinatorial central limit theorem using Stein pairs essentially shows that φ_A is a right eigenfunction for P with eigenvalue $\lambda = 1 - \frac{2}{n-1}$ (provided that φ_A is not the zero map). To see that this is indeed the case, we compute

$$\begin{aligned}
[P\varphi_A](\pi) &= \sum_{\sigma \in S_n} P(\pi, \sigma)\varphi_A(\sigma) = \binom{n}{2}^{-1} \sum_{\tau \in \mathcal{T}_n} \varphi_A(\tau\pi) = \frac{2}{n(n-1)} \sum_{\tau \in \mathcal{T}_n} \sum_{k=1}^{n} a_{k,\tau(\pi(k))} \\
&= \frac{2}{n(n-1)} \sum_{i<j} \left[\left(\sum_{k \neq \pi^{-1}(i), \pi^{-1}(j)} a_{k,\pi(k)} \right) + a_{\pi^{-1}(i),j} + a_{\pi^{-1}(j),i} \right] \\
&= \frac{2}{n(n-1)} \sum_{i<j} \left(\varphi_A(\pi) - a_{\pi^{-1}(i),i} - a_{\pi^{-1}(j),j} + a_{\pi^{-1}(i),j} + a_{\pi^{-1}(j),i} \right) \\
&= \varphi_A(\pi) - \frac{2}{n(n-1)} \sum_{i<j} \left(a_{\pi^{-1}(i),i} + a_{\pi^{-1}(j),j} - a_{\pi^{-1}(i),j} - a_{\pi^{-1}(j),i} \right).
\end{aligned}$$

Because the summands are symmetric in i and j, we may write

$$\begin{aligned}
[P\varphi_A](\pi) &= \varphi_A(\pi) - \frac{2}{n(n-1)} \sum_{i<j} \left(a_{\pi^{-1}(i),i} + a_{\pi^{-1}(j),j} - a_{\pi^{-1}(i),j} - a_{\pi^{-1}(j),i} \right) \\
&= \varphi_A(\pi) - \frac{1}{n(n-1)} \sum_{\substack{i,j: \\ i \neq j}} \left(a_{\pi^{-1}(i),i} + a_{\pi^{-1}(j),j} - a_{\pi^{-1}(i),j} - a_{\pi^{-1}(j),i} \right) \\
&= \varphi_A(\pi) - \frac{1}{n(n-1)} \sum_{\substack{k,l: \\ k \neq l}} \left(a_{k,\pi(k)} + a_{l,\pi(l)} - a_{k,\pi(l)} - a_{l,\pi(k)} \right) \\
&= \varphi_A(\pi) - \frac{2}{n(n-1)} \sum_{\substack{k,l: \\ k \neq l}} \left(a_{k,\pi(k)} - a_{k,\pi(l)} \right)
\end{aligned}$$

$$= \varphi_A(\pi) - \frac{2}{n(n-1)} \sum_{k=1}^{n} \left(\sum_{\substack{l: \\ l \neq k}} a_{k,\pi(k)} - \sum_{\substack{l: \\ l \neq k}} a_{k,\pi(l)} \right)$$

$$= \varphi_A(\pi) - \frac{2}{n(n-1)} \sum_{k=1}^{n} \big((n-1) a_{k,\pi(k)} - (-a_{k,\pi(k)}) \big)$$

$$= \varphi_A(\pi) - \frac{2}{n-1} \varphi_A(\pi)$$

where the third equality reindexed the sum by $k = \pi^{-1}(i)$, $l = \pi^{-1}(j)$ and the penultimate equality used the fact that the rows of A sum to zero. Therefore, if A is any real $n \times n$ matrix whose rows sum to zero, then $P\varphi_A = \left(1 - \frac{2}{n-1}\right) \varphi_A$.

At this point, we observe that we may as well work with matrices whose columns also sum to zero since if $A \in M_n(\mathbb{R})$ is any row-sum zero matrix, then taking $A' = [a'_{i,j}]_{i,j=1}^n$ with $a'_{i,j} = a_{i,j} - a_{\bullet j}$, we have

$$\begin{aligned}
\varphi_{A'}(\pi) &= \sum_{k=1}^{n} a'_{k,\pi(k)} = \sum_{k=1}^{n} (a_{k,\pi(k)} - a_{\bullet\pi(k)}) = \sum_{k=1}^{n} a_{k,\pi(k)} - \sum_{k=1}^{n} a_{\bullet\pi(k)} \\
&= \sum_{k=1}^{n} a_{k,\pi(k)} - \sum_{k=1}^{n} \sum_{i=1}^{n} a_{i,\pi(k)} = \sum_{k=1}^{n} a_{k,\pi(k)} - \sum_{i=1}^{n} \sum_{k=1}^{n} a_{i,\pi(k)} \\
&= \sum_{k=1}^{n} a_{k,\pi(k)} - \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j} = \sum_{k=1}^{n} a_{k,\pi(k)} - \sum_{i=1}^{n} 0 \\
&= \sum_{k=1}^{n} a_{k,\pi(k)} = \varphi_A(\pi).
\end{aligned}$$

Now, define

$$\mathcal{A} = \{A = [a_{i,j}]_{i,j=1}^n \in M_n(\mathbb{R}) : \sum_{k=1}^{n} a_{i,k} = \sum_{k=1}^{n} a_{k,j} = 0 \text{ for all } i, j \in [n]\}$$

and let $\mathcal{F} = \{\varphi_A : A \in \mathcal{A}\}$. It follows from linearity that $\varphi_A - \varphi_B = \varphi_{A-B}$ and $c\varphi_A = \varphi_{cA}$ for all $A, B \in \mathcal{A}$, $c \in \mathbb{R}$, so, since $\mathcal{A}$ is closed under linear combinations, $\mathcal{F}$ is a vector space. We also observe that the nonzero constant functions are right eigenfunctions of

P having eigenvalue 1, so means that φ_A is constant in π if and only if $\varphi_A \equiv 0$. One checks that if π is uniform over S_n and Y is the random variable $Y(\pi) = \varphi_A(\pi)$ for any $A \in M_n(\mathbb{R})$, then $\mathrm{Var}(Y) = \frac{1}{n-1}\sum_{i,j=1}^{n} |a_{i,j} - a_{i\bullet} - a_{\bullet j} + a_{\bullet\bullet}|^2$. It follows that if $A \in \mathcal{A}$, then φ_A has variance $\frac{1}{n-1}\sum_{i,j=1}^{n} |a_{i,j}|^2$ under the uniform distribution on S_n, so φ_A is nonconstant if and only if $a_{i,j} = 0$ for all $i, j \in [n]$. Accordingly, $\varphi_A = 0$ if and only if $A = 0$, so for all $A, B \in \mathcal{A}$, $\varphi_A = \varphi_B$ if and only if $0 = \varphi_A - \varphi_B = \varphi_{A-B}$ if and only if $A - B = 0$. Therefore, $\dim(\mathcal{F}) = \dim(\mathcal{A}) = (n-1)^2$, a basis for the latter being $\{E^{(i,j)}\}_{i,j=1}^{n-1}$ with

$$E^{(i,j)}_{k,j} = \begin{cases} 1, & i = k, j = l \\ -1, & i = k, j = n \text{ or } i = n, j = l \\ 0, & \text{otherwise} \end{cases} .$$

In order to find a basis for $\mathcal{F}$, note that for any $A \in \mathcal{A}$, the fact that the rows of A sum to 0 implies

$$\begin{aligned} \varphi_A(\pi) &= \sum_{k=1}^{n} a_{k,\pi(k)} = \sum_{k=1}^{n-1}(a_{k,\pi(k)} - a_{n,\pi(k)}) = \sum_{i=1}^{n-1}\sum_{j=1}^{n}(a_{i,j} - a_{n,j})1\{\pi(i) = j\} \\ &= \sum_{i=1}^{n-1}\sum_{j=1}^{n-1}(a_{i,j} - a_{n,j})1\{\pi(i) = j\} + \sum_{i=1}^{n-1}(a_{i,n} - a_{n,n})1\{\pi(i) = n\} \\ &= \sum_{i=1}^{n-1}\sum_{j=1}^{n-1}(a_{i,j} - a_{n,j})1\{\pi(i) = j\} - \sum_{i=1}^{n-1}\sum_{j=1}^{n-1}(a_{i,j} - a_{n,j})1\{\pi(i) = n\} \\ &= \sum_{i=1}^{n-1}\sum_{j=1}^{n-1}(a_{i,j} - a_{n,j})(1\{\pi(i) = j\} - 1\{\pi(i) = n\}) = \sum_{i,j=1}^{n-1} a'_{i,j} e^{(i,j)}(\pi) \end{aligned}$$

where $a'_{i,j} = a_{i,j} - a_{n,j}$ and $e^{(i,j)}(\pi) = 1\{\pi(i) = j\} - 1\{\pi(i) = n\}$ for $i, j \in [n-1]$. Consequently, $\mathcal{F}$ is spanned by the $(n-1)^2$ functions $\{e^{(i,j)}\}_{i,j=1}^{n-1}$, so $\{e^{(i,j)}\}_{i,j=1}^{n-1}$ is a basis for $\mathcal{F}$. We will see shortly that the dimension of the eigenspace corresponding to $1 - \frac{2}{n-1}$

is $(n-1)^2$, so $\{e^{(i,j)}\}_{i,j=1}^{n-1}$ is actually a basis for the $\left(1-\frac{2}{n-1}\right)$-eigenspace of the random transposition walk. (Note that the above argument carries through directly if we work over $\mathbb{C}$ instead of $\mathbb{R}$. However, this is not an issue since the irreducible characters of S_n are rational-valued, so it will follow from the ensuing arguments that the eigenvalues of random walks on S_n which are driven by measures that are constant on conjugacy classes are all real-valued.)

At this point we change gears for a moment and consider a fascinating result due to Peter Matthews which was alluded to in section 3.2. Recall that for any probability measure μ on a finite group $G = \{s_1, ..., s_N\}$, one can construct a Markov chain on G by defining the transition matrix $Q(s_i, s_j) = \mu(s_j s_i^{-1})$ - the random transposition walk being the special case where $G = S_n$ and μ is uniform over $\mathcal{T}_n$. In this setting, let $\rho_1, ..., \rho_K$ denote the irreducible representations of G with respect to a basis such that $\rho_j(s_i)$ is unitary for each i, j. Let $\chi_1, ..., \chi_K$ denote their respective characters and $d_1, ..., d_K$ their degrees (so $\chi_j(s) = \mathrm{Tr}(\rho_j(s))$, $d_j = \chi_j(id)$), and, through a slight abuse of notation, write the Fourier transform of μ at the representation ρ as $\widehat{Q}(\rho) = \sum_{i=1}^{N} \mu(s_i)\rho(s_i)$. For each $k = 1, ..., K$, define the $d_k^2 \times d_k^2$ block diagonal matrix

$$M_k = I \otimes \widehat{Q}(\rho_k) = \begin{bmatrix} \widehat{Q}(\rho_k) & & 0 \\ & \ddots & \\ 0 & & \widehat{Q}(\rho_k) \end{bmatrix}$$

and define the $N \times N$ block diagonal matrix

$$M = \begin{bmatrix} M_1 & & & 0 \\ & M_2 & & \\ & & \ddots & \\ 0 & & & M_K \end{bmatrix}.$$

(The dimensions check out since the sum of the squares of the degrees of the irreducible complex representations of a finite group equals the order of the group.)

Let

$$\psi_k(s) = \sqrt{\frac{d_k}{N}} \begin{bmatrix} \rho_k(s)_{1,1} & \cdots & \rho_k(s)_{d_k,1} & \cdots & \rho_k(s)_{1,d_k} & \cdots & \rho_k(s)_{d_k,d_k} \end{bmatrix}^T$$

be the column vector in $\mathbb{C}^{d_k^2}$ obtained by stacking the columns of $\rho_k(s)$ and normalizing by $\sqrt{\frac{d_k}{N}}$, and let

$$\psi(s) = [\ \psi_1(s)^T \quad \psi_2(s)^T \quad \cdots \quad \psi_K(s)^T]^T$$

be the column vector in $\mathbb{C}^N$ obtained by likewise concatenating the $\psi_k(s)$'s. It is worth observing that we can write $\psi_k(s) = \sqrt{\frac{d_k}{N}}\mathrm{vec}(\rho_k(s))$ where vec is the vectorization operator appearing in the formula from matrix calculus [41]

$$\mathrm{Tr}(A^*BC) = \mathrm{vec}(A)^*\mathrm{vec}(BC) = \mathrm{vec}(A)^*(I \otimes B)\mathrm{vec}(C).$$

Define the $N \times N$ matrix ϕ by

$$\phi = [\ \psi(s_1) \quad \psi(s_2) \quad \cdots \quad \psi(s_N)]\ .$$

Then

$$\phi^* = \begin{bmatrix} \psi(s_1)^* \\ \psi(s_2)^* \\ \vdots \\ \psi(s_N)^* \end{bmatrix} = \begin{bmatrix} \psi_1(s_1)^* & \psi_2(s_1)^* & \cdots & \psi_K(s_1)^* \\ \psi_1(s_2)^* & \psi_2(s_2)^* & \cdots & \psi_K(s_2)^* \\ \vdots & \vdots & & \vdots \\ \psi_1(s_N)^* & \psi_2(s_N)^* & \cdots & \psi_K(s_N)^* \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{\frac{d_1}{N}}\overline{\rho_1(s_1)_{1,1}} & \cdots & \sqrt{\frac{d_1}{N}}\overline{\rho_1(s_1)_{d_1,d_1}} & \cdots & \sqrt{\frac{d_K}{N}}\overline{\rho_K(s_1)_{d_K,d_K}} \\ \sqrt{\frac{d_1}{N}}\overline{\rho_1(s_2)_{1,1}} & \cdots & \sqrt{\frac{d_1}{N}}\overline{\rho_1(s_2)_{d_1,d_1}} & \cdots & \sqrt{\frac{d_K}{N}}\overline{\rho_K(s_2)_{d_K,d_K}} \\ \vdots & & \vdots & & \vdots \\ \sqrt{\frac{d_1}{N}}\overline{\rho_1(s_N)_{1,1}} & \cdots & \sqrt{\frac{d_1}{N}}\overline{\rho_1(s_N)_{d_1,d_1}} & \cdots & \sqrt{\frac{d_K}{N}}\overline{\rho_K(s_N)_{d_K,d_K}} \end{bmatrix}.$$

Recall that if we define an inner product on the space of functions $\{f : G \to \mathbb{C}\}$ by $\langle f, g\rangle_G = \frac{1}{N}\sum_{s\in G} f(s)\overline{g(s)}$, then it follows from the Schur orthogonality relations that the matrix entries of the irreducible representations, which we have taken to be unitary, satisfy $\langle(\rho_a)_{i,j}, (\rho_b)_{k,l}\rangle_G = 0$ for all i, j, k, l if $a \neq b$, and $\langle(\rho_a)_{i,j}, (\rho_a)_{k,l}\rangle = \frac{1}{d_a}\delta_{i,k}\delta_{j,l}$. Consequently, the columns of ϕ^* are orthonormal with respect to the standard inner product on $\mathbb{C}^N$, so the matrix ϕ is unitary.

Using Fourier analysis, Matthews showed that we can write $Q = \phi^* M^* \phi$. To see that this is true, note that by Fourier inversion we have

$$\begin{aligned} Q(s_i, s_j) = \mu(s_j s_i^{-1}) &= \frac{1}{N}\sum_{k=1}^{K} d_k \mathrm{Tr}\left(\rho_k(s_i s_j^{-1})\widehat{Q}(\rho_k)\right) \\ &= \frac{1}{N}\sum_{k=1}^{K} d_k \mathrm{Tr}\left(\rho_k(s_i)\rho_k(s_j^{-1})\widehat{Q}(\rho_k)\right) \\ &= \frac{1}{N}\sum_{k=1}^{K} d_r \mathrm{Tr}\left(\rho_k(s_i)\rho_k(s_j)^*\widehat{Q}(\rho_k)\right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N}\sum_{k=1}^{K} d_r \operatorname{Tr}\left(\rho_k(s_j)^* \widehat{Q}(\rho_k)\rho_k(s_i)\right) \\
&= \sum_{k=1}^{K} \sqrt{\frac{d_r}{N}}\operatorname{vec}\left(\sqrt{\frac{d_k}{N}}\rho_k(s_j)\right)^* \left(I \otimes \widehat{Q}(\rho_k)\right)\sqrt{\frac{d_r}{N}}\operatorname{vec}\left(\sqrt{\frac{d_k}{N}}\rho_k(s_i)\right) \\
&= \sum_{k=1}^{K} \psi_k(s_j)^* M_k \psi_k(s_i) = \left(\phi^* M \phi\right)(s_j, s_i) \\
&= \left(\phi^* M \phi\right)^T (s_i, s_j) = \overline{\left(\phi^* M \phi\right)^* (s_i, s_j)} = \overline{\left(\phi^* M^* \phi\right)(s_i, s_j)},
\end{aligned}
$$

thus $Q = \overline{(\phi^* M^* \phi)}$ and the result follows upon taking conjugates and keeping in mind that the entries of Q are real valued. The author finds it more instructive to leave this decomposition in the form $Q = \overline{(\phi^* M^* \phi)} = \phi^T M^T \overline{\phi} = \Phi M^T \Phi^{-1}$ where

$$
\Phi = \phi^T =
$$

$$
\begin{bmatrix}
\sqrt{\frac{d_1}{N}}\rho_1(s_1)_{1,1} & \cdots & \sqrt{\frac{d_1}{N}}\rho_1(s_1)_{d_1,d_1} & \cdots & \sqrt{\frac{d_K}{N}}\rho_K(s_1)_{d_K,d_K} \\
\sqrt{\frac{d_1}{N}}\rho_1(s_2)_{1,1} & \cdots & \sqrt{\frac{d_1}{N}}\rho_1(s_2)_{d_1,d_1} & \cdots & \sqrt{\frac{d_K}{N}}\rho_K(s_2)_{d_K,d_K} \\
\vdots & & \vdots & & \vdots \\
\sqrt{\frac{d_1}{N}}\rho_1(s_N)_{1,1} & \cdots & \sqrt{\frac{d_1}{N}}\rho_1(s_N)_{d_1,d_1} & \cdots & \sqrt{\frac{d_K}{N}}\rho_K(s_N)_{d_K,d_K}
\end{bmatrix}.
$$

(Since ϕ is unitary, $\phi^{-1} = \phi^*$, so $\Phi^{-1} = (\phi^T)^{-1} = (\phi^{-1})^T = (\phi^*)^T = \overline{\phi}$.)

The utility of this decomposition becomes especially evident when the measure μ driving the random walk is constant on conjugacy classes. This is because Fourier transforms of class functions are homotheties. Specifically, if $f(s) = f(t)$ whenever s and t are conjugate, then, for ρ a unitary irreducible representation of G, $\widehat{f}(\rho) = \lambda I$ with I the $d_\rho \times d_\rho$ identity matrix and $\lambda = \frac{1}{d_\rho}\sum_{s\in G} f(s)\chi_\rho(s)$. It follows that M is a diagonal matrix (and thus equals its transpose) whenever μ is constant on conjugacy classes, so Q decomposes as $Q = \Phi M \Phi^{-1}$ where M is a diagonal matrix and Φ is a unitary matrix. As such, the

columns of Φ give an orthonormal basis of eigenvectors and the diagonal entries of M give the corresponding eigenvalues. We record this result as

Theorem 3.3.8. *Suppose that Q is the transition matrix for a random walk on a finite group G which is driven by a measure μ that is constant on the conjugacy classes of G. If $\rho_1, ..., \rho_K$ are the irreducible (complex, unitary) representations of G where ρ_k has degree d_k and character χ_k, then for each $k = 1, ..., K$, Q has an eigenvalue $\lambda_k = \frac{1}{d_k} \sum_{s \in G} \mu(s)\chi_k(s)$ occurring with multiplicity d_k^2. An orthonormal basis of eigenfunctions for the eigenspace corresponding to λ_k is given by the (normalized) matrix entries of ρ_k, $\{\sqrt{\frac{d_k}{|G|}}\rho_k(s)_{i,j}\}_{i,j=1}^{d_k}$.*

Theorem 3.3.5 has appeared in the literature under various guises for at least 50 years (see [22] for some history), but the author feels that the result is stated most conveniently above and that the use of the *vect* operator makes the proof more transparent than many other treatments. Also, it is worth explicitly stating that the eigenfunctions of Q depend only on the irreducible representations of G, so every random walk on G driven by a measure which is constant on conjugacy classes has the same set of eigenspaces (though the eigenvalues will vary with the measure defining the walk). Thus if we have determined an eigenbasis corresponding to the representation ρ_k for one such Markov chain, then we have found that eigenbasis for all of them.

Specializing to the case where Q is the transition matrix for the random transposition walk on S_n and using the fact that $\rho_{def} = \rho_{triv} \oplus \rho_{std}$ where ρ_{def} is the n-dimensional defining representation whose character is given by $\chi_{def}(\pi) = fp(\pi) = \{\#$ of fixed points of $\pi\}$, $\rho_{triv} \equiv 1$ is the 1-dimensional trivial representation, and ρ_{std}

is the $(n-1)$-dimensional standard representation, the latter two being irreducible, we see that the $\chi_{std}(\tau) = (n-2) - 1$ for every transposition $\tau \in S_n$, so the eigenvalue of Q parametrized by ρ_{std} is $\lambda = \frac{n-3}{n-1} = 1 - \frac{2}{n-1}$ occurring with multiplicity $(n-1)^2$. Also, it follows from the proof of Lemma 1 in chapter 3D of [21] that there are no other irreducible representations whose character ratio at a transposition is $\lambda = \frac{n-3}{n-1}$, so the dimension of the eigenspace corresponding to λ is exactly $(n-1)^2$. By Theorem 3.3.5, an orthonormal basis of eigenfunctions for λ is given by $\{\sqrt{\frac{n-1}{n!}}\rho_{std}(\pi)_{i,j}\}_{i,j=1}^{n-1}$.

Recalling that the $(n-1)^2$ functions $e^{(i,j)}(\pi) = 1\{\pi(i) = j\} - 1\{\pi(i) = n\}$, $i, j \in [n-1]$, form a basis for an $(n-1)^2$-dimensional space of $\left(1 - \frac{2}{n-1}\right)$-eigenfunctions for the random transposition walk, we have established

Theorem 3.3.9. *Suppose that Q is the transition matrix for a random walk on S_n which is driven by a measure μ that is constant on conjugacy classes. Then, letting $fp(\pi)$ denote the number of fixed points of a permutation π, Q has an eigenvalue $\lambda = \sum_{\pi \in S_n} \mu(\pi)\frac{fp(\pi)-1}{n-1}$ occurring with multiplicity $(n-1)^2$. A basis for the corresponding eigenspace is given by $e^{(i,j)}(\pi) = 1\{\pi(i) = j\} - 1\{\pi(i) = n\}$, $i, j \in [n-1]$.*

Because $\{\rho_{std}(\pi)_{i,j}\}_{i,j=1}^{n-1}$ and $\{e^{(i,j)}(\pi)\}_{i,j=1}^{n-1}$ are both bases for the $\frac{n-3}{n-1}$-eigenspace for the random transposition walk, we have

Proposition 3.3.3. *The matrix entries of the standard representation of S_n form a basis for $\{\varphi_A : A \in \mathcal{A}\}$.*

It may be of interest to observe that the standard representation is the restriction of the natural action of S_n on $\mathbb{C}^n$ (i.e. $\pi e_i = e_{\pi(i)}$) to the $(n-1)$-dimensional subspace of vectors whose coordinates sum to zero, while the description of the eigenvectors from the

CCLT involves matrices whose rows and columns sum to zero. The author is not sure what to do with Proposition 3.3.3, but finds it very intriguing nonetheless.

Chapter 4

Concluding Remarks

The primary contribution of this thesis has been the general derivation of the top eigenfunctions for Markov chains which can be represented as random walks on the chambers of hyperplane arrangements. By adopting a perspective which focuses on the sign sequence coordinates of the states and the update mechanisms, it is quite natural to employ ideas related to lumping and products in this analysis. A broad lesson to be taken away from this work is that when studying a random process, it may be fruitful to consider whether it can be broken down into simpler components and examined from that point of view. One advantage is that the constituent processes may be easier to work with, possibly because of a reduction in the size of the effective state space as was the case here. Also, since the time needed for the original chain to be within some fixed distance of its stationary distribution is necessarily greater than the corresponding time for any of its components, one may be able to find nice lower bounds using such a decomposition. In addition, this shift in focus may help reveal the driving force behind the chain's mixing behavior. For example, the rows of colored tiles paradigm underscores the fact that approach to stationarity for hyperplane walks is largely a matter of coupon collecting. A variety of

upper bounds on Markov chain mixing times - particularly those involving coupling and strong stationary times - use coupon collector arguments (see [21, 53] for example) and in many of these cases the "coupons" correspond to components of the states in some representation. In the present work, we were able to study card shuffling schemes by representing the states in terms of vectors of indicators of inversions, but other possibilities include letting the components correspond to indicators of the position of each card or to the coordinates of the Lehmer code of the associated permutation.

In light of the above observations, it seems useful to pursue some of the ideas in Section 1.3 in greater detail. For example, it would be nice to know more about the relationship between products and projections. It is interesting that several of the chains considered here were amenable to investigation using either approach while others seemed more difficult to study in terms of product chains. In addition, the fact that one could recover eigenfunctions for the random-to-top shuffle using the product chain perspective even though the component chains were not simultaneously diagonalizable suggests that this approach is more generally applicable than the statement of Theorem 1.3.3 would suggest. The author also intends to explore the ramifications of Theorem 1.3.2 for random walks on groups more thoroughly. This idea was given only passing attention here as it was tangential to the primary focus, but it is intriguing that the technique seems to apply so broadly without any assumptions on the measure driving the walk. Like hyperplane chamber walks, random walks on finite groups provide a diverse assortment of interesting Markov chains and further contributions to the theory will almost certainly prove propitious.

On a related note, in the setting of Theorem 1.3.1, one would like to know more about when different equivalence relations yield linearly independent eigenfunctions. As remarked at the end of Section 3.1, it seems likely that ideas related to lumping could be used to recover parts of Theorem 2.2.1 in a more intuitive fashion and a large part of the obstacle seems to be establishing when eigenfunctions corresponding to incomparable flats are linearly independent. Though a variety of compelling proofs of Theorem 2.2.1 have been given and, in particular, the eigenvalue argument from [5] and the diagonalizability argument from [14] are quite simple, the known methods are somewhat unsatisfying in that they rely on technical results more so than intuition concerning the fundamental dynamics. In contrast, the projection paradigm does a partial job explaining why the eigenvalues are indexed by flats and has the possibility of establishing linear independence of certain eigenfunctions using the fact that they are constant over different equivalence classes. Though the author doubts that the full strength of Theorem 2.1.1 can be recovered in this manner, this does seem to be an issue warranting further investigation.

Another important avenue for future research involves constructing further examples of chamber walks and exploring others in greater detail. From a theoretical point of view, more examples provide a richer testing ground for various techniques and greater opportunities to compare and contrast the saliency of different features of related processes. From a more practical point of view, if one can situate problems of interest within this framework, then they can call upon the many known results to simplify their investigations. The author is especially interested in the possibilities afforded by the connection between the braid arrangement and voting systems and intends to give this matter more serious

thought in the future. Also, it seems likely that one could build interesting models using rows of colored tiles with more than two opaque colors. In particular, this may provide a convenient setting for studying certain product chains under less stringent conditions than those given in Theorem 1.3.3. Finally, though the results from Chapter 3 extend quite naturally to oriented matroids and rows of colored tiles, it is less clear how to use these ideas for walks on left-regular bands. It would be nice to have a more concrete description of the eigenvectors than is given in [62] and it is likely that this could be accomplished by considering lumped chains. To motivate this direction of inquiry, it would again be useful to have more examples of such chains (other than hyperplane chamber walks) like those given in [13].

Yet another possible research project would be to compute eigenfunctions corresponding to intersections of multiple hyperplanes. Preliminary investigations and results from [32] concerning smaller eigenvalues for riffle shuffles suggest that such eigenfunctions will involve indicators of more involved inversion relations. One might then be able to construct other permutation statistics than those given here (perhaps related to pattern avoiding permutations or peaks and valleys) and study their behavior using Stein's method techniques as in Subsection 3.3.2. This is admittedly tedious work, but it has the potential to produce interesting and useful theorems. Likewise, one may be able to study other distributions, such as the Luce model, by constructing Stein type pairs from appropriate eigenfunctions.

It would also be productive to try to use Theorem 2.2.3 to find improved upper bounds on the mixing times of some of the walks considered here. Exploratory computations suggest that it will yield decidedly better estimates than the truncated bound in

certain cases, but the author has been unable to obtain definitive results as of yet. It is generally not too hard to explicitly write down the bound, but it seems that fairly clever combinatorial arguments are required to effectively analyze the resulting sums. One of the author's more immediate research plans is to devote some serious attention to these calculations.

Perhaps the most promising extension of the material in this thesis is a generalization of Theorem 3.2.1 to arbitrary finite Coxeter groups. Though the argument given here relied on specific properties of the symmetric group, it seems that similar ideas should apply more generally: One has notions of inversions for all such groups and the symmetries involving sign sequence coordinates should persist in the broader setting. It is also likely that analogous theorems exist for other eigenspaces as well. Moreover, the type of reasoning used to show that the algebraic multiplicity of the top eigenvalue is $\binom{n}{2}$ may be applicable in determining which flats contribute a common eigenvalue. This in turn should prove useful in explaining aspects of Theorem 2.1.1 in terms of projections.

The theory of hyperplane chamber walks offers a broad and unifying perspective on many interesting Markov chains. The utility of this generality lies in the fact that one can simultaneously establish many particular results by analyzing the broader framework (as was done here in finding eigenfunctions and setting up Wilson bounds), and also in establishing reference points with which to compare the differences in related processes which are most important for various questions of interest. For example, comparing the Tsetlin library with the random-to-top shuffle highlights the importance of symmetry, both in regards to simplifying arguments and in contributing to high multiplicity of the subdominant eigenvalue (which is believed to be of central importance in establishing

cut-off [25]). The same applies to random walk on the hypercube with or without bias. In a similar vein, the author finds it instructive to contrast the performance of the techniques considered here for hypercube walks, random-to-top shuffles, and riffle shuffles. For example, hypercube walks admit an almost complete description in terms of product chains, random-to-top shuffles can be studied to some extent within this framework, and riffle shuffles seem to resist all such efforts. In this case, it seems that the operative difference is the extent to which various components interact: Hypercube walks change at most one coordinate at each time step, top-to-random shuffles change only in terms of relationships involving a single card, and the number of pairwise relative orderings of cards which change after a single riffle shuffle can be on the order of the effective state space. Likewise, Wilson's method gives the correct lower bound for hypercube walks, the correct order for random-to-top shuffles, and is useless for riffle shuffles. This seems primarily to be a function of the size of the subdominant eigenvalue. Finally, the upper bounds coming from coupling/strong stationary times (both in terms of the truncated bound from Theorem 2.2.2 and the bound using stochastic monotonicity) are pretty good for all three chains, but while giving the correct mixing rates for hypercube walks on random-to-top shuffles, the bounds for riffle shuffles are again found to be lacking and it is not as clear why this is so. The issue here does not seem to be the size of the top eigenvalue, and all three chains are highly symmetric. The latter two even have the same subdominant eigenspace. To get the correct bound for riffle shuffles, one must consider more subtle aspects such as how the number of rising sequences changes after a shuffle [6]. Though this information is encoded in the top eigenfunctions, the author has been

unable to exploit this fact to obtain optimal bounds. However, this does provide some indication that further uses of eigenfunctions are awaiting discovery.

As alluded to in the preceding paragraph, much of the motivation for examining particular Markov chains or classes thereof is the hope that more sweeping principles will reveal themselves in the process. For example, we still do not have a general method of determining whether a given Markov chain presents total variation cut-off without finding explicit bounds. One would like at least to give conditions under which Peres' condition [33] is sufficient as was done for the L^p distance by Chen and Saloff-Coste [16]. In fact, the original aim of this thesis was to investigate the cut-off phenomenon for hyperplane chamber walks, and the author's focus on the behavior of sign sequence coordinates stemmed from studying the problem in this light. Another general concern is the development of new techniques for bounding mixing times. A variety of approaches are known at present, but their applicability varies from case to case and there are many examples for which optimal bounds remain elusive. In particular, there is a shortage of general methods for obtaining lower bounds. It seems to be the consensus in the Markov chain community that lower bounds are generally the easy part because of the definitions of total variation in terms of maxima, but it has been the author's experience that the task is quite formidable when no distinguished events or statistics immediately present themselves. This is largely due to the lack of machinery, eigenvector-based or otherwise, for tackling this problem, and one hopes that by obtaining increasingly complete descriptions of certain classes of Markov chains, it will be possible to test and develop new strategies. It would be especially nice to have spectral lower bounds which take the multiplicity of the subdominant eigenvalue into account, and the author suspects that

information concerning the corresponding eigenbasis would play a role in the derivation if such bounds exist. A final observation is that many of the classical results in Markov chain theory are specific to reversible chains and most investigations have been restricted to this setting accordingly. An advantage of studying hyperplane walks is that they provide a sort of middle ground in that they are diagonalizable with real eigenvalues despite being nonreversible in general. As such, they provide an excellent opportunity to generalize results originally stated for reversible chains, and it would be an interesting exercise to test whether some of these results also hold in the hyperplane setting.

Bibliography

[1] M. Aigner. A course in enumeration. Graduate Texts in Mathematics, 238, *Springer, Berlin,* 2007.

[2] D. Aldous. Random walks on finite groups and rapidly mixing Markov chains. *Seminar on probability, XVII,* 243-297, Lecture Notes in Math., 986, *Springer, Berlin,* 1983.

[3] D. Aldous and P. Diaconis. Shuffling cards and stopping times, *Amer. Math. Monthly,* **93** (1986), 333-348.

[4] D. Aldous and P. Diaconis. Strong uniform times and finite random walks, *Adv. in Appl. Math.* **8** (1987), no. 1, 69-97.

[5] C.A. Athanasiadis and P. Diaconis. Functions of random walks on hyperplane arrangements, *Adv. in Appl. Math.* **45** (2010), no. 3, 410-437.

[6] D. Bayer and P. Diaconis. Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.* **2** (1992), no. 2, 294-313.

[7] P. Bidigare. Hyperplane Arrangement Face Algebras and Their Associated Markov Chains. PhD Thesis, University of Michigan, 1997.

[8] P. Bidigare, P. Hanlon and D. Rockmore. A combinatorial description of the spectrum for the Tsetlin library and its generalization to hyperplane arrangements, *Duke Math. J.* **99** (1999), no. 1, 135-174.

[9] A. Björner, M. Las Vergnas, B. Sturmfels, N. White, and G.M. Ziegler. Oriented matroids. Second edition. Encyclopedia of Mathematics and its Applications, 46. *Cambridge University Press, Cambridge,* 1999. xii+548 pp.

[10] A. Björner. Random walks, arrangements, cell complexes, greedoids, and self-organizing libraries, "Building Bridges - The Lovász Festschrift," *Bolyai Soc. Math. Studies* **19** (2008), *Springer (Berlin) and Janos Bolyai Math. Sci. (Budapest),* 165-203.

[11] E. Bolthausen. An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete* **66** (1984), no. 3, 379-386.

[12] K.S. Brown, P. Diaconis. Random walks and hyperplane arrangements, *Ann. Probab.* **26** (1998), no. 4, 1813-1854.

[13] K.S. Brown, Semigroups, rings, and Markov chains. *J. Theoret. Probab.* **13** (2000), no. 3, 871-938.

[14] K.S. Brown. Semigroup and ring theoretical methods in probability. *Representations of finite dimensional algebras and related topics in Lie theory and geometry,* 3-26, Fields Inst. Commun., 40, *Amer. Math. Soc., Providence, RI,* 2004.

[15] G.Y. Chen. The Cutoff Phenomenon for Finite Markov chains. PhD Thesis, Cornell University, 2006.

[16] G.Y. Chen and L. Saloff-Coste. The cutoff phenomenon for ergodic Markov processes, *Electronic Journal of Probability,* **13** (2008) 26-78.

[17] L.H.Y. Chen, L. Goldstein, and Q.-M. Shao. Normal approximation by Stein's method. Probability and its Applications (New York). *Springer, Heidelberg,* 2011. xii+405 pp.

[18] F. De Mari, M.A. Shayman, Generalized Eulerian numbers and the topology of the Hessenberg variety of a matrix. *Acta Appl. Math.* **12** (1988), no. 3, 213-235.

[19] G. Denham. Eigenvectors for a random walk on a hyperplane arrangement, *Adv. in Appl. Math.* **48** (2012), no. 2, 312-324.

[20] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions, *Z. Wahrsch. Verw. Gebiete* **57** (1981), no. 2, 159-179.

[21] P. Diaconis. Group representations in probability and statistics. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. *Institute of Mathematical Statistics, Hayward, CA,* 1988.

[22] P. Diaconis. Patterned matrices. *Matrix theory and applications (Phoenix, AZ, 1989),* 37-58, Proc. Sympos. Appl. Math., 40, *Amer. Math. Soc., Providence, RI,* 1990.

[23] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains, *Ann. Appl. Probab.* **1** (1991), no. 1, 36-61.

[24] P. Diaconis, J.A. Fill, and J. Pitman. Analysis of top to random shuffles. *Combin. Probab. Comput.* 1 (1992), no. 2, 135-155.

[25] P. Diaconis. The cutoff phenomenon in finite Markov chains, *Proc. Nat. Acad. Sci. U.S.A.* **93** (1996), no. 4, 1659-1964.

[26] P. Diaconis and S. Holmes, editors. Stein's method: expository lectures and applications. Papers from the Workshop on Stein's Method held at Stanford University, Stanford, CA, 1998. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 46. *Institute of Mathematical Statistics, Beachwood, OH,* 2004. vi+139 pp.

[27] P. Diaconis and D. Freedman. Iterated random functions. *SIAM Rev.* **41** (1999), no. 1, 45-76.

[28] P. Diaconis. Mathematical developments from the analysis of riffle shuffling. *Groups, combinatorics & geometry (Durham, 2001),* 73-97, *World Sci. Publ., River Edge, NJ,* 2003.

[29] P. Diaconis and L. Saloff-Coste. Separation cut-offs for birth and death chains. *Ann. Appl. Probab.* **16** (2006), no. 4, 2098-2122.

[30] P. Diaconis and J. Fulman. Carries, shuffling, and an amazing matrix. *Amer. Math. Monthly* **116** (2009), no. 9, 788-803.

[31] P. Diaconis, K. Khare, and L. Saloff-Coste. Gibbs sampling, conjugate priors and coupling. *Sankhya A* **72** (2010), no. 1, 136-169.

[32] P. Diaconis, A. Pang and A. Ram. Hopf algebras and Markov chains: Two examples and a theory. arXiv:1206.3620v1. 2012.

[33] J. Ding, E. Lubetsky and Y. Peres. Total variation cutoff in birth-and-death chains, *Probability Theory and Related Fields,* **146** (2010), 61-85.

[34] R. Durrett. Probability: theory and examples. Fourth edition. Cambridge Series in Statistical and Probabilistic Mathematics. *Cambridge University Press, Cambridge,* 2010.

[35] P.H. Edelman. A partial order on the regions of $\mathbb{R}^n$ dissected by hyperplanes. *Trans. Amer. Math. Soc.* **283** (1984), no. 2, 617-631.

[36] J.A. Fill. Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process, *Ann. Appl. Probab.* **1** (1991), no. 1, 62-87.

[37] J. Fulman. The combinatorics of biased riffle shuffles. *Combinatorica* **18** (1998), no. 2, 173-184.

[38] J. Fulman. Descent algebras, hyperplane arrangements, and shuffling cards. *Proc. Amer. Math. Soc.* **129** (2001), no. 4, 965-973.

[39] J. Fulman. Stein's method and non-reversible Markov chains. *Stein's method: expository lectures and applications,* 69-77, IMS Lecture Notes Monogr. Ser., 46, *Inst. Math. Statist., Beachwood, OH,* 2004.

[40] J. Fulman. Convergence rates of random walk on irreducible representations of finite groups. *J. Theoret. Probab.* **21** (2008), no. 1, 193-211.

[41] A. Graham. Kronecker products and matrix calculus: with applications. Ellis Horwood Series in Mathematics and its Applications. *Ellis Horwood Ltd., Chichester; Halsted Press [John Wiley & Sons, Inc.], New York,* 1981. 130 pp.

[42] R.A. Horn and C.R. Johnson. Matrix analysis. Corrected reprint of the 1985 original. *Cambridge University Press, Cambridge,* 1990.

[43] J.E. Humphreys. Reflection groups and Coxeter groups. Cambridge Studies in Advanced Mathematics, 29. *Cambridge University Press, Cambridge,* 1990.

[44] U. Islak. Descent-inversion statistics in riffle shuffles. arXiv:1303.2387v1. 2013.

[45] J. Jonasson. The overhand shuffle mixes in $\Theta\left(n^2 \log(n)\right)$ steps. *Ann. Appl. Probab.* **16** (2006), no. 1, 231-243.

[46] J. Jonasson. Biased random-to-top shuffling. *Ann. Appl. Probab.* **16** (2006), no. 2, 1034-1058.

[47] K. Khare and N. Mukherjee. Convergence analysis of some multivariate Markov chains using stochastic monotonicity, *Ann. Appl. Probab.* **23** (2013), no.2, 811-833.

[48] D. Levin, Y. Peres and E. Wilmer. Markov Chains and Mixing Times, *American Mathematical Society, Providence, RI,* 2009.

[49] B. Mann. How many times should you shuffle a deck of cards? *Topics in contemporary probability and its applications,* 261-289, Probab. Stochastics Ser., *CRC, Boca Raton, FL,* 1995.

[50] R. Montenegro and P. Tetali. Mathematical Aspects of Mixing Times in Markov Chains, *Foundations and Trends in Theoretical Computer Science,* **1** (2006), no. 3, 237-354.

[51] R. Montenegro. cited 2012: Lecture Notes for Math 8843 at University of Massachusetts, Lowell. [Available online at http://faculty.uml.edu/rmontenegro/8843/index.html].

[52] P. Orlik and H. Terao. Arrangements of hyperplanes. Grundlehren der Mathematischen Wissenschaften, 300. *Springer-Verlag, Berlin,* 1992.

[53] I. Pak. Random Walks on Groups: Strong Uniform Time Approach. PhD Thesis, Harvard University, 1997.

[54] J. Pike. Convergence rates for generalized descents. *Electron. J. Combin.* **18** (2011), no. 1, Paper 236, 14 pp.

[55] J. Richter-Gebert and G.M. Ziegler. Oriented matroids. *Handbook of discrete and computational geometry,* 111-132, CRC Press Ser. Discrete Math. Appl., *CRC, Boca Raton, FL,* 1997.

[56] Y. Rinott, and V. Rotar, On coupling constructions with rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics, *Ann. Appl. Probab.* **7** (1997), 1080-1105.

[57] A. Röllin, A note on the exchangeability condition in stein's method, *Statistics and Probability Letters,* Volume 78 (2008), 1800-1806.

[58] N. Ross. Step size in Stein's method of exchangeable pairs. *Combin. Probab. Comput.* **18** (2009), no. 6, 979-1017.

[59] N. Ross. Fundamentals of Stein's method. *Probab. Surv.* **8** (2011), 210-293.

[60] B.E. Sagan. The symmetric group. Representations, combinatorial algorithms, and symmetric functions. Second edition. Graduate Texts in Mathematics, 203. *Springer-Verlag, New York,* 2001.

[61] F.V. Saliola. The face semigroup algebra of a hyperplane arrangement. *Canad. J. Math.* **61** (2009), no. 4, 904-929.

[62] F.V. Saliola. Eigenvectors for a random walk on a left-regular band. *Adv. in Appl. Math.* **48** (2012), no. 2, 306-311.

[63] L. Saloff-Coste. Random walks on finite groups. *Probability on discrete structures,* 263-346, Encyclopaedia Math. Sci., 110, *Springer, Berlin,* 2004.

[64] R.P. Stanley. An introduction to hyperplane arrangements. *Geometric combinatorics,* 389-496, IAS/Park City Math. Ser., 13, *Amer. Math. Soc., Providence, RI,* 2007.

[65] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory,* pp. 583-602. *Univ. California Press, Berkeley, Calif.,* 1972.

[66] C. Stein. Approximate computation of expectations. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. *Institute of Mathematical Statistics, Hayward, CA,* 1986. iv+164 pp.

[67] B. Steinberg. Möbius functions and semigroup representation theory, *J. Combin. Theor. Ser. A.* **113** (2006), 866-881.

[68] H. Terao. Chambers of arrangements of hyperplanes and Arrow's impossibility theorem. *Adv. Math.* **214** (2007), no. 1, 366-378.

[69] H. Whitney. On the Abstract Properties of Linear Dependence. *Amer. J. Math.* **57** (1935), no. 3, 509-533.

[70] D.B. Wilson. Mixing times of lozenge tiling and card shuffling Markov chains, *Ann. Appl. Probab.* **14** (2004), no.1, 274-325.

[71] T. Zaslavsky. Facing up to arrangements: Face-count formulas for partitions of spaces by hyperplanes, PhD Thesis, Massachusetts Institute of Technology, 1974.

[72] H. Zhou. Examples of multivariate Markov chains with orthogonal polynomial eigenfunctions. PhD Thesis, Stanford University. ProQuest LLC, Ann Arbor, MI, 2008.