ABSTRACT

STANLEY, CAPRICE RAYN. Markov Chain Mixing Times. (Under the direction of Seth Sullivant).

A Markov chain is a random process that satisfies the memoryless property, that is, the conditional distribution of future states depends only on the present state and not on any events occurring prior. The mixing time of a Markov chain is the number of steps of the chain required in order for the t-step distribution to be close to its stationary distribution. Markov chains appear in many application areas including Monte Carlo simulations, sampling algorithms, and approximate counting algorithms. In this thesis, we consider two distinct problems that are connected by the common theme of Markov chain mixing times.

In Chapter 2, we seek to determine the mixing behavior for a family of random walks associated to a linear recurrence. Let $(G_i)_{i=1}^{\infty}$ be a positive integer sequence satisfying a linear recurrence $G_n = \sum_{i=1}^{d} \alpha_i G_{n-i}$, with $G_1 = 1$. For each n we have a random walk whose state space is $\mathbb{Z}_{G_n} = \{0, 1, 2, \ldots, G_n - 1\}$ and where the state $x_{t+1} \equiv x_t + z \mod G_n$ for z chosen randomly from $\{0, 1, G_2, \ldots, G_{n-1}\}$.

We show that for general linear recurrences with exponential growth, the mixing time is bounded above by $\kappa_1 n^2$ and below by $\kappa n / \log n$, where κ_1 and κ_2 are constants that depend on the sequence. We further show that in the special case of first order recurrences that the mixing time is between $\gamma_1 n$ and $\gamma_2 n \log n$, where γ_1 and γ_2 are also constants that depend on the sequence.

In Chapters 3 and 4 we consider the problem of generating uniform samples from $\mathcal{F} = \mathcal{P} \cap \mathbb{Z}^n$ where $\mathcal{P}$ is a polytope. The motivation for this sampling problem arises from independence testing in statistics. In Chapter 3, which is joint work with Tobias Windisch, the approach taken is to define a structure on $\mathcal{F}$, and using a Markov basis, define a Markov chain on $\mathcal{F}$ called the *simple fiber walk*. We prove that the simple fiber walk does not enjoy good mixing behavior. We also briefly discuss modifications to the graph structure that might improve mixing.

In Chapter 4 we consider a relaxation of the problem of sampling the lattice points $\mathcal{F}$ that follows the strategies of Morris [22] and Dyer, Kannan, and Mount [12]. There we implement a continuous sampling algorithm on a polytope $\tilde{P}$ that contains P, and then round to the nearest lattice point, repeating the process until a point in $\mathcal{F}$ is generated. For this approach, there are choices to be made about $\tilde{P}$ and the continuous sampling algorithm. We discuss those choices, prove a result to bound the rejection rate, and implement the algorithms in `R`.

Markov Chain Mixing Times

by
Caprice Rayn Stanley

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Applied Mathematics

Raleigh, North Carolina

2019

APPROVED BY:

Min Kang	Justin Post
Nathan Reading	Seth Sullivant Chair of Advisory Committee

DEDICATION

For Quincy

BIOGRAPHY

The author was born in Bethesda, Maryland to military parents Roslyn and Craig Stanley. She was raised mostly in Fayetteville, North Carolina and educated in the Cumberland County School System, graduating from Pine Forest High in 2009. She received a B.S. in Mathematics from the George Washington University in Washington, DC. Before coming to North Carolina State University in the Fall of 2014, she participated in MSRI-Up in 2013 and the EDGE Program for women in 2014. After graduation, the author looks forward to joining the staff at Johns Hopkins University Applied Physics Laboratory.

ACKNOWLEDGEMENTS

There are many people to thank for supporting me on this journey. First, I would like to thank my research advisor, Dr. Seth Sullivant. Over the past four years he has provided me with encouragement, challenged me to work harder, and has been patient with me as I adjusted to balance graduate school with motherhood. I have learned a great deal from him. I would like to thank my committee, Dr. Nathan Reading, Dr. Min Kang, and Dr. Justin Post for their constructive comments and suggestions. I would like to thank my co-author, Tobias Windisch. I learned a lot about collaborative research through working on our joint project. I would also like to thank Dr. Candice Price and Dr. Joseph Bonin for encouraging me to pursue graduate study.

Finally, I would like to thank my family and friends. Thank you to my mother, Roslyn, for her love and strength. Thank you to my father, Craig, for his guidance. Thank you to my sisters for their support and their humor. Thank you to my partner, Kenneth Burrs Jr. for his love and support. Thank you to little Quincy for being a light. Last but not least, thank you to Anila, Nadine, and Nesi, who each started out as a math friend but turned into a life long friend.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

This thesis covers two projects whose unifying theme is that of Markov chains. Put briefly, in the first project, covered in Chapter 2, we look at a certain family of Markov chains associated to an integer sequence and investigate the time required for the chain to converge to its long-term distribution. In the second project, covered in Chapters 3 and 4, we consider the problem of sampling from a discrete subset of a convex continuous set in $\mathbb{R}^n$. The approach in Chapter 3 involves defining a graph structure on the discrete set and then using that structure to construct a Markov chain with a desired long-term distribution. In Chapter 4, the approach is to use Markov chains to sample from the continuous set first, then round to the nearest element of the discrete set. In both settings we analyze the time required for convergence.

In this chapter we introduce Markov Chains as a special instance of Markov processes. What we present here is necessary to make sense of the problems under consideration and results, but is not by any stretch an exhaustive survey. Where they are either nice or short, we have included the proofs of fundamental results. And we refer the reader to Chapters 1,4, and 12 in [20] and to Chapters 1 in [23] for a more involved proofs and a complete treatment of Markov chains.

1.1 Markov Chains

A Markov process is a model for a random process through time. The time at which observations of the process are made can occur at either discrete or continuous intervals. The state space Ω of points x represents the possible observations. The possible events for which a probability is well-defined are elements of a Borel-algebra $\mathcal{A}$ of subsets of Ω. Acting as the generating mechanism of a Markov process is its transition probability function or kernel $\kappa_t(x, A)$ which can either change or remain stable with time. What distinguishes a Markov process from other random processes is the property of being memoryless in the sense that the future distribution of the process, given the present and past states, only depends on the present state, and not the realized path of states taken to arrive at the present state. This distinguishing property is called the Markov property. When a Markov process occurs in discrete time, we refer to it as a Markov chain. In this thesis, the Markov chains that we encounter will be time-homogeneous, in that the kernel is stable, and have either continuous or discrete state spaces.

Definition 1.1.1. A *time-homogeneous Markov chain with continuous state space* Ω is a sequence $(X_t)_{t=0,1,2,...}$ of random variables taking on values in Ω. The probability of transitioning from $x \in \Omega$ to a Borel-measurable set $A \subset \Omega$ is given by the kernel $\kappa(x, A)$ and is independent of time. Additionally for all $x_0, x_1, \dots, x_t \in \Omega$ and measurable $A \subset \Omega$, Equation 1.1 is satisfied.

$$\Pr\left(X_{t+1} \in A \mid X_t = x_t, X_{t-1} = x_{t-1}, \dots, X_0 = x_0\right) = \Pr\left(X_{t+1} \in A \mid X_t = x_t\right) = \kappa(x, A) \quad (1.1)$$

When Ω is countable or finite the definition is analogous. The Markov property is stated in terms of one-step transition probabilities between pairs of states: For all $x, y, x_0, \dots, x_{t-1} \in \Omega$ and for all $t \geq 0$, the following equation holds:

$$\Pr\left(X_{t+1} = y \mid X_t = x, X_{t-1} = x_{t-1}, \dots, X_0 = x_0\right) = \Pr\left(X_{t+1} = y \mid X_t = x\right).$$

Further if Ω is finite, then the one-step transition probabilities $\Pr\left(X_{t+1} = y \mid X_t = x\right)$ are stored in an $|\Omega| \times |\Omega|$ transition matrix P, that is $P(x, y) = \Pr\left(X_{t+1} = y \mid X_t = x\right)$. With this construction, we can easily derive general t-step transition probabilities. The next result demonstrates that for a transition matrix P, the (x, y)-entry of the matrix P^t is the probability that the Markov chain transitions from state x to state y in exactly t steps.

Theorem 1.1.2 (Chapman-Kolmogorov)**.** *Let P be the transition matrix of a time-homogenous Markov chain with finite Ω. Then $P^t(x, y) = \Pr(X_t = y \mid X_0 = x)$.*

Proof. Proceeding by induction, notice that the $t = 1$ case holds by construction. Now suppose $P^k(x, y) = \Pr(X_k = y \mid X_0 = x)$ holds for all $0 \leq k < t$. Observe the following equalities:

$$\begin{aligned}
\Pr\left(X_t = y \mid X_0 = x\right) &= \sum_{z \in \Omega} \Pr\left(X_{t-1} = z \mid X_0 = x\right) \Pr(X_t = y \mid X_{t-1} = z) \\
&= \sum_{z \in \Omega} P^{t-1}(x, z) P(z, y) \\
&= P^t(x, y).
\end{aligned}$$

□

For the rest of this section, unless otherwise noted, we will assume the Markov chains we consider are time-homogeneous with finite state space. In these cases we often identify the Markov chain by its transition matrix alone, since all of the important information about

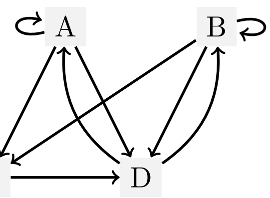

Figure 1.1: The underlying directed graph of the Markov chain in Example 1.1.3

the dynamics is stored there. At times we may find it advantageous to visually represent the Markov chain. For this we look at its underlying directed graph. We will see more about graphs in Section 1.1.1. The visual representation of a Markov chain consists of a collection of vertices each representing a state in Ω and an arrow from state x to state y when the probability $P(x, y)$ is positive.

Example 1.1.3. Consider the Markov chain with state space $\Omega = \{A, B, C, D\}$ and transition matrix given below:

$$P = \begin{array}{c|cccc} & A & B & C & D \\ \hline A & \frac{1}{6} & 0 & \frac{1}{2} & \frac{1}{3} \\ B & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ C & 0 & 0 & 0 & 1 \\ D & \frac{1}{2} & \frac{1}{2} & 0 & 0 \end{array}$$

Suppose at step t in the Markov chain that the current state is $X_t = A$. Then in the next step the possible states are A, C, and D taken with the probabilities $\frac{1}{6}, \frac{1}{2}$, and $\frac{1}{3}$ respectively. Notice that the sum of each row in P is one. In general such matrices, namely nonnegative, square, with rows summing to one, are called *stochastic matrices.* When the sum of each column is also one, then the matrix is called *doubly stochastic.*

Another notable feature of this example is that the Markov chain is in a sense "connected". Notice that from the state A the probability of transitioning to state B in one time step is 0, however, the probability of transitioning to state B in exactly two time steps is positive, in fact the probability is $\frac{1}{6}$. It is more easily seen from the underlying directed graph in Figure 1.1 that between any two states there is a directed path of edges of positive probability. This connected property is a desirable property for Markov chains that will be enjoyed by the chains we encounter. We formalize the notion with a definition.

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

In addition to being irreducible, the Markov chains we consider will also have the property of being aperiodic.

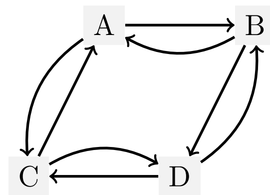

Figure 1.2: The underlying directed graph of the Markov chain in Example 1.1.8

Definition 1.1.5. Let $\tau(x) := \{t \geq 1 : P^t(x,x) > 0\}$ be the set of times when it is possible for the chain to return to starting position x. The *period of state* x is $\gcd \tau(x)$.

Lemma 1.1.6. *If P is irreducible, then* $\gcd \tau(x) = \gcd \tau(y)$ *for all* $x, y \in \Omega$.

Definition 1.1.7. For an irreducible chain, the *period of the chain* is the period which is common to all states. The chain is *aperiodic* if all states have period 1. If a chain is not aperiodic, then it is *periodic.*

The Markov chain in Example 1.1.3 is irreducible and aperiodic.

Example 1.1.8. Consider the Markov chain on $\{ A, B, C, D \}$ whose transition matrix is

$$P = \begin{array}{c|cccc} & A & B & C & D \\ \hline A & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ B & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ C & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ D & \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{array} .$$

The period of the state A is two, since from state A, the chain can return in two or some multiple of 2 steps. We can demonstrate that this Markov chain is irreducible by looking at the underlying graph. This Markov chain experiences periodic behavior. The period of each state is two.

Next we define probability distributions, which we use to describe the distribution of the random variable X_t in a Markov chain. Put another way, probability distributions will be used to describe the relative likelihood of observing a particular state of the chain at the t-th step.

Definition 1.1.9. A *probability distribution* μ on Ω is a function $\mu : \Omega \to [0, 1]$ such that

$$\sum_{x \in \Omega} \mu(x) = 1.$$

Example 1.1.10. The vector $\mu = \frac{1}{n}\mathbf{1}_{1\times n}$ represents the uniform probability distribution on Ω whenever $|\Omega| = n$.

As a side note, when the state space $\Omega \subset \mathbb{R}$ is a continuous set, the object analogous to a distribution is a probability density function f defined on $\mathbb{R}$. The function $f : \mathbb{R} \to (0, \infty)$ is a probability density function if $\int_{\mathbb{R}} f dx = 1$.

Definition 1.1.11. Let P be a transition matrix and μ_0 be a probability distribution on Ω. The *t-step distribution* μ_t of a Markov chain with initial distribution μ_0, is the distribution of the random variable X_t, given the distribution of X_0 is μ_0. The distribution μ_t is computed by the product, $\mu_t = \mu_0 P^t$.

Definition 1.1.12. A distribution π on Ω is *stationary* if $\pi = \pi P$.

The next couple of results are well known and detail conditions on the spectrum of P. We will see in Section 1.2 that these results play an important role in bounding the distance between the t-step and stationary π distributions. Proofs are included when nice.

Lemma 1.1.13. *If λ is an eigenvalue for a transition matrix P, then $|\lambda| \leq 1$.*

The strategy of the proof is from [20].

Proof. First we show that for a function $f : \Omega \to \mathbb{R}$, the infinity norm $\|f\|_\infty := \max_{x\in\Omega} |f(x)|$ satisfies

$$\|Pf\|_\infty \leq \|f\|_\infty .$$

Observe the sequence of inequalities

$$\begin{aligned}
\|Pf\|_\infty &= \max_{x\in\Omega} |Pf(x)| \\
&= \max_{x\in\Omega} |\sum_{y\in\Omega} P(x,y)f(y)| \\
&\leq \max_{x\in\Omega} \sum_{y\in\Omega} P(x,y)|f(y)|
\end{aligned}$$

where the last inequality follows since $P(x, y) \geq 0$ for all $x, y \in \Omega$. Now suppose $y_* \in \Omega$ such that $\|f\|_\infty = |f(y_*)|$. Then

$$\begin{aligned}
\|Pf\|_\infty &\leq \max_{x\in\Omega} \sum_{y\in\Omega} P(x,y)|f(y_*)| \\
&\leq \|f\|_\infty
\end{aligned}$$

since $\sum_{x\in\Omega} P(x, y) = 1$. Now suppose (u, λ) is an eigenpair for P. It follows that $\|\lambda u\|_\infty = \|Pu\|_\infty \leq \|u\|_\infty$. Hence $|\lambda| \|u\|_\infty \leq \|u\|_\infty$ which implies that $|\lambda| \leq 1$ as desired. $\square$

Lemma 1.1.14. *If P is a transition matrix, then 1 is an eigenvalue with right eigenvector $\mathbf{1} = (1, 1, \ldots, 1)^T$.*

Continuing with Example 1.1.3, we can check that the distribution

$$\pi = \left(\frac{6}{29}, \frac{15}{58}, \frac{11}{58}, \frac{10}{29} \right)$$

is stationary and is, in fact, the only stationary distribution. With little effort we can further show that the eigenvalues of P are $\lambda = 1, -0.3814211 \pm 0.5265428i$, and 0.2628422 with magnitudes $|\lambda| = 1, 0.403055, 0.403055$, and 0.2628422, respectively.

The following result guarantees nice properties for the Markov chain considered in this thesis.

Theorem 1.1.15. *Let P be the transition matrix for an aperiodic, irreducible Markov chain.*

1. *Then there exists a unique probability distribution π on Ω such that $\pi = \pi P$ and $\pi(x) > 0$ for all $x \in \Omega$, this is the left eigenvalue $\pi P = \pi$ with eigenvalue 1.*
2. *The value 1 is an eigenvalue and the corresponding eigenspace is 1 dimensional.*
3. *There are no other eigenvalues λ whose magnitude $|\lambda| = 1$.*

Theorem 1.1.15 follows from Perron-Frobenius theorem for an $|\Omega| \times |\Omega|$ nonnegative, aperiodic, irreducible matrix with spectral radius $\rho = 1$.

1.1.1 Random Walk on a Graph

In this section we present a classic type of Markov chain called a random walk on a graph. Markov chains of this type arise in many different contexts. One interesting example that we will see at the end of the section models card shuffles.

An undirected finite *graph* $G = (V, E)$ is a collection of vertices $V = \{x_1, x_2, \ldots, x_n\}$ along with a finite collection of edges $x_i x_j \in E$ joining the vertices in some configuration. A graph is often represented either visually or by its associated adjacency matrix A_G, which captures the graph's structure. The rows and columns of A_G are indexed by the vertices of G and the (i, j)-entry of A_G is the number of edges whose endpoints are exactly x_i and x_j. By construction, A_G is a symmetric matrix. The *degree* of a vertex $x_i \in V$, denoted $\deg(x_i)$, is the total number of edges incident to x_i. In terms of the adjacency matrix $\deg(x_i) = \sum_{j=1}^{n} A_G(i, j)$.

A *walk* in the graph G is an alternating sequence of vertices and edges that starts and ends at a vertex and where each edge in the sequence is preceded and succeeded by its two endpoints. A *path* is a walk with no repeated vertices or edges. We say that G is *connected* if there exists a path from x_i to x_j for any pair of vertices. The *distance* $d(x_i, x_j)$ between

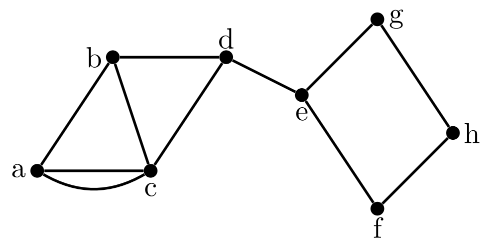

Figure 1.3: A visual representation of the graph from Example 1.1.16

vertices x_i and x_j is the number of edges in a path of shortest length that starts at x_i and ends at x_j. If no such path exists, then by convention we set $d(x_i, x_j) = \infty$. The *diameter* of G is $\text{diam}(G) = \max_{x_i, x_j \in V} d(x_i, x_j)$, that is, the maximum distance over all pairs of vertices in V.

Example 1.1.16. Suppose G is the graph displayed in Figure 1.3. Then G has as its vertex set $V = \{a, b, c, d, e, f, g, h\}$, has 11 edges, and is connected. Notice that if the edge de is removed, then the resulting graph is disconnected. The sequence (a, ac, c, cd, d, de, e) is a path from a to e. The adjacency matrix of G is

$$A_G = \begin{bmatrix} 0 & 1 & 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}.$$

Suppose G is a connected graph with vertex set $V = \{x_1, \ldots, x_n\}$. We can define a Markov chain with state space V, where from the current state $x_i \in V$, the next state is generated by choosing an edge incident to x_i uniformly at randomly then traversing that edge. The one-step transition probabilities are given by $P(x_i, x_j) = \frac{A_G(i,j)}{\deg(x_i)}$. It follows from the connectivity of the graph that the Markov chain is irreducible. The equations

$$\begin{aligned}(\pi P)(x_i) &= \sum_{j=1}^{n} \pi(x_j) P(j,i) \\ &= \sum_{j=1}^{n} \frac{\deg(x_j)}{2|E|} P(j,i) \\ &= \frac{1}{2|E|} \sum_{j=1}^{n} \deg(x_j) \frac{A_G(j,i)}{\deg(x_j)} \\ &= \frac{\deg(x_i)}{2|E|}\end{aligned} \tag{1.2}$$

demonstrate that the distribution $\pi(x_i) = \frac{\deg(x_i)}{2|E|}$ is stationary.

Example 1.1.17 (Card Shuffles)**.** A sequence of card shuffles can by modeled as a random walk on a graph. Let $\sigma = (c_1, c_2, \ldots, c_N)$ represent a deck of N cards. A simple way to generate from σ a random permutation σ' of the deck is to choose a pair of indices $1 \leq i, j \leq N$ at random then transpose cards c_i and c_j. By repeating this process many times, the deck of cards will be slowly shuffled and the initial ordering of the deck forgotten. Let G_N be the graph whose vertex set consists of the $N!$ permutations of the deck, and where σ_i and σ_j are adjacent if there exists a transposition that takes σ_i to σ_j. Then this method of shuffling cards corresponds to the random walk on G_N. This random walk is irreducible, aperiodic, and the stationary distribution is uniform over all permutations. It is shown in Chapter 8 Section 2 of [20], using techniques that involve strong stationary times, that the number of transpositions required for the ordering of the deck to be near uniformly distributed over all possible permutations, in other words mixing time of the random walk, is at most on the order of $N \log N$.

A more natural way to shuffle cards is via riffle shuffles. To do a riffle shuffle, we choose a location $1 \leq i \leq N$, at random and split the deck between the i-th and $i+1$-st cards, resulting in two smaller decks. Then, alternating between the decks, we drop a number of cards from the bottom of each into one pile and repeat. This method of generating card shuffles can also be characterized as a random walk on graph. However, in this case, edges exist between permutations σ_i and σ_j if either, σ_i can be obtained by applying a riffle shuffle to σ_j, or vice-versa, and modifications are made to include appropriate edge weights as, from a particular permutation σ_i of the deck, the set of adjacent permutations are not equally likely.

Anyone who has played with a deck of cards as a leisurely past-time or a serious professional, has been confronted with the question: "How many riffle-shuffles are sufficient to shuffle the deck?" It was shown in [3] that for a standard deck of 52 cards, the answer is more or less 7 and after that, more riffle shuffles does not increase the randomness.

Section 1.1.2 demonstrates another important application of Markov chains.

1.1.2 Markov Chain Monte Carlo

Markov Chain Monte Carlo refers to a class of algorithms for sampling from a target probability distribution. For more details see [20] and [24]. Suppose we have a countable set Ω of states, some irreducible Markov chain with transition matrix P, and a target distribution π on Ω from which we would like to sample. A new Markov chain, often called the Metropolis chain, whose long-term distribution is the target π, can be constructed. The idea is that from some current state $X_t = x$, we choose a state y according to the distribution $P(x, \cdot)$. Instead of moving immediately, we accept y with a certain probability that depends only on the pair of states x and y, and reject otherwise. The transition matrix Q for the Metropolis chain is given by

$$Q(x,y) = \begin{cases} P(x,y) \min\left\{ \frac{\pi(y)P(y,x)}{\pi(x)P(x,y)}, 1 \right\} & \text{if } y \neq x \\ 1 - \sum_{z \neq x} P(x,z) \min\left\{ \frac{\pi(z)P(z,x)}{\pi(x)P(x,z)}, 1 \right\} & \text{otherwise.} \end{cases} \tag{1.3}$$

Markov Chain Monte Carlo is a powerful tool that can be used in statistical settings and for numerical approximations. For example, we find Markov chain Monte Carlo appearing in hill climb algorithms for optimizing functions. Suppose f is a real-valued function defined on a finite state space Ω. Letting $\lambda > 1$ be some fixed parameter, we can specify a target distribution $\pi(x) = \frac{\lambda^{f(x)}}{\sum_{y \in \Omega} \lambda^{f(y)}}$ whose mass is centered on the maximizers of f. Replacing the acceptance probability $\frac{\pi(y)P(y,x)}{\pi(x)P(x,y)}$ with $\frac{\lambda^{f(y)}P(y,x)}{\lambda^{f(x)}P(x,y)}$ in Equation 1.3, the Markov chain defined by Q can be implemented to search for states that optimize f. Similar to the question posed in Example 1.1.17, a very important question to ask is: "How many steps of the Metropolis chain are needed before the distribution is near π?" The question is developed more in the next section.

1.2 Mixing Times of Markov Chains

In this section, we discuss the long-term behavior of irreducible, aperiodic Markov chains. Recall that by Theorem 1.1.15, for such Markov chains there exist a unique stationary distribution π. The distribution π is stable in the sense that if the chain moves forward one step starting from a state chosen randomly from π, then the distribution of the new state is again π. In this section, we will see that regardless of the initial state, as the chain progresses the t-step distribution converges to π. From there, we formalize the notion of *mixing time*, the theme of this thesis, which is concerned with the rate at which the t-step distribution converges to the stationary distribution. The importance of mixing time is appreciated when we need to sample from or approximate target distributions as suggested in Section 1.1.2 on Markov chain Monte

Carlo or Example 1.1.17 on card shuffles. To see this in action, we revisit the simpler situation from Example 1.1.3.

Example 1.2.1 (Continuing 1.1.3)**.** Suppose we need to generate a random variable from the set $\{A, B, C, D\}$ according to the distribution $\pi = \left(\frac{6}{29}, \frac{15}{58}, \frac{11}{58}, \frac{10}{29}\right)$. One way to proceed is to run the Markov chain from Example 1.1.3 from some arbitrary starting state X_0 for some prescribed number of steps τ. Then take the state X_τ returned after τ steps to be the random variable. To ensure that X_τ has the desired distribution we need to determine a reasonable choice for τ. Recall that given an initial distribution μ_0 the t- step distribution is $\mu_0 P^t$. If we let the initial state $X_0 = A$ then $\mu_0 = (1, 0, 0, 0)$ and the sequence of distributions after the first five time steps are given, along with the $\|\cdot\|_2$-norm distance to π, in the table below.

t	$\mu_0 P^t$	$\left\|\mu_0 P^t - \pi\right\|_2$
1	$(0.16667, 0, 0.5, 0.33333)$	0.40614
2	$(0.19444, 0.16667, 0.08333, 0.55556)$	0.25362
3	$(0.31019, 0.33333, 0.15278, 0.20370)$	0.19372
4	$(0.15355, 0.21296, 0.26620, 0.36728)$	0.10628
5	$(0.20923, 0.25463, 0.14776, 0.38837)$	0.06063
$\vdots$	$\vdots$	$\vdots$

From the table, we see that the t- step distribution gets closer and closer to π with each time step. This is not surprising since the Markov chain is irreducible and aperiodic. Depending on the level of tolerance acceptable for the application, we may take X_5 as the generated random variable since its distance to π is less than 0.1.

In Example 1.1.3 where the problem is small and discrete, there are more attractive ways to generate a random variable with the desired distribution. But the strategy described can be applied in any context where one wishes to sample according to a target distribution. In Chapter 4, we see that it becomes particularly useful when we need to generate random vectors supported on convex, continuous sets in higher dimensions with intricate geometries. For the purposes of practical implementation, we will need an understanding of the chain's mixing time. The main goal of this thesis is to analyze the mixing time of certain Markov chains and also to understand how mixing times play a major role in the efficiency of sampling algorithms. In this section, we develop the notion of mixing time formally and present some common tools that are often used to analyze mixing time.

We arbitrarily chose to use the $\|\cdot\|_2$-norm in Example 1.2.1 to compare the t-step and the stationary distributions. However, there is a particular metric that is more commonly used to measure the distance between two probability distributions.

Definition 1.2.2. The *total variation distance* between probability distributions μ and η defined on Ω is

$$\|\mu - \eta\|_{TV} = \max_{A \subset \Omega} |\mu(A) - \eta(A)|. \tag{1.4}$$

In other words, the total variation distance is the maximum difference in probability that μ and η assign to a fixed event A. It follows that $\|\mu - \eta\|_{TV}$ is always at most one. There is an equivalent formulation of Equation 1.4 that expresses the total variation distance as a scaled $\|\cdot\|_1$-distance. This alternate formulation is often easier to use.

Proposition 1.2.3. *The total variation distance between probability distributions μ and η defined on Ω is*

$$\|\mu - \eta\|_{TV} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \eta(x)|.$$

Proof. First we show that $\frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \eta(x)| \leq \max_{A \subset \Omega} |\mu(A) - \eta(A)|$. Observe the following inequalities:

$$\begin{aligned}
\sum_{x \in \Omega} |\mu(x) - \eta(x)| &= \left| \sum_{x : \mu(x) \geq \eta(x)} \mu(x) - \eta(x) \right| + \left| \sum_{x : \mu(x) < \eta(x)} \mu(x) - \eta(x) \right| \\
&\leq 2 \cdot \max \left\{ \left| \sum_{x : \mu(x) \geq \eta(x)} \mu(x) - \eta(x) \right|, \left| \sum_{x : \mu(x) < \eta(x)} \mu(x) - \eta(x) \right| \right\} \\
&\leq 2 \cdot \max_{A \subset \Omega} |\mu(A) - \eta(A)|.
\end{aligned}$$

To finish the proof we show that the inequality also goes the other way. For any set $A \subset \Omega$,

$$|\mu(A) - \eta(A)| \leq \left| \sum_{x \in \Omega : \mu(x) \geq \eta(x)} \mu(x) - \eta(x) \right| = \sum_{x \in \Omega : \mu(x) \geq \eta(x)} |\mu(x) - \eta(x)|$$

and

$$|\eta(A) - \mu(A)| \leq \left| \sum_{x \in \Omega : \mu(x) < \eta(x)} \eta(x) - \mu(x) \right| = \sum_{x \in \Omega : \mu(x) \geq \eta(x)} |\eta(x) - \mu(x)|.$$

By adding both sides, we conclude $2|\mu(A) - \eta(A)| \leq \sum_{x \in \Omega} |\mu(x) - \eta(x)|$. The result follows. $\square$

Instead of taking the supremum of the value $|\mu(A) - \nu(A)|$ over all subsets $A \subset \Omega$ in Definition 1.4, when Ω is a continuous subset of $\mathbb{R}$ we let A range over measurable sets and μ and ν are replaced with probability measures.

Before we are able to determine how quickly the distribution of a Markov chain approaches its stationary distribution we must first guarantee that the distribution actually converges.

Theorem 1.2.4 (Convergence Theorem)**.** *Suppose that P is irreducible and aperiodic with stationary distribution π. Then there exists constants $\alpha \in (0,1)$ and $C > 0$ such that*

$$\max_{x\in\Omega} ||P^t(x,\cdot) - \pi||_{TV} \leq C\alpha^t.$$

Theorem 1.2.4 states that regardless of the starting state, the total variation distance between the t-step distribution and the stationary distribution converges to zero. It also states that as a function of t, the rate that the total variation distance decreases is bounded by an exponential function. The mixing time of an irreducible and aperiodic Markov chain is the first time t that the t-step distribution $P^t(x,\cdot)$ is "ϵ-close"' to the stationary distribution π.

Definition 1.2.5. For an irreducible and aperiodic Markov chain with transition matrix P and some small parameter $0 < \epsilon < 1$, the *mixing time* is

$$t_{mix}(\epsilon) := \min_{t\in\mathbb{Z}} \left\{ \max_{x\in\Omega} \left\| P^t(x,\cdot) - \pi \right\|_{TV} \leq \epsilon \right\},$$

where π is the stationary distribution.

The parameter ϵ in Definition 1.2.5 is a user-defined tolerance, which quantifies how close to the stationary distribution is "close" enough. It is common practice to let $\epsilon = \frac{1}{4}$ and abbreviate the mixing time as $t_{mix} = t_{mix}(\frac{1}{4})$.

Also notice that Definition 1.2.5 is used to describe the convergence of a single Markov chain. There are contexts in which we have a collection of related Markov chains and would like to describe how the mixing time behaves with respect to the size of the chains. We can imagine this being relevant if, for example, we decide to implement a Markov chain Monte Carlo algorithm to generate a sequence of random vectors $X \in \mathbb{R}^n$ with increasing parameter n. In a business setting, for this task it would be greatly beneficial to have an understanding of how mixing time of the chain grows with n as time and storage costs are relevant considerations.

The next definition says that a family of Markov chains has fast mixing if the mixing time grows at most polynomially with respect to the size of the state space.

Definition 1.2.6. Suppose there is a family of Markov chains indexed by $\mathcal{I}$ with transition matrices $(P_i)_{i\in\mathcal{I}}$ and state spaces $(\Omega_i)_{i\in\mathcal{I}}$. Letting τ_i be the mixing time for P_i, we say that the family $(P_i)_{i\in\mathcal{I}}$ is *rapidly mixing* if there exists a polynomial $p \in \mathbb{Q}_{\geq 0}[t]$ such that $\tau_i \leq p(\log|\Omega_i|)$.

From here we pivot to discuss common techniques and tools that have been developed in order to look at mixing times. Perhaps the most important tool that we will see and actually use relies on the knowledge of the spectrum of the transition matrix P.

In Section 1.1 we saw that 1 is an eigenvalue of any transition matrix P and all the eigenvalues λ are bounded in magnitude by one. By Theorem 1.1.15, when P is irreducible and aperiodic then the multiplicity of the eigenvalue 1 is one, and the magnitude of all non-trivial eigenvalues is strictly less than one. The mixing time of an irreducible aperiodic Markov chain is determined by the non-trivial eigenvalues.

Definition 1.2.7. For a transition matrix P, the *second largest eigenvalue modulus (SLEM)* is $\lambda_* = \max\{|\lambda| : \lambda \neq 1, \lambda \text{ is an eigenvalue of P }\}$.

With the next definitions and results, we build up to a decomposition of certain transition matrices in terms of their eigenvalues and eigenfunctions $f \in \mathbb{R}^{|\Omega|}$. We refer the reader to Chapter 12 in [20] for more details.

Definition 1.2.8. A Markov chain with transition matrix P is *reversible* with respect to the distribution π if for all $x, y \in \Omega$,

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

When P is reversible with respect to π then if an initial state $X_0 = x_0$ is chosen according to π then the probability of any realization of the chain is equal to the probability of its time-reversal. That is, we can show inductively that

$$\pi(x_0)P(x_0,x_1)\dots P(x_{n-1},x_n) = \pi(x_n)P(x_n,x_{n-1})\dots P(x_1,x_0).$$

Proposition 1.2.9. *Let P be reversible with respect to π. Then $\pi(x_0)P(x_0,x_1)\dots P(x_{n-1},x_n) = \pi(x_n)P(x_n,x_{n-1})\dots P(x_1,x_0)$.*

Proof. By definition of reversible, $\pi(x_0)P(x_0,x_1) = \pi(x_1)P(x_1,x_0)$. Now suppose that

$$\pi(x_0)P(x_0,x_1)\dots P(x_{n-1},x_n) = \pi(x_n)P(x_n,x_{n-1})\dots P(x_1,x_0).$$

Then

$$\begin{aligned}\pi(x_0)P(x_0,x_1)\dots P(x_{n-1},x_n)P(x_n,x_{n+1}) &= P(x_n,x_{n+1})\pi(x_n)P(x_n,x_{n-1})\dots P(x_1,x_0)\\ &= \pi(x_{n+1})P(x_{n+1},x_n)P(x_n,x_{n-1})\dots P(x_1,x_0).\end{aligned}$$ $\square$

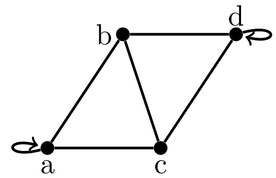

Figure 1.4: A graph on four nodes for Example 1.2.10.

Also when P is reversible with respect to π, by the equations

$$\begin{aligned}(\pi P)(x) &= \sum_{y\in\Omega} \pi(y)P(y,x) \\ &= \sum_{y\in\Omega} \pi(x)P(x,y) \\ &= \pi(x)\end{aligned} \tag{1.5}$$

it follows that π is stationary. Thus if P is additionally aperiodic and irreducible, then π is the unique stationary distribution. The converse is not true. Example 1.1.3 is aperiodic and irreducible with unique stationary distribution $\pi = (\frac{6}{29}, \frac{15}{58}, \frac{11}{58}, \frac{10}{29})$. Substituting $x = A$, and $y = B$ in the equation $\pi(x)P(x,y) = \pi(y)P(y,x)$ we see that P is not reversible with respect to the stationary distribution.

Example 1.2.10. The following matrix P is the transition matrix for the random walk on the graph displayed in Figure 1.4. Let $\pi = (0.25, 0.25, 0.25, 0.25)$ be the uniform distribution. Since P is symmetric it is clear that P is reversible with respect to π.

$$P = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

In this next result we consider the inner product space $(\mathbb{R}^\Omega, \langle\cdot,\cdot\rangle_\pi)$ with the following inner product

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

Lemma 1.2.11. *Let P be reversible with respect to π. Then*

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

2. *The matrix P can be decomposed as*

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

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

$$\frac{P^t(x,y)}{\pi(y)} = 1 + \sum_{j=2}^{\Omega} f_j(x) f_j(y) \lambda_j^t.$$

Definition 1.2.12. A Markov chain with transition matrix P is *transitive* if for any pair $x, y \in \Omega$ there exists a permutation $\sigma_{x,y} : \Omega \to \Omega$ that maps x to y and preserves the one step transition probabilities. In other words, $\sigma_{x,y}(x) = y$ and for all $u, v \in \Omega$,

$$P(u,v) = P(\sigma_{x,y}(u), \sigma_{x,y}(v)).$$

We can think of a transitive Markov chain as one whose underlying directed graph is regular and, if labels on the vertices are ignored, the same configuration occurs at each vertex. Notice that the periodic Markov chain from Example 1.1.8 is transitive. For transitive chains, since the dynamics are the same at any state, the stationary distribution is uniform over the states.

Lemma 1.2.13. *Let P be a reversible transition matrix, with eigenvalues*

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{|\Omega|} \geq -1$$

and associated eigenfunctions $\{f_j\}$, orthonormal with respect to $\langle \cdot, \cdot \rangle_\pi$. Then

$$4 \left\| P^t(x,\cdot) - \pi \right\|_{TV}^2 \leq \left\| \frac{P^t(x,\cdot)}{\pi(\cdot)} - 1 \right\|_2^2 = \sum_{j=2}^{|\Omega|} f_j(x)^2 \lambda_j^{2t}.$$

If the chain is transitive, then

$$4 \left\| P^t(x,\cdot) - \pi \right\|_{TV}^2 \leq \left\| \frac{P^t(x,\cdot)}{\pi(\cdot)} - 1 \right\|_2^2 = \sum_{j=2}^{|\Omega|} \lambda_j^{2t}.$$

Lemmas 1.2.11 and 1.2.13 will be used in Chapter 2 when we look at a certain random walk on a finite abelian group. The Markov chain itself will not be reversible necessarily but we will still have an orthonormal basis of eigenfunctions. The main take away from Lemmas 1.2.11 and 1.2.13 is an upper bound on the total variation distance to stationarity in terms of the non-trivial eigenvalues of the transition matrix.

For the Markov chains seen in Chapter 2 we would like to sandwich the mixing times with both an upper and lower bound. Some tools that we will use to determine lower bounds involve some knowledge of the spectral gap and relaxation time.

Definition 1.2.14. For a Markov chain with transition matrix P the *absolute spectral gap* denoted γ_* is the difference between 1 and the SLEM. That is, $\gamma_* = 1 - \lambda_*$.

Recall by Theorem 1.1.15, when P is the transition matrix for an aperiodic and irreducible Markov chain then $\lambda \neq -1$ is not an eigenvalue. In which case the absolute spectral gap is positive.

Definition 1.2.15. The *relaxation time* t_{rel} of a reversible Markov chain with absolute spectral gap γ_* is $t_{rel} = (\gamma_*)^{-1}$.

The relaxation time is inversely proportional to the distance between the SLEM and 1, so when the SLEM is small, for instance, t_{rel} is large.

Theorem 1.2.16. *Suppose that $\lambda \neq 1$ is an eigenvalue for the transition matrix P of an irreducible and aperiodic Markov chain. Then*

$$t_{mix}(\epsilon) \geq \left(\frac{1}{1-|\lambda|} - 1\right) \log\left(\frac{1}{2\epsilon}\right).$$

If P is also reversible then

$$t_{mix}(\epsilon) \geq (t_{rel} - 1) \log\left(\frac{1}{2\epsilon}\right).$$

Proof. (Follows [20]) Since P is irreducible and aperiodic we can let f be an eigenfunction of P with eigenvalue $\lambda \neq 1$ such that $Pf = \lambda f$. Since the eigenfunctions are orthonormal with respect to $\langle \cdot, \cdot \rangle_\pi$ and $\mathbf{1}$ is an eigenfunction, then

$$\langle f, \mathbf{1} \rangle_\pi = \sum_{x \in \Omega} \pi(x) f(x) = 0.$$

Observe the following inequalities

$$\begin{aligned}
|\lambda^t f(x)| &= |P^t f(x)| \\
&= \left| \sum_{y\in\Omega} P^t(x,y) f(y) - \sum_{y\in\Omega} \pi(y) f(y) \right| \\
&= \left| \sum_{y\in\Omega} [P^t(x,y) - \pi(y)] f(y) \right|
\end{aligned}$$

$$\leq \sum_{y\in\Omega} \left|P^t(x,y) - \pi(y)\right| |f(y)|$$
$$\leq \sum_{y\in\Omega} \left|P^t(x,y) - \pi(y)\right| \|f\|_\infty$$
$$\leq 2d(t)\,\|f\|_\infty\,,$$

where $d(t) = \max_{x\in\Omega}\left\{ \frac{1}{2}\sum_{y\in\Omega}\left|P^t(x,y) - \pi(y)\right| \right\}$. Choosing $x \in \Omega$ such that $|f(x)| = \|f\|_\infty$ it follows that $|\lambda|^t \leq 2d(t)$. By substituting $t = t_{mix}(\epsilon)$ we obtain $|\lambda|^{t_{mix}(\epsilon)} \leq 2d(t_{mix}(\epsilon)) = 2\epsilon$. From which it follows that

$$t_{mix}(\epsilon)\left(\frac{1}{|\lambda|} - 1\right) \geq t_{mix}(\epsilon)\log\left(\frac{1}{|\lambda|}\right) \geq \log\left(\frac{1}{2\epsilon}\right).$$

Then by dividing through by the quantity $\left(\frac{1}{|\lambda|} - 1\right)$ we arrive at the lower bound

$$t_{mix}(\epsilon) \geq \left(\frac{1}{1-|\lambda|} - 1\right)\log\left(\frac{1}{2\epsilon}\right).$$

Finally, if P is irreducible then we can choose λ such that $|\lambda| = \lambda_*$, to obtain

$$t_{mix}(\epsilon) \geq (t_{rel} - 1)\log\left(\frac{1}{2\epsilon}\right). \square$$

The connection between mixing time and SLEM can be used to characterize the rapid mixing property of a family of Markov chains as a statement about the growth of the corresponding SLEM. By Theorem 1.2.4 and Lemma 1.2.13, we have that for Markov chains that are irreducible, aperiodic, and reversible, the total variation distance to stationarity can be bounded

$$\max_{x\in\Omega}\left\|P^t(x,\cdot) - \pi\right\|_{TV} \leq C\lambda_*^t,$$

where $C > 0$ is some constant. If we force the right-hand side to be bounded above by ϵ and rearrange, then we see that

$$t_{mix}(\epsilon) \leq \log\left(\frac{C}{\epsilon}\right)\frac{1}{\log(\frac{1}{\lambda_*})}.$$

If we let $C'\frac{1}{\log(\frac{1}{\lambda_*})}$ serve as a proxy for the mixing time, and use the fact that $\log(\frac{1}{x}) \sim 1 - x$ when $x \in [0,1]$, then we get the following characterization of rapid mixing:

Definition 1.2.17. Let the sequence of transition matrices $(P_i)_{i\in\mathcal{I}}$ represent a family of irreducible, aperiodic, and reversible Markov chains and suppose $(\lambda_*^i)_{i\in\mathcal{I}}$ represents the corresponding SLEMs. Then the family $(P_i)_{i\in\mathcal{I}}$ of Markov chains is *rapidly mixing* if there exists a polynomial $p \in \mathbb{Q}_{\geq 0}[t]$ such that $\lambda_*^i \leq 1 - \frac{1}{p(\log|\Omega_i|)}$.

We conclude this section by introducing the final tool that we will use to analyze mixing time of Markov chains. Suppose that there is a set $A \subset \Omega$ of the states where the probability of transitioning from A to $\Omega \backslash A$ is low. Then if the Markov chain lands in A we would expect for the chain to bounce around within A for a while before escaping. In which case we refer to A as a "bottleneck" and the effect is that the convergence to the stationary distribution is slowed. On the other hand, if for any subset $A \subset \Omega$ of states, there is a high probability of transitioning to $\Omega \backslash A$, then we would expect good mixing properties.

Definition 1.2.18. Let P be the transition matrix for an irreducible and aperiodic Markov chain whose stationary distribution is π. For a set $A \subset \Omega$ let $A^c = \Omega \backslash A$. The *bottleneck ratio or conductance* of the Markov chain is

$$\Phi_* := \min_{A \subset \Omega, \pi(A) \leq \frac{1}{2}} \frac{\sum_{x \in A, y \in A^c} \pi(x) P(x,y)}{\pi(A)}.$$

Example 1.2.19. Let G be a finite graph that is d- regular, that is, each vertex has degree d. We can compute the conductance of the simple random walk defined on G, as described in Section 1.1.1. Recall that the stationary distribution π is uniform on Ω. Then the conductance,

$$\begin{aligned}
\Phi_* &= \min_{A \subset \Omega, \pi(A) \leq \frac{1}{2}} \frac{\sum_{x \in A, y \in A^c} \pi(x) P(x,y)}{\pi(A)} \\
&= \min_{\substack{A \subset \Omega, \\ 0 < 2|A| \leq |\Omega|}} \sum_{\substack{x \in A, y \in A^c, \\ x \sim y}} \frac{1}{d|A|} \\
&= \min_{\substack{A \subset \Omega, \\ 0 < 2|A| \leq |\Omega|}} \sum_{\substack{x \in A, y \in A^c, \\ x \sim y}} \frac{1}{d|A|} \\
&= \frac{1}{d} \min_{\substack{A \subset \Omega, \\ 0 < 2|A| \leq |\Omega|}} \frac{e(A, A^c)}{|A|},
\end{aligned}$$

where $e(A, A^c)$ is the number of edges with exactly one endpoint in A and the other in A^c. Notice that in this case $\Phi_* = \frac{1}{d} h(G)$, where $h(G)$ is the edge expansion of G.

Lemma 1.2.20 shows that the diameter of a d- regular graph G and the conductance of the random walk on G are related. We will make use of this fact in Chapter 3. The proof of Lemma 1.2.20 we follow is from [18] Chapter 4 Section 2.

Lemma 1.2.20. *Let $G = (V, E)$ be a finite connected d- regular graph. The conductance Φ_* of* *the random walk on G satisfies the following inequality:*

$$diam(G) \leq \frac{2\log|V|}{\log(1+\Phi_*)}$$

Proof. First observe that for any set $S \subset V$ of vertices such that $0 < |S| < \frac{1}{2}|V|$ the number of edges $e(S, S^c)$ across S is at least $d|S|\Phi_*$. Moreover the number of neighbors

$$|\left\{ y \in S^c : d(x,y) = 1 \text{ for some } x \in S \right\}|$$

is at least $|S|\Phi_*$ since G is d- regular.

For a vertex $x \in V$, let $\mathcal{B}_r(x) := \left\{ y \in V \mid d(x,y) \leq r \right\}$ be the closed ball of radius r centered at x. It follows from the observation that if r_x is the least positive integer such that $|\mathcal{B}_{r_x}(x)| > \frac{1}{2}|V|$ then we must have that $|\mathcal{B}_{r_x-1}(x)| \leq \frac{1}{2}|V|$ and so $|\mathcal{B}_{r_x}(x)| \geq (1+\Phi_*)^{r_x}$.

Now for any $y \in V$ with $x \neq y$, let r_y be analogously defined. Then the intersection $\mathcal{B}_{r_x}(x) \cap \mathcal{B}_{r_y}(y) \neq \emptyset$. Say $w \in \mathcal{B}_{r_x}(x) \cap \mathcal{B}_{r_y}(y)$ then a path from x to y can be constructed by joining a path from x to w and a path from w to y. So the graph distance between x and y satisfies the following:

$$d(x,y) \leq r_x + r_y \leq \frac{\log|\mathcal{B}_{r_x}(x)|}{\log(1+\Phi_*)} + \frac{\log|\mathcal{B}_{r_y}(y)|}{\log(1+\Phi_*)} \leq \frac{2\log|V|}{\log(1+\Phi_*)}.$$

The result follows since the choice of x and y are arbitrary. □

Finally, the conductance of a Markov chain is related to its mixing time.

Theorem 1.2.21. *If Φ_* is the conductance of an irreducible aperiodic Markov chain then,* $t_{mix} \leq \frac{1}{4\Phi_*}$.

For proof of Theorem 1.2.21 see Chapter 7 of [20].

1.3 Polytopes

In this section we introduce polytopes and related tools in preparation for Chapters 3 and 4. Here we see that a polytope is a convex set in $\mathbb{R}^n$ with flat sides that can be described by vertices and by a finite collection of half-spaces of $\mathbb{R}^n$.

Definition 1.3.1. For a pair of points $x, y \in \mathbb{R}^n$ the *line segment* $\overline{xy}$ is the set

$$\overline{xy} = \{\lambda x + (1-t)y \mid \lambda \in [0,1]\}.$$

Definition 1.3.2. A set S is *convex* if for any pair of points $x, y \in S$, $\overline{xy}$ is contained is S.

Given a point set in $\mathbb{R}^n$, we can consider its convex hull, which as the name suggests, is a convex set.

Definition 1.3.3. Let $\mathcal{X}$ be a point set in $\mathbb{R}^n$. The *convex hull* of $\mathcal{X}$, denoted $\operatorname{conv}(\mathcal{X})$, is the intersection of all convex sets containing $\mathcal{X}$.

When $\mathcal{X}$ is a finite point set then $\operatorname{conv}(\mathcal{X})$ is equivalently represented by the set of all convex combinations of the points in $\mathcal{X}$. In symbols, if $\mathcal{X} = \{x_1, \ldots, x_d\}$ then,

$$\operatorname{conv}(\mathcal{X}) = \left\{ \sum_{i=1}^{d} \lambda_i x_i \mid \lambda_i \geq 0 \quad \text{for all } i \text{ and } \sum_{i=1}^{d} \lambda_i = 1 \right\}.$$

In addition to the convex hull we define the affine hull of a set $\mathcal{X}$. The affine hull is an affine subspace and we borrow the notion of dimension of affine subspace to later define dimension of a polytope.

Definition 1.3.4. The *affine hull* of a set $\mathcal{X}$ is the set of all affine combinations of its points:

$$\operatorname{aff}(\mathcal{X}) = \left\{ \sum_{i=1}^{d} \lambda_i x_i \mid d > 0 \text{ and } \sum_{i=1}^{d} \lambda_i = 1 \right\}.$$

There are two complementary ways to represent, and therefore define, a polytope.

Definition 1.3.5 ($\mathcal{V}$- Representation)**.** A *$\mathcal{V}$-polytope* P is the convex hull of a finite point set V. If each point in V is necessary, meaning $\operatorname{conv}(V - \{v\}) \subsetneq \operatorname{conv}(V)$ for all $v \in V$, then V is the *vertex set* of P, denoted $\operatorname{vert}(P) = V$.

Example 1.3.6. Let $\mathcal{V} = \{(-1, 11), (1, 4), (2, 3), (5, 0), (6, 6), (7, 7), (8, -2)\}$ be a finite point set in the plane. The convex hull of $\mathcal{V}$, displayed in Figure 1.5 is a pentagon. Notice that the vertices of the pentagon are present in $\mathcal{V}$ along with some additional points. The points $(2, 3)$ and $(6, 6)$ can be removed from $\mathcal{V}$ with no consequence. Let $\mathcal{V}' = \{(-1, 11), (1, 4), (5, 0), (7, 7), (8, -2)\}$ then $\operatorname{conv}(\mathcal{V}) = \operatorname{conv}(\mathcal{V}')$. If any points are further removed from $\mathcal{V}'$ then the resulting convex hull is a proper subset of $\operatorname{conv}(\mathcal{V})$. Hence $\mathcal{V}'$ is the vertex set of the polytope $\operatorname{conv}(\mathcal{V})$.

Definition 1.3.7. An *$\mathcal{H}$-polyhedron* P is the solution set of finitely many linear inequalities and thus can be represented as

$$P = P(A, b) = \{x \in \mathbb{R}^n \mid Ax \leq b\},$$

for some $m \times n$ matrix A and vector $b \in \mathbb{R}^m$. The prefix $\mathcal{H}$- refers to the fact that the solution set is also the intersection of finitely many half-spaces.

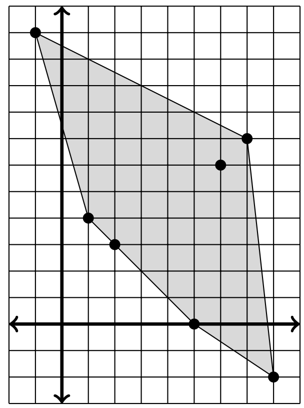

Figure 1.5: The convex hull of the set $\mathcal{V}$ from Example 1.3.6.

Example 1.3.8. Let the matrix $A \in \mathbb{R}^{3\times 2}$ and vector $b \in \mathbb{R}^3$ be the following.

$$A = \begin{bmatrix} 1 & 2 \\ -7 & -2 \\ -2 & -3 \end{bmatrix}, \quad b = \begin{bmatrix} 21 \\ -15 \\ -10 \end{bmatrix}$$

Then the polyhedron $P = P(A, b)$ is the solution to the system of linear equations Eq 1.6.

$$\begin{aligned} x + 2y &\leq 21 \\ -7x - 2y &\leq -15 \\ -2x - 3y &\leq -10 \end{aligned} \tag{1.6}$$

Geometrically, the polyhedron is the shaded region in Figure 1.6. From the figure we observe that P is a convex set with "flat sides". This feature is common to all polyhedra. We further notice that P is unbounded since it contains a ray, in particular $\{(4, 4) + t(1, -1), t \geq 0\}$, which is also displayed in Figure 1.6. Later in Chapters 3 and 4 we will be concerned with bounded polyhedron and look at Markov chain-based methods to sample from these sets.

Definition 1.3.9 ($\mathcal{H}$-Representation)**.** An *$\mathcal{H}$-polytope* P is a bounded $\mathcal{H}$-polyhedron.

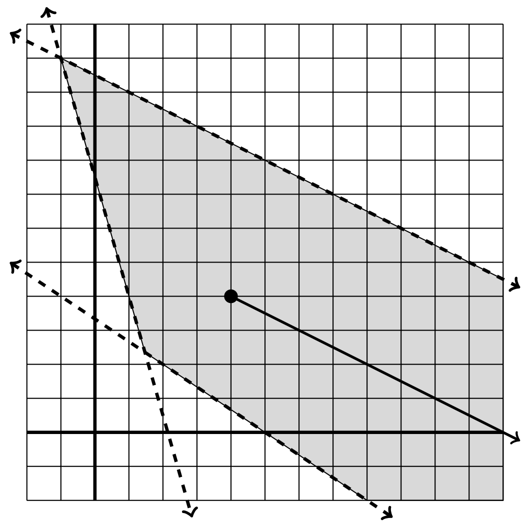

Figure 1.6: The shaded region is the polyhedron $P(A, b)$ from Example 1.3.8. The polyhedron contains a ray, in particular $(4, 4) + \lambda(1, -2)$ for $\lambda \geq 0$, therefore the polyhedron is unbounded.

Example 1.3.10. Suppose we add rows to the matrix A and vector b from Example 1.3.8. Let

$$A' = \begin{bmatrix} 1 & 2 \\ -7 & -2 \\ -2 & -3 \\ -1 & -1 \\ 9 & 1 \end{bmatrix}, \quad b' = \begin{bmatrix} 21 \\ -15 \\ -10 \\ -5 \\ 70 \end{bmatrix}.$$

By graphing the corresponding system of inequalities we find that $P = P(A', b')$ is a bounded set, therefore P is an $\mathcal{H}$-polytope. Also P is the same as the $\mathcal{V}$-polytope in Figure 1.5 from Example 1.3.6.

It is no coincidence that the $\mathcal{V}$-polytope in Example 1.3.6 can also be represented as the $\mathcal{H}$-polytope in Example 1.3.10. Rather, it is an instance of the Minkowski-Weyl Theorem, a fundamental result in the theory of polyhedra, which implies that any subset of $\mathbb{R}^n$ that can be represented as a $\mathcal{V}$-polytope can also be represented as an $\mathcal{H}$-polytope, and vice versa.

Definition 1.3.11. For a point set $\mathcal{X}$ the *cone* denoted $\text{cone}(\mathcal{X})$ is the set of all nonnegative combinations of the points in $\mathcal{X}$, that is,

$$\text{cone}(\mathcal{X}) = \left\{ \sum_{i=1}^{d} \lambda_i x_i : d > 0, x_i \in \mathcal{X}, \ \ \lambda_i \geq 0 \right\}.$$

Definition 1.3.12. The *Minkowski sum* of two sets P and Q in $\mathbb{R}^n$ is

$$P + Q = \{x + y \mid x \in P \text{ and } y \in Q\}.$$

Theorem 1.3.13 (Minkowski- Weyl Theorem)**.** *For a subset P of $\mathbb{R}^n$, the following statements are equivalent:*

1. *$P = P(A, b)$ for some $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$.*
2. *There exist vectors $x_1, \ldots, x_k$ and $v_1, \ldots, v_s \in \mathbb{R}^n$ such that*

$$P = \mathrm{conv}(\{x_1, \ldots, x_k\}) + \mathrm{cone}(\{v_1, \ldots, v_s\}).$$

We refer the reader to [29] for a detailed proof of this result and for a more complete theory of polytopes. In Chapters 3 and 4, our work will include both representations of polytopes depending on which is more convenient. For the $\mathcal{H}$-representation of a polytope we will often assume that the system of linear inequalities that defines a polytope does not contain any redundant inequalities. At times we may refer to those inequalities individually. Let a_i^T represent the i-th row vector of the matrix A.

Definition 1.3.14. For a matrix $A \in \mathbb{R}^{m\times n}$ and vector $b \in \mathbb{R}^m$, the i-th inequality $a_i^T x \leq b_i$ of the system $Ax \leq b$ is *redundant* if $P(A, b) = P(A_{-i}, b_{-i})$, where A_{-i} and b_{-i} are the result of removing the i-th row from A and b respectively. A system $Ax \leq b$ is *irredundant* if it contains no redundant inequalities.

At times we will refer to the dimension of a polytope. In these cases the notion of dimension is consistent with the usual notion of dimension for a convex subset of $\mathbb{R}^n$.

Definition 1.3.15. The *dimension*, denoted $\dim(P)$, of a polytope P is the dimension of its affine hull.

Most of the polytopes that we work with are full-dimensional, meaning that their dimension is equal to the dimension of the ambient space being considered. We may refer to an n-dimensional polytope as an n-polytope.

The most significant features of a polytope, that determine its combinatorial and geometrical structure are called faces. Informally, the faces of a 2- polytope, like the one in Figure 1.5, include its vertices and the edges. For a 3-polytope, the faces include the vertices, edges, and 2-dimensional sides.

Definition 1.3.16. A linear inequality $cx \leq d$ is *valid* for P if it is satisfied by all points $x \in P$.

Definition 1.3.17. A *face* F of a polytope P is any set of the form $F = P \cap \{x \in \mathbb{R}^n \mid cx = d\}$ for any valid inequality $cx \leq d$ for P.

For any polytope P the inequality $\mathbf{0}x \leq \mathbf{1}$ is valid and the set $P \cap \{x \in \mathbb{R}^n \mid \mathbf{0}x = \mathbf{1}\}$ is empty. Hence $\emptyset$ is always a face of P. On the other extreme, the inequality $\mathbf{0}x \leq \mathbf{0}$ is valid and the set $P \cap \{x \in \mathbb{R}^n \mid \mathbf{0}x = \mathbf{0}\} = P$. Hence P is also always a face of P.

Suppose $P = P(A, b)$ is a polytope and $F = P \cap \{x \in \mathbb{R}^n : cx = d\}$ is a face of P. Then F is represented by the system in Equation 1.7. Consequently, F is itself a polytope. We can further show that the vertices of F are exactly the vertices of P contained in F.

$$\begin{bmatrix} A \\ c \\ -c \end{bmatrix} x \leq \begin{bmatrix} b \\ d \\ d \end{bmatrix} \tag{1.7}$$

For an n-polytope it is standard practice to refer to the 0-, $1-$, $(n-2)-$, and $(n-1)$-dimensional faces as vertices, edges, ridges, and facets, respectively. We often refer to any inequality $cx \leq d$ that defines a facet as *facet-defining* and the solution to the corresponding equation $cx = d$ as a *facet-defining hyperplane*. For a polytope $P = P(A, b)$ if the system $Ax \leq b$ is irredundant then each inequality $a_i^T x \leq b_i$ is facet-defining.

Example 1.3.18. Let $\mathcal{C}_3$ be the unit cube situated in the positive octant of $\mathbb{R}^3$. It is straightforward to see that $\mathcal{C}_3$ is a 3-polytope since it can be represented as the convex hull $\text{conv}(\{0,1\}^3)$ and by $P(A, b)$ with the following matrix and vector:

$$A = \begin{bmatrix} I_3 \\ -I_3 \end{bmatrix}, \quad b = \begin{bmatrix} \mathbf{1}_3 \\ \mathbf{0}_3 \end{bmatrix}$$

In addition to $\emptyset$ and $\mathcal{C}_3$, the faces of $\mathcal{C}_3$ include 8 vertices enumerated by $\{0,1\}^3$, 12 edges corresponding to the valid inequalities displayed in Equation 1.8 for $1 \leq i < j \leq 3$, and 6 facets corresponding to the valid inequalities $0 \leq x_i \leq 1$ for $i = 1, 2, 3$.

$$\begin{aligned} x_i + x_j &\leq 2 \\ -x_i + x_j &\leq 1, \\ x_i - x_j &\leq 1, \\ -x_i - x_j &\leq 0 \end{aligned} \tag{1.8}$$

The problem addressed in Chapters 3 and 4 concerns algorithms for sampling from polytopes. For the most part, we attempt to keep arbitrary polytopes in mind, however, the motivation for our work comes from the specific class of polytopes that arise from contingency tables.

The next example is a small taste of these polytopes that we revisit in Chapters 3 and 4.

Example 1.3.19. Consider the set $\mathcal{P}_{3,4}(r,c)$ of 3×4 matrices with nonnegative real values and whose row and columns sums are given by the vectors $r = (4, 12, 10)$ and $c = (7, 7, 6, 6)$, respectively. Then $\mathcal{P}_{3,4}(r,c)$ represents a polytope in $\mathbb{R}^{12}$, which we can demonstrate with the appropriate $\mathcal{H}$-representation. For $X \in \mathcal{P}_{3,4}(r,c)$, X satisfies the following:

$$X_{ij} \geq 0 \quad \text{for all } 1 \leq i \leq 3, 1 \leq j \leq 4, \tag{1.9}$$

$$\begin{aligned} \sum_{i=1}^{3} X_{ij} &= c_j \text{ for all } 1 \leq j \leq 4, \\ \sum_{j=1}^{4} X_{ij} &= r_i \text{ for all } 1 \leq 1 \leq 3. \end{aligned} \tag{1.10}$$

Notice though, that due to the equalities of the latter two statements, those matrices satisfying Equation 1.10 form a 6-dimensional set in $\mathbb{R}^{12}$. So $\mathcal{P}_{3,4}(r,c)$ can be realized as a full-dimensional polytope in the ambient space $\mathbb{R}^6$. To do so, notice that for $X \in \mathcal{P}_{3,4}(r,c)$ the last row and column of X can be expressed in terms of the entries X_{ij}, with $1 \leq i < 3$ and $1 \leq j < 4$:

$$\begin{aligned} X_{3j} &= c_j - X_{1j} - X_{2j} \quad \text{for } 1 \leq j < 4 \\ X_{i4} &= r_i - \sum_{j=1}^{3} X_{ij} \quad \text{for } i = 1, 2 \\ X_{34} &= c_4 - r_1 - r_2 + \sum_{j=1}^{3} X_{1j} + \sum_{j=1}^{3} X_{2j} \end{aligned}$$

Then in $\mathbb{R}^6$ we can represent $\mathcal{P}_{3,4}(r,c)$ by the irredundant system

$$\begin{aligned} -X_{ij} &\leq 0 \quad \text{for all } 1 \leq i < 3, 1 \leq j < 4 \\ X_{1j} + X_{2j} &\leq c_j \quad \text{for } 1 \leq j < 4 \\ \sum_{j=1}^{3} X_{ij} &\leq r_i \quad \text{for } i = 1, 2 \\ -\sum_{j=1}^{3} X_{1j} - \sum_{j=1}^{3} X_{2j} &\leq c_4 - r_1 - r_2. \end{aligned}$$

Remark 1.3.20. *If we let the matrix*

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

and reformat matrices to vectors, then each point $X \in \mathcal{P}_{3,4}(r,c)$ *is a nonnegative solution to* $AX = (r,c)^T$. *Looking ahead to Chapter 3, we call* A *the* configuration matrix *for* 3×4 *contingency tables.*

1.4 Survey of Results in Thesis

In this thesis we look at a few settings where Markov chains arise and we explore this mixing time question. In Chapter 2 we seek to determine the mixing behavior for a family of random walks associated to a linear recurrence. Let $\{G_i\}_{i=1}^{\infty}$ be a positive integer sequence satisfying a linear recurrence $G_n = \sum_{i=1}^{d} \alpha_i G_{n-i}$, with $G_1 = 1$. For each n we have a random walk whose state space is $\mathbb{Z}_{G_n} = \{0, 1, 2, \ldots, G_n - 1\}$, and whose transition probabilities given by

$$P(x,y) = \begin{cases} \frac{1}{n} & \text{if } (y-x) \in \{0, 1, G_2, \ldots, G_{n-1}\} \\ 0 & \text{otherwise.} \end{cases}$$

We show that for general linear recurrences with exponential growth, the mixing time is bounded above by $\kappa_1 n^2$ and below by $\kappa n / \log n$, where κ_1 and κ_2 are constants that depend on the sequence. We further show that in the special case of first order recurrences that the mixing time is between $\gamma_1 n$ and $\gamma_2 n \log n$, where γ_1 and γ_2 are also constants that depend on the sequence.

Random walks on the integers modulo p have been examined most notably as it relates to the problem of pseudo-random number generation. In such cases the dynamics of these random walks are given by the recurrence $X_{t+1} = aX_t + b \mod p$ where p is some prime number and a and b can be given by a variety of schemes. For example, [6] shows that if $a = 1$ and $b = 0, -1,$ or 1 each with probability $\frac{1}{3}$ then the mixing times is bounded by κp^2. The situation improves, mixing time is bounded by $\kappa \log p \log \log p$ if $a = 2$ and again $b = 0, -1,$ or 1 with equal probability. Our setting differs since the number of available moves at each step grows with n.

Chapters 3 and 4 concern the problem of sampling lattice points of polytope. The approach of Chapter 3 is to define a fiber graph on the points in question, and define a Markov basis, called the *simple fiber walk* on the lattice points. By analyzing the diameter of the underlying graphs, we show the simple fiber walk does not exhibit rapid mixing.

In Chapter 4 we consider a relaxation of the problem of sampling the lattice points $\mathcal{F}$ that follows the strategies of Morris [22] and Dyer, Kannan, and Mount [12]. There we implement a continuous sampling algorithm on a polytope $\tilde{P}$ that contains P, and then round to the nearest lattice point, repeating the process until a point in $\mathcal{F}$ is generated. For this approach, there are choices to be made about $\tilde{P}$ and the continuous sampling algorithm. We discuss those choices, prove a result to bound the rejection rate, and implement the algorithms in `R`.

1.5 Notation

We make every attempt to maintain consistent notation throughout this thesis. The set of natural numbers $\mathbb{N} := \{ 1, 2, 3, \dots \}$ does not include zero. For a number $n \in \mathbb{N}$, the set $[n] := \{ 1, 2, \dots, n \}$. For logarithmic functions $\log(x) = \log_{10}(x)$ denotes the common logarithm while $\ln(x)$ is used to denote the natural logarithm. The vector $\mathbf{1}_{1\times n}$ and $\mathbf{0}_{1\times n}$ represents the all-ones vector and all-zeroes vector, and dimensions are given by context. When describing the limiting behavior of a real-valued function f, we say that f *is dominated by the function* g and write $f(x) = O(g(x))$ if there exists a constant c and $x_0 \in \mathbb{R}$ such that for all $x \geq x_0$, we have $|f(x)| \leq cg(x)$. We say that f *is asymptotically bound below by the function* g if there exists a constant c and $x_0 \in \mathbb{R}$ such that for all $x \geq x_0$, $f(x) \geq c \cdot g(x)$. At times, we also adopt the soft-O notation, as used in [11]. We write $f(x) = O^*(g(x))$ when $f(x) = O(g(x))$, where logarithmic factors of x have been suppressed.

Chapter 2

Linear Recurrence Random Walk

2.1 Introduction

Let $(G_n)_{n\geq 1}$ be a positive increasing integer sequence given by the linear recurrence with constant coefficients

$$G_n = \alpha_1 G_{n-1} + \alpha_2 G_{n-2} + \cdots + \alpha_d G_{n-d}, \tag{2.1}$$

and $G_1 = 1$. This sequence determines a family of random walks.

Definition 2.1.1. The *linear recurrence random walk* associated to the sequence $(G_n)_{n\geq 1}$ is the Markov chain $(X_t)_{t\geq 0}$ whose state space is $\Omega = \mathbb{Z}_{G_n}$. The initial state is $X_0 = 0$ and from the current state X_t, the next state is

$$X_{t+1} \equiv X_t + z_t \bmod G_n,$$

where z_t is chosen from the set $\mathcal{M} = \{G_1, G_2, \ldots, G_n\}$ uniformly at random. The transition matrix P for the linear recurrence random walk is

$$P(x,y) = \begin{cases} \frac{1}{n} & \text{if } y - x \equiv G_i \bmod G_n \text{ for some } 1 \leq i \leq n \\ 0 & \text{otherwise.} \end{cases}$$

So for each n we have a random walk on the finite abelian group $(\mathbb{Z}_{G_n}, +)$. By the assumption $G_1 = 1$, the set $\mathcal{M}$ generates the group and hence the random walk is irreducible. Further as $G_n \in \mathcal{M}$, the walk is aperiodic. The stationary distribution π, to which the random walk converges, is uniform over Ω. In this chapter, we seek to answer the following question.

Problem 2.1.2. *What is the mixing time of the linear recurrence random walk associated to the sequence* $(G_n)_{n\geq 1}$*?*

Our approach to Problem 2.1.2 is to leverage the relationship between the mixing time and the second largest eigenvalue modulus of the transition matrix. In Section 2.4 we use explicit formulas for the eigenvalues of the transition matrix to prove that for a random walk arising from $(G_n)_{n\geq 1}$ subject to certain conditions, at most κn^2 steps will suffice where κ is some constant that depends on $(G_n)_{n\geq 1}$. Section 2.5 focuses on random walks arising from first order recurrences. In that case we show that $\gamma n \log n$ steps will suffice, where γ is also some constant that depends on $(G_n)_{n\geq 1}$.

Our results on the eigenvalues of these Markov chains also allow us to derive lower bounds on the mixing times in the case that G_n grows like an exponential function. For general linear recurrences of exponential growth, we have the lower bound of the form $\kappa n/\log n$ and in the first order case we get a lower bound of the form κn.

Though we have proven these upper and lower bounds on the mixing times we suspect from simulations that the mixing time grows like n instead of $n \log n$ or n^2. The table below displays the mixing times for random walks arising from three integer sequences.

Mixing Times for Three Sequences						
n	$G_n = 2^{n-1}$	t_{mix}	$G_n = 3^{n-1}$	t_{mix}	$G_n = 3G_{n-1} - G_{n-2}$	t_{mix}
1	1	0	1	0	1	0
2	2	1	3	2	3	2
3	4	2	9	3	8	3
4	8	2	27	3	21	3
5	16	3	81	4	55	3
6	32	3	243	4	144	4
7	64	3	729	4	377	4
8	128	4	2187	5	987	4
9	256	4	6561	5	2584	4

Random walks on the integers modulo some n have been studied frequently, as they are a prototypical example of a Markov chain on a group, and are amenable to techniques based on discrete Fourier analysis. In his review article [25], Saloff-Coste considers random walks on $\mathbb{Z}_n$ given by $X_{t+1} \equiv X_t + z_t \bmod n$ where $\Pr(z_t = a) = \Pr(z_t = b) = \frac{1}{2}$ for some choice of $a, b \in \mathbb{Z}_n$. Hildebrand [14] considers walks on $\mathbb{Z}_n$ given by the $X_{t+1} \equiv X_t + z_t \bmod n$ where z_t is uniform on a set of k random elements of $\mathbb{Z}_n$. He shows that if n is prime then it suffices to take $\kappa n^{2/(k-1)}$ steps to be close to uniformly distributed for almost all choices of k elements. Hildebrand also considers the case where the size of the random step set grows with n, and the situation studied in this paper provides an interesting deterministic boundary case between Theorems 3 and 4 of [14]. Diaconis [8] discusses various random walks on $\mathbb{Z}_n$ given by the $X_{t+1} \equiv a_t X_t + z_t \bmod n$, where a_t and z_t are subject to various restrictions. Our work seems to be the first that considers a family of steps in the Markov chain on $\mathbb{Z}_n$ where the set of possible steps increases with n.

2.2 Generalization of the Abelian Sandpile Markov Chain

Our attention to Problem 2.1.2 arose from a project in which we set out to generalize the abelian sandpile Markov chain introduced in [15] by Jerison, Levine, and Pike. We first summarize the relevant results and then outline our trajectory.

Let $G = (V, E)$ be a simple connected graph on n vertices and identify a special sink vertex v_s. A *sandpile* on G is a distribution of "sand grains" over the vertices of G. A *configuration* of the sandpile is a function $\sigma : V \backslash \{v_s\} \to \mathbb{N}$. The configuration is *stable* if $\sigma(v) < \deg(v)$ for all $v \in V \backslash \{v_s\}$. Any configuration can be stabilized by iteratively "toppling" non-sink vertices where $\sigma(v) \geq \deg(v)$.

In [15] the authors introduce a Markov chain on the set of stable configurations of the sandpile. From current state σ_t, pick a vertex $v \in V$ uniformly at random. Add one grain of sand at v and then stabilize the configuration to obtain the next state σ_{t+1}. The chain is irreducible, aperiodic, and the stationary distribution is uniform over the set of recurrent states. The recurrent states of the chain form a finite abelian group called the *abelian sandpile group* of (G, v_s). The *abelian sandpile Markov chain* is the result of restricting the state space of the aforementioned chain to the abelian sandpile group.

The abelian sandpile Markov chain can also be recognized as a random walk on a lattice. Let Δ be the reduced Laplacian of G. That is, Δ is an $(n-1) \times (n-1)$ integer matrix whose rows and columns are indexed by the non-sink vertices of the graph and

$$\Delta_{ij} = \begin{cases} \deg(v_i) & \text{if } i = j \\ -1 & \text{if } v_i \sim v_j \\ 0 & \text{else.} \end{cases}$$

Letting $\Delta\mathbb{Z}^{n-1} = \{\Delta z : z \in \mathbb{Z}^{n-1}\}$, the chain can be characterized as having as its state space the quotient $\mathbb{Z}^{n-1}/\Delta\mathbb{Z}^{n-1}$. The dynamics of the chain are restated: From the current state x_t, to generate the next state choose some z from the set $\{\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_{n-1}}, \mathbf{0}\}$ uniformly at random, where e_i's are the standard basis vectors and $\mathbf{e_i} = e_i + \Delta\mathbb{Z}^{n-1}$. The next state is then $x_{t+1} = x_t + z$. In [15], the mixing time of various instances of these chains were analyzed. For example, when $G = C_n$ is the cycle graph on n vertices, the chain enjoys very fast mixing,

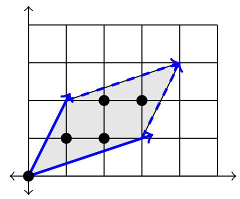

Figure 2.1: Elements of the quotient $\mathcal{Q} = \mathbb{Z}^2/A_0\mathbb{Z}^2$ from Example 2.2.1

in particular, the chain reaches stationarity after one step. On the other hand, when $G = K_n$ is the complete graph on n vertices the order of the mixing time is $kn^3 \log n$, hence the chain exhibits significantly slower mixing behavior.

We were interested in determining what type of mixing behavior is exhibited by a variation of these chains. Let A be an invertible $n \times n$ matrix and consider the lattice quotient $\mathcal{Q} := \mathbb{Z}^n/A\mathbb{Z}^n$. Then elements of $\mathcal{Q}$ are the integer points of the parallelepiped $\{ Ax : x \in [0,1)^n \}$ and for arbitrary $x \in \mathbb{Z}^n$, we let $[x]$ denote the equivalence class $x + A\mathbb{Z}^n$. The size of the quotient $\mathcal{Q}$ is given by the magnitude of the determinant $|\det(A)|$. Addition on $\mathcal{Q}$ is given by usual coset addition, that is, $[x] + [y] = [x + y]$ for all $x, y \in \mathbb{Z}^n$. Then $(\mathcal{Q}, +)$ forms a finite abelian group.

Example 2.2.1. For the invertible matrix $A_0 = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$, the quotient $\mathcal{Q} = \mathbb{Z}^2/A_0\mathbb{Z}^2$, illustrated in Figure 2.1, is given by $\mathcal{Q} = \{(0,0), (1,1), (2,1), (2,2), (3,2)\}$. The Cayley table, which displays the group structure, is

+	(0,0)	(1,1)	(2,1)	(2,2)	(3,2)
(0,0)	(0,0)	(1,1)	(2,1)	(2,2)	(3,2)
(1,1)	(1,1)	(2,2)	(3,2)	(2,1)	(0,0)
(2,1)	(2,1)	(3,2)	(1,1)	(0,0)	(2,2)
(2,2)	(2,2)	(2,1)	(0,0)	(3,2)	(1,1)
(3,2)	(3,2)	(0,0)	(2,2)	(1,1)	(2,1) .

We can define a Markov chain on $\mathcal{Q}$, that is analogous to the abelian sandpile Markov chain, by letting the equivalence classes represented by each standard basis vector and the zero vector represent moves, roughly giving us a way to walk around within a cell of the integer lattice.

Definition 2.2.2. For an invertible $n \times n$ matrix A, the *lattice walk* on $\mathcal{Q} = \mathbb{Z}^n/A\mathbb{Z}^n$ is the

Markov chain with transition matrix

$$P([x],[y]) = \begin{cases} \frac{1}{n+1} & \text{if } x-y \in A\mathbb{Z}^n \\ \frac{1}{n+1} & \text{if } x-y \in e_i + A\mathbb{Z}^n \text{ for some } i=1,\dots,n \\ 0 & \text{otherwise.} \end{cases}$$

It is clear that the lattice walk is aperiodic since $P([x],[x]) > 0$. The lattice walk is irreducible since for any $[x],[y] \in \mathcal{Q}$, there exists a representative $[x'] = [x]$ such that the difference $x' - y$ is a positive integer combination of the standard basis vectors. The stationary distribution of the lattice walk is uniform. Arguments in Section 2 of [15] can be modified appropriately in order to get a formula for the eigenvalues of the transition matrix for the lattice walk. For a function $h : [n] \to D$, where D is an arbitrary set, we use the shorthand h_i to denote the function value $h(i)$. Also, recall that we let $\mathbb{T}$ denote the unit circle in the complex plane.

Definition 2.2.3. Let A_i denote the i-th column of the matrix A. Then a function $h : [n] \to \mathbb{T}$ is *harmonic with respect to* A if $h^{A_i^+} = h^{A_i^-}$ for all $i = 1,\dots,n$. Let $\mathcal{H}_A$ denote the set of harmonic functions with respect to A.

Given a function $h \in \mathcal{H}_A$, we can define a homomorphism $\chi_h : \mathcal{Q} \to \mathbb{T}$, where

$$\chi_h([x]) = \prod_{i=1}^{n} h_i^{x_i}. \tag{2.2}$$

Using the correspondence defined in Equation 2.2, we can show that the harmonic functions for a matrix A are in 1-1 correspondence with the elements of the dual group $\hat{\mathcal{Q}} = \mathrm{Hom}(\mathcal{Q},\mathbb{T})$, which consists of homomorphism from the group $\mathcal{Q}$ to $\mathbb{T}$.

Proposition 2.2.4. *Let P be the transition matrix for the lattice walk on $\mathcal{Q}$. Then P has an orthonormal basis $\{\chi_h : h \in \mathcal{H}_A\}$ of eigenfunctions and the corresponding eigenvalues are $\lambda_h = \hat{\mu}(\chi_h) := \sum_{[x]\in\mathcal{Q}} \mu([x])\chi_h([x])$ for each $h \in \mathcal{H}_A$. Further, $\lambda_h = \frac{1}{n+1}(\sum_{i=1}^n h_i + 1)$.*

Proof. The first statement is Lemma 2.1 of [15], which only relies on the finite abelian structure of $\mathcal{Q}$. The latter holds by the 1-1 correspondence between $\mathcal{H}_A$ and $\widehat{\mathcal{Q}}$ defined in Equation 2.2. □

The significant takeaway from Proposition 2.2.4 is the characterization of the eigenvalues in terms of the harmonic functions $\mathcal{H}_A$. It is worth mentioning that in certain cases, the formula for the eigenvalues can be arrived in another way. Due to the structure, the lattice walk on $\mathcal{Q}$ is equivalent to a random walk on a finite abelian group where moves are generated uniformly from the set $\{ \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n, \mathbf{0} \}$. If the equivalent group is cyclic, then the transition matrix of the lattice walk is a circulant matrix, the eigenvalues of which can be described nicely.

Definition 2.2.5. An $n \times n$ matrix A is *circulant* if each row is a circular shift of the first row. In other words, A has the form

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_{n-1} & a_0 \\ a_0 & a_1 & \cdots & a_{n-2} & a_{n-1} \\ & \ddots & \ddots & & \\ a_2 & a_3 & & \cdots & a_1 \end{bmatrix}.$$

For a vector $\mathbf{a}$, the matrix $\operatorname{circ}(\mathbf{a})$ is the circulant matrix whose first row is $\mathbf{a}$.

Lemma 2.2.6. *Suppose* $\mathbf{a} = (a_1, a_2, \ldots, a_{n-1}, a_0) \in \mathbb{R}^n$. *For* $j = 0, 1, \ldots, n-1$, *let* $\xi_j = \exp\left(\frac{2\pi j}{n}\mathbf{i}\right)$ *be an* n*-th root of unity. (Here* i *is the imaginary unit, not an index). Then the eigenvalues of the matrix* $\operatorname{circ}(\mathbf{a})$ *are*

$$\lambda_j = \sum_{i=1}^{n} a_{i (\bmod n)} \xi_j^{i-1}, \ \textit{for } j = 0, 1, \ldots, n-1$$

and the corresponding eigenvectors are $f_j = (1, \xi_j, \xi_j^2, \ldots, \xi_j^{n-1})$.

Proof. Let $A = \operatorname{circ}(\mathbf{a})$. Then the k-th entry of the vector (Af_j) is

$$\begin{aligned} (Af_j)_k &= \sum_{i=1}^{n} a_{i-k+1 (\bmod n)} \xi_j^{i-1} \\ &= \xi_j^{k-1} \cdot \sum_{i=1}^{n} a_{i-k+1 (\bmod n)} \xi_j^{i-k} \end{aligned}$$

Since ξ_j is an n-th root of unity we have that, for each $1 \leq k \leq n$, the list $(a_{i-k+1 (\bmod n)} \xi_j^{i-k})_{i=1}^{n}$ is a circular shift of the list $(a_{i (\bmod n)} \xi_j^{i-1})_{i=1}^{n}$. The result follows. □

Example 2.2.7. Let T_n be the $n \times n$ tridiagonal matrix with 3's on the diagonal, -1's on the super- and sub-diagonals, and zeros elsewhere. The lattice walk on $\mathbb{Z}^n / T_n \mathbb{Z}^n$ is a natural generalization of the sandpile random walk for the cycle graph, whose reduced Laplacian is the matrix with 2's on the diagonal and -1's on super- and sub-diagonals.

The size $|\mathbb{Z}^n / T_n \mathbb{Z}^n|$ of the state space of the lattice walk is F_{2n}, the $2n$-th Fibonacci number, where we assume the seed values $F_0 = 1$ and $F_1 = 2$. This fact can be shown inductively by computing the determinant of the matrix T_n. First observe that $\det(T_0) = 1$ and $\det(T_1) = 3$. For $n \geq 2$, by expanding the determinant along the first row it follows that $\det(T_n) = 3(-1)^{(1+1)} d_{1,1}^n - 1(-1)^{(1+2)} d_{1,2}^n$, where $d_{i,j}^n$ is the determinant of the $(n-1) \times (n-1)$ sub-matrix of T_n obtained by removing the i-th row and j-th column. Then, $\det(T_n) = 3d_{1,1}^n + d_{1,2}^n =$

$3\det(T_{n-1}) - \det(T_{n-2})$, where the last equality comes from expanding the determinant $d^n_{1,2}$ along the first column of the appropriate sub-matrix.

Using the Structure Theorem for Finitely Generated Modules over a PID, we can classify the quotient $\mathbb{Z}^n/T_n\mathbb{Z}^n$. For convenience, we let the sequence $(D_n)_{n\in\mathbb{N}}$ denote the bisection of the Fibonacci sequence given by the recurrence $D_{n+1} = 3D_n - D_{n-1}$. Then $D_n = F_{2n}$ for all $n \in \mathbb{Z}_{\geq 0}$.

Let $\varphi : \mathbb{Z}^n \to \mathbb{Z}^n$ be the map given by $\varphi(z) = T_n z$. Then the image of φ is the set of generators of $T_n\mathbb{Z}^n$. The matrix representation of φ with respect to the standard basis $\mathcal{E}$ is T_n. Computing the Smith normal form of T_n reveals bases $\mathcal{B}$ and $\mathcal{C}$ with respect to which the matrix of φ is diagonal. If R and S are the $n \times n$ unimodular matrices,

$$R = \begin{bmatrix} 0 & -1 & 0 & 0 & \dots & 0 \\ 0 & 0 & -1 & 0 & \dots & 0 \\ & & & & \ddots & \\ 0 & 0 & 0 & 0 & \dots & -1 \\ 1 & 3 & 8 & 21 & \dots & D_{n-1} \end{bmatrix} \quad S = \begin{bmatrix} 1 & 3 & 8 & 21 & \dots & D_{n-1} \\ 0 & 1 & 3 & 8 & \dots & D_{n-2} \\ 0 & 0 & 1 & 3 & \dots & D_{n-3} \\ & & & \ddots & & \\ 0 & 0 & 0 & 0 & \dots & 1 \end{bmatrix},$$

then $RT_nS = I_n + (D_n - 1)e_ne_n^T$ is diagonal and by interpreting R and S as change of basis matrices, it follows that $R = [id]_{\mathcal{C},\mathcal{E}}$ and $S = [id]_{\mathcal{E},\mathcal{B}}$ where the bases $\mathcal{B}$ and $\mathcal{C}$ can be expressed in terms of $\mathcal{E}$:

$$\mathcal{B} := \{b_1 = e_1,\ b_2 = 3e_1 + e_2, \dots,\ b_n = D_{n-1}e_1 + D_{n-2}e_2 + \dots + e_n\} \quad \text{and,}$$

$$\mathcal{C} := \{c_1 = 3e_1 - e_2,\ c_2 = 8e_1 - e_3,\ \ \dots, c_{n-1} = D_{n-1}e_1 - e_n,\ c_n = e_1\}.$$

The map φ sends b_i to c_i for $i = 1, \dots, n-1$ and b_n to D_nc_n. By the Structure Theorem, the group $T_n\mathbb{Z}^n$ is isomorphic to $\{\,0\,\} \bigoplus \cdots \bigoplus \{\,0\,\} \bigoplus \mathbb{Z}_{D_n}$ and so the quotient $\mathbb{Z}^n/T_n\mathbb{Z}^n$ is isomorphic to $\mathbb{Z}/D_n\mathbb{Z}$.

To see the latter isomorphism explicitly, for each $x \in \mathbb{Z}^n$, let $x = \sum_{i=1}^n x_ic_i$ give the coordinates of x with the respect to the $\mathcal{C}$ basis. Then the function $\phi([x]) := x_n \mod (D_n)$ is an isomorphism. In particular, ϕ maps the standard basis vectors $\phi([e_i]) = D_{i-1} \bmod D_n$. Hence the lattice walk on $\mathbb{Z}^n/T_n\mathbb{Z}^n$ can be characterized simply as the random walk on $\mathbb{Z}/D_n\mathbb{Z}$ with moves chosen from $\{0, 1, D_2, D_3, \dots, D_{n-1}\}$. The main problem studied in this chapter is inspired by this connection. With this nicer characterization of the lattice walk we can easily derive the following proposition.

Proposition 2.2.8. *The transition matrix for the lattice walk on $\mathbb{Z}^n/T_n\mathbb{Z}^n$ with moves generated from the set $\{\,\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n, \mathbf{0}\,\}$ is the $D_n \times D_n$ circulant matrix whose first row is given by the vector $\frac{1}{n+1}(1, 1, 0, 1, 0, \dots, 0)$ where the nonzero entries are in positions $1 + D_i (\bmod D_n)$*

for $i = 0, 1, 2, \ldots, n$.

The eigenvalues follow from Lemma 2.2.6. Let ξ_{D_n} be a primitive D_n-th root of unity. The eigenvalues $\lambda_1^n, \lambda_2^n, \ldots, \lambda_{D_n}^n$ of the transition matrix for the lattice walk on $\mathbb{Z}^n/T_n\mathbb{Z}^n$ are given by

$$\lambda_k^n = \frac{1}{n+1}\sum_{j=0}^{n} \xi_{D_n}^{k\cdot D_j}. \tag{2.3}$$

2.3 Preliminary Results

This section collects additional results that will be applied to prove results about the linear recurrence random walk. We explicitly state the formula for the eigenvalues of the transition matrix, we recall a theorem about the growth of an integer sequence given by certain linear recurrences, and we state a bounds on the mixing time for Markov chains on groups. More details on the importance of the group structure for analyzing eigenvalues of Markov chains appear in [8].

Lemma 2.3.1. *Let $(P_n)_{n\geq 1}$ be the sequence of transition matrices of the linear recurrence random walk associated to the positive increasing integer sequence $(G_n)_{n\geq 1}$ that satisfies Equation 2.1 and $G_1 = 1$. Let $\xi_{G_n}^k = \exp(\frac{2\pi k}{G_n}\mathbf{i})$ be a primitive G_n-th root of unity, where $\mathbf{i} = \sqrt{-1}$. Then the eigenvalues of P_n are*

$$\lambda_k = \frac{1}{n}\sum_{j=1}^{n} \xi_{G_n}^{kG_j} \quad \textit{for } k = 1, 2, \ldots, G_n. \tag{2.4}$$

Proof. The rows and columns of P_n are indexed by the elements of the cyclic group $\mathbb{Z}_{G_n}$. If we let the vector $\mathbf{g}^{(n)} \in \mathbb{R}^n$ be defined by

$$\mathbf{g}_j^{(n)} = \begin{cases} \frac{1}{n} & \text{if } j-1 \bmod G_n \in \{\,0, 1, G_2, \ldots, G_{n-1}\,\} \\ 0 & \text{otherwise,} \end{cases}$$

then we can let $P_n = \operatorname{circ}(\mathbf{g}^{(n)})$. By Lemma 2.2.6, the eigenvalue λ_k is

$$\lambda_k = \sum_{j=1}^{G_n} \mathbf{g}_{j(\bmod G_n)}^{(n)} \xi_{G_n}^{k\cdot(j-1)}.$$

Since the coordinates of $\mathbf{g}^{(n)}$ are nonzero, in fact they are $\frac{1}{n}$, exactly when $j-1 \in \{\,0, 1, G_2, \ldots, G_{n-1}\,\}$, then $\lambda_k = \frac{1}{n}\sum_{j=1}^{n} \xi_{G_n}^{kG_j}$. ∎

A standard theorem of elementary combinatorics characterizes the solutions of linear recurrence relations (see, e.g. [27, Chapter 4]):

Theorem 2.3.2. *The sequence* $\{G_n\}_{n\geq 1}$ *satisfies*

$$G_n - \alpha_1 G_{n-1} - \alpha_2 G_{n-2} - \cdots - \alpha_d G_{n-d} = 0$$

exactly when for all $n \geq 0$,

$$G_n = \sum_{i=1}^{l} P_i(n)\gamma_i^n$$

where $1 - \alpha_1 x - \alpha_2 x^2 - \cdots - \alpha_d x^d = \prod_{i=1}^{l}(1-\gamma_i x)^{d_i}$, *the* γ_i*'s are distinct and nonzero, and each* $P_i(n)$ *is a polynomial of degree less than* d_i.

A consequence, of which we make frequent use, is that there exists a $\kappa_1 > 0$ such that $\log G_n \leq \kappa_1 n$ for all n. We say that the sequence $(G_n)_{n\in\mathbb{N}}$ exhibits *exponential growth* if there exists $\kappa_2 > 0$ such that $\kappa_2 n \leq \log G_n$ for all sufficiently large n.

The following lemma is a rephrasing of the Upper Bound Lemma which allows us to use a sum involving the eigenvalues of the transition matrix as an approximation for the distance to stationarity at time t.

Lemma 2.3.3 (Upper Bound Lemma, [10]). *Let* P_0^t *be the* t*-step distribution of the linear recurrence random walk associated to* $(G_n)_{n\geq 1}$ *and let* π *be the uniform distribution over* $\mathbb{Z}_{G_n}$. *Then,*

$$\|P_0^t - \pi\|_{TV}^2 \leq \frac{1}{4}\sum_{k=1}^{G_n - 1}|\lambda_k|^{2t},$$

where λ_k*'s are nontrivial eigenvalues of the transition matrix of the random walk.*

Lemma 2.3.3, combined with bounds on the eigenvalues of the transition matrices can be used to get upper bounds on the mixing times of random walks over our finite group. Similarly, lower bounds on the largest nontrivial eigenvalue modulus can give lower bounds on the mixing time:

Lemma 2.3.4. *For the linear recurrence random walk associated to* $(G_n)_{n\geq 1}$ *with transition matrix* P,

$$t_{mix}(\epsilon) \geq (\tfrac{1}{1-\lambda_*} - 1)\log(\tfrac{1}{2\epsilon})$$

where $\lambda_* = \max\{|\lambda| : \lambda$ *is an eigenvalue of* $P, \lambda \neq 1\}$.

Lemma 2.3.4 is a special case of Theorem 1.2.16. Proof that the result of Theorem 1.2.16 holds for reversible, irreducible, aperiodic Markov chains can be found in [20]. As noted in

[15], the proof from [20] also applies to the linear recurrence random walk since P has an orthonormal basis of eigenfunctions with respect to the standard complex inner product $\langle f, g\rangle = \frac{1}{G_n}\sum_{x\in\mathbb{Z}_{G_n}} f(x)\overline{g(x)}$.

2.4 General Linear Recurrences

In this section, we prove bounds on nontrivial eigenvalue moduli for linear recurrence relations of arbitrary order. From this we are able to deduce lower and upper bounds on the mixing time of the Markov chain. In the next section, we specialize to the case of first order linear recurrences, where we are able to prove stronger upper and lower bounds.

The main result of this section is the following:

Theorem 2.4.1. *For the random walk determined by the linear recurrence $\{G_n\}_{n\geq 1}$ with $G_1 = 1$, the mixing time satisfies:*

$$t_{mix}(\epsilon) \leq \kappa n \log(G_n - 1) - \kappa n \log(4\epsilon^2), \quad \textit{where } \kappa = \frac{1}{4 - 4\cos(\frac{\pi}{s+1})}.$$

Note that for large n, there is a constant κ_1 such that $\log(G_n - 1) \leq \kappa_1 n$. So from this bound we have the following corollary.

Corollary 2.4.2. *For the random walk determined by the linear recurrence $\{G_n\}_{n\geq 1}$ with $G_1 = 1$, $t_{mix} \leq \gamma n^2$ for some γ.*

The overall strategy to prove Theorem 2.4.1 is to bound the modulus of the eigenvalues of the transition matrix and then appeal to Lemma 2.3.3. We first establish a few lemmas.

Lemma 2.4.3. *Let $a > 0$ be some real number. If $\theta \in [\frac{2\pi}{a+1}, \frac{2\pi a}{a+1}]$ then*

$$|1 + \exp(\theta\mathbf{i})| \leq |1 + \exp(\tfrac{2\pi\mathbf{i}}{a+1})|.$$

Proof. If $\theta \in [\frac{2\pi}{a+1}, \frac{2\pi a}{a+1}]$ then $\cos(\theta) \leq \cos(\frac{2\pi}{a+1})$ so

$$\begin{aligned}|1 + \exp(\theta\mathbf{i})| &= \sqrt{2 + 2\cos(\theta)} \\ &\leq \sqrt{2 + 2\cos(\tfrac{2\pi}{a+1})} \\ &= |1 + \exp(\tfrac{2\pi\mathbf{i}}{a+1})|.\end{aligned}$$ □

Now for each G_i we identify a subset A_i of $[0, 2\pi]$. Let

$$A_i := \bigcup_{m=0}^{G_i - 1}\left[\frac{2\pi}{(s+1)G_i} + \frac{2\pi m}{G_i}, \frac{2\pi s}{(s+1)G_i} + \frac{2\pi m}{G_i}\right], \text{ where } s = \sum_{j:\alpha_j > 0} \alpha_j.$$

Notice that each A_i satisfies the property that if the angle $\frac{2\pi k}{G_n}$ is in A_i, then $\frac{2\pi k G_i}{G_n} \mod 2\pi \in [\frac{2\pi}{s+1}, \frac{2\pi s}{s+1}]$.

Lemma 2.4.4. *If* $\mathcal{A} = \cup_{i=1}^{n-1} A_i$ *then* $\mathcal{A} = \left[\frac{2\pi}{(s+1)G_{n-1}}, \frac{2\pi((s+1)G_{n-1}-1)}{(s+1)G_{n-1}}\right]$.

Proof. First note that $A_1 = [\frac{2\pi}{(s+1)G_1}, \frac{2\pi s}{(s+1)G_1}]$. Now suppose $\cup_{i=1}^{m} A_i$ is an interval, for some $1 \leq m < n$. Since $G_i \leq G_{i+1}$ and $G_i + 1 \leq sG_{i+1}$ for all i, then inequalities (2.5) and (2.6) hold:

$$\frac{2\pi}{(s+1)G_{i+1}} \leq \frac{2\pi}{(s+1)G_i} \leq \frac{2\pi s}{(s+1)G_{i+1}} \leq \frac{2\pi s}{(s+1)G_i} \tag{2.5}$$

$$\frac{2\pi}{(s+1)G_{i+1}} + \frac{2\pi(G_i-1)}{G_i} \leq \frac{2\pi}{(s+1)G_i} + \frac{2\pi(G_{i+1}-1)}{G_{i+1}} \leq \frac{2\pi s}{(s+1)G_{i+1}} + \frac{2\pi(G_i-1)}{G_i} \leq \frac{2\pi s}{(s+1)G_i} + \frac{2\pi(G_{i+1}-1)}{G_{i+1}}. \tag{2.6}$$

It follows that the first and last intervals in the set A_{m+1} extend the endpoints of the interval $\cup_{i=1}^{m} A_i$. □

Lemma 2.4.5. *The angle* $\frac{2\pi k}{G_n} \mod 2\pi$ *is in* $\mathcal{A} = \cup_{i=1}^{n-1} A_i$ *for each* $k = 1, 2, \ldots, G_n - 1$.

Proof. It suffices to show that $[\frac{2\pi}{G_n}, \frac{2\pi(G_n-1)}{G_n}] \subset \mathcal{A}$. Since $G_n \leq (s+1)G_{n-1}$, then inequality (2.7) holds:

$$\frac{2\pi}{(s+1)G_{n-1}} \leq \frac{2\pi}{G_n} \leq \frac{2\pi(G_n-1)}{G_n} \leq \frac{2\pi s}{(s+1)G_{n-1}} + \frac{2\pi(G_{n-1}-1)}{G_{n-1}}. \tag{2.7}$$

□

Lemma 2.4.6. *For* $n \geq 2$ *and each* $k = 1, 2, \ldots, G_n - 1$, *the eigenvalue modulus* $|\lambda_k|$ *satisfies the following:*

$$|\lambda_k| \leq 1 - \tfrac{2}{n}(1 - |\cos(\tfrac{\pi}{s+1})|) \text{ where } s = \sum_{j:\alpha_j > 0} \alpha_j.$$

Proof. We will show that for each k there exists some $j \in \{1, 2, \ldots, n-1\}$ such that

$$|\xi_{G_n}^{kG_j} + 1| \leq \sqrt{2 + 2\cos(2\pi/s+1)}. \tag{2.8}$$

Then assuming (2.8) holds it follows that

$$\begin{aligned}
|\lambda_k| &= \tfrac{1}{n}|\sum_{i=1}^{n} \xi_{G_n}^{kG_i}| \\
&\leq \tfrac{1}{n}\left(|\xi_{G_n}^{kG_j} + \xi_{G_n}^{kG_n}| + \sum_{i \neq j,n} |\xi_{G_n}^{kG_i}|\right) \\
&\leq \tfrac{1}{n}\left(n - 2 + \sqrt{2 + 2\cos(\tfrac{2\pi}{s+1})}\right)
\end{aligned}$$

$$= 1 - \tfrac{2}{n}\left(1 - |\cos(\tfrac{\pi}{s+1})|\right).$$

Thus it only remains to show that (2.8) holds. By Lemma 2.4.3 it suffices to show that there exists some $j \in \{1, 2, \ldots, n-1\}$ such that $\frac{2\pi k G_j}{G_n} \mod 2\pi$ is in the interval $[\frac{2\pi}{s+1}, \frac{2\pi s}{s+1}]$. By Lemma 2.4.5, the angle $\frac{2\pi k}{G_n} \mod 2\pi$ is in $\mathcal{A}$ therefore we can let j be the integer such that $1 \leq j < n$ and $\frac{2\pi k}{G_n}$ is in $\mathcal{A}_j$. Then we have $\frac{2\pi k G_j}{G_n} \mod 2\pi \in [\frac{2\pi}{s+1}, \frac{2\pi s}{s+1}]$ and hence $|\xi_{G_n}^{kG_j} + 1| \leq \sqrt{2 + 2\cos(\frac{2\pi}{s+1})}$. □

We now prove Theorem 2.4.1.

Proof of Theorem 2.4.1. By Lemma 2.3.3, the distance to stationarity after t steps is less than ϵ when $\sum_{k=1}^{G_n-1} |\lambda_k|^{2t} \leq 4\epsilon^2$. If $\kappa = \frac{1}{4-4\cos(\frac{\pi}{s+1})}$ then by Lemma 2.4.6,

$$\sum_{k=1}^{G_n-1} |\lambda_k|^{2t} \leq \sum_{k=1}^{G_n-1} (1 - \tfrac{1}{2\kappa n})^{2t} \leq (G_n - 1)\exp(-\tfrac{t}{\kappa n}). \tag{2.9}$$

Notice the right hand side of (2.9) is bounded above by $4\epsilon^2$ when $t \geq n\kappa \log(\frac{G_n-1}{4\epsilon^2})$. □

To conclude this section, we prove a lower bound for t_{mix} in the case of general linear recurrences where (G_n) satisfies the exponential growth condition.

Theorem 2.4.7. *For the random walk determined by the linear recurrence $(G_n)_{n\geq 1}$ with $G_1 = 1$, satisfying the exponential growth condition, if $n > 1$*

$$t_{mix}(\epsilon) \geq \frac{n - \gamma \log n}{\gamma \log n} \log(\tfrac{1}{2\epsilon})$$

where γ is some constant.

Proof. We will show that λ_* satisfies the inequality $\lambda_* \geq 1 - \frac{\gamma \log n}{n}$ then appeal to Lemma 2.3.4.

Let $m : \mathbb{N} \to \mathbb{N} \cup \{0\}$ be the function

$$m(n) = \begin{cases} \max_{j\in\{1,\ldots,n-1\}}\{\frac{G_{n-j}}{G_n} > \frac{1}{n}\} & \text{if } \frac{G_{n-1}}{G_n} > \frac{1}{n} \\ 0 & \text{otherwise} \end{cases}$$

Recall that one of the eigenvalues λ_1 has the form:

$$\lambda_1 = \frac{1}{n}\sum_{i=1}^{n} \xi_{G_n}^{G_i}.$$

We will use the function $m(n)$ to give a lower bound on $|\lambda_1|$. The modulus of λ_1 is bounded from below by the real part of λ_1. This real part is

$$\sum_{i=1}^{n} \cos\left(\frac{2\pi G_i}{G_n}\right).$$

We can bound this sum from below to see that

$$|\lambda_1| \geq \frac{1 + (n - m(n) - 1)\cos(\frac{2\pi}{n}) - m(n)}{n}$$

by replacing all summands $\cos\left(\frac{2\pi G_i}{G_n}\right)$ where $G_i/G_n < 1/n$ by $\cos(\frac{2\pi}{n})$ and replacing all summands where $G_i/G_n > 1/n$ by -1.

Further, since $\cos(x) \geq 1 - \frac{x^2}{2}$, it follows that

$$\begin{aligned}
|\lambda_1| &\geq 1 - \frac{2m(n)}{n} - \frac{2\pi^2}{n^2} + \frac{2\pi^2(m(n)+1)}{n^3} \\
&\geq 1 - \frac{2m(n)}{n} - \frac{2\pi^2}{n^2}.
\end{aligned}$$

Now let $\eta_1, \eta_2 > 1$ be constants and p be a polynomial such that $\eta_1^n p(n) \leq G_n \leq \eta_2^n p(n)$ for all n. Then we observe that $\frac{G_{n-j}}{G_n} > \frac{1}{n}$ holds when the inequality $\frac{\eta_1^{(n-j)} p(n-j)}{\eta_2^n p(n)} \geq \frac{1}{n}$ holds.

By rearranging, this occurs when

$$\begin{aligned}
j &< \frac{\log n}{\log \eta_1} + \frac{n(\log \eta_1 - \log \eta_2)}{\log \eta_1} + \log\left(\frac{p(n-j)}{p(n)}\right) && (2.10) \\
&\leq \frac{\log n}{\log \eta_1} && (2.11)
\end{aligned}$$

(since the two dropped terms are negative). It follows that $m(n) \leq \frac{\log n}{\log \eta_1}$ and so

$$\begin{aligned}
|\lambda_1| &\geq 1 - \frac{2\log n}{n \log \eta_1} - \frac{2\pi^2}{n^2} \\
&\geq 1 - \frac{\log n}{n}\left(\tfrac{2}{\log \eta_1} + \tfrac{2\pi^2}{n \log n}\right)
\end{aligned}$$

For $n \geq 2$, the term $\frac{2\pi^2}{n \log n}$ is bounded above by $\frac{\pi^2}{\log 2}$.

$$|\lambda_1| \geq 1 - \frac{\log n}{n}\left(\tfrac{2}{\log \eta_1} + \tfrac{\pi^2}{\log 2}\right).$$

This shows that $\lambda_* \geq 1 - \frac{\gamma \log n}{n}$ where $\gamma = \frac{2}{\log \eta_1} + \frac{\pi^2}{\log 2}$. Then by Lemma 2.3.4, $t_{mix}(\epsilon) \geq$

$\frac{n-\gamma\log n}{\gamma\log n}\log\left(\frac{1}{2\epsilon}\right)$. $\square$

2.5 First Order Recurrences

This section considers sequences generated by first order recurrences $G_n = cG_{n-1}$, that is, geometric series of the form $1, c, c^2, c^3, \ldots$, where $c > 1$ is a positive integer. For these sequences, we show that the order of the mixing time of associated family of random walks is between n and $n\log n$. The main result of this section is the following upper bound on mixing time:

Theorem 2.5.1. *For the random walk determined by the sequence $\{c^{n-1}\}_{n\geq 1}$, where $c > 1$ is an integer,*

$$t_{mix}(\epsilon) \leq \kappa n\log((n-1)(c-1)) - \kappa n\log(\log(4\epsilon^2+1)), \quad \textit{where } \kappa = \frac{1}{1-\cos(\pi/c)}.$$

The easier lower bound will be proven in Theorem 2.5.4 at the end of the section. The key to proving Theorem 2.5.1 will be to exploit the following relationship between the eigenvalues of the n-th random walk and the $(n+1)$-th random walk. Let $\tilde{\lambda}_{n,k}$ denote the k-th unnormalized eigenvalue of the n-th random walk determined by $\{c^{n-1}\}_{n\geq 1}$. That is,

$$\tilde{\lambda}_{n,k} = \sum_{i=1}^{n} \xi_{c^{n-1}}^{kc^{i-1}} = \sum_{i=0}^{n-1} \xi_{c^i}^{k}.$$

Observation. For each $k = 1, 2, \ldots, G_n$, we "lift" the unnormalized eigenvalue $\tilde{\lambda}_{n,k}$ to the set

$$\mathcal{L}_{n,k} = \{\tilde{\lambda}_{n+1,k+jc^{n-1}} : j = 0, 1, \ldots, c-1\}$$

of c unnormalized eigenvalues in the $(n+1)$-th random walk. Each element of $\mathcal{L}_{n,k}$ is equal to $\tilde{\lambda}_{n,k}$ plus some value of the form $\xi_{c^n}^{N}$. That is,

$$\tilde{\lambda}_{n+1,k+jc^{n-1}} = \sum_{i=0}^{n} \xi_{c^i}^{k+jc^{n-1}} = \tilde{\lambda}_{n,k} + \xi_{c^n}^{k+jc^{n-1}}.$$

Over the course of the next two lemmas, we use Observation 2.5 and show that each $|\tilde{\lambda}_{n,k}|$ is bounded above by a value of the form $n + \frac{m}{2}(1-\cos(\frac{\pi}{c}))$, for some $m \in \{0, 1, \ldots, n-1\}$. Once that is established, to prove Theorem 2.5.1 we will apply the Upper Bound Lemma.

Lemma 2.5.2. *Let $c > 1$ be an integer, $z \in \mathbb{C}$, and define sets $\mathcal{A}$ and $\mathcal{B}$ as follows:*

$$\mathcal{A} = \{|z + \exp(\tfrac{2\pi j\mathbf{i}}{c})| : j = 0, 1, \ldots, c-1\}$$

$$\mathcal{B} = \{|z|+1\} \cup \{\sqrt{|z|^2 + 2|z|\cos(\tfrac{(2j-1)\pi}{c}) + 1} : j = 1, 2, \ldots, \lfloor \tfrac{c}{2} \rfloor\}$$

There exists a function $f : \mathcal{A} \to \mathcal{B}$ such that $x \leq f(x)$ for all $x \in \mathcal{A}$.

Proof. Let α be the angle between z and the vector nearest to z from the set $\{\exp(\frac{2\pi j \mathbf{i}}{c}) : j = 0, 1, \ldots, c-1\}$ in the complex plane. So α satisfies the inequality $0 \leq \alpha \leq \frac{\pi}{c}$. We illustrate an example in Figure 2.2.

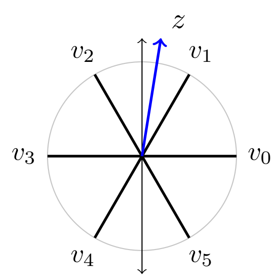

Figure 2.2: Suppose $c = 6$, $v_j = \exp(\pi j \mathbf{i}/3)$, and $z \in \mathbb{C}$ as shown. Then α is the angle between z and v_1. Lemma 2.5.2 gives an upper bound on $|z + v_j|$ for each j.

When c is even,

$$\mathcal{A} = \{ \sqrt{|z|^2 \pm 2|z|\cos(\alpha) + 1}\} \cup \{\sqrt{|z|^2 \pm 2|z|\cos(\tfrac{2j\pi}{c} \pm \alpha) + 1} : j = 1, 2, \ldots, \tfrac{c}{2} - 1\}.$$

We define the function f as follows:

$$f(x) = \begin{cases} |z| + 1 & \text{if } x = \sqrt{|z|^2 + 2|z|\cos(\alpha) + 1}, \\ \sqrt{|z|^2 + 2|z|\cos(\pi - \frac{\pi}{c}) + 1} & \text{if } x = \sqrt{|z|^2 - 2|z|\cos(\alpha) + 1}, \\ \sqrt{|z|^2 + 2|z|\cos(\frac{(2j-1)\pi}{c}) + 1} & \text{if } x = \sqrt{|z|^2 \pm 2|z|\cos(\frac{2j\pi}{c} \pm \alpha) + 1}, \text{ for } 1 \leq j \leq \frac{c}{2} - 1. \end{cases}$$

It is clear that $f(\mathcal{A}) \subset \mathcal{B}$. Now to check that $x \leq f(x)$ for each $x \in \mathcal{A}$ we consider the three cases. First, since $0 \leq \alpha \leq \pi$, then

$$\sqrt{|z|^2 + 2|z|\cos(\alpha) + 1} \leq |z| + 1.$$

Second, since $\pi - \alpha \geq \pi - \frac{\pi}{c}$, then $-\cos(\alpha) = \cos(\pi - \alpha) \leq \cos(\pi - \frac{\pi}{c})$. Hence,

$$\sqrt{|z|^2 - 2|z|\cos(\alpha) + 1} \leq \sqrt{|z|^2 + 2|z|\cos(\pi - \tfrac{\pi}{c}) + 1}.$$

Third, for each $j = 1, 2, \ldots, \frac{c}{2} - 1$, the inequality $\frac{2j\pi}{c} \pm \alpha \geq \frac{(2j-1)\pi}{c}$ holds. Hence,

$$\sqrt{|z|^2 \pm 2|z|\cos(\tfrac{2j\pi}{c} \pm \alpha) + 1} \leq \sqrt{|z|^2 + 2|z|\cos(\tfrac{(2j-1)\pi}{c}) + 1}.$$

When c is odd,

$$\mathcal{A} = \{\ \sqrt{|z|^2 + 2|z|\cos(\alpha) + 1}\} \cup \{\sqrt{|z|^2 \pm 2|z|\cos(\tfrac{2j\pi}{c} \pm \alpha) + 1} : j = 1, 2, \ldots, \tfrac{c-1}{2}\}.$$

In this case we define the function f as

$$f(x) = \begin{cases} |z| + 1 & \text{if } x = \sqrt{|z|^2 + 2|z|\cos(\alpha) + 1}, \\ \sqrt{|z|^2 + 2|z|\cos(\frac{(2j-1)\pi}{c}) + 1} & \text{if } x = \sqrt{|z|^2 \pm 2|z|\cos(\frac{2j\pi}{c} \pm \alpha) + 1}, \text{ for } 1 \leq j \leq \frac{c-1}{2}. \end{cases}$$

By the same arguments used in the even case, $x \leq f(x)$ for all $x \in \mathcal{A}$. □

Notice that Lemma 2.5.2 still holds when we instead define $\mathcal{A} = \{|z + \exp(\frac{2\pi(j+l)\mathbf{i}}{c})| : j = 0, 1, \ldots, c-1\}$, for some fixed integer $l > 0$, since this change corresponds to rotating each $v \in \{\exp(\frac{2\pi j\mathbf{i}}{c}) : j = 0, 1, \ldots, c-1\}$ about the origin through the same fixed angle.

Lemma 2.5.3. *For $n > 1$, define the sets $\mathcal{U}_n$ and $\mathcal{V}_n$ as follows:*

$$\mathcal{U}_n = \{|\tilde{\lambda}_{n,k}| : k = 1, 2, \ldots, c^{n-1}\}$$
$$\mathcal{V}_n = \{n + \tfrac{m}{2}(\cos(\tfrac{\pi}{c}) - 1) : m = 0, 1, \ldots, n-1\}$$

There exists a function $h_n : \mathcal{U}_n \to \mathcal{V}_n$ such that,

1. *$u \leq h_n(u)$ for all $u \in \mathcal{U}_n$, and*
2. *$\#h_n^{-1}(n + \frac{m}{2}(\cos(\frac{\pi}{c}) - 1)) = \binom{n-1}{m}(c-1)^m$, for $m = 0, 1, \ldots, n-1$.*

Proof. Here we use induction. Let $n = 2$. By observation 2.5, the set $\mathcal{U}_2$ is $\{|\tilde{\lambda}_{1,1} + \xi_c^{1+j}| : j = 0, 1, \ldots, c-1\}$. Note that $\tilde{\lambda}_{1,1} = 1$ and

$$\{\xi_c^{1+j} : j = 0, 1, \ldots, c-1\} = \{\exp(\frac{2\pi j\mathbf{i}}{c}) : j = 0, 1, \ldots, c-1\}.$$

So we can let

$$f : \mathcal{U}_2 \to \{2\} \cup \{\sqrt{2 + 2\cos(\tfrac{(2j-1)\pi}{c})} : j = 1, 2, \ldots, \lfloor \tfrac{c}{2} \rfloor\}$$

be as described in proof of Lemma 2.5.2 where $u \leq f(u)$ for all $u \in \mathcal{U}_2$ and define h_2 as follows:

$$h_2(u) = \begin{cases} 2 & \text{if } u \in f^{-1}(2) \\ 2 + \frac{1}{2}(\cos(\frac{\pi}{c}) - 1) & \text{otherwise.} \end{cases}$$

Since $\#f^{-1}(2) = 1$, then $\#h_2^{-1}(2) = 1$ and $\#h_2^{-1}(2 + \frac{1}{2}(\cos(\frac{\pi}{c}) - 1)) = c - 1$, so h_2 satisfies condition (2). For $u \in h_2^{-1}(2)$, the inequality $u \leq h_2(u)$ holds by the triangle inequality. If $u \in h_2^{-1}(2 + \frac{1}{2}(\cos(\frac{\pi}{c}) - 1))$, then $u = |\tilde{\lambda}_{1,1} + \xi_c^{1+j}|$ for some j such that the angle between $\tilde{\lambda}_{1,1}$ and ξ_c^{1+j}, when plotted in the complex plane, is greater than or equal to $\frac{\pi}{c}$. As a consequence of Lemma 2.5.2, $u \leq \sqrt{2 + 2\cos(\frac{\pi}{c})}$. Now

$$2 + 2\cos(\tfrac{\pi}{c}) \leq (2 + \tfrac{1}{2}(\cos(\tfrac{\pi}{c}) - 1))^2$$

since $\frac{1}{4}(\cos(\frac{\pi}{c}) - 1)^2 \geq 0$ and hence h_2 also satisfies condition (1).

Now suppose the Lemma 2.5.3 holds for some $n > 1$. We will define a function

$$h_{n+1} : \{|\tilde{\lambda}_{n+1,k}| : k = 1, 2, \ldots, c^n\} \to \{n + 1 + \tfrac{m}{2}(\cos(\tfrac{\pi}{c}) - 1) : m = 0, 1, \ldots, n\}$$

that satisfies conditions (1) and (2) assuming there exists a function

$$h_n : \{|\tilde{\lambda}_{n,k}| : k = 1, 2, \ldots, c^{n-1}\} \to \{n + \tfrac{m}{2}(\cos(\tfrac{\pi}{c}) - 1) : m = 0, 1, \ldots, n - 1\}$$

that satisfies those conditions.

For each $k = 1, 2, \ldots, c^{n-1}$, let

$$U_{n+1,k} = \{|\tilde{\lambda}_{n+1,k+jc^{n-1}}| : j = 0, 1, \ldots, c - 1\}.$$

Then by Observation 2.5,

$$U_{n+1,k} = \{|\tilde{\lambda}_{n,k} + \xi_{c^n}^{k+jc^{n-1}}| : j = 0, 1, \ldots, c - 1\}$$

and $\mathcal{U}_{n+1} = \cup_{k=1}^{c^{n-1}} U_{n+1,k}$. For each k, the set

$$\{\xi_{c^n}^{k+jc^{n-1}} : j = 0, 1, \ldots, c - 1\} = \{\exp(\tfrac{2\pi k\mathbf{i}}{c^n})\exp(\tfrac{2\pi j\mathbf{i}}{c}) : j = 0, 1, \ldots, c - 1\}$$

is a rotation of the set $\{\exp(\frac{2\pi j\mathbf{i}}{c}) : j = 0, 1, \ldots, c - 1\}$ about the origin in the complex plane. So we can let $|\tilde{\lambda}_{n,k} + \xi_{c^n}^{k+j'c^{n-1}}|$ be an element of $U_{n+1,k}$ such that the vector nearest to $\tilde{\lambda}_{n,k}$ from the set $\{\xi_{c^n}^{k+jc^{n-1}} : j = 0, 1, \ldots, c - 1\}$ is $\xi_{c^n}^{k+j'c^{n-1}}$. Now set

$$h_{n+1}(|\tilde{\lambda}_{n,k} + \xi_{c^n}^{k+j'c^{n-1}}|) = h_n(|\tilde{\lambda}_{n,k}|) + 1$$

and for the remaining $|\tilde{\lambda}_{n,k} + \xi_{c^n}^{k+jc^{n-1}}| \in U_{n+1,k}$, set

$$h_{n+1}(|\tilde{\lambda}_{n,k} + \xi_{c^n}^{k+jc^{n-1}}|) = h_n(|\tilde{\lambda}_{n,k}|) + \tfrac{1}{2}(\cos(\tfrac{\pi}{c}) + 1).$$

By repeating for each k, we define h_{n+1} on all of $\mathcal{U}_{n+1}$.

It remains to show that h_{n+1} satisfies conditions (1) and (2). We first show that $u \leq h_{n+1}(u)$ for all $u \in U_{n+1}$:

For $u \in \mathcal{U}_{n+1}$, $u = |\tilde{\lambda}_{n,k} + \xi_{c^n}^{k+jc^{n-1}}|$ for some $k \in \{1, 2, \ldots, c^{n-1}\}$ and some $j \in \{0, 1, \ldots, c-1\}$. If $h_{n+1}(u) = h_n(|\tilde{\lambda}_{n,k}|) + 1$, then $u \leq h_{n+1}(u)$ by the triangle inequality. On the other hand suppose $h_{n+1}(u) = h_n(|\tilde{\lambda}_{n,k}|) + \frac{1}{2}(\cos(\frac{\pi}{c}) + 1)$ and say $h_n(|\tilde{\lambda}_{n,k}|) = n + \frac{m'}{2}(\cos(\frac{\pi}{c}) - 1)$ for some $0 \leq m' \leq n-1$. Then $|\tilde{\lambda}_{n,k}| \leq n + \frac{m'}{2}(\cos(\frac{\pi}{c}) - 1)$ and $h_{n+1}(u) = n + 1 + \frac{m'+1}{2}(\cos(\frac{\pi}{c}) - 1)$. As a corollary to Lemma 2.5.2,

$$\begin{aligned}
u &\leq \sqrt{|\tilde{\lambda}_{n,k}|^2 + 2|\tilde{\lambda}_{n,k}|\cos(\tfrac{\pi}{c}) + 1} \\
&\leq \sqrt{(n + \tfrac{m'}{2}(\cos(\tfrac{\pi}{c}) - 1))^2 + 2(n + \tfrac{m'}{2}(\cos(\tfrac{\pi}{c}) - 1))\cos(\tfrac{\pi}{c}) + 1} \\
&\leq n + 1 + \tfrac{m'+1}{2}(\cos(\tfrac{\pi}{c}) - 1)
\end{aligned}$$

The last step follows since

$$m'\cos(\tfrac{\pi}{c})(\cos(\tfrac{\pi}{c}) - 1) \leq m'\cos(\tfrac{\pi}{c}) + (n+1)(\cos(\tfrac{\pi}{c}) - 1) + \tfrac{2m'+1}{4}(\cos(\tfrac{\pi}{c}) - 1).$$

Finally we show that $\#h_{n+1}^{-1}(n + 1 + \frac{m}{2}(\cos(\frac{\pi}{c}) - 1)) = \binom{n}{m}(c-1)^m$, for $m = 0, 1, \ldots, n$. By inductive hypothesis $h_n^{-1}(n + \frac{m}{2}(\cos(\frac{\pi}{c}) - 1)) = \binom{n-1}{m}(c-1)^m$, for $m = 0, 1, \ldots, n-1$.

We note that $\#h_{n+1}^{-1}(n+1) = \#h_n^{-1}(n) = 1$ and for m' satisfying $1 \leq m' \leq n$,

$$\begin{aligned}
\#h_{n+1}^{-1}(n + 1 + \tfrac{m}{2}(\cos(\tfrac{\pi}{c}) - 1)) &= \#h_n^{-1}(n + \tfrac{m}{2}(\cos(\tfrac{\pi}{c}) - 1)) + \#h_n^{-1}(n + \tfrac{m-1}{2}(\cos(\tfrac{\pi}{c}) - 1)) \cdot (c-1) \\
&= \binom{n-1}{m}(c-1)^m + \binom{n-1}{m-1}(c-1)^m \\
&= (c-1)^m \binom{n}{m}
\end{aligned}$$

which concludes the proof. □

Proof of Theorem 2.5.1. Recall that $\lambda_k = \frac{1}{n}\sum_{i=1}^{n} \xi_{c^{n-1}}^{kc^{i-1}}$ is the k-th eigenvalue of the n-th random walk. So $|\lambda_k| = \frac{1}{n}|\tilde{\lambda}_{n,k}|$. By Lemma 2.3.3, to find t such that $\|P_0^t - \pi\|_{TV} \leq \epsilon$, it suffices to find t such that $\sum_{k=1}^{c^{n-1}-1} |\lambda_k|^{2t} \leq 4\epsilon^2$.

If $\kappa = \frac{1}{1-\cos(\pi/c)}$ then by Lemma 2.5.3 we have

$$\sum_{k=1}^{c^{n-1}-1} |\lambda_k|^{2t} = \sum_{k=1}^{c^{n-1}-1} \left(\frac{1}{n}|\tilde{\lambda}_{n,k}|\right)^{2t} \leq \sum_{m=1}^{n-1} \binom{n-1}{m} (c-1)^m \left(1 - \tfrac{m}{2\kappa n}\right)^{2t} \tag{2.12}$$

The right hand side of (2.12) can also be bounded above

$$\leq \sum_{m=1}^{n-1} \binom{n-1}{m} (c-1)^m \exp\left(-\tfrac{t}{\kappa n}\right)^m$$

and by the Binomial Theorem,

$$= \left(1 + (c-1)\exp(-\tfrac{t}{\kappa n})\right)^{n-1} - 1 \leq \exp\left((c-1)(n-1)\exp(-\tfrac{t}{\kappa n})\right) - 1. \tag{2.13}$$

Finally, the right hand side of (2.13) $\leq 4\epsilon^2$ when

$$t \geq \kappa n \log((n-1)(c-1)) - \kappa n \log(\log(4\epsilon^2 + 1)). \qquad \square$$

We conclude this section with a lower bound on mixing time.

Theorem 2.5.4. *For the random walk determined by the sequence $\{c^{n-1}\}_{n\geq 1}$, where $c > 1$ is an integer,*

$$t_{mix}(\epsilon) \geq (\gamma n - 1)\log(\tfrac{1}{2\epsilon}), \quad \textit{where } \gamma = \tfrac{1}{1-\cos(2\pi/c)}.$$

Proof. For fixed $n > 1$, the modulus of the $k = c^{n-2}$-th eigenvalue satisfies the inequality

$$|\lambda_{c^{n-2}}| = \tfrac{1}{n}|\xi_c + n - 1| \leq 1 - \tfrac{1-\cos(2\pi/c)}{n}.$$

So $\lambda_* = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } P, \lambda \neq 1\} \geq 1 - \frac{1-\cos(2\pi/c)}{n}$, thus by Lemma 2.3.4,

$$t_{mix}(\epsilon) \geq \left(\tfrac{n}{1-\cos(2\pi/c)} - 1\right)\log(\tfrac{1}{2\epsilon}). \qquad \square$$

2.6 Conclusion

We have shown that the order of the mixing time of random walks determined by a general linear recurrence exhibiting exponential growth is between $n/\log n$ and n^2. A situation that requires further study is the special case where the integer sequence defined by the linear recurrence exhibits polynomial growth instead. This occurs when the characteristic equation of the recurrence is $(1-x)^d$ for some $d \in \mathbb{N}$. For this case, the result and proof of Theorem 2.4.1 still holds and the corresponding upper bound on mixing time is on the order of $n \log n$.

However based on the computations of certain examples, we expect that the true mixing time of these random walks are likely bounded by a function of $\log n$.

Proving mixing times results for sequences of polynomial growth seems to be related to some classic problems in number theory. For example, consider the following special case:

For fixed $k \in \mathbb{N}_{>1}$ and $n > 1$ ranging, describe the mixing behavior of the random walk $(X_t)_{t\geq 0}$ with state space $\mathcal{S} = \mathbb{Z}_{n^k}$, initial state $X_0 = 0$, and where from the current state X_t, the next state is given by

$$X_{t+1} \equiv X_t + z^k \mod n^k,$$

with z chosen from the set $\mathcal{M} = \{1, 2, \ldots, n\}$ uniformly at random.

The Hilbert-Waring theorem [13] (which says that there is a function $g(k)$ such that every nonnegative integer can we written as a sum of at most $g(k)$ k-th powers) guarantees that this Markov chain has a bounded diameter for all n. The mixing time of the Markov chain appears to be related to the problem of determining the number of ways that a number can be written as the sum of l k-th powers. This has complicated relations to theta functions.

Chapter 3

Sampling Lattice Points of Polytopes via Fiber Walks

3.1 Motivation

We now turn our attention to the problem of sampling lattice points of polytopes. The motivation arises from independence testing in statistics where the task is as follows: Imagine we would like to explore the relationship between two categorical variables and determine if they are dependent on each other. To test this we may collect data from a sample population and from perform a hypothesis test of independence. Example 3.1.1 details this process.

Example 3.1.1. Suppose we have randomly polled 120 people on their favorite color and favorite board game, and then recorded and organized the count data into the contingency table seen in Table 3.1.

The rows of Table 3.1 are indexed by the games $G =$ (Monopoly, Catan, Scrabble, Clue) and the columns are indexed by colors $C =$ (Red, Blue, Green, Black). The entry T^o_{ij} of the table is the number of people whose favorite game is g_i and whose favorite color c_j. The null hypothesis that we seek to test is that a person's favorite color is independent of their favorite board game. Assuming the null hypothesis is true, if a person is randomly chosen then the probability that their favorite game is g_i and favorite color is c_j is given by $\Pr(g_i, c_j) = \Pr(g_i)\Pr(c_j)$.

Given the null hypothesis, the maximum likelihood estimation of the contingency table is

$$T^e_{ij} = \frac{1}{N}\sum_{k=1}^{|C|} T^o_{ik} \sum_{k=1}^{|G|} T^o_{kj}.$$

It is clear that the observed table T^o and the expected table T^e are not identical, and we would not expect them to be even if the null hypothesis is true, due to the randomness that

Table 3.1: The 4×4 contingency table organizes the observed data from a random sample from Example 3.1.1.

Game \ Color	Red	Blue	Green	Black	Total
Monopoly	12	11	6	7	36
Catan	10	8	10	9	37
Scrabble	6	13	5	11	35
Clue	3	5	2	2	12
Total	31	37	23	29	120

Table 3.2: Expected contingency table if a person's favorite color and favorite board game are independent

Game \ Color	Red	Blue	Green	Black	Total
Monopoly	9.3	11.1	6.9	8.7	36
Catan	9.56	11.41	7.09	8.94	37
Scrabble	9.04	10.79	6.71	8.46	35
Clue	3.1	3.7	2.3	2.9	12
Total	31	37	23	29	120

arises from sampling. The question to consider is whether the difference between the two tables is significant. To this end, we can use the volume test, introduced by Diaconis and Efron [9] as an alternative to the classical chi-square test for independence. For the volume test, we compute the chi-square statistic $\chi^2(x)$ on the observed table to obtain a measure of the distance between T^o and T^e, that is,

$$\chi^2(T^o) = \sum_{i,j} \frac{(T^o_{ij} - T^e_{ij})^2}{T^e_{ij}}.$$

We then consider all tables $T \in \mathbb{N}^{4\times 4}$ whose row and column sums are $(36, 37, 35, 12)$ and $(31, 37, 23, 29)$, respectively, and for each compute the statistic $\chi^2(T)$. The significance level of the volume test is the proportion of tables T satisfying $\chi^2(T) \geq \chi^2(T^o)$ to all tables with the same margins.

Using `LattE integrale` [2] to count, we see that there are approximately 9.35 billion non-negative integer tables whose row sums are $(36, 37, 35, 12)$ and column sums are $(31, 37, 23, 29)$, and in practice, it is infeasible to enumerate each of these for the purposes of computing the chi-square statistic for each. What we do instead is generate a uniform sample of these contingency tables to approximate the significance level.

Recall in Example 1.3.19 we saw that the set of contingency tables with specified margins are exactly the integer points $P \cap \mathbb{Z}^n$ of a certain polytope P. Chapters 3 and 4 are concerned with finding efficient methods for sampling near-uniformly from the set of lattice points of general polytopes.

The approach in this chapter is to define a graph structure on a set $\mathcal{F} = P \cap \mathbb{Z}^n$ for some polytope P where edges of the graph correspond to a finite set of moves $\mathcal{M}$. With the graph in hand, a variation of a random walk on the graph with uniform stationary distribution is then implemented as a means to sample from the set $\mathcal{F}$. In Section 3.2 we discuss a general graph construction where lattice points are vertices and edges correspond to moves. In Section 3.3 we define the simple walk on that graph. Though the particular walk is a natural one to define, we show that in certain instances the mixing behavior is slow. Finally, in Section 3.4, we explore the heat-bath random walk which can be thought of as the discrete analog of the hit-and-run random wall which is often used to sample from general convex, continuous sets. The hit-and-run algorithm is discussed in greater detail in Chapter 4. We will see that moving to the heat-bath random walk does not necessarily improve the mixing behavior.

3.2 Fiber Graphs

The discussion and results presented in the remaining sections of this chapter are the result of joint work with Tobias Windisch. The project began with providing an alternate proof of a result in [28] stating that the simple walk, a Markov chain defined on a fiber graph, is not rapidly mixing. In this section we formally define fiber graphs.

Definition 3.2.1. Let $\mathcal{F} \subset \mathbb{Z}^n$ be a finite set and let $\mathcal{M} \subset \mathbb{Z}^n$ be a set of moves. The graph $\mathcal{F}(\mathcal{M})$ has vertex set $\mathcal{F}$ and two nodes $x, y \in \mathcal{F}$ are adjacent if $x - y \in \mathcal{M}$ or $y - x \in \mathcal{M}$.

Definition 3.2.2. Let $\mathcal{F} \subset \mathbb{Z}^n$ and $\mathcal{M} \subset \mathbb{Z}^n$ be finite sets. If the graph $\mathcal{F}(\mathcal{M})$ is connected then $\mathcal{M}$ is a *Markov basis* for $\mathcal{F}$. When $\mathcal{P}$ is a collection of finite subsets of $\mathbb{Z}^n$, we say that $\mathcal{M}$ is a Markov basis for $\mathcal{P}$, if for all $\mathcal{F} \in \mathcal{P}$, the graph $\mathcal{F}(\mathcal{M})$ is a connected.

Example 3.2.3. Let $\mathcal{F}_d = [4] \times [d]$ be the rectangular grid and let $\mathcal{P} = \{F_d : d \in \mathbb{N}\}$. The set $\mathcal{M}_1 = \{(0,1),(1,0)\}$ is a Markov basis for $\mathcal{P}$. On the other hand the set $\mathcal{M}_2 = \{(1,1)\}$ is not. In Figure 3.1 we fix $d = 3$ and display both $\mathcal{F}_3(\mathcal{M}_1)$ and $\mathcal{F}_3(\mathcal{M}_2)$.

If there exists a polytope $P \subset \mathbb{R}^n$ such that $\mathcal{F} = P \cap \mathbb{Z}^n$ then the set $\mathcal{F}$ is *normal*. Definition 3.2.5 introduces a particular type of normal set.

Definition 3.2.4. For a matrix $A \in \mathbb{Z}^{m \times n}$, the set $\mathbb{N}A$ consists of all nonnegative integer combinations of the columns of A. That is, $\mathbb{N}A := \{Az : z \in \mathbb{Z}^n_{\geq 0}\}$.

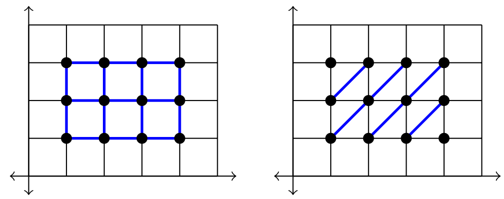

Figure 3.1: On the left is $\mathcal{F}_3(\mathcal{M}_1)$ and on the right is $\mathcal{F}_3(\mathcal{M}_2)$ from Example 3.2.3.

Definition 3.2.5. Let $A \in \mathbb{Z}^{m\times n}$ be a matrix and $b \in \mathbb{N}A$ a vector. The *b-fiber of A* is the set $\mathcal{F}_{A,b} := \{x \in \mathbb{N}^n \mid Ax = b\}$. The collection of all b-fibers of A is denoted $\mathcal{P}_A := \{\mathcal{F}_{A,b} \mid b \in \mathbb{N}A\}$.

The b-fiber of $A \in \mathbb{Z}^{m\times n}$ is normal since it is the set of integer points of the polytope

$$P = P\left(\begin{bmatrix} A \\ -A \\ -I_n \end{bmatrix}, \begin{bmatrix} b \\ -b \\ \mathbf{0}_n \end{bmatrix}\right).$$

Definition 3.2.6. Let $A \in \mathbb{Z}^{m\times n}$ be a matrix, $b \in \mathbb{N}A$ a vector, and $\mathcal{M} \subset \mathbb{Z}^n$ a set of moves. The graph $\mathcal{F}_{A,b}(\mathcal{M})$ is called a *fiber graph*.

Example 3.2.7. Suppose we have the matrix $A \in \mathbb{Z}^{2\times 5}$, the vector $b \in \mathbb{Z}^2$, and set $\mathcal{M} \subset \mathbb{Z}^5$ defined as follows:

$$A = \begin{bmatrix} 1 & 3 & 1 & 4 & 5 \\ 2 & -1 & 1 & 0 & 3 \end{bmatrix}, \quad b = \begin{bmatrix} 18 \\ 5 \end{bmatrix},$$

$$\text{and } \mathcal{M} = \{\mathbf{m}_1 = (2,1,0,0,-1)^T, \mathbf{m}_2 = (0,1,1,-1,0)^T, \mathbf{m}_3 = (2,3,2,-2,-1)^T\}.$$

Then the b-fiber of A is the set,

$$\begin{aligned} \mathcal{F}_{A,b} &= \{\, x_1, x_2, \ldots, x_8 \,\} \\ &= \left\{ \begin{bmatrix} 3 \\ 4 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 0 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 3 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 3 \\ 1 \end{bmatrix} \right\}. \end{aligned}$$

Figure 3.2 shows the fiber graph $\mathcal{F}_{A,b}(\mathcal{M})$.

For a matrix $A \in \mathbb{Z}^{m\times n}$, there is at least one set of moves that is a Markov basis for

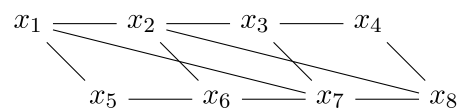

Figure 3.2: The fiber graph from Example 3.2.7

the collection $\mathcal{P}_A$ of b-fibers and thus we can always construct a connected fiber graph. With the next group of definitions we introduce the Graver basis. Later we will see some properties satisfied by the Graver basis. For more complete details on the Graver basis and its applications, we refer the reader to [1].

Definition 3.2.8. The *integer kernel* of a matrix $A \in \mathbb{Z}^{m\times n}$ is the set

$$\ker_{\mathbb{Z}}(A) := \{ x \in \mathbb{Z}^n : Ax = 0 \}.$$

If $x \in \ker_{\mathbb{Z}}(A)$ then we call x a *move* of the matrix A.

Definition 3.2.9. For a point $x \in \mathbb{Z}^n$, the *support of* x, denoted $\operatorname{supp}(x) = \left\{ i : x(i) \neq 0 \right\}$ is the set of indices corresponding to nonzero coordinates. The *positive part of* x, denoted x^+, is the point given coordinate-wise by $(x^+)_i := \max(0, x_i)$. Analogously, each coordinate of the *negative part of* x, is $(x^-)_i := \max(0, -x_i)$.

Definition 3.2.10. The sum $x = x_1 + x_2 + \cdots + x_k$ is *conformal* if $\operatorname{supp}(x^+) = \bigcup_{j=1}^k \operatorname{supp}(x_j^+)$ and $\operatorname{supp}(x^-) = \bigcup_{j=1}^k \operatorname{supp}(x_j^-)$.

For example, $(1, 0, -2, -2, 2) = (1, 0, -1, -1, 1) + (0, 0, -1, -1, 1)$ is a conformal sum while $(1, 1, 0, 0, 0) = (1, 0, -1, -1, 1) + (0, 1, 1, 1, -1)$ is not.

Definition 3.2.11. Let x be a move of A. We say that x is *(conformally) primitive* if there does not exist two nonzero moves y and z of A such that $x = y + z$ is a conformal sum.

Definition 3.2.12. The *Graver basis of* $A \in \mathbb{Z}^{m\times n}$, denoted $\mathcal{G}_A$, is the set of conformally primitive moves of A.

Example 3.2.13. Suppose $\mathcal{P}_{m,n}(r, c)$ is the set of $m \times n$ contingency tables with row sums r and column sums c. If we let A be the configuration matrix for $m \times n$ contingency tables (recall Remark 1.3.20) then $\mathcal{P}_{m,n}(r, c)$ represents the nonnegative points in the (r, c)-fiber of the configuration matrix A.

For $2 \leq k \leq \min\{ m, n \}$, let $i_{[k]} = (i_1, \ldots, i_k)$ be a vector of distinct row indices and $j_{[k]} = (j_1, \ldots, j_k)$ be a vector of distinct column entries. A *loop of degree* k is a move $z_k(i_{[k]}, j_{[k]}) \in$

$\{0, \pm 1\}^{m\times n}$ where the nonzero entries are

$$z_{i_1 j_1} = z_{i_2 j_2} = \cdots = z_{i_k j_k} = 1$$
$$z_{i_1 j_2} = z_{i_2 j_3} = \cdots = z_{i_k j_1} = -1.$$

Then the set of loops of degree k, where $2 \leq k \leq \min\{m, n\}$ forms the Graver basis for A. For proof see Section 4.6 of [1].

Proposition 3.2.14. *For $A \in \mathbb{Z}^{m\times n}$, the Graver basis $\mathcal{G}_A$ is finite. Further, for any $x, y \in \mathbb{Z}^n$ such that $Ax = Ay$, there exists moves $g_1, \ldots, g_k \in \mathcal{G}_A$ and constants $\kappa_j \in \{\pm 1\}$ such that $x - y = \sum_{j=1}^{k} \kappa_j g_j$.*

Proof. If $x, y \in \mathbb{Z}^n$ such that $Ax = Ay$, then $(x - y)$ is a move of A. So either $(x - y)$ is itself a primitive move or can be recursively decomposed as a conformal sum of primitive moves. For proof that $\mathcal{G}_A$ is finite, see Section 5.4.3 in [1] of Hilbert Basis Theorem. $\square$

By Proposition 3.2.14, if we have x and y in the set of lattice points $\mathcal{F} = \{x \in \mathbb{Z}^n : Ax = b\}$, then using a sequence of moves in the Graver basis of A we can walk from x and y. Later with Proposition 3.3.9, we also see that the sequence of moves can be chosen so that at each step the walk remains in $\mathcal{F}$, thus proving that $\mathcal{G}_A$ is a Markov basis.

3.3 Simple Fiber Walk

By implementing a random walk on the graph $\mathcal{F}_{A,b}(\mathcal{M})$ we can explore and sample from the set $\mathcal{F}_{A,b}$. The simple walk, which we formally define for arbitrary $\mathcal{F} \subset \mathbb{Z}^n$ next, is the random walk on $\mathcal{F}$ where from the current state x, a move $\mathbf{m} \in \mathcal{M}$ is chosen at random. If $x + \mathbf{m} \in \mathcal{F}$ then the chain moves there. Otherwise the chain remains at x. The simple walk is a slight variation of the random walk on a graph described in Section 1.1.1 in that the probability of remaining at a point x can be positive even if $\mathbf{0} \notin \mathcal{M}$.

Definition 3.3.1. Let $\mathcal{F} \subset \mathbb{Z}^n$ and $\mathcal{M} \subset \mathbb{Z}^n$ be two finite sets with $\mathbf{0} \notin \mathcal{M}$. The *simple walk* is the Markov chain with state space $\mathcal{F}$ whose transition probabilities are given as follows:

$$P(x, y) = \begin{cases} \frac{1}{|\pm\mathcal{M}|} & \text{if } x - y \in \pm\mathcal{M}, \\ \frac{|\{\mathbf{m}\in\pm\mathcal{M} : x+\mathbf{m}\notin\mathcal{F}_{A,b}\}|}{|\pm\mathcal{M}|} & \text{if } x = y \\ 0 & \text{otherwise} \end{cases}$$

for all $x, y \in \mathcal{F}$.

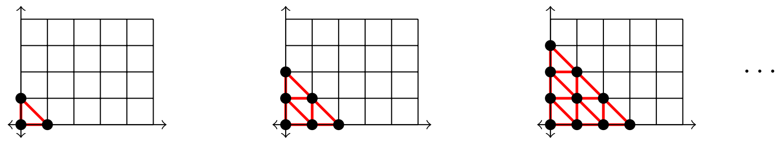

Figure 3.3: A sequence of fiber graphs $\mathcal{F}_{A,ib}(\mathcal{M})$ where $i \in \mathbb{N}$ ranges.

If $\mathcal{M}$ is a Markov basis for $\mathcal{F}$ then the simple walk is irreducible. Since $\mathcal{F}$ is finite, there must exist some $x \in \mathcal{F}$ and $\mathbf{m} \in \mathcal{M}$ such that $x + \mathbf{m} \notin \mathcal{F}$ and so the simple walk is also aperiodic. Also notice that the transition matrix is symmetric since, for distinct $x, y \in \mathcal{F}$, if $\pm(x-y) \notin \mathcal{M}$ then $P(x,y) = 0$. Otherwise if x and y are connected by a move in $\mathcal{M}$ then $P(x,y) = P(y,x) = \frac{1}{|\pm\mathcal{M}|}$. Thus when $\mathcal{M}$ is a Markov basis, the stationary distribution π to which the simple walk converges is uniform on $\mathcal{F}$.

The simple walk provides one way to sample points from $\mathcal{F}$. To actually implement this random walk requires computationally simple means of randomly generating moves from $\mathcal{M}$ and checking if an arbitrary point x is in $\mathcal{F}$. If $\mathcal{F}$ is a normal set expressed as a system of linear inequalities then the latter can be checked efficiently with matrix- vector products. What remains is to focus on the mixing behavior of the simple walk. The main theorem in this chapter states that the simple walk on a sequence of normal sets $\{ \mathcal{F}_i \}$, arising as the set of lattice points of an integer dilation of a fixed polytope P, does not mix rapidly.

Theorem 3.3.2. *Let $A \in \mathbb{Z}^{m\times n}$ with $ker_{\mathbb{Z}}(A) \cap \mathbb{Z}^n_{\geq 0} = \{0\}$, let $b \in \mathbb{N}A$, and let $\mathcal{M}$ be a Markov basis for $\mathcal{P}_A$. The simple walk on $\{\mathcal{F}_{A,ib}(\mathcal{M})\}_{i\in\mathbb{N}}$ is not rapidly mixing.*

The proof strategy relies on the growth of the diameter of the underlying graphs of the random walk, which happens to coincide with the fiber graph. We will see that the diameter grows linearly then relate the diameter to the conductance of the random walk.

For a finite set $\mathcal{M} \subset \mathbb{Z}^n$ and any norm $\|\cdot\|$ on $\mathbb{R}^n$, let $\|\mathcal{M}\| := \max_{m\in\mathcal{M}}\|m\|$.

Lemma 3.3.3. *Let $\mathcal{F} \subset \mathbb{Z}^n$ and $\mathcal{M} \subset \mathbb{Z}^n$ be finite sets, then*

$$diam(\mathcal{F}(\mathcal{M})) \geq \frac{1}{\|\mathcal{M}\|} \max\left\{ \|x-y\| : x, y \in \mathcal{F} \right\}.$$

Proof. Let $x', y' \in \mathcal{F}$ such that $\left\|x'-y'\right\| = \max\left\{ \|x-y\| : x,y \in \mathcal{F} \right\}$ and let $p = \sum_{i=1}^r \mathbf{m}_i$ be a minimal path of length r in $\mathcal{F}(\mathcal{M})$ from x' to y'. Hence $y' = x' + \sum_{i=1}^r \mathbf{m}_i$. Then by the triangle inequality $\left\|x'-y'\right\| = \|p\| \leq \sum_{i=1}^r \|\mathbf{m}_i\|$ moreover $\left\|x'-y'\right\| \leq \sum_{i=1}^r \|\mathbf{m}_i\| \leq r\|\mathcal{M}\|$. The result follows since r is the graph distance $d(x', y')$ and so $r \leq \mathrm{diam}(\mathcal{F}(\mathcal{M}))$. $\square$

Intuitively, Lemma 3.3.3 says that the number of edges in a path between $x, y \in \mathcal{F}$ is at least $\|x-y\|$ divided by size of the largest step in $\mathcal{M}$. Using Lemma 3.3.3 we show that the diameter of the fiber graph $\mathcal{F}_i(\mathcal{M})$ where $\mathcal{F}_i = (i \cdot P \cap \mathbb{Z}^n)$ grows at least linearly with i.

Definition 3.3.4. A polytope $P \subset \mathbb{R}^n$ is *rational* if all of its vertices have rational coordinates.

Proposition 3.3.5. *Let $P \subset \mathbb{R}^n$ be a rational polytope with $|P \cap \mathbb{Z}^n| > 1$ and let $\mathcal{M}$ be a Markov basis for $\mathcal{F}_i = (i \cdot P) \cap \mathbb{Z}^n$. Then there exists a constant $C \in \mathbb{Q}_{>0}$ such that $C \cdot i \leq diam(\mathcal{F}_i(\mathcal{M}))$ for all $i \in \mathbb{N}$.*

Proof. Since $|P \cap \mathbb{Z}^n| > 1$ we can choose distinct x', y' in $P \cap \mathbb{Z}^n$. Then for all $i \in \mathbb{N}$, the points ix' and iy' are in $\mathcal{F}_i$ and we have that $\|ix' - iy'\| = i\|x' - y'\| \leq \max\left\{ \|x-y\| : x, y \in \mathcal{F}_i \right\}$. By Lemma 3.3.3 it follows that $\operatorname{diam}(\mathcal{F}_i(\mathcal{M})) \geq i\frac{\|x'-y'\|}{\|\mathcal{M}\|}$. $\square$

Though it is not necessary for the proof of Theorem 3.3.2, we show that if the set $\mathcal{M}$ satisfies further conditions then the diameter of $\mathcal{F}_i(\mathcal{M})$ grows at most linearly in i.

Definition 3.3.6. Let $\mathcal{P}$ be a collection of finite subsets of $\mathbb{Z}^n$. A finite set $\mathcal{M} \subset \mathbb{Z}^n$ is *norm-like* if there exists a constant $C \in \mathbb{N}$ such that for all $\mathcal{F} \in \mathcal{P}$ and for all $x, y \in \mathcal{F}$, $d(x, y) \leq C\|x - y\|$. The set $\mathcal{M}$ is $\|\cdot\|$*-norm-reducing* for $\mathcal{P}$ if for all $\mathcal{F} \in \mathcal{P}$ and all $x, y \in \mathcal{F}$ there exists $\mathbf{m} \in \mathcal{M}$ such that $x + \mathbf{m} \in \mathcal{F}$ and $\|x + \mathbf{m} - y\| < \|x - y\|$.

The property of being norm-like does not depend on the norm whereas being norm-reducing does. Norm-reducing sets are always norm-like and norm-like sets are Markov bases, since for any $\mathcal{F}$ and $x, y \in \mathcal{F}$ the distance $d(x, y) < \infty$. The converse of each statement is not true in general.

Example 3.3.7. For $i \in \mathbb{N}$ consider the normal set $\mathcal{F}_i = \{[2] \times [i] \times \{0\}\} \cup \left\{ (2, i, 1) \right\}$ along with the Markov basis $\mathcal{M} = \left\{ (0, 1, 0), (0, 0, 1), (1, 0, 1) \right\}$. Then the diameter of $\mathcal{F}_i(\mathcal{M})$ is equal to the distance $d((1, 1, 0), (2, 1, 0)) = 2i$. Hence $\mathcal{M}$ is a Markov basis for $\{ \mathcal{F}_i : i \in \mathbb{N} \}$ but it is not norm-like.

Example 3.3.8. Suppose $P \subset \mathbb{R}^2$ is the polytope given by the system of inequalities $x \geq 0, y \geq 0$, and $x + y \leq 2$ and $\mathcal{M} = \left\{ \mathbf{m}_1 = (1, -1), \mathbf{m}_2 = (0, 1) \right\}$. If we let $\mathcal{F}_i = (i \cdot P) \cap \mathbb{Z}^2$ then $\mathcal{M}$ is a Markov basis for $\{ \mathcal{F}_i : i \in \mathbb{N} \}$. Notice that $\mathcal{M}$ is not $\|\cdot\|_1$- norm-reducing since, in particular, for $x = (0, 0)$ and $y = (1, 0)$, any move from x increases the 1- norm distance to y. The set $\mathcal{M}$ is however norm-like, as we can show that for any $i \in \mathbb{N}$ and any $x, y \in \mathcal{F}_i$ the distance $d(x, y)$ in $\mathcal{F}_i(\mathcal{M})$ is at most $2\|x - y\|_1$.

Proposition 3.3.9. *For $\mathbb{A} \in \mathbb{Z}^{m \times n}$, the Graver basis $\mathcal{G}_A$ is $\|\cdot\|_1$-norm-reducing for the collection $\mathcal{P}_A = \left\{ \mathcal{F}_{A,b} : b \in \mathbb{N}A \right\}$.*

Proof. Suppose $x, y \in \mathcal{F}_{A,b}$ are points in the same b-fiber of A. Then the difference $(x-y)$ is a move of A and we can write $x - y = g_1 + \cdots g_k$ as a conformal sum of nonzero elements of $\mathcal{G}_A$. As the sum is conformal, the support of the positive parts and the negative parts are compatible and so it follows that

$$\|x-y\|_1 = \left\|\sum_{j=1}^{k} g_j\right\|_1 = \sum_{j=1}^{k} \|g_j\|_1 .$$

The point $(x - g_1)$ remains in $\mathcal{F}_{A,b}$ and we can show that the sum $(x-y) - g_1 = \sum_{j=2}^{k} g_j$ is also conformal: First note $\mathrm{supp}(((x-y)-g_1)^+) \subseteq \mathrm{supp}((x-y)^+)$ since $x - y = g_1 + \cdots g_k$ is a conformal sum. If the index $l \in \mathrm{supp}((x-y)^+)$ but $l \notin \mathrm{supp}(((x-y)-g_1)^+)$, then $(x-y)_l = (g_1)_l$. This means that $\mathrm{supp}(((x-y)-g_1)^+) \subseteq \bigcup_{j=2}^{k} \mathrm{supp}(g_j^+)$, moreover, if index $l \in \bigcup_{j=2}^{k} \mathrm{supp}(g_j^+)$ then $(x-y)_l > (g_1)_l$ and so $\mathrm{supp}(((x-y)-g_1)^+) = \bigcup_{j=2}^{k} \mathrm{supp}(g_j^+)$. Using a similar argument, $\mathrm{supp}(((x-y)-g_1)^-) = \bigcup_{j=2}^{k} \mathrm{supp}(g_j^-)$. Therefore,

$$\|x-(y+g_1)\|_1 = \left\|\sum_{j=2}^{k} g_j\right\|_1 = \sum_{j=2}^{k} \|g_j\|_1 < \|x-y\|_1 .$$

□

Proposition 3.3.10. *Let $P \subset \mathbb{R}^n$ be a rational polytope with $|P \cap \mathbb{Z}^n| > 1$ and let $\mathcal{M}$ be a Markov basis for $\mathcal{F}_i = (i \cdot P) \cap \mathbb{Z}^n$. If $\mathcal{M}$ is norm-like for $\{\, \mathcal{F}_i : i \in \mathbb{N} \,\}$, then there exists a constant $C \in \mathbb{Q}_{>0}$ such that diam$(\mathcal{F}_i(\mathcal{M})) \leq C \cdot i$ for all $i \in \mathbb{N}$.*

Proof. If $\mathcal{M}$ is norm-like then there exists a constant C such that for all $i \in \mathbb{N}$,

$$\mathrm{diam}(\mathcal{F}_i(\mathcal{M})) = \max_{x,y \in \mathcal{F}_i} d(x,y) \leq C \max_{x,y\in\mathcal{F}_i} \|x-y\| .$$

Now it suffices to show that there exists a constant C_0 such that $\max_{x,y\in\mathcal{F}_i} \|x-y\| \leq C_0 \cdot i$ for all $i \in \mathbb{N}$. Let $v_1, \ldots v_r \in \mathbb{Q}^n$ such that $P = \mathrm{conv}(v_1, \ldots, v_r)$ and define $C_0 = \max\left\{\, \|v_s - v_t\| : s \neq t \,\right\}$. Since $\mathcal{F}_i = (i \cdot P) \cap \mathbb{Z}^n \subset \mathrm{conv}(iv_1, \ldots, iv_r)$ for all $i \in \mathbb{N}$, we have $\max\left\{\, \|x-y\| : x,y \in \mathcal{F}_i \,\right\} \leq \max\left\{\, \|iv_s - iv_t\| : s \neq t \,\right\} = C_0 \cdot i$. □

Proposition 3.3.11. *Let $A \in \mathbb{Z}^{m\times n}$ with $\ker_{\mathbb{Z}}(A) \cap \mathbb{N}^n = \{0\}$, $b \in \mathbb{N}A$, and $\mathcal{M}$ be a Markov basis for $\mathcal{P}_A$. There exists constants $C, C' \in \mathbb{Q}_{>0}$ such that*

$$C' \cdot i \leq \mathit{diam}(\mathcal{F}_{A,ib}(\mathcal{M})) \leq C \cdot i.$$

Proof. The lower bound follows from Proposition 3.3.5. For the upper bound, we will show that

$\mathcal{M}$ is norm-like for $\mathcal{P}_A$. Then the upper bound follows from Proposition 3.3.10.

By Proposition 3.3.9, the Graver basis $\mathcal{G}_A$ for A is a finite set which is $\|\cdot\|_1$- norm-reducing for $\mathcal{P}_A$. Let $C_0 = \max_{g \in \mathcal{G}_A} \operatorname{diam}(\mathcal{F}_{A,Ag^+}(\mathcal{M}))$, where g^+ is the positive part of g. Now pick $x, y \in \mathcal{F}_{A,b}$ arbitrarily and let $x = y + \sum_{j=1}^r g_j$ be a walk from x to y in $\mathcal{F}_{A,b}(\mathcal{G}_A)$ of minimal length. Since the Graver basis is norm-reducing for $\mathcal{F}_{A,b}$ there always exists a path of length at most $\|x-y\|_1$ and hence $r \leq \|x-y\|_1$. Every g_j can be replaced by a path in $\mathcal{F}_{A,Ag_j^+}(\mathcal{M})$ of length at most C_0 and these paths stay in $\mathcal{F}_{A,b}$. This gives a path of length $C_0 \cdot r$, hence the graph distance in $\mathcal{F}_{A,b}(\mathcal{M})$ from x to y satisfies $d(x,y) \leq C_0\|x-y\|_1$. □

We conclude this section with the proof of Theorem 3.3.2 which states that the simple walk on $\mathcal{F}_{A,ib}(\mathcal{M})$ is not rapidly mixing. As mentioned previously the proof strategy takes the lower bound on the diameter of the fiber graph and relates it to the conductance of the simple walk.

Proof of Theorem 3.3.2. The lower bound from Proposition 3.3.11 combined with Lemma 1.2.20 imposes an upper bound on the conductance of the simple walk on $\mathcal{F}_{A,ib}(\mathcal{M})$, namely

$$\Phi_* \leq \exp\left(\frac{2\log|\mathcal{F}_{A,ib}|}{C' \cdot i}\right) - 1$$

for some constant C'. This upper bound on conductance along with Cheeger's inequality implies a lower bound on the SLEM of the form $\lambda_* \geq 1 - 2\Phi_*$. By Ehrhart's theory, the number of integer points in the i-th dilation of a rational polytope is given by a quasi-polynomial in i. More specifically, we have $|\mathcal{F}_{A,ib}| \in \Omega(i^n)$, where n is the dimension of the polytope (see Section 3.7 in [4]). What follows is the bound on the SLEM

$$\lambda_* \geq 1 - 2\left[\exp\left(\frac{2\log(Ci^n)}{C' \cdot i}\right) - 1\right]$$

for some constants C and C'. By Definition 1.2.17, since λ_* approaches 1 quickly relative to the size of the state space $|\mathcal{F}_{A,ib}|$, the simple walk is not rapidly mixing. □

3.4 Heat-Bath Random Walk

We saw by Theorem 3.3.2 that the simple walk on $\mathcal{F}_{A,ib}(\mathcal{M})$ is not rapidly mixing. A natural question to consider in response is whether the situation improves if more moves are added to the Markov basis $\mathcal{M}$. In particular, we can consider a variation of the simple walk on $\mathcal{F}(\mathcal{M})$ where from the current state $x \in \mathcal{F}$, a move $\mathbf{m} \in \mathcal{M}$ is chosen randomly, then a new state y is chosen randomly from the ray $(x + \mathbf{m} \cdot \mathbb{Z}) \cap \mathcal{F}$. This modification of the simple walk is a special case of the heat-bath random walk which can be thought of as the discrete version of a hit-and-run algorithm. Intuitively, if we compare the heat-bath random walk to the simple walk, then we expect for the mixing to improve since at each step in the chain, the pool of candidates for the next state is increased. Thus the chain should theoretically "see" more states quickly. In this section, we formally define and explore the heat-bath random walk. As in the previous section we will also pay attention to the underlying graph.

Definition 3.4.1. Let $\mathcal{F} \subset Z^n$ be a finite set and $\mathbf{m} \in \mathbb{Z}^n$. For $x \in \mathcal{F}$, the *ray in* $\mathcal{F}$ *through* x *along* $\mathbf{m}$ is the set $\mathcal{R}_{\mathcal{F},\mathbf{m}}(x) := (x + \mathbf{m} \cdot \mathbb{Z}) \cap \mathcal{F}$.

Definition 3.4.2. Let $\mathcal{F}, \mathcal{M} \subset \mathbb{Z}^n$ be finite sets, and let $\pi : \mathcal{F} \to (0, 1]$ and $f : \mathcal{M} \to [0, 1]$ be probability distributions. The *heat-bath random walk* is the Markov chain with state space $\mathcal{F}$ where from the current state $X_t = x$, a move $\mathbf{m} \in \mathcal{M}$ is chosen with probability $f(\mathbf{m})$. The next state $X_{t+1} = y \in \mathcal{R}_{\mathcal{F},\mathbf{m}}(x)$ is chosen with probability $\frac{\pi(y)}{\pi(\mathcal{R}_{\mathcal{F},\mathbf{m}}(x))}$. Let the matrix

$$H^{\pi}_{\mathcal{F},\mathbf{m}}(x, y) = \begin{cases} \frac{\pi(y)}{\pi(\mathcal{R}_{\mathcal{F},\mathbf{m}}(x))} & \text{if } y \in \mathcal{R}_{\mathcal{F},\mathbf{m}} \\ 0 & \text{otherwise} \end{cases}$$

describe the transition probability when the chain is restricted to a single move $\mathbf{m}$. Then the transition matrix for the heat-bath random walk is $H^{\pi,f}_{\mathcal{F},\mathcal{M}} := \sum_{\mathbf{m}\in\mathcal{M}} f(\mathbf{m})H^{\pi}_{\mathcal{F},\mathbf{m}}$.

The desirable properties of Markov chains hold for the heat-bath random walk under mild conditions. Irreducibility follows when the set $\{\mathbf{m} \in \mathcal{M} : f(\mathbf{m}) > 0\}$ is a Markov basis for $\mathcal{F}$. The heat-bath random walk is aperiodic since the probability $H^{\pi,f}_{\mathcal{F},\mathcal{M}}(x, x)$ is positive, for all $x \in \mathcal{F}$. The stationary distribution is π and the heat-bath random walk is reversible with respect to π. We note here that the underlying graph of the heat-bath random walk is the compression of the graph $\mathcal{F}(\mathcal{M})$, which is essentially $\mathcal{F}(\mathcal{M})$ along with additional edges that arise from allowing scalar multiples of the moves of $\mathcal{M}$.

Definition 3.4.3. Let $\mathcal{F} \subset \mathbb{Z}^d$ and $\mathcal{M} \subset \mathbb{Z}^d$ be finite sets. The *compression* of the graph $\mathcal{F}(\mathcal{M})$ is the graph $\mathcal{F}^c(\mathcal{M}) := \mathcal{F}(\mathbb{Z} \cdot \mathcal{M})$.

Example 3.4.4. Suppose $\mathcal{F}_{A,b}, \mathcal{M} \subset \mathbb{Z}^5$ are sets as defined in Example 3.2.7. For each $x \in \mathcal{F}_{A,b}$ the ray through x along m_2 and m_3 contains at most two vertices. The compressed graph $\mathcal{F}^c_{A,b}(\mathcal{M})$ is displayed in Figure 3.4

When more edges are added to the fiber graph, the diameter is decreased. In fact, the following theorem says that the diameter of the compressed fiber graph $\mathcal{F}^c_{A,b}(\mathcal{M})$ can be bounded by a constant for all $b \in \mathbb{N}A$.

Proposition 3.4.5. *Let $A \in \mathbb{Z}^{m\times d}$ with $\ker_{\mathbb{Z}}(A) \cap \mathbb{N}^d = \{0\}$ and let $\mathcal{M}$ be a Markov basis for $\mathcal{P}_A$. There exists a constant $C \in \mathbb{N}$ such that diam$(\mathcal{F}^c(\mathcal{M})) \leq C$ for all $\mathcal{F} \in \mathcal{P}_A$.*

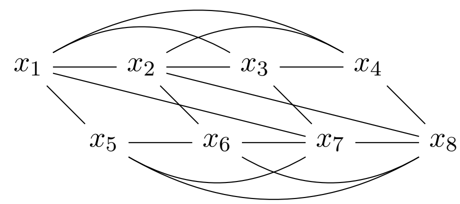

Figure 3.4: The compressed fiber graph from Example 3.4.4

For proof see Section 3 in [26].

While a low diameter on the underlying graph is necessary for rapid mixing it is not sufficient. For example, consider the graph $K_n + K_n$ obtained by joining two complete graphs by a single edge. The diameter of $K_n + K_n$ is 3 however the conductance of the random walk on K_n is $\Phi_* = \frac{1}{n^2}$. That Φ_* is small implies that the mixing is slow.

In the following example, we demonstrate a case where the heat-bath random walk does not improve the mixing behavior.

Example 3.4.6. For $n \in \mathbb{N}$, consider the set $F_n \subset \mathbb{Z}^{2\times n}$ defined as follows:

$$\mathcal{F}_n := \left\{ \begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 & \cdots & 1 \\ 0 & 1 & 0 & \cdots & 0 \end{bmatrix}, \ldots, \begin{bmatrix} 1 & 1 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \right\}.$$

We say that $\mathcal{F}_n$ is the set of $2 \times n$ contingency tables with row sums $(n-1, 1)$ and columns sums $\mathbf{1}_n = (1, \ldots, 1)$. For each n, the set $\mathcal{M}_n := \{ x - y : x, y \in \mathcal{F}_n \} \setminus \{ \mathbf{0} \}$ is a Markov basis for $\mathcal{F}_n$ and, by construction, $\mathcal{F}_n(\mathcal{M}_n) = K_n$. Suppose that f and π are the uniform distribution on $\mathcal{M}_n$ and $\mathcal{F}_n$, respectively. Since $|\mathcal{M}_n| = n(n-1)$ and $\mathbf{0} \notin \mathcal{M}$, the transition matrix for the simple walk is

$$H^{\pi,f}_{\mathcal{F}_n,\mathcal{M}} = \frac{1}{n(n-1)} \mathbf{1}_{n\times n} + \frac{(n(n-1)-n)}{n(n-1)} I_n,$$

where $\mathbf{1}_{n\times n}$ is the $n \times n$ all- ones matrix. Now for any $x \in \mathcal{F}_n$ and $\mathbf{m} \in \mathcal{M}_n$, the ray $\mathcal{R}_{F_n,\mathbf{m}}(x)$ through x along $\mathbf{m}$ contains only two vertices. Thus it follows that the $\mathcal{F}_n(M_n) = \mathcal{F}^c_n(\mathcal{M}_n)$ and the transition matrices for the heat-bath random walk and the simple walk coincide.

In order to compute the SLEM of the transition matrix we first observe that the matrix $\mathbf{1}_{n\times n}$ is diagonalizable, in particular, there exists an invertible $n \times n$ matrix S such that $\mathbf{1}_{n\times n} = S^{-1}DS$ where D is the diagonal matrix whose only nonzero entry is $D_{n,n} = n$.

Then we can write

$$H^{\pi,f}_{\mathcal{F}_n,\mathcal{M}} = S^{-1} \left(\frac{1}{n(n-1)} D + \frac{(n(n-1)-n)}{n(n-1)} I_n \right) S.$$

With this expression we identify the eigenvalues of $H^{\pi,f}_{\mathcal{F}_n,\mathcal{M}}$ as the diagonal entries of the matrix $\frac{1}{n(n-1)}D + \frac{(n(n-1)-n)}{n(n-1)}I_n$. The SLEM of both random walks is $\lambda_n = 1 - \frac{1}{n-1}$ which can not be bounded by $1 - \frac{1}{p(|F_n|)}$, where $p(x)$ is a polynomial. Neither the simple fiber walk nor the heat-bath random walk are rapidly mixing.

Example 3.4.6 shows that implementing the heat-bath random walk does not necessarily improve the mixing behavior. In Section 4 of [26] it is shown that the SLEM of the heat-bath random walk can be bounded when more conditions are imposed on $\mathcal{M}$. For the remainder of this section we briefly mention one of those conditions.

By Proposition 3.4.5, the diameter of the compressed fiber graph $\mathcal{F}^c(\mathcal{M})$ is bounded above by a constant for all $\mathcal{F} \in \mathcal{P}_A$, where A is an integer matrix. Depending on the geometry, paths between distinct $x, y \in \mathcal{M}$ may require that a move $\mathbf{m}$ be used more than once. Consequently the diameter may be larger than the number $|\mathcal{M}|$ of available moves. When we consider graphs $\mathcal{F}^c(\mathcal{M})$ where for any $x, y \in \mathcal{F}$, a path from x to y of minimal length uses each move at most once, then the upper bound on the diameter is lowered to $|\mathcal{M}|$.

Definition 3.4.7. Suppose $\mathcal{F} \subset \mathbb{Z}^n$ and $\mathcal{M} = \{m_1, \dots, m_d\} \subset \mathbb{Z}^n$ are finite sets. An *augmenting path* of length r between distinct $x, y \in \mathcal{F}$ is a path in the compressed graph $\mathcal{F}^c(\mathcal{M})$ of the following form,

$$x \to x + \lambda_{i_1} m_{i_1} \to x + \lambda_{i_1} m_{i_1} + \lambda_{i_2} m_{i_2} \to \cdots y = x + \sum_{k=1}^{r} \lambda_{i_k} m_{i_k}.$$

An augmenting path from x to y is *minimal* if there exists no shorter augmenting path in $\mathcal{F}^c(\mathcal{M})$.

Definition 3.4.8. Suppose $\mathcal{F} \subset \mathbb{Z}^n$ and $\mathcal{M} = \{m_1, \dots, m_d\} \subset \mathbb{Z}^n$ is a Markov basis for $\mathcal{F}$. We say that $\mathcal{M}$ is an *augmenting Markov basis* if there is an augmenting path between any distinct $x, y \in \mathcal{F}$. The *augmentation length* $\mathcal{A}_{\mathcal{M}}(\mathcal{F})$ of an augmenting Markov basis $\mathcal{M}$ is the maximum length of all minimal augmenting paths in $\mathcal{F}^c(\mathcal{M})$.

Example 3.4.9. For fixed $n, r \in \mathbb{N}$ let $\mathcal{C}_{n,r} := \{ x \in \mathbb{Z}^n : \|x\|_1 \le r \}$ be the set of integer points of the n-dimensional cross-polytope with radius r. First we show that the set $\mathcal{M}_n = \{ e_1, \dots, e_n \}$ of standard basis vectors is an augmenting Markov basis for $\mathcal{C}_{n,r}$ for any $r \in \mathbb{N}$. For distinct $x, y \in \mathcal{C}_{n,r}$ it suffices to show that there exists an intermediate point in $\mathcal{C}_{n,r}$ that can be obtained from x by changing a single coordinate x_i to y_i. In other words, we will show that there exists an index $i \in [n]$ such that $x + (y_i - x_i)e_i \in \mathcal{C}_{n,r}$. Let $\mathcal{S}_{xy} \subset [n]$ be the set of indices where x and y differ and let $s = r - \|x\|_1$. If $|\mathcal{S}_{xy}| = 1$, then the result is clear so suppose $|\mathcal{S}_{xy}| \ge 2$. If the result does not hold then for each $i \in \mathcal{S}_{xy}$, we have $\left\|x + (y_i - x_i)e_i\right\|_1 > r$ and so $|y_i| - |x_i| > s$. It follows that

$$\|y\|_1 = \sum_{i \notin \mathcal{S}_{xy}} |x_i| + \sum_{i \in \mathcal{S}_{xy}} |y_i| > \sum_{i \notin \mathcal{S}_{xy}} |x_i| + \sum_{i \in \mathcal{S}_{xy}} (s + |x_i|) = |\mathcal{S}_{xy}|s + \|x\|_1 = (|\mathcal{S}_{xy}| - 1)s + r.$$

This is a contradiction since we assumed $y \in \mathcal{C}_{n,r}$.

When the heat-bath random walk is implemented with an augmenting Markov basis then, using the augmentation length and a given distribution on the moves $\mathcal{M}$, the SLEM can be bounded away from one (see Theorem 5.8 from [26].) As a corollary we get the following result.

Proposition 3.4.10. *Let $\mathcal{F} \subset \mathbb{Z}^n$ be finite and let $\mathcal{M} = \{ m_1, \dots, m_k \}$ be an augmenting Markov basis. Let π be the uniform and f a positive distribution on $\mathcal{F}$ and $\mathcal{M}$ respectively. For $i \in [k]$, let $r_i := \max \left\{ \ |\mathcal{R}_{\mathcal{F},\mathbf{m}_i}(x)| : x \in \mathcal{F} \right\}$ and suppose that $r_1 \geq r_2 \geq \cdots \geq r_k$. Then*

$$\lambda(\mathcal{H}^{\pi,f}_{\mathcal{M},\mathcal{F}}) \leq 1 - \frac{|\mathcal{F}| \cdot \min(f)}{\mathcal{A}_{\mathcal{M}}(F) \cdot \mathcal{A}_{\mathcal{M}}(F)! \cdot 3^{\mathcal{A}_{\mathcal{M}}(F)-1} \cdot 2^{|\mathcal{M}|} \cdot r_1 r_2 \cdots r_{\mathcal{A}_{\mathcal{M}}(\mathcal{F})}}.$$

Example 3.4.11. For a fixed dimension n, let π be uniform on $C_{n,r}$ and f a positive distribution on $\mathcal{M} = \{ e_1, \dots, e_n \}$. The size of $C_{n,r}$ is given by

$$|\mathcal{C}_{n,r}| = \sum_{j=0}^{n} \binom{r+1}{n-j} \binom{n}{j} 2^{n-j},$$

and the binomial inequality $(\frac{n}{k})^k \leq \binom{n}{k} \leq (\frac{ne}{k})^k$ for all $1 \leq k \leq n$ gives $|C_{n,r}| = \Theta(r^n)$. For each $e_i \in \mathcal{M}$, the size of the largest ray along e_i is $r_i = (2r+1)$. Then by Proposition 3.4.10 the SLEM of the heat-bath walk $\mathcal{H}^{\pi,f}_{\mathcal{M},C_{n,r}}$ can be bounded away from one as the radius $r \to \infty$.

Chapter 4

Sampling Lattice Points of Polytopes via Continuous Relaxation

4.1 Introduction

In this chapter, we continue the discussion of sampling lattice points of polytopes. In Chapter 3, the methods for sampling involved first defining a graph structure on the set of lattice points and then implementing a random walk on that graph. Here we follow the strategy of Morris [22] and Dyer, Kannan, and Mount [12] in that we leverage continuous sampling algorithms for convex sets in $\mathbb{R}^n$. The steps of this strategy can be summarized: First sample from a convex set $\tilde{P}$ that contains P to obtain a point $x \in \tilde{P}$. Round x to the nearest lattice point $\mathrm{rd}(x)$. If $\mathrm{rd}(x) \in P \cap \mathbb{Z}^n$, then we are done. If not, then discard $\mathrm{rd}(x)$ and start again.

One immediate question that arises is, What necessitates the intermediate set $\tilde{P}$? Notice that if we collect sample points from P directly and round, then lattice points near or on the boundary of P are less likely to be sampled. Instead, we consider a larger polytope $\tilde{P}$ where the volume of points in $\tilde{P}$ rounding to a given lattice point $x \in P$ is close to one. One of the two main tasks that we tackle in the chapter is how to determine an appropriate choice of $\tilde{P}$? Ideally $\tilde{P}$ is large enough so that the volume of points rounding to any one lattice point in P is close to one but at the same time, $\tilde{P}$ should be as small as possible to reduce the number of rejections.

The second main task is to decide how to sample from $\tilde{P}$. While there are many sampling algorithms on the market, we will focus on two random walk-based sampling algorithms, more specifically based on the ball walk and Dikin ellipsoid random walk, taking advantage of proven results about their efficiency. The ball walk is a random walk that can be used to sample from a general convex set K. From a point $x_t \in K$, one step of the ball walk is to first choose a point y uniformly from the ball $\gamma\mathcal{B}$ centered at x_t with radius $\gamma > 0$ and letting $x_{t+1} = y$ if $y \in K$.

To sample from K via the ball walk, this process is repeated for some predetermined number of steps T, taking x_T as the generated sample point. There are some pre-processing steps often applied to deal with convex sets that have complicated geometries, for instance, tight corners or long and skinny shapes. In Section 4.4, we follow the version of the ball walk described by Kannan, Lovász, and Simonovits in [16].

In Section 4.5 we look at the Dikin ellipsoid algorithm introduced by Kannan and Narayanan in [17]. This is another random walk based method for sampling that is applied specifically to polytopes of the form $P(A, \mathbf{1})$, where $\mathbf{1}$ is the all-ones vector. From a point x_t, in one step of the Dikin walk, a candidate point y is chosen uniformly from the Dikin ellipsoid centered at x_t. If x_t is contained in the Dikin ellipsoid centered at y, then a transition to $x_{t+1} = y$ is accepted with a certain probability. Again, the process is repeated for a predetermine number of steps.

In Sections 4.4 and 4.5, we pin down these sampling schemes more formally. We will see that the major difference between the random walks lie in the fact that the geometry of the Dikin ellipsoid centered at $x \in P$ depends on the geometry of P and on x, where as the ball $\gamma\mathcal{B}$, clearly does not. In Sections 4.4 and 4.5 we also present the results on the required number of steps to obtain a near-uniform sampling. For now, let us assume that we have decided on the algorithm to do continuous sampling on $\tilde{P}$. We generically refer to that algorithm as Algorithm A. Algorithm 4.1.1 formally states the continuous sampling plus round algorithm.

Algorithm 4.1.1 (Continuous Sample plus Round)**.**

- INPUT: matrix $A \in \mathbb{R}^{m \times n}$, vectors $b \in \mathbb{R}^m$ and $\delta \in \mathbb{R}^m_{\geq 0}$, and initial point $x_0 \in P(A, b)$.
- OUTPUT: point $x \in P(A, b) \cap \mathbb{Z}^n$.

1. With initial point x_0, do Algorithm A on the set $P(A, b + \delta)$ to generate a point $X \in P(A, b + \delta)$.
2. Round X to the nearest integer point $\mathrm{rd}(X)$.
3. If $\mathrm{rd}(X) \in P(A, b)$, then output $\mathrm{rd}(X)$ and stop. Otherwise, return to Step 1.

4.2 How to Choose $\tilde{P}$

This section addresses the task of finding an appropriate polytope $\tilde{P}$, such that $P \subset \tilde{P}$, on which we carry out the continuous sampling plus rounding step. Our general strategy follows that of Morris in [22], where the polytopes considered are those whose lattice points are the contingency tables with fixed margins.

Notice that if we do not first choose $\tilde{P}$, and instead carry out continuous sampling on P, then the lattice points near the boundary are less likely to be generated relative to those lattice points that are sufficiently in the interior of P. The goal is to choose $\tilde{P}$ large enough so that the volume of points rounding to each lattice point in P is near one.

Recall that the matrix $A \in \mathbb{R}^{m\times n}$ and vector $b \in \mathbb{R}^m$ give the $\mathcal{H}$-representation of the polytope $P = P(A, b)$. Each row vector a_i^T of A and corresponding entry b_i of b encodes one inequality that is satisfied by points in P, that is, $a_i^T x \leq b_i$ for all $x \in P$. To define $\tilde{P}$, we add a vector $\delta \in \mathbb{R}^m_{\geq 0}$ to the right-hand side of the system and say $\tilde{P} := P(A, b + \delta)$. Geometrically, this change corresponds to pushing each facet of P out by some positive distance. By choosing δ with arbitrarily large coordinates, we can easily achieve the requirement that the volume of points rounding to a lattice point in P is one. However, this action would also increase the rejection rate of the procedure and thus slow down the process. So when choosing a vector δ, we must balance two competing desires, namely, to obtain a sampling scheme that is near-uniform over all lattice points of P while simultaneously being time-efficient.

Definition 4.2.1. For $x \in \mathbb{Z}^n$, the *n-dimensional cube centered at x*, is the set

$$\text{cube}(x) = \left\{ y \in \mathbb{R}^n : -\frac{1}{2} < x_i - y_i \leq \frac{1}{2} \text{ for } i = 1, \ldots, n \right\}.$$

For $x \in \mathcal{F}$, the set cube(x) represents the set of points that round to x.

Problem 4.2.2. *For small error parameter $0 < \epsilon_0 < \frac{1}{2}$, find a vector δ such that for all lattice points $x \in \mathcal{F} = P \cap \mathbb{Z}^n$, the volume of points in $\tilde{P} = P(A, b + \delta)$ rounding to x satisfies* $\text{vol}(\text{cube}(x) \cap \tilde{P}) \geq 1 - \epsilon_0$.

In the remainder of this section we present a method, Proposition 4.2.3, to choose δ that satisfies the conditions of Problem 4.2.2, we look at a couple of examples to see the method in action, and we present the proof of Proposition 4.2.3. For a vector $c \in \mathbb{R}^n$ and constant $d \in \mathbb{R}$, we let $H(c, d)$ denote the hyperplane $\left\{ x \in \mathbb{R}^n : c^T x = d \right\}$.

Proposition 4.2.3. *Suppose $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ such that $P = P(A, b)$ is a polytope.*

- *Let $\mathcal{I} \subset [m]$ be the set of indices corresponding to inequalities of the form $\pm x_j \leq b_i$ for some $i \in [m]$ and $j \in [n]$.*
- *Let m' be the maximum number of hyperplanes, none of which corresponding to inequalities in $\mathcal{I}$, that intersect a fixed* cube(x)*, over all lattice points $x \in P \cap \mathbb{Z}^n$. That is,*

$$m' := \max_{x\in\mathcal{F}} \# \left\{ i \in [m] : i \notin \mathcal{I}, \;\; H(a_i^T, b_i) \cap \text{cube}(x) \neq \emptyset \right\}$$

So $m' \leq m - |\mathcal{I}|$.

- *Let* $z_i \sim U(-\frac{1}{2}, \frac{1}{2})$ *be i.i.d. random variables that are uniform on the interval* $[-\frac{1}{2}, \frac{1}{2}]$ *and* $Z = (z_1, \ldots, z_n)$ *a random vector.*

If we define the vector $\delta \in \mathbb{R}^m$ *coordinate-wise by*

$$\delta_i = \begin{cases} \frac{1}{2}, & i \in \mathcal{I} \\ F^{-1}_{a_i^T Z}(1 - \frac{\epsilon_0}{m'}), & i \notin \mathcal{I}, \end{cases}$$

where $F_{a_i^T Z}(t)$ *is the cumulative distribution function for the random variable* $a_i^T Z$*, and define* $\tilde{P} = P(A, b + \delta)$*, then for all lattice points* $x \in \mathcal{F}$*,*

$$\mathrm{vol}(\mathrm{cube}(x) \cap \tilde{P}) \geq 1 - \epsilon_0.$$

Proposition 4.2.3 says that if we have the $\mathcal{H}$-representation of the polytope P, then we can determine a vector δ so that $\tilde{P}$ satisfies our volume requirement, by solving at most m-many probability equations. Each of those probability equations involve the computation, or at least approximation, of the cumulative distribution function for random variables $a_i^T Z$. Appendix A includes an example of computing the cumulative distribution function for a finite sum of z_i's. Proposition 4.2.3 is demonstrated in Examples 4.2.4 and 4.2.5.

Example 4.2.4. Let P be the triangle with vertices $(1, -1), (1, -1)$, and $(1, 21)$. Then

$$P = P\left(\begin{bmatrix} 0 & -1 \\ 1 & 0 \\ -11 & 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 10 \end{bmatrix}\right).$$

Observe that the unit cube centered at the lattice point $x_0 = (1, 19) \in P$ intersects two of the three facet-defining hyperplanes, namely $H_2 = \left\{ (x, y) \in \mathbb{R}^2 : x = 1 \right\}$ and

$$H_3 = \left\{ (x, y) \in \mathbb{R}^2 : -11x + y = 10 \right\}.$$

On the other hand, cube$((0, -1))$ only intersects $H_1 = \left\{ (x, y) \in \mathbb{R}^2 : -y = 1 \right\}$ and cube$((0, 0))$ does not intersect any of the facet-defining hyperplanes. As a result, notice that if we replace the inequality $-y \leq 1$ with $-y \leq \frac{3}{2}$, then the inequality is satisfied by the entire cube centered at $(0, -1)$ however this action has no effect on the cube centered as $(1, 19)$. Similarly, replacing either inequality $x \leq 1$ or $-11x + y \leq 10$ with $x \leq 1 + c$ or $-11x + y \leq 10 + c$, where $c > 0$, will have no effect on the cube centered at the point $(0, -1)$. So if we want to push out a facet

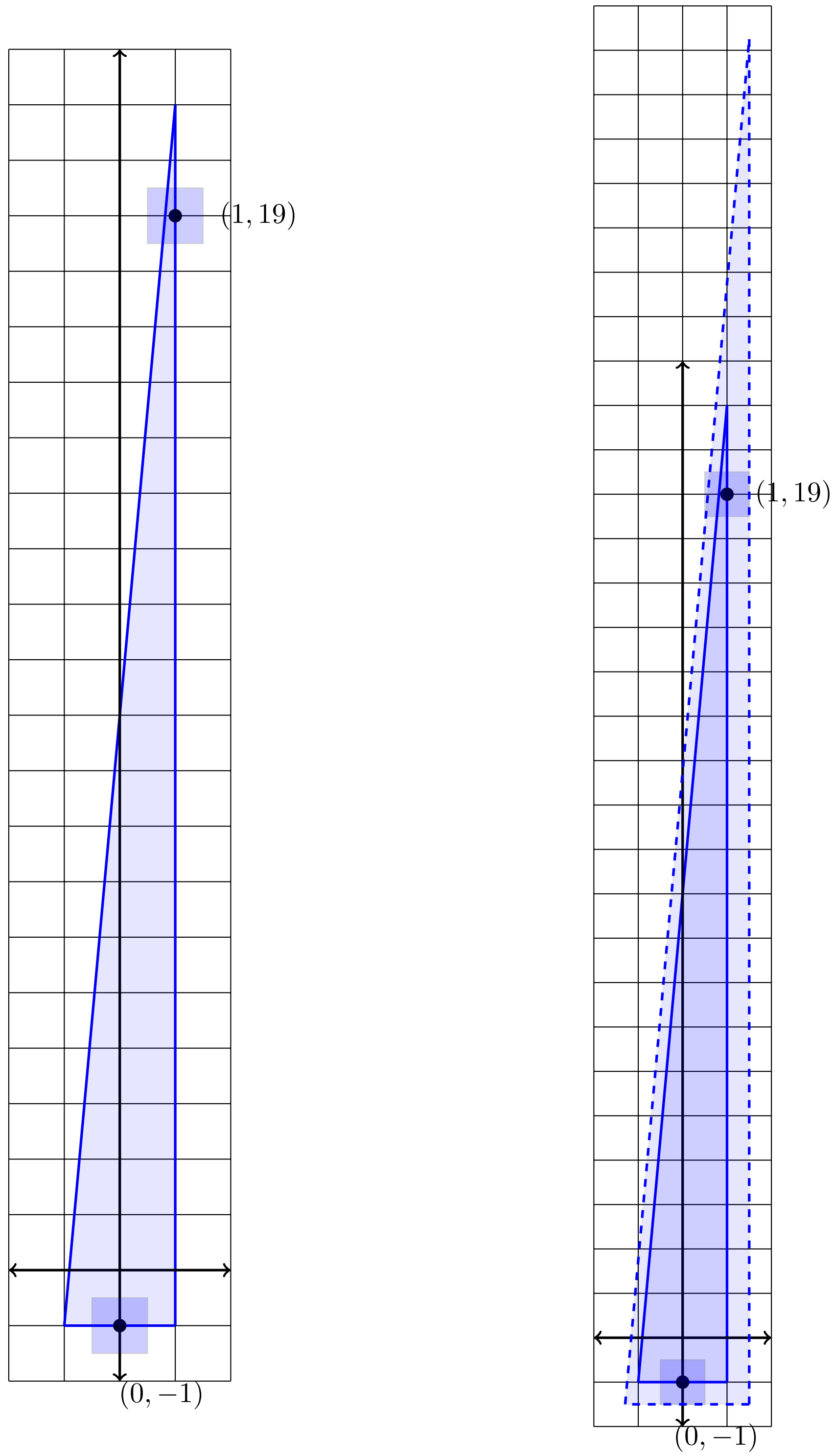

Figure 4.1: Pushing out facets of polytope from Example 4.2.4.

in order to contain the cube centered at a particular lattice point, we only need to consider the facet-defining hyperplanes that intersect that cube. The second thing we notice, as suggested by Proposition 4.2.3 refers to the inequalities that impose constraints on exactly one coordinate. Notice that if we replace $x \leq 1$ with $x \leq \frac{3}{2}$ and $-y \leq 1$ with $-y \leq \frac{3}{2}$ then points in the new area are contained in a cube centered at lattice points of P.

Finally, suppose we want to push out the facets so that at least three-quarters of every cube centered at a lattice point is contained. We determine that the cumulative distribution function for $-11z_1 + z_2$ is

$$F_{-11z_1+z_2}(t) = \begin{cases} 0 & \text{if } t \leq -6 \\ \frac{1}{22}(t+6)^2 & \text{if } -6 \leq t < -5 \\ \frac{1}{22}(2t+11) & \text{if } -5 \leq t < 5 \\ -\frac{1}{22}(t^2 - 12t + 14) & \text{if } 5 \leq t < 6 \\ 1 & \text{if } 6 \leq t. \end{cases}$$

According to Proposition 4.2.3, we can let δ_3 be the solution to $F_{-11z_1+z_2}(t) = \frac{3}{4}$, that is let $\delta_3 = 2.75$, and set $\delta = (\frac{1}{2}, \frac{1}{2}, 2.75)$. In this case, the polytope

$$\tilde{P} = P\left(\begin{bmatrix} 0 & -1 \\ 1 & 0 \\ -11 & 1 \end{bmatrix}, \begin{bmatrix} \frac{3}{2} \\ \frac{3}{2} \\ 12.75 \end{bmatrix}\right)$$

has the property that $\text{vol}(\text{cube}(x) \cap \tilde{P}) \geq \frac{3}{4}$ for all lattice points $x \in P \cap \mathbb{Z}^2$. Both polytopes P and $\tilde{P}$ are displayed in Figure 4.1.

Example 4.2.5. Suppose $P = P(A, b) \subset \mathbb{R}^3$ is the polytope with

$$A = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ 10 & 0 & 1 \\ 0 & 10 & 1 \end{bmatrix} \text{ and } b = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 10 \\ 10 \end{bmatrix}.$$

First we find a polytope $\tilde{P}$ such that $\text{vol}(\text{cube}(x) \cap \tilde{P}) = 1$ for all $x \in \mathcal{F} = P \cap \mathbb{Z}^3$. Afterwards, using Proposition 4.2.3, we can compare that to the choice of $\tilde{P}$ if we only require

$$\text{vol}(\text{cube}(x) \cap \tilde{P}) \geq 1 - \epsilon_0$$

for some small parameter $0 < \epsilon_0 < \frac{1}{2}$.

Recall that if $y \in \text{cube}(x)$ then $|y_j - x_j| \leq \frac{1}{2}$ for each coordinate $j = 1, \ldots, n$. So if we want to choose $\tilde{P}$ such that $\text{cube}(x) \subset \tilde{P}$ for each $x \in \mathcal{F}$, then it suffices to choose a vector $\delta \in \mathbb{R}^5_{\geq 0}$ such that $Ay \leq b + \delta$. If we look at each inequality individually, then we require

$$a_i^T y \leq b_i + \delta_i, \text{ for } i = 1, \ldots, 5. \tag{4.1}$$

For $y \in \text{cube}(x)$, to satisfy Equation 4.1, we need

$$a_i^T x + \frac{1}{2} \sum_{j=1}^{n} |(a_i^T)_j| > a_i^T y \geq a_i^T x - \frac{1}{2} \sum_{j=1}^{n} |(a_i^T)_j|$$

which implies $|a_i^T y - a_i^T x| \leq \frac{1}{2} \sum_{j=1}^{n} |(a_i^T)_j|$. If we set $\delta^0 = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{10}{2}, \frac{10}{2})$ and let $\tilde{P}^0 = P(A, b + \delta^0)$ then $\text{vol}(\text{cube}(x) \cap \tilde{P}^0) = 1$ for all $x \in \mathcal{F}$. We can see how this modification of the right-hand side of the system of inequalities affects the size of the polytope by computing the volume of both P and $\tilde{P}^0$, which we do in `LattE integrale`, see [2]. By computing the volumes, we find that $\text{vol}(P) = \frac{10}{3}$ and $\text{vol}(\tilde{P}^0) = 30.87$, thus the volume increases by a factor of 9.261. Given that

$$\text{vol}(\text{cube}(x) \cap \tilde{P}^0) = 1$$

and by counting the lattice points, we determine that the volume of points in $\tilde{P}^0$ that do not round to points in P is 16.87. This means that if Y is a random vector uniform over $\tilde{P}^0$, then the probability that $\text{rd}(Y) \in P$ is 0.4535.

Now suppose that we loosen the restrictions on $\tilde{P}$ and require $\text{vol}(\text{cube}(x) \cap \tilde{P}) = 1 - \epsilon_0$ for all $x \in \mathcal{F}$ where $\epsilon_0 = \frac{1}{4}$. Observe that the first three inequalities of the system that defines P are non-negativity constraints. Then by Proposition 4.2.3, we can define the vector $\delta^{1/4}$ where $\delta_i^{1/4} = \frac{1}{2}$ for $i = 1, 2, 3$. To determine the remaining coordinates of $\delta^{1/4}$, we need to determine the cumulative distribution function $F_{10z_1+z_2}(t) = \Pr(10z_1 + z_2 \leq t)$ and solve $F_{10z_1+z_2}(t) = 1 - \frac{\epsilon_0}{2}$ for t.

The function $F_{10z_1+z_2}(t)$ is obtained by first extracting the density function $f_{10z_1+z_2}(t)$ from the moment generating function and then integrating over the support. In Appendix A, this same process is taken to compute the cumulative distribution function of the sum of random variables $\sum_{j=1}^{n} z_j$. We find that

$$F_{10z_1+z_2}(t) = \begin{cases} 0 & \text{if } t \leq -\frac{11}{2} \\ \frac{1}{80}(2t+11)^2 & \text{if } -\frac{11}{2} \leq t < -\frac{9}{2} \\ \frac{1}{20}(2t+10) & \text{if } -\frac{9}{2} \leq t < \frac{9}{2} \\ 1-\frac{1}{80}(2t-11)^2 & \text{if } \frac{9}{2} \leq t < \frac{11}{2} \\ 1 & \text{if } \frac{11}{2} \leq t \end{cases},$$

and since $F_{10z_1+z_2}(3.75) = \frac{7}{8}$, we can let $\delta^{1/4} = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 3.75, 3.75)$. In this case the ratio of volumes is 7.1333.

We conclude this section with the proof of Proposition 4.2.3.

Proof of Proposition 4.2.3. We first point out that the volume can be characterized as a probability. For $x \in \mathcal{F}$, the volume $\text{vol}(\text{cube}(x) \cap \tilde{P})$ can be expressed as a probability:

$$\text{vol}(\text{cube}(x) \cap \tilde{P}) = \Pr(x + Z \in \tilde{P}).$$

Then the equality $\text{vol}(\text{cube}(x) \cap \tilde{P}) \geq 1 - \epsilon_0$ is satisfied exactly when $\Pr(x + Z \notin \tilde{P}) \leq \epsilon_0$, that is, the probability that $x + Z$ falls outside of $\tilde{P}$ is at most ϵ_0.

The event that $x + Z$ falls outside of $\tilde{P}$ occurs when at least one of the inequalities that defines $\tilde{P}$ is not satisfied. Hence the event $\Pr(x + Z \notin \tilde{P})$ can be expressed as the union of events $\bigcup_{i=1}^{m}(a_i^T(x + Z) > b_i + \delta_i)$.

If the i-th inequality of P has the form $\pm x_j \leq b_i$, then $\pm(x_j + z_j) \leq b_i + \frac{1}{2}$. In this case, we let $\delta_i = \frac{1}{2}$ and then $x + Z$ will always satisfy the i-th inequality of $\tilde{P}$. Now for a lattice point x, let $\mathcal{H}_x := \left\{ i \in [m] : H(a_i^T, b_i) \cap \text{cube}(x) \neq \emptyset \right\}$ be the set of indices corresponding to facet-defining hyperplanes that intersect the cube(x). Then,

$$
\begin{aligned}
\Pr(x+Z\notin\tilde{P}) &= \Pr\left(\bigcup_{i\in\mathcal{I}}(a_i^T(x+Z)_i > b_i+\frac{1}{2})\right) + \Pr\left(\bigcup_{i\notin\mathcal{I}}(a_i^T(x+Z)_i > b_i+\delta_i)\right)\\
&= \Pr\left(\bigcup_{i\notin\mathcal{I}}(a_i^T(x+Z)_i > b_i+\delta_i)\right)\\
&\le \sum_{i\in\mathcal{H}_x,\ i\notin\mathcal{I}} \Pr(a_i^T(x+Z) > b_i+\delta_i)\\
&= \sum_{i\in\mathcal{H}_x,\ i\notin\mathcal{I}} \Pr(a_i^T Z > (b_i - a_i^T x)+\delta_i)\\
&\le \sum_{i\in\mathcal{H}_x,\ i\notin\mathcal{I}} \Pr(a_i^T Z > \delta_i).
\end{aligned}
\tag{4.2}
$$

The final inequality of Equation 4.2 follows since $(b_i - a_i^T x)\ge 0$. The result is obtained if, for each $i\notin\mathcal{I}$, we choose the coordinate δ_i so that

$$
\Pr(a_i^T Z > \delta_i) = \frac{\epsilon_0}{\max\{\ |\ \mathcal{H}_x|\ \} : x\in P\cap\mathbb{Z}^n}.
$$

□

4.3 Rejection Rate

In this section we assume that δ is chosen and $\tilde{P}=P(A,b+\delta)$ is fixed. The focus here is to measure the efficiency of the sample and round procedure, of which, there are two components. First, the algorithms we consider for sampling on $\tilde{P}$ involve an implementation of a Markov chain with a uniform stationary distribution on a continuous state space. As such, we require some understanding of the mixing time of thes Markov chains that generate a uniform sample Y from $\tilde{P}$. These mixing time questions will be addressed in Sections 4.4 and 4.5. The second component is the rejection rate of the rounded point rd(Y).

Let τ be the number of times that a point $Y\in\tilde{P}$ is generated until rd$(Y)\in\mathcal{F}$. Then τ is a geometric random variable and the expected value $E[\tau]=\frac{1}{\Pr(\text{rd}(Y)\in\mathcal{F})}$. Let

$$
\sum_{x\in\mathcal{F}} \text{vol}(\text{cube}(x)\cap\tilde{P})
$$

denote the total volume of points in $\tilde{P}$ that round to a lattice point in $\mathcal{F}$. It follows that

$$E[\tau] = \frac{\operatorname{vol}(\tilde{P})}{\sum_{x\in\mathcal{F}} \operatorname{vol}(\operatorname{cube}(x)\cap\tilde{P})}.$$

The expected value $E[\tau]$ can be bounded when P is "closed" under rounding.

Definition 4.3.1. A polytope P is *neat* if, for all points x in the interior of P, the rounded point $\operatorname{rd}(x)$ remains in P.

Example 4.3.2. Let $\mathbf{a} \in \mathbb{Z}^n_{\geq 0}$ and $\mathbf{b}, \mathbf{c} \in \mathbb{Z}^{\binom{n}{2}}_{\geq 0}$ be integral vectors indexed by pairs i, j with $1 \leq i < j \leq n$. The polytope $P = \left\{ x \in \mathbb{R}^n : 0 \leq x_i \leq \mathbf{a}_i,\ \ \mathbf{b}_{ij} \leq x_i - x_j \leq \mathbf{c}_{ij} \right\}$ is neat. It is easy to see that if $0 < x_i < \mathbf{a}_i$, then $0 \leq \operatorname{rd}(x_i) \leq \mathbf{a}_i$. So suppose there exists a pair of indices i, j such that $\mathbf{b}_{ij} < x_i - x_j < \mathbf{c}_{ij}$ but $\operatorname{rd}(x_i) - \operatorname{rd}(x_j) > \mathbf{c}_{ij}$. Then, since $\mathbf{c}_{ij}$ and $\operatorname{rd}(x_i) - \operatorname{rd}(x_j)$ are both integers,

$$(\operatorname{rd}(x_i) - x_i) + (x_j - \operatorname{rd}(x_j)) > 1,$$

which is a contradiction since $0 \leq |\operatorname{rd}(s) - s| \leq \frac{1}{2}$ for all real numbers s. It follows that if $\mathbf{b}_{ij} < x_i - x_j < \mathbf{c}_{ij}$, then we must have $\operatorname{rd}(x_i) - \operatorname{rd}(x_j) \leq \mathbf{c}_{ij}$. Using the same arguments, we can show that if $\mathbf{b}_{ij} < x_i - x_j < \mathbf{c}_{ij}$ then $\mathbf{b}_{ij} \leq \operatorname{rd}(x_i) - \operatorname{rd}(x_j)$.

When P is neat,

$$\operatorname{vol}(P) \leq \sum_{x\in\mathcal{F}} \operatorname{vol}(\operatorname{cube}(x)\cap\tilde{P})$$

and so the expected value of τ is bounded above by the ratio $\frac{\operatorname{vol}(\tilde{P})}{\operatorname{vol}(P)}$. Even when P is not neat, we can still obtain a bound on τ that is a ratio of volumes: If $S \subset \mathbb{R}^n$ is a full-dimensional set such that $\operatorname{rd}(x) \in P$ for all $x \in S$, then $\frac{\operatorname{vol}(\tilde{P})}{\operatorname{vol}(S)}$. For the remainder of this section we work to prove an upper bound on the expected value of τ. For a vector $c \in \mathbb{R}^n$ and constant $d \in \mathbb{R}$, we let $H(c, d) := \left\{ x \in \mathbb{R}^n : c^T x = d \right\}$ be a hyperplane.

Definition 4.3.3. Let $P \subset \mathbb{R}^n$ be an $(n-1)$-dimensional polytope contained in the hyperplane $H(c, d)$ and let $x_0 \in \mathbb{R}^n$ be a point not in $H(c, d)$. The *pyramid with base P and peak x_0*, denoted $\operatorname{pyr}(x_0, P) := \operatorname{conv}(\{ x_0 \} \cup P)$, is the convex hull of x_0 taken along with the points in P.

Definition 4.3.4. Let $P \subset \mathbb{R}^n$ be an $(n-1)$-dimensional polytope contained in the hyperplane $H(c, d)$ and let $x_0 \in \mathbb{R}^n$ be a point not in $H(c, d)$. The *height* of $\operatorname{pyr}(x_0, P)$ is the distance from x_0 to the plane $H(c, d)$, that is,

$$\operatorname{ht}(\operatorname{pyr}(x_0, P)) := \frac{|c^T x_0 - d|}{||c||_2}.$$

Example 4.3.5. Consider the line segment $P = t(1,2) + (1-t)(\frac{3}{2}, 1)$ for $t \in [0,1]$ in $\mathbb{R}^2$ which is contained in the line $2x + y = 4$. If $x_0 = (1,1)$, then $\text{pyr}(x_0, P)$ is the triangle whose vertices are the points $(1,1), (1,2)$, and $(\frac{3}{2}, 1)$. The height of the triangle with respect to P is $\frac{1}{\sqrt{5}}$.

Example 4.3.6. Suppose $P = \text{conv}(\{ (1,1,0), (1,-1,0), (-1,1,0), (-1,-1,0) \})$ be the square in $\mathbb{R}^3$ that is contained in the plane $z = 0$. For $x = (0,0,t)$, with $t \neq 0$, the polytope $\text{pyr}(x_0, P)$ is the pyramid with square base and whose height is $|t|$.

Lemma 4.3.7. *Let $P \subset \mathbb{R}^n$ be an $(n-1)$- dimensional polytope, with vertex set $V(P) = \{ v_1, \ldots, v_k \}$, such that P that is contained in the hyperplane $H(c,d)$. Suppose $x_0 \in \mathbb{R}^n$ such that $c^T x_0 < d$. Let $P_{pyr} = \text{pyr}(x_0, P)$ be the pyramid with base P and peak x_0 and $\tilde{P}_{pyr} = x_0 + \text{cone}(v_1 - x_0, \ldots, v_k - x_0) \cap P(c, d+\delta)$ where $\delta > 0$. Then*

1. *$\tilde{P}_{pyr}$ is a pyramid, and*
2. *the ratio of volumes $\frac{\text{vol}(\tilde{P}_{pyr})}{\text{vol}(P_{pyr})} = \left(1 + \frac{\delta}{d - c^T x_0}\right)^n$.*

During the course of the proof of Lemma 4.3.7, the vertices of the pyramid $\tilde{P}_{pyr}$ will be determined explicitly. Each vertex lies on the ray through x_0 and a vertex v_j of P. We make a definition then proceed to the proof of Lemma 4.3.7.

Definition 4.3.8. For points $x, y \in \mathbb{R}^n$, let $r_{\vec{xy}}$ denote the ray that begins at x and passes through y, that is,

$$r_{\vec{xy}} := \{ (1-t)x + ty : t \in [0, \infty) \}.$$

Proof of Lemma 4.3.7. To prove (1) we will find the vertices of $\tilde{P}_{pyr}$: For each $v_j \in V(P)$ define

$$v'_j = r_{\vec{x_0 v_j}} \cap H(c, d+\delta) = -\frac{\delta}{d - c^T x_0} x_0 + (1 + \frac{\delta}{d - c^T x_0}) v_j. \tag{4.3}$$

Since each $v'_j \in H(c, d+\delta)$, for $j = 1, \ldots, k$, it suffices to show that $\tilde{P}_{pyr} = \text{conv}(\{ x_0, v'_1, \ldots, v'_k \})$.

$\tilde{P}_{pyr} \subseteq \text{conv}(\{ x_0, v'_1, \ldots, v'_k \})$: Suppose $x \in \tilde{P}_{pyr}$. Then

$$x = x_0 + \sum_{j=1}^{k} \alpha_j (v_j - x_0), \text{ with } \alpha_j \geq 0. \tag{4.4}$$

We can rearrange the right-hand side of Equation 4.3 to write each vector $v_j - x_0$ in terms of v'_j and x_0:

$$v_j - x_0 = \frac{d - c^T x_0}{\delta + d - c^T x_0}(v'_j - x_0).$$

Now substituting into Equation 4.4 we see

$$x = (1 - \sum_{j=1}^{k} \frac{\alpha_j}{t})x_0 + \sum_{j=1}^{k} \frac{\alpha_j}{t} v'_j, \text{ where } t = 1 + \frac{\delta}{d - c^T x_0}.$$

Since we assumed $c^T x_0 < d$ and $\delta > 0$, we get that each $\frac{\alpha_j}{t} \geq 0$. To show that $x \in \operatorname{conv}\{ x_0, v'_1, \ldots, v'_k \}$, it remains to show that $\sum_{j=1}^{k} \frac{\alpha_j}{t} \leq 1$. For contradiction, suppose that $\sum_{j=1}^{k} \frac{\alpha_j}{t} > 1$. Then we can let $\sum_{j=1}^{k} \frac{\alpha_j}{t} = 1 + p$ for some $p > 0$. Now

$$\begin{aligned} c^T x &= -pc^T x_0 + \sum_{j=1}^{k} \frac{\alpha_j}{t} c^T v'_j \\ &= -pc^T x_0 + \sum_{j=1}^{k} \frac{\alpha_j}{t} (d + \delta) \\ &= -pc^T x_0 + (d + \delta)(1 + p) \\ &= d + \delta + p(\delta + d - c^T x_0) \\ &> d + \delta, \text{ which is a contradiction.} \end{aligned}$$

$\operatorname{conv}(\{ x_0, v'_1, \ldots, v'_k \}) \subseteq \tilde{P}_{pyr}$: Conversely, suppose $x \in \operatorname{conv}\{ x_0, v'_1, \ldots, v'_k \}$. Since each x_0 and each v'_j are contained in the convex set $P(c, d+\delta)$, then $x \in P(c, d+\delta)$. Also by assumption, we can write

$$x = \alpha_0 x_0 + \sum_{j=1}^{k} \alpha_j v'_j$$

where $\alpha_j \geq 0$ for $j = 0, 1, \ldots, k$ and $\sum_{j=0}^{k} \alpha_j = 1$. Again using Equation 4.3, we can rewrite

$$\begin{aligned} x &= \alpha_0 x_0 + \sum_{j=1}^{k} \alpha_j ((1 - t)x_0 + tv_j), \\ &= x_0 + \sum_{j=1}^{k} \alpha_j t(v_j - x_0), \end{aligned}$$

where, again, $t = 1 + \frac{\delta}{d - c^T x_0}$. As $t, \alpha_j \geq 0$, we have $x \in \tilde{P}_{pyr}$, completing the proof of (1). We note here that $\tilde{P}_{pyr} = \operatorname{pyr}(x_0, F' = \{ v'_1, \ldots, v'_k \})$. For (2), we notice that the pyramid $\tilde{P}_{pyr}$ is a dilation of P_{pyr}, so the ratio of volumes is given in terms of the scale factor, namely,

$$\frac{\operatorname{vol}(\tilde{P}_{pyr})}{\operatorname{vol}(P_{pyr})} = \left(\frac{\mathrm{ht}(\mathrm{pyr}(\mathrm{x}_0, \mathrm{F}'))}{\mathrm{ht}(\mathrm{pyr}(\mathrm{x}_0, \mathrm{F}))} \right)^n = \left(\frac{|c^T x_0 - (d + \delta)|}{|c^T x_0 - d|} \right)^n$$

The result (2) follows since $c^T x_0 - d < 0$ and $\delta > 0$. $\square$

Lemma 4.3.7 allows us to bound the expected value of τ, assuming the original polytope, from which we want to sample, is a pyramid. This situation is actually the worst-case scenario and can be used to bound the expected value of τ for arbitrary polytopes.

Theorem 4.3.9. *Suppose $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ such that $P = P(A, b)$ is a full dimensional polytope. Let $\tilde{P} = P(A, b + \delta_i e_i)$, where $\delta_i \geq 0$, be the polytope obtained by modifying the i-th inequality of the system. Let $F_i := P \cap H(a_i^T, b_i)$ and $\tilde{F}_i := \tilde{P} \cap H(a_i^T, b_i + \delta_i)$ be the facets of P and $\tilde{P}$, respectively, that correspond to the i-th inequality in the defining system. For a fixed point $x_0 \in P \backslash F_i$, let $P_{pyr} = \mathrm{pyr}(x_0, F_i)$ and $\tilde{P}_{pyr} = \mathrm{pyr}(x_0, \tilde{F}_i)$ be the pyramids with bases F_i and $\tilde{F}_i$, respectively. Then $\frac{\mathrm{vol}(\tilde{P})}{\mathrm{vol}(P)} \leq \frac{\mathrm{vol}(\tilde{P}_{pyr})}{\mathrm{vol}(P_{pyr})}$.*

Theorem 4.3.9 allows us to bound the ratio of volumes of polytopes where one polytope is obtained from the other by shifting a single facet. Moreover, the result posits that the worst-case scenario occurs when the polytopes in question are pyramids.

Example 4.3.10. Let $P \subset \mathbb{R}^2$ be the polytope from Example 1.3.10 whose $\mathcal{H}$-representation is given by

$$A = \begin{bmatrix} 1 & 2 \\ -7 & -2 \\ -2 & -3 \\ -1 & -1 \\ 9 & 1 \end{bmatrix} \text{ and } b = \begin{bmatrix} 21 \\ -15 \\ -10 \\ -5 \\ 70 \end{bmatrix}.$$

Let $\tilde{P}$ be the polytope obtained by changing the fifth inequality in $Ax \leq b$ from $9x_1 + x_2 \leq 70$ to $9x_1 + x_2 \leq 76$. In other words, $\tilde{P}$ is obtained when we push the hyperplane $H(a_5^T, b_5)$ that defines the facet $F_5 = \{ t(7,7) + (1-t)(8,-2) : t \in [0,1] \}$ of P outwards. We can write $\tilde{P} = P(A, b + 6e_5)$. Both P and $\tilde{P}$ are displayed on the left-hand side of Figure 4.2.

Now let $x_0 = (1,4)$, a point in P but not in F_5, and consider the following two triangles. The first, which we call $P_{pyr} = \mathrm{conv}(\{ x_0, v_1 = (7,7), v_2 = (8,-2) \})$ is obtained by taking the convex hull of x_0 together with the vertices of F_5. For the second triangle, let $v_1' = r_{x_0 \vec{v}_1} \cap H(a_5^T, 76)$ and $v_2' = r_{x_0 \vec{v}_2} \cap H(a_5^T, 76)$ be the intersection points of the rays $r_{x_0 \vec{v}_1}$ and $r_{x_0 \vec{v}_2}$ with the hyperplane $H(a_5^T, 76)$, respectively, and set $\tilde{P}_{pyr} = \mathrm{conv}(\{ x_0, v_1', v_2' \})$. The triangles P_{pyr} and $\tilde{P}_{pyr}$ are displayed on the right-hand side of Figure 4.2.

As each of the polytopes live in the plane, their volumes are not difficult to compute. Using `LattE integrale`, we compute the exact volumes

$$\mathrm{vol}(P) = \frac{109}{2}, \ \mathrm{vol}(\tilde{P}) = \frac{51461}{850}, \ \mathrm{vol}(P_{pyr}) = \frac{57}{2}, \text{ and } \ \mathrm{vol}(\tilde{P}_{pyr}) = \frac{1323}{38}.$$

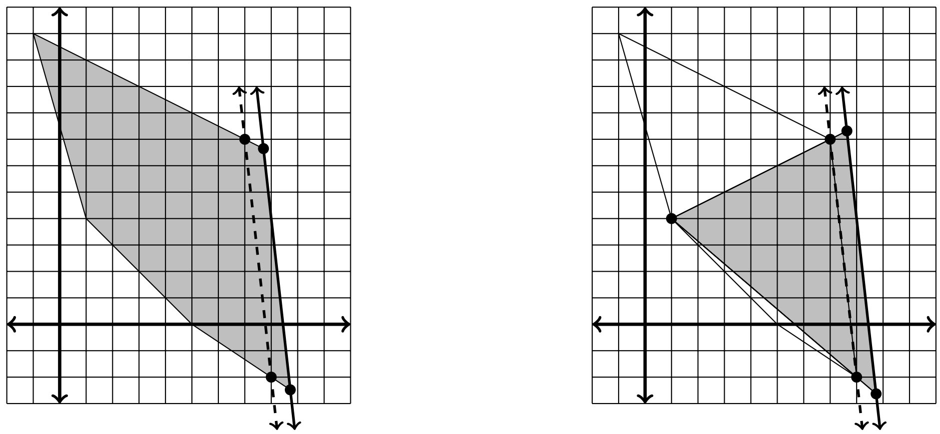

Figure 4.2: Polytopes P and $\tilde{P}$ compared to the pyramids P_{pyr} and $\tilde{P}_{pyr}$ from Example 4.3.10

Therefore $\frac{\mathrm{vol}(\tilde{P})}{\mathrm{vol}(P)} = 1.11087 \leq \frac{\mathrm{vol}(\tilde{P}_{pyr})}{\mathrm{vol}(P_{pyr})} = 1.22161$.

We now prove Theorem 4.3.9.

Proof of Theorem 4.3.9. Since $P \subset \tilde{P}$, we can rewrite the ratio of volumes as

$$\frac{\mathrm{vol}(\tilde{P})}{\mathrm{vol}(P)} = \frac{\mathrm{vol}(\tilde{P}) + \mathrm{vol}(\tilde{P} \backslash P)}{\mathrm{vol}(P)}.$$

An analogous statement can be written for P_{pyr} and $\tilde{P}_{pyr}$. The inequality $\frac{\mathrm{vol}(\tilde{P})}{\mathrm{vol}(P)} \leq \frac{\mathrm{vol}(\tilde{P}_{pyr})}{\mathrm{vol}(P_{pyr})}$ occurs exactly when

$$\frac{\mathrm{vol}(\tilde{P}) + \mathrm{vol}(\tilde{P} \backslash P)}{\mathrm{vol}(P)} \leq \frac{\mathrm{vol}(\tilde{P}_{pyr}) + \mathrm{vol}(\tilde{P}_{pyr} \backslash P_{pyr})}{\mathrm{vol}(P_{pyr})}.$$

By construction $P_{pyr} \subseteq P$, so $\mathrm{vol}(P_{pyr}) \leq \mathrm{vol}(P)$ and it suffices to show

$$\mathrm{vol}(\tilde{P} \backslash P) \leq \mathrm{vol}(\tilde{P}_{pyr} \backslash P_{pyr}).$$

If we let $V(F_i) = \{ v_1, \ldots, v_k \}$ denote the vertex set of F_i, then

$$\tilde{P}_{pyr} = \big(x_0 + cone(\{ v_1 - x_0, \ldots, v_k - x_0 \})\big) \bigcap P(a_i^T, b_i + \delta_i).$$

Suppose $y \in \tilde{P} \backslash P$. Since $x_0 \in P \backslash F_i$, the line segment with endpoints x_0 and y must intersect F_i at a point y'. If we let $V(F_i) = \{ v_1, \ldots, v_k \}$ denote the vertex set of F_i then $y' = (1-t)x_0 + ty = \sum_{j=1}^{k} \alpha_j v_j$ for some $t \in (0,1)$, $\alpha_j \geq 0$, and $\sum_{j=1}^{k} \alpha_j = 1$. By rearranging the equation for y, we see that

$$\begin{aligned} y &= \frac{1}{t}y' - \frac{1-t}{t}x_0 \\ &= x_0 + \sum_{j=1}^{k} \frac{\alpha_j}{t}(v_j - x_0), \end{aligned}$$

which means that $y \in x_0 + cone(\{ v_1 - x_0, \ldots, v_k - x_0 \})$. And since $y \in \tilde{P} \backslash P$, we have $b_i < a_i^T y < b_i + \delta_i$, so $y \in \tilde{P}_{pyr} \backslash P_{pyr}$. This shows that $\tilde{P} \backslash P \subset \tilde{P}_{pyr} \backslash P_{pyr}$ and the result follows. □

Theorem 4.3.9 is concerned with the relationship between a pair of polytopes that can be obtained from each other by shifting a facet. The result can be leveraged to deal with any pair of polytopes of the form $P = P(A, b)$ and $\tilde{P} = P(A, b + \delta)$ since $\tilde{P}$ can be obtained from P by a sequence of facet shifts. We can define a sequence of polytopes that correspond to these shifts. In particular, let $P_0 = \tilde{P}$ and for each $i = 1, \ldots, m$, define $P_i := P_{i-1} \cap P(a_i^T, b_i)$. Then $P_m = P$ and the following containment relationship holds:

$$P = P_m \subset P_{m-1} \subset \cdots \subset P_1 \subset P_0 = \tilde{P}.$$

The ratio of volumes can then be expressed as

$$\frac{\text{vol}(\tilde{P})}{\text{vol}(P)} = \frac{\text{vol}(\tilde{P})}{\text{vol}(P_1)} \cdot \frac{\text{vol}(P_1)}{\text{vol}(P_2)} \cdots \frac{\text{vol}(P_{m-1})}{\text{vol}(P)}. \tag{4.5}$$

The next result, Corollary 4.3.11 combines Theorem 4.3.9, Lemma 4.3.7, and Equation 4.5 to bound the ratio $\frac{\text{vol}(\tilde{P})}{\text{vol}(P)}$. The idea to express the ratio of volumes as the product in Equation 4.5 comes from [22].

Corollary 4.3.11. *Suppose $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are such that $P = P(A, b)$ is a full-dimensional neat polytope. Further suppose $\tilde{P} = P(A, b + \delta)$, where $\delta \in \mathbb{R}^m_{\geq 0}$. For each $i = 1, \ldots, m$,*

- *let $F_i = P \cap H(a_i^T, b_i)$ be a facet of P, and*
- *let $y_i \in P \backslash F_i$ be a point that maximizes the distance $\frac{|a_i^T x - b_i|}{||a_i^T||_2}$ from the hyperplane $H(a_i^T, b_i)$ ranging over all $x \in P \backslash F_i$ and let $h_i = \frac{|a_i^T y_i - b_i|}{||a_i^T||_2}$ denote the distance.*

Then the expected value of τ is bounded above:

$$E[\tau] \leq \prod_{i=1}^{m} \left(1 + \frac{\delta_i}{h_i}\right)^n.$$

Proof. If we set $P_0 = \tilde{P}$ and for each $i = 1, \ldots, m$, define $P_i := P_{i-1} \cap P(a_i^T, b_i)$, then by Equation 4.5, it suffices to bound each ratio $\frac{\mathrm{vol}(P_{i-1})}{\mathrm{vol}(P_i)}$. Recall that the $\mathcal{H}$-representations are $P_i = P(A, b + \sum_{j=i+1}^m \delta_j e_j)$ and $P_{i-1} = P(A, b + \sum_{j=i}^m \delta_j e_j)$. Let $G_i = P_i \cap H(a_i^T, b_i)$ be a facet of P_i and notice that $F_i \subseteq G_i$. Similarly let $G_i' = P_{i-1} \cap H(a_i^T, b_i + \delta_i)$ be a facet of P_{i-1} that is obtained by pushing the i-th facet of P_i outwards. Let $z \in P_i$ be a point such that $a_i^T z < b_i$ whose distance to the hyperplane $H(a_i^T, b_i)$ is maximized over all points $x \in P_i$. Then by Theorem 4.3.9 and Lemma 4.3.7

$$\frac{\mathrm{vol}(P_{i-1})}{\mathrm{vol}(P_i)} \leq \frac{\mathrm{vol}(\mathrm{pyr}(z, G_i'))}{\mathrm{vol}(\mathrm{pyr}(z, G_i))} = \left(1 + \frac{\delta_i}{b_i - a_i^T z}\right)^n.$$

Since $P \subseteq P_i$ and $F_i, G_i \subset H(a_i^T, b_i + \delta)$, we have $b_i - a_i^T y_i \leq b_i - a_i^T z$, and therefore,

$$\frac{\mathrm{vol}(P_{i-1})}{\mathrm{vol}(P_i)} \leq \left(1 + \frac{\delta_i}{h_i}\right)^n.$$

□

Example 4.3.12. We saw earlier in Example 4.3.2 the polytope P

$$P = \left\{ x \in \mathbb{R}^n : 0 \leq x_i \leq \mathbf{a}_i, \;\; \mathbf{b}_{ij} \leq x_i - x_j \leq \mathbf{c}_{ij} \right\},$$

where $\mathbf{a} \in \mathbb{Z}_{\geq 0}^n$ and $\mathbf{b}, \mathbf{c} \in \mathbb{Z}_{\geq 0}^{\binom{n}{2}}$ are integral vectors. Polytopes of this form are called Alcove polytopes and they arise from Coxeter arrangements, see [19]. In Example 4.3.2 we showed that P is neat. So when we choose $\tilde{P}$ and carry out Algorithm 4.1.1, the expected number $E[\tau]$ of points generated in $\tilde{P}$ before landing in P is bounded above by $\frac{\mathrm{vol}(\tilde{P})}{\mathrm{vol}(P)}$. Now given a small parameter $0 < \epsilon_0 < \frac{1}{2}$, we want to choose $\tilde{P}$ such that

$$\mathrm{vol}(\mathrm{cube}(x) \cap \tilde{P}) \geq 1 - \epsilon_0$$

for all $x \in P \cap \mathbb{Z}^n$. Observe that when each inequality $0 \leq x_i \leq \mathbf{a}_i$ is replaced with $-\frac{1}{2} \leq x_i \leq \mathbf{a}_i + \frac{1}{2}$ in the description of P, the resulting polytope does not collect "bad area". In other words, for

$$P' = \left\{ x \in \mathbb{R}^n : -\frac{1}{2} \leq x_i \leq \mathbf{a}_i + \frac{1}{2}, \;\; \mathbf{b}_{ij} \leq x_i - x_j \leq \mathbf{c}_{ij} \right\},$$

if $x \in \mathrm{int}(P')$ then the nearest lattice point $\mathrm{rd}(x)$ is in P.

The problem of choosing $\tilde{P}$ reduces to determining how far to push out the facets corresponding to the inequalities $\mathbf{b}_{ij} \leq x_i - x_j \leq \mathbf{c}_{ij}$. Following the proof of Proposition 4.2.3, we can let δ_{ij} be the solution to

$$\Pr(z_1 + z_2 \leq \delta) = 1 - \frac{\epsilon_0}{n(n-1)},$$

where $z_1, z_2 \sim U(-\frac{1}{2}, \frac{1}{2})$ are i.i.d. random variables. The cumulative distribution function of $Z = z_1 + z_2$ is $F_Z(t) = -\frac{1}{2}(t-1)^2 + 1$ for $t \in (0,1)$. So we can let $\delta_{ij} = 1 - \sqrt{\frac{4\epsilon}{n^2-n}}$, and set

$$\tilde{P} = \left\{ X \in \mathbb{R}^n : -\frac{1}{2} \leq x_i \leq \mathbf{a}_i + \frac{1}{2} \;\; \mathbf{b}_{ij} - \delta_{ij} \leq x_i - x_j \leq \mathbf{c}_{ij} + \delta_{ij} \right\}.$$

Finally we note that, with respect to the facet $\{x \in P : x_i - x_j = \mathbf{c}_{ij}\}$, the width of P is $\frac{c_{ij}-b_{ij}}{\sqrt{2}}$. Thus, by Corollary 4.3.11, the expected value $E[\tau]$ is bounded,

$$E[\tau] \leq \prod_{\{i,j\} \subset [n]} \left(1 + \frac{\sqrt{2} - \sqrt{8\epsilon/(n^2-n)}}{c_{ij} - b_{ij}}\right)^{2n} \leq \exp\left(2n \cdot \sum_{\{i,j\} \subset [n]} \frac{\sqrt{2} - \sqrt{8\epsilon/(n^2-n)}}{c_{ij} - b_{ij}}\right).$$

If the sum $\sum_{\{i,j\} \subset [n]} \frac{1}{c_{ij}-b_{ij}}$ of values $c_{ij} - b_{ij}$ is $\Omega(\frac{1}{n})$ then this upper bound on $E[\tau]$ is a constant.

4.4 Sampling via Ball Walk

In this section, we discuss the ball walk-based algorithm for near-uniform sampling from a convex set K. Here we follow the implementation that comes from Kannan, Lovász, and Simonovits [16]. Their algorithm is a piece of a larger solution to the problem of determining the volume of a convex body. The main result and algorithm of [16] to approximate the volume of a convex body fits in a line of successive improvements on the polynomial time randomized algorithm by voulme by Dyer, Freize, and Kannan [11]. Their near-uniform sampling subroutine uses $O^*(n^3)$ oracle calls, an improvement over the sampling algorithm by Lovász and Simonovits [21].

Both the ball walk and the Dikin ellipsoid random walk provide an alternative to the also popular hit-and-run random walk. The basic idea of the hit-and-run random walk is to start from some point $x_0 \in K$, choose a direction L at random, and choose a new point $x_1 \in K$ from the line along L through x_0, repeating this process for some predetermined number of steps. The stationary distribution of the hit-and-run walk is uniform and the number of steps required to get a sample point that is near-uniformly distributed over K is a mixing time question, that among other things, depends on the geometry of K and the choice of starting point.

Example 4.4.1. We use the following `R` code to implement the basic hit-and-run algorithm on the triangle with vertices $(-1,-1), (1,-1)$, and $(1, 21)$. This particular triangle is long and skinny. Figure 4.3 plots each step of the hit-and-run walk, where we begin at the origin. We carry out three separate trials in which we consider $N = 100, 500$, or 1000 steps. The results of our trial, displayed in Figure 4.3 suggests that we may have to wait over 1000 steps before we visit any states in the corner.

```
unif_ball<-function(n, rad){
  u=rnorm(n)
  sca=runif(1)^(1/n)
  u=(rad*sca*u/norm(u,"2"))
  return(rad*sca*u/norm(u,"2"))
}
#
A=matrix(c(0,1,-11,-1,0,1),nrow=3)
b=c(1,1,10)
#x0=c(95/100,20)
x0=c(0,0)

N=1000
x=x0
samples=matrix(0,2,N)
for (i in 1:N){
  u=unif_sphere(2,1) #new direction
  #
  vec=(b-A%*%x)/(A%*%u) #determine length of ray
  pos=min(vec[which(vec>0)])
  neg=max(vec[which(vec<0)])
  sca=runif(1,min=neg,max=pos)
  #
  x=sca*u+x
  samples[,i]=x
}
#
plot(samples[1,],samples[2,],xlab="x",ylab="y",xlim=c(-2,2),
ylim=c(-1,22),pch=20, cex=.5,main=paste(N,"steps",sep=" "))
```

We focus on the ball walk and the Dikin ellipsoid walk is to address these issues.

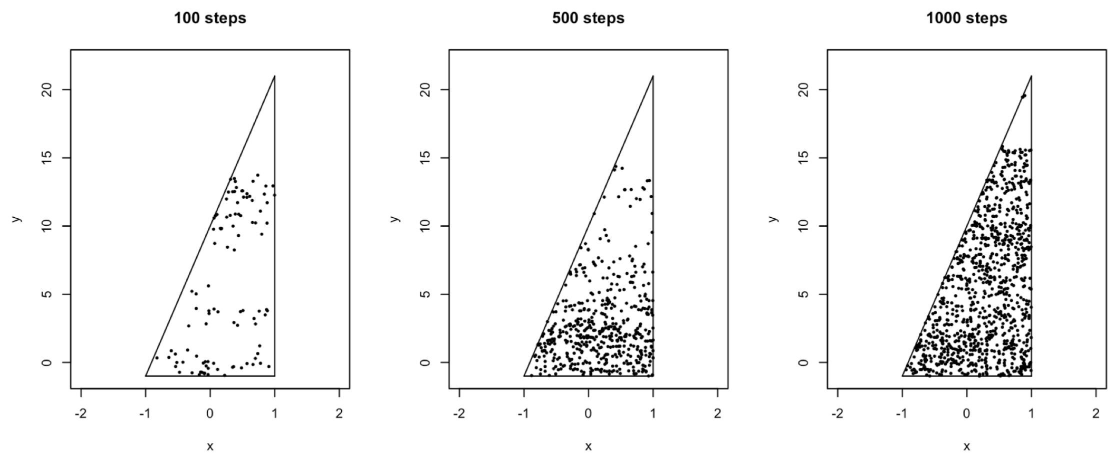

Figure 4.3: Basic Hit-and-Run on a triangle. The number of steps of the random walk considered are 100, 500, and 1000.

Definition 4.4.2. Let $K \subset \mathbb{R}^n$ be a convex set that satisfies $\mathcal{B} \subset K \subset d\mathcal{B}$, where

$$mathcalB := \left\{ x \in \mathbb{R}^n : \|x\|_2 \leq 1 \right\}$$

is the unit ball and $d > 1$ is some fixed constant. The *lazy random walk with γ-steps* on K is defined as follows: From the current state $x_t \in K$, flip a fair coin. If heads then $x_{t+1} = x_t$. If tails, then generate a point $u \in \gamma\mathcal{B}$ uniformly at random. We let $x_{t+1} = x_t + u$ if $x_t + u \in K$ and call this a *proper step*. Otherwise $x_{t+1} = x_t$.

For our sampling algorithm we will let the step size γ be a function of the dimension n and the dilation factor d, and choose γ small enough so that the random walk avoids getting stuck in a corner of K, but still large enough to cover significant portions of K quickly.

Another walk that we define is the speedy random walk. This random walk and its stationary distribution, which is called the speedy distribution, is useful as it can be leveraged to approximate the uniform distribution.

Definition 4.4.3. Let $K \subseteq \mathbb{R}^n$ be a convex set that satisfies $\mathcal{B} \subset K \subset d\mathcal{B}$ for some constant $d > 1$. The *speedy random walk with γ-steps* on K is defined as follows: From the current state $x_t \in K$, flip a fair coin. If heads, then $x_{t+1} = x_t$. Otherwise x_{t+1} is chosen from the uniform distribution on $(x_t + \gamma B) \cap K$.

As noted in [16], the speedy random walk on K can be implemented by doing the lazy random walk, where we only choose proper steps that correspond to points that are different from the previous point, or those that correspond to flipping heads.

Definition 4.4.4. The *speedy distribution* $\hat{\mathcal{Q}}$ is the stationary distribution of the speedy random walk with γ-steps. For a measurable set A, the speedy distribution is

$$\hat{\mathcal{Q}}(A) = \frac{\int_A \operatorname{vol}((x + \gamma\mathcal{B}) \cap K)dx}{\int_K \operatorname{vol}((x + \gamma\mathcal{B}) \cap K)dx}.$$

Theorem 4.4.6 states that we can generate near-uniform, near-independent samples from a convex set K.

Definition 4.4.5. A collection of random points $x_1, \ldots, x_k \in K$ is an *ϵ-good sample for a distribution μ* if

1. for the distribution μ_i of x_i, we have $\|\mu_i - \mu\|_{TV} \le \epsilon$, and
2. for all $1 \le i < j \le k$, the random points x_i and x_j are ϵ-independent, meaning that for any measurable sets A and B,

$$|\Pr(x_i \in A, x_j \in B) - \Pr(x_i \in A)\Pr(x_j \in B)| \le \epsilon.$$

Theorem 4.4.6 (Kannan, Lovasz, and Simonovits [16])**.** *Given a convex set $K \subset \mathbb{R}^n$ satisfying $\mathcal{B} \subset K \subset d\mathcal{B}$, a positive integer N and $\epsilon > 0$, we can generate a set of N random points $\{ x_1, x_2, \ldots, x_N \} \subset K$ that are an ϵ-good sample for the uniform distribution. The algorithm uses only $O^*(n^3d^2 + Nn^2d^2)$ calls on the separation oracle.*

The proof of Theorem 4.4.6 (see Section 4 of [16]) contains the steps of the sampling algorithm. The focus here is to present the steps and only sketch the arguments for the proof of the theorem. First we introduce the M-distance, an alternate way to measure distance between probability distributions.

Definition 4.4.7. Let P and Q be two probability distributions on the same σ-algebra $(\Omega, \mathcal{A})$. The *M-distance* from P to Q is

$$M(P, Q) := \sup_S \frac{P(S) - Q(S)}{\sqrt{Q(S)}},$$

where S ranges over all P- and Q-measurable sets with $Q(S) > 0$.

Sketch of Proof of Theorem 4.4.6. Let $x_0 \in K$ be a random point chosen from some distribution $\mathcal{Q}_0$ that satisfies $M(\mathcal{Q}_0, \hat{\mathcal{Q}}) < \infty$. Starting at x_0, do the speedy random walk on K with γ-steps.

The distribution $\mathcal{Q}_t$ of the t-th step satisfies

$$M(\mathcal{Q}_t, \hat{\mathcal{Q}}) \le M(\mathcal{Q}_0, \hat{\mathcal{Q}}) \exp\left(-\frac{t\gamma^2}{800d^2n}\right).$$

Further, the random points x_0 and x_t are τ-independent, where

$$\tau = [M(\mathcal{Q}_0, \hat{\mathcal{Q}}) + 1] \exp\left(-\frac{t\gamma^2}{800d^2n}\right).$$

So for the sampling algorithm, we set $m = \lceil n \log d \rceil$ and $0 < \epsilon < \frac{1}{4m}$, and let $x_0 \in K$ be a random point whose distribution Q_1 satisfies $M(\mathcal{Q}_1, \hat{\mathcal{Q}}) < 2$, where $\hat{\mathcal{Q}}$ is the speedy distribution with

$$\gamma = \frac{1}{10\sqrt{n \log m \backslash \epsilon}}.$$

The first step: Starting from x_0, do the lazy random walk with γ-steps on K. Let x_1 be the point obtained immediately after $T = \lceil 801 n \ln \frac{5}{\epsilon}(\frac{d}{\gamma})^2 \rceil$ proper steps. From x_t, continue the lazy random walk for T proper steps to obtain x_{t+1} and repeat this process to obtain the sequence $x_1, x_2, \ldots, x_{3N}$.

The collection $\mathcal{S} = \{ x_1, x_2, \ldots, x_{3N} \}$ is an ϵ-good sample for a distribution μ whose total variation distance to the speedy distribution $\hat{\mathcal{Q}}$ is bounded by ϵ. For each point x_i in the collection $\mathcal{S}$, if

$$v_i := \frac{2n}{2n-1} x_i \in K,$$

then distribution μ_i' of v_i satisfies $\left\|\mu_i' - Q\right\|_{TV} < 10\epsilon$, where $\mathcal{Q}$ is the uniform distribution on K. So we take $\mathcal{S}' := \left\{ v_i = \frac{2n}{2n-1} x_i : x_i \in \mathcal{S}, v_i \in K \right\}$ to be the sample set. With high probability $|\mathcal{S}'| \ge N$. $\square$

Remark 4.4.8. *The first step of the algorithm in the proof of Theorem 4.4.6 requires a random point x_0 whose distribution is near the speedy distribution. Once again we refer the reader to [16] for the details on generating such a point.*

The number of proper steps T that pass before we record the state of the chain depends on the scalar factor d of the ball containing the convex set. It is therefore important if we do not know d exactly, that we are able to give a tight upper bound on d. When K has a skewed shape, like in Example 4.4.1, it is beneficial to "round out" the convex set through some affine transformation, thereby minimizing the smallest bounding ball, before carrying out the ball walk.

Definition 4.4.9. A convex set $K \subset \mathbb{R}^n$ is in *isotropic position* if

- the mean $\mu_K = \frac{1}{\text{vol}(K)} \int_K x dx$ is the origin,
- and for each pair of indices i, j, we have

$$\frac{1}{\text{vol}(K)} \int_K x_i x_j dx = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

Definition 4.4.9 is based on the notion of isotropic random variables and can be restated: A convex set K is in isotropic position if, for a random variable X that is uniform over K, the mean of X is 0 and the covariance matrix of X is the identity I_n. The Whitening Transformation in Proposition 4.4.10, which transforms a random variable into white noise, allows us to transform a convex set K into isotropic position.

Proposition 4.4.10. *[Whitening Transformation] Let X be a random vector that is uniform over the convex set $K \subset \mathbb{R}^n$. Let μ_X and Σ_X denote the mean and covariance matrix of X, respectively. Assuming Σ_X is invertible, if $R = \Sigma^{-\frac{1}{2}}$ then the random vector $W(X) := R(X - E[X])$ is isotropic.*

Proof. Since expected value is a linear operator,

$$\mu_{W(X)} = E[R(X - \mu_X)] = R(E[X] - \mu_X) = 0.$$

As a result, the covariance matrix

$$\Sigma_{W(X)} = E[W(X)((W(X))^T] = E[R(X - \mu_X)(X - \mu_X)^T R^T].$$

Again by linearity of expected value,

$$\Sigma_{W(X)} = R\Sigma_X R^T.$$

Finally, since Σ_X is an invertible covariance matrix, then it is symmetric and positive definite. It follows that R is also symmetric and so $\Sigma_{W(X)} = I_n$. $\square$

The affine transformation $W(x) = R(x - \mu_X)$ of Proposition 4.4.10 allows us to transform a polytope P into isotropic position so long as its mean and covariance can be computed. Section 5 of [16] describes an algorithm that takes samples points from the convex set in order to approximate the Whitening transformation.

Finally, Algorithm 4.4.11 is summary of the ball walk that follows from the proof of Theorem 4.4.6. Here we assume that the input is a polytope.

Algorithm 4.4.11 (Sampling via ball walk)**.**

- INPUT: Polytope $P = P(A, b)$ satisfying $\mathcal{B} \subset P \subset d\mathcal{B}$ and number of proper steps T
- OUTPUT: sample points x_T

1. Let $m = \lceil n \log d \rceil$, $0 < \epsilon < \frac{1}{4m}$, and $\gamma = \frac{1}{10\sqrt{n \log m \backslash \epsilon}}$.
2. Let $x_0 \in P$ be a random point from a distribution Q_0 within M distance 2 of the speedy distribution $\hat{\mathcal{Q}}$.
3. counter=0
4. **While** counter $< T$: Flip a fair coin;
   (a) if **Heads**, then $x_{t+1} = x_t$.
   (b) if **Tails**, then generate a point $u \in \gamma B$ uniformly at random. If $x_t + u \in P$ then $x_{t+1} = x_t + u$ and counter= counter+1. Otherwise $x_{t+1} = x_t$.
5. If $\frac{2n}{2n-1} x_t \in P$ then **Output** x_t. Otherwise return to Step 2, continuing the lazy random walk from x_t.

4.5 Sampling via Dikin Ellipsoid Walk

In this section we describe the Dikin ellipsoid walk for uniform sampling from a polytope. This walk is similar to the ball walk, in the sense that at each step, the next point is generated from an ellipsoid centered at the current state. The main distinction is that the geometry of the ellipsoid depends on the center and on the polytope. We consider this particular Markov chain since it was shown in [17] that the mixing time, when starting from a central point, is strongly polynomial in the dimension n and the number of inequalities m that define the polytope. We refer the reader to [17] for more complete details.

Definition 4.5.1. Let $A \in \mathbb{R}^{m \times n}$ such that $P = P(A, \mathbf{1}) \subset \mathbb{R}^n$ is a full-dimensional polytope. For a point $x \in \operatorname{int}(P)$, let $D(x)$ be the $m \times m$ diagonal matrix with entries $d_{ii} = \left(\frac{1}{1 - a_i^T x}\right)$, where a_i^T is the i-th row of A. The *Dikin ellipsoid centered at x with radius r* is the set

$$\mathcal{D}_x^r = \{y \in \mathbb{R}^n : \left\| D(x) A (y - x) \right\|_2 \leq r\}.$$

The Dikin walk takes polytopes in the form $P = P(A, \mathbf{1})$. However, this does not impose too rigid of restrictions when P does not have this form, so long as we can easily apply an appropriate translation to P as a pre-processing step.

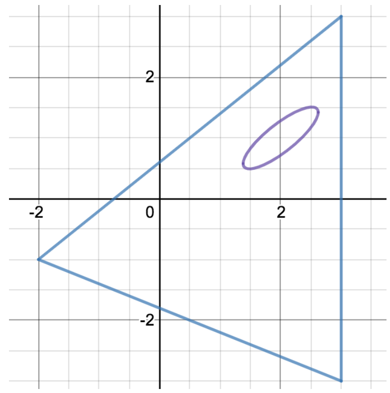

Figure 4.4: The triangle $\mathcal{T}$ and the Dikin ellipsoid centered at $x_0 = (2,1)$ form Example 4.5.2.

Example 4.5.2. The matrix $A = \begin{bmatrix} \frac{1}{3} & 0 \\ -\frac{4}{3} & \frac{5}{3} \\ -\frac{2}{9} & -\frac{5}{9} \end{bmatrix}$ defines the triangle $\mathcal{T} = P(A, \mathbf{1})$ with vertex set $V(T) = \left\{ (-2,-1), (3,-3), (3,3) \right\}$. Notice the point $x_0 = (2,1)$ is in the interior of T. The Dikin ellipsoid centered at x_0 with radius $r = 1$ is the set

$$\mathcal{D}^{\frac{1}{2}}(x_0) = \left\{ y \in \mathbb{R}^2 : \frac{1}{81}\left[673(y_1-2)^2 - 1360(y_1-2)(y_2-1) + 1000(y_2-1)^2\right] \leq 1 \right\},$$

which is displayed in Figure 4.4.

At each step in the Dikin walk, we generate a candidate point from $\mathcal{D}_x^r$. Lemma 4.5.3 shows that the chain will remain in the polytope when the radius is at most one.

Lemma 4.5.3. *If $x \in \operatorname{int}(P)$ and $r \leq 1$, then the Dikin ellipsoid $\mathcal{D}_x^r$ is contained in P.*

Proof. The containment $\mathcal{D}_x^{r_1} \subseteq \mathcal{D}_x^{r_2}$ holds whenever $0 < r_1 \leq r_2$ so it suffices to show $\mathcal{D}_x^1 \subset P$.

For $y \in \mathcal{D}_x^1$, by definition, the value $\left\|D(x)A(y-x)\right\|_2^2$ is bounded above by one, which occurs exactly when

$$\sum_{i=1}^{m} \left(\frac{a_i^T(y-x)}{1-a_i^T x} \right)^2 \leq 1. \tag{4.6}$$

Since each summand of Equation 4.6 is nonnegative and $a_i^T x < 1$, then for each $i = 1, \ldots, m$,

$$\left| a_i^T(y-x) \right| \leq 1 - a_i^T x.$$

This implies that $2a_i^T x - 1 \leq a_i^T y \leq 1$, for each $i = 1, \ldots, m$. $\square$

The Dikin walk is a Metropolis chain, recall Section 1.1.2, that modifies a certain Markov chain by introducing an acceptance probability at each step. For Definitions 4.5.4 and 4.5.5 assume that each ellipsoid $\mathcal{D}_x$ has radius $r = \frac{1}{40}$.

Definition 4.5.4. For a polytope $P = P(A, \mathbf{1}) \subset \mathbb{R}^n$, let $p(x, y)$ be the one-step transition density function for the following Markov chain on P: from some current state x, flip a fair coin. If Heads, then remain at x. If tails, then the next state y is chosen uniformly from $\mathcal{D}_x$. That is, for $x \neq y$,

$$p(x,y) = \begin{cases} \frac{1}{2\mathrm{vol}(\mathcal{D}_x)} & \text{if } y \in \mathcal{D}_x, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 4.5.5. For a polytope $P = P(A, \mathbf{1}) \subset \mathbb{R}^n$, the *Dikin walk* is the Metropolis chain taking $p(x, y)$ with the uniform density π. The one-step transition density function, for $x \neq y$, is

$$q(x,y) = \begin{cases} \min\left\{ \frac{1}{2\mathrm{vol}(\mathcal{D}_x)}, \frac{1}{2\mathrm{vol}(\mathcal{D}_y)}) \right\} & \text{if } x \in \mathcal{D}_y \text{ and } y \in \mathcal{D}_x \\ 0 & \text{otherwise} \end{cases},$$

and $q(x, x) = 1 - \int_{y \in P} dp(x, y)dy$.

A point $x \in P$ is *central* if $\ln s$, where s is the function defined in Theorem 4.5.6, is polynomial in m. It is noted in [17] that if the Dikin walk starts at a central point of P, then the chain mixes in time that is strongly polynomial in the arguments.

Theorem 4.5.6 (Kannan and Narayanan)**.** *Let n be greater than some universal constant. Let $x_0 \in P$ and let*

$$s = \sup_{\overline{pq}} \frac{|p - x_0|}{|q - x_0|}$$

where the supremum is over all chords $\overline{pq}$ passing through the point x_0 and let $\epsilon > 0$ be the desired variation distance to the uniform distribution. Let

$$T > 7 \times 10^8 mn(n \ln(20s\sqrt{m}) + \ln(\tfrac{32}{\epsilon}))$$

and let $x_0, x_1, \ldots,$ be a Dikin walk in which the radius is $\frac{1}{40}$. Then for any measurable set $S \subset P$ the distribution of x_T satisfies $|\Pr(x_T \in S) - \frac{\mathrm{vol}(S)}{\mathrm{vol}(P)}| < \epsilon$.

Definition 4.5.7. Let $P = P(A, b) \subset \mathbb{R}^n$ be a polytope. The point $x \in P$ is an *analytic center* of P if it is a solution to the problem

$$\max_{x \in \mathrm{int}(P)} \sum_{i=1}^{m} \ln(b_i - a_i^T x).$$

Example 4.5.8. The point $(\frac{4}{3}, -\frac{1}{30})$ is the analytic center of the triangle $\mathcal{T}$ from Example 4.5.2.

When we implement the Dikin walk, see Section 4.7, we let the initial state be the analytic center of P since it can be found simply by solving an optimization problem and, as noted in [17], the number of steps required to be within ϵ-total variation distance to stationarity is $O(mn(n \log m + \log \frac{1}{\epsilon}))$.

We conclude this section with Algorithm 4.5.9 which outlines the process for near-uniform sampling via the Dikin walk. For the input polytope $P = P(A, \mathbf{1})$ and $x \in \mathrm{int}(P)$, define the $n \times n$ matrix $H(x) := A^T D^2(x) A$. Assume that the radius of each Dikin ellipsoid is $r = \frac{1}{40}$.

Algorithm 4.5.9 (Sampling via Dikin Walk)**.**

- INPUT: $m \times n$ matrix A, initial solution $x_0 \in P(A, \mathbf{1})$, number of steps T
- OUTPUT: sample point x

1. For $t = 0, 1, \ldots, T-1$:
   (a) Flip a fair coin. If **Heads** then $x_{t+1} = x_t$. If **Tails** then generate a random point $y \in D_{x_t}$.
   (b) If $x_t \in D_y$ then accept y, that is set $x_{t+1} = y$, with probability $\min(1, \sqrt{\det(H(y))/\det(H(x_t))}$. **Otherwise** $x_{t+1} = x_t$.
2. **Return** x_T

4.6 Sampling from a General Lattice

In this section we outline pre-processing steps that must be taken in order to sample points from $P \cap \Lambda$ where $\Lambda \subset \mathbb{R}^n$, $\Lambda \neq \mathbb{Z}^n$ is a general lattice. Notice that if we naively implement Algorithm 4.1.1 on P then we may generate an integer point that is in P but not in Λ. To sample lattice points appropriately, we modify the rounding scheme so that sampled points are rounded to the nearest point in Λ. The motivation for this section comes from situations where we would like to sample from a d-dimensional polytope in $\mathbb{R}^n$ where $d < n$, which occurs when the polytope is fully contained in a hyperplane. We begin with an example to demonstrate the pre-processing steps.

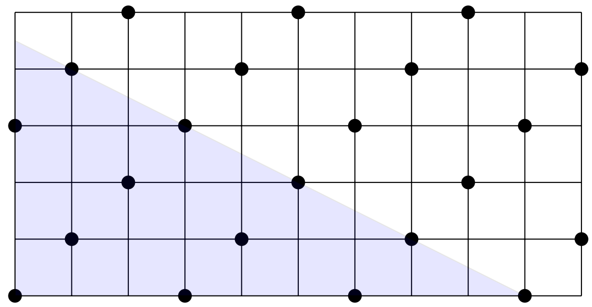

Figure 4.5: Polytope and lattice from Example 4.6.1

Example 4.6.1. Consider the polytope $P = \{X \in \mathbb{R}^2 : 0 \le x_i, 2x_1 + 4x_2 \le 18\}$ and lattice $\Lambda = \{X \in \mathbb{R}^2 : 2x_1 + 4x_2 \equiv 0 \pmod 3\}$, as illustrated in Figure 4.5. The set $\mathcal{B} = \{b_1 = (3, 0), b_2 = (1, 1)\}$ is a basis for Λ. We partition $\mathbb{R}^2$ into cells centered at lattice points in Λ where each cell is the n-dimensional parallelepiped generated by $\mathcal{B}$. For $Y \in \Lambda$, let $[Y]_{\mathcal{B}}$ be the coordinate vector of Y relative to $\mathcal{B}$, then the cell centered at Y is

$$cell(Y) = \{X \in \mathbb{R}^2 : [Y]_{\mathcal{B}} - [X]_{\mathcal{B}} \in (-\tfrac{1}{2}, \tfrac{1}{2}]^2\}.$$

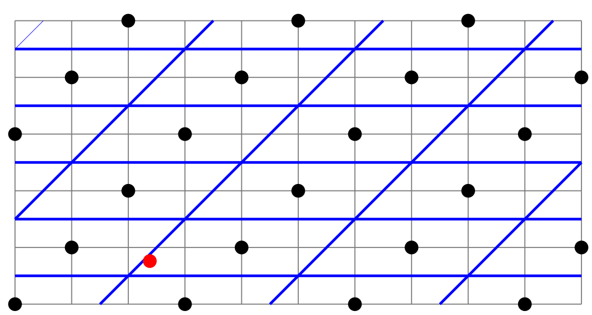

Figure 4.6: A partition of R^2 into cells centered at points in $\Lambda = \{X \in \mathbb{R}^2 : 2x_1 + 4x_2 \equiv 0 \pmod 3\}$. The point $(2.38, 0.76)$ is also plotted in red.

When we sample a point X from the interior of P, instead of rounding each coordinate as usual, we round X to the nearest point in Λ by determining which cell in the partition contains X. This is achieved by simply rounding the coordinates of X relative to $\mathcal{B}$. For instance, suppose that the result of sampling in P gives $X = (2.38, 0.76)$. The coordinate vector relative to $\mathcal{B}$ is $[X]_{\mathcal{B}} = [0.54, 0.76]$ and the rounded coordinate vector is $rd([X]_{\mathcal{B}}) = [1, 1]$ which is the lattice point $b_1 + b_2 = (4, 1) \in \Lambda$. In other words, the lattice point in Λ nearest to $(2.38, 0.76)$ to which our algorithm should round is $(4, 1)$. This scenario is illustrated in Figure 4.6

Following Example 4.6.1 we have a general guideline to follow. Given some polytope $P = P(A, b) \subset \mathbb{R}^n$ and lattice Λ with basis $\mathcal{B}$, to implement Algorithm 4.1.1 we need to look at a new polytope P' whose points represent the coordinate vectors $[X]_{\mathcal{B}}$ relative to $\mathcal{B}$ for $X \in P$. What exactly should P be? Recall that P is given by the inequality $CX \leq d$ where $C \in \mathbb{Z}^{m\times n}$ and $d \in \mathbb{Z}^m$. If $Q = [id]_{\mathcal{B},\mathcal{E}}$ is the change of basis matrix that takes the basis $\mathcal{B}$ to the standard basis $\mathcal{E}$, then we can simply let $P' = P(AQ, b)$.

Once P' is computed and the small parameter ϵ fixed, the steps to determine $\widetilde{P'}$ are unchanged. The remaining step that must be added is that once a point Y is generated from $P' \cap \mathbb{Z}^n$, we must take QY in order to obtain point in $P \cap \Lambda$. The modified workflow is summarized as follows:

Example 4.6.2. Let P and Λ be as defined in Example 4.6.1. To generate a random point $P \cap \Lambda$ note that $Q = \begin{bmatrix} 3 & 1 \\ 0 & 1 \end{bmatrix}$ is the change of basis matrix that takes the basis $\mathcal{B}$ to the standard basis $\mathcal{E}$ so we carry out Algorithm 4.1.1 on the polytope

$$P' = \{X \in \mathbb{R}^2 : -3x_1 - x_2 \leq 0, -x_2 \leq 0, \text{ and } 6x_1 + 6x_2 \leq 18\}$$

and if for example $\epsilon = 0$ then we let $\delta = [2 \quad \frac{1}{2} \quad 6]^T$. Suppose we generate the point $Y = (1, 2)$ then $QY = (5, 2)$ is the sampled point in Λ.

The next example comes from [7].

Example 4.6.3. In this example we are interested in sampling from the integer solutions to a certain knapsack problem. We will see that in this situation the rejection rates for samples generated by Algorithm 4.1.1 are high.

Let $a' = [12223 \quad 12224 \quad 36674 \quad 61119 \quad 85569]$ and $b = 89643482$. We are looking for points $X \in \mathbb{Z}^n$ that satisfy $X \geq \mathbf{0}$ and $a'X = b$. To see the solution set as a full-dimensional polytope, we project along the first coordinate. Let $a = [12224 \quad 36674 \quad 61119 \quad 85569]$ and consider the polytope $P = \{X \in \mathbb{R}^4 : 0 \leq x_i, aX \leq b\}$.

When we implement Algorithm 4.1.1 to generate a point Y we will recover the missing coordinate y_m by letting $y_m = \frac{b-aY}{12223}$. To ensure that y_m is an integer we need to consider the lattice Λ given by $aX \equiv 0 \pmod{b}$ and sampling from $P \cap \Lambda$.

The set $\mathcal{B} = \{b_1 = [12223 \quad 0 \quad 0 \quad 0], b_2 = [12218 \quad 1 \quad 0 \quad 0], b_3 = [12219 \quad 0 \quad 1 \quad 0], b_4 = [12215 \quad 0 \quad 0 \quad 1], \}$ is a basis of Λ. Using $\mathcal{B}$ to construct the change of basis matrix, we determine that the new polytope P' should be given by the system

$$\begin{bmatrix} -12223 & -12218 & -12219 & 12215 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 149413952 & 149389506 & 149426175 & 149401729 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \leq \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 89643482 \end{bmatrix}.$$

If we let the parameter $\epsilon_0 = 0$, then to define $\widetilde{P'}$, the polytope on which we do continuous sampling that contains all cubes centered at a lattice point of interest, we let $\delta^{\epsilon_0} = \begin{bmatrix} 24437.5 & 0.5 & 0.5 & 0.5 & 298815681 \end{bmatrix}^T$. Finally, we use the ratio of volumes $\frac{\mathrm{vol}(\widetilde{P'})}{\mathrm{vol}(P')} = 3454.99$ as an estimate for the rejection rate of Algorithm 4.1.1.

4.7 `R` Codes and Examples

This section contains `R` scripts to carry out sampling via both the ball walk and the Dikin ellipsoid walk. In addition, we show these scripts in action through a few examples.

Subsection 4.7.1 contains Algorithm 4.5.9 which outputs sample points from a polytope $P(A, \mathbf{1})$ when given the matrix $A \in \mathbb{R}^{m\times n}$, the analytic center, the number of steps for one trial of the Dikin walk, and the radius of each Dikin ellipsoid $\mathcal{D}_x^r$. Appendix B, details the method for generating a random point uniform on $\mathcal{D}_x^r$.

Subsection 4.7.2 contains Algorithm 4.4.11 which outputs sample points from a general polytope $P(A, b)$ when given the matrix $A \in \mathbb{R}^{m\times n}$, vector $b \in \mathbb{R}^m$, an initial solution $x_0 \in P(A, b)$, radius d of a ball containing $P(A, b)$, and the number of steps T for one trial of the ball walk.

In Subsection 4.7.3 we implement Algorithm 4.5.9 on a small two-dimensional example, particularly the triangle from Example 4.5.2. The purpose is to show that the programs behaves as we expect. In the remainder of the section, we look at two examples where we perform the steps of Algorithm 4.1.1, using both the ball walk and the Dikin ellipsoid walk, to generate a random sample of the lattice points of a polytope. Some things to notice, not only for the example in Subsection 4.7.3 but also for examples in other subsections, is that in practice we set the number of steps for each random walk to be significantly lower than what the results from the literature suggest is required for good mixing.

4.7.1 `R` Code for Dikin Sampler

Let $A \in \mathbb{R}^{m\times n}$ be a matrix so that $P(A, \mathbf{1})$ is polytope. Let x_{ac} be the analytic center of $P(A, \mathbf{1})$. Optional parameters are

1. `de_st` is the number of steps of the Dikin walk. By default, this value is 10000

2. `rad` is the radius of each Dikin ellipsoid. By default this value is rad= $\frac{1}{40}$

3. `no_samples` is the desired number of samples

```
dikin_sampler<- function(A,x_ac,de_st,rad,no_samples){
  m=dim(A)[1] #no. inequalities
  n=dim(A)[2] #dimension
  b=c(matrix(1,nrow=1,ncol=m))
  Dmat<-function(x,A){
    diag(c((b-A%*%x)^(-1)))
  }
  Hx<-function(x){
    t(A)%*%Dmat(x,A)^2%*%A
  }
  if (missing(de_st)){
    de_st=10000 #desired number of steps
  }
  if (missing(rad)){
    rad=1/40 #desired radius
  }
  x=x_ac
  samples=matrix(0,n,no_samples)
  for (j in 1:no_samples){
    step_no=0
    while (step_no < de_st){
      if (sample.int(2,size=1)==1){
        hx=Hx(x)
        decomp=eigen(hx,symmetric=TRUE)
        R=t(decomp$vectors)
        D=diag(decomp$values)
        E=sqrt(D)
        ER=E%*%R
        #
        #generate point in ball of radius r
        u=rnorm(n)
        sca=runif(1)^(1/n)
        u=(rad*sca*u/norm(u,"2"))
        v=solve(ER,u)
```

```
        #
        Dy=Dmat(v+x,A)
        #ask if v+x in D_y
        if (norm(Dy%*%A%*%(-v),"2") <=rad ){
          #determine probabilities
          dHy=det(Hx(v+x))
          dHx=det(hx)
          proba=min(1,sqrt(dHy/dHx))
          if (sample.int(2,size=1,prob=c(proba,1-proba))==1){
            x=v+x
            #accept y
          }
        }
      }
      step_no=step_no+1
    }
    samples[,j]=x
  }
  return(samples)
}
```

4.7.2 R Code for Ball Walk Sampler

Let $P = P(A, b) \subset \mathbb{R}^n$ be a polytope such that $\mathcal{B} \subset P \subset d\mathcal{B}$ for some constant $d > 0$. Further assume that P is close to isotropic position. Other parameters

1. no_samples is the desired number of samples
2. T_prop is the number of steps of the ball walk before capturing the state
3. x_0 is the initial state

```
kls_ball<- function(A,b,no_samples,d,T_prop,x0){
  #set parameters
  n=dim(A)[2]
  m=ceiling(n*log10(d))
  eps=1/(4*m+1)
  gamma_ss=(100*n*log10(m/eps))^(-.5)
  #
```

```
  x=x0
  #
  samples=matrix(0,n,no_samples)
  sf=2*n/(2*n-1)
  j=0
  while (j < no_samples){
    counter=0
    while (counter<T_prop){
      u=rnorm(n)
      sca=runif(1)^(1/n)
      u=(gamma_ss*sca*u/norm(u,"2"))
      if (prod(A%*%(x+u)<=b)==1){
        x=x+u
        counter=counter+1
      }
    }
    if (prod(A%*%(sf*x)<=b)==1){
      j=j+1
      samples[,j]=sf*x
    }
  }
  return(samples)
}
```

4.7.3 Sampling Points from $\mathcal{T}$

In this subsection we continue Example 4.5.2 by implementing the `R` codes for the Dikin ellipsoid and ball walk- based sampling algorithms from Subsections 4.7.1 and 4.7.2. The goal of this small example is to see the algorithms for continuous sampling in action. After this example, in Subsection 4.7.4, we put all the pieces together and implement Algorithm 4.1.1 for sampling lattice points on a couple of examples in higher dimension.

Recall from 4.5.2 that $\mathcal{T} = P(A, \mathbf{1})$ is the triangle with vertices $(-2,-1), (3,-3)$, and $(3,3)$. Using `Latte integrale` [2], we determine the true mean and covariance matrix of a random vector uniformly distributed over $\mathcal{T}$.

1. The mean $\mu = \frac{1}{\mathrm{vol}(\mathcal{T})} \int_{\mathcal{T}} x dx = (4/3, -1/3)$, and

Table 4.1: This table displays the $\|\cdot\|_2$-distance between the sample means and the true mean for both the ball walk and Dikin ellipsoid-based sampling algorithms for various choices of the step number.

$\|\cdot\|_2$-distance to true mean								
	10 steps	50	100	500	10^3	$5*10^3$	10^4	$5*10^4$
γ-ball	0.80589	0.77037	0.58698	0.30993	0.19511	0.02127		
dikin			1.29629	0.58828	0.75742	0.19363	0.19329	0.07574

Table 4.2: This table displays the $\|\cdot\|_2$-distance between the sample covariances and the true covariance for both the ball walk and Dikin ellipsoid-based sampling algorithms for various choices of the step number.

$\|\cdot\|_2$-distance to true covariance								
	10 steps	50	100	500	10^3	$5*10^3$	10^4	$5*10^4$
hit-and-run	0.84725	0.49005	0.57261	0.27530	0.12151	0.20014		
dikin			1.10228	0.41477	0.56815	0.59557	0.35444	0.11613

2. the covariance $\mathrm{Cov}(x_1, x_2) = \begin{bmatrix} 25/18 & 5/18 \\ 5/18 & 14/9 \end{bmatrix}$.

In this experiment, we generate 500 random points from $\mathcal{T}$ using both the ball walk and Dikin ellipsoid algorithms. For each algorithm, we consider a range of choices for the number of steps that pass before recording the state (that is, `T_prop` for the ball walk and `de_st` in the Dikin walk.)

Theorem 4.5.6 suggests that we should set the number of steps to be 5.939×10^{10}. Notice from Tables 4.1 and 4.2 that when we implement the Dikin walk, we get a sample set whose mean and covariance are both within 0.25 of the true mean and true covariance taking 5×10^4 steps. The ball walk from the proof of Theorems 4.4.6 suggests letting the number of steps be approximately 4.29×10^7. Again notice from Tables 4.1 and 4.2 that we can get a good sample set whose mean and covariance are near the true values. The plots in Figures 4.8 and 4.7 are the result of running the `R` codes and plotting the sample points.

4.7.4 Sampling Lattice Points Examples

In this section we look at two examples where we perform the steps of Algorithm 4.1.1 to generate a random sample of the lattice points of a polytope. The first polytope that we consider is a truncated cube in $\mathbb{R}^3$ obtained by removing a pair of opposite corners. In the second example, we seek to generate a sample of contingency tables with given table margins. For both examples,

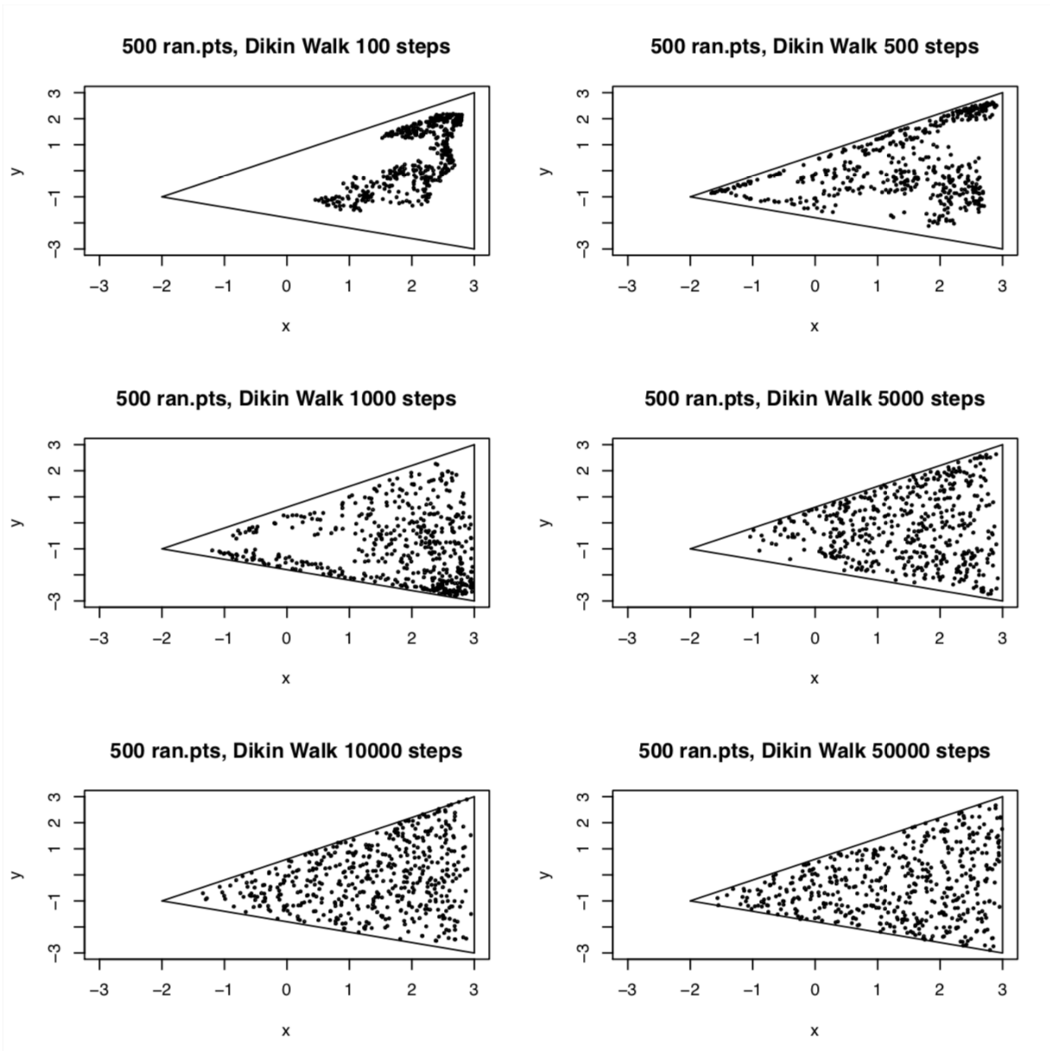

Figure 4.7: Plot of 500 sample points from $\mathcal{T}$, using Dikin algorithm, with various number of step between

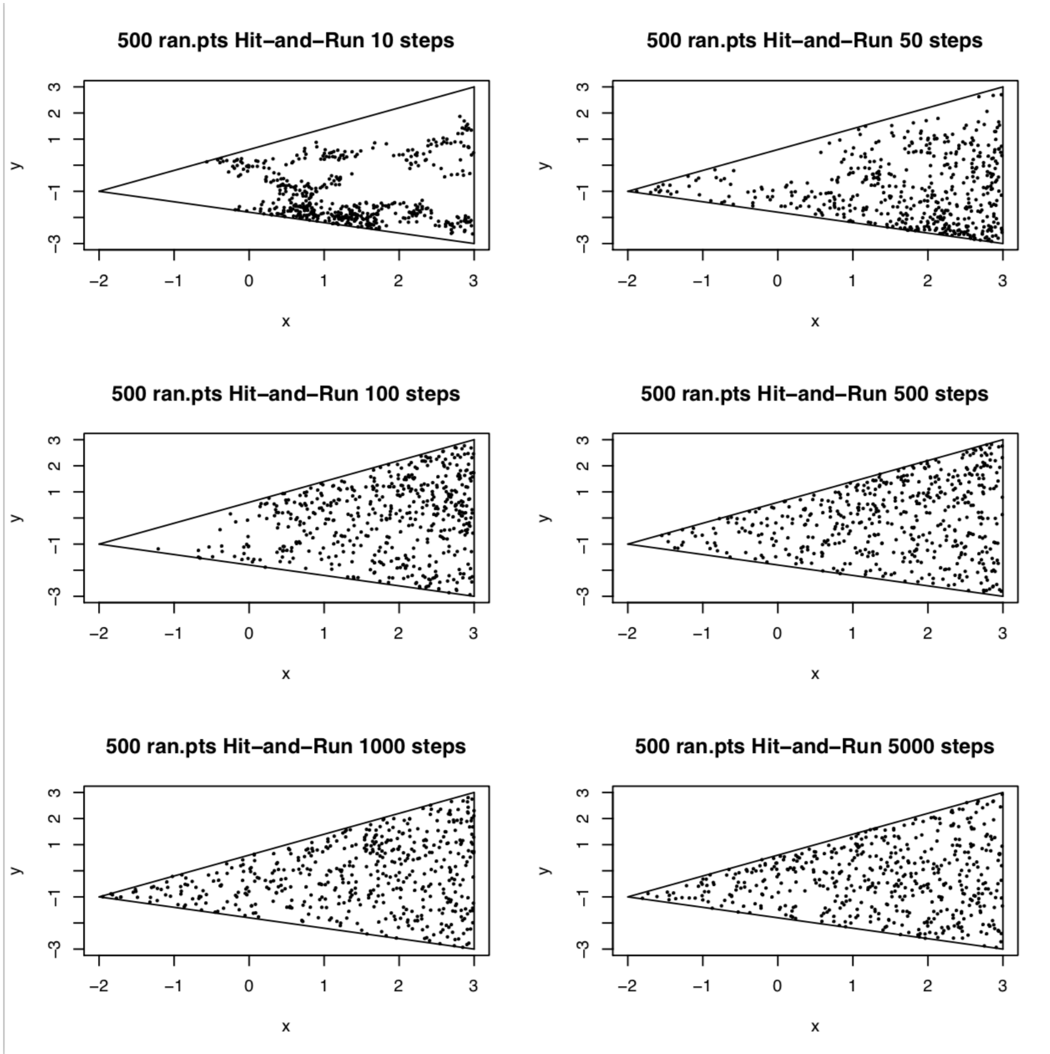

Figure 4.8: Plot of 500 sample points from $\mathcal{T}$, using the ball walk with various number of step between

we generate random samples using the ball walk and Dikin ellipsoid walks and, for the sake of comparison, we consider a range of steps of the random walk taken between sample points. The general outline for both examples is the following:

Let $P = P(A, b) \subset \mathbb{R}^n$ be the polytope whose lattice points we would like to sample. Following Algorithm 4.1.1, we first need to choose a polytope $\tilde{P}$ and choose a method to sample continuously from $\tilde{P}$. In our examples we will compare the performance of Algorithm 4.1.1 when we consider different options:

1. Choice of $\tilde{P}$: We will choose vectors δ^0 and $\delta^{\frac{1}{4}}$ in $\mathbb{R}^m_{\geq 0}$ (using Proposition 4.2.3) so that the polytope $\tilde{P}_0 := P(A, b + \delta^0)$ contains each cube centered at any given lattice point in $\mathcal{F}$ and the polytope $\tilde{P}_{\frac{1}{4}} := P(A, b + \delta^{\frac{1}{4}})$ contains at least three-quarters of each cube centered at a lattice point in $\mathcal{F}$.

2. Choice of sampling algorithm: both the ball (Section 4.4) and the Dikin ellipsoid walk algorithm (Section 4.5) will be implemented.

3. Number of steps: for both the ball and Dikin ellipsoid walk algorithm, we specify the number of steps taken by the random walk between sample points. We generate sample sets using the following number of steps: 100, 500, 1e03, 5e03, 1e04, 5e04, 1e05, and 5e05.

For each choice of parameters, we use the `R` codes to generate a sample of 500 random lattice points in $\mathcal{F}$. We then compute the sample mean, sample covariance, the run time to generate each sample set, and the acceptance rate, that is, the sample size divided by the number of trials of continuous sampling on $\tilde{P}$ before we have 500 lattice points in $\mathcal{F}$. We have also computed the true mean and the true covariance of $\mathcal{F}$ by enumerating all lattice points of P. The true mean and true covariance are

$$\mu_{\mathcal{F}} = \frac{1}{|\mathcal{F}|} \sum_{x \in \mathcal{F}} x,$$
$$\Sigma_{\mathcal{F}} = \frac{1}{|\mathcal{F}|} \sum_{x \in \mathcal{F}} (x - \mu_{\mathcal{F}})^T (x - \mu_{\mathcal{F}}).$$

With this information, the goal is to determine which choice of parameters is optimal. Further, we would like to obtain a rough estimate for the number of steps required in the random walk before the sample statistics are near the true statistics.

To use the Dikin algorithm, we first shift $\tilde{P}$ so that the lattice point nearest to the analytic center becomes the origin. Formally, we transform $\tilde{P}$ to the system

$$P(A, (b + \delta) - A(\mathrm{rd}(x_{ac}))) = \left\{ y \in \mathbb{R}^n : y + \mathrm{rd}(x_{ac}) \in \tilde{P} \right\},$$

where x_{ac} is the analytic center of $\tilde{P}$. The purpose of this shift is to work on a polytope that contains the origin and thus can be expressed as $P(A', \mathbf{1})$, for some matrix A'. The choice to shift $\tilde{P}$ by the rounded point $\mathrm{rd}(x_{ac})$ is so that lattice points in $\tilde{P}$ correspond to lattice points in the shifted polytope $P(A, (b+\delta) - A(\mathrm{rd}(x_{ac})))$.

For the ball walk, we also shift $\tilde{P}$ so that $\mathrm{rd}(x_{ac})$ becomes the origin. Then the scalar factor d, of the ball containing $\tilde{P}$, is approximated by solving the optimization problem, $\max_{x \in P(A,(b+\delta)-A(\mathrm{rd}(x_{ac})))} \|x\|_2$ as a pre-processing step.

Example 4.7.1 (Truncated Cube)**.** Suppose we want to sample lattice points in the truncated cube $\mathcal{C}$, where

$$\mathcal{C} = \left\{ x \in \mathbb{R}^3 : -10 \leq x_1, x_2, x_3 \leq 10, \;\; -2 \leq x_1 - x_2 \leq 10 \right\}.$$

In other words, $\mathcal{C} = P(A, b)$ where

$$A = \begin{bmatrix} & I_3 & \\ & -I_3 & \\ 1 & -1 & 0 \\ -1 & 1 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 10 \cdot \mathbf{1}_6 \\ 10 \\ 2 \end{bmatrix}.$$

Following Algorithm 4.1.1, we need to choose a larger polytope $\tilde{\mathcal{C}}$ such that $\mathcal{C} \subseteq \tilde{\mathcal{C}}$. If we let $\delta^0 = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 1, 1)^T$ then the polytope $\tilde{\mathcal{C}_0} = P(A, b + \delta^0)$ has the property that $\mathcal{C} \subset \tilde{\mathcal{C}_0}$ and $\mathrm{vol}(\mathrm{cube}(x) \cap \tilde{\mathcal{C}_0}) = 1$ for all $x \in \mathcal{F}$.

Using Proposition 4.2.3, we also choose a polytope that satisfies $\mathrm{vol}(\mathrm{cube}(x) \cap \tilde{\mathcal{C}}) \geq \frac{3}{4}$ for all $x \in \mathcal{F}$. There are only two inequalities in the defining system of $\mathcal{C}$ that do not have the form $\pm x_j \leq b_i$. And since the hyperplanes $H(a_7^T, 10)$ and $H(a_8^T, 2)$ are far enough apart, there are no cubes centered at a lattice point in $\mathcal{F}$ that intersect both hyperplanes. From Appendix A, we determine that the cumulative distribution function for the sum $z_1 + z_2$, where z_1, z_2 are independent $U(-\frac{1}{2}, \frac{1}{2})$ random variables, is given by

$$F_{z_1+z_2}(t) = \begin{cases} 0, & t \leq -1 \\ \frac{1}{2}(t+1)^2, & -1 < t \leq 0 \\ -\frac{1}{2}t^2 + t + \frac{1}{2}, & 0 < t < 1 \\ 1, & 1 \leq t \end{cases}.$$

Table 4.3: Truncated Cube: This table displays the $\|\cdot\|_2$-distance between the sample means and the true mean for both the ball walk and Dikin ellipsoid-based sampling algorithms for various choices of the step number.

$\|\cdot\|_2$-distance to true mean						
	100	500	10^3	$5*10^3$	10^4	$5*10^4$
ball, $\epsilon_0 = 0$	5.472734	6.754784	5.398435	4.829845	0.9768404	1.954941
ball, $\epsilon_0 = \frac{1}{4}$	1.271215	4.810954	8.358417	2.306616	3.108179	0.7753852
dikin,$\epsilon_0 = 0$	3.20219	5.70497	1.50822	2.29012	0.98878	0.70390
dikin, $\epsilon_0 = \frac{1}{4}$	5.73197	8.28852	5.81650	2.13970	1.67759	0.59985

Table 4.4: Truncated Cube: This table displays the $\|\cdot\|_2$-distance between the sample covariances and the true covariance for both the ball walk and Dikin ellipsoid-based sampling algorithms for various choices of the step number.

$\|\cdot\|_2$-distance to true covariance						
	100	500	10^3	$5*10^3$	10^4	$5*10^4$
ball, $\epsilon_0 = 0$	45.57909	41.47734	38.8652	11.92179	5.514269	8.091575
ball, $\epsilon_0 = \frac{1}{4}$	46.3134	29.89801	43.28793	11.09837	10.27829	3.276207
dikin,$\epsilon_0 = 0$	39.96622	20.53888	29.32092	13.96456	11.80100	5.12891
dikin, $\epsilon_0 = \frac{1}{4}$	47.29601	43.24955	30.89206	7.16252	7.25398	3.47414

Then by solving the equation $F_{z_1+z_2}(t) = \frac{3}{4}$, we find that if we set

$$\delta^{\frac{1}{4}} = (\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, 0.293, 0.293)^T$$

and $\tilde{\mathcal{C}}_{\frac{1}{4}} := P(A, b + \delta^{\frac{1}{4}})$, then $\mathcal{C} \subset \tilde{\mathcal{C}}_{\frac{1}{4}}$ and $\mathrm{vol}(\mathrm{cube}(x) \cap \tilde{\mathcal{C}}_{\frac{1}{4}}) \geq \frac{3}{4}$ for all $x \in \mathcal{F}$. With $\tilde{\mathcal{C}}_0$ and $\tilde{\mathcal{C}}_{\frac{1}{4}}$ chosen, we perform Algorithm 4.1.1 for a total of 24 times, each time using different options for the input to generate a set of 500 random lattice points of $\mathcal{C}$. Those options correspond to the different combinations of the following choices:

- $\tilde{\mathcal{C}}_0$ or $\tilde{\mathcal{C}}_{\frac{1}{4}}$?
- ball or Dikin ellipsoid random walk?
- 100, 500, 1000, 5000, 1e04, or 5e04 steps

For each generated sample set of lattice points we record the sample mean and the sample covariance, and we compare those values to the true values that we obtain by enumerating all lattice points with the following `R` code:

Table 4.5: Truncated Cube: This table displays the $\|\cdot\|_2$-distance between the sample means and the true mean for both the ball walk and Dikin ellipsoid-based sampling algorithms for various choices of the step number.

Acceptance Rate						
	100	500	10^3	$5*10^3$	10^4	$5*10^4$
ball, $\epsilon_0=0$	0.93458	0.95785	0.91575	0.94340	0.94162	0.93110
ball, $\epsilon_0=\frac{1}{4}$	1	0.99404	0.99602	0.99206	0.99404	0.99602
dikin,$\epsilon_0=0$	1	0.93110	0.97656	0.95238	0.94518	0.93633
dikin, $\epsilon_0=\frac{1}{4}$	1	1	1	1	0.95057	0.99206

```
> #lattice points tcube
> lp=matrix(0,3,1)
> for (I in -10:10){
+   for (J in -10:10){
+     for (K in -10:10){
+       if (I-J<=10){
+         if (I-J>=-2){
+           lp=cbind(lp,c(I,J,K))
+         }
+       }
+     }
+   }
+ }
>
> lattice_points=lp[,2:dim(lp)[2]]
> truemean=apply(lattice_points,1,mean)
> truecov=cov(t(lattice_points))
> truemean
[1]  1.655814 -1.655814  0.000000
> truecov
         [,1]     [,2]     [,3]
[1,] 28.44598 21.93728  0.00000
[2,] 21.93728 28.44598  0.00000
[3,]  0.00000  0.00000 36.67479
```

For each sample set we also record the acceptance rate which is the sample size, 500, divided by the number of trials of continuous sampling on $\tilde{C}$ before obtaining those 500 lattice points in $\mathcal{C}$. The result of these actions are summarized by Tables 4.3, 4.4, and 4.5.

We first note that the acceptance rates are slightly higher when we perform Algorithm 4.1.1 letting $\tilde{\mathcal{C}} = \tilde{\mathcal{C}}_{\frac{1}{4}}$ instead of $\tilde{\mathcal{C}}_0$. That the difference is slight is expected since the dimension of the polytope and the ambient space is not large.

Next, if we just look at the sample statistics for sample sets generated when we let $\tilde{\mathcal{C}} = \tilde{\mathcal{C}}_{\frac{1}{4}}$, then it is not clear that either choice of the ball or Dikin ellipsoid walk significantly outperforms the other. However, Tables 4.3 and 4.4 do suggest that we may be able to generate a sample set whose sample statistics are near the true values if we let the number of steps be on the order of 5×10^4 or 10^5. We compare this to the results for the Dikin random walk which suggests that the number of steps should be at least 2.387×10^{11}, and the results for the ball walk suggests that we choose at least 1.796×10^9 steps.

As a final comment, we noted from Tables 4.3 and 4.4 that it is not clear whether the Dikin ellipsoid walk performs drastically better than the ball walk for any given choice of parameters but, we did observe that the computational time for Algorithm 4.1.1 is larger when we use the Dikin ellipsoid walk to do continuous sampling. For instance, generating the sample set with $5e04$ steps, using the Dikin walk, took 8 hours. On the other hand, generating the sample set again with $5e04$ steps, using the ball walk, took only 3 hours. The reason for the time differential, is that in each step of the Dikin ellipsoid walk, to generate a point in the Dikin ellipsoid requires that we construct a matrix and then solve a matrix equation. See Appendix B.

In this next example, we repeat the process of Example 4.7.1 but for a polytope corresponding to 3×4 contingency tables with fixed row and column sums.

Example 4.7.2 (Two-way Contingency Table)**.** Suppose we are interested in generating a sample of 3×4 contingency tables with row sums $r = (33, 27, 21)$ and column sums $c = (22, 18, 19, 22)$. This set of contingency tables is the set of lattice points $\mathcal{F} = P \cap \mathbb{Z}^6$ of the polytope $P(A, b)$ where

$$A = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 33 \\ 27 \\ 22 \\ 18 \\ 19 \\ -38 \end{bmatrix}.$$

We know that this set is nonempty since, for instance, the point $(10, 6, 9, 7, 9, 1)$ is in $\mathcal{F}$. For this round of test in `R`, we let $\tilde{P}_0 := P(A, b + \delta^0)$ where the vector δ^0 is

$$\delta^0 = (\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{3}{2}, 1, 1, 1, 3).$$

As in the previous example, this choice of $\tilde{P}_0$ contains all cubes centered at a lattice point in $\mathcal{F}$. Now to choose a polytope $\tilde{P}$ that contains at least $\frac{3}{4}$ of every cube centered at a lattice point, notice that there are 6 defining inequalities that are not of the form $\pm x_j \leq b_i$, and so following Proposition 4.2.3, we can let $m' = 6$ be the maximum number of facet-defining hyperplanes, none of the form $\left\{ x \in \mathbb{R}^n : \pm x_j = b_i \right\}$, that intersect any cube centered at a given lattice point in $\mathcal{F}$. We compute the cumulative distribution function $F_{Z_{n'}}(t)$ for $n' = 2, 3$, and 6 where, in general, the random variable $Z_{n'} = \sum_{i=1}^{n'} z_i$ of $U(-\frac{1}{2}, \frac{1}{2})$ i.i.d random variables. We then solve each function for $1 - \frac{\epsilon_0}{6} = \frac{23}{24}$. Since,

$$\begin{aligned} F_{Z_2}(t) &= \frac{23}{24} \implies t = 0.7113 \\ F_{Z_3}(t) &= \frac{23}{24} \implies t = 0.87 \\ F_{Z_6}(t) &= \frac{23}{24} \implies t = 1.2447, \end{aligned} \tag{4.7}$$

we can let

$$\delta^{\frac{1}{4}} = (\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, 0.87, 0.87, 0.7113, 0.7113, 0.7113, 1.2247)$$

and then the polytope $\tilde{P}_{\frac{1}{4}} := P(A, b + \delta^{\frac{1}{4}})$ contains at least three-quarters of each cube centered at a lattice point in $\mathcal{F}$. See Appendix A for details of Equations 4.7.

Table 4.6: Two-way Contingency Table: This table displays the $\|\cdot\|_2$-distance between the sample means and the true mean for both the Dikin and hit-and-run algorithms for various choices of the step number.

$\|\cdot\|_2$-distance to true mean								
	100	500	1e03	5e03	1e04	5e04	1e05	5e05
ball, $\epsilon_0 = 0$	2.9842	3.0877	8.3720	7.8133	3.4329	1.4209	2.9453	0.5585
ball, $\epsilon_0 = \frac{1}{4}$	4.8949	6.1500	1.6690	2.1058	4.1438	2.0696	1.0708	1.0627
dikin,$\epsilon_0 = 0$	6.8451	4.5846	3.7366	2.3054	2.9069	0.6060	0.5585	0.7492
dikin, $\epsilon_0 = \frac{1}{4}$	5.7313	8.3500	8.2319	4.2147	2.1128	0.4519	0.6875	0.2813

We generate sample sets consisting of 500 random lattice points using both the ball and the Dikin ellipsoid walks, we let $\tilde{P}$ be either $\tilde{P}_0$ or $\tilde{P}_{\frac{1}{4}}$, and we let the number of steps be 1e03, 5e03, 1e04, 5e04, and 1e05.

We make some observations based on the results reported in Figure 4.9 and Tables 4.6, 4.7, and 4.8.

First we focus on the sample sets generated when we let $\tilde{P} = \tilde{P}_{\frac{1}{4}}$. From Tables 4.6, 4.7 it seems that, using the Dikin walk, a sample set of lattice points can be generated if the let the number of steps of the Dikin walk be on the order of 10^5 or 10^6. We compare this to Theorem 4.5.6 which suggests that we should choose at least 2.519×10^{13} steps. And if we use the ball walk, then the results from Theorem 4.4.6 direct us to choose at least 1.783×10^{10} steps. From the sample statistics in Tables 4.6 and 4.7 it is not clear if we can generate a good sample set using fewer steps.

The mitigating factor that works in favor of the ball walk is computation time. On average, generating a sample set via the Dikin walk takes 1.5 times longer than generating a sample set via the ball walk assuming all other parameters are equal. This time differential is significant when we set the number of steps of either random walk to be over 10^5 where the computing time to generate a sample set takes over an hour. For example, generating a sample set using 5×10^4 steps took 8.12 hours with the Dikin walk, whereas the ball walk only required 3.7 hours.

Finally, Table 4.8 shows that the acceptance rate is significantly larger when we let $\tilde{P} = \tilde{P}_{\frac{1}{4}}$ instead of $\tilde{P} = \tilde{P}_0$, which speaks to the tangible benefit of relaxing the requirement on $\tilde{P}$ to $\mathrm{vol}(\mathrm{cube}(x) \cap \tilde{P}) \geq 1 - \epsilon_0$ for all $x \in \mathcal{F}$.

.

Table 4.7: Two-way Contingency Table: This table displays the $\|\cdot\|_2$-distance between the sample covariances and the true covariance for both the Dikin and hit-and-run algorithms for various choices of the step number.

$\|\cdot\|_2$-distance to true covariance								
	100	500	1e03	5e03	1e04	5e04	1e05	5e05
ball, $\epsilon_0 = 0$	48.3545	46.0120	39.6203	24.8661	25.2947	11.2617	10.3096	10.1980
ball, $\epsilon_0 = \frac{1}{4}$	48.0543	43.8976	43.2418	32.3943	36.9701	12.9152	19.4854	15.0922
dikin,$\epsilon_0 = 0$	39.6126	33.8709	20.9273	30.8996	8.0076	8.1195	8.0932	10.9406
dikin, $\epsilon_0 = \frac{1}{4}$	37.4560	28.5128	40.6693	15.2506	11.4101	6.5604	9.1686	9.5600

Table 4.8: Two-way Contingency Table: This table displays the $\|\cdot\|_2$-distance between the sample means and the true mean for both the Dikin and hit-and-run algorithms for various choices of the step number.

Acceptance Rate								
	100	500	1e03	5e03	1e04	5e04	1e05	5e05
ball, $\epsilon_0 = 0$	0.9960	0.7800	0.5855	0.7364	0.6935	0.6329	0.6859	0.6596306
ball, $\epsilon_0 = \frac{1}{4}$	0.6337	0.8446	0.8993	0.8278	0.7962	0.8306	0.8026	0.8104
dikin,$\epsilon_0 = 0$	0.9940	0.5787	0.6974	0.7429	0.6614	0.6711	0.6329	0.6150
dikin, $\epsilon_0 = \frac{1}{4}$	0.9980	0.7072	0.9881	0.8489	0.8503	0.8052	0.8210	0.8210

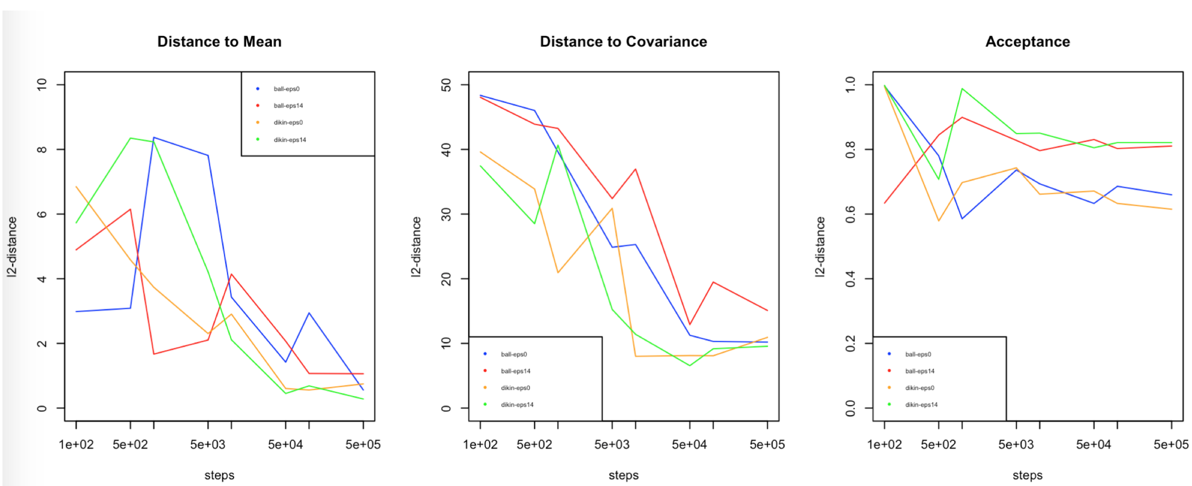

Figure 4.9: The figure summarizes the results of the codes in R. The plot shows, for different number of steps, the distance between sample mean and true mean, the distance between sample covariance and true covariance, and acceptance rate for the Dikin and hit-and-run algorithms, letting $\epsilon_0 = 0$, and $\frac{1}{4}$.

REFERENCES

[1] Satoshi Aoki, Hisayuki Hara, and Akimichi Takemura. *Markov bases in algebraic statistics*. Springer Series in Statistics. Springer, New York, 2012.

[2] V. Baldoni, N. Berline, J.A. De Loera, B. Dutra, M. Koppe, S. Moreinis, G. Pinto, M. Vergne, and J. Wu. *A User's Guide for LattE integrale v1.7.2*, 2013.

[3] Dave Bayer and Persi Diaconis. Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.*, 2(2):294–313, 1992.

[4] Matthias Beck and Sinai Robins. *Computing the continuous discretely*. Undergraduate Texts in Mathematics. Springer, New York, second edition, 2015. Integer-point enumeration in polyhedra, With illustrations by David Austin.

[5] William E. Boyce and Richard C. DiPrima. *Elementary differential equations and boundary value problems*. John Wiley & Sons, Inc., New York-London-Sydney, 1965.

[6] F. R. K. Chung, Persi Diaconis, and R. L. Graham. Random walks arising in random number generation. *Ann. Probab.*, 15(3):1148–1165, 1987.

[7] J. A. De Loera, D. Haws, R. Hemmecke, P. Huggins, and R. Yoshida. A computational study of integer programming algorithms based on Barvinok's rational functions. *Discrete Optim.*, 2(2):135–144, 2005.

[8] Persi Diaconis. *Group representations in probability and statistics*, volume 11 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1988.

[9] Persi Diaconis and Bradley Efron. Testing for independence in a two-way table: new interpretations of the chi-square statistic. *Ann. Statist.*, 13(3):845–913, 1985. With discussions and with a reply by the authors.

[10] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981.

[11] Martin Dyer, Alan Frieze, and Ravindran Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *J. Assoc. Comput. Mach.*, 38(1):1–17, 1991.

[12] Martin Dyer, Ravi Kannan, and John Mount. Sampling contingency tables. *Random Structures Algorithms*, 10(4):487–506, 1997.

[13] David Hilbert. Beweis für die Darstellbarkeit der ganzen Zahlen durch eine feste Anzahl n^{ter} Potenzen (Waringsches Problem). *Math. Ann.*, 67(3):281–300, 1909.

[14] Martin Hildebrand. Random walks supported on random points of $\mathbf{Z}/n\mathbf{Z}$. *Probab. Theory Related Fields*, 100(2):191–203, 1994.

[15] Daniel C. Jerison, Lionel Levine, and John Pike. Mixing time and eigenvalues of the abelian sandpile Markov chain. *arXiv e-prints*, page arXiv:1511.00666, Nov 2015.

[16] Ravindran Kannan, László Lovász, and Miklós Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures Algorithms*, 11(1):1–50, 1997.

[17] Ravindran Kannan and Hariharan Narayanan. Random walks on polytopes and an affine interior point method for linear programming. *Math. Oper. Res.*, 37(1):1–20, 2012.

[18] Mike Krebs and Anthony Shaheen. *Expander families and Cayley graphs*. Oxford University Press, Oxford, 2011. A beginner's guide.

[19] Thomas Lam and Alexander Postnikov. Alcoved polytopes. I. *Discrete Comput. Geom.*, 38(3):453–478, 2007.

[20] David A. Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017. Second edition of [MR2466937], With contributions by

Elizabeth L. Wilmer, With a chapter on “Coupling from the past” by James G. Propp and David B. Wilson.

[21] L. Lovász and M. Simonovits. Random walks in a convex body and an improved volume algorithm. *Random Structures Algorithms*, 4(4):359–412, 1993.

[22] Ben J. Morris. Improved bounds for sampling contingency tables. *Random Structures Algorithms*, 21(2):135–146, 2002.

[23] Murray Rosenblatt. *Markov processes. Structure and asymptotic behavior*. Springer-Verlag, New York-Heidelberg, 1971. Die Grundlehren der mathematischen Wissenschaften, Band 184.

[24] K.W. S, L. Faming, and W. Jian-sheng. *Markov Chain Monte Carlo: Innovations And Applications*. Lecture Notes Series, Institute For Mathematical Sciences, National University Of Singapore. World Scientific Publishing Company, 2005.

[25] Laurent Saloff-Coste. Random walks on finite groups. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pages 263–346. Springer, Berlin, 2004.

[26] Caprice Stanley and Tobias Windisch. Heat-bath random walks with Markov bases. *Adv. in Appl. Math.*, 92:122–143, 2018.

[27] Richard P. Stanley. *Enumerative combinatorics. Volume 1*, volume 49 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2012.

[28] Tobias Windisch. Rapid mixing and Markov bases. *SIAM J. Discrete Math.*, 30(4):2130–2145, 2016.

[29] Günter M. Ziegler. *Lectures on polytopes*, volume 152 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.

APPENDICES

Appendix A

The CDF of a Sum of I.I.D Uniform Random Variables

Let $P = P(A, b) \subset \mathbb{R}^n$ and $\tilde{P} = P(A, b + \delta)$ be polytopes where $\delta \in \mathbb{R}^m_{\geq 0}$. In Section 4.2 the question under consideration is, Given a small parameter $0 < \epsilon_0 \leq \frac{1}{2}$, how can we choose a vector δ such that $\mathrm{vol}(\mathrm{cube}(x) \cap \tilde{P}) \geq 1 - \epsilon_0$, for all $x \in P \cap \mathbb{Z}^n$? Recall that a_i^T is the i-th row of A, and $Z = (z_1, \dots, z_n)$ is a random vector whose coordinates z_j are i.i.d random variables uniform on the interval $[-\frac{1}{2}, \frac{1}{2}]$. Proposition 4.2.3 provides a method for choosing δ coordinate by coordinate and the proof involves solving each equation

$$\Pr(a_i^T Z \leq \delta_i) = 1 - \frac{\epsilon_0}{m}, \tag{A.1}$$

for δ_i. The left hand side of Equation A.1 is the cumulative distribution function $F_{a_i^T z}(\delta_i)$ for the random variable $a_i^T Z$. So to solve for δ_i, we need to derive an algebraic expression for $F_{a_i^T z}(t)$.

Lemma A.0.1 gives the cumulative distribution function of $\sum_{i=1}^n z_i$. The steps in the proof can be followed to derive the cumulative distribution function for any $a_i^T Z$.

Lemma A.0.1. *Let $z_i \sim U(-\frac{1}{2}, \frac{1}{2})$ be i.i.d. random variables uniform on $[-\frac{1}{2}, \frac{1}{2}]$. If $Z_n = \sum_{i=1}^n z_i$, then for $t \in \mathbb{R}$,*

$$F_{Z_n}(t) = \Pr(Z_n \leq t) = \frac{1}{n!} \sum_{j=0}^{\lfloor t + \frac{n}{2} \rfloor} \binom{n}{j} (-1)^j (t + \frac{n}{2} - j)^n.$$

To fully understand the proof requires some knowledge of Laplace transforms. We briefly comment on Laplace transforms and refer the reader to Chapter 6 of [5] for more complete details.

Definition A.0.2. For a function $f(t)$ defined for $t \geq 0$, the *Laplace transform of f*, often denoted by $\mathcal{L}\{f(t)\}$ or $F(s)$, is

$$\mathcal{L}\{f(t)\} = F(s) = \int_0^\infty \exp(-st)f(t)dt,$$

assuming the integral exists, where s is a complex number.

Example A.0.3. The Laplace transform of $f(t) = t$ is $\mathcal{L}\{t\}(s) = \int_0^\infty t\exp(-st)dt = \frac{1}{s^2}$.

It follows from linearity of integration that the Laplace transform is a linear operator. That is, for functions $f(t)$ and $g(t)$ defined for $t \geq 0$ and constant $c \in \mathbb{R}$, the Laplace transform of $(af + g)(t)$ is

$$\mathcal{L}\{(af + g)(t)\}(s) = a\mathcal{L}\{f(t)\}(s) + \mathcal{L}\{g(t)\}(s).$$

From a transformed function $F(s)$, we can recover the function f by inverting the process. Table 6.2.1 of [5] contains the Laplace transforms for some elementary functions. This table is used to quickly invert functions $F(s)$, in other words, to determine the inverse Laplace transform. Laplace transforms are useful here due to the connection to moment generating functions of random variables.

Definition A.0.4. For a random variable X with probability density function $f_X(x)$, the *moment generating function of* X, denoted $m_X(t)$, is

$$m_X(t) = E(\exp(tX)) = \int_{-\infty}^\infty \exp(tx)f_X(x)dx,$$

for $t \in \mathbb{R}$.

Within probability theory, we refer to $\mathcal{L}\{f_X\}(s) = E(\exp(-sX))$ as the Laplace transform of the random variable X. With the substitution $s = -t$, the Laplace transform of X is the moment generating function. So if we know the moment generating function for a random variable X, then using inverse Laplace transforms, we can recover the probability density function $f_X(x)$.

Proof of Lemma A.0.1. For $z_i \sim U(-\frac{1}{2}, \frac{1}{2})$, the probability density function is $f_{z_i}(t) = 1$ for $t \in [-\frac{1}{2}, \frac{1}{2}]$ and $f_{z_i}(t) = 0$ otherwise. Hence the moment generating function is

$$\begin{aligned} m_{z_i}(t) &= \mathbb{E}(\exp(tz_i)) \\ &= \int_{-\infty}^\infty \exp(tz)f_{z_i}(z)dz \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp(tz)dz \end{aligned}$$

$$= \frac{\exp(\frac{t}{2}) - \exp(-\frac{t}{2})}{t}.$$

Since the z_i's are independent random variables, the moment generating function for Z_n is

$$m_{Z_n}(t) = \mathbb{E}(\exp(tZ_n)) = \prod_{i=1}^{n} \mathbb{E}(\exp(tz_i)),$$

which can be expressed in terms of the moment generating function of the z_i's. In particular

$$m_{Z_n}(t) = \prod_{i=1}^{n} m_{z_i}(t) = \left(\frac{\exp(t/2) - \exp(-t/2)}{t}\right)^n.$$

By substituting $t = -s$, we obtain the Laplace transform of Z_n

$$m_{Z_n}(-s) = \left(\frac{\exp(s/2) - \exp(-s/2)}{s}\right)^n = \frac{1}{s^n} \sum_{i=0}^{n} \binom{n}{i} (-1)^i \exp\left(\frac{s(n-2i)}{2}\right).$$

We let $c = i - \frac{n}{2}$ and then by applying inverse Laplace transform (see Table 6.2.1 in [5]) we recover the probability distribution function $f_{Z_n}(t)$ for Z_n, namely,

$$f_{Z_n}(t) = \frac{1}{(n-1)!} \sum_{i=0}^{n} \binom{n}{i} (-1)^i u_c(t)(t + \frac{n}{2} - i)^{n-1},$$

where $u_c(t)$ is the Heaviside step function, $u_c(t) = 1$ if $t \geq c$ and $u_c(t) = 0$ otherwise.

Integrating over the probability density function $f_{Z_n}(t)$ gives the cumulative distribution function,

$$\begin{aligned} F_{Z_n}(t) &= \int_{-\infty}^{t} f_{Z_n}(z)dz \\ &= \int_{-\infty}^{t} \frac{1}{(n-1)!} \sum_{i=0}^{n} \binom{n}{i} (-1)^i u_c(z)(z + \frac{n}{2} - i)^{n-1} dz \\ &= \frac{1}{(n-1)!} \sum_{i=0}^{n} \binom{n}{i} (-1)^i \int_{-\infty}^{t} u_c(z)(z + \frac{n}{2} - i)^{n-1} dz. \end{aligned}$$

Since $u_c(t) = 1$ when $t \geq i - \frac{n}{2}$, equivalently when $t + \frac{n}{2} \geq i$ and $u_c(t) = 0$ otherwise, then

$$F_{Z_n}(t) = \frac{1}{(n-1)!} \sum_{i=0}^{\lfloor t + \frac{n}{2} \rfloor} \binom{n}{i} (-1)^i \int_{i - \frac{n}{2}}^{t} (z + \frac{n}{2} - i)^{n-1} dz$$

$$= \frac{1}{n!} \sum_{i=0}^{\lfloor t+\frac{n}{2} \rfloor} \binom{n}{i} (-1)^i (z + \frac{n}{2} - i)^n \Big|_{z=i-\frac{n}{2}}^{z=t}$$

$$= \frac{1}{n!} \sum_{i=0}^{\lfloor t+\frac{n}{2} \rfloor} \binom{n}{i} (-1)^i (t + \frac{n}{2} - i)^n.$$

$\square$

Appendix B

To Generate a Random Point in a Dikin Ellipsoid

Let $A \in \mathbb{R}^{m \times n}$ be a matrix such that $P = P(A, \mathbf{1}) \subset \mathbb{R}^n$ is a polytope. Recall that for a point x in the interior of P, the Dikin ellipsoid $\mathcal{D}_x^r$ centered at x with radius r is the set

$$\mathcal{D}_x^r = \left\{ y \in \mathbb{R}^n : \left\| D(x)A(y-x) \right\|_2 \leq r \right\},$$

where $D(x) = \operatorname{diag}(\frac{1}{1-a_i^T x})$ is a diagonal $m \times m$ matrix. Further recall that each step of the Dikin walk requires that we choose a point y from $\mathcal{D}_x^r$ uniformly at random. So to sample points from P via the Dikin walk (Algorithm 4.5.9) we need a practical method for generating such points. For each $y \in \mathcal{D}_x^r$, we can write $y = z + x$ where z satisfies $\left\| D(x)Az \right\|_2 \leq r$. This demonstrating that sampling from $\mathcal{D}_x^r$ is equivalent to sampling from the ellipse $\mathcal{E} = \{z \in \mathbb{R}^n : \left\| D(x)Az \right\|_2 \leq r\}$.

The underlying idea that allows us to sample from $\mathcal{E}$ is the fact that every ellipsoid is the image of a Euclidean ball under some linear transformation. Suppose $A^T D^2(x) A = R^T D R$ is an eigenvalue decomposition. Specifically R is an $n \times n$ orthogonal matrix and D is a diagonal matrix which contains the eigenvalues of $A^T D^2(x) A$. By the positive definite structure of $A^T D^2(x) A$, the entries of D are nonnegative and so we can let $E = \sqrt{D}$. For $z \in \mathcal{E}$, the image ERz is contained in the ball $r\mathcal{B} := \{y \in \mathbb{R}^n : \|y\|_2 \leq r\}$. In fact, since ER is invertible, it defines an endomorphism on $\mathbb{R}^n$. It follows that to generate a random point from $\mathcal{E}$:

- Let $X_i \sim \mathcal{N}(0,1)$ be i.i.d. standard normal random variables and let $U \sim \mathcal{U}(0,1)$ be a random variable uniform on the interval $[0,1]$. Then

  $$X = rU^{\frac{1}{n}} \cdot \frac{(X_1, \dots, X_n)}{\sqrt{X_1^2 + \cdots + X_n^2}}$$

  is uniformly distributed over $r\mathcal{B} \subset \mathbb{R}^n$.

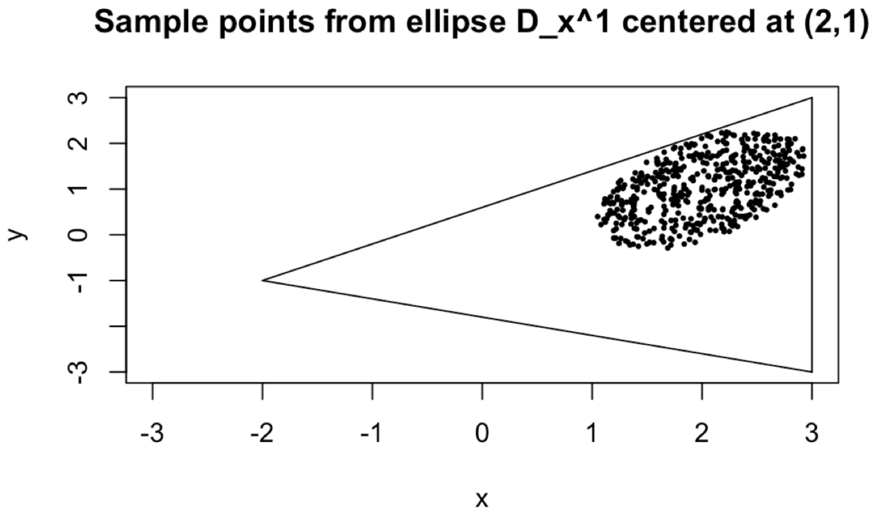

Figure B.1: Continuing Example 4.5.2, in R, we generated $N = 500$ random points from $\mathcal{D}^1_{x_0}$ where $x_0 = (2, 1)$.

- The point $Y = (ER)^{-1}X$ is then uniformly distributed over $\mathcal{E}$.

Example B.0.1. Consider the triangle from Example 4.5.2. Here we use R to generate $N = 500$ random points uniformly from $\mathcal{D}^1_{x_0}$ where $x_0 = (2, 1)$. The result is plotted in Figure B.1.

```
#Generating uniform random points in Dikin Ellipse

A0=matrix(c(1/3,-4/3,-2/9,0,5/3,-5/9),nrow=3)
b=c(1,1,1)

Dmat<-function(x,A){
  #Let Ax<=1 define a polytope. Dmat is D(x)
  b=c(matrix(1,nrow=1,ncol=dim(A)[1]))
  return (diag(c((b-A%*%x)^(-1))))
}

n=dim(A0)[2]
N=500 #number of samples
samples=matrix(0,nrow=N, ncol=n)
rad=1 #radius
```

```
center=c(2,1)

Hx<-function(x){
  t(A0)%*%Dmat(x,A0)^2%*%A0
}

#eigendecomp
decomp=eigen(Hx(center),symmetric=TRUE)
R=t(decomp$vectors)
D=diag(decomp$values)
E=sqrt(D)
ER=E%*%R

for (i in 1:N){
  u=rnorm(n)
  sca=runif(1)^(1/n)
  u=(rad*sca*u/norm(u,"2"))
  samples[i,]=solve(ER,u)+center
}
#plot sample points
plot(samples[,1],samples[,2],xlim=c(-3,3),ylim=c(-3,3),
main="Sample points from ellipse D_x^1 centered at (2,1)")
```