# Mixing Time of a Rook's Walk

Steven S. Kim

April 2012

Advisor:
Michael Damron

Independent work for undergraduate certificate,
Program in Applied and Computational Mathematics

# Contents

# 0 Introduction

Markov chains, or Markov processes with finite (or countable) state space and discrete time scale, have long been studied for their many applications in engineering and the sciences. Ergodic chains in particular are known to eventually reach stationary distribution, but a natural question arises of how quickly a Markov chain gets "close" to its invariant distribution.

In this work, I will first briefly review the theory of mixing times, and then apply one particular technique of bounding mixing time to a Markov chain called the "rook's walk". Consider the movement of a chess piece on an otherwise empty board, and at each time, let it move randomly to any of its valid positions. I consider the rook in particular for three reasons.

i. There is a potentially interesting application to a specific MCMC sampling algorithm that will be discussed in a later section.
ii. It is particularly natural to generalize the movement of a rook to larger chessboards and higher dimensions. Compare this to a piece like a queen, for which the "diagonal" movement would have to be specified in higher dimensions.
iii. The rook does not suffer from the "boundary" conditions that a knight or a king might. For example, a rook on a standard $8 \times 8$ chessboard has 15 valid positions to move to, even if it is at the edge of the board; on the other hand, a knight or a king has very different valid moves when it is at the center as compared to the edge.

In preparing this independent work requirement for the Program of Applied and Computational Mathematics, most of my literature review was done in Spring 2011 (my junior year), as independent reading with Ramon van Handel from ORFE. I then chose to continue a small project with Mike Damron, who has provided many helpful hints in our conversations.

# 1 Review of Mixing Times

Consider a finite or countable set $S$ called the *state space*. A vector $(\alpha_x)_{x\in S}$ is called a *distribution* if $\alpha_x \geq 0$ and $\sum_{x\in S} \alpha_x = 1$. A matrix $p = (p_{xy})_{x,y\in S}$ is called *stochastic* if all of its rows are distributions.

Definition 1.1 (Markov chain). In a probability space $(\Omega, \mathcal{F}, \mathrm{P})$, a collection of random variables $(X_t)_{t\in\mathbb{N}}$ where $X_t : (\Omega, \mathcal{F}) \to (S, 2^S)$ is called a *Markov chain* with initial distribution $\alpha$ and transition matrix $p$ if

- $\mathrm{P}(X_0 = x) = \alpha_x$ for all $x \in S$
- $\mathrm{P}(X_{t+1} = x_{t+1} | X_0 = x_0, \cdots, X_t = x_t) = p_{x_t x_{t+1}}$ for all $x_0, \cdots, x_{t+1} \in S$

As shorthand, we write $(X_t)_{t\in\mathbb{N}} \sim \mathrm{Mkv}(\alpha, p)$.

Note that a Markov chain is just a special case of a Markov process. That is, for $s < t$,

and any $A \in 2^S$,

$$\mathrm{P}\left(X_t \in A | \mathcal{F}_s\right) = \mathrm{P}\left(X_t \in A | X_s\right) \tag{1.1}$$

The behavior of this process is fully specified by $\mathrm{P}\left(X_t \in A | X_s\right)$, which defines the transition matrix $p$. While continuous-time, uncountable state space Markov processes have their own elegance about them, the convenience of being able to define a Markov chain through its finite transition matrix $p$ leads to some nice results, particular regarding the invariant distribution.

DEFINITION 1.2. Let $(X_t)_{t\in\mathbb{N}} \sim \mathrm{Mkv}(\alpha, p)$. Write $p_{xy}^{(t)}$ for the element in the $x$-th row, $y$-th column of $p^t$.

- $(X_t)$ is *aperiodic* if for each state $x \in S$, there exists some $t_0$ such that $p_{xx}^{(t)} > 0$ for all $t \geq t_0$.
- $(X_t)$ is *irreducible* if for any two states $x, y \in S$, there exists some $t$ such that $p_{xy}^{(t)} > 0$.
- $(X_t)$ is *ergodic* if it is aperiodic and irreducible.
- A vector $\pi$ that satisfies $\sum_i \pi_i = 1$ and $\pi p = \pi$ is called an *invariant distribution.*

PROPOSITION 1.3 ([LPW09], see Corollary 1.17). *If $(X_t)_{t\in\mathbb{N}} \sim \mathrm{Mkv}(\alpha, p)$ is irreducible, and the state space $S$ is finite, there exists a unique invariant distribution.*

THEOREM 1.4. *If $(X_t)_{t\in\mathbb{N}} \sim \mathrm{Mkv}(\alpha, p)$ is ergodic with invariant distribution $\pi$, then for all $x, y \in S$,*

$$\lim_{t\to\infty} p_{xy}^{(t)} = \pi_y \tag{1.2}$$

*Proof.* A proof can be found in [LPW09], one that is based mostly on linear algebra manipulations. There is, however, a well-known alternative proof that relies on the idea of coupling, of which we will provide a sketch here.

Regardless of the initial state $X_0$, consider a different Markov chain $(Y_t)_{t\in\mathbb{N}} \sim \mathrm{Mkv}(\pi, p)$. For some state $z \in S$, let $\tau = \inf\{t \geq 0 : X_t = Y_t = z\}$. First, note that $\mathrm{P}(\tau < \infty) = 1$ since $(X_t, Y_t)_{t\in\mathbb{N}}$ is a Markov chain that's aperiodic and irreducible, just with state space $S \times S$. Now consider the process $(Z_t)_{t\in\mathbb{N}}$ where $Z_t = X_t$ for $t \leq \tau$, and $Z_t = Y_t$ for $t > \tau$. Note that $(Z_t)_{t\in\mathbb{N}} \sim \mathrm{Mkv}(\alpha, p)$, so

$$\begin{aligned}
|\mathrm{P}(X_t = x) - \pi_x| &= |\mathrm{P}(Z_t = x) - \mathrm{P}(Y_t = x)| && (1.3)\\
&= |\mathrm{P}(X_t = x, t < \tau) - \mathrm{P}(Y_t = x, t < \tau)| && (1.4)\\
&\leq \mathrm{P}(t < \tau) && (1.5)\\
&\xrightarrow{t\to\infty} 0 && (1.6)
\end{aligned}$$

□

Let us return to the question of "closeness" to stationarity. The concepts of invariance and stationary are greatly simplified in the case of Markov chains due to the finite state space. We will simply use a concept of distance that looks at how different $p^t$ is from $\pi$ in the worst case.

DEFINITION 1.5. For two probability measures $\mu, \nu$ on a measurable space $(\Omega, \mathcal{F})$, the

*total variation distance* is defined to be

$$\|\mu - \nu\|_{\mathrm{TV}} := \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)| \tag{1.7}$$

PROPOSITION 1.6. *For two probability measures $\mu, \nu$ on a measurable space $(S, 2^S)$ where $S$ is finite,*

$$\|\mu - \nu\|_{\mathrm{TV}} = \tfrac{1}{2} \sum_{x \in S} |\mu(x) - \nu(x)| \tag{1.8}$$

DEFINITION 1.7. Where $(p^t)_x$ is the $x$-th row of a transition matrix $p^t$ with invariant distribution $\pi$, we define the *distance to stationarity*

$$D(t) := \max_{x \in S} \|(p^t)_x - \pi\|_{\mathrm{TV}} \tag{1.9}$$

DEFINITION 1.8. For $\epsilon > 0$, we define the *mixing time* of a Markov chain,

$$t_{\mathrm{mix}}(\epsilon) := \inf\{t \in \mathbb{N} : D(t) < \epsilon\} \tag{1.10}$$

# 2 The Rook's Walk

Imagine the random walk of a rook on a chessboard. At each time step, it moves from its current position to any of its valid positions with equal probability. The random walk of a king would be similar to a simple symmetric random walk, in that it could only move to nearest neighbor points. The convenience of the rook is that we can actually think of the chessboard as a torus rather than a square, because the rook's ability to go to any point in the same row essentially "connects" the ends of each row together. While 3-dimensional (or any sort of higher-dimensional) chess is not a commonly played game, we will consider the behavior of a rook moving randomly on a more general chessboard nonetheless. This type of walk first appeared in [AK91], but our main goals will be: to precisely define the rook's walk, and to work out the details of a mixing time bound that was mentioned but not fully discussed in [AF02].

DEFINITION 2.1. We denote by $\mathbb{Z}_n^d := \{1, \cdots, n\}^d$ the *d-dimensional integer torus of length n*.

DEFINITION 2.2. The *rook's walk on* $\mathbb{Z}_n^d$ is the Markov chain $(X_t)_{t \in \mathbb{N}} \sim \mathrm{Mkv}(\alpha, p)$ on $S = \mathbb{Z}_n^d$, where

$$p_{xy} = \tfrac{1}{d(n-1)} \mathbf{1}_{\{\|y-x\|_0 = 1\}} \tag{2.1}$$

where $\|x\|_0 = \sum_{i=1}^d \mathbf{1}_{\{x_i \neq 0\}}$ is the Hamming distance.

PROPOSITION 2.3. *For the rook's walk on $\mathbb{Z}_n^d$ where $n > 2$,*

$$\tfrac{d(n-1)}{n} \log\left(\tfrac{1}{2\epsilon}\right) \leq t_{\mathrm{mix}}(\epsilon) \leq \tfrac{d(n-1)}{n} \log\left(\tfrac{n^d}{\epsilon}\right) \tag{2.2}$$

There are several well-studied techniques and concepts used to bound distance and find mixing times, including: coupling, conductance, bounding by hitting times and strong stationary times. Many of these arguments are discussed in-depth in [AF02] [LPW09]. However, our proof of Proposition 2.3 will use the idea of spectral gap and relaxation time

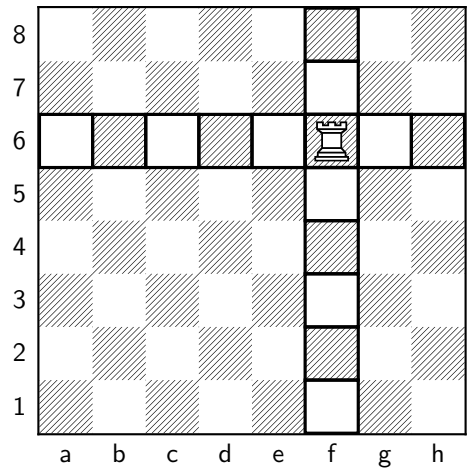



Figure 1: Rook on a standard $8 \times 8$ chessboard, and its possible moves

discussed in [AF02]. Let's first introduce the concept of Markov chains that are in some way the same "forwards" as they are "backwards".

DEFINITION 2.4. If $(X_t)_{t\in\mathbb{N}} \sim \text{Mkv}(\alpha, p)$ has invariant distribution $\pi$ such that

$$\pi_x p_{xy} = \pi_y p_{yx} \tag{2.3}$$

for all $x, y \in S$, the Markov chain is said to be *reversible*.

LEMMA 2.5. *For $n > 2$, the rook's walk on $\mathbb{Z}_n^d$ is ergodic, reversible, and has unique invariant distribution $\pi$ where $\pi_x = \frac{1}{n^d}$ for each $x \in S$.*

*Proof.* Let's show these properties step-by-step.

- *ergodicity:* The rook's walk is trivially aperiodic, since for $n > 2$, the rook can move from state $x$ back to state $x$ in any amount of time $t > 1$ simply by moving back and forth between two states that are in the same row as $x$ until time $t$ is reached. It is also trivially irreducible, since a rook can move from state $x$ to state $y$ in at most $d$ steps (where $d$ is the dimension of the board).
- *invariant distribution:* The rook's walk is symmetric, since if two vertices $x, y \in \mathbb{Z}_n^d$ are in the same "row", then $p_{xy} = p_{yx} = \frac{1}{d(n-1)}$. If two vertices do not share a row, then $p_{xy} = p_{yx} = 0$. Thus, for any $y \in S$, we know $\sum_x p_{xy} = \sum_x p_{yx} = 1$. That is, all the columns of $p$ sum to 1. Thus, the invariant distribution is $\pi$ since for any $x \in S$,

$$\sum_x \pi_x p_{xy} = \tfrac{1}{n^d} \sum_x p_{xy} = \tfrac{1}{n^d} = \pi_y \tag{2.4}$$

  We know this $\pi$ is unique since $S$ is finite.
- *reversibility:* For any $x, y \in S$, we know $\pi_x = \pi_y$ and $p_{xy} = p_{yx}$, so the rook's walk is reversible.

□

LEMMA 2.6 ([LPW09], see Lemma 12.1). *For the Markov chain* $(X_t)_{t\in\mathbb{N}} \sim \text{Mkv}(\alpha, p)$ *on finite* $S$*,*

- *If* $\lambda$ *is an eigenvalue of* $p$*, then* $|\lambda| \leq 1$*.*
- *If* $p$ *is irreducible, the vector space of eigenvectors corresponding to the eigenvalue 1 is the one-dimensional space generated by* $(1, 1, \cdots, 1)^T$*.*
- *If* $p$ *is ergodic, then* $-1$ *is not an eigenvalue of* $p$*.*

DEFINITION 2.7. Let $p$ be a reversible transition matrix, with eigenvalues indexed in decreasing order

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{|S|} \geq -1 \tag{2.5}$$

Let $\lambda_* := \max\{|\lambda| : \lambda \text{ eigenvalue of } p,\ \lambda \neq 1\}$. We call $\gamma_* := 1 - \lambda_*$ the *absolute spectral gap*. We call $t_{\text{rel}} := 1/\gamma_*$ the *relaxation time*.

Note that the Lemma above implies that for ergodic Markov chains, $\lambda_* \neq -1$, so $\gamma_* > 0$. There is also a similar concept that does not rely on the absolute value, the *spectral gap* $\gamma := 1 - \lambda_2$. It's easy to show a connection between $\gamma$ and $\gamma_*$. For example, consider a "lazy" version of a Markov chain. That is, a new chain where at each time step, we stay with probability $\frac{1}{2}$, and move as the old chain would with probability $\frac{1}{2}$. This is a Markov chain with transition matrix $q = \frac{p+I}{2}$. For any eigenvector $v$ and eigenvalue $\lambda$ of $q$,

$$\lambda v = qv = \tfrac{pv+v}{2} \Rightarrow (2\lambda - 1)v = pv \tag{2.6}$$

Thus, $(2\lambda - 1)$ is an eigenvalue of $p$, but by the Lemma above, we must have $|2\lambda - 1| \leq 1$, meaning $\lambda \geq 0$. That is, for this lazy version of the Markov chain with transition matrix $q = \frac{p+I}{2}$, we know all the eigenvalues are positive, meaning $\gamma_* = \gamma$.

THEOREM 2.8 ([LPW09], Theorem 12.3, 12.4). *For a reversible and irreducible Markov chain with state space* $S$*,*

$$t_{\text{mix}}(\epsilon) \leq t_{\text{rel}} \log\left(\tfrac{1}{\epsilon \min_{x\in S} \pi_x}\right) \tag{2.7}$$

*For a reversible and ergodic Markov chain,*

$$t_{\text{mix}}(\epsilon) \geq (t_{\text{rel}} - 1) \log\left(\tfrac{1}{2\epsilon}\right) \tag{2.8}$$

With the above theorem in mind, all we need to do to prove Proposition 2.3 is to calculate $t_{\text{rel}}$ for the rook's walk. To do this, we can think of the rook's walk on $\mathbb{Z}_n^d$ as $d$ simultaneous random walks on complete graphs $K_n$. Each dimension of the chessboard looks like a complete graph to the rook, since within a row, a rook can move to any point in just one move.

To be more precise, consider for each $j = 1, \cdots, d$ an irreducible transition matrix $\tilde{p}_j$ on the state space $\tilde{S}_j$ with invariant distribution $\tilde{\pi}_j$. Let $w$ be a probability distribution on $\{1, \cdots, d\}$. Consider the *product chain* on $S = \tilde{S}_1 \times \cdots \times \tilde{S}_d$ which selects $j$ according to distribution $w$, and moves only in the $j$-th coordinate with transition probabilities $\tilde{p}_j$.

For the rook's walk, let's write $X_t = (X_t^1, \cdots, X_t^d)$ where $X_t^j$ represents the $j$-th coordinate of the rook's position at time $t$. Note that $(X_t^1, \cdots, X_t^d)$ is a Markov chain on $S = \tilde{S}_1 \times \cdots \times \tilde{S}_d$ where $\tilde{S}_j = \mathbb{Z}_n$ for each $j$. The rook's walk selects a coordinate

$j \in \{1, \cdots, d\}$ uniformly at random, meaning $w_j = \frac{1}{d}$, and then changes only the $j$-th coordinate according to the transition matrix $\tilde{p}_j$ where for $x, y \in \mathbb{Z}_n$,

$$\tilde{p}_{j,xy} = \tfrac{1}{n-1}\mathbf{1}_{\{x \neq y\}} \tag{2.9}$$

It's clear that each component chain is just a random walk on a complete graph of order $n$, which has nice properties. The following theorem tells us what we can say about the product chain given what we know about each component chain.

LEMMA 2.9 ([LPW09], Lemma 12.11). *Consider transition matrices $\tilde{p}_j$ on space $\tilde{S}_j$ for $j = 1, \cdots, d$. Let $w$ be a probability distribution on $\{1, \cdots, d\}$. Define the transition matrix $p$ where for $x, y \in S = \tilde{S}_1 \times \cdots \times \tilde{S}_d$,*

$$p_{xy} = \sum_{j=1}^{d} w_j \tilde{p}_{j,x_j y_j} \prod_{i \neq j} \mathbf{1}_{\{x_i = y_i\}} \tag{2.10}$$

*Then eigenvalues of $p$ are of the form $\lambda = \sum_{j=1}^{d} w_j \tilde{\lambda}_j$ where $\tilde{\lambda}_j$ is an eigenvalue of $\tilde{p}_j$.*

LEMMA 2.10. *The eigenvalues of the transition matrix of the random walk on the complete graph of order $n$ – that is, the stochastic process $(X_t)_{t \in \mathbb{N}} \sim \mathrm{Mkv}(\alpha, \tilde{p})$ on $\tilde{S} = \mathbb{Z}_n$ where $\tilde{p}$ is as in (2.9) – has eigenvalue 1 with multiplicity 1, and eigenvalues $\frac{-1}{n-1}$ with multiplicity $n-1$.*

*Proof.* Consider the $n \times n$ matrix $q$ of all ones,

$$q = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \tag{2.11}$$

Note that $\tilde{p} = \frac{1}{n-1}(q - I)$. Thus, if $q$ has eigenvalue $\lambda$ for eigenvectors $v_1, \cdots, v_j$, then $\tilde{p}$ has eigenvalue $\frac{1}{n-1}(\lambda - 1)$ for the same eigenvectors $v_1, \cdots, v_j$.

Let's consider possible values of $\lambda$ where $qv = \lambda v$ for $v \in \mathbb{R}^n$. Note that we can write $qv = (v_1 + \cdots + v_n)\mathbf{1}_n$ where $\mathbf{1}_n$ is an $n$-dimensional column vector of ones. One possible solution is $\lambda = n$ with eigenvector $v$ where $v_1 = v_2 = \cdots = v_n$, a space spanned by a single vector. Another possible solution is $\lambda = 0$, meaning $v_1 + \cdots + v_n = 0$, a solution space with dimension $n-1$. Thus, $q$ has eigenvalue $n$ with multiplicity 1, and 0 with multiplicity $n-1$, proving our result.

□

*Proof (of Proposition 2.3).* Note that the transition matrix $p$ for the rook's walk as defined in (2.1) is of the form (2.10) with $w_j = \frac{1}{n^d}$ and $\tilde{p}_j$ as in (2.9), since

$$p_{xy} = \tfrac{1}{d(n-1)}\mathbf{1}_{\{\|y-x\|_0 = 1\}} = \tfrac{1}{d(n-1)} \sum_{j=1}^{d} \mathbf{1}_{\{x_j \neq y_j\}} \prod_{i \neq j} \mathbf{1}_{\{x_i = y_i\}} = \sum_{j=1}^{d} w_j \tilde{p}_{j,x_j y_j} \prod_{i \neq j} \mathbf{1}_{\{x_i = y_i\}}$$

By Lemmas 2.9 and 2.10, the second largest eigenvalue of $p$ (both in absolute value and in

natural order) is

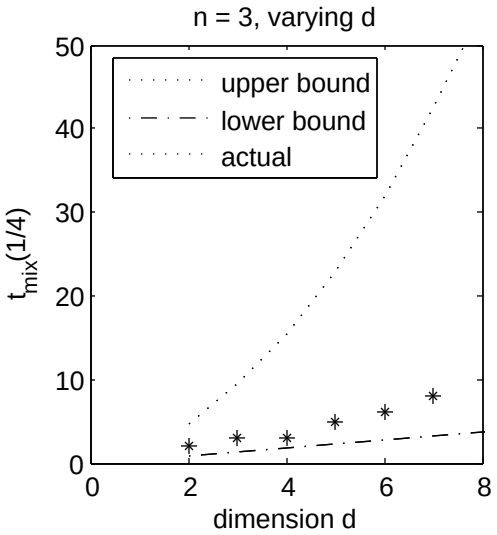
$$\lambda_* = \lambda_2 = \tfrac{1}{d}\left[\tfrac{-1}{n-1} + \sum_{i=1}^{d-1} 1\right] = \tfrac{1}{d}\left[d - \left(1 + \tfrac{1}{n-1}\right)\right] = 1 - \tfrac{n}{d(n-1)}$$

Thus, $\gamma_* = \frac{n}{d(n-1)}$, and $t_{\text{rel}} = \frac{d(n-1)}{n}$. Since the rook's walk is reversible and ergodic for $n > 2$ by Lemma 2.5, and we know $\min_{x \in S} \pi_x = \frac{1}{n^d}$, then by Theorem 2.8, the result of Proposition 2.3 follows. □

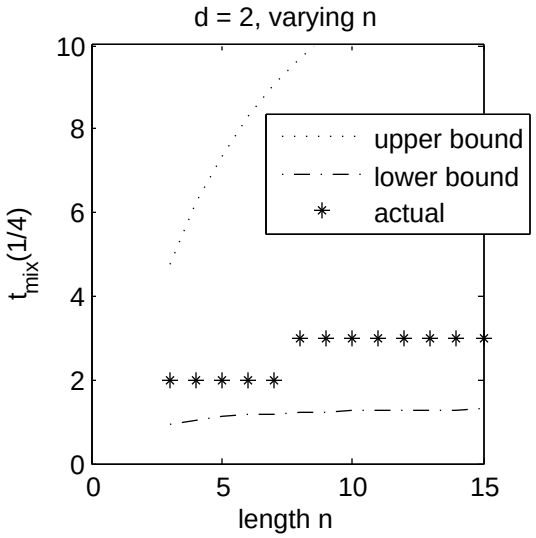
Our method for bounding $t_{\text{mix}}$ differs slightly from the method by Aldous and Fill in [AF02]. They consider only the case $d = 2$, and then transform the rook's walk into a continuous time version by "Poissonizing". That is, where $N(t)$ is a Poisson process with rate 1, they consider the process $Y_t := X_{N(t)}$. Using the concept of product chains in continuous time (essentially, all that needs to be done to multiply $d$ walks together is to slow down each random walk by a factor of $d$), they explicitly derive formulae for the conditional probability $\mathrm{P}(Y_t = j | Y_0 = i)$, and a modified distance $\bar{D}(t) := \max_{i,j \in S} \|(p^t)_i - (p^t)_j\|_{\text{TV}}$. Their formula for $\bar{D}(t)$ is of the form $\bar{D}(t) = a_1 e^{b_1 t/2} + a_2 e^{b_2 t}$, which they then inverted to find $t_{\text{mix}}$. However, this result is limited, since it is much nastier to invert such a function for $d > 2$.

# 3 Computations

In this section we will show a few computations for actual values of mixing time, obtained simply by matrix multiplication to explicitly determine $t_{\text{mix}}$ for $\epsilon = \frac{1}{4}$. Our bound from Proposition 2.3 certainly looks nice in that it is polynomial and relatively low order in both $n$ and $d$, but the plots below will show that this bound is fairly loose, and the actual rook's walk seems to mix quite quickly.



For increasing $d$, the Rook's walk seems to scale nicely, but our bound (of order $d^2$) doesn't seem to be too tight.




For increasing $n$, the Rook's walk seems to mix in at most $d = 3$ time steps, regardless of how big $n$ gets. This makes sense for the rook, since on the $\mathbb{Z}_n^d$ chessboard, it can reach any other point in at most $d$ steps. Especially for larger $n$, this should be quite different from the "king's walk", the simple nearest-neighbor random walk on $\mathbb{Z}_n^d$.

The code for computing the mixing times was written in MATLAB, and is attached below:

```
function [ tmix ] = RookSim(n,d)

% n = side length of chessboard
% d = dimensions
states = n^d; % number of positions
indices = n*ones(1,d); % indices per dimension
p = zeros(states, states); % initialize transition matrix

% fill in transition matrix p
for i=1:states
    ivec = cell(1,d); % convert one-dim index i to d-dim indices
    [ivec{:}] = ind2sub(indices,i);
    for dim=1:d
        for ind=1:n
            % change p_{ij} iff j is accessible from i by a rook
            jvec = [ivec{:}];
            jvec(dim) = ind;
            if sum(jvec - [ivec{:}]) ~= 0
                j = polyval(jvec-1,n) + 1;
                p(i,j) = 1/(d*(n-1));
            end
        end
    end
```

```
end

% measure mixing time
T = 50;
Dis = zeros(1,T);
invpi = 1/states * ones(1,states);
q = eye;
for t=1:T
    q = q * p;
    Dis(t) = max(1/2 * sum(abs(q-ones(states,1)*invpi),2));
    if Dis(t) < 0.25
       tmix = t;
       break;
    end
end
```

# 4 Applications

The simplest way to sample from a one-dimensional distribution $F$ is just to simulate a uniform random variable $U$ on $[0, 1)$, and to take the inverse cdf $F^{-1}(U)$. There is also a technique known as *rejection sampling*, which samples some $z$ uniformly from a cube, and accepts $z$ if it is in the area under the graph of the density $f(x) = F'(x)$, and rejects otherwise. Neither of these techniques works particularly well for higher-dimensional distributions.

In [AK91], Applegate and Kannan describe and prove the efficiency of an MCMC algorithm that samples from multi-dimensional, log-concave distributions; essentially the algorithm is to run a random walk on a cube like $\mathbb{Z}_n^d$, and accept that point with a certain probability.

Consider a cube $A \in \mathbb{R}^n$ and a function $f : A \to \mathbb{R}_+$ where $\phi(x) = \log f(x)$ is concave and smooth ([AK91] introduces parameters $\alpha, \beta$ to explicitly quantify the level of concavity and smoothness). Note that the Gaussian distribution and members of the exponential family fall under the condition of log-concavity. In particular, the Gaussian distribution is basically defined on a compact interval, since for points sufficiently far from the mean, their probability can be taken as zero.

Say I want to sample according to $f$; that is, pick $x$ such that the probability $p(x)$ of picking $x$ is proportional to $f(x)$. For $\epsilon > 0$, we want for all $x \in A$ and some $c$ constant,

$$|p(x) - cf(x)| < \epsilon \tag{4.1}$$

For $\gamma > 0$, let $L_\gamma = \{x : x_i \text{ integer multiple of } \gamma \quad \forall i\}$; we want to generate samples from $L \cap A$. Note that we want $\gamma$ to be as small as possible to let $L_\gamma$ be as fine as possible. Divide $A$ into cubes of side length $\delta$ (in [AK91], $\delta$ is inversely proportional to the "smoothness" level $\alpha$). Let $V$ be the set of centers of those cubes. Let $C_x$ be the cube in $\mathbb{R}^n$ with center $x$. Now let's introduce a random walk on $V$ with transition matrix $p$ such that for $x, y \in V$,

- $p_{xy} = 0$ if $C_x$ and $C_y$ don't share a face
- $p_{xy} = \frac{1}{4n} \min\left(1, \frac{f(y)}{f(x)}\right)$ if they share a face, and $x \neq y$
- $p_{xx} = 1 - \sum_{y \neq x} p_{xy}$

The invariant distribution is $\pi_x = \frac{f(x)}{\sum_{y \in V} f(y)}$. Note that this is essentially a lazy version of a nearest-neighbor random walk (a "king's walk").

Pick $x_0 \in V$, run the random walk above until the time is greater than $t_{\text{mix}}(\epsilon)$ for some predetermined level of accuracy $\epsilon > 0$, to get some point $x$. Pick point $z$ uniformly from $L \cap C_x$. With probability $\frac{f(z)}{ef(x)}$, output $z$; otherwise, start the random walk over again.

Applegate and Kannan prove that the probability of obtaining $z$ is indeed proportional to $f(z)$, and this algorithm is polynomial time. Indeed, they show a similar algorithm can be used to integrate $f(x)$ over the cube $A$, and also to find the volume of any convex body in $\mathbb{R}^n$. However, note that one of the fundamental steps is running the random walk until mixing time, and it's possible that the rook's walk would mix faster for higher dimensions.

The bound for the simple nearest-neighbor random walk in [AK91] is found using *conductance* arguments, which essentially attaches a weight (the probability of the Markov chain going across that edge) to each edge, and bounding mixing time through that. This is bounded through a "weighted isoperimetry" argument, which compares the square of circumference of a closed curve to the area of the region it encompasses. Unfortunately, this relationship between the geometric surface area and conductance is a little trickier for the rook's walk, since the rook's walk can essentially "skip" to non-adjacent boxes.

# 5 Future Directions

Given more time to explore, there are many directions to go from here. The first would be to try to bound $t_{\text{mix}}$ using a conductance argument as in [AK91], but with some other isoperimetric inequality. There is literature on isoperimetric inequalities for products of graphs [CT98], and isoperimetric inequalities to bound random walks on graphs [Bar10], that was not fully explored in this project.

There is also an interesting question of sharp threshold for Markov chains: that is, for a sequence of Markov chains $(X_n)$, for large $n$, does the mixing time show a sharp threshold, such that before $t_{\text{mix}}$, it is not very mixed, but afterwards, it is? A sequence of Markov chains with mixing time $t_{\text{mix}}^{(k)}(\epsilon)$ for the $k$-th chain is defined to have *cutoff* if for any $\epsilon > 0$,

$$\lim_{k \to \infty} \frac{t_{\text{mix}}^{(k)}(\epsilon)}{t_{\text{mix}}^{(k)}(1-\epsilon)} = 1 \tag{5.1}$$

A known result is that such a sequence has cutoff if and only if

$$\lim_{k \to \infty} D_k(ct_{\text{mix}}^{(k)}) = \begin{cases} 1 & \text{if } c < 1 \\ 0 & \text{if } c > 1 \end{cases} \tag{5.2}$$

where $D_k$ is the distance to stationarity for the $k$-th in a sequence of Markov chains. With our current bound in Proposition 2.3, it's not possible to say whether the family of rook's walks displays the cutoff phenomenon (either for increasing $n$ or increasing $d$). If the rook's walks do in fact have a cutoff, there are additional questions to consider, such as the concept of "window width", the order of magnitude of the time a rook's walk on $\mathbb{Z}_n^d$ takes to go from not very mixed to very mixed.

# References


[AF02] D. Aldous and J. Fill. Reversible markov chains and random walks on graphs, 2002.

[AK91] D. Applegate and R. Kannan. Sampling and integration of near log-concave functions. In *Proceedings of the Twenty-Third Annual ACM Symposium on Theory of Computing*, pages 156–163. ACM, 1991.

[Bar10] M.T. Barlow. Random walks on graphs: a brief introduction. Lecture notes from RIMS Kyoto and Cornell Probability Summer School, 2010.

[CT98] F.R.K. Chung and P. Tetali. Isoperimetric inequalities for cartesian products of graphs. *Combinatorics Probability and Computing*, 7(2):141–148, 1998.

[LPW09] D.A. Levin, Y. Peres, and E.L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2009.