involve
a journal of mathematics

Mixing times for the rook's walk via path coupling

Cam McLeman, Peter T. Otto, John Rahmani and Matthew Sutter

msp

2017 vol. 10, no. 1

INVOLVE 10:1 (2017)
dx.doi.org/10.2140/involve.2017.10.51

Mixing times for the rook's walk via path coupling

Cam McLeman, Peter T. Otto, John Rahmani and Matthew Sutter

(Communicated by John C. Wierman)

The mixing time of a convergent Markov chain measures the number of steps required for the state distribution to be within a prescribed distance of the stationary distribution. In this paper, we illustrate the strength of the probabilistic technique called coupling and its extension, path coupling, to bound the mixing time of Markov chains. The application studied is the rook's walk on an n^d-chessboard, for which the mixing time has recently been studied using the spectral method. Our path-coupling result improves the previously obtained spectral bounds and includes an asymptotically tight upper bound in n for the two-dimensional case.

1. Introduction

In the standard game of chess, a rook occupies one of the 8^2 squares of the board, and moves by translating any distance along the row or column of its current position.

Suppose one came across a chessboard as in Figure 1 and was told that two legal rook moves had just been completed. Clearly there is no way to definitively deduce from this scant information what the position of the rook had been before the two moves — any of the 64 squares are possibilities — but could one guess the correct square in such a way as to have better odds than $\frac{1}{64}$? Certainly! Of the $14^2 = 196$ two-move sequences that a rook can undertake from a given position, there are 14 that would return it to its original square, six more that would take it to any given square in the same row or column, and only two such sequences to take the rook to any of the 49 remaining squares. The optimal strategy, over four times better than guessing a square uniformly at random, is to guess that the two moves left the rook in the same square as it started.

In the language of probabilistic processes, we might say that the chessboard is *unmixed* after only two moves — the location of the rook after two moves is heavily influenced by its starting position. In many applications of such processes, one is interested in the time it takes — the *mixing time* — in order for the influence of the starting position to be marginalized. In our chessboard example, this would

MSC2010: 60J10.
Keywords: Markov chains, mixing time, rook's walk, path coupling.

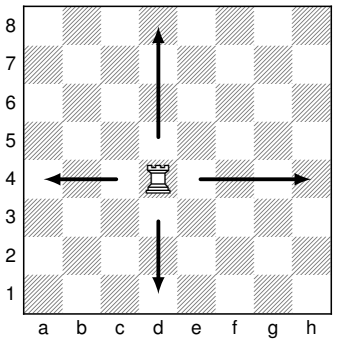

Figure 1. The 14 legal rook moves on a standard 8×8 chessboard.

be the number t such that after t random rook moves, its position is essentially governed by the uniform distribution. (The word "essentially" is needed here since the effect of the starting position never *completely* disappears — we insist only that the distribution be sufficiently close to the uniform distribution, using metrics introduced in the next section).

A standard approach to finding and bounding mixing times for processes like the rook's walk is via the spectral method: one writes down a 64×64 transition matrix encoding the squares of the chessboard and the probabilities of making moves between them, and one can then bound the mixing time in terms of the eigenvalues of this matrix. Such a method was used in [Kim 2012] to bound the mixing time of a rook's walk on the more general n^d-chessboard — a d-dimensional chessboard comprised of n squares in each dimension, on which a rook can move along precisely one dimension. Subsequently, in [Li and Tucker 2014], the exact mixing time for the rook's walk on n^2-chessboards was derived by direct computation, and the exact results showed how loose the spectral-method bounds were.

Here we take another approach to bounding mixing times from above. Namely, we use the *path coupling method* rather than spectral methods, leading to a much tighter upper bound. In Section 2, we introduce the basic theory of mixing times and Markov chains, including the metric one uses to measure the distance between distributions. This leads us to a formal definition of the mixing time. Section 3 specializes this material to the context of the rook's walk, and surveys the current state of knowledge concerning mixing times for the rook's walk. Sections 4 and 5 respectively introduce and implement path coupling in the context of the rook's walk, and Section 6 consists of an analysis and exhibition of the mixing times thus achieved.

2. Mixing times of Markov chains

The rook's walk studied in this paper is an example of a finite, discrete-time Markov chain, which, by definition, is a sequence of random variables $(X_t)_{t=0}^{\infty}$ that take

values in a finite state space Ω and satisfy the *Markov* (or memoryless) property

$$\mathbf{P}\big(X_{t+1}=y \mid X_0=x_0, X_1=x_1, \ldots, X_t=x\big) = \mathbf{P}(X_{t+1}=y \mid X_t=x).$$

The Markov property means that what happens next (i.e., in the $(t+1)$-th step) only depends on the state of the current step ($X_t = x$), and not the previous steps. Consequently, Markov chains are fully defined by their one-step transition probabilities $\mathbf{P}(X_{t+1}=y \mid X_t=x)$. We will also only be interested in time-homogeneous Markov chains, i.e., those for which

$$\mathbf{P}(X_{t+1}=y \mid X_t=x) = \mathbf{P}(X_{s+1}=y \mid X_s = x) \quad \text{for all } s, t.$$

Therefore, when discussing the one-step transition probabilities, one can just consider the case $t = 0$, that is, $\mathbf{P}(X_1=y \mid X_0=x)$. See [Kemeny et al. 1976] for a reference on the general theory of Markov chains.

The $|\Omega| \times |\Omega|$ matrix P consisting of entries $P(x, y) = \mathbf{P}(X_1 = y \mid X_0 = x)$ is called the (one-step) *transition matrix* of the Markov chain. Note that P is a stochastic matrix which means that its entries are nonnegative and

$$\sum_{y\in\Omega} P(x, y) = 1 \quad \text{for all } x \in \Omega.$$

The x-th row of the transition matrix P is the distribution $P(x, \cdot)$ on Ω. By the Chapman–Kolmogorov equations, we know that the t-step transition probabilities $\mathbf{P}(X_t=y \mid X_0=x)$ are just the entries in the matrix P^t.

Two important properties of Markov chains are irreducibility and periodicity. A Markov chain with transition matrix P is *irreducible* if for any two states $x, y \in \Omega$, there exists t (possibly depending on x and y) such that $P^t(x, y) > 0$. In other words, for an irreducible Markov chain, it is possible to reach any state from any other state in the state space Ω. Let $\mathcal{T}_x = \{t \geq 1 : P^t(x, x) > 0\}$ be the set of times when it is possible for a Markov chain to return to its starting state x. We call the greatest common divisor of $\mathcal{T}_x$ the *period* of state x. If a Markov chain consists of only states with period 1 it is called *aperiodic*; otherwise, it is called a *periodic chain*. The assumptions of irreducibility and aperiodicity lead to the fundamental result of Markov chains, called the convergence theorem, stated below.

Theorem 2.1 [Levin et al. 2009, Theorem 4.9]. *For an irreducible and aperiodic Markov chain on a finite state space Ω with transition matrix P, there exists a unique probability distribution π on Ω, called its* stationary distribution, *such that $\pi P = \pi$ and $P^t(x, \cdot)$ converges to π as $t \to \infty$ for all initial states x.*

By the definition of the stationary distribution, if P is symmetric, then π is the uniform distribution on Ω since P is a stochastic matrix.

The mixing time is then a measure of the convergence rate of the chain to its stationary distribution, quantified in terms of a choice of a metric. As per the standard convention, we adopt the *total variation distance*: given two distributions μ and ν on a common state space Ω, the total variation distance is defined by

$$\|\mu - \nu\|_{\mathrm{TV}} = \sup_{A \subset \Omega} |\mu(A) - \nu(A)| = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|. \tag{2-1}$$

Using this metric, we define the *maximal distance to stationary* of a Markov chain with transition matrix P to its stationary distribution π to be

$$d(t) = \max_{x \in \Omega} \| P^t(x, \cdot) - \pi \|_{\mathrm{TV}}.$$

Then, given any $\varepsilon > 0$, the *mixing time* of the Markov chain is defined by

$$t_{\mathrm{mix}}(\varepsilon) = \min\{t : d(t) \le \varepsilon\}.$$

We note that it is somewhat conventional to fix a particular value of ε, and often specifically the value $\varepsilon = \frac{1}{4}$, when comparing mixing-time results. We too will adhere to this convention at times.

Finally, as we will show in Section 4, rather than obtaining bounds on $d(t)$, it is sometimes more convenient to bound the *standardized maximal distance* defined by

$$\bar{d}(t) := \max_{x,y \in \Omega} \| P^t(x, \cdot) - P^t(y, \cdot) \|_{\mathrm{TV}}, \tag{2-2}$$

which satisfies the following result.

Lemma 2.2 [Levin et al. 2009, Lemma 4.11]. *With $d(t)$ and $\bar{d}(t)$ defined above, we have*

$$d(t) \le \bar{d}(t) \le 2\, d(t).$$

See [Levin et al. 2009] for more on mixing times of Markov chains.

3. The rook's walk

The rook's walk describes a rook moving on an n^d-chessboard. The rook moves according to a uniformly random distribution so that any square available to it has an equal likelihood of being selected in the next move. The following definition introduces notation to formalize the intuitive notion of a legal move on the n^d-chessboard.

Definition 3.1. For integers $n \ge 3$ and $d \ge 1$, the n^d *rook's walk* is the irreducible, aperiodic, and symmetric Markov chain on the d-fold Cartesian product $\{1, \ldots, n\}^d$ with the transition probabilities

$$P(x, y) = \frac{1}{d(n-1)} \mathbf{1}_{\{\|y - x\|_0 = 1\}},$$

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

By Theorem 2.1, the n^d rook's walk converges to a unique stationary distribution. Moreover, since the transition matrix for the n^d rook's walk is symmetric, the stationary distribution is the uniform distribution on the n^d possible states.

As mentioned in the introduction, one approach to finding the mixing time of this Markov chain is the spectral method, bounding the mixing time in terms of the eigenvalues of the transition matrix P. Following such a method, the bounds below for the rook's walk mixing times were obtained in [Kim 2012].

Proposition 3.2. *For the n^d rook's walk with $n \geq 3$, we have*

$$\frac{d(n-1)}{n} \log \frac{1}{2\varepsilon} \leq t_{\text{mix}}(\varepsilon) \leq \frac{d(n-1)}{n} \log \frac{n^d}{\varepsilon}. \tag{3-1}$$

In [Li and Tucker 2014], by direct computation of the maximal distance to stationary $d(t)$ for the rook's walk in two dimensions, the authors derived the following result.

Proposition 3.3. *For the n^2 rook's walk with $n \geq 3$, we have*

$$t_{\text{mix}}\left(\tfrac{1}{4}\right) = \begin{cases} 2 & \text{for } 3 \leq n \leq 7, \\ 3 & \text{for } n \geq 8. \end{cases}$$

That the mixing time for the rook's walk is asymptotically constant (in n) is rather intuitive: since the rook can move arbitrarily far along any row or column, increasing the length of the board does not increase the number of moves required to reach any square. This renders the length of the board to be of little consequence.

4. Coupling and path coupling methods

One of the advantages of defining the mixing time of a Markov chain in terms of the total variation distance (2-1) is that the total variation distance can be expressed in terms of couplings of distributions, which provide a powerful probabilistic tool in the analysis of mixing times.

A *coupling* of two distributions μ and ν is a pair (X, Y) of random variables defined on a single probability space such that the marginal distribution of X is μ and the marginal distribution of Y is ν. The relationship between couplings and total variation distance between two distributions is given by the following proposition:

Proposition 4.1 [Levin et al. 2009, Proposition 4.7]. *Let μ and ν be two distributions on the state space Ω. Then*

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

Moreover, a coupling (X, Y) which attains the infimum exists and we call such a coupling the optimal coupling.

There are several applications of couplings to the analysis of probability distributions (see, e.g., [Lindvall 2002]), but we will be more interested in the extension of this idea to coupling Markov chains.

Definition 4.2. We define a *coupling of a Markov chain* with transition matrix P to be a process $(X_t, Y_t)_{t=0}^{\infty}$ with the property that both (X_t) and (Y_t) are Markov chains with common transition matrix P.

Given a coupling of a Markov chain with initial states $X_0 = x_0$ and $Y_0 = y_0$, we will write $\mathbf{P}_{x_0,y_0}$ for the probability on the product space of the process (X_t, Y_t). From Proposition 4.1, we get the following bound on the total variation distance of coupled Markov chains.

Theorem 4.3. *If (X_t, Y_t) is a coupling of a Markov chain transition matrix P and initial states with $X_0 = x_0$ and $Y_0 = y_0$, then*

$$\|P^t(x_0, \cdot) - P^t(y_0, \cdot)\|_{\mathrm{TV}} \le \mathbf{P}_{x_0,y_0}(X_t \neq Y_t).$$

The *coupling method* for bounding mixing times of Markov chains is a process for bounding the value $\mathbf{P}_{x_0,y_0}(X_t \neq Y_t)$. There are a couple of standard techniques to bound $\mathbf{P}_{x_0,y_0}(X_t \neq Y_t)$. The approach we employ is in terms of a given a metric ρ on the state space Ω, and in our setting a natural such metric arises. Namely, given a symmetric, irreducible, and aperiodic Markov chain, we can view the elements of its state space Ω as vertices of a connected, undirected graph, with an edge between x and y precisely when $P(x, y) > 0$. Every such graph has a natural metric associated to it: for $x, y \in \Omega$, we define $\rho(x, y)$ to be the *geodesic distance* from x to y, that is, the length (in number of edges) of the shortest path from x to y in Ω. The presence of this geodesic metric ρ allows for the reformulation of several key ideas: legal moves are encoded as pairs of states with $\rho(x, y) = 1$; we can make sense of the diameter of Ω, defined by $\operatorname{diam}\Omega = \max_{x,y\in\Omega} \rho(x, y)$; and, since the metric only takes values that are natural numbers, the event considered in Theorem 4.3 is encoded as

$$\mathbf{P}_{x_0,y_0}(X_t \neq Y_t) = \mathbf{P}_{x_0,y_0}(\rho(X_t, Y_t) \ge 1). \tag{4-1}$$

Combining Theorem 4.3, Lemma 2.2, (4-1), and (2-2), we see that to bound mixing times it suffices to bound the probability $\mathbf{P}_{x_0,y_0}(\rho(X_t, Y_t) \ge 1)$. Generally speaking, a strong bound on this probability for arbitrary time step t is not feasible, and so we turn to the alternative method of deriving a bound of the form

$$\mathbf{E}_{x,y}[\rho(X_1, Y_1)] = \mathbf{E}\big[\rho(X_t, Y_t) \mid X_{t-1}=x, Y_{t-1}=y\big] \le e^{-\alpha}\rho(x, y)$$

for some $\alpha > 0$. We refer to the above behavior as *contraction* of the coupled Markov chains. Then iterating over the time steps t yields a bound on the maximal distance to stationary $d(t)$. This result is stated in the following proposition.

Proposition 4.4. *Suppose (X_t, Y_t) is a coupling of an aperiodic and irreducible Markov chain on a finite state space Ω, and let ρ be an integer-valued metric on Ω. If for all $(x, y) \in \Omega \times \Omega$,*

$$\mathbf{E}_{x,y}[\rho(X_1, Y_1)] \le e^{-\alpha}\rho(x, y)$$

for some real number $\alpha > 0$, then

$$t_{\text{mix}}(\varepsilon) \le \left\lceil \frac{1}{\alpha} \log \frac{\operatorname{diam}\Omega}{\varepsilon} \right\rceil.$$

Proof. Let $x_0, y_0 \in \Omega$. By Theorem 4.3 and (4-1), we have

$$\|P^t(x_0, \cdot) - P^t(y_0, \cdot)\|_{\text{TV}} \le \mathbf{P}_{x_0,y_0}(X_t \ne Y_t) = \mathbf{P}_{x_0,y_0}(\rho(X_t, Y_t) \ge 1)$$

and Markov's inequality yields

$$\mathbf{P}_{x_0,y_0}(\rho(X_t, Y_t) \ge 1) \le \mathbf{E}_{x_0,y_0}[\rho(X_t, Y_t)].$$

By the contraction assumption $\mathbf{E}_{x,y}[\rho(X_1, Y_1)] \le e^{-\alpha}\rho(x, y)$, we have

$$\begin{aligned}\mathbf{E}_{x_0,y_0}[\rho(X_t, Y_t)] &= \mathbf{E}_{x_0,y_0}\big[\mathbf{E}_{x_0,y_0}[\rho(X_t, Y_t) \mid (X_{t-1}, Y_{t-1})]\big] \\ &\le e^{-\alpha}\,\mathbf{E}_{x_0,y_0}[\rho(X_{t-1}, Y_{t-1})].\end{aligned}$$

Repeated iteration over t time steps yields

$$\mathbf{E}_{x_0,y_0}[\rho(X_t, Y_t)] \le e^{-\alpha t}\rho(x_0, y_0).$$

Therefore, since x_0 and y_0 were arbitrary, by Lemma 2.2 we conclude that

$$d(t) \le e^{-\alpha t} \operatorname{diam}\Omega. \qquad \square$$

The above coupling method for bounding mixing times of Markov chains requires bounding the coupling distance probability $\mathbf{P}_{x,y}(\rho(X_t, Y_t) \ge 1)$ for *all* states x and y which is often nontrivial. For contrast, the *path coupling* method derived by Bubley and Dyer [1997], essentially a combination of the coupling method with the triangle inequality, requires only the demonstration of the contraction bound for *adjacent* states. The full path-coupling result is stated in the following proposition.

Proposition 4.5. *Suppose (X_t, Y_t) is a coupling of an irreducible and aperiodic Markov chain on a finite state space Ω with its geodesic metric ρ. If for all adjacent states x and y we have*

$$\mathbf{E}_{x,y}[\rho(X_1, Y_1)] \le e^{-\alpha}$$

for some constant $\alpha > 0$, then

$$t_{\text{mix}}(\varepsilon) \le \left\lceil \frac{1}{\alpha} \log \frac{\operatorname{diam}\Omega}{\varepsilon} \right\rceil.$$

Proof. Let $x, y \in \Omega$, let $r = \rho(x, y)$, and choose a length-r path from x to y

$$x = x_0, x_1, \ldots, x_{r-1}, x_r = y.$$

Such a path exists by the definition of ρ, and we have $\rho(x_i, x_{i-1}) = 1$ for all $i = 1, \ldots, r$. If we denote by $(X_{1,i}, X_{1,i-1})$ the coupling corresponding to (x_i, x_{i-1}), then by the hypotheses of the theorem, there is a positive constant α such that

$$\mathbf{E}_{x_i, x_{i-1}}[\rho(X_{1,i}, X_{1,i-1})] \leq e^{-\alpha}$$

for each i. Then by the triangle inequality we have

$$\mathbf{E}_{x,y}[\rho(X_1, Y_1)] \leq \sum_{i=1}^{r} \mathbf{E}_{x_i, x_{i-1}}[\rho(X_{1,i}, X_{1,i-1})] \leq \rho(x, y)e^{-\alpha}.$$

The result then follows from Proposition 4.4. □

In the case of rooks moving on an n^d-chessboard, the geodesic metric is described easily in terms of the motion of the rooks — the distance between two states/squares is simply the minimum number of rook moves to connect two such squares. Alternatively, viewing our chessboard as being coordinatized by the d principal axes, the distance between two states is simply the number of coordinates in which the two states' components differ (i.e., the Hamming distance between their coordinate vectors). That is, if we let $x = [x^1, \ldots, x^d]$ and $y = [y^1, \ldots, y^d]$ be squares in the d-dimensional chessboard state space Ω, then the geodesic metric ρ on Ω is given by

$$\rho(x, y) = \sum_{i=1}^{d} \mathbf{1}_{\{y^i - x^i \neq 0\}}.$$

It is worth briefly unpacking this in the language of two rooks occupying two states on the standard two-dimensional board. The case $\rho = 2$ means that the two rooks have neither a column nor row in common. Similarly $\rho = 1$ means that the two rooks share either a row or a column, and we have $\rho = 0$ if and only if two rooks occupy the same space on the board. Note that this trichotomy of metric values is precisely the trichotomy we used in the introduction to count two-move sequences.

5. Coupling of the rook's walk

The key to obtaining a good bound on the mixing time of a Markov chain using the coupling method is to construct an optimal or near-optimal coupling, i.e., one with very small values of $P(X_t \neq Y_t)$. In other words, you want to construct a coupling that encourages the coordinate processes of (X_t, Y_t) to meet as fast as possible. Below we will construct a Markov chain Y that accomplishes this for the rook's walk

on the d-dimensional chessboard Ω, and use the path coupling technique derived in the previous section to bound the mixing time. We will continue to use subscripts to denote the time step of the chain, and superscripts to denote the components of the vector describing the state; i.e., X_t^i denotes the i-th component of X at time t.

Definition 5.1 (rook's walk coupling). Let (X_t) denote the rook's walk, and (X_t, Y_t) be a process with current states $X_t = x_t$ and $Y_t = y_t$, and let $X_{t+1} = x_{t+1}$. Then for $t \geq 0$, we define Y_{t+1} using x_t, y_t, and x_{t+1}. Since $\rho(x_t, x_{t+1}) = 1$ for all values of x_{t+1}, there is a unique l such that $x_t^l \neq x_{t+1}^l$.

- If $x_{t+1}^l \neq y_t^l$, set
$$Y_{t+1}^i = \begin{cases} y_t^i & \text{for } i \neq l, \\ x_{t+1}^i & \text{for } i = l. \end{cases}$$
- Suppose $x_{t+1}^l = y_t^l$. If $x_{t+1} = y_t$, set $Y_{t+1} = x_t$. Otherwise choose uniformly one of the indices m for which $y_t^m \neq x_{t+1}^m$, and set
$$Y_{t+1}^i = \begin{cases} y_t^i & \text{for } i \neq m, \\ x_{t+1}^i & \text{for } i = m. \end{cases}$$

Let us give a more informal and intuitive explanation in the language of rooks moving on an n-dimensional chessboard. Slightly abusing terminology and notation, let X and Y denote two rooks moving on a board, with X moving according to the standard rook's walk. Then the movement of Y is to respond a move of X as directed by the following rule:

> When X moves in a given dimension, move Y along that same dimension so as to match X's component in that dimension, unless those components already match, in which case:
>
> - randomly choose another dimension for Y to move along, and match X's component in that dimension; unless
> - X moves to the square currently inhabited by Y, then lacking anything more clever to do, Y moves to X's previous location.

In short, Y will move so as to decrease $\rho(X, Y)$ by 1 whenever possible, moving along the same dimension as X unless X and Y agree in that dimension. In this case, X moves along a dimension chosen uniformly between all other axes in which their components differ. Once Y occupies the same square as X, it will move to mirror X's move every turn. In particular, note that $\rho(X, Y)$ is nonincreasing, and decreases by 0 or 1 each move until the two rooks occupy the same state.

Example 5.2. Figure 2 shows a two-dimensional walk according to this coupling for $n = 4$. The white rook moves at random via the uniform distribution on the set of legal moves, and the black rook moves deterministically via the rules of

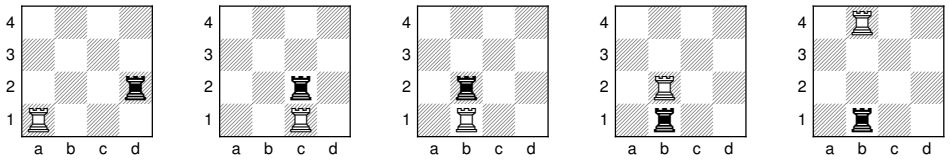

Figure 2. An instance of the coupling, with $d = 2$ and $n = 4$.

Definition 5.1. In algebraic chess notation, the move order would read 1. `Rc1 Rc2` 2. `Rb1 Rb2` 3. `Rb2 Rb1` 4. `Rb4`. The move following the last diagram will be 4. ... `Rb4`, at which point the two rooks' positions will coincide from then on.

The content of the following theorem can be paraphrased rather nicely in this language: if X moves according to the uniform distribution on legal moves, and Y "follows" X as prescribed by the rules above, then Y's position also satisfies the uniform distribution.

Theorem 5.3. *Let (X_t) be the rook's walk with transition matrix P. Then the process (Y_t) described above is also a Markov chain with transition matrix P; i.e., (X_t, Y_t) forms a coupling of the rook's walk Markov chain.*

Proof. The transition probabilities for X are given by the uniform distribution on legal squares, and we need to show the same is true for Y. By translational symmetry, we can without loss of generality suppose that $Y_t = [1, 1, \ldots, 1]$. For $2 \le k \le n$ and $1 \le m \le d$, let $y_{k,m}$ denote the state whose coordinates are all 1, except for a k in the m-th component. We need to show that

$$P\big(Y_{t+1} = y_{k,m} \mid Y_t = [1, 1, \ldots, 1]\big) = \frac{1}{d(n-1)}$$

for each of the $d(n-1)$ such choices, regardless of the value of X_t. Given such an X_t, let $r = \rho(X_t, Y_t)$. Relabeling the axes if needed, we can also without loss of generality assume that $X_t = [x^1, x^2, \ldots, x^r, 1, 1, \ldots, 1]$ with each $x^i > 1$.

Now, following the rules set forth in Definition 5.1:

- For $m > r$, we have $Y_{t+1} = y_{k,m}$ if and only if
$$X_{t+1} = [x^1, x^2, \ldots, x^r, 1, \ldots, k, \ldots, 1],$$
with the k occurring in the m-th slot, which occurs with probability $1/(d(n-1))$ by the transition probabilities for X.
- If $r = 1$ and $X_t = [k, 1, 1, \ldots, 1]$, then $Y_{t+1} = y_{k,1}$ if and only if $X_{t+1} = Y_t$, which occurs with probability $1/(d(n-1))$.
- If $r > 1$, $m \le r$ and $x_m \ne k$, we have $Y_{t+1} = y_{k,m}$ if and only if
$$X_{t+1} = [x^1, \ldots, x^{m-1}, k, x^{m+1}, \ldots, x^r, 1, \ldots, 1],$$
which again occurs with probability $1/(d(n-1))$.

- Finally, if $r > 1$, $m \le r$ and $x_m = k$, then Y_{t+1} can only be $y_{k,m}$ if $l \le r$, $l \ne m$, and $X^l_{t+1} = 1$, which occurs with probability $(r-1)/(d(n-1))$. In this case, $\rho(Y_t, X_{t+1}) = r-1$, and so Y_{t+1} is obtained from Y_t by uniformly choosing one of the $r-1$ components in which x_{t+1} and y_t differ for a total probability of

$$\frac{1}{r-1}\frac{r-1}{d(n-1)} = \frac{1}{d(n-1)}.$$

□

6. Mixing time of the rook's walk

Having established that the rook's walk coupling defined in Definition 5.1 does indeed provide us a coupling, we will now compute the expected contraction factor α (in the language of Proposition 4.5) for the coupling, which gives the following result for the mixing time of the rook's walk.

Theorem 6.1. *For the n^d rook's walk, we have*

$$t_{\text{mix}}(\varepsilon) \le \left\lceil \frac{\log \frac{d}{\varepsilon}}{\log \frac{d(n-1)}{(d-1)(n-1)+1}} \right\rceil. \tag{6-1}$$

Proof. Let x and y be adjacent states. Then since $\rho(x, y) = 1$, there exists a unique component l for which $x^l \ne y^l$. Then, in the language of rooks, from Definition 5.1, $\rho(X_t, Y_t) = 0$ if and only if X makes one of the $n-2$ moves along that axis and *not* onto y, and $\rho(X_t, Y_t) = 1$ otherwise. This gives an expected value of

$$\mathbf{E}\big[\rho(X_t, Y_t) \mid X_{t-1}{=}x, Y_{t-1}{=}y\big] = 0 \cdot \frac{n-2}{d(n-1)} + 1 \cdot \frac{d(n-1)-(n-2)}{d(n-1)}.$$

In the notation of Proposition 4.4, this says that we may take the value

$$\alpha = \log \frac{d(n-1)}{(d-1)(n-1)+1}.$$

Thus, combined with the observation that $\operatorname{diam}\Omega = d$, we obtain the result from Proposition 4.4. □

Taking $\varepsilon = \frac{1}{4}$, we have the numerical data in Table 1.
We note that:

- For fixed d, the upper bound in (3-1) grows logarithmically with n as $n \to \infty$, whereas (6-1) is asymptotically constant, approaching

$$\left\lceil \frac{\log \frac{d}{\varepsilon}}{\log \frac{d}{(d-1)}} \right\rceil.$$

$d\downarrow$	$n\to$ 3	4	5	6	7	8	100	1000
2	8	6	5	5	4	4	4	4
3	14	10	9	9	8	8	7	7
4	21	16	14	13	12	12	10	10
5	29	21	19	18	17	16	14	14
10	72	54	48	45	43	42	36	36
100	1196	896	796	746	716	697	603	597

Table 1. Coupling bounds for $t_{\text{mix}}\left(\frac{1}{4}\right)$ for various n and d.

- For fixed n, the upper bound in (3-1) grows quadratically with d as $d\to\infty$, whereas our bound

$$\left\lceil \frac{\log \frac{d}{\varepsilon}}{\log \frac{d(n-1)}{(d-1)(n-1)+1}} \right\rceil \approx \frac{1}{1-\frac{\log(d-1)}{\log d}}$$

  grows like $d\log d$, regardless of n. (Try this as an exercise!) It is interesting to note that this value of $d\log d$ agrees with the asymptotic solution to the coupon collector's problem with d coupons. This suggests that a necessary and sufficient condition for the rook's walk to be thoroughly mixed is that our rook has moved at least once in each of the d dimensions.

As the motivating example, let us pay extra attention to generalizations of the standard 8^2-chessboard. For the case where we fix the dimension as $d=2$, the actual values of t_{mix} (as always, for $\varepsilon=\frac{1}{4}$) are known from Proposition 3.3 for all lengths n, and so serve as a litmus test for the accuracy of a given bound. Let us denote by t_c the "unceilinged" version of the coupling bound given in Theorem 6.1, and by t_s the upper bound obtained by the spectral method in Proposition 3.2. Figure 3 demonstrates the two upper bounds t_s and t_c in conjunction with the exact values of $t_{\text{mix}}\left(\frac{1}{4}\right)$ from Proposition 3.3.

In particular, it is easy to show that for $d=2$, we have

$$\lim_{n\to\infty} t_c \le 1-\log_2 \varepsilon,$$

and so for $\varepsilon=\frac{1}{4}$, we get an asymptotic mixing time of precisely 3, making the upper bound (6-1) asymptotically tight.

On the other hand, if we fix the length as $n=8$ and vary instead the dimension of the board, it quickly becomes computationally difficult to evaluate exact values — we were only able to easily compute the exact mixing times for $d=2,3,4$ (with respective mixing times of 3, 5, and 7). See Figure 4.

In short, it seems that the coupling method provides a stronger alternative to the standard spectral method for bounding mixing times of rook's walk Markov chains

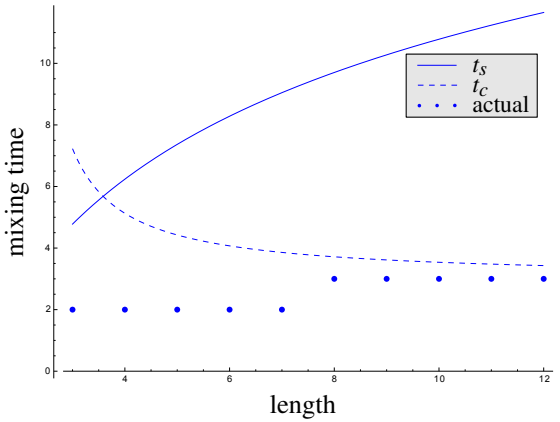

Figure 3. Bounds and actual values for $t_{\text{mix}}\left(\frac{1}{4}\right)$, $d = 2$, n varying.

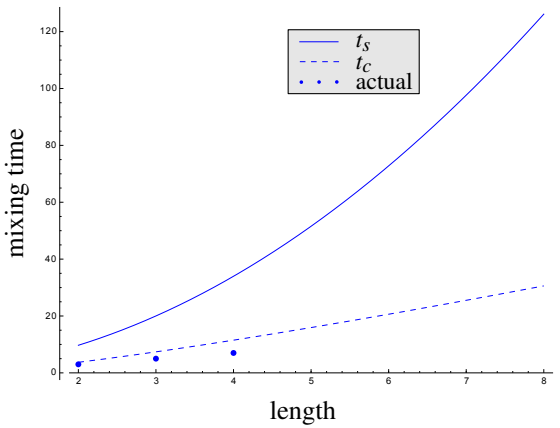

Figure 4. Bounds and actual values for $t_{\text{mix}}\left(\frac{1}{4}\right)$, $n = 8$, d varying.

from above. Further, in the two-dimensional case, the coupling bounds are asymptotically tight. There are natural extensions to the rook's walk that would be interesting to pursue. In particular, future work will address the case where the rook's moves are restricted to traversing at most k squares per move, and the case where the transition probabilities to all the allowable squares are not uniform.

Acknowledgments

This research was partially supported by NSF grant #1157105 funding the WiVaM REU-RET in Mathematics. We would like to thank the reviewer for a very careful reading of the article and the helpful comments.

References

[Bubley and Dyer 1997] R. Bubley and M. E. Dyer, "Path coupling: a technique for proving rapid mixing in Markov chains", pp. 223–231 in *Proceedings of the 38th IEEE Symposium on Foundations of Computer Science* (Miami, FL, 1997), Institute of Electrical and Electronics Engineers, Los Alamitos, CA, 1997.

[Kemeny et al. 1976] J. G. Kemeny, J. L. Snell, and A. W. Knapp, *Denumerable Markov chains*, 2nd ed., Graduate Texts in Mathematics **40**, Springer, New York, NY, 1976. MR Zbl

[Kim 2012] S. S. Kim, "Mixing time of a rook's walk", Undergraduate certificate paper, Princeton University, 2012, available at http://www.pacm.princeton.edu/documents/Kim.pdf.

[Levin et al. 2009] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, Providence, RI, 2009. MR Zbl

[Li and Tucker 2014] Y. Li and K. Tucker, "Investigating the rook's walk", Senior thesis, Willamette University, Salem, OR, 2014.

[Lindvall 2002] T. Lindvall, *Lectures on the coupling method*, Dover, Mineola, NY, 2002. MR Zbl

Received: 2015-07-07 Revised: 2015-12-18 Accepted: 2015-12-19

mclemanc@umflint.edu *Department of Mathematics, University of Michigan–Flint, Flint, MI 48502 United States*

potto@willamette.edu *Department of Mathematics, Willamette University, Salem, OR 97301, United States*

johnr4@vt.edu *Department of Mathematics, Virginia Tech, Blacksburg, VA 24061, United States*

masutter@umflint.edu *Department of Mathematics, University of Michigan–Flint, Flint, MI 48502 United States*

involve

msp.org/involve

INVOLVE YOUR STUDENTS IN RESEARCH

Involve showcases and encourages high-quality mathematical research involving students from all academic levels. The editorial board consists of mathematical scientists committed to nurturing student participation in research. Bridging the gap between the extremes of purely undergraduate research journals and mainstream research journals, *Involve* provides a venue to mathematicians wishing to encourage the creative involvement of students.

See inside back cover or msp.org/involve for submission instructions. The subscription price for 2017 is US $175/year for the electronic version, and $235/year (+$35, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to MSP.

Involve (ISSN 1944-4184 electronic, 1944-4176 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

Involve peer review and production are managed by EditFlow® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

involve
2017 vol. 10 no. 1