Probab. Theory and Relat. Fields 108, 441–457 (1997)

Probability Theory and Related Fields

Random random walks on $\mathbb{Z}_2^d$*

David Bruce Wilson

University of California, 387 Soda Hall Berkeley, CA 94720-1776, USA
e-mail: dbwilson@alum.mit.edu

Summary. We consider random walks on classes of graphs defined on the d-dimensional binary cube $\mathbb{Z}_2^d$ by placing edges on n randomly chosen parallel classes of vectors. The mixing time of a graph is the number of steps of a random walk before the walk forgets where it started, and reaches a random location. In this paper we resolve a question of Diaconis by finding exact expressions for this mixing time that hold for all $n > d$ and almost all choices of vector classes. This result improves a number of previous bounds. Our method, which has application to similar problems on other Abelian groups, uses the concept of a universal hash function, from computer science.

Mathematics Subject Classification (1991): Primary 60J15; Secondary 60B15

1. Introduction

We choose a set S consisting of n vectors chosen uniformly at random in the binary d-cube $\mathbb{Z}_2^d$, and consider the graph on the vertices of $\mathbb{Z}_2^d$ with edges between pairs of vertices whose difference is in S. A random walk on any such graph consists of a sequence of steps, starting from some initial point. In each step one moves from a vertex to a neighboring vertex chosen at random from the set of neighbors.

Random walks of this kind have been studied because they represent tractable models on which to test and simulate behavior on more complicated Markov structures. The d-cube itself is an example with $n = d$. This case has been intensively studied because of its connection to the Ehrenfest urn of statistical mechanics. Diaconis (1988), Diaconis, Graham, and Morrison (1990), and Sinclair (1993) describe and reference many results on such walks.

Diaconis (1993) has posed the following question about the class of graphs defined above: How many steps are required in such a walk before the

Key words and phrases. Random walk, hypercube, mixing time, threshold.

* This research was done while the author was at MIT, and was supported in part by an ONR-NDSEG fellowship, and in part by DARPA contract N00014-92-J-1799.

probability distribution as a function of initial vertex loses all memory of its origin, by becoming essentially uniform on the vertices of the d-cube? The purpose of this paper is to answer this question. The answer exhibits threshold behavior at a certain number, $T(d,n)$, of steps. Up until $(1-\varepsilon)T(d,n)$ steps, the distribution is always far from uniform. After $(1+\varepsilon)T(d,n)$ steps, however, for almost all S with $n > d$, the probability is close to uniform. This threshold number of steps is called the *mixing time* of the random walk.

To be more precise, let $Q^{*k}(x)$ be the chance that the random walk started at 0 is at vertex x after k steps. (Because of symmetry, starting the walk at another vertex amounts to relabelling the vertices of the d-cube.) Under mild regularity conditions on S, $Q^{*k}(x)$ approaches the uniform distribution $U(x) = 1/2^d$ as k goes to ∞. We will measure convergence in total variation (TV):

$$\|Q^{*k} - U\|_{\mathrm{TV}} \equiv \max_{A \subseteq \mathbb{Z}_2^d} \left|Q^{*k}(A) - U(A)\right| \;=\; \frac{1}{2}\sum_x \left|Q^{*k}(x) - 2^{-d}\right| \;.$$

The function $T(d,n)$ is defined by

$$T(d,n) \;=\; \frac{n}{2}\ln\frac{1}{1-2H^{-1}(d/n)} \;,$$

where H is the binary entropy function given by

$$H(x) = x\log_2\frac{1}{x} + (1-x)\log_2\frac{1}{1-x} \qquad 0 \le x \le 1 \;,$$

and lg is the base-2 logarithm. The main result of this paper is

Theorem 1. *For any $\varepsilon > 0$, for all sufficiently large d and all $n > d$, the random walk satisfies*

1) *For any choice of S, if $k \le (1-\varepsilon)T(d,n)$, then $\|Q^{*k} - U\|_{\mathrm{TV}} > 1-\varepsilon$.*
2) *For almost any choice of S, if $k \ge (1+\varepsilon)T(d,n)$ then $\|Q^{*k} - U\|_{\mathrm{TV}} < \varepsilon$, provided the Markov chain is ergodic.*

Note that almost all Markov chains will be ergodic anyway when $n-d$ is sufficiently large.

We obtain the lower bound on the mixing time by considering a related random walk on $\mathbb{Z}_2^n$, and counting the number of states likely to be reached after k steps. The upper bound is obtained by considering the map from $\mathbb{Z}_2^n$ to $\mathbb{Z}_2^d$ as a *universal hash function*, and making use of various norms. The arguments are presented in the next two sections. The same ideas can be applied to similar problems in other larger lattices, and to a lesser extent to similar problems in general Abelian groups (Wilson, 1994).

In Figure 1 we illustrate the behavior of $1-2H^{-1}(x)$ as a function of x, and in the appendix we derive the asymptotic properties of this function that we will use in this article. When the mixing time threshold $T(d,n)$ is normalized by d, it is a function of the ratio n/d only. In Figure 2 we show its

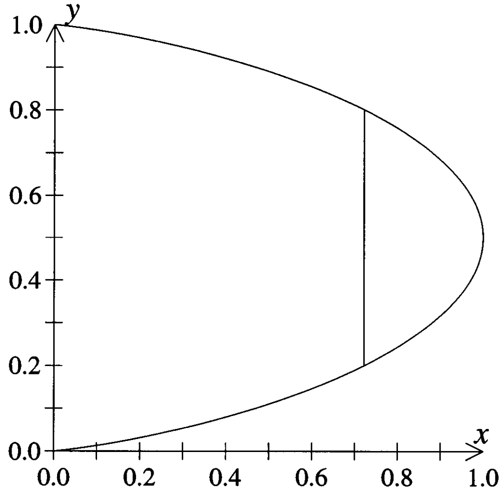

Fig. 1. Here $x = H(y)$. $H^{-1}(x)$ is the inverse of H in the range $0 \le H^{-1}(x) \le 1/2$. The function $1 - 2H^{-1}(x)$ is represented by the height of the vertical line

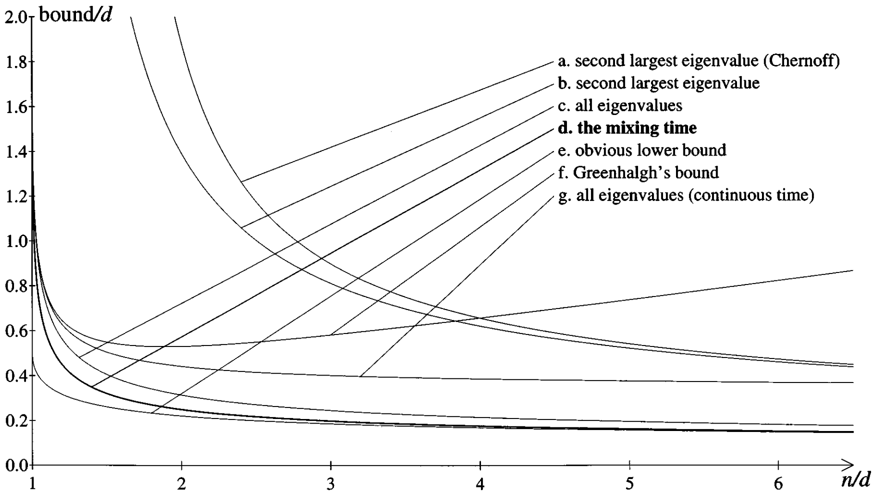

Fig. 2. Bounds on the mixing time for $\mathbb{Z}_2^d$. $T(d, n)$ is shown in bold. All upper bounds (besides $T(d, n)$) were derived from the Diaconis-Shahshahani upper bound lemma. The upper bounds hold for almost all $v_1, \dots, v_n$, and the lower bounds hold for all $v_1, \dots, v_n$. All bounds except **f** and **g** are tight when $n/d \to \infty$, and all bounds except **a**, **b**, and **e** are tight when $n/d \to 1$

dependence on this variable explicitly, with, for comparison some other bounds that have been obtained on this parameter.

In the classical case, the set S consists of the standard basis vectors together with 0, the 0 vector being included to avoid parity problems. Diaconis and Shahshahani (1987) have shown that the threshold for mixing occurs at

$\frac{1}{4}d\ln d$ steps. As we shall see, $T(d, d+1) \doteq \frac{1}{4}d\ln d$. By contrast, if $|S| = 2d$ then $T(d, 2d) \doteq 0.24853d$. In general, when n is linear in d with $n = ad$ and $a > 1$, the function $T(d, n)$, and therefore the mixing time, is linear in d.

In the appendix we see that $H^{-1}(x) \sim x/\lg(1/x)$ as $x \to 0$. When n is superlinear in d, we have $H^{-1}(d/n) \sim (d/n)/\lg(n/d) \ll 1$, so that

$$T(d,n) \sim \frac{d}{\lg(n/d)} \quad .$$

Thus for $n = d^a$ (fixed $a > 1$) we have

$$T(d,n) \sim \frac{a}{a-1}\frac{d}{\lg n} \quad ,$$

while if n is superpolynomial in d, $T(d,n)$ behaves as $d/\lg n$ for large d.

Aldous and Diaconis (1985) conjectured that for any large group G, for n suitably large, the mixing time of almost any random walk on G generated by n group elements is $\ln|G|/\ln n$. Dou (1992) confirmed this conjecture for n superpolylogarithmic in $|G|$, which implies the superpolynomial case above. Dou also considered random walks where, when one moves from a vertex to a neighboring vertex, the neighbor need not be chosen uniformly at random. In the more general setting for $\mathbb{Z}_2^d$, he showed that if $n - d$ is large but $n \le d^2$, then the mixing time is at most $(1/2 + \varepsilon)n\ln n$. Hildebrand (1993) obtained the polynomial result. Recently Dou and Hildebrand (1996) showed that when $n = d^a$, $a/(a-1)\ln|G|/\ln n$ upper bounds the mixing time of all groups, and is tight for many groups including Abelian groups. Alon & Roichman (1994) have studied the related problem of the eigenvalues of random graphs based on groups, and Roichman (1996) then gave an alternative proof to the Dou-Hildebrand result.

In the other limit, if $n = d + \gamma = d + o(d)$, from the appendix we see that $1 - 2H^{-1}(d/n) = 1 - 2H^{-1}(1 - \gamma/n) \sim \sqrt{(\ln 4)\gamma/n}$, so

$$T(d,n) \sim \frac{n}{2}\ln\sqrt{n/\gamma\ln 4} \sim \frac{d}{4}\ln(d/\gamma) \quad . \tag{1}$$

Greenhalgh (1993) showed that almost all random walks on $\mathbb{Z}_2^d$ converge to uniformity in $n\ln[n/(\gamma\ln 2)] + \varepsilon n$ steps, for large $\gamma = n - d$. When $\gamma = o(d)$, this implies that Equation 1 upper bounds the mixing time. If $\gamma = 1$ then Equation 1 reduces to the well-known mixing time threshold of $\frac{1}{4}d\ln d$ of the usual random walk on $\mathbb{Z}_2^d$.

2. Mixing time lower bound

In this section we prove a lower bound on the mixing time of the random walk on $\mathbb{Z}_2^d$ whose support is $v_1, \ldots, v_n$. The bound holds for all choices of supporting vectors. We assume $n > d$. To prove the lower bound on the mixing time, we consider the random walk on $\mathbb{Z}_2^n$ where the translating vectors are $e_1, \ldots, e_n$. This walk is a lifting of the random walk on $\mathbb{Z}_2^d$: a translation by e_i in $\mathbb{Z}_2^n$ corresponds to a translation by v_i in the walk on $\mathbb{Z}_2^d$, and the current state in $\mathbb{Z}_2^n$ determines the current state in $\mathbb{Z}_2^d$. The idea is to show that after k steps, the state in $\mathbb{Z}_2^n$ is very far from looking like a random state, so the state in $\mathbb{Z}_2^d$ is far from looking like a random state.

Let LOWER(d, n, ε) denote the assertion "If

$$k \leq (1-\varepsilon)\frac{n}{2}\ln\frac{1}{1-2H^{-1}(d/n)} \ ,$$

then

$$\|Q^{*k} - U\|_{\mathrm{TV}} \geq 1 - \varepsilon \ ."$$

This assertion is well-defined only when $n > d$. The farthest that a probability distribution can be from uniformity is $1 - 2^{-d}$, so for a given ε, LOWER(d, n, ε) will be true only if d is big enough.

Theorem 2. *For any* $\varepsilon > 0$, *if* d *is big enough, then for any* $n > d$, LOWER(d, n, ε).

To prove the theorem, we use different bounds which are effective when $n \gg d$, $n = d + \Theta(d)$, and $n = d + o(d)$. Each case is proved as a separate lemma.

Case $n \gg d$: After k steps of the walk, the state in $\mathbb{Z}_2^n$ can take on only $\binom{n}{k} + \binom{n}{k-1} + \cdots + \binom{n}{0}$ possible values. Appendix A of Peterson & Weldon [1972] gives $\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{\alpha n} \leq 2^{H(\alpha)n}$ for $\alpha \leq 1/2$. (This inequality follows from an entropy argument; the entropy of a bit vector of size n is at most the sum of the entropies of its bits, the entropy of each bit being at most $H(\alpha)$ if its probability of being 1 is at most α.) Let A be the set of corresponding states in $\mathbb{Z}_2^d$: $Q^{*k}(A) = 1$. If $k \leq n/2$ and $2^{nH(k/n)} \leq 2^{d-y}$, then at most a 2^{-y} fraction of the states in $\mathbb{Z}_2^d$ are reachable: $U(A) \leq 2^{-y}$. When

$$k \leq nH^{-1}(d/n - y/n)$$

we get $\|Q^{*k} - U\|_{\mathrm{TV}} \geq 1 - 2^{-y}$.

It is straightforward to verify that if $0 \leq \delta \ll x \leq 1$ then $H^{-1}(x-\delta) \sim H^{-1}(x)$. This gives the "obvious lower bound" of $nH^{-1}(d/n)$ in Figure 2. But for small values of x,

$$H^{-1}(x) \sim \frac{1}{2}\ln\frac{1}{1-2H^{-1}(x)} \ ,$$

giving

Lemma 3. *For any* $\varepsilon > 0$, *if* d *is big enough and* $d \ll n$, *then* LOWER(d, n, ε).

Case $n = O(d)$: As in the proof of the lower bound of the usual random walk on $\mathbb{Z}_2^d$, here it is useful to count the number of 1's in the Markov chain state. The strategy is to show that even after k steps of the random walk on $\mathbb{Z}_2^n$, with high probability at most m of the n bits are set. For this range of n, m can be significantly smaller than k, since after a while some vectors are chosen multiple times. Here A is the set of states in $\mathbb{Z}_2^d$ which correspond to states in $\mathbb{Z}_2^n$ with at most m bits set. As before we require $m \leq n/2$ and

$nH(m/n) \leq d - y$ so that $U(A) \leq 2^{-y}$. Now we need to find a large value of k for which $Q^{*k}(A) \approx 1$.

In the random walk on $\mathbb{Z}_2^n$, let I_k denote the number of 1's in the state at time k. $I_0 = 0$, and I_k drifts to $n/2$ for large k. It is not hard to check that $E[I_k] = (n/2)(1 - (1 - 2/n)^k)$, and $\mathrm{Var}[I_k] \leq n/2$. Diaconis [1988] gives a proof, but for the reader's convenience we rederive it here: Let X_k denote the number of 0's in the state, minus $n/2$; $X_0 = n/2$, and X_k drifts to 0 for large k.

$$\begin{aligned} E[X_{k+1}|X_k] &= \frac{n/2 + X_k}{n}(X_k - 1) + \frac{n/2 - X_k}{n}(X_k + 1) \\ &= X_k - 2X_k/n \\ E[X_{k+1}] &= X_0(1 - 2/n)^{k+1} \end{aligned}$$

and

$$\begin{aligned} E\left[X_{k+1}^2|X_k\right] &= \frac{n/2 + X_k}{n}(X_k - 1)^2 + \frac{n/2 - X_k}{n}(X_k + 1)^2 \\ &= X_k^2 + 1 - \left(4X_k^2\right)/n \\ E\left[X_{k+1}^2 - n/4|X_k\right] &= \left(X_k^2 - n/4\right)(1 - 4/n) \\ E\left[X_{k+1}^2\right] &= \left(X_0^2 - n/4\right)(1 - 4/n)^{k+1} + n/4 \ , \end{aligned}$$

so

$$\begin{aligned} \mathrm{Var}[X_k] &= E\left[X_k^2\right] - E[X_k]^2 \\ &= \left(n^2/4 - n/4\right)(1 - 4/n)^k + n/4 - \left[n/2(1 - 2/n)^k\right]^2 \\ &= \left(n^2/4\right)\left[(1 - 4/n)^k - \left(1 - 4/n + 4/n^2\right)^k\right] \\ &\quad + (n/4)\left[1 - (1 - 4/n)^k\right] \\ &\leq n/4 \end{aligned}$$

for $n \geq 4$, $n = 1$, or k odd. When $n = 2, 3$, we still have $\mathrm{Var}[X_k] \leq n/2$. Hence $E[I_k] = (n/2)(1 - (1 - 2/n)^k)$, and $\mathrm{Var}[I_k] \leq n/2$.

Since the variance isn't too big, odds are that I_k is close to $E[I_k]$. If $m = n/2 - (n/2)(1 - 2/n)^k + x\sqrt{n}$ $(x > 0)$, then

$$\Pr[I_k > m] \leq 1/\left(2x^2\right) \ .$$

(Martingale inequalities can be used to show that this probability actually goes to zero much more quickly as x grows; McDiarmid (1989) gives a survey of these techniques.)

So $Q^{*k}(A) \geq 1 - 1/(2x^2)$. Our other constraints are $m \leq n/2$ and

$$\begin{aligned} d - y &\geq nH(m/n) \\ d - y &\geq nH\left(\left[1 - (1 - 2/n)^k\right]/2 + x/\sqrt{n}\right) \\ H^{-1}((d - y)/n) &\geq \left[1 - (1 - 2/n)^k\right]/2 + x/\sqrt{n} \end{aligned}$$

$$\begin{aligned}(1-2/n)^k &\geq 1-2H^{-1}((d-y)/n)+2x/\sqrt{n}\\ k &\leq \frac{\ln\left[1-2H^{-1}((d-y)/n)+2x/\sqrt{n}\right]}{\ln[1-2/n]}\end{aligned}$$

The constraint $m \leq n/2$ translates to $(1-2/n)^k \geq 2x/\sqrt{n}$, which is implied by the previous constraint on k.

Assuming that k satisfies this constraint, then with probability at least $1-1/(2x^2)$, the kth state is contained in subset A which occupies a 2^{-y} fraction of $\mathbb{Z}_2^d$. That is

$$\|Q^{*k}-U\|_{\mathrm{TV}} \geq Q^{*k}(A)-U(A) \geq 1-1/(2x^2)-2^{-y} \ .$$

Case $n=d+\Theta(d)$: Suppose that constants $1<a_l<a_u$, and $x>0, y>0$ are given. If d is big enough, and $a_l d \leq n \leq a_u d$, then the y/n and $2x/\sqrt{n}$ terms are clearly insignificant. For any $\varepsilon>0$, if d is big enough, then k can be taken to be

$$(1-\varepsilon)\frac{-n}{2}\ln\left[1-2H^{-1}(d/n)\right] \ .$$

So we have

Lemma 4. *For any* $\varepsilon>0$ *and* $a_u>a_l>1$, *if* d *is big enough and* $a_l d \leq n \leq a_u d$, *then* LOWER(d,n,ε).

Case $n=d+o(d)$: Suppose x and y are given. Suppose that $\delta>0$ is small enough, and $n=d+\gamma$ where $\gamma \leq \delta d$. Assume d is big enough.

The appendix derives the behavior of H^{-1} as its argument approaches 1, and from this we see

$$\begin{aligned}H^{-1}\left(\frac{d-y}{n}\right) &= H^{-1}\left(1-\frac{\gamma+y}{n}\right)=\frac{1}{2}-\sim\sqrt{\frac{\ln 2}{2}\frac{\gamma+y}{n}}\\ \ln\left[1-2H^{-1}\left(\frac{d-y}{n}\right)+2x/\sqrt{n}\right] &= \ln\left[\sim\sqrt{\ln 4\frac{\gamma+y}{n}}+2x/\sqrt{n}\right]\\ &= -\ln\sqrt{n}+\ln\left[\sim\sqrt{\ln 4(\gamma+y)}+2x\right]\\ &= -\ln\sqrt{n}+\ln\sqrt{\gamma}+O(1)\\ &= \sim\ln\sqrt{\gamma/n}\quad \text{as } \gamma \ll n\\ &= \sim\ln\left[1-2H^{-1}(d/n)\right]\end{aligned}$$

Lemma 5. *For any* $\varepsilon>0$, *if* d *is big enough and* $n/d-1$ *is small enough, then* LOWER(d,n,ε).

Proof of Theorem 2: Immediate from Lemmas 3, 5, and 4. □

3 Mixing time upper bound

Let UPPER(d, n, ε) be the assertion "If

$$k \le (1+\varepsilon)\frac{n}{2}\ln\frac{1}{1-2H^{-1}(d/n)} \ ,$$

then with probability at least $1-\varepsilon$ a random $v_1, \ldots, v_n$ satisfies

$$\|Q^{*k} - U\|_{\mathrm{TV}} \le \varepsilon\text{''} \ .$$

For a given ε, UPPER(d, n, ε) can only be true if $n-d$ is big enough, since otherwise there is a finite chance that the Markov chain will not be ergodic.

Theorem 6. *For any $\varepsilon > 0$, if d is big enough and $n - d = \omega(\log d)$, then* UPPER(d, n, ε).

As before we consider the random walk on $\mathbb{Z}_2^n$, and map it to the walk on $\mathbb{Z}_2^d$. Let M be the matrix whose columns are $v_1, \ldots, v_n$. Then a state x in $\mathbb{Z}_2^n$ is mapped to state $u = Mx$. After enough steps, with high probability the state in $\mathbb{Z}_2^n$ has many 1's, and the number of such states is $2^{\approx d}$. Since the map amounts to a universal hash function, these states are mapped to $\mathbb{Z}_2^d$ approximately evenly, yielding a distribution close to uniformity.

We use the second moment method to argue that the distribution Q^{*k} is close to uniformity. Since it isn't likely that the state in $\mathbb{Z}_2^n$ will have an unusual number of 1's, these states collectively don't contribute much to the distance from uniformity. However, these states contribute more than their fair share to the second moment. The straightforward application of the second moment method yields the same upper bound as that given by the L_2 norm in Section 5. So the states with a typical number of 1's and the states with an unusual number of 1's need to be dealt with separately.

Define $R_i(x)$ on $\mathbb{Z}_2^n$ with

$$R_i(x) = \begin{cases} Q_n^{*k}(x) & w(x) = i \\ 0 & \text{otherwise} \end{cases} ,$$

where $w(x)$ is the Hamming weight (number of ones) of state x.

Let m be a map from $\mathbb{Z}_2^n$ to $\mathbb{Z}_2^d$. If f_x is the function on $\mathbb{Z}_2^n$ which is one at x and zero elsewhere, then let mf_x denote the function on $\mathbb{Z}_2^d$ which is one at $m(x)$ and zero elsewhere. Extend this definition linearly so that m maps arbitrary real-valued functions on $\mathbb{Z}_2^n$ to functions on $\mathbb{Z}_2^d$. If Q_n denotes the walk on $\mathbb{Z}_2^n$, then $Q^{*k} = (x \mapsto Mx)Q_n^{*k}$, where $x \mapsto y$ denotes the function which maps x to y.

With this notation, $Q^{*k} = \sum_{i=0}^n (x \mapsto Mx)R_i$. Let $r_i = \|R_i\|_1$. Then by the triangle inequality,

$$\begin{aligned} \|Q^{*k} - U\|_1 &\le \sum_{i=0}^n \|(x \mapsto Mx)R_i - r_i U\|_1 \\ &= \sum_{i=0}^n \|(x \mapsto Mx + b)R_i - r_i U\|_1 \end{aligned}$$

for any $b \in \mathbb{Z}_2^d$. For convenience we will assume that b is chosen uniformly at random, and will denote $(x \mapsto Mx + b)R_i$ by S_i.

Let $1_{x \mapsto u}$ be the random variable which is one if $u = Mx + b$, and zero otherwise. Since b is chosen at random, $E[1_{x \mapsto u}] = 2^{-d}$, and since M is chosen at random,

$$E[1_{x \mapsto u} 1_{y \mapsto u}] = \begin{cases} 2^{-2d} & x \neq y \\ 2^{-d} & x = y \end{cases} .$$

We have

$$\begin{aligned} S_i(u) &= \sum_{x \in \mathbb{Z}_2^n} R_i(x) 1_{x \mapsto u} \\ E[S_i(u)] &= \sum_{x \in \mathbb{Z}_2^n} R_i(x) E[1_{x \mapsto u}] \\ E[S_i(u)] &= r_i 2^{-d} \end{aligned}$$

and

$$\begin{aligned} E\left[S_i(u)^2\right] &= \sum_{x,y \in \mathbb{Z}_2^n} R_i(x) R_i(y) E[1_{x \mapsto u} 1_{y \mapsto u}] \\ &= 2^{-2d} \sum_{x,y} R_i(x) R_i(y) + \left(2^{-d} - 2^{-2d}\right) \sum_x R_i(x) R_i(x) \\ E\left[S_i(u)^2\right] &\leq 2^{-2d} r_i^2 + 2^{-d} r_i \max_x R_i(x) \\ \operatorname{Var}[S_i(u)] &\leq r_i^2 2^{-d} \Big/ \binom{n}{i} . \end{aligned}$$

Next we use $E[Z]^2 \leq E\left[Z^2\right]$ with $Z = |S_i(u) - E[S_i(u)]|$ to get

$$\begin{aligned} E[|S_i(u) - E[S_i(u)]|]^2 &\leq \operatorname{Var}[S_i(u)] \\ E\left[\left|S_i(u) - r_i 2^{-d}\right|\right] &\leq \left[r_i^2 2^{-d} \Big/ \binom{n}{i}\right]^{1/2} \\ E\left[\sum_u \left|S_i(u) - r_i 2^{-d}\right|\right] &\leq \left[r_i^2 2^{d} \Big/ \binom{n}{i}\right]^{1/2} \\ E[\|S_i - r_i U\|_1] &\leq r_i \left[2^d \Big/ \binom{n}{i}\right]^{1/2} . \end{aligned}$$

On the other hand, $\|S_i - r_i U\|_1 \leq 2 r_i$.

Case $n = O(d)$: Let I_k be the number of 1's in the kth state in $\mathbb{Z}_2^n$. Recall that $E[I_k] = n/2(1 - (1 - 2/n)^k)$, and $\operatorname{Var}[I_k] \leq n/2$. By Chebychev,

$$\begin{aligned} \Pr\left[|I_k - E[I_k]| > c\sqrt{n}\right] &\leq 1/\left(2c^2\right) \\ \sum_{i:|i - E[I_k]| > c\sqrt{n}} \|S_i - r_i U\|_1 &\leq 1/c^2 . \end{aligned}$$

But

$$\begin{aligned}\sum_{i:|i-E[I_k]|\le c\sqrt{n}} E[\|S_i - r_i U\|_1] &\le \left[2^d \Big/ \binom{n}{\lceil E[I_k] - c\sqrt{n}\,\rceil}\right]^{1/2}\\ &\le \sqrt{2}^{d-nH(E[I_k]/n-c/\sqrt{n})}\sqrt[4]{\pi n/2}\,e^{1/12}\\ &\le \sqrt{2}^{-y}\sqrt[4]{\pi/2}\,e^{1/12}\end{aligned}$$

when

$$\begin{aligned}d + y + (\lg n)/2 &\le nH(E[I_k]/n - c/\sqrt{n})\\ 2H^{-1}((d + y + (\lg n)/2)/n) + 2c/\sqrt{n} &\le 2E[I_k]/n\\ k\ln[1 - 2/n] &\le \ln\left[1 - 2H^{-1}((d + y + (\lg n)/2)/n) - 2c/\sqrt{n}\,\right]\\ k &\ge \frac{\ln[1 - 2H^{-1}((d + y + (\lg n)/2)/n) - 2c/\sqrt{n}\,]}{\ln[1 - 2/n]}\end{aligned}$$

Then assuming that k satisfies this constraint,

$$\Pr\left[\|Q^{*k} - U\|_1 \ge 1/c^2 + 2^{-y/4}\right] \le \sqrt[4]{\pi/2}\,e^{1/12}2^{-y/4}\ ,$$

i.e. almost all Markov chains will have mixed after k steps.

Case $n = d + \Theta(d)$: Same as the $n = d + \Theta(d)$ case in the lower bound, except with $(1 + \varepsilon)$ rather than $(1 - \varepsilon)$.

Case $n = d + o(d) = d + \omega(\ln d)$: Similar to the $n = d + o(d)$ case in the lower bound, switching y with $-y$ and x with $-c$. This time we need γ big enough, e.g. $\gamma \ge y + (\lg n)/2$.

Case $n \gg d$: This case is most readily solved by using the L_2 norm bounds proven in the next few sections, and noting that these bounds are asymptotic to the claimed result when $n \gg d$.

Proof of Theorem 6. Immediate from the above cases.

Case $n = d + O(\ln d)$: Here we use the condition on the Markov chain being ergodic. Consider the random walk on $\mathbb{Z}_2^n$. Since the chain is ergodic, 2^{n-d-1} even-parity states in $\mathbb{Z}_2^n$ map to each state in $\mathbb{Z}_2^d$, and similarly for the odd-parity states. After $\sim 1/4n\ln n$ steps, except for the parity of the state, the state in $\mathbb{Z}_2^n$ is very close to being random, and the state in $\mathbb{Z}_2^d$ is at least as close to being random. But since $\gamma = n - d$ is sub-polynomial in d, $1/4n\ln(n/\gamma) \sim 1/4d\ln d$. See also Ross & Xu (1993) for a study of worst case mixing time when the Markov chain is ergodic and the v_i are distinct.

*Proof of Theorem 1***.** Immediate from Theorems 2 and 6, and the above case. □

Remark. We assumed that M is a random 0-1 matrix, but the only property of M that we used is that

$$\Pr[Mz = 0] \le 2^{-d} \quad \text{when } z \neq 0 \ .$$

Choosing M at random requires nd random bits. We can use an upper-triangular Toeplitz matrix to get by with only $n-1$ random bits. Using Wozencraft's ensemble of randomly shifted codes (see Massey (1963, page 21)) we can get the same bounds using only $m = \max(d, n-d)$ random bits and a primitive irreducible polynomial of degree m. In fact we don't need the polynomial to be primitive, so we can compute it deterministically using an algorithm due to Shoup (1990). Using either of these ensembles will guarantee that the Markov chain is connected. We can guarantee ergodicity by using one less random bit. The expected values above can at most double, but they still vanish.

4. Using the second largest eigenvalue

For general time-reversible Markov chains, the second-largest eigenvalue of the state transition matrix is frequently used to upper bound the mixing time of the random walk. See Diaconis and Stroock (1991). If the Markov chain has m states and converges to uniformity, then

$$\|Q^{*k} - U\|_1^2 \le (m-1)\lambda_*^{2k} \ , \tag{2}$$

where λ_* is the second largest eigenvalue in absolute value. This section determines the value that λ_* approximates with high probability, and gives the upper bound on the mixing time which follows from Equation 2. When $n \gg d$ this bound is tight, but when $n \approx d$, this bound is off by a large factor. The next section gives the mixing time upper bound derived using all the eigenvalues.

For the random walk on $\mathbb{Z}_2^d$ where the probability of adding v to the current state is $Q(v)$, the eigenvalues are associated with group elements: for $u \in \mathbb{Z}_2^d$, λ_u is an eigenvalue of the random walk given by

$$\lambda_u = \sum_v Q(v)(-1)^{u \cdot v}$$

(see for instance Diaconis (1988)). λ_* is just $\max_{u \neq 0} |\lambda_u|$. Since

$$\|Q^{*k} - U\|_1 < 2^{d/2}\lambda_*^k = 2^{d/2}e^{-k \ln \lambda_*^{-1}} \ ,$$

after

$$k = \frac{d(\ln 2)/2 + c}{\ln \lambda_*^{-1}}$$

steps, $\|Q^{*k} - U\|_1 < e^{-c}$.

Next we derive bounds on λ_* assuming the support of Q is randomly chosen. Let S_n denote the probability distribution of a random variable which is the sum of n independent identically distributed random variables which are ± 1 with probability $1/2$. Suppose vectors $v_1, \ldots, v_n$ are chosen uniformly at random. If $u \neq 0$, then the random variables $u \cdot v_1, \ldots, u \cdot v_n$ are i.i.d. 0-1 random variables with probability $1/2$. If the translating vector is chosen uniformly at random from $v_1, \ldots, v_n$, then λ_u is distributed according to S_n/n. Chernoff's bound gives

$$\Pr[|\lambda_u| \geq t] \leq 2e^{-t^2 n/2} \enspace .$$

With $n = a(d+b)$ and $t = \sqrt{\ln 4/a}$, we have

$$\Pr[\lambda_* \geq t] < 2^d 2e^{-t^2 n/2} = 2^{1-b} \enspace .$$

Thus with probability more than $1 - 2^{1-b}$, after

$$\sim \frac{d + 2c/\ln 2}{\lg[(n/(d+b))/\ln 4]}$$

steps $\|Q^{*k} - U\|_1 < e^{-c}$. In order for b to be negligible in this expression, we need $b \ll d$, and also $\lg[(d+b)/d] \ll \lg[n/(d \ln 4)]$. If $d \gg M \gg b$ and $n - d\ln 4 \geq M$, then

$$\lg \frac{n}{d \ln 4} \geq \lg\left[1 + \frac{M}{d \ln 4}\right] \sim \frac{M}{d \ln 4} \gg \frac{b}{d} \sim \lg \frac{d+b}{d} \enspace .$$

The mixing time bound we get is therefore

$$\sim \frac{d}{\lg((n/d)/\ln 4)} \enspace ,$$

when $d \gg c$, $d \gg b$, and $n - d \ln 4 \gg b$.

We can get a better bound using

$$2^n \Pr[|\lambda_u| \geq t]/2 \leq \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{\lfloor n(1-t)/2 \rfloor} \leq 2^{nH\left(\frac{1-t}{2}\right)} \enspace .$$

With $t = 1 - 2H^{-1}(1 - 1/a)$,

$$\Pr[\lambda_* \geq t] < 2^{1-b} \enspace ,$$

yielding an upper bound of

$$\sim \frac{d}{2 \lg 1/[1 - 2H^{-1}(1 - d/n)]} \tag{3}$$

almost always if $d \gg c$, $d \gg b$, and $n - d \gg b$.

When $n \gg d$, both these bounds are tight.

Now we argue that bound 3 for average random walks is the best possible using only Equation 2. Specifically we show that it is a rare event that λ_* is less than $1 - 2H^{-1}(1 - d/n)$ by more than a factor of $1 + o(1)$. Let T be the number of eigenvalues λ_u ($u \neq 0$) that are larger than a certain threshold t, and let p_t be the probability that a particular eigenvalue exceeds t. Consider two vectors u and w, $0 \neq u \neq w \neq 0$, and the eigenvalues λ_u and λ_w. The values $u \cdot v_i$ and $w \cdot v_i$ are 0-1 r.v.'s with probability $1/2$, and they are independent because $(u - w) \cdot v_i$ is a 0-1 r.v. with probability $1/2$. Thus λ_u is independent of λ_w. With the eigenvalues being pairwise independent, we can compute the variance of T:

$$\begin{aligned}\operatorname{Var}[T] &= E\left[T^2\right] - E[T]^2 \\ &= \left(2^d-1\right)\left(2^d-2\right)p_t^2 + \left(2^d-1\right)p_t - \left(2^d-1\right)^2 p_t^2 \\ &= \left(2^d-1\right)\left(p_t - p_t^2\right) \le E[T] \ .\end{aligned}$$

In particular, if $E[T]$ is large, Chebychev's inequality implies (among other things) that with high probability T is nonzero, so $\lambda_* > t$. From Stirling's formula we see that for fixed $\delta > 0$, the expected number of eigenvalues exceeding $(1-\delta)(1-2H^{-1}(1-d/n))$ goes to infinity.

5 Using all the eigenvalues

An important tool for bounding the mixing time of random walks on groups is the Diaconis-Shahshahani upper bound lemma (Diaconis 1988). When applied to the random walk on $\mathbb{Z}_2^d$, the lemma says

$$\|Q^{*k} - U\|_1 \le \sqrt{2^d}\|Q^{*k} - U\|_2 = \left[\sum_{u\neq 0} \lambda_u^{2k}\right]^{1/2} . \tag{4}$$

This section derives the upper bound on the mixing time which follows from all the eigenvalues in Equation 4.

We need only consider the case $n = d + \Theta(d)$, because Equation 4 is at least as strong as the second largest eigenvalue bound and Greenhalgh's bound, and these bounds are tight when $n = \omega(1)d$ and $n = (1+o(1))d$ respectively. Consider $E\left[\sum_{u\neq 0} \lambda_u^{2(k+k')}\right] \le (2^d-1)E\left[(S_n/n)^{2k}\right]\lambda_*^{2k'}$.

$$E\left[(S_n/n)^{2k}\right] = \sum_{i=0}^{n} \binom{n}{i} 2^{-n}\left(1 - \frac{2i}{n}\right)^{2k} .$$

Let $\gamma = 1 - 2i/n$.

$$E\left[(S_n/n)^{2k}\right] \le (n+1)2^{-n} \max_{0\le\gamma\le 1} 2^{nH((1-\gamma)/2)}\gamma^{2k} .$$

To find the maximum we differentiate:

$$[nH((1-\gamma)/2) + 2k\lg\gamma]' = n\lg\left[\frac{1}{(1-\gamma)/2} - 1\right]\left(\frac{-1}{2}\right) + \frac{2k}{\gamma\ln 2} = 0 \ ,$$

which is solved with

$$\frac{k}{n} = \frac{1}{4}\gamma\ln\frac{1+\gamma}{1-\gamma} . \tag{5}$$

Setting

$$2^{nH((1-\gamma)/2)}\gamma^{2k} = 2^{n-d}$$

gives

$$\frac{d}{n} = 1 - H\left(\frac{1-\gamma}{2}\right) - 2\frac{k}{n}\lg\gamma . \tag{6}$$

Given n and d, γ is determined by Equations 5 and 6. Find k with Equation 5 and round it up. So then $E\left[(S_n/n)^{2k}\right] \leq n+1$. From Section 4, with high probability we can pick k' to be $O(\log n)$ to get $\|Q^{*(k+k')} - U\|_1 = o(1)$. But k is linear in n, so $k' \ll k$. Equations 5 and 6 have been used to parametrically plot the upper bound in Figure 2.

It is impossible to derive a bound which is better by more than a factor of $1+o(1)$ using only Equation 4, at least on average. The reason is as in Section 4, the eigenvalues are pairwise independent. Assume γ is bounded away from 1, since otherwise it is trivially impossible to derive a better bound, as there is a matching lower bound. The expected number of eigenvalues that are $\geq \gamma$ is $2^{nH((1-\gamma)/2)-n+d-O(\log n)} = \gamma^{-2k}/n^{O(1)} \to \infty$. Since the mean dominates the variance, Chebychev's inequality implies that with high probability about that many eigenvalues are $\geq \gamma$.

6. Continuous time

In this section we briefly consider the continuous time version of the random walk. The continuous time version is similar to the discrete time version except that 1) the proofs are easier, 2) there is no threshold when $T(d,n) = O(1)$ (i.e. when n is exponentially large in d), and 3) the eigenvalue bounds are worse when $n/d = 1 + \Omega(1)$. In fact, even when all the eigenvalues are used, the upper bound is never better than $\sim (\ln 2/2)d$, no matter how large n is, at least in the average case. Here we derive the upper bound given by all the eigenvalues.

This derivation follows the proof of Greenhalgh's (1993) bound, but uses one less approximation. In continuous time, the eigenvalues are given by

$$\lambda_u = \exp\left[-1 + \sum_{i=1}^{n} Q(v_i)(-1)^{u\cdot v_i}\right] \ .$$

It is now easier to evaluate $E\left[\lambda_u^{2k}\right]$. For $u \neq 0$,

$$\begin{aligned} E\left[\lambda_u^{2k}\right] &= 2^{-n}\sum_{i=0}^{n}\binom{n}{i}e^{-4ki/n} \\ &= 2^{-n}\left[1+e^{-4k/n}\right]^n \ . \end{aligned}$$

To get $2^d E\left[\lambda_u^{2k}\right] = 1$, we need

$$\begin{aligned} 2^{n-d} &\geq \left[1+e^{-4k/n}\right]^n \\ 2^{1-d/n} - 1 &\geq e^{-4k/n} \\ n/4 \ln 1/\left[2^{1-d/n} - 1\right] &\leq k \ . \end{aligned}$$

If $n \gg d$, then

$$\begin{aligned} 2^{1-d/n} &= 2(1- \sim (\ln 2)d/n) \\ 2^{1-d/n} - 1 &= 1- \sim (d\ln 4)/n \end{aligned}$$

$$\begin{aligned} \ln 1/\left[2^{1-d/n}-1\right] &= \sim (d\ln 4)/n \\ n/4 \ln 1/\left[2^{1-d/n}-1\right] &= \sim d\ln 2/2 \ . \end{aligned}$$

7. Discussion

For the ordinary random walk on the hypercube, the mixing time threshold is $\frac{1}{4}d\ln d$. When n is large compared with d, the random walk on $\mathbb{Z}_2^d$ based on n random group elements has a mixing time threshold of $d/\lg(n/d)$. These are two limiting cases of a threshold function valid for all $n > d$, but whose interesting behavior occurs when n is linear d. Coincidentally, the Diaconis-Shahshahani upper bound lemma yields tight upper bounds when $n/d \to 1$ and when $n/d \to \infty$, but not when $n/d \to c \in (1, \infty)$.

The techniques used here generalize to give tight bounds for the mixing time of random random walks on $\mathbb{Z}_b^d$ (fixed b), but tight bounds for other Abelian groups, such as $\mathbb{Z}_2^d \times \mathbb{Z}_4^d$, are not known. In a forthcoming paper it will be shown that the mixing time threshold for $\mathbb{Z}_2^d$ upper bounds the mixing time of all Abelian groups, i.e. part 2 of Theorem 1 holds for Abelian groups G with $\lg|G|$ replacing d. Since Abelian groups are generally slow to mix compared with other groups, it is natural to conjecture that $\mathbb{Z}_2^d$ is the most slowly mixing of groups with respect to random random walks:

Conjecture 7 *For all $\varepsilon > 0$, for all sufficiently large groups G, if $n - \lg|G| \gg 1$, then almost any $S = \{g_1, \dots, g_n\} \subset G$ defines a random walk $G(S)$ which satisfies $\|Q^{*k} - U\|_{\mathrm{TV}} \le \varepsilon$ for*

$$k \ge (1+\varepsilon)\frac{n}{2}\ln\frac{1}{1-2H^{-1}(\lg|G|/n)} \ .$$

Acknowledgements. I would like to thank Persi Diaconis for posing such a delightful problem, and for his encouragement. I thank Martin Hildebrand for encouraging me to improve on my earlier results. I also thank Diaconis and Dan Kleitman for much needed help with the exposition. Bonnie Berger, Peter Shor, Brad Spiers, Richard Stanley, and David Zuckerman provided helpful comments.

A The inverse binary entropy function

In this appendix we derive those properties of $H^{-1}(x)$ and $1 - 2H^{-1}(x)$ that we use in the article, in particular the asymptotic behavior as $x \to 0$ or $x \to 1$.

We will make use of some higher-order derivatives of the binary entropy function H.

$$\begin{aligned} H(x) &= x\lg 1/x + (1-x)\lg 1/(1-x) \\ -\ln 2H(x) &= x\ln x + (1-x)\ln(1-x) \\ -\ln 2H'(x) &= \ln x + 1 - \ln(1-x) - 1 = \ln(x/(1-x)) \end{aligned}$$

$$\begin{aligned}
-\ln 2H''(x) &= 1/x + 1/(1-x)\\
-\ln 2H'''(x) &= -1/x^2 + 1/(1-x)^2\\
-\ln 2H''''(x) &= 2/x^3 + 2/(1-x)^3
\end{aligned}$$

From these it is clear that $H(x)$ increases monotonically from 0 at $x = 0$ to 1 at $x = 1/2$, and then decreases back to 0 at $x = 1$. For $0 \le y \le 1$ we define $H^{-1}(y)$ to be the value of x between 0 and 1/2 for which $H(x) = y$.

By Taylor's theorem,

$$H\left(\frac{1}{2} - \varepsilon\right) = 1 - \frac{2}{\ln 2}\varepsilon^2 + O\left(\varepsilon^4\right) = 1 - \delta \ ,$$

where the constant in the big-oh notation is between 4/3 and 6 when $|\varepsilon| \le 1/4$. Thus $H^{-1}(1-\delta) = 1/2 - \varepsilon$ where $\delta \sim (2/\ln 2)\varepsilon^2$ when $\varepsilon \to 0$, and we conclude

Lemma 8.

$$\begin{aligned}
H^{-1}(1-\delta) &= \frac{1}{2} - \sim \sqrt{\frac{(\ln 2)\delta}{2}}\\
1 - 2H^{-1}(1-\delta) &\sim \sqrt{(\ln 4)\delta}
\end{aligned}$$

when $\delta \to 0$.

To get the other limit we let $y = w/\lg(1/w)$, and assume $w \to 0$. Then

$$\begin{aligned}
H(y) =&\ \frac{w}{\lg(1/w)}\lg\frac{\lg(1/w)}{w} + \left(1 - \frac{w}{\lg(1/w)}\right)\lg\frac{1}{1 - w/\lg(1/w)}\\
H(y) =&\ w + \frac{w \lg\lg(1/w)}{\lg(1/w)} + (1 - o(1))\frac{w}{\lg(1/w)}\\
H(y) =&\ (1 + o(1))w
\end{aligned}$$

For any fixed $\varepsilon > 0$, if x is small enough we have

$$H\left(\frac{(1-\varepsilon)x}{-\lg((1-\varepsilon)x)}\right) < x < H\left(\frac{(1+\varepsilon)x}{-\lg((1+\varepsilon)x)}\right) \ ,$$

so the monotonicity of H implies

Lemma 9.

$$H^{-1}(x) \sim \frac{x}{\lg(1/x)}$$

as $x \to 0$.

References

Aldous, D., Diaconis, P.: Shuffling cards and stopping times. Technical Report 231, Department of Statistics, Stanford University (1985)

Alon, N., Roichman, Y.: Random Cayley graphs and expanders. Random Structures and Algorithms, **5**(2), 271–284 (1994)

Diaconis, P.: Group Representations in Probability and Statistics. Institute of Mathematical Statistics (1988)

Diaconis, P.: Personal communication (1993)

Diaconis, P., Graham, R.L., Morrison, J.A.: Asymptotic analysis of a random walk on a hypercube with many dimensions. Random Structures and Algorithms, **1**(1), 51–72 (1990)

Diaconis, P., Shahshahani, M.: Time to reach stationarity in the Bernoulli-Laplace diffusion model. SIAM Journal on Mathematical Analysis, **18**(1), 208–218 (1987)

Diaconis, P., Stroock, D.: Geometric bounds for eigenvalues of Markov chains. The Annals of Applied Probability, **1**(1), 36–61 (1991)

Dou, C., Hildebrand, M.: Enumeration and random random walks on finite groups. The Annals of Probability, **24**(2), 987–1000 (1996)

Dou, C.C.Z.: Studies of Random Walks on Groups and Random Graphs. PhD thesis, Department of Mathematics, MIT (1992)

Greenhalgh, A.S.: A model for random random-walks on finite groups. Submitted (1993)

Hildebrand, M.: Personal communication (1963)

Massey, M.: Threshold Decoding. M.I.T. Press (1963)

McDiarmid, C.: On the method of bounded differences. In Surveys in Combinatorics, 1989 pages 148–188 (1989)

Peterson, W.W., Weldon, Jr., E.J.: Error-Correcting Codes. The MIT Press, second edition (1972)

Roichman, Y.: On random random walks. The Annals of Applied Probability, **24**(2), 1001–1011 (1996)

Ross, K.A., Xu, D.: A comparison theorem on convergence rates of random walks on groups. Journal of Theoretical Probability, **6**(2), 323–343 (1993)

Shoup, V.: New algorithms for finding irreducible polynomials over finite fields. Mathematics of Computation, **54**(189), 435–447 (1990)

Sinclair, A.: Algorithms for Random Generation and Counting: A Markov Chain Approach. Birkhäuser (1993)

Wilson, D.B.: Random random walks on Abelian groups. Manuscript (1994)