Adv. Appl. Prob. **49,** 1288–1299 (2017)
*doi:*10.1017/apr.2017.42
© *Applied Probability Trust* 2017

A NON-LOCAL RANDOM WALK ON THE HYPERCUBE

EVITA NESTORIDI,* *Stanford University*

Abstract

In this paper we study the random walk on the hypercube $(\mathbb{Z}/2\mathbb{Z})^n$ which at each step flips k randomly chosen coordinates. We prove that the mixing time for this walk is of the order $(n/k)\log n$. We also prove that if $k = o(n)$ then the walk exhibits cutoff at $(n/2k)\log n$ with window $n/2k$.

Keywords: Hypercube; coupling; random walk; Ehrenfest urn model

2010 Mathematics Subject Classification: Primary 60J10
Secondary 60C05

1. Introduction

Consider two urns, one containing zero balls and the second containing n balls. At each step, pick k total balls at random and move each of them to the opposite urn. This is a generalization of the Ehrenfest urn model, which works for $k = 1$.

The above Markov chain can be also viewed as a random walk on $(\mathbb{Z}/2\mathbb{Z})^n$, where at each step we flip k random coordinates (for some fixed k). For the walk to be transitive, k needs to be odd, and to avoid parity problems it is simplest to consider the lazy version of this walk. In other words, at each step, do nothing with probability $\frac{1}{2}$ and with probability $\frac{1}{2}$ choose a random set of k coordinates and flip them. The main question is to find the mixing time of this walk for the total variation distance.

This non-local walk implies a big change at each step. Almost all of the tools and examples developed over the past years give careful rates for local Markov chains, such as nearest neighbor random walks on graphs and the usual implementation of the Gibbs sampler and Metropolis algorithm. On the other hand, there are a host of much more global algorithms such as Swedsen–Wang, Wolff, hit and run, and, in card shuffling, the riffle shuffles and hyperplane arrangement walks, where the chain moves quite far in one step. See Andersen and Diaconis [4] for a survey of many such algorithms. Cheeger and path arguments are not set up to deal with these non-local Markov chains. In this paper we carry out a careful study of this kind of long-range walk via techniques capable of giving sharp results.

There is a second reason why this particular random walk is interesting. Fix $x \in (\mathbb{Z}/2\mathbb{Z})^n$. Let P denote the transition matrix of the random walk and let $P^{*\ell}$ denote the ℓth power of P. There are two different approaches to finding the mixing time of this walk. The first approach is developed in Section 5. It involves finding the eigenvalues of the walk using representation theory and using the Fourier transform to give bounds on the ℓ^2-norm of the difference $P_x^{*\ell} - U$, where U is the uniform measure and $P_x^{*\ell}$ is the row of $P^{*\ell}$ corresponding to x. For the $k = 1$ case, this technique works nicely and gives a sharp upper bound on the mixing time. However,

Received 17 June 2016; revision received 21 July 2017.
* Current address: Department of Mathematics, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544-1000, USA. Email address: exn@princeton.edu

https://doi.org/10.1017/apr.2017.42 Published online by Cambridge University Press

for $k = n/2$, it turns out that the bound obtained via the L^2-norm does not give a sharp upper bound on the mixing time, which is defined in terms of the total variation distance (L^1-norm).

A second argument via coupling is introduced in Section 3. It provides a solution to the general case and makes the difference between the L^2-norm and the total variation distance clear. This coupling argument is a generalization of one used by Aldous [2] for the $k = 1$ case. See [1] for more results of Aldous concerning the hypercube. The lower bound uses the first two eigenvectors and eigenvalues of the random walk and the second moment technique. This method was firstly introduced by Diaconis and Shahshahani [8]. The authors managed to prove a lower bound for the $k = 1$ case that matched Aldous' upper bound, proving in this way the existence of a cutoff at $\frac{1}{4}(n+1)\log n$. Another way to find a lower bound was proved by Saloff-Coste [14] using Wilson's lemma, which is Lemma 2 of [16]. In [9], the authors used Fourier analysis directly to derive the exact behavior of the error for the nearest neighbor random walk.

It is easy to see that the mixing time for the k model and the $n-k$ model will be the same, therefore we will focus on the $k \le n/2$ case. The results of this paper are as follows.

Theorem 1. *For the lazy walk changing $k \le n/2$ coordinates on the hypercube, the following hold for every $x \in (\mathbb{Z}/2\mathbb{Z})^n$.*

(a) *For $\ell = (n^2/2k(n-k))\log n + c(n^2/k(n-k))$, we have*

$$\| P_x^{*\ell} - U \|_{\mathrm{TV}} \le \mathrm{e}^{-c} + 2^{-c},$$

where $c > 0$.

(b) *For $\ell = (n/2k)\log n - c(n/k)$, where $0 < c \le \frac{1}{4}\log n$ and, for x being the identity element, we have*

$$\| P_{\mathrm{id}}^{*\ell} - U \|_{\mathrm{TV}} \ge 1 - \frac{B}{\mathrm{e}^{4c}}$$

for a uniformly bounded constant $B > 0$.

In the following corollary we discuss cutoff.

Corollary 1. *If $k = o(n)$ then the walk exhibits cutoff at $(n/2k)\log n$ with window $n/2k$.*

Section 7 contains the analysis for the L^2-mixing time of the random walk on $(\mathbb{Z}/m\mathbb{Z})^n$ generated by the measure

$$Q(a_{i_1}e_{i_1} + a_{i_2}e_{i_2} + \cdots + a_{i_k}e_{i_k}) = 1 \Big/ \binom{n}{k} m^k, \tag{1}$$

where $\{e_j\}_{j=1}^n$ is the standard base, $a_{i_j} \in \mathbb{Z}/m\mathbb{Z}$, and $\{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots, n\}$. Let $Q(x, xg) = Q(g)$ for every $x, g \in (\mathbb{Z}/m\mathbb{Z})^n$. The main result of Section 7 is the following theorem.

Theorem 2. *For the walk generated by Q, if $\ell = ((n+1)/2k)\log(mn) + c(n+1)/2k$ then, for every $x \in (\mathbb{Z}/m\mathbb{Z})^n$, we have*

$$4\| Q_x^{*\ell} - U \|_{\mathrm{TV}}^2 \le \mathrm{e}^{-c}.$$

It is known that the total variation mixing time is faster than that (using the fact that the first time that we have touched all coordinates is a strong stationary time) but the above result

https://doi.org/10.1017/apr.2017.42 Published online by Cambridge University Press

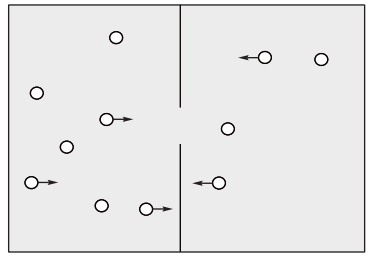


FIGURE 1: Eleven particles in two containers, five of each changing containers.

holds for the L^2-norm, which allows us to use comparison theory to provide bounds for the L^2-mixing times of the walk on $(\mathbb{Z}/m\mathbb{Z})^n$ generated by

$$\tilde{\mathbb{P}}(\pm e_1) = \frac{1}{4n}, \qquad \tilde{\mathbb{P}}(\mathrm{id}) = \frac{1}{2}.$$

More precisely, in Section 7.2 we prove the bound $m^2(((n+1)/2)\log(mn) + c(n+1)/2)$ for the mixing time of the last random walk. The analysis of the L^2-norm of the last walk has already been carried out by Diaconis and Saloff-Coste [7], where they proved an upper bound of order $m^2 n \log n$ and, following this, Saloff-Coste proved the cutoff [13].

2. The history of the Ehrenfest urn model

The Ehrenfest urn model was introduced by Ehrenfest and Ehrenfest [10] in order to study the second law of thermodynamics. This is a model for n particles distributed in two containers and each particle changes container independently from the others (see Figure 1). This process is repeated several times and the question is to find the limiting distribution of the process. Kac [11] approached this problem by finding the eigenvectors and eigenvalues of the transition matrix. He also proved that if the initial system state is not at equilibrium then the entropy is increasing.

This problem can also be viewed as a random walk on $(\mathbb{Z}/2\mathbb{Z})^n$, where the number of 1s in the binary vector represent the number of particles in the right-hand container. Flipping one (or k) coordinates of the binary vector corresponds to moving one (or k) particles to the other container. But now the Markov chain problem can be studied through a random walk on an abelian group, where representation theory is quite simple to use. As Diaconis writes in [6, Chapter 3], Kac posed the question: when can a Markov chain be lifted to a random walk on a group?

3. Coupling argument

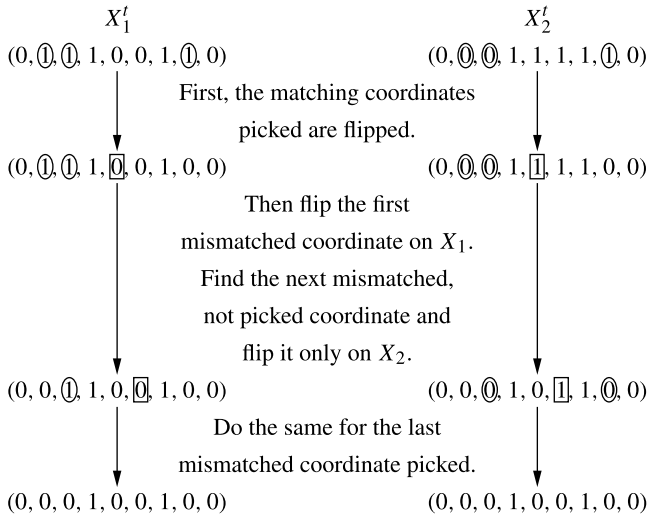
Consider the following measure on $(\mathbb{Z}/2\mathbb{Z})^n$:

$$\mathbb{P}(g) = \begin{cases} \frac{1}{2} & \text{if } g = \mathrm{id}, \\ 1 \Big/ 2\binom{n}{k} & \text{if } g \in (\mathbb{Z}/2\mathbb{Z})^n \text{ has } k \text{ 1s and } n-k \text{ 0s}, \end{cases}$$

and note that $P(x, xg) = P(g)$ for every $x, g \in (\mathbb{Z}/2\mathbb{Z})^n$.

What follows is the coupling argument which will provide an upper bound for the mixing time for $k \leq n/2$. Start with two different copies of the Markov chain. At time t, denote the

https://doi.org/10.1017/apr.2017.42 Published online by Cambridge University Press



In each chain, three coordinates were flipped, but $y(t) = 4$, while $y(t+1) = 0$.

FIGURE 2: A representation of how the coupling works.

state of each as X_1^t and X_2^t. We say that X_1 will start at a deterministic $x \in (\mathbb{Z}/2\mathbb{Z})^n$ and X_2 will start at a random configuration. At time t, let

$$y(t) = \|X_1^t - X_2^t\|_1 = \sum_i |X_1^t(i) - X_2^t(i)|,$$

where $X^t(i)$ denotes the ith coordinate of the corresponding vector. Then consider the following cases.

1. If $y(t)$ is odd then take one independent step on each chain according to the probability measure $\mathbb{P}$.
2. If $y(t)$ is even then with probability $\frac{1}{2}$ it stays fixed in both chains. With probability $1/2\binom{n}{k}$, choose k coordinates and denote them by $i_1 < \cdots < i_k$. Flip $X_1^t(i_1), \ldots, X_1^t(i_k)$. To determine the move on X_2^t, we need the following definition.

Definition 1. Denote by $a(t)$ the number of the mismatching coordinates among the k 1s selected.

If $a(t) > y(t)/2$ then flip $X_2^t(i_1), \ldots, X_2^t(i_k)$. If $a(t) \le y(t)/2$, for i_1 find the first $i_1' \notin \{i_1, \ldots, i_k\}$ such that $X_2^t(i_1') \neq X_1^t(i_1')$. Flip $X_2^t(i_1')$. Then find the second $i_2' \notin \{i_1, \ldots, i_k\}$ (cyclically) such that $X_2^t(i_2') \neq X_1^t(i_2')$, flip $X_2^t(i_1')$, and so on. In this way we determine which coordinates of X_2^t will be flipped.

Definition 2. If $x \in (\mathbb{Z}/2\mathbb{Z})^n$ is the starting configuration of X_1 then let T_x denote the first time the two chains match.

In the following lemma, which can be found as Lemma 5 of [6, Chapter 4], we state how the above coupling can be used to obtain an upper bound for the total variation distance.

https://doi.org/10.1017/apr.2017.42 Published online by Cambridge University Press

Lemma 1. *For every starting configuration $x \in (\mathbb{Z}/2\mathbb{Z})^n$, we have*

$$\|P_x^{*\ell} - U\|_{\mathrm{TV}} \le \mathbb{P}(T_x > \ell).$$

Remark 1. The proof of Theorem 1 does not depend on the starting configuration, since we will be bounding the quantity $\max_x \mathbb{P}(T_x > \ell)$.

4. Proof of Theorem 1

4.1. Upper bound

Proof of Theorem 1(a). At first, in the case that the starting number of the mismatched coordinates of the chains is odd, wait until the coupling suggests staying fixed at one of them and take a step on the other to turn the difference even. Call t_0 the time the above happens. Then t_0 follows a geometric distribution with probability of success $\frac{1}{2}$. Therefore,

$$\mathbb{P}\left(t_0 > c\frac{n^2}{2k(n-k)}\right) \le \frac{1}{2^{c(n^2/2k(n-k))}} \le 2^{-c}.$$

The coupling ensures that the distance remains even. To analyze the coupling, we imitate path coupling techniques as introduced by Bubley and Dyer [5]. As explained in [5, Section 2], we need to consider a new graph structure on $(\mathbb{Z}/2\mathbb{Z})^n$ and examine how the distance behaves after one step if X_1^ℓ and X_2^ℓ are in adjacent positions.

Let $p(x, y) = \|x - y\|/2$ and x be adjacent to y if and only if $p(x, y) = 1$. If $X_1^{\ell-1}$ and $X_2^{\ell-1}$ differ only at two coordinates, that is, $p(X_1^\ell, X_2^\ell) = 1$, then the probability of having $X_1^\ell = X_2^\ell$ is $2k(n-k)/n(n-1)$, which yields

$$\mathbb{E}\{\|X_1^\ell - X_2^\ell\| \mid X_1^{\ell-1}, X_2^{\ell-1}\} \le \left(1 - \frac{2k(n-k)}{n(n-1)}\right)\|X_1^{\ell-1} - X_2^{\ell-1}\|.$$

From path coupling, we see that this extends to

$$\mathbb{E}\{\|X_1^\ell - X_2^\ell\| \mid X_1^{t_0}, X_2^{t_0}\} \le n\left(1 - \frac{2k(n-k)}{n(n-1)}\right)^{\ell-t_0},$$

where $p(X_1^\ell, X_2^\ell)$ is not necessarily 1, as explained thoroughly in Theorem 6.2 of [15].

If $\ell = (n^2/2k(n-k))\log n + c(n^2/k(n-k))$, Markov's inequality yields

$$\begin{aligned}
\mathbb{P}(T > \ell) &= \mathbb{P}\left(T > \ell \;\middle|\; t_0 \le c\frac{n^2}{2k(n-k)}\right)\mathbb{P}\left(t_0 \le c\frac{n^2}{2k(n-k)}\right) \\
&\quad + \mathbb{P}\left(T > \ell \;\middle|\; t_0 > c\frac{n^2}{2k(n-k)}\right)\mathbb{P}\left(t_0 > c\frac{n^2}{2k(n-k)}\right) \\
&\le \mathbb{P}\left(\|X_1^\ell - X_2^\ell\| \ge 1 \;\middle|\; t_0 \le c\frac{n^2}{2k(n-k)}\right) + \mathbb{P}\left(t_0 > c\frac{n^2}{2k(n-k)}\right) \\
&\le \mathbb{E}\left\{\|X_1^\ell - X_2^\ell\| \;\middle|\; X_1^0, X_2^0, t_0 \le c\frac{n^2}{2k(n-k)}\right\} + 2^{-c} \\
&\le n\left(1 - \frac{2k(n-k)}{n(n-1)}\right)^{\ell - c(n^2/2k(n-k))} + 2^{-c} \\
&\le \mathrm{e}^{-c} + 2^{-c}.
\end{aligned}$$

□

https://doi.org/10.1017/apr.2017.42 Published online by Cambridge University Press

4.2. Lower bound

The lower bound will be proved using the eigenvectors and eigenvalues for this Markov chain. In this section, the random walk will begin at the identity. From Theorem 6 of [6, p. 49], we see that the eigenvalues are the Krawtchuck polynomials and the eigenvectors are the normalized Krawtchuck polynomials. To see this, note that the irreducible representations of $(\mathbb{Z}/2\mathbb{Z})^n$ are indexed by vectors $\boldsymbol{a} \in (\mathbb{Z}/2\mathbb{Z})^n$ so that

$$\rho_{\boldsymbol{a}}(\boldsymbol{v}) = (-1)^{\boldsymbol{a}\cdot\boldsymbol{v}},$$

where $\boldsymbol{v} \in (\mathbb{Z}/2\mathbb{Z})^n$. Therefore, the Fourier transform of P at $\rho_{\boldsymbol{a}}$ is

$$\widehat{P}(\rho_{\boldsymbol{a}}) \coloneqq \sum_{\boldsymbol{v}} \rho_{\boldsymbol{a}}(\boldsymbol{v})\mathbb{P}(\boldsymbol{v}) = \frac{1}{2} + \frac{1}{2}\sum_{b=0}^{k}(-1)^b\left(\binom{j}{b}\binom{n-j}{k-b}\Big/\binom{n}{k}\right),$$

where j denotes the number of coordinates of $\boldsymbol{a}$ that are equal to 1. According to Theorem 6 of [6], the eigenvalues of the transition matrix are exactly the $\widehat{P}(\rho_{\boldsymbol{a}})$, $\boldsymbol{a} \in (\mathbb{Z}/2\mathbb{Z})^n$. The corresponding (nonnormalized) eigenfunction is $f_{\boldsymbol{a}}(\boldsymbol{x}) = (-1)^{\boldsymbol{x}\cdot\boldsymbol{a}}$. Note that all $\boldsymbol{a} \in (\mathbb{Z}/2\mathbb{Z})^n$ that have the same number of 0s yield the same eigenvalue. Thus, if $|\boldsymbol{x}|$ denotes the number of ones of $\boldsymbol{x}$, the jth Krawtchuck polynomials

$$f_j(\boldsymbol{x}) = \sum_{b=0}^{|\boldsymbol{x}|}(-1)^b\left(\binom{|\boldsymbol{x}|}{b}\binom{n-|\boldsymbol{x}|}{j-b}\Big/\binom{n}{j}\right)$$

are eigenfunctions and their normalized form will be used to compute the lower bound for the mixing time.

Proof of Theorem 1(b). Recall the definition of the total variation distance

$$\|P - Q\|_{\mathrm{TV}} = \sup_A |P(A) - Q(A)|.$$

A specific set A will provide a lower bound. To find this lower bound, consider the normalized Krawtchuck polynomial of degree one $f(x) = \sqrt{n}(1 - 2x/n)$ and the nonnormalized Krawtchuck polynomial of degree two $f_2(x) = 1 - 4x/(n-1) + 4x^2/(n^2 - n)$. Then consider $A_\alpha = \{x : |f(x)| \le \alpha\}$. A specific choice for α will guarantee the correct lower bound.

The orthogonality relations that the normalized Krawtchuck polynomials satisfy yield that if Z is a point chosen uniformly in $X = \{0, 1, 2, \ldots, n\}$ then

$$\mathbb{E}\{f(Z)\} = 0 \quad \text{and} \quad \mathrm{var}\{f(Z)\} = 1.$$

If $Z_\ell = \|X_\ell\|$ then we have

$$\mathbb{E}\{f(Z_\ell)\} = \sqrt{n}\left(\frac{1}{2} + \frac{1}{2}\left(1 - \frac{2k}{n}\right)\right)^\ell = \sqrt{n}\left(1 - \frac{k}{n}\right)^\ell,$$

because f is an eigenfunction of the Markov chain corresponding to the eigenvalue $1 - k/n$. Again, under the convolution measure, we have

$$\begin{aligned}\mathrm{var}\{f(Z_\ell)\} &= \frac{n}{n} + \frac{n(n-1)}{n}\left(1 + \frac{2k^2 - 2kn}{n^2 - n}\right)^\ell - n\left(1 - \frac{k}{n}\right)^{2\ell} \\ &= 1 + (n-1)\left(1 - \frac{2kn - 2k^2}{n^2 - n}\right)^\ell - n\left(1 - \frac{k}{n}\right)^{2\ell}.\end{aligned}$$

https://doi.org/10.1017/apr.2017.42 Published online by Cambridge University Press

Recall that the first three (nonnormalized) eigenfunctions of this Markov chain are

$$f_0(x) = 1, \qquad f_1(x) = 1 - \frac{2x}{n}, \qquad f_2(x) = 1 - \frac{4x}{n-1} + \frac{4x^2}{n^2-n}.$$

By direct computation, we have

$$f_1^2(x) = \frac{1}{n} f_0(x) + \frac{n-1}{n} f_2(x).$$

Combining this and the fact that f_2 corresponds to the eigenvalue $1 + (2k^2 - 2kn)/(n^2 - n)$ yields the claimed variance. Now take ℓ of the form $(n/2k)\log n - c(n/k)$ (where $c > 0$). We consider two cases.

Case 1. Consider $k = n/d$, where d is a constant:

$$\mathbb{E}(f(Z_\ell)) = \sqrt{n}\left(1 - \frac{k}{n}\right)^\ell = \sqrt{n}\left(1 - \frac{1}{d}\right)^\ell$$

and

$$\begin{aligned}
\operatorname{var}\{f(Z_\ell)\} &= 1 + (n-1)\left(1 - \frac{2kn - 2k^2}{n^2 - n}\right)^\ell - n\left(1 - \frac{k}{n}\right)^{2\ell} \\
&= 1 + (n-1)\left(1 - \frac{2n(1-1/d)}{d(n-1)}\right)^\ell - n\left(1 - \frac{1}{d}\right)^{2\ell} \\
&\le 1 + (n-1)\left(1 - \frac{2(1-1d)}{d}\right)^\ell - n\left(1 - \frac{1}{d}\right)^{2\ell} \\
&= 1 + (n-1)\left(\frac{d-1}{d}\right)^{2\ell} - n\left(\frac{d-1}{d}\right)^{2\ell}.
\end{aligned}$$

In this case, if $\ell = \frac{1}{2}\log_{d/(d-1)}(n) - c$ then

$$\mathbb{E}\{f(Z_\ell)\} \ge \left(\frac{d}{d-1}\right)^c,$$

so if $0 < c < \frac{1}{2}\log_{d/(d-1)}(n)$, the expectation $\mathbb{E}\{f(Z_\ell)\}$ can get large while

$$\operatorname{var}\{f(Z_\ell)\} \le 2.$$

Case 2. If there is $0 < \varepsilon < 1$, so that $k = O(n^\varepsilon)$, then the mean becomes

$$\mathbb{E}\{f(Z_\ell)\} = \exp\left(c + O\left(\frac{k\log n}{n}\right) + O\left(\frac{ck}{n}\right)\right),$$

which means that, for $0 < c < \frac{1}{4}\log(n/k)$, this expectation is large. Similarly for the variance, we obtain

$$\begin{aligned}
\operatorname{var}\{f(Z_\ell)\} &= 1 + (n-1)\exp\left(c\frac{2n-2k}{n-1} - \frac{n-k}{n-1}\log n + O\left(\frac{k}{n}\log n\right) + O\left(\frac{ck}{n}\right)\right) \\
&\quad - \exp\left(2c + O\left(\frac{k\log n}{n}\right) + O\left(\frac{ck}{n}\right)\right) \\
&\le 1 + O\left(\frac{1}{n}\right) + \mathrm{e}^{2c}\left(O\left(\frac{k\log n}{n}\right) + O\left(\frac{ck}{n}\right)\right).
\end{aligned}$$

Therefore, the variance is uniformly bounded for $0 \le c < \frac{1}{4}\log(n/k)$.

https://doi.org/10.1017/apr.2017.42 Published online by Cambridge University Press

In both cases, Chebyshev's inequality yields that, for the set $A_\alpha = \{x\colon |f(x)| \le \alpha\}$, we have

$$U(A_\alpha) \ge 1 - \frac{1}{\alpha^2} \quad \text{and} \quad P^{*\ell}(A_\alpha) < \frac{B}{(\mathrm{e}^{2c} - \alpha)^2},$$

where B is uniformly bounded when $0 \le c \le \frac{1}{4}\log n$. Therefore, we have

$$\|P_{\mathrm{id}}^{*\ell} - U\|_{\mathrm{TV}} \ge 1 - \frac{1}{\alpha^2} - \frac{B}{(\mathrm{e}^{2c} - \alpha)^2}.$$

Now take $\alpha = \mathrm{e}^{2c}/2$, which completes the proof. □

5. Fourier transform arguments

In this section a different approach is introduced. It combines the representation theory of the hypercube and the Fourier transform to provide a bound for the mixing time. All of the irreducible representations of the hypercube are one-dimensional and they are indexed by $z \in (\mathbb{Z}/2\mathbb{Z})^n$ in the following way:

$$\rho_z(\boldsymbol{w}) = (-1)^{z\cdot \boldsymbol{w}},$$

where $z \cdot \boldsymbol{w}$ is the inner product of $z, \boldsymbol{w} \in (\mathbb{Z}/2\mathbb{Z})^n$. The Fourier transform of a probability $\mathbb{P}$ at a representation ρ is defined as

$$\widehat{P}(\rho) = \sum_{g\in(\mathbb{Z}/2\mathbb{Z})^n} \mathbb{P}(g)\rho(g),$$

which in our case means that

$$\widehat{P}(\rho_z) = p + (1-p)\sum_{a=0}^{j}(-1)^a \left(\binom{j}{a}\binom{n-j}{k-a}\Big/\binom{n}{k}\right) = p + (1-p)K_j^n(k),$$

where j is the number of 1s that z has and $K_j^n(k)$ is the jth Krawtchuck polynomial evaluated at k.

The upper bound lemma [6, Chapter 3, Lemma 1], deals with how to use the Fourier transform of the representations of a group to find an upper bound for the mixing time of a walk on the group. More precisely, the upper bound lemma in the case of the hypercube (or, in general, for $(\mathbb{Z}/p\mathbb{Z})^n$, can be written as follows.

Lemma 2. *(Upper bound lemma) For a random walk on the hypercube, after ℓ steps, we have*

$$4\|P_x^{*\ell} - U\|_{\mathrm{TV}}^2 \le \sum_{z\ne 0}(\widehat{P}(\rho_z))^{2\ell} \quad \textit{for all } x \in (\mathbb{Z}/2\mathbb{Z})^n.$$

Lemma 2 will be used for the $k = n/2$ case and for a the walk generated by Q as defined in (1).

6. The $k = n/2$ case

In the case where n is even with $n = 2k$, where k is a positive, odd integer, the following facts hold.

https://doi.org/10.1017/apr.2017.42 Published online by Cambridge University Press

Lemma 3. *For $k = n/2$, the Fourier transform of representation ρ_z is given by*

$$\widehat{P}(\rho_z) = \begin{cases} \frac{1}{2} & \text{if } j \text{ is odd,} \\ \frac{1}{2} + (-1)^i \binom{n/2}{i} \Big/ 2\binom{n}{2i} & \text{if } j = 2i, \end{cases}$$

where j is the number of 1*s belonging to z.*

Proof. Acccording to Koekoek and Swarttouw [12], the jth Krawtchuck polynomial satisfies the following recurrence relation:

$$-kK_j^n(k) = \frac{1}{2}(n-i)K_{j+1}^n(k) - \frac{n}{2}K_j^n(k) + \frac{n}{2}K_{j-1}^n(k). \tag{2}$$

If $k = n/2$ then (2) yields

$$K_j^n\left(\frac{n}{2}\right) = \begin{cases} 0 & \text{if } j \text{ is odd,} \\ (-1)^i \binom{n/2}{i} \Big/ \binom{n}{2i} & \text{if } j = 2i, \end{cases}$$

since $K_0^n(k) = 1$ and $K_1^n(k) = 1$. □

The next step is to bound the eigenvalues and use the upper bound lemma to actually obtain an upper bound for the L^2-norm.

Lemma 4. *For every representation ρ_z, where $z \neq \mathbf{0}$,*

$$|\widehat{P}(\rho_z)| \leq \tfrac{3}{4}.$$

Proof. Let j be the number of 1s that z has. For j odd, the theorem holds because $\widehat{P}(\rho_z) = \frac{1}{2}$. If $j = 2i$, the quantity

$$\frac{1}{2} + (-1)^i \binom{n/2}{i} \Big/ 2\binom{n}{2i}$$

is the main concern. For i odd, the second term is negative but greater than $-\frac{1}{2}$, therefore, the quantity is positive less than $\frac{1}{2}$. For i even, it turns out that

$$c_i = \binom{n/2}{i} \Big/ \binom{n}{i}$$

is maximized for $i = n/2 - 1$ and at most $1/(n-1)$ by a simple argument. Thus, for i even,

$$\frac{1}{2} + (-1)^i \binom{n/2}{i} \Big/ 2\binom{n}{2i} \leq \frac{3}{4}.$$ □

Theorem 3. *For $k = n/2$ and for $\ell = (n \log 2 - \log \varepsilon)/\log(4/3)$ with $0 < \varepsilon < 1$,*

$$4\| P_x^{*\ell} - U \|_{\mathrm{TV}}^2 < \varepsilon \quad \textit{for all } x \in (\mathbb{Z}/2\mathbb{Z})^n.$$

Proof. After having computed the Fourier transform of each representation and bounded them in Lemma 4, the upper bound lemma yields

$$4\| P_x^{*\ell} - U \|_{\mathrm{TV}}^2 \leq \sum_{j=1}^{n} \binom{n}{j} b_j^{2\ell} \leq 2^n \left(\frac{3}{4}\right)^{2\ell} \leq \varepsilon.$$ □

https://doi.org/10.1017/apr.2017.42 Published online by Cambridge University Press

Remark 2. Note that, for the ℓ^2-norm, we have

$$|G|\|P_x^{*\ell} - U\|_2^2 \geq \left(\frac{1}{2}\right)^{2\ell} \sum_{i \text{ is odd}} \binom{n}{i} = 2^{(n-1)/2-2\ell},$$

so indeed the L^2-norm cannot give a better upper bound for the mixing time. However, the L^1-norm may still be small for smaller ℓ.

7. Similar random walks

In this section we focus on similar random walks on $(\mathbb{Z}/m\mathbb{Z})^n$. We use Fourier transforms and comparison theory methods to bound the mixing times.

7.1. A random walk on $(\mathbb{Z}/m\mathbb{Z})^n$

Consider the walk generated by the measure

$$Q(a_{i_1}e_{i_1} + a_{i_2}e_{i_2} + \cdots + a_{i_k}e_{i_k}) = 1 \Big/ \binom{n}{k} m^k,$$

where $a_{i_j} \in \mathbb{Z}/m\mathbb{Z}$ and $\{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots, n\}$.

Proof of Theorem 2. Let ρ_a denote a representation of $(\mathbb{Z}/m\mathbb{Z})^n$, where $a \in (\mathbb{Z}/m\mathbb{Z})^n$. If $g = (g_1, \ldots, g_n) \in (\mathbb{Z}/m\mathbb{Z})^n$ then

$$\rho_a(g) = \exp\left(\frac{2\pi \mathrm{i} \sum_{j=1}^n a_j g_j}{m}\right).$$

Then the Fourier transform of Q at this representation with respect to Q is

$$\widehat{Q}(\rho_a) = \left(1 \Big/ \binom{n}{k} m^k\right) \sum_{|g| \leq k} \exp\left(\frac{2\pi \mathrm{i} \sum_{j=1}^n a_j g_j}{m}\right),$$

where $|g|$ denotes the number of positions that g_i is not equal to 0. Now

$$\sum_{b \in \mathbb{Z}/m\mathbb{Z}} \exp\left(\frac{2\pi \mathrm{i} a_j b}{m}\right) = m\delta_{0,a_j},$$

with $\delta_{0,a_j} = 1$ if $a_j = 0$ and $\delta_{0,a_j} = 0$ otherwise. If $n - |a| \geq k$ then

$$\widehat{Q}(\rho_a) = \binom{n-|a|}{k} \Big/ \binom{n}{k},$$

otherwise $\widehat{Q}(\rho_a) = 0$. These are the eigenvalues of the random walk that are not equal to 1.

Now note that all of the eigenvalues are nonnegative and, in particular,

$$\begin{aligned}\binom{n-|a|}{k} \Big/ \binom{n}{k} &= \left(1 - \frac{k}{n}\right)\left(1 - \frac{k}{n-1}\right) \cdots \left(1 - \frac{k}{n-|a|+1}\right) \\ &\leq \exp\left(-k \sum_{j=n-|a|+1}^{n} \frac{1}{j}\right)\end{aligned}$$

https://doi.org/10.1017/apr.2017.42 Published online by Cambridge University Press

$$\leq \exp\left(-k \log \frac{n+1}{n-|a|+1}\right)$$
$$= \left(1 - \frac{|a|}{n+1}\right)^k.$$

Then, the upper bound lemma (Lemma 2) yields

$$\begin{aligned} 4\|Q_x^{*\ell} - U\|_{\mathrm{TV}}^2 &\leq \sum_{j=1}^{n-k} \binom{n}{j} (m-1)^j \left(1 - \frac{j}{n+1}\right)^{2kl} \\ &\leq \sum_{j=1}^{n-k} \frac{n^j}{j!} (m-1)^j \left(1 - \frac{j}{n+1}\right)^{2kl} \\ &\leq e^{-c}, \end{aligned}$$

where $\ell = (n+1)/2k \log(mn) + c(n+1)/2k$. □

Remark 3. Note that the first time that all coordinates have been touched is a strong stationary time, which implies that the total variation distance needs of the order $(n/k)\log n$ steps to become small. To see this, one can imitate the calculation for the coupon collector problem as presented in Lemma 2 of [3]. Further, note that

$$|G|\|Q_x^{*\ell} - U\|_2^2 \geq n(m-1)\left(1 - \frac{1}{n}\right)^{2kl},$$

which means that the L^2-norm needs at least $(n/2k)\log(mn) + c(n+1)/2k$ steps to become small. Therefore, there is a gap between the separation distance and the L^2-norm mixing times.

7.2. Comparison theory application

Comparison theory can help provide an upper bound for the following example.

Example 1. With the notation as in Theorem 2, consider the $k = 1$ case. Then

$$Q(be_i) = \frac{1}{mn} \quad \text{for } b \in \mathbb{Z}/m\mathbb{Z} \text{ and } 1 \leq i \leq n.$$

The walk suggests to pick a coordinate at random and randomize it. Theorem 2 states that if $\ell = \frac{1}{2}((n+1)\log(mn) + c(n+1))$ then, for every $x \in \mathbb{Z}/m\mathbb{Z}$, we have

$$4\|Q_x^{*\ell} - U\|_{\mathrm{TV}}^2 \leq \mathrm{e}^{-c}.$$

Comparison theory yields the following theorem for the mixing time of the random walk generated by

$$\tilde{\mathbb{P}}(\pm e_1) = \frac{1}{4n}, \qquad \tilde{\mathbb{P}}(\mathrm{id}) = \frac{1}{2},$$

and let $\tilde{P}(x, gx) = \tilde{P}(g)$ for every $x, g \in (\mathbb{Z}/m\mathbb{Z})^n$.

Theorem 4. *Let $\ell = \frac{1}{2}m^2((n+1)\log(mn) + c(n+1)$ then, for every $x \in (\mathbb{Z}/m\mathbb{Z})^n$, we have*

$$4\|\tilde{P}_x^{*\ell} - U\|_{\mathrm{TV}}^2 \leq \left(1 + \frac{1}{n^n}\right)\mathrm{e}^{-c}.$$

https://doi.org/10.1017/apr.2017.42 Published online by Cambridge University Press

Proof. Let $S = \{\pm e_j, \mathrm{id}\}$ and $S' = \{be_j, b \in \mathbb{Z}/m\mathbb{Z}\}$. According to Diaconis and Saloff-Coste [7], if we represent each $z \in S'$ as a product of elements of S that has odd length and

$$A = \max_{s \in S} \frac{1}{\tilde{P}(s)} \sum_{z \in S'} \|z\| N(z, s) Q(z),$$

where $\|z\|$ is the length of this representation and $N(z, s)$ is the number of times s is inside the representation of z, then

$$4\|\tilde{P}_x^{*\ell} - U\|_{\mathrm{TV}}^2 \leq m^n \mathrm{e}^{-\ell/A} + m^n \|Q^{*\ell/2A} - U\|_2^2.$$

It is straightforward to show that $A \leq \max\{m^2, 2/m + 2m\} = m^2$. Therefore, if $\ell = (m^2/2)((n+1)\log(mn) + c(n+1))$,

$$4\|\tilde{P}_x^{*\ell} - U\|_{\mathrm{TV}}^2 \leq \frac{\mathrm{e}^{-c}}{n^n} + \mathrm{e}^{-c}. \qquad \square$$

Acknowledgements

The author would like to thank Persi Diaconis and Graham White for their helpful comments and suggestions.

References

[1] Aldous, D. (1983). Minimization algorithms and random walk on the d-cube. *Ann. Prob.* **11,** 403–413.
[2] Aldous, D. (1983). Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on Probability, XVII* (Lecture Notes Math. **986**), Springer, pp. 243–297.
[3] Aldous, D. and Diaconis, P. (1986). Shuffling cards and stopping times. *Amer. Math. Monthly* **93,** 333–348.
[4] Andersen, H. C. and Diaconis, P. (2007). Hit and run as a unifying device. *J. Soc. Fr. Statist. Rev. Statist. Appl.* **148,** 5–28.
[5] Bubley, R. and Dyer, M. (1997). Path coupling: a technique for proving rapid mixing in Markov chains. In *Proceedings of the 38th Annual Symposium on Foundations of Computer Science,* IEEE, pp. 223–231
[6] Diaconis, P. (1988). *Group Representations in Probability and Statistics* (Inst. Math. Statist. Lecture Notes Monogr. Ser. **11**). Institute of Mathematical Statistics, Hayward, CA.
[7] Diaconis, P. and Saloff-Coste, L. (1993). Comparison techniques for random walk on finite groups. *Ann. Prob.* **21,** 2131–2156.
[8] Diaconis, P. and Shahshahani, M. (1987). Time to reach stationarity in the Bernoulli–Laplace diffusion model. *SIAM J. Math. Anal.* **18,** 208–218.
[9] Diaconis, P., Graham, R. L. and Morrison, J. A. (1990). Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures Algorithms* **1,** 51–72.
[10] Ehrenfest, P. and Ehrenfest, T. (1907). Über zwei bekannte einwände gegen das Boltzmannsche H-theorem. *Phys. Z.* **8,** 311–314.
[11] Kac, M. (1947). Random walk and the theory of Brownian motion. *Amer. Math. Monthly* **54,** 369–391.
[12] Koekoek, R. and Swarttouw, R. F. (1996). The Askey-scheme of hypergeometric orthogonal polynomials and its q-analogue. Preprint. Available at https://arxiv.org/abs/maths/9602214.
[13] Saloff-Coste, L. (2004). Random walks on finite groups. In *Probability on Discrete Structures*. Springer, Berlin, pp. 263–346.
[14] Saloff-Coste, L. (2004). Total variation lower bounds for finite Markov chains: Wilson's lemma. In *Random Walks and Geometry*. De Gruyter, Berlin, pp. 515–532.
[15] Sinclair, A. (2009). Markov chain Monte Carlo: foundations & applications. Available at https://people.eecs.-berkeley.edu/~sinclair/cs294/f09.html.
[16] Wilson, D. B. (2003). Mixing time of the Rudvalis shuffle. *Electron. Commun. Prob.* **8,** 77–85.

https://doi.org/10.1017/apr.2017.42 Published online by Cambridge University Press