PUBLISHER CORRECTION

Correction to: Speeding up Markov chains with deterministic jumps

Sourav Chatterjee[1] [ORCID] **· Persi Diaconis[1]**

In memory of Harry Kesten

Correction to: Probability Theory and Related Fields (2020) 178:1193–1214
https://doi.org/10.1007/s00440-020-01006-4

This article was inadvertently published in Volume 178(3–4) December 2020, https://doi.org/10.1007/s00440-020-01006-4, when it should have been published in the special issue honouring Professor Harry Kesten.

Springer Nature apologizes for this inconvenience.

The original article can be found online at https://doi.org/10.1007/s00440-020-01006-4.

✉ Sourav Chatterjee
   souravc@stanford.edu

   Persi Diaconis
   diaconis@math.stanford.edu

[1]   Departments of Mathematics and Statistics, Stanford University, Stanford, CA, USA

Abstract

We show that the convergence of finite state space Markov chains to stationarity can often be considerably speeded up by alternating every step of the chain with a deterministic move. Under fairly general conditions, we show that not only do such schemes exist, they are numerous.

1 Introduction

This paper started from the following example. Consider the simple random walk on $\mathbb{Z}_n$ (the integers mod n):

$$X_{k+1} = X_k + \epsilon_{k+1} \pmod{n},$$

with $X_0 = 0$, and $\epsilon_1, \epsilon_2, \ldots$ i.i.d. with equal probabilities of being 0, 1 or -1. This walk takes order n^2 steps to reach its uniform stationary distribution in total variation distance. This slow, diffusive behavior is typical of low dimensional Markov chains ("Diameter^2 steps are necessary and sufficient"—see Diaconis and Saloff-Coste [19] for careful statements and many examples). Now change the random walk by deterministic doubling:

$$X_{k+1} = 2X_k + \epsilon_{k+1} \pmod{n}.$$

This walk has the "same amount of randomness". Results discussed below show that it takes order $\log n$ steps to mix (at least for almost all n).

We would like to understand this speedup more abstractly—hopefully to be able to speed up real world Markov chains. Our main result shows a similar speedup occurs for fairly general Markov chains with deterministic doubling replaced by almost any bijection of the state space into itself. We proceed to a careful statement in the next section. In Sect. 3 a literature review offers pointers to related efforts to beat diffusive behavior. A sketch of the proof is in Sect. 4. Proofs are in Sects. 5 and 6. A different kind of deterministic speedup,

$$X_{k+1} = X_k + X_{k-1} + \epsilon_{k+1} \pmod{n},$$

is analyzed in Sect. 7. The last section has applications and some interesting open questions.

2 Speedup by deterministic functions

Let S be a finite set and $P = (p_{ij})_{i,j \in S}$ be a Markov transition matrix on S. We assume the following conditions on P:

(1) The Markov chain generated by P is irreducible and aperiodic.
(2) If $p_{ij} > 0$ for some $i, j \in S$, then $p_{ji} > 0$.
(3) For each $i \in S$, $p_{ii} > 0$.
(4) The uniform distribution on S is the stationary measure of P. We will call it μ.

In addition to the above assumptions, we define

$$n := |S|$$

and

$$\delta := \min\{p_{ij} : i, j \in S, p_{ij} > 0\}. \tag{2.1}$$

Let $f : S \to S$ be a bijection. Consider a new Markov chain defined as follows. Apply f, and then take one step according to P. This is one step of the new chain. The transition matrix for the new chain is

$$Q = \Pi P,$$

where $\Pi = (\pi_{ij})_{i,j \in S}$ is the permutation matrix defined by f. (That is, $\pi_{ij} = 1$ if $j = f(i)$ and 0 otherwise.) Clearly, μ is a stationary measure for the new chain too. The main questions that we will try to answer are the following: (a) Under what conditions on f does this new chain mix quickly—let's say, in time $\log n$? (b) Is it always possible to find f satisfying these conditions? (c) If so, are such functions rare or commonplace?

Our first theorem answers question (a), by giving a sufficient condition for fast mixing of Q.

Theorem 2.1 *For any $A \subseteq S$, let $\mathcal{E}(A)$ be the set of $j \in S$ such that $p_{ij} > 0$ for some $i \in A$. Suppose that there is some $\epsilon \in (0, 1)$ such that for any A with $|A| \leq n/2$,*

$$|\mathcal{E} \circ f \circ \mathcal{E}(A)| \geq (1 + \epsilon)|A|. \tag{2.2}$$

Let $X_0, X_1, \ldots$ be a Markov chain with transition matrix Q. For each $i \in S$ and $k \geq 1$ let μ_i^k be the law of X_k given $X_0 = i$. Then

$$\|\mu_i^k - \mu\|_{TV} \leq \frac{\sqrt{n}}{2}\left(1 - \frac{\epsilon^2 \delta^8}{2}\right)^{(k-2)/4},$$

where δ is the quantity defined in (2.1) *and TV stands for total variation distance.*

This result shows that if we have a sequence of problems where $n \to \infty$ but ϵ and δ remain fixed, and the condition (2.2) is satisfied, then the mixing time of the Q-chain is of order $\log n$.

Let us now try to understand the condition (2.2). The set $\mathcal{E}(A)$ is the 'one-step expansion' of A, that is, the set of all states that are accessible in one step by the P-chain if it starts in A. By the condition that $p_{ii} > 0$ for all i, we see that $\mathcal{E}(A)$ is a superset of A. The condition (2.2) says that for any A of size $\leq n/2$, if we apply the one-step expansion, then apply f and then again apply the one-step expansion, the resulting set must be larger than A by a fixed factor.

We found it difficult to produce explicit examples of functions that satisfy (2.2) with ϵ independent of n. Surprisingly, the following theorem shows that they are extremely abundant, even in our general setting, thereby answering both the questions (b) and (c) posed earlier.

Theorem 2.2 *Let all notation be as in Theorem* 2.1. *There are universal constants* $C_1, C_2 > 0$ *and* $C_3 \in (0, 1)$ *such that all but a fraction* $C_1 n^{-C_2/\delta}$ *of bijections f satisfy* (2.2) *with* $\epsilon = C_3\delta$. *Consequently, for all but a fraction* $C_1 n^{-C_2/\delta}$ *of bijections f, the corresponding Q-chain satisfies, for all* $i \in S$ *and* $k \geq 1$,

$$\|\mu_i^k - \mu\|_{TV} \leq \frac{\sqrt{n}}{2}\left(1 - \frac{C_3^2\delta^{10}}{2}\right)^{(k-2)/4}.$$

For instance, if the P-chain is the lazy random walk on $\mathbb{Z}_n$ with equal probabilities of taking one step to the left, or one step to the right, or not moving, then $\delta = 1/3$. So in this example this result shows that for all but a fraction $C_1 n^{-3C_2}$ of bijections of $\mathbb{Z}_n$, the corresponding Q-chain mixes in time $O(\log n)$. This is a striking improvement over the lazy random walk, which mixes in time $O(n^2)$.

Interestingly, the doubling random walk discussed in Sect. 1 *does not* satisfy the expansion condition (2.2) with ϵ independent of n. To see this, take $n = 4m - 1$ for large enough m, and define

$$A = \{1, 2, \ldots, m-1\} \cup \{2m+1, \ldots, 3m-1\}.$$

Then

$$\mathcal{E}(A) = \{0, 1, \ldots, m\} \cup \{2m, \ldots, 3m\}.$$

So if f is the doubling map, an easy calculation shows that

$$f \circ \mathcal{E}(A) = \{0, 1, 2, \ldots, 2m+1\},$$

and therefore

$$\mathcal{E} \circ f \circ \mathcal{E}(A) = \{0, 1, \ldots, 2m+2, n-1\}.$$

So we have $|A| = 2m - 2 \leq n/2$ and $|\mathcal{E} \circ f \circ \mathcal{E}(A)| = 2m + 4$, which means that the condition (2.2) cannot hold with $\epsilon > 6/(2m - 2)$. Therefore the speedup of the doubling walk is *not* happening due to expansion, although as Theorem 2.2 shows, expansion is the reason for speedup for "most" bijections.

3 Related previous work

The deterministic doubling example comes from work of Chung et al. [9]. They found that for almost all n, $1.02 \log_2 n$ steps suffice. Hildebrand [32] showed that $1.004 \log_2 n$ steps are not enough for mixing. In a recent tour-de-force, Eberhard and Varjú [23] show that for almost all n, there is a cutoff at $c \log_2 n$ where

$c = 1.01136\ldots$. Two companion papers by Breuillard and Varjú [7, 8] relate this problem to irreducibility of random polynomials and very deep number theory. All of these authors consider $X_{k+1} = aX_k + \epsilon_{k+1} \pmod{n}$ for general a and ϵ_i's having more general distributions. Related work is in Hildebrand [32] and the Ph.D. thesis of Neville [38]. One natural extension is that if G is a finite group with $\epsilon_1, \epsilon_2, \ldots$ i.i.d. from a probability μ on G, the random walk $X_{k+1} = X_k \epsilon_{k+1}$ may be changed to $X_{k+1} = A(X_k)\epsilon_{k+1}$ where A is an automorphism of G. This is studied in Diaconis and Graham [15, 16] who treat $G = \mathbb{Z}_p^n$ with a matrix as automorphism. Similar speed-ups occur.

A very rough heuristic to explain the speedup in our Theorem 2.2 goes as follows. Composing the original random walk step with a random bijection mimics a jump along the edge of a random graph. Since random graphs are usually expanders, and random walks on expanders with n vertices mix in $O(\log n)$ steps, we may naturally expect something like Theorem 2.2 to be true. Remarkable work has been done on rates of convergence of random walks on d-regular expanders. For instance, Lubetzky and Sly [35] show that random walks on d-regular random graphs have a cutoff for total variation mixing at $\frac{d}{d-2} \log_{d-1} n$. Lubetzky and Peres [34] prove a similar result for Ramanujan graphs, with cutoff at $\frac{d}{d-1} \log_{d-1} n$. Related work has recently appeared in Hermon et al. [30], who show that if you begin with any bounded degree graph on $2n$ vertices (for which all connected components have size at least 3) and add a random perfect matching, the resulting walk has a cutoff and mixes in time $O(\log n)$.

A setup very similar to ours, with different results, has been considered recently by Bordenave et al. [6]. This paper studies a random bistochastic matrix of the form ΠP, where Π is a uniformly distributed permutation matrix and P is a given bistochastic matrix. Under sparsity and regularity assumptions on P, the authors prove that the second largest eigenvalue of ΠP is essentially bounded by the normalized Hilbert–Schmidt norm of P. They apply the result to random walks on random regular digraphs.

A different, related, literature treats piecewise deterministic Markov processes. Here, one moves deterministically for a while, then at a random time (depending on the path) a random big jump is made. This models the behavior of biological systems and chemical reactions. A splendid overview is in Malrieu [36] (see also Benaïm [3]). There is a wide swath of related work in a non-stochastic setting where various extra "stirring" processes are shown to speed up mixing and defeat diffusive behavior. See Constantin et al. [11], Ottino [39] and Rallabandi et al. [41]. The papers of Ding and Peres [22], Hermon [28] and Hermon and Peres [29] are Markov chain versions of these last results.

There are a host of other procedures that aim at defeating diffusive behavior. For completeness, we highlight four:

- Event chain Monte Carlo.
- Hit and run.
- Lifting/non-reversible variants.
- Adding more moves.

Each of these is in active development; searching for citations to the literature below should bring the reader up to date.

While "random walk" type algorithms are easy to implement, it is by now well known that they suffer from "diameter^2" behavior. To some extent that remains true for general reversible Markov chains. Indeed, Neal [37] showed that the spectral gap of *any* reversible Markov chain can be improved by desymmetrizing. Of course, the spectral gap is a crude measure of convergence. A host of sharpenings are in Diaconis and Miclo [18].

Non-reversible versions of Hamiltonian Monte Carlo introduced in Diaconis et al. [17] have morphed into a variety of "lifting algorithms" culminating in the event chain algorithm of Kapfer and Krauth [33]. This revolutionized the basic problem of sampling hard disk configurations in statistical mechanics. The hard disk problem was the inspiration for three basic tools: The Metropolis algorithm, the Gibbs sampler (Glauber dynamics) and molecular dynamics. All gave wrong conclusions for the problems of interest; the new algorithms showed that entirely new pictures of phase transitions are needed. It is an important open problem to generalize these algorithms to more general stationary distributions.

Recent, closely related work is in Bordenave and Lacoin [5] and Conchon–Kerjan [10]. Written independently, both papers study 'lifting' random walks to covering graphs to speed up mixing and both prove that most liftings have cutoffs. The first paper also studies non-reversible chains. The paper of Gerencsér and Hendrickx [25] adds 'long distance edges' (as in small world graphs) and non-reversibility to get speedups. It too has close connections with the present project.

A final theme is "hit and run": Taking long steps in a chosen direction instead of just "going ± 1". For a survey, see Anderson and Diaconis [1]. It has been difficult to give sharp running time estimates for these algorithms (even though they are "obviously better"). For a recent success, see the forthcoming manuscript of Boardman and Saloff-Coste [4].

It is natural to try to speed up convergence by adding more generators. This is a largely open problem. For example, consider the simple random walk on the hypercube $\mathbb{Z}_2^d$. The usual generating set is $\{0, e_1, \ldots, e_d\}$, where e_i is the ith standard basis vector. At each step, one of these vectors is added with probability $1/(d+1)$. It is well known that $\frac{1}{4}d \log d + O(d)$ steps are necessary and sufficient for mixing (Diaconis and Shahshahani [20]). Suppose you want to add more generators (say $2d + 1$ generators)—what should you choose? Wilson [42] studied this problem and showed that almost all generating sets of size $2d + 1$ mix in time $0.2458d$ steps. Mirroring our results above, his methods do not give an explicit choice effecting this improvement.

These problems are interesting, and largely open, even for the cyclic group $\mathbb{Z}_p$ when p is a prime. As above, the generating set $\{0, \pm 1\}$ take order p^2 steps to mix. On the other hand, almost all sets of 3 generators mix in order p steps. Greenhalgh [26] found 3 generators (roughly, 1, $p^{1/3}$ and $p^{2/3}$) that do the job. References to this literature can be found in a survey by Hildebrand [31].

There has been vigorous work on adding and deleting random edges at each stage (dynamical configuration model). See the paper of Avena et al. [2], which also has pointers to another well studied potential speedup—the non-backtracking random walk.

Finally we note that natural attempts at speeding things up may fail. Cutting cards between shuffles and systematic versus random scan in the Gibbs sampler are examples. For details, see Diaconis [13].

4 Proof sketch

For the proof of Theorem 2.1, we define an auxiliary chain with kernel $R = L^2(L^T)^2$, where $L := P\Pi$. This kernel is symmetric and positive semidefinite, and defines a reversible Markov chain. The proof has two steps—first, relate the rate of convergence of the R-chain with that of the Q-chain, and second, use the reversibility of the R-chain to invoke Cheeger's bound for the spectral gap. A lower bound on the Cheeger constant is then obtained using condition (2.2).

The proof of Theorem 2.2 has its root in a beautiful idea from a paper of Ganguly and Peres [24] (see also the related earlier work of Pymar and Sousi [40]). In our context, the idea translates to showing the following: Let f be a uniform random bijection. First, show that "most" sets A of size $\leq 3n/4$ have the property that $|\mathcal{E}(A)| \geq (1 + \epsilon)|A|$ for some suitable fixed $\epsilon > 0$. Since $|\mathcal{E}(B)| \geq |B|$ and $|f(B)| = |B|$ for all B, this implies that (2.2) is satisfied for "most" A of size $\leq n/2$. But we want (2.2) to hold for *all* A of size $\leq n/2$. The Ganguly–Peres idea is to circumvent this difficulty in the following ingenious way. Since we know that $|\mathcal{E}(B)| < (1 + \epsilon)|B|$ for only a small number of B's of size $\leq 3n/4$, it is very unlikely that for any given set B, $f(B)$ has this property. Consequently, for a given set A such that $|\mathcal{E}(A)| \leq 3n/4$, it is very unlikely that $f(\mathcal{E}(A))$ has this property (taking $B = \mathcal{E}(A)$). Thus, for any such set A, it is very likely that

$$|\mathcal{E} \circ f \circ \mathcal{E}(A)| \geq (1 + \epsilon)|f(\mathcal{E}(A))| \geq (1 + \epsilon)|A|.$$

Since the above event is very likely for any given A, it remains very likely if we take the intersection of such events over a suitably small set of A's. So we now take this small set to be precisely the set of A's that did not satisfy $|\mathcal{E}(A)| \geq (1 + \epsilon)|A|$— that is, our rogue set that spoiled (2.2). By some calculations, we can show that this indeed happens, and so (2.2) holds for all A with $|A| \leq n/2$.

In [24], the Ganguly–Peres idea was applied to solve a different problem for a random walk on points arranged in a line. As noted above, the key step is to show that the set of all A of size $\leq 3n/4$ that violate $|\mathcal{E}(A)| \geq (1 + \epsilon)|A|$ is a small set. Since we are working in a general setup here, the proof of this step from [24] does not generalize; it is proved here by a new method.

5 Proof of Theorem 2.1

Let $L := P\Pi$. Since the uniform distribution on S is the stationary measure of P, we deduce that P is a doubly stochastic matrix. Since Π is a permutation matrix, it is automatically doubly stochastic. Since the product of doubly stochastic matrices is doubly stochastic, we see that L is also doubly stochastic. Thus, L and L^T are both

stochastic matrices, and so is any power of either of these matrices. Consequently, the matrix

$$R := L^2(L^T)^2 = P\Pi P\Pi\Pi^T P^T \Pi^T P^T = P\Pi PP^T \Pi^T P^T.$$

is a stochastic matrix. (This identity explains why we symmetrize L^2 instead of L. Indeed, $LL^T = P\Pi\Pi^T P^T = PP^T$, and the effect of Π disappears.)

Lemma 5.1 *The Markov chain with transition matrix R is irreducible, aperiodic and reversible, with uniform stationary distribution. Moreover, R is symmetric and positive semidefinite.*

Proof It is evident from the definition of R that R is symmetric and positive semidefinite. Since R is symmetric, the Markov chain with transition matrix R is reversible with respect to the uniform distribution on the state space.

Since $p_{ii} > 0$ for any $i \in S$, it follows that

$$r_{ii} \geq p_{ii}\pi_{if(i)}p_{f(i)f(i)}p_{f(i)f(i)}\pi_{if(i)}p_{ii} > 0. \tag{5.1}$$

Thus, R is aperiodic. Next, note that if $p_{ij} > 0$, then

$$r_{ij} \geq p_{ij}\pi_{if(j)}p_{f(j)f(j)}p_{f(j)f(j)}\pi_{if(j)}p_{jj} > 0. \tag{5.2}$$

Since P is irreducible, this shows that R is also irreducible. □

Corollary 5.2 *The principal eigenvalue of R is 1, the principal eigenvector is the vector of all 1's (which we will henceforth denote by 1), and the principal eigenvalue has multiplicity one. The second largest eigenvalue of R (which we will henceforth call λ_2), is strictly less than 1. All eigenvalues of R are nonnegative.*

Proof Since the Markov chain defined by R is irreducible and aperiodic (on a finite state space), and R is symmetric, all but the last claim follow from the general theory of Markov chains. The eigenvalues are nonnegative because R is a positive semidefinite matrix. □

The information about R collected in the above corollary allows us to prove the next lemma.

Lemma 5.3 *For any $k \geq 1$, and any $x \in \mathbb{R}^n$ that is orthogonal to the vector 1,*

$$\|L^k x\| \leq \lambda_2^{(k-1)/4}\|x\|.$$

Proof Take any vector $x \in \mathbb{R}^n$ that is orthogonal to 1. By the properties of R given in Corollary 5.2, it is easy to see that

$$x^T R x \leq \lambda_2 \|x\|^2.$$

But notice that

$$x^T R x = x^T L^2 (L^T)^2 x = \|(L^T)^2 x\|^2.$$

So for any x that is orthogonal to 1,

$$\|(L^T)^2 x\|^2 \le \lambda_2 \|x\|^2. \tag{5.3}$$

But if x is orthogonal to 1, then the stochasticity of L implies that

$$1^T (L^T)^2 x = 1^T x = 0.$$

Thus, $(L^T)^2 x$ is also orthogonal to 1. So we can apply (5.3) iteratively to get

$$\|(L^T)^{2k} x\|^2 \le \lambda_2^k \|x\|^2$$

for any positive integer k. But L^T, being a stochastic matrix, is an ℓ^2-contraction. Therefore

$$\|(L^T)^{2k+1} x\|^2 \le \|(L^T)^{2k} x\|^2 \le \lambda_2^k \|x\|^2.$$

This completes the proof of the lemma. $\qquad\square$

Lemma 5.3 yields the following corollary, which relates the rate of convergence of the Q-chain with the spectral gap of the R-chain.

Corollary 5.4 *For any $k \ge 2$, and any $x \in \mathbb{R}^n$ that is orthogonal to the vector 1,*

$$\|Q^k x\| \le \lambda_2^{(k-2)/4} \|x\|.$$

Proof Note that $Q^k = (\Pi P)^k = \Pi L^{k-1} P$. Being stochastic matrices, Π and P are both ℓ^2-contractions. Moreover, if $1^T x = 0$, then $1^T P x = 1^T x = 0$ since P is doubly stochastic. Therefore by Lemma 5.3,

$$\|Q^k x\| = \|\Pi L^{k-1} P x\| \le \|L^{k-1} P x\|$$
$$\le \lambda_2^{(k-2)/4} \|P x\| \le \lambda_2^{(k-2)/4} \|x\|.$$

This completes the proof of the corollary. $\qquad\square$

Lastly, we need the following upper bound on λ_2.

Lemma 5.5 *Let λ_2 be the second largest eigenvalue of R. Then*

$$\lambda_2 \le 1 - \frac{\epsilon^2 \delta^8}{2},$$

where ϵ is the constant from condition (2.2).

Proof Define a probability measure ν on $S \times S$ as

$$v(i,j) = \mu(i)r_{ij} = \frac{r_{ij}}{n}.$$

Recall that the Cheeger constant for the Markov chain with transition matrix R is

$$\Phi = \min_{A \subseteq S, \, \mu(A) \leq 1/2} \frac{v(A \times A^c)}{\mu(A)},$$

where $A^c := S \setminus A$. Since μ is the uniform distribution on V, this simplifies to

$$\Phi = \min_{A \subseteq S, \, |A| \leq n/2} \frac{1}{|A|} \sum_{i \in A, j \in A^c} r_{ij}.$$

Now, if $r_{ij} > 0$, then from the inequality (5.2) and the assumption that $p_{kl} \geq \delta$ for all nonzero p_{kl}, we get $r_{ij} \geq \delta^4$. Therefore, if A' is the set of all vertices that are attainable in one step from some $i \in A$ by the Markov chain with transition matrix R, then

$$\sum_{i \in A, j \in A^c} r_{ij} \geq \delta^4 |A' \cap A^c| = \delta^4(|A'| - |A|), \tag{5.4}$$

where the last identity holds because $A \subseteq A'$ (by (5.1)). Now, since $p_{ij} > 0$ if and only if $p_{ji} > 0$, the set A' defined above can be written as

$$A' = \mathcal{E} \circ f^{-1} \circ \mathcal{E} \circ \mathcal{E} \circ f \circ \mathcal{E}(A).$$

Application of $\mathcal{E}$ cannot decrease the size of a set, and application of f^{-1} does not alter the size. Therefore by condition (2.2),

$$|A'| \geq |\mathcal{E} \circ f \circ \mathcal{E}(A)| \geq (1 + \epsilon)|A|,$$

provided that $|A| \leq n/2$. Plugging this into (5.4), we get $\Phi \geq \epsilon \delta^4$. By the well-known bound

$$\lambda_2 \leq 1 - \frac{\Phi^2}{2},$$

this completes the proof of the lemma.   □

We are now ready to complete the proof of Theorem 2.1

Proof of Theorem 2.1 Now take any $i \in S$ and $k \geq 2$. Let $x \in \mathbb{R}^n$ be the vector whose i^{th} component is $1 - 1/n$ and the other components are all equal to $-1/n$. Then it is easy to see that x is orthogonal to 1, and that $(Q^T)^k x$ is the vector whose j^{th} component is $\mu_i^k(j) - \mu(j)$. Thus by Corollary 5.4,

$$\|\mu_i^k - \mu\|_{TV} = \frac{1}{2} \sum_{j \in S} |\mu_i^k(j) - \mu(j)|$$

$$\leq \frac{1}{2} \left(n \sum_{j \in S} (\mu_i^k(j) - \mu(j))^2 \right)^{1/2}$$

$$= \frac{\sqrt{n}}{2} \|(Q^T)^k x\| \leq \frac{\sqrt{n}}{2} \lambda_2^{(k-2)/4} \|x\|.$$

Since $\|x\| \leq 1$, this proves that

$$\|\mu_i^k - \mu\|_{TV} \leq \frac{\sqrt{n}}{2} \lambda_2^{(k-2)/4}. \tag{5.5}$$

Using the bound on λ_2 from Lemma 5.5, this completes the proof when $k \geq 2$. When $k = 1$, the bound is greater than 1 and therefore holds trivially. $\square$

6 Proof of Theorem 2.2

For this proof, it will be convenient to define a graph structure on S. Join two distinct points (vertices) $i, j \in S$ by an edge whenever $p_{ij} > 0$. Since $p_{ij} > 0$ if and only if $p_{ji} > 0$, this defines an undirected graph on S. The irreducibility of the P-chain implies that this graph is connected. A *path* in this graph is simply a sequence of vertices such that each is connected by an edge to the one following it.

Define the *external boundary* of a set $A \subseteq S$ as

$$\partial A := \mathcal{E}(A) \backslash A.$$

The following lemma proves a crucial property of the graph defined above.

Lemma 6.1 *Take any $A \subseteq S$. Then the number of sets B such that $\partial B = A$ is at most $2^{|A|/\delta}$.*

The proof of this lemma is divided into several steps. Fix a set A. Define an equivalence relation on A^c as follows. Say that two vertices $i, j \in A^c$ are equivalent, and write $i \sim j$, if either $i = j$, or there is a path from i to j that avoids the set A (meaning that there is no vertex in the path that belongs to A). It is easy to see that this is an equivalence relation, because it is obviously reflexive and symmetric, and if $i \sim j$ and $j \sim k$, then there is a path from i to k that avoids A. To prove Lemma 6.1, we need to prove two facts about this equivalence relation.

Lemma 6.2 *If B is a set of vertices such that $\partial B = A$, then B must be the union of some equivalence classes of the equivalence relation defined above.*

Proof Take any B such that $\partial B = A$. Take any $i \in B$ and $j \in A^c$ such that $i \sim j$. To prove the lemma, we have to show that $j \in B$.

Let m be the length of the shortest path from i to j that avoids A. Since $i \sim j$, we know that there is at least one such path, and therefore m is well-defined. We will now prove the claim that $j \in B$ by induction on m. If $m = 1$, then j is a neighbor of i. Since $i \in B$, j is a neighbor of i and $j \notin \partial B$, the only possibility is that $j \in B$. This proves the claim when $m = 1$.

Let us now assume that the claim holds for all $m' < m$. Take a path of length m from i to j that avoids A. Let k be the vertex on this path that comes immediately before j. Then $k \in A^c$, and k can be reached from i by a path of length $m - 1$ that avoids A. Therefore by the induction hypothesis, $k \in B$. Applying the case $m = 1$ to k, we see that j must be a member of B. This completes the induction step. $\qquad\square$

Lemma 6.3 *If D is any equivalence class of the equivalence relation defined above, then there is at least one element of D that is adjacent to some element of A.*

Proof Take any $i \in D$ and $j \in A$. Since the graph is connected, there is a path from i to j. Along this path, let k be the first vertex that belongs to A. (This is well-defined because at least one vertex of the path, namely j, is in A.) Let l be the vertex that immediately precedes k in the path. Then we see that there is a path from i to l that avoids A, and so $i \sim l$. Therefore $l \in D$ and l is adjacent to an element of A, which proves the lemma. $\qquad\square$

Lastly, we need an upper bound on the maximum degree of our graph on S.

Lemma 6.4 *The maximum degree of the graph defined on S is at most $1/\delta$.*

Proof Since $p_{ij} \geq \delta$ for every i and j that are connected by an edge, and

$$\sum_{j \in S} p_{ij} = 1$$

because P is a stochastic matrix, the number of neighbors of i must be $\leq 1/\delta$. $\qquad\square$

We are now ready to prove Lemma 6.1.

Proof of Lemma 6.1 Since each equivalence class of our equivalence relation has at least one element that is adjacent to an element of A (by Lemma 6.3), the number of equivalence classes can be at most the size of ∂A. But by Lemma 6.4, $|\partial A| \leq |A|/\delta$. Thus, there are at most $|A|/\delta$ equivalence classes.

But by Lemma 6.2, if B is any set such that $\partial B = A$, then B must be a union of equivalence classes. Consequently, the number of such B is at most the size of the power set of the set of equivalence classes. By the previous paragraph, this is bounded above by $2^{|A|/\delta}$. $\qquad\square$

Lemma 6.1 has the following corollary, which shows that not too many sets can have a small external boundary.

Corollary 6.5 *Take any* $1 \le k \le n$*. The number of sets* $B \subseteq S$ *such that* $|\partial B| = k$ *is at most* $\binom{n}{k} 2^{k/\delta}$*.*

Proof We can choose the external boundary of B in at most $\binom{n}{k}$ ways. Having chosen the external boundary, Lemma 6.1 says that B can be chosen in at most $2^{k/\delta}$ ways. Thus, the number of ways of choosing B with the given constraint is at most $\binom{n}{k} 2^{k/\delta}$. $\qquad\square$

For each $\epsilon > 0$, let $\mathcal{A}_\epsilon$ be the set of all $A \subseteq S$ such that $|\mathcal{E}(A)| < (1 + \epsilon)|A|$. For each $1 \le m \le n$, let $\mathcal{A}_{\epsilon,m}$ be the set of all sets $A \in \mathcal{A}_\epsilon$ that are of size m. Let

$$\mathcal{A}'_{\epsilon,m} := \bigcup_{1 \le k \le m} \mathcal{A}_{\epsilon,k}.$$

The following lemma uses the bound from Corollary 6.5 to get a bound on the size of $\mathcal{A}'_{\epsilon,m}$.

Lemma 6.6 *For any* $\epsilon > 0$ *and* $1 \le m \le n$*,*

$$|\mathcal{A}'_{\epsilon,m}| \le \sum_{1 \le k < \epsilon m} \binom{n}{k} 2^{k/\delta}.$$

Proof Take any $A \in \mathcal{A}'_{\epsilon,m}$. Then note that

$$|\partial A| = |\mathcal{E}(A)| - |A| < \epsilon|A| \le \epsilon m.$$

The claim is now proved by Corollary 6.5. $\qquad\square$

Now fix some $\epsilon \in (0, 1/2)$. Let f be a *random* bijection chosen uniformly from the set of all bijections of S. Define an event

$$E := \{f \circ \mathcal{E}(A) \in \mathcal{A}_\epsilon \text{ for some } A \in \mathcal{A}_\epsilon \text{ with } 1/\epsilon \le |A| \le n/2\}.$$

We want to show that E is a rare event. The following lemma is the first step towards that.

Lemma 6.7 *Let E be the event defined above. Then*

$$\mathbb{P}(E) \le \sum_{1/\epsilon \le m \le 3n/4} \frac{1}{\binom{n}{m}} \left(\sum_{1 \le k < \epsilon m} \binom{n}{k} 2^{k/\delta} \right)^2.$$

Proof Take any $A \in \mathcal{A}_\epsilon$ with $1/\epsilon \le |A| \le n/2$. Let $B = \mathcal{E}(A)$. Then observe the following about B:

- Since f is a uniform random bijection, the random set $f(B)$ is uniformly distributed among all subsets of size $|B|$.
- Since $A \in \mathcal{A}_\epsilon$, $\epsilon \in (0, 1/2)$, and $1/\epsilon \le |A| \le n/2$,

$$|B| < (1 + \epsilon)|A| < \frac{3}{2}|A| \le \frac{3n}{4},$$

and $|B| \ge |A| \ge 1/\epsilon$.

These two observations imply that

$$\mathbb{P}(E) \le \sum_{1/\epsilon \le m \le 3n/4} \sum_{\substack{A \in \mathcal{A}_\epsilon, \\ |\mathcal{E}(A)| = m}} \mathbb{P}(f(\mathcal{E}(A)) \in \mathcal{A}_\epsilon)$$

$$= \sum_{1/\epsilon \le m \le 3n/4} \sum_{\substack{A \in \mathcal{A}_\epsilon, \\ |\mathcal{E}(A)| = m}} \frac{|\mathcal{A}_{\epsilon,m}|}{\binom{n}{m}}$$

$$\le \sum_{1/\epsilon \le m \le 3n/4} \frac{|\mathcal{A}'_{\epsilon,m}||\mathcal{A}_{\epsilon,m}|}{\binom{n}{m}}.$$

Since $\mathcal{A}'_{\epsilon,m}$ is a superset of $\mathcal{A}_{\epsilon,m}$, plugging in the bound from Lemma 6.6 completes the proof. □

The bound in Lemma 6.7 is not straightforwardly understandable. The following lemma clarifies the matter.

Lemma 6.8 *There are universal constants C_0, C_1, C_2 and C_3 such that if $\epsilon \le C_0\delta$, then*

$$\sum_{1/\epsilon \le m \le 3n/4} \frac{1}{\binom{n}{m}} \left(\sum_{1 \le k < \epsilon m} \binom{n}{k} 2^{k/\delta} \right)^2 \le C_1 e^{-C_2\sqrt{n}} + C_3^{1/\epsilon} n^{2-1/4\epsilon}.$$

Proof Recall the well-known inequalities

$$\frac{n^k}{k^k} \le \binom{n}{k} \le \frac{e^k n^k}{k^k}.$$

Also, check that the map $x \mapsto (en/x)^x$ is increasing in $[1, n]$. Thus,

$$\sum_{1/\epsilon \leq m \leq 3n/4} \frac{1}{\binom{n}{m}} \left(\sum_{k < \epsilon m} \binom{n}{k} 2^{k/\delta} \right)^2$$

$$\leq \sum_{1/\epsilon \leq m \leq 3n/4} \frac{m^m}{n^m} n^2 \left(\frac{en2^{1/\delta}}{\epsilon m} \right)^{2\epsilon m}$$

$$= n^2 \sum_{1/\epsilon \leq m \leq 3n/4} \left(\frac{m^{1-2\epsilon}}{n^{1-2\epsilon}} 2^{2\epsilon/\delta} e^{2\epsilon(1+\log(1/\epsilon))} \right)^m.$$

Now choose ϵ so small that $1 - 2\epsilon \geq 1/2$ and

$$2^{2\epsilon/\delta} e^{2\epsilon(1+\log(1/\epsilon))} < \sqrt{\frac{8}{7}}.$$

Note that this can be ensured by choosing ϵ to be less than some universal constant times δ (noting that $\delta \leq 1$). With such a choice of ϵ, we get

$$n^2 \sum_{\sqrt{n} \leq m \leq 3n/4} \left(\frac{m^{1-2\epsilon}}{n^{1-2\epsilon}} 2^{2\epsilon/\delta} e^{2\epsilon(1+\log(1/\epsilon))} \right)^m$$

$$\leq n^2 \sum_{\sqrt{n} \leq m \leq 3n/4} \left(\sqrt{\frac{3}{4}} \sqrt{\frac{8}{7}} \right)^m$$

$$= n^2 \sum_{\sqrt{n} \leq m \leq 3n/4} \left(\frac{6}{7} \right)^{m/2} \leq C_1 e^{-C_2 \sqrt{n}},$$

where C_1 and C_2 are universal constants. On the other hand, if $1/\epsilon \leq m \leq \sqrt{n}$, then

$$\frac{m^{1-2\epsilon}}{n^{1-2\epsilon}} \leq n^{-1/4},$$

and so

$$n^2 \sum_{1/\epsilon \leq m \leq \sqrt{n}} \left(\frac{m^{1-2\epsilon}}{n^{1-2\epsilon}} 2^{2\Delta\epsilon} e^{2\epsilon(1+\log(1/\epsilon))} \right)^m$$

$$\leq n^2 \sum_{1/\epsilon \leq m \leq \sqrt{n}} \left(n^{-1/4} \sqrt{\frac{8}{7}} \right)^m \leq C_3^{1/\epsilon} n^{2-1/4\epsilon}.$$

Adding up the two parts, we get the required bound. $\square$

We are now ready to finish the proof of Theorem 2.2.

Proof By Lemmas 6.7 and 6.8, we see that there are universal constants $C_4, C_5 > 0$ and $C_6 \in (0, 1)$ such that if we choose $\epsilon = C_6 \delta$, then

$$\mathbb{P}(E) \le C_4 n^{-C_5/\delta}. \tag{6.1}$$

Suppose that E does not happen. Then for any $A \in \mathcal{A}_\epsilon$ with $1/\epsilon \le |A| \le n/2$, we have $f(\mathcal{E}(A)) \notin \mathcal{A}_\epsilon$ and hence

$$|\mathcal{E} \circ f \circ \mathcal{E}(A)| \ge (1 + \epsilon)|f \circ \mathcal{E}(A)| \ge (1 + \epsilon)|A|.$$

On the other hand, if $A \notin \mathcal{A}_\epsilon$, then

$$|\mathcal{E} \circ f \circ \mathcal{E}(A)| \ge |\mathcal{E}(A)| \ge (1 + \epsilon)|A|.$$

Finally, if $|A| \le 1/\epsilon$ and $|A| \le n/2$, then since the P-chain is irreducible,

$$|\mathcal{E} \circ f \circ \mathcal{E}(A)| \ge |\mathcal{E}(A)| \ge |A| + 1 \ge (1 + \epsilon)|A|.$$

So if E does not happen, then the random bijection f satisfies the condition (2.2). By (6.1), this completes the proof of Theorem 2.2. □

7 A different speedup

Going back to the original example (simple random walk on $\mathbb{Z}_n$), there is a different way to speed this up. Define a process $X_0, X_1, \ldots$ on $\mathbb{Z}_n$ by $X_0 = 0$, $X_1 = 1$ and

$$X_{k+1} = X_k + X_{k-1} + \epsilon_{k+1} \pmod{n},$$

where ϵ_i are independent, taking values 0, 1 and -1 with equal probabilities. Let $P_k(j) := \mathbb{P}(X_k = j)$ and $U(j) := 1/n$ for $j \in \mathbb{Z}_n$.

Theorem 7.1 *For any $n \ge 22$ and $k = 5[(\log n)^2 + c \log n]$,*

$$\|P_k - U\|_{TV} \le 1.6e^{-c/2}.$$

Remark The best lower bound we have is that at least $\log n$ steps are required. It is natural to suspect that this is the right answer, but numerical experiments do not make a clear case. At any rate, we find it interesting that a simple recurrence speeds things up from order n^2 to order $(\log n)^2$.

By running the recurrence, the chain can be represented as

$$X_k = F_k + F_{k-1}\epsilon_2 + F_{k-2}\epsilon_3 + \cdots F_1\epsilon_k \pmod{n},$$

with F_k the usual Fibonacci numbers 0, 1, 1, 2, 3, 5, … (so $F_5 = 5$). The (mod n) Fourier transform of P_k is

$$\hat{P}_k(a) = \mathbb{E}(e^{2\pi i a X_k/n}) = e^{2\pi i a F_k/n} \prod_{b=1}^{k-1} \left(\frac{1}{3} + \frac{2}{3}\cos(2\pi a F_b/n)\right). \tag{7.1}$$

We will use the inequality (see Diaconis [12, Chapter 3])

$$4\|P_k - U\|_{TV}^2 \leq \sum_{a=1}^{n-1} |\hat{P}_k(a)|^2 \tag{7.2}$$

to obtain an upper bound on the total variation distance between P_k and U. We thus need to know about the distribution of Fibonacci numbers mod n. We were surprised that we couldn't find what was needed in the literature (see Diaconis [14]). The following preliminary proposition is needed. Let $x_0, x_1, x_2, \ldots$ be any sequence of integers satisfying the Fibonacci recursion

$$x_k = x_{k-1} + x_{k-2}.$$

Take any n such that at least one x_i is not divisible by n. Let b_k be the remainder of x_k modulo n. We will prove the following property of this sequence.

Proposition 7.2 *For any j, there is some $j \leq k \leq j + 8 + 3\log_{3/2} n$ such that $b_k \in [n/3, 2n/3]$.*

We need several lemmas to prove this proposition.

Lemma 7.3 *There cannot exist k such that $b_k = b_{k+1} = 0$.*

Proof If $b_k = b_{k+1} = 0$ for some k, then x_{k+1} and x_k are both divisible by n. So the Fibonacci recursion implies that x_j is divisible by n for all j. But we chose n such that at least one x_i is not divisible by n. Thus, we get a contradiction. $\square$

Lemma 7.4 *If $b_j, b_{j+1} \in [1, n/3)$ for some j, then there is some $j + 2 \leq k < j + 2\log_2 n$ such that $b_k \in [n/3, 2n/3]$.*

Proof Since the b_i's satisfy the Fibonacci recursion modulo n, it follows that if b_i and b_{i+1} are both in $[1, n/3)$ for some i, then $b_{i+2} = b_{i+1} + b_i \in [1, 2n/3)$. So there exists an index k which is the first index bigger than $j + 1$ such that $b_k \in [n/3, 2n/3)$. We claim that for any $i \in [j + 2, k]$, $b_i \geq 2^{(i-j)/2}$. To see this, first note that for any $i \in [j + 2, k]$, $b_i = b_{i-1} + b_{i-2}$. Therefore, since $b_j, b_{j+1} \geq 1$, the claim is true for $i = j + 2$. Suppose that it is true up to $i - 1$. Then

$$b_i = b_{i-1} + b_{i-2} \geq 2^{(i-1-j)/2} + 2^{(i-2-j)/2}$$
$$= 2^{(i-j)/2}(2^{-1/2} + 2^{-1}) \geq 2^{(i-j)/2}.$$

In particular, $b_k \geq 2^{(k-j)/2}$. But we know that $b_k < 2n/3$. Combining these two inequalities gives $k - j < 2\log_2 n$. $\square$

Lemma 7.5 *If $b_j, b_{j+1} \in (2n/3, n-1]$ for some j, then there is some $j + 2 \leq k < j + 2\log_2 n$ such that $b_k \in [n/3, 2n/3]$.*

Proof Define $c_i := n - b_i$ for each i. Then the c_i's also satisfy the Fibonacci recursion modulo n. Moreover, $c_j, c_{j+1} \in [1, n/3)$. Therefore by the proof of Lemma

7.4, there exist $k < j + 2 \log_2 n$ such that $c_k \in [n/3, 2n/3]$. But this implies that $b_k \in [n/3, 2n/3]$.   □

Lemma 7.6 *If $b_j \in [1, n/3)$ and $b_{j+1} \in (2n/3, n-1]$ for some j, then there is some $j + 2 \le k < j + 6 + 3 \log_{3/2} n$ such that $b_k \in [n/3, 2n/3]$.*

Proof For each i, let

$$d_i := \begin{cases} b_i & \text{if } b_i \in [0, n/2], \\ b_i - n & \text{if } b_i \in (n/2, n-1]. \end{cases}$$

Clearly, the d_i's also satisfy the Fibonacci recursion modulo n, and $|d_i| \le n/2$ for each i. Take any i such that

- $d_i > 0$, $d_{i+1} < 0$, $d_{i+2} > 0$, $d_{i+3} < 0$ and $d_{i+4} > 0$, and
- $|d_i|$ and $|d_{i+1}|$ are less than $n/3$.

Under the above conditions, $d_i \in (0, n/3)$ and $d_{i+1} \in (-n/3, 0)$. This implies that $d_i + d_{i+1} < d_i < n/3$ and $d_i + d_{i+1} > d_{i+1} > -n/3$. Thus, $|d_i + d_{i+1}| < n/3$. But we know that $d_{i+2} \equiv d_i + d_{i+1} \pmod{n}$, and $|d_{i+2}| \le n/2$. Thus, we must have that d_{i+2} is actually equal to $d_i + d_{i+1}$, and therefore also that $|d_{i+2}| < n/3$. Similarly, $|d_{i+3}|$ and $|d_{i+4}|$ are also less than $n/3$, and satisfy the equalities $d_{i+3} = d_{i+2} + d_{i+1}$ and $d_{i+4} = d_{i+3} + d_{i+2}$. Thus,

$$0 < d_{i+4} = 3d_{i+1} + 2d_i,$$

which gives

$$|d_{i+1}| = -d_{i+1} < \frac{2d_i}{3} = \frac{2|d_i|}{3}.$$

Similarly, if $d_i < 0$, $d_{i+1} > 0$, $d_{i+2} < 0$, $d_{i+3} > 0$ and $d_{i+4} < 0$, and $|d_i|$ and $|d_{i+1}|$ are less than $n/3$, then also all of the absolute values are less than $n/3$, and

$$0 > d_{i+4} = 3d_{i+1} + 2d_i,$$

which gives

$$|d_{i+1}| = d_{i+1} < -\frac{2d_i}{3} = \frac{2|d_i|}{3}.$$

Now let j be as in the statement of the lemma. Then $d_j > 0$, $d_{j+1} < 0$, and $|d_j|$ and $|d_{j+1}|$ are both less than $n/3$. Let l be an index greater than j such that $d_j, d_{j+1}, \ldots, d_l$ are all nonzero, with alternating signs. Suppose that $l \ge j + 4$. The above deductions show that $|d_i| < n/3$ for all $i \in [j, l]$ and $|d_{i+1}| < 2|d_i|/3$ for all $i \in [j, l-4]$. Since $|d_{l-4}| \ge 1$ and $|d_j| < n/3$, this proves that l cannot be greater than $j + 4 + \log_{3/2} n$. Thus, if we define l to be the largest number greater than j with the above properties, then l is well-defined and is $\le j + 4 + \log_{3/2} n$.

By the definition of l, it follows that either $d_{l+1} = 0$, or d_{l+1} has the same sign as d_l. We already know that d_l and d_{l-1} are nonzero, have opposite signs, and are in $(-n/3, n/3)$. So, if $d_{l+1} = 0$, then $d_{l+2} = d_{l+3} = d_l \in (-n/3, -1] \cup [1, n/3)$, and if d_{l+1} is nonzero and has the same sign as d_l, then $|d_{l+1}| < n/3$. In the first situation, either both b_{l+2} and b_{l+3} are in $[1, n/3)$ or both are in $(2n/3, n-1]$. In the second situation, we can make the same deduction about b_l and b_{l+1}. The claim now follows by Lemmas 7.4 and 7.5. $\qquad\square$

Lemma 7.7 *If $b_j \in (2n/3, n-1]$ and $b_{j+1} \in [1, n/3)$ for some j, then there is some $j+2 \le k < j + 6 + 3\log_{3/2} n$ such that $b_k \in [n/3, 2n/3]$.*

Proof The proof is exactly the same as for Lemma 7.6. $\qquad\square$

We are now ready to prove Proposition 7.2.

Proof of Proposition 7.2 Take any j. If one of b_j and b_{j+1} is in $[n/3, 2n/3]$, there is nothing to prove. If b_j and b_{j+1} are both in $[1, n/3) \cup (2n/3, n-1]$, then one of Lemmas 7.4–7.7 can be applied to complete the proof. If $b_j = 0$ and $b_{j+1} \ne 0$, then $b_{j+2} = b_{j+1} \ne 0$, and so we can again apply one of the four lemmas. If $b_j \ne 0$ and $b_{j+1} = 0$, then $b_{j+2} = b_{j+3} = b_j \ne 0$, and so again one of the four lemmas can be applied. Finally, note that by Lemma 7.3, we cannot have $b_j = b_{j+1} = 0$. $\qquad\square$

Having proved Proposition 7.2, we can now complete the proof of Theorem 7.1.

Proof of Theorem 7.1 By (7.1) and (7.2), we get

$$4\|P_k - U\|_{TV}^2 \le \sum_{a=1}^{n-1} \prod_{b=1}^{k-1} \left(\frac{1}{3} + \frac{2}{3} \cos(2\pi a F_b/n) \right)^2.$$

Now take any $1 \le a \le n-1$. The sequence $aF_1, aF_2, \dots$ satisfies the Fibonacci recursion, and the first term of the sequence is not divisible by n since $a < n$. Thus, Proposition 7.2 is applicable to this sequence. Letting $m = 8 + 3\log_{3/2} n$, we get that at least $[(k-1)/m]$ among $aF_1, \dots, aF_{k-1}$ are in $[n/3, 2n/3]$ modulo n. Now if $x \in [n/3, 2n/3]$, then $\cos(2\pi x/n) \in [-1, -1/2]$, and so

$$\frac{1}{3} + \frac{2}{3} \cos(2\pi a F_b/n) \in [-1/3, 0].$$

Combining these observations, we get

$$4\|P_k - U\|_{TV}^2 \le n 9^{-[(k-1)/m]}.$$

It is easy to verify numerically that $30 \le m \le 10\log n$ for $n \ge 22$, and also that $\log(9)/10 \ge 1/5$. Thus, for $n \ge 22$ and $k = 5[(\log n)^2 + c\log n]$,

$$4\|P_k - U\|_{TV}^2 \le n9^{-(k-1)/m+1}$$
$$\le n9^{-k/m+31/30} \le 9^{31/30}e^{-c}.$$

It can now be numerically verified that the claimed bound holds. □

8 Applications and open problems

One class of problems where uniform sampling is needed arises from exponential families. Let $\mathcal{X}$ be a finite set and $T : \mathcal{X} \to \mathbb{R}^d$ be a given statistic. For $\theta \in \mathbb{R}^d$, let p_θ be the probability density

$$p_\theta(x) = Z(\theta)^{-1}e^{\theta \cdot T(x)},$$

where $Z(\theta)$ is the normalizing constant. The family $\{p_\theta\}_{\theta \in \mathbb{R}^d}$ is called an exponential family of probability densities with sufficient statistic T. If $X \sim p_\theta$, the conditional distribution of X given $T(X) = t$ is the uniform distribution on $\mathcal{X}_t := \{x \in \mathbb{R}^d : T(x) = t\}$. Such models appear in myriad statistical applications such as contingency tables and graphical models. They also appear in physics as Ising and related models. Uniform sampling on $\mathcal{X}_t$ is required to test if a given dataset fits the model. An overview is in Diaconis and Sturmfels [21], who introduced Gröbner basis techniques to do the sampling. These are typically diffusive and it would be wonderful to have speedups. A second use for uniform sampling comes from drawing samples from the original p_θ. For low-dimensional T, this can be done by sampling from the marginal distribution of T, and along the way, sampling X given $T(X) = t$.

Our main result shows that almost every bijection gives a speedup. This is not the same as having a specific bijection (such as $x \mapsto ax \pmod{n}$). Finding these, even in simple examples, seems challenging. One case where lots of bijections can be specified comes from permutation polynomials. Let us work mod p for a prime p. Then a permutation polynomial is a polynomial f with coefficients mod p such that $j \mapsto f(j)$ is one-to-one mod p. Large classes of these are known. The Wikipedia entry is useful and the article by Guralnick and Müller [27] shows how these can be found in several variables to map varieties (mod p) to themselves.

As an example, suppose that $(3, p - 1) = 1$. Then the map $j \mapsto j^3 \pmod{p}$ is one-to-one. The corresponding walk is

$$X_{k+1} = X_k^3 + \epsilon_{k+1} \pmod{p}.$$

We have no idea how to work with this but (weakly) conjecture that it mixes in order $\log p$ steps.

Our colleague Kannan Soundararajan has suggested $f(0) = 0$, $f(j) = j^{-1} \pmod{p}$ for $j \neq 0$. Preliminary exploration did not reveal this as an easy problem.

It is natural to try to generalize the Fibonacci recurrence walk further. The following generalization was proved by Jimmy He, who kindly allowed us to include his result in this paper.

Let X be a finite set and P be a Markov kernel on X. Let $f : X^n \to X$ be a function such that $f(\cdot, x_2, \ldots, x_n)$ is a bijection for all $x_2, \ldots, x_n \in X$. Define a Markov chain P_f on X^n by moving from the state $(X_1, \ldots, X_n)$ to the state $(X_2, \ldots, X_n, f(X_1, \ldots, X_n))$, and then taking a step from P in the last coordinate. Here the last coordinate can be viewed as a higher order Markov chain on X, which depends on the previous n steps of the walk.

To see that this is a generalization of the Fibonacci walk, note that if $n = 2$, $X = G$ is a finite Abelian group, P is a random walk on G generated by a measure Q on G, and $f(x_1, x_2) = x_1 + x_2$, then the walk moves from (X_1, X_2) to $(X_2, X_1 + X_2 + \varepsilon)$ where ε is drawn from Q, and this is exactly the original Fibonacci walk.

Proposition 8.1 (Due to Jimmy He, personal communication) *Assume that P is lazy and ergodic, and has a uniform stationary distribution. Then P_f is ergodic, and has a uniform stationary distribution.*

Proof First, note that the assumption that $f(\cdot, x_2, \ldots, x_n)$ is a bijection for all $x_2, \ldots, x_n \in X$ implies that the function $g : X^n \to X^n$ defined by

$$g(x_1, \ldots, x_n) = (x_2, \ldots, x_n, f(x_1, \ldots, x_n))$$

is a bijection. Then P_f can be described as applying the function g, followed by a step from P in the last coordinate.

Now we describe how to mimic steps from P using P_f. Note that since g is a permutation, g^m is the identity permutation for some m. Since P is lazy, we can alternate applying g, and then remaining stationary when taking a step from P. Doing so $m - 1$ times, and then applying g one last time, we can then take a step from P in the last coordinate. Thus, it is possible to move from $(x_1, \ldots, x_{n-1}, x_n)$ to $(x_1, \ldots, x_{n-1}, x'_n)$ in m steps of P_f, where $x'_n \in X$ is some state for which $P(x_n, x'_n) > 0$. We call this mimicking a step from P.

To show that P_f is irreducible, we simply repeat the above procedure. Since P is irreducible, we can reach any state in X in the last coordinate, while keeping the first $n - 1$ coordinates fixed. But now, we can apply g once, and then repeat the above procedure, and now the last two coordinates can be made arbitrary (since the procedure described fixes the first $n - 1$ coordinates). Repeating this n times allows any state in X^n to be reached.

Now we show that P_f is aperiodic. Starting from any state, we can return in m steps by mimicking a lazy step from P using the above procedure. But we can also first take a single step from P_f, and then use the same procedure as in the proof of irreducibility to return to the initial state, using a path of length km for some k. This gives a path from the state to itself of length $km + 1$. As m and $km + 1$ are coprime, this implies that P_f is aperiodic.

Finally, it is clear that the stationary distribution is uniform, since g is a bijection and so preserves the uniform distribution, and taking a step from P in the last coordinate also preserves the uniform distribution.   □

Proposition 8.1 allows us to generalize the Fibonacci walk to finite non-Abelian groups. If G is a finite group and Q is a probability on G with support generating G and having nonzero mass at the identity, then Proposition 8.1 implies that the second order Markov chain which proceeds by $X_{k+1} = X_k X_{k-1} \epsilon_{k+1}$ (with the ϵ_i drawn i.i.d. from Q), is ergodic with uniform stationary distribution.

Proposition 8.1 also allows generalization of Fibonacci walks to nonlinear recurrences on finite fields. For example, take a prime p such that $(3, p - 1) = 1$, and consider the random walk on $\mathbb{F}_p$ which proceeds as

$$X_{k+1} = X_k^3 + X_{k-1} + \epsilon_{k+1},$$

where ϵ_i are i.i.d. uniform from $\mathbb{F}_p$. Proposition 8.1 implies that this has uniform stationary distribution, since $f(x, y) = x^3 + y$ is a bijection in x for every fixed y.

Getting rates of convergence in any of the above examples seems like a challenging problem.

References

1. Anderson, H.C., Diaconis, P.: Hit and run as a unifying device. J. Soc. Francaise Statist. **148**, 5–28 (2007)
2. Avena, L., Güldaş, H., van der Hofstad, R., den Hollander, F.: Mixing times of random walks on dynamic configuration models. Ann. Appl. Probab. **28**(4), 1977–2002 (2018)
3. Benaïm, M., Le Borgne, S., Malrieu, F., Zitt, P.-A.: Qualitative properties of certain piecewise deterministic Markov processes. Ann. Inst. Henri Poincaré Probab. Stat. **51**(3), 1040–1075 (2015)
4. Boardman, S., Saloff-Coste, L.: The hit-and-run version of top-to-random (2020, in preparation)
5. Bordenave, C., Lacoin, H.: Cutoff at the entropic time for random walks on covered expander graphs (2018). Preprint arXiv:1812.06769
6. Bordenave, C., Qiu, Y., Zhang, Y.: Spectral gap of sparse bistochastic matrices with exchangeable rows with application to shuffle-and-fold maps (2018). Preprint arXiv:1805.06205
7. Breuillard, E., Varjú, P.: Irreducibility of random polynomials of large degree. Acta Math. **223**(2), 195–249 (2019)
8. Breuillard, E., Varjú, P.: Cut-off phenomenon for the ax+b Markov chain over a finite field (2019). Preprint arXiv:1909.09053
9. Chung, F.R.K., Diaconis, P., Graham, R.L.: Random walks arising in random number generation. Ann. Probab. **15**(3), 1148–1165 (1987)
10. Conchon–Kerjan, G.: Cutoff for random lifts of weighted graphs (2019). Preprint arXiv:1908.02898
11. Constantin, P., Kiselev, A., Ryzhik, L., Zlatoš, A.: Diffusion and mixing in fluid flow. Ann. Math. (2) **168**(2), 643–674 (2008)
12. Diaconis, P.: Group representations in probability and statistics. Institute of Mathematical Statistics, Hayward, CA (1988)

13. Diaconis, P.: Some things we've learned (about Markov chain Monte Carlo). Bernoulli **19**(4), 1294–1305 (2013)
14. Diaconis, P.: Probabilizing Fibonacci Numbers. Connections in Discrete Mathematics. Cambridge University Press, Cambridge (2018)
15. Diaconis, P., Graham, R.: An affine walk on the hypercube. Asymptotic methods in analysis and combinatorics. J. Comput. Appl. Math. **41**(1–2), 215–235 (1992)
16. Diaconis, P., Graham, R.: Binomial coefficient codes over GF(2). A collection of contributions in honour of Jack van Lint. Discret. Math. **106/107**, 181–188 (1992)
17. Diaconis, P., Holmes, S., Neal, R.M.: Analysis of a nonreversible Markov chain sampler. Ann. Appl. Probab. **10**(3), 726–752 (2000)
18. Diaconis, P., Miclo, L.: On the spectral analysis of second-order Markov chains. Ann. Fac. Sci. Toulouse Math. (6) **22**(3), 573–621 (2013)
19. Diaconis, P., Saloff-Coste, L.: Moderate growth and random walk on finite groups. Geom. Funct. Anal. **4**(1), 1–36 (1994)
20. Diaconis, P., Shahshahani, M.: Time to reach stationarity in the Bernoulli–Laplace diffusion model. SIAM J. Math. Anal. **18**(1), 208–218 (1987)
21. Diaconis, P., Sturmfels, B.: Algebraic algorithms for sampling from conditional distributions. Ann. Stat. **26**(1), 363–397 (1998)
22. Ding, J., Peres, Y.: Sensitivity of mixing times. Electron. Commun. Probab.**18**(88), 6 pp (2013)
23. Eberhard, S., Varjú, P.: Mixing time of the Chung–Diaconis–Graham random process (2020). Preprint arXiv:2003.08117
24. Ganguly, S., Peres, Y.: Permuted random walk exits typically in linear time. In: ANALCO14—Meeting on Analytic Algorithmics and Combinatorics, pp. 74–81. SIAM, Philadelphia (2014)
25. Gerencsér, B., Hendrickx, J.M.: Improved mixing rates of directed cycles by added connection. J. Theoret. Probab. **32**(2), 684–701 (2019)
26. Greenhalgh, A.: On a model for random random-walks on finite groups (1990). Unpublished manuscript
27. Guralnick, R.M., Müller, P.: Exceptional polynomials of affine type. J. Algebra **194**(2), 429–454 (1997)
28. Hermon, J.: On sensitivity of uniform mixing times. Ann. Inst. Henri Poincaré Probab. Stat. **54**(1), 234–248 (2018)
29. Hermon, J., Peres, Y.: On sensitivity of mixing times and cutoff. Electron. J. Probab. **23**, Paper No. 25, 34 pp (2018)
30. Hermon, J., Sly, A., Sousi, P.: Universality of cutoff for quasi-random graphs (2020). Preprint arXiv:2008.08564
31. Hildebrand, M.: A survey of results on random random walks on finite groups. Probab. Surv. **2**, 33–63 (2005)
32. Hildebrand, M.: A lower bound for the Chung–Diaconis–Graham random process. Proc. Am. Math. Soc. **137**(4), 1479–1487 (2009)
33. Kapfer, S.C., Krauth, W.: Irreversible local Markov chains with rapid convergence towards equilibrium. Phys. Rev. Lett. **119**(24), 240603 (2017)
34. Lubetzky, E., Peres, Y.: Cutoff on all Ramanujan graphs. Geom. Funct. Anal. **26**(4), 1190–1216 (2016)
35. Lubetzky, E., Sly, A.: Cutoff phenomena for random walks on random regular graphs. Duke Math. J. **153**(3), 475–510 (2010)
36. Malrieu, F.: Some simple but challenging Markov processes. Ann. Fac. Sci. Toulouse Math. (6) **24**(4), 857–883 (2015)
37. Neal, R.M.: Improving asymptotic variance of mcmc estimators: non-reversible chains are better (2004). Preprint arXiv:math/0407281
38. Neville, R., III.: On lower bounds of the Chung–Diaconis–Graham random process. Ph.D. thesis, State University of New York at Albany (2011)
39. Ottino, J.M.: The Kinematics of Mixing: Stretching, Chaos, and Transport. Cambridge University Press, Cambridge (1989)
40. Pymar, R., Sousi, P.: A permuted random walk exits faster. ALEA Lat. Am. J. Probab. Math. Stat. **11**(1), 185–195 (2014)

41. Rallabandi, B., Wang, C., Hilgenfeldt, S.: Analysis of optimal mixing in open-flow mixers with time-modulated vortex arrays. Phys. Rev. Fluids **2**(6), 064501 (2017)
42. Wilson, D.B.: Random random walks on $\mathbb{Z}_2^d$. Probab. Theory Related Fields **108**(4), 441–457 (1997)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.