The-Square-and-Add Markov Chain

PERSI DIACONIS, JIMMY HE, AND I. MARTIN ISAACS

In memory of John Conway

Let us begin with a problem we cannot solve. If q is a prime power, we write $\mathbf{F}_q$ to denote the field with q elements, so if p is prime, $\mathbf{F}_p$ is the field of integers modulo p. A simple random walk (drunkard's walk) on $\mathbf{F}_p$ goes from j to $j+1$ or $j-1$ with probability $1/2$. As time goes on, this converges to the uniform distribution on $\mathbf{F}_p$. This means that after a long time, the probability that the random walk will be at some $\alpha \in \mathbf{F}_p$ is about $1/p$. It takes about p^2 steps for this convergence to kick in. This is slow: if $p = 101$, then $p^2 = 10\,201$. These informal statements are explained more carefully after Theorem 1 below.

One attempt to speed things up intersperses deterministic doubling with the random ±1 steps. If X_n denotes the position of the walk after n steps (say starting from $X_0 = 0$), then this new walk is

$$X_n = 2X_{n-1} + \varepsilon_n \pmod{p},$$

with $\varepsilon_n = \pm 1$ with probability $1/2$, independently from step to step.

In [3], it is shown that order-log(p) steps are necessary and sufficient for convergence (log will always refer to the natural logarithm). See [8] for amazing refinements and [2] for other applications to deterministic speedup.

Seeking to understand such speedups, we consider the random walk

$$X_n = X_{n-1}^2 + \varepsilon_n \pmod{p}.$$

This is the problem we cannot solve! We do not understand the stationary distribution—numerical evidence at the end of this paper shows that it is wildly nonuniform. We do not even know its support, much less rates of convergence to stationarity.

Squaring defines an automorphism of a finite field of 2-power order, so we decided to study the corresponding problem over the field $\mathbf{F}_q$, where $q = 2^d$. To be specific, we choose a basis $\mathcal{B}$ for $\mathbf{F}_q$ over its prime subfield $\mathbf{F}_2$, so $|\mathcal{B}| = d$, and we consider the random walk on the elements of $\mathbf{F}_q$ defined by setting $X_0 = 0$ and

$$X_n = X_{n-1}^2 + \epsilon_n \tag{1}$$

for $n > 0$. Here ϵ_n is randomly chosen from the set $\{0\} \cup \mathcal{B}$, where the probability that $\epsilon_n = 0$ is $1/2$, and for each element $\alpha \in \mathcal{B}$, the probability that $\epsilon_n = \alpha$ is $\frac{1}{2d}$. The unique stationary distribution for this walk is the uniform distribution $\pi(\alpha) = 1/2^d$. (Random walks, or in more formal language, Markov chains, are discussed in greater detail below.)

If we were to omit the squaring and simply take $X_n = X_{n-1} + \epsilon_n$, it is not hard to see that the behavior of the resulting walk would be independent of the choice of the basis $\mathcal{B}$ that defines it. Surprisingly, however, the walk we defined above (which includes squaring) does depend on the choice of the basis. To illustrate this, we compute the transition matrices for the square-and-add Markov chains on $\mathbf{F}_8$ defined using two different bases. As we shall see, these matrices have different eigenvalues.

First, we explain what we mean by the transition matrix for a Markov chain on a finite set X. This is a square matrix M, with rows and columns indexed by the members of X, where for $x, y \in X$, the entry $M(x, y)$ in row x and column y is the probability of arriving at y in one step, starting at x.

To see the relevance of the transition matrix, let v_n denote the row vector having entries indexed by the elements of X, where the entry at position x in v_n is the probability that the random walk has arrived at x at time n. It is easy to see that $v_{n+1} = v_n M$, so $v_n = v_0 M^n$, and thus the convergence of the Markov chain is controlled by the powers of the transition matrix M.

To compute transition matrices for our walks on $\mathbf{F}_8$, we need to name the elements of this field, and to do this, we take advantage of the fact that in general, the multiplicative

https://doi.org/10.1007/s00283-021-10058-w

group of the finite field $\mathbf{F}_q$ is cyclic of order $q-1$. If we fix a generator r for this group (so r is a primitive element), we see that the elements of the field are 0 and r^i for $0 \le i \le q-2$.

Once we have named the field elements in this way, it is trivial to see how to compute the product of two field elements, but it is not clear how to determine their sum. In fact, more information is needed before this is possible: it suffices, for example, to know the minimal polynomial $f(x)$ of r over the prime subfield of $\mathbf{F}_q$. Taking $q = 2^d$, we see that f is an irreducible polynomial of degree d over $\mathbf{F}_2$, so if $q = 8$, we can assume that $f(x) = x^3 + x + 1$, and with this information, the arithmetic in $\mathbf{F}_8$ is completely determined.

The transition matrix M for a Markov chain on $\mathbf{F}_8$ is an 8×8 matrix whose rows and columns are indexed by the field elements, and we choose to write these elements in the order 0, 1, r, r^2, r^3, r^4, r^5, r^6, and we recall that the entry $M(\alpha, \beta)$ is the probability that one step of the chain goes from α to β.

If we take the basis $\mathcal{B} = \{1, r, r^2\}$, it is not hard to compute that the matrix is

$$\begin{array}{c|cccccccc|}
 & 0 & 1 & r & r^2 & r^3 & r^4 & r^5 & r^6 \\
0 & \frac{1}{2} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 \\
1 & \frac{1}{6} & \frac{1}{2} & 0 & 0 & \frac{1}{6} & 0 & 0 & \frac{1}{6} \\
r & \frac{1}{6} & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{6} & 0 & \frac{1}{6} \\
r^2 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{2} & \frac{1}{6} & 0 \\
r^3 & 0 & \frac{1}{6} & 0 & \frac{1}{6} & 0 & 0 & \frac{1}{6} & \frac{1}{2} \\
r^4 & \frac{1}{6} & 0 & \frac{1}{2} & 0 & \frac{1}{6} & \frac{1}{6} & 0 & 0 \\
r^5 & 0 & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{2} & 0 & \frac{1}{6} & 0 \\
r^6 & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{2} & \frac{1}{6}
\end{array}.$$

The eigenvalues of this matrix are 0, 0, 0, 2/3, 1, and the three cube roots of 4/27.

If instead we take $\mathcal{B} = \{r^3, r^5, r^6\}$, the transition matrix is

$$\begin{array}{c|cccccccc|}
 & 0 & 1 & r & r^2 & r^3 & r^4 & r^5 & r^6 \\
0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} \\
1 & 0 & \frac{1}{2} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & 0 & 0 \\
r & 0 & \frac{1}{6} & 0 & \frac{1}{2} & \frac{1}{6} & 0 & \frac{1}{6} & 0 \\
r^2 & 0 & \frac{1}{6} & 0 & 0 & \frac{1}{6} & \frac{1}{2} & 0 & \frac{1}{6} \\
r^3 & \frac{1}{6} & 0 & \frac{1}{6} & 0 & 0 & \frac{1}{6} & 0 & \frac{1}{2} \\
r^4 & 0 & \frac{1}{6} & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} \\
r^5 & \frac{1}{6} & 0 & 0 & \frac{1}{6} & \frac{1}{2} & \frac{1}{6} & 0 & 0 \\
r^6 & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & 0 & 0 & \frac{1}{2} & 0
\end{array},$$

and the eigenvalues of this matrix are 0, 1, the three cube roots of 1/27, and the three cube roots of 8/27.

Since these two Markov chains on $\mathbf{F}_8$ have transition matrices with different sets of eigenvalues, we see that random walks determined by different bases for $\mathbf{F}_8$ can have different long-term behaviors. We do not know, however, the extent to which the choice of a basis for $\mathbf{F}_q$ can affect the rate of convergence of the corresponding Markov chain.

The second of our two bases for $\mathbf{F}_8$, namely $\{r^3, r^5, r^6\}$, consists of an orbit under the automorphism group of $\mathbf{F}_8$, which is the group generated by the squaring map. In fact, for every prime power q, there always exists a basis for $\mathbf{F}_q$ that forms an orbit under the automorphism group of the field. Such a basis is said to be a normal basis, and it happens that our basis $\{r^3, r^5, r^6\}$ is the unique normal basis for $\mathbf{F}_8$. (The set $\{r, r^2, r^4\}$ is also an orbit under the automorphism group, but it is not a basis, because $r + r^2 + r^4 = 0$, since r is a root of the polynomial $x^3 + x + 1$.)

Although the properties of a Markov chain on $\mathbf{F}_q$ defined by choosing a basis can depend on the chosen basis, it can be proved that the transition matrices for chains defined by normal bases are identical up to an appropriate renaming of the field elements. It follows that the corresponding random walks are essentially the same. In fact, if $q = 2^d$, it is not hard to show that after a multiple of d steps, the probability distribution of a square-and-add walk defined on $\mathbf{F}_q$ using a normal basis is exactly the same as the distribution for the walk on $\mathbf{F}_q$ without squaring. Also, this is the same as a simple random walk on the binary hypercube, and it is well known that this walk takes $\frac{1}{2}d(\log(d) + c)$ steps to converge [5].

A Conway Digression

Once we realized that the way the field is represented matters, our thoughts turned to one of the hundreds of magical mathematical gems that John Conway left to us: Conway polynomials. (These are unrelated to the Conway–Alexander polynomials in knot theory.)

Fix a prime p, and let $f(x) \in \mathbf{F}_p[x]$ be an irreducible polynomial of degree n. Then the quotient ring $\mathbf{F}_p[x]/(f)$ is a field of order p^n, and all choices of f yield the same (up to isomorphism) field $\mathbf{F}_{p^n}$. The p^n elements of this field can be represented as polynomials $h(x)$ of degree at most $n-1$, and adding field elements in this representation is easy, but multiplication is more tedious. Alternatively, we can take advantage of the fact that the multiplicative group of a finite field is cyclic, so there is a generator r of the multiplicative group $(\mathbf{F}_{p^n})^*$ of order $p^n - 1$. (Any such generator is referred to as a primitive element of $\mathbf{F}_{p^n}$.) Given a primitive element r, therefore, the distinct nonzero elements of $\mathbf{F}_{p^n}$ are r^i for $0 \le i \le p^n - 2$, and thus

$$\mathbf{F}_{p^n} = \{0, 1, r, r^2, \ldots, r^{p^n-2}\},$$

and we see that with this representation of the field elements, multiplication is a triviality, but unfortunately, addition can be quite difficult.

An irreducible polynomial $f \in \mathbf{F}_p[x]$ of degree n is said to be a primitive polynomial if one of its roots—and hence all of them—is a primitive element of the field $\mathbf{F}_{p^n}$, and in this case, the element of $\mathbf{F}_p[x]/(f)$ represented by the polynomial $h(x) = x$ is a primitive element. If we had constructed the field $\mathbf{F}_{p^n}$ using an imprimitive irreducible polynomial f, it would not be so clear which elements $h(x)$ were primitive.

If m is a divisor of n, then $\mathbf{F}_{p^n}$ contains a unique subfield $\mathbf{F}_{p^m}$ of order p^m, and the multiplicative group $(\mathbf{F}_{p^m})^*$ is the unique subgroup of index $(p^n-1)/(p^m-1)$ in the cyclic group $(\mathbf{F}_{p^n})^*$. Thus if r is a primitive element of $\mathbf{F}_{p^n}$, then $r^{(p^n-1)/(p^m-1)}$ is one of the primitive elements of $\mathbf{F}_{p^m}$.

Suppose we construct the fields $\mathbf{F}_{p^n}$ and $\mathbf{F}_{p^m}$ using primitive polynomials f and g in $\mathbf{F}_p[x]$, having degrees n and m respectively. If $r \in \mathbf{F}_{p^n}$ is a root of f and $s \in \mathbf{F}_{p^m}$ is a root of g, then r and s are primitive elements of $\mathbf{F}_{p^n}$ and $\mathbf{F}_{p^m}$, so they generate the multiplicative groups $(\mathbf{F}_{p^n})^*$ and $(\mathbf{F}_{p^m})^*$. Viewing $\mathbf{F}_{p^m}$ as a subfield of $\mathbf{F}_{p^n}$, we know that $r^{(p^n-1)/(p^m-1)}$ is a primitive element of $\mathbf{F}_{p^m}$, but there is no reason to believe that it is equal to s, or to any other root of g. Wouldn't it be nice if we could choose the primitive polynomials f and g with degrees n and m in such a way that if r is a root of f, then $r^{(p^n-1)/(p^m-1)}$ is a root of g? That would make it much easier to see how $\mathbf{F}_{p^m}$ is embedded as a subfield of $\mathbf{F}_{p^n}$.

Given primitive polynomials f and g in $\mathbf{F}_p[x]$, where f has degree n, g has degree m, and m divides n, we say that f and g are compatible if for every root r of f in $\mathbf{F}_{p^n}$, the element $r^{(p^n-1)/(p^m-1)}$ is a root of g. Equivalently, f and g are compatible if the polynomial $g\left(x^{(p^n-1)/(p^m-1)}\right)$ is divisible by f. (Note that this latter formulation of the compatibility condition can be checked using nothing more sophisticated than polynomial long division, and in particular, it does not require finding roots of polynomials.) Conway has given us a way to construct compatible primitive polynomials. Even better, Conway showed how to construct, for each prime p, a family of primitive polynomials $\{f_n\}$, one for each positive integer n, such that f_n has degree n, and whenever m divides n, the polynomials f_m and f_n are compatible.

It does not seem obvious that such families of compatible polynomials exist, but in fact, they do, and there is an abundance of riches: there is more than one compatible family for each prime p, even if we require that all of the polynomials f_n be monic.

Conway described a somewhat arbitrary procedure that would uniquely determine a specific compatible family of monic primitive polynomials. Conway's procedure stuck, and the resulting polynomials, now referred to as the Conway polynomials, are the default, and they are used in such computer algebra systems as Magma and GAP. We can give a taste of Conway's procedure by considering the linear (degree-1) case. Every polynomial of the form $x-a$, where a is a primitive root modulo p, is a primitive polynomial. Among these, Conway chose the one for which a is minimal in the ordering $1 < 2 < \cdots < p-1$. For example, the smallest primitive root modulo 7 is 3, so the Conway polynomial of degree 1 for $p=7$ is $x-3$. On consulting tables of Conway polynomials, we see that the degree-2 Conway polynomial for $p=7$ is x^2+6x+3. Taking $m=1$ and $n=2$, we have $(7^2-1)/(7-1)=48/6=8$, so we can verify the compatibility condition by checking that x^8-1 is divisible by x^2+6x+3 in $\mathbf{F}_7(x)$.

A rough description of a recursive algorithm to compute the Conway polynomial f_n for a prime p is as follows. If $n=1$, then as we have already mentioned, $f_n(x)=x-a$, where a is the "smallest" primitive root modulo p. Assume now that $n>1$ and that we have already found all of the Conway polynomials f_m for m a proper divisor of n. Consider the set $\mathcal{S}_n$ of all degree-n monic primitive polynomials f such that f is compatible with all of the polynomials f_m for proper divisors m of n. It is not obvious, but it is true, that the set $\mathcal{S}_n$ is nonempty, so we can define f_n to be the smallest member of $\mathcal{S}_n$ with respect to a specific linear ordering of polynomials defined by Conway. We mention that to determine whether a polynomial f lies in $\mathcal{S}_n$, it is not necessary to check whether f is compatible with f_m for all the proper divisors m of n; it suffices to consider only those divisors of the form $m=n/q$, where q is prime. The reason for this is that if l divides m, and m divides n, and we have established that f is compatible with f_m, then f is guaranteed to be compatible with f_l. This follows easily from the fact that f_m is compatible with f_l.

For more on Conway polynomials, see [11, 16]. A detailed listing of available Conway polynomials can also be found on Frank Lübeck's website [16]. Let us end this digression by admitting that we have not (yet) found that Conway polynomials mesh with our study of "square-and-add" random walks. Our problem gave us the excuse, however, to marvel at Conway's magic, and that is almost as good as finding a new theorem.

What We Can Prove

There is one situation in which a sharp analysis of the square-and-add Markov chain on a field of 2-power order is possible. Following a suggestion of Amol Aggarwal, we let p be a prime such that 2 is a primitive root modulo p, which means that 2 generates the multiplicative group of $\mathbf{F}_p$. (According to the Artin primitive root conjecture, which was proved conditionally on the generalized Riemann hypothesis [12], these have positive density among all primes.)

Then for $d=p-1$, the cyclotomic polynomial

$$f(x)=x^d+x^{d-1}+\cdots+x+1 \tag{2}$$

is irreducible over $\mathbf{F}_2$. (These polynomials are discussed below in the subsection on cyclotomic polynomials.) With these assumptions, the field $\mathbf{F}_2[x]/(f)$ has order 2^d, and a basis is

$$\{1,x,x^2,\ldots,x^{d-1}\}. \tag{3}$$

(Note that because $x^d=1$, x is not a primitive element of this field, and so x does not have order 2^d.) The following result says roughly that about $\frac{1}{2}d\log(d)$ steps are necessary and sufficient for convergence of the Markov chain determined by this basis on the field $\mathbf{F}_2[x]/(f)$.

If K denotes the transition matrix for a Markov chain, let $K^n(\alpha,\beta)$ denote the probability of moving from α to β in n steps. Let

$$\|P-Q\|_{TV}=\frac{1}{2}\sum_{\alpha\in\mathbf{F}_q}|P(\alpha)-Q(\alpha)|$$

denote the total variation distance of probability measures.

Theorem 1. *Let p be a prime with 2 a primitive root in $\mathbf{F}_p$ and let $d = p - 1$. In $\mathbf{F}_q$, with $q = 2^d$, the Markov chain (1), defined by the basis (3), satisfies, for $n = \frac{1}{2}d(\log(d) + c)$ with $c > 0$,*

$$\|K^n(0,\cdot) - \pi\|_{TV} \le ae^{-bc},$$

and for $n = \frac{1}{2}d(\log(d) - c)$ with $c > 0$, it satisfies

$$\|K^n(0,\cdot) - \pi\|_{TV} \ge 1 - d'e^{-bc}$$

for universal constants d', a, and b, where π denotes the uniform measure.

Informally, the precise upper and lower bounds in Theorem 1 can be phrased thus: about $\frac{1}{2}d\log(d)$ steps are necessary and sufficient for convergence.

The heart of the proof is some magical combinatorics for the Frobenius map of repeated squaring. It is the kind of magic John Conway enjoyed.

Remark 2: Theorem 1 holds in more generality. As long as d is even, f defined by (2) has no repeated factors, and so the random walk can be defined on the quotient $\mathbf{F}_2[x]/(f)$, which will be a direct sum of fields (with componentwise addition and multiplication). Squaring is still an isomorphism in this case. This is proved in Lemma 5. The same bounds hold in this case. Theorem 1 can also be extended (although with weaker estimates) to general primes p, with the random walk being $X_{n+1} = X_n^p + \varepsilon_{n+1}$.

The combinatorics of combining addition and multiplication in finite fields is currently a hot topic in additive combinatorics; see [9]. The problems studied here seem different.

Background

This section contains some needed background on Markov chains, finite fields, and Fourier analysis over $(\mathbf{F}_2)^d$. It presents these topics in a form needed to prove Theorem 1, which will be proved in the section following. A final section returns to the square-and-add walk over $\mathbf{F}_p$ and has some computed examples and open questions.

Markov Chains

A Markov chain is a sequence of random variables X_n taking values in some finite set X, so that X_{n+1} depends on $X_1, \ldots, X_n$ solely through X_n. We will assume that our Markov chains are homogeneous, which means that the chance of moving from one state to another at step n doesn't depend on n. Such a process can be represented using a matrix P indexed by X, whose entries $P(x, y)$ encode the chance of moving from x to y. Here, by convention, probability distributions are written as row vectors, and P acts on the right, so if $\mu_n(x)$ is the chance of being at x after n steps of the Markov chain, then $\mu_n = \mu_{n-1}P$.

A stationary distribution for the Markov chain defined by P is some probability measure π on X such that $\pi P = \pi$. A Markov chain is said to be irreducible if for any two states $x, y \in X$, there is some positive integer t such that $P^t(x, y) > 0$. This means that it is possible to reach any state from any other in the chain. A Markov chain is said to be aperiodic if $P^t(x, x) > 0$ for all sufficiently large t. Note that a sufficient condition for P to be aperiodic is $P^s(x, x) > 0$ and $P^t(x, x) > 0$ for some $s, t \ge 1$ with $(s, t) = 1$. By the Perron–Frobenius theorem, an aperiodic irreducible Markov chain has a unique stationary distribution.

Finite Fields

The classical subject of finite fields is exhaustively developed in [15]. Throughout, we take $q = p^d$, where p is prime, and we write $\mathbf{F}_q$ to denote the unique field with q elements. If f is an arbitrary irreducible degree-d polynomial with coefficients in $\mathbf{F}_p$, then

$$\mathbf{F}_q \cong \mathbf{F}_p[x]/(f),$$

and if we represent $\mathbf{F}_q$ in this way, we see that the set $\{1, x, x^2, \ldots, x^{d-1}\}$ is a basis for $\mathbf{F}_q$ over its prime subfield $\mathbf{F}_p$.

Even if $f(x)$ is not irreducible, $\mathbf{F}_p[x]/(f)$ is still an algebra over $\mathbf{F}_p$, and $1, x, \ldots, x^{d-1}$ is still a basis. This algebra is readily identified, provided that f has no repeated factors.

Lemma 3. *Let $f(x) \in \mathbf{F}_p[x]$ have no repeated factors. Suppose that $f = \prod f_i$, where the degree of f_i is d_i. Then:*

1. *$\mathbf{F}_p[x]/(f)$ is isomorphic to the direct sum of the fields $\mathbf{F}_p[x]/(f_i) \cong \mathbf{F}_{p^{d_i}}$.*
2. *The map $y \mapsto y^p$ is an automorphism on $\mathbf{F}_p[x]/(f)$.*

Proof The first claim is a restatement of the Chinese remainder theorem, and the second claim follows from the first, since the map $y \mapsto y^p$ is an automorphism for each factor. □

The random walk (1) can be defined on the algebra $\mathbf{F}_p[x]/(f)$ using the basis $\mathcal{B} = \{1, x, x^2, \ldots, x^{d-1}\}$, even if the polynomial f is not irreducible. And provided that f has no repeated factors, this walk has a uniform stationary distribution. In the following lemma, we take $p = 2$.

Lemma 4. *Let $f \in \mathbf{F}_2[x]$, where f has no repeated factors. Then the Markov chain on $\mathbf{F}_2[x]/(f)$ defined as in (1) with respect to the basis $\mathcal{B} = \{1, x, x^2, \ldots, x^{d-1}\}$ is irreducible, aperiodic, and has a unique stationary distribution, which is uniform.*

Proof Factor the transition matrix for the random walk as $K = PT$, where T is the transition matrix for the walk defined by $X_n = X_{n-1} + \varepsilon_n$ and P is the permutation matrix encoding the bijection $y \mapsto y^2$ on $\mathbf{F}_2[x]/(f)$.

Since P is a permutation matrix, it has some finite order, so $P^n = I$ for some $n > 0$. First, we show that $K^n(\alpha, \beta) > 0$ if $T(\alpha, \beta) > 0$. It will be useful to view a step from K as applying P followed by a step from T. Since T is lazy, we can always apply P and then remain stationary for the step from T; so do this $n - 1$ times. At the very last step, instead of remaining stationary, take a step from T. The result is moving according to P exactly n times, returning to the initial state, and then a step from T, and so $K^n(\alpha, \beta) > 0$ if $T(\alpha, \beta) > 0$.

To see that K is irreducible, observe first that T is irreducible, so there exists a path using steps from T that goes from α to β. Since each step of T can be mimicked by a block of n steps from K, it follows that there is a path from α to β using steps from K.

To see that the Markov chain is aperiodic, start by taking a single step from K, say going from α to β, and then take steps in blocks of size n, going from β back to α, which is possible because T is irreducible and $K^n(\alpha, \beta) > 0$ if $T(\alpha, \beta) > 0$. This means that $K^{kn+1}(\alpha, \alpha) > 0$ for some k. Also, since T is lazy, $K^n(\alpha, \alpha) > 0$. Because $kn + 1$ and n are coprime, the Markov chain is aperiodic.

Finally, since T and P both preserve the uniform distribution, so does K. Irreducibility and aperiodicity imply uniqueness of the stationary distribution. □

Cyclotomic Polynomials

Fix $n \in \mathbf{N}$ and let the cyclotomic polynomials $\Phi_n(x) \in \mathbf{Z}[x]$ be defined by

$$\Phi_n(x) = \prod_{\substack{1 \le k \le n \\ \gcd(k,n)=1}} \left(x - e^{2\pi i k/n}\right).$$

The following facts are well known (see [15, Section 2.4], for example):

- $\Phi_n(x)$ has degree $\phi(n)$ (ϕ denotes the Euler totient function).
- The coefficients of $\Phi_n(x)$ lie in $\mathbb{Z}$.
- $\Phi_n(x)$ is irreducible over $\mathbb{Q}$.
- If p is prime, then $\Phi_p(x) = 1 + x + \cdots + x^{p-1}$.
- $\Phi_{p^n}(x) = \Phi_p(x^{p^{n-1}})$.

A primitive element, or primitive root, of $\mathbf{Z}/n\mathbf{Z}$ is an element that generates the group of units $(\mathbf{Z}/n\mathbf{Z})^{\times}$. A primitive polynomial over $\mathbf{F}_p$ is the minimal polynomial of some primitive element $\alpha \in \mathbf{F}_q$. The following result (see [15, Theorem 2.47], for example) is useful.

Lemma 5. *Let n be a positive integer relatively prime to a prime power q, and let d be the order of q modulo n. Since the cyclotomic polynomial Φ_n has coefficients in $\mathbb{Z}$, it can be viewed as a polynomial in $\mathbf{F}_q[x]$, and as such, it has $\phi(n)/d$ distinct irreducible factors, each of which has degree d.*

From now on, we work over $\mathbf{F}_2$, and we observe that if n is an odd integer and 2 is a primitive root modulo n, then Lemma 5 guarantees that the cyclotomic polynomial Φ_n is irreducible. For example, $1 + x + x^2 + x^3 + x^4 = \Phi_5(x)$ and $1 + x^3 + x^6 = \Phi_9(x)$ are both irreducible over $\mathbf{F}_2$.

Trinomials

A huge collection of explicit trinomials $x^n + x^m + 1$ that are primitive and irreducible over $\mathbf{F}_2$ is available; see [1] and [15, Section 3.5]. Consider $x^n + x + 1$. Some computations suggest that it is often irreducible (but certainly not for every value of n). It has the following useful property, however.

Lemma 6. *For all $n \ge 2$, the polynomial $x^n + x + 1$ has no repeated factors over $\mathbf{F}_2$.*

Proof A polynomial has repeated factors if and only if it shares a common factor with its formal derivative. If n is even and $f(x) = x^n + x + 1$, then $f'(x) = 1$, and so f' has no common factor with f. If n is odd, then $f'(x) = x^{n-1} + 1$. If r denotes a root of $f'(x)$ (in some splitting field), we have $r^{n-1} = 1$, so $r^n = r$, and thus $f(r) = r^n + r + 1 = 1$. It follows that r is not a root of f, so f and f' cannot share any common factors. □

Fourier Analysis over $(\mathbf{F}_2)^d$

Let $(\mathbf{F}_2)^d$ be the abelian group of length-d binary vectors under coordinatewise addition. The characters of $(\mathbf{F}_2)^d$ are indexed by $\beta \in (\mathbf{F}_2)^d$:

$$\chi_\beta(\alpha) = (-1)^{\alpha \cdot \beta},$$

where $\alpha \cdot \beta$ denotes the number of coordinates i for which $\alpha_i = \beta_i = 1$ (alternatively, it can be thought of as a dot product over $\mathbf{F}_2$).

If $Q(\alpha)$ is a probability distribution on $(\mathbf{F}_2)^d$ (or more generally, any function $(\mathbf{F}_2)^d \to \mathbf{C}$), its Fourier transform at $\beta \in (\mathbf{F}_2)^d$ is

$$\widehat{Q}(\beta) = \sum_{\alpha \in (\mathbf{F}_2)^d} Q(\alpha)(-1)^{\alpha \cdot \beta}.$$

It is easy to see that $\widehat{Q}(0) = 1$. The uniform distribution $U(\alpha) = 1/2^d$ for all $\alpha \in (\mathbf{F}_2)^d$ has the Fourier transform

$$\widehat{U}(0) = 1, \quad \widehat{U}(\alpha) = 0, \ \alpha \neq 0.$$

The convolution of two probabilities Q_1, Q_2 is

$$(Q_1 * Q_2)(\alpha) = \sum_{\gamma} Q_1(\gamma) Q_2(\alpha + \gamma).$$

Note that if X_1 and X_2 are independent random variables in $(\mathbf{F}_2)^d$ with distributions Q_1 and Q_2 respectively, then $X_1 + X_2$ has $Q_1 * Q_2$ as its distribution. The Fourier transform turns convolution into product, with

$$\widehat{Q_1 * Q_2}(\beta) = \widehat{Q}_1(\beta)\widehat{Q}_2(\beta).$$

The measure Q can be recovered from its Fourier transform via the inversion formula

$$Q(\alpha) = \frac{1}{2^d} \sum_{\beta} (-1)^{\alpha \cdot \beta} \widehat{Q}(\beta).$$

Finally, Plancherel's theorem relates the L^2 norm of Q with $\widehat{Q}$ and states that

$$2^d \sum_{\alpha \in (\mathbf{F}_2)^d} |Q(\alpha)|^2 = \sum_{\beta \in (\mathbf{F}_2)^d} \left|\widehat{Q}(\beta)\right|^2.$$

The following upper bound lemma is the key to establishing the upper bound in Theorem 1. It is a direct consequence of Plancherel's theorem.

Lemma 7. *Let $Q(\alpha)$ be a probability on $(\mathbf{F}_2)^d$ and let $U(\alpha)$ be the uniform distribution. Then*

$$4\|Q - U\|_{TV}^2 \leq 2^d \sum_{\alpha} (Q(\alpha) - U(\alpha))^2 = \sum_{\beta \neq 0} |\widehat{Q}(\beta)|^2.$$

Proof The inequality follows by Cauchy–Schwarz, and the equality follows from Plancherel's theorem and the fact that $\widehat{U}(0) = \widehat{Q}(0) = 1$ and $\widehat{U}(\alpha) = 0$ for $\alpha \neq 0$.

To set up the application of Lemma 7 to the proof of Theorem 1, let $e_1, \ldots, e_d$ be the standard basis for $(\mathbf{F}_2)^d$. Let

$$Q(\alpha) = \begin{cases} \frac{1}{2}, & \alpha = 0, \\ \frac{1}{2d}, & \alpha = e_i, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\widehat{Q}(\beta) = \sum_{\alpha} Q(\alpha)(-1)^{\alpha \cdot \beta} = \frac{1}{2} + \frac{1}{2d} \sum_{i=1}^{d} (-1)^{\beta_i} = 1 - \frac{|\beta|}{d},$$

where $|\beta|$ denotes the number of nonzero entries in β (with respect to the standard basis).

Let $A : (\mathbf{F}_2)^d \to (\mathbf{F}_2)^d$ be a linear map, and consider the Markov chain starting from $X_0 = 0$, and

$$X_n = AX_{n-1} + \varepsilon_n, \tag{4}$$

with $\mathrm{P}(\varepsilon_n = \alpha) = Q(\alpha)$ for all $\alpha \in (\mathbf{F}_2)^d$ and the ε_n independent. Iterating yields $X_0 = 0$, $X_1 = \varepsilon_1$, $X_2 = A\varepsilon_1 + \varepsilon_2$, and so on, and so

$$X_n = A^{n-1}\varepsilon_1 + A^{n-2}\varepsilon_2 + \cdots + \varepsilon_n. \tag{5}$$

Since this is a sum of independent random variables, if $Q_n(\alpha) = \mathrm{P}(X_n = \alpha)$, then

$$\widehat{Q}_n(\beta) = \prod_{j=0}^{n-1} \left(1 - \frac{|(A^t)^j \beta|}{d}\right). \tag{6}$$

In our application, A will be the matrix of squaring (which is linear in characteristic 2), $A^d = I$, and the product becomes tractable.

In [6, 7], this technique was used on $(\mathbf{F}_2)^d$ with

$$A = \begin{pmatrix} 1 & & & \\ 1 & 1 & & \\ & \ddots & \ddots & \\ & & 1 & 1 \end{pmatrix}$$

(1's along the diagonal and lower subdiagonal and 0 otherwise) to get sharp results. See [5] for applications to nonabelian groups.

The following proposition shows that adding deterministic mixing in this situation cannot slow things down. It gives one way of proving the upper bound in Theorem 1.

Proposition 8. *Let $A : (\mathbf{F}_2)^d \to (\mathbf{F}_2)^d$ be an invertible linear map, and consider the walk (4). Let P_n be the walk $X_n = X_{n-1} + \varepsilon_n$ without applying A. Then*

$$\|Q_n - U\|_2^2 \leq \|P_n - U\|_2^2,$$

where the L^2 norm is defined by

$$\|P - Q\|_2^2 = 2^d \sum_{\alpha \in (\mathbf{F}_2)^d} |P(\alpha) - Q(\alpha)|^2.$$

Proof Note that

$$\begin{aligned} \|Q_n - U\|_2^2 &= \sum_{\beta \neq 0} \prod_{j=0}^{n-1} \left(1 - \frac{|(A^t)^j \beta|}{d}\right) \\ &\leq \sum_{\beta \neq 0} \prod_{j=0}^{n-1} \left(1 - \frac{|\beta|}{d}\right) = \|P_n - U\|_2^2, \end{aligned}$$

where the middle inequality is an application of the rearrangement inequality [17], noting that an invertible linear map acts as a permutation on the nonzero elements of $(\mathbf{F}_2)^d$ and all factors are nonnegative. □

Remark 9. Proposition 8 says that applying a deterministic bijection between steps of the random walk on the hypercube cannot slow the mixing of the Markov chain (at least in an L^2 sense). While this is not very helpful if the resulting chain is supposed to mix faster, squaring fails to speed up the mixing (see Remark 13), and so Proposition 8 gives one way of proving the upper bound in Theorem 1.

Proof of Theorem 1

Throughout this section, p is a prime such that 2 is a primitive root in $\mathbf{F}_p$, and $d = p - 1$. By Lemma 5, the cyclotomic polynomial $\Phi_p(x) = 1 + x + \cdots + x^d$ is irreducible over $\mathbf{F}_2$. Represent $\mathbf{F}_{2^d} \cong \mathbf{F}_2[x]/(\Phi_p)$. The random walk defined by (1) with basis (3) can be represented as (4) with the basis $e_i = x^{i-1}$, with A the matrix of squaring with respect to this basis. We will index the rows and columns of matrices starting from 0 rather than 1, to match the exponents in the powers of x.

Example 10. Consider the case $p = 5$. The matrix A representing the linear map $x \mapsto x^2$ on $\mathbf{F}_{16}$ (viewed as an $\mathbf{F}_2$-vector space) with respect to the standard basis $1, x, x^2, x^3$ is

$$A = \begin{array}{c} \\ 1 \\ x \\ x^2 \\ x^3 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & x & x^2 & x^3 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array},$$

$$A^2 = \begin{array}{c} \\ 1 \\ x \\ x^2 \\ x^3 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & x & x^2 & x^3 \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \end{array},$$

$$A^3 = \begin{array}{c} \\ 1 \\ x \\ x^2 \\ x^3 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & x & x^2 & x^3 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array},$$

and $A^4 = I$. Note that A^j is a permutation matrix with one column replaced by a column of all ones. If this column is j^*, then $j^* = (p-1)/2^j$. The following result shows that this holds for all primes p such that 2 is a primitive root in $\mathbf{F}_p$.

Proposition 11. *Suppose that $\Phi_p(x) = 1 + x + \cdots + x^d$ is irreducible over $\mathbf{F}_2$. Then the matrix A^j, $1 \leq j \leq d-1$, of squaring j times, with respect to the basis $1, x, \ldots, x^{d-1}$, is a permutation matrix in which the column $j^* = (p-1)/2^j$ (starting the indexing from 0) is replaced by all ones.*

Proof Note that since $x^p - 1 = (x-1)(x^{p-1} + \cdots + 1) = 0$ in $\mathbf{F}_{2^d}$, it follows that $x^i = x^j$ if $i = j \pmod{p}$. The matrix A^j of squaring j times sends x^i to $x^{2^j i}$ for all i.

Since 2 is a primitive root modulo p, we have that as j goes from 1 to $p-2$, 2^j runs over all elements of $\mathbf{F}_p^\times$ except 1. If $2^j i = p - 1 \pmod{p}$, then $x^{2^j i} = x^{p-1} + \cdots + 1$, and otherwise, it is equal to some x^k with $1 \leq k \leq p-2$. This means that each column except j^* has exactly one nonzero entry, which is 1. Moreover, since 2^j is invertible modulo p, all rows can have at most one nonzero entry off the column j^*. □

Next, consider (5) with $n = dm$ for some positive integer m. From (6), we have

$$\widehat{Q}_n(\beta) = \prod_{j=0}^{d-1} \left(1 - \frac{|(A^t)^j \beta|}{d}\right)^m. \tag{7}$$

The next result determines these values.

Proposition 12. *Let β_i denote the coefficient of x^i in β. The Fourier transform of the square-and-add Markov chain, (7), after $n = dm$ steps satisfies $\widehat{Q}_n(\beta) = \widehat{Q}_d(\beta)^m$ and*

$$\widehat{Q}_d(\beta) = \begin{cases} \left(1 - \frac{|\beta|}{d}\right)^{d-|\beta|} \left(1 - \frac{|\beta|-1}{d}\right)^{|\beta|}, & |\beta| \text{ is even, } \beta_0 = 0, \\ \left(1 - \frac{|\beta|}{d}\right)^{|\beta|+1} \left(1 - \frac{|\beta|+1}{d}\right)^{d-|\beta|-1}, & |\beta| \text{ is odd, } \beta_0 = 0, \\ \left(1 - \frac{|\beta|}{d}\right)^{d-|\beta|+1} \left(1 - \frac{|\beta|-1}{d}\right)^{|\beta|-1}, & |\beta| \text{ is even, } \beta_0 = 1, \\ \left(1 - \frac{|\beta|}{d}\right)^{|\beta|} \left(1 - \frac{|\beta|+1}{d}\right)^{d-|\beta|}, & |\beta| \text{ is odd, } \beta_0 = 1. \end{cases}$$

Proof The key point is that the matrix A^j is a permutation matrix except for one column of all ones. The all-ones column j^* occurs exactly once in the positions $1, 2, \ldots, d-1$ as j varies in $\{1, 2, \ldots, d-1\}$. The argument then follows by considering the four separate cases.

For example, when $|\beta|$ is even and $\beta_0 = 0$, there are exactly $|\beta|$ nonzero entries in the vector β among the coefficients of $x, \ldots, x^{d-1}$. When j^* is among the indices where β is nonzero, $(A^t)^j \beta$ has one fewer nonzero entry (since one of the 1's was replaced by $\beta \cdot (1, \ldots, 1) = 0$). This occurs exactly $|\beta|$ times. Otherwise, the number of nonzero entries remains the same. This gives the desired expression.

The other cases are similar. □

Proof of Theorem 1 From the upper bound lemma (Lemma 7), for $n = dm$, we have

$$2^d \sum_{\alpha \in \mathbf{F}_{2^d}} |Q_n(\alpha) - U(\alpha)|^2 = \sum_{\beta \neq 0} \widehat{Q}_d(\beta)^{2m}. \tag{8}$$

For the four cases in Proposition 12, the sum in (8) breaks into four sums:

$$\Sigma_I = \sum_{j\text{even}} \left(1-\frac{j}{d}\right)^{2m(d-j)} \left(1-\frac{j-1}{d}\right)^{2mj} \binom{d-1}{j},$$
$$\Sigma_{II} = \sum_{j\text{odd}} \left(1-\frac{j}{d}\right)^{2m(j+1)} \left(1-\frac{j+1}{d}\right)^{2m(d-j-1)} \binom{d-1}{j},$$
$$\Sigma_{III} = \sum_{j\text{even}} \left(1-\frac{j}{d}\right)^{2m(d-j+1)} \left(1-\frac{j-1}{d}\right)^{2m(j-1)} \binom{d-1}{j-1},$$
$$\Sigma_{IV} = \sum_{j\text{odd}} \left(1-\frac{j}{d}\right)^{2mj} \left(1-\frac{j+1}{d}\right)^{2m(d-j)} \binom{d-1}{j-1}. \tag{9}$$

Let us use the expressions in (9) to prove an L^2 lower bound. Because of the equality (8), the L^2 norm is bounded below by any single term. Choose $j = 2$ in Σ_I. This is

$$\begin{aligned}&\left(1-\frac{2}{d}\right)^{2m(d-2)} \left(1-\frac{1}{d}\right)^{4m} \binom{d-1}{2} \\ &= \left(1+\frac{1}{d-2}\right)^{4m} \left(1-\frac{2}{d}\right)^{2md} \binom{d-1}{2}.\end{aligned} \tag{10}$$

Choose $m = \frac{1}{2}(\log(d) - c)$. For d large, we have

$$\left(1+\frac{1}{d-2}\right)^{4m} = 1 + o(1),$$
$$\left(1-\frac{2}{d}\right)^{2md} \sim e^{-4m} = d^{-2}e^{2c},$$
$$\binom{d-1}{2} \sim \frac{d^2}{2}.$$

Thus, the right-hand side of (10) is asymptotic to $e^{2c}/2$. It follows that Q_n is exponentially far from uniform if $n = d(\log(d) - c)/2$. A similar argument shows, for this n, that the total variation distance to uniform is exponentially close to 1; this uses the (available) second moment method; see [14, Proposition 7.14].

We now proceed to the upper bound. By the upper bound lemma and Proposition 8,

$$4\|Q_n - U\|_{TV}^2 \leq \|P_n - U\|_2^2,$$

where P_n is the distribution of the random walk $X_n = X_{n-1} + \varepsilon_n$ on $(\mathbf{F}_2)^d$ after n steps. It is known (see [5], for example) that if $n = \frac{1}{2}d(\log(d) + c)$, then

$$\|P_n - U\|_2^2 \leq e^{e^{-c}} - 1,$$

and $e^{e^{-c}} - 1$ goes to zero like e^{-c} when c is large, which gives the desired upper bound. □

Remark 13. Note that the random walk $X_n = X_{n-1} + \varepsilon_n$ on $(\mathbf{F}_2)^d$ without squaring also takes $\frac{1}{2}d(\log(d) + c)$ steps to equilibrate. Thus in this case, squaring does not introduce a dramatic speedup.

Remark 14. The upper bound can also be proved directly from Proposition 12. These more detailed calculations yield essentially the same answers as the rearrangement bounds.

Remark 15. All the arguments given when $p = 2$ extend to the case of a general prime p, with squaring replaced by taking the pth power. An upper bound on the L^2 distance needed to apply Proposition 8 can be found in [4], which would show that the Markov chain mixes after order-$p^2 d \log(d)$ steps.

Back to Squaring and Adding on $\mathbf{F}_p$

We return to our motivating problem

$$X_n = X_{n-1}^2 + \varepsilon_n \pmod{p}, \tag{11}$$

where p is a prime and ε_n is 1 or -1, independently, with probability 1/2. To showcase the difference, consider the following problem: what is the stationary distribution of this Markov chain? Call this stationary distribution π_p.

A look at the data shows that for $p \geq 7$, there are many j with $\pi_p(j) = 0$ such that for some p, the nonzero $\pi_p(j)$ vary wildly in magnitude, while for other p, $\pi_p(j)$ is roughly uniform.

The data below are normalized so that $\tilde{\pi}_p$ is the left eigenvector for the eigenvalue 1, scaled so all entries are integers.

Example 16 ($p = 29$)

$$\begin{aligned}\tilde{\pi}_{29} = (&4,2,2,2,0,8,2,6,7,0,5,0,4,0,4,0,0,0,0,\\ &3,0,5,0,2,8,0,8,2,2).\end{aligned}$$

Example 17 ($p = 31$)

$$\begin{aligned}\tilde{\pi}_{31} = (&2,3,2,4,2,2,4,2,4,4,2,2,0,2,0,4,0,4,2,\\ &4,2,2,0,0,2,0,2,2,0,2,1).\end{aligned}$$

Here (and in fact, for all $p = 3 \pmod 4$; see Theorem 20), the smallest nonzero entry is 1, the largest is 4, and 1 and 3 appear only once.

Example 18 ($p = 101$)

$$\begin{aligned}\widetilde{\pi}_{101} = (&66056, 33028, 33028, 33028, 0, 33028, 0, 0, 48868,\\ &0, 48868, 0, 7376, 48200, 7376, 62952, 21038,\\ &14752, 21038, 0, 32951, 0, 68115, 0, 85876, 0,\\ &50712, 0, 0, 16514, 0, 16514, 34236, 0, 34236,\\ &14752, 0, 14752, 0, 0, 0, 0, 3688, 0, 34700, 0,\\ &32856, 0, 3688, 0, 1844, 34236, 0, 53012, 0, 26152,\\ &0, 7376, 0, 0, 0, 0, 0, 33028, 0, 33028, 0, 27788,\\ &0, 62164, 51958, 34376, 51958, 0, 0, 18040, 0, 18040,\\ &0, 68115, 0, 96465, 0, 44864, 0, 16514, 7376, 0, 7376,\\ &0, 0, 17188, 0, 17188, 3688, 29504, 68396,\\ &29504, 64708, 33028, 33028).\end{aligned}$$

Here, the ratio of the largest to smallest nonzero entries is large ($\max / \min \approx 52$). There appears to be unbounded fluctuation for larger p with $p = 1 \pmod 4$.

Example 19 ($p = 103$)

$$\begin{aligned}\widetilde{\pi}_{103} = (&2, 3, 2, 4, 0, 2, 2, 2, 4, 2, 2, 0, 2, 2, 4, 4, 4, 4, 4,\\ &2, 2, 0, 2, 0, 4, 2, 2, 4, 2, 4, 2, 4, 2, 4, 2, 4, 0,\\ &4, 0, 2, 2, 0, 2, 0, 0, 2, 0, 2, 2, 2, 2, 4, 0, 2, 2, 2,\\ &2, 4, 2, 4, 4, 2, 4, 2, 2, 4, 0, 4, 0, 2, 0, 2, 0, 2, 0,\\ &2, 0, 2, 2, 0, 4, 2, 4, 2, 2, 0, 0, 0, 0, 0, 2, 2, 4, 2,\\ &2, 0, 2, 2, 2, 4, 0, 2, 1).\end{aligned}$$

In all cases we looked at, the Markov chain was ergodic (had a unique eigenvector with eigenvalue 1). We are unable to prove this in general.

There is some sense to be made. Observe that if j has both $j - 1$ and $j + 1$ nonsquares modulo p, then $\pi_p(j) = 0$. Classical number theory (see [13, Chapter 5, Exercise 8], for example) shows that asymptotically, this accounts for a quarter of all j. This matches the data when $p = 3 \pmod 4$. For example, when $p = 103$, then $\pi_{103}(j) = 0$ for 25 values of j. However, when $p = 1 \pmod 4$, there are further forced zeros, with $\pi_{101}(j) = 0$ for 44 values of j.

Ron Graham and Steve Butler observed the following:

- When $p = 3 \pmod 4$, these $j \pm 1$ nonresidues exactly match the zeros (for all $p \leq 10\,000$).
- When $p = 1 \pmod 4$, the proportion of zeros appears to be converging to approximately 42%.

We record one further piece of mathematical progress, which explains the first point.

Theorem 20 (He, [10]) If $p = 3 \pmod 4$, then the square-and-add Markov chain (11) is irreducible, aperiodic, and has a unique stationary distribution given by

$$\pi_p(j) = \frac{\left|\left\{k \in \mathbf{F}_p \mid k^2 \pm 1 = j\right\}\right|}{2p}.$$

Acknowledgments

We thank Amol Aggarwal, Steve Butler, Ron Graham, Bob Guralnick, David Kazhdan, Laurent Miclo, Yuval Peres, and Kannan Soundararajan for their help. The first author was partially supported by NSF grant DMS 1954042. The second author was partially supported by NSERC.

Persi Diaconis
Department of Mathematics
Stanford University
Stanford, CA 94305
USA
e-mail: diaconis@math.stanford.edu

Jimmy He
Department of Mathematics
Stanford University
Stanford, CA 94305
USA
e-mail: jimmyhe@stanford.edu

I. Martin Isaacs
Department of Mathematics
University of Wisconsin
Madison, WI 53706
USA
e-mail: isaacs@math.wisc.edu

References

[1] Richard P. Brent and Paul Zimmermann. The great trinomial hunt. *Notices Amer. Math. Soc.* 58:2 (2011), 233–239.

[2] Sourav Chatterjee and Persi Diaconis. Speeding up Markov chains with deterministic jumps. *Probab. Theory Related Fields* 178:3-4 (2020), 1193–1214.

[3] Fan R. K. Chung, Persi Diaconis, and Ron L. Graham. Random walks arising in random number generation. *Ann. Probab.* 15:3 (1987), 1148–1165.

[4] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* 6:3 (1996), 695–750.

[5] Persi Diaconis. *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics Lecture Notes, Monograph Series, vol. 11, Institute of Mathematical Statistics, Hayward, CA, 1988.

[6] Persi Diaconis and Ron Graham. An affine walk on the hypercube. *J. Comput. Appl. Math.* 41:1–2 (1992), 215–235.

[7] Persi Diaconis and Ron Graham. Binomial coefficient codes over GF(2). *Discrete Math.* 106/107 (1992), 181–188.

[8] Sean Eberhard and Péter P. Varjú. Mixing time of the Chung–Diaconis–Graham random process. *Probab. Relat. Fields* 179 (2021), 317–344.

[9] Ben Green. Finite field models in additive combinatorics. In *Surveys in Combinatorics 2005*, London Math. Soc. Lecture Note Ser., vol. 327, pp. 1–27. Cambridge Univ. Press, 2005.

[10] Jimmy He. Markov chains on finite fields with deterministic jumps. arXiv:2010.10668 [math.PR], 2020.

[11] Lenwood S. Heath and Nicholas A. Loehr. New algorithms for generating Conway polynomials over finite fields. *J. Symbolic Comput.* 38:2 (2004), 1003–1024.

[12] Christopher Hooley. On Artin's conjecture. *J. Reine Angew. Math.* 225 (1967), 209–220.

[13] Kenneth Ireland and Michael Rosen. *A Classical Introduction to Modern Number Theory*, second ed., Graduate Texts in Mathematics, vol. 84. Springer, 1990.

[14] David A. Levin and Yuval Peres. *Markov Chains and Mixing Times*, second edition. American Mathematical Society, 2017.

[15] Rudolf Lidl and Harald Niederreiter. *Finite Fields*, second ed., Encyclopedia of Mathematics and Its Applications, vol. 20. Cambridge University Press, 1997.

[16] Frank Lübeck, Conway polynomials for finite fields. Available online at http://www.math.rwth-aachen.de/~Frank.Luebeck/data/ConwayPol/index.html.

[17] Harry D. Ruderman. Two new inequalities. *Amer. Math. Monthly* 59 (1952), 29–32.

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