Markov chains on finite fields with deterministic jumps[*]

Jimmy He[†]

Abstract

We study the Markov chain on $\mathbf{F}_p$ obtained by applying a function f and adding $\pm\gamma$ with equal probability. When f is a linear function, this is the well-studied Chung–Diaconis–Graham process. We consider two cases: when f is the extension of a non-linear rational function which is bijective, and when $f(x) = x^2$. In the latter case, the stationary distribution is not uniform and we characterize it when $p = 3 \pmod 4$. In both cases, we give an almost linear bound on the mixing time, showing that the deterministic function dramatically speeds up mixing. The proofs involve establishing bounds on exponential sums over the union of short intervals.

Keywords: Markov chain; mixing time; Cheeger constant; spectral gap.
MSC2020 subject classifications: Primary 60J10, Secondary 11T23; 05C81.
Submitted to EJP on June 24, 2021, final version accepted on February 14, 2022.

1 Introduction

Recent work of Chatterjee and Diaconis studied how Markov chains can be sped up by interspersing a deterministic function between steps of the Markov chain [5]. For "most" bijective functions, the mixing time becomes logarithmic in the size of the state space. They asked for concrete examples where this speedup takes place. Currently, one of the only known examples is the Chung–Diaconis–Graham process, which is a Markov chain on the finite field $\mathbf{F}_p$ given by

$$X_{n+1} = aX_n + \varepsilon_{n+1}, \tag{1.1}$$

where the ε_n are uniform on $\{-1, 0, 1\}$ and independent. While the walk $X_{n+1} = X_n + \varepsilon_{n+1}$ mixes in order p^2 steps, for certain $a \in \mathbf{F}_p$ and almost all p, the walk defined by (1.1) mixes in order $\log(p)$ steps [11].

The purpose of this paper is to study the mixing times of non-linear analogues of (1.1). In particular, for bijections which are extensions of rational functions on $\mathbf{F}_p$, we

---

[†]Department of Mathematics, MIT, Cambridge, MA. E-mail: jimmyhe@mit.edu

are able to show that the mixing time is faster than $p^{1+\varepsilon}$ for any $\varepsilon > 0$. This provides more examples of Markov chains which mix faster after adding deterministic jumps. We also consider the case when $f(x) = x^2$. We consider a slightly different version of the Markov chains considered in [5], and so we also explain how to apply our methods to the Markov chains appearing there.

1.1 Main results

Let us now define the class of functions f that we will work with. For a prime p and $d \in \mathbf{N}$, let $\mathcal{B}(p,d)$ denote the set of functions $f : \mathbf{F}_p \to \mathbf{F}_p$ which are bijections, and for which there exist coprime $P, Q \in \mathbf{F}_p[x]$ polynomials of degree at most d such that $f = P/Q$ except at the zeroes of Q, and such that P/Q is not a linear or constant function.

Remark 1.1. Some examples of functions f to keep in mind are permutation polynomials, which are polynomials $P \in \mathbf{F}_p[x]$ that are also bijections, and the function $f(x) = x^{-1}$ for $x \neq 0$ and $f(0) = 0$. A specific family of permutation polynomials is given by $f(x) = x^3$ for primes $p > 3$ with $(3, p - 1) = 1$.

Theorem 1.2. *Let p be a prime and $d \in \mathbf{N}$. Let $f \in \mathcal{B}(p,d)$, and let $\gamma \in \mathbf{F}_p$ be non-zero. Let ε_i denote independent random variables uniform on $\pm\gamma$. Consider the lazy Markov chain defined by*

$$X_{n+1} = \begin{cases} f(X_n) + \varepsilon_{n+1} & \text{with probability } 1/2, \\ X_n & \text{with probability } 1/2. \end{cases}$$

Then if π denotes the uniform distribution on $\mathbf{F}_p$, for any $\varepsilon > 0$, there is some constant $C = C(\varepsilon, d)$ such that

$$\sup_{X_0 \in \mathbf{F}_p} \| \mathrm{P}(X_n \in \cdot) - \pi \|_{TV} \leq e^{-C \frac{n}{p^{1+\varepsilon}}}.$$

Remark 1.3. The *mixing time* $t_{mix}(\varepsilon)$ of an ergodic Markov chain P is the minimum n such that

$$\sup_{X_0 \in \mathbf{F}_p} \| \mathrm{P}(X_n \in \cdot) - \pi \|_{TV} \leq \varepsilon,$$

where π is the stationary distribution. The total variation bound in Theorem 1.2 (and the corresponding bounds in Theorems 1.8 and 4.5) imply that $t_{mix}(\varepsilon)$ is $O(p^{1+\delta})$ for any $\delta > 0$, assuming that f is the extension of rational functions of bounded degree.

We also remark that the $O(p^{1+\delta})$ could be improved to $O(p \log^C p)$ for some large constant C (probably $C = 100$ would work). Since the bound we obtain is most likely far from optimal (see Remark 1.6), we do not bother to explicitly state the stronger bound.

Remark 1.4. The assumption that the ε_i are uniform on $\pm\gamma$ is easily relaxed. By considering the bijection $f + \gamma_0$, we can choose the ε_i to be supported on any two point set. Since the bound from Cheeger's inequality can only get worse as edges are removed, this extends to any distribution μ for the ε_i as long as μ is supported on at least two points, with the constant depending on $\min_{\gamma \in \text{supp} \, \mu} \mu(\gamma)$.

The laziness can also be removed at the cost of some assumptions on the distribution of the ε_i, see Theorem 4.5.

Remark 1.5. This random walk is a slight modification of the walks considered in [5], which always applies the function f, but possibly adds 0 instead of $\pm\gamma$. A small modification to the argument allows a similar result to be established in this setting as well, see Theorem 4.5. This gives a large class of examples where adding a deterministic bijection dramatically speeds up the mixing time. Finding such examples was a question raised in [5].

Remark 1.6. By an entropy argument, it's easy to see that at least order $\log(p)$ steps are necessary in order for the random walk to be close to uniform. This leaves a large gap between the upper and lower bounds and determining the correct order of the mixing time seems to be a challenging problem.

The proof of Theorem 1.2 involves applying a version of Cheeger's inequality for directed graphs to reduce the problem to giving a lower bound for the number of solutions to $y = f(x) \pm \gamma$ with $x \in S$ and $y \in S^c$. This bound is obtained using some ideas from analytic number theory, namely exponential sum bounds, with a key ingredient being the Weil bound. While applications of the Weil bound have previously appeared in the study of expanders (see [7] for example), our approach does not require the eigenvalues to be exactly identified. This provides a more flexible approach, but gives a weaker bound.

The main technical tools are exponential sum bounds of the form

$$\sum_{j=1}^{J} \left| \sum_{x \in I_j} \exp(2\pi i k f(x)/p) \right| \leq C p \log^2(L), \tag{1.2}$$

where the I_j are disjoint intervals of length L, and C is some constant depending on f, along with bounds for linear exponential sums. These bounds appear to be new, although the proof of (1.2) closely follows ideas of Browning and Haynes [4], who proved the bound when $f(x) = 1/x$.

Finally, we also give some results for the square-and-add Markov chain, which is the case when $f(x) = x^2$. This is obviously not a bijection, but when $p = 3 \pmod 4$, the stationary distribution can be determined and a similar mixing time bound can be established.

Theorem 1.7. *Let p be a prime, $p = 3 \pmod 4$, and let $\gamma \in \mathbf{F}_p$ be non-zero. Let ε_i denote independent random variables uniform on $\pm\gamma$. Define a Markov chain on $\mathbf{F}_p$ by $X_{n+1} = X_n^2 + \varepsilon_{n+1}$. Then the chain has a unique recurrent communicating class, and thus a unique stationary distribution. The stationary distribution is given by*

$$\pi(\alpha) = \frac{|\{\beta \in \mathbf{F}_p \mid \beta^2 \pm \gamma = \alpha\}|}{2p}.$$

Furthermore, the Markov chain is aperiodic if $\gamma = 1$.

Theorem 1.8. *Let p be a prime, $p = 3 \pmod 4$, and let $\gamma \in \mathbf{F}_p$ be non-zero. Let ε_i denote independent random variables uniform on $\pm\gamma$. Consider the lazy Markov chain defined by*

$$X_{n+1} = \begin{cases} X_n^2 + \varepsilon_{n+1} & \text{with probability } 1/2, \\ X_n & \text{with probability } 1/2. \end{cases}$$

Then if π denotes the stationary distribution on $\mathbf{F}_p$, for any ε, there is some constant $C = C(\varepsilon)$ such that

$$\sup_{X_0 \in \mathbf{F}_p} \| \mathrm{P}(X_n \in \cdot) - \pi \|_{TV} \leq e^{-C \frac{n}{p^{1+\varepsilon}}}.$$

These results partially answer some questions raised in [10]. The case of $p = 1 \pmod 4$ seems to be more mysterious and even the stationary distribution is not well-understood. Some heuristics and a conjecture are given in Section 2.2.

1.2 Related work

The Markov chains considered in this paper can be viewed as a non-linear analogue of the Chung–Diaconis–Graham process, defined by $X_{n+1} = aX_n + \varepsilon_{n+1}$ on $\mathbf{F}_p$ (or more

generally on $\mathbf{Z}/(p)$ for composite p). This was first introduced in [8], and studied in [15, 16, 20, 17, 3]. Recent work of Eberhard and Varjú established cutoff for this Markov chain for many values of a [11]. Specifically, they show that when $a = 2$, the Markov chain has cutoff at $c \log_2 p$ for an explicit $c \approx 1.01136$ for almost all p, and extend this to many other values of a. These works all rely heavily on the linear nature of the problem.

Chatterjee and Diaconis studied how deterministic jumps can increase the speed of convergence for Markov chains [5]. They showed that for a state space of size n, applying a (fixed) random bijection gives a mixing time of order $\log(n)$ with high probability, but noted that specific examples were hard to find. The random walks treated in this paper provide some progress in this direction. This fits into a broader theme of finding ways to accelerate the convergence of Markov chains to their stationary distribution [14, 2, 1].

The Markov chain when $f(x) = x^2$ studied in this paper can be viewed as a random version of a discrete dynamical system on $\mathbf{F}_p$. The dynamical system defined by iterating the map $x \mapsto x^2 + c$ has been well-studied although many questions remain, see [21, 22, 18].

The exponential sum bounds are obtained using arguments that closely follows that of [4], who considered the case when $f(x) = 1/x$. Their work is based on an argument of Heath-Brown [13]. These arguments work for any finite field, not just prime fields. It would be interesting to see if these ideas could be useful in studying the mixing times of random walks on $\mathbf{F}_q$, for q a prime power.

1.3 Outline

Section 2 establishes some basic properties of the square-and-add Markov chain, including the proof of Theorem 1.7. In Section 3, we prove the exponential sum bounds needed, which are the main technical tool. Finally, in Section 4, we apply the bounds from Section 3 to prove Theorem 1.2, and explain how to extend the arguments to establish Theorem 1.8 and to study some of the random walks defined in [5].

1.4 Notation

Let $e_p(x) = \exp(2\pi i x/p)$. We will use C to denote a constant that may change from line to line.

2 The square-and-add Markov chain

In this section, we establish some basic properties of the Markov chain when $f(x) = x^2$. The reason this case is more involved is that f is not a bijection, and so even the stationary distribution is not obvious. While the case of $p = 1 \pmod 4$ seems wild, surprisingly the case of $p = 3 \pmod 4$ is tractable.

2.1 The case $p = 3 \pmod 4$

The reason that the case of $p = 3 \pmod 4$ is much easier is that in this case, exactly one of α and $-\alpha$ is a quadratic residue mod p. Using this fact, we are able to prove all essential properties about the Markov chain.

Lemma 2.1. *A stationary distribution for the Markov chain defined in Theorem 1.7 is given by*

$$\pi(\alpha) = \frac{|\{\beta \in \mathbf{F}_p \mid \beta^2 \pm \gamma = \alpha\}|}{2p}.$$

Proof. The result can be shown by checking that

$$\sum_{\alpha^2 \pm \gamma = \beta} \frac{1}{2}\pi(\alpha) = \pi(\beta)$$

for all $\beta \in \mathbf{F}_p$, which is equivalent to showing that

$$\sum_{\alpha^2 \pm \gamma = \beta} |\{\rho \mid \rho^2 \pm \gamma = \alpha\}| = 2|\{\rho \mid \rho^2 \pm \gamma = \beta\}|. \tag{2.1}$$

Note that

$$|\{\rho \mid \rho^2 \pm \gamma = \alpha\}| + |\{\rho \mid \rho^2 \pm \gamma = -\alpha\}| = 4$$

because exactly one of $\alpha + \gamma$ and $-\alpha - \gamma$ is a quadratic residue, and similarly for $\alpha - \gamma$ and $-\alpha + \gamma$ (if $\alpha = \pm\gamma$, the equation also holds since the solution to $\rho^2 = 0$ is counted twice). Since the sum in (2.1) is over pairs $\pm\alpha$, this implies (2.1) holds for $\beta \neq \pm\gamma$.

If $\beta = \pm\gamma$, the only difference is there is an extra term $|\{\rho \mid \rho^2 = \pm\gamma\}| = 2$ in the left hand side of (2.1) corresponding to $\alpha = 0$ which is not paired. But of course this corresponds to the solution to $\rho^2 = 0$ on the right hand side of (2.1). $\qquad\square$

Lemma 2.2. *The Markov chain defined in Theorem 1.7 has a unique recurrent communicating class, and the restriction to this component is aperiodic if $\gamma = 1$.*

Proof. Any recurrent state must be of the form $\alpha^2 \pm \gamma$, and so must be contained in $\operatorname{supp} \pi$. Conversely, since π is a stationary distribution, every element of $\operatorname{supp} \pi$ is recurrent. It thus suffices to show that $\operatorname{supp} \pi$ contains a single communicating class. We do so as follows.

We claim for any $\alpha, \beta \in \operatorname{supp} \pi$, we can find a sequence of moves $\alpha \mapsto -\alpha$ and $\alpha \mapsto \alpha \pm 2\gamma$ taking us from α to β, while staying entirely inside $\operatorname{supp} \pi$. We'll then show that these moves stay within a single communicating class, from which we can conclude that $\operatorname{supp} \pi$ contains a single communicating class.

To see that we can go from α to β using these moves while staying in $\operatorname{supp} \pi$, we argue as follows. Either $\alpha, \alpha + 2\gamma \in \operatorname{supp} \pi$, or $-\alpha, -\alpha - 2\gamma \in \operatorname{supp} \pi$, as one of $\alpha + \gamma$ and $-\alpha - \gamma$ will be a quadratic residue. A similar statement holds for α and $\alpha - 2\gamma$. In particular, at least one of α or $-\alpha$ lies in $\operatorname{supp} \pi$. We thus partition $\mathbf{F}_p$ into sets $\{\pm\alpha\}$ and $\{0\}$, and note that each set contains an element of $\operatorname{supp} \pi$. Starting from $\alpha \in \operatorname{supp} \pi$, we can then move to an element in $\{\pm(\alpha + 2\gamma)\}$ while staying inside $\operatorname{supp} \pi$, by either moving directly from α to $\alpha + 2\gamma$ if it lies in $\operatorname{supp} \pi$, or moving from α to $-\alpha$, and then $-\alpha - 2\gamma$ which are both guaranteed to lie in $\operatorname{supp} \pi$ if $\alpha + 2\gamma$ does not. Similarly, we can move to an element in $\{\pm(\alpha - 2\gamma)\}$. In this way, we can reach an element in any set of the form $\{\pm(\alpha + 2k\gamma)\}$. But this exhausts $\mathbf{F}_p$, since 2γ is non-zero as $p \neq 2$, so we can reach either β or $-\beta$. If we reach $-\beta$, then as β by assumption also lies in $\operatorname{supp} \pi$, we can move from $-\beta$ to β.

Now we show the claims that if $\alpha, -\alpha \in \operatorname{supp} \pi$, then α and $-\alpha$ belong to the same communicating class, and that if $\alpha, \alpha + 2\gamma \in \operatorname{supp} \pi$, then they belong to the same communicating class.

Suppose that α and $-\alpha$ both lie in $\operatorname{supp} \pi$. Then as there is a path from α to $\alpha^2 + \gamma$, and $\alpha \in \operatorname{supp} \pi$ so it must be recurrent, there must be a path from $\alpha^2 + \gamma$ to α. But then there is a path from $-\alpha$ to $\alpha^2 + \gamma$, and then α. Thus, α and $-\alpha$ must belong to the same communicating class.

Now suppose that α and $\alpha + 2\gamma$ both lie in $\operatorname{supp} \pi$. If $\alpha + \gamma$ is a quadratic residue, then note that one of its square roots lies in $\operatorname{supp} \pi$. This square root must then be recurrent, and so from α, we must be able to reach it, after which we can move to $\alpha + 2\gamma$. If $\alpha + \gamma$ is not a quadratic residue, then $-\alpha - \gamma$ is a quadratic residue, and so both $-\alpha$ and $-\alpha - 2\gamma$ lie in $\operatorname{supp} \pi$ and by the same argument lie in the same communicating class. We can then use the fact that $-\alpha$ and α lie in the same communicating class, as well as $-\alpha - 2\gamma$ and $\alpha + 2\gamma$, to conclude that α and $\alpha + 2\gamma$ lie in the same communicating class.

To see that the random walk is aperiodic if $\gamma = 1$, first note that there is a cycle of length 2, namely $0 \to 1 \to 0$, and so it suffices to find an odd cycle. Now if a path from

$\alpha^2 + 1$ to α were of even length, we would immediately be done because then after taking the step $\alpha \to \alpha^2 + 1$, we would obtain an odd cycle, so assume otherwise. Then moving from $-\alpha$ to $\alpha^2 + 1$, and then back to α, is a path of even length. If $\alpha + 1$ and is a quadratic residue, we obtain an odd length path going from $\alpha + 2$ to $\sqrt{\alpha + 1}$ and then moving to α gives an even length path from $\alpha + 2$ to α. If $-\alpha - 1$ were a quadratic residue, we would obtain an even length path from $-\alpha - 2$ to $-\alpha$. We can similarly assume paths from $\alpha^2 - 1$ to α are odd length, and obtain even length paths from either $\alpha - 2$ to α or $-\alpha + 2$ to $-\alpha$. But then using these we can construct an even length path from $\alpha^2 + 1$ back to α, since we proved above that we can move from any element in $\operatorname{supp} \pi$ to any other using these moves. Thus, there must be an odd cycle. $\qquad\square$

Theorem 1.7 follows immediately from Lemmas 2.1 and 2.2.

Example 2.3. Consider the case $p = 11$ and $\gamma = 1$. Then the stationary distribution is given by

$$\pi = \left(\frac{2}{22}, \frac{1}{22}, \frac{4}{22}, \frac{2}{22}, \frac{4}{22}, \frac{2}{22}, \frac{2}{22}, \frac{0}{22}, \frac{2}{22}, \frac{0}{22}, \frac{3}{22} \right),$$

and it can be checked that the numerators are indeed given by counting solutions to $\beta^2 \pm 1 = \alpha$.

2.2 The case $p = 1 \pmod 4$

The case when $p = 1 \pmod 4$ is more wild, but there is some evidence to suggest that much of the behavior is similar to that of a random directed graph with certain degree constraints.

The following heuristic was suggested by Alex Cowan (private communication). For convenience, we work with the graph whose edges are given by (α, β) for $\beta = (\alpha \pm 1)^2$. Let π' denote the stationary distribution of this walk. It's clear that the stationary distribution of the original walk can be recovered from the new one by $\pi(x) = \pi'(x-1)/2 + \pi'(x+1)/2$.

The idea is that the degree distribution of the graph is determined, with about half the vertices having indegree 0 and half having indegree 4 (we just ignore 0, 1, and -1). Now the key observation is that the vertices with indegree 0 come in pairs, as if α is not a quadratic residue, then neither is $-\alpha$. This means after removing all vertices with indegree 0, if the edges were random subject to this constraint then we have $1/4$ of the vertices have indegree 0, $1/2$ have indegree 2 and $1/4$ have indegree 4. If we assume that the resulting graph is random, then by applying work of Cooper and Frieze [9], we can obtain a precise conjecture for the size of the support.

Conjecture 2.4. *Let ε_n be uniformly distributed on ± 1 and independent. Consider the Markov chain defined by $X_{n+1} = X_n^2 + \varepsilon_{n+1}$ on $\mathbf{F}_p$ with $p = 1 \pmod 4$. Then conjecturally it has a unique stationary distribution π, and*

$$\frac{|\operatorname{supp} \pi|}{p} \to 1 - \frac{1}{4}(1 + \alpha)^2,$$

where $\alpha \approx 0.2956$ is the smallest positive root of $x^4 + 2x^2 - 4x + 1$.

This conjecture agrees with computer computations done by Steve Butler, which suggests that asymptotically the support of π contains 58% of the elements in $\mathbf{F}_p$.

3 Counting solutions with exponential sum bounds

To obtain lower bounds on the Cheeger constant, after a reduction we will need to show that $f(x) \pm \gamma = y$ has many solutions for $x \in S$ and $y \in S'$, where S and S' are a disjoint union of arithmetic progressions. This will be shown following the argument of

Browning–Haynes [4], who actually proved the needed result in the case of $f(x) = x^{-1}$ (the sum was restricted to $\mathbf{F}_p^\times$ but this is unimportant). We give the modifications needed to generalize to the class of functions $\mathcal{B}(p, d)$ as well as the function $f(x) = x^2$. The key is the Weil bound for exponential sums. We use the following convenient formulation given in [12, Lemma 3].

Lemma 3.1. *Let $P, Q \in \mathbf{F}_p[x]$ be coprime polynomials such that $P(x)/Q(x)$ is not a constant function on $\mathbf{F}_p$. Then*

$$\left| \sum_{\substack{x \in \mathbf{F}_p \\ Q(x) \neq 0}} e_p(P(x)/Q(x)) \right| \leq C p^{1/2},$$

where the constant C depends only on the degree of P and Q.

We now give bounds for averaged exponential sums over intervals. A separate elementary argument is given when $f(x) = x^2$.

Lemma 3.2. *Let $f \in \mathcal{B}(p, d)$. Then for any interval I, and any $k \in \mathbf{F}_p^\times$,*

$$\sum_{n=1}^{p} \left| \sum_{x \in I} e_p(kf(x + n)) \right|^2 \leq C p |I|,$$

where the constant C depends only on d.

Proof. We have

$$\sum_{n=1}^{p} \left| \sum_{x \in I} e_p(kf(x + n)) \right|^2$$

$$= \sum_{x, y \in I} \sum_{n=1}^{p} e_p(k(f(x + n) - f(y + n)))$$

$$= \sum_{x, y \in I} \sum_{\alpha=1}^{p} \sum_{a, b=1}^{p} p^{-1} e_p(k(a - b)) e_p(\alpha(f^{-1}(a) - f^{-1}(b) - x + y))$$

$$= p^{-1} \sum_{\alpha=1}^{p} \left| \sum_{a=1}^{p} e_p(\alpha a + kf(a)) \right|^2 \left| \sum_{x \in I} e_p(\alpha x) \right|^2,$$

where the second equality follows from the fact that the sum over α enforces the condition $f^{-1}(a) - f^{-1}(b) = x - y$ (taking $n = f^{-1}(a) - x = f^{-1}(b) - y$, this is equivalent to $a = f(x + n)$ and $b = (y + n)$) and the third equality uses that f is bijective. Now $\alpha a + kf(a)$ is not constant as a function of a, because $f(a)$ is not linear. Then

$$\left| \sum_{a=1}^{p} e_p(\alpha a + kf(a)) \right| \leq C p^{1/2}$$

for a constant C depending only on d, by Lemma 3.1. Since there are a bounded number of poles for f, the additional terms introduced by extending to $\mathbf{F}_p$ are negligible up to changing C.

Note that for $\alpha \in [-p/2, p/2]$

$$\left| \sum_{x \in I} e_p(\alpha x) \right| \leq \min(|I|, p/2|\alpha|).$$

Thus,

$$p^{-1} \sum_{\alpha=1}^{p} \left| \sum_{a=1}^{p} e_p(\alpha a + k f(a)) \right|^2 \left| \sum_{x \in I} e_p(\alpha x) \right|^2 \le p^{-1} C \sum_{|\alpha| \le p/|I|} p|I|^2 + p^{-1} C \sum_{|\alpha| > p/|I|} \frac{p^3}{\alpha^2}$$

$$\le C p |I|,$$

using the bound

$$\sum_{\alpha=n}^{\infty} \frac{1}{\alpha^2} \le \frac{1}{n-1}. \qquad \square$$

Lemma 3.3. *For all $k \in \mathbf{F}_p^{\times}$, we have*

$$\sum_{n=1}^{p} \left| \sum_{x \in I} e_p(k(x+n)^2) \right|^2 = p|I|.$$

Proof. Expanding the square, we have

$$\sum_{n=1}^{p} \left| \sum_{x \in I} e_p(k(x+n)^2) \right|^2 = \sum_{x,y \in I} \sum_{n=1}^{p} e_p(2k(x-y)n + k(x^2 - y^2))$$

$$= p|I|. \qquad \square$$

These averaged estimates feed into an argument originally due to Heath-Brown [13], to give the following bound. We include the proof for completeness.

Proposition 3.4. *Let I_j be a collection of J disjoint arithmetic progressions of length L, with common difference $\delta \in \mathbf{F}_p^{\times}$. Suppose that $f \in \mathcal{B}(p,d)$ or $f(x) = x^2$. Then for all $k \in \mathbf{F}_p^{\times}$,*

$$\sum_{j=1}^{J} \left| \sum_{x \in I_j} e_p(k f(x)) \right|^2 \le C p \log^2(L+1),$$

where the constant depends only on d.

Proof. By replacing $f(x)$ with $f(\delta x)$ (or in the case of $f(x) = x^2$, absorbing δ^2 into k), we may work with intervals I_j instead of arithmetic progressions. The proof then proceeds exactly as in the proof of Theorem 2 in [4] (this argument is originally due to Heath-Brown [13]).

Let $S(n,h) = \sum_{x=n+1}^{n+h} e_p(k f(x))$. Let x_j be defined such that the first element in the interval I_j is $x_j + 1$. Then the sum we are interested in is simply

$$\sum_{j=1}^{J} |S(x_j, L)|^2.$$

Now for any $1 \le l \le L$ and $x_j - L < n \le x_j$, we have

$$|S(x_j, l)| = |S(n, x_j - n + l) - S(n, x_j - n)| \le 2 \max_{l \le 2L} |S(n, l)|,$$

and so

$$|S(x_j, L)| \le \frac{2}{L} \sum_{x_j - L < n \le x_j} \max_{l \le 2L} |S(n, l)|.$$

Then by Cauchy's inequality,

$$|S(x_j, L)|^2 \le \frac{4}{L} \sum_{x_j - L < n \le x_j} \max_{l \le 2L} |S(n, l)|^2,$$

and summing over j gives

$$\sum_{j=1}^{J} |S(x_j, L)|^2 \leq \frac{4}{L} \sum_{n \in \mathbf{F}_p} \max_{l \leq 2L} |S(n,l)|^2,$$

where we use the fact that the intervals are disjoint. Now pick t an integer so that $2L \leq 2^t \leq 4L$, and for each $n \in \mathbf{F}_p$, pick l^* depending on n so that $|S(n,l^*)| = \max_{l \leq 2L} |S(n,l)|$. Then writing the binary expansion $l^* = \sum_{s \in D} 2^{t-s}$, we have

$$S(n,l^*) = \sum_{s \in D} S(n + v_{n,s} 2^{t-s}, 2^{t-s}),$$

where

$$v_{n,s} = \sum_{u \in D, u < s} 2^{s-u} < 2^s.$$

Then we have

$$|S(n,l^*)|^2 \leq |D| \sum_{s \in D} |S(n + v_{n,s} 2^{t-s}, 2^{t-s})|^2$$

$$\leq (t+1) \sum_{0 \leq s \leq t} \sum_{0 \leq v < 2^s} |S(n + v 2^{t-s}, 2^{t-s})|^2.$$

Finally, we can sum over n and use Lemma 3.2 or Lemma 3.3 to obtain

$$\frac{4}{L} \sum_{n \in \mathbf{F}_p} \max_{l \leq 2L} |S(n,l)|^2 \leq \frac{Cp(t+1)}{L} \sum_{0 \leq s \leq t} \sum_{0 \leq v < 2^s} 2^{t-s}$$

$$\leq \frac{Cp(t+1)^2 2^t}{L}$$

$$\leq Cp \log^2(L+1). \qquad \square$$

We now prove a bound for linear exponential sums over a union of intervals.

Proposition 3.5. *Let I_j be a collection of J disjoint arithmetic progressions of length L, with common difference $\delta \in \mathbf{F}_p^\times$. Then*

$$\sum_{k=1}^{p-1} \left| \sum_{j=1}^{J} \sum_{x \in I_j} e_p(kx) \right| \leq C J^{1/2} p \log^{3/2}(p),$$

for some constant $C > 0$.

Proof. First, note that the sum is the same if we replace the arithmetic progressions I_j by intervals I_j / δ, so we may assume that $\delta = 1$ and the I_j are all intervals. Let x_j be defined such that $I_j = I_0 + x_j$ for $I_0 = \{1, 2, \ldots, L\}$. By relabeling the intervals if necessary, we will assume that the x_j are ordered. Since the intervals are disjoint, if $j \neq j'$ then $|x_j - x_{j'}| \geq L$.

The idea is to now break the sum over k up into intervals of length p/L, and take advantage of the different scales at which the oscillations for the sums over j and x occur. For $1 \leq l \leq (L-1)/2$ let I_l' denote the interval $[lp/L, (l+1)p/L]$ and let $I_{-l}' = -I_l'$. Then

$$\sum_{k=1}^{p-1} \left| \sum_{j=1}^{J} \sum_{x \in I_j} e_p(kx) \right|$$

$$\leq L \sum_{k=-p/L}^{p/L} \left| \sum_{j=1}^{J} e_p(kx_j) \right| + \sum_{\substack{l=-(L-1)/2 \\ l \neq 0}}^{(L-1)/2} \frac{L}{|l|} \sum_{k \in I_l'} \left| \sum_{j=1}^{J} e_p(kx_j) \right|, \qquad (3.1)$$

where we use $\left|\sum_{x \in I_0} e_p(kx)\right| \le \min(L, p/2|k|)$. Now note that for any interval I' of length p/L,

$$\sum_{k \in I'} \left|\sum_{j=1}^{J} e_p(kx_j)\right| \le (p/L)^{1/2} \left(\sum_{k \in I'} \left|\sum_{j=1}^{J} e_p(kx_j)\right|^2\right)^{1/2}. \tag{3.2}$$

Now

$$\sum_{k \in I'} \left|\sum_{j=1}^{J} e_p(kx_j)\right|^2 = \sum_{j,j'=1}^{J} \sum_{k \in I'} e_p(k(x_j - x_{j'}))$$

$$\le \sum_{j,j'=1}^{J} \min(p/L, p/|x_j - x_{j'}|)$$

where $x_j - x_{j'}$ is the representative mod p between $-p/2$ and $p/2$. But now by the spacing condition and the ordering of the x_j, we have $|x_j - x_{j'}| \ge |j - j'|L$ (where again $j - j'$ is the representative between $-J/2$ and $J/2$ mod J). Thus,

$$\sum_{j,j'=1}^{J} \min(p/L, p/|x_j - x_{j'}|) \le \frac{Jp}{L} + \sum_{m=1}^{J/2} \sum_{|j-j'|=m} \frac{p}{mL} \le 3\frac{Jp}{L}\log(p),$$

and combined with (3.2) this gives

$$\sum_{k \in I'} \left|\sum_{j=1}^{J} e_p(kx_j)\right| \le 3\frac{pJ^{1/2}}{L}\log^{1/2}(p).$$

Finally, this together with (3.1) gives the desired inequality. □

Using the exponential bounds established in this section, we can prove the following estimate for the number of solutions to $f(x) = y$ with $x \in S$ and $y \in S'$ when S and S' are the union of arithmetic progressions.

Proposition 3.6. *Let $S \subseteq \mathbf{F}_p$ be a disjoint union of J arithmetic progressions I_j of length L and common difference $\delta \in \mathbf{F}_p^\times$, and let S' be a disjoint union of J arithmetic progressions of length L' and common difference δ. Suppose that $f \in \mathcal{B}(p, d)$ or $f(x) = x^2$, and suppose that $JLL' \ge p^{3/2+\varepsilon}$ for some $\varepsilon > 0$. Then for large enough p,*

$$|\{x \in S | f(x) \in S'\}| \ge c\frac{|S||S'|}{p},$$

where $c > 0$ is some constant depending only on ε and d.

Proof. We have

$$|\{x \in S | f(x) \in S'\}| = \frac{1}{p}\sum_{k=1}^{p} \sum_{x \in S} \sum_{y \in S'} e_p(k(f(x) - y))$$

$$= \frac{|S||S'|}{p} + R,$$

where

$$R = \frac{1}{p}\sum_{k=1}^{p-1} \sum_{x \in S} \sum_{y \in S'} e_p(k(f(x) - y)).$$

By the triangle inequality,

$$|R| \leq \frac{1}{p} \sum_{k=1}^{p-1} \left| \sum_{x \in S} e_p(kf(x)) \right| \left| \sum_{y \in S'} e_p(-ky) \right|.$$

Now by Cauchy–Schwarz,

$$\left| \sum_{x \in S} e_p(kf(x)) \right| \leq \sum_{j=1}^{J} \left| \sum_{x \in I_j} e_p(kf(x)) \right| \leq J^{1/2} \left(\sum_{j=1}^{J} \left| \sum_{x \in I_j} e_p(kf(x)) \right|^2 \right)^{1/2}.$$

Since $k \neq p$, Proposition 3.4 implies,

$$\left| \sum_{x \in S} e_p(kf(x)) \right| \leq J^{1/2} C p^{1/2} \log(L).$$

By Proposition 3.5,

$$\sum_{k=1}^{p-1} \left| \sum_{y \in S'} e_p(-ky) \right| \leq C J^{1/2} p \log^{3/2}(p).$$

Thus,

$$|R| \leq C J p^{1/2} \log^{5/2}(p).$$

Now if $JLL' \geq p^{3/2+\varepsilon}$, then $|S||S'|/p \geq Jp^{1/2+\varepsilon}$ which dominates R. Finally, as C depends only on d, we can choose $c > 0$, depending only on the degree d and ε, so that for large enough p, we have $|\{x \in S | f(x) \in S'\}| \geq c|S||S'|/p$. $\qquad\square$

4 Mixing time bounds

The mixing time bounds for the cases when $f \in \mathcal{B}(p,d)$ and when $f(x) = x^2$ are both established using the same overall argument. This involves using Cheeger's inequality for directed graphs to reduce to a problem of counting edges between two sets, and then utilizing the results in Section 3 to show that there are many edges. The argument for the case when $f(x) = x^2$ is complicated by the fact that the stationary distribution does not have full support, and we explain how to work around this.

Let P be an ergodic Markov chain on X with stationary distribution π. Define the *Cheeger constant* $h(P)$ by

$$h(P) = \min_{S \subseteq X} \frac{\sum_{x \in S, y \in S^c} \pi(x) P(x,y)}{\min(\pi(S), \pi(S^c))},$$

where S ranges over all non-trivial subsets of X. Note that if P is simple random walk on a k-regular (i.e. both in and out-degrees are k) directed graph G, then this definition reduces to

$$h(P) = \min_{S \subseteq X} \frac{e(S, S^c)}{k \min(|S|, |S^c|)}, \tag{4.1}$$

where $e(S, S^c)$ is the number of edges going from S to S^c in G. When f is a bijection, our random walk falls into this case. Note also that because π is the stationary distribution,

$$\sum_{x \in S, y \in S^c} \pi(x) P(x,y) = \sum_{y \in S, x \in S^c} \pi(x) P(x,y),$$

and in particular for a random walk on a regular directed graph, $e(S, S^c) = e(S^c, S)$. Thus, we may consider edges in both directions up to a factor of 2.

The key tool for the bounds on the mixing times is Cheeger's inequality for non-reversible Markov chains. The following theorem follows from some standard facts about the spectral theory of Markov chains. See for example, page 23 of [19], noting that $(I + P)/2$ is lazy, and $h((I + P)2) = h(P)/2$ (see also [6] for the special case of random walk on a directed graph).

Theorem 4.1. *Let P be an ergodic Markov chain on X with stationary distribution π. Consider the lazy chain X_n with transition matrix $(I + P)/2$, and starting from some deterministic X_0. Then if*

$$n \geq 4h(P)^{-2}(\max_{x \in X} \log(\pi(x)^{-1}) + 2c)$$

for some $c > 0$, we have

$$\sup_{X_0 \in X} \| \mathrm{P}(X_n \in \cdot) - \pi\|_{TV} \leq e^{-c}.$$

4.1 Proof of Theorem 1.2

We show that in the supremum over S, the set S can be assumed to have some structure, at the cost of a constant. Since we are only interested in the order of the mixing time, the loss of a constant is okay.

First, we define a decomposition of any set $S \subseteq \mathbf{F}_p$ into certain types of arithmetic progressions. We will call an arithmetic progression with difference d a d-*AP*. A 2γ-*AP decomposition* of S is a decomposition $S = \bigsqcup I_k$ where the I_k are 2γ-APs, the number of I_k's is minimal. Such a decomposition always exists and is unique.

It is easy to see that if S has a 2γ-AP decomposition into J 2γ-APs, then the same is true of S^c.

We now show that when computing $\min_S \frac{e(S, S^c)}{\min(|S|, |S^c|)}$, we may assume that there are at most $|S|p^{-1/2-\varepsilon/2}$ 2γ-APs in the 2γ-AP decomposition of $|S|$. Thus, let $\mathcal{P}$ denote the set of S whose 2γ-AP decomposition contains at most $|S|p^{-1/2-\varepsilon/2}$ 2γ-APs.

Lemma 4.2. *We have for large enough p,*

$$\min_S \frac{e(S, S^c)}{\min(|S|, |S^c|)} \geq \min\left(p^{-1/2-\varepsilon/2}, \min_{S \in \mathcal{P}} \frac{e(S, S^c)}{\min(|S|, |S^c|)} \right).$$

Proof. Assume that $|S| \leq |S^c|$. Write $S = \bigsqcup I_k$ the 2γ-AP decomposition. Note that each right endpoint of a 2γ-AP contributes an edge from either S to S^c or from S^c to S, because if x is a right endpoint, then $x + 2\gamma \in S^c$ and x and $x + 2\gamma$ are both connected to $f^{-1}(x + \gamma)$. If there were more than $|S|p^{-1/2-\varepsilon/2}$ many 2γ-APs, then this would imply

$$\frac{e(S, S^c)}{\min(|S|, |S^c|)} \geq p^{-1/2-\varepsilon/2}. \qquad \square$$

The 2γ-AP decomposition of sets in $\mathcal{P}$ have arithmetic progressions of different lengths. The following lemma lets us reduce to the case when all arithmetic progressions have the same length.

Lemma 4.3. *Let $S = \bigsqcup_{j=1}^{J} I_j \subseteq \mathbf{F}_p$ be a disjoint union of arithmetic progressions with common difference $\delta \in \mathbf{F}_p^{\times}$. Let $L = |S|/J$ denote the average length of the I_j. Then S contains J disjoint arithmetic progressions I_j' with common difference δ and length $\lfloor L/4 \rfloor$.*

Proof. First, note that it suffices to prove the result for intervals. We have

$$\sum_{|I_j| \geq L/2} |I_j| \geq |S|/2,$$

and by splitting the intervals I_j with $|I_j| \geq L/2$ up into intervals of length exactly $\lfloor L/4 \rfloor$, we throw away at most $JL/4$ points. Thus, we obtain disjoint intervals of length $\lfloor L/4 \rfloor$, whose total length is at least $|S|/4$, and so there must be at least J such intervals. $\qquad\square$

We are now in a position to prove Theorem 1.2 using the bounds given by Proposition 3.6.

Proof of Theorem 1.2. By Cheeger's inequality, it suffices to show that the Cheeger constant is bounded from below by $p^{-1/2-\varepsilon/2}$. By Lemma 4.2, it suffices to give a lower bound for

$$\min_{S \in \mathcal{P}} \frac{e(S, S^c)}{\min(|S|, |S^c|)}$$

of order $p^{-1/2-\varepsilon/2}$ (in fact when $S \in \mathcal{P}$ we'll give a constant order lower bound). Thus, take $S \in \mathcal{P}$ and assume that $|S| \leq |S^c|$.

Using Lemma 4.3, we can find subsets $S_1 \subseteq S$ and $S_2 \subseteq S^c$ which are a disjoint union of at most $J \leq |S|p^{-1/2-\varepsilon/2}$ 2γ-APs of fixed lengths L_1 and L_2, with L_1.

Now $J \leq |S|p^{-1/2-\varepsilon/2}$, and so $L_1 \geq p^{1/2+\varepsilon/2}/4$. Similarly, $JL_2 \geq J\lfloor|S^c|/J\rfloor \geq p/8$. Then $JL_1L_2 \geq p^{3/2+\varepsilon/2}/32$, and so Proposition 3.6 shows that

$$e(S, S^c) \geq e(S_1, S_2) \geq c\frac{|S||S^c|}{p}$$

for some positive constant c depending only on d and ε. This implies that

$$\frac{e(S, S^c)}{\min(|S|, |S^c|)} \geq c\frac{|S^c|}{p} \geq \frac{c}{2}.$$

Since the estimate is uniform over $S \in \mathcal{P}$, this completes the proof. $\qquad\square$

4.2 Proof of Theorem 1.8

We now sketch the adjustments that have to be made to handle the case of $f(x) = x^2$. The main difficulty lies in the fact that the support of the stationary distribution $\operatorname{supp} \pi_p$ is not all of $\mathbf{F}_p$. To adapt the argument from the bijective case, we take advantage of the fact that when $p = 3 \pmod 4$, at least one of α and $-\alpha$ lies in $\operatorname{supp} \pi$, and that π is close to uniform on its support.

Consider now the random walk restricted to $\operatorname{supp} \pi$. Since the walk enters this set after a single step, it suffices to bound the mixing time for the restricted walk. Now although the walk is not a simple random walk on a regular graph, the transition probabilities are of constant order in p and by Theorem 1.7 the stationary distribution is essentially constant up to a factor of 4. Thus, (4.1) holds up to some constant factor, and so we provide a lower bound for $e(S, S^c)/\min(|S|, |S^c|)$.

We now adapt the 2γ-AP decomposition to this setting. A set $S \subseteq \operatorname{supp} \pi$ is *symmetric* if it is of the form $\widetilde{S} \cap \operatorname{supp} \pi$ for a set $\widetilde{S} \subseteq \mathbf{F}_p$ with $\widetilde{S} = -\widetilde{S}$. A *symmetric 2γ-AP* is a set of the form $(J \cup -J) \cap \operatorname{supp} \pi$, where J is a 2γ-AP contained in $\{0, 1, \ldots, (p-1)/2\}$. A *$2\gamma$-AP decomposition* of a symmetric set $S \subseteq \operatorname{supp} \pi$ is a decomposition $S = \bigsqcup I_k$ where the I_k are of the form $I_k = \operatorname{supp} \pi \cap (J_k \cup -J_k)$, and the J_k are 2γ-APs contained in $\{0, 1, \ldots, (p-1)/2\}$, such that the number of I_k's is minimal. It is clear that this exists and is unique, and moreover the 2γ-APs J_k are uniquely determined as well.

The idea is to use this as a replacement for the 2γ-AP decomposition used in the proof for $f \in \mathcal{B}(p, d)$.

Let $\mathcal{P}$ denote the set of subsets $S \subseteq \operatorname{supp} \pi$ which are symmetric, and whose 2γ-AP decomposition contains at most $J \leq |S|p^{-1/2-\varepsilon/2}$ 2γ-APs.

Lemma 4.4. *We have for large enough p,*

$$\min_{S} \frac{e(S, S^c)}{\min(|S|, |S^c|)} \geq \max\left(p^{-1/2-\varepsilon/2}, \frac{1}{16} \min_{S \in \mathcal{P}} \frac{e(S, S^c)}{\min(|S|, |S^c|)}\right).$$

Proof. Assume that $|S| \leq |S^c|$, which can be done since $e(S, S^c)$ and $e(S^c, S)$ differ by at most a multiplicative constant, since π is essentially constant on $\operatorname{supp} \pi$ up to a factor of 4. Note that if $\alpha \in S$ and $-\alpha \in S^c$, then as both are connected to $\alpha^2 \pm \gamma$, this contributes 2 edges from S to S^c. Thus, if there are more than $|S|^{1/2-\varepsilon/2}$ such pairs, we have

$$\frac{e(S, S^c)}{\min(|S|, |S^c|)} \geq |S|^{-1/2-\varepsilon/2} \geq p^{-1/2-\varepsilon/2}.$$

Otherwise, replace S with the set $S' = S \cup \{\alpha \in S^c \mid -\alpha \in S\}$. Let $n = |S'| - |S|$, and note that $n \leq |S|^{1/2-\varepsilon/2}$ and $e(S, S^c) \geq 2n$. For large enough p, we have

$$\min(|S|, |S^c|) = |S| \leq 2\min(|S'|, |(S')^c|), \tag{4.2}$$

since either $\min(|S'|, |(S')^c|) = |S'|$ for which the statement obviously holds, or $|S| \geq |\operatorname{supp} \pi|/2 - n = (p-1)/4 - n$, in which case we have

$$|(S')^c| = |S^c| - n \geq |S| - |S|^{1/2-\varepsilon/2},$$

and for large enough p this also implies (4.2).

The numerator increases by at most $6n$, since each vertex has degree at most 6, and so for large enough p

$$e(S', (S')^c) \leq e(S, S^c) + 6n \leq 8e(S, S^c).$$

This implies that the minimum can be restricted to symmetric S, at the cost of the $1/16$ factor.

As in the proof of Lemma 4.2, each right endpoint of a 2γ-AP appearing in the decomposition of S contributes an edge from S to S^c. This is because if x is the right endpoint of a 2γ-AP J appearing in the decomposition, one of $x + \gamma$ and $-x - \gamma$ is a quadratic residue, and thus either $x \in S$ and $x + 2\gamma \in S^c$ contributes an edge, or $-x \in S$ and $-x - 2\gamma \in S^c$ contributes an edge. $\square$

With this replacement for Lemma 4.2, we can now prove Theorem 1.8.

Proof of Theorem 1.8. As in the proof of Theorem 1.2, it suffices to give a lower bound for $e(S, S^c)/|S|$ when $S \in \mathcal{P}$ and $|S| \leq |S^c|$. A small extension of Lemma 4.3 then reduces to sets S_1 and S_2 both of comparable size to S and S^c respectively, with S_1 and S_2 being the disjoint union of symmetric 2γ-APs of lengths L_1 and L_2 respectively.

Note that if J is a 2γ-AP appearing in the decomposition of S_1, then if $x \in J$, at least one of x and $-x$ belong to S_1. If there is an edge from $x \in J$ to S_2, there is also an edge from $-x$ to S_2, so if $S_1 = \bigsqcup (J_j \cup -J_j) \cap \operatorname{supp} \pi$, then $e(S_1, S_2) \geq e(\bigsqcup J_j, S_2)$. Further, if $S_2 = \bigsqcup (K_j \cup -K_j) \cap \operatorname{supp} \pi$, then any edge landing in K_j must also land in $\operatorname{supp} \pi$, because $\operatorname{supp} \pi$ contains all elements of the form $\alpha^2 \pm \gamma$, and so $e(\bigsqcup J_j, S_2) \geq e(\bigsqcup J_j, \bigsqcup K_j)$.

Since the sets have been written as a disjoint union of 2γ-APs, we can now apply Proposition 3.6 and proceed as before. $\square$

4.3 A variant of the lazy chain

Here, we sketch the adjustments needed to deal with the original Markov chains appearing in [5]. In the Markov chains considered there, the function f is always applied, but $\varepsilon_n = 0$ with positive probability. Note that the arguments in [5] only work when f is bijective, even though the Markov chain when $f(x) = x^2$ still makes sense, so we restrict to the setting where $f \in \mathcal{B}(p, d)$.

Theorem 4.5. *Let p be a prime and $d \in \mathbf{N}$. Let $f \in \mathcal{B}(p, d)$, and let $\gamma \in \mathbf{F}_p$ be non-zero. Let ε_i denote independent random variables uniform on $\{-\gamma, 0, \gamma\}$. Consider the Markov chain defined by*

$$X_{n+1} = f(X_n) + \varepsilon_{n+1}.$$

Then if π denotes the uniform distribution on $\mathbf{F}_p$, for any ε, there is some constant $C = C(\varepsilon, d)$ such that

$$\sup_{X_0 \in \mathbf{F}_p} \| \operatorname{P}(X_n \in \cdot) - \pi \|_{TV} \leq e^{-C \frac{n}{p^{1+\varepsilon}}}.$$

By the arguments of Section 2 in [5], it suffices to bound the second largest eigenvalue of a symmetrization using Cheeger's inequality, which reduces to showing that

$$\frac{e(S, S^c)}{\min(|S|, |S^c|)}$$

is large in the graph determined by the random walk $Pf^{-1}P^2fP$, where P is the transition matrix for taking a random step of either 0 or $\pm\gamma$ with equal probability, and f is the permutation matrix corresponding to the bijection f. Actually to make an argument later work, we instead work with $PfP^2f^{-1}P$, which has the same (non-zero) spectrum.

Now certainly, it cannot hurt to throw away edges, and so we instead consider the bijective function $g(x) = f(f^{-1}(x) + \gamma)$, and consider the graph with edges connecting x to $g(x) \pm \gamma$ (we are forcing a lazy first step from P, and after appling f^{-1} one lazy step from P and then adding γ).

Now from the arguments in Section 4.1, if the set S has a 2γ-AP decomposition with too many intervals, then $e(S, S^c)$ will be large since every interval contributes an edge, and otherwise we can apply the exponential sum bounds of Section 3, with an analogous version of Lemma 3.2. We first need the following result to apply the Weil bound in this situation.

Lemma 4.6. *Let $P, Q \in \mathbf{F}_p[x]$ be coprime and of degree at most $p/4$, and suppose that $P(x)/Q(x)$ is not constant or linear. Then for all $\alpha, \beta, \gamma \in \mathbf{F}_p$ with $\beta, \gamma \neq 0$, the function*

$$\alpha P(x + \gamma)/Q(x + \gamma) + \beta P(x)/Q(x)$$

is not a constant function.

Proof. Suppose that $\alpha P(x + \gamma)/Q(x + \gamma) + \beta P(x)/Q(x) = c$ away from the poles of $Q(x)$ and $Q(x + \gamma)$. Then $\alpha P(x + \gamma)Q(x) + \beta P(x)Q(x + \gamma) = cQ(x)Q(x + \gamma)$ for more than $p/2$ points, and so both sides must be equal as polynomials since they are polynomials of degree at most $p/2$. Then as P and Q are coprime, $Q(x)$ must divide $Q(x + \gamma)$ and $Q(x + \gamma)$ must divide $Q(x)$, and by comparing the coefficient of the highest degree term it's clear that $Q(x) = Q(x + \gamma)$, so $Q(x)$ is a constant. Then $\alpha P(x + \gamma) + \beta P(x) = c$, and if P is not a constant, this implies $\alpha = -\beta$ by comparing the highest degree term. Then the coefficient of the second highest degree term on the left is non-zero, so P must be a linear function. $\square$

Now, we prove the analogue of Lemma 3.2.

Lemma 4.7. *Let $f \in \mathcal{B}(p, d)$ and let $g(x) = f(f^{-1}(x) + \gamma)$. Then for any interval I, and any $k \in \mathbf{F}_p^\times$,*

$$\sum_{n=1}^{p} \left| \sum_{x \in I} e_p(kg(x + n)) \right|^2 \leq Cp|I|,$$

where the constant C depends only on d.

Proof. Since the constant C can depend on d, if we take C to be at least $4d$, then if $d \geq p/4$, the statement is trivially true as the left hand side is at most $p|I|^2 \leq p^2|I|$. Thus, we assume that $d < p/4$.

Proceeding as in the proof of Lemma 3.2, we have

$$\sum_{n=1}^{p} \left| \sum_{x \in I} e_p(kg(x + n)) \right|^2$$

$$= \sum_{x,y \in I} \sum_{\alpha=1}^{p} \sum_{a,b=1}^{p} p^{-1} e_p(k(a - b)) e_p(\alpha(g^{-1}(a) - g^{-1}(b) - x + y))$$

$$= p^{-1} \sum_{\alpha=1}^{p} \left| \sum_{a=1}^{p} e_p(\alpha f(a - \gamma) + kf(a)) \right|^2 \left| \sum_{x \in I} e_p(\alpha x) \right|^2,$$

where the substitution $a \mapsto f(a)$ is made. Then as $\alpha f(a - \gamma) + kf(a)$ is not constant in a by Lemma 4.6, we can apply Lemma 3.1. The proof then proceeds as before. $\square$

The rest of the arguments in Section 3 are unchanged, leading to a version of Proposition 3.6 for the function g, which completes the argument.

References

[1] N. Alon, I. Benjamini, E. Lubetzky, and S. Sodin, *Non-backtracking random walks mix faster*, Commun. Contemp. Math. **9** (2007), no. 4, 585–603. MR2348845

[2] C. Bordenave, Y. Qiu, and Y. Zhang, *Spectral gap of sparse bistochastic matrices with exchangeable rows*, Ann. Inst. Henri Poincaré Probab. Stat. **56** (2020), no. 4, 2971–2995. MR4164862

[3] E. Breuillard and P. P. Varjú, *Cut-off phenomenon for the ax+b Markov chain over a finite field*, 2019. arXiv:1909.09053

[4] T. D. Browning and A. Haynes, *Incomplete Kloosterman sums and multiplicative inverses in short intervals*, Int. J. Number Theory **9** (2013), no. 2, 481–486. MR3005560

[5] S. Chatterjee and P. Diaconis, *Speeding up Markov chains with deterministic jumps*, Probab. Theory Related Fields **178** (2020), no. 3-4, 1193–1214. MR4168397

[6] F. Chung, *Laplacians and the Cheeger inequality for directed graphs*, Ann. Comb. **9** (2005), no. 1, 1–19. MR2135772

[7] F. R. K. Chung, *Diameters and eigenvalues*, J. Amer. Math. Soc. **2** (1989), no. 2, 187–196. MR965008

[8] F. R. K. Chung, P. Diaconis, and R. L. Graham, *Random walks arising in random number generation*, Ann. Probab. **15** (1987), no. 3, 1148–1165. MR893921

[9] C. Cooper and A. Frieze, *The size of the largest strongly connected component of a random digraph with a given degree sequence*, Combin. Probab. Comput. **13** (2004), no. 3, 319–337. MR2056402

[10] P. Diaconis, J. He, and I. M. Isaacs, *The-square-and-add Markov chain*, 2021, pp. 27–36. MR4278471

[11] S. Eberhard and P. P. Varjú, *Mixing time of the Chung-Diaconis-Graham random process*, Probab. Theory Related Fields **179** (2021), no. 1-2, 317–344. MR4221660

[12] J. Gutierrez and I. E. Shparlinski, *Expansion of orbits of some dynamical systems over finite fields*, Bull. Aust. Math. Soc. **82** (2010), no. 2, 232–239. MR2685147

[13] D. R. Heath-Brown, *Burgess's bounds for character sums*, Number theory and related fields, Springer Proc. Math. Stat., vol. 43, Springer, New York, 2013, pp. 199–213. MR3081042

[14] J. Hermon, A. Sly, and P. Sousi, *Universality of cutoff for graphs with an added random matching*, 2020. arXiv:2008.08564

[15] M. Hildebrand, *On the Chung-Diaconis-Graham random process*, Electron. Comm. Probab. **11** (2006), 347–356. MR2274529

[16] M. Hildebrand, *A lower bound for the Chung-Diaconis-Graham random process*, Proc. Amer. Math. Soc. **137** (2009), no. 4, 1479–1487. MR2465674

[17] M. Hildebrand, *On a lower bound for the Chung-Diaconis-Graham random process*, Statist. Probab. Lett. **152** (2019), 121–125. MR3953053

[18] S. V. Konyagin, F. Luca, B. Mans, L. Mathieson, M. Sha, and I. E. Shparlinski, *Functional graphs of polynomials over finite fields*, J. Combin. Theory Ser. B **116** (2016), 87–122. MR3425238

[19] R. Montenegro and P. Tetali, *Mathematical aspects of mixing times in Markov chains*, Found. Trends Theor. Comput. Sci. **1** (2006), no. 3, x+121. MR2341319

[20] R. Neville, III, *On lower bounds of the Chung-Diaconis-Graham random process*, ProQuest LLC, Ann Arbor, MI, 2011, Thesis (Ph.D.)–State University of New York at Albany. MR2873460

[21] A. Peinado, F. Montoya, J. Muñoz, and A. J. Yuste, *Maximal periods of $x^2 + c$ in $\mathbb{F}_q$*, Applied algebra, algebraic algorithms and error-correcting codes (Melbourne, 2001), Lecture Notes in Comput. Sci., vol. 2227, Springer, Berlin, 2001, pp. 219–228. MR1913468

[22] T. Vasiga and J. Shallit, *On the iteration of certain quadratic maps over* $\mathrm{GF}(p)$, Discrete Math. **277** (2004), no. 1-3, 219–240. MR2033734

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm