Electron. J. Probab. **22** (2017), no. 90, 1–49.
ISSN: 1083-6489 DOI: 10.1214/17-EJP108

# Mixing and cut-off in cycle walks*

Robert Hough†

**Abstract**

Given a sequence $(\mathfrak{X}_i, \mathscr{K}_i)_{i=1}^{\infty}$ of Markov chains, the cut-off phenomenon describes a period of transition to stationarity which is asymptotically lower order than the mixing time. We study mixing times and the cut-off phenomenon in the total variation metric in the case of random walk on the groups $\mathbb{Z}/p\mathbb{Z}$, $p$ prime, with driving measure uniform on a symmetric generating set $A \subset \mathbb{Z}/p\mathbb{Z}$.

**Keywords:** random walk on a group; random lattice; cut-off phenomenon; embedded hypercube.
**AMS MSC 2010:** Primary 60B10, Secondary 60B15; 60G50; 60J60; 11H06; 11A63.
Submitted to EJP on June 20, 2016, final version accepted on September 15, 2017.
Supersedes arXiv:1512.00571.

## 1 Introduction

The mixing analysis of random walk on a finite abelian group is a classical problem of probability theory, with widespread applications; the Ehrnfest urn and sandpile models of statistical mechanics are motivating examples [8, 17, 26]. Among the early results in this area is a theorem of Greenhalgh [15], which shows that for generating set of size $k$ contained in $\mathbb{Z}/n\mathbb{Z}$, the mixing time of the corresponding random walk satisfies $t^{\text{mix}} \gg_k n^{\frac{2}{k-1}}$. A set of size $k$ with mixing time bounded by $\ll_k n^{\frac{2}{k-1}} \log n$ is also exhibited. Dou, Hildebrand and Wilson [13], [16], [28] consider the mixing of measures driven by typical generating sets on cyclic and more general groups. Among the results of [16] is that typical generating sets of size $k = (\log n)^a$, $a > 1$ produce a random walk satisfying the cut-off phenomenon. We confine our attention to cyclic groups and symmetric generating sets which are smaller than logarithmic size in the order of the group, and prove a number of refined results on the mixing behavior. Our results are in a similar spirit to those of Diaconis and Saloff-Coste [5] proven in the more general context of random walk on groups of polynomial growth, but in narrowing our focus we emphasize

*This material is based upon work supported by the National Science Foundation under agreement No. DMS-1128155. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.
The author was partially supported by a Ric Weiland Graduate Research Fellowship at Stanford University.

†School of Mathematics, Institute of Advanced Study, 1 Einstein Drive, Princeton, NJ, 08540, USA.
Current address: Dept. of Mathematics, Stony Brook University, 100 Nicolls Road, Stony Brook, NY 11794
E-mail: robert.hough@stonybrook.edu

strong uniformity in the number of generators of the random walk. Note that in the context of random walk on nilpotent groups, the mixing of the walk projected to the abelianization often controls the mixing in the group as a whole, see [14], [9].

To briefly summarize the results, Theorem 1.1 gives spectral upper and lower bounds for the mixing time in a sharper form than previous results which have appeared in the literature. A natural conjecture regarding random walk on a connected graph is that the total variation mixing time is bounded by the maximum degree times the diameter squared. A highlight of our work is Theorem 1.4, which verifies the conjecture for the mixing time of random walk on the Cayley graph of $\mathbb{Z}/p\mathbb{Z}$ with a small symmetric generating set. Theorem 1.6 gives a lower bound for the period of transition to uniformity relative to the mixing time – a lower bound on the cut-off window. Theorem 1.7 determines the generic and worst case mixing behavior for a sequence of typical symmetric random walks. We conclude by analyzing the mixing time of a walk which may be considered an approximate embedding of the hypercube $(\mathbb{Z}/2\mathbb{Z})^d$ into the cycle, demonstrating a cut-off phenomenon.

### 1.1 Precise statement of results

Let $\mathscr{P}$ be the set of primes. Given $p \in \mathscr{P}$ let $A \subset \mathbb{Z}/p\mathbb{Z}$ be symmetric ($x \in A$ if and only if $-x \in A$), lazy ($0 \in A$) and generating ($|A| > 1$). Write $\mathscr{A}(p)$ be the collection of symmetric, lazy, generating subsets of $\mathbb{Z}/p\mathbb{Z}$, and for $k \in \mathbb{Z}_{>0}$ write $\mathscr{A}(p,k) \subset \mathscr{A}(p)$ be those sets of size $2k+1$. Given $A \in \mathscr{A}(p)$ let $\mu_A$ denote the uniform measure on $A$,

$$\mu_A = \frac{1}{|A|} \sum_{x \in A} \delta_x.$$

The distribution at step $n \geq 1$ of random walk driven by $\mu_A$ is given by the convolution power

$$\mu_A^{*1} = \mu_A, \qquad \mu_A^{*n} = \mu_A^{*(n-1)} * \mu_A, \ n > 1.$$

As $n \to \infty$, $\mu_A^{*n}$ converges to the uniform measure $\mathbb{U}_{\mathbb{Z}/p\mathbb{Z}}$ on $\mathbb{Z}/p\mathbb{Z}$ and we consider asymptotic behavior of this convergence for large $p$. In particular, the behavior of these walks as $k = k(p)$ varies as a function of $p$, and as $A$ varies in the set $\mathscr{A}(p,k)$ is studied.

Given measure space $(\mathfrak{X}, \mathscr{B})$, a norm $\|\cdot\|$ on the space $\mathscr{M}(\mathfrak{X})$ of probability measures on $\mathfrak{X}$, a Markov chain $P^n(\cdot)$ with stationary measure $\nu \in \mathscr{M}(\mathfrak{X})$, and $0 < \epsilon < 1$, define the $\epsilon$-mixing time

$$t^{\mathrm{mix}}(\epsilon) = \inf \left\{ n : \sup_{\mu \in \mathscr{M}(\mathfrak{X})} \|P^n(\mu) - \nu\| \leq \epsilon \right\}$$

and the standard mixing time $t^{\mathrm{mix}} = t^{\mathrm{mix}}\left(\frac{1}{e}\right)$. In the cases considered $\mathfrak{X}$ is a (finite, compact, locally compact) abelian group, and, due to the symmetry of the walk, it is sufficient to take for $\mu$ the point mass at 0. Of primary interest is the total variation norm, which for $\mu, \nu \in \mathscr{M}(\mathfrak{X})$ is given by

$$\|\mu - \nu\|_{\mathrm{TV}(\mathfrak{X})} = \sup_{S \in \mathscr{B}} |\mu(S) - \nu(S)|.$$

The mixing time with respect to this norm is indicated $t_1^{\mathrm{mix}}$. Two further important parameters in considering reversible Markov chains are the spectral gap of the transition kernel

$$\mathrm{gap} = 1 - \sup\{|\lambda| : \lambda \in \mathrm{spec}(P) \setminus \{\pm 1\}\}$$

and the relaxation time

$$t^{\mathrm{rel}} = \frac{1}{-\log(1 - \mathrm{gap})} \approx \frac{1}{\mathrm{gap}}.$$

In stating our results we let $\tau_0$ denote the ratio $\frac{t_1^{\mathrm{mix}}}{t^{\mathrm{rel}}}$ of the one dimensional Gaussian diffusion

$$\theta(x,t) = \sum_{j\in\mathbb{Z}} \exp(-2\pi^2 t j^2) e^{2\pi i j x} \tag{1.1}$$

on $(\mathbb{R}/\mathbb{Z}, dx)$; $2\pi^2 t = \tau_0$ solves the equation

$$\int_0^1 |\theta(x,t) - 1| dt = \frac{2}{e}$$

and has numerical value[1]

$$\tau_0 = 0.56161265(1). \tag{1.2}$$

In the context of random walk on $\mathbb{Z}/p\mathbb{Z}$ with small symmetric generating sets, the relaxation and total variation mixing times are related as follows.[2]

**Theorem 1.1.** *Let $p$ be prime, let $1 \le k \le \frac{\log p}{\log\log p}$ and let $A \in \mathscr{A}(p,k)$. Denote $t^{\mathrm{rel}}, t_1^{\mathrm{mix}}$ the relaxation time and total variation mixing time of $\mu_A$ on $\mathbb{Z}/p\mathbb{Z}$. We have*

$$\frac{\tau_0 e}{4\pi} p^{\frac{2}{k}} \lesssim_k \tau_0 t^{\mathrm{rel}} \lesssim_p t_1^{\mathrm{mix}} \lesssim_k 0.163 k t^{\mathrm{rel}}.$$

*Also, uniformly in $k$,*

$$\frac{2k+1}{16\pi\Gamma\left(\frac{k}{2}+1\right)^{\frac{2}{k}}} p^{\frac{2}{k}} \lesssim_p t^{\mathrm{rel}}.$$

**Remark 1.2.** The relationship $\frac{t_1^{\mathrm{mix}}}{t^{\mathrm{rel}}} \gtrsim \tau_0$ exhibits Gaussian diffusion on $\mathbb{R}/\mathbb{Z}$ as asymptotically extremal for the ratio between the mixing and relaxation times.

**Remark 1.3.** The lower bound gives an explicit dependence on $k$ in Greenhalgh's theorem. An upper bound of this type may be extracted from [5], Theorem 1.2, but the $k$ dependence there is, in worst case, exponential.

Theorem 1.1 relates the mixing time to spectral data, but in some cases it is more desirable to understand the mixing time geometrically. Given symmetric generating set $A \subset \mathbb{Z}/p\mathbb{Z}$ denote $\mathscr{C}(A,p)$ the Cayley graph with vertices $V = \mathbb{Z}/p\mathbb{Z}$ and edge set $E = \{(n_1,n_2) \in (\mathbb{Z}/p\mathbb{Z})^2 : n_1 - n_2 \in A\}$. Write $\mathrm{diam}(\mathscr{C}(A,p))$ for the graph-theoretic diameter of $\mathscr{C}(A,p)$. Since $\mathbb{Z}/p\mathbb{Z}$ is abelian there is a more geometric notion of diameter

$$\mathrm{diam}_{\mathrm{geom}}(\mathscr{C}(A,p)) = \max_{x\in\mathbb{Z}/p\mathbb{Z}} \min\left(\|\underline{n}\|_2 : \underline{n} \in \mathbb{Z}^k,\ \exists \underline{a} \in A^k,\ \underline{n}\cdot\underline{a} \equiv x \bmod p\right).$$

One has (the second inequality is given in Lemma 2.4)[3]

$$\mathrm{diam}(\mathscr{C}(A,p)) \ge \mathrm{diam}_{\mathrm{geom}}(\mathscr{C}(A,p)) \gg \sqrt{\frac{t^{\mathrm{rel}}}{k}}.$$

Random walk driven by $\mu_A$ on $\mathbb{Z}/p\mathbb{Z}$ may be interpreted as random walk on $\mathscr{C}(A,p)$ in which at each step the walker chooses a uniform edge leaving its current position.

**Theorem 1.4.** *Let $p$ be an odd prime and let $A \in \mathscr{A}(p)$ with $|A| = 2k+1$, $1 \le k \le \frac{\log p}{\log\log p}$. The mixing time $t_1^{\mathrm{mix}}$ of random walk driven by $\mu_A$ satisfies, as $p \to \infty$,*

$$t_1^{\mathrm{mix}} \ll k \cdot \mathrm{diam}_{\mathrm{geom}}(\mathscr{C}(A,p))^2.$$

[1] We use parentheses to indicate the last significant digit of numerical constants.

[2] We write $A(x) \lesssim_x B(x)$ meaning that there is a non-increasing function $f : \mathbb{R}^+ \to \mathbb{R}^+$ with $\lim_{x\to\infty} f(x) = 1$ such that $A(x) \le f(x)B(x)$, thus indicating the parameter which must grow for the asymptotic to hold.

[3] The notation $A \gg B$ means $B = O(A)$.

**Remark 1.5.** In the context of random walk on a cycle, Theorem 1.4 refines in two ways the much more general Theorem 1.2 of [5], which applies in the context of groups of moderate growth. The dependence on the number of generators $k$ there is, in worst case, exponential. Also, we replace the diameter there with the smaller geometric diameter here. See also [27].

Given a sequence of triples $(\mathfrak{X}_i, P_i, \nu_i)_{i=1}^{\infty}$ where $\mathfrak{X}_i$ is a measure space and $P_i$ is a Markov kernel on $\mathfrak{X}_i$ which has $\nu_i \in \mathscr{M}(\mathfrak{X}_i)$ as its stationary distribution, the sequence exhibits the cut-off phenomenon in total variation if for all $0 < \epsilon < \frac{1}{2}$,

$$\lim_{i\to\infty} \frac{t_{1,i}^{\mathrm{mix}}(\epsilon)}{t_{1,i}^{\mathrm{mix}}(1-\epsilon)} = 1.$$

The cut-off phenomenon is frequently observed in natural families of Markov chains including the hypercube walk of [8] and riffle shuffling viewed as a random walk on the symmetric group [1]. Especially in total variation, the cut-off phenomenon is still imperfectly understood, so that there is significant interest in deciding its occurrence in specific examples, see for instance [10], [12], [4], [2], [21].

One necessary condition for cut-off in total variation to occur is

$$\lim_{i\to\infty} \frac{t_{1,i}^{\mathrm{mix}}}{t_i^{\mathrm{rel}}} = \infty,$$

see Chapter 18.3 of [20]. In particular, by Theorem 1.1 any sequence of walks generated by $\{A_p \bmod p \subset \mathbb{Z}/p\mathbb{Z}\}_{p\in\mathscr{P}}$ for which $|A_p|$ remains bounded does not have cut-off, a result first obtained in [5]. We give a different proof of this result found independently by the author, which gives further information on the period of transition to uniformity.

**Theorem 1.6.** *Let $p \geq 3$ be prime, let $1 \leq k \leq \frac{\log p}{\log\log p}$ and let $A \in \mathscr{A}(p,k)$. For any $0 < \epsilon < \frac{1}{e}$ the total variation mixing times of $\mu_A$ on $\mathbb{Z}/p\mathbb{Z}$ satisfy*

$$t_1^{\mathrm{mix}}(\epsilon) - t_1^{\mathrm{mix}}(1-\epsilon) \gg_\epsilon \frac{t_1^{\mathrm{mix}}}{k}.$$

In contrast to Theorem 1.6, our next theorem shows that the generic behavior when $|A_p|$ grows slowly is for there to be a sharp transition to uniformity with infrequent exceptions. This Theorem answers a question raised in [6].

**Theorem 1.7.** *Let $k : \mathscr{P} \to \mathbb{Z}_{>0}$ tend to $\infty$ with $p$ in such a way that $k(p) \leq \frac{\log p}{\log\log p}$. Let sets $\{A_p \bmod p\}_{p\in\mathscr{P}}$ be chosen independently with $A_p$ chosen uniformly from $\mathscr{A}(p,k(p))$. The following hold with probability 1.*

1. *Let $\rho : \mathscr{P} \to \mathbb{R}^+$ satisfy $\sum_p \frac{1}{\rho(p)^k} = \infty$. There is an infinite subsequence $\mathscr{P}_0 \subset \mathscr{P}$ such that for $p$ increasing through $\mathscr{P}_0$,*

$$t^{\mathrm{rel}}(p) \gtrsim \frac{e}{\pi}\rho(p)^2 p^{\frac{2}{k(p)}} \qquad \text{and} \qquad t_1^{\mathrm{mix}}(p) \sim \tau_0 t^{\mathrm{rel}}(p).$$

   *In particular, the cut-off phenomenon does not occur for $(\mathbb{Z}/p\mathbb{Z}, \mu_{A_p}, \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}})_{p\in\mathscr{P}}$.*

2. *Let $\rho : \mathscr{P} \to \mathbb{R}^+$ satisfy $\sum_p \frac{1}{\rho(p)^k} < \infty$. Then*

$$t_1^{\mathrm{mix}}(p) \lesssim \frac{\tau_0 e}{\pi}\rho(p)^2 p^{\frac{2}{k(p)}}.$$

3. *For any sequence $\{\epsilon(p)\}_{p\in\mathscr{P}} \subset \mathbb{R}_{>0}$ satisfying $\epsilon(p)\sqrt{k(p)} \to \infty$ there is a density 1 subset $\mathscr{P}_0 \subset \mathscr{P}$ such that in the family $(\mathbb{Z}/p\mathbb{Z}, \mu_{A_p}, \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}})_{p\in\mathscr{P}_0}$ we have*

$$t_1^{\mathrm{mix}}(p) \sim \frac{k(p)}{2\pi e} p^{\frac{2}{k(p)}},$$

*and as $p$ increases through $\mathscr{P}_0$*

$$\lim \left\| \mu_{A_p}^{(1-\epsilon)t_1^{\mathrm{mix}}(p)} - \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}} \right\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})} = 1, \qquad \lim \left\| \mu_{A_p}^{(1+\epsilon)t_1^{\mathrm{mix}}(p)} - \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}} \right\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})} = 0.$$

*In particular, the cut-off phenomenon occurs.*

**Remark 1.8.** Since $\sum_p \frac{1}{p} = \infty$, items (1) and (3) of Theorem 1.7 demonstrate that almost surely among a sequence of walks, infinitely often there are slowly mixing walks which are slower than the typical behavior by a factor of $\gg \frac{p^{\frac{2}{k(p)}}}{k(p)}$.

**Remark 1.9.** (3) of Theorem 1.7 gives a cut-off sequence with, for $0 < \epsilon < \frac{1}{2}$, period of transition between $t_1^{\mathrm{mix}}(1-\epsilon)$ and $t_1^{\mathrm{mix}}(\epsilon)$ of length $O_\epsilon\left(\frac{t_1^{\mathrm{mix}}}{\sqrt{k}}\right)$. While this is longer than the lower bound $\frac{t_1^{\mathrm{mix}}}{k}$ given in Theorem 1.6, it is much shorter than the true transition period for many known examples giving cut-off. For instance, the transition period of random walk on the hypercube is faster than the mixing time by a factor which is logarithmic in the number of generators.

Our proofs of Theorems 1.1–1.7 approximate the distribution of random walk on the cycle $\mathbb{Z}/p\mathbb{Z}$ with that of a Gaussian diffusion on $\mathbb{R}^k/\Lambda$ where $\Lambda$ is a co-volume $p$ lattice. In making the transition between these models we use the following quantitative normal approximation lemma for which we don't know an easy reference in the literature. A proof is included in Appendix A.

**Lemma 1.10.** *Let $n, k(n) \geq 1$ with $k^2 = o(n)$ for large $n$. Let $\nu_k$ be the measure on $\mathbb{R}^k$ which is uniform on $\{0, \pm e_i,\ 1 \leq i \leq k\}$, where $e_i$ denotes the $i$th standard basis vector. For $\sigma > 0$ set*

$$\eta_k(\sigma, \underline{x}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{k}{2}} \exp\left(-\frac{\|\underline{x}\|_2^2}{2\sigma^2}\right)$$

*the standard Gaussian density. As $n \to \infty$ we have*

$$\left\| \nu_k^{*n} * \mathbf{1}_{\left[-\frac{1}{2},\frac{1}{2}\right)^k} - \eta_k\left(\sqrt{\frac{2n}{2k+1}}, \cdot\right) \right\|_{\mathrm{TV}(\mathbb{R}^k)} = o(1).$$

After transition to the diffusion model, the measure on lattices induced from the random choice in Theorem 1.7 is close to the uniform measure on the (rescaled) $p$-Hecke points, which are the index $p$ lattices of $\mathbb{Z}^k$. It is known that, after rescaling to volume 1, as $p \to \infty$ these lattices are equidistributed with respect to the induced Haar measure in the space $\mathrm{SL}_k(\mathbb{Z})\backslash \mathrm{SL}_k(\mathbb{R})$ of all volume 1 lattices in $\mathbb{R}^k$. Statistics regarding correlations of vectors in a random lattice are well-known, see for instance [25] for a modern treatment. Although we estimate somewhat different quantities, the results considered there may be useful in understanding our argument.

We conclude by giving an example of random walk on the cycle which has cut-off. This may be considered an approximate embedding of the classical hypercube walk into the cycle.

**Theorem 1.11.** *For $p \in \mathscr{P}$ let $\ell_2(p) = \lceil \log_2 p \rceil$ (logarithm base 2) and let the power-of-2 set be $A_{2,p} = \{0, \pm 1, \pm 2, ..., \pm 2^{\ell_2(p)-1}\} \subset \mathbb{Z}/p\mathbb{Z}$. Set*

$$c_0 = \sum_{j=1}^{\infty} \left(1 - \cos\frac{2\pi}{2^j}\right) = 3.394649802(1).$$

*The power-of-2 walk $(\mathbb{Z}/p\mathbb{Z}, \mu_{A_{2,p}}, \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}})_{p\in\mathscr{P}}$ has cut-off in total variation at mixing time*

$$t_1^{\mathrm{mix}}(p) \sim \frac{\ell_2(p)\log\ell_2(p)}{2c_0}.$$

### 1.2 Discussion of method

Our arguments view random walk on the cycle $\mathbb{Z}/p\mathbb{Z}$ with symmetric generating set $A$, $|A| = 2k+1$ as random walk on an index $p$ quotient of $\mathbb{Z}^k$, in which a standard basis vector is assigned to each non-zero symmetric pair $\{x, -x\}$ of generators. The index $p$ lattice is the set $\Lambda = \{\underline{n} \in \mathbb{Z}^k : \sum_x n_x x \equiv 0 \bmod p\}$. In the case of Theorem 1.11, the corresponding lattice is approximately cubic, and the argument is a perturbation of the Fourier analytic analysis of the hypercube walk in [11]. In particular, the mixing time and cut-off are the same in total variation and in $L^2$.

For $k \le \frac{\log p}{\log\log p}$, a random index $p$ lattice gives a mixing time in total variation which is less than the $L^2$ mixing time by a constant, and thus the $L^2$ methods of proving cut-off are not immediately suitable. Thus in our first four Theorems the arguments are made initially in time domain by first applying Lemma 1.10 to replace the discrete random walk with a diffusion on $\mathbb{R}^k/\Lambda$. This initial step is the reason for the restriction on the size of $k$ since the corresponding approximation fails for $k > (1+\epsilon)\frac{\log p}{\log\log p}$. For larger $k$ there is a standard method of correcting the approximation using the saddle point method, but we have not made an attempt to do so.

After having made the Gaussian approximation, Theorem 1.1 combines standard spectral estimates with bounds for the shortest vector in a lattice (the lower bound) and for sphere packing (the upper bound). Theorem 1.4 goes through in time domain, using convexity. Theorem 1.6 goes through in time domain, and uses an estimate for the derivative of the density in time.

Parts (1) and (2) of Theorem 4 study rare events in which the random lattice is essentially one dimensional due to the presense of many short vectors. We study these cases in frequency space. The dual lattice of an index $p$ lattice of $\mathbb{Z}^k$ is $\Lambda^\vee = \mathbb{Z}^k + \ell$ where

$$\ell = \ell_v = \{av : 0 < a < p\}, \qquad v \in \frac{1}{p}\mathbb{Z}^k \setminus \mathbb{Z}^k$$

is a line. We are able to show that with high probability the large Fourier coefficients arise from frequencies which are small multiples of a single vector. The analysis restricts attention to primitive vectors, and their multiples by *Farey fractions modulo* $p$, which are residues $bq^{-1} \bmod p$ in which $b$ and $q$ are bounded.

Part (3) of Theorem 4 is proven in time domain again. After removing a small $L^1$ error, the modified density may be estimated using a variance bound. In particular, our argument requires averages concerning pairs of short vectors in a random lattice which are discrete analogues of the averages performed by Siegel and Rogers [23], [22] regarding the distribution of vectors in a random lattice.

### 1.3 Possible extensions

From the point of view of mixing of Markov chains, an attractive open problem is to decide the Peres conjecture

$$\text{cut-off} \;\Leftrightarrow\; t^{\text{mix}}/t^{\text{rel}} \to \infty$$

for random walk on a cycle.

Abelian groups are prevalent in arithmetic, and there would be interest in extending the results to random walks on more general abelian groups. The class group of an imaginary quadratic field grows like the discriminant to the power $\frac{1}{2} + o(1)$, so a reweighting of Theorem 1.7 with roughly $d$ groups of order $d$ would be of interest. The techniques presented should translate without any great difficulty to studying random walk on cycles of composite order. The general case has not been considered, but see [28] for a study of random random walk on the hypercube.

To model abelian sandpiles, asymmetric generating sets should be considered.

## Notation and conventions

Given groups $G, H$, $H < G$ indicates that $H$ is a subgroup of $G$ and $[G : H]$ denotes the index. $\mathfrak{S}_k$ is the symmetric group on $k$ letters and we write $(\mathbb{Z}/2\mathbb{Z})^k \rtimes \mathfrak{S}_k = O_k(\mathbb{Z})$ for the $k \times k$ orthogonal group over $\mathbb{Z}$. For ring $R = \mathbb{Z}, \mathbb{Z}/p\mathbb{Z}$, $\mathrm{GL}_n(R)$ and $\mathrm{SL}_n(R)$ are the usual linear groups with entries in $R$. We denote $e(x) = e^{2\pi i x}$ the standard additive character on $\mathbb{R}/\mathbb{Z}$.

Given measure space $(\mathfrak{X}, \mathscr{B})$, $\mathscr{M}(\mathfrak{X})$ indicates the Borel probability measures on $\mathfrak{X}$. When $\mathfrak{X}$ is a finite set, $\mathbb{U}_{\mathfrak{X}}$ denotes the uniform probability measure on $\mathfrak{X}$ and when $\mathfrak{X}$ is a compact abelian group, $\mathbb{U}_{\mathfrak{X}}$ denotes the probability Haar measure. In either case expectation and variance with respect to $\mathbb{U}_{\mathfrak{X}}$ are indicated $\mathbf{E}_{\mathfrak{X}}$ and $\mathbf{Var}_{\mathfrak{X}}$. $\|\cdot\|_{\mathrm{TV}(\mathfrak{X})}$ indicates the total variation norm on $\mathscr{M}(\mathfrak{X})$.

Unless otherwise stated, $\|\cdot\|$ indicates the $\ell^2$-norm on $\mathbb{R}^k$, $k \geq 1$, $\|\cdot\|_p$ denotes the $\ell^p$ norm, $p \geq 1$, and $\|\cdot\|_{(\mathbb{R}/\mathbb{Z})^k}$ denotes the $\ell^2$ distance to the nearest integer lattice point. $\mathbb{S}^{k-1}$ is the unit sphere in $\mathbb{R}^k$, $\mathbb{S}^{k-1} = \{\underline{x} \in \mathbb{R}^k : \|\underline{x}\|_2 = 1\}$. Given $\underline{x} \in \mathbb{R}^k$, $R \in \mathbb{R}_{>0}$, and $p \geq 1$, $B_p(\underline{x}, R)$ denotes the $\ell^p$ ball

$$B_p(\underline{x}, R) = \left\{\underline{y} \in \mathbb{R}^k : \|\underline{y} - \underline{x}\|_p \leq R\right\},$$

the ambient dimension being clear from the context. If $p$ is not stated $\ell^2$ is assumed. Given further parameter $0 < \tau < 1$, $S(\underline{x}, R, \tau)$ indicates the spherical shell

$$S(\underline{x}, R, \tau) = \left\{\underline{y} \in \mathbb{R}^k : \|\underline{x} - \underline{y}\|_2 \in [(1-\tau)R, (1+\tau)R]\right\}.$$

For $k \geq 1$,

$$R_k = \left(\frac{\Gamma\left(\frac{k}{2}+1\right)}{\pi^{\frac{k}{2}}}\right)^{\frac{1}{k}} = \left(1 + \frac{\log(k+1)}{2k} + O\left(\frac{1}{k}\right)\right)\sqrt{\frac{k}{2\pi e}}$$

is the radius of an $\ell^2$ ball of unit volume in $\mathbb{R}^k$. One may check that $R_k > \sqrt{\frac{k}{2\pi e}}$ for all $k \geq 1$.

For $k \geq 1$, given $\underline{x} \in \mathbb{R}^k$ and $\sigma \in \mathbb{R}_{>0}$, $\eta_k(\sigma, \underline{x})$ denotes the density at $\underline{x}$ of a symmetric centered Gaussian distribution scaled by $\sigma$,

$$\eta_k(\sigma, \underline{x}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{k}{2}} \exp\left(-\frac{\|\underline{x}\|_2^2}{2\sigma^2}\right).$$

By default, quantities considered depend upon a large prime parameter $p$ varying over a set of primes $\mathscr{P}_0$. We use the Vinogradov notation $A \ll B$ with the same meaning as $A(p) = O(B(p))$. $A \asymp B$ means $A \ll B$ and $B \ll A$. For positive parameters $A, B$, $A \sim B$ means $\lim_{p\to\infty} \frac{A(p)}{B(p)} = 1$ and $A \lesssim B$, resp. $A \gtrsim B$ means $\limsup \frac{A(p)}{B(p)} \leq 1$, resp. $\liminf \frac{A(p)}{B(p)} \geq 1$. We also use the non-standard notation already introduced in the introduction $A \lesssim_x B$, with the meaning that there is a non-increasing function $f : \mathbb{R}^+ \to \mathbb{R}^+$ with $\lim_{x\to\infty} f(x) = 1$ such that $A(x) \leq f(x)B(x)$.

## 2 Background

This section collects together several statements regarding classical probability theory and lattice theory on $\mathbb{R}^k$, $k \geq 1$.

### 2.1 Classical probability

See [6] for background regarding random walk on a group and [20] for a thorough treatment of Markov chains. We have provided proofs of the statements which we use for the reader's convenience.

We have already introduced the total variation distance between two probability measures $\mu, \nu$ on a measure space $(\mathfrak{X}, \mathscr{B})$, by

$$\|\mu - \nu\|_{\mathrm{TV}(\mathfrak{X})} = \sup_{S \in \mathscr{B}} |\mu(S) - \nu(S)| .$$

In the case when $\mu$ has a density with respect to $\nu$, equivalent characterizations are

$$\|\mu - \nu\|_{\mathrm{TV}(\mathfrak{X})} = \frac{1}{2} \int_{\mathfrak{X}} \left| \frac{d\mu}{d\nu} - 1 \right| d\nu = \int_{\mathfrak{X}} \left( \frac{d\mu}{d\nu} - 1 \right) \mathbf{1} \left( \frac{d\mu}{d\nu} > 1 \right) d\nu.$$

When $\mu$ is the distribution of a Markov chain with stationary measure $\nu$ define the $L^2(d\nu)$ distance to stationarity by

$$\|\mu - \nu\|_{L^2(d\nu)} = \frac{1}{2} \left( \int \left( \frac{d\mu}{d\nu} - 1 \right)^2 d\nu \right)^{\frac{1}{2}}$$

with the convention that the norm is infinite if $\frac{d\mu}{d\nu}$ is not in $L^2(d\nu)$. The factor of $\frac{1}{2}$ is for consistency with the interpretation of total variation distance as half the $L^1(d\nu)$ norm. For $\epsilon > 0$ denote $t_2^{\mathrm{mix}}(\epsilon)$ the $\epsilon$-mixing time of the $L^2(d\nu)$ norm.

**Lemma 2.1.** *Convolution with a probability measure is a contraction in the total variation norm. Also, given symmetric probability measure $\mu$ on finite or compact abelian group $G$, for any $0 < \epsilon < 1$ the total variation mixing time of random walk driven by $\mu$ satisfies $t^{\mathrm{rel}} \log \frac{1}{2\epsilon} \le t_1^{\mathrm{mix}}(\epsilon) \le t_2^{\mathrm{mix}}(\epsilon)$ and $\frac{2\pi^2}{27} \epsilon^3 t^{\mathrm{rel}} \lesssim t_1^{\mathrm{mix}}(1-\epsilon)$ as $\epsilon \downarrow 0$.*

*Proof.* The contraction property follows from the triangle inequality.

To prove $t_1^{\mathrm{mix}}(\epsilon) \le t_2^{\mathrm{mix}}(\epsilon)$, use the $L^1$ characterization of the total variation metric and Cauchy-Schwarz

$$\|\mu^{*n} - \mathbb{U}_G\|_{\mathrm{TV}(G)} = \frac{1}{2} \int_G \left| \frac{d\mu^{*n}}{d\mathbb{U}_G} - 1 \right| d\mathbb{U}_G \le \|\mu^{*n} - \mathbb{U}_G\|_{L^2(d\mathbb{U}_G)}.$$

To prove the lower bounds regarding $t^{\mathrm{rel}}$, observe that the eigenvalues of the transition kernel for the random walk are given by

$$\mathrm{spec}(\mu) = \left\{ \mathbf{E}_\mu[\chi] : \chi \in \widehat{G} \right\},$$

where $\widehat{G}$ denotes the set of characters of $G$. Let $\chi_1$ generate the spectral gap. Since $\|\chi_1\|_\infty \le 1$, we have, for any $n \ge 1$,

$$\|\mu^{*n} - \mathbb{U}_G\|_{\mathrm{TV}(G)} \ge \frac{1}{2} \left| \mathbf{E}_{\mu^{*n}}[\chi_1] \right| = \frac{1}{2} \left| (\mathbf{E}_\mu[\chi_1])^n \right|,$$

so that the first mixing time bound follows by taking logarithms.

To obtain the bound for $t_1^{\mathrm{mix}}(1-\epsilon)$, let $\epsilon_0 > \epsilon_1$ be small parameters, satisfying, for some $A, B > 0$, $\epsilon_0 = A\epsilon^2$, $\epsilon_1 = B\epsilon^3$. Let $n$ be maximal such that $\mathbf{E}_{\mu^{*n}}[\chi_1] \ge 1 - \epsilon_1$. Set $S = \{g \in G : \mathrm{Re}(\chi_1(g)) \ge 1 - \epsilon_0\}$ and $\alpha = \mu^{*n}(S)$. Bounding $\mathrm{Re}\,(\chi|_S) \le 1$ and $\mathrm{Re}\,(\chi|_{S^c}) \le 1 - \epsilon_0$,

$$(1 - \epsilon_1) \le \mathbf{E}_{\mu^{*n}}[\chi_1] \le \alpha + (1 - \epsilon_0)(1 - \alpha)$$

whence $\alpha \ge 1 - \frac{\epsilon_1}{\epsilon_0}$. According to uniform measure, $\mathrm{Re}(\chi)$ has the same distribution as $\cos(2\pi x)$ on $(\mathbb{R}/\mathbb{Z}, dx)$, so that

$$\mathbb{U}_G(S) = \frac{\cos^{-1}(1 - \epsilon_0)}{\pi} = \frac{\sqrt{2\epsilon_0}}{\pi}(1 + O(\epsilon_0)).$$

It follows that

$$\begin{aligned}\|\mu^{*n} - \mathbb{U}_G\|_{\mathrm{TV}(G)} &\geq \mu^{*n}(S) - \mathbb{U}_G(S) \\ &\geq 1 - \frac{\epsilon_1}{\epsilon_0} - \frac{\cos^{-1}(1-\epsilon_0)}{\pi} = 1 - \left(\frac{B}{A} + \frac{\sqrt{2A}}{\pi} + O(\epsilon^2)\right)\epsilon.\end{aligned}$$

Imposing the constraint $\mu^{*n}(S) - \mathbb{U}_G(S) \geq 1 - \epsilon$ gives $t_1^{\mathrm{mix}}(1-\epsilon) \geq n+1$. As $\epsilon \downarrow 0$, one obtains the constraint $\left(\frac{B}{A} + \frac{\sqrt{2A}}{\pi} + O(\epsilon^2)\right) < 1$, which gives the asymptotic claimed with $A \sim \frac{2\pi^2}{9}$, $B \sim \frac{2\pi^2}{27}$. □

Define the standard symmetric centered normal distribution on $\mathbb{R}^k$ scaled by $\sigma \in \mathbb{R}_{>0}$ to be

$$\eta_k(\sigma, \underline{x}) = \frac{1}{(2\pi\sigma^2)^{\frac{k}{2}}} \exp\left(\frac{-\|\underline{x}\|_2^2}{2\sigma^2}\right).$$

For $t \in \mathbb{R}_{>0}$, $\eta(\sqrt{t}\sigma, \underline{x})$ is its $t$-fold convolution. We use several results regarding concentration of the Gaussian measure.

**Lemma 2.2.** *Let $k \geq 1$ and $\sigma > 0$. There are positive constants $C, \{C_p\}_{2 \leq p < \infty}$ such that, for any $t > C$,*

$$\int_{\underline{x} \in \mathbb{R}^k} \eta_k(\sigma, \underline{x}) \mathbf{1}\left(\left|\|\underline{x}\|_2 - \sigma\sqrt{k}\right| > \sigma t\right) d\underline{x} \leq \exp\left(-\frac{(t-C)^2}{2}\right),$$

*and, for all $t > 0$, for all $2 \leq p < \infty$,*

$$\int_{\underline{x} \in \mathbb{R}^k} \eta_k(\sigma, \underline{x}) \mathbf{1}\left(\|\underline{x}\|_p > C_p \sigma k^{\frac{1}{p}} + t\sigma\right) d\underline{x} \leq \exp\left(-\frac{t^2}{2}\right).$$

*Proof.* All quantities scale with $\sigma$ so we may assume $\sigma = 1$. Let $\gamma_k$ denote the measure on $\mathbb{R}^k$ with density $\gamma_k(\underline{x}) = \frac{1}{(2\pi)^{\frac{k}{2}}} \exp\left(-\frac{\|\underline{x}\|_2^2}{2}\right)$. Let $M_p$, $2 \leq p < \infty$ denote the median with respect to $\gamma_k$ of $\|\cdot\|_p$, that is, $\gamma_k\left(\underline{x} : \|\underline{x}\|_p \leq M_p\right) = \frac{1}{2}$. Since $\|\cdot\|_p$ is 1-Lipschitz on $(\mathbb{R}^k, \|\cdot\|_2)$ for $p \geq 2$, Talagrand's inequality ([19], p.21) gives, for any $t > 0$,

$$\gamma_k\left(\underline{x} : |\|\underline{x}\|_p - M_p| > t\right) \leq \exp\left(-\frac{t^2}{2}\right).$$

The first statement follows, since the mean, root mean square, and median of $\|\cdot\|_2$ differ by constants, as is evident from the concentration around the median. The second statement follows since $M_p \ll k^{\frac{1}{p}}$. □

## 2.2 Lattices

Siegel's *Lectures on the Geometry of Numbers* [24] are a recommended reference. A lattice $\Lambda < \mathbb{R}^k$ is a discrete finite co-volume subgroup of $\mathbb{R}^k$. Write

$$\mathrm{vol}(\Lambda) = \int_{\mathbb{R}^k/\Lambda} d\underline{x}$$

for its co-volume. Fixing the usual inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^k$, the dual lattice of lattice $\Lambda$ is

$$\Lambda^\vee = \left\{\lambda' \in \mathbb{R}^k : \forall\, \lambda \in \Lambda, \langle \lambda', \lambda \rangle \in \mathbb{Z}\right\}.$$

This satisfies $\mathrm{vol}(\Lambda) \cdot \mathrm{vol}(\Lambda^\vee) = 1$. For instance, the dual lattice to $\Lambda = 2\mathbb{Z}$ is $\frac{1}{2}\mathbb{Z}$. More generally, if $\Lambda = Q\mathbb{Z}^k$ for some $Q \in \mathrm{GL}_k(\mathbb{R})$, then $\Lambda^\vee = (Q^{-1})^t\mathbb{Z}^k$. We reserve $\lambda^*$ for the shortest non-zero vector of $\Lambda^\vee$.

Given lattice $\Lambda < \mathbb{R}^k$, its norm-minimal fundamental domain (Voronoi cell) is

$$\mathscr{F}(\Lambda) = \{\underline{x} \in \mathbb{R}^k : \forall\, \lambda \in \Lambda \setminus \{0\}, \|\underline{x}\| < \|\underline{x} - \lambda\|\}.$$

One may choose a set $\mathscr{F}^0(\Lambda)$

$$\mathscr{F}(\Lambda) \subset \mathscr{F}^0(\Lambda) \subset \overline{\mathscr{F}(\Lambda)}$$

such that every $\underline{x} \in \mathbb{R}^k/\Lambda$ has a unique representative in $\mathscr{F}^0(\Lambda)$.

Minkowski's geometry of numbers gives an upper bound for the shortest non-zero vector in a lattice.

**Theorem 2.3** (Minkowski's Theorem)**.** *Let $\Lambda \subset \mathbb{R}^k$ be a lattice and let $C$ be a convex symmetric body, i.e. $\underline{x} \in C \Leftrightarrow -\underline{x} \in C$. If*

$$\mathrm{vol}(C) > 2^k \mathrm{vol}(\Lambda)$$

*then $C$ contains a non-zero vector in $\Lambda$. In particular*

$$\min_{\lambda \in \Lambda \setminus \{0\}} \|\lambda\|_2 \leq \frac{2}{\sqrt{\pi}} \left(\Gamma\left(\frac{k}{2}+1\right)\mathrm{vol}(\Lambda)\right)^{\frac{1}{k}} \sim \sqrt{\frac{2k}{\pi e}}\,\mathrm{vol}(\Lambda)^{\frac{1}{k}}.$$

*with the asymptotic holding as $k \to \infty$.*

For lattice $\Lambda$, the diameter of the norm-minimal fundamental domain and the shortest non-zero vector in the dual lattice are related as follows.

**Lemma 2.4.** *Let $\Lambda$ be a lattice with norm-minimal fundamental domain $\mathscr{F}$ and dual lattice $\Lambda^\vee$. Let $\lambda^*$ be the shortest non-zero vector in $\Lambda^\vee$. We have*

$$\|\lambda^*\|_2 \cdot \mathrm{diam}(\mathscr{F}) \geq 1.$$

*Proof.* Let $\underline{v} = \frac{\lambda^*}{\|\lambda^*\|_2}$ and choose $\underline{x}$ the point on the boundary of $\mathscr{F}$ on the ray determined by $\underline{v}$. Write $\underline{x} = x_0\underline{v}$. Since $\underline{x} \in \partial(\mathscr{F})$ we may find $\underline{y} \in \Lambda \setminus \{0\}$ with $\left|\langle \underline{x}, \underline{y}\rangle\right| = \frac{1}{2}\|\underline{y}\|_2^2$. Set $\underline{y} = y_0\underline{v} + \underline{v}'$ where $\langle \underline{v}, \underline{v}'\rangle = 0$. In particular, $y_0 \neq 0$ so $\left|\langle \underline{y}, \lambda^*\rangle\right| = \|\lambda^*\|_2|y_0| \geq 1$. Since $|x_0y_0| \geq \frac{1}{2}y_0^2$ it follows that $\|\underline{x}\|_2 \cdot \|\lambda^*\|_2 \geq \frac{1}{2}$. The diameter is at least as large as $2\|\underline{x}\|_2$. □

Given $\underline{x} \in \mathbb{R}^k$ and $R > 0$, let $B_2(\underline{x}, R)$ denote the ball

$$B_2(\underline{x}, R) = \{\underline{y} \in \mathbb{R}^k : \|\underline{x} - \underline{y}\|_2 \leq R\}.$$

The following is an easy estimate for the number of lattice points contained in a ball.

**Lemma 2.5.** *Let $k \in \mathbb{Z}_{>0}$, let $\underline{x} \in \mathbb{R}^k$ and let $R > k^{\frac{3}{2}}$. Then*

$$\left|\mathbb{Z}^k \cap B_2(\underline{x}, R)\right| = \left(1 + O\left(\frac{k^{\frac{3}{2}}}{R}\right)\right)\mathrm{vol}(B_2(\underline{x}, R)).$$

*Proof.* Let $\mu_{\underline{x},R} = \sum_{\underline{n} \in \mathbb{Z}^k \cap B_2(\underline{x},R)} \delta_{\underline{n}}$. Since the hypercube $\left[-\frac{1}{2}, \frac{1}{2}\right)^k$ has diameter $\sqrt{k}$,

$$\mathbf{1}_{B_2(\underline{x}, R-\sqrt{k})} \leq \mu_{\underline{x},R} * \mathbf{1}_{\left[-\frac{1}{2},\frac{1}{2}\right)^k} \leq \mathbf{1}_{B_2(\underline{x}, R+\sqrt{k})}$$

and thus

$$\begin{aligned}
\left|\mathbb{Z}^k \cap B_2(\underline{x}, R)\right| &= \int_{\mathbb{R}^k} \mu_{\underline{x},R} * \mathbf{1}_{\left[-\frac{1}{2},\frac{1}{2}\right)^k} \\
&= \left(1 + O\left(\frac{\sqrt{k}}{R}\right)\right)^k \mathrm{vol}(B_2(\underline{x}, R)) \\
&= \left(1 + O\left(\frac{k^{\frac{3}{2}}}{R}\right)\right)\mathrm{vol}(B_2(\underline{x}, R)).
\end{aligned}$$

□

We also use the following estimate counting lattice points of a more general lattice.

**Lemma 2.6.** *Let $\Lambda < \mathbb{R}^k$ be a lattice with shortest non-zero vector $\lambda^*$. For any $t \geq 1$,*

$$\log |\Lambda \cap B_2(0, t\|\lambda^*\|)| \lesssim_k k \left[ \frac{1+\sin\theta}{2\sin\theta} \log \frac{1+\sin\theta}{2\sin\theta} - \frac{1-\sin\theta}{2\sin\theta} \log \frac{1-\sin\theta}{2\sin\theta} \right],$$

*where $\theta = 2\sin^{-1}\left(\frac{1}{2t}\right)$.*

*Proof.* This follows from [18], see [3] for a nice exposition and related results. We sketch the argument.

Write $B_{2,j}$ for a $\ell^2$ ball in $\mathbb{R}^j$. By rescaling we may assume $\|\lambda^*\|_2 = 1$. View $\mathbb{R}^k$ as a hyperplane through zero in $\mathbb{R}^{k+1}$, and consider the ball $\tilde{B} = B_{2,k+1}(0,t)$ in $\mathbb{R}^{k+1}$. Project $\Lambda \cap B_{2,k}(0,t)$ orthogonally onto $\tilde{B}$. The points remain 1-spaced and thus satisfy an angular spacing of at least $\theta = 2\sin^{-1}(\frac{1}{2t})$. Let, as in [18], $A(n,\theta)$ denote the largest set $S \subset \mathbb{S}^{n-1}$ which is separated by angle $\theta$ as above. Thus

$$\Lambda \cap B_{2,k}(0,t) \leq A(k+1,\theta).$$

The claimed estimate for $A(k+1,\theta)$ is the main result of [18]. □

Given a probability measure $\mu \in \mathscr{M}(G)$, $G = \mathbb{Z}^k$ or $G = \mathbb{R}^k$, and a lattice $\Lambda < G$ the quotient measure $\mu_\Lambda$ is defined for $f \in C(G/\Lambda)$ by

$$\langle f, \mu_\Lambda \rangle_{G/\Lambda} = \langle f, \mu \rangle_G.$$

Quotienting commutes with convolution and contracts the total variation norm. For lattice $\Lambda < \mathbb{R}^k$, $t \in \mathbb{R}_{>0}$ and $\underline{x} \in \mathbb{R}^k$, the quotient measure of Gaussian $\eta_k\left(\sqrt{t}, \cdot\right)$ is the theta function

$$\Theta(\underline{x}, t; \Lambda) = \sum_{\lambda \in \Lambda} \eta_k\left(\sqrt{t}, \underline{x} + \lambda\right).$$

This has a representation in frequency space as

$$\Theta(\underline{x}, t; \Lambda) = \frac{1}{\mathrm{vol}(\Lambda)} \sum_{\lambda \in \Lambda^\vee} \exp\left(-2\pi^2 t \|\lambda\|_2^2\right) e(\lambda \cdot \underline{x}).$$

To check the expansion, Fourier expand $\Theta$ in the orthonormal basis $\left\{ \frac{e(\lambda \cdot \underline{x})}{\sqrt{\mathrm{vol}(\Lambda)}} \right\}_{\lambda \in \Lambda^\vee}$ for $L^2(\mathbb{R}^k/\Lambda)$ (this is the usual proof of the Poisson summation formula). In the case of a cubic lattice, where for some $\alpha \in \mathbb{R}_{>0}$, $\Lambda = \alpha\mathbb{Z}^k$, the theta function is particularly pleasant.

**Lemma 2.7.** *Let $k \in \mathbb{Z}_{>0}$, $\alpha, t \in \mathbb{R}_{>0}$ and $\underline{x} \in \mathbb{R}^k$. We have*

$$\Theta\left(\underline{x}, t; \alpha\mathbb{Z}^k\right) = \prod_{i=1}^k \Theta\left(x_i, t; \alpha\mathbb{Z}\right).$$

*The one dimensional theta function $\Theta(x, t; \alpha\mathbb{Z})$ satisfies*

$$\begin{aligned}\Theta(x, t; \alpha\mathbb{Z}) &= \frac{\exp\left(-\frac{\alpha^2 \left\| \frac{x}{\alpha} \right\|_{\mathbb{R}/\mathbb{Z}}^2}{2t}\right)}{\sqrt{2\pi t}} + O\left( \frac{\exp\left(-\frac{\alpha^2}{8t}\right)}{\sqrt{2\pi t}\left(1 - \exp\left(-\frac{\alpha^2}{8t}\right)\right)} \right) \\ &= \frac{1}{\alpha} + O\left( \frac{\exp\left(-\frac{2\pi^2 t}{\alpha^2}\right)}{\alpha\left(1 - \exp\left(-\frac{2\pi^2 t}{\alpha^2}\right)\right)} \right).\end{aligned}$$

*Proof.* The factorization is immediate from the definition of $\Theta$. The first estimate for $\Theta(x,t;\alpha\mathbb{Z})$ is the result of pulling out the largest term and bounding the remaining terms by a geometric progression. For the second, apply the Poisson summation formula,

$$\sum_{n\in\mathbb{Z}} \eta_1\left(\sqrt{t}, x+\alpha n\right) = \frac{1}{\alpha}\sum_{n\in\mathbb{Z}} \exp\left(-\frac{2\pi^2 t n^2}{\alpha^2}\right) e\left(\frac{xn}{\alpha}\right)$$

and bound the $n \neq 0$ terms by a geometric progression. □

### 2.3 Identification between generating sets and lattices

Our proofs of Theorems 1.1– 1.7 approximate random walk on $\mathbb{Z}/p\mathbb{Z}$ with symmetric generating set $A$, $|A| = 2k+1$ with a Gaussian diffusion on $\mathbb{R}^k/\Lambda$ where $\Lambda$ is a co-volume $p$ lattice. The reduction is as follows.

Let $O_k(\mathbb{Z}) \cong (\mathbb{Z}/2\mathbb{Z})^k \rtimes \mathfrak{S}_k$ be the orthogonal group over $\mathbb{Z}$ consisting of signed $k\times k$ permutation matrices, which acts naturally on $\mathbb{R}^k$. Let

$$L = L(p,k) = \{\Lambda < \mathbb{Z}^k : [\mathbb{Z}^k : \Lambda] = p\}$$
$$\mathscr{L} = \mathscr{L}(p,k) = O_k(\mathbb{Z})\backslash L(p,k)$$

be the set of index-$p$ lattices of $\mathbb{Z}^k$, resp. those lattices up to $O_k(\mathbb{Z})$-equivalence. The action is matrix multiplication on the left applied to lattice vectors. Define subsets

$$L^0(p,k) = \left\{\Lambda \in L(p,k) : \lambda \in \Lambda\setminus\{0\} \;\Rightarrow\; \|\lambda\|_2^2 > 2\right\}$$
$$\mathscr{L}^0(p,k) = O_k(\mathbb{Z})\backslash L^0(p,k).$$

Let

$$A(p,k) = \left\{\underline{a} \in (\mathbb{F}_p^\times)^k : \forall 1 \le i < j \le k,\, a_i \neq \pm a_j\right\}.$$

$\mathscr{A}(p,k)$ may be identified with $O_k(\mathbb{Z})\backslash A(p,k)$ by interpreting the factors of $(\mathbb{Z}/2\mathbb{Z})^k$ as flipping signs, and the factor of $\mathfrak{S}_k$ as rearranging the order of the coordinates in the vector. Evidently the action is free, so that uniform measure on $A(p,k)$ descends to uniform measure on $\mathscr{A}(p,k)$.

$\mathbb{F}_p^\times$ acts freely on $A(p,k)$ dilating all coordinates simultaneously. $\mathbb{F}_p^\times\backslash A(p,k)$ and $L^0(p,k)$ are in bijection via the map

$$A(p,k) \ni \underline{a} \overset{\phi}{\mapsto} \Lambda(\underline{a}) = \left\{\underline{n} \in \mathbb{Z}^k : \sum_{i=1}^k n_i a_i \equiv 0 \bmod p\right\} \in L^0(p,k).$$

The map in the reverse direction is

$$\Lambda \overset{\phi}{\mapsto} \underline{a}(\Lambda) = \{1, a_2, \cdots, a_k : \forall i,\, e_1 - a_i e_i \equiv 0 \bmod p\}.$$

It follows that uniform measure on $A(p,k)$ pushes forward to uniform measure on $L^0(p,k)$. $O_k(\mathbb{Z})$ acts on $L^0(p,k)$, and we obtain a map $\mathbb{F}_p^\times\backslash\mathscr{A}(p,k) \overset{\bar{\phi}}{\mapsto} \mathscr{L}^0(p,k)$ which we write as $\Lambda(A)$. Note that the joint action of $\mathbb{F}_p^\times \times O_k$ on $A(p,k)$ need not be free, but this will not concern us. We write $\mathbb{U}_L, \mathbb{U}_{L^0}$ for uniform measure on $L$ and $L^0$.

Let $\nu = \nu_k \in \mathscr{M}(\mathbb{Z}^k)$ be the uniform measure on $S_k = \{0, \pm e_1, ..., \pm e_k\}$, $e_i$ the $i$th standard basis vector. Let $A \in \mathscr{A}(p,k)$. For any $n \ge 1$ the law of $\mu_A^{*n}$ on $\mathbb{Z}/p\mathbb{Z}$ and $(\nu_k^{*n})_{\Lambda(A)}$ on $\mathbb{Z}^k/\Lambda(\underline{a})$ are equal. The above observations imply that we may sample the laws of $\mu_A^{*n}$ with $A$ chosen according to $\mathbb{U}_{\mathscr{A}(p,k)}$ by instead sampling the laws of $(\nu_k^{*n})_\Lambda$ with $\Lambda$ drawn according to $\mathbb{U}_{L^0(p,k)}$.

Combining this discussion with Minkowski's theorem has the following consequence.

**Lemma 2.8.** *Let $p$ be a large prime, let $1 \le k < \frac{2\log p}{\log\log p}$ and let $A \in \mathscr{A}(p,k)$. Let $\Lambda < \mathbb{Z}^k$ be any lattice in the class of $\Lambda(A) \in \mathscr{L}$, and let*

$$\ell(A) = \min\{\|\lambda\|_2 : 0 \neq \lambda \in \Lambda^\vee\}.$$

*The relaxation time of random walk driven by $\mu_A$ on $\mathbb{Z}/p\mathbb{Z}$ satisfies*

$$t^{\mathrm{rel}} \sim \frac{2k+1}{4\pi^2 \ell(A)^2}.$$

*Proof.* The characters of $\mathbb{Z}^k/\Lambda$ are given by the dual group, $\Lambda^\vee/\mathbb{Z}^k$. Let $\lambda^* = (\lambda_1, ..., \lambda_k)$ be a vector of minimal length in $\Lambda^\vee \setminus \{0\}$. The claim follows on noting that the spectral gap is given by

$$1 - \hat{\nu}_\Lambda(\lambda^*) = \frac{1}{2k+1}\sum_{j=1}^{k}(2 - 2\cos(2\pi\lambda_j)) = \frac{4\pi^2}{2k+1}\sum_{j=1}^{k}\left(\lambda_j^2 + O(\lambda_j^4)\right).$$

The error is of lower order since $\|\lambda^*\|_\infty \ll \sqrt{k}p^{-\frac{1}{k}} = o(1)$ by Minkowski's Theorem. □

Lemma 1.10 from the introduction has the following consequence.

**Lemma 2.9.** *Let $p \ge 3$ be a prime, let $1 \le k \le \frac{\log p}{\log\log p}$ and let $A \in \mathscr{A}(p,k)$ with $\Lambda < \mathbb{Z}^k$ any representative of $\Lambda(A) \in \mathscr{L}^0(p,k)$. There is a function $\epsilon : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ with $\lim_{x\to\infty}\epsilon(x) = 0$, such that, for $n \ge 1$*

$$\left\|\mu_A^{*n} - \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}}\right\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})} = \left\|\Theta\left(\cdot, \frac{2n}{2k+1}; \Lambda\right) - \mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)} + O\left(\epsilon\left(\frac{n}{k^2}\right)\right).$$

*Proof.* Write $\Lambda = \Lambda(A)$ and $\mathbf{1}_{\left[-\frac{1}{2},\frac{1}{2}\right)^k}$ for the indicator function of the cube $\left[-\frac{1}{2},\frac{1}{2}\right)^k \subset \mathbb{R}^k$. We have

$$\left\|\mu_A^{*n} - \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}}\right\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})} = \left\|\nu_\Lambda^{*n} - \mathbb{U}_{\mathbb{Z}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{Z}^k/\Lambda)} = \left\|\nu_\Lambda^{*n} * \mathbf{1}_{\left[-\frac{1}{2},\frac{1}{2}\right)^k} - \mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)}$$

and

$$\begin{aligned}
\left|\left\|\nu_\Lambda^{*n} * \mathbf{1}_{\left[-\frac{1}{2},\frac{1}{2}\right)^k} - \mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)} - \left\|\Theta\left(\underline{x}, \frac{2n}{2k+1}; \Lambda\right) - \mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)}\right| & \\
\le \left\|\nu_\Lambda^{*n} * \mathbf{1}_{\left[-\frac{1}{2},\frac{1}{2}\right)^k} - \Theta\left(\underline{x}, \frac{2n}{2k+1}; \Lambda\right)\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)} & \\
\le \left\|\nu^{*n} * \mathbf{1}_{\left[-\frac{1}{2},\frac{1}{2}\right)^k} - \eta_k\left(\sqrt{\frac{2n}{2k+1}}, \cdot\right)\right\|_{\mathrm{TV}(\mathbb{R}^k)} &
\end{aligned}$$

by two applications of the triangle inequality. The bound now follows from Lemma 1.10. □

Combining the pieces above we prove the following lemma which is the main reduction in this section.

**Lemma 2.10.** *Let $0 < \epsilon < 1$, and let $k = k(p)$ satisfy $1 \le k \le \frac{\log p}{\log\log p}$. For any set $A \in \mathscr{A}(p,k)$ with uniform measure $\mu_A$ of total variation mixing time $t_1^{\mathrm{mix}}(\epsilon)$, we have, as $p \to \infty$, for all $n \ge t_1^{\mathrm{mix}}(\epsilon)$*

$$\left\|\mu_A^{*n}(x) - \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}}\right\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})} = \left\|\Theta\left(\underline{x}, \frac{2n}{2k+1}; \Lambda(A)\right) - \mathbb{U}_{\mathbb{R}^k/\Lambda(A)}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda(A))} + o_\epsilon(1).$$

*Proof.* By Minkowski's geometry of numbers, the shortest non-zero vector in the dual lattice $\Lambda(A)^\vee$ has length

$$\ell(A) \ll \sqrt{k} p^{\frac{-1}{k}}$$

so that Lemmas 2.1 and 2.8 give for the discrete walk $t_1^{\mathrm{mix}}(\epsilon) \gg t^{\mathrm{rel}} \gg p^{\frac{2}{k}}$. The claim now follows from Lemma 2.9, since $k = o\left(p^{\frac{1}{k}}\right)$. □

## 3 Mixing time estimates

Let $p$ be prime, $A \in \mathscr{A}(p)$ with $|A| = 2k+1$ and $1 \le k \le \frac{\log p}{\log\log p}$. Let $\Lambda = \Lambda(A)$ be any lattice associated to $A$ in $\mathbb{Z}^k$, as above.

*Proof of Theorem 1.1.* Theorem 1.1 is contained in the set of estimates

$$\tau_0 \frac{2k+1}{16\pi\Gamma\left(\frac{k}{2}+1\right)^{\frac{2}{k}}} p^{\frac{2}{k}} \lesssim_p \tau_0 t^{\mathrm{rel}} \lesssim_p t_1^{\mathrm{mix}} \lesssim_k 0.163 k t^{\mathrm{rel}}.$$

since

$$\frac{2k+1}{16\pi\Gamma\left(\frac{k}{2}+1\right)^{\frac{2}{k}}} \to \frac{e}{4\pi}, \qquad k \to \infty.$$

Combining Lemma 2.8 and Minkowski's theorem gives

$$t^{\mathrm{rel}} \sim \frac{2k+1}{4\pi^2 \ell(A)^2} \ge \frac{2k+1}{16\pi\Gamma\left(\frac{k}{2}+1\right)^{\frac{2}{k}}} p^{\frac{2}{k}}.$$

The estimate $t^{\mathrm{rel}}(1-\log 2) \le t_1^{\mathrm{mix}}$ is given in Lemma 2.1. To replace $(1-\log 2)$ with the larger constant $\tau_0$, consider the theta function $\Theta\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right)$, which has asymptotically the same relaxation time as $\mu_A$ by Lemma 2.8. Let $\lambda^*$ be a shortest non-zero vector in the dual space, and consider

$$\Theta_0\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right) = \frac{1}{p} \sum_{j \in \mathbb{Z}} \exp\left(\frac{-4\pi^2 t}{2k+1} \|\lambda^*\|_2^2 j^2\right) e(j\lambda^* \cdot \underline{x}),$$

which is found by projecting $\Theta$ in frequency space onto the line determined by $\lambda^*$. Equivalently, identify $\mathbb{R}^{k-1}$ with $\mathbb{R}^k \cap (\lambda^*)^\perp$ and let $\eta_{k-1}(T, \cdot)$ denote a Gaussian of covariance matrix $T^2 I$ on this space. Write $\lambda \in \Lambda^\vee$ as $\lambda = \lambda_1 + \lambda_2$ where $\lambda_1$ is the projection to the span of $\lambda^*$ and $\lambda_2$ is orthogonal to $\lambda^*$. One has, for $T > 0$,

$$\begin{aligned}
&\int_{\mathbb{R}^k \cap (\lambda^*)^\perp} \eta_{k-1}(T, \underline{y}) \Theta\left(\underline{x} + \underline{y}, \frac{2t}{2k+1}; \Lambda\right) d\underline{y} \\
&= \sum_{\lambda \in \Lambda^\vee} \exp\left(-\frac{4\pi^2 t}{2k+1} \|\lambda\|_2^2 - 2\pi^2 T^2 \|\lambda_2\|_2^2\right) e(\lambda_1 \cdot x)
\end{aligned}$$

and thus

$$\Theta_0\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right) = \lim_{T \to \infty} \int_{\mathbb{R}^k \cap (\lambda^*)^\perp} \eta_{k-1}(T, \underline{y}) \Theta\left(\underline{x} + \underline{y}, \frac{2t}{2k+1}; \Lambda\right) d\underline{y}.$$

The convergence is uniform in $\underline{x}$ as the error at $T$ is dominated by the case in which $\underline{x}$ is orthogonal to $\lambda^*$ so that all the terms are positive. This justifies exchanging the limit and integral in the following calculation. Let $\mathscr{F}$ be a fundamental domain for $\mathbb{R}^k/\Lambda$.

$$\begin{aligned}
&\left\| \Theta_0\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right) - \mathbb{U}_{\mathbb{R}^k/\Lambda} \right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)} = \frac{1}{2} \int_{\mathscr{F}} \left| \Theta_0\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right) - \frac{1}{p} \right| d\underline{x} \\
&= \lim_{T \to \infty} \frac{1}{2} \int_{\mathscr{F}} \left| \int_{\mathbb{R}^k \cap (\lambda^*)^\perp} \eta_{k-1}(T, \underline{y}) \left( \Theta\left(\underline{x} + \underline{y}, \frac{2t}{2k+1}; \Lambda\right) - \frac{1}{p} \right) d\underline{y} \right| d\underline{x}
\end{aligned}$$

Applying the triangle inequality,

$$\begin{aligned}\left\|\Theta_0 - \mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)} &\le \lim_{T\to\infty} \frac{1}{2}\int_{\mathscr{F}}\int_{\mathbb{R}^k\cap(\lambda^*)^\perp} \eta_{k-1}(T,\underline{y})\left|\Theta\left(\underline{x}+\underline{y},\frac{2t}{2k+1};\Lambda\right)-\frac{1}{p}\right| d\underline{y}d\underline{x}\\ &= \lim_{T\to\infty}\int_{\mathbb{R}^k\cap(\lambda^*)^\perp}\eta_{k-1}(T,\underline{y})\left\|\Theta-\mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)} d\underline{y}\\ &= \left\|\Theta-\mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)}.\end{aligned}$$

Let

$$\theta(x,t)=\sum_{j\in\mathbb{Z}}\exp(-2\pi^2 tj^2)e(jx)$$

denote the time $t$ Gaussian diffusion on $\mathbb{R}/\mathbb{Z}$. For $t>0$,

$$\left\|\Theta_0\left(\cdot,\frac{t}{\|\lambda^*\|_2^2};\Lambda\right)-\mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)} = \|\theta(\cdot,t)-\mathbb{U}_{\mathbb{R}/\mathbb{Z}}\|_{\mathrm{TV}(\mathbb{R}/\mathbb{Z})}.$$

Since the latter distance is monotonically decreasing and smooth, and since for $n\ge t_1^{\mathrm{mix}}$,

$$\left\|\mu_A^{*n}-\mathbb{U}_{\mathbb{Z}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{Z}^k/\Lambda)} = \left\|\Theta\left(\cdot,\frac{2n}{2k+1};\Lambda\right)-\mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)}+o(1)$$

by Lemma 2.10, it follows that $t_1^{\mathrm{mix}}\gtrsim\tau_0 t^{\mathrm{rel}}$.

To give the spectral upper bound for $t_1^{\mathrm{mix}}$, again consider instead the distance from uniformity of $\Theta\left(\cdot,\frac{2n}{2k+1};\Lambda\right)$ on $\mathbb{R}^k/\Lambda$. For $t>0$,

$$\left\|\Theta\left(\cdot,\frac{2t}{2k+1};\Lambda\right)-\mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|^2_{\mathrm{TV}(\mathbb{R}^k/\Lambda)} \le \frac{1}{4}\sum_{\lambda\in\Lambda^\vee\setminus\{0\}}\exp\left(-\frac{8\pi^2 t\|\lambda\|_2^2}{2k+1}\right)$$

Writing the sum as a Stieltjes integral, then integrating by parts, the right hand side becomes

$$\begin{aligned}&\frac{1}{4}\int_{s=1^-}^{\infty}\exp\left(-\frac{8\pi^2 t\|\lambda^*\|_2^2 s^2}{2k+1}\right)d\left(|\Lambda^\vee\cap B_2(0,s\|\lambda^*\|)|\right) \qquad (3.1)\\ &=\frac{4\pi^2 t\|\lambda^*\|_2^2}{2k+1}\int_{1-}^{\infty} s\exp\left(-\frac{8\pi^2 t\|\lambda^*\|_2^2 s^2}{2k+1}\right)|\Lambda^\vee\cap B_2(0,s\|\lambda^*\|)|\,ds.\end{aligned}$$

Set $t=\tau\frac{2k+1}{4\pi^2\|\lambda^*\|_2^2}$ so that $\tau\sim\frac{t}{t^{\mathrm{rel}}}$. Thus (3.1) simplifies to

$$\begin{aligned}(3.1) &= \tau\int_{1-}^{\infty} s\exp\left(-2\tau s^2\right)|\Lambda^\vee\cap B_2(0,s\|\lambda^*\|)|\,ds\\ &\le \tau\int_{1-}^{\infty} s\exp\left(-2\tau s^2+(1+\varepsilon(k))kF(s)\right)ds\end{aligned}$$

where $\varepsilon(k)\to 0$ as $k\to\infty$, and

$$F(s)=\left[\frac{1+\sin\theta}{2\sin\theta}\log\frac{1+\sin\theta}{2\sin\theta}-\frac{1-\sin\theta}{2\sin\theta}\log\frac{1-\sin\theta}{2\sin\theta}\right],\qquad \theta(s)=2\sin^{-1}\left(\frac{1}{2s}\right)$$

see Lemma 2.6. The maximum of $\frac{F(s)}{s^2}$ in $s\ge 1$ occurs at $s=1.260816271(1)$ with maximum $<0.324908241$ and $\frac{F(s)}{s^2}\to 0$ as $s\to\infty$. Thus, choosing $2\tau=(0.325+\tilde{\varepsilon}(k))k$ for an appropriate function $\tilde{\varepsilon}(k)$ tending to 0 as $k\to\infty$ the $L^2$ distance is negligible so that $\tau t^{\mathrm{rel}}$ is an upper bound for $t_2^{\mathrm{mix}}\ge t_1^{\mathrm{mix}}$. □

### 3.1 Geometric mixing time bound, proof of Theorem 1.4

Let $p$, $A$ and $\Lambda$ as above, and let $\mathscr{F}$ be the Voronoi cell for $\mathbb{R}^k/\Lambda$. Note that $\mathbb{Z}^k \cap \overline{\mathscr{F}}$ contains a system of representatives for $\mathbb{Z}^k/\Lambda$, and that the Cayley graph $\mathscr{C}(A,p)$ is isomorphic to $\mathscr{C}(\{0,\pm e_i : 1 \le i \le k\}, \mathbb{Z}^k/\Lambda)$. Thus

$$\mathrm{rad}(\mathscr{F}) := \sup\{\|\underline{x}\|_2 : \underline{x} \in \mathscr{F}\} = \mathrm{diam}_{\mathrm{geom}}(\mathscr{C}(A,p)).$$

*Proof of Theorem 1.4.* Write $D = \mathrm{diam}_{\mathrm{geom}}(\mathscr{C}(A,p))$ and assume, as we may, that $t > kD^2$. In view of Lemma 2.4, which proves $D \ge \frac{1}{\ell(A)}$, we have $t \gg t^{\mathrm{rel}}$, and thus as in Lemma 2.10

$$\left\|\mu_A^{*t} - \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}}\right\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})} + o(1) = \left\|\Theta\left(\cdot, \frac{2t}{2k+1}; \Lambda\right) - \mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)},$$

so we will estimate the right hand side.

Since, for any $\underline{x}$, $t$, $\mathbf{E}_{\underline{y}\in\mathscr{F}}\left[\Theta\left(\underline{x}+\underline{y}, \frac{2t}{2k+1}; \Lambda\right)\right] = \frac{1}{p}$, we may estimate using the triangle inequality

$$\begin{aligned}
&\left\|\Theta - \mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)} = \frac{1}{2}\int_{\underline{x}\in\mathscr{F}} \left|\Theta\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right) - \mathbf{E}_{\underline{y}\in\mathscr{F}}\left[\Theta\left(\underline{x}+\underline{y}, \frac{2t}{2k+1}; \Lambda\right)\right]\right| d\underline{x} \\
&\le \frac{1}{2}\int_{\underline{x}\in\mathscr{F}} \sum_{\lambda\in\Lambda} \left|\eta_k\left(\sqrt{\frac{2t}{2k+1}}, \underline{x}-\lambda\right) - \mathbf{E}_{\underline{y}\in\mathscr{F}}\left[\eta_k\left(\sqrt{\frac{2t}{2k+1}}, \underline{x}+\underline{y}-\lambda\right)\right]\right| d\underline{x}.
\end{aligned}$$

Now use the inequality $|1-e^x| \le e^{|x|}-1$ to obtain

$$\begin{aligned}
&\left\|\Theta - \mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)} \\
&\le \frac{1}{2}\int_{\underline{x}\in\mathscr{F}} \sum_{\lambda\in\Lambda} \eta_k\left(\sqrt{\frac{2t}{2k+1}}, \underline{x}-\lambda\right) \mathbf{E}_{\underline{y}\in\mathscr{F}}\left[\exp\left(\frac{2k+1}{4t}\left(\|\underline{y}\|_2^2 + 2|\langle \underline{x}-\lambda, \underline{y}\rangle|\right)\right) - 1\right] d\underline{x}.
\end{aligned}$$

Fold together the sum over $\lambda$ and the integral over $\underline{x}$, then integrate away all directions in $\underline{x}$ orthogonal to $\underline{y}$ to obtain

$$\begin{aligned}
&\left\|\Theta\left(\cdot, \frac{2t}{2k+1}; \Lambda\right) - \mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)} \\
&\le \frac{1}{2}\int_{x\in\mathbb{R}} \eta_1\left(\sqrt{\frac{2t}{2k+1}}, x\right) \mathbf{E}_{\underline{y}\in\mathscr{F}}\left[\exp\left(\frac{2k+1}{4t}\left(\|\underline{y}\|_2^2 + \|\underline{y}\|_2|x|\right)\right) - 1\right] dx \\
&\ll D\sqrt{\frac{k}{t}}.
\end{aligned}$$

The last estimate follows on using $\frac{1}{\sqrt{2\pi}}\int_{x\in\mathbb{R}} e^{-\frac{x^2}{2}+\delta|x|}dx = 1 + O(\delta)$ as $\delta \downarrow 0$. □

## 4 Transition window bound, proof of Theorem 1.6

We prove the following somewhat more general theorem.

**Theorem 4.1.** *Let $p$ be a large prime and let $k \le \frac{\log p}{\log\log p}$. Let $A \subset \mathbb{Z}/p\mathbb{Z}$ be a lazy symmetric generating set of size $|A| = 2k+1$. For any $1 > \epsilon_1 > \epsilon_2 > 0$, for all $n < \exp\left(\frac{2\epsilon_2}{k}\right) \cdot t_1^{\mathrm{mix}}(\epsilon_1)$ we have*

$$\left\|\mu_A^{*n} - \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}}\right\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})} \ge \epsilon_1 - \epsilon_2 + o_{\epsilon_1,\epsilon_2}(1).$$

*Proof.* Let $\Lambda < \mathbb{Z}^k$ be any lattice representing the class of $\Lambda(A) \in \mathscr{L}$. By Lemma 2.10 we may replace $\left\|\mu_A^{*n} - \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}}\right\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})}$ with $\left\|\Theta\left(\underline{x}, \frac{2n}{2k+1}; \Lambda\right) - \mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)}$ making error $o(1)$.

Write $n = \sigma t_1^{\mathrm{mix}}(\epsilon_1)$. Differentiating under the sum in the $\theta$ function,

$$\frac{d}{d\sigma}\Theta\left(\underline{x}, \frac{2\sigma t_1^{\mathrm{mix}}(\epsilon_1)}{2k+1}; \Lambda\right)\Big|_{\sigma=\sigma'} \geq -\frac{k}{2\sigma'}\Theta\left(\underline{x}, \frac{2\sigma' t_1^{\mathrm{mix}}(\epsilon_1)}{2k+1}; \Lambda\right). \tag{4.1}$$

Also, $\left\|\Theta\left(\underline{x}, \frac{2\sigma t_1^{\mathrm{mix}}(\epsilon_1)}{2k+1}; \Lambda\right) - \mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)}$ is a decreasing function of $\sigma > 0$. Define

$$P(\sigma) = \left\{\underline{x} \in \mathbb{R}^k/\Lambda : \Theta\left(\underline{x}, \frac{2\sigma t_1^{\mathrm{mix}}(\epsilon_1)}{2k+1}; \Lambda\right) > \frac{1}{p}\right\}.$$

Now for any $\sigma, \sigma_0 > 0$,

$$\begin{aligned}\left\|\Theta\left(\cdot, \frac{2\sigma t_1^{\mathrm{mix}}(\epsilon_1)}{2k+1}; \Lambda\right) - \mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)} &= \int_{P(\sigma)} \Theta\left(\underline{x}, \frac{2\sigma t_1^{\mathrm{mix}}(\epsilon_1)}{2k+1}; \Lambda\right) - \frac{1}{p} d\underline{x} \\ &\geq \int_{P(\sigma_0)} \Theta\left(\underline{x}, \frac{2\sigma t_1^{\mathrm{mix}}(\epsilon_1)}{2k+1}; \Lambda\right) - \frac{1}{p} d\underline{x}.\end{aligned}$$

Thus for $\sigma > \sigma_0$,

$$\begin{aligned}&\left\|\Theta\left(\cdot, \frac{2\sigma t_1^{\mathrm{mix}}(\epsilon_1)}{2k+1}; \Lambda\right) - \mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)} - \left\|\Theta\left(\cdot, \frac{2\sigma_0 t_1^{\mathrm{mix}}(\epsilon_1)}{2k+1}; \Lambda\right) - \mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)} \\ &\quad \geq \int_{P(\sigma_0)} \Theta\left(\underline{x}, \frac{2\sigma t_1^{\mathrm{mix}}(\epsilon_1)}{2k+1}; \Lambda\right) - \Theta\left(\underline{x}, \frac{2\sigma_0 t_1^{\mathrm{mix}}(\epsilon_1)}{2k+1}; \Lambda\right) d\underline{x}\end{aligned} \tag{4.2}$$

Differentiate under the integral, then apply (4.1) and finally drop the restriction to $P(\sigma_0)$ to obtain the estimate

$$\begin{aligned}(4.2) &= \int_{P(\sigma_0)} \int_{\sigma_0}^{\sigma} \frac{d}{ds}\Theta\left(\underline{x}, \frac{2s t_1^{\mathrm{mix}}(\epsilon_1)}{2k+1}; \Lambda\right)\Big|_{s=\sigma'} d\sigma' d\underline{x} \\ &\geq -\frac{k}{2}\int_{\sigma_0}^{\sigma} \frac{1}{\sigma'} \int_{P(\sigma_0)} \Theta\left(\underline{x}, \frac{2\sigma' t_1^{\mathrm{mix}}(\epsilon_1)}{2k+1}; \Lambda\right) d\underline{x} d\sigma' \\ &\geq \frac{-k}{2}\log\frac{\sigma}{\sigma_0}.\end{aligned} \tag{4.3}$$

Note that $k = o\left(t_1^{\mathrm{mix}}(\epsilon_1)\right)$. Applying (4.3) with $\sigma_0 = 1 - \frac{1}{t_1^{\mathrm{mix}}(\epsilon_1)}$ and $\sigma = 1$, which corresponds to the random walk at the mixing time and the step before, we deduce

$$\left\|\Theta\left(\cdot, \frac{2t_1^{\mathrm{mix}}(\epsilon_1)}{2k+1}; \Lambda\right) - \mathbb{U}_{\mathbb{R}^k/\Lambda}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)} = \epsilon_1 + o_{\epsilon_1}(1).$$

Applying (4.3) again, but now with $\sigma_0 = 1$, $\sigma = \exp(\frac{2\epsilon_2}{k})$, we obtain in the range $t_1^{\mathrm{mix}}(\epsilon_1) < n < \exp(\frac{2\epsilon_2}{k}) \cdot t_1^{\mathrm{mix}}(\epsilon_1)$,

$$\left\|\mu_A^{(n)} - \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}}\right\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})} \geq \epsilon_1 - \epsilon_2 + o_{\epsilon_1,\epsilon_2}(1). \qquad \square$$

## 5 Random random walk, proof of Theorem 1.7

We record several facts regarding the uniform measure $\mathbb{U}_L$ on the set $L(p,k)$ of index $p$ lattices in $\mathbb{Z}^k$.

**Lemma 5.1.** *When $\Lambda$ is chosen uniformly from $L(p,k)$, the dual lattice $\Lambda^\vee$ has the distribution of*

$$\{0,1,\cdots,p-1\}\frac{\underline{v}}{p}+\mathbb{Z}^k$$

*where $\underline{v}$ is a uniform random vector in $(\mathbb{Z}/p\mathbb{Z})^k\setminus\{0\}$.*

*When $\Lambda$ is chosen uniformly from $L^0(p,k)$, the dual lattice $\Lambda^\vee$ has the distribution of*

$$\{0,1,\cdots,p-1\}\frac{\underline{v}}{p}+\mathbb{Z}^k$$

*where $\underline{v}$ is chosen uniformly from*

$$\mathscr{D}=\{\underline{v}\in(\mathbb{Z}/p\mathbb{Z}\setminus\{0\})^k:\forall 1\le i<j\le k, v_i\neq\pm v_j\}.$$

*Proof.* In the case of $L(p,k)$, the structure follows from $[\Lambda:\mathbb{Z}^k]=p$ and $\frac{1}{p}\mathbb{Z}^k<\Lambda$, while the uniformity follows from the fact that $\mathrm{SL}_k(\mathbb{Z}/p\mathbb{Z})$ acts transitively on the space of dual lattices. This holds since any non-zero vector may be completed to a basis for $(\mathbb{Z}/p\mathbb{Z})^k$.

The further conditions imposed in the case of $L^0(p,k)$ are those necessary to ensure that $\Lambda$ does not contain a vector $\lambda$ with $\|\lambda\|_2^2\in\{1,2\}$. □

**Lemma 5.2.** *Let $p$ be prime, let $k\ge 2$ and let $\underline{v}\neq\underline{w}\in\mathbb{Z}^k$. We have*

$$\mathbb{U}_L(\Lambda:\underline{v},\underline{w}\in\Lambda)=\begin{cases}1 & \underline{v},\underline{w}\in(p\mathbb{Z})^k\\ \frac{p^{k-1}-1}{p^k-1} & |\mathbb{Z}\underline{v}+\mathbb{Z}\underline{w}\bmod p|=p\\ \frac{p^{k-2}-1}{p^k-1} & |\mathbb{Z}\underline{v}+\mathbb{Z}\underline{w}\bmod p|=p^2\end{cases}.$$

*In particular, $\mathbb{U}_L(L^0(p,k))\ge 1-O\left(\frac{k^2}{p}\right)$.*

*Proof.* These follow immediately from the distribution of the dual group. □

### 5.1 Summary of argument

As the calculations in the remainder of this section are somewhat involved, we pause to sketch the main ideas.

Theorem 1.7 has three claims, the first two of which consider the worst case mixing time behavior, with the third considering typical behavior. When considering the walk as a diffusion on $\mathbb{R}^k/\Lambda$ where $\Lambda$ is a lattice, the spectrum of the transition kernel is determined by the dual lattice $\Lambda^\vee$. In general, it is difficult to work on the spectral side due to the high concentration of eigenvalues near the spectral gap, but in the worst case regime we are able to show that for all behavior that persists, the dual lattice is essentially one dimensional. When this occurs the mixing and relaxation times are proportional and we obtain a slow transition.

In typical behavior the walk has a sharp transition to uniformity. The analysis in this regime consists of separate arguments estimating the distance to uniformity at times $(1\pm\epsilon)t_1^{\mathrm{mix}}$. When considering the walk at time $(1-\epsilon)t_1^{\mathrm{mix}}$ we study the diffusion $\Theta\left(\underline{x},\frac{2t}{2k+1};\Lambda\right)$ on the norm-minimal fundamental domain $\mathscr{F}(\Lambda)$ for $\mathbb{R}^k/\Lambda$. For a particular lattice $\Lambda$, $\mathscr{F}(\Lambda)$ is a highly complex convex body determined by a number of hyperplanes, but in a statistical sense, for the purpose of the lower bound, $\mathscr{F}(\Lambda)$ behaves very much like the volume $p$ ball of $\mathbb{R}^k$ centered at the origin. A Gaussian in $\mathbb{R}^k$ centered at the origin is concentrated on a thin spherical shell (see Lemma 2.2), and the mixing time is essentially the time needed for this spherical shell to expand to the boundary of the volume $p$ ball. At time $(1-\epsilon)t_1^{\mathrm{mix}}$ we are then able to show that the diffusion is typically concentrated on a small measure part of $\mathscr{F}(\Lambda)$.

For the upper bound at time $(1+\epsilon)t_1^{\mathrm{mix}}$, we note that $p\mathbb{Z}^k < \Lambda$, and we show that the distribution of values of $\Theta\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right)$ is concentrated near 1 when $\underline{x}$ is chosen uniformly from $\mathbb{R}^k/p\mathbb{Z}^k$ and $\Lambda$ is chosen at random from $L(p,k)$. This is the most delicate part of the argument. For instance, it is not sufficient to consider the expectation of $\left(\Theta\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right) - 1\right)^2$ as this gives an upper bound which is too weak, so we split $\Theta$ into an $L^2$-concentrated piece $\Theta_M$ plus a small $L^1$ error $\Theta_E$.

### 5.2 Slow mixing behavior

We prove Theorem 1.7 in two parts. In this section we prove parts (1) and (2) which concern rare slow mixing walks. In Section 5.3 we prove part (3) regarding the typical behavior. The main estimate regarding slow mixing behavior is the following theorem.

**Theorem 5.3.** *Let $p$ be a large prime, and let $k = k(p)$ tending to $\infty$ with $p$ in such a way that $k \leq \frac{\log p}{\log\log p}$. For any $\delta > 0$, for all $p$ sufficiently large, uniformly in $\delta \frac{p^{\frac{1}{k}}}{\sqrt{k}} < \rho < \frac{(p\log p)^{\frac{1}{k}}}{\delta}$, the following hold*

1. $$\mathbf{P}_{\mathscr{A}(p,k)}\left[t^{\mathrm{rel}} \geq \frac{e\rho^2 p^{\frac{2}{k}}}{\pi}\right] = \frac{\exp(o(k))}{\rho^k}. \tag{5.1}$$

2. *Let, as in Theorem 1.7, $\tau_0$ be the ratio between total variation mixing time and relaxation time for Gaussian diffusion on $\mathbb{R}/\mathbb{Z}$. For any $C \geq 1$, and $\frac{\delta p^{\frac{4}{k}}}{k} \leq J \leq \frac{p^{\frac{4}{k}}(\log p)^{\frac{2}{k}}}{\delta}$*

$$\mathbf{P}_{\mathscr{A}(p,k)}\left[t_1^{\mathrm{mix}} \geq C(\tau_0+\delta)t^{\mathrm{rel}} \text{ and } \frac{J}{2} \leq t^{\mathrm{rel}} \leq J\right] \leq \exp\left(\frac{k}{2}\log\frac{k}{C} + O_\delta(k)\right)\frac{p^2}{J^k} \tag{5.2}$$
$$\mathbf{P}_{\mathscr{A}(p,k)}\left[t_1^{\mathrm{mix}} \leq (\tau_0-\delta)t^{\mathrm{rel}} \text{ and } \frac{J}{2} \leq t^{\mathrm{rel}} \leq J\right] \leq \exp\left(\frac{k}{2}\log k + O_\delta(k)\right)\frac{p^2}{J^k}.$$

*Deduction of Theorem 1.7, parts (1) and (2).* Before giving the proof of the Theorem we prove an auxiliary claim.

Let $\delta > 0$ be an arbitrarily small fixed quantity. We claim that with probability 1, only finitely many of the events

$$B_p = \left\{t_1^{\mathrm{mix}}(p) \geq \delta p^{\frac{4}{k}} \text{ and } \left|\frac{t_1^{\mathrm{mix}}(p)}{\tau_0 t^{\mathrm{rel}}(p)} - 1\right| \geq \delta\right\}$$

occur. Note that by Theorem 1.1, $t_1^{\mathrm{mix}}(p) \geq \delta p^{\frac{4}{k}}$ implies $t^{\mathrm{rel}}(p) \gg \delta\frac{p^{\frac{4}{k}}}{k}$. Thus, combining (5.1) and (5.2),

$$\mathbf{P}(B_p) \leq \frac{\exp\left(k\log k + O_\delta(k)\right)}{p^2} + \frac{1}{p\log p},$$

where the first term is handled with (5.2) and covers the range $t^{\mathrm{rel}} \ll p^{\frac{4}{k}}(\log p)^{\frac{2}{k}}$, the worst case occuring when $t^{\mathrm{rel}} \ll \frac{p^{\frac{4}{k}}}{k}$ is minimized. Note $k \leq \frac{\log p}{\log\log p}$ from which it follows

$$\sum_p \mathbf{P}(B_p) \leq \sum_p \left(\frac{\exp\left(-\frac{\log p}{\log\log p}\left(\log\log\log p + O_\delta(1)\right)\right)}{p} + \frac{1}{p\log p}\right) < \infty,$$

so that the claim holds by the Borel-Cantelli Lemma.

We now prove the Theorem.

(1) Replace $\rho(p)$ with $\rho(p) := \max\left(\rho(p), p^{\frac{1}{k}}\right)$ without altering the divergence of $\sum_p \rho(p)^{-k}$. Estimating with (5.1), by Borel-Cantelli, with probability 1 there is an infinite sequence $\mathscr{P}_0 \subset \mathscr{P}$ such that, for $p \to \infty$ through $\mathscr{P}_0$,

$$t^{\mathrm{rel}}(p) \gtrsim \frac{e}{\pi}\rho(p)^2 p^{\frac{2}{k}}.$$

The above remarks guarantee that, for this sequence, $t_1^{\mathrm{mix}}(p) \sim \tau_0 t^{\mathrm{rel}}(p)$.

(2) Let $\delta > 0$ be fixed. Estimate with (5.1) to obtain that with probability 1, for all but finitely many $p$,

$$t^{\mathrm{rel}}(p) \leq \left(1 + \frac{\delta}{2}\right)\frac{e}{\pi}\rho(p)^2 p^{\frac{2}{k}}.$$

Since $\rho(p) \geq p^{\frac{1}{k}}$ eventually, the remarks above imply that with probability 1

$$t_1^{\mathrm{mix}}(p) \leq (1+\delta)\frac{e\tau_0}{\pi}\rho(p)^2 p^{\frac{2}{k}}$$

for all but finitely many $p$. □

In proving Theorem 5.3 we introduce two commonly used pieces of terminology from the theory of lattices. Let $p$ be a prime and let $k \geq 1$. Say that $\lambda \in \mathbb{Z}^k$ is *reduced* (at $p$) if $\lambda \in \left[-\frac{p}{2}, \frac{p}{2}\right)^k$. Any class $\lambda \in (\mathbb{Z}/p\mathbb{Z})^k$ has a unique reduced representative $r(\lambda) \in \mathbb{Z}^k$. Say that $\lambda = (\lambda_1, ..., \lambda_k) \in \mathbb{Z}^k$ is *primitive* if $\lambda \neq 0$ and $\mathrm{GCD}(\lambda_i : 1 \leq i \leq k) = 1$.

Our proof of Theorem 5.3 depends upon the following two estimates, the first of which estimates a mean concerning pairs of short vectors in the dual space.

**Proposition 5.4.** *Let $\delta > 0$ be a fixed constant. Let $p$ and $k(p)$ tend to $\infty$ in such a way that $k \leq \frac{\log p}{\log\log p}$. Let $\frac{\delta p^{\frac{1}{k}}}{\sqrt{k}} \leq \rho \leq \frac{1}{\delta}(p\log p)^{\frac{1}{k}}$. For any $\frac{\delta}{\sqrt{k}} \leq C \leq \frac{\sqrt{k}}{\delta}$, for any $\epsilon > 0$,*

$$\mathbf{E}_{L^0(p,k)}\left[\sum_{\substack{\lambda_1 \neq \pm\lambda_2 \in \Lambda^\vee \setminus \{0\} \\ p\lambda_i \text{ primitive}}} \eta_k\left(\frac{1}{\rho p^{\frac{1}{k}}}, \lambda_1\right)\eta_k\left(\frac{1}{C\rho p^{\frac{1}{k}}}, \lambda_2\right)\right] \leq p^2 + O_\epsilon\left(p^{\frac{3}{2}+\frac{4}{k}+\epsilon}\right). \tag{5.3}$$

**Remark 5.5.** This proposition should be interpretted as expressing the approximate independence of the appearance of a pair of short primitive vectors in the dual space.

*Proof.* It is enough to estimate with respect to $\mathbb{U}_{L(p,k)}$ since this introduces a relative error $1 + O\left(\frac{k^2}{p}\right)$, which is smaller than the error claimed.

Let $\mathscr{S} \subset (\mathbb{Z}/p\mathbb{Z})^k \times \mathbb{Z}/p\mathbb{Z}$ denote the set of pairs $(\lambda, a)$ such that $\lambda \in (\mathbb{Z}/p\mathbb{Z})^k$, $a \in \mathbb{Z}/p\mathbb{Z} \setminus \{0, \pm 1\}$ and both reduced vectors $r(\lambda)$ and $r(a\lambda)$ are primitive. Also denote for $\lambda \in (\mathbb{Z}/p\mathbb{Z})^k$, $\mathscr{S}(\lambda) \subset \mathbb{Z}/p\mathbb{Z}$ the fiber over $\lambda$.

Lemma 5.1 gives

$$\begin{aligned}&\mathbf{E}_{L(p,k)}\left[\sum_{\substack{\lambda_1 \neq \pm\lambda_2 \in \Lambda^\vee \setminus \{0\} \\ p\lambda_i \text{ primitive}}} \eta_k\left(\frac{1}{\rho p^{\frac{1}{k}}}, \lambda_1\right)\eta_k\left(\frac{1}{C\rho p^{\frac{1}{k}}}, \lambda_2\right)\right] \\ &\leq \frac{p^{2k}(p-1)}{p^k - 1}\sum_{\substack{\lambda \in \left(\mathbb{Z} \cap \left(-\frac{p}{2}, \frac{p}{2}\right]\right)^k \\ \text{primitive}}}\sum_{a \in \mathscr{S}(\lambda)} \Phi_1(\lambda)\Phi_C(a\lambda) + o(1),\end{aligned} \tag{5.4}$$

where

$$\Phi_c(\underline{x}) = \sum_{\underline{n} \in \mathbb{Z}^k} \eta_k\left(\frac{p^{1-\frac{1}{k}}}{c\rho}, \underline{x} + p\underline{n}\right).$$

To briefly explain this formula, the factor of $p^{2k}$ results from scaling both the variable and the standard deviation in the Gaussians by $p$. The condition $\lambda_1 \equiv a\lambda_2 \bmod p\mathbb{Z}^k$ for some $a$ follows from the characterization of $\Lambda^\vee$. The error term $o(1)$ covers summation over pairs $\lambda_1, \lambda_2$ for which one of $\lambda_1, \lambda_2$ is not reduced but both are primitive. The summation in this case is bounded by, for some $c > 0$ and all $B > 0$

$$\begin{aligned} &\ll p^{-k+1} \sum_{\substack{\lambda_1,\lambda_2 \in \left(\frac{1}{p}\mathbb{Z}\right)^k \\ \max(\|\lambda_1\|_\infty, \|\lambda_2\|_\infty > \frac{1}{2})}} \eta_k\left(\frac{1}{\rho p^{\frac{1}{k}}}, \lambda_1\right) \eta_k\left(\frac{1}{C\rho p^{\frac{1}{k}}}, \lambda_2\right) \\ &\ll p^{O(k)} \exp\left(-cp^{\frac{2}{k}}\right) = O_B(p^{-B}), \end{aligned} \tag{5.5}$$

since $p^{\frac{2}{k}}$ dominates $k \log p$.

We make several modifications to the sum of (5.4) which make it easier to estimate. First we may exclude from $\mathscr{S}$ any pairs $(\lambda, a)$ for which

$$\max\left(\rho p^{\frac{1}{k}} \left\|\frac{\lambda}{p}\right\|_{(\mathbb{R}/\mathbb{Z})^k}, C\rho p^{\frac{1}{k}} \left\|\frac{a\lambda}{p}\right\|_{(\mathbb{R}/\mathbb{Z})^k}\right) \geq Q := \sqrt{\log p \log\log p}$$

as these contribute, for any $B > 0$, $O_B(p^{-B})$. To obtain this, note that the cardinality of the summation set is $O(p^{k+1})$ since we have replaced summation over $\lambda_2$ with summation over $a$. Thus it suffices to show that for excluded pairs, $\Phi_1(\lambda)\Phi_C(a\lambda) \ll_B p^{-2k-2-B}$; to see this, note that $\Phi$ is controlled by the contribution of the summand nearest 0.

Let $\mathscr{S}'$ be those choices of $(\lambda, a)$ which remain. Denote by $\mathscr{F}(Q)$ the collection of Farey fractions modulo $p$ (the definition is non-standard since the numerator and denominator are bounded by different quantities),

$$\mathscr{F}(Q) = \left\{bq^{-1} \bmod p : \max\left(|b|, \frac{|q|}{C}\right) \leq \frac{\rho p^{\frac{1}{k}}}{2Q}, q \neq 0\right\}.$$

We claim that for any reduced $\lambda$, $\mathscr{S}'(\lambda \bmod p) \subset \mathbb{Z}/p\mathbb{Z} \setminus \mathscr{F}(Q)$. Indeed, suppose otherwise and let $a = bq^{-1} \in \mathscr{S}'(\lambda \bmod p) \cap \mathscr{F}(Q)$. Let $\eta \equiv a\lambda \bmod p$ with $\eta$ reduced. Then $b\lambda \equiv q\eta \bmod p$, but the norm condition implies that in fact $b\lambda = q\eta$, which contradicts the primitivity.

Replace $\mathscr{S}'(\lambda)$ with $\mathbb{Z}/p\mathbb{Z} \setminus \mathscr{F}(Q)$ and complete the sum over $\lambda$ to obtain

$$(5.4) \leq O_B(p^{-B}) + \frac{p^{2k}(p-1)}{p^k - 1} \sum_{\lambda \in (\mathbb{Z}/p\mathbb{Z})^k} \sum_{a \in \mathscr{F}(Q)^c} \Phi_1(\lambda)\Phi_C(a\lambda).$$

Applying Plancherel on $(\mathbb{Z}/p\mathbb{Z})^k$, we obtain

$$(5.4) < O_B(p^{-B}) + \frac{p^k(p-1)}{p^k - 1} \sum_{a \in \mathscr{F}(Q)^c} \sum_{\xi \in (\mathbb{Z}/p\mathbb{Z})^k} \hat{\Phi}_1(a\xi)\overline{\hat{\Phi}_C(\xi)}$$

where

$$\hat{\Phi}_c(\xi) = \sum_{\underline{n} \in \mathbb{Z}^k} \eta_k\left(\frac{p^{1-\frac{1}{k}}}{c\rho}, \underline{n}\right) \exp\left(\frac{2\pi i \xi \cdot \underline{n}}{p}\right) = \sum_{\underline{n} \in \mathbb{Z}^k} \exp\left(-\frac{2\pi^2 p^{2-\frac{2}{k}}}{c^2\rho^2} \left\|\frac{\xi}{p} + \underline{n}\right\|_2^2\right).$$

All but one term from the sum over $\underline{n}$ is negligible, and we obtain, for any $\epsilon > 0$,

$$\begin{aligned} (5.4) =& O_B(p^{-B}) \\ &+ \frac{p^k(p-1)}{p^k - 1} \sum_{a \in \mathscr{F}(Q)^c} \sum_{\xi \in (\mathbb{Z}/p\mathbb{Z})^k} \exp\left(-\frac{2\pi^2 p^{2-\frac{2}{k}}}{\rho^2} \left(\frac{1}{C^2} \left\|\frac{\xi}{p}\right\|_{(\mathbb{R}/\mathbb{Z})^k}^2 + \left\|\frac{a\xi}{p}\right\|_{(\mathbb{R}/\mathbb{Z})^k}^2\right)\right). \end{aligned}$$

Due to the decay in the exponential, we may truncate summation over $a$ and $\xi$ to $\left\|\frac{\xi}{p}\right\|_{(\mathbb{R}/\mathbb{Z})^k}, \left\|\frac{a\xi}{p}\right\|_{(\mathbb{R}/\mathbb{Z})^k} \ll_\epsilon p^{-1+\frac{1}{k}+\epsilon}$ with negligible error.

From $\xi = 0$ pull out a term $\sim p^2$. To treat the remaining terms, suppose $k \geq 3$, and let $\xi = q\xi^0$ for $q \in \mathbb{Z}_{>0}$ and $\xi_0$ primitive. Write $a\xi \equiv \zeta \bmod p$ where $\|\zeta\|_{\mathbb{R}^k} \ll_\epsilon p^{\frac{1}{k}+\epsilon}$. It follows for $1 < i \leq k$, $\xi_1^0\zeta_i \equiv \xi_i^0\zeta_1 \bmod p$, and in fact, $\xi_1^0\zeta_i = \xi_i^0\zeta_1$ so $\zeta = b\xi_0$ for some $b \in \mathbb{Z}$. The sum is thus bounded by

$$\frac{p^k(p-1)}{p^k-1} \sum_{\substack{\xi\in\mathbb{Z}^k \\ 1\leq\|\xi\|\ll p^{\frac{1}{k}+\epsilon}}} \sum_{\max(|b|,\frac{|q|}{C})>\frac{\rho p^{\frac{1}{k}}}{2Q}} \exp\left(-\frac{2\pi^2}{\rho^2 p^{\frac{2}{k}}}\left(\left(\frac{q^2}{C^2}+b^2\right)\|\xi\|_2^2\right)\right).$$

We may estimate this sum crudely by truncating summation over $b, q$ at $|b|, \frac{|q|}{C} \leq \rho p^{\frac{1}{k}+\epsilon}$ with error $O_B(p^{-B})$. The total number of such $b, q$ is $\ll k^{O(1)}\rho^2 p^{\frac{2}{k}+2\epsilon} \ll_{\epsilon'} p^{\frac{4}{k}+\epsilon'}$. For all such $b, q$, summation over $\xi$ is bounded by (see Lemma 2.7)

$$p\sum_{0\neq\xi\in\mathbb{Z}^k} \exp\left(\frac{-\pi^2\|\xi\|_2^2}{2Q^2}\right) < p\left(\sum_{\xi\in\mathbb{Z}} \exp\left(\frac{-\pi^2\xi^2}{2Q^2}\right)\right)^k \ll p(2Q)^k \ll_\epsilon p^{\frac{3}{2}+\epsilon}. \qquad \square$$

Next we determine the distribution of the shortest vector in the dual lattice. Recall that $R_k = \left(\frac{\Gamma\left(\frac{k}{2}+1\right)}{\pi^{\frac{k}{2}}}\right)^{\frac{1}{k}}$ is the radius of a volume 1 ball.

**Proposition 5.6.** *Let $\delta > 0$ be a fixed constant, and let $p$, $k$ and $\rho$ be such that $k \leq \frac{\log p}{\log\log p}$, and $\frac{\delta p^{\frac{1}{k}}}{\sqrt{k}} \leq \rho \leq \frac{1}{\delta}(p\log p)^{\frac{1}{k}}$. Given $\Lambda \in L^0(p,k)$ denote $\lambda^*$ the shortest non-zero vector of the dual lattice. One has*

$$\mathbf{P}_{L^0(p,k)}\left[\|\lambda^*\|_2 \leq \frac{R_k}{\rho p^{\frac{1}{k}}}\right] = \frac{1}{2\rho^k}\left(1+O\left(\frac{e^{O(k)}}{\rho^k}+\frac{k^2\rho}{p^{1-\frac{1}{k}}}\right)\right).$$

*Proof.* By Lemma 2.5

$$\begin{aligned}\mathbf{E}_{L(p,k)}\left[\#\left\{0\neq\lambda\in\Lambda^\vee\cap B_2\left(0,\frac{R_k}{\rho p^{\frac{1}{k}}}\right)\right\}\right] &= \frac{p-1}{p^k-1}\#\left\{0\neq\lambda\in\mathbb{Z}^k\cap B_2\left(0,\frac{R_k p^{1-\frac{1}{k}}}{\rho}\right)\right\} \\ &= \frac{1}{\rho^k}\left(1+O\left(\frac{k\rho}{p^{1-\frac{1}{k}}}\right)\right).\end{aligned}$$

By counting vectors $\lambda$ with $\lambda_1 = 0$ or $\lambda_1 = \pm\lambda_2$ one finds

$$\mathbf{E}_{L^0(p,k)}\left[\#\left\{0\neq\lambda\in\Lambda^\vee\cap B_2\left(0,\frac{R_k}{\rho p^{\frac{1}{k}}}\right)\right\}\right] = \frac{1}{\rho^k}\left(1+O\left(\frac{k^2\rho}{p^{1-\frac{1}{k}}}\right)\right). \tag{5.6}$$

Let $0 < \tau < 1$ and observe that for all $(1-\tau)\frac{\sqrt{k}}{p^{\frac{1}{k}}\rho} < \|\underline{x}\|_2 \leq (1+\tau)\frac{\sqrt{k}}{p^{\frac{1}{k}}\rho}$,

$$p\rho^k\exp\left(-\frac{k}{2}\left((1+\tau)^2+\log 2\pi\right)\right) \leq \eta_k\left(\frac{1}{p^{\frac{1}{k}}\rho},\underline{x}\right) \leq p\rho^k\exp\left(-\frac{k}{2}\left((1-\tau)^2+\log 2\pi\right)\right). \tag{5.7}$$

Choosing $C = 1$ in Proposition 5.4 and inserting these bounds, one finds

$$\mathbf{P}_{L^0(p,k)}\left[\|\lambda^*\|_2 \leq \frac{R_k}{\rho p^{\frac{1}{k}}}\right] = \frac{1}{2\rho^k}\left(1+O\left(\frac{e^{O(k)}}{\rho^k}+\frac{k^2\rho}{p^{1-\frac{1}{k}}}\right)\right),$$

by subtracting the contribution to (5.6) from lattices with pairs of primitive short vectors, and accounting for the factor of 2 from counting $\pm\lambda^*$. $\square$

*Proof of Theorem 5.3.* The estimate (5.1) regarding the distribution of $t^{\mathrm{rel}}$ follows from Proposition 5.6 together with $R_k \sim \sqrt{\frac{k}{2\pi e}}$ and $t^{\mathrm{rel}} \sim \frac{k}{2\pi^2\|\lambda^*\|_2^2}$ as $k \to \infty$.

For (5.2), choose $\rho = 2^{\frac{n}{2}}$ such that $\rho^2 p^{\frac{2}{k}} \asymp J$. Equivalently, consider $\Lambda$ for which the shortest non-zero vector $\lambda^*$ of $\Lambda^\vee$ satisfies $\frac{\sqrt{k}}{\rho p^{\frac{1}{k}}} \asymp \|\lambda^*\|_2$. For such $\lambda^*$,

$$\eta_k\left(\frac{1}{\rho p^{\frac{1}{k}}}, \lambda^*\right) = p\rho^k e^{O(k)}. \tag{5.8}$$

This majorant is used in what follows.

Let

$$\Theta_0\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right) = \frac{1}{p}\sum_{j\in\mathbb{Z}} \exp\left(\frac{-4\pi^2 t}{2k+1}\|\lambda^*\|_2^2 j^2\right) e\left(j\lambda^* \cdot \underline{x}\right)$$

denote the projection of $\Theta\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right)$ in frequency space onto the line determined by $\lambda^*$. If $\left|\frac{t_1^{\mathrm{mix}}}{\tau_0 t^{\mathrm{rel}}} - 1\right| > \epsilon$ then there is some $t = (1+O(\epsilon))t_1^{\mathrm{mix}}$ such that

$$\left\|(\Theta - \Theta_0)\left(\cdot, \frac{2t}{2k+1}; \Lambda\right)\right\|_{L^1(\mathbb{R}^k/\Lambda)} \gg_\epsilon 1. \tag{5.9}$$

Apply Cauchy-Schwarz to obtain

$$1 \ll_\epsilon \sum_{\lambda\in\Lambda^\vee\setminus\mathbb{Z}\cdot\lambda^*} \exp\left(\frac{-8\pi^2 t}{2k+1}\|\lambda\|_2^2\right) \ll \sum_{\substack{\lambda\in\Lambda^\vee\setminus\{\pm\lambda^*\}\\ p\lambda \text{ primitive}}} \exp\left(\frac{-8\pi^2 t}{2k+1}\|\lambda\|_2^2\right). \tag{5.10}$$

The latter sum may be written as

$$\sum_{\substack{\lambda\in\Lambda^\vee\setminus\{\pm\lambda^*\}\\ p\lambda \text{ primitive}}} \left(\frac{8\pi t}{2k+1}\right)^{\frac{k}{2}} \eta_k\left(\frac{1}{4\pi}\sqrt{\frac{2k+1}{t}}, \lambda\right).$$

Since $t \gg Ct^{\mathrm{rel}} \asymp C\frac{k}{\|\lambda^*\|_2^2} \asymp C\rho^2 p^{\frac{2}{k}}$ (take $C \asymp 1$ in the case of the second estimate of (5.2)) there is $c \asymp C$ such that

$$\sum_{\substack{\lambda\in\Lambda^\vee\setminus\{\pm\lambda^*\}\\ p\lambda \text{ primitive}}} \eta_k\left(\sqrt{\frac{k}{c}}\frac{1}{\rho p^{\frac{1}{k}}}, \lambda\right) \gg p\rho^k \left(\frac{C}{k}\right)^{\frac{k}{2}} e^{O(k)}.$$

Applying Proposition 5.4,

$$\mathbf{E}_{L^0(p,k)}\left[\sum_{\substack{\lambda_1\neq\pm\lambda_2\in\Lambda^\vee\setminus\{0\}\\ p\lambda_i \text{ primitive}}} \eta_k\left(\frac{1}{\rho p^{\frac{1}{k}}}, \lambda_1\right)\eta_k\left(\sqrt{\frac{k}{c}}\frac{1}{\rho p^{\frac{1}{k}}}, \lambda_2\right)\right] \leq p^2 + O_\epsilon\left(p^{\frac{3}{2}+\frac{4}{k}+\epsilon}\right)$$

and thus, by specializing to $\lambda_1 = \lambda^*$ and applying Markov's inequality,

$$\begin{aligned}&\mathbf{P}_{L^0(p,k)}\left[\|\lambda^*\|_2 \asymp \frac{\sqrt{k}}{\rho p^{\frac{1}{k}}} \text{ and } 1 \ll_\epsilon \sum_{\lambda\in\Lambda^\vee\setminus\mathbb{Z}\cdot\lambda^*} \exp\left(\frac{-8\pi^2 t}{2k+1}\|\lambda\|_2^2\right)\right]\\ &\ll \rho^{-2k}\exp\left(\frac{k}{2}\log\frac{k}{C} + O(k)\right).\end{aligned}$$

This verifies (5.2). □

### 5.3 Analysis of typical mixing behavior

We turn to analysis of the mixing behavior for $A$ in the bulk of $\mathscr{A}(p,k)$ proving the following theorem.

**Theorem 5.7.** *Let $p$ be prime, let $0 < \epsilon = \epsilon(p) < \frac{1}{2}$ and let $1 \le k \le \frac{\log p}{\log\log p}$. Set $\overline{t_1^{\mathrm{mix}}} = \frac{k}{2\pi e} p^{\frac{2}{k}}$. There is a function $\theta = \theta(\epsilon, k) > 0$ tending to 0 as $\epsilon^2 k \to \infty$ and a set $\mathscr{A}^*(p,k) \subset \mathscr{A}(p,k)$ satisfying*

$$|\mathscr{A}^*(p,k)| \ge (1 - o(1))\,|\mathscr{A}(p,k)|\,,$$

*such that, for all $A \in \mathscr{A}^*_{p,k}$,*

$$\forall\, n < (1-\epsilon)\overline{t_1^{\mathrm{mix}}}, \qquad \left\|\mu_A^{*n} - \frac{1}{p}\right\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})} \ge 1 - \theta(\epsilon,k) + o(1),$$
$$\forall\, n > (1+\epsilon)\overline{t_1^{\mathrm{mix}}}, \qquad \left\|\mu_A^{*n} - \frac{1}{p}\right\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})} \le \theta(\epsilon,k) + o(1)$$

*where all quantities $o(1)$ tend to zero as $p \to \infty$ uniformly in $k$.*

We can now conclude our proof of Theorem 1.7.

*Deduction of Theorem 1.7, part (3).* For each $j = 1, 2, \ldots$, let $E(p,j)$ be the event that

$$\forall\, n < (1-\epsilon(p))\overline{t_1^{\mathrm{mix}}}, \qquad \left\|\mu_A^{*n} - \frac{1}{p}\right\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})} > 1 - 2^{-j},$$
$$\forall\, n > (1+\epsilon(p))\overline{t_1^{\mathrm{mix}}}, \qquad \left\|\mu_A^{*n} - \frac{1}{p}\right\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})} < 2^{-j}.$$

For a fixed $p$, the events $E(p,j)$ are nested in $j$. For each $j \in \mathbb{Z}_{>0}$, let $N_j$ be minimal such that for all $p > N_j$, $\mathbb{U}_{\mathscr{A}(p,k)}[E(p,j)] \ge 1 - 2^{-j}$. This is finite by Theorem 5.7. Define $E^*(p) = \bigcap_{j : N_j < p} E(p,j)$ and let $p \in \mathscr{P}_0$ if and only if $E^*(p)$ occurs. Since $\mathbb{U}_{\mathscr{A}(p,k)}[E^*(p)] \to 1$ as $p \to \infty$ and the events are independent, we have $\mathscr{P}_0$ has density 1 with probability 1, as desired. □

In the remainder of this section we shall frequently be concerned with counting lattice points within Euclidean balls $B_2(\underline{x}, R) \subset \mathbb{R}^k$. It is useful to bear in mind that the radius $R_k$ of a ball of unit volume in $\mathbb{R}^k$ satisfies

$$R_k = \left(\frac{\Gamma\left(\frac{k}{2}+1\right)}{\pi^{\frac{k}{2}}}\right)^{\frac{1}{k}} = \sqrt{\frac{k}{2\pi e}}\left(1 + \frac{\log k}{2k} + O\left(\frac{1}{k}\right)\right).$$

Let $\epsilon = \epsilon(p)$ as in the theorem and set $\delta = \frac{1}{2}(1 - \sqrt{1-\epsilon})$. Recall that, given lattice $\Lambda < \mathbb{R}^k$, $\mathscr{F}(\Lambda)$ is the norm-minimal fundamental domain of $\Lambda$,

$$\mathscr{F}(\Lambda) = \left\{\underline{x} \in \mathbb{R}^k : \forall\, \lambda \in \Lambda \setminus \{0\}, \|\underline{x}\| < \|\lambda - \underline{x}\|\right\}.$$

Let $k = k(p)$ and set $t = t(p,k) = (1-\epsilon)\overline{t_1^{\mathrm{mix}}} \sim (1-\epsilon)R_k^2 p^{\frac{2}{k}}$.

**Lemma 5.8.** *As $k, p \to \infty$ in such a way that $k \le \frac{\log p}{\log\log p}$ we have*

$$\mathbf{E}_{L(p,k)}\left[\int_{\underline{x} \in B_2\left(0,(1-\delta)R_k p^{\frac{1}{k}}\right) \cap \mathscr{F}(\Lambda)} \Theta\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right) d\underline{x}\right] = 1 - o(1). \tag{5.11}$$

*Proof.* Since $\Theta\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right) \geq \eta_k\left(\sqrt{\frac{2t}{2k+1}}, \underline{x}\right)$,

$$\mathbf{E}_{L(p,k)}\left[\int_{\underline{x}\in B_2\left(0,(1-\delta)R_k p^{\frac{1}{k}}\right)\cap\mathscr{F}(\Lambda)} \Theta\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right) d\underline{x}\right]$$

$$\geq \int_{\|\underline{x}\|\leq(1-\delta)R_k p^{\frac{1}{k}}} \eta_k\left(\sqrt{\frac{2t}{2k+1}}, \underline{x}\right) d\underline{x} \tag{5.12}$$

$$-\mathbf{E}\left[\int_{\|\underline{x}\|\leq(1-\delta)R_k p^{\frac{1}{k}}} \eta_k\left(\sqrt{\frac{2t}{2k+1}}, \underline{x}\right) \mathbf{1}\left(\exists\, \lambda\in\Lambda\setminus\{0\} : \|\lambda-\underline{x}\| < \|\underline{x}\|\right) d\underline{x}\right]. \tag{5.13}$$

Since $\delta \sim \frac{\epsilon}{2}$ as $\epsilon \downarrow 0$, (5.12) $= 1 - o(1)$ follows from concentration of the norm of a Gaussian vector on scale $\frac{1}{\sqrt{k}}$ times its median length, see Lemma 2.2.

We estimate

$$(5.13) \leq \int_{\|\underline{x}\|\leq(1-\delta)R_k p^{\frac{1}{k}}} \eta_k\left(\sqrt{\frac{2t}{2k+1}}, \underline{x}\right) \mathbf{E}\left[\sum_{\lambda\in\Lambda\setminus\{0\}} \mathbf{1}\left(\|\lambda-\underline{x}\| < \|\underline{x}\|\right)\right] d\underline{x}.$$

For $k$ sufficiently large, any $\lambda$ counted in the expectation satisfies $\|\lambda\| < p$, and thus, by Lemma 5.2,

$$\mathbf{E}\left[\sum_{\lambda\in\Lambda\setminus\{0\}} \mathbf{1}\left(\|\lambda-\underline{x}\| < \|\underline{x}\|\right)\right] = \frac{p^{k-1}-1}{p^k-1}\#\{\lambda\in\mathbb{Z}^k : \|\lambda-\underline{x}\| < \|\underline{x}\|\}.$$

For any $\underline{x}\in\mathbb{R}^k$, any lattice point $\tilde{x}\in\mathbb{Z}^k$ which is the vertex of the unit lattice cube containing $\underline{x}$ satisfies $\|\tilde{\underline{x}}\| = \left(1+O\left(\frac{\sqrt{k}}{\|\underline{x}\|}\right)\right)\|\underline{x}\|$. Since $k^{\frac{3}{2}} = o(R_k p^{\frac{1}{k}})$, it follows that for all $\|\underline{x}\| \leq (1-\delta)R_k p^{\frac{1}{k}}$ we have

$$\frac{p^{k-1}-1}{p^k-1}\#\{\lambda\in\mathbb{Z}^k : \|\lambda-\underline{x}\| < \|\underline{x}\|\} \leq \left(1-\delta+o\left(\frac{1}{k}\right)\right)^k = o(1),$$

and thus

$$(5.13) = o\left(\int_{\|\underline{x}\|\leq(1-\delta)R_k p^{\frac{1}{k}}} \eta_k\left(\sqrt{\frac{2t}{2k+1}}, \underline{x}\right) d\underline{x}\right) = o(1). \qquad \square$$

*Proof of Theorem 5.7, lower bound.* For $n \geq \frac{t(p,k)}{2}$, Lemma 2.9 gives

$$\mathbf{E}_{\mathscr{A}(p,k)}\left[\left\|\mu_A^{*n} - \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}}\right\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})}\right] = o(1) + \mathbf{E}_{L^0(p,k)}\left[\left\|\Theta\left(\cdot, \frac{2n}{2k+1}; \Lambda\right) - \frac{1}{p}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)}\right]$$

while, for all $n < t(p,k)$,

$$(1+o(1))\mathbf{E}_{L^0(p,k)}\left[\left\|\Theta\left(\cdot, \frac{2n}{2k+1}; \Lambda\right) - \frac{1}{p}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)}\right]$$

$$\geq \mathbf{E}_{L(p,k)}\left[\int_{\underline{x}\in B_2\left(0,(1-\delta)R_k p^{\frac{1}{k}}\right)\cap\mathscr{F}(\Lambda)} \Theta\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right) - \frac{1}{p} d\underline{x}\right].$$

By Lemma 5.8, the expectation of the integral against $\Theta$ is $1-o(1)$, while the expectation of the integral against $\frac{1}{p}$ is bounded by

$$\int_{\underline{x}\in B_2\left(0,(1-\delta)R_k p^{\frac{1}{k}}\right)} \frac{1}{p} d\underline{x} = (1-\delta)^k = o(1). \qquad \square$$

### 5.3.1 Proof of Theorem 5.7, upper bound

The main proposition of the upper bound is as follows.

**Proposition 5.9.** *Let $p$ and $k(p)$ tend to $\infty$ in such a way that $k \le \frac{\log p}{\log\log p}$, and let $0 < \epsilon(p) < 1$ with $\epsilon(p)^2 k(p) \to \infty$. Set $t = t(p) = (1+\epsilon)\frac{k}{2\pi e} p^{\frac{2}{k}}$. For any fixed $\delta > 0$*

$$\mathbb{U}_{L(p,k)\times(\mathbb{R}/p\mathbb{Z})^k}\left[(\Lambda, \underline{x}) \in L(p,k)\times(\mathbb{R}/p\mathbb{Z})^k : \left|\Theta\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right) - \frac{1}{p}\right| < \frac{\delta}{p}\right] = (1 + o_\delta(1)). \tag{5.14}$$

*Deduction of Theorem 5.7, upper bound.* For any $\Lambda \in L^0(p,k)$ we have $p\mathbb{Z}^k < \Lambda$, and thus

$$\begin{aligned}\left\|\Theta\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right) - \frac{1}{p}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)} &= \int_{\mathbb{R}^k/\Lambda} \frac{1}{p} - \min\left(\Theta\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right), \frac{1}{p}\right) d\underline{x} \\ &= p^{-k+1} \int_{\underline{x}\in(\mathbb{R}/p\mathbb{Z})^k} \frac{1}{p} - \min\left(\Theta\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right), \frac{1}{p}\right) d\underline{x}\end{aligned}$$

and so

$$\begin{aligned}&(1+o(1))\mathbf{E}_{L^0(p,k)}\left[\left\|\Theta\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right) - \frac{1}{p}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)}\right] \\ &= \mathbf{E}_{L(p,k)}\left[\left\|\Theta\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right) - \frac{1}{p}\right\|_{\mathrm{TV}(\mathbb{R}^k/\Lambda)}\right] < \delta + o(1). \qquad \square\end{aligned}$$

Let $\tau = \tau(p) = \frac{\epsilon(p)}{2}$. Given $\underline{x} \in \mathbb{R}^k$, define spherical shell

$$S(\underline{x}, R, \tau) = \left\{\underline{y} \in \mathbb{R}^k : \|\underline{y} - \underline{x}\| \in [(1-\tau)R, (1+\tau)R]\right\}.$$

We use several times the estimate for $\underline{x} \in S(0, \sqrt{t}, \tau)$

$$\eta_k\left(\sqrt{\frac{2t}{2k+1}}, \underline{x}\right) \le \frac{1}{p}\exp\left(-\left(\frac{3\epsilon^2}{4} + O(\epsilon^3)\right)\frac{k}{2}\right) = o\left(\frac{1}{p}\right). \tag{5.15}$$

The critical part of $\Lambda$ when considering $\Theta\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right)$ in $L^1$ is $\Lambda_c(\underline{x}) = \Lambda \cap S(\underline{x}, \sqrt{t}, \tau)$. We split

$$\begin{aligned}\Theta\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right) &= \Theta_M\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right) + \Theta_E\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right); \\ \Theta_M\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right) &= \sum_{\lambda\in\Lambda_c(\underline{x})} \eta_k\left(\sqrt{\frac{2t}{2k+1}}, \lambda - \underline{x}\right).\end{aligned}$$

**Lemma 5.10.** *For all $\underline{x} \in (\mathbb{R}/p\mathbb{Z})^k$,*

$$\mathbf{E}_{L(p,k)}\left[\Theta_M\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right)\right] = \frac{1}{p}(1 + o(1)).$$

*Proof.* If $p$ is sufficiently large then there is at most one point of $p\mathbb{Z}^k$ contained in $\Lambda_c(\underline{x})$, and so (5.15) gives

$$\begin{aligned}&\mathbf{E}_{L(p,k)}\left[\Theta_M\left(\underline{x}, \frac{2t}{2k+1}; \Lambda\right)\right] \\ &= o\left(\frac{1}{p}\right) + \frac{p^{k-1}-1}{p^k-1}\sum_{\lambda\in\mathbb{Z}^k} \eta_k\left(\sqrt{\frac{2t}{2k+1}}, \lambda - \underline{x}\right)\mathbf{1}\left(\lambda \in S\left(\underline{x}, \sqrt{t}, \tau\right)\right)\end{aligned}$$

Let $\underline{v} \in \mathbb{R}^k$ be a unit vector, and let $D_{\underline{v}}$ denote the directional derivative in the $\underline{x}$ variable in direction $\underline{v}$. For any $\lambda \in S\left(\underline{x}, \sqrt{t}, 2\tau\right)$ we have

$$\left| D_{\underline{v}} \left( \log \eta_k \left( \sqrt{\frac{2t}{2k+1}}, \lambda - \underline{x} \right) \right) \right| \ll \frac{k}{\sqrt{t}} \ll \frac{\sqrt{k}}{p^{\frac{1}{k}}}.$$

In particular, for any $\underline{y} \in \left[-\frac{1}{2}, \frac{1}{2}\right)^k$, since $\|\underline{y}\|_2 \leq \frac{\sqrt{k}}{2}$, we have

$$\eta_k \left( \sqrt{\frac{2t}{2k+1}}, \lambda - \underline{x} \right) = (1 + o(1)) \eta_k \left( \sqrt{\frac{2t}{2k+1}}, \lambda - \underline{x} - \underline{y} \right).$$

Thus the sum may be approximated with an integral, and the result follows. □

**Lemma 5.11.** *We have the following estimates.*

$$\mathbf{E}_{L(p,k)\times(\mathbb{R}/p\mathbb{Z})^k} \left[ \Theta_M \left( \underline{x}, \frac{2t}{2k+1}; \Lambda \right) \right] = \frac{1}{p} \left(1 + o(1)\right)$$
$$\mathbf{E}_{L(p,k)\times(\mathbb{R}/p\mathbb{Z})^k} \left[ \Theta_E \left( \underline{x}, \frac{2t}{2k+1}; \Lambda \right) \right] = o\left(\frac{1}{p}\right)$$

*and for $k > 2$,*

$$\mathbf{E}_{(\mathbb{R}/p\mathbb{Z})^k} \left[ \mathbf{Var}_{L(p,k)} \left[ \Theta_M \left( \underline{x}, \frac{2t}{2k+1}; \Lambda \right) \right] \right] = o\left(\frac{1}{p^2}\right).$$

*Proof.* The evaluations of the means follow from Lemma 5.10.

In evaluating the variance term, we write, for $\lambda_1, \lambda_2 \in \mathbb{Z}^k$, $\lambda_1 \sim \lambda_2$ if $\lambda_2 \equiv a\lambda_1 \bmod p$ for some $0, 1 \not\equiv a \bmod p$. We have the following evaluations (see Lemma 5.2):

$$\begin{aligned} &\mathbb{U}_L\left(\Lambda : \lambda_1, \lambda_2 \in \Lambda\right) - \mathbb{U}_L\left(\Lambda : \lambda_1 \in \Lambda\right) \mathbb{U}_L\left(\Lambda : \lambda_2 \in \Lambda\right) \\ &\quad = \begin{cases} 0 & \lambda_1 \in p\mathbb{Z}^k \text{ or } \lambda_2 \in p\mathbb{Z}^k \\ O\left(p^{-k}\right) & \lambda_1, \lambda_2 \in \mathbb{Z}^k \setminus p\mathbb{Z}^k,\ \lambda_1 \neq \lambda_2, \lambda_1 \not\sim \lambda_2 \\ O\left(p^{-1}\right) & \lambda_1, \lambda_2 \in \mathbb{Z}^k \setminus p\mathbb{Z}^k,\ \lambda_1 \sim \lambda_2 \text{ or } \lambda_1 = \lambda_2 \end{cases}. \end{aligned}$$

The variance thus evaluates to

$$\begin{aligned} &\mathbf{E}_{(\mathbb{R}/p\mathbb{Z})^k} \left[ \mathbf{Var}_{L(p,k)} \left[ \Theta_M \left( \underline{x}, \frac{2t}{2k+1}; \Lambda \right) \right] \right] \ll \\ &\frac{1}{p^k} \sum_{\lambda_1, \lambda_2 \in \mathbb{Z}^k \setminus (p\mathbb{Z})^k} \left( \mathbf{1}(\lambda_1 \sim \lambda_2) \frac{1}{p} + O(p^{-k}) \right) \qquad (5.16) \\ &\times \int_{\underline{x} \in [-\frac{p}{2}, \frac{p}{2})^k \cap S(\lambda_1, \sqrt{t}, \tau) \cap S(\lambda_2, \sqrt{t}, \tau)} \eta_k \left( \sqrt{\frac{2t}{2k+1}}, \lambda_1 - \underline{x} \right) \eta_k \left( \sqrt{\frac{2t}{2k+1}}, \lambda_2 - \underline{x} \right) d\underline{x} \\ &+ \frac{1}{p^{k+1}} \sum_{\lambda \in \mathbb{Z}^k \setminus (p\mathbb{Z})^k} \int_{\underline{x} \in \left[-\frac{p}{2}, \frac{p}{2}\right)^k \cap S(\lambda, \sqrt{t}, \tau)} \eta_k \left( \sqrt{\frac{2t}{2k+1}}, \lambda - \underline{x} \right)^2 d\underline{x}. \qquad (5.17) \end{aligned}$$

The term (5.17) captures $\lambda_1 = \lambda_2$. Replacing one Gaussian by the bound (5.15) and then estimating as for the mean of $\Theta_M$ gives a bound for this term of

$$(5.17) = o\left(\frac{1}{p^2}\right).$$

The error term $O(p^{-k})$ of (5.16) may be bounded by omitting the restriction on $\|\lambda_2 - \underline{x}\|$ and summing over $\lambda_2$, the summation being bounded by a constant. The remaining

summation over $\lambda_1$ and integral over $\underline{x}$ are then evaluated as for the mean, and give an error of $O(p^{-k})$.

It remains to treat those terms from (5.16) with $\lambda_1 \sim \lambda_2$. Let $R(\tau) = 2(1+\tau)\sqrt{t}$. Any $\lambda_1 \sim \lambda_2$ contributing to the variance satisfies $\lambda = \lambda_1 - \lambda_2 \in B(0, R(\tau)) \setminus \{0\}$ and $\lambda_1 \equiv (a+1)\lambda \bmod p\mathbb{Z}^k$, $\lambda_2 \equiv a\lambda \bmod p\mathbb{Z}^k$ for some $a \bmod p$. Arranging the summation over $\lambda$ and $a$, we find that the contribution of terms with $\lambda_1 \sim \lambda_2$ to (5.16) is bounded by (by expanding the integral, this is now independent of $a$, which we pull out)

$$\ll \frac{1}{p^k} \sum_{\lambda \in \mathbb{Z}^k \cap B(0,R(\tau)) \setminus \{0\}} \int_{\underline{x} \in S(0,\sqrt{t},\tau) \cap S(\lambda,\sqrt{t},\tau)} \eta_k\left(\sqrt{\frac{2t}{2k+1}}, \underline{x}\right) \eta_k\left(\sqrt{\frac{2t}{2k+1}}, \lambda - \underline{x}\right) d\underline{x}.$$

The total number of such $\lambda$ is $\ll 2^k(1+\tau)^k(1+\epsilon)^k p$ by estimating with the volume of the ball, see Lemma 2.5. Putting in the bound (5.15) for one Gaussian and integrating the second over all of $\mathbb{R}^k$, we obtain an estimate from the terms with $\lambda_1 \sim \lambda_2$ of $\ll \frac{8^k}{p^k}$. □

*Proof of Proposition 5.9.* Consider separately the cases

$$|\Theta_E|, \left|\mathbf{E}_{L(p,k)}[\Theta_M] - \frac{1}{p}\right|, |\Theta_M - \mathbf{E}_{L(p,k)}[\Theta_M]| > \frac{\delta}{3p}$$

and apply Markov's inequality. □

# 6 The power-of-2 random walk

## 6.1 A Chebyshev cut-off criterion

We begin by describing a commonly used second moment method for proving cut-off, which we apply in analyzing the power-of-2 random walk. The following is a variant of the lower bound method from [11], see also Wilson's lemma in [20].

Given a probability measure $\mu$ on $\mathbb{Z}/p\mathbb{Z}$ and frequency $\xi \in \mathbb{Z}/p\mathbb{Z}$, define the Fourier coefficient of $\mu$ at $\xi$ to be

$$\hat{\mu}(\xi) = \sum_{x \in \mathbb{Z}/p\mathbb{Z}} \mu(x) e\left(\frac{\xi x}{p}\right).$$

Define, as before, the $L^2$ mixing time by

$$t_2^{\mathrm{mix}} = \inf\left\{ n \in \mathbb{Z}_{>0} : \sum_{0 \neq \xi \in \mathbb{Z}/p\mathbb{Z}} |\hat{\mu}(\xi)|^{2n} \leq \frac{4}{e^2} \right\}$$

and the spectral gap

$$\mathrm{gap} = 1 - \max_{0 \neq \xi \in \mathbb{Z}/p\mathbb{Z}} |\hat{\mu}(\xi)|.$$

**Proposition 6.1.** *Let $\{A_p \subset \mathbb{Z}/p\mathbb{Z}\}_{p \in \mathscr{P}}$ be a sequence of symmetric, lazy, generating sets for $\mathbb{Z}/p\mathbb{Z}$, with $\mu_{A_p}$ the corresponding uniform probability measure. Assume that the spectral gap tends to 0 with increasing $p$.*

*Suppose the following holds for each fixed $\epsilon > 0$. For each $p \in \mathscr{P}$ there exists symmetric subset $0 \notin B_p \subset \widehat{\mathbb{Z}/p\mathbb{Z}}$ such that as $p \to \infty$,*

- *For all $\xi \in B_p$,*
$$\hat{\mu}_{A_p}(\xi) = 1 - o(1). \tag{6.1}$$
- *For all $n < (1-\epsilon)t_2^{\mathrm{mix}}(p)$*
$$\frac{1}{\sqrt{|B_p|}} \sum_{\xi \in B_p} \hat{\mu}_{A_p}^n(\xi) \to \infty \tag{6.2}$$

- *For all* $n < (1-\epsilon)t_2^{\mathrm{mix}}(p)$

$$\sum_{\xi_1,\xi_2\in B_p} \hat{\mu}_{A_p}^n(\xi_1-\xi_2) \le (1+o(1)) \sum_{\xi_1,\xi_2\in B_p} \hat{\mu}_{A_p}^n(\xi_1)\hat{\mu}_{A_p}^n(\xi_2). \tag{6.3}$$

*Then the sequence* $\{(\mathbb{Z}/p\mathbb{Z}, \mu_{A_p}, \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}})\}$ *converges to uniform in total variation distance with a cut-off at* $t_1^{\mathrm{mix}}(p) \sim t_2^{\mathrm{mix}}(p)$ *if and only if the condition*

$$t_2^{\mathrm{mix}}(p)\,\mathrm{gap}(p) \to \infty \qquad \text{as } p\to\infty \tag{6.4}$$

*is satisfied.*

**Remark 6.2.** The condition (6.3) is in fact equivalent to

$$\sum_{\xi_1,\xi_2\in B_p} \hat{\mu}_{A_p}^n(\xi_1-\xi_2) = (1+o(1)) \sum_{\xi_1,\xi_2\in B_p} \hat{\mu}_{A_p}^n(\xi_1)\hat{\mu}_{A_p}^n(\xi_2) \tag{6.5}$$

since

$$\sum_{\xi_1,\xi_2\in B_p} \hat{\mu}_{A_p}^n(\xi_1-\xi_2) \ge \sum_{\xi_1,\xi_2\in B_p} \hat{\mu}_{A_p}^n(\xi_1)\hat{\mu}_{A_p}^n(\xi_2)$$

by the following application of Cauchy-Schwarz:

$$\begin{aligned}
\sum_{\xi_1,\xi_2\in B_p} \hat{\mu}_{A_p}^n(\xi_1)\hat{\mu}_{A_p}^n(\xi_2) &= \left|\sum_{\xi\in B_p} \hat{\mu}_{A_p}^n(\xi)\right|^2 \\
&= \left|\sum_{x \bmod p} \mu_{A_p}^{*n}(x) \sum_{\xi\in B_p} e\left(\frac{\xi x}{p}\right)\right|^2 \\
&\le \left(\sum_{x \bmod p} \mu_{A_p}^{*n}(x)\right)\left(\sum_{x \bmod p} \mu_{A_p}^{*n}(x) \sum_{\xi_1,\xi_2\in B_p} e\left(\frac{(\xi_1-\xi_2)x}{p}\right)\right) \\
&= \sum_{\xi_1,\xi_2\in B_p} \hat{\mu}_{A_p}^n(\xi_1-\xi_2).
\end{aligned}$$

*Proof of Proposition 6.1.* Since $t_1^{\mathrm{mix}} \le t_2^{\mathrm{mix}}$, if the condition $\mathrm{gap}(p)\cdot t_2^{\mathrm{mix}}(p) \to \infty$ fails then there is no cut-off in total variation, so we may assume that this condition holds. Let $\epsilon > 0$ be fixed. For $n > (1+\epsilon)t_2^{\mathrm{mix}}$, by Cauchy-Schwarz,

$$\begin{aligned}
\left\|\mu_{A_p}^{*n} - \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}}\right\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})}^2 &\le \frac{1}{4} \sum_{\xi\not\equiv 0 \bmod p} |\hat{\mu}_{A_p}(\xi)|^{(2+2\epsilon)t_2^{\mathrm{mix}}} \\
&\le \frac{1}{4}(1-\mathrm{gap})^{2t_2^{\mathrm{mix}}\epsilon} \sum_{\xi\not\equiv 0 \bmod p} \left|\hat{\mu}_{A_p}(\xi)\right|^{2t_2^{\mathrm{mix}}} \to 0
\end{aligned} \tag{6.6}$$

since $\mathrm{gap}\cdot t_2^{\mathrm{mix}} \to \infty$.

To prove the lower bound, let $n < (1-\epsilon)t_2^{\mathrm{mix}}$. Define function $f_p(x)$ on $\mathbb{Z}/p\mathbb{Z}$ by $f_p(x) = \frac{1}{\sqrt{|B_p|}} \sum_{\xi\in B_p} \hat{\mu}_{A_p}(\xi) e\left(\frac{-\xi x}{p}\right)$. Writing $\mathbf{E}_\mu$, $\mathbf{Var}_\mu$ for expectation and variance with respect to probability measure $\mu$, we have

$$\mathbf{E}_{\mathbb{U}_{\mathbb{Z}/p\mathbb{Z}}}[f_p] = \frac{1}{p} \sum_{x \bmod p} \frac{1}{\sqrt{|B_p|}} \sum_{\xi\in B_p} \hat{\mu}_{A_p}(\xi) e\left(\frac{-\xi x}{p}\right) = \frac{1}{\sqrt{|B_p|}} \sum_{\xi\in B_p} \hat{\mu}_{A_p}(\xi)\delta_{\xi=0} = 0 \tag{6.7}$$

since $0 \notin B_p$, and

$$\begin{aligned}\mathbf{Var}_{\mathbb{U}_{\mathbb{Z}/p\mathbb{Z}}}[f_p] &= \frac{1}{p}\sum_{x \bmod p}\frac{1}{|B_p|}\sum_{\xi_1,\xi_2\in B_p}\hat{\mu}_{A_p}(\xi_1)\hat{\mu}_{A_p}(\xi_2)e\left(\frac{-(\xi_1-\xi_2)x}{p}\right) \\ &= \frac{1}{|B_p|}\sum_{\xi\in B_p}\hat{\mu}_{A_p}(\xi)^2 \le 1.\end{aligned} \tag{6.8}$$

Meanwhile

$$\begin{aligned}\mathbf{E}_{\mu_{A_p}^{*n}}[f_p] &= \sum_{x \bmod p}\frac{1}{\sqrt{|B_p|}}\sum_{\xi\in B_p}\hat{\mu}_{A_p}(\xi)e\left(\frac{-\xi x}{p}\right)\mu_{A_p}^{*n}(x) \\ &= \frac{1}{\sqrt{|B_p|}}\sum_{x \bmod p}\sum_{\xi\in B_p}\hat{\mu}_{A_p}(\xi)e\left(\frac{-\xi x}{p}\right)\frac{1}{p}\sum_{\eta \bmod p}\hat{\mu}_{A_p}^n(\eta)e\left(\frac{-\eta x}{p}\right) \\ &= \frac{1}{\sqrt{|B_p|}}\sum_{\xi\in B_p}\hat{\mu}_{A_p}(\xi)\hat{\mu}_{A_p}^n(-\xi) \\ &= (1+o(1))\frac{1}{\sqrt{|B_p|}}\sum_{\xi\in B_p}\hat{\mu}_{A_p}^n(\xi)\end{aligned} \tag{6.9}$$

and

$$\begin{aligned}\mathbf{E}_{\mu_{A_p}^{*n}}\left[f_p^2\right] &= \sum_{x \bmod p}\frac{1}{|B_p|}\sum_{\xi_1,\xi_2\in B_p}\hat{\mu}_{A_p}(\xi_1)\hat{\mu}_{A_p}(\xi_2)e\left(\frac{-(\xi_1-\xi_2)x}{p}\right)\mu_{A_p}^{*n}(x) \\ &= \frac{1}{|B_p|}\sum_{\xi_1,\xi_2\in B_p}\hat{\mu}_{A_p}(\xi_1)\hat{\mu}_{A_p}(\xi_2)\hat{\mu}_{A_p}^n(-(\xi_1-\xi_2)) \\ &= (1+o(1))\frac{1}{|B_p|}\sum_{\xi_1,\xi_2\in B_p}\hat{\mu}_{A_p}^n(\xi_1-\xi_2).\end{aligned} \tag{6.10}$$

It follows by condition (6.3) that

$$\mathbf{Var}_{\mu_{A_p}^{*n}}[f_p] = \mathbf{E}_{\mu_{A_p}^{*n}}\left[f_p^2\right] - \mathbf{E}_{\mu_{A_p}^{*n}}[f_p]^2 = o\left(\mathbf{E}_{\mu_{A_p}^{*n}}\left[f_p^2\right]\right). \tag{6.11}$$

Let $X_p$ be the subset of $\mathbb{Z}/p\mathbb{Z}$ defined by

$$X_p = \left\{x : f_p(x) \le \frac{1}{2}\mathbf{E}_{\mu_{A_p}^{*n}}[f_p]\right\}.$$

By (6.9) and condition (6.2),

$$\mathbf{E}_{\mu_{A_p}^{*n}}[f_p] \to \infty. \tag{6.12}$$

Hence Chebyshev's inequality, (6.7), (6.8) and (6.12) imply

$$\mathbb{U}_{\mathbb{Z}/p\mathbb{Z}}[X_p] = 1 - o(1)$$

while Chebyshev, (6.11) and (6.12) imply

$$\mu_{A_p}^{*n}(X_p) = o(1).$$

We conclude

$$\left\|\mu_{A_p}^{*n} - \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}}\right\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})} \ge \left|\mathbb{U}_{\mathbb{Z}/p\mathbb{Z}}(X_p) - \mu_{A_p}^{*n}(X_p)\right| = 1 - o(1). \quad \square$$

### 6.2 Proof of Theorem 1.11, lower bound

Recall that we set $\ell = \ell_2(p) = \lceil \log_2 p \rceil$ and

$$c_0 = \sum_{j=1}^{\infty} \left(1 - \cos \frac{2\pi}{2^j}\right).$$

We prove the lower bound of Theorem 1.11 conditional on $t_2^{\mathrm{mix}} \lesssim \frac{\ell \log \ell}{2c_0}$, which is proven in the next section. The proof of the lower bound is a reduction to the conditions of Proposition 6.1.

Let $J = o(\log\log p)$ be a parameter. With an eye toward applying Proposition 6.1, set

$$B_p = \left\{ \xi \in \widehat{\mathbb{Z}/p\mathbb{Z}} : \exists\, 1 \le j_1 \ne j_2 \le \ell,\ \left\| \frac{\xi}{p} - 2^{-j_1} + 2^{-j_2} \right\|_{\mathbb{R}/\mathbb{Z}} \le 2^{-\ell-J} \right\}.$$

**Lemma 6.3.** $|B_p| \gg \frac{\ell^2}{2^J} - \ell$.

*Proof.* Let

$$S = \left\{ \xi \bmod p : \exists\, 1 \le j \le \ell,\ \left\| \frac{\xi}{p} - \frac{1}{2^j} \right\| \le 2^{-\ell} \right\}.$$

We have $\ell \le |S| \le 2\ell$. For each $s \in S$ write $\frac{s}{p}$ in binary

$$\frac{s}{p} = *.s_1 s_2 s_3 \ldots.$$

Partition $S$ into $2^{J+1}$ sets $S_1, S_2, \ldots, S_{2^{J+1}}$ according to the digits $s_\ell s_{\ell+1} \ldots s_{\ell+J}$. To each pair $s \ne s' \in S_i$ we obtain $r = s - s' \in B_p$. The multiplicity with which a given such $r$ arises in this way is $O(1)$. Hence

$$|B_p| \gg \sum_{j=1}^{2^{J+1}} |S_j|(|S_j| - 1) = -|S| + \sum_{j=1}^{2^{J+1}} |S_j|^2.$$

By Cauchy-Schwarz,

$$|S|^2 = \left( \sum_j |S_j| \right)^2 \le 2^{J+1} \sum_j |S_j|^2$$

so

$$|B_p| \gg \frac{|S|^2}{2^{J+1}} - |S| \ge \frac{\ell^2}{2^{J+1}} - \ell. \quad \square$$

**Lemma 6.4.** *For $\xi \in B_p$, $\hat{\mu}_{A_{2,p}}(\xi) \ge 1 - \frac{4c_0}{2\ell+1} - O\left(\frac{1}{2^J \ell}\right)$.*

*Proof.* After possibly replacing $\xi$ with $-\xi$ we may take $\xi = 2^{-j_1} - 2^{-j_2} + O\left(2^{-\ell-J}\right)$ with $j_1 < j_2$. Then

$$\begin{aligned}
1 - \hat{\mu}_{A_{2,p}}(\xi) &= \frac{2}{2\ell+1} \sum_{i=0}^{\ell-1} \left(1 - \cos\left(2\pi \left(2^{i-j_1} - 2^{i-j_2} + O\left(2^{i-\ell-J}\right)\right)\right)\right) \\
&= O\left(\frac{1}{2^J \ell}\right) + \frac{2}{2\ell+1} \left[ \sum_{i=0}^{j_1-1} \left(1 - \cos\left(2\pi \left(2^{i-j_1} - 2^{i-j_2}\right)\right)\right) + \sum_{i=j_1}^{j_2-1} \left(1 - \cos\left(2\pi 2^{i-j_2}\right)\right) \right] \\
&\le O\left(\frac{1}{2^J \ell}\right) + \frac{2}{2\ell+1} \left[ \sum_{i=-\infty}^{j_1-1} \left(1 - \cos\left(2\pi 2^{i-j_1}\right)\right) + \sum_{i=-\infty}^{j_2-1} \left(1 - \cos\left(2\pi 2^{i-j_2}\right)\right) \right] \\
&= \frac{4c_0}{2\ell+1} + O\left(\frac{1}{2^J \ell}\right).
\end{aligned}$$

$\square$

**Lemma 6.5.** *For all but $O\left(J\ell^3\right)$ pairs $\xi_1 \neq \xi_2 \in B_p$*

$$\hat{\mu}_{A_{2,p}}(\xi_1 - \xi_2) = 1 - \frac{8c_0}{2\ell+1} + O\left(\frac{1}{2^J\ell}\right).$$

*For all but $O\left(J^2\ell^2\right)$ pairs $\xi_1 \neq \xi_2 \in B_p$,*

$$\hat{\mu}_{A_{2,p}}(\xi_1 - \xi_2) \leq 1 - \frac{6c_0}{2\ell+1} + O\left(\frac{1}{2^J\ell}\right).$$

*For all but $O\left(J^3\ell\right)$ pairs $\xi_1 \neq \xi_2 \in B_p$,*

$$\hat{\mu}_{A_{2,p}}(\xi_1 - \xi_2) \leq 1 - \frac{4c_0}{2\ell+1} + O\left(\frac{1}{2^J\ell}\right).$$

*Proof.* Write $\frac{\xi_1}{p} = 2^{-j_1} - 2^{-j_2} + O\left(2^{-\ell-J}\right)$, $\frac{\xi_2}{p} = 2^{-j_3} - 2^{-j_4} + O\left(2^{-\ell-J}\right)$. By excluding at most $O\left(J\ell^3\right)$ quadruples $(j_1, j_2, j_3, j_4)$ we may assume $j_i > J$ for all $i$ and $|j_i - j_k| \geq J$ for all $i \neq k$ in $\{1,2,3,4\}$. Then

$$\begin{aligned}
1 - \hat{\mu}_{A_{2,p}}(\xi) &= \frac{2}{2\ell+1}\sum_{i=0}^{\ell-1}\left(1 - \cos\left(2\pi\left(2^{i-j_1} - 2^{i-j_2} - 2^{i-j_3} + 2^{i-j_4} + O\left(2^{i-\ell-J}\right)\right)\right)\right) \\
&= O\left(\frac{1}{2^J\ell}\right) + \frac{2}{2\ell+1}\left[\sum_{k=1}^{4}\sum_{i=j_k-J}^{j_k-1}\left(1 - \cos\left(2\pi 2^{i-j_k}\right)\right)\right] \\
&= O\left(\frac{1}{2^J\ell}\right) + \frac{8c_0}{2\ell+1}.
\end{aligned}$$

For the second statement, by excluding $O\left(J^2\ell^2\right)$ tuples $(j_1, j_2, j_3, j_4)$ we may assume that three of $j_1, j_2, j_3, j_4$ are larger than $J$ and mutually separated by at least $J$. One argues as before, using the additional calculation that for $1 < j < J$,

$$\sum_{i=1}^{J}\left(1 - \cos\left(2\pi\left(2^{-i} \pm 2^{j-i}\right)\right)\right) \geq \sum_{i=1}^{J}\left(1 - \cos\left(2\pi 2^{-i}\right)\right) = c_0 + O\left(2^{-J}\right),$$

which holds since for all $i, j > 0$,

$$\|2^{-i} \pm 2^{j-i}\|_{\mathbb{R}/\mathbb{Z}} \geq \|2^{-i}\|_{\mathbb{R}/\mathbb{Z}}.$$

The third statement is similar. □

*Proof of Theorem 1.11, lower bound.* Let $\epsilon > 0$ be given, and suppose that $n < (1-\epsilon)\frac{\ell\log\ell}{2c_0}$.

Set $J = 2\log\log\ell$ and define $B_p$ as above. It suffices to show that $B_p$ satisfies conditions (6.2) and (6.5) of Proposition 6.1.

By Lemma 6.4

$$\begin{aligned}
\frac{1}{\sqrt{|B_p|}}\sum_{\xi\in B_p}\left(\hat{\mu}_{A_{2,p}}(\xi)\right)^n &\geq \sqrt{|B_p|}\exp\left[(1-\epsilon)\frac{\ell\log\ell}{2c_0}\left(\frac{-4c_0}{2\ell+1} + O\left(\frac{1}{2^J\ell}\right)\right)\right] \\
&\geq \sqrt{|B_p|}\ell^{\epsilon-1}\left(1 + O\left(\frac{\log\ell}{2^J}\right)\right).
\end{aligned}$$

In particular Lemma 6.3 implies

$$\frac{1}{\sqrt{|B_p|}}\sum_{\xi\in B_p}\left(\hat{\mu}_{A_{2,p}}(\xi)\right)^n \gg \frac{\ell^\epsilon}{2^{\frac{J}{2}}} = \ell^{\epsilon - o(1)}$$

and condition (6.2) is satisfied.

To check (6.5), split $\xi_1, \xi_2 \in B_p$ according as $\xi_1 = \xi_2$, or $\xi_1, \xi_2$ fall into one of the several cases enumerated in Lemma 6.5. This gives

$$\begin{aligned}\sum_{\xi_1,\xi_2\in B_p} \left(\hat{\mu}_{A_{2,p}}(\xi_1-\xi_2)\right)^n \le |B_p| + |B_p|^2 \exp\left[(1-\epsilon)\frac{\ell\log\ell}{2c_0}\left(-\frac{4c_0}{\ell}+O\left(\frac{1}{2^J\ell}\right)\right)\right] \\ + O\left(J\ell^3\right)\exp\left[(1-\epsilon)\frac{\ell\log\ell}{2c_0}\left(\frac{-3c_0}{\ell}+O\left(\frac{1}{2^J\ell}\right)\right)\right] \\ + O\left(J^2\ell^2\right)\exp\left[(1-\epsilon)\left(\frac{\ell\log\ell}{2c_0}\left(\frac{-2c_0}{\ell}+O\left(\frac{1}{2^J\ell}\right)\right)\right)\right] \\ + O\left(J^3\ell\right).\end{aligned}$$

By Lemma 6.3, $|B_p| = \ell^{2-o(1)}$, and thus all but the second term is an error term. Condition (6.5) holds, since

$$|B_p|^2 \exp\left[(1-\epsilon)\frac{\ell\log\ell}{2c_0}\left(-\frac{4c_0}{\ell}\right)\right] \le (1+o(1))\left(\sum_{\xi\in B_p}(\hat{\mu}_{A_{2,p}}(\xi))^n\right)^2. \qquad \square$$

### 6.3 Proof of Theorem 1.11, upper bound

We prove the following somewhat more precise estimate.

**Proposition 6.6.** *For all $0 < \beta < \log\ell$, for all $n \ge \frac{\ell}{2c_0}(\log\ell + \beta)$ we have*

$$\|\mu_{A_{2,p}}^{*n} - \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}}\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})}^2 \ll e^{-\beta} + \frac{e^{-\frac{\beta}{c_0}}\log\ell}{\ell^{\frac{1}{c_0}}}.$$

**Remark 6.7.** The second term results from a discrepancy between the eigenvalue generating the spectral gap and the bulk of the large spectrum which determines the mixing time. With more effort, the factor of $\log\ell$ could be removed.

The proof uses the following frequently used application of the Cauchy-Schwarz inequality, see [6] for an introduction to these types of estimates, also [7].

**Lemma 6.8.** *Let $\mu$ be a probability measure on finite abelian group $G$. We have the upper bound*

$$\|\mu - \mathbb{U}_G\|_{\mathrm{TV}(G)} \le \frac{1}{2}\left(\sum_{0\neq\chi\in\widehat{G}}|\hat{\mu}(\chi)|^2\right)^{\frac{1}{2}}.$$

*In particular,*

$$\|\mu_{A_{2,p}}^{*n} - \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}}\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})} \le \frac{1}{2}\left(\sum_{0\neq\xi \bmod p}\left|\hat{\mu}_{A_{2,p}}(\xi)\right|^{2n}\right)^{\frac{1}{2}}. \tag{6.13}$$

*Proof.* We have

$$\|\mu - \mathbb{U}_G\|_{\mathrm{TV}(G)} = \frac{1}{2}\sum_{x\in G}|\mu(x) - \mathbb{U}_G(x)|.$$

Hence, by Cauchy-Schwarz,

$$\|\mu - \mathbb{U}(G)\|_{\mathrm{TV}(G)} \le \frac{1}{2}\left(|G|\sum_{x\in G}|\mu(x)-\mathbb{U}_G(x)|^2\right)^{\frac{1}{2}} = \frac{1}{2}\left(\sum_{0\neq\chi\in\widehat{G}}|\hat{\mu}(\chi)|^2\right)^{\frac{1}{2}}. \qquad \square$$

The above lemma reduces to estimation of the size of the Fourier coefficients $\hat{\mu}_{A_{2,p}}(\xi)$. In estimating these coefficients it will be convenient to use the following modified binary expansion of $\frac{\xi}{p}$.

**Lemma 6.9.** *Let $p \geq 3$ be prime. For each $0 \not\equiv \xi \bmod p$ there is an increasing sequence $\mathscr{I} = \{i_j\}_{j=1}^{\infty} \subset \mathbb{Z}_{>0}$, and $\epsilon = \pm 1$ such that*

$$\frac{\xi}{p} \equiv \epsilon \sum_{j=1}^{\infty} (-1)^j 2^{-i_j} \bmod 1.$$

*This representation is unique.*

*Proof.* Write $-\frac{\xi}{p}$ in binary as $*.s_1 s_2 s_3 ...$ with each $s_i \in \{0,1\}$, then write $\frac{\xi}{p} = -\frac{\xi}{p} - \left(-\frac{2\xi}{p}\right)$ where $\left(-\frac{2\xi}{p}\right)$ is obtained by a left shift, and then the subtraction is performed bitwise.

The uniqueness follows because any two distinct such representations $(\epsilon, \{i_j\}), (\epsilon', \{i'_j\})$ differ by $\gg 2^{-J}$, where $J$ is $\min(i_1, i'_1)$ if $\epsilon \neq \epsilon'$, and otherwise is the least integer which appears in the symmetric difference $\{i_j\} \Delta \{i'_j\}$. □

### 6.3.1 Index sequences

We introduce several notions which will be useful in the remainder of the argument.

Given a real parameter $J > 0$, define a *$J$-sequence* of non-negative integers to be an ordered set $A \subset \mathbb{Z}_{\geq 0}$, with members enumerated $A = a_1 < a_2 < ...$ such that any pair of consecutive elements differ by at most $J$. $|A|$ denotes the cardinality. Set $i(A) = a_1$, $t(A) = \sup(A)$. A $J$-sequence with $a_1 = 0$ is called *normalized*. Given $J$-sequence $A = a_1 < a_2 < ...$, its *off-set sequence* is the normalized $J$-sequence $A' = 0 < a_2 - a_1 < a_3 - a_1 < ....$ For instance,

$$1, 3, 7, 8, 10, 14$$

is a 4-sequence with offset sequence

$$0, 2, 6, 7, 9, 13.$$

A $J$-sequence is called *non-trivial* if it contains a pair of elements that differ by more than 1. We denote $\mathscr{J}$ the set of $J$-sequences, $\mathscr{J}_0$ the set of normalized $J$-sequences and $\mathscr{J}'_0 = \mathscr{J}_0 \setminus \{\{0\}, \{0,1\}\}$ the set of non-trivial normalized $J$-sequences.

A $J$-sequence $A$ contained in sequence $B \subset \mathbb{Z}_{\geq 0}$ is called a $J$-subsequence. We say that $J$-subsequence $A \subset B$ is *maximal* if it is not properly contained in another $J$-subsequence $A' \subset B$. Given parameter $J$, one easily checks that any $B \subset \mathbb{Z}_{\geq 0}$ has a unique partition into maximal $J$-subsequences. For instance, in the first sequence above,

$$1, 3; \quad 7, 8, 10; \quad 14$$

is a partition into maximal 2-subsequences.

We write $\mathscr{C}(B)$ for the set of maximal $J$-subsequences of $B$. The $J$-sequences in $\mathscr{C}(B)$ are *$J$-separated* in the sense that if $A_1 \neq A_2 \in \mathscr{C}(B)$ and $x_1 \in A_1, x_2 \in A_2$ then $|x_1 - x_2| > J$. The sequences in $\mathscr{C}(B)$ are naturally ordered by, for $A_1, A_2 \in \mathscr{C}(B)$, $A_1 < A_2$ if and only if for any $x_1 \in A_1, x_2 \in A_2$, $x_1 < x_2$.

In the remainder of the argument we think of the non-zero bits in the expansion of $\frac{\xi}{p}$ above as partitioned into maximal $J$-sequences. These $J$-separated parts do not interact significantly in calculating the Fourier transform. The argument that follows quantifies the interaction.

Let $J \geq \log_2 \ell$ be a parameter. Given $\xi \bmod p$, represent $\xi$ as $(\mathscr{I}(\xi), \epsilon(\xi))$ as above. Truncate $\mathscr{I}(\xi)$ to $\mathscr{I}'(\xi) = \mathscr{I}(\xi) \cap (0, \ell]$ (note that $\epsilon$ and $\mathscr{I}'$ determine $\xi$) and set

$$\sigma(\xi) = |\mathscr{I}'(\xi)|, \qquad \mathscr{C}(\xi) = \mathscr{C}(\mathscr{I}'(\xi)). \tag{6.14}$$

We call $\mathscr{C}(\xi)$ the set of *clumps* of $\xi$, each clump being a $J$-sequence. If there exists $C \in \mathscr{C}(\xi)$ with $i(C) \leq J$ we say that $C$ is *initial*. A clump $C$ with $t(C) > \ell - J$ is *final*. We write $C_{\rm init}(\xi)$, $C_{\rm fin}(\xi)$ for the initial and final clump, with the convention that $C_{\rm init} = \emptyset$ if there is no initial clump, and similarly $C_{\rm fin}$. A clump is *typical* if it is neither initial nor final. $\mathscr{C}_0(\xi) \subset \mathscr{C}(\xi)$ is the subset of typical clumps.

Given frequency $\xi$, define the *savings* of $\xi$ to be

$$\mathrm{sav}(\xi) = \frac{2\ell+1}{2}\left(1 - \hat{\mu}_{A_{2,p}}(\xi)\right) = \sum_{l=0}^{\ell-1}\left(1 - \cos\left(2\pi\left(\sum_{k=1}^{\infty}(-1)^k 2^{l-i_k}\right)\right)\right). \tag{6.15}$$

For a typical clump $C \in \mathscr{C}_0(\xi)$ also define

$$\mathrm{sav}(C) = \sum_{i(C)-J \leq l < t(C)}\left[1 - \cos\left(2\pi \sum_{i_k \in C}(-1)^k 2^{l-i_k}\right)\right]. \tag{6.16}$$

**Lemma 6.10.** *We have*

$$\mathrm{sav}(\xi) \geq \sum_{C \in \mathscr{C}_0(\xi)} \mathrm{sav}(C) + |C_{\rm init}(\xi)| + |C_{\rm fin}(\xi)| + O\left(2^{-J}|\mathscr{C}|\right).$$

*Proof.* Since the clumps $C \in \mathscr{C}$ are $J$-separated, we have

$$\begin{aligned}\mathrm{sav}(\xi) \geq \sum_{C \in \mathscr{C}_0(\xi)} \sum_{i(C)-J \leq l < t(C)}&\left[1 - \cos\left(2\pi \sum_{i_k \in C}(-1)^k 2^{l-i_k}\right) + O\left(2^{-J-t(C)+l}\right)\right] \\ &+ \sum_{i \in C_{\rm init}(\xi)}\left(1 - \cos\left(\frac{\pi}{2}\right)\right) + \sum_{i \in C_{\rm fin}(\xi)}\left(1 - \cos\left(\frac{\pi}{2}\right)\right),\end{aligned}$$

where in the last two sums we specialize to $j = i - 1$, and note that for any fixed $l$ $\frac{1}{2} \geq \left|\sum_{i_j \geq l}(-1)^{i_j} 2^{l-i_j-1}\right| \geq \frac{1}{4}$. □

In a similar spirit we have the following crude estimate for savings.

**Lemma 6.11.** *Let $0 \not\equiv \xi \in \widehat{\mathbb{Z}/p\mathbb{Z}}$ and let $C \in \mathscr{C}_0(\xi)$. We have* $\mathrm{sav}(C) \geq |C|$.

*Proof.* Write $C = i_1 < \cdots < i_j$. We have

$$\mathrm{sav}(C) = \sum_{l=i_1-J}^{i_j-1}\left[1 - \cos\left(2\pi \sum_{m=1}^{j}(-1)^m 2^{l-i_m}\right)\right] \geq \sum_{m=1}^{j}\left(1 - \cos\left(\frac{\pi}{2}\right)\right)$$

by specializing to $l = i_m - 1$, $m = 1, ..., j$. □

**Lemma 6.12.** *Let $0 \not\equiv \xi \in \widehat{\mathbb{Z}/p\mathbb{Z}}$. For fixed $\delta_3 > 0$, for $\sigma(\xi)$ as in (6.14),*

$$|\hat{\mu}(\xi)| \leq \max\left(1 - \frac{2\sigma(\xi)}{2\ell+1}, 1 - \delta_3\right).$$

*Proof.* The bound $\hat{\mu}(\xi) \leq 1 - \frac{2\sigma(\xi)}{2\ell+1}$ follows from Lemmas 6.10 and 6.11. The bound $\hat{\mu}(\xi) \geq 1 - \delta_3$ follows from $\frac{1}{2}(\cos\theta + \cos 2\theta) \geq -1 + c$ for a fixed $c > 0$. □

For typical clumps we require slightly stronger estimates.

**Lemma 6.13.** *Let $0 \not\equiv \xi \in \widehat{\mathbb{Z}/p\mathbb{Z}}$. Let $C \in \mathscr{C}_0(\xi)$, and suppose that $|C| = j > 1$. Enumerate $C = i_1 < i_2 < ... < i_j$. There exists fixed $\delta_1 > 0$ such that if $i_2 > i_1 + 1$ then $\mathrm{sav}(C) > \mathrm{sav}(C') + \delta_1$ where $C'$ is the $J$-sequence formed by $i_2 - 1, i_2, ..., i_j$, i.e. by shifting $i_1$ to the place adjacent to $i_2$.*

*Proof.* We have

$$\mathrm{sav}(C) - \mathrm{sav}(C') = \sum_{i_1 - J \le l < i_2 - 1} \left[1 - \cos\left(2\pi \sum_{m=1}^{j} (-1)^m 2^{l - i_m}\right)\right] - \sum_{i_2 - J - 1 \le l < i_2 - 1} \left[1 - \cos\left(2\pi\left(-2^{l+1-i_2} + \sum_{m=2}^{j} (-1)^m 2^{l-i_m}\right)\right)\right].$$

Set

$$x = \sum_{m=3}^{j} (-1)^{m-1} 2^{-i_m + i_2 - 1}, \qquad 0 \le x \le \frac{1}{4}.$$

Take only the first $J$ terms of the first sum, to obtain for some $\delta_1 > 0$

$$\mathrm{sav}(C) - \mathrm{sav}(C') \ge \sum_{l=1}^{J} \left[\cos\left(\frac{2\pi}{2^l}\left(\frac{-3}{4} - \frac{x}{2}\right)\right) - \cos\left(\frac{2\pi}{2^l}\left(\frac{-1}{2} - x\right)\right)\right] > \delta_1$$

by noting that the worst case is $i_1 = i_2 - 2$. □

**Lemma 6.14.** *Let $0 \not\equiv \xi \in \widehat{\mathbb{Z}/p\mathbb{Z}}$. Let $C \in \mathscr{C}_0(\xi)$, and suppose that $|C| = j > 1$, $C = i_1 < \cdots < i_j$ with $i_2 = i_1 + 1$. Then $\mathrm{sav}(C) \ge \mathrm{sav}(C')$ where $C'$ is the $J$-sequence formed by $i_2, ..., i_j$, i.e. by dropping $i_1$. Furthermore, if $j \ge 3$ and $i_3 = i_1 + 2$ then there exists fixed $\delta_2 > 0$ such that $\mathrm{sav}(C) > \mathrm{sav}(C') + \delta_2$.*

*Proof.* We have

$$\mathrm{sav}(C) - \mathrm{sav}(C') = \sum_{l=i_1-J}^{i_1-1} \left[1 - \cos\left(2\pi \sum_{m=1}^{j} (-1)^m 2^{l-i_m}\right)\right] - \sum_{l=i_1-J+1}^{i_1-1} \left[1 - \cos\left(2\pi \sum_{m=2}^{j} (-1)^m 2^{l-i_m}\right)\right].$$

Replace $l$ with $i_1 - 1 - l$ and set $x = \sum_{m=2}^{j} (-1)^m 2^{i_m - i_1 - 1}$, $\frac{1}{8} \le x \le \frac{1}{4}$ to obtain

$$\mathrm{sav}(C) - \mathrm{sav}(C') \ge \sum_{l=0}^{J-2} \left[\cos\left(\frac{2\pi x}{2^l}\right) - \cos\left(\frac{2\pi(x - \frac{1}{2})}{2^l}\right)\right]. \tag{6.17}$$

In the case $j \ge 3$ and $i_3 = i_1 + 2$ we have $x \le \frac{3}{16}$ which proves

$$(6.17) \ge \cos\left(\frac{3\pi}{8}\right) - \cos\left(\frac{5\pi}{8}\right). \quad \square$$

The previous two lemmas imply the following one.

**Lemma 6.15.** *Let $0 \not\equiv \xi \in \widehat{\mathbb{Z}/p\mathbb{Z}}$. Let $C \in \mathscr{C}_0(\xi)$ be a typical clump of $\mathscr{C}(\xi)$ with digits $i_1 < i_2 < ... < i_j$. If $j = 1$ or $j = 2$ and $i_2 = i_1 + 1$ we have*

$$\mathrm{sav}(C) \ge c_0 + O(2^{-J}).$$

*Furthermore, there is a $\delta > 0$ such that, if $j \ge 3$ or $j = 2$ and $i_2 > i_1 + 1$ then*

$$\mathrm{sav}(C) \ge c_0 + \delta j.$$

*Proof.* By a sequence of steps in which we either (i) move the first index of $C$ adjacent to the second, or (ii) delete the first, we reduce to case of $C_0$ containing a single element, which satisfies $\mathrm{sav}(C_0) = c_0 - O(2^{-J})$. □

We collect together several easy combinatorial estimates. Given frequency $\xi$ we are most interested in typical clumps $C \in \mathscr{C}_0(\xi)$ which consist of a single index, or a pair of adjacent indices. Let the number of these be $x_1(\xi)$ and $x_2(\xi)$. Let $x_3(\xi) = |\mathscr{C}_0(\xi)| - x_1(\xi) - x_2(\xi)$ be the number of non-trivial clumps in $\mathscr{C}_0(\xi)$, and let $m = \sigma(\xi) - |C_{\mathrm{init}}| - |C_{\mathrm{fin}}| - x_1(\xi) - 2x_2(\xi)$ be the number of indices contained in the clumps counted in $x_3(\xi)$.

Given $m \geq 0$ and $x_3 \geq 0$, let

$$\mathscr{T}(m, x_3) = \left\{ \underline{A} \in (\mathscr{J}_0')^{x_3} : \sum_{i=1}^{x_3} |A_j| = m \right\}$$

be the collection of $x_3$-tuples of non-trivial normalized $J$-sequences of total cardinality $m$. Given initial and final clumps $C_{\mathrm{init}}$ and $C_{\mathrm{fin}}$, $T \in \mathscr{T}(m, x_3)$ and integers $x_1, x_2 \geq 0$, let $\mathscr{N}(C_{\mathrm{init}}, C_{\mathrm{fin}}, x_1, x_2, T)$ denote the number of $\xi$ with initial clump $C_{\mathrm{init}}$, final clump $C_{\mathrm{fin}}$, $x_1$ typical clumps with a single index, $x_2$ typical clumps which consist of a pair of consecutive indices and $x_3$ non-trivial typical clumps, whose offsets taken in order are given by $T$. For any $j \geq 0$, let $I(j)$ (resp. $F(j)$) be the number of $J$-sequences on $j$ indices which may appear as the initial (resp. final) clump of $\mathscr{J}(\xi)$, $\xi \in \widehat{\mathbb{Z}/p\mathbb{Z}} \setminus \{0\}$.

**Lemma 6.16.** *Let $x_1, x_2, x_3, m, T$ be as above and let $C_{\mathrm{init}}$, $C_{\mathrm{fin}}$ be any initial and final clumps (possibly empty). We have the bounds*

$$|\mathscr{T}(m, x_3)| \leq (J+1)^{m-1}$$

*and, for any $T \in \mathscr{T}(m, x_3)$,*

$$\mathscr{N}(C_{\mathrm{init}}, C_{\mathrm{fin}}, x_1, x_2, T) \leq 2\frac{\ell^{x_1+x_2+x_3}}{x_1!x_2!x_3!}.$$

*Also, for any $j \geq 0$,*

$$I(j), F(j) \leq J^j.$$

*Proof.* To bound $|\mathscr{T}|$, neglecting $x_3$ and the non-triviality condition, choose for each index $1 \leq j < m$ a distance $1 \leq d(j) \leq J+1$ between $j$ and $j+1$ in the arrangement, with a distance of $J+1$ indicating that a new clump begins with $j+1$.

Similarly, the bound for $I(j)$ follows on choosing a first index in one of at most $J$ ways, and then choosing sequentially distances between the consecutive indices. For $F(j)$, choose counting from the back instead.

The bound for $\mathscr{N}(C_{\mathrm{init}}, C_{\mathrm{fin}}, x_1, x_2, T)$ follows on choosing a first index for each clump, the factor of 2 coming from choosing the sign. □

Our results on savings may be summarized as follows.

**Lemma 6.17.** *Let $0 \not\equiv \xi \in \widehat{\mathbb{Z}/p\mathbb{Z}}$ have parameters $x_1, x_2, x_3, m, C_{\mathrm{init}}, C_{\mathrm{fin}}$ as above. There is a fixed $0 < \delta < \frac{1}{2}$ such that*

$$\mathrm{sav}(\xi) \geq c_0(x_1 + x_2 + x_3) + \delta m + |C_{\mathrm{init}}| + |C_{\mathrm{fin}}| - O\left(\frac{x_1 + x_2 + x_3}{2^J}\right).$$

*Proof of Proposition 6.6.* Let $\log J = o(\log \ell)$ and fix some $\theta$, $1 - \frac{1}{c_0} < \theta < 1$. By Lemma 6.8

$$\begin{aligned}\left\|\mu_{A_{2,p}}^{*n} - \mathbb{U}_{\mathbb{Z}/p\mathbb{Z}}\right\|_{\mathrm{TV}(\mathbb{Z}/p\mathbb{Z})}^2 &\le \frac{1}{4}\sum_{\xi \not\equiv 0 \bmod p} \left|\hat{\mu}_{A_{2,p}}(\xi)\right|^{2n} \\ &= \frac{1}{4}\left\{\sum_{1\le\sigma(\xi)<\ell^\theta} + \sum_{\ell^\theta\le\sigma(\xi)<\delta_3\ell} + \sum_{\delta_3\ell\le\sigma(\xi)}\right\}\left|\hat{\mu}_{A_{2,p}}(\xi)\right|^{2n} \\ &= \frac{1}{4}\left(\mathscr{S}_1 + \mathscr{S}_2 + \mathscr{S}_3\right).\end{aligned}$$

By Lemma 6.12, for some $c > 0$,

$$\mathscr{S}_3 \le 2^\ell \left(1 - \frac{2\ell\delta_3}{2\ell+1}\right)^{\frac{\ell}{c_0}(\log\ell+\beta)} = O(e^{-c\ell\log\ell}). \tag{6.18}$$

By Lemma 6.12, again for some $c > 0$,

$$\begin{aligned}\mathscr{S}_2 &\le 2\sum_{\ell^\theta\le j<\delta_3\ell}\binom{\ell}{j}\left(1-\frac{2j}{2\ell+1}\right)^{\frac{\ell}{c_0}(\log\ell+\beta)} \\ &\ll \sum_{\ell^\theta\le j<\delta_3\ell}\exp\left(j\log\ell - j\log j + j - \frac{2\ell j}{(2\ell+1)c_0}\log\ell - \frac{2j\ell\beta}{c_0(2\ell+1)}\right) \\ &\ll e^{-c\ell^\theta\beta}\sum_{\ell^\theta\le j<\delta_3\ell}\exp\left(\left(-\theta+1-\frac{1}{c_0}\right)j\log\ell + j\right) = O\left(e^{-c\ell^\theta(\beta+\log\ell)}\right).\end{aligned} \tag{6.19}$$

Conditioning on $x_1(\xi), x_2(\xi), x_3(\xi), m$ as in Lemma 6.16 and $i = |C_{\mathrm{init}}|$, $f = |C_{\mathrm{fin}}|$ we find

$$\begin{aligned}\mathscr{S}_1 \le 2 &\sum_{1\le j<\ell^\theta}\sum_{x_1+2x_2+m+i+f=j}\sum_{|C_{\mathrm{init}}|=i,|C_{\mathrm{fin}}|=j}\sum_{x_3\le\lfloor\frac{m}{2}\rfloor} \\ &\times\sum_{T\in\mathscr{T}(m,x_3)}\mathscr{N}(C_{\mathrm{init}},C_{\mathrm{fin}},x_1,x_2,T)\left(1-\frac{2\,\mathrm{sav}}{2\ell+1}\right)^{\frac{\ell}{c_0}(\log\ell+\beta)}\end{aligned}$$

where $\mathrm{sav} = c_0(x_1+x_2+x_3) + \delta m + i + f - O\left(\frac{x_1+x_2+x_3}{2^J}\right)$. Inserting the estimates for $|\mathscr{T}|$ and $\mathscr{N}$ from Lemma 6.16, we obtain

$$\begin{aligned}\mathscr{S}_1 \ll &\sum_{1\le j<\ell^\theta}\sum_{x_1+2x_2+m+i+f=j}(J+1)^{m+i+f}\frac{\ell^{x_1+x_2}}{x_1!x_2!} \\ &\sum_{\substack{x_3\le\lfloor\frac{m}{2}\rfloor\\ m>0\Rightarrow x_3>0}}\frac{\ell^{x_3}}{x_3!}\exp\left(-\left(x_1+x_2+x_3+\frac{\delta m+i+f}{c_0}-O\left(\frac{x_1+x_2+x_3}{2^J}\right)\right)(\log\ell+\beta)\right).\end{aligned}$$

Assume that $\frac{\log\ell}{2^J} = o(1)$. Then, when $m \ge 1$ we find that the sum over $x_3$ is

$$O\left(\exp(-\beta)\right).$$

The terms for which $x_1 = x_2 = 0$ thus contribute $O\left(\frac{J}{e^\beta \ell^{\frac{\delta}{c_0}+o(1)}} + \frac{J}{e^{\frac{\beta}{c_0}}\ell^{\frac{1}{c_0}}}\right)$. When $x_1+x_2 \ne 0$ summation over $i, f, m$ reduces to $1 + o(1)$. Thus

$$\mathscr{S}_1 \le O\left(\frac{J}{e^{\frac{\beta}{c_0}}\ell^{\frac{1}{c_0}}}\right) + O\left(\exp\left(-\beta + O\left(\frac{\log\ell+\beta}{2^J}\right)\right)\right).$$

Choose $2^J = \ell$ to complete the proof. □

## A Local limit theorem on $\mathbb{R}^k$

For $k > 1$ recall that we define the measure on $\mathbb{Z}^k$,

$$\nu_k = \frac{1}{2k+1}\left(\delta_0 + \sum_{j=1}^{k}(\delta_{e_j} + \delta_{-e_j})\right)$$

and that we write

$$\eta_k(\sigma, \underline{x}) = \frac{1}{(2\pi\sigma^2)^{\frac{k}{2}}}\exp\left(-\frac{\|\underline{x}\|_2^2}{2\sigma^2}\right)$$

for the density of the centered standard normal distribution on $\mathbb{R}^k$. In this appendix we prove Lemma 1.10, which we recall for convenience.

**Lemma.** *Let $n, k(n) \geq 1$ with $k^2 = o(n)$ for large $n$. As $n \to \infty$ we have*

$$\left\| \nu_k^{*n} * \mathbf{1}_{[-\frac{1}{2},\frac{1}{2})^k} - \eta_k\left(\sqrt{\frac{2n}{2k+1}}, \cdot\right)\right\|_{\mathrm{TV}(\mathbb{R}^k)} = o(1).$$

We actually prove a stronger estimate, which is a local limit theorem on $\mathbb{R}^k$ for which we don't know an easy reference.

**Lemma A.1.** *Let $n, k(n) \geq 1$ with $k^2 = o(n)$ for large $n$. Uniformly for $\underline{\alpha} \in \mathbb{Z}^k$ such that $\|\underline{\alpha}\|_2^2 \leq \frac{2kn}{2k+1} + \frac{n\log n}{\sqrt{k}}$, and $\|\underline{\alpha}\|_4^4 \ll \frac{n^2}{k}\left(1 + \frac{\log n}{\sqrt{k}}\right)$, as $n \to \infty$,*

$$\nu_k^{*n}(\alpha) = \{1 + o(1)\}\,\eta_k\left(\sqrt{\frac{2n}{2k+1}}, \alpha\right).$$

The deduction of Lemma 1.10 is as follows.

*Proof of Lemma 1.10.* We have, for any $A, \delta > 0$, and for some $C > 0$,

$$\int_{\|\underline{x}\|_2^2 > \frac{2kn}{2k+1} + \delta\frac{n\log n}{\sqrt{k}}} \eta_k\left(\sqrt{\frac{2n}{2k+1}}, \underline{x}\right) d\underline{x} = O_{\delta,A}\left(n^{-A}\right)$$

$$\int_{\|\underline{x}\|_4^4 > C\frac{n^2}{k}\left(1+\frac{\log n}{\sqrt{k}}\right)} \eta_k\left(\sqrt{\frac{2n}{2k+1}}, \underline{x}\right) d\underline{x} = O_A\left(n^{-A}\right)$$

see Lemma 2.2, so it suffices to estimate the difference

$$\nu_k^{*n} * \mathbf{1}_{[-\frac{1}{2},\frac{1}{2})^k}(\underline{x}) - \eta_k\left(\sqrt{\frac{2n}{2k+1}}, \underline{x}\right)$$

for $\|\underline{x}\|_2^2 \leq \frac{2kn}{2k+1} + O\left(\frac{n\log n}{\sqrt{k}}\right)$ and $\|\underline{x}\|_4^4 \ll \frac{n^2}{k}\left(1 + \frac{\log n}{\sqrt{k}}\right)$.

For $\underline{x} \in \mathbb{Z}^k$ satisfying this upper bound and for $\underline{y} \in [-\frac{1}{2}, \frac{1}{2})^k$,

$$\begin{aligned}\eta_k\left(\sqrt{\frac{2n}{2k+1}}, \underline{x} + \underline{y}\right) &= \eta_k\left(\sqrt{\frac{2n}{2k+1}}, \underline{x}\right)\exp\left(-\frac{2k+1}{4n}\left(2\underline{x}\cdot\underline{y} + \|\underline{y}\|_2^2\right)\right)\\ &= (1+o(1))\eta_k\left(\sqrt{\frac{2n}{2k+1}}, \underline{x}\right)\exp\left(-\frac{(2k+1)\underline{x}\cdot\underline{y}}{2n}\right).\end{aligned}$$

Therefore

$$\int_{\left[-\frac{1}{2},\frac{1}{2}\right)^k} \left|\eta_k\left(\sqrt{\frac{2n}{2k+1}},\underline{x}+\underline{y}\right)-\eta_k\left(\sqrt{\frac{2n}{2k+1}},\underline{x}\right)\right| d\underline{y}$$
$$= \eta_k\left(\sqrt{\frac{2n}{2k+1}},\underline{x}\right)\left(o(1)+(1+o(1))\int_{\left[-\frac{1}{2},\frac{1}{2}\right)^k}\left|\exp\left(-\frac{(2k+1)\underline{x}\cdot\underline{y}}{2n}\right)-1\right| d\underline{y}\right).$$

We claim that for all $\|\underline{x}\|_2^2 \ll \frac{2kn}{2k+1}+\frac{n\log n}{\sqrt{k}}$,

$$\int_{\left[-\frac{1}{2},\frac{1}{2}\right)^k}\left|\exp\left(-\frac{(2k+1)\underline{x}\cdot\underline{y}}{2n}\right)-1\right| d\underline{y} = o(1). \tag{A.1}$$

To see that this suffices for the proof, let

$$\mathscr{B} = \left\{\underline{x}\in\mathbb{Z}^k : \|\underline{x}\|_2^2 \le \frac{2kn}{2k+1}+\frac{n\log n}{\sqrt{k}},\ \|x\|_4^4 \le C\frac{n^2}{k}\left(1+\frac{\log n}{\sqrt{k}}\right)\right\}$$

and estimate

$$\sum_{\underline{x}\in\mathscr{B}}\int_{\left[-\frac{1}{2},\frac{1}{2}\right)^k}\left|\eta_k\left(\sqrt{\frac{2n}{2k+1}},\underline{x}+\underline{y}\right)-\eta_k\left(\sqrt{\frac{2n}{2k+1}},\underline{x}\right)\right| d\underline{y}$$
$$\le o(1)\sum_{\underline{x}\in\mathscr{B}}\eta_k\left(\sqrt{\frac{2n}{2k+1}},\underline{x}\right)$$
$$= o(1)\sum_{\underline{x}\in\mathscr{B}}(1+o(1))\eta_k^{*n}(\underline{x}) = o(1)$$

where in the last line we apply Lemma A.1. Since

$$\sum_{\underline{x}\in\mathscr{B}}\int_{\left[-\frac{1}{2},\frac{1}{2}\right)^k}\eta_k\left(\sqrt{\frac{2n}{2k+1}},\underline{x}+\underline{y}\right)d\underline{y} = 1+o(1)$$

it follows that $\sum_{\underline{x}\in\mathscr{B}}\eta_k^{*n}(\underline{x}) = 1+o(1)$ so that

$$\left\|\nu_k^{*n} * \mathbf{1}_{\left[-\frac{1}{2},\frac{1}{2}\right)^k} - \eta_k\left(\sqrt{\frac{2n}{2k+1}},\cdot\right)\right\|_{\mathrm{TV}(\mathbb{R}^k)}$$
$$= \sum_{\underline{x}\in\mathbb{Z}^k}\int_{\left[-\frac{1}{2},\frac{1}{2}\right)^k}\left|\eta_k\left(\sqrt{\frac{2n}{2k+1}},\underline{x}+\underline{y}\right)-\eta_k\left(\sqrt{\frac{2n}{2k+1}},\underline{x}\right)\right| d\underline{y} = o(1)$$

by bounding both terms in the sum over $\underline{x}\in\mathscr{B}^c$ separately.

To prove (A.1), choose a parameter $A = A(n,k)\to\infty$ with $n$ such that $A = o\left(\frac{\sqrt{n}}{k}\right)$ and partition $\left[-\frac{1}{2},\frac{1}{2}\right)^k = \mathscr{S}_{\mathrm{good}}\sqcup\mathscr{S}_{\mathrm{bad}}$ with

$$\mathscr{S}_{\mathrm{good}} = \left\{\underline{y}\in\left[-\frac{1}{2},\frac{1}{2}\right)^k : |\underline{x}\cdot\underline{y}| \le A\sqrt{n}\left(1+\frac{\log n}{\sqrt{k}}\right)\right\}.$$

By Azuma's inequality, for some fixed $C>0$,

$$\mathrm{meas}\left(\underline{y}\in\left[-\frac{1}{2},\frac{1}{2}\right)^k : |\underline{x}\cdot\underline{y}| > t\sqrt{n}\left(1+\frac{\log n}{\sqrt{k}}\right)\right) \le 2\exp\left(-\frac{t^2}{C}\right),$$

and thus $\mathrm{meas}(\mathscr{S}_{\mathrm{good}}) = 1 - o(1)$. Since $\exp\left(-\frac{(2k+1)\underline{x}\cdot\underline{y}}{2n}\right) = 1 + o(1)$ for all $\underline{y} \in \mathscr{S}_{\mathrm{good}}$ we have

$$\int_{\underline{y}\in\mathscr{S}_{\mathrm{good}}} \left|\exp\left(-\frac{(2k+1)\underline{x}\cdot\underline{y}}{2n}\right) - 1\right| d\underline{y} = o(1).$$

Meanwhile,

$$\begin{aligned}
&\int_{\underline{y}\in\mathscr{S}_{\mathrm{bad}}} \left|\exp\left(-\frac{(2k+1)\underline{x}\cdot\underline{y}}{2n}\right) - 1\right| d\underline{y} \leq \int_{\underline{y}\in\mathscr{S}_{\mathrm{bad}}} \exp\left(\left|\frac{(2k+1)\underline{x}\cdot\underline{y}}{2n}\right|\right) d\underline{y} \\
&= -\int_A^\infty \exp\left(\frac{t(2k+1)\left(1+\frac{\log n}{\sqrt{k}}\right)}{\sqrt{n}}\right) d\,\mathrm{meas}\left(\underline{y} : |\underline{y}\cdot\underline{x}| > t\sqrt{n}\left(1+\frac{\log n}{\sqrt{k}}\right)\right) \\
&\ll \exp\left(\frac{-A^2}{C} + o(1)\right) = o(1).
\end{aligned}$$

□

The proof of Lemma A.1 is a standard application of the saddle point method. As there are several intermediate lemmas, it may help the reader to skip ahead to first read the eventual proof. Associate to $\nu_k$ the generating function

$$f(z_1, ..., z_k) = \frac{1}{2k+1}\left(1 + z_1 + z_1^{-1} + ... + z_k + z_k^{-1}\right),$$

so that $\nu_k(\alpha) = C_\alpha[f]$, where for Laurent series in multiple variables

$$g(z_1, ..., z_k) = \sum_{n_1,...,n_k=-\infty}^{\infty} a_{n_1,...,n_k} z_1^{n_1} ... z_k^{n_k}$$

we write $C_{\underline{\alpha}}[g] = a_{\underline{\alpha}}$. The generating function associated to $\nu_k^{*n}$ is thus $f^n$.

By symmetry we may assume $\underline{\alpha} \geq 0$ coordinatewise. By Cauchy's theorem, for any $R_1, ..., R_k > 0$

$$\begin{aligned}
\nu_k^{*n}(\alpha) &= \left(\frac{1}{2\pi i}\right)^k \int_{|z_1|=R_1} \cdots \int_{|z_k|=R_k} \frac{f(z_1, ..., z_k)^n}{z_1^{\alpha_1} ... z_k^{\alpha_k}} \frac{dz_1}{z_1} \cdots \frac{dz_k}{z_k} \\
&= \frac{1}{R_1^{\alpha_1} ... R_k^{\alpha_k}} \int_{(\mathbb{R}/\mathbb{Z})^k} f_0(\theta_1, ..., \theta_k)^n e\left(-\underline{\alpha}\cdot\underline{\theta}\right) d\underline{\theta}, \qquad \text{(A.2)}
\end{aligned}$$

where

$$f_0(\theta_1, ..., \theta_k) = f(R_1 e(\theta_1), ..., R_k e(\theta_k)).$$

and $\underline{\alpha}\cdot\underline{\theta}$ is the usual dot product on $\mathbb{R}^k$. The asymptotic in Lemma A.1 is derived by choosing $R_1, ..., R_k$ such that the phase in $f_0(\underline{\theta})^n$ is approximately equal to $e(\underline{\alpha}\cdot\underline{\theta})$ for $\underline{\theta}$ near 0. The main contribution of the integral then comes from small $\underline{\theta}$.

Let $\mathscr{D}_{\mathrm{sm}}$ be the domain

$$\mathscr{D}_{\mathrm{sm}} = \left\{\underline{\theta} \in (\mathbb{R}/\mathbb{Z})^k : \|\underline{\theta}\|_\infty \leq \frac{1}{12}\right\}.$$

For $\underline{\theta} \in \mathscr{D}_{\mathrm{sm}}$ define

$$\begin{aligned}
F(\underline{\theta}) = n\log\left[\frac{1}{2k+1}\left(1 + R_1 e(\theta_1) + \frac{e(-\theta_1)}{R_1} + ... + R_k e(\theta_k) + \frac{e(-\theta_k)}{R_k}\right)\right] \\
- (\alpha_1 \log R_1 + ... + \alpha_k \log R_k) - 2\pi i \underline{\alpha}\cdot\underline{\theta}.
\end{aligned}$$

Evidently $F(\underline{\theta})$ gives a continuous definition of $\log \frac{f_0(\underline{\theta})^n e(-\underline{\alpha}\cdot\underline{\theta})}{R_1^{\alpha_1} ... R_k^{\alpha_k}}$ on $\mathscr{D}_{\mathrm{sm}}$.

**Lemma A.2.** *The first few partial derivatives of $F(\underline{\theta})$ are given as follows*

$$D_jF(\underline{\theta}) = 2\pi i\left[\frac{n}{2k+1}\frac{R_je(\theta_j)-\frac{e(-\theta_j)}{R_j}}{f_0(\underline{\theta})}-\alpha_j\right]$$

$$D_{j_1}D_{j_2}F(\underline{\theta}) = \frac{4\pi^2 n}{(2k+1)^2}\frac{\left(R_{j_1}e(\theta_{j_1})-\frac{e(-\theta_{j_1})}{R_{j_1}}\right)\left(R_{j_2}e(\theta_{j_2})-\frac{e(-\theta_{j_2})}{R_{j_2}}\right)}{f_0(\underline{\theta})^2},\quad j_1\neq j_2$$

$$D_j^2F(\underline{\theta}) = -\frac{4\pi^2 n}{2k+1}\frac{R_je(\theta_j)+\frac{e(-\theta_j)}{R_j}}{f_0(\underline{\theta})}+\frac{4\pi^2 n}{(2k+1)^2}\frac{\left(R_je(\theta_j)-\frac{e(-\theta_j)}{R_j}\right)^2}{f_0(\underline{\theta})^2}$$

$$D_{j_1}D_{j_2}D_{j_3}F(\underline{\theta}) = \frac{-16\pi^3 in}{(2k+1)^3}\frac{\left(R_{j_1}e(\theta_{j_1})-\frac{e(-\theta_{j_1})}{R_{j_1}}\right)\left(R_{j_2}e(\theta_{j_2})-\frac{e(-\theta_{j_2})}{R_{j_2}}\right)\left(R_{j_3}e(\theta_{j_3})-\frac{e(-\theta_{j_3})}{R_{j_3}}\right)}{f_0(\underline{\theta})^3},$$

$j_1, j_2, j_3$ *distinct*

$$D_{j_1}^2D_{j_2}F(\underline{\theta}) = \frac{8\pi^3 in}{(2k+1)^2}\frac{\left(R_{j_1}e(\theta_{j_1})+\frac{e(-\theta_{j_1})}{R_{j_1}}\right)\left(R_{j_2}e(\theta_{j_2})-\frac{e(-\theta_{j_2})}{R_{j_2}}\right)}{f_0(\underline{\theta})^2} - \frac{16\pi^3 in}{(2k+1)^3}\frac{\left(R_{j_1}e(\theta_{j_1})-\frac{e(-\theta_{j_1})}{R_{j_1}}\right)^2\left(R_{j_2}e(\theta_{j_2})-\frac{e(-\theta_{j_2})}{R_{j_2}}\right)}{f_0(\underline{\theta})^3},\quad j_1\neq j_2$$

$$D_j^3F(\underline{\theta}) = -\frac{8\pi^3 in}{2k+1}\frac{R_je(\theta_j)-\frac{e(-\theta_j)}{R_j}}{f_0(\underline{\theta})} + \frac{24\pi^3 in}{(2k+1)^2}\frac{\left(R_je(\theta_j)+\frac{e(-\theta_j)}{R_j}\right)\left(R_je(\theta_j)-\frac{e(-\theta_j)}{R_j}\right)}{f_0(\underline{\theta})^2} - \frac{16\pi^3 in}{(2k+1)^3}\frac{\left(R_je(\theta_j)-\frac{e(-\theta_j)}{R_j}\right)^3}{f_0(\underline{\theta})^3}.$$

Choose $R_j$ by solving the stationary phase equation, for each $j$, $D_jF(0)=0$, thus

$$\frac{n}{2k+1}\frac{R_j-\frac{1}{R_j}}{f_0(0)}-\alpha_j = 0. \tag{A.3}$$

**Lemma A.3.** *Let $n, k(n)\in\mathbb{Z}_{>0}$ with $k^2=o(n)$ as $n\to\infty$. Let $\underline{\alpha}\in\mathbb{Z}^k$ and assume $\|\underline{\alpha}\|_2^2\leq n\left(1+\frac{\log n}{\sqrt{k}}\right)$ and $\|\underline{\alpha}\|_4^4\ll\frac{n^2}{k}\left(1+\frac{\log n}{\sqrt{k}}\right)$. The stationary phase equations (A.3) have a solution and the solution satisfies*

$$f_0(0) = 1+\frac{2k+1}{4n^2}\|\underline{\alpha}\|_2^2+O\left(\frac{k^3}{n^4}\|\underline{\alpha}\|_4^4\right) \tag{A.4}$$

$$R_j+\frac{1}{R_j} = 2+\left(f_0(0)\alpha_j\frac{2k+1}{2n}\right)^2+O\left(\frac{k^4\alpha_j^4}{n^4}\right) \tag{A.5}$$

$$= 2+O\left(\frac{\alpha_j^2k^2}{n^2}\right)$$

$$\log R_j = \frac{2k+1}{2n}f_0(0)\alpha_j+O\left(\frac{k^3\alpha_j^3}{n^3}\right). \tag{A.6}$$

*Proof.* We have the system of equations

$$f_0(0) = 1 + \frac{1}{2k+1}\sum_{j=1}^{k}\left(R_j + \frac{1}{R_j} - 2\right). \tag{A.7}$$

and

$$R_j + \frac{1}{R_j} = \sqrt{4 + \left(f_0(0)\alpha_j \frac{2k+1}{n}\right)^2}. \tag{A.8}$$

Beginning from an initial guess $f_0(0) = 2$, solve for each $R_j$ in $R_j \geq 1$ according to (A.8), sequentially update $f_0(0)$ and then the $R_j$. This produces a decreasing sequence of guesses for $f_0(0)$ and for each $R_j$, as is evident since the first step is decreasing, e.g. since

$$R_j + \frac{1}{R_j} - 2 \leq \frac{f_{0,\mathrm{old}}(0)^2\alpha_j^2(2k+1)^2}{4n^2}$$

and therefore,

$$f_{0,\mathrm{new}}(0) \leq 1 + \frac{2k+1}{4n^2} f_{0,\mathrm{old}}(0)^2 \|\underline{\alpha}\|_2^2 \leq 1 + O\left(\frac{1}{n}\right).$$

As $f_0(0)$ is bounded below, the sequence necessarily converges.

To verify the asymptotics, note that $f_0(0) = O(1)$ leads to

$$\begin{aligned} R_j + \frac{1}{R_j} &= 2 + \frac{\left(f_0(0)\alpha_j \frac{2k+1}{n}\right)^2}{2 + \sqrt{4 + \left(f_0(0)\alpha_j \frac{2k+1}{n}\right)^2}} \\ &= 2 + \left(f_0(0)\alpha_j \frac{2k+1}{2n}\right)^2 + O\left(\frac{k^4\alpha_j^4}{n^4}\right), \end{aligned}$$

which satisfies the claimed asymptotic.

Inserted into the formula for $f_0(0)$, this yields

$$f_0(0) = 1 + \frac{2k+1}{4n^2}\|\underline{\alpha}\|_2^2 f_0(0)^2 + O\left(\frac{k^3}{n^4}\|\underline{\alpha}\|_4^4\right).$$

The error introduced by the factor of $f_0(0)^2$ may be absorbed into the last error term, since $\|\underline{\alpha}\|_2^4 \leq k\|\underline{\alpha}\|_4^4$.

Combining the stationary phase equation with (A.5) we find

$$R_j = 1 + \frac{2k+1}{2n} f_0(0)\alpha_j + \frac{1}{2}\left(\frac{2k+1}{2n} f_0(0)\alpha_j\right)^2 + O\left(\frac{k^4}{n^4}\alpha_j^4\right),$$

so

$$\log R_j = \frac{2k+1}{2n} f_0(0)\alpha_j + O\left(\frac{k^3\alpha_j^3}{n^3}\right) \square$$

**Lemma A.4.** *Let $n, k(n) \in \mathbb{Z}_{>0}$ with $k(n)^2 = o(n)$ as $n \to \infty$. Let $\underline{\alpha} \in \mathbb{Z}^k$ and assume $\|\underline{\alpha}\|_2^2 \leq n\left(1 + \frac{\log n}{\sqrt{k}}\right)$ and $\|\underline{\alpha}\|_4^4 \ll \frac{n^2}{k}\left(1 + \frac{\log n}{\sqrt{k}}\right)$. Let $R_j$ be determined by the saddle*

*point equations (A.3). For $\underline{\theta} \in \mathscr{D}_{\mathrm{sm}}$ we have*

$$\begin{aligned}
F(0) &= -\frac{2k+1}{4n}\|\underline{\alpha}\|_2^2 + O\left(\frac{k^3}{n^3}\|\underline{\alpha}\|_4^4\right)\\
D_jF(0) &= 0\\
D_{j_1}D_{j_2}F(0) &= \frac{4\pi^2\alpha_{j_1}\alpha_{j_2}}{n},\ j_1 \neq j_2\\
D_j^2F(0) &= \frac{-8\pi^2 n}{2k+1} + \frac{2\pi^2\|\underline{\alpha}\|_2^2}{n} + O\left(\frac{k\alpha_j^2}{n}\right) + O\left(\frac{k^2}{n^3}\|\underline{\alpha}\|_4^4\right)\\
D_{j_1}D_{j_2}D_{j_3}F(\underline{\theta}) &= O\left(\frac{n}{k^3}\left(|\theta_{j_1}| + \frac{k|\alpha_{j_1}|}{n}\right)\left(|\theta_{j_2}| + \frac{k|\alpha_{j_2}|}{n}\right)\left(|\theta_{j_3}| + \frac{k|\alpha_{j_3}|}{n}\right)\right),\\
&\qquad j_1, j_2, j_3 \text{ distinct}\\
D_{j_1}^2D_{j_2}F(\underline{\theta}) &= O\left(\frac{n}{k^2}\left(|\theta_{j_2}| + \frac{k|\alpha_{j_2}|}{n}\right)\right),\ \ j_1 \neq j_2\\
D_j^3F(\underline{\theta}) &= O\left(\frac{n}{k}\left(|\theta_j| + \frac{k|\alpha_j|}{n}\right)\right).
\end{aligned}$$

*Proof.* We have

$$\begin{aligned}
F(0) &= n\log f(0) - \sum_j \alpha_j \log R_j\\
&= \frac{2k+1}{4n}\|\underline{\alpha}\|_2^2 - \frac{2k+1}{2n}f_0(0)\|\underline{\alpha}\|_2^2 + O\left(\frac{k^3}{n^3}\|\underline{\alpha}\|_4^4\right)\\
&= -\frac{2k+1}{4n}\|\underline{\alpha}\|_2^2 + O\left(\frac{k^3}{n^3}\|\underline{\alpha}\|_4^4\right).
\end{aligned}$$

At the saddle point, the first derivatives vanish. The mixed derivatives are evaluated by plugging in

$$R_j - \frac{1}{R_j} = \frac{2k+1}{n}f_0(0)\alpha_j.$$

We have

$$\begin{aligned}
D_j^2F(0) &= -\frac{4\pi^2 n}{2k+1}\frac{R_j + \frac{1}{R_j}}{f_0(0)} + \frac{4\pi^2\alpha_j^2}{n}\\
&= -\frac{8\pi^2 n}{2k+1}\frac{1}{f_0(0)} + O\left(\frac{k\alpha_j^2}{n}\right)\\
&= -\frac{8\pi^2 n}{2k+1} + \frac{2\pi^2\|\underline{\alpha}\|_2^2}{n} + O\left(\frac{k\alpha_j^2}{n}\right) + O\left(\frac{k^2}{n^3}\|\underline{\alpha}\|_4^4\right).
\end{aligned}$$

The triple derivatives are estimated by Taylor expanding $e(\theta)$ to degree 1 in the numerator, using $R_j - \frac{1}{R_j} \ll \frac{k\alpha_j}{n}$ and $R_j + \frac{1}{R_j}, f_0(\underline{\theta}) \asymp 1$. □

**Lemma A.5.** *Let $n, k(n) \in \mathbb{Z}_{>0}$ with $k^2 = o(n)$ as $n \to \infty$. Let $\underline{\alpha} \in \mathbb{Z}^k$ and assume $\|\underline{\alpha}\|_2^2 \leq n\left(1 + \frac{\log n}{\sqrt{k}}\right)$ and $\|\underline{\alpha}\|_4^4 \ll \frac{n^2}{k}\left(1 + \frac{\log n}{\sqrt{k}}\right)$. Let $\underline{\theta} \in \mathscr{D}_{\mathrm{sm}}$. We have*

$$\begin{aligned}
F(\underline{\theta}) - F(0) = \frac{-4\pi^2 n}{2k+1}\|\underline{\theta}\|_2^2 &+ O\left(\left(1 + \frac{\log n}{\sqrt{k}}\right)\|\underline{\theta}\|_2^2 + \sqrt{k}\left(1 + \frac{\log n}{\sqrt{k}}\right)\|\underline{\theta}\|_4^2\right)\\
&+ O\left(\frac{n}{k^3}\|\underline{\theta}\|_2^6 + \frac{n}{k^2}\|\underline{\theta}\|_2^4 + \frac{\sqrt{n}}{k}\left(1 + \frac{\log n}{\sqrt{k}}\right)\|\underline{\theta}\|_2^3\right)\\
&+ O\left(\frac{n}{k}\|\underline{\theta}\|_4^4 + \frac{\sqrt{n}}{k^{\frac{1}{4}}}\left(1 + \frac{\log n}{\sqrt{k}}\right)\|\underline{\theta}\|_4^3\right).
\end{aligned}$$

*In particular, for any fixed constants $c_2, c_4 > 0$, for $\|\underline{\theta}\|_2 \le c_2 \frac{k}{n^{\frac{1}{2}}}$ and $\|\underline{\theta}\|_4 \le c_4 \frac{k^{\frac{3}{4}}}{n^{\frac{1}{2}}}$ we have*

$$F(\underline{\theta}) - F(0) + \frac{4\pi^2 n}{2k+1}\|\underline{\theta}\|_2^2 = o(1), \tag{A.9}$$

*while for $\|\underline{\theta}\|_2 \le c_2 \frac{k}{n^{\frac{1}{2}}}$ and $\|\underline{\theta}\|_4 = o(1)$,*

$$F(\underline{\theta}) - F(0) + \frac{4\pi^2 n}{2k+1}\|\underline{\theta}\|_2^2 \ll o(1) + \frac{n}{k}\|\underline{\theta}\|_4^4 + \left(1 + \frac{\log n}{\sqrt{k}}\right)\left(\sqrt{k}\|\underline{\theta}\|_4^2 + \frac{\sqrt{n}}{k^{\frac{1}{4}}}\|\underline{\theta}\|_4^3\right), \tag{A.10}$$

*and in general for $\|\underline{\theta}\|_\infty < \delta < \frac{1}{12}$,*

$$F(\underline{\theta}) - F(0) + \frac{4\pi^2 n}{2k+1}\|\underline{\theta}\|_2^2 \ll \delta^2 \frac{n}{k}\|\underline{\theta}\|_2^2 + \left(1 + \frac{\log n}{\sqrt{k}}\right)\left(\sqrt{\frac{n}{k}}\|\underline{\theta}\|_2 + \frac{\sqrt{n}}{k^{\frac{1}{4}}}\|\underline{\theta}\|_2^{\frac{3}{2}}\right). \tag{A.11}$$

*Proof.* By Taylor's theorem, for $\underline{\theta} \in \mathscr{D}_{\mathrm{sm}}$, for some $0 \le t_{\underline{\theta}} \le 1$,

$$F(\underline{\theta}) - F(0) = \frac{1}{2}\mathscr{D}^2(0)(\underline{\theta}, \underline{\theta}) + \frac{1}{6}\mathscr{D}^3(t_{\underline{\theta}}\underline{\theta})(\underline{\theta}, \underline{\theta}, \underline{\theta})$$

where $\mathscr{D}^2$ and $\mathscr{D}^3$ represent the second and third derivatives of $F$. Write

$$\mathscr{D}^2(0) = \frac{-8\pi^2 n}{2k+1} I_k + \tilde{\mathscr{D}}^2(0)$$

We have

$$\begin{aligned}\tilde{\mathscr{D}}^2(0)(\underline{\theta}, \underline{\theta}) &\ll \frac{\|\underline{\alpha}\|_2^2\|\underline{\theta}\|_2^2}{n} + \frac{k\|\underline{\alpha}\|_4^2\|\underline{\theta}\|_4^2}{n} + \frac{k^2}{n^3}\|\underline{\alpha}\|_4^4\|\underline{\theta}\|_2^2 \\ &\ll \left(1 + \frac{\log n}{\sqrt{k}}\right)\|\underline{\theta}\|_2^2 + \sqrt{k}\left(1 + \frac{\log n}{\sqrt{k}}\right)\|\underline{\theta}\|_4^2\end{aligned}$$

Also,

$$\begin{aligned}\left|\mathscr{D}^3(t_{\underline{\theta}}\underline{\theta})(\underline{\theta}, \underline{\theta}, \underline{\theta})\right| &\ll \frac{n}{k^3}\left(\|\underline{\theta}\|_2^6 + \frac{k^3}{n^3}\|\underline{\theta}\|_2^3\|\underline{\alpha}\|_2^3\right) \\ &\quad + \frac{n}{k^2}\left(\|\underline{\theta}\|_2^4 + \frac{k}{n}\|\underline{\theta}\|_2^3\|\underline{\alpha}\|_2\right) \\ &\quad + \frac{n}{k}\left(\|\underline{\theta}\|_4^4 + \frac{k}{n}\|\underline{\theta}\|_4^3\|\underline{\alpha}\|_4\right) \\ &\ll \frac{n}{k^3}\|\underline{\theta}\|_2^6 + \frac{n}{k^2}\|\underline{\theta}\|_2^4 + \frac{\sqrt{n}}{k}\left(1 + \frac{\log n}{\sqrt{k}}\right)\|\underline{\theta}\|_2^3 \\ &\quad + \frac{n}{k}\|\underline{\theta}\|_4^4 + \frac{\sqrt{n}}{k^{\frac{1}{4}}}\left(1 + \frac{\log n}{\sqrt{k}}\right)\|\underline{\theta}\|_4^3.\end{aligned}$$

For (A.11) use $\|\underline{\theta}\|_4 \le \delta^{\frac{1}{2}}\|\underline{\theta}\|_2^{\frac{1}{2}}$, and $\|\underline{\theta}\|_2 \le \delta\sqrt{k}$. □

**Lemma A.6.** *Keep the same assumptions on $k, n$ and $\underline{\alpha}$ as in Lemma A.5. We have*

$$|\mathrm{Im}\, f_0(\underline{\theta})| \ll \frac{\|\underline{\alpha}\|_2\|\underline{\theta}\|_2}{n}.$$

*Moreover, there is a constant $c > 0$ such that, if $\mathrm{Re}(f_0(\underline{\theta})) > 0$ then*

$$\left|\frac{f_0(\underline{\theta})}{f_0(0)}\right| \le 1 - \frac{c}{k}\|\underline{\theta}\|_{(\mathbb{R}/\mathbb{Z})^k}^2,$$

*and if* $\mathrm{Re}(f_0(\underline{\theta})) < 0$ *then*

$$\left|\frac{f_0(\underline{\theta})}{f_0(0)}\right| \leq 1 - \frac{c}{k} - \frac{c}{k}\left\|\underline{\theta} - \left(\frac{1}{2}\right)_k\right\|^2_{(\mathbb{R}/\mathbb{Z})^k},$$

*where* $\left(\frac{1}{2}\right)_k$ *denotes the vector of* $\mathbb{R}^k$*, all of whose coordinates are* $\frac{1}{2}$*.*

*Proof.* We have

$$\begin{aligned}|\mathrm{Im}\, f_0(\underline{\theta})| &= \left|\frac{1}{2k+1}\sum_{j=1}^{k}\left(R_j - \frac{1}{R_j}\right)\sin(2\pi\theta_j)\right| \\ &= \frac{f_0(0)}{n}\left|\sum_{j=1}^{k}\alpha_j \sin(2\pi\theta_j)\right| \\ &\ll \frac{\|\underline{\alpha}\|_2\|\underline{\theta}\|_2}{n}.\end{aligned}$$

If $\mathrm{Re}(f_0(\underline{\theta})) > 0$ then

$$\begin{aligned}\left|\frac{\mathrm{Re}\, f_0(\underline{\theta})}{f_0(0)}\right| &= 1 - \frac{1}{f_0(0)(2k+1)}\sum_{j=1}^{k}\left(R_j + \frac{1}{R_j}\right)(1-\cos(2\pi\theta_j)) \\ &\leq 1 - \frac{c}{k}\|\underline{\theta}\|_2^2.\end{aligned}$$

If, instead, $\mathrm{Re}(f_0(\underline{\theta})) < 0$ then

$$\begin{aligned}\left|\frac{\mathrm{Re}\, f_0(\underline{\theta})}{f_0(0)}\right| &= 1 - \frac{1}{(2k+1)f_0(0)} - \frac{1}{(2k+1)f_0(0)}\sum_{j=1}^{k}\left(R_j + \frac{1}{R_j}\right)(1+\cos(2\pi\theta_j)) \\ &\geq 1 - \frac{c}{k} - \frac{c}{k}\left\|\underline{\theta} - \left(\frac{1}{2}\right)_k\right\|^2_{(\mathbb{R}/\mathbb{Z})^k}.\end{aligned}$$

The bound for $\left|\frac{f_0(\underline{\theta})}{f_0(0)}\right|$ in the case $\mathrm{Re}(f_0(\underline{\theta})) > 0$ follows from, for some $c' > 0$,

$$\left|\frac{f_0(\underline{\theta})}{f_0(0)}\right|^2 \leq 1 - \frac{c'}{k}\|\underline{\theta}\|_2^2 + O\left(\left(1 + \frac{\log n}{\sqrt{k}}\right)\frac{\|\underline{\theta}\|_2^2}{n}\right),$$

and the claim in the case $\mathrm{Re}(f(\underline{\theta})) < 0$ is similar. □

We give our final estimate.

*Proof of Lemma A.1.* Let $0 < \delta < \frac{1}{12}$ be a constant to be chosen.

$$\begin{aligned}\nu_k^{*n}(\underline{\alpha}) &= \int_{(\mathbb{R}/\mathbb{Z})^k}\frac{f_0(\underline{\theta})^n}{R_1^{\alpha_1}\ldots R_k^{\alpha_k}}d\underline{\theta} \\ &= \frac{f_0(0)^n}{R_1^{\alpha_1}\ldots R_k^{\alpha_k}}\left[\int_{\|\underline{\theta}\|_\infty \leq \delta} e^{F(\underline{\theta})-F(0)}d\underline{\theta} + \int_{\|\underline{\theta}\|_\infty > \delta}\left(\frac{|f_0(\underline{\theta})|}{f_0(0)}\right)^n d\underline{\theta}\right].\end{aligned}$$

By Lemma A.4,

$$\frac{f_0(0)^n}{R_1^{\alpha_1}\ldots R_k^{\alpha_k}} = e^{F(0)} = e^{-\frac{2k+1}{4n}\|\underline{\alpha}\|_2^2}\left[1 + O\left(\frac{k^3}{n^3}\|\underline{\alpha}\|_4^4\right)\right] \sim e^{-\frac{2k+1}{4n}\|\underline{\alpha}\|_2^2} \tag{A.12}$$

since $\|\underline{\alpha}\|_4^4 \ll \frac{n^2}{k}\left(1 + \frac{\log n}{\sqrt{k}}\right)$.

To treat the integral over $\|\underline{\theta}\|_\infty < \delta$, write

$$\int_{\|\underline{\theta}\|_\infty \le \delta} e^{F(\underline{\theta})-F(0)} d\underline{\theta} = \left(\frac{2k+1}{4\pi n}\right)^{\frac{k}{2}} \int_{\|\underline{\theta}\|_\infty \le \delta} \eta_k\left(\sqrt{\frac{2k+1}{8\pi^2 n}}, \underline{\theta}\right) \exp\left(G(\underline{\theta})\right) d\underline{\theta}.$$

Partition $B_\infty(0,\delta)$ by choosing for some parameters $c_2, c_4$,

$$\begin{aligned} B_\infty(0,\delta) &= B \sqcup E_1 \sqcup E_2 \\ B &= B_\infty(0,\delta) \cap B_2\left(0, c_2\frac{k}{\sqrt{n}}\right) \cap B_4\left(0, c_4\frac{k^{\frac{2}{3}}}{n^{\frac{1}{2}}}\right) \\ E_1 &= B_\infty(0,\delta) \cap B_2\left(0, c_2\frac{k}{\sqrt{n}}\right) \setminus B_4\left(0, c_4\frac{k^{\frac{2}{3}}}{n^{\frac{1}{2}}}\right) \\ E_2 &= B_\infty(0,\delta) \setminus B_2\left(0, c_2\frac{k}{\sqrt{n}}\right). \end{aligned}$$

The parameters $c_2, c_4$ are considered fixed, but may be arbitrarily large.

On $B$, (A.9) gives $G(\underline{\theta}) = o(1)$ and we find

$$\begin{aligned} \int_{\underline{\theta}\in B} \eta_k\left(\sqrt{\frac{2k+1}{8\pi^2 n}}, \underline{\theta}\right) \exp\left(G(\underline{\theta})\right) d\underline{\theta} &= (1+o(1)) \int_{\underline{\theta}\in B} \eta_k\left(\sqrt{\frac{2k+1}{8\pi^2 n}}, \underline{\theta}\right) d\underline{\theta} \\ &= 1 + \varepsilon(c_2, c_4), \end{aligned}$$

where $\varepsilon(c_2,c_4) \to 0$ as $\min(c_2,c_4) \to \infty$, as follows by Lemma 2.2 ($c_2$ and $c_4$ only need be taken growing if $k$ does not grow).

In treating $E_1$ and $E_2$, let $C_2, C_4$ be the constants of Lemma 2.2. To treat $E_1$, note that with respect to the Gaussian measure $\gamma = \eta_k\left(\sqrt{\frac{2k+1}{8\pi^2 n}}, \underline{\theta}\right)$, the event $\underline{\theta} \in B_\infty(0,\delta) \cap B_2\left(0, c_2\frac{k}{\sqrt{n}}\right)$ has probability $\asymp 1$, and thus, even after conditioning on this event, the probability of $\|\underline{\theta}\|_4 > C_4 k^{\frac{1}{4}}\sqrt{\frac{2k+1}{8\pi^2 n}} + t$ is, for some $C > 0$, $O\left(\exp\left(-\frac{n}{k}\frac{t^2}{C}\right)\right)$. The bound $\|\underline{\theta}\|_4 \le \delta^{\frac{1}{2}}\|\underline{\theta}\|_2^{\frac{1}{2}}$ implies that on $E_1$, $\|\underline{\theta}\|_4 = o(1)$. Set $t = \|\underline{\theta}\|_4 - C_4 k^{\frac{1}{4}}\sqrt{\frac{2k+1}{8\pi^2 n}}$ and assume $c_4$ is larger than a sufficiently large multiple of $C_4$, so that $t \gg \frac{k^{\frac{3}{4}}}{\sqrt{n}}$. By (A.10) we find that for $\underline{\theta} \in E_1$,

$$\begin{aligned} &G(\underline{\theta}) \le g(t) \\ &g(t) \ll o(1) + \frac{n}{k}t^4 + \left(1 + \frac{\log n}{\sqrt{k}}\right)\left(\sqrt{\frac{n}{k}}t^2 + \frac{\sqrt{n}}{k^{\frac{1}{4}}}t^3\right). \end{aligned}$$

Then

$$\begin{aligned} &\int_{E_1} \eta_k\left(\sqrt{\frac{2k+1}{8\pi^2 n}}, \underline{\theta}\right) \exp(G(\underline{\theta})) d\underline{\theta} \\ &\le -\int_{\frac{k^{\frac{3}{4}}}{\sqrt{n}} \ll t = o(1)} \exp\left(g(t)\right) d\,\mathrm{meas}\left(\|\underline{\theta}\|_4 \ge C_4 k^{\frac{1}{4}}\sqrt{\frac{2k+1}{8\pi^2 n}} + t\right). \end{aligned}$$

Integrating by parts, we find that this integral is $o(1)$ as $c_4 \to \infty$.

To treat $E_2$, set $s = \|\underline{\theta}\|_2$ and appeal to (A.11) to find

$$\begin{aligned} &G(\underline{\theta}) \le h(s) \\ &h(s) \ll o(1) + \delta^2\frac{n}{k}s^2 + \left(1 + \frac{\log n}{\sqrt{k}}\right)\left(\sqrt{\frac{n}{k}}s + \frac{\sqrt{n}}{k^{\frac{1}{4}}}s^{\frac{3}{2}}\right). \end{aligned}$$

Also, for some $C > 0$,

$$\mathrm{meas}\left(\|\underline{\theta}\|_2 > C_2 k^{\frac{1}{2}}\sqrt{\frac{2k+1}{8\pi^2 n}} + s\right) \ll \exp\left(-\frac{n}{k}\frac{s^2}{C}\right).$$

We conclude

$$\begin{aligned}&\int_{\underline{\theta}\in E_2} \eta_k\left(\sqrt{\frac{2k+1}{8\pi^2 n}}, \underline{\theta}\right)\exp(G(\underline{\theta}))d\underline{\theta}\\ &\leq -\int_{\frac{k}{\sqrt{n}}\ll s\leq \delta\sqrt{k}} \exp(h(s))d\,\mathrm{meas}\left(\|\underline{\theta}\|_2 > \sqrt{\frac{2k+1}{8\pi^2 n}} + s\right).\end{aligned}$$

If $\delta$ is sufficiently small, this integral is in fact $o(1)$ as $c_2 \to \infty$, as may be checked again by integration by parts.

It remains to bound the integral over $\|\underline{\theta}\|_\infty > \delta$. Consider first the case $\mathrm{Re}(f(\underline{\theta})) > 0$. Let $S \subset [k]$ be the collection of $\theta_j$ with $|\theta_j| > \delta$. Write $\underline{\theta}_S$ for the variables in $S$ and $\underline{\theta}_{S^c}$ for the variables in $S^c$. Appealing to Lemma A.6, we see that if $|S| \gg \log n$ then the integral is negligible. Using $1 - x < e^{-x}$ in the remaining range we obtain a bound, for some fixed $c > 0$,

$$\begin{aligned}&\ll \sum_{1\leq j\ll \log n}\sum_{S\subset[k],|S|=j} \exp\left(-\frac{cjn}{k}\right)\int_{\|\underline{\theta}_{S^c}\|_\infty<\delta}\exp\left(-\frac{cn}{k}\|\underline{\theta}_{S^c}\|_2^2\right)\\ &\ll \left(\frac{2k+1}{4\pi nc}\right)^{\frac{k}{2}}\sum_{1\leq j\ll\log n}\binom{k}{j}\left(\frac{4\pi nc}{2k+1}\right)^{\frac{j}{2}}\exp\left(-\frac{cjn}{k}\right)\\ &= o\left(\left(\frac{2k+1}{4\pi n}\right)^{\frac{k}{2}}\right).\end{aligned}$$

The terms for which $\mathrm{Re}(f(\underline{\theta})) < 0$ are handled similarly. □

## References


[1] David Aldous and Persi Diaconis, *Shuffling cards and stopping times*, Amer. Math. Monthly **93** (1986), no. 5, 333–348. MR-841111

[2] Michael Bate and Stephen Connor, *Cutoff for a random walk on the integers mod $n$*, 2014.

[3] Henry Cohn and Yufei Zhao, *Sphere packing bounds via spherical codes*, Duke Math. J. **163** (2014), no. 10, 1965–2002. MR-3229046

[4] Amir Dembo, Jian Ding, Jason Miller, and Yuval Peres, *Cut-off for lamplighter chains on tori: dimension interpolation and phase transition*, 2013.

[5] P. Diaconis and L. Saloff-Coste, *Moderate growth and random walk on finite groups*, Geom. Funct. Anal. **4** (1994), no. 1, 1–36. MR-1254308

[6] Persi Diaconis, *Group representations in probability and statistics*, Lecture Notes-Monograph Series **11** (1988), i–192. MR-0964069

[7] Persi Diaconis, *Threads through group theory*, Character theory of finite groups **524** (2010), 33–47. MR-2731916

[8] Persi Diaconis, R. L. Graham, and J. A. Morrison, *Asymptotic analysis of a random walk on a hypercube with many dimensions*, Random Structures Algorithms **1** (1990), no. 1, 51–72. MR-1068491

[9] Persi Diaconis and Robert Hough, *Random walk on unipotent matrix groups*, arXiv:1512.06304 (2015).

[10] Persi Diaconis and Laurent Saloff-Coste, *Separation cut-offs for birth and death chains*, Ann. Appl. Probab. **16** (2006), no. 4, 2098–2122. MR-2288715

[11] Persi Diaconis and Mehrdad Shahshahani, *Time to reach stationarity in the Bernoulli–Laplace diffusion model*, SIAM Journal on Mathematical Analysis **18** (1987), no. 1, 208–218. MR-0871832

[12] Jian Ding, Eyal Lubetzky, and Yuval Peres, *Total variation cutoff in birth-and-death chains*, Probab. Theory Related Fields **146** (2010), no. 1-2, 61–85. MR-2550359

[13] Carl Dou and Martin Hildebrand, *Enumeration and random random walks on finite groups*, Ann. Probab. **24** (1996), no. 2, 987–1000. MR-1404540

[14] Ben Green and Terence Tao, *The quantitative behaviour of polynomial orbits on nilmanifolds*, Annals of Mathematics **175** (2012), no. 2, 465–540. MR-2877065

[15] Andrew Simon Greenhalgh, *Random walks on groups with subgroup invariance properties*, ProQuest LLC, Ann Arbor, MI, 1989, Thesis (Ph.D.)–Stanford University. MR-2637557

[16] Martin Hildebrand, *Random walks supported on random points of* $\mathbf{Z}/n\mathbf{Z}$, Probab. Theory Related Fields **100** (1994), no. 2, 191–203. MR-1296428

[17] Daniel Jerison, Lionel Levine, and John Pike, *Mixing time and eigenvalues of the abelian sandpile Markov chain*, arXiv:1511.00666 (2015).

[18] G. A. Kabatjanskiĭ and V. I. Levenšteĭn, *Bounds for packings on the sphere and in space*, Problemy Peredači Informacii **14** (1978), no. 1, 3–25. MR-0514023

[19] Michel Ledoux and Michel Talagrand, *Probability in Banach spaces*, Classics in Mathematics, Springer-Verlag, Berlin, 2011, Isoperimetry and processes, Reprint of the 1991 edition. MR-2814399

[20] DA Levin, Yuval Peres, and Elizabeth Wilmer, *Markov chains and mixing times*, American Mathematical Society, 2009. MR-2466937

[21] Eyal Lubetzky and Yuval Peres, *Cutoff on all Ramanujan graphs*, arXiv:1507.04725 (2015). MR-3558308

[22] C. A. Rogers, *Mean values over the space of lattices*, Acta Math. **94** (1955), 249–287. MR-0075243

[23] Carl Ludwig Siegel, *A mean value theorem in geometry of numbers*, Ann. of Math. (2) **46** (1945), 340–347. MR-0012093

[24] Carl Ludwig Siegel, *Lectures on the geometry of numbers*, Springer Science & Business Media, 2013. MR-1020761

[25] Anders Södergren, *On the distribution of angles between the N shortest vectors in a random lattice*, Journal of the London Mathematical Society (2011), jdr032. MR-2855800

[26] Andrey Sokolov, Andrew Melatos, Tien Kieu, and Rachel Webster, *Memory on multiple time-scales in an abelian sandpile*, Physica A: Statistical Mechanics and its Applications **428** (2015), 295–301.

[27] Bálint Virág, *Random walks on finite convex sets of lattice points*, J. Theoret. Probab. **11** (1998), no. 4, 935–951. MR-1660887

[28] David Bruce Wilson, *Random random walks on* $\mathbf{Z}_2^d$, Probab. Theory Related Fields **108** (1997), no. 4, 441–457. MR-1465637

**Acknowledgments.** The author thanks Persi Diaconis, Yuval Peres, David Wilson, Lionel Levine, Elon Lindenstrauss and Daniel Jerison for useful conversations.

# 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`