Asymptotic Enumeration of Difference Matrices over Cyclic Groups

AARON M. MONTGOMERY

Department of Mathematics, Baldwin Wallace University, Berea, Ohio, 44017, USA
(e-mail: `amontgom@bw.edu`)

We identify a relationship between a certain family of random walks on Euclidean lattices and difference matrices over cyclic groups. We then use the techniques of Fourier analysis to estimate the return probabilities of these random walks, which in turn yields the asymptotic number of difference matrices over cyclic groups as the number of columns increases.

1. Introduction

This paper will explore a connection between random walks and a certain class of combinatorial designs known as difference matrices. Difference matrices have long been a part of the combinatorial design literature, and they are related to many other types of designs such as orthogonal arrays, transversal designs, pairwise-balanced designs, and more. For instance, a difference matrix over $\mathbb{Z}_2$ is also a partial Hadamard matrix (see, for instance, [5]). A comprehensive overview of the existing literature on difference matrices and their relationships with other types of designs can be found in [4].

The main result of this work will be to provide an asymptotic count of the difference matrices over cyclic groups as the number of their columns increases. Enumerating combinatorial designs directly is often quite challenging due to the computational complexity involved, and difference matrices are no exception. We will instead relate these matrices to a certain family of random walks on Euclidean lattices, and we will use the tools of Fourier analysis to estimate the return probability of these random walks. We will then exploit the connection between the two problems to obtain the desired result about difference matrices. One advantage of this approach is that it permits an asymptotic enumeration of these matrices without ever requiring the explicit construction of any such matrix, which can be computationally difficult when the parameters are large.

Definition 1. Let $(G, \odot)$ be a group of order g. A $(g, k; \lambda)$-*difference matrix over G* is a $k \times g\lambda$ matrix $D = [d_{ij}]$ with entries from G, so that for each $1 \leqslant i < j \leqslant k$, the multiset

$$\left\{ d_{i\ell} \odot d_{j\ell}^{-1} : 1 \leqslant \ell \leqslant g\lambda \right\}$$

contains every element of G exactly λ times.

In order to simplify the Fourier analysis, in this work we will consider only cases where $G = \mathbb{Z}_g$. Accordingly, we will prefer to use notation such as $d_{i\ell} - d_{j\ell}$ in place of $d_{i\ell} \odot d_{j\ell}^{-1}$. The goal of this paper is to prove the following theorem.

Theorem 1.1. *Let $g \geqslant 2$ be fixed, and suppose (k, λ) is a sequence of ordered pairs such that $k \geqslant 3$, $\lambda \to \infty$, and if g is even then each λ is also even. Suppose also that there exists some $\varepsilon_0 > 0$ so that*

$$k < \left(\frac{1}{6} - \varepsilon_0 \right) \frac{\log(\lambda g)}{\log(g)}$$

for all pairs in the sequence. Then along the sequence (k, λ) the number of $(g, k; \lambda)$-difference matrices over $\mathbb{Z}_g$ is

$$\frac{g^{k\lambda g + (3k-4)(k-1)/4}}{(2\pi\lambda)^{\binom{k}{2}(g-1)/2}} [1 + o(1)].$$

We remark here that the prohibition on the existence of a $(g, k; \lambda)$-difference matrix over $\mathbb{Z}_g$ with $k > 2$, g even, and λ odd is due to Drake [6], although the Fourier analysis involved in the proof of Theorem 1.1 will suggest this as well. We also note that Jungnickel showed that the existence of a $(g, k; \lambda)$-difference matrix over $\mathbb{Z}_g$ requires that $k \leqslant \lambda g$ [12]. Consequently, an ideal version of Theorem 1.1 would permit k to grow as a near-linear function of λ, whereas this version permits k only to grow as a logarithmic function of λ. We do not claim any technical or number-theoretic reason for this restriction; rather, it is made as a matter of practicality to complete the Fourier analysis, and it could almost surely be improved.

To prove Theorem 1.1, we will consider a randomly generated $k \times g\lambda$ matrix with entries chosen uniformly and independently from $\mathbb{Z}_g$; we will then compute the probability that such a matrix satisfies Definition 1. The columns of the matrix will correspond to the steps of a random walk on a certain high-dimensional Euclidean lattice, and the existence of a $(g, k; \lambda)$-difference matrix over $\mathbb{Z}_g$ will correspond to a return path of the random walk to the origin. We will prove a local central limit theorem to provide estimates on the return probability of the random walk, which will in turn yield estimates on the numbers of these matrices.

We pause to remark that this type of analysis is certainly not new to the study of combinatorial designs. De Launey and Levin used this tactic to study partial Hadamard matrices [5], and their enumeration results can be recognized as the particular case of Theorem 1.1 where $g = 2$, although their work shows the formula to be valid so long as the pairs (k, λ) satisfy the more generous condition $k \leqslant (2\lambda)^{1/(12+\varepsilon_0)}$. Additionally, the author of this work has used this strategy to count balanced incomplete block design incidence matrices [15, Theorem 2.3]. Kuperberg, Lovett and Peled used a slightly different random walk approach to enumerate simple

orthogonal arrays and simple t-designs [13]. Many other works exist which perform asymptotic enumerations of combinatorial structures by finding suitable estimates on complex integrals; see [1, 3, 8, 11, 14] for a few of the numerous examples.

We now define the random walk and identify its correspondence to difference matrices. With $(\mathbb{Z}_g)^k$ regarded as a column vector with entries in $\mathbb{Z}_g$, we will define a map $Z\colon (\mathbb{Z}_g)^k \to \mathbb{R}^{\binom{k}{2}\cdot(g-1)}$. The Euclidean space is to be regarded as indexed by two coordinates, where the first coordinate is an unordered pair of rows (*i.e.* $\{i,j\}$ with $1 \leqslant i < j \leqslant k$) and the second coordinate is an element of $\mathbb{Z}_g$ besides 0. The ordering of the indices will be assumed to be lexicographic. Let $\vec{x} = (x_1, \ldots, x_k)^T$ be an element of $(\mathbb{Z}_g)^k$; then the map Z is defined by

$$[Z(\vec{x})]_{\{i,j\},a} = \begin{cases} 1 - 1/g & \text{if } x_i - x_j = a, \\ -1/g & \text{if } x_i - x_j \neq a. \end{cases} \tag{1.1}$$

As an example, if $g = 3$ and $k = 4$, then the codomain $\mathbb{R}^{12}$ of the map Z is understood to be indexed in the order $(\{1,2\},1),(\{1,2\},2),(\{1,3\},1),\ldots,(\{3,4\},2)$; the vector $(2,1,0,2)^T$ would map to

$$\frac{1}{3} \cdot (2,-1,-1,2,-1,-1,2,-1,-1,2,2,-1)^T$$

under Z. We remark that in the index scheme, for each pair of rows $\{i,j\}$, the coordinate $(\{i,j\},0)$ is intentionally omitted. The purpose of the function Z is that if $D = [\vec{x}_1, \ldots, \vec{x}_t]$ is a $k \times t$ matrix with entries in $\mathbb{Z}_g$, then a given coordinate $\{i,j\},a$ of the expression $Z(\vec{x}_1) + \cdots + Z(\vec{x}_t)$ will be $\vec{0}$ if and only if t is a multiple of g and the multiset $\{x_{i\ell} - x_{j\ell} : 1 \leqslant \ell \leqslant t\}$ contains the element a exactly t/g times. Consequently, we have $Z(\vec{x}_1) + \cdots + Z(\vec{x}_t) = \vec{0}$ if and only if every element of $\{1, \ldots, g-1\} \subset \mathbb{Z}_g$ appears exactly t/g times in the multiset $\{x_{i\ell} - x_{j\ell} : 1 \leqslant \ell \leqslant t\}$ for every pair $\{i,j\}$; the latter condition also implies that 0 also appears exactly t/g times.

Definition 2. Let $\{X_t\}$ be the random walk defined with increments drawn uniformly and independently from $\{Z(\vec{x}) : \vec{x} \in (\mathbb{Z}_g)^k\}$.

If $D = [\vec{x}_1, \ldots, \vec{x}_t]$ is a $k \times t$ matrix with elements taken uniformly and independently from $\mathbb{Z}_g$, then the preceding remarks imply that D is a difference matrix if and only if $Z(\vec{x}_1) + \cdots + Z(\vec{x}_t) = \vec{0}$; that is, the corresponding random walk has returned to the origin. Since Z is invariant under the action of adding any element to each entry of $\vec{x}$, there is a g^t to 1 correspondence between $\mathrm{Mat}_{k \times t}(\mathbb{Z}_g)$ and paths of the random walk $\{X_t\}$. There is also a g^t to 1 correspondence between $(g,k;\lambda)$-difference matrices over $\mathbb{Z}_g$ and return paths of $\{X_t\}$ to $\vec{0}$, which implies the following relationship:

$$\frac{\#(g,k;\lambda)\text{-difference matrices over } \mathbb{Z}_g}{|\mathrm{Mat}_{k \times \lambda g}(\mathbb{Z}_g)|} = \frac{\#\text{ return paths of } X_t \text{ to } \vec{0}}{\#\text{ all paths of } X_t}.$$

The right side of that equation is merely $\mathbb{P}(X_t = \vec{0})$, and the denominator on the left is $g^{k\lambda g}$. Therefore, we have

$$\#(g,k;\lambda)\text{-difference matrices over } \mathbb{Z}_g = g^{k\lambda g}\mathbb{P}(X_t = \vec{0}). \tag{1.2}$$

Hence, in order to estimate the number of $(g, k; \lambda)$-differences matrices over $\mathbb{Z}_g$, we need only to estimate the return probability of $\{X_t\}$ to $\vec{0}$.

We will estimate $\mathbb{P}(X_t = \vec{0})$ by using the standard tactics of Fourier analysis. In all that follows, we let $d = \binom{k}{2} \cdot (g - 1)$. We define the characteristic function $\Phi : \mathbb{R}^d \to \mathbb{C}$ by

$$\Phi(\vec{\theta}) = \mathbb{E}[e^{i\vec{\theta} \cdot X_1}] = \sum_{\vec{x} \in (\mathbb{Z}_g)^k} g^{-k} e^{i\vec{\theta} \cdot Z(\vec{x})}.$$

One can verify that $|\Phi(\vec{\theta})|$ is 2π-periodic, and that if t is a multiple of g, then $\Phi(\vec{\theta})^t$ is also 2π-periodic. Consequently, if t is a multiple of g, then the walk X_t is supported on the integer lattice $\mathbb{Z}^d \subset \mathbb{R}^d$, which permits use of the Fourier inversion formula (see, for instance, [16, P3, p. 57]):

$$\mathbb{P}(X_t = \vec{0}) = (2\pi)^{-d} \int_{[-\pi, \pi)^d} \Phi(\vec{\theta})^t \, \mathrm{d}\vec{\theta}. \tag{1.3}$$

To estimate this integral, we will partition $[-\pi, \pi)^d$ based on the value of $|\Phi(\vec{\theta})|$ by dealing separately with regions where $|\Phi(\vec{\theta})|$ is close to 1 and those where it is not. Intuitively, as $t \to \infty$, the contributions to the integral from the former regions should become dominant, while those from the latter regions should become negligible.

It is noteworthy that, in principle, combining (1.2) with (1.3) provides an exact count of the number of $(g, k; \lambda)$-difference matrices over $\mathbb{Z}_g$. However, in practice this is difficult to exploit because the complicated nature of $\Phi(\vec{\theta})^t$ makes the integral intractably difficult. Consequently, it will be preferable to estimate $\Phi(\vec{\theta})^t$ instead of calculating it directly. We also note that a common approach to the general problem of estimating return probabilities of a random walk is to transform the walk to a strongly aperiodic random walk on an integer lattice. However, this tactic is difficult here because of the complicated structure of the increment set $\{Z(\vec{x}) : \vec{x} \in (\mathbb{Z}_g)^k\}$. These challenges motivate the Fourier-analytic approach used in this work.

The outline of the rest of this paper is as follows: in Section 2, we discuss how to decompose the integral in (1.3) into manageable pieces. Sections 3 and 4 are devoted to finding estimates on the integral where $|\Phi(\vec{\theta})|$ is and is not close to 1, respectively. Finally, in Section 5 we combine all the pieces to prove Theorem 1.1.

2. Anatomy of the integral

Let $\Lambda = \{\vec{\theta} \in \mathbb{R}^d : |\Phi(\vec{\theta})| = 1\}$. Our goal in this section is to characterize the set Λ, and then to use this characterization to partition the integral in (1.3) in a suitable way. We observe that

$$\vec{\theta} \in \Lambda \iff \text{for all } \vec{x}, \vec{y} \in (\mathbb{Z}_g)^k, \ e^{i\vec{\theta} \cdot Z(\vec{x})} = e^{i\vec{\theta} \cdot Z(\vec{y})},$$

or equivalently

$$\vec{\theta} \in \Lambda \iff \text{for all } \vec{x}, \vec{y} \in (\mathbb{Z}_g)^k, \ \vec{\theta} \cdot Z(\vec{x}) \equiv \vec{\theta} \cdot Z(\vec{y}) \pmod{2\pi}. \tag{2.1}$$

This expression shows that Λ is closed under addition and under negation, so Λ is a subgroup of the additive group $\mathbb{R}^d$. Henceforth, any reference to vectors in $\mathbb{R}^d$ being equivalent modulo 2π should be understood to mean that their respective components are equivalent to one another modulo 2π. We remark that Φ^t is 2π-periodic in this sense if t is a multiple of g.

In order to characterize Λ, we first observe a useful fact.

Lemma 2.1. *Let $a, b \in \mathbb{Z}_g$ and let $1 \leqslant i < j \leqslant k$. Suppose that $\vec{\theta} \in \mathbb{R}^d$ and $\varepsilon > 0$ have the property that for any pair of vectors $\vec{x}, \vec{y} \in (\mathbb{Z}_g)^k$, there are $z \in \mathbb{Z}$ and ε_1 with $|\varepsilon_1| < \varepsilon$ such that $\vec{\theta} \cdot Z(\vec{x}) - \vec{\theta} \cdot Z(\vec{y}) = 2\pi z + \varepsilon_1$. Then there exists ε_2 with $|\varepsilon_2| < 2\varepsilon$ such that*

$$\theta_{\{i,j\},a} + \theta_{\{i,j\},b} \equiv \theta_{\{i,j\},a+b} + \varepsilon_2 \pmod{2\pi},$$

where, if necessary, the undefined value $\theta_{\{i,j\},0}$ is understood to be 0.

The interpretation of this lemma is that if $\vec{\theta} \cdot Z(\vec{x})$ is nearly independent (modulo 2π) of $\vec{x}$, then the coordinates of $\vec{\theta}$ nearly satisfy a certain group homomorphism property. This lemma is quite technical, and we delay its proof to examine several useful corollaries. The first corollary is also quite technical and will be used in Section 3.

Corollary 2.2. *Suppose $\vec{\theta} \in \mathbb{R}^d$ and $\varepsilon > 0$ have the property that for any pair of vectors $\vec{x}, \vec{y} \in (\mathbb{Z}_g)^k$, there are $z \in \mathbb{Z}$ and ε_1 with $|\varepsilon_1| < \varepsilon$ such that $\vec{\theta} \cdot Z(\vec{x}) - \vec{\theta} \cdot Z(\vec{y}) = 2\pi z + \varepsilon_1$. Then, for any $\{i, j\}, a$, there are $z \in \mathbb{Z}$ and ε_3 with $|\varepsilon_3| < 2\varepsilon$ such that*

$$\theta_{\{i,j\},a} - \varepsilon_3 = \frac{2\pi}{g} z.$$

Proof. If $\vec{\theta}$ and ε satisfy the hypotheses of Lemma 2.1, then since $ga = 0$ it follows inductively that $g\theta_{\{i,j\},a} = 2\pi z + \varepsilon_g$, where $z \in \mathbb{Z}$ and $|\varepsilon_g| < 2(g-1)\varepsilon$ by the triangle inequality. If we set $\varepsilon_3 = \varepsilon_g/g$, then

$$\theta_{\{i,j\},a} = \frac{2\pi}{g} z + \varepsilon_3,$$

where

$$|\varepsilon_3| < 2\frac{g-1}{g}\varepsilon < 2\varepsilon. \qquad \square$$

The second corollary can be obtained by letting $\varepsilon \to 0$ in Lemma 2.1.

Corollary 2.3. *Let $a, b \in \mathbb{Z}_g$ and let $1 \leqslant i < j \leqslant k$. If $\vec{\theta} \in \Lambda$, then*

$$\theta_{\{i,j\},a} + \theta_{\{i,j\},b} \equiv \theta_{\{i,j\},a+b} \pmod{2\pi}$$

where, if necessary, the undefined value $\theta_{\{i,j\},0}$ is understood to be 0.

The third corollary is an immediate consequence of Corollary 2.3.

Corollary 2.4. *If $\vec{\theta} \in \Lambda$, then for every $\{i, j\}, a$, we have $\theta_{\{i,j\},a} \equiv 0 \pmod{2\pi/v}$, where v is the order of a. In particular, for all non-zero elements a, it holds that $\theta_{\{i,j\},a} \equiv 0 \pmod{2\pi/g}$.*

Proof of Lemma 2.1. Fix elements $a, b \in \mathbb{Z}_g$ and integers i, j with $1 \leqslant i < j \leqslant k$. If either a or b is 0, then the result is trivial; we assume first that a, b, and $a + b$ are all non-zero. We define four vectors in $(\mathbb{Z}_g)^k$:

$$\vec{x}_1 = (0, \ldots, 0, a, 0, \ldots, 0, 0, 0, \ldots, 0)^T,$$
$$\vec{x}_2 = (0, \ldots, 0, 0, 0, \ldots, 0, -b, 0, \ldots, 0)^T,$$
$$\vec{x}_3 = (0, \ldots, 0, a, 0, \ldots, 0, -b, 0, \ldots, 0)^T,$$
$$\vec{x}_4 = (0, \ldots, 0, 0, 0, \ldots, 0, 0, 0, \ldots, 0)^T.$$

The a in $\vec{x}_1$ and $\vec{x}_3$ occurs in the ith position, and the $-b$ in $\vec{x}_2$ and $\vec{x}_3$ occurs in the jth position. Let $\vec{\theta} \in \mathbb{R}^d$. We will use $\vec{1}$ to represent the vector in $\mathbb{R}^d$ of all ones. The following calculations are straightforward from Definition 1.1:

$$\vec{\theta} \cdot Z(\vec{x}_1) = \sum_{m<i} \theta_{\{m,i\},-a} + \sum_{m>i} \theta_{\{i,m\},a} - \frac{1}{g} \vec{\theta} \cdot \vec{1},$$

$$\vec{\theta} \cdot Z(\vec{x}_2) = \sum_{m<j} \theta_{\{m,j\},b} + \sum_{m>j} \theta_{\{j,m\},-b} - \frac{1}{g} \vec{\theta} \cdot \vec{1},$$

$$\vec{\theta} \cdot Z(\vec{x}_3) = \sum_{m<i} \theta_{\{m,i\},-a} + \sum_{\substack{m>i \\ m \neq j}} \theta_{\{i,m\},a} + \sum_{\substack{m<j \\ m \neq i}} \theta_{\{m,j\},b} + \sum_{m>j} \theta_{\{j,m\},-b} + \theta_{\{i,j\},a+b} - \frac{1}{g} \vec{\theta} \cdot \vec{1},$$

$$\vec{\theta} \cdot Z(\vec{x}_4) = -\frac{1}{g} \vec{\theta} \cdot \vec{1}.$$

By assumption, there are integers z, z' and error terms $\varepsilon_1, \varepsilon_1'$ in $(-\varepsilon, \varepsilon)$ such that

$$\vec{\theta} \cdot Z(\vec{x}_1) - \vec{\theta} \cdot Z(\vec{x}_3) = 2\pi z + \varepsilon_1 \quad \text{and} \quad \vec{\theta} \cdot Z(\vec{x}_2) - \vec{\theta} \cdot Z(\vec{x}_4) = 2\pi z' + \varepsilon_1'.$$

If $z'' = z + z'$ and $\varepsilon_2 = \varepsilon_1 + \varepsilon_1'$, then

$$\vec{\theta} \cdot Z(\vec{x}_1) + \vec{\theta} \cdot Z(\vec{x}_2) - \vec{\theta} \cdot Z(\vec{x}_3) - \vec{\theta} \cdot Z(\vec{x}_4) = 2\pi z'' + \varepsilon_2$$

and $|\varepsilon_2| < 2\varepsilon$ by the triangle inequality. Using the calculations of the dot product terms and cancelling all the relevant terms gives the desired result.

Finally, in the case where $a, b \neq 0$ but $a + b = 0$, the above proof still holds if the $\theta_{\{i,j\},a+b}$ term is omitted from the calculation of $\vec{\theta} \cdot Z(\vec{x}_3)$. $\square$

We now collect another lemma regarding the structure of Λ.

Lemma 2.5. *Let a be a non-zero element of $\mathbb{Z}_g$, and let i be some fixed number between 1 and k (inclusively). If $\vec{\theta} \in \Lambda$, then*

$$\sum_{1 \leqslant m < i} \theta_{\{m,i\},-a} + \sum_{i < m \leqslant k} \theta_{\{i,m\},a} \equiv 0 \pmod{2\pi},$$

where the appropriate empty sum is 0 in the case that $i = 1$ or $i = k$.

Proof. We define two vectors in $(\mathbb{Z}_g)^k$:

$$\vec{x}_1 = (0, \ldots, 0, a, 0, \ldots, 0)^T,$$
$$\vec{x}_2 = (0, \ldots, 0, 0, 0, \ldots, 0)^T.$$

The a that appears in $\vec{x}_1$ does so in the ith position. If $\vec{\theta} \in \mathbb{R}^d$, then again with $\vec{1}$ representing the vector in $\mathbb{R}^d$ of all ones, we have

$$\vec{\theta} \cdot Z(\vec{x}_1) = \sum_{1 \leqslant m < i} \theta_{\{m,i\}, -a} + \sum_{i < m \leqslant k} \theta_{\{i,m\}, a} - \frac{1}{g} \vec{\theta} \cdot \vec{1},$$
$$\vec{\theta} \cdot Z(\vec{x}_2) = -\frac{1}{g} \vec{\theta} \cdot \vec{1},$$

so if $\vec{\theta} \in \Lambda$, then by (2.1) these two expressions are equivalent modulo 2π, as desired. $\qquad \square$

Remark 1. We point out that the assumption that the underlying group G is cyclic is not meaningfully used in the proofs of Lemma 2.1 or any of its corollaries, nor is it used in the proof of Lemma 2.5. In fact, even if G is non-abelian (and the map Z is redefined appropriately), these proofs require only trivial modifications such as exchanging $0 \in \mathbb{Z}_g$ for the identity element of G.

In order to characterize Λ, we define a collection of 'building block' vectors.

Definition 3. Fix a pair $\{i, j\}$ with $1 \leqslant i < j < k$ (note the strict inequality $j < k$). Let $\mathbb{1}_{\{i,j\},a}$ denote the vector in $\mathbb{R}^d$ with a 1 in the $\{i, j\}, a$-component and 0 elsewhere. We define the vector $\vec{\alpha}^{\{i,j\}}$ as follows:

$$\vec{\alpha}^{\{i,j\}} = \sum_{n=1}^{g-1} \frac{2\pi n}{g} \mathbb{1}_{\{i,j\},n} + \sum_{n=1}^{g-1} \frac{2\pi(g-n)}{g} \mathbb{1}_{\{i,k\},n} + \sum_{n=1}^{g-1} \frac{2\pi n}{g} \mathbb{1}_{\{j,k\},n}$$

Here, n is regarded to be an integer in $[0, g-1]$, except where it appears in the subscript of $\mathbb{1}$ as the corresponding element of $\mathbb{Z}_g$.

Proposition 2.6. *Any $\vec{\alpha}^{\{i,j\}}$ vector as defined in Definition 3 is an element of Λ.*

Proof. Our goal is to show that the expression $Z(\vec{x}) \cdot \vec{\alpha}^{\{i,j\}}$ does not depend on $\vec{x}$. Using (1.1) and Definition 3, we see that

$$Z(\vec{x}) \cdot \vec{\alpha}^{\{i,j\}} \equiv \frac{2\pi(x_i - x_j)}{g} + \frac{2\pi(g - (x_i - x_k))}{g} + \frac{2\pi(x_j - x_k)}{g} - 3 \sum_{n=1}^{g-1} \frac{2\pi n}{g^2}$$

$$\equiv -3 \sum_{n=1}^{g-1} \frac{2\pi n}{g^2} \pmod{2\pi},$$

which gives the desired result. As an addendum, we remark that

$$-\sum_{n=1}^{g-1} \frac{2\pi n}{g^2} = -\frac{2\pi}{g^2} \frac{g(g-1)}{2} = -\frac{\pi(g-1)}{g},$$

whence

$$Z(\vec{x}) \cdot \vec{\alpha}^{\{i,j\}} \equiv -\frac{3\pi(g-1)}{g} \quad (\text{mod } 2\pi), \tag{2.2}$$

which we preserve for later use. $\qquad \square$

We are now prepared to fully characterize the set Λ. Since $|\Phi|$ is 2π-periodic, it will suffice to characterize Λ on the region $[-\pi, \pi)^d$; to that end, we let Λ_0 denote $\Lambda \cap [-\pi, \pi)^d$.

Lemma 2.7. *Let $\vec{\theta} \in \Lambda_0$ and let $\vec{\alpha}^{\{i,j\}}$ be as defined in Definition 3. There are constants $c_{\{i,j\}}$ with $1 \leqslant i < j < k$ such that $c_{\{i,j\}} \in \{0, 1, \dots, g-1\}$ and*

$$\vec{\theta} \equiv \sum_{1 \leqslant i < j < k} c_{\{i,j\}} \vec{\alpha}^{\{i,j\}} \quad (\text{mod } 2\pi).$$

Moreover, this representation of $\vec{\theta}$ is unique.

Proof. Let $\vec{\theta} \in \Lambda_0$. Set each coefficient $c_{\{i,j\}}$ to be

$$c_{\{i,j\}} = \begin{cases} \dfrac{g}{2\pi} \theta_{\{i,j\},1} & \text{if } \theta_{\{i,j\},1} \geqslant 0, \\[2mm] \dfrac{g}{2\pi} \theta_{\{i,j\},1} + g & \text{if } \theta_{\{i,j\},1} < 0, \end{cases}$$

and let

$$\vec{x} = \left[\sum_{1 \leqslant i < j < k} c_{\{i,j\}} \vec{\alpha}^{\{i,j\}} \right] - \vec{\theta}.$$

We will show that $\vec{x} \equiv 0 \pmod{2\pi}$. We first note that setting the $c_{\{i,j\}}$ terms in this way implies that each is in the set $\{0, 1, \dots, g-1\}$ by Corollary 2.4. Since Λ is a subgroup of $\mathbb{R}^d$, and since the bracketed term is in Λ by Proposition 2.6, so also is $\vec{x}$. Among the collection $\{\vec{\alpha}^{\{i,j\}} : i < j < k\}$, only $\vec{\alpha}^{\{1,2\}}$ has a non-zero $\{1,2\},1$-component, analogous comments apply for any other fixed pair $\{i,j\}$ with $i < j < k$. Thus, for $i < j < k$, we have that $x_{\{i,j\},1}$ is either 0 (if $\theta_{\{i,j\},1} \geqslant 0$) or 2π (if $\theta_{\{i,j\},1} < 0$). By Corollary 2.3, this inductively implies that $x_{\{i,j\},a} \equiv 0 \pmod{2\pi}$ for all a and all pairs $\{i,j\}$ with $i < j < k$. Hence, modulo 2π, the only possible non-zero coordinates of $\vec{x}$ are those of the form $x_{\{i,k\},a}$ for some i and a.

Next, fix i between 1 and $k-1$ (inclusively). By Lemma 2.5, we see that

$$x_{\{i,k\},1} \equiv -\sum_{1 \leqslant m < i} x_{\{m,i\},-1} - \sum_{i < m < k} x_{\{i,m\},1} \quad (\text{mod } 2\pi).$$

We have already established that each summand on the right-hand side is $0 \pmod{2\pi}$, since it is not of the form $x_{\{i,k\},a}$. Thus, $x_{\{i,k\},1} \equiv 0$ as well. By Corollary 2.3 again, this establishes that $x_{\{i,k\},a} \equiv 0$ for all a. The fact that this occurs for all i between 1 and $k-1$ completes the argument that $\vec{x} \equiv 0$. Finally, the uniqueness of this expression is immediate from the fact that each distinct vector $\vec{\alpha}^{\{i,j\}}$ is the unique contributor to the $\theta_{\{i,j\},1}$-component. $\qquad \square$

This characterization of Λ_0 motivates how we will break up the integral in (1.3); the primary contribution to the integral will be the regions in $[-\pi, \pi)^d$ that are close to Λ. For $\vec{\theta} \in [-\pi, \pi)^d$,

we define the box

$$B_\delta(\vec{\theta}) = \{\vec{\mu} \in [-\pi, \pi)^d : \vec{\mu} \equiv \vec{\theta} + \vec{\zeta} \text{ with } |\zeta_{\{i,j\},a}| < \delta \text{ for all } (\{i,j\}, a)\},$$

where the equivalence is taken modulo 2π. The parameter δ is assumed to be small and positive. We remark that if $\delta < \pi/g$, and $\vec{\theta}_1, \vec{\theta}_2$ are distinct elements of Λ_0, then $B_\delta(\vec{\theta}_1)$ and $B_\delta(\vec{\theta}_2)$ are disjoint by Corollary 2.4. Since we expect the bulk of the integral to be contributed from regions of the form $\{B_\delta(\vec{\theta}) : \vec{\theta} \in \Lambda_0\}$, we define the 'remainder set'

$$R_\delta = [-\pi, \pi)^d \setminus \bigcup_{\vec{\theta} \in \Lambda_0} B_\delta(\vec{\theta}). \tag{2.3}$$

Proposition 2.8. *Suppose $\delta < \pi/g$ and that t is a multiple of g. If g is odd, or if both g and t/g are even, then*

$$\mathbb{P}(X_t = \vec{0}) = \frac{g^{\binom{k-1}{2}}}{(2\pi)^d} \int_{B_\delta(\vec{0})} \Phi(\vec{\theta})^t \, d\vec{\theta} + \frac{1}{(2\pi)^d} \int_{R_\delta} \Phi(\vec{\theta})^t \, d\vec{\theta},$$

whereas if g is even and t/g is odd, then

$$\mathbb{P}(X_t = \vec{0}) = \frac{1}{(2\pi)^d} \int_{R_\delta} \Phi(\vec{\theta})^t \, d\vec{\theta}.$$

Proof. We first note that $\delta < \pi/g$ implies that distinct vectors $\vec{\eta} \in \Lambda_0$ have disjoint boxes $B_\delta(\vec{\eta})$. From this fact and (1.3), we have

$$\mathbb{P}(X_t = \vec{0}) = (2\pi)^{-d} \sum_{\vec{\eta} \in \Lambda_0} \int_{B_\delta(\vec{\eta})} \Phi(\vec{\theta})^t \, d\vec{\theta} + (2\pi)^{-d} \int_{R_\delta} \Phi(\vec{\theta})^t \, d\vec{\theta}. \tag{2.4}$$

Next, we consider a non-zero $\vec{\eta} \in \Lambda_0$. By (2.1), we note that $\vec{\eta} \cdot Z(\vec{x})$ is deterministic in that it does not depend on the random vector $\vec{x} \in (\mathbb{Z}_g)^k$. Hence, $e^{i\vec{\eta} \cdot X_1}$ is also deterministic, and we have $\Phi(\vec{\eta}) = \mathbb{E}[e^{i\vec{\eta} \cdot X_1}] = e^{i\vec{\eta} \cdot X_1}$. For any other $\vec{\zeta} \in \mathbb{R}^d$, it follows that

$$\begin{aligned}
\Phi(\vec{\eta} + \vec{\zeta}) &= \mathbb{E}[e^{i(\vec{\eta} + \vec{\zeta}) \cdot X_1}] \\
&= e^{i\vec{\eta} \cdot X_1} \mathbb{E}[e^{i\vec{\zeta} \cdot X_1}] \\
&= \Phi(\vec{\eta})\Phi(\vec{\zeta}).
\end{aligned}$$

Since Φ^t is 2π-periodic when t is a multiple of g, then for $\vec{\eta} \in \Lambda_0$ we have

$$\int_{B_\delta(\vec{\eta})} \Phi(\vec{\theta})^t \, d\vec{\theta} = \Phi(\vec{\eta})^t \int_{B_\delta(\vec{0})} \Phi(\vec{\theta})^t \, d\vec{\theta},$$

so (2.4) becomes

$$\mathbb{P}(X_t = \vec{0}) = \frac{\sum_{\vec{\eta} \in \Lambda_0} \Phi(\vec{\eta})^t}{(2\pi)^d} \int_{B_\delta(\vec{0})} \Phi(\vec{\theta})^t \, d\vec{\theta} + \frac{1}{(2\pi)^d} \int_{R_\delta} \Phi(\vec{\theta})^t \, d\vec{\theta}, \tag{2.5}$$

and our only remaining task is therefore to evaluate the sum $\sum_{\vec{\eta} \in \Lambda_0} \Phi(\vec{\eta})^t$.

Since $\vec{\eta} \in \Lambda_0$, by Lemma 2.7 there are constants $c_{\{i,j\}}$ and $\vec{z} \in (2\pi\mathbb{Z})^d$ so that

$$\vec{\eta} = \sum_{1 \leqslant i < j < k} c_{\{i,j\}} \vec{\alpha}^{\{i,j\}} + \vec{z}.$$

Combining this with (2.2) shows that

$$\vec{\eta} \cdot Z(\vec{x}) \equiv \sum_{1 \leqslant i < j < k} -c_{\{i,j\}} \frac{3\pi(g-1)}{g} + \vec{z} \cdot Z(\vec{x}) \pmod{2\pi}.$$

If $t = \lambda g$, then

$$\Phi(\vec{\eta})^t = e^{it\vec{\eta} \cdot X_1} = \exp\left[i\left(-3\pi(g-1)\lambda \sum_{1 \leqslant i < j < k} c_{\{i,j\}} + \lambda g \vec{z} \cdot Z(\vec{x})\right)\right].$$

Because $Z(\vec{x}) \in ((1/g)\mathbb{Z})^d$ and $\vec{z} \in (2\pi\mathbb{Z})^d$, it follows that $\vec{z} \cdot Z(\vec{x})$ is an integer multiple of $2\pi/g$. Therefore, $\lambda g \vec{z} \cdot Z(\vec{x})$ is an integer multiple of 2π, and thus

$$\Phi(\vec{\eta})^t = \exp\left[i\left(-3\pi(g-1)\lambda \sum_{1 \leqslant i < j < k} c_{\{i,j\}}\right)\right].$$

In the case where g is odd, or if $\lambda = t/g$ is even, then the term in parentheses is an integer multiple of 2π, whence $\Phi(\vec{\eta})^t = 1$ for every $\vec{\eta} \in \Lambda_0$. The uniqueness of the representation in Lemma 2.7 shows that $|\Lambda_0| = g^{\binom{k-1}{2}}$, which is equal to $\sum_{\vec{\eta} \in \Lambda_0} \Phi(\vec{\eta})^t$ in (2.5), as desired.

On the other hand, if g is even and λ is odd, then $-3\pi(g-1)\lambda \sum c_{\{i,j\}}$ must be congruent modulo 2π to 0 or π. In this case, half of the choices for the coefficients $c_{\{i,j\}}$ have odd parity and the other half have even parity. The same is true of the collection of all possible sums $\sum c_{\{i,j\}}$, so half the terms in $\sum_{\vec{\eta} \in \Lambda_0} \Phi(\vec{\eta})^t$ will be 1 and the other half will be -1, and the sum will be 0. $\square$

3. Bounds in the remainder region

This section is devoted to obtaining upper bounds on $|\Phi(\vec{\theta})|$ for $\vec{\theta} \in R_\delta$ and leveraging them to obtain an upper bound on the corresponding integral in Proposition 2.8. In all that follows, we assume that $\delta < \pi/g$. We begin by defining a useful set:

$$L = \{\vec{\theta} \in [-\pi, \pi)^d : \theta_{\{i,j\},a} \equiv 0 \pmod{2\pi/g} \text{ for all } \{i,j\}, a\}.$$

We recall from Corollary 2.4 that $\Lambda_0 \subset L$. Using the boxes defined in (2.3), we note that if $\vec{\eta}_1, \vec{\eta}_2$ are distinct elements of L, then $\delta < \pi/g$ implies that $B_\delta(\vec{\eta}_1) \cap B_\delta(\vec{\eta}_2) = \emptyset$. We set $R_\delta^A = \bigcup_{\vec{\eta} \in L \setminus \Lambda_0} B_\delta(\vec{\eta})$ and $R_\delta^B = [-\pi, \pi)^d \setminus \bigcup_{\vec{\eta} \in L} B_\delta(\vec{\eta})$, which yields

$$R_\delta = R_\delta^A \cup R_\delta^B. \tag{3.1}$$

Intuitively, R_δ^A is the portion of the remainder region that is close to satisfying the modular condition of Corollary 2.4 (but is not near Λ), and R_δ^B is the region that is far from satisfying the modular condition.

Lemma 3.1. *Suppose $\delta < \frac{8}{5}g^{-k-3}k^{-2}$. Then if $\vec{\theta} \in R_\delta^A$, we have*

$$|\Phi(\vec{\theta})| \leqslant 1 - \frac{1}{10}g^{-k-2}.$$

Proof. Suppose $\vec{\eta} \in L \setminus \Lambda_0$ and that $\vec{\theta} \in B_\delta(\vec{\eta})$; then $\vec{\theta}$ is equivalent (modulo 2π) to $\vec{\eta} + \vec{\zeta}$ where $|\zeta_{\{i,j\},a}| < \delta$ for all $\{i,j\}, a$. Since $\eta \in L \setminus \Lambda_0$, by (2.1) there are $\vec{x}, \vec{y} \in (\mathbb{Z}_g)^k$ such that

$$\vec{\eta} \cdot (Z(\vec{x}) - Z(\vec{y})) \equiv \frac{2\pi n}{g} \quad (\text{mod } 2\pi)$$

with $n \in \{1, \ldots, g-1\}$. It follows that

$$|\Phi(\vec{\eta})| = \left| \sum_{\vec{w} \in (\mathbb{Z}_g)^k} g^{-k} e^{i\vec{\eta} \cdot Z(\vec{w})} \right| \leqslant g^{-k} \left[|e^{i\vec{\eta} \cdot Z(\vec{x})} + e^{i\vec{\eta} \cdot Z(\vec{y})}| + \left| \sum_{\vec{w} \neq \vec{x}, \vec{y}} e^{i\vec{\eta} \cdot Z(\vec{w})} \right| \right],$$

and since $|e^{ia} + e^{ib}|^2 = 2 + 2\cos(a-b)$, we have

$$|\Phi(\vec{\eta})| \leqslant g^{-k} \left[\sqrt{2 + 2\cos\left(\frac{2\pi}{g}\right)} + g^k - 2 \right].$$

Applying the bounds $\sqrt{x} \leqslant 1 + x/4$ and $\cos(x) \leqslant 1 - x^2/2 + x^4/24$ to this expression yields

$$|\Phi(\vec{\eta})| \leqslant 1 - g^{-k} \left[\left(\frac{\pi}{g}\right)^2 - \frac{(\pi/g)^4}{3} \right]. \tag{3.2}$$

Together with the fact that

$$\left(\frac{\pi}{g}\right)^4 \leqslant \left(\frac{\pi}{2}\right)^2 \left(\frac{\pi}{g}\right)^2$$

when $g \geqslant 2$, we have

$$|\Phi(\vec{\eta})| \leqslant 1 - g^{-k} \left(\frac{\pi}{g}\right)^2 \left[1 - \frac{(\pi/2)^2}{3} \right].$$

Hence,

$$|\Phi(\vec{\eta})| \leqslant 1 - \frac{1}{10} g^{-k} \left(\frac{\pi}{g}\right)^2 \leqslant 1 - \frac{9}{10} g^{-k-2}. \tag{3.3}$$

We also note that $\vec{\theta} = \vec{\eta} + \vec{\zeta} + \vec{z}$ with $\vec{z} \in (2\pi\mathbb{Z})^d$, and since $|\Phi|$ is 2π-periodic, it follows that

$$|\Phi(\vec{\theta})| = |\Phi(\vec{\eta} + \zeta)|. \tag{3.4}$$

We now state a pair of remainder bounds on Taylor polynomials for e^z that we will use here and elsewhere: if $a \geqslant 0$ and b is real, then

$$\left| e^{-a} - \sum_{s=0}^{j} \frac{(-a)^s}{s!} \right| \leqslant \min\left\{ \frac{2|a|^j}{j!}, \frac{|a|^{j+1}}{(j+1)!} \right\}, \tag{3.5}$$

$$\left| e^{ib} - \sum_{s=0}^{j} \frac{(ib)^s}{s!} \right| \leqslant \min\left\{ \frac{2|b|^j}{j!}, \frac{|b|^{j+1}}{(j+1)!} \right\}. \tag{3.6}$$

For a reference, (3.5) can be found as [2, equation 26.4]; (3.6) is proved similarly. From the triangle inequality, we see that

$$|\Phi(\vec{\eta}+\vec{\zeta})-\Phi(\vec{\eta})| = g^{-k}\left|\sum_{\vec{w}\in(\mathbb{Z}_g)^k}e^{i(\vec{\eta}+\vec{\zeta})\cdot Z(\vec{w})}-e^{i\vec{\eta}\cdot Z(\vec{w})}\right|$$

$$\leqslant g^{-k}\sum_{\vec{w}\in(\mathbb{Z}_g)^k}|e^{i\vec{\eta}\cdot Z(\vec{w})}||e^{i\vec{\zeta}\cdot Z(\vec{w})}-1|.$$

Since $|e^{i\vec{\eta}\cdot Z(\vec{w})}| = 1$, applying (3.6) to $|e^{i\vec{\zeta}\cdot Z(\vec{w})}-1|$ shows that

$$|\Phi(\vec{\eta}+\vec{\zeta})-\Phi(\vec{\eta})| \leqslant g^{-k}\sum_{\vec{w}\in(\mathbb{Z}_g)^k}|\vec{\zeta}\cdot Z(\vec{w})|. \tag{3.7}$$

When considering an individual term $|\vec{\zeta}\cdot Z(\vec{w})|$, we observe that $|\zeta_{\{i,j\},a}| < \delta$, and each coordinate of $Z(\vec{w})$ is at most $1-1/g$. These vectors have $(g-1)\binom{k}{2}$ coordinates, so

$$|\vec{\zeta}\cdot Z(\vec{w})| \leqslant (g-1)\binom{k}{2}\delta\left(1-\frac{1}{g}\right) = \frac{1}{g}(g-1)^2\frac{k(k-1)}{2}\delta.$$

There are g^k terms in the summation in (3.7), and therefore we have

$$|\Phi(\vec{\eta}+\vec{\zeta})-\Phi(\vec{\eta})| \leqslant \frac{1}{g}(g-1)^2\frac{k(k-1)}{2}\delta. \tag{3.8}$$

Our assumption that $\delta \leqslant \frac{8}{5}g^{-k-3}k^{-2}$ implies that

$$\frac{1}{g}(g-1)^2\binom{k}{2}\delta \leqslant \frac{4}{5}g^{-k-2}.$$

Combining this with (3.4), (3.3) and (3.8) shows that

$$|\Phi(\vec{\theta})| = |\Phi(\vec{\eta}+\vec{\zeta})| \leqslant |\Phi(\vec{\eta})| + |\Phi(\vec{\eta}+\vec{\zeta})-\Phi(\vec{\eta})| \leqslant 1 - \frac{9}{10}g^{-k-2} + \frac{4}{5}g^{-k-2},$$

as desired. $\qquad\square$

Next, we bound $|\Phi(\vec{\theta})|$ on the region R^B_δ.

Lemma 3.2. *Suppose $\delta < \pi/g$. Then if $\vec{\theta}\in R^B_\delta$, we have*

$$|\Phi(\vec{\theta})| \leqslant 1 - \frac{11}{48}g^{-k}\left(\frac{\delta}{2}\right)^2.$$

Proof. If $\vec{\theta}\in R^B_\delta$, then there is some $\{i,j\},a$ such that $\theta_{\{i,j\},a}$ is not expressible as $2\pi n/g+\zeta$ with $n\in\mathbb{Z}$ and $|\zeta| < \delta$. By Corollary 2.2, this means that there is a pair of vectors $\vec{x},\vec{y}\in(\mathbb{Z}_g)^k$ such that for any $z\in\mathbb{Z}$, the equation $\vec{\theta}\cdot Z(\vec{x})-\vec{\theta}\cdot Z(\vec{y}) = 2\pi z+\varepsilon_1$ requires that $|\varepsilon_1| > \delta/2$; in other words, $\vec{\theta}\cdot Z(\vec{x})$ and $\vec{\theta}\cdot Z(\vec{y})$ are not within $\delta/2$ of each other when taken modulo 2π. If we choose z and ε_1 so that $|\varepsilon_1| \leqslant \pi$, then it follows that

$$\cos(\vec{\theta}\cdot Z(\vec{x})-\vec{\theta}\cdot Z(\vec{y})) = \cos(\varepsilon_1) \leqslant \cos(\delta/2).$$

By repeating the same arguments that led to (3.2) with this $\vec{x}$ and $\vec{y}$, it follows that

$$|\Phi(\vec{\theta})| \leqslant 1 - g^{-k}\left[\frac{(\delta/2)^2}{4} - \frac{(\delta/2)^4}{48}\right],$$

and because $\delta < 2$, the desired result follows.   □

Finally, we put together the bounds on $|\Phi(\vec{\theta})|$ over R_δ^A and R_δ^B to obtain the bound on the R_δ-component of Proposition 2.8.

Proposition 3.3. *If $\delta < \frac{8}{5}g^{-k-3}k^{-2}$, then*

$$\left|\frac{1}{(2\pi)^d}\int_{R_\delta}\Phi(\vec{\theta})^t\,d\vec{\theta}\right| \leqslant \exp\left(-\frac{11}{192}g^{-k}t\delta^2\right).$$

Proof. Suppose $\delta < \frac{8}{5}g^{-k-3}k^{-2}$. Let $\vec{\theta} \in R_\delta$; since $\delta < \pi/g$, by (3.1) either $\vec{\theta} \in R_\delta^A$ or $\vec{\theta} \in R_\delta^B$, and the hypotheses of both Lemmas 3.1 and 3.2 apply. The assumptions on δ imply also that $\delta < 1/g$, so the bound on $|\Phi(\vec{\theta})|$ in Lemma 3.2 is higher than that of Lemma 3.1; hence, we can assert that

$$|\Phi(\vec{\theta})| \leqslant 1 - \frac{11}{48}g^{-k}\left(\frac{\delta}{2}\right)^2.$$

Since $R_\delta \subset [-\pi,\pi)^d$ and $1 - x \leqslant e^{-x}$, we conclude that

$$\begin{aligned}
\left|\frac{1}{(2\pi)^d}\int_{R_\delta}\Phi(\vec{\theta})^t\,d\vec{\theta}\right| &\leqslant \frac{1}{(2\pi)^d}\int_{R_\delta}|\Phi(\vec{\theta})^t|\,d\vec{\theta}\\
&\leqslant \left[1 - \frac{11}{48}g^{-k}\left(\frac{\delta}{2}\right)^2\right]^t\\
&\leqslant \exp\left(-\frac{11}{192}tg^{-k}\delta^2\right),
\end{aligned}$$

as desired.   □

4. Bounds in the primary region

The goal of this section is to obtain sharp estimates on $|\Phi(\vec{\theta})|$ in the region $B_\delta(\vec{0})$. Our first task is to calculate and estimate quantities of the form $\mathbb{E}[(\vec{\theta}\cdot Z(\vec{x}))^p]$ for $p = 1,2,3,4$, where the expectation denotes that $\vec{x}$ is drawn randomly and uniformly from $(\mathbb{Z}_g)^k$.

Proposition 4.1. *For any $\vec{\theta} \in \mathbb{R}^d$, $\mathbb{E}[\vec{\theta}\cdot Z(\vec{x})] = 0$.*

Proof. Fix some $\{i,j\},a$. That component of the dot product will be $(1-1/g)\theta_{\{i,j\},a}$ if $x_i - x_j = a$ and $-(1/g)\theta_{\{i,j\},a}$ if $x_i - x_j \neq a$. If $\vec{x}$ is chosen randomly and uniformly from $(\mathbb{Z}_g)^k$, then $x_i - x_j$

will be a random, uniformly distributed element of G. Thus, the expected value of the $\{i,j\}$, a-component of the dot product will be

$$\frac{1}{g}\left(1-\frac{1}{g}\right)\theta_{\{i,j\},a} - \frac{g-1}{g}\frac{1}{g}\theta_{\{i,j\},a} = 0. \qquad \square$$

To describe the second moment, we will first define a $d \times d$ matrix M which is indexed in the same way as $\mathbb{R}^d$:

$$M_{(\{i,j\},a),(\{m,n\},b)} = \begin{cases} (g-1)/g^2 & \text{if } \{i,j\} = \{m,n\} \text{ and } a = b, \\ -1/g^2 & \text{if } \{i,j\} = \{m,n\} \text{ and } a \neq b, \\ 0 & \text{if } \{i,j\} \neq \{m,n\}. \end{cases} \tag{4.1}$$

Proposition 4.2. *For any $\vec{\theta} \in \mathbb{R}^d$, we have $\mathbb{E}[(\vec{\theta} \cdot Z(\vec{x}))^2] = \vec{\theta}^T M \vec{\theta}$.*

Proof. For convenience of notation, we will let $\vec{\xi} = Z(\vec{x})$; that is, $\vec{\xi}$ is a vector in $\mathbb{R}^d$ chosen randomly and uniformly from the collection $\{Z(\vec{x}) : \vec{x} \in (\mathbb{Z}_g)^k\}$. As in the proof of Proposition 4.1, for a fixed $\{i,j\}, a$, the random variable $\xi_{\{i,j\},a}$ is $1 - 1/g$ with probability $1/g$ and is $-1/g$ with probability $1 - 1/g$. We note that

$$\mathbb{E}[(\vec{\theta} \cdot \vec{\xi})^2] = \mathbb{E}\left[\left(\sum_{\{i,j\},a} \theta_{\{i,j\},a}\xi_{\{i,j\},a}\right)^2\right]$$

$$= \sum_{(\{i,j\},a),(\{m,n\},b)} \theta_{\{i,j\},a}\theta_{\{m,n\},b}\mathbb{E}[\xi_{\{i,j\},a}\xi_{\{m,n\},b}].$$

Our goal is therefore to show that the expected value $\mathbb{E}[\xi_{\{i,j\},a}\xi_{\{m,n\},b}]$ agrees with the $(\{i,j\},a)$, $(\{m,n\},b)$ entry of M.

We first consider the case where $\{i,j\} = \{m,n\}$ and $a = b$. To compute $\mathbb{E}[(\xi_{\{i,j\},a})^2]$, we recall that the random variable $(\xi_{\{i,j\},a})^2$ is $(1 - 1/g)^2$ with probability $1/g$ and is $1/g^2$ with probability $1 - 1/g$; hence, its expectation is $(g-1)/g^2$, as desired. Next, we consider the case where $\{i,j\} = \{m,n\}$, but $a \neq b$. To compute $\mathbb{E}[\xi_{\{i,j\},a}\xi_{\{i,j\},b}]$, we note that there are two possibilities: if $x_i - x_j \in \{a,b\}$, then $\xi_{\{i,j\},a}\xi_{\{i,j\},b} = -(1 - 1/g)1/g$, and if $x_i - x_j \notin \{a,b\}$, then $\xi_{\{i,j\},a}\xi_{\{i,j\},b} = 1/g^2$. The former will occur with probability $2/g$, and the latter will occur with probability $1 - 2/g$. Hence, the expected value is $-1/g^2$.

Finally, if the pairs $\{i,j\}$ and $\{m,n\}$ are not the same, we claim that the variables $\xi_{\{i,j\},a}$ and $\xi_{\{m,n\},b}$ are independent. We will show this by separately considering the cases where $|\{i,j\} \cap \{m,n\}|$ is 0 or 1. If $\{i,j\}$ and $\{m,n\}$ are disjoint, then the expressions $x_i - x_j$ and $x_m - x_n$ are clearly independent of one another since $\vec{x}$ is chosen uniformly from $(\mathbb{Z}_g)^k$. If the pairs are of the form $\{i,j\}$ and $\{i,m\}$, then since x_i, x_j, x_m, $x_i - x_j$, and $x_i - x_m$ are all uniformly distributed on

$\mathbb{Z}_g$ and the first three are independent of each other, then for any $g_1, g_2 \in \mathbb{Z}_g$, we have

$$\mathbb{P}(x_i - x_j = g_1, \ x_i - x_m = g_2) = \sum_{c \in \mathbb{Z}_g} \mathbb{P}(x_i = c, \ x_j = c - g_1, \ x_m = c - g_2)$$

$$= \sum_{c \in \mathbb{Z}_g} \mathbb{P}(x_i = c) \mathbb{P}(x_j = c - g_1) \mathbb{P}(x_m = c - g_2)$$

$$= \sum_{c \in \mathbb{Z}_g} \frac{1}{g^3} = \frac{1}{g^2},$$

which is also equal to $\mathbb{P}(x_i - x_j = g_1) \mathbb{P}(x_i - x_m = g_2)$. A similar argument can be made for any configuration of $\{i, j\}, \{m, n\}$ with exactly one shared element. Hence, whether $|\{i, j\} \cap \{m, n\}|$ is 0 or 1, the variables $\xi_{\{i,j\},a}$ and $\xi_{\{m,n\},b}$ are independent, and consequently

$$\mathbb{E}[\xi_{\{i,j\},a} \xi_{\{m,n\},b}] = \mathbb{E}[\xi_{\{i,j\},a}] \mathbb{E}[\xi_{\{m,n\},b}] = 0$$

as in Proposition 4.1. □

We will not need to compute the third and fourth moments of $\vec{\theta} \cdot Z(\vec{x})$ explicitly; rather, we will only require estimates of those moments. However, we will need to compute the determinant of M. In all that follows, we will use I_n to denote the $n \times n$ identity matrix. We will use the following well-known identity, which can be found (for instance) in [10, Corollary 18.1.2].

Lemma 4.3 (Sylvester's Determinant Identity). *For any $n \times m$ matrix S and $m \times n$ matrix U, we have $\det(I_n + SU) = \det(I_m + US)$.*

Proposition 4.4. *With M as defined by* (4.1)*, we have $\det(M) = g^{-g\binom{k}{2}}$.*

Proof. We note that M is a block diagonal matrix with $\binom{k}{2}$ repeated copies of the same $(g-1) \times (g-1)$ submatrix S with $(g-1)/g^2$ in the diagonal entries and $-1/g^2$ in the off-diagonal entries. We can therefore express S as $g^{-1} I_{g-1} - g^{-2} \vec{1} \vec{1}^T$, where $\vec{1}$ is the column vector of length $g-1$ consisting of all ones. By Sylvester's Determinant Identity,

$$\det(S) = g^{-(g-1)} \det(I_{g-1} - g^{-1} \vec{1} \vec{1}^T) = g^{-(g-1)} \det(1 - g^{-1} \vec{1}^T \vec{1}) = g^{-(g-1)} \left(1 - \frac{g-1}{g}\right) = g^{-g}.$$

The desired result follows from the block structure of M. □

Finally, we state the lemma that gives the desired estimates on $|\Phi(\vec{\theta})|$ in the region $B_\delta(\vec{0})$.

Lemma 4.5. *Let $\delta > 0$ and $\vec{\theta} \in B_\delta(\vec{0})$. Then there is a function $\varepsilon : B_\delta(\vec{0}) \to \mathbb{R}$ such that*

$$\mathrm{Re}(\Phi(\vec{\theta})) = e^{-\frac{1}{2} \vec{\theta}^T M \vec{\theta}} (1 + \varepsilon(\vec{\theta})) \tag{4.2}$$

and

$$|\varepsilon(\vec{\theta})| \leqslant \frac{1}{6} (d\delta)^4 e^{\frac{1}{2} d^2 \delta^2}.$$

Moreover,

$$|\text{Im}(\Phi(\vec{\theta}))| \leqslant \frac{(d\delta)^3}{6}. \tag{4.3}$$

Further, if $d\delta < 1$, then

$$\text{Re}(\Phi(\vec{\theta})) > \frac{1}{3}. \tag{4.4}$$

Proof. We will mimic the proof of Lemma 3.1 of [5]. First, using (3.5) with $j = 1$ shows that

$$\left| e^{-\frac{1}{2}\vec{\theta}^T M \vec{\theta}} - \left(1 - \frac{1}{2}\vec{\theta}^T M \vec{\theta}\right) \right| \leqslant \frac{1}{8}(\vec{\theta}^T M \vec{\theta})^2. \tag{4.5}$$

The coefficients of M are bounded between -1 and 1, and the components of $\vec{\theta}$ are bounded between $-\delta$ and δ. Hence, by the triangle inequality,

$$\begin{aligned} |\vec{\theta}^T M \vec{\theta}| &\leqslant \sum_{(\{i,j\},a),(\{m,n\},b)} |\theta_{\{i,j\},a}\theta_{\{m,n\},b}M_{(\{i,j\},a),(\{m,n\},b)}| \\ &\leqslant \sum_{(\{i,j\},a),(\{m,n\},b)} \delta^2 = d^2\delta^2. \end{aligned} \tag{4.6}$$

Putting this together with (4.5) shows that

$$\left| e^{-\frac{1}{2}\vec{\theta}^T M \vec{\theta}} - \left(1 - \frac{1}{2}\vec{\theta}^T M \vec{\theta}\right) \right| \leqslant \frac{1}{8}d^4\delta^4. \tag{4.7}$$

Next, let $\vec{x} \in (\mathbb{Z}_g)^k$. Using (3.6) with $j = 3$ gives

$$\left| e^{i\vec{\theta}\cdot Z(\vec{x})} - \left[1 + i\vec{\theta}\cdot Z(\vec{x}) - \frac{(\vec{\theta}\cdot Z(\vec{x}))^2}{2} - \frac{i(\vec{\theta}\cdot Z(\vec{x}))^3}{6}\right] \right| \leqslant \frac{1}{24}(\vec{\theta}\cdot Z(\vec{x}))^4.$$

By examining only the real part of the term in the absolute value and recalling that $|\text{Re}(z)| \leqslant |z|$ for any $z \in \mathbb{C}$, we have

$$\left| \text{Re}(e^{i\vec{\theta}\cdot Z(\vec{x})}) - \left[1 - \frac{(\vec{\theta}\cdot Z(\vec{x}))^2}{2}\right] \right| \leqslant \frac{1}{24}(\vec{\theta}\cdot Z(\vec{x}))^4.$$

If $\vec{x}$ is chosen randomly and uniformly from $(\mathbb{Z}_g)^k$, then this shows that

$$\begin{aligned} &\left| \mathbb{E}[\text{Re}(e^{i\vec{\theta}\cdot Z(\vec{x})})] - \mathbb{E}\left[1 - \frac{(\vec{\theta}\cdot Z(\vec{x}))^2}{2}\right] \right| \\ &\qquad \leqslant \mathbb{E}\left| \text{Re}(e^{i\vec{\theta}\cdot Z(\vec{x})}) - \left[1 - \frac{(\vec{\theta}\cdot Z(\vec{x}))^2}{2}\right] \right| \\ &\qquad \leqslant \frac{1}{24}\mathbb{E}[(\vec{\theta}\cdot Z(\vec{x}))^4]. \end{aligned}$$

From the linearity of the Re operator, we have $\mathbb{E}[\text{Re}(e^{i\vec{\theta}\cdot Z(\vec{x})})] = \text{Re}(\Phi(\vec{\theta}))$. Thus, Proposition 4.2 shows that

$$\left| \text{Re}(\Phi(\vec{\theta})) - \left[1 - \frac{1}{2}\vec{\theta}^T M \vec{\theta}\right] \right| \leqslant \frac{1}{24}\mathbb{E}[(\vec{\theta}\cdot Z(\vec{x}))^4]. \tag{4.8}$$

Next, we seek to obtain a similar bound on $\text{Im}(\Phi(\vec{\theta}))$. Using (3.6) again with $j = 2$ shows that

$$\left| e^{i\vec{\theta}\cdot Z(\vec{x})} - \left[1 + i\vec{\theta}\cdot Z(\vec{x}) - \frac{1}{2}(\vec{\theta}\cdot Z(\vec{x}))^2 \right] \right| \leqslant \frac{1}{6}|\vec{\theta}\cdot Z(\vec{x})|^3,$$

so examining only the imaginary part in the absolute value and using the fact that $|\text{Im}(z)| \leqslant |z|$ gives

$$\left| \text{Im}(e^{i\vec{\theta}\cdot Z(\vec{x})}) - \vec{\theta}\cdot Z(\vec{x}) \right| \leqslant \frac{1}{6}|\vec{\theta}\cdot Z(\vec{x})|^3.$$

By the same argument as for the real part, if $\vec{x} \in (\mathbb{Z}_g)^k$ is chosen randomly and uniformly, then

$$\left| \text{Im}(\Phi(\vec{\theta})) - \mathbb{E}[\vec{\theta}\cdot Z(\vec{x})] \right| \leqslant \frac{1}{6}\mathbb{E}[|\vec{\theta}\cdot Z(\vec{x})|^3],$$

so by Proposition 4.1,

$$\left| \text{Im}(\Phi(\vec{\theta})) \right| \leqslant \frac{1}{6}\mathbb{E}[|\vec{\theta}\cdot Z(\vec{x})|^3]. \tag{4.9}$$

The next step is to bound the expectations in (4.8) and (4.9). For any $\vec{x} \in (\mathbb{Z}_g)^k$, the components of $Z(\vec{x})$ all have absolute value less than 1, and the components of $\vec{\theta}$ all have absolute value at most δ. Hence,

$$|\vec{\theta}\cdot Z(\vec{x})| \leqslant \sum_{\{i,j\},a} |\theta_{\{i,j\},a}| \leqslant d\delta. \tag{4.10}$$

Combining this with (4.9) yields (4.3). Similarly, combining (4.10) with (4.8) yields

$$\left| \text{Re}(\Phi(\vec{\theta})) - \left[1 - \frac{1}{2}\vec{\theta}^T M\vec{\theta} \right] \right| \leqslant \frac{(d\delta)^4}{24},$$

so (4.7) and the triangle inequality give

$$\left| \text{Re}(\Phi(\vec{\theta})) - e^{-\frac{1}{2}\vec{\theta}^T M\vec{\theta}} \right| \leqslant \frac{(d\delta)^4}{6}.$$

Dividing both sides by $e^{-\frac{1}{2}\vec{\theta}^T M\vec{\theta}}$ then applying (4.6) to the right side shows that

$$\left| \frac{\text{Re}(\Phi(\vec{\theta}))}{e^{-\frac{1}{2}\vec{\theta}^T M\vec{\theta}}} - 1 \right| \leqslant \frac{1}{6}(d\delta)^4 e^{\frac{1}{2}d^2\delta^2}.$$

Therefore, we can write

$$\text{Re}(\Phi(\vec{\theta})) = e^{-\frac{1}{2}\vec{\theta}^T M\vec{\theta}} \left[\frac{\text{Re}(\Phi(\vec{\theta}))}{e^{-\frac{1}{2}\vec{\theta}^T M\vec{\theta}}} \right] = e^{-\frac{1}{2}\vec{\theta}^T M\vec{\theta}}(1 + \varepsilon(\vec{\theta}))$$

where

$$|\varepsilon(\vec{\theta})| \leqslant \frac{1}{6}(d\delta)^4 e^{\frac{1}{2}d^2\delta^2}.$$

This establishes (4.2).

Finally, we observe by (4.2) that

$$\text{Re}(\Phi(\vec{\theta})) \geqslant e^{-\frac{1}{2}\vec{\theta}^T M\vec{\theta}} \left(1 - \frac{1}{6}(d\delta)^4 e^{\frac{1}{2}d^2\delta^2} \right)$$

so by (4.6) and the assumption that $d\delta < 1$, we have

$$\mathrm{Re}(\Phi(\vec{\theta})) \geqslant \frac{1 - (1/6)(d\delta)^4 e^{\frac{1}{2}d^2\delta^2}}{e^{\frac{1}{2}\vec{\theta}^T M\vec{\theta}}} \geqslant \frac{1 - (1/6)e^{1/2}}{e^{1/2}} > 1/3$$

which proves (4.4). $\qquad\square$

5. Proof of main theorem

Our final task is to put all the pieces together to obtain suitable estimates on the return probability of the random walk. We first gather an assortment of technical lemmas.

Proposition 5.1. *There is a symmetric matrix P such that $P^2 = M$. Moreover, there are positive constants D_1, D_2 which depend only on g such that, for all $\delta > 0$,*

$$[-D_1\delta, D_1\delta]^d \subset P[-\delta, \delta]^d \subset [-D_2\delta, D_2\delta]^d,$$

where $P[\delta, \delta]^d = \{P\vec{\mu} : \vec{\mu} \in [-\delta, \delta]^d\}$.

Proof. Propositions 4.2 and 4.4 imply that M is positive definite. As noted in the proof of Proposition 4.4, M is a block diagonal matrix with $\binom{k}{2}$ repeated copies of the same $(g-1) \times (g-1)$ submatrix $S = g^{-1}I_{g-1} - g^{-2}\vec{1}\vec{1}^T$. It follows that S is also positive definite, so there is a symmetric, positive definite matrix (call it Q) such that $Q^2 = S$. The linear transformation corresponding to Q maps $[-1, 1]^{g-1}$ to a non-degenerate subset of $\mathbb{R}^{g-1}$; thus, there are constants D_1, D_2 such that

$$[-D_1, D_1]^{g-1} \subset Q[-1, 1]^{g-1} \subset [-D_2, D_2]^{g-1}.$$

Since S and Q depend only on g, the same is true of D_1 and D_2.

The block diagonal matrix P consisting of $\binom{k}{2}$ repeated copies of Q is therefore a symmetric, positive definite matrix for which $P^2 = M$. Moreover, it follows that $[-D_1, D_1]^d \subset P[-1, 1]^d \subset [-D_2, D_2]^d$, where D_1 and D_2 are the same constants as above which depend only on g. Since the transformation associated to P is linear, scaling by δ completes the result. $\qquad\square$

For $z \in \mathbb{C}$, we set $\beta(z) = \mathrm{Im}(z)/\mathrm{Re}(z)$ and $\alpha(z, t) = 1 - \binom{t}{2}\beta(z)^2$.

Lemma 5.2. *Let $t \geqslant 2$ be an integer, and let $z \in \mathbb{C}$ with $\mathrm{Re}(z) > 0$ and $\alpha(z, t) > 0$. Then*

$$\mathrm{Re}(z^t) \leqslant \mathrm{Re}(z)^t (1 + [\beta(z)]^2)^{t/2} \tag{5.1}$$

and

$$\mathrm{Re}(z^t) \geqslant \mathrm{Re}(z)^t (1 + [\beta(z)]^2)^{t/2} \left(1 + \left[\frac{t}{\alpha(z,t)}\right]^2 [\beta(z)]^2\right)^{-1/2}. \tag{5.2}$$

Proof. This requires only trivial modifications to parts (i) and (iv) of Proposition A.1 in [5]. $\qquad\square$

Lemma 5.3. *Let ρ be a positive real number. Then*

$$\sqrt{2\pi(1 - e^{-\rho^2/2})} \leqslant \int_{-\rho}^{\rho} e^{-\frac{1}{2}x^2} \, dx \leqslant \sqrt{2\pi(1 - e^{-\rho^2})}.$$

Proof. By multiplying two copies of the integral together, applying Fubini's theorem, and converting to polar coordinates, we have

$$\int_0^{\rho} 2\pi r e^{-\frac{1}{2}r^2} \, dr < \int_{[-\rho,\rho]^2} e^{-\frac{1}{2}(x^2+y^2)} \, dy \, dx < \int_0^{\sqrt{2}\rho} 2\pi r e^{-\frac{1}{2}r^2} \, dr.$$

Computing the left and right sides and taking square roots gives the result. □

With D_1, D_2 as defined in Proposition 5.1, we define

$$L(g,k,t,\delta) = [1 + t^2(d\delta)^6]^{-1/2} \left[1 - \frac{1}{3}(d\delta)^4\right]^t [1 - e^{-\frac{t}{2}(D_1\delta)^2}]^{d/2},$$

$$U(g,k,t,\delta) = \left[1 + \frac{1}{4}(d\delta)^6\right]^{t/2} \left[1 + \frac{1}{3}(d\delta)^4\right]^t [1 - e^{-t(D_2\delta)^2}]^{d/2}.$$

Theorem 5.4. *Suppose that $\delta < \frac{8}{5}g^{-k-3}k^{-2}$, and let t be any positive integer multiple of g such that $t < 2(d\delta)^{-3}$. If g is odd, or if g and t/g are both even, then*

$$\mathbb{P}(X_t = \vec{0}) \leqslant \frac{g^{\frac{g}{2}\binom{k}{2} + \binom{k-1}{2}}}{(2\pi t)^{d/2}} U(g,k,t,\delta) + \exp\left(-\frac{11}{192}g^{-k}t\delta^2\right) \tag{5.3}$$

and

$$\mathbb{P}(X_t = \vec{0}) \geqslant \frac{g^{\frac{g}{2}\binom{k}{2} + \binom{k-1}{2}}}{(2\pi t)^{d/2}} L(g,k,t,\delta) - \exp\left(-\frac{11}{192}g^{-k}t\delta^2\right). \tag{5.4}$$

Henceforth, δ will be chosen to vary with t in such a way that the exponential term above will tend to 0 and the U and L terms will tend to 1, which will complete the proof of Theorem 1.1.

Proof. If g is odd or t/g is even, Propositions 2.8 and 3.3 show that

$$\left| \mathbb{P}(X_t = \vec{0}) - \frac{g^{\binom{k-1}{2}}}{(2\pi)^d} \int_{B_\delta(\vec{0})} \Phi(\vec{\theta})^t \, d\vec{\theta} \right| \leqslant \exp\left(-\frac{11}{192}g^{-k}t\delta^2\right).$$

Therefore, to prove (5.3) and (5.4), it will suffice to show that

$$\frac{L(g,k,t,\delta)g^{\frac{g}{2}\binom{k}{2}}}{(2\pi t)^{d/2}} \leqslant (2\pi)^{-d} \int_{B_\delta(\vec{0})} \Phi(\vec{\theta})^t \, d\vec{\theta} \leqslant \frac{U(g,k,t,\delta)g^{\frac{g}{2}\binom{k}{2}}}{(2\pi t)^{d/2}}. \tag{5.5}$$

Moreover, since $\Phi(-\vec{\theta})$ and $\Phi(\vec{\theta})$ are complex conjugates and $B_\delta(\vec{0})$ is closed under negation, we have

$$\int_{B_\delta(\vec{0})} \Phi(\vec{\theta})^t \, d\vec{\theta} = \int_{B_\delta(\vec{0})} \text{Re}(\Phi(\vec{\theta})^t) \, d\vec{\theta}. \tag{5.6}$$

Our strategy will be to relate $\text{Re}(\Phi(\vec{\theta})^t)$ to $[\text{Re}(\Phi(\vec{\theta}))]^t$ by using Lemma 5.2.

We note that $\delta < \frac{8}{5}g^{-k-3}k^{-2} < 2g^{-1}k^{-2} < d^{-1}$; the relationship $d\delta < 1$ will be referenced repeatedly throughout the proof. Using again the definitions $\beta(z) = \text{Im}(z)/\text{Re}(z)$ and $\alpha(z,t) = 1 - \binom{t}{2}\beta(z)^2$, we note by (4.3) and (4.4) that for $\vec{\theta} \in B_\delta(\vec{0})$ we have

$$|\beta(\Phi(\vec{\theta}))| \leqslant \frac{(d\delta)^3/6}{1/3} = \frac{(d\delta)^3}{2}. \tag{5.7}$$

Since we assume $t < 2(d\delta)^{-3}$, it follows that

$$\binom{t}{2}\beta(\Phi(\vec{\theta}))^2 \leqslant \binom{t}{2}\frac{(d\delta)^6}{4} < \frac{1}{2},$$

whence $\alpha(\Phi(\vec{\theta}),t) > 1/2$. In particular, since $\alpha(\Phi(\vec{\theta}),t) > 0$ and since $\text{Re}(\Phi(\vec{\theta})) > 0$ by (4.4), we can use Lemma 5.2. From (5.1) and (5.7) we have

$$\text{Re}(\Phi(\vec{\theta})^t) \leqslant [\text{Re}(\Phi(\vec{\theta}))]^t \left(1 + \frac{(d\delta)^6}{4}\right)^{t/2} \tag{5.8}$$

and from (5.2) we have

$$\text{Re}(\Phi(\vec{\theta})^t) \geqslant [\text{Re}(\Phi(\vec{\theta}))]^t (1 + [\beta(\Phi(\vec{\theta}))]^2)^{t/2} \left(1 + t^2\left[\frac{\beta(\Phi(\vec{\theta}))}{\alpha(\Phi(\vec{\theta}),t)}\right]^2\right)^{-1/2}$$

$$\geqslant [\text{Re}(\Phi(\vec{\theta}))]^t \left(1 + t^2\left[\frac{\beta(\Phi(\vec{\theta}))}{\alpha(\Phi(\vec{\theta}),t)}\right]^2\right)^{-1/2}. \tag{5.9}$$

Since $\alpha(\Phi(\vec{\theta}),t) \geqslant 1/2$ and $\beta(\Phi(\vec{\theta})) \leqslant (d\delta)^3/2$, it follows that

$$\left[\frac{\beta(\Phi(\vec{\theta}))}{\alpha(\Phi(\vec{\theta}),t)}\right]^2 \leqslant (d\delta)^6,$$

so (5.8) and (5.9) combine to give

$$[1 + t^2(d\delta)^6]^{-1/2} \int_{B_\delta(\vec{0})} [\text{Re}(\Phi(\vec{\theta}))]^t \, d\vec{\theta}$$

$$\leqslant \int_{B_\delta(\vec{0})} \text{Re}(\Phi(\vec{\theta})^t) \, d\vec{\theta}$$

$$\leqslant \left[1 + \frac{(d\delta)^6}{4}\right]^{t/2} \int_{B_\delta(\vec{0})} [\text{Re}(\Phi(\vec{\theta}))]^t \, d\vec{\theta}. \tag{5.10}$$

We recall from Lemma 4.5 that there exists a function $\varepsilon : B_\delta(\vec{0}) \to \mathbb{R}$ such that, for $\vec{\theta} \in B_\delta(\vec{0})$,

$$[\text{Re}(\Phi(\vec{\theta}))]^t = e^{-\frac{t}{2}\vec{\theta}^T M\vec{\theta}}(1 + \varepsilon(\vec{\theta}))^t$$

and

$$|\varepsilon(\vec{\theta})| < \frac{1}{6}(d\delta)^4 e^{\frac{1}{2}(d\delta)^2}.$$

Because $d\delta < 1$, it follows that $e^{\frac{1}{2}(d\delta)^2} < 2$, so

$$|\varepsilon(\vec{\theta})| < \frac{1}{3}(d\delta)^4.$$

Therefore,

$$e^{-\frac{t}{2}\vec{\theta}^T M \vec{\theta}}\left[1 - \frac{1}{3}(d\delta)^4\right]^t \leqslant [\text{Re}(\Phi(\vec{\theta}))]^t \leqslant e^{-\frac{t}{2}\vec{\theta}^T M \vec{\theta}}\left[1 + \frac{1}{3}(d\delta)^4\right]^t$$

and substituting these bounds into (5.10) gives

$$[1 + t^2(d\delta)^6]^{-1/2}\left[1 - \frac{1}{3}(d\delta)^4\right]^t \int_{B_\delta(\vec{0})} e^{-\frac{t}{2}\vec{\theta}^T M \vec{\theta}}\,\mathrm{d}\vec{\theta}$$

$$\leqslant \int_{B_\delta(\vec{0})} \text{Re}(\Phi(\vec{\theta})^t)\,\mathrm{d}\vec{\theta}$$

$$\leqslant \left[1 + \frac{(d\delta)^6}{4}\right]^{t/2}\left[1 + \frac{1}{3}(d\delta)^4\right]^t \int_{B_\delta(\vec{0})} e^{-\frac{t}{2}\vec{\theta}^T M \vec{\theta}}\,\mathrm{d}\vec{\theta}. \tag{5.11}$$

To verify (5.5) and thus complete the proof, by (5.11) and (5.6) it suffices to show that

$$[1 - e^{-\frac{t}{2}(D_1\delta)^2}]^{d/2}\left(\frac{2\pi}{t}\right)^{d/2} g^{\frac{g}{2}\binom{k}{2}}$$

$$\leqslant \int_{B_\delta(\vec{0})} e^{-\frac{t}{2}\vec{\theta}^T M \vec{\theta}}\,\mathrm{d}\vec{\theta}$$

$$\leqslant [1 - e^{-t(D_2\delta)^2}]^{d/2}\left(\frac{2\pi}{t}\right)^{d/2} g^{\frac{g}{2}\binom{k}{2}} \tag{5.12}$$

so we now turn our attention to the integral in the middle.

We recall from Proposition 5.1 that there is a symmetric matrix P such that $P^2 = M$. Since $B_\delta(\vec{0}) = [-\delta, \delta]^d$, we have

$$\int_{B_\delta(\vec{0})} e^{-\frac{t}{2}\vec{\theta}^T M \vec{\theta}}\,\mathrm{d}\vec{\theta} = \int_{[-\delta,\delta]^d} e^{-\frac{1}{2}(\sqrt{t}P\vec{\theta})^T(\sqrt{t}P\vec{\theta})}\,\mathrm{d}\vec{\theta},$$

so if we apply a change of variables with $\vec{\eta} = \sqrt{t}P\vec{\theta}$, we have

$$\int_{B_\delta(\vec{0})} e^{-\frac{t}{2}\vec{\theta}^T M \vec{\theta}}\,\mathrm{d}\vec{\theta} = \frac{1}{t^{d/2}\det(P)}\int_{P[-\sqrt{t}\delta,\sqrt{t}\delta]^d} e^{-\frac{1}{2}\vec{\eta}^T\vec{\eta}}\,\mathrm{d}\vec{\eta}.$$

Since the integrand is positive, Proposition 5.1 also implies that

$$\frac{1}{t^{d/2}\det(P)}\int_{[-D_1\sqrt{t}\delta,D_1\sqrt{t}\delta]^d} e^{-\frac{1}{2}\vec{\eta}^T\vec{\eta}}\,\mathrm{d}\vec{\eta}$$

$$< \int_{B_\delta(\vec{0})} e^{-\frac{t}{2}\vec{\theta}^T M \vec{\theta}}\,\mathrm{d}\vec{\theta}$$

$$< \frac{1}{t^{d/2}\det(P)}\int_{[-D_2\sqrt{t}\delta,D_2\sqrt{t}\delta]^d} e^{-\frac{1}{2}\vec{\eta}^T\vec{\eta}}\,\mathrm{d}\vec{\eta}.$$

Because

$$\vec{\eta}^T\vec{\eta} = \sum \eta_{\{i,j\},a}^2,$$

we can regard the integrals in the lower and upper bounds as the product of d integrals of the form $\int e^{-\frac{1}{2}x^2}\,dx$. Using the estimates in Lemma 5.3 gives

$$\frac{1}{t^{d/2}\det(P)}[2\pi(1-e^{-\frac{1}{2}t(D_1\delta)^2})]^{d/2}$$

$$< \int_{B_\delta(\vec{0})} e^{-\frac{t}{2}\vec{\theta}^T M\vec{\theta}}\,d\vec{\theta}$$

$$< \frac{1}{t^{d/2}\det(P)}[2\pi(1-e^{-t(D_2\delta)^2})]^{d/2}$$

and since Proposition 4.4 shows that $\det(P)=\sqrt{\det(M)}=g^{-\frac{1}{2}g\binom{k}{2}}$, this yields (5.12) and completes the proof. □

We pause to remark that if g is even and t/g is odd, then Propositions 2.8 and 3.3 show that

$$\mathbb{P}(X_t=\vec{0}) \leqslant \exp\left(-\frac{11}{192}g^{-k}t\delta^2\right).$$

This hints at the fact that there are no difference matrices over cyclic groups with such parameters [6]; however, as stated, it does not actually constitute a proof of that result (even asymptotically), since the g^{kt} factor found in (1.2) causes the product

$$g^{kt}\exp\left(-\frac{11}{192}g^{-k}t\delta^2\right)$$

not to converge to 0. This result could potentially be obtained by tightening the error term estimate, but such endeavours are not necessary for our purposes.

Proof of Theorem 1.1. Let $g \geqslant 2$ be fixed. Suppose (k,t) is a sequence of ordered pairs such that $k \geqslant 3$, $t \to \infty$, each value of t is a positive integer multiple of g, and there exists some $\varepsilon_0 > 0$ so that

$$k < \left(\frac{1}{6}-\varepsilon_0\right)\frac{\log(t)}{\log(g)}$$

for all pairs in the sequence. We recall that if g is even and $\lambda = t/g$ is odd, then there is no $(g,k;\lambda)$-difference matrix over $\mathbb{Z}_g$ [6]; we therefore assume that g is odd or that λ is even for every pair in the sequence. We define the sequences δ,ε by

$$k = \left(\frac{1}{6}-\varepsilon\right)\frac{\log(t)}{\log(g)}, \qquad \delta = g^{-5k/(2-12\varepsilon)}.$$

We remark that these definitions imply that $\delta = t^{-5/12}$ and that our assumptions on the sequences (k,t) imply that $\varepsilon \in (\varepsilon_0, 1/6)$. With these definitions, we have three goals: we wish to argue that the hypotheses of Theorem 5.4 hold for all but finitely many pairs (k,t), that the bracketed components inside U and L tend to 1, and that in (5.3) and (5.4) the exponential error terms become small in comparison to the coefficients on the U and L terms.

First, we verify that $\delta < \frac{8}{5}g^{-k-3}k^{-2}$. If the sequence of k values is bounded above, then $\delta = t^{-5/12}$ will certainly be less than $g^{-k-3}k^{-2}$ for sufficiently large t. If the sequence of k values is

unbounded, then we note that

$$\delta = g^{-5k/(2-12\varepsilon)} = g^{-(k+(3+12\varepsilon)/(2-12\varepsilon)k)} < g^{-(k+(3/2)k+6\varepsilon_0 k)},$$

and because $k > 2$ we have $\delta < g^{-k-3}g^{-6\varepsilon_0 k}$, which is smaller than

$$g^{-k-3}\left(\frac{8}{5}k^{-2}\right)$$

for sufficiently large k. Also, since $k < \log(t)$, we see that

$$t(d\delta)^3 = t(g-1)^3 k^3 (k-1)^3 \delta^3/8 < tg^3[\log(t)]^6 t^{-5/4}/8,$$

which is less than 2 for sufficiently large t. Hence, we have verified that the hypotheses of Theorem 5.4 hold for all but finitely many pairs (k,t).

Next, we consider the factors of L and U. Since

$$t^2(d\delta)^6 = (g-1)^6 \binom{k}{2}^6 t^{-1/2} < (g-1)^6 (\log(t))^{12} t^{-1/2} \to 0$$

as $t \to \infty$, we have $[1 + t^2(d\delta)^2]^{-1/2} \to 1$. On the other hand, because

$$t(d\delta)^6 < (g-1)^6 (\log(t))^{12} t^{-3/2} \to 0,$$

it follows that $[1 + \frac{1}{4}(d\delta)^6]^{t/2} \to 1$. Similarly, since

$$t(d\delta)^4 < (g-1)^4 (\log(t))^8 t^{-2/3} \to 0,$$

the $[1 - \frac{1}{3}(d\delta)^4]^t$ and $[1 + \frac{1}{3}(d\delta)^4]^t$ terms each tend to 1. Finally, since the constants D_1 and D_2 depend only upon g, we see that

$$de^{-\frac{t}{2}(D_1\delta)^2} \leqslant g(\log(t))^2 e^{-\frac{1}{2}D_1^2 t^{1/6}} \to 0, \quad de^{-t(D_2\delta)^2} \leqslant g(\log(t))^2 e^{-D_2^2 t^{1/6}} \to 0$$

whence $[1 - e^{\frac{t}{2}(D_1\delta)^2}]^{d/2}$ and $[1 - e^{-t(D_2\delta)^2}]^{d/2}$ both tend to 1. Thus, as $t \to \infty$, $L(g,k,t,t^{-5/12})$ and $U(g,k,t,t^{-5/12})$ both converge to 1.

Finally, from (5.3) and (5.4), we see that

$$\liminf_{t\to\infty} \frac{\mathbb{P}(X_t = \vec{0})}{\left[\frac{g^{\frac{g}{2}\binom{k}{2}+\binom{k-1}{2}}}{(2\pi t)^{d/2}}\right]} \geqslant \lim_{t\to\infty} \left[L(g,k,t,t^{-5/12}) - \frac{e^{-\frac{11}{192}g^{-k}t\delta^2}}{\left[\frac{g^{\frac{g}{2}\binom{k}{2}+\binom{k-1}{2}}}{(2\pi t)^{d/2}}\right]}\right],$$

$$\limsup_{t\to\infty} \frac{\mathbb{P}(X_t = \vec{0})}{\left[\frac{g^{\frac{g}{2}\binom{k}{2}+\binom{k-1}{2}}}{(2\pi t)^{d/2}}\right]} \leqslant \lim_{t\to\infty} \left[U(g,k,t,t^{-5/12}) + \frac{e^{-\frac{11}{192}g^{-k}t\delta^2}}{\left[\frac{g^{\frac{g}{2}\binom{k}{2}+\binom{k-1}{2}}}{(2\pi t)^{d/2}}\right]}\right]$$

and because $L, U \to 1$, if we can show that

$$\frac{e^{-\frac{11}{192}g^{-k}t\delta^2}}{\left[\frac{g^{\frac{g}{2}\binom{k}{2}+\binom{k-1}{2}}}{(2\pi t)^{d/2}}\right]} \to 0 \tag{5.13}$$

then combining the above with (1.2) and substituting $t = \lambda g$ will complete the proof. If the sequence of k values is bounded above, then the term in (5.13) is at most

$$C_1 e^{-C_2 t \delta^2} t^{C_3} = C_1 e^{-C_2 t^{1/6}} t^{C_3}$$

for positive constants C_1, C_2, C_3; consequently, this term tends to 0 as $t \to \infty$. On the other hand, suppose that $k \to \infty$; in particular, assume that $k \geqslant 4$. Then

$$\frac{(2\pi)^{d/2}}{g^{\frac{g}{2}\binom{k}{2} + \binom{k-1}{2}}} = \left(\frac{(2\pi)^{(g-1)/2}}{g^{\frac{g}{2} + \frac{k-2}{k}}} \right)^{\binom{k}{2}} \leqslant \left(\frac{(2\pi)^{g-1}}{g^{g+1}} \right)^{\binom{k}{2}/2}.$$

If $g \geqslant 7$, then clearly $(2\pi)^{g-1}/g^{g+1} < 1$, and it can be easily verified that this also holds for $g = 2, \ldots, 6$. Hence, the fraction in (5.13) is at most

$$t^{d/2} \exp\left(-\frac{11}{192} g^{-k} t \delta^2 \right),$$

and since $g^k < t^{1/6 - \varepsilon_0}$, the aforementioned fraction is at most

$$\exp\left(\frac{d}{2} \log(t) - \frac{11}{192} t^{\varepsilon_0 - 1/6} t^{1/6} \right) < \exp\left((\log(t))^3/2 - \frac{11}{192} t^{\varepsilon_0} \right),$$

which tends to 0 as $t \to \infty$. This verifies (5.13), as desired. $\qquad \square$

6. Conclusion

By adopting the perspective and tactics of random walks to the problem of difference matrices over cyclic groups, we have developed a formula for the asymptotic number of such matrices as the number of columns (or equivalently, the row inner product λ) grows large. There are a number of related projects which require further efforts that we leave for future work. One immediate question is what occurs when the underlying cyclic group is replaced with an arbitrary group; most of the changes required to the proof would be to the latter half of Section 2.

We also note the relationship between difference matrices and orthogonal arrays, as defined in [4, III.6.1]. Any $(g, k; \lambda)$-difference matrix over $\mathbb{Z}_g$ can be used to construct an $\mathrm{OA}_\lambda(k, g)$ [4, Remark VI.17.7]; conversely, any $\mathrm{OA}_\lambda(k, g)$ can be viewed as a $(g, k; \lambda g)$-difference matrix over an arbitrary group G of order g [4, Theorem VI.17.10]. These facts intertwine the number of orthogonal arrays and the number of difference matrices in such a way that one can obtain crude estimates on the number of orthogonal arrays. However, these estimates do not yield the exact asymptotics for the number of such arrays, which could be obtained by reinventing the Fourier analysis in this work for those designs. (We remark that Kuperberg, Lovett, and Peled have already completed this analysis for orthogonal arrays with no repeated columns [13].)

Another common direction for this type of work is to find bounds that guarantee the existence of $(g, k; \lambda)$-difference matrices over $\mathbb{Z}_g$. For any suitable configuration of g, k, t, δ such that the expression for $\mathbb{P}(X_t = \vec{0})$ in (5.4) is positive, the existence of a $(g, k; t/g)$-difference matrix over $\mathbb{Z}_g$ is assured. The estimates provided in this work seem not to be sufficient to provide a non-trivial bound of this type, but many of these estimates could be greatly improved with some effort, perhaps to an extent that would yield a non-trivial lower bound on the probability. Existence questions of difference matrices remain an active area of research, and while any bounds on parameters obtained in this way would likely be far from optimal, they may nonetheless be novel.

In this work, we have also ignored the question of equivalence classes of difference matrices. Commonly, two difference matrices are regarded as equivalent if one can be obtained from another by exchanging rows or columns, rotating an entire row or column by a group element, or applying an automorphism of the underlying group to every element in the matrix. Such actions are difficult to capture with the random walk enumeration scheme described in this paper, and the questions of counting the raw number of matrices and counting the equivalence classes are non-trivially different when rows or columns can be repeated.

Finally, we remark that enumeration and existence results are each perhaps most interesting in the case when λ is small. To illustrate, we recall that in order for a $(g, 3; 1)$-difference matrix over $\mathbb{Z}_g$ to exist, it is necessary for g to be odd [6]; however, this condition is also sufficient [9]. The general question of the existence of a $(g, k; 1)$-difference matrix over $\mathbb{Z}_g$ with g odd and $k \geqslant 4$ remains open. For example, it is known that both a $(5, 4; 1)$-difference matrix and a $(7, 4; 1)$-difference matrix over cyclic groups exist [7], and a computer search has shown that a $(9, 4; 1)$-difference matrix over $\mathbb{Z}_9$ does not [9], which illustrates that the $k = 4$ case is not as tidy as the $k = 3$ case. In principle, combining (1.3) with (1.2) shows that one can obtain the the exact number of $(g, k; \lambda)$-difference matrices over $\mathbb{Z}_g$ by evaluating an integral, so one might hope to obtain interesting results about relatively small λ by minimizing all the error terms in the preceding Fourier analysis. Analogously, one might hope to use this sort of tactic to resolve the Hadamard conjecture, or to count the number of Steiner triple system incidence matrices. Of course, such efforts have thus far fallen short of those lofty goals, but interesting existence and enumeration results of this nature have been derived for other combinatorial designs. While it is likely too much to hope that this Fourier analysis can address the enumeration or existence of $(g, k; \lambda)$-difference matrices over $\mathbb{Z}_g$ when $k = g\lambda$, perhaps it can yield results when λ grows slowly as a function of k.

Acknowledgements

References

[1] Barvinok, A. and Hartigan, J. A. (2012) An asymptotic formula for the number of non-negative integer matrices with prescribed row and column sums. *Trans. Amer. Math. Soc.* **364** 4323–4368.

[2] Billingsley, P. (1995) *Probability and Measure*, third edition, Wiley.

[3] Canfield, E. R. and McKay, B. D. (2011) Asymptotic enumeration of integer matrices with large equal row and column sums. *Combinatorica* **30** 655–680.

[4] Colbourn, C. J. and Dinitz, J. H., eds (2006) *Handbook of Combinatorial Designs*, second edition, Discrete Mathematics and its Applications, CRC Press.

[5] De Launey, W. and Levin, D. A. (2010) A Fourier-analytic approach to counting partial Hadamard matrices. *Cryptogr. Commun.* **2** 307–334.

[6] Drake, D. A. (1979) Partial λ-geometries and generalized Hadamard matrices over groups. *Canad. J. Math.* **31** 617–627.

[7] Evans, A. B. (2002) On orthogonal orthomorphisms of cyclic and non-abelian groups. *Discrete Math.* **243** 229–233.

[8] Gao, Z., McKay, B. D. and Wang, X. (2000) Asymptotic enumeration of tournaments with a given score sequence containing a specified digraph. *Random Struct. Alg.* **16** 47–57.

[9] Ge, G. (2005) On $(g, 4; 1)$-difference matrices. *Discrete Math.* **301** 164–174.

[10] Harville, D. A. (1997) *Matrix Algebra from a Statistician's Perspective*, Springer.

[11] Isaev, M. (2011) Asymptotic behaviour of the number of Eulerian circuits. *Electron. J. Combin.* **18** #219.

[12] Jungnickel, D. (1979) On difference matrices, resolvable transversal designs and generalized Hadamard matrices. *Math. Z.* **167** 49–60.

[13] Kuperberg, G., Lovett, S. and Peled, R. (2013) Probabilistic existence of regular combinatorial structures. https://arxiv.org/pdf/1302.4295

[14] McKay, B. D. and Wormald, N. C. (1990) Asymptotic enumeration by degree sequence of graphs of high degree. *European J. Combin.* **11** 565–580.

[15] Montgomery, A. (2013) Topics in random walks. PhD thesis, University of Oregon.

[16] Spitzer, F. (1976) *Principles of Random Walk*, second edition, Vol. 34 of Graduate Texts in Mathematics, Springer.