Probabilistic existence of rigid combinatorial structures

(extended abstract version)

Greg Kuperberg * Shachar Lovett † Ron Peled ‡

May 28, 2012

Abstract

We show the existence of rigid combinatorial objects which previously were not known to exist. Specifically, for a wide range of the underlying parameters, we show the existence of non-trivial orthogonal arrays, t-designs, and t-wise permutations. In all cases, the sizes of the objects are optimal up to polynomial overhead. The proof of existence is probabilistic. We show that a randomly chosen such object has the required properties with positive yet tiny probability. The main technical ingredient is a special local central limit theorem for suitable lattice random walks with finitely many steps.

1 Introduction

We introduce a new framework for establishing the existence of rigid combinatorial structures, such as orthogonal arrays, t-designs and t-wise permutations. Let B be a finite set and let V be a vector space of functions from B to the rational numbers $\mathbb{Q}$. We study when there is a small subset $T \subset B$ satisfying

$$\frac{1}{|T|}\sum_{t\in T} f(t) = \frac{1}{|B|}\sum_{b\in B} f(b) \quad \text{for all } f \text{ in } V. \tag{1}$$

In probabilistic terminology, equation (1) means that if t is a uniformly random element in T and b is a uniformly random element in B then

$$\mathbb{E}[f(t)] = \mathbb{E}[f(b)] \quad \text{for all } f \text{ in } V, \tag{2}$$

where $\mathbb{E}$ denotes expectation. Of course, (1) holds trivially when $T = B$. Our goal is to find conditions on B and V that yield a small subset T that satisfies (1), where in our situations, small will mean polynomial in the dimension of V. (In many natural problems one might encounter a function space V over $\mathbb{R}$ or $\mathbb{C}$ instead. However, since (1) is a rational equation, we can always reduce to the case of rational vector spaces.)

Our main theorem, Theorem 2.1, gives sufficient conditions for the existence of a small subset T satisfying (1). We apply the theorem to establish results in three interesting cases of the general framework: orthogonal arrays, t-designs, and t-wise permutations. These are detailed in the next sections. Our methods solve an open problem, whether there exist non-trivial t-wise permutations for every t. They strengthen Teirlinck's theorem [Tei87], which was the first theorem to show the existence of t-designs for every t. And they improve existence results for orthogonal arrays, when the size of the alphabet is divisible by many

*University of California, Davis. E-mail: greg@math.ucdavis.edu. Supported by NSF grant CCF-1013079.

†Institute for Advanced Study. E-mail: slovett@math.ias.edu. Supported by NSF grant DMS-0835373.

‡Tel Aviv University, Israel. E-mail: peledron@post.tau.ac.il. Supported by an ISF grant and an IRG grant.

distinct primes. Moreover, in all three cases considered, we show the existence of a structure whose size is optimal up to polynomial overhead.

Our approach to the problem is via probabilistic arguments. In essence, we prove that a random subset of B satisfies equation (1) with positive, albeit tiny, probability. Thus our method is one of the few known methods for showing existence of rare objects. This class includes such other methods as the Lovász local lemma [EL75] and Spencer's "six deviations suffice" method [Spe85]. However, our method does not rely on these previous approaches. Instead, our technical ingredient is a special version of the (multi-dimensional) local central limit theorem with only finitely many available steps. Since only finitely many steps are available, and since we can only gain access to more steps by increasing the dimension of the random walk, we cannot use any "off the shelf" local central limit theorem, not even one enhanced by a Berry-Esseen-type estimate of the rate of convergence. Instead, we prove the local central limit theorem that we need directly using Fourier analysis. Section 1.4 gives an overview of our approach.

We also mention that efficient randomized algorithm versions of the Lovász local lemma [Mos09, MT10] and Spencer's method [Ban10] have recently been found. Relative to these new algorithms, the objects that they produce are no longer rare. Our method is the only one that we know that shows the existence of rare combinatorial structures, which are still rare relative to any known, efficient, randomized algorithm.

1.1 Orthogonal arrays

A subset $T \subset [q]^n$ is an *orthogonal array of alphabet size q, length n and strength t* if it yields all strings of length t with equal frequency if restricted to any t coordinates. In other words, for any distinct indices $i_1, \ldots, i_t \in [n]$ and any (not necessarily distinct) values $v_1, \ldots, v_t \in [q]$,

$$|\{x \in T : x_{i_1} = v_1, \ldots, x_{i_t} = v_t\}| = q^{-t}|T|.$$

Equivalently, choosing $x = (x_1, \ldots, x_n) \in T$ uniformly, the distribution of $x \in [q]^n$ is t-wise independent. For an introduction to orthogonal arrays see [HSS99].

Orthogonal arrays fit into our general framework as follows. We take B to be $[q]^n$ and V to be the space spanned by all functions of the form

$$f_{(I,v)}(x_1, \ldots, x_n) = \begin{cases} 1 & x_i = v_i \text{ for all } i \in I \\ 0 & \text{Otherwise} \end{cases}, \tag{3}$$

with $I \subset [n]$ a subset of size t and $v \in [q]^I$. With this choice, a subset $T \subset B$ satisfying (1) is precisely an orthogonal array of alphabet size q, length n and strength t.

It is well known that if $T \subset [q]^n$ is t-wise independent then $|T| \geq \left(\frac{cqn}{t}\right)^{t/2}$ for some universal constant $c > 0$ (see, e.g., [Rao73]). Matching constructions of size $|T| \leq q^{ct}\left(\frac{n}{t}\right)^{c_q t}$ are known, however, as these rely on finite field properties the constant c_q generally tends to infinity with the number of prime factors of q. Our technique provides the first upper bound on the size of orthogonal arrays in which the constant in the exponent is independent of q.

Theorem 1.1 (Existence of orthogonal arrays)**.** *For all integers $q \geq 2$, $n \geq 1$ and $1 \leq t \leq n$ there exists an orthogonal array T of alphabet size q, length n and strength t satisfying $|T| \leq (qn)^{ct}$ for some universal constant $c > 0$.*

1.2 Designs

A (simple) t-(v,k,λ) *design* is a family of distinct subsets of $[v]$, where each set is of size k, such that each t elements belong to exactly λ sets. In other words, denoting by $\binom{[v]}{k}$ the family of all subsets of $[v]$ of size k, a

set $T \subset \binom{[v]}{k}$ is a t-design if for any distinct elements $i_1, \ldots, i_t \in [v]$,

$$|\{s \in T : i_1, \ldots, i_t \in s\}| = \frac{\binom{k}{t}}{\binom{v}{t}}|T| = \lambda. \tag{4}$$

For an introduction to combinatorial designs see [CD07].

Our general framework includes t-designs as follows. We take B to be $\binom{[v]}{k}$ and V to be the space spanned by all functions of the form

$$f_a(b) = \begin{cases} 1 & a \subset b \\ 0 & \text{Otherwise} \end{cases}, \tag{5}$$

with $a \in \binom{[v]}{t}$. With this choice, a subset $T \subset B$ satisfying (1) is precisely a simple $t-(v,k,\lambda)$ design.

Although t-designs have been investigated for many years, the basic question of existence of a design for a given set of parameters t, v, k and λ remains mostly unanswered unless t is quite small. The case $t = 2$ is known as a block design and much more is known about it than for larger t. Explicit constructions of t-designs for $t \geq 3$ are known for various specific constant settings of the parameters (e.g. 5-$(12,6,1)$ design). The breakthrough result of Teirlinck [Tei87] was the first to establish the existence of non-trivial t-designs for $t \geq 7$. In Teirlinck's construction, $k = t+1$ and v satisfies congruences that grow very quickly as a function of t. Other sporadic and infinite examples have been found since then (see [CD07] or [Mag09] and the references within), however, the set of parameters which they cover is still very sparse. Moreover, it follows from (4) that any $t-(v,k,\lambda)$ design T has size $|T| = \lambda\binom{v}{t}/\binom{k}{t} \geq (v/k)^t$. Even when existence has been shown, the designs obtained are often inefficient in the sense that their size is far from this lower bound. One of the main results of our work is to establish the existence of efficient t-designs for a wide range of parameters.

Theorem 1.2 (Existence of t-designs)**.** *For all integers $v \geq 1$, $1 \leq t \leq v$ and $t \leq k \leq v$ there exists a t-(v,k,λ) design whose size is at most v^{ct} for some universal constant $c > 0$.*

1.3 Permutations

A family of permutations $T \subset S_n$ is called a *t-wise permutation* if its action on any t-tuple of elements is uniform. In other words, for any distinct elements $i_1, \ldots, i_t \in [n]$ and distinct elements $j_1, \ldots, j_t \in [n]$,

$$|\{\pi \in T : \pi(i_1) = j_1, \ldots, \pi(i_t) = j_t\}| = \frac{1}{n(n-1)\cdots(n-t+1)}|T|. \tag{6}$$

Our general framework includes t-wise permutations as follows. We take $B = S_n$ and V to be the space spanned by all functions of the form

$$f_{(i,j)}(b) = \begin{cases} 1 & b(i_1) = j_1, \ldots, b(i_t) = j_t \\ 0 & \text{Otherwise} \end{cases},$$

where $i = (i_1, \ldots, i_t)$ and $j = (j_1, \ldots, j_t)$ are t-tuples of distinct elements in $[n]$. With this choice, a subset $T \subset B$ satisfying (1) is precisely a t-wise permutation.

Constructions of families of t-wise permutations are known only for $t = 1, 2, 3$: the group of cyclic shifts $x \mapsto x + a$ modulo n is a 1-wise permutation; the group of invertible affine transformations $x \mapsto ax + b$ over a finite field $\mathbb{F}$ yields a 2-wise permutation; and the group of Möbius transformations $x \mapsto (ax+b)/(cx+d)$ with $ad - bc = 1$ over the projective line $\mathbb{F} \cup \{\infty\}$ yields a 3-wise permutation. For $t \geq 4$ (and n large enough), however, no t-wise permutation is known, other then the full symmetric group S_n and the alternating group A_n [KNR05, AL11]. In fact, it is known (c.f., e.g., [Cam95], Theorem 5.2) that for $n \geq 25$ and $t \geq 4$ there

are no other *subgroups* of S_n which form a t-wise permutation. (On other words, there are no other t-transitive subgroups of S_n for $t \geq 4$ and $n \geq 25$.) One of our main results is to show existence of small t-wise permutations for all t.

Theorem 1.3 (Existence of t-wise permutations)**.** *For all integers $n \geq 1$ and $1 \leq t \leq n$ there exists a t-wise permutation $T \subset S_n$ satisfying $|T| \leq \exp(t^c)n^{ct}$ for some universal constant $c > 0$.*

It is clear from the definition (6) above that any t-wise permutation T must satisfy $|T| \geq n(n-1)\cdots(n-t+1) = n^{\Omega(t)}$. Thus, for fixed t, the t-wise permutations we exhibit are of optimal size up to polynomial overhead. For t growing with n these t-wise permutations may be larger, but still no larger than n^{t^c} for some universal constant $c > 0$.

1.4 Proof overview

The idea of our approach is as follows. Let T be a random multiset of B of some fixed size N chosen by sampling B uniformly and independently N times (with replacement). Let $(\phi_a)_{a\in A}$ be a spanning set of integer-valued functions for V (where A is some finite index set). Observe that T satisfies (1) if and only if

$$\sum_{t\in T} \phi_a(t) = \frac{N}{|B|} \sum_{b\in B} \phi_a(b) = \mathbb{E}\left[\sum_{t\in T} \phi_a(t)\right] \quad \text{for all } a \text{ in } A. \tag{7}$$

Thus defining an integer-valued random variable

$$X_a := \sum_{t\in T} \phi_a(t)$$

and $X := (X_a)_{a\in A} \in \mathbb{Z}^A$ we see that existence of a subset of size N satisfying (1) will follow if we can show that $\mathbb{P}[X = \mathbb{E}[X]] > 0$. To this end we examine more closely the distribution of X. Let $t_1, \ldots, t_N$ be the random elements chosen in forming T. The spanning set $(\phi_a)_{a\in A}$ defines a mapping $\phi : B \to \mathbb{Z}^A$ by the trivial

$$\phi(b)_a := \phi_a(b).$$

Observe that our choice of random model implies that the vectors $(\phi(t_i))_{i\in[N]}$ are independent and identically distributed. Hence,

$$X = \sum_i \phi(t_i) \tag{8}$$

may be viewed as the end position of an N-step random walk in the lattice $\mathbb{Z}^{|A|}$. Thus we may hope that if N is sufficiently large, then X has an approximately (multi-dimensional) Gaussian distribution by the central limit theorem. If the relevant local central limit theorem holds as well, then the probability $\mathbb{P}[X = x]$ also satisfies a Gaussian approximation. In particular, since a (non-degenerate) Gaussian always has positive density at its expectation, we could conclude that $\mathbb{P}[X = \mathbb{E}[X]] > 0$ as desired.

The above description is the essence of our approach. The main obstacle is, of course, pointed out in the last step. We must control the rate of convergence of the local central limit theorem well enough that the convergence error does not outweigh the probability density of the Gaussian distribution at $\mathbb{E}[X]$. Recall that the order of magnitude of such a density is typically $c^{-|A|}$ for some constant $c > 1$, and recall that $|A|$ is at least the dimension of V, which is the main parameter of our problem. So we indeed have very small probabilities. For this reason, and because we want convergence when N is only polynomial in the dimension of V, we were unable to use any standard local central limit theorem. Instead, we develop an ad hoc version using direct Fourier analysis.

In our proof of the main theorem, we modify the above description in one respect. It is technically more convenient to work with a slightly different probability model. Instead of choosing T as above, we set

$p := N/|B|$ and define T by taking each element of B into T independently with probability p. This has the benefit of guaranteeing that T is a proper set instead of a multiset. However, it has also the disadvantage that it does not guarantee that $|T| = N$. To remedy this, we assume that the space V contains the constant function $h(b) = 1$; or if not, we can add it to V at the minor cost of increasing the dimension of V by 1. With this assumption, we note that

$$\mathbb{E}\big[\sum_{t\in T} h(t)\big] = \mathbb{E}[|T|] = N.$$

Thus (7), or equivalently $X = \mathbb{E}[X]$, also implies that $|T| = N$ as required. Another disadvantage is that in this new probability model, the vector X is no longer a sum of identically distributed variables. However, since the summands in (8) are still independent, we can continue to use Fourier analysis methods in our proof.

We cannot expect there to always be a small subset T that satisfies (1). For instance, Alon and Vu [AV97] found a regular hypergraph with n vertices and $\approx n^{n/2}$ edges, with no regular sub-hypergraph. Here, the degree of a vertex is the number of hyperedges incident to it and a regular hypergraph is one in which the degrees of all vertices are equal. We may describe their example in our language by letting B be the set of edges of this hypergraph, A be its vertex set, and define $\phi : B \to \{0,1\}^A$ by letting $\phi(b)$ be the indicator function of the set of vertices incident to b. The result of [AV97] implies that while the vector $\sum_{b\in B}\phi(b)$ is constant, this property is not shared by $\sum_{t\in T}\phi(t)$ for any non-empty, proper subset $T \subset B$. Thus, we need to impose certain conditions on B and V, or equivalently on the map ϕ. We start by requiring certain divisibility, boundedness and symmetry assumptions.

Divisibility: N is such that $\frac{N}{|B|}\sum_{b\in B}\phi(b)$ is an integer vector. This property is clearly necessary for (7) to hold and is typically a mild restriction on N.

Boundedness: The entries of ϕ must be small. More precisely, $\max_{a\in A, b\in B}|\phi(b)_a|$ is bounded by a polynomial in $\dim V$, since our method requires N to be at least some polynomial in this maximum.

Symmetry: A *symmetry of* ϕ is a pair consisting of a permutation $\pi \in S_B$ and a linear transformation $\tau \in \mathrm{GL}(V)$ which satisfies $\phi(\pi(b)) = \tau(\phi(b))$ for all $b \in B$. The set of symmetries (π,τ) of ϕ is a subgroup of $S_B \times \mathrm{GL}(V)$. We require that the projection to B of the group of symmetries is transitive. In other words, that for any $b_1, b_2 \in B$ there exists a symmetry (π,τ) of ϕ satisfying $\pi(b_1) = b_2$.

It is not hard to verify that the third condition is intrinsic to the structure of V and does not depend on the specific choice of spanning set (ϕ_a). In our applications it follows easily from the overall symmetry of the setup.

However, we also have a fourth assumption which is more technical than the others. First, we require that $(\phi_a)_{a\in A}$ forms a basis of V. This implies that for any $a \in A$, we may express e_a, the unit vector with 1 at its a'th coordinate, as a linear combination of the form $\sum_{b\in B} c_b\phi(b)$. We call any such linear combination an *isolating combination for* a. We assume that for each $a \in A$, there are many isolating combinations supported on disjoint subsets of B. Moreover, we require the coefficients of these combinations to have small norm and to be rational with a small common denominator. This is the most difficult assumption to verify in our applications. Section 2 gives more details about all of these assumptions.

Our main theorem shows that these four conditions yield the existence of a small solution of (1).

Theorem (Main theorem - informal statement)**.** *Let B be a finite set and let V be a vector space of functions from B to $\mathbb{Q}$ which contains the constant functions. If there exists a basis $(\phi_a)_{a\in A}$ of V, consisting of integer-valued functions, which satisfies the boundedness, symmetry and isolation conditions above. Then there is a small subset $T \subset B$ such that*

$$\frac{1}{|T|}\sum_{t\in T} f(t) = \frac{1}{|T|}\sum_{b\in B} f(b)$$

for all f in V.

We note that the size $N = |T|$ of the subset obtained must satisfy the divisibility condition above. The existence theorems for orthogonal arrays, t-designs and t-wise permutations follow by showing that for the choice of B and V detailed in Sections 1.1 through 1.3 there exists a choice of basis $\{\phi_a\}$ and small N for which all four conditions above hold.

1.5 Related work

In the probabilistic formulation (2) of our problem we seek a small subset $T \subset B$ such that the uniform distribution over T simulates the uniform distribution over B with regards to certain tests. There are two ways to relax the problem to make its solution easier, and raise new questions regarding explicit solutions.

One relaxation is to allow a set T with a non-uniform distribution μ. For many practical applications of t-designs and t-wise permutations in statistics and computer science, but not quite every application, this relaxation is as good as the uniform question. The existence of a solution with small support is guaranteed by Carathéodory's theorem, using the fact that the constraints on μ are all linear equalities and inequalities. Moreover, such a solution can be found efficiently, as was shown by Karp and Papadimitriou [KP82] and in more general settings by Koller and Megiddo [KM94]. Alon and Lovett [AL11] give a strongly explicit analog of this in the case of t-wise permutations and more generally in the case of group actions.

A different relaxation is to require the uniform distribution on T to only approximately satisfy equation (2). Then it is trivial that a sufficiently large random subset $T \subset B$ satisfies the requirement with high probability, and the question is to find an explicit solution. For instance, we can relax the problem of t-wise permutations to *almost* t-wise permutations. For this variant an optimal solution (up to polynomial factors) was achieved by Kaplan, Naor and Reingold [KNR05], who gave a construction of such an almost t-wise permutation of size $n^{O(t)}$. Alternatively, one can start with the constant size expanding set of S_n given by Kassabov [Kas07] and take a random walk on it of length $O(t \log n)$.

1.6 Paper organization

We give a precise description of the general framework and our main theorem in Section 2. We apply it to show the existence of orthogonal arrays and t-designs in Section 3. The case of t-wise permutations requires a detour to the representation theory of the symmetric group, and we defer it to the full version of this paper. The proof of our main theorem is given in Section 4. We summarize and give some open problems in Section 5.

2 Main Theorem

Let B be a finite set and let V be a vector space of functions from B to $\mathbb{Q}$. We ask for conditions for the existence of a small set $T \subset B$ for which (1) holds. Our theorem uses the following notation.

For a basis $(\phi_a)_{a \in A}$ (where A is some finite index set) of V we define $\phi : B \to \mathbb{Z}^A$ by $\phi(b)_a = \phi_a(b)$. This definition is extended linearly to $\phi : \mathbb{Z}^B \to \mathbb{Z}^A$ by setting $\phi(\gamma) = \sum_{b \in B} \gamma_b \phi(b)$. In the same manner, a set $T \subset B$ is identified with its indicator vector so that $\phi(T) = \sum_{t \in T} \phi(t)$. Finally, we recall from Section 1.4 that a symmetry of ϕ is a pair $\pi \in S_B$ and $\tau \in \mathrm{GL}(V)$ such that $\phi(\pi(b)) = \tau(\phi(b))$ for all b in B. We now state formally our main theorem.

Theorem 2.1 (Main Theorem)**.** *Let B be a finite set and V be a vector space of functions from B to $\mathbb{Q}$ which contains the constant functions. Suppose that there exist integers $m, c_0 \geq 1$, real numbers $c_1, c_2, c_3 > 0$ and a basis $(\phi_a)_{a \in A}$ of V consisting of integer-valued functions such that:*

Divisibility: $\frac{c_0}{|B|}\phi(B)$ *is an integer vector.*

Boundedness: $\|\phi(b)\|_2 \le c_1$ *for all* $b \in B$.

Symmetry: *For each* $b_1, b_2 \in B$ *there exists a symmetry* (π, τ) *of* ϕ *such that* $\pi(b_1) = b_2$.

Isolation: *For any* $a \in A$ *there exist vectors* $\gamma_1, \ldots, \gamma_r \in \mathbb{Z}^B$ *for* $r \ge |B|/c_2$ *such that*

- $\phi(\gamma_i) = m \cdot e_a$ *for all* $i \in [r]$.
- *The vectors* $\gamma_1, \ldots, \gamma_r$ *have disjoint supports, where the support of a vector* $\gamma \in \mathbb{Z}^B$ *is the set of coordinates on which it is nonzero.*
- $\|\gamma_i\|_2 \le c_3$ *for all* $i \in [r]$.

Then there exists a subset $T \subset B$ *with* $|T| \le \mathrm{poly}(|A|, m, c_0, c_1, c_2, c_3)$ *such that*

$$\frac{1}{|T|}\sum_{t\in T} f(t) = \frac{1}{|B|}\sum_{b\in B} f(b) \quad \text{for all } f \text{ in } V.$$

We prove Theorem 2.1 in Section 4. A careful examination of the proof shows that we can choose $|T| = N$ for any $N \ge 1$ which satisfies the following constraints:

- $c_0 m$ divides N;
- $N \ge \Omega(1) \cdot \max(m^3, |A|^2 m^2 \log^2(|A| m c_0 c_1 c_2 c_3), |A|^6 c_1^6 c_2^3 c_3^6 \log^3(|A| m c_0 c_1 c_2 c_3))$;
- $N \le O(\sqrt{|B|})$.

Of course, if the parameters are so large so that the second and third conditions contradict each other, then our theorem remains trivially true by taking $T = B$.

3 Applications

In this section we apply our main theorem, Theorem 2.1, to prove the existence results for orthogonal arrays and t-designs, Theorems 1.1 and 1.2. The existence result for t-wise permutations, Theorem 1.3, is more complicated because it requires a discussion of the representation theory of the symmetric group. We defer it to the full version of this paper.

3.1 Orthogonal arrays

We use the choice of B and V described in Section 1.1 and recall the definition (3) of the functions $f_{(I,v)}$ of that section. We note that for every subset I we have $\sum_{v\in[q]^I} f_{(I,v)} \equiv 1$. Thus V contains the constant functions as Theorem 2.1 requires. We start by choosing a convenient basis for V of integer-valued functions. Recall that the alphabet is $[q] = \{1, \ldots, q\}$ and let $[q-1] = \{1, \ldots, q-1\}$ be all symbols other than q. Extend the definition (3) of $f_{(I,v)}$ to apply to all subsets I with $|I| \le t$ and $v \in [q]^I$. Here, we mean that $f_{(\emptyset,\emptyset)}$ is the constant function 1. Finally, let

$$A := \{(I, v) : |I| \le t, v \in [q-1]^{|I|}\}$$

and for $a = (I, v) \in A$ set $\phi_a := f_{(I,v)}$.

Claim 3.1. *The span of the functions* $\{\phi_a\}_{a\in A}$ *is* V.

Proof. Clearly $\phi_a \in V$ for all $a \in A$. To see that $\{\phi_a\}_{a\in A}$ spans V, we will show that any $f_{(I,v)}$ with $|I| \le t$ and $v \in [q]^I$ is spanned by $\{\phi_a\}_{a\in A}$. We do this by induction on the number of elements in v which are equal to q. First, if $v \in [q-1]^I$ then $f_{(I,v)} = \phi_{(I,v)}$. Otherwise, let $I = \{i_1, \ldots, i_r\}$ with $r \le t$, $v \in [q]^I$ and assume WLOG that $v_{i_1} = q$. Then

$$f_{(I,v)} = f_{(\{i_2,\ldots,i_r\},(v_{i_2},\ldots,v_{i_r}))} - \sum_{m=1}^{q-1} f_{(I,(m,v_{i_2},\ldots,v_{i_r}))}$$

and by induction, the right hand side belongs to the linear span of $\{\phi_a\}_{a\in A}$. □

Recall that $\phi : B \to \mathbb{Z}^A$ is defined as $\phi(b)_a = \phi_a(b)$. We now choose integers $m, c_0 \ge 1$ and real numbers $c_1, c_2, c_3 > 0$ such that the conditions of divisibility, boundedness, symmetry and isolation required by Theorem 2.1 are satisfied. First, let $a = (I,v) \in A$. Note that $\frac{1}{|B|}\phi(B)_a = q^{-|I|}$. Thus we set $c_0 = q^t$ so that $\frac{c_0}{|B|}\phi(B)$ is an integer vector. Second, we clearly have for any $b \in B$ that $\|\phi(b)\|_2^2 = \sum_{i=0}^{t} \binom{n}{i} \le (n+1)^t$. Hence we set $c_1 = (n+1)^{t/2}$.

Third, to witness the symmetry condition, fix $x \in [q]^n$ and consider the permutation $\pi \in S_B$ given by $\pi(b) = b + x \pmod q$. We need to show that there exists a linear map τ acting on V such that $\phi(\pi(b)) = \tau(\phi(b))$ for all $b \in B$. This holds since for $a = (I,v) \in A$ we have

$$\phi(\pi(b))_a = f_{I,v}(b + x \,(\mathrm{mod}\, q)) = f_{I,v-x \,(\mathrm{mod}\, q)}(b)$$

and $f_{I,v-x \,(\mathrm{mod}\, q)} \in V$ is in the linear span of $\{\phi_a\}_{a\in A}$ by Claim 3.1.

The fourth condition we need to verify is the existence of many disjoint isolation vectors for each $a \in A$. Note that this condition also implies that $\{\phi_a\}_{a\in A}$ is a basis for V. This is established in the following lemma.

Lemma 3.2. *Let $a \in A$. There exist disjoint vectors $\gamma_1, \ldots, \gamma_r \in \mathbb{Z}^B$ with $r \ge |B|/(q^t n^{2t})$ and $\|\gamma_i\|_2 \le 2^{3t/2} n^t$ such that $\phi(\gamma_i) = e_a$.*

We prove Lemma 3.2 in two steps. First we fix some notations. Let $K \subset [n]$ be of size $|K| \le t$, and let $K^c = [n] \setminus K$. For $x \in [q]^n$ let $x|_K \in [q]^K$ be the restriction of x to the coordinates of K. Abusing notation, we also think of $x|_K \in [q]^n$ by setting coordinates outside K to zero. Note that in this notation, $f_{I,v}(x) = \mathbf{1}\{x|_I = v\}$. We define the vector $\delta_{x,K} \in \mathbb{Z}^B$ as

$$\delta_{x,K} := \sum_{J\subseteq K} (-1)^{|K|-|J|} e_{x|_{J\cup K^c}},$$

where we recall that for $b \in B$, $e_b \in \{0,1\}^B$ is the corresponding unit vector. Note that if $K = \emptyset$ then $\delta_{x,\emptyset} = e_x$.

Claim 3.3. *Let $a = (I,v) \in A$. Then*

$$\phi(\delta_{x,K})_a = \begin{cases} 0 & \text{if } K \not\subseteq I \\ 0 & \text{if } a|_K \ne x|_K \\ 1 & \text{if } a|_K = x|_K \end{cases}$$

Proof. We compute the value of $\phi(\delta_{x,K})$ in coordinate $a = (I,v) \in A$. We have

$$\phi(\delta_{x,K})_a = \sum_{J\subseteq K} (-1)^{|K|-|J|} \mathbf{1}\{(x|_{J\cup K^c})|_I = v\}.$$

Suppose first that $K \not\subseteq I$. Then there exists $j \in K \setminus I$. Flipping the j-th element in J doesn't change the expression $\mathbf{1}\{(x|_{J\cup K^c})|_I = v\}$ and hence the alternating sign sum cancels. We thus assume from now on that $K \subseteq I$. We thus have

$$\mathbf{1}\{(x|_{J\cup K^c})|_I = v\} = \mathbf{1}\{x|_J = v|_K \text{ and } x|_{K^c\cap I} = v|_{K^c}\}.$$

This expression evaluates to 1 only if $J = K$ and $x|_I = v$. □

We next prove Lemma 3.2, showing that we can build many disjoint isolation vectors for any $a \in A$. The proof uses the vectors $\delta_{x,K}$ we just analyzed.

Proof of Lemma 3.2. Fix $a = (I, v)$. Let $x \in [q]^n$ be such that $x|_I = v$. We will construct a vector $\gamma_{x,I}$ such that $\phi(\gamma_{x,I}) = e_a$. We will do so by backward induction on $|I| \leq t$. If $|I| = t$ we take

$$\gamma_{x,I} := \delta_{x,I},$$

and if $|I| < t$ we construct recursively

$$\gamma_{x,I} := \delta_{x,I} - \sum_{K \supsetneq I, |K| \leq t, x_K \in [q-1]^K} \gamma_{x,K}.$$

It is easy to verify using Claim 3.3 that indeed $\phi(\gamma_{x,I}) = e_a$ as claimed. We further claim that $\|\gamma_{x,I}\|_2 \leq 2^{t/2}(2n)^{t-|I|}$. This clearly holds if $|I| = t$. If $|I| < t$ we bound by induction

$$\begin{aligned}
\|\gamma_{x,I}\|_2 &\leq \|\delta_{x,I}\|_2 + \sum_{k=|I|+1}^{t} \sum_{K \supset I, |K|=k} \|\gamma_{x,K}\|_2 \\
&\leq 2^{t/2} \left(1 + \sum_{k=|I|+1}^{t} \binom{n-|I|}{k-|I|} (2n)^{t-k} \right) \\
&\leq 2^{t/2} \left(1 + \sum_{k=|I|+1}^{t} n^{k-|T|} (2n)^{t-k} \right) \\
&\leq 2^{t/2} n^{t-|I|} \left(1 + \sum_{k=|I|+1}^{t} (2)^{t-k} \right) = 2^{t/2}(2n)^{t-|I|}.
\end{aligned}$$

To conclude, we need to show that by choosing different values for x such that $x|_I = v$ we can achieve many disjoint vectors which isolate a. The key observation is that $\gamma_{x,I}$ is supported on elements $b \in B$ whose hamming distance from x is at most t. Thus, if we choose $x_1, \ldots, x_r \in [q]^n$ such that $(x_i)|_I = v$ and such that the hamming distance between each pair x_i, x_j is at least $2t+1$, we get that $\gamma_{x_1,I}, \ldots, \gamma_{x_r,I}$ have disjoint supports. We can achieve $r \geq q^{n-t}/n^{2t}$ by a simple greedy process: choose $x_1, \ldots, x_r$ iteratively; after choosing x_i delete all elements in $[q]^n$ whose hamming distance from x_i is at most $2t$. Since the number of these elements is bounded by $\sum_{i=1}^{2t} \binom{n}{i} \leq n^{2t}$ the claim follows. □

We now have all the conditions to apply Theorem 2.1. We have $|A| = \sum_{i=0}^{t} \binom{n}{i}(q-1)^i \leq (q(n+1))^t, c_0 = q^t, c_1 = (n+1)^{t/2}, c_2 = q^t n^{2t}, c_3 = 2^{3t/2} n^t$ and $m = 1$. Hence we obtain that there exists an orthogonal array $T \subset [q]^n$ of strength t and size $|T| \leq (qn)^c$ for some universal constant $c > 0$.

3.2 Designs

In this section, we prove Theorem 1.2. It suffices to prove the theorem for $k > 2t$, since if $k \leq 2t$ then the complete design (the design containing all subsets of size k) establishes the theorem. We use the choice of B and V described in Section 1.2 and recall the definition (5) of the functions f_a of that section. We set $A = \binom{[v]}{t}$ and note that $\sum_{a \in A} f_a \equiv \binom{[k]}{t}$ and thus V contains the constant functions as Theorem 2.1 requires. As a convenient basis for V of integer-valued functions, we take $\{\phi_a\}_{a \in A}$ with $\phi_a = f_a$. By definition, $\{\phi_a\}_{a \in A}$ spans V and the fact that $\{\phi_a\}_{a \in A}$ is a basis for V will be implied by showing the isolation condition of Theorem 2.1.

We choose integers $m, c_0 \geq 1$ and real numbers $c_1, c_2, c_3 > 0$ to satisfy the conditions of divisibility, boundedness, symmetry and isolation in Theorem 2.1. First, $\frac{1}{|B|}\phi(B) = \binom{k}{t} / \binom{v}{t} \cdot (1, \ldots, 1)$ and hence we set

$c_0 = \binom{v}{t}$ so that $\frac{c_0}{|B|}\phi(B)$ is an integer vector. Second, $\|\phi(b)\|_2^2 \le |A| \le v^t$. Hence we set $c_1 = v^{t/2}$. Third, the symmetry condition also follows simply: let $\sigma \in S_v$ be a permutation on $[v]$. It acts naturally on B and A (by permuting subsets of $[v]$) and gives two permutations $\pi \in S_B$ and $\tilde{\pi} \in S_A$ that satisfy $\phi(\pi(b))_a = \phi(b)_{\tilde{\pi}^{-1}(a)}$. The linear transformation $\tau \in \mathrm{GL}(V)$ then corresponds to the permutation $\tilde{\pi}^{-1}$.

Finally, we need to show that for each $a \in A$ there exist many disjoint vectors which isolate it. This is accomplished in the following lemma.

Lemma 3.4. *Assume $k > 2t$. For any $a \in A$ there exist vectors $\gamma_1, \ldots, \gamma_r \in \mathbb{Z}^B$ with $r \ge |B|/(vk)^{2t}$ such that $\phi(\gamma_i) = \frac{k!}{(k-t)!} \cdot e_a$. Moreover, $\gamma_1, \ldots, \gamma_r$ have disjoint supports and $\|\gamma_i\|_2 \le (2k)^{3t/2}$ for $i \in [r]$.*

We will need the following technical claim for the proof of Lemma 3.4. In the following we consider binomial coefficients $\binom{n}{m} = 0$ whenever $n < m$.

Claim 3.5. *Let $a > b \ge 0$ and $c \ge 0$. Then*

$$\sum_{i=0}^{a} (-1)^i \binom{a}{i}\binom{c+i}{b} = 0.$$

Proof. Let $f(a,b,c) = \sum_{i=0}^{a}(-1)^i\binom{a}{i}\binom{c+i}{b}$. If $b,c > 0$ we have $\binom{c+i}{b} = \binom{c-1+i}{b} + \binom{c-1+i}{b-1}$ and hence $f(a,b,c) = f(a,b,c-1) + f(a,b-1,c-1)$. So, it is enough to verify the claim whenever $b = 0$ or $c = 0$. If $b = 0$ then $f(a,0,c) = \sum_{i=0}^{a}(-1)^i\binom{a}{i} = 0$ since $a \ge 1$. If $c = 0$ then $f(a,b,0) = \sum_{i=b}^{a}(-1)^i\binom{a}{i}\binom{i}{b} = \binom{a}{b}\sum_{i=b}^{a}(-1)^i\binom{a-b}{i-b} = 0$. □

Proof of Lemma 3.4. Let $a \in A = \binom{[v]}{t}$ be a coordinate we wish to isolate. Let $x \in \binom{[v]}{k}$ be a set disjoint from a and let $0 \le j \le t$. Define $\delta_{x,a,j} \in \mathbb{Z}^B$ to be the indicator vector for all subsets $b \in B = \binom{[v]}{k}$ such that $b \subset a \cup x$ and $|a \cap b| = j$, that is

$$\delta_{x,a,j} := \sum_{b \subset a\cup x, |b|=k, |a\cap b|=j} e_b.$$

We define vectors $\gamma_{a,x} \in \mathbb{Z}^B$ as

$$\gamma_{x,a} := \sum_{j=0}^{t} (-1)^{t-j} \frac{j!(k-j-1)!}{(k-t-1)!}\delta_{x,a,j}.$$

We will shortly show that

$$\phi(\gamma_{x,a}) = \frac{k!}{(k-t)!}e_a.$$

First we bound the norm of $\gamma_{x,a}$ and show the existence of many disjoint vectors. It is easy to check that $\|\gamma_{x,a}\|_2 \le (2k)^{3t/2}$. Also, the vector $\gamma_{x,a}$ is supported on coordinates $y \in B$ such that $|y \cap x| \ge k - t$. Thus, if we choose $x_1, \ldots, x_r \in B$ such that $|x_i \cap x_j| \le k - 2t - 1$ we get that the vectors $\gamma_{x_1,a}, \ldots, \gamma_{x_r,a}$ have disjoint support. We can choose $r \ge |B|/(vk)^{2t}$ by a simple greedy argument: choose $x_1, \ldots, x_r$ iteratively, where in each step after choosing x_i we remove all subsets $y \in B$ whose intersection with x_i is at least $k - 2t$. The number of subsets eliminated in each step is at most $(vk)^{2t}$ hence we will get $r \ge |B|/(vk)^{2t}$.

To conclude the proof, we need to compute $\phi(\gamma_{x,a})$. Let $a' \in A$. Clearly if $a' \not\subseteq a \cup x$ then $\phi(\gamma_{x,a})_{a'} = 0$. We thus assume that $a' \subset a \cup x$. Let $\ell = |a \cap a'|$ where $0 \le \ell \le t$. We have that $\phi(\delta_{x,a,j})_{a'} = 0$ if $j < \ell$, and that

$$\begin{aligned}\phi(\delta_{x,a,j})_{a'} &= |\{b \in B : a' \subset b \subset a \cup x, |a \cap b| = j\}| \\ &= \binom{t-\ell}{t-j}\binom{k-t+\ell}{j}.\end{aligned}$$

Hence we have that

$$\phi(\gamma_{x,a})_{a'} = \frac{(k-1)!}{(k-t-1)!}\sum_{j=\ell}^{t}(-1)^{t-j}\frac{\binom{t-\ell}{t-j}\binom{k-t+\ell}{j}}{\binom{k-1}{j}} \tag{9}$$

If $a' = a$ then $\phi(\gamma_{x,a})_a = k!/(k-t)!$ as claimed. To conclude we need to prove that if $a' \neq a$ then $\phi(\gamma_{x,a})_{a'} = 0$. We have $\ell = |a \cap a'| < t$ and let $s = t - \ell > 0$. Thus

$$\begin{aligned}
\phi(\gamma_{x,a})_{a'} &= (-1)^t \frac{(k-1)!}{(k-t-1)!}\sum_{j=\ell}^{t}(-1)^{j}\frac{\binom{t-\ell}{t-j}\binom{k-t+\ell}{j}}{\binom{k-1}{j}} \\
&= (-1)^s \frac{(k-1)!}{(k-t-1)!}\sum_{j=0}^{s}(-1)^{j}\frac{\binom{s}{j}\binom{k-s}{j+\ell}}{\binom{k-1}{j+\ell}} \\
&= (-1)^s \frac{(k-1)!}{(k-t-1)!\binom{k-1}{s-1}}\sum_{j=0}^{s}(-1)^{j}\binom{s}{j}\binom{k-\ell-1-j}{s-1} \\
&= (-1)^s \frac{(k-1)!}{(k-t-1)!\binom{k-1}{s-1}}\sum_{j=0}^{s}(-1)^{j}\binom{s}{j}\binom{k-\ell-1-s+j}{s-1}.
\end{aligned}$$

We now apply Claim 3.5 with $a = s, b = s-1, c = k-\ell-1-s$ and conclude that $\phi(\gamma_{x,a})_{a'} = 0$. □

We are now ready to apply Theorem 2.1. We have $|A| = \binom{v}{t}, c_0 = \binom{v}{t}, c_1 = v^{t/2}, c_2 = (vk)^{2t}, c_3 = (2k)^{3t/2}$ and $m = k!/(k-t)!$. Thus the theorem implies the existence of a $t-(v,k,\lambda)$ design $T \subset B$ with $|T| \leq v^{ct}$ for some universal constant $c > 0$.

4 Proof of Main Theorem

We prove Theorem 2.1 in this section. We recall the settings: B is a finite set and V is a vector space of functions from B to $\mathbb{Q}$. We assume the space V is spanned by integer valued functions $\{\phi_a : B \to \mathbb{Z}\}_{a\in A}$, where A is a finite index set. We also assume that the constant functions belong to V.

The proof strategy is conceptually simple: choose T randomly and show that this choice is successful with positive probability. Let N be the target size of T, to be chosen later. Let each $b \in B$ be chosen to be in T independently with probability $p := N/|B|$. Identifying T with its indicator vector in $\{0,1\}^B$, we have that $T_b \in \{0,1\}$ with $\mathbb{P}[T_b = 1] = p$. Define $X = \phi(T) \in \mathbb{Z}^A$ and note that $\mathbb{E}[X] = p \cdot \phi(B)$. In order to prove Theorem 2.1 we need to show that

$$\mathbb{P}[X = \mathbb{E}[X]] > 0. \tag{10}$$

We make two notes: first, since we assume that constant functions belong to V we have that if $X = \mathbb{E}[X]$ then in particular $|X| = p|B| = N$. Second, in order for (10) to hold we must have that $\mathbb{E}[X]$ is an integer vector. Thus, we must choose N to be divisible by c_0.

The difficulty with establishing (10) comes from the fact that we require A different events to occur simultaneously: for all $a \in A$ we require that $X_a = \mathbb{E}[X_a]$. To better explain the challenge, consider momentarily for simplicity the case where $\phi(b) \in \{0,1\}^A$ for all $b \in B$ and that for each $a \in A$, $\mathbb{P}_{b\in B}[\phi(b)_a = 1] = q$ (that is, all columns of ϕ have qB ones). Then each individual X_a is binomially distributed, $X_a \sim \mathrm{Bin}(|B|, pq)$, and it is not hard to see that

$$\mathbb{P}[X_a = \mathbb{E}[X_a]] \approx \frac{1}{\sqrt{qN}}.$$

However, we need the events $X_a = \mathbb{E}[X_a]$ to occur simultaneously for all $a \in A$. The problem arises because these events are dependent, and general techniques for handling such dependencies (for example, the Lovász

local lemma) only work when each event depends only on a few other events (which is not the case here) and where each event holds with sufficiently high probability (which is also not the case here). What we show is that, under the conditions of Theorem 2.1, if we choose N large enough (but only polynomially large in $|A|, m, c_0, c_1, c_2, c_3$) then all the events $X_a = \mathbb{E}[X_a]$ become essentially independent, and we show that

$$\mathbb{P}[X = \mathbb{E}[X]] \approx \prod_{a \in A} \mathbb{P}[X_a = \mathbb{E}[X_a]] \approx (\frac{1}{\sqrt{qN}})^{|A|}.$$

The actual expression we get is somewhat more complicated as it also involves pairwise correlations between the different events X_a, but conceptually it is of a similar flavor.

Our main technique to study the distribution of the random variable $X \in \mathbb{Z}^A$ is Fourier analysis. We recall some basic facts about Fourier analysis on $\mathbb{Z}^A$.

Fact 4.1 (Fourier analysis on $\mathbb{Z}^A$)**.** *Let $X \in \mathbb{Z}^A$ be a random variable. The Fourier coefficients of X live in the A-dimensional torus. Let $\mathbb{T} = [-1/2, 1/2)$ denote the torus. The Fourier coefficients $\widehat{X}(\theta)$ for $\theta \in \mathbb{T}^A$ are given by*

$$\widehat{X}(\theta) = \mathbb{E}_X[e^{2\pi i \langle X, \theta \rangle}],$$

where $\langle X, \theta \rangle = \sum_{a \in A} X_a \theta_a$. The probability that $X = \lambda$ for $\lambda \in \mathbb{Z}^A$ is given by the Fourier inversion formula

$$\mathbb{P}[X = \lambda] = \int_{\theta \in \mathbb{T}^A} \widehat{X}(\theta) e^{-2\pi i \langle \lambda, \theta \rangle} d\theta.$$

Recall that our goal is to understand the probability that $X = \mathbb{E}[X]$. Applying the Fourier inversion formula for $\lambda = \mathbb{E}[X]$ gives

$$\mathbb{P}[X = \mathbb{E}[X]] = \int_{\theta \in \mathbb{T}^A} \widehat{X}(\theta) e^{-2\pi i \langle \mathbb{E}[X], \theta \rangle} d\theta. \tag{11}$$

Thus, our goal from now on is to understand the Fourier coefficients of X. We first give an explicit formula for the Fourier coefficients.

Claim 4.2. *We have*

$$\widehat{X}(\theta) = \prod_{b \in B} (1 - p + p e^{2\pi i \cdot \langle \phi(b), \theta \rangle}).$$

Proof. By definition $X = \phi(T) = \sum_{b \in B} T_b \phi(b)$, where $T_b \in \{0, 1\}$ are independent with $\mathbb{P}[T_b = 1] = p$. Thus

$$\begin{aligned} \widehat{X}(\theta) &= \mathbb{E}_X[e^{2\pi i \langle X, \theta \rangle}] = \mathbb{E}_{\{T_b : b \in B\}}[e^{2\pi i \sum_{b \in B} T_b \langle \phi(b), \theta \rangle}] \\ &= \prod_{b \in B} \mathbb{E}_{T_b}[e^{2\pi i \, T_b \langle \phi(b), \theta \rangle}] = \prod_{b \in B} (1 - p + p e^{2\pi i \langle \phi(b), \theta \rangle}). \end{aligned}$$

□

Clearly all Fourier coefficients of X have absolute value at most 1. The first step is to understand the maximal Fourier coefficients of X, that is θ for which $|\widehat{X}(\theta)| = 1$.

Claim 4.3. *Let $L := \{\theta \in \mathbb{T}^A : \widehat{X}(\theta) = 1\}$. Then*

- *If $\theta \notin L$ then $|\widehat{X}(\theta)| < 1$.*
- *If $\theta \in L, \theta' \in \mathbb{T}^A$ then $\widehat{X}(\theta + \theta') = \widehat{X}(\theta')$. In particular, L is a subgroup of $\mathbb{T}^A$.*

Proof. Both claims follow immediately from the observation that $\theta \in L$ iff $\langle \phi(b), \theta \rangle \in \mathbb{Z}$ for all $b \in B$. □

In fact, the isolation conditions in Theorem 2.1 imply that L is a discrete subgroup of $\mathbb{T}^A$ (i.e. a lattice). Let $M := (1/m \cdot \mathbb{Z})^A$ be the lattice in $\mathbb{T}^A$ of all elements whose coordinates are integer multiples of $1/m$. We show that L is a sublattice of M.

Claim 4.4. $L \subseteq M$.

Proof. Let $\theta \in L$. We need to show that $m\theta_a \in \mathbb{Z}$ for all $a \in A$. By the isolation condition of Theorem 2.1, there exists $\gamma \in \mathbb{Z}^B$ such that $\phi(\gamma) = me_a$. Since $\theta \in L$ we have that $\langle \phi(b), \theta \rangle \in \mathbb{Z}$ for all $b \in B$. Hence also $\langle \phi(\gamma), \theta \rangle \in \mathbb{Z}$, i.e. $m\theta_a \in \mathbb{Z}$ as claimed. □

The first step we take is to approximate the Fourier coefficients of X near the lattice L. This will assume very little about ϕ, essentially only boundedness. The second (and more complex) step will be to show that all other Fourier coefficients are negligible, and in fact the contribution to (11) all come from Fourier coefficients near L. The second part will heavily utilize the symmetry of the map ϕ and the existence of many disjoint isolation vectors. Theorem 2.1 then follows by a careful setting of parameters and a routine calculation.

Formally, we will use ℓ_2 distance on $\mathbb{T}^A$. For $x \in \mathbb{T}$ define its absolute value $|x| = |x \pmod 1|$ to be the minimal absolute value of x modulo 1 (that is, we take $x \pmod 1 \in [-1/2, 1/2]$). Define the distance between $\theta', \theta'' \in \mathbb{T}^A$ by

$$d(\theta', \theta'') := \sqrt{\sum_{a \in A} |\theta'_a - \theta''_a|^2}.$$

The distance between $\theta \in \mathbb{T}^A$ and $L \subset \mathbb{T}^A$ is given by

$$d(\theta, L) := \min_{\alpha \in L} d(\theta, \alpha).$$

The following three lemmas are the main technical ingredients of the proof. The first lemma gives a good approximation for the Fourier coefficients of X near zero (and by Claim 4.3, near any point in L).

Lemma 4.5 (Estimating Fourier coefficients near zero)**.** *Assume the conditions of Theorem 2.1 and fix $\varepsilon \leq O(1/(c_1 N^{1/3}))$. Let $\theta \in \mathbb{T}^A$ be such that $\|\theta\|_2 \leq \varepsilon$. Then*

$$\widehat{X}(\theta) = e^{2\pi i \langle \mathbb{E}[X], \theta \rangle} e^{-4\pi^2 p \cdot \theta^T R \theta} (1 + \delta)$$

where R is the $A \times A$ pairwise-correlation matrix of ϕ given by $R_{a',a''} = \sum_{b \in B} \phi(b)_{a'} \phi(b)_{a''}$, and where $|\delta| = O(N^2/|B| + N c_1^3 \varepsilon^3)$.

The second lemma bounds the Fourier coefficients of X far from the lattice M.

Lemma 4.6 (Bounding Fourier coefficients far from M)**.** *Assume the conditions of Theorem 2.1. Let $\theta \in \mathbb{T}^A$ be such that $d(\theta, M) \geq \varepsilon$. Then*

$$|\widehat{X}(\theta)| \leq \exp\left(-N\varepsilon^2 \cdot \frac{m^2}{|A| c_2 c_3^2}\right).$$

The third lemma bounds the remaining Fourier coefficients which are near M but far from L. In the following let $M \setminus L$ denote the set of elements in M but not in L.

Lemma 4.7 (Bounding Fourier coefficients near M but far from L)**.** *Assume the conditions of Theorem 2.1 and fix $\varepsilon \leq 1/(2c_1 m)$. Let $\theta \in \mathbb{T}^A$ be such that $d(\theta, M \setminus L) \leq \varepsilon$. Then*

$$|\widehat{X}(\theta)| \leq \exp\left(-N \cdot \frac{O(1)}{m^2 |A| \log(c_1 |A|)}\right).$$

We prove Lemmas 4.5, 4.6 and 4.7 in Sections 4.1, 4.2 and 4.3, respectively. We combine them to prove Theorem 2.1 in Section 4.4.

4.1 Estimating Fourier coefficients near zero

Let $\theta \in \mathbb{T}^A$ be such that $\|\theta\|_2 \le \varepsilon$. We may assume that $\varepsilon \le O(1/(c_1 N^{1/3}))$ otherwise the conclusion of the lemma is trivial. We decompose

$$e^{-2\pi i\langle \mathbb{E}[X],\theta\rangle} \cdot \widehat{X}(\theta) = \prod_{b\in B}\left(e^{-2\pi i\cdot p\langle \phi(b),\theta\rangle}\cdot(1-p+pe^{2\pi i\langle\phi(b),\theta\rangle})\right). \tag{12}$$

Let $v_b := \langle \phi(b),\theta\rangle$ where the inner product is taken over $\mathbb{R}$. Since we assume $\|\theta\|_2 \le \varepsilon$ we can bound $|v_b| \le \|\phi(b)\|_2\|\theta\|_2 \le c_1\varepsilon \ll 1$. Thus we can approximate the terms in (12) by their Taylor series. The following claim gives a cubic approximation.

Claim 4.8. *Let $f:\mathbb{R}\to\mathbb{C}$ be given by $f(x) := e^{-ipx}(1-p+pe^{ix})$. Then for $|x|\le 1$ we have*

$$f(x) = e^{-px^2}(1+\delta),$$

where $|\delta| \le O(p^2x^2+px^3)$.

Proof. We compute the cubic approximation for $f(x)$ as a polynomial in p,x. In the following we use shorthand expression $x = y+O(z)$ for $|x-y| = O(z)$. We have

$$\begin{aligned} f(x) &= (1-p)e^{-ipx} + pe^{i(1-p)x} \\ &= (1-p)(1-ipx+O(p^2x^2)) + p(1+i(1-p)x - x^2 \pm O(px^2+x^3)) \\ &= 1-px^2+O(p^2x^2+px^3) \\ &= e^{-px^2}+O(p^2x^2+px^3). \end{aligned}$$

□

We next apply the approximation given in Claim 4.8 to each of the terms appearing in (12). Summing up the errors, and using the fact that each term is bounded in absolute value by 1, we get that

$$\widehat{X}(\theta) = e^{2\pi i\langle\mathbb{E}[X],\theta\rangle}e^{-4\pi^2 p\cdot\sum_{b\in B}v_b^2}(1+\delta) \tag{13}$$

where $|\delta| \le O(p^2\sum_{b\in B}v_b^2 + p\sum_{b\in B}v_b^3)$. To conclude the proof, note that

$$\sum_{b\in B}v_b^2 = \sum_{b\in B}\langle\phi(b),\theta\rangle^2 = \theta^T R\theta,$$

where we recall that $R_{a',a''} = \sum_{b\in B}\phi(b)_{a'}\phi(b)_{a''}$. To bound the error term, recall that $|v_b| \le c_1\varepsilon \ll 1$ hence

$$|\delta| \le O(p^2|B| + p|B|(c_1\varepsilon)^3) = O(N^2/|B| + Nc_1^3\varepsilon^3).$$

4.2 Bounding Fourier coefficients far from M

Let $\theta\in\mathbb{T}^A$ be such that $d(\theta,M)\ge\varepsilon$. Thus, there exists at least on coordinate θ_a whose distance from multiples of $1/m$ is at least $\varepsilon/\sqrt{|A|}$. Otherwise put, there exists $a\in A$ such that

$$|m\theta_a \pmod 1| \ge \varepsilon m/\sqrt{|A|}. \tag{14}$$

Recall that the Fourier coefficient $\widehat{X}(\theta)$ is given by

$$\widehat{X}(\theta) = \prod_{b\in B}(1-p+pe^{2\pi i\langle\phi(b),\theta\rangle}).$$

Hence, to get a bound on $|\widehat{X}(\theta)|$ essentially we need to show that $\langle \phi(b), \theta \rangle$ is far from integer for many $b \in B$. Note that we cannot longer assume, as in the proof of Lemma 4.5, that $\langle \phi(b), \theta \rangle$ is small in absolute value, since we assume no upper bound on $\|\theta\|_2$. Thus, it may be the case that $\langle \phi(b), \theta \rangle$ is large but still approximately integer. Let $\nu_b := \langle \phi(b), \theta \rangle \pmod 1$ where $|\nu_b| \leq 1/2$. Our goal is to show that $|\nu_b|$ is noticeably large for many values $b \in B$. This will then imply the required upper bound on $|\widehat{X}(\theta)|$.

We will show this using the isolation vectors guaranteed by Theorem 2.1. Let $\gamma \in \mathbb{Z}^B$ be an isolation vector for a with modulus m; that is $\phi(\gamma) = m \cdot e_a$. We first show that it cannot be that $\nu_b \approx 0$ for all $b \in \mathrm{Supp}(\gamma)$.

Claim 4.9. *Let $\gamma \in \mathbb{Z}^B$ be such that $\phi(\gamma) = m \cdot e_a$. Then*

$$\sum_{b \in \mathrm{Supp}(\gamma)} |\nu_b|^2 \geq \frac{\varepsilon^2 m^2}{|A| \|\gamma\|_2^2}.$$

Proof. Using the isolation property of γ we get that

$$\begin{aligned} \sum_{b \in \mathrm{Supp}(\gamma)} \gamma_b \nu_b \pmod 1 &= \sum_{b \in \mathrm{Supp}(\gamma)} \gamma_b \langle \phi(b), \theta \rangle \pmod 1 \\ &= \langle \phi(\gamma), \theta \rangle \pmod 1 = m\theta_a \pmod 1. \end{aligned}$$

Hence by (14) we get that $|\sum_{b \in \mathrm{Supp}(\gamma)} \gamma_b \nu_b \pmod 1| \geq \varepsilon m / \sqrt{|A|}$. On the other hand, we can bound

$$|\sum_{b \in \mathrm{Supp}(\gamma)} \gamma_b \nu_b \pmod 1| \leq |\sum_{b \in \mathrm{Supp}(\gamma)} \gamma_b \nu_b| \leq \|\gamma\|_2 \sqrt{\sum_{b \in \mathrm{Supp}(\gamma)} |\nu_b|^2}.$$

Combining the two bounds, we get that $\sum_{b \in \mathrm{Supp}\gamma} |\nu_b|^2 \geq \varepsilon^2 m^2 / |A| \|\gamma\|_2^2$ as claimed. □

We now use the assumption of Theorem 2.1 on the existence of many vectors which isolate a with disjoint support. Recall that by assumption we have $r \geq |B|/c_2$ vectors $\gamma_1, \ldots, \gamma_r \in \mathbb{Z}^B$ such that: (1) each γ_i isolates a with modulus m; (2) The vectors $\gamma_1, \ldots, \gamma_r$ have disjoint supports; and (3) $\|\gamma_i\| \leq c_3$ for all $i \in [r]$. Applying Claim 4.9 to each vector γ_i independently we derive that

$$\sum_{b \in B} |\nu_b|^2 \geq \varepsilon^2 |B| \cdot \frac{m^2}{|A| c_2 c_3^2}. \tag{15}$$

To conclude the proof of the lemma, we apply (15) to derive an upper bound on $|\widehat{X}(\theta)|$. The following claim is simple.

Claim 4.10. *Let $p \leq 1/2$ and $|x| \leq 1/2$. Then*

$$|1 - p + pe^{2\pi i x}| \leq \exp(-px^2).$$

Applying Claim 4.10 we derive the bound

$$|\widehat{X}(\theta)| = \prod_{b \in B} |1 - p + pe^{2\pi i \cdot \nu_b}| \leq \exp\left(-p \sum_{b \in B} |\nu_b|^2\right) \leq \exp\left(-\varepsilon^2 N \cdot \frac{cm^2}{|A| c_2 c_3^2}\right).$$

4.3 Bounding Fourier coefficients near M but far from L

Let $\theta \in \mathbb{T}^A$ be such that $d(\theta, M \setminus L) \leq \varepsilon$. That is, there exists $\alpha \in M \setminus L$ such that $d(\theta, \alpha) \leq \varepsilon$. Since $\alpha \notin L$ there must exist $b^* \in B$ such that $\langle \phi(b^*), \alpha \rangle \notin \mathbb{Z}$. We will show using the symmetry of ϕ that in fact this holds for many $b \in B$. Moreover, since $\alpha \in M$ we have that if $\langle \phi(b), \alpha \rangle \notin \mathbb{Z}$ is must be at least $1/m$ far from the integers. This will allow us to give strong upper bounds on the Fourier coefficient $\widehat{X}(\alpha)$ and by continuity also on $\widehat{X}(\theta)$.

Let $\mathscr{L}$ denote the lattice generated by $\{\phi(b) : b \in B\}$. In other words, $\mathscr{L}$ is the subgroup of $\mathbb{Z}^A$ whose elements are all possible integer combinations of $\{\phi(b) : b \in B\}$. We first show that any subset of B which generates the lattice $\mathscr{L}$ must contain b for which $\langle \phi(b), m\alpha \rangle \neq 0$.

Claim 4.11. *Let $K \subset B$ be a set which generates the lattice $\mathscr{L}$. Then there must exist $b \in K$ for which $\langle \phi(b), m\alpha \rangle \neq 0$.*

Proof. By assumption since K generates the lattice $\mathscr{L}$, we can express $\phi(b^*)$ as an integer combination of $\{\phi(b) : b \in K\}$. That is, there exist integer coefficient α_b for $b \in K$ such that

$$\phi(b^*) = \sum_{b \in K} \alpha_b \phi(b).$$

Thus, as $\langle \phi(b^*), m\alpha \rangle \neq 0$, there must exist $b \in K$ for which $\langle \phi(b), m\alpha \rangle \neq 0$ as well. □

We next claim that there must exist at least one small set $K \subset B$ which generates $\mathscr{L}$. We will later use symmetry to generate from it many such sets.

Claim 4.12. *There exists $K \subset B$ of size $|K| \leq O(|A| \log(c_1|A|))$ such that $\{\phi(b) : b \in K\}$ generates the lattice $\mathscr{L}$.*

Proof. Let K be a minimal subset of B such that $\{\phi(b) : b \in K\}$ generates the lattice $\mathscr{L}$. We claim that the minimality of K implies that all partial sums $\phi(K')$ for $K' \subseteq K$ must be distinct. Otherwise, assume that there exist two distinct subsets $K_1, K_2 \subseteq K$ for which $\phi(K_1) = \phi(K_2)$. We can assume w.l.o.g that K_1, K_2 are disjoint by removing common elements from both. Thus we have

$$\sum_{b \in K_1} \phi(b) - \sum_{b \in K_2} \phi(b) = 0.$$

In particular, we can express any $b' \in K_1 \cup K_2$ as an integer combination of $\{\phi(b) : b \in K \setminus \{b'\}\}$. Thus, we can remove b' from K and maintain the property that the resulting set generates $\mathscr{L}$. This contradicts the minimality of K.

We thus know that all sums $\{\phi(K') : K' \subseteq K\}$ are distinct. We now apply the assumption that ϕ is bounded. By the assumptions of Theorem 2.1 we know that $\|\phi(b)\|_\infty \leq \|\phi(b)\|_2 \leq c_1$. Hence we conclude that

$$\{\phi(K') : K' \subseteq K\} \subseteq [-c_1 K, c_1 K]^A,$$

which imply that

$$2^K \leq (2c_1 K + 1)^{|A|}.$$

It is easy to verify that this gives the bound $K \leq O(|A| \log(c_1|A|))$ as claimed. □

The next step is to use the symmetry of ϕ to generate many small sets which span $\mathscr{L}$.

Claim 4.13. *Let $K \subset B$ be a set such that $\{\phi(b) : b \in K\}$ generates the lattice $\mathscr{L}$. Let $(\pi, \tau) \in S_B \times \mathrm{GL}(V)$ be a symmetry of ϕ. Let $K_\pi := \{\pi(b) : b \in K\}$ be a shift of K by π. Then $\{\phi(b) : b \in K_\pi\}$ also generates the lattice $\mathscr{L}$.*

Proof. Let $b' \in B$. We need to show that we can express $\phi(b')$ as integer combination of $\{\phi(\pi(b)) : b \in K\}$. Consider $\pi^{-1}(b')$. By assumption the image of ϕ on elements of K generates the lattice $\mathscr{L}$, hence there exist coefficients $\alpha_b \in \mathbb{Z}$ for $b \in K$ such that

$$\phi(\pi^{-1}(b')) = \sum_{b \in K} \alpha_b \phi(b).$$

Applying the assumption that (π, τ) is a symmetry of ϕ we get that

$$\phi(b') = \phi(\pi(\pi^{-1}(b'))) = \tau(\phi(\pi^{-1}(b'))) = \sum_{b \in K} \alpha_b \cdot \tau(\phi(b)) = \sum_{b \in K} \alpha_b \phi(\pi(b)). \quad \square$$

We combine Claims 4.11, 4.12 and 4.13 to derive that $\langle \phi(b), \alpha \rangle \neq 0$ for many $b \in B$. Let $\widetilde{B} = \{b \in B : \langle \phi(b), \alpha \rangle \neq 0\}$.

Corollary 4.14. $|\widetilde{B}| \geq \Omega\left(\frac{|B|}{|A|\log(c_1|A|)}\right)$.

Proof. Let K be the set guaranteed by Claim 4.12 where $|K| \leq O(|A|\log(c_1|A|))$. Let $K_\pi = \{\pi(b) : b \in K\}$. We know by Claim 4.13 that for any symmetry (π, τ) of ϕ we have

$$|K_\pi \cap \widetilde{B}| \geq 1.$$

Let H be the subgroup of permutations on B given by symmetries of ϕ. That is, $H = \{\pi : (\pi, \tau) \text{ symmetry of } \phi\}$. We know by the assumptions of Theorem 2.1 that H acts transitively on B. Thus, for any fixed $b \in B$, if we choose $\pi \in H$ uniformly we have that $\pi(b)$ is uniformly distributed in B. Thus,

$$\mathbb{E}_{\pi \in H}[K_\pi \cap \widetilde{B}] = \sum_{b \in K} \mathbb{P}_{\pi \in H}[\pi(b) \in \widetilde{B}] = \frac{|K||\widetilde{B}|}{|B|}.$$

We thus conclude that we must have $|\widetilde{B}| \geq |B|/|K|$. $\square$

We conclude the proof of Lemma 4.7 by establishing an upper bound of $\widehat{X}(\alpha)$. For any $b \in \widetilde{B}$ we have that $\langle \phi(b), \alpha \rangle \pmod 1 \neq 0$, hence since $\alpha \in (1/m \cdot \mathbb{Z})^A$ we have

$$|\langle \phi(b), \alpha \rangle \pmod 1| \geq 1/m.$$

Recall that by assumption $\|\phi(b)\|_2 \leq c_1$ and $\|\alpha - \theta\| \leq 1/(2c_1 m)$. Thus $|\langle \phi(b), \alpha - \theta \rangle| \leq 1/2m$ by Cauchy-Schwarz and we get that

$$|\langle \phi(b), \theta \rangle \pmod 1| \geq 1/2m.$$

We thus conclude with an upper bound on $|\widehat{X}(\theta)|$. Applying Claim 4.10 we have

$$|\widehat{X}(\theta)| \leq \prod_{b \in \widetilde{B}} |1 - p + pe^{2\pi i \langle \phi(b), \theta \rangle}| \leq \exp(-p(1/2m)^2 |\widetilde{B}|) \leq \exp\left(-N \cdot \frac{O(1)}{m^2 |A| \log(c_1|A|)}\right).$$

4.4 Proof of Theorem 2.1 from Lemmas 4.5, 4.6 and 4.7

We now deduce Theorem 2.1 from Lemmas 4.5, 4.6 and 4.7. Recall that we have

$$\mathbb{P}[X = \mathbb{E}[X]] = \int_{\theta \in \mathbb{T}^A} \widehat{X}(\theta) e^{-2\pi i \langle \mathbb{E}[X], \theta \rangle} d\theta. \tag{16}$$

Let $N = \text{poly}(|A|, m, c_0, c_1, c_2, c_3)$ large enough to be chosen later. We would assume throughout that N is a multiple of $c_0 m$. If $|B| = O(N^2)$ then the set B is small to begin with, so assume that $|B| \gg N^2$. We set

$\varepsilon \approx N^{-1/3}$ so that the conditions for Lemmas 4.5 and 4.7 hold. More explicitly, we set $\varepsilon := O(1/c_1 N^{1/3})$ so that the conditions for Lemma 4.5 hold with $|\delta| \leq 1/2$; and we assume that $N \geq \Omega(m^3)$ so that $\varepsilon \leq 1/(2c_1 m)$ and the conditions for Lemma 4.7 also hold.

We decompose the integral in (16) into three integrals: over points which are ε close to L; over points which are ε close to $M \setminus L$; and over points which are ε far from M. Our choice of $\varepsilon < 1/2m$ also guarantees that balls of radius ε around distinct points in M are disjoint. We thus have that $\mathbb{P}[X = \mathbb{E}[X]] = I_1 + I_2 + I_3$ where

$$I_1 := \sum_{\alpha \in L} \int_{\theta \in \mathbb{T}^A : d(\theta,\alpha) \leq \varepsilon} \widehat{X}(\theta) e^{-2\pi i \langle \mathbb{E}[X], \theta \rangle} d\theta,$$
$$I_2 := \sum_{\alpha \in M \setminus L} \int_{\theta \in \mathbb{T}^A : d(\theta,\alpha) \leq \varepsilon} \widehat{X}(\theta) e^{-2\pi i \langle \mathbb{E}[X], \theta \rangle} d\theta,$$
$$I_3 := \int_{\theta \in \mathbb{T}^A : d(\theta,M) > \varepsilon} \widehat{X}(\theta) e^{-2\pi i \langle \mathbb{E}[X], \theta \rangle} d\theta.$$

We first lower bound I_1.

Claim 4.15.

$$I_1 \geq |L| \left(\frac{\Omega(1)}{c_1 N^{1/2} |A|^{1/2}} \right)^{|A|}.$$

Proof. We first use the assumption that N divides $c_0 m$ to reduce computing I_1 to an integral around 0. We claim that the assumption that $c_0 m | N$ implies that $\langle \mathbb{E}[X], \alpha \rangle \in \mathbb{Z}$ for all $\alpha \in L$. This is since this choice implies that all entries of $\mathbb{E}[X]$ are divisible by m since

$$\mathbb{E}[X] = \frac{N}{|B|} \phi(B) = (N/c_0 m) \cdot m \cdot \frac{c_0}{|B|} \phi(B) \in m\mathbb{Z}^A.$$

Moreover, since $\alpha \in L \subset M$ we have that $m\alpha \in \mathbb{Z}^A$, hence $\langle \mathbb{E}[X], \alpha \rangle \in \mathbb{Z}$. Combining this with Claim 4.3 which states that the Fourier coefficients of X are invariants to shifts by $\alpha \in L$, we deduce that

$$I_1 = |L| \int_{\theta \in \mathbb{T}^A : \|\theta\|_2 \leq \varepsilon} \widehat{X}(\theta) e^{-2\pi i \langle \mathbb{E}[X], \theta \rangle} d\theta.$$

Recall that by Lemma 4.5 and our choice of parameters, if $\|\theta\|_2 \leq \varepsilon$ then

$$\widehat{X}(\theta) = \widetilde{X}(\theta)(1 + \delta(\theta))$$

where $\widetilde{X}(\theta) = e^{2\pi i \langle \mathbb{E}[X], \theta \rangle} e^{-4\pi^2 p \cdot \theta^T R \theta}$ and where $|\delta(\theta)| \leq 1/2$. Hence

$$I_1 = |L| \int_{\theta \in \mathbb{T}^A : \|\theta\|_2 \leq \varepsilon} e^{-4\pi^2 p \cdot \theta^T R \theta} (1 + \delta) d\theta.$$

Consider

$$I_1' = |L| \int_{\theta \in \mathbb{T}^A : \|\theta\|_2 \leq \varepsilon} e^{-4\pi^2 p \cdot \theta^T R \theta} d\theta.$$

We claim that $|I_1| \geq |I_1'|/2$, hence it suffices to lower bound $|I_1'|$ in order to lower bound $|I_1|$. To see that, note that I_1' is an integral of a real positive function; that we can always lower bound $|I_1|$ by its real part $\mathrm{Re}(I_1)$; and that $\mathrm{Re}(1 + \delta) \geq 1/2$ since $|\delta| \leq 1/2$. Thus

$$|I_1| \geq \mathrm{Re}(I_1) \geq \mathrm{Re}(I_1')/2 = I_1'/2.$$

We next lower bound I_1'. Note first that we can bound $\theta^T R\theta \le Bc_1^2\|\theta\|_2^2$. This is because

$$\theta^T R\theta = \sum_{b\in B}\langle\theta,\phi(b)\rangle^2 \le \sum_{b\in B}\|\theta\|_2^2\|\phi(b)\|_2^2 \le Bc_1^2\|\theta\|_2^2.$$

Thus we get that

$$I_1' \ge |L|\int_{\theta\in\mathbb{T}^A:\|\theta\|_2\le\varepsilon} e^{-4\pi^2 c_1^2 N\|\theta\|_2^2}d\theta.$$

We bound I_1' from below by the volume of the region in which the integrand is constant. This occurs whenever $\|\theta\|_2 \le \varepsilon' = O(1/(c_1N^{1/2}))$. Recall that we chose $\varepsilon = O(1/(c_1N^{1/3})) \gg \varepsilon'$. Hence the ball of radius ε' is contained in the area over which we integrate, so we obtain the lower bound

$$|I_1| \ge I_1'/2 \ge |L|\cdot O(1)\cdot \mathrm{Vol}(\mathrm{Ball}(0,\varepsilon')) = L\cdot\left(\frac{\Omega(1)}{\varepsilon'|A|^{1/2}}\right)^{|A|} = L\cdot\left(\frac{\Omega(1)}{c_1N^{1/2}|A|^{1/2}}\right)^{|A|}. \quad \square$$

The next steps are to bound I_2 and I_3 from above. We bound them by the maximal value that $|\widehat{X}(\theta)|$ can achieve in their integral domains. Lemma 4.7 gives a bound on I_2,

$$|I_2| \le \max\{|\widehat{X}(\theta)| : d(\theta, M\setminus L)\le\varepsilon\} \le \exp\left(-N\cdot\frac{O(1)}{m^2|A|\log(c_1|A|)}\right),$$

and Lemma 4.6 and our choice of $\varepsilon = O(1/(c_1N^{1/3}))$ gives a bound on I_3,

$$|I_3| \le \max\{|\widehat{X}(\theta)| : d(\theta,M)\ge\varepsilon\} \le \exp\left(-N\varepsilon^2\cdot\frac{m^2}{|A|c_2c_3^2}\right) = \exp\left(-O(N^{1/3})\cdot\frac{m^2}{|A|c_1^2c_2c_3^2}\right).$$

We now need to choose N large enough so that $I_1 \gg |I_2|, |I_3|$. This can be accomplished since I_1 decays polynomially with N, while $|I_2|, |I_3|$ decay exponentially fast. It is not hard to verify that this is guaranteed whenever

$$N \ge \Omega(1)\cdot\max(A^2m^2\log^2(mAc_0c_1c_2c_3), A^6c_1^6c_2^3c_3^6\log^3(mAc_0c_1c_2c_3)).$$

5 Summary and open problems

Our main theorem guarantees the existence of a small subset $T\subset B$ for which (1) holds. The conditions we require are boundedness, divisibility, symmetry and isolation. The first three conditions seem natural for this type of problems, but the fourth seems artificial, as it depends on the specific basis we choose for V. Thus, we wonder if this condition can be removed. In particular, the following question captures much of the difficulty. Let G be a group that acts transitively on a set X. A subset $T\subset G$ is *X-uniform* (or an *X-design*) if it acts on X exactly as G does. That is, for any $x,y\in X$,

$$\frac{1}{|T|}|\{g\in T : g(x)=y\}| = \frac{1}{|G|}|\{g\in G : g(x)=y\}| = \frac{1}{|X|}.$$

In our language we may take $B=G$ and V to be the space spanned by all functions $\phi_{(x,y)} : B\to\{0,1\}$ of the form $\phi_{(x,y)}(b) = \mathbf{1}_{\{b(x)=y\}}$ for $x,y\in X$. Then T is X-uniform if and only if (1) holds. Taking A to be some subset of $X\times X$ for which $(\phi_a)_{a\in A}$ forms a basis of V, the boundedness, divisibility and symmetry conditions are clearly satisfied. However, it is not clear whether the isolation condition is satisfied as well. If indeed the isolation condition is redundant, one may conjecture that:

Conjecture 5.1. *Let G be a group that acts transitively on a set X. Then there exists an X-uniform subset $T\subset G$ such that $|T|\le|X|^c$ for some universal constant $c>0$.*

A second question is whether one can apply our techniques to get *minimal* objects. Recall that the size of the objects we achieve is only minimal up to polynomial factors. For example, one of the main open problems in design theory is whether there exists a Steiner system (i.e. a t-design with $\lambda = 1$) for any $t > 5$. Another major open problem of a similar spirit is the existence of Hadamard matrices of all orders $n = 4m$, or equivalently, 2-$(4m-1, 2m-1, m-1)$ designs. Empirical estimates for $n \leq 32$ suggest that there are $\exp(O(n(\log n)))$ Hadamard matrices of order $n = 4m$. Since are so many of them, and since the logarithm of their number grows at a regular rate, we suspect that they exist for some purely statistical reason. However, the Gaussian local limit model seems to be false for Hadamard matrices interpreted as t-designs; it does not accurately estimate how many there are.

A third question is whether there exists an algorithmic version of our work, similar to the algorithmic Moser [Mos09] and Moser-Tardos [MT10] versions of the Lovász local lemma [EL75], and the algorithmic Bansal [Ban10] version of the six standard deviations method of Spencer [Spe85]. If an efficient randomized algorithm of our method were found, then we could no longer indisputably claim that we have a low-probability version of the probabilistic method. On the other hand it would be strange, from the viewpoint of computational complexity theory, if low-probability existence can always be converted to high-probability existence. Maybe our construction is fundamentally a low-probability construction.

References

[AL11] Noga Alon and Shachar Lovett, *Almost k-wise vs k-wise independent permutations, and uniformity for general group actions*, 2011, ECCC TR11-049.

[AV97] Nogal Alon and Van H. Vu, *Anti-Hadamard matrices, coin weighing, threshold gates and indecomposable hypergraphs*, J. Combin. Theory Ser. A **79** (1997), no. 1, 133–160.

[Ban10] Nikhil Bansal, *Constructive algorithms for discrepancy minimization*, Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science, FOCS '10, IEEE Computer Society, 2010, arXiv:1002.2259, pp. 3–10.

[Cam95] P. J. Cameron, *Permutation groups*, Handbook of combinatorics, Vol. 1, 2, Elsevier, 1995, pp. 611–645.

[CD07] Charles J. Colbourn and Jeffrey H. Dinitz (eds.), *The CRC handbook of combinatorial designs*, 2nd ed., Discrete Mathematics and its Applications, Chapman & Hall/CRC, 2007.

[EL75] Paul Erdős and László Lovász, *Problems and results on 3-chromatic hypergraphs and some related questions*, Infinite and Finite Sets, Coll. Math. Soc. J. Bolyai, no. 11, North-Holland, 1975, pp. 609–627.

[HSS99] A. S. Hedayat, N. J. A. Sloane, and John Stufken, *Orthogonal arrays: Theory and applications*, Springer-Verlag, 1999.

[Kas07] M. Kassabov, *Symmetric groups and expanders*, Invent. Math. **170** (2007), no. 2, 327–354, arXiv: math/0503204.

[KM94] Daphne Koller and Nimrod Megiddo, *Constructing small sample spaces satisfying given constants*, SIAM J. Discrete Math. **7** (1994), no. 2, 260–274.

[KNR05] E. Kaplan, M. Naor, and O. Reingold, *Derandomized constructions of k-wise (almost) independent permutations*, Approximation, randomization and combinatorial optimization (C. Chekuri,

K. Jansen, J. D. P. Rolim, and L. Trevisan, eds.), Lecture Notes in Computer Science, vol. 3624, Springer, 2005, pp. 354–365.

[KP82] Richard M. Karp and Christos H. Papadimitriou, *On linear characterizations of combinatorial optimization problems*, SIAM J. Comput. **11** (1982), no. 4, 620–632.

[Mag09] Spyros S. Magliveras, *Large sets of t-designs from groups*, Mathematica Slovaca **59** (2009), no. 1, 1–20.

[Mos09] Robin A. Moser, *A constructive proof of the Lovász local lemma*, Proceedings of the 41st annual ACM symposium on Theory of computing, STOC, ACM, 2009, arXiv:0810.4812, pp. 343–350.

[MT10] Robin A. Moser and Gábor Tardos, *A constructive proof of the general Lovász local lemma*, J. ACM **57** (2010), no. 2, 11:1–11:15, arXiv:0903.0544.

[Rao73] C. Radhakrishna Rao, *Some combinatorial problems of arrays and applications to design of experiments*, Survey of combinatorial theory (J. N. Srivastava, ed.), North-Holland, 1973, pp. 349–359.

[Spe85] Joel Spencer, *Six standard deviations suffice*, Trans. Amer. Math. Soc. **289** (1985), no. 2, 679–706.

[Tei87] Luc Teirlinck, *Non-trivial t-designs without repeated blocks exist for all t*, Discrete Math. **65** (1987), no. 3, 301–311.