Electronic Colloquium on Computational Complexity, Report No. 70 (2017)

Probabilistic Existence of Large Sets of Designs

Shachar Lovett
University of California San Diego
9500 Gilman Drive, La Jolla, CA 92093
slovett@ucsd.edu

Sankeerth Rao
University of California San Diego
9500 Gilman Drive, La Jolla, CA 92093
sankeerth1729@gmail.com

Alexander Vardy
University of California San Diego
9500 Gilman Drive, La Jolla, CA 92093
avardy@ucsd.edu

April 3, 2017

Abstract

A new probabilistic technique for establishing the existence of certain regular combinatorial structures has been introduced by Kuperberg, Lovett, and Peled (STOC 2012). Using this technique, it can be shown that under certain conditions, a randomly chosen structure has the required properties of a t-(n, k, λ) combinatorial design with tiny, yet positive, probability.

Herein, we strengthen both the method and the result of Kuperberg, Lovett, and Peled as follows. We modify the random choice and the analysis to show that, under the same conditions, not only does a t-(n, k, λ) design exist but, in fact, with positive probability there exists a *large set* of such designs — that is, a partition of the set of k-subsets of $[n]$ into t-(n, k, λ) designs. Specifically, using the probabilistic approach derived herein, we prove that for all sufficiently large n, large sets of t-(n, k, λ) designs exist whenever $k > 9t$ and the necessary divisibility conditions are satisied. This resolves the existence conjecture for large sets of designs for all $k > 9t$.

*The research of Shachar Lovett was supported by the National Science Foundation under grant CCF-1614023. The research of Sankeerth Rao and Alexander Vardy was supported by the National Science Foundation under grant CCF-1405119.

ISSN 1433-8092

1. Introduction

Let $[n] = \{1,2,\ldots,n\}$. A k-set is a subset of $[n]$ of size k. A t-(n,k,λ) *combinatorial design* is a collection $\mathcal{D}$ of distinct k-sets of $[n]$, called *blocks*, such that every t-set of $[n]$ is contained in exactly λ blocks. A *large set of designs* of size l, denoted $\mathsf{LS}(l;\, t,k,n)$, is a set of l disjoint t-(n,k,λ) designs $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_l$ such that $\mathcal{D}_1 \cup \mathcal{D}_2 \cup \cdots \cup \mathcal{D}_l$ is the set of all k-sets of $[n]$. That is, $\mathsf{LS}(l;\, t,k,n)$ is a partition of the set of k-sets of $[n]$ into t-(n,k,λ) designs, where necessarily $\lambda = \binom{n-t}{k-t}/l$.

The ***existence problem*** for large sets of designs can be phrased as follows: for which values of l,t,k,n do $\mathsf{LS}(l;\, t,k,n)$ large sets exist? The ***existence conjecture*** for large sets, formulated for example in [23, Conjecture 1.4], asserts that for every fixed l,t,k with $k \geqslant t+1$, a large set $\mathsf{LS}(l;\, t,k,n)$ exists for all sufficiently large n that satisfy the obvious divisibility constraints (see Section 1.2). However, according to [23, p. 564] as well as more recent surveys, "not many results about $\mathsf{LS}(l;\, t,k,n)$ with $k > t+1$ are known." One of our main results herein is a proof of the foregoing existence conjecture for all $k > 9t$.

1.1. Large sets of designs

Combinatorial design theory can be traced back to the work of Euler, who introduced the famous "36 officers problem" in 1782. Euler's ideas were further developed in the mid-19th century by Cayley, Kirkman, and Steiner. In particular, the existence problem for large sets of designs was first considered in 1850 by Cayley [1], who found two disjoint 2-$(7,3,1)$ designs and showed that no more exist. The first nontrivial large set, namely $\mathsf{LS}(7;\, 2,3,9)$, was constructed by Kirkman [8] in the same year. Following these results, the existence problem for large sets of type $\mathsf{LS}(n{-}2;\, 2,3,n)$ — that is, large sets of Steiner triple systems — attracted considerable research attention. Nevertheless, this problem remained open until the 1980s, when it was settled by Lu [10,11] and Teirlinck [22]. Specifically, it is shown in [10,11,22] that $\mathsf{LS}(n{-}2;\, 2,3,n)$ exist for all $n \geqslant 9$ with $n \equiv 1,3 \pmod 6$. In 1987, came the celebrated work of Teirlinck [20], who proved that nontrivial t-(n,k,λ) designs exist for all values of t. In fact, Teirlinck's proof of this theorem in [20] proceeds by constructing for all $t \geqslant 1$, a large set $\mathsf{LS}(l;\, t,t+1,n)$, where $l = (n-t)/(t+1)!^{(2t+1)}$. His results in [20,21] further imply that for all fixed t,k with $k \geqslant t{+}1$, nontrivial large sets $\mathsf{LS}(l;\, t,k,n)$ exist for infinitely many values of n. However, as mentioned earlier, it is unknown whether such large sets exist for *all sufficiently large values of* n that satisfy the necessary divisibility constraints. For much more on the history of the problem and the current state of knowledge, see the surveys [6,7,23] and references therein.

There are numerous applications of large sets of designs in discrete mathematics and computer science. For example, large sets of Steiner systems were used to construct perfect secret-sharing schemes by Stinson and Vanstone [19], and others [4,18]. An application of general large sets of designs to threshold secret-sharing schemes was proposed by Chee [2]. As another example, Chee and Ling [3] showed how large sets can be used to construct infinite families of optimal constant weight codes. As yet another example, large sets of 1-designs (also known as one-factorizations) have been used extensively in various kinds of scheduling problems — see [15, pp. 51–53] and references therein.

1.2. Divisibility constraints and our existence theorem

Consider a t-(n,k,λ) design with N blocks. It is very easy to see that every such design must satisfy certain natural divisibility constraints. For instance, every k-set of $[n]$ contains exactly $\binom{k}{t}$ many t-sets, and since every t-set is covered exactly λ times by the N blocks, we have $N\binom{k}{t} = \lambda\binom{n}{t}$. In particular, this implies that $\binom{k}{t}$ should divide $\lambda\binom{n}{t}$. Now let us fix a positive integer $s \leqslant t-1$ and restrict our attention only to those N'

blocks that contain a specific s-set of $[n]$. Since the fixed s-set can be extended to a t-set in $\binom{n-s}{t-s}$ ways and each of these t-sets is covered λ times by the N' blocks, a similar argument yields $N'\binom{k-s}{t-s} = \lambda\binom{n-s}{t-s}$. Thus $\binom{k-s}{t-s}$ should divide $\lambda\binom{n-s}{t-s}$. Altogether, this simple counting argument produces t divisibility constraints:

$$\binom{k-s}{t-s} \;\Big|\; \lambda\binom{n-s}{t-s} \qquad \text{for all } \; s = 0, 1 \ldots, t-1. \tag{1}$$

The above leads to the following natural question. Are these t divisibility conditions also *sufficient* for the existence of t-(n,k,λ) designs, at least when n is large enough? This is one of the central questions in combinatorial design theory. In a remarkable achievement, Keevash [5] was able to answer this question positively, thereby settling the *existence conjecture for combinatorial designs*. Specifically, Keevash proved that for any $k > t \geqslant 1$ and $\lambda \geqslant 1$, there is a sufficiently large $n_0 = n_0(t,k,\lambda)$ such that the following holds: for all $n \geqslant n_0$ such that n,t,k,λ satisfy the divisibility conditions in (1), there exists a t-(n,k,λ) design.

Let us now consider the divisibility conditions for large sets. A large set $\mathsf{LS}(l;\, t,k,n)$ is a partition of all k-sets of $[n]$ into t-(n,k,λ) designs. Clearly, each of these designs consists of $N = \binom{n}{k}/l = \lambda\binom{n}{t}/\binom{k}{t}$ blocks. This can be used to specify λ in terms of n,t,k,l as follows:

$$\lambda \;=\; \frac{\binom{n}{k}\binom{k}{t}}{l\binom{n}{t}} \;=\; \frac{1}{l}\binom{n-t}{k-t} \tag{2}$$

With this, the divisibility constraints (1) for the l component designs of a large set $\mathsf{LS}(l;\, t,k,n)$ can be rewritten in terms of n,t,k,l. Altogether, we conclude that the parameters of a large set $\mathsf{LS}(l;\, t,k,n)$ must satisfy the following $t+1$ divisibility constraints:

$$l\binom{k-s}{t-s} \;\Big|\; \binom{n-t}{k-t}\binom{n-s}{t-s} \qquad \text{for all } \; s = 0, 1 \ldots, t. \tag{3}$$

Note that the constraint for $s = t$ simply refers to the condition that l must divide $\binom{n-t}{k-t}$, which is clearly necessary in view of (2). Once again, this leads to the following natural question. Are these $t+1$ divisibility conditions also *sufficient* for the existence of $\mathsf{LS}(l;\, t,k,n)$ large sets, at least when n is large enough?

One of our main results in this paper is a positive answer to this question for all $k > 9t$, which settles the existence conjecture for large sets for such values of k. We formulate this result as the following theorem.

Theorem 1. *For any $t \geqslant 1, k > 9t$ and $l \geqslant 1$, there is an $n_0 = n_0(t,k,l)$ such that the following holds: for all $n \geqslant n_0$ such that n,t,k,l satisfy the divisibility conditions in* (3)*, there exists an $\mathsf{LS}(l;\, t,k,n)$ large set.*

In fact, Theorem 1 follows as a special case of a more general statement — namely, Theorem 9 of Section 1.4. Theorem 9 itself follows by extending and strengthening the probabilistic argument of Kuperberg, Lovett, and Peled [9]. We begin by describing the general framework for this probabilistic argument below.

1.3. General framework

Throughout this work, we will use the notation of the Kuperberg, Lovett, and Peled paper [9], which we shorthand as KLP. Let B, A be finite sets and let $\phi : B \to \mathbb{Z}^A$ be a vector valued function. One can think of

ϕ as described by a $|B| \times |A|$ matrix where the rows correspond to the evaluation of the function ϕ on the elements in B. In this setting [9] gives sufficient conditions for the existence of a small set $T \subset B$ such that

$$\frac{1}{|T|}\sum_{t\in T}\phi(t) = \frac{1}{|B|}\sum_{b\in B}\phi(b). \tag{4}$$

In the context of designs we can think of B as all the k-sets of $[n]$ and A as all the t-sets of $[n]$. ϕ denotes the inclusion function, that is $\phi(b)_a = 1_{a\subset b}$ where b is a k-set of $[n]$ and a is a t-set of $[n]$. Equation (5) is then equivalent to T being a t-(n,k,λ) design for an appropriate λ.

Next, we present the conditions under which KLP showed that there is a solution for (5). We start with a few useful notations. For $a \in A$ we denote by $\phi_a \in \mathbb{Z}^B$ the a-column of the matrix described by ϕ, namely $(\phi_a)_b = \phi(b)_a$. Let $V \subset \mathbb{Q}^B$ be the vector space spanned by the columns of this matrix $\{\phi_a : a \in A\}$. Observe that (5) depends only on V and not on $\{\phi_a : a \in A\}$, which is a specific choice of basis for V. We identify $f \in V$ with a function $f : B \to \mathbb{Q}$. Thus, we may reformulate (5) as

$$\frac{1}{|T|}\sum_{t\in T} f(t) = \frac{1}{|B|}\sum_{b\in B} f(b) \quad \forall f \in V. \tag{5}$$

In particular, we may assume without loss of generality that $\dim(V) = |A|$.

The conditions and results outlined below will depend only on the subspace V. However, it will be easier to present some of them with a specific choice of basis. We may assume this to be an integer basis. Thus, we assume throughout that $\phi : B \to \mathbb{Z}^A$ is a map whose coordinate projections $\phi_a : B \to \mathbb{Z}$ are a basis for V.

1.3.1. Divisibility conditions

For T to be a valid set for (5) with $|T| = N$, we must have

$$\sum_{t\in T} f(t) = \frac{N}{|B|}\sum_{b\in B} f(b) \quad \forall f \in V.$$

In particular there must exist $\gamma \in \mathbb{Z}^B$ such that

$$\sum_{b\in B} \gamma_b f(b) = \frac{N}{|B|}\sum_{b\in B} f(b) \quad \forall f \in V. \tag{6}$$

The set of integers N satisfying (6) consists of all integer multiples of some minimal positive integer c_1. This is because if N_1 and N_2 are solutions then so is $N_1 - N_2$. Thus it follows that $|T|$ must be an integer multiple of c_1. This is the divisibility condition and c_1 is the divisibility parameter of V.

We can rephrase (6) as $\frac{N}{|B|}\sum_{b\in B}\phi(b)$ belongs to the lattice spanned by $\{\phi(b) : b \in B\}$.

Definition 2 (Lattice spanned by ϕ)**.** *We define $\mathcal{L}(\phi)$ to be the lattice spanned by $\{\phi(b) : b \in B\}$.*

$$\mathcal{L}(\phi) = \Big\{ \sum_{b\in B} n_b \cdot \phi(b) : n_b \in \mathbb{Z} \Big\} \subset \mathbb{Z}^A.$$

Note that since we assume that $\dim(V) = |A|$ we have that $\mathcal{L}(\phi)$ is a full rank lattice.

Definition 3 (Divisibility parameter c_1)**.** *The divisibility parameter of V is the minimal integer $c_1 \geqslant 1$ that satisfies $\frac{c_1}{|B|}\sum_{b\in B}\phi(b) \in \mathcal{L}(\phi)$. Note that it does not depend on the choice of basis for V which defines ϕ.*

1.3.2. Boundedness conditions

The second condition is about boundedness conditions for integer vectors which span V and its orthogonal dual. We start with some general definitions. Let $1 \leqslant p \leqslant \infty$. The ℓ_p norm of a vector $\gamma \in \mathbb{Z}^B$ is $\|\gamma\|_p = (\sum_{b \in B} |\gamma_b|^p)^{1/p}$. Below we restrict our attention to $\|\gamma\|_1 = \sum_{b \in B} |\gamma_b|$ and $\|\gamma\|_\infty = \max_{b \in B} |\gamma_b|$.

Definition 4 (Bounded integer basis)**.** *Let $W \subset \mathbb{Q}^B$ be a vector space. For $1 \leqslant p \leqslant \infty$, we say that W has a c-bounded integer basis in ℓ_p if W is spanned by integer vectors whose ℓ_p norm is at most c. That is, if*

$$Span(\{\gamma \in W \cap \mathbb{Z}^B : \|\gamma\|_p \leqslant c\}) = W.$$

Recall that $V \subset \mathbb{Q}^B$ is the vector space spanned by $\{\phi_a : a \in A\}$. We denote by $V^\perp$ the orthogonal complement of V in $\mathbb{Q}^B$, that is,

$$V^\perp := \{g \in \mathbb{Q}^B : \sum_{b \in B} f(b)g(b) = 0 \quad \forall f \in V\}.$$

Definition 5 (Boundedness parameters c_2, c_3)**.** *We impose two boundedness conditions:*

- *Let $c_2 \geqslant 1$ be such that V has a c_2-bounded integer basis in ℓ_∞.*
- *Let $c_3 \geqslant 1$ be such that $V^\perp$ has a c_3-bounded integer basis in ℓ_1.*

1.3.3. Symmetry conditions

Next we require some symmetry conditions from the space V. Given a permutation $\pi \in S_B$ and a vector $f \in \mathbb{Q}^B$, we denote by $\pi(f) \in \mathbb{Q}^B$ the vector obtained by permuting the coordinates of f, namely $\pi(f)_b = f_{\pi(b)}$.

Definition 6 (Symmetry group of V)**.** *The symmetry group of V, denoted $Sym(V)$, is the set of all permutations $\pi \in S_B$ which satisfy that $\pi(f) \in V$ for all $f \in V$.*

It is easy to verify that $\mathrm{Sym}(V)$ is a subgroup of S_B, the symmetric group of permutations on B. Note that the condition $\pi \in \mathrm{Sym}(V)$ can be equivalently case as the existence of an invertible linear map $\tau : \mathbb{Q}^A \to \mathbb{Q}^A$ such that

$$\phi(\pi(b)) = \tau(\phi(b)) \qquad \forall\, b \in B.$$

Definition 7 (Transitive symmetry group)**.** *The symmetry group of V is said to be transitive if it acts transitively on B. That is, for every $b_1, b_2 \in B$ there is $\pi \in Sym(V)$ such that $\pi(b_1) = b_2$.*

1.3.4. Constant functions condition

The last condition is very simple: we require that the constant functions belong to V.

1.3.5. Main theorem of KLP

We are now at a position to state the main theorem of KLP [9].

Theorem 8 (KLP Theorem)**.** *Let B be a finite set and let $V \subset \mathbb{Q}^B$ be the subspace of functions. Assume that the following holds for some integers $c_1, c_2, c_3 \geqslant 1$:*

- *Divisibility:* c_1 *is the divisibility parameter of* V.
- *Boundedness of* V: V *has a* c_2*-bounded integer basis in* ℓ_∞.
- *Boundedness of* $V^\perp$: $V^\perp$ *has a* c_3*-bounded integer basis in* ℓ_1.
- *Symmetry: The symmetry group of* V *is transitive.*
- *Constant functions: The constant functions belong to V.*

Let N *is an integer multiple of* c_1 *satisfying*

$$min(N, |B| - N) \geqslant C \cdot c_2 c_3^2 dim(V)^6 \log(2c_3 dim(V))^6,$$

where $C > 0$ *is an absolute constant. Then there exists a subset* $T \subset B$ *of size* $|T| = N$ *satisfying*

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

1.4. Our main theorem

Our main result is an extension of the KLP theorem (Theorem 8) to large sets. It will have many of the same conditions, except that we need to update the divisibility condition to require the size of each design to be $N = |B|/\ell$. Thus the new divisibility condition is

$$\frac{1}{l}\sum_{b\in B}\phi(b) \in \mathcal{L}(\phi).$$

Note that as before, this condition depends only on V; it does not depend on the choice of basis for V which defines ϕ.

Theorem 9 (Main theorem)**.** *Let* B *be a finite set and let* $V \subset \mathbb{Q}^B$ *be the subspace of functions. Let also* $l \geqslant 1$ *be an integer. Assume that the following holds for some integers* $c_2, c_3 \geqslant 1$*:*

- *Divisibility:* $\frac{1}{l}\sum_{b\in B}\phi(b) \in \mathcal{L}(\phi)$.
- *Boundedness of* V: V *has a* c_2*-bounded integer basis in* ℓ_∞.
- *Boundedness of* $V^\perp$: $V^\perp$ *has a* c_3*-bounded integer basis in* ℓ_1.
- *Symmetry: The symmetry group of* V *is transitive.*
- *Constant functions: The constant functions belong to V.*

Assume furthermore that

$$|B| \geqslant C \dim(V)^6 l^6 c_3^3 \log^3(\dim(V) c_2 c_3 l),$$

for some absolute constant $C > 0$*. Then there exists a partition of* B *to* $T_1, \ldots, T_l$*, each of size* $|T_i| = |B|/l$ *such that*

$$\sum_{t\in T_i}\phi(t) = \frac{1}{l}\sum_{b\in B}\phi(b) \quad \textit{for all} \quad i = 1, \ldots, l.$$

Theorem 1 follows as a special case of Theorem 9.

Proof of Theorem 1. To recall, in this setting we have B the set of all k-sets of $[n]$, A the set of all t-sets of $[n]$, $\phi : B \to \{0,1\}^A$ given by inclusion $\phi(b)_a = 1_{a \subset b}$ for $a \in A, b \in B$ and V the subspace spanned by $\{\phi_a : a \in A\}$.

KLP [9] showed (see Section 3.3 in the arxiv version) that in this setting, the subspace V has a transitive symmetric group, it contains the constant functions, and it has boundedness parameters $c_2 = 1, c_3 \leqslant (4en/t)^t$. Furthermore, the condition that the vector $\bar{\lambda} = (\lambda, \ldots, \lambda) \in \mathcal{L}(\phi)$ is equivalent to the set of conditions

$$\binom{k-s}{t-s} \Big| \lambda \binom{n-s}{t-s} \quad \text{for all} \quad s = 0, \ldots, t.$$

(see Theorem 3.7 in [9]). In particular in our case $\lambda = \binom{n-t}{k-t}/l$ and hence the divisibility conditions in Theorem 9 are equivalent to the necessary divisibility conditions given in (3). To obtain the lower bound on $|B|$, lets fix k, t, l and let n be large enough. Then $|B| \approx n^k$, $\dim(V) \approx n^t$ and $c_3 \approx n^t$. Then if $k > 9t$ and n is large enough the lower bound on B holds. $\square$

1.5. Proof overview

The high level idea, similar to [9], is to analyze the natural random process and show that with positive (yet exponentially small) probability a desired event occurs.

Say that a subset $T \subset B$ is "uniform" if

$$\frac{1}{|T|} \sum_{b \in T} \phi(b) = \frac{1}{|B|} \sum_{b \in B} \phi(b).$$

Equivalently, the "tests" defined by V cannot distinguish the uniform distribution over T from the uniform distribution over B.

Let $\tau : B \to [l]$ be a uniform partition of B into l sets. Let $T_i = \tau^{-1}(i)$ be the induced partition for $i = 1, \ldots, l$. We would like to analyze the event that each part is uniform. That is, we would like to show that

$$\Pr[T_1, \ldots, T_l \text{ are uniform}] > 0. \tag{7}$$

Notice that under the same notations, the main result of [9] can be formulated as

$$\Pr[T_1 \text{ is uniform}] > 0.$$

The random process can be modeled as a random walk on a lattice. For $i = 1, \ldots, l$ let $X_i = \sum_{b \in T_i} \phi(b)$ be random variables taking values in $\mathbb{Z}^A$. Let $\lambda = \mathbb{E}[X_1] = \ldots = \mathbb{E}[X_l] \in \mathbb{Q}^{|A|}$. Note that if $X_1 = \ldots = X_{l-1} = \lambda$ then also $X_l = \lambda$. Let $X = (X_1, \ldots, X_{l-1}) \in \mathbb{Z}^{(l-1)|A|}$. Thus we can reformulate (7) as

$$\Pr[X = \mathbb{E}[X]] > 0. \tag{8}$$

Recall that each random variable X_i takes values in a full-dimensional sub-lattice of $\mathbb{Z}^A$ which we denoted $\mathcal{L}(\phi)$. One can show that X takes values in the lattice $\mathcal{L}(\phi)^{\otimes(l-1)}$, which is a full dimensional lattice in $\mathbb{Q}^{(l-1)|A|}$. In order to study the distribution of X, we apply a local central limit theorem. The same approach was applied in [9] in order to analyze the individual distribution of each X_i. Here, we extend the method to analyze their joint distribution, namely the distribution of X. This is accomplished by a careful analysis of the Fourier coefficients of X, which in turn relies on “coding theoretic” properties of the space V. Given this coding theoretic properties, we show that $\Pr[X = \mathbb{E}[X]]$ can be approximated by the density of a gaussian process with the same first and second moment as X at the point $\mathbb{E}[X]$. In particular, it is positive, which establishes the existence result.

1.6. Broader perspective

The current work falls into the regime of “rare events” in probabilistic analysis. It is very common that the probabilistic method, when applied to show that certain combinatorial objects exist (such as expander graphs, error correcting codes, etc) shows that a random sample succeeds with high probability. The challenge then shifts to obtaining explicit constructions of such objects, with efficient algorithmic procedures whenever relevant (e.g. efficient decoding algorithms for codes).

However, there are several scenarios where the “vanilla” probabilistic method fails, and one is forced to develop much more fine tuned techniques to prove existence of the desired combinatorial objects. The current work falls into the regime where the random process is the natural one, but the analysis is much more delicate. Other examples of similar instances are the constructive proof of the Lovász local lemma (see e.g. [16,17]), the works on interlacing families of polynomials (see e.g. [13,14]), and the entire field of discrepancy theory (see e.g. the book [12]). In each such instance, new methods were developed to prove existence of the relevant objects, that go beyond simple probabilistic analysis.

There are several families of problems in combinatroics, for which the only known constructions are explicit and of algebraic or combinatorial nature. For example, this is the case for all types of local codes (such as locally testable, decodable, or correctable codes; PIR schemes; batch codes, and so on). It is also the case for Zarenkiewicz-type Ramsey problems in graph theory, about maximal bipartite graphs without certain induced subgraphs. Another well known example is the existence of Hadamard matrices. The lack of a probabilistic model for a solution may be seen as the reason why the existential results known for these problems are very sparse and ad-hoc.

In the current work, we show that for the problem of existence of large sets, one can move beyond explicit ad-hoc constructions, such as the one of Teirlinck [22], to a more rigorous understanding of when existence of large sets is possible. Of course, the next step in this line of research, after existence has been established, is to find explicit constructions. We leave this question for future research. Another question is whether the existence result can be established to the full spectrum of parameters, namely $k \geqslant t+1$ and any $\ell \geqslant 1$ (recall that our result requires that $k > 9t$). This seems to be possible by replacing the gaussian estimate by an estimate which uses higher moments of the distribution of the random variable being analyzed. We leave this also for future research.

2. Preliminaries

Recall that $\phi : B \to \mathbb{Z}^A$ is a map, whose coordinate projections are $\phi_a : B \to \mathbb{Z}$. We defined V to be the subspace of $\mathbb{Q}^B$ spanned by $\{\phi_a : a \in A\}$. We may assume that that these form a basis for V, and hence $\dim(V) = |A|$.

Let $\tau : B \to [l]$ be a mapping that partitions B into l bins. Let $T_i := \{b \in B : \tau(b) = i\}$ for $i \in [\ell]$ be the induced partition of B. In order to prove Theorem 9 we are looking for a τ for which

$$\sum_{b \in T_i} \phi(b) = \frac{1}{l} \sum_{b \in B} \phi(b) \quad \text{for all} \quad i = 1, \ldots, l. \tag{9}$$

Note that it suffices to require that (9) holds for $i = 1, \ldots, l-1$, as then it automatically also holds for $i = l$. So from now on we only require that (9) holds for the first $l-1$ bins. We will choose a uniformly random mapping τ, and show that (9) holds with a positive probability.

We start with some definitions. Let $\Phi : B \times [l] \to \mathbb{Z}^{(l-1)|A|}$ be defined as follows. $\Phi(b, i) = (x_1, \ldots, x_{l-1})$, where $x_1, \ldots, x_{l-1} \in \mathbb{Z}^A$ are given by $x_j = \phi(b) \cdot 1_{i=j}$. Note that in particular $\Phi(b, l) = 0$. Next, define a random variable $X \in \mathbb{Z}^{(l-1)|A|}$ as

$$X := \sum_{b \in B} \Phi(b, \tau(b)).$$

The mean of X is

$$\mathbb{E}[X] = \left(\frac{1}{l} \sum_{b \in B} \phi(b), \ldots, \frac{1}{l} \sum_{b \in B} \phi(b) \right) \in \mathbb{Q}^{(l-1)|A|}.$$

Thus, proving Theorem 9 is equivalent to showing that

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

We start by computing the covariance matrix of X.

Claim 10. *The covariance matrix of X is the $(l-1)|A| \times (l-1)|A|$ matrix*

$$\Sigma[X] = R \otimes M$$

where R is the $|A| \times |A|$ matrix

$$R_{a,a'} = \sum_{b \in B} \phi(b)_a \phi(b)_{a'}$$

and M is the $(l-1) \times (l-1)$ matrix

$$M = \frac{1}{l^2} \begin{bmatrix} (l-1) & -1 & \ldots & -1 \\ -1 & (l-1) & \ldots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \ldots & (l-1) \end{bmatrix}.$$

Proof. The random variables $\{\Phi(b, \tau(b)) : b \in B\}$ are independent, thus their contribution to the covariance matrix of X is additive. Fix $b \in B$. We compute the contribution of $\Phi(b, \tau(b))$ to the $(a, i), (a', i')$ entry of $\Sigma[X]$, where $a, a' \in A$ and $i, i' \in [l-1]$. The second moment is

$$\mathbb{E}_{\tau}[\Phi(b, \tau(b))_{a,i} \cdot \Phi(b, \tau(b))_{a',i'}] = \frac{1}{l} \phi(b)_a \phi(b)_{a'} \cdot 1_{i=i'}.$$

The expectation product is

$$\mathbb{E}_{\tau}[\Phi(b, \tau(b))_{a,i}] \cdot \mathbb{E}_{\tau}[\Phi(b, \tau(b))_{a',i'}] = \frac{1}{l^2} \phi(b)_a \phi(b)_{a'}.$$

Thus

$$\Sigma[X]_{(a,i),(a',i')} = \sum_{b\in B} \phi(b)_a \phi(b)_{a'} \left(\frac{1}{l}\cdot 1_{i=i'} - \frac{1}{l^2}\right) = R_{a,a'}\cdot M_{i,i'} = (R\otimes M)_{(a,i),(a',i')}.$$

□

Similar to the proof in KLP we would be interested in the lattice in which X resides. Recall that $\mathcal{L}(\phi)$ is the lattice in $\mathbb{Z}^{|A|}$ spanned by the image of ϕ. We similarly define $\mathcal{L}(\Phi)$.

Definition 11 (Lattice spanned by Φ)**.** *We define $\mathcal{L}(\Phi)$ to be the lattice spanned by $\{\Phi(b,i) : b \in B, i \in [l]\}$. Namely,*

$$\mathcal{L}(\Phi) := \left\{ \left(\sum_{b_1\in B} n_{b_1}\cdot\phi(b_1), .., \sum_{b_j\in B} n_{b_j}\cdot\phi(b_j), .., \sum_{b_{l-1}\in B} n_{b_{l-1}}\cdot\phi(b_{l-1}) \right) : n_{b_j}\in\mathbb{Z}, j\in[l-1] \right\}.$$

Note that since $\dim(V) = |A|$ then $\mathcal{L}(\phi)$ is a full rank lattice in $\mathbb{Z}^{|A|}$. Hence $\mathcal{L}(\Phi) = \mathcal{L}(\phi)^{\otimes(l-1)}$ is a full rank lattice in $\mathbb{Z}^{(l-1)|A|}$.

Similar to KLP we use Fourier analysis to study the distribution of X. The Fourier transform of X is the function $\widehat{X} : \mathbb{R}^{(l-1)|A|} \to \mathbb{C}$ defined by

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

Note that $\widehat{X}$ is periodic. Concretely, let $L(\Phi)$ denote the dual lattice to $\mathcal{L}(\Phi)$,

$$L(\Phi) := \left\{\Theta\in\mathbb{R}^{(l-1)A} : \langle\Lambda,\Theta\rangle\in\mathbb{Z} \quad \forall\Lambda\in\mathcal{L}(\Phi)\right\}.$$

Note that if $\Theta\in L(\Phi)$ then $\widehat{X}(\Theta+\Theta') = \widehat{X}(\Theta')$ for all $\Theta'\in\mathbb{R}^{(l-1)|A|}$, and $\widehat{X}(\Theta) = 1$ iff $\Theta\in L(\Phi)$. As $\mathcal{L}(\Phi)$ is a full rank lattice it follows that $L(\Phi)$ is also be a full rank lattice and $\det(\mathcal{L}(\Phi))\det(L(\Phi)) = 1$. Thus studying $\widehat{X}$ on any fundamental domain of $L(\Phi)$ would be sufficient to study the behavior of $\widehat{X}$ on $\mathbb{R}^{(l-1)|A|}$. Similar to KLP we work with a natural fundamental domain defined by a norm related to the covariance matrix of X.

Definition 12 (R-norm)**.** *For $\Theta = (\theta_1, ..., \theta_{l-1}) \in \mathbb{R}^{(l-1)A}$ we define the norm $\|\cdot\|_R$ as*

$$\|\Theta\|_R := \max_{j\in[l-1]} \left(\frac{1}{|B|}\theta_j^t R\theta_j\right)^{1/2} = \max_{j\in[l-1]} \left(\frac{1}{|B|}\sum_{b\in B}\langle\phi(b),\theta_j\rangle^2\right)^{1/2}.$$

We define two related notions. Balls around zero in the R-norm are defined as

$$\mathcal{B}_R(\varepsilon) := \{\Theta\in\mathbb{R}^{(l-1)A} : \|\Theta\|_R \leqslant \varepsilon\}.$$

The Voronoi cell of 0 in the R-norm, with respect to the dual lattice $L(\Phi)$, is

$$D := \left\{\Theta\in\mathbb{R}^{(l-1)A} : \|\Theta\|_R < \|\Theta-\alpha\|_R \quad \forall\alpha\in L(\Phi)\setminus\{0\}\right\}.$$

Observe that D is a fundamental domain of $L(\Phi)$ up to a set of measure zero (its boundary), which we can ignore in our calculations. Then we have the following Fourier inversion formula over lattices: for every $\Gamma \in \mathcal{L}(\Phi)$ it holds that

$$\Pr[X = \Gamma] = \frac{1}{\mathrm{vol}(D)} \int_D \widehat{X}(\Theta) e^{-2\pi i \langle \Gamma, \Theta \rangle} d\Theta = \det(\mathcal{L}(\Phi)) \int_D \widehat{X}(\Theta) e^{-2\pi i \langle \Gamma, \Theta \rangle} d\Theta. \tag{11}$$

Note that this formula holds true for any fundamental region of $L(\Phi)$ but we chose it to be the Voronoi cell D arising from the norm $\|\cdot\|_R$ because it would help in the computations later on. In order to prove (10), we specialize (11) to $\Gamma = \mathbb{E}[X]$ and obtain

$$\Pr[X = \mathbb{E}[X]] = \det(\mathcal{L}(\Phi)) \int_D \widehat{X}(\Theta) e^{-2\pi i \langle \mathbb{E}[X], \Theta \rangle} d\Theta. \tag{12}$$

In the next section, we approximate this by a Gaussian estimate.

3. Gaussian estimate

In order to estimate (12), let Y be a Gaussian random variable in $\mathbb{R}^{(l-1)|A|}$ with the same mean and covariance as X. The density f_Y of Y is given by

$$f_Y(x) = \frac{\exp(-\frac{1}{2}(x - \mathbb{E}[X])^t \Sigma[X]^{-1} (x - \mathbb{E}[X]))}{(2\pi)^{\frac{(l-1)|A|}{2}} \sqrt{\det(\Sigma[X])}}. \tag{13}$$

The Fourier transform of Y equals

$$\widehat{Y}(\Theta) := \mathbb{E}[e^{2\pi i \langle Y, \Theta \rangle}] = e^{2\pi i \langle \mathbb{E}[X], \Theta \rangle - 2\pi^2 \Theta^t \Sigma[X] \Theta}. \tag{14}$$

The inverse Fourier transform applied to Y yields

$$f_Y(x) = \int_{\mathbb{R}^{(l-1)A}} \widehat{Y}(\Theta) e^{-2\pi i \langle x, \Theta \rangle} d\Theta \qquad \forall x \in \mathbb{R}^{(l-1)A}. \tag{15}$$

We show that $\Pr[X = \mathbb{E}[X]]$ can be approximated by an appropriate scaling of $f_Y(\mathbb{E}[X])$. By (12) we have

$$\frac{\Pr[X = \mathbb{E}[X]]}{\det(\mathcal{L}(\Phi))} - f_Y(\mathbb{E}[X]) = \int_D \widehat{X}(\Theta) e^{-2\pi i \langle \mathbb{E}[X], \Theta \rangle} d\Theta - \int_{\mathbb{R}^{(l-1)A}} \widehat{Y}(\Theta) e^{-2\pi i \langle \mathbb{E}[X], \Theta \rangle} d\Theta.$$

Note that by plugging $x = \mathbb{E}[X]$ in (13) we obtain that

$$f_Y(\mathbb{E}[X]) = \frac{1}{(2\pi)^{\frac{(l-1)|A|}{2}} \sqrt{\det(\Sigma[X])}}. \tag{16}$$

We will show that $|\frac{\Pr[X=\mathbb{E}[X]]}{\det(\mathcal{L}(\Phi))} - f_Y(\mathbb{E}[X])| \ll f_Y(\mathbb{E}[X])$. For $\varepsilon > 0$ to be chosen later, we will bound it by

$$\left| \frac{\Pr[X = \mathbb{E}[X]]}{\det(\mathcal{L}(\Phi))} - f_Y(\mathbb{E}[X]) \right| \leqslant \underbrace{\int_{\mathcal{B}_R(\varepsilon)} |\widehat{X}(\Theta) - \widehat{Y}(\Theta)| d\Theta}_{=I_1} + \underbrace{\int_{D \setminus \mathcal{B}_R(\varepsilon)} |\widehat{X}(\Theta)| d\Theta}_{=I_2} + \underbrace{\int_{\mathbb{R}^{(l-1)A} \setminus \mathcal{B}_R(\varepsilon)} |\widehat{Y}(\Theta)| d\Theta}_{=I_3}. \tag{17}$$

At a high level, the upper bound is obtained by comparing $\widehat{X}(\Theta)$ and $\widehat{Y}(\Theta)$ in a small enough ball; and upper bounding their absolute value outside this ball. Observe that we need ε to be small enough so that $\mathcal{B}_R(\varepsilon) \subset D$.

3.1. Norms on $\mathbb{R}^{|A|}$ induced by ϕ

The following key technical lemmas from [9] are very useful in bounding the integrals. We begin with defining few norms which are all functions of ϕ.

Definition 13 (Norms on $\mathbb{R}^{|A|}$ induced by ϕ)**.** *For $\theta \in \mathbb{R}^{|A|}$ define the following norms:*

- $\|\theta\|_{\phi,\infty} = \max_{b\in B} |\langle \phi(b), \theta\rangle|$.
- $\|\theta\|_{\phi,2} = \left(\frac{1}{|B|}\sum_{b\in B} |\langle \phi(b), \theta\rangle|^2\right)^{1/2}$.

Furthermore, for $b \in B$ let $\langle \phi(b), \theta\rangle = n_b + r_b$ where $n_b \in \mathbb{Z}$ and $r_b \in [-1/2, 1/2)$. Define

- $|||\theta|||_{\phi,\infty} = \max_{b\in B} |r_b|$.
- $|||\theta|||_{\phi,2} = \left(\frac{1}{|B|}\sum_{b\in B} |r_b|^2\right)^{1/2}$.

Note that if $\theta' \in L(\phi)$ then $\langle \phi(b), \theta + \theta'\rangle - \langle \phi(b), \theta\rangle \in \mathbb{Z}$ for all $b \in B$. In particular, $|||\theta + \theta'|||_{\phi,\infty} = |||\theta|||_{\phi,\infty}$ and $|||\theta + \theta'|||_{\phi,2} = |||\theta|||_{\phi,2}$. The following lemmas from [9] relates the above norms.

Lemma 14 (Lemma 4.4 in [9])**.** *For every $\theta \in \mathbb{R}^A$ it holds that*

$$\|\theta\|_{\phi,\infty} \leqslant M\|\theta\|_{\phi,2}$$

and

$$|||\theta|||_{\phi,\infty} \leqslant M|||\theta|||_{\phi,2}.$$

Here, $M := C\left(|A|\log(2c_2|A|)\right)^{3/2}$ for some absolute constant $C > 0$.

Lemma 15 (Claim 4.12 in [9])**.** *Assume that for $\theta \in \mathbb{R}^A$ it holds that*

$$|||\theta|||_{\phi,\infty} \leqslant \frac{1}{c_3}.$$

Then there exists $\theta' \in L(\phi)$ such that $\langle \theta - \theta', \phi(b)\rangle \in [-1/2, 1/2]$ for all $b \in B$. In particular

$$\left\|\theta - \theta'\right\|_{\phi,2} = |||\theta|||_{\phi,2}.$$

3.2. Norms on $\mathbb{R}^{(l-1)|A|}$ induced by Φ

We extend the previous definitions to norms on $\mathbb{R}^{(l-1)|A|}$ induced by Φ, and prove related lemmas relating the different norms.

Definition 16 (Generalizing the norms to $\mathbb{R}^{(l-1)|A|}$)**.** *For $\Theta = (\theta_1, \ldots, \theta_{l-1}) \in \mathbb{R}^{(l-1)|A|}$ define the following norms:*

- $\|\Theta\|_{\Phi,\infty} = \max_{j\in[l-1]} \left\|\theta_j\right\|_{\phi,\infty}$
- $\|\Theta\|_{\Phi,2} = \max_{j\in[l-1]} \left\|\theta_j\right\|_{\phi,2}$

- $|||\Theta|||_{\Phi,\infty} = \max_{j\in[l-1]} |||\theta_j|||_{\phi,\infty}$
- $|||\Theta|||_{\Phi,2} = \max_{j\in[l-1]} |||\theta_j|||_{\phi,2}$

Observe that $\|\cdot\|_{\Phi,2}$ is the same as the R-norm $\|\cdot\|_R$ we defined before. Similar to before, if $\Theta' \in L(\Phi)$ then $|||\Theta+\Theta'|||_{\Phi,\infty} = |||\Theta|||_{\Phi,\infty}$ and $|||\Theta+\Theta'|||_{\Phi,2} = |||\Theta|||_{\Phi,2}$.

The following extends Lemma 14 and Lemma 15 to the norms induced by Φ.

Lemma 17. *For the same M defined in Lemma 14, for every $\Theta \in \mathbb{R}^{(l-1)|A|}$ it holds that*

$$\|\Theta\|_{\Phi,\infty} \leqslant M\|\Theta\|_{\Phi,2}$$

and

$$|||\Theta|||_{\Phi,\infty} \leqslant M|||\Theta|||_{\Phi,2}.$$

Proof. Let $\Theta = (\theta_1, \ldots, \theta_{l-1})$. Then using Lemma 14 we have

$$\|\Theta\|_{\Phi,\infty} = \max_{j\in[l-1]} \|\theta_j\|_{\phi,\infty} \leqslant \max_{j\in[l-1]} M\|\theta_j\|_{\phi,2} = M\|\Theta\|_{\Phi,2}$$

and

$$|||\Theta|||_{\Phi,\infty} = \max_{j\in[l-1]} |||\theta_j|||_{\phi,\infty} \leqslant \max_{j\in[l-1]} M|||\theta_j|||_{\phi,2} = M|||\Theta|||_{\Phi,2}.$$

□

Lemma 18. *Assume that for $\Theta \in \mathbb{R}^{(l-1)A}$ it holds that*

$$|||\Theta|||_{\Phi,\infty} \leqslant \frac{1}{c_3}.$$

Then there exists $\Theta' \in L(\Phi)$ such that $\langle \Theta - \Theta', \Phi(b,j)\rangle \in [-1/2, 1/2]$ for all $b \in B, j \in [l-1]$. In particular

$$\|\Theta - \Theta'\|_{\Phi,2} = |||\Theta|||_{\Phi,2}.$$

Proof. Let $\Theta = (\theta_1, \ldots, \theta_{l-1})$. We have $|||\theta_j|||_{\phi,\infty} \leqslant \frac{1}{c_3}$ for all $j \in [l-1]$. Then using Lemma 15 we get that there exist $\theta'_1, \ldots, \theta'_{l-1} \in L(\phi)$ such that $\langle \theta_j - \theta'_j, \phi(b)\rangle \in [-1/2, 1/2]$ for all $b \in B$. The lemma follows for $\Theta' = (\theta'_1, \ldots, \theta'_{l-1}) \in L(\Phi)$. □

3.3. Estimates for balls in the Voronoi cell

To recall, we need $\varepsilon > 0$ to be small enough so that $\mathcal{B}_R(\varepsilon)$ is contained in the Voronoi cell D. The following Lemma utilizes Lemma 17 to achieve that.

Lemma 19. *If $\varepsilon < \frac{1}{2M}$ then $\mathcal{B}_R(\varepsilon) \subset D$.*

Proof. Let $\Theta = (\theta_1, \ldots, \theta_{l-1}) \in L(\Phi) \setminus \{0\}$. By definition $\langle \phi(b), \theta_j\rangle \in \mathbb{Z}$ for all $b \in B, j \in [l-1]$. Since $\mathcal{L}(\phi)$ is of full rank and $\Theta \neq 0$, there exists some $b \in B, j \in [l-1]$ for which $|\langle \phi(b), \theta_j\rangle| \geqslant 1$. Thus

$$\|\Theta\|_{\Phi,\infty} \geqslant 1.$$

By Lemma 17 if follows that

$$\|\Theta\|_R = \|\Theta\|_{\Phi,2} \geqslant 1/M.$$

Thus, if $\Theta' \in \mathcal{B}_R(\varepsilon)$ for $\varepsilon < 1/2M$ then

$$\|\Theta - \Theta'\|_R \geqslant \|\Theta\|_R - \|\Theta'\|_R \geqslant 1/M - \varepsilon > 1/2M \geqslant \|\Theta'\|_R.$$

Hence $\mathcal{B}_R(\varepsilon) \subset D$ for any $\varepsilon < \frac{1}{2M}$. □

Let $\Theta \in D \setminus \mathcal{B}_R(\varepsilon)$. Clearly, its $\|\cdot\|_{\Phi,2}$ norm is noticeable (at least ε). We show that also its $|||\cdot|||_{\Phi,2}$ norm is noticeable. This will later be useful in bounding $\hat{X}(\Theta)$ in $D \setminus \mathcal{B}_R(\varepsilon)$.

Lemma 20. *Assume that $c_3 \geqslant 2$ and $\varepsilon < 1/c_3 M$. Let $\Theta \in D \setminus \mathcal{B}_R(\varepsilon)$. Then $|||\Theta|||_{\Phi,2} > \varepsilon$.*

Proof. Note that the conditions of Lemma 19 hold, and so $\mathcal{B}_R(\varepsilon) \subset D$. Assume towards contradiction that $|||\Theta|||_{\Phi,2} \leqslant \varepsilon$. Applying Lemma 17 gives $|||\Theta|||_{\Phi,\infty} \leqslant \varepsilon M \leqslant \frac{1}{c_3}$. Applying Lemma 18, this implies that there exists $\Theta' \in L(\Phi)$ for which $\|\Theta - \Theta'\|_{\Phi,2} = |||\Theta|||_{\Phi,2} \leqslant \varepsilon$. However, as $\Theta \in D$ we have $\|\Theta\|_{\Phi,2} \leqslant \|\Theta - \Theta'\|_{\Phi,2} \leqslant \varepsilon$, which gives that $\Theta \in \mathcal{B}_R(\varepsilon)$, a contradiction. □

3.4. Bounding the integrals

The following lemmas provide the necessary bounds on the integrals I_1, I_2, I_3, as defined in (17). The proofs are deferred to Section 4.

Lemma 21. *Assume that $\varepsilon \leqslant (CM|B|)^{-1/3}$. Then*

$$I_1 \leqslant \frac{Cl^3 M|A|^{3/2}}{|B|^{1/2}} \cdot f_Y(\mathbb{E}[X]).$$

Here $C > 0$ is some large enough absolute constant.

Lemma 22. *Assume that $c_3 \geqslant 2$ and $\varepsilon \leqslant 1/c_3 M$. Then*

$$I_2 \leqslant \frac{1}{\det(\mathcal{L}(\Phi))} \exp\left(-\frac{|B|\varepsilon^2}{l^2}\right)$$

Lemma 23. *For any $\varepsilon > 0$ it holds that*

$$I_3 \leqslant f_Y(\mathbb{E}[X]) \cdot (l-1)2^{|A|/2} \exp\left(-\frac{\pi^2|B|\varepsilon^2}{l^2}\right).$$

3.5. Putting it all together

Let $C_1, C_2, \ldots$ be unspecified absolute constants below. By choosing an appropriate basis for V which is c_2-bounded in ℓ_∞, we may assume that $\phi : B \to \mathbb{Z}^A$ where $|\phi(b)|_a \leqslant c_2$ for all $a \in A, b \in B$.

Set $\varepsilon = (C_1 MB)^{-1/3}$ so that we may apply Lemma 21, and assume that $\varepsilon \leqslant 1/c_3 M$ so that we may apply Lemma 22. We thus have

$$\Pr[X = \mathbb{E}[X]] = \det(\mathcal{L}(\Phi)) f_Y(\mathbb{E}[X])(1 + \alpha_1 + \alpha_3) + \alpha_2,$$

where

$$|\alpha_1| = \frac{C_1 l^3 M |A|^{3/2}}{|B|^{1/2}},$$
$$|\alpha_2| = \exp\left(-\frac{|B|\varepsilon^2}{l^2}\right) = \exp\left(-C_2 \frac{|B|^{1/3}}{l^2 M^{2/3}}\right),$$
$$|\alpha_3| = (l-1)2^{|A|/2} \exp\left(-\frac{\pi^2 |B|\varepsilon^2}{l^2}\right) \leqslant l 2^{|A|} \exp\left(-C_3 \frac{|B|^{1/3}}{l^2 M^{2/3}}\right).$$

We would like that $|\alpha_1|, |\alpha_3| \leqslant 1/4$, which requires that

$$|B| \geqslant C_4 |A|^3 M^2 l^6 c_3^3$$

Thus

$$\Pr[X = \mathbb{E}[X]] \geqslant \frac{1}{2} \det(\mathcal{L}(\Phi)) f_Y(\mathbb{E}[X]) + \alpha_2.$$

We assume that $\phi : B \to \mathbb{Z}^A$, so $\mathcal{L}(\Phi)$ is an integer lattice and hence $\det(\mathcal{L}(\Phi)) \geqslant 1$. We next lower bound $f_Y(\mathbb{E}[X])$. We have by (16) that

$$f_Y(\mathbb{E}[X]) = \frac{1}{(2\pi)^{\frac{(l-1)|A|}{2}} \sqrt{\det(\Sigma[X])}}.$$

We assume that ϕ is spanned by integer vectors of maximum entry c_2, so we can bound each entry of $\Sigma[X]$ by

$$|\Sigma[X]_{(a,i),(a',i')}| \leqslant \sum_{b \in B} |\phi(b)_a \phi(b)_{a'}| \leqslant |B| c_2^2.$$

Thus

$$\det(\Sigma[X]) \leqslant (|A||B|c_2^2)^{|A|}.$$

In order to require $\alpha_2 \leqslant (1/4) f_Y(\mathbb{E}[X])$, say, we need to require that

$$|B| \geqslant C_5 |A|^3 M^2 l^6 \log(|A|Ml).$$

Putting it all together, and plugging in the value of M from Lemma 14, as long as

$$|B| \geqslant C|A|^6 l^6 c_3^3 \log^3(|A| c_2 c_3 l),$$

we have that

$$\Pr[X = \mathbb{E}[X]] \geqslant \frac{1}{4} \det(\mathcal{L}(\Phi)) f_Y(\mathbb{E}[X]) > 0.$$

4. Bounding the integrals

4.1. Bounding I_1

Recall that $I_1 = \int_{\mathcal{B}_R(\varepsilon)} |\hat{X}(\Theta) - \hat{Y}(\Theta)| d\Theta$. We will bound it by bounding pointwise the difference $|\hat{X}(\Theta) - \hat{Y}(\Theta)|$ and integrating it.

We first compute an exact formula for $\widehat{X}(\Theta)$. Recall that $X = \sum_{b \in B} \Phi(b, \tau(b))$ where $\tau(b) \in [l]$ are independently chosen. Thus

$$\widehat{X}(\Theta) = \mathbb{E}_X \left[e^{2\pi i \langle X, \Theta \rangle} \right] = \prod_{b \in B} \left[\frac{1}{l} \left(1 + \sum_{j=1}^{l-1} e^{2\pi i \langle \phi(b), \theta_j \rangle} \right) \right]. \tag{18}$$

Fix $\Theta = (\theta_1, \ldots, \theta_{l-1})$. To simplify notations, let $x_{b,j} = 2\pi \langle \phi(b), \theta_j \rangle$ and $\mathbf{x}_b = (x_{b,1} \ldots x_{b,l-1}) \in \mathbb{R}^{l-1}$. Define the function $f : \mathbb{R}^{l-1} \to \mathbb{C}$ given by $f(\mathbf{x}) = \frac{1}{l} \left(1 + \sum_{j=1}^{l-1} e^{ix_j} \right)$. Then we can simplify (18) as

$$\widehat{X}(\Theta) = \prod_{b \in B} f(\mathbf{x}_b). \tag{19}$$

We next approximate $\log f(\mathbf{x})$. We use the shorthand $O(z)$ to denote a (possible complex) value, whose absolute value is bounded by Cz for some unspecified absolute constant $C > 0$. For $\mathbf{x} = (x_1, \ldots, x_{l-1})$ we denote $|\mathbf{x}| = \max_j |x_j|$.

Claim 24. *Let* $\mathbf{x} = (x_1, \ldots, x_{l-1}) \in \mathbb{R}^{l-1}$ *with* $|\mathbf{x}| \leqslant 1$. *Then*

$$f(\mathbf{x}) = \exp\left(i\frac{1}{l} \sum_j x_j - \frac{1}{2l}\left(1 - \frac{1}{l}\right) \sum_j x_j^2 + \frac{1}{2l^2} \sum_{j \neq j'} x_j x_{j'} + O\left(|\mathbf{x}|^3\right) \right).$$

Proof. Let $y = \frac{1}{l} \sum_{j=1}^{l-1} (e^{ix_j} - 1)$ so that $f(\mathbf{x}) = 1 + y$. The condition $|\mathbf{x}| \leqslant 1$ guarantees that $|y| < 1$, so the Taylor expansion for $\log(1 + y)$ converges and gives

$$\log(f(\mathbf{x})) = \log(1 + y) = y - \frac{y^2}{2} + O(|y|^3).$$

One can verify that $|y| \leqslant O(|\mathbf{x}|)$, that

$$y = i\frac{1}{l} \sum_j x_j - \frac{1}{2l} \sum_j x_j^2 + O\left(|\mathbf{x}|^3\right).$$

and that

$$y^2 = -\frac{1}{l^2} \left(\sum_j x_j \right)^2 + O\left(|\mathbf{x}|^3\right).$$

Combining these gives the required result. $\square$

Applying Claim 24 to (19) allows us to approximate $\widehat{X}(\Theta)$ as

$$\widehat{X}(\Theta) = \exp\left(\frac{2\pi i}{l} \sum_{\substack{b \in B \\ j \in [l-1]}} \langle \phi(b), \theta_j \rangle - \frac{2\pi^2}{l}(1 - \frac{1}{l}) \sum_{\substack{b \in B \\ j \in [l-1]}} \langle \phi(b), \theta_j \rangle^2 + \frac{2\pi^2}{l^2} \sum_{\substack{b \in B \\ j \neq j'}} \langle \phi(b), \theta_j \rangle \langle \phi(b), \theta_{j'} \rangle + \delta(\Theta) \right),$$

which can be rephrased as

$$\widehat{X}(\theta) = \exp\left(2\pi i \langle \mathbb{E}[X], \Theta \rangle - 2\pi^2 \Theta^t \Sigma[X] \Theta + \delta(\Theta) \right). \tag{20}$$

The error term $\delta(\Theta)$ is bounded by

$$\begin{aligned}\delta(\Theta) &\leqslant O\left(\sum_{b\in B}|\mathbf{x}_b|^3\right) = O\left(\sum_{b\in B}\max_{j\in[l-1]}|\langle\phi(b),\theta_j\rangle|^3\right)\\ &\leqslant \left(\max_{b\in B, j\in[l-1]}|\langle\phi(b),\theta_j\rangle|\right)\left(\sum_{b\in B}\max_{j\in[l-1]}|\langle\phi(b),\theta_j\rangle|^2\right)\\ &= \|\Theta\|_{\Phi,\infty}\cdot|B|\|\Theta\|_{\Phi,2}^2.\end{aligned}$$

By Lemma 17 we have $\|\Theta\|_{\Phi,\infty} \leqslant M\|\Theta\|_{\Phi,2}$, and hence as $\|\Theta\|_{\Phi,2} = \|\Theta\|_R$ we conclude that

$$\delta(\Theta) \leqslant C_1 M|B|\|\Theta\|_R^3, \tag{21}$$

where $C_1 > 0$ is some absolute constant.

Next, we apply these estimates to bound the integral I_1. Recall that by (14) we have

$$\widehat{Y}(\Theta) := \exp(2\pi i\langle\mathbb{E}[X],\Theta\rangle - 2\pi^2\Theta^t\Sigma[X]\Theta).$$

Thus we can bound I_1 by

$$I_1 = \int_{\mathcal{B}_R(\varepsilon)}|\hat{X}(\Theta)-\hat{Y}(\Theta)|d\Theta \leqslant \int_{\mathcal{B}_R(\varepsilon)} e^{-2\pi^2\Theta^t\Sigma[X]\Theta}|e^{\delta(\Theta)}-1|d\Theta.$$

We assume that $\varepsilon > 0$ is small enough so that $C_1M|B|\varepsilon^3 \leqslant 1$, so that for all for $\Theta \in \mathcal{B}_R(\varepsilon)$ we have

$$|e^{\delta(\Theta)}-1| \leqslant 2\delta(\Theta) \leqslant 2C_1M|B|\|\Theta\|_R^3.$$

Thus

$$I_1 \leqslant 2C_1M|B|\int_{\mathcal{B}_R(\varepsilon)} e^{-2\pi^2\Theta^t\Sigma[X]\Theta}\|\Theta\|_R^3 d\Theta \leqslant 2C_1M|B|\int_{\mathbb{R}^{(l-1)A}} e^{-2\pi^2\Theta^t\Sigma[X]\Theta}\|\Theta\|_R^3 d\Theta.$$

Next, we evaluate the integral on the right. Let Z be a Gaussian random variable in $\mathbb{R}^{(l-1)|A|}$ with mean zero and covariance matrix $\frac{1}{4\pi^2}\Sigma[X]^{-1}$. Then the density of Z is

$$f_Z(\Theta) = (2\pi)^{\frac{(l-1)|A|}{2}}\sqrt{\det(\Sigma)}e^{-2\pi^2\Theta^t\Sigma[X]\Theta} = \frac{1}{f_Y(\mathbb{E}[X])}e^{-2\pi^2\Theta^t\Sigma[X]\Theta},$$

where we have used (16). Hence

$$\int_{\mathbb{R}^{(l-1)A}} e^{-2\pi^2\Theta^t\Sigma[X]\Theta}\|\Theta\|_R^3 d\Theta = f_Y(\mathbb{E}[X])\cdot\mathbb{E}[\|Z\|_R^3].$$

Let $G \in \mathbb{R}^{(l-1)|A|}$ be standard multivariate Gaussian with mean zero and identity covariance matrix. Recall that by Claim 10 we have $\Sigma[X] = R\otimes M$. In particular, $\Sigma[X]$ is positive definite, so its root exists. So we may take $Z = \frac{1}{2\pi}\Sigma[X]^{-1/2}G$. We have

$$\Sigma[X] = R\otimes M = R\otimes(U^tDU)$$

where D is a diagonal matrix with diagonal $(1/l^2, 1/l, \ldots, 1/l)$ and U is an orthogonal matrix. Thus

$$\Sigma[X]^{-1/2} = R^{-1/2}\otimes(U^tD^{-1/2}U).$$

Note that $D^{-1/2}$ is a diagonal matrix with diagonal $(l, \sqrt{l}, \ldots, \sqrt{l})$.

Let $G = (G_1, \ldots, G_{l-1})$ with $G_i \in \mathbb{R}^{|A|}$ and similarly $Z = (Z_1, \ldots, Z_{l-1})$ with $Z_i \in \mathbb{R}^{|A|}$. We can express $Z_1, \ldots, Z_{l-1}$ as

$$
\begin{aligned}
Z_1 &= \frac{l}{2\pi} R^{-1/2} \sum_{k=1}^{l-1} U_{1,k} G_k \\
Z_j &= \frac{\sqrt{l}}{2\pi} R^{-1/2} \sum_{k=1}^{l-1} U_{j,k} G_k \qquad j = 2, \ldots, l-1.
\end{aligned}
$$

Let $G^j = \sum_{k=1}^{l-1} U_{j,k} G_k$. Since U is an orthogonal matrix, we have that $(G^1, \ldots, G^{l-1})$ is also a standard multivariate Gaussian $\mathbb{R}^{(l-1)|A|}$ with mean zero and identity covarince matrix. Thus we have

$$
\begin{aligned}
Z_1 &= \frac{l}{2\pi} R^{-1/2} G^1 \\
Z_j &= \frac{\sqrt{l}}{2\pi} R^{-1/2} G^j \qquad j = 2, \ldots, l-1.
\end{aligned}
$$

That is, $Z_1, \ldots, Z_{l-1}$ are independent Gaussian random variables with mean zero, where Z_1 has covariance matrix $\frac{l^2}{4\pi^2} R^{-1}$ and for $j = 2, \ldots, l-1$ we have that Z_j has covariance matrix $\frac{l}{4\pi^2} R^{-1}$. We may thus bound

$$
\begin{aligned}
\mathbb{E}_Z \left[\|Z\|_R^3 \right] &= \mathbb{E}_Z \left[\max_j \left(\frac{1}{|B|} Z_j^t R Z_j \right)^{3/2} \right] \\
&\leqslant \mathbb{E}_Z \left[\sum_j \left(\frac{1}{|B|} Z_j^t R Z_j \right)^{3/2} \right] = \sum_j \mathbb{E}_Z \left[\left(\frac{1}{|B|} Z_j^t R Z_j \right)^{3/2} \right] \\
&= \left(\left(\frac{l^2}{4\pi^2 |B|} \right)^{\frac{3}{2}} + (l-2) \left(\frac{l}{4\pi^2 |B|} \right)^{\frac{3}{2}} \right) \mathbb{E} \left[\|G'\|_2^3 \right] \\
&\leqslant \frac{2l^3}{(4\pi^2)^{3/2} |B|^{3/2}} \mathbb{E} \left[\|G'\|_2^3 \right]
\end{aligned}
$$

where G' is a standard multivariate Gaussian random vector in $\mathbb{R}^A$ with mean zero and identity covariance matrix. Note that by Jensen's inequality $\mathbb{E}[\|G'\|_2^3] \leqslant \mathbb{E}[\|G'\|_2^4]^{3/4} \leqslant 4^{3/4} |A|^{3/2}$. Thus we can summarize that

$$
I_1 \leqslant O\left(\frac{l^3 M |A|^{3/2}}{|B|^{1/2}} \right) \cdot f_Y(\mathbb{E}[X]).
$$

4.2. Bounding I_2

Recall that $I_2 = \int_{D \setminus \mathcal{B}_R(\varepsilon)} |\widehat{X}(\Theta)| d\Theta$. We upper bound I_2 by proving an upper bound on $|\hat{X}(\Theta)|$ in $D \setminus \mathcal{B}_R(\varepsilon)$.

Fix $\Theta = (\theta_1, \ldots, \theta_{l-1}) \in D$ where we assume $\|\Theta\|_{\Phi,2} = \|\Theta\|_R \geqslant \varepsilon$. Our goal is to upper bound $\hat{X}(\Theta)$. Let $\langle \phi(b), \theta_j \rangle = n_{b,j} + r_{b,j}$ where $n_{b,j} \in \mathbb{Z}$ and $r_b \in [-1/2, 1/2)$. By (19) we have

$$\widehat{X}(\Theta) = \prod_{b \in B} \left[\frac{1}{l} \left(1 + \sum_{j=1}^{l-1} e^{2\pi i \langle \theta_j, \phi(b) \rangle} \right) \right] = \prod_{b \in B} \left[\frac{1}{l} \left(1 + \sum_{j=1}^{l-1} e^{2\pi i \cdot r_{b,j}} \right) \right] = \prod_{b \in B} f(2\pi \cdot \mathbf{r}_b),$$

where $f(\mathbf{x}) = \frac{1}{l} \left(1 + \sum_{j=1}^{l-1} e^{ix_j} \right)$ and $\mathbf{r}_b = (r_{b,1}, \ldots, r_{b,l-1})$. Recall that $|\mathbf{x}| = \max |x_j|$.

Claim 25. *Let $\mathbf{x} \in \mathbb{R}^{l-1}$ be with $|\mathbf{x}| \leqslant \pi$. Then $|f(\mathbf{x})| \leqslant \exp(-|\mathbf{x}|^2/8l)$.*

Proof. Let $x_j = |\mathbf{x}|$. Then $|f(\mathbf{x})| \leqslant \frac{l-2}{l} + \frac{2}{l} |\frac{1+e^{ix_j}}{2}|$. If $z \in [-\pi, \pi]$ then $|\frac{1+e^{iz}}{2}| \leqslant e^{-z^2/8}$. One can verify that

$$\log |f(\mathbf{x})| \leqslant \log \left(1 - \frac{2}{l} \left(e^{-|\mathbf{x}|^2/8} - 1 \right) \right) \leqslant -\frac{|\mathbf{x}|^2}{8l}.$$

□

Thus we have

$$\log |\widehat{X}(\Theta)| \leqslant -\frac{4\pi^2}{8l} \sum_{b \in B} |\mathbf{r}_b|^2 \leqslant -\frac{1}{l^2} \sum_{b \in B, j \in [l-1]} r_{b,j}^2 = -\frac{|B|}{l^2} |||\Theta|||_{\Phi,2}^2.$$

Next, assume that $\varepsilon \leqslant 1/c_3 M$. By Lemma 20 we have that $|||\Theta|||_{\Phi,2} \geqslant \varepsilon$. Thus

$$|\widehat{X}(\Theta)| \leqslant \exp(-|B|\varepsilon^2/l^2).$$

Thus we may bound

$$I_2 \leqslant \mathrm{vol}(D) \exp(-|B|\varepsilon^2/l^2) = \frac{1}{\det(\mathcal{L}(\Phi))} \exp(-|B|\varepsilon^2/l^2).$$

4.3. Bounding I_3

Recall that

$$I_3 = \int_{\mathbb{R}^{(l-1)A} \setminus \mathcal{B}_R(\varepsilon)} |\hat{Y}(\Theta)| d\Theta = \int_{\mathbb{R}^{(l-1)A} \setminus \mathcal{B}_R(\varepsilon)} e^{-2\pi^2 \Theta^t \Sigma[X] \Theta} d\Theta.$$

As in the calculation of the bound for I_1, let $Z \in \mathbb{R}^{(l-1)|A|}$ be Gaussian random variable with mean zero and covariance matrix $\frac{1}{4\pi^2} \Sigma[X]^{-1}$. Then

$$I_3 = f_Y(\mathbb{E}[X]) \cdot \Pr\left[\|Z\|_R > \varepsilon \right].$$

Recall that we showed that if we set $Z = (Z_1, \ldots, Z_{l-1})$, then $Z_1, \ldots, Z_{l-1} \in \mathbb{R}^A$ are independent Gaussian random variables in with mean zero, where Z_1 has covariance matrix $\frac{l^2}{4\pi^2} R^{-1}$ and Z_j has covariance matrix $\frac{l}{4\pi^2} R^{-1}$ for $j = 2, \ldots, l-1$. We may thus bound

$$\begin{aligned}\Pr\left[\|Z\|_R > \varepsilon\right] &= \Pr_Z\left[\max_j\left(\frac{1}{|B|}Z_j^t R Z_j\right) > \varepsilon^2\right] \leqslant \sum_j \Pr_{Z_j}\left[\left(\frac{1}{|B|}Z_j^t R Z_j\right) > \varepsilon^2\right] \\ &= \Pr_{G'}\left[\|G'\|_2^2 > \frac{4\pi^2|B|\varepsilon^2}{l^2}\right] + (l-2)\Pr_{G'}\left[\|G'\|_2^2 > \frac{4\pi^2|B|\varepsilon^2}{l}\right] \\ &\leqslant (l-1)\Pr_{G'}\left[\|G'\|_2^2 > \frac{4\pi^2|B|\varepsilon^2}{l^2}\right],\end{aligned}$$

where $G' \in \mathbb{R}^A$ is a Gaussian random variable with mean zero and identity covariance matrix.

In order to bound $\Pr_{G'}\left[\|G'\|_2^2 > \rho\right]$ we note that for any $t < 1/2$, it holds that $\mathbb{E}\left[e^{t\|G'\|_2^2}\right] = (1-2t)^{-|A|/2}$. Fixing $t = 1/4$ and apply the Markov inequality gives

$$\Pr_{G'}\left[\|G'\|_2^2 > \rho\right] \leqslant \frac{\mathbb{E}\left[e^{\|G'\|_2^2/4}\right]}{e^{\rho/4}} = 2^{|A|/2}e^{-\rho/4}.$$

So

$$I_3 \leqslant f_Y(\mathbb{E}[X]) \cdot (l-1)2^{|A|/2}e^{-\frac{\pi^2|B|\varepsilon^2}{l^2}}.$$

Acknowledgment

We are grateful to Yeow Meng Chee and Tuvi Etzion for helpful discussions regarding the history and the current state of knowledge about the existence problem for large sets of designs.

References

[1] A. Cayley, On the triadic arrangements of seven and fifteen things, *London Edinburgh and Dublin Philos. Mag. and J. Sci.* **37** (1850), 50–53.

[2] Y.M. Chee, *The Basis Reduction Algorithm and Existence of Combinatorial Designs*, M. Math. Thesis, Department of Computer Science, University of Waterloo, Ontario, Canada, 1989.

[3] Y.M. Chee and S. Ling, Constructions for q-ary constant-weight codes, *IEEE Trans. Inform. Theory* **53** (2007), 135–146.

[4] T. Etzion, On threshold schemes from large sets, *J. Combin. Designs* **4** (1996), 323–338.

[5] P. Keevash, The existence of designs, **arXiv:1401.3665**, January 2014.

[6] G.B. Khosrovshahi and R. Laue, t-designs with $t \geqslant 3$, in C.J. Colbourn and J.H. Dinitz, Eds., *CRC Handbook of Combinatorial Designs*, 2-nd ed., CRC Press, Boca Raton, FL., 2006, pp. 79–100.

[7] G.B. Khosrovshahi and B. Tayfeh-Rezaie, Large sets of t-designs through partitionable sets: A survey, *Discrete Math.* **306** (2006), 2993–3004.

[8] T.P. KIRKMAN, Note on an unanswered prize question, *Cambridge and Dublin Mathematical Journal* **5** (1850), 255–262.

[9] G. KUPERBERG, S. LOVETT, and R. PELED, Probabilistic existence of regular combinatorial structures, **arXiv:1302.4295**, October 2013, also in *Proc. 44-th ACM Symp. Theory of Computing* (STOC), New York, May 2012, pp. 1091–1106.

[10] J.X. LU, On large sets of disjoint Steiner triple systems: Parts I, II, and III, *J. Combin. Theory Ser. A* **34** (1983), 140–146, 147–155, and 156–182.

[11] J.X. LU, On large sets of disjoint Steiner triple systems: Parts IV, V, and VI, *J. Combin. Theory Ser. A* **37** (1984), 136–163, 164–188, and 189–192.

[12] J. MATOUSEK, Geometric discrepancy: An illustrated guide. Vol. 18. Springer Science & Business Media, 2009.

[13] A. MARCUS, D. SPIELMAN and N. SRIVASTAVA, Interlacing families I: Bipartite Ramanujan graphs of all degrees, in *Proc. of the 54-th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 529-537. IEEE, 2013.

[14] A. MARCUS, D. SPIELMAN and N. SRIVASTAVA, Interlacing families II: Mixed characteristic polynomials and the Kadison-Singer problem, arXiv preprint arXiv:1306.3969 (2013).

[15] E. MENDELSOHN and A. ROSA, One-factorizations of the complete graph: A survey, *J. Graph Theory* **9** (1985), 43–65.

[16] R. MOSER, A constructive proof of the Lovász local lemma, in *Proc. of the 41-st annual ACM symp. on Theory of computing* (STOC), 2009, pp. 343-350.

[17] R. MOSER and G. TARDOS, A constructive proof of the general Lovász Local Lemma, *Journal of the ACM (JACM)* **57**, no. 2 (2010): 11.

[18] P.J. SCHELLENBERG and D.R. STINSON, Threshold schemes from combinatorial designs, *J. Combin. Math. Combin. Comput.* **5** (1989), 143–160.

[19] D.R. STINSON and S.A. VANSTONE, A combinatorial approach to threshold schemes, *SIAM J. Discrete Math.* **1** (1988), 230–236.

[20] L. TEIRLINCK, Non-trivial t-designs without repeated blocks exist for all t, *Discrete Math.* **65** (1987), 301–311.

[21] L. TEIRLINCK, Locally trivial t-designs and t-designs without repeated blocks, *Discrete Math.* **77** (1989), 345–356.

[22] L. TEIRLINCK, A completion of Lu's determination of the spectrum of large sets of disjoint Steiner triple systems, *J. Combin. Theory Ser. A* **57** (1991), 302–305.

[23] L. TEIRLINCK, Large sets of disjoint designs and related structures, in J.H. DINITZ and D.R. STINSON, Eds., *Contemporary Design Theory*, Wiley, New York, 1992, pp. 561–592.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il