ANTI-CONCENTRATION IN MOST DIRECTIONS

ANUP RAO AND AMIR YEHUDAYOFF

Abstract. We prove anti-concentration bounds for the inner product of two independent random vectors. For example, we show that if A, B are subsets of the cube $\{\pm 1\}^n$ with $|A| \cdot |B| \geq 2^{1.01n}$, and $X \in A$ and $Y \in B$ are sampled independently and uniformly, then the inner product $\langle X, Y \rangle$ takes on any fixed value with probability at most $O(\frac{1}{\sqrt{n}})$. Extending Halász work, we prove stronger bounds when the choices for x are unstructured. We also describe applications to communication complexity, randomness extraction and additive combinatorics.

1. Introduction

Anti-concentration bounds establish that the distribution of outcomes of a random process is not concentrated in any small region. No single outcome is obtained too often. Anti-concentration plays an important role in mathematics and computer science. It is used in the study of roots of random polynomials [16], random matrix theory [13, 23], communication complexity [4, 26, 19], quantum computation [1], and more.

A well-known example is the sum of independent identically distributed random variables. If $Y \in \{\pm 1\}^n$ is uniformly distributed, then the probability that $\sum_{j=1}^{n} Y_j$ takes any specific value is at most $\binom{n}{\lceil n/2 \rceil}/2^n = O(\frac{1}{\sqrt{n}})$.

This was studied and generalized by Littlewood and Offord [16], Erdös [7], and many others. The classical Littlewood-Offord problem is about understanding the anti-concentration of the inner product $\langle x, Y \rangle = \sum_{j=1}^{n} x_j Y_j$, for arbitrary $x \in \mathbb{R}^n$ and $Y \in \{\pm 1\}^n$ chosen uniformly. Various generalizations were studied by Frankl and Furedi [10], Halász [11] and others.

It is interesting to understand the most general conditions under which anti-concentration holds (see [22] and references within). Anti-concentration certainly fails when the entries of Y are not independent.

A.R. is supported by the National Science Foundation under agreement CCF-1420268. A.Y. is partially supported by ISF grant 1162/15. This work was done while the authors were visiting the Simons Institute for the Theory of Computing.

We can, for example, sample Y uniformly from the set of strings with exactly $\lceil n/2 \rceil$ entries that are 1. Then $\sum_j Y_j$ is always the same, yet Y has almost full entropy.

Can we somehow recover anti-concentration? A natural setting is to consider the inner-product $\langle X, Y \rangle$ of two independent variables. This indeed recovers anti-concentration, as the following theorem shows.

Theorem (Chakrabarti and Regev [4])**.** *There is a constant $c > 0$ such that if $A, B \subseteq \{\pm 1\}^n$ are each of size at least $2^{(1-c)n}$ and $X \in A, Y \in B$ are sampled uniformly and independently, then*

$$\Pr[|\langle X, Y \rangle| \leq c\sqrt{n}] \leq 1 - c.$$

The theorem shows that $\langle X, Y \rangle$ does not land in an interval of length much smaller than $\sqrt{n}$ with high probability. When studying anti-concentration, however, what we are ultimately interested in is proving point-wise estimates. We would like to control the *concentration probability*

$$\max_{k \in \mathbb{Z}} \Pr[\langle X, Y \rangle = k];$$

see [23] and references within.

Here we prove a sharp bound on the concentration probability. The bound holds for an overwhelming majority of directions x.

Theorem 1. *For every $\beta > 0$ and $\delta > 0$, there exists $C > 0$ such that the following holds. If $B \subseteq \{\pm 1\}^n$ is of size $2^{\beta n}$, then for all but $2^{n(1-\beta+\delta)}$ directions $x \in \{\pm 1\}^n$,*

$$\max_{k \in \mathbb{Z}} \Pr_Y[\langle x, Y \rangle = k] \leq \tfrac{C}{\sqrt{n}}.$$

The theorem is sharp in the following sense. As mentioned above, the $O(\frac{1}{\sqrt{n}})$ bound is tight even when A and B are $\{\pm 1\}^n$. To see that the bound on the number of bad directions is sharp, observe that if $B \subset \{\pm 1\}^n$ is the set of y's where the first $(1-\beta)n$ coordinates are set to 1 and $\sum_{j > (1-\beta)n} y_j = 0$, and $A \subset \{\pm 1\}^n$ is the set of x's where the last βn coordinates are set to 1 and $\sum_{j \leq (1-\beta)n} x_j = 0$, then

$$|B| \approx \tfrac{1}{\sqrt{n}} 2^{\beta n} \quad \& \quad |A| \approx \tfrac{1}{\sqrt{n}} 2^{(1-\beta)n},$$

yet $\langle x, y \rangle = 0$ for every $x \in A$ and $y \in B$. There is no anti-concentration in this case, although $|A| \cdot |B|$ is roughly 2^n.

Our proof builds a flexible framework for proving anti-concentration results in discrete domains. We use this framework to prove a more general theorem, stated below. The theorem extends results from [8, 18, 20, 11] from the uniform distribution on $\{\pm 1\}^n$ to the case of non-uniform distributions.

When Y is uniformly distributed, the additive structure of the entries in the direction vector x controls anti-concentration [10]. If x is unstructured, we get stronger anti-concentration bounds for $\langle x, Y\rangle$. This idea is instrumental when analyzing random matrices [13, 23].

We choose the direction x from sets of the following form. We call a set $A \subset \mathbb{Z}^n$ a two-cube if $A = A_1 \times A_2 \times \cdots A_n$, where each $A_j = \{u_j, v_j\}$ consists of two distinct integers. The differences of A are the numbers $d_j = u_j - v_j$ for $j \in [n]$.

The following theorem describes three cases that yield different anti-concentration bounds. It shows that the additive structure of A is deeply related to the bounds we obtain. The less structured A is, the stronger the bounds are.

The first bound in the theorem holds for arbitrary two-cubes. The second bound holds when all the differences $d_1, \ldots, d_n$ are distinct. The third bound applies in more general settings where the set of differences is unstructured. This is captured by the following definition. A set $S \subset \mathbb{N}$ of size n is called a Sidon set, or a Golomb ruler, if the number of solutions to the equation $s_1 + s_2 = s_3 + s_4$ for $s_1, s_2, s_3, s_4 \in S$ is $4 \cdot \binom{n}{2} + n$. In other words, every pair of integers has a distinct sum. Sidon sets were defined by Erdös and Turán [9] and have been studied by many others since. We say that $S \subset \mathbb{Z}$ is a weak Sidon set if the number of solutions to the equation $\epsilon_1 s_1 + \epsilon_2 s_2 = \epsilon_3 s_3 + \epsilon_4 s_4$ for $\epsilon_1, \ldots, \epsilon_4 \in \{\pm 1\}$ and $s_1, \ldots, s_4 \in S$ is at most $100n^2$. The number 100 can be replaced by any other constant, we use it here just to be concrete.

Theorem 2. *For every $\beta > 0$ and $\delta > 0$, there exists $C > 0$ such that the following holds. Let $A \subset \mathbb{Z}^n$ is a two-cube with differences $d_1, \ldots, d_n$. Let $B \subseteq \{\pm 1\}^n$ be of size $2^{\beta n}$. Let Y be uniformly distributed in B.*

(1) For all but $2^{n(1-\beta+\delta)}$ directions $x \in A$,

$$\max_{k \in \mathbb{Z}} \Pr_Y [\langle x, Y\rangle = k] \leq C\sqrt{\ln(n)} n^{-0.5}.$$

(2) If $d_1, \ldots, d_n$ are distinct, then for all but $2^{n(1-\beta+\delta)}$ directions $x \in A$,

$$\max_{k \in \mathbb{Z}} \Pr_Y [\langle x, Y\rangle = k] \leq C\sqrt{\ln(n)} n^{-1.5}.$$

(3) If $\{d_1, \ldots, d_n\}$ is a weak Sidon set of size n, then for all but $2^{n(1-\beta+\delta)}$ directions $x \in A$,

$$\max_{k \in \mathbb{Z}} \Pr_Y [\langle x, Y\rangle = k] \leq C\sqrt{\ln(n)} n^{-2.5}.$$

The first bound in Theorem 2 nearly implies Theorem 1. It is weaker by a factor of $\sqrt{\ln(n)}$. However, it holds for all two-cubes, not just

the hypercube $\{\pm 1\}^n$. The second bound almost matches the sharp $O(n^{-1.5})$ bound that holds when $(u_j, v_j) = (j, -j)$ for each j and Y is uniform in the hypercube [18, 20]. We believe that the $\sqrt{\ln(n)}$ factor is not needed, but were not able to eliminate it. The theorem is, in fact, part of a more general phenomenon. We postpone the full technical description to Section 4.

The proof. Chakrabarti and Regev's proof uses the deep connection between the discrete cube and Gaussian space. They proved a geometric correlation inequality in Gaussian space, and then translated it to the cube. Vidick [26] later simplified part of their argument, but stayed in the geometric setting. Sherstov [19] found a third proof that uses Talagrand's inequality from convex geometry [21] and ideas of Babai, Frankl and Simon from communication complexity [2].

There are several differences between our argument and the ones in [4, 26, 19]. The main difference is that the arguments from [4, 26, 19] are based, in one way or another, on the geometry of Euclidean space. The arguments in [4, 26] prove a correlation inequality in Gaussian space and translate it to the discrete world. It seems that such an argument can not yield point-wise bounds on the concentration probability. A common ingredient in [4, 19] is a step showing that every set of large enough measure contains many almost orthogonal vectors (this is called 'identifying the hard core' in [19]). In [26] this part of the argument is replaced by a statement about a relevant matrix. Our argument does not contain such steps.

Let us briefly discuss our proof at a high level. The proof is based on harmonic analysis (Section 2). The argument consists of two parts. In the first part, we analyze the Fourier behavior of $\langle x, Y \rangle$ for x fixed and Y random. We are able to identify a collection of good x's for which the Fourier spectrum of the distribution of $\langle x, Y \rangle$ decays rapidly. In the second part, we show that the number of bad x's is small by giving an explicit encoding of all of them.

Although the proofs of the two theorems follow a similar strategy, we were not able to completely merge them. The proof for the hypercube (Theorem 1) is carried in Section 3. The proof for general two-cubes (Theorem 2) is given in Section 4.

Remark. *Theorem 2 can be used as a black box to prove the same bounds when $A_1, \ldots, A_n$ are pairs of real numbers. To see this, think of the relevant real numbers as a finite dimensional vector space over the rationals.*

Applications.

Communication Complexity. Chakrabarti and Regev's main motivation was understanding the randomized communication complexity of the gap-hamming problem. The gap-hamming problem was introduced by Indyk and Woodruff in the context of streaming algorithms [12]. Proving lower bounds on its communication complexity was a central open problem for almost ten years, until Chakrabarti and Regev solved it [4]. Vidick [26] and Sherstov [19] later simplified the proof.

Our results also imply the lower bound for the randomized communication complexity of the gap-hamming problem (see e.g. [19]). As opposed to [4, 26, 19], the proof presented here lies entirely in the discrete domain. The underlying ideas may therefore be of independent interest.

Pseudorandomness. Randomness is a computational resource [25]. There are many sources of randomness, and some of them are *weak* or imperfect. Randomness extractors allow to use weak sources of randomness as if they were perfect.

The study of randomness extractors is about constructing explicit maps that transform weak sources of randomness to almost uniform outputs. The main goal is generating a uniform output in the most general scenario possible. This often requires ingenious constructions.

The scenario described above fits nicely in the context of *two-source extractors*. A two-source extractor maps two independent random variables X and Y with significant min-entropy to a single almost uniform output.

Chor and Goldreich [6] used Lindsey's lemma to show that inner product modulo two is a two-source extractor. Namely, the distribution of the bit $\langle X, Y \rangle \bmod 2$ is close to uniform as long as $|A| \cdot |B| \gg 2^n$. Bourgain [3], Raz [17] and Chattopadhyay and Zuckerman [5] constructed two-source extractors with much better parameters. In our work, we study a related but somewhat different question.

The high-level suggestion is to investigate other pseudorandom properties satisfied by known extractors. We already know that inner product is an excellent two-source extractor. Our work shows that the inner product is anti-concentrated over the integers. Theorems 1 and 2, in fact, imply a stronger statement. It is the analog of "strong" extraction in extractor lingo. Not only is $\langle X, Y \rangle$ anti-concentrated, but an overwhelming majority of fixings $X = x$ lead to $\langle x, Y \rangle$ being anti-concentrated.

Additive Combinatorics. Additive combinatorics studies the behavior of sets under algebraic operations [24]. It has many deep results, and connections to other areas of mathematics, as well as many applications in computer science. Our main result can be interpreted as showing that Hamming spheres are far from being sum-sets. Our results give quantitative bounds on the size of the intersection of any Hamming sphere with a sum-set.

Replace $\{\pm 1\}$ by the field $\mathbb{F}_2$ with two elements. The sum-set of $A \subseteq \mathbb{F}_2^n$ and $B \subseteq \mathbb{F}_2^n$ is

$$A + B = \{x + y : x \in A, y \in B\}.$$

If X and Y are sampled uniformly at random from A and B, then $X + Y$ is supported on $A + B$.

The cube $\mathbb{F}_2^n$ is endowed with a natural metric—the Hamming distance $\Delta(x, y)$. The sphere around 0 is the collection of all vectors with a fixed number of ones in them (a.k.a. a slice). The inner product $I = \sum_j (-1)^{X_j}(-1)^{Y_j}$ is similar to the inner product studied above (here $X_j, Y_j \in \{0, 1\}$). The inner product is related to the Hamming distance by $I(X, Y) = n - 2\Delta(X, Y)$. We saw that if $|A| \cdot |B| > 2^{1.01n}$, then I is anti-concentrated. We can conclude that the distribution of the Hamming distance of $X + Y$ is anti-concentrated. The set $A + B$ is far from any sphere.

2. Harmonic Analysis

We are interested in proving anti-concentration for integer-valued random variables. Harmonic analysis is a natural framework for studying such random variables [11].

Let Y be distributed in $\{\pm 1\}^n$. Let $x \in \mathbb{Z}^n$ be a direction. Let θ be uniformly distributed in $[0, 1]$, independently of Y. The idea is to use

$$\Pr_Y[\langle x, Y\rangle = k] = \mathop{\mathbb{E}}_Y \left[\mathop{\mathbb{E}}_\theta \left[\exp(2\pi i \theta \cdot (\langle x, Y\rangle - k)) \right] \right]$$

to bound

$$(\star) \qquad \max_{k \in \mathbb{Z}} \Pr_Y[\langle x, Y\rangle = k] \leq \mathop{\mathbb{E}}_\theta \left[\left| \mathop{\mathbb{E}}_Y \left[\exp(2\pi i \theta \cdot \langle x, Y\rangle) \right] \right| \right].$$

This inequality is useful for two reasons. First, the left hand side is a maximum over k, while the right hand side is not. So, there is one less quantifier to worry about. Secondly, the right hand side lives in the Fourier world, where it is easier to argue about the underlying operators. For example, when the coordinates of Y are independent, the expectation over Y breaks into a product of n simple terms.

3. The Hypercube

We start by considering directions in the hypercube $\{\pm 1\}^n$. The following theorem controls the Fourier coefficients in most directions.

Theorem 3. *For every $\beta > 0$ and $\delta > 0$, there is $c > 0$ so that the following holds. Let $B \subseteq \{\pm 1\}^n$ be of size $2^{\beta n}$. For each $\theta \in [0,1]$, for all but $2^{n(1-\beta+\delta)}$ directions $x \in \{\pm 1\}^n$,*

$$\left| \mathop{\mathbb{E}}_{Y} \left[\exp(2\pi i \theta \cdot \langle x, Y \rangle) \right] \right| > 2 \exp(-cn \sin^2(4\pi\theta))$$

The rest of this section is devoted for proving this theorem (the proof appears in Section 3.3).

Proof of Theorem 1. Theorem 3 promises that for each $\theta \in [0,1]$, the size of

$$A_\theta = \left\{ x \in \{\pm 1\}^n : \left| \mathop{\mathbb{E}}_{Y} \left[\exp(2\pi i \theta \cdot \langle x, Y \rangle) \right] \right| > 2 \exp(-cn \sin^2(4\pi\theta)) \right\}$$

is at most $2^{n(1-\beta+\delta)}$. For each x, define

$$S_x = \{\theta \in [0,1] : x \in A_\theta\}.$$

Since

$$\mathop{\mathbb{E}}_{x} |S_x| = \mathop{\mathbb{E}}_{\theta} \tfrac{|A_\theta|}{2^n} \leq 2^{n(-\beta+\delta)},$$

by Markov's inequality, the number of $x \in A$ for which $|S_x| > 2^{-\delta n}$ is at most $2^{(1-\beta+2\delta)n}$. Fix x such that $|S_x| \leq 2^{-\delta n}$. Bound

$$\begin{aligned} \Pr_Y[\langle x, Y \rangle = k] &\leq \mathop{\mathbb{E}}_{\theta} \left[\left| \mathop{\mathbb{E}}_{Y} \left[\exp(2\pi i \theta \cdot \langle x, Y \rangle) \right] \right| \right] \\ &\leq 2^{-\delta n} + 2 \int_0^1 \exp(-cn \sin^2(4\pi\theta)) \, \mathrm{d}\theta. \end{aligned}$$

Each term in the integral occurs eight times; it circles the origin twice. Using the inequality $\sin(\zeta) > \frac{\zeta}{\pi}$ for $0 < \zeta < \frac{\pi}{2}$, we can bound it by

$$\begin{aligned} &\leq 2^{-\delta n} + 16 \int_0^{1/8} \exp(-16cn\theta^2) \, \mathrm{d}\theta \\ &\leq 2^{-\delta n} + \frac{8}{\sqrt{n}} \cdot \int_{-\infty}^{\infty} \exp(-16c\zeta^2) \, \mathrm{d}\zeta. \end{aligned}$$

The integral converges, so this quantity is at most $\frac{C}{\sqrt{n}}$.

□

3.1. **A Single Direction.** In this section we analyze the behavior of $\langle x, Y\rangle$ for a single direction $x \in \mathbb{Z}^n$. We also focus on a single Fourier coefficient $\mathbb{E}_Y\left[\exp(i\eta \langle x, Y\rangle)\right]$ for a fixed angle $\eta \in [0, 2\pi]$.

We reveal the entropy of Y coordinate by coordinate. To keep track of this entropy, define the following functions $\gamma_1, \ldots, \gamma_n$ from $B = \mathrm{supp}(Y)$ to $\mathbb{R}$. For each $j \in [n]$, let

$$\gamma_j(y) = \gamma_j(y_{<j}) = \min_{\epsilon \in \{\pm 1\}} \Pr[Y_j = \epsilon | Y_{<j} = y_{<j}].$$

To understand the interaction between x and y, we use the following n measurements. For $j \in [n-1]$, define $\phi_j(x, y)$ to be half of the phase of the complex number

$$\mathop{\mathbb{E}}_{Y_{>j}|Y_j=1, Y_{<j}=y_{<j}}\left[\exp(i\eta \langle x_{>j}, Y_{>j}\rangle)\right] \cdot \overline{\mathop{\mathbb{E}}_{Y_{>j}|Y_j=-1, Y_{<j}=y_{<j}}\left[\exp(i\eta \langle x_{>j}, Y_{>j}\rangle)\right]}.$$

This quantity is not defined when $\gamma_j(y) = 0$. In this case, set $\phi_j(x, y)$ to be zero. Define $\phi_n(x, y)$ to be zero. The number $\phi_j(x, y)$ is determined by $y_{<j}$ and $x_{>j}$.

In the following we think of x as fixed, and of γ_j and ϕ_j as random variables that are determined by the random variable Y.

Lemma 4. *For each $x \in \mathbb{R}^n$, every random variable Y over $\{\pm 1\}^n$, and every angle $\eta \in \mathbb{R}$,*

$$\left|\mathop{\mathbb{E}}_{Y}\left[\exp(i\eta \langle x, Y\rangle)\right]\right|^2 \leq \mathop{\mathbb{E}}_{Y}\left[\prod_{i \in [n]} (1 - \gamma_j \sin^2(\phi_j + x_j \eta))\right].$$

Proof. The proof is by induction on n. We prove the base case of the induction and the inductive step simultaneously. Express

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{Y}\left[\exp(i\eta \langle x, Y\rangle)\right]\right|^2 &= \left|\mathop{\mathbb{E}}_{Y_1}\left[\exp(i\eta x_1 Y_1) \cdot \mathop{\mathbb{E}}_{Y_{>1}|Y_1}\left[\exp(i\eta \langle x_{>1}, Y_{>1}\rangle)\right]\right]\right|^2 \\ &= \left|p_1 \exp(i\eta x_1) Z_1 + p_{-1} \exp(-i\eta x_1) Z_{-1}\right|^2,\end{aligned}$$

where for $\epsilon \in \{\pm 1\}$,

$$p_\epsilon = \Pr[Y_1 = \epsilon] \qquad \& \qquad Z_\epsilon = \mathop{\mathbb{E}}_{Y|Y_1=\epsilon}\left[\exp(i\eta \langle x_{>1}, Y_{>1}\rangle)\right].$$

When $n = 1$, we have $Z_1 = Z_{-1} = 1$. Rearranging,

$$\begin{aligned}&\left|p_1 \exp(i\eta x_1) Z_1 + p_{-1} \exp(-i\eta x_1) Z_{-1}\right|^2 \\ &= p_1^2 |Z_1|^2 + p_{-1}^2 |Z_{-1}|^2 + p_1 p_{-1}(Z_1 \overline{Z_{-1}} \exp(i2\eta x_1) + \overline{Z_1} Z_{-1} \exp(-i2\eta x_1)) \\ &= p_1^2 |Z_1|^2 + p_{-1}^2 |Z_{-1}|^2 + 2p_1 p_{-1} |Z_1||Z_{-1}| \cos(2\phi_1 + 2x_1\eta).\end{aligned}$$

The last equality holds by the definition of ϕ_1.

There are two cases to consider. When $\cos(2\phi_1 + 2x_1\eta) < 0$, we continue to bound

$$\begin{aligned}
&< p_1^2|Z_1|^2 + p_{-1}^2|Z_{-1}|^2 \\
&\leq (p_1|Z_1|^2 + p_{-1}|Z_{-1}|^2)(1-\gamma_1) \\
&\leq (p_1|Z_1|^2 + p_{-1}|Z_{-1}|^2)(1-\gamma_1\sin^2(\phi_1 + x_1\eta)).
\end{aligned}$$

Recall that γ_1 and ϕ_1 do not depend on Y. When $\cos(2\phi_1 + 2x_1\eta) \geq 0$, using the inequality $a^2 + b^2 \geq 2ab$, we bound

$$\begin{aligned}
&\leq p_1^2|Z_1|^2 + p_{-1}^2|Z_{-1}|^2 + p_1p_{-1}(|Z_1|^2 + |Z_{-1}|^2)\cos(2\phi_1 + 2x_1\eta) \\
&= p_1|Z_1|^2(p_1 + p_{-1}\cos(2\phi + 2x_1\eta)) + p_{-1}|Z_{-1}|^2(p_{-1} + p_1\cos(2\phi_1 + 2x_1\eta)) \\
&\leq (p_1|Z_1|^2 + p_{-1}|Z_{-1}|^2)(1 - \gamma_1 + \gamma_1\cos(2\phi_1 + 2x_1\eta)) \\
&= \mathop{\mathbb{E}}_{Y_1}\left[|Z_{Y_1}|^2\right](1 - 2\gamma_1\sin^2(\phi_1 + x_1\eta)).
\end{aligned}$$

When $n = 1$, we have proved the base case of the induction. When $n > 1$, apply induction on $|Z_\epsilon|^2$. □

3.2. **A Few Bad Directions.** Lemma 4 suggests proving that the expression $\sum_j \gamma_j \sin^2(\phi_j + x_j\eta)$ is typically large. Namely, we aim to show that there are usually many coordinates j for which both γ_j and $\sin^2(\phi_j + x_j\eta)$ are bounded away from zero. Our approach is to explicitly encode the cases where this fails to hold.

Recall that Y is uniformly distributed in a set B of size $|B| = 2^{\beta n}$. Let $1 \geq \lambda > 1/n$ be a parameter. Set $0 < \kappa < \frac{1}{2}$ and $1 \geq \tau > 0$ to be parameters satisfying the conditions

$$H\left(\tfrac{1}{\log(1/\kappa)}\right) = \tau + H(\tau) = \lambda, \tag{1}$$

where H is the binary entropy function:

$$H(\xi) = \xi\log(1/\xi) + (1-\xi)\log(1/(1-\xi)).$$

The encoding is based on the following two sets:

$$J(y) = J_{B,\kappa}(y) = \{j \in [n] : \gamma_j(y) \geq \kappa\}$$

and

$$G(x,y) = G_{B,\kappa,\theta}(x,y) = \left\{j \in J(y) : \sin^2(\phi_j(x,y) + x_j\eta) \geq \tfrac{\sin^2(2\eta)}{4}\right\}.$$

We start by showing that there are few y's for which $|J(y)|$ is small.

Lemma 5. *The number of $y \in B$ with $|J(y)| \leq n(\beta - 3\lambda)$ is at most $2^{n(\beta-2\lambda)}$.*

Proof. If $3\lambda > \beta$, the statement is trivially true. So, in the rest of the proof, assume that $3\lambda \leq \beta$. Each $y \in B$ with $|J(y)| \leq n(\beta - 3\lambda)$ can be uniquely encoded by the following data:

- An vector $q \in \{\pm 1\}^t$ with $t = \lfloor n(\beta - 3\lambda) \rfloor$.
- A subset $S \subseteq [n]$ of size $|S| \leq \frac{n}{\log(1/\kappa)}$.

Let us describe the encoding. The vector q encodes the values taken by y in the coordinates $J(y)$. We do not encode $J(y)$ itself, only the values of y in the coordinates corresponding to $J(y)$. The set S includes $j \in [n]$ if and only if $\gamma_j(y) < \kappa$. Each string $y \in B$ has probability at least 2^{-n}. This implies that $\kappa^{|S|} \geq 2^{-n}$.

We can reconstruct y from q and S by iteratively computing y_1, then y_2, and so on, until we get to y_n. Whether or not $1 \in J(y)$ is determined even before we know y. If $1 \in J(y)$ then q tells us what y_1 is. If $1 \notin J(y)$ and $1 \in S$ then y_1 is the least likely value between ± 1. If $1 \notin J(y)$ and $1 \notin S$ then y_1 is the more likely value. Given the value of y_1, we can continue in the same way to compute the rest of y.

The number of choices for q is at most $2^{n(\beta-3\lambda)}$. The number of choices for S is at most $2^{nH(1/\log(1/\kappa))} = 2^{\lambda n}$. □

Next, we argue that there are few x's for which there are many y's with small $G(x, y)$.

Lemma 6. *The number of $x \in A$ for which*

$$\Pr_Y[|G(x, Y)| \leq \tau n] \geq 2^{-\lambda n}$$

is at most $2^{n(1-\beta+6\lambda)}$.

Proof. The lemma is proved by double-counting the edges in a bipartite graph. Let $\mathcal{X}$ be the set we are interested in:

$$\mathcal{X} = \left\{x : \Pr_Y[|G(x, Y)| \leq \tau n] \geq 2^{-\lambda n}\right\}.$$

The left side of the bipartite graph is $\mathcal{X}$ and the right side is B. Connect $x \in \mathcal{X}$ to $y \in B$ by an edge if and only if $G(x, y) \leq \tau n$. Let E denote the set of edges in this graph.

First, we bound the number of edges from below. The number of edges that touch each $x \in \mathcal{X}$ is at least $2^{-\lambda n}|B|$. It follows that

$$|E| \geq 2^{-\lambda n} \cdot |\mathcal{X}| \cdot |B|.$$

Next, we bound the number of edges from above. By Lemma 5, the number of $y \in B$ so that $|J(y)| \leq n(\beta - 3\lambda)$ is at most $2^{-2\lambda n}|B|$. We shall prove that the number of edges that touch each y with $|J(y)| > n(\beta - 3\lambda)$ is at most $2^{n(1-\beta+4\lambda)}$. It follows that

$$|E| \leq 2^{-2\lambda n} \cdot |\mathcal{X}| \cdot |B| + |B| \cdot 2^{n(1-\beta+4\lambda)}.$$

We can conclude that

$$\begin{aligned} 2^{-\lambda n} \cdot |\mathcal{X}| \cdot |B| &\leq 2^{-2\lambda n} \cdot |\mathcal{X}| \cdot |B| + |B| \cdot 2^{n(1-\beta+4\lambda)} \\ \Rightarrow |\mathcal{X}| &\leq 2^{n(1-\beta+6\lambda)}, \end{aligned}$$

since $\lambda n > 1$.

It remains to fix y so that $|J(y)| > n(\beta - 3\lambda)$ and bound its degree from above. This too is achieved by an encoding argument. Encode each x that is connected to y by an edge using the following data:

- A vector $q \in \{\pm 1\}^t$ with $t = \lfloor n(1 - \beta + 3\lambda) \rfloor$.
- The set $G(x, y)$.
- A vector $r \in \{\pm 1\}^s$ with $s = \lfloor \tau n \rfloor$.

Let us describe the encoding. The vector q specifies the values of x on coordinates not in $J(y)$. There are at most $n - n(\beta - 3\lambda) = n(1 - \beta + 3\lambda)$ such coordinates. The size of $G(x, y)$ is at most τn. The vector r specifies the values of x in the coordinates of $G(x, y)$, written in descending order.

The decoding of x from q, S and r is done as follows. Decode the coordinates of x in descending order from n to 1. If $n \notin J(y)$ then we read the value of x_n from q. If $n \in J(y)$ and $n \in G(x, y)$, we decode x_n by reading its value from r. If $n \in J(y)$ and $n \notin G(x, y)$, then

$$\sin^2(\phi_n(x, y) + x_n \eta) \leq \tfrac{\sin^2(2\eta)}{4}.$$

The number $\phi_n(x, y)$ does not depend on x. The following claim implies that there is at most one value of x_n that satisfies this property.

Claim 7. *For all $\varphi \in \mathbb{R}$ and $u, v \in \mathbb{Z}$,*

$$\max\{|\sin(\varphi + \eta u)|, |\sin(\varphi + \eta v)|\} \geq \tfrac{|\sin(\eta(u-v))|}{2}.$$

Proof. Consider the map

$$\varphi \mapsto g(\varphi) = \max\{|\sin(\varphi + \eta u)|, |\sin(\varphi + \eta v)|\}.$$

The minimum of this map is attained when

$$|\sin(\varphi + \eta u)| = |\sin(\varphi + \eta v)|.$$

This happens only when $\varphi = -\frac{\eta(u+v)}{2} \mod \frac{\pi}{2}\mathbb{Z}$. By symmetry,

$$\begin{aligned} g(\varphi) &\geq \min\{g(-\eta(u+v)/2), g(-\eta(u+v)/2 + \pi/2)\} \\ &\geq |\sin(\eta(u-v)/2) \cdot \cos(\eta(u-v)/2)| \\ &= \frac{|\sin(\eta(u-v))|}{2}. \end{aligned}$$

□

The claim implies that we can indeed reconstruct x_n. Given x_n, we can similarly reconstruct x_{n-1}, since ϕ_{n-1} depends only on y and x_n. Continuing in this way, we can reconstruct $x_{n-2}, \ldots, x_1$. The total number of choices for q, S, r is at most $2^{n(1-\beta+3\lambda)+nH(\tau)+\tau n} = 2^{n(1-\beta+4\lambda)}$. □

3.3. **Putting It Together.**

Proof of Theorem 3. Set $\lambda = \frac{\delta}{6}$. By Lemma 4,

$$\left|\mathop{\mathbb{E}}_{Y}\left[\exp(2\pi i\theta \left\langle x, Y\right\rangle)\right]\right| \leq \sqrt{\mathop{\mathbb{E}}_{Y}\left[\exp\left(-\sum_{j=1}^{n}\gamma_j \sin^2(\phi_j + 2\pi\theta x_j)\right)\right]}.$$

Whenever x is such that

$$\Pr_{Y}[G(x,Y) \leq \tau n] < 2^{-\lambda n}, \tag{2}$$

we can bound

$$\mathop{\mathbb{E}}_{Y}\left[\exp\left(-\sum_{j=1}^{n}\gamma_j \sin^2(\phi_j + 2\pi\theta x_j)\right)\right] \leq \exp(-\tfrac{\kappa}{4}n\tau \sin^2(4\pi\theta)) + 2^{-\lambda n}.$$

Since $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for $a, b \geq 0$, for such an x we can bound

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{Y}\left[\exp(2\pi i\theta \left\langle x, Y\right\rangle)\right]\right| &\leq \exp(-\tfrac{\kappa}{8}n\tau \sin^2(4\pi\theta)) + 2^{-\lambda n/2} \\ &\leq 2\exp(-cn\sin^2(4\pi\theta)).\end{aligned}$$

Lemma 6 promises that there are at most $2^{n(1-\beta+\delta)}$ choices for x that does not satisfy (2).

□

4. General Two-Cubes

Now we move to the setting where the direction x is chosen from an arbitrary two-cube $A \subset \mathbb{Z}^n$ with differences $d_1, \ldots, d_n$. The way we measure the structure of A follows ideas of Halász [11]. For an integer $\ell > 0$, define $r_\ell(A)$ to be the number of elements $(\epsilon, j) \in \{\pm 1\}^{2\ell} \times [n]^{2\ell}$ that satisfy

$$\epsilon_1 \cdot d_{j_1} + \cdots + \epsilon_{2\ell} \cdot d_{j_{2\ell}} = 0.$$

The smaller $r_\ell(A)$ is, the less structured A is.

The theorem below shows that $r_\ell(A)$ allows us to control the concentration probability. More concretely, for $C > 0$ and $\ell > 0$, define

$$R_{C,\ell}(A) = \frac{C^\ell r_\ell(A)}{n^{2\ell+1/2}} + \exp(-\tfrac{n}{C}).$$

Define

$$R_C(A) = \inf\{R_{C,\ell}(A) : \ell \in \mathbb{N}\}.$$

This is essentially the bound on the concentration probability that Halász obtained in [11] when Y is uniform in $\{\pm 1\}^n$. Our upper bounds are slightly weaker. Let

$$\mu_C(A) = \inf\left\{\mu \in [0,1] : \exists \nu \in (0,1] \;\; \mu^{(1+\nu)^2} \geq 3\exp(-\tfrac{\nu n}{C}) + \tfrac{R_C(A)}{50\sqrt{\nu}}\right\},$$

where we adopt the convention that the infimum of the empty set is 1. Before stating the theorem, let us go over the three examples from Theorem 2:

(1) For arbitrary A, since $r_1(A) \leq O(n^2)$, we get[1] $\mu_C(A) \leq O(\frac{\sqrt{\ln n}}{\sqrt{n}})$ with $\nu = \frac{1}{\ln(1/R_{C,1}(A))}$.
(2) When all the differences are distinct, since $r_1(A) \leq O(n)$, we get $\mu_C(A) \leq O(n^{-1.5}\sqrt{\ln n})$ with $\nu = \frac{1}{\ln(1/R_{C,1}(A))}$.
(3) When $\{\pm d_1, \ldots, \pm d_n\}$ is a Sidon set, since $r_2(A) \leq O(n^2)$, we get $\mu_C(A) \leq O(n^{-2.5}\sqrt{\ln n})$ with $\nu = \frac{1}{\ln(1/R_{C,2}(A))}$.

More generally, when $R_C(A)$ is bound from below by some polynomial in $\frac{1}{n}$ then $\mu_C(A)$ is at most $O(R_C(A)\sqrt{\log(4/R_C(A))})$.

Theorem 8. *For every $\beta > 0$ and $\delta > 0$, there is $C > 0$ so that the following holds. Let $B \subseteq \{\pm 1\}^n$ be of size $2^{\beta n}$. Let Y be uniformly distributed in B. Let $A \subset \mathbb{Z}^n$ be a two-cube. Then, for all but $2^{n(1-\beta+\delta)}$ directions $x \in A$,*

$$\mathop{\mathbb{E}}_{\theta}\left[\left|\mathop{\mathbb{E}}_{Y}\left[\exp(2\pi i\theta \cdot \langle x, Y\rangle)\right]\right|\right] \leq \mu_C(A).$$

Before moving on, we discuss a fourth extreme example. When $A_j = \{2^j, -2^j\}$ for each $j \in [n]$, we have $r_\ell(A) \leq (2\ell n)^\ell$. In this case, setting $\ell = \Omega(n)$ gives exponentially small anti-concentration with $\nu = 1$. This result is trivial, but it illustrates that the mechanism underlying the proof yields strong bound in many settings.

By $(\star)$ and the explanation above, we see that Theorem 8 implies Theorem 2. The rest of this section is devoted to the proof of Theorem 8. The high-level structure of the proof is similar to that of Theorem 3. However, there are several new technical challenges that we need to overcome.

The main technical challenge that needs to be overcome has to do with the definition of the set G. The G defined in the previous section depends on the angle θ. This is problematic for the proof in the

[1]Here and below the big O notation hides a constant that may depend on C.

generality we are working with now. So, we need to find a different set of *good* coordinates, one that depends only on x and y. Our solution is based on the following claim, which quantifies the strict convexity of the map $\zeta \mapsto \zeta^{1+\nu}$ for $\nu > 0$. We defer the proof to Appendix A.

Claim 9. *For every $\kappa > 0$, there is a constant $c_1 > 0$ so that the following holds. For every random variable $W \in \{\pm 1\}$ such that*

$$\min\left\{ \Pr[W = 1], \Pr[W = -1] \right\} \geq \kappa,$$

every $\alpha_1 \geq 2\alpha_{-1} \geq 0$ and every $0 < \nu \leq 1$,

$$\mathbb{E}\left[\alpha_W\right]^{1+\nu} \leq (1 - c_1\nu)\,\mathbb{E}\left[\alpha_W^{1+\nu}\right].$$

4.1. **A Single Direction.** The following lemma generalizes Lemma 4. Recall the definition of γ_j, ϕ_j and $J(y)$ from Sections 3.1 and 3.2.

Lemma 10. *For every $\kappa > 0$, there is a constant $c_0 > 0$ so that the following holds. For every $0 < \nu \leq 1$, every angle $\eta \in \mathbb{R}$, every direction $x \in \mathbb{Z}^n$, and every random variable Y over $\{\pm 1\}^n$,*

$$\left| \mathop{\mathbb{E}}_{Y} \left[\exp(i\eta \left\langle x, Y \right\rangle)\right] \right|^{1+\nu} \leq \mathop{\mathbb{E}}_{Y} \left[\prod_{j \in J} (1 - c_0\nu \sin^2(\phi_j + x_j\eta) \right].$$

Proof. The proof is by induction on n. If $1 \notin J$, the proof holds by induction. The base case of $n = 1$ is trivial. So assume that $1 \in J$. Express

$$\mathop{\mathbb{E}}_{Y} \left[\exp(i\eta \left\langle x, Y \right\rangle)\right] = p_1 \exp(i\eta x_1) Z_1 + p_{-1} \exp(-i\eta x_1) Z_{-1},$$

where for $\epsilon \in \{\pm 1\}$,

$$p_\epsilon = \Pr[Y_1 = \epsilon] \qquad \& \qquad Z_\epsilon = \mathop{\mathbb{E}}_{Y | Y_1 = \epsilon} \left[\exp(i\eta \left\langle x_{>1}, Y_{>1} \right\rangle)\right].$$

When $n = 1$, we have $Z_1 = Z_{-1} = 1$. Using the definition of ϕ_1,

$$\begin{aligned}
&|p_1 \exp(i\eta x_1) Z_1 + p_{-1} \exp(-i\eta x_1) Z_{-1}|^2 \\
&= p_1^2 |Z_1|^2 + p_{-1}^2 |Z_{-1}|^2 + p_1 p_{-1} (Z_1 \overline{Z_{-1}} \exp(i2\eta x_1) + \overline{Z_1} Z_{-1} \exp(-i2\eta x_1)) \\
&= p_1^2 |Z_1|^2 + p_{-1}^2 |Z_{-1}|^2 + 2p_1 p_{-1} |Z_1||Z_{-1}| \cos(2\phi_1 + 2x_1\eta) \\
&= p_1^2 |Z_1|^2 + p_{-1}^2 |Z_{-1}|^2 + 2p_1 p_{-1} |Z_1||Z_{-1}| \\
&\qquad - 2p_1 p_{-1} |Z_1||Z_{-1}| (1 - \cos(2\phi_1 + 2x_1\eta)) \\
&= \mathbb{E}\left[|Z_{Y_1}|\right]^2 - 4p_1 p_{-1} |Z_1||Z_{-1}| \sin^2(\phi_1 + x_1\eta),
\end{aligned}$$

Without loss of generality, assume that $|Z_1| \geq |Z_{-1}|$. There are two cases to consider. The first case is that Z_1 and Z_{-1} are comparable in

magnitude: $|Z_1| \le 2|Z_{-1}|$. In this case, we can continue the bound by

$$\begin{aligned}
&\le \mathbb{E}\left[|Z_{Y_1}|\right]^2 - 2p_1p_{-1}|Z_1|^2\sin^2(\phi_1 + x_1\eta) \\
&\le \mathbb{E}\left[|Z_{Y_1}|\right]^2 (1 - 2\kappa(1-\kappa)\sin^2(\phi_1 + x_1\eta)),
\end{aligned}$$

since $1 \in J$. This gives

$$\begin{aligned}
&|p_1\exp(i\eta x_1)Z_1 + p_{-1}\exp(-i\eta x_1)Z_{-1}|^{1+\nu} \\
&\le \mathbb{E}\left[|Z_{Y_1}|\right]^{1+\nu} (1 - 2\kappa(1-\kappa)\sin^2(\phi_1 + x_1\eta))^{(1+\nu)/2} \\
&\le \mathbb{E}\left[|Z_{Y_1}|^{1+\nu}\right] (1 - \kappa(1-\kappa)\sin^2(\phi_1 + x_1\eta)),
\end{aligned}$$

since the map $\zeta \mapsto \zeta^{1+\nu}$ is convex.

The second case is when $|Z_1| > 2|Z_{-1}|$. Recall that we have already shown

$$\begin{aligned}
&|p_1\exp(i\eta x_1)Z_1 + p_{-1}\exp(-i\eta x_1)Z_{-1}|^2 \\
&= \mathbb{E}\left[|Z_{Y_1}|\right]^2 - 4p_1p_{-1}|Z_1||Z_{-1}|\sin^2(\phi_1 + x_1\eta) \\
&\le \mathbb{E}\left[|Z_{Y_1}|\right]^2.
\end{aligned}$$

Claim 9 implies that

$$\begin{aligned}
&|p_1\exp(i\eta x_1)Z_1 + p_{-1}\exp(-i\eta x_1)Z_{-1}|^{1+\nu} \\
&\le \mathbb{E}\left[|Z_{Y_1}|\right]^{1+\nu} \\
&\le (1 - c_1\nu)\cdot\mathbb{E}\left[|Z_{Y_1}|^{1+\nu}\right] \\
&\le (1 - c_1\nu\sin^2(\phi_j + x_j\eta))\cdot\mathbb{E}\left[|Z_{Y_1}|^{1+\nu}\right].
\end{aligned}$$

Finally, setting $c_0 = \min\{c_1, \kappa(1-\kappa)\}$, we get a bound that applies in both cases:

$$\left|\mathop{\mathbb{E}}_{Y}\left[\exp(i\eta\left\langle x, Y\right\rangle)\right]\right|^{1+\nu} \le (1 - c_0\nu\sin^2(\phi_j + x_j\eta))\cdot\mathbb{E}\left[|Z_{Y_1}|^{1+\nu}\right].$$

This proves the base case of the induction and also allows to perform the inductive step.

□

4.2. **An Average Direction.** In this section we analyze the bound from the previous section for an average direction X in a two-cube $A \subset \mathbb{Z}^n$. This step has no analogy in the proof of Theorem 3. To compute the expectation over an average direction, we reveal the entropy of X coordinate by coordinate in reverse order (from the n'th coordinate to the first one).

In analogy with $\gamma_1, \ldots, \gamma_n$, define the following functions $\mu_1, \ldots, \mu_n$. For each $j \in [n]$, let

$$\mu_j(x) = \mu_j(x_{>j}) = \min_{\epsilon \in A_j} \Pr[X_j = \epsilon | X_{>j} = x_{>j}];$$

this is well-defined for x in $A = \mathrm{supp}(X)$. In analogy with the definition of $J(y)$, let

$$J'(x) = \{j \in [n] : \mu_j(x) \geq \kappa\}.$$

In this section, we define the set G differently, but use the same notation. Let

$$G(x, y) = G_{A,B,\kappa}(x, y) = J'(x) \cap J(y).$$

Recall that γ_j, ϕ_j and $J(\cdot)$ depend on the set B, on $y \in B$ and on $x \in \mathbb{Z}^n$. In the following lemma, we fix an arbitrary $y \in B$, and take the expectation over a random $X \in A$. We allow G to be a random set that depends on X, and ϕ_j to be a random variable that depends on $X_{>j}$.

Lemma 11. *For every $\kappa > 0$ and $0 < c_0 \leq 1$, there is a constant $c > 0$ so that the following holds. For every $0 < \nu \leq 1$, every angle $\eta \in \mathbb{R}$, every $B \subseteq \{\pm 1\}^n$, every $y \in B$, every random variable X taking values in a two-cube $A \subseteq \mathbb{Z}^n$ with differences $d_j = u_j - v_j$,*

$$\mathop{\mathbb{E}}_{X} \left[\prod_{j \in J} (1 - c_0 \nu \sin^2(\phi_j + X_j \eta)) \right]^{1+\nu} \leq \mathop{\mathbb{E}}_{X} \left[\exp\Big(- c\nu \sum_{j \in G} \sin^2(d_j \eta) \Big) \right].$$

Proof. The proof is by induction on n. Recall that ϕ_j and μ_j is determined by $x_{>j}$. In particular, whether or not $n \in G(x, y)$ does not depend on x. If $n \notin G(x, y)$, the proof holds by induction, or is trivially true for $n = 1$. So assume that $n \in G(x)$. Start with

$$\begin{aligned} &\mathop{\mathbb{E}}_{X} \left[\prod_{j \in J} (1 - c_0 \zeta \sin^2(\phi_j + X_j \eta)) \right] \\ &= \mathop{\mathbb{E}}_{X_n} \left[(1 - c_0 \zeta \sin^2(\phi_n + X_n \eta)) Z_{X_n} \right], \end{aligned}$$

where for $a \in A_n := \{u, v\}$,

$$Z_a = \mathop{\mathbb{E}}_{X | X_n = a} \left[\prod_{j \in J : j < n} (1 - c_0 \sin^2(\phi_j + X_j \eta)) \right].$$

If $n = 1$, then $Z_u = Z_v = 1$. Assume without loss of generality that $Z_u \geq Z_v$. There are two cases to consider. The first case is that

$Z_u > 2Z_v$. In this case, Claim 9 implies

$$\begin{aligned}\mathop{\mathbb{E}}_{X}\left[\prod_{j\in J}(1-c_0\nu\sin^2(\phi_j+X_j\eta))\right]^{1+\nu} &\le \mathbb{E}\left[Z_{X_n}\right]^{1+\nu}\\ &\le (1-c_1\nu)\,\mathbb{E}\left[Z_{X_n}^{1+\nu}\right]\\ &\le \exp(-c_1\nu)\,\mathbb{E}\left[Z_{X_n}^{1+\nu}\right].\end{aligned}$$

The second case is when $Z_u \le 2Z_v$. By Claim 7,

$$\max\left\{|\sin(\phi_n+u\eta)|,|\sin(\phi_n+v\eta)|\right\} \ge \tfrac{\sin(d_n\eta)}{2}.$$

Since $\mu_n(x) \ge \kappa$,

$$\begin{aligned}&\mathop{\mathbb{E}}_{X_n}\left[(1-c_0\nu\sin^2(\phi_n+X_n\eta))Z_{X_n}\right]^{1+\nu}\\ &\le (\mathop{\mathbb{E}}_{X_n}\left[Z_{X_n}\right]-\kappa c_0\nu\tfrac{\sin^2(d_n\eta)}{4}\tfrac{Z_u}{2})^{1+\nu}\\ &\le (\mathop{\mathbb{E}}_{X_n}\left[Z_{X_n}\right](1-\tfrac{\kappa c_0\nu}{8}\sin^2(d_n\eta)))^{1+\nu}\\ &\le \mathop{\mathbb{E}}_{X_n}\left[Z_{X_n}^{1+\nu}\right]\exp(-\tfrac{c_0\kappa\nu}{8}\sin^2(d_n\eta)).\end{aligned}$$

In both cases,

$$\mathop{\mathbb{E}}_{X}\left[\prod_{j\in J}(1-c_0\sin^2(\phi_j+X_j\eta))\right]^{1+\nu} \le \exp(-c\nu\sin^2(d_n\eta))\mathop{\mathbb{E}}_{X_n}\left[Z_{X_n}^{1+\nu}\right],$$

for some constant $c(\kappa,c_0)>0$. This proves the base case of the induction and also allows to perform the inductive step.

□

4.3. **Putting It Together.**

Proof of Theorem 8. Let $\mu > 0$ and $0 < \nu \le 1$ be so that

$$\mu^{(1+\nu)^2} \ge 3\exp(-\tfrac{\nu n}{C}) + \tfrac{R_C(A)}{50\sqrt{\nu}};$$

if no such μ,ν exist then the theorem is trivially true. Let

$$A_0 = \Big\{x\in A : \mathop{\mathbb{E}}_{\theta}\left[\Big|\mathop{\mathbb{E}}_{Y}\left[\exp(2\pi i\theta\cdot\langle x,Y\rangle)\right]\Big|\right] \ge \mu\Big\}.$$

Denote the size of A_0 by $2^{\alpha n}$. Assume towards a contradiction that $\alpha+\beta \ge 1+\delta$. Let X be uniformly distributed in A_0, independently of

Y and θ. Let $\lambda = \frac{\delta}{7}$, and let κ be as in (1). By Lemma 10,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{X,\theta}\left[\left|\mathop{\mathbb{E}}_{Y}\left[\exp(i2\pi\theta\left\langle x,Y\right\rangle)\right]\right|\right]^{(1+\nu)^2}\\
&\leq \mathop{\mathbb{E}}_{X,\theta}\left[\left|\mathop{\mathbb{E}}_{Y}\left[\exp(i2\pi\theta\left\langle x,Y\right\rangle)\right]\right|^{1+\nu}\right]^{1+\nu}\\
&\leq \mathop{\mathbb{E}}_{X,\theta}\left[\mathop{\mathbb{E}}_{Y}\left[\prod_{j\in J}(1-c_0\nu\sin^2(\phi_j+x_j2\pi\theta)\right]\right]^{1+\nu}.
\end{aligned}$$

By Lemma 11, we can continue

$$\begin{aligned}
&= \mathop{\mathbb{E}}_{Y,\theta}\left[\mathop{\mathbb{E}}_{X}\left[\prod_{j\in J}(1-c_0\nu\sin^2(\phi_j+x_j2\pi\theta)\right]\right]^{1+\nu}\\
&\leq \mathop{\mathbb{E}}_{Y,\theta}\left[\mathop{\mathbb{E}}_{X}\left[\prod_{j\in J}(1-c_0\nu\sin^2(\phi_j+x_j2\pi\theta)\right]^{1+\nu}\right]\\
&\leq \mathop{\mathbb{E}}_{X,Y,\theta}\left[\exp(-c\nu D(\theta))\right],
\end{aligned}$$

where

$$D(\theta) = D_{x,y}(\theta) = \sum_{j\in G(x,y)}\sin^2(2\pi\theta d_j).$$

By Lemma 5, $|J(y)| > n(\beta-3\lambda)$ for all but $2^{n(\beta-2\lambda)}$ choices for y. Similarly, $|J'(x)| > n(\alpha-3\lambda)$ for all but $2^{n(\alpha-2\lambda)}$ choices of x. By assumption, $\beta-3\lambda+\alpha-3\lambda \geq \lambda$. Since $|G(x,y)| \geq |J(y)|+|J'(x)|-n$,

$$\begin{aligned}
&\Pr[|G(X,Y)| \leq \lambda n]\\
&\leq \Pr[|J(Y)| \leq n(\beta-3\lambda)] + \Pr[|J'(X)| \leq n(\alpha-3\lambda)]\\
&\leq 2^{-2\lambda n} + 2^{-2\lambda n}.
\end{aligned}$$

Claim. *Let x, y be so that $G(x,y) \geq \lambda n$. For every $0 \leq \rho \leq \frac{\lambda n}{4}$ and integer $\ell > 0$,*

$$\Pr_{\theta}[D(\theta) < \rho] \leq \frac{4r_\ell(A)}{(\lambda n)^{2\ell+1/2}}\sqrt{\rho}.$$

Given the claim, for every x, y so that $G(x,y) \geq \lambda n$ and $\ell > 0$,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\theta}[\exp(-c\nu D(\theta))] \\
&= \int_0^1 \Pr_{\theta}[\exp(-c\nu D(\theta)) > t]\,\mathrm{d}t \\
&\leq \exp(-\tfrac{c\nu\lambda n}{4}) + \int_{\exp(-c\nu\lambda n/4)}^1 \Pr_{\theta}[D(\theta) < -\tfrac{\ln t}{c\nu}]\,\mathrm{d}t \\
&\leq \exp(-\tfrac{c\nu\lambda n}{4}) + \frac{4r_\ell(A)}{(\lambda n)^{2\ell+1/2}} \int_0^1 \sqrt{-\tfrac{\ln t}{c\nu}}\,\mathrm{d}t.
\end{aligned}$$

The integral $\int_0^1 \sqrt{-\ln t}\,\mathrm{d}t \leq 1$ converges to a constant. For an appropriate $C = C(\beta,\delta) > 0$ and $\ell > 0$, we get the desired contradiction.

$$\begin{aligned}
\mu^{(1+\nu)^2} &\leq 2 \cdot 2^{-2\lambda n} + \exp(-\tfrac{c\nu\lambda n}{4}) + \frac{4r_\ell(A)}{\sqrt{c\nu}(\lambda n)^{2\ell+1/2}} \\
&< 3\exp(-\tfrac{\nu n}{C}) + \tfrac{R_C(A)}{50\sqrt{\nu}}.
\end{aligned}$$

Proof of Claim. Let $G = G(x,y)$. Observe that

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\theta}\left[(|G| - 2D(\theta))^{2\ell}\right] \\
&= \mathop{\mathbb{E}}_{\theta}\left[\Big(\sum_{j\in G}\cos(4\pi d_j\theta)\Big)^{2\ell}\right] \\
&= 2^{-2\ell}\mathop{\mathbb{E}}_{\theta}\left[\Big(\sum_{j\in G}\exp(4\pi i d_j\theta) + \exp(-4\pi i d_j\theta)\Big)^{2\ell}\right] \\
&\leq 2^{-2\ell} r_\ell(A);
\end{aligned}$$

the last equality follows from the fact that of the $\leq (2|G|)^{2\ell}$ terms in the expansion, the only ones that survive are the ones with phase 0. There are at most $r_\ell(A)$ such terms, and each contributes 1.

By Markov's inequality, since $|G| \geq \lambda n$,

$$\Pr_{\theta}[D(\theta) \leq \tfrac{\lambda n}{4}] \leq \Pr_{\theta}\left[(|G| - 2D(\theta))^{2\ell} \geq (\tfrac{\lambda n}{2})^{2\ell}\right] \leq \frac{2^{-2\ell}r_\ell(A)}{(\lambda n/2)^{2\ell}} = \frac{r_\ell(A)}{(\lambda n)^{2\ell}}.$$

This proves the claim for $\rho = \frac{\lambda n}{4}$.

It remains to prove the claim for $\rho < \frac{\lambda n}{4}$. This part uses Kemperman's theorem [14] from group theory (in fact Kneser's theorem [15] for abelian groups suffices). Think of $[0,1)$ as the group $\mathbb{R}/\mathbb{Z}$. Let

$$S_\rho = \{\theta \in \mathbb{R}/\mathbb{Z} : D(\theta) \leq \rho\}.$$

We claim that the m-fold sum $S_\rho + S_\rho + \cdots + S_\rho \subseteq \mathbb{R}/\mathbb{Z}$ is contained in $S_{\rho m^2}$. Indeed,

$$\begin{aligned} |\sin(\eta_1 + \eta_2)| &= |\sin(\eta_1)\cos(\eta_2) + \sin(\eta_2)\cos(\eta_1)| \\ &\leq |\sin(\eta_1)| + |\sin(\eta_2)|, \end{aligned}$$

and so

$$\begin{aligned} \sin^2(\eta_1 + \cdots + \eta_m) &\leq (|\sin(\eta_1)| + \cdots + |\sin(\eta_m)|)^2 \\ &\leq m(\sin^2(\eta_1) + \cdots + \sin^2(\eta_m)). \end{aligned}$$

It follows that

$$\begin{aligned} D(\theta_1 + \theta_2 + \cdots + \theta_m) &\leq m(D(\theta_1) + D(\theta_2) + \cdots + D(\theta_m)) \\ &\leq m^2 \max\{D(\theta_1), D(\theta_2), \ldots, D(\theta_m)\}. \end{aligned}$$

Kemperman's theorem thus implies that

$$|S_{\rho m^2}| \geq |S_\rho + \cdots + S_\rho| \geq m|S_\rho|,$$

as long as $S_{\rho m^2}$ is not all of $\mathbb{R}/\mathbb{Z}$. Since

$$\mathop{\mathbb{E}}_{\theta}[D(\theta)] = \sum_{j \in G} \mathop{\mathbb{E}}_{\theta}\left[\sin^2(2\pi\theta d_j)\right] = \frac{|G|}{2},$$

we can deduce that $|S_{\lambda n/4}| = \Pr_\theta[D(\theta) \leq \frac{\lambda n}{4}]$ is strictly less than one. Hence, $S_{\lambda n/4}$ is not the full group $\mathbb{R}/\mathbb{Z}$. Setting m to be the largest integer so that $m^2\rho \leq \frac{\lambda n}{4}$, we can conclude

$$\Pr_\theta[D(\theta) \leq \rho] \leq \tfrac{1}{m} \Pr_\theta[D(\theta) \leq \rho m^2] \leq \tfrac{1}{m} \Pr_\theta[D(\theta) \leq \tfrac{\lambda n}{4}]. \qquad \square$$

$\square$

Acknowledgements. We wish to thank James Lee, Oded Regev, Avishay Tal and David Woodruff for helpful conversations.

References

[1] S. Aaronson and A. Arkhipov. The computational complexity of linear optics. In *STOC*, pages 333–342, 2011.

[2] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *FOCS*, pages 337–347, 1986.

[3] J. Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory* 1(1), pages 1–32, 2005.

[4] A. Chakrabarti and O. Regev. An optimal lower bound on the communication complexity of gap-hamming-distance. *SIAM Journal on Computing* 41(5), pages 1299–1317, 2012.

[5] E. Chattopadhyay and D. Zuckerman. Explicit two-source extractors and resilient functions. In *STOC*, pages 670–683, 2016.

[6] B. Chor and O. Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing* 17(2), pages 230–261, 1988.
[7] P. Erdös. On a lemma of littlewood and offord. *Bulletin of the American Mathematical Society* 51(12), pages 898–902, 1945.
[8] P. Erdös. Extremal problems in number theory. *Proc. Sympos. Pure Math.* VIII, 181—189, 1965.
[9] P. Erdös and P. Turán. On a problem of Sidon in additive number theory and on some related problems. *J. London Math. Soc.* 16, pages 212–215, 1941.
[10] P. Frankl and Z. Furedi. Solution of the littlewood-offord problem in high dimensions. *Annals of Mathematics*, pages 259–270, 1988.
[11] G. Halász. Estimates for the concentration function of combinatorial number theory and probability. *Periodica Mathematica Hungarica* 8(3-4), pages 197–211, 1977.
[12] P. Indyk and D. Woodruff. Tight lower bounds for the distinct elements problem. In *FOCS*, pages 283–288, 2003.
[13] J. Kahn, J. Komlos and E. Szemeredi. On the probability that a random ± 1-matrix is singular. *J. Amer. Math. Soc.* 8, pages 223–240, 1995.
[14] J. Kemperman. On products of sets in a locally compact group. *Fundamenta Mathematicae* 56 (1964), 51-68
[15] M. Kneser. Abschatzungen der asymptotischen Dichte von Summenmengen. *Math. Z* 58, pages 459–484, 1953.
[16] J. E. Littlewood and A. C. Offord. On the number of real roots of a random algebraic equation (iii). *Rec. Math. (Mat. Sbornik) N.S* 12(3), pages 277–286, 1943.
[17] R. Raz. Extractors with weak random seeds. In *STOC*, pages 11–20, 2005.
[18] A. Sarkö zyand and E. Szemeredi. Über ein Problem von Erdös und Moser. *Acta Arithmetica* 11, pages 205–208, 1965.
[19] A. A. Sherstov. The communication complexity of gap hamming distance. *Theory of Computing* 8(1), pages 197–208, 2012.
[20] R. Stanley. Weyl groups, the hard Lefschetz theorem, and the Sperner property. *SIAM J. Algebraic Discrete Methods* 1, pages 168–184, 1980.
[21] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques* 81(1), pages 73–205, 1995.
[22] T. Tao and V. Vu. A sharp inverse Littlewood-Offord theorem. *Random Structures & Algorithms* 37(4), pages 525–539, 2010.
[23] T. Tao and V. H. Vu. Inverse Littlewood-Offord theorems and the condition number of random discrete matrices. *Annals of Mathematics*, pages 595–632, 2009.
[24] T. Tao and V. H. Vu. Additive Combinatorics. *Cambridge University Press.*
[25] L. Trevisan and S. Vadhan. Extracting randomness from samplable distributions. In *FOCS*, pages 32–42, 2000.
[26] T. Vidick. A concentration inequality for the overlap of a vector on a large set. *Chicago Journal of Theoretical Computer Science*, 1, pages 1–12, 2012.
[27] A. C. Yao. Lower bounds by probabilistic arguments. In *FOCS*, pages 420–428, 1983.

Appendix A. Strict Convexity

Proof of Claim 9. If $\alpha_1 = 0$, then the claim is trivially true. So, assume that $\alpha_1 > 0$. Without loss of generality, we may also assume that $\kappa > 0$ is small enough so that $4^\kappa > \exp(\kappa + \kappa^2)$.

Let $p = \Pr[W = 1] \in [\kappa, 1-\kappa]$ and $\xi = \frac{\alpha_{-1}}{\alpha_1} \in [0, \frac{1}{2}]$. So,

$$\frac{\mathbb{E}\left[\alpha_W\right]^{1+\nu}}{\mathbb{E}\left[\alpha_W^{1+\nu}\right]} = \frac{(p+(1-p)\xi)^{1+\nu}}{p+(1-p)\xi^{1+\nu}}.$$

We need to upper bound this ratio by $1 - c_1\nu$, for some constant c_1 that depends only on κ. Let

$$\Phi(\xi, p, \nu) = (p+(1-p)\xi^{1+\nu}) - (p+(1-p)\xi)^{1+\nu}.$$

We shall argue that there is a constant $c_1 = c_1(\kappa) > 0$ such that $\Phi(\xi, p, \nu) \geq c_1\nu$. This completes the proof, since

$$\frac{(p+(1-p)\xi)^{1+\nu}}{(p+(1-p)\xi^{1+\nu})} = 1 - \frac{\Phi(\xi, p, \nu)}{(p+(1-p)\xi^{1+\nu})} < 1 - c_1\nu.$$

First, we show that for every ν and ξ, the function $\Phi(\xi, p, \nu)$ is minimized when $p = \kappa$. Consider

$$\begin{aligned}\frac{\partial \Phi}{\partial p} &= 1 - \xi^{1+\nu} - (1+\nu)(p+(1-p)\xi)^\nu(1-\xi) \\ &\geq 1 - \xi^{1+\nu} - (1+\nu)(1-\xi) \\ &\geq \xi(1+\nu-\xi^\nu) > 0,\end{aligned}$$

since $\xi^\nu < 1$. So, the minimum is achieved when $p = \kappa$.

Second, we claim that for every ν and p, the function $\Phi(\xi, p, \nu)$ is minimized when $\xi = \frac{1}{2}$. Consider

$$\begin{aligned}\frac{\partial \Phi}{\partial \xi} &= (1-p)(1+\nu)\xi^\nu - (1+\nu)(p+(1-p)\xi)^\nu(1-p) \\ &= (1-p)(1+\nu)(\xi^\nu - (p+(1-p)\xi)^\nu) < 0,\end{aligned}$$

since $p + (1-p)\xi > \xi$. So, the minimum is achieved when $\xi = 1/2$.

Third, we control the derivative with respect to ν for $\xi = \frac{1}{2}$ and $p = \kappa$. Consider

$$\begin{aligned}\frac{\partial \Phi}{\partial \nu}(\tfrac{1}{2}, \kappa, \nu) &= (1-\kappa)\ln(\tfrac{1}{2})(\tfrac{1}{2})^{1+\nu} - \ln(\tfrac{1+\kappa}{2})(\tfrac{1+\kappa}{2})^{1+\nu} \\ &\geq (\tfrac{1}{2})^2((1-\kappa)\ln(\tfrac{1}{2}) - \ln(\tfrac{1+\kappa}{2})(1+\kappa)^{1+\nu}),\end{aligned}$$

since $\nu \leq 1$. The expression

$$(1-\kappa)\ln(\tfrac{1}{2}) - \ln(\tfrac{1+\kappa}{2})(1+\kappa)^{1+\nu}$$

only increases with ν. When $\nu = 0$, this expression is

$$\ln(\tfrac{2^{2\kappa}}{(1+\kappa)^{1+\kappa}}) \geq \ln(\tfrac{4^{\kappa}}{\exp(\kappa(1+\kappa))}) > 0,$$

since $4^{\kappa} > \exp(\kappa + \kappa^2)$. This proves that $\frac{\partial \Phi}{\partial \nu}(\frac{1}{2}, \kappa, \nu) > c_1$ for some constant $c_1(\kappa) > 0$.

Finally,

$$\Phi(\xi, p, \nu) \geq \Phi(\tfrac{1}{2}, \kappa, \nu) = \int_0^{\nu} \frac{\partial \Phi}{\partial \nu}(\tfrac{1}{2}, \kappa, \zeta)\, \mathrm{d}\zeta \geq \int_0^{\nu} c_1\, \mathrm{d}\zeta = c_1 \nu. \quad \square$$

School of Computer Science, University of Washington
E-mail address: anuprao@cs.washington.edu

Department of Mathematics, Technion-IIT
E-mail address: amir.yehudayoff@gmail.com